ISBA 2012 World Meeting

Kyoto, Japan

June 25–29, 2012

Abstract book

Last update: June 28, 2012

Sponsors

ISBA
INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

助成 独立行政法人日本万国博覧会記念機構
Supported by the Commemorative Organization for the Japan World Exposition('70).
この助成金は、日本万国博覧会の収益を基にしています。

JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE · FOUNDED 1932 ·
學振

日本政府観光局 (JNTO)
Japan National Tourism Organization

ASA
AMERICAN STATISTICAL ASSOCIATION
Section on Bayesian Statistical Science

Google

Microsoft®

PASCAL 2

NSF

Collegio Carlo Alberto

Research Organization of Information and Systems
The Institute of Statistical Mathematics

Also supported by Grants-in-Aid for Scientific Research (A) (KAKENHI, #21243018, #21243030, #20243017), Grant-in-Aid for Young Scientists (B) (KAKENHI #23730213) and endorsed by: Science Council of Japan, Institute for Monetary and Economic Studies (Bank of Japan), Economic and Social Research Institute (Cabinet Office, Government of Japan), Behaviormetric Society of Japan, Biometric Society of Japan, Japan Institute of Marketing Science, Japan Association for Research on Testing, Japanese Classification Society, Japanese Society of Applied Statistics, Japanese Society of Computational Statistics, and Japan Statistical Society.

Foreword

ISBA 2012 meeting is the first of the renowned bi-yearly World Meetings of the International Society for Bayesian Analysis and successor of the Valencia Meeting. There were ten ISBA World Meetings in the past 30 years, but this is the first to be held in East Asia. Japan and East Asia have been the home of many well known Bayesians, and the number of Bayesians in the region has been growing steadily over the years. As a sign of the era of globalization and international collaboration, Bayesians too are currently expanding their activities across national boundaries, fostering friendship and partnership with statisticians in Asian societies.

Professor Hirotugu Akaike, was among those who attended the first Valencia Meeting in Spain, in 1979, organized by Profs. Morris DeGroot, Dennis Lindley, Adrian Smith and Jose Bernardo. Prof. Akaike remembered fondly the deep discussions about AIC that took place at the meeting. He was invited by Professor Arnold Zellner to organize the first International Bayesian meeting in Fuji, Japan in 1993, which has been a predecessor of the World Meetings. The meeting was a great success, and it presented the new graduate students with an exciting window into the world of Bayesian statistics.

We hope that ISBA2012 in Kyoto will present a similar opportunity for the young researchers and students in Japan and neighboring Asian countries. Almost 30 years have passed since 1979 and the world has experienced so many changes, but there still remain many problems to be solved. We researchers have exchanged many ideas to find better solutions. Through our discussion without reserve, we expect young researchers and students to broaden their intellectual horizons.

We have received many wonderful “special topic sessions” submissions. To our regret, due to the limit of presentation capacities, we had to decline several applications and reorganize the sessions. Hence, the scientific program includes 4 ISBA Lectures on Bayesian Foundations, 5 keynote speakers, 203 invited and special topic talks and over 240 posters. Also, as part of an initiative set forth by the Institute of Statistical Mathematics to promote Bayesian Computational methods, a two days satellite meeting in Tokyo will precede the main event. It is an honor for Japan to have the opportunity to host ISBA2012 in the authentic Japanese city of Kyoto.

Lastly we would like to thank all scientific associations, companies, institutions and research councils, which made this event possible. In particular, the Japan Society for the Promotion of Science, the Japan World Exposition 1970 Commemorative Fund, Grants-in-Aid for Scientific Research (KAKENHI, #21243018, #21243030, #20243017, #23730213), the Japan National Tourism Organization, the Science Council of Japan, the Institute for Monetary and Economic Studies (Bank of Japan), the Economic and Social Research Institute (Cabinet Office, Government of Japan), Google, Microsoft, the US National Science Foundation, Collegio Carlo Alberto, Pascal 2, the Section on Bayesian Statistical Science of the American Statistical Association, the Behaviormetric Society of Japan, the Biometric Society of Japan, the Japan Institute of Marketing Science, the Japan Association for Research on Testing, the Japanese Classification Society, the Japanese Society of Applied Statistics, the Japanese Society of Computational Statistics and the Japan Statistical Society.

Welcome to Kyoto!

Hajime Wago
The Local Organizing Committee

Program Committee

Ming-Hui Chen (University of Conneticut, USA) • Andres Christen (CIMAT, Mexico) • Vanja Dukic (University of Colorado-Boulder, USA) • Ed George (University of Pennsylvania, USA) • Jayanta Ghosh (Purdue University, USA) • Michael Jordan (University of California Berkeley, USA) • Jaeyong Lee (Seoul National University, Korea) • Kerrie Mengersen (Queensland University of Technology, Australia) • Peter Mueller (University of Texas Austin, USA) • Yasuhiro Omori (University of Tokyo, Japan) • Igor Pruenster (University of Torino, Italy) • Fernando Quintana (Pontificia Universidad Catolica de Chile, Chile) • Gareth Roberts (University of Warwick, UK) • Fabrizio Ruggeri (CNR-IMATI Milan, Italy) • Alexandra Schmidt (Federal University of Rio de Janeiro, Brazil) • Scott Sisson (University of New South Wales, Australia) • David Stephens (McGill University, Canada) • Hajime Wago (Kyoto Sangyo University, Japan)

Organizing Committee

Makoto Abe (University of Tokyo, Tokyo, Japan) • Tomohiro Ando (Keio University, Tokyo, Japan) • Michikazu Aoi (Meiji University, Tokyo, Japan) • Manabu Asai (Soka University, Tokyo, Japan) • Kentaro Fukumoto (Gakushuin University, Tokyo, Japan) • Hikaru Hasegawa (Hokkaido University, Hokkaido, Japan) • Yu Hayakawa (Waseda University, Tokyo, Japan) • Takahiro Hoshino (Nagoya University, Aichi, Japan) • Yukito Iba (Institute of Statistical Mathematics, Tokyo, Japan) • Kosuke Imai (Princeton University, New Jersey, USA) • Shigeru Iwata (University of Kansas, Kansas, USA) • Yuichiro Kanazawa (University of Tsukuba, Ibaraki, Japan) • Sotaro Katsumata (University of Tokyo, Tokyo, Japan) • Hirohisa Kishino (University of Tokyo, Tokyo, Japan) • Genshiro Kitagawa (Institute of Mathematical Statistics, Tokyo, Japan) • Toru Kitagawa (University College London, London, UK) • Atsuyuki Kogure (Keio University, Tokyo, Japan) • Fumiyasu Komaki (University of Tokyo, Tokyo, Japan) • Fumiyo Kondo (University of Tsukuba, Ibaraki, Japan) • Fumiyo Kondo (University of Tsukuba, Ibaraki, Japan) • Sadanori Konishi (Chuo University, Tokyo, Japan) • Robert Leon-Gonzalez (National Graduate Institute for Policy Studies, Tokyo, Japan) • Yuzo Maruyama (University of Tokyo, Tokyo, Japan) • Nozomu Matsubara (Seigakuin University, Saitama, Japan) • Haruhisa Nishino (Chiba University, Chiba, Japan) • Takashi Oga (Chiba University, Chiba, Japan) • Yoshihiko Ogata (Institute of Statistical Mathematics, Tokyo, Japan) • Yasuo Ohashi (University of Tokyo, Tokyo, Japan) • Taisuke Otsu (Yale University, Conneticut, USA) • Kosuke Oya (Osaka University, Osaka, Japan) • Tadahiko Sato (University of Tsukuba, Ibaraki, Japan) • Takuya Satomura (Keio University, Tokyo, Japan) • Akira Sekiya (Keio University, Tokyo, Japan) • Kenichiro Shimatani (Institute of Statistical Mathematics, Tokyo, Japan) • Hiroko Solvang (Oslo University Hospital, Oslo, Norway) • Shintaro Sono (Hokkaido University, Hokkaido, Japan) • Katsuhiro Sugita (University of Ryukyus, Okinawa, Japan) • Hiroshi Sunada (Yamagata University, Yamagata, Japan) • Kunio Tanabe (Waseda University, Tokyo, Japan) • Masami Tanemura (Institute of Statistical Mathematics, Tokyo, Japan) • Hiroki Tsurumi (Rutgers University, New Jersey, USA) • Teruaki Watabe (Kochi University, Kochi, Japan) • Michiko Watanabe (Toyo University, Tokyo, Japan) • Takashi Yanagawa (Kurume University, Fukuoka, Japan) • Takemi Yanagimoto (Chuo University, Tokyo, Japan) • Yasuto Yoshizoe (Aoyama Gakuin University, Tokyo, Japan)

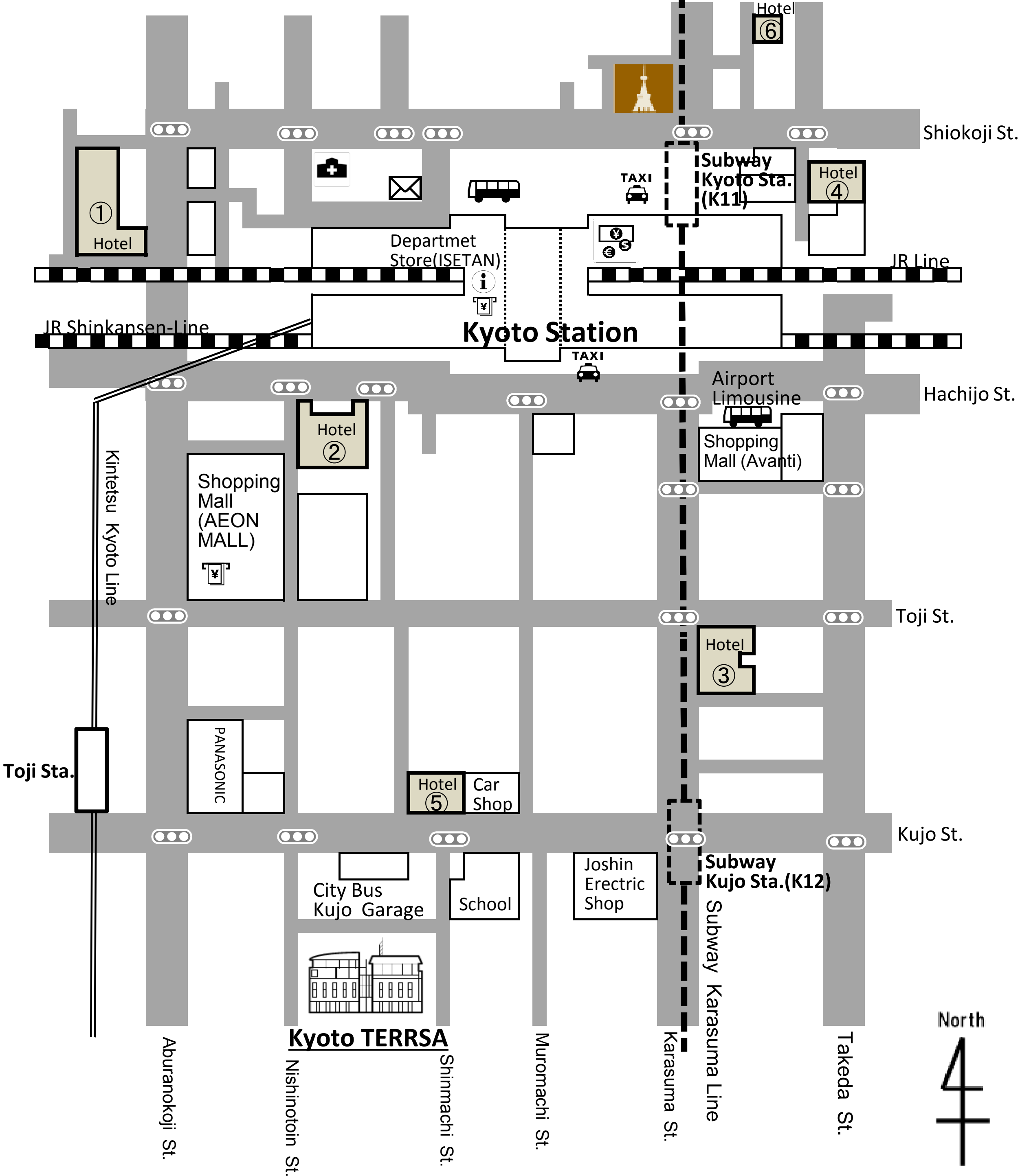
Executive Committee

Hajime Wago (Kyoto Sangyo University, Kyoto, Japan) • Kazuho Shigemasu (Teikyo University, Tokyo, Japan) • Yasuhiro Omori (University of Tokyo, Tokyo, Japan) • Mototsugu Fukushige (Osaka University, Osaka, Japan) • Kazuhiko Kakamu (Chiba University, Chiba, Japan) • Hideo Kozumi (Kobe University, Kobe, Japan) • Koji Miyawaki (National Institute for Environment Studies, Ibaraki, Japan) • Seisho Sato (Institute of Mathematical Statistics, Tokyo, Japan) • Hisashi Tanizaki (Osaka University, Osaka, Japan) • Nobuhiko Terui (Tohoku University, Miyagi, Japan) • Toshiaki Watanabe (Hitotsubashi University, Tokyo, Japan) • Teruo Nakatsuma (Keio University, Tokyo, Japan)

Contents

Kyoto TERRSA Access Map

Address Kyoto TERRSA Shinmachi -Kujo Minami-ku Kyoto Jaoan
Tel +81-75-692-3400 Fax +81-75-692-3402



Official Hotel	To Kyoto Station	To Kyoto TERRSA
① RIHGA Royal Hotel Kyoto	･8 min on foot ･Free shuttle bus（5 min ride） http://www.rihga.com/kyoto/location.html	･12min on foot
② New Miyako Hotel	･2min on foot	･9min on foot
③ Daiwa Roynet Hotel Kyoto-hachijoguchi	･5min on foot	･6min on foot
④ Kyoto Dai-ni Tower Hotel	･1min on foot	･10min. by Subway Karasuma Line
⑤ Kyoto Daiichi Hotel	･8min on foot ･Free shuttle bus(4 min ride) http://w w w .kyoto-1-hotel.jp/english/access.html#pickup	･1min on foot
⑥ Nishikiro-Ryokan	･3min on foot	･13min. by Subway Karasuma Line

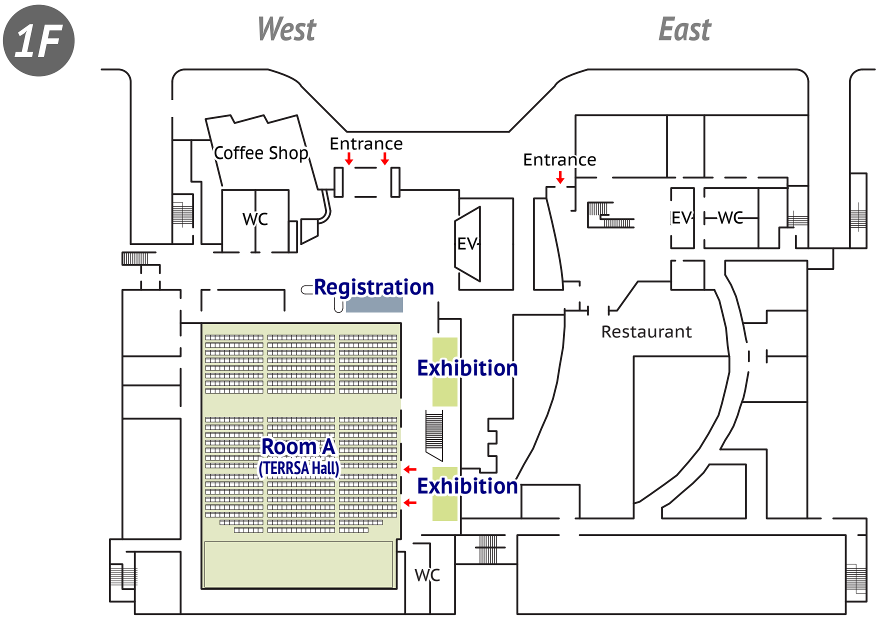
Free shuttle service is available

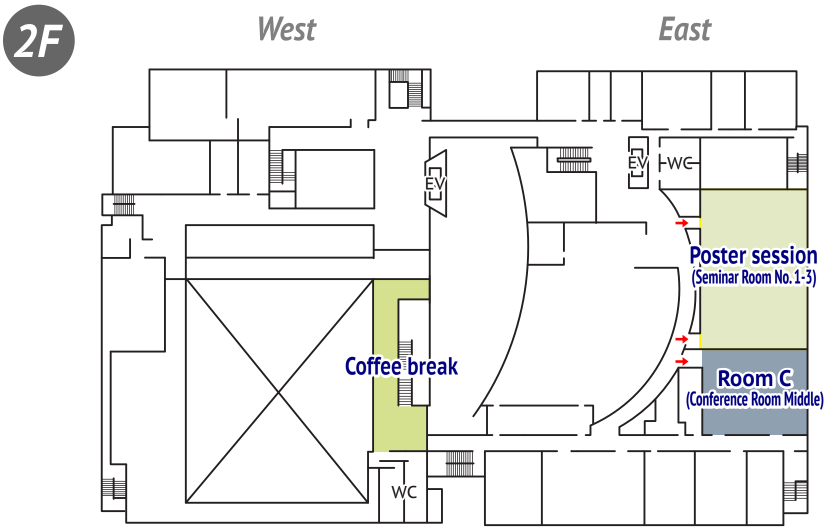
Mon 25 June / for Welcome Reception	
Kyoto TERRSA to Garden Oriental Kyoto	Bus Pick-up time 18:40
Garden Oriental Kyoto to Official Hotel	Bus Pick-up time 21:40

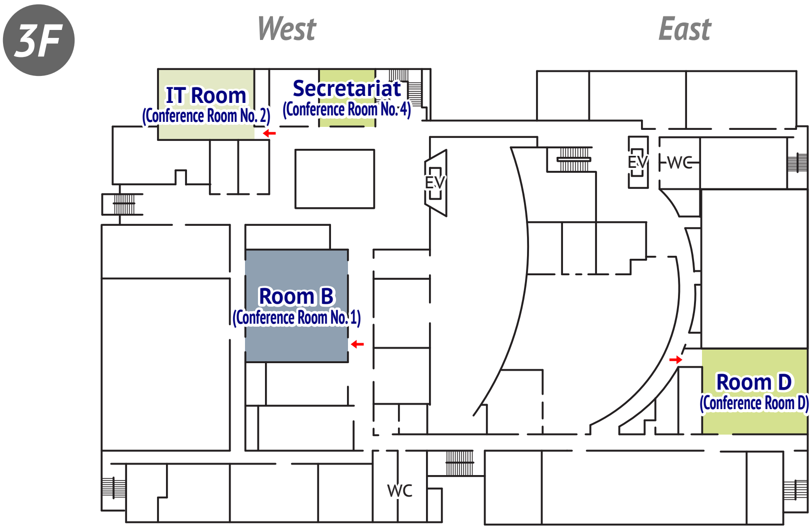
Tue - Thu 26-28 June	
RIHGA Royal Hotel Kyoto to Kyoto TERRSA	Bus Pick-up time 8:00 -9:00 (about every 20min)
Kyoto TERRSA to RIHGA Royal Hotel	Bus Pick-up time 21:00 -22:00 (about every 20min)

Fri 29 June	
RIHGA Royal Hotel Kyoto to Kyoto TERRSA	Bus Pick-up time 8:00 -9:00 (about every 20min)
Kyoto TERRSA to RIHGA Royal	Bus Pick-up time 17:30

Floor map


1F
West
East
Coffee Shop
Entrance
Entrance
WC
EV
EV
WC
Registration
Exhibition
Restaurant
Room A
(TERRSA Hall)
Exhibition
WC


2F
West
East
EV
EV
WC
Poster session
(Seminar Room No. 1-3)
Coffee break
Room C
(Conference Room Middle)
WC


3F
West
East
IT Room
(Conference Room No. 2)
Secretariat
(Conference Room No. 4)
EV
EV
WC
Room B
(Conference Room No. 1)
Room D
(Conference Room D)
WC

Monday 25th	Tuesday 26th	Wednesday 27th	Thursday 28th	Friday 29th
	Registration Opening 8:30 hrs.	Registration	Registration	Registration
	Parallel sessions 9:00 hrs.	***Parallel sessions 9:00 hrs.***	***Parallel sessions 9:00 hrs.***	***Parallel sessions 9:00 hrs.***
	Advances in Gaussian processes Liang; Kaufman; Lee A	**Being simultaneously Bayesian and frequentist** Lumley; Mukherjee; Rice A	**Applications of particle filtering and sequential updating** Cornebise; Sisson; Kohn A	**Savage Award Session** Jang; Mandel; Martínez-Ovando; Scheipl A
	Bayesian methods in Biostatistics Riebler; Mueller; Rossell B	**Bayesian applications** Carvalho; Draper; Richardson; Prado B	**Bayesian graphical and factor models: structure, sparsity and dimension** Yoshida; Cron; H. Wang; Hahn B	**Topics in Bayesian Statistics** Johnson; Daniels; Kuensch B
	Bayesian Econometrics IV Gerlach; Nakatsuma; Forbes; Chen C	**Beta processes: extensions and applications** Kim; Williamson; Broderick; Hjort C	**Bayesian methods in biological, environmental and ecological systems** Telesca; Muthukumarana; Berrocal; Mondal C	**Bayesian spatio-temporal disease mapping: new frontiers** Cramb; White; Li; Farah C
	Bayesian empirical likelihood Lazar; Basu; Ghosh; Chaudhuri D	**Auxiliary variable and particle MCMC methods** Chopin; Lee; Jacob D	**Bayesian semi-parametric analysis: theory** van Zanten; Castillo; De Blasi D	**Bayes modeling in marketing** Allenby; Abe; Kondo; Terui D
	Coffee break (10:30 hrs.)	*Coffee break (10:30 hrs.)*	*Coffee break (10:30 hrs.)*	*Coffee break (10:30 hrs.)*
	Parallel sessions 11:00 hrs.	***Parallel sessions 11:00 hrs.***	***Parallel sessions 11:00 hrs.***	***Parallel sessions 11:00 hrs.***
	Model selection George; Scott; Bayarri A	**Scaling Bayesian computation to handle big data: methods and feasibility** Jordan; Kou; Atchade; Woodard A	**Approximate Bayesian computation : likelihood-free Bayesian inference II** Nott; Blum; Marin; Ratman A	**Bayesian models for high-dimensional complex-structured data** Brown; Ferreira; Morris; Banerjee A
Registration	**Approximate Bayesian computation : likelihood-free Bayesian inference I** Stumpf; Prangle; Rousseau; Drovandi B	**Partial identification and causal inference: what can Bayes bring to the table?** Kitagawa; C. Wang; Gustafson B	**Time Series analysis and Finance** Mira; Nakajima; Omori B	**S'Bayes: constructing and using subjective priors for Bayesian modelling** Goldstein; Low-Choy; Gill B
ISBA lecture on Bayesian foundations 13:00 hrs. **Confidence in nonparametric credible sets?** Aad van der Vaart Ⓐ	**Optimal Bayesian experimental design** Kim; Huan; Morita; Solonen C	**Bayesian analysis of network data: from network determination to network modeling** Baladandayuthapani; Pillai; Rodriguez C	**Bayesian model assessment** Clyde; Hatfield; García-Donato; Mulder C	**Bayesian analysis of inverse problems** Haario; Herbei; C. Fox C
	Bayesian Econometrics Frühwirth-Schnatter; Norets; Geweke D	**Applied Bayesian Econometrics** Lopes; Smith; Villani; Wagner D	**Problem-driven developments in Bayesian nonparametrics** Teh; Bacallado; Petrone D	**Applications of non- and semi-parametric Bayesian methods** Argiento; Griffin; Kottas; Quintana D
	Lunch (12:30 hrs.)	*Lunch (12:30 hrs.)*	*Lunch (12:30 hrs.)*	*Lunch (12:30 hrs.)*
ISBA lecture on Bayesian foundations 14:15 hrs. **Bayesian dynamic modelling** Mike West Ⓐ	***Keynote lecture 14:00 hrs.*** **Data assimilation and sequential Bayes filters: information fusion with numerical simulation** Tomoyuki Higuchi Ⓐ	***Keynote lecture 14:00 hrs.*** **Bayesian methods in cancer genomics** Chris C. Holmes Ⓐ	***Keynote lecture 14:00 hrs.*** **Bayesian nonparametrics and conditional distributions** Stephen G. Walker Ⓐ	***Keynote lecture 14:00 hrs.*** **Bayesian inference for intractable state-space models** Arnaud Doucet Ⓐ
	Parallel sessions 15:00 hrs.	***Parallel sessions 15:00 hrs.***	***Parallel sessions 15:00 hrs.***	***Keynote lecture 14:45 hrs.*** **Demographic analysis of forest dynamics using stochastic integral projection models** Alan Gelfand Ⓐ
	Hierarchies of Bayesian nonparametric processes E. Fox; Kim; Favaro; Wood A	**Bayesian methods for Spatial Statistics** Haran; Sang; Huerta; Yang A	**Case studies of Bayesian success stories: babies, trials and ratings** Saria; Carlin; Ji; Glickman A	
Coffee break (15:30 hrs.)	**Networks and relational data** McCormick; Thomas; Lunagomez B	**Bayesian analysis of protein structure and evolution** Challis; Wu; Czogiel; Schmidler B	**Parallel processing in Bayesian computing** Suchard; Scott; Niemi; McAlinn B	*Coffee break (15:30 hrs.)*
ISBA lecture on Bayesian foundations 16:00 hrs. **Approximate Bayesian computation: advances and questions** Christian Robert Ⓐ	**Recent advances in Bayesian causal inference** Mealli; Li; Elliott; Gutman C	**On the uses of random probabilities in Bayesian inference** Orbanz; Mena; Lijoi C	**Adaptive Bayesian function estimation** Belitser; Tokdar; Lian C	**General meeting** **16:00–17:30 hrs.** Ⓐ
	Adaptive Monte Carlo Fort; Vihola; Salakhutdinov; Z. Wang D	**Bayesian Econometrics II** Jacobi; Strachan; Leon-Gonzalez; Chan D	**Bayesian Econometrics III** Ando; Watanabe; Basturk; Choi D	
	Coffee break (16:30 hrs.)	*Coffee break (16:30 hrs.)*	*Coffee break (16:30 hrs.)*	
	Parallel sessions 17:00 hrs.	***Parallel sessions 17:00 hrs.***	***Parallel sessions 17:00 hrs.***	
	Predictive inference and Bayes methods Clarke; Parry; Komaki A	**Bayesian analysis of astronomical data** Trotta; Mahabal; Philip; Stoica, Martinez, Saar A	**Spatial state-space models** Sanso; Manolopoulou; Strickland A	
ISBA lecture on Bayesian foundations 17:15 hrs. **Slowly but surely, Bayesian ideas revolutionize medical research** Donald A. Berry Ⓐ	**Beyond MCMC methods in Bayesian inference** Braun; Andrieu; Stroud; Kalli B	**Recent advances in Bayesian variable selection** Hans; Tadesse; Leman; Ghosh B	**Bayesian modeling and its applications in social science** Sun; Nicholls; He; Lu B	
	Bayesian inference in science: the pursuit of a synergy Held; Forster; Lemos; Rigat C	**Nonparametric Bayes applications in Biostatistics** Guindani; Trippa; Nieto-Barajas C	**High dimensional graphical models in genomics** Stingo; Liverani; Lenkoski; Bottolo C	
	Bayesian approaches to design and model comparison McGree; Taddy; Bornkamp; MacEachern D	**Advances in honest Monte Carlo** Flegal; Huber; Jones; Beskos D	**Bayesian methods in reliability** Mazzuchi; Soyer; Christen; Wilson D	
Welcome reception at Garden Oriental 19:30 hrs.	***Poster session*** 18:30–21:00 hrs. *Seminar Room No. 1–3*	***Poster session*** 18:30–21:00 hrs. *Seminar Room No. 1–3*	***Poster session*** 18:30–21:00 hrs. *Seminar Room No. 1–3*	*Banquet, Cabaret and Afterparty 18:00–1:00 hrs.*

Detailed schedule

Monday, June 25th

12:30–13:00 Registration

13:00–15:30 **ISBA Lectures on Bayesian Foundations: Session 1**

Room A Chair: Michael Jordan (University of California, Berkeley, USA)

Speakers:

- Aad van der Vaart (University of Leiden, Netherlands)
  Confidence in nonparametric credible sets?
- Mike West (Duke University, USA)
  Bayesian dynamic modelling

15:30–16:00 Coffee break

16:00–18:30 **ISBA Lectures on Bayesian Foundations: Session 2**

Room A Chair: Peter Mueller (University of Texas at Austin, USA)

Speakers:

- Christian Robert (Université Paris Dauphine, France)
  Approximate Bayesian computation (ABC): advances and questions
- Donald A. Berry (University of Texas MD Anderson Cancer Center, USA)
  Slowly but surely, Bayesian ideas revolutionize medical research

19:30–21:30 Welcome reception at Garden Oriental

Tuesday, June 26th

08:00–08:30 Registration

08:30–09:00 Opening

09:00–10:30

Room A *Invited session.* **Advances in Gaussian processes**

Organizer: Herbie Lee (University of California, Santa Cruz, USA)

Chair: Robert Gramacy (University of Chicago Booth School of Business, USA)

Speakers:

- Waley Liang (University of California, Santa Cruz, USA)
  Bayesian nonstationary Gaussian process models via treed process convolutions
- Cari Kaufman (University of California, Berkeley, USA)
  Efficient emulators of computer experiments using compactly supported correlation functions
- Herbie Lee (University of California, Santa Cruz, USA)
  Gaussian process modeling of derivative curves

Discussant: Kate Calder (Ohio State University, USA)

Room B *Invited session.* **Bayesian methods in Biostatistics**

Organizer: Valen E. Johnson (University of Texas MD Anderson Cancer Center, USA)

Chair: David Rossell (IRB Barcelona, Spain)

Speakers:

- Andrea Riebler (University of Zurich, Switzerland)
  A novel empirical Bayes approach for profiling immunoprecipitation-based DNA methylation
- Peter Mueller (University of Texas at Austin, USA)
  A Bayesian graphical model for ChIP-Seq data on histone modifications
- David Rossell (IRB Barcelona, Spain)
  Bayesian de novo characterization of alternative splicing using high-throughput sequencing

Discussant: Donatello Telesca (University of California, Los Angeles, USA)

Room C *Special topic session.* **Bayesian Econometrics IV**

Organizer: Cathy W. S. Chen (Feng Chia University, Taiwan)

Chair: Cathy W. S. Chen (Feng Chia University, Taiwan)

Speakers:

- Richard Gerlach (University of Sydney, Australia)
  Bayesian semi-parametric forecasting of expected shortfall

- Teruo Nakatsuma (Keio University, Japan)
  Bayesian risk assessment with threshold mixture extreme value models
- Catherine Forbes (Monash University, Australia)
  Efficient filtering of stochastic volatility models
- Cathy W. S. Chen (Feng Chia University, Taiwan)
  A Bayesian perspective on backtesting value-at-risk models

Room D *Special topic session.* **Bayesian empirical likelihood**

Organizer: Sanjay Chaudhuri (National University of Singapore, Singapore)

Chair: Sanjay Chaudhuri (National University of Singapore, Singapore)

Speakers:

- Nicole Lazar (University of Georgia, USA)
  Empirical likelihood weighting
- Sanib Basu (Northern Illinois University, USA)
  Bayesian empirical likelihood
- Malay Ghosh (University of Florida, USA)
  Bayesian empirical likelihood for longitudinal data with small area application
- Sanjay Chaudhuri (National University of Singapore, Singapore)
  Empirical likelihood based Bayesian methodology for complex survey datasets

---

10:30–11:00 Coffee break

11:00–12:30

Room A *Invited session.* **Model selection**

Organizer: Susie Bayarri (University of Valencia, Spain)

Chair: David Conesa (University of Valencia, Spain)

Speakers:

- Ed George (University of Pennsylvania, USA)
  Shrinkage adjustment for model selection
- James Scott (University of Texas at Austin, USA)
  A new look at logit likelihoods
- Susie Bayarri (University of Valencia, Spain)
  TESS: the effective sample size for linear models

Discussant: Gonzalo García-Donato (University of Castilla-La Mancha, Spain)

Room B *Special topic session.* **Approximate Bayesian computation (ABC): likelihood-free Bayesian inference I**

Organizers: Christian Robert (Université Paris Dauphine, France), Scott Sisson (University of New South Wales, Australia)

Chair: Scott Sisson (University of New South Wales, Australia)

Speakers:

- Michael Stumpf (Imperial College London, UK)
  Considerate approaches to ABC model selection
- Dennis Prangle (Lancaster University, UK)
  Summary statistics for ABC model choice
- Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)
  Assessing the influence of the choice of summary statistics for ABC: application to model choice
- Christopher Drovandi (Queensland University of Technology, Australia)
  Approximate Bayesian computation using indirect inference

Room C *Special topic session.* **Optimal Bayesian experimental design**

Organizers: Thomas J. Loredo (Cornell University, USA), Youssef M. Marzouk (Massachusetts Institute of Technology, USA)

Chair: Youssef M. Marzouk (Massachusetts Institute of Technology, USA)

Speakers:

- Woojae Kim (Ohio State University, USA)
  Optimal experimental design for model discrimination in cognitive science: a nonparametric extension
- Xun Huan (Massachusetts Institute of Technology, USA)
  Optimal sequential Bayesian experimental design via approximate dynamic programming
- Satoshi Morita (Yokohama City University, Japan)
  Bayesian study designs for new drug development
- Antti Solonen (Lappeenranta University of Technology, Finland)
  Simulation-based optimal design using MCMC

Room D *Invited session.* **Bayesian Econometrics**

Organizer: John Geweke (University of Technology, Sydney, Australia)

Chair: Yasuhiro Omori (University of Tokyo, Japan)

Speakers:

- Sylvia Frühwirth-Schnatter (Vienna University of Economics and Business, Austria)
  Flexible econometric modelling based on sparse finite mixtures
- Andriy Norets (Princeton University, USA)
  Bayesian regression with nonparametric heteroskedasticity
- John Geweke (University of Technology, Sydney, Australia)
  Massively parallel sequential Monte Carlo for Bayesian inference

Discussant: Robert Kohn (University of New South Wales, Australia)

12:30–14:00 Lunch

14:00–14:45 **Keynote Lecture**

Room A Chair: Hajime Wago (Institute of Statistical Mathematics, Japan)

Speaker: Tomoyuki Higuchi (Institute of Statistical Mathematics, Japan)
Data assimilation and sequential Bayes filters: information fusion with numerical simulation

15:00–16:30

Room A *Special topic session.* **Hierarchies of Bayesian nonparametric processes**

Organizer: Lancelot F. James (Hong Kong University of Science and Technology, Hong Kong)

Chair: Peter Orbanz (Columbia University, USA)

Speakers:

- Emily Fox (University of Pennsylvania, USA)
  Multiresolution Gaussian process regression
- Dohyun Kim (Seoul National University, Korea)
  Hierarchical random measure
- Stefano Favaro (University of Torino, Italy)
  On the stick-breaking representation for Gibbs-type priors
- Frank Wood (Columbia University, USA)
  Infinite structured explicit duration hidden Markov models

Room B *Invited session.* **Networks and relational data**

Organizer: Peter Hoff (University of Washington, USA)

Chair: Perla Reyes (University of California, Santa Cruz, USA)

Speakers:

- Tyler McCormick (University of Washington, USA)
  A Bayesian framework for indirect sampling of relational data
- Andrew Thomas (Carnegie Mellon University, USA)
  Marginal-additive models and processes for network-correlated outcomes
- Simon Lunagomez (Harvard University, USA)
  Bayesian inference from non-ignorable network sampling designs

Discussant: Abel Rodriguez (University of California, Santa Cruz, USA)

Room C *Special topic session.* **Recent advances in Bayesian causal inference**

Organizer: Fan Li (Duke University, USA)

Chair: Fan Li (Duke University, USA)

Speakers:

- Fabrizia Mealli (University of Florence, Italy)
  Using multivariate outcomes in Bayesian causal inference with noncompliance
- Fan Li (Duke University, USA)
  Bayesian multivariate inference for a non-standard fuzzy regression discontinuity design
- Michael R. Elliott (University of Michigan, USA)
  Assessing the causal effect of treatment dosages in the presence of self-selection
- Roee Gutman (Brown University, USA)
  A robust outcome-free procedure for interval estimation of causal effects

Room D *Special topic session.* **Adaptive Monte Carlo**

Organizers: Nando de Freitas (University of British Columbia, Canada), Christophe Andrieu (University of Bristol, UK), Arnaud Doucet (University of Oxford, UK)

Chair: Nando de Freitas (University of British Columbia, Canada)

Speakers:

- Gersende Fort (LTCI, CNRS and Telecom ParisTech, France)
  Adaptive equi-energy samplers
- Matti Vihola (University of Jyväskylä, Finland)
  Robustness in adaptive MCMC
- Russ Salakhutdinov (University of Toronto, Canada)
  Learning high-dimensional deep Boltzmann machines using adaptive MCMC
- Ziyu Wang (University of British Columbia, Canada)
  Adaptive MCMC with bandits

16:30–17:00 Coffee break

17:00–18:30

Room A *Invited session.* **Predictive inference and Bayes methods**

Organizer: Fumiyasu Komaki (University of Tokyo, Japan)

Chair: Yuzo Maruyama (University of Tokyo, Japan)

Speakers:

- Bertrand Clarke (University of Miami, USA)
  Bayesian clustering stability
- Matthew Parry (University of Otago, New Zealand)
  Proper local scoring rules
- Fumiyasu Komaki (University of Tokyo, Japan)
  On the use of latent information priors

Discussant: Xinyi Xu (Ohio State University, USA)

Room B *Special topic session.* **Beyond MCMC methods in Bayesian inference**

Organizer: Paul Damien (University of Texas at Austin, USA)

Chair: Yee Whye Teh (University College London, UK)

Speakers:

- Michael Braun (Massachusetts Institute of Technology, USA)
  Generalized direct sampling for Bayesian hierarchical models
- Christophe Andrieu (University of Bristol, UK)
  The pseudo-marginal approach and exact approximations for efficient MC
- Jonathan Stroud (George Washington University, USA)
  Sequential estimation in dynamic spatio-temporal models
- Maria Kalli (University of Kent, UK)
  Understanding the slice sampler

Room C *Special topic session.* **Bayesian inference in science: the pursuit of a synergy**

Organizer: Fabio Rigat (University of Warwick, UK)

Chair: Fabio Rigat (University of Warwick, UK)

Speakers:

- Leonhard Held (University of Zurich, Switzerland)
  Assessing the impact of a movement network on the spatiotemporal spread of infectious diseases
- Jonathan Forster (University of Southampton, UK)
  Mortality projection incorporating model uncertainty
- Ricardo Lemos (University of Queensland, Australia)
  Getting the facts right on the climate change debate
- Fabio Rigat (University of Warwick, UK)
  On the construction of parametric hierarchical predictive distributions

Room D *Special topic session.* **Bayesian approaches to design and model comparison**

Organizer: Marina Savelieva (Novartis Pharma AG, Switzerland)

Chair: Luis Nieto-Barajas (ITAM, Mexico)

Speakers:

- James McGree (Queensland University of Technology, Australia)
  A particle filter for Bayesian sequential design
- Matt Taddy (University of Chicago Booth School of Business, USA)
  Design of text mining experiments
- Bjoern Bornkamp (Novartis Pharma AG, Switzerland)
  Functional uniform prior distributions for nonlinear regression
- Steven MacEachern (Ohio State University, USA)
  The calibrated Bayes factor for model comparison

18:30–21:00 *Seminar Room No. 1–3:* **Poster session**

Wednesday, June 27th

08:00–09:00 Registration

09:00–10:30

Room A *Invited session.* **Being simultaneously Bayesian and frequentist**

Organizer: Kenneth Rice (University of Washington, USA)

Chair: Fumiyasu Komaki (University of Tokyo, Japan)

Speakers:

- Thomas Lumley (University of Auckland, New Zealand)
  Likelihood of the empirical CDF: a coarsening approach to complex samples
- Bhramar Mukherjee (University of Michigan, USA)
  Prediction in high dimensional datasets: an ensemble of shrinkage approaches
- Kenneth Rice (University of Washington, USA)
  Interpreting tests which are both Bayesian and frequentist

Discussant: David Stephens (Mc Gill University, Canada)

Room B *Special topic session.* **Bayesian applications**

Organizer: David Draper (University of California, Santa Cruz, USA)

Chair: David Draper (University of California, Santa Cruz, USA)

Speakers:

- Carlos Carvalho (University of Texas at Austin, USA)
  Bayesian testing of factor asset pricing models
- David Draper (University of California, Santa Cruz, USA)
  Power-intrinsic Bayesian variable selection in Gaussian linear models
- Sylvia Richardson (Imperial College London, UK)
  Hierarchically related regression structures for integrative genomics analysis
- Raquel Prado (University of California, Santa Cruz, USA)
  Bayesian models for multiple physiological time series in multi-subject studies

Room C *Special topic session.* **Beta processes: extensions and applications**

Organizer: Nils Lid Hjort (University of Oslo, Norway)

Chair: Sonia Petrone (Bocconi University, Italy)

Speakers:

- Yongdai Kim (Seoul National University, Korea)
  Analysis of credit risk history data using a mixture of beta-Dirichlet processes prior
- Sinead Williamson (Carnegie Mellon University, USA)
  Inducing dependency between beta processes using a Poisson process representation
- Tamara Broderick (University of California, Berkeley, USA)
  The beta process, exchangeable feature models, and the feature paintbox
- Nils Lid Hjort (University of Oslo, Norway)
  A brief history of beta processes

Room D *Invited session.* **Auxiliary variable and particle MCMC methods**

Organizer: Omiros Papaspiliopoulos (University Pompeu Fabra, Spain)

Chair: Arnaud Doucet (University of Oxford, UK)

Speakers:

- Nicolas Chopin (CREST-ENSAE, France)
  ABC-EP: expectation propagation for likelihood-free Bayesian computation
- Anthony Lee (University of Warwick, UK)
  Auxiliary variables, active particles and locally adaptive Monte Carlo methods
- Pierre Jacob (CEREMADE, Université Paris Dauphine, France)
  SMC^2*: an algorithm for exact sequential inference in state space models*

Discussant: Arnaud Doucet (University of Oxford, UK)

10:30–11:00 Coffee break

11:00–12:30

Room A *Special topic session.* **Scaling Bayesian computation to handle big data: methods and feasibility**

Organizer: Dawn Woodard (Cornell University, USA)

Chair: Galin Jones (University of Minnesota, USA)

Speakers:

- Michael Jordan (University of California, Berkeley, USA)
  Massive data and the bootstrap
- Samuel Kou (Harvard University, USA)
  Multi-resolution inference of stochastic models from partially observed data
- Yves Atchade (University of Michigan, USA)
  Potts-modulated Gaussian processes for high-dimensional output computer models
- Dawn Woodard (Cornell University, USA)
  Scalability of Markov chain methods for genomic motif discovery

Room B *Invited session.* **Partial identification and causal inference: what can Bayes bring to the table?**

Organizer: Paul Gustafson (University of British Columbia, Canada)

Chair: Fan Li (Duke University, USA)

Speakers:

- Toru Kitagawa (University College London, UK)
  Estimation and inference for set-identified parameters using posterior lower probability
- Chenguang Wang (Johns Hopkins University, USA)
  A Bayesian causal effect model with weak stochastic assumption for clinical trials with incomplete longitudinal outcomes
- Paul Gustafson (University of British Columbia, Canada)
  Bayesian inference in partially identified models

Discussant: David Stephens (Mc Gill University, Canada)

Room C *Invited session.* **Bayesian analysis of network data: from network determination to network modeling**

Organizer: Abel Rodriguez (University of California, Santa Cruz, USA)

Chair: Michele Guindani (University of Texas MD Anderson Cancer Center, USA)

Speakers:

- Veera Baladandayuthapani (University of Texas MD Anderson Cancer Center, USA)
  Bayesian sparse graphical models and their applications
- Natesh Pillai (Harvard University, USA)
  On a class of shrinkage priors for covariance matrix estimation
- Abel Rodriguez (University of California, Santa Cruz, USA)
  Dynamic models for financial trading networks

Discussant: Andrew Thomas (Carnegie Mellon University, USA)

Room D *Special topic session.* **Applied Bayesian Econometrics**

Organizer: Sylvia Frühwirth-Schnatter (Vienna University of Economics and Business, Austria)

Chair: Sylvia Frühwirth-Schnatter (Vienna University of Economics and Business, Austria)

Speakers:

- Hedibert Lopes (University of Chicago Booth School of Business, USA)
  On the long run volatility of stocks
- Michael Smith (Melbourne Business School, Australia)
  Estimation of copula models with discrete margins via Bayesian data augmentation

- Mattias Villani (Linköping University, Sweden)
  Bayesian mixture models for large microeconomic panels
- Helga Wagner (University of Linz, Austria)
  Bayesian treatment effects models for panel outcomes

12:30–14:00 Lunch

12:30–14:00 *Room A:* **Young Bayesian Meeting**

Conveners: Andrew Cron (Duke University, USA), Marian Farah (MRC Biostatistics Unit, Cambridge, UK), Francesca Ieva (Politecnico di Milano, Italy)

13:00–14:00 *Room B:* **George Casella Memorial Session**

Convener: Christian Robert (Université Paris Dauphine, France)

14:00–14:45 **Keynote Lecture**

Room A Chair: Merlise Clyde (Duke University, USA)

Speaker: Chris C. Holmes (University of Oxford, UK)

Bayesian methods in cancer genomics

15:00–16:30

Room A *Special topic session.* **Bayesian methods for Spatial Statistics**

Organizer: Mikyoung Jun (Texas A&M University, USA)

Chair: Bruno Sanso (University of California, Santa Cruz, USA)

Speakers:

- Murali Haran (Pennsylvania State University, USA)
  Computer model calibration with high dimensional multivariate space-time observations
- Huiyan Sang (Texas A&M University, USA)
  Covariance approximation for large multivariate spatial datasets
- Gabriel Huerta (Indiana University, USA)
  Estimating parametric uncertainties in computationally enabled strategies for climate model development
- Hongxia Yang (IBM Watson Research, USA)
  Adaptive sampling for Bayesian geospatial models

Room B *Special topic session.* **Bayesian analysis of protein structure and evolution**

Organizer: Scott Schmidler (Duke University, USA)

Chair: Scott Schmidler (Duke University, USA)

Speakers:

- Christopher Challis (Duke University, USA)
  Bayesian evolutionary modeling of protein structures
- Chieh-Hsi Wu (University of Auckland, New Zealand)
  Revealing the across site heterogeneity of nucleotide substitution patterns using Dirichlet process mixture model and Bayesian model selection
- Irina Czogiel (Max Planck Institute for Molecular Genetics, Germany)
  Bayesian alignment of unlabeled marked point sets using random fields - molecular shape analysis
- Scott Schmidler (Duke University, USA)
  Some theoretical bounds on Bayesian estimation of evolutionary distance

Room C *Invited session.* **On the uses of random probabilities in Bayesian inference**

Organizer: Antonio Lijoi (University of Pavia, Italy)

Chair: Fernando Quintana (Pontificia Universidad Católica de Chile, Chile)

Speakers:

- Peter Orbanz (Columbia University, USA)
  Constructions of priors on probability measures
- Ramses Mena (UNAM, Mexico)
  An EPPF from independent sequences of geometric random variables
- Antonio Lijoi (University of Pavia, Italy)
  Discrete random probabilities for Bayesian inference on species variety

Discussant: Julyan Arbel (CREST, Université Paris Dauphine, France)

Room D *Special topic session.* **Bayesian Econometrics II**

Organizer: Roberto Leon-Gonzalez (National Graduate Institute for Policy Studies, Japan)

Chair: Rodney Strachan (Australian National University, Australia)

Speakers:

- Liana Jacobi (University of Melbourne, Australia)
  Consumption decisions in markets with limited accessibility: the case of cannabis
- Rodney Strachan (Australian National University, Australia)
  Time varying dimension models
- Roberto Leon-Gonzalez (National Graduate Institute for Policy Studies, Japan)
  Fat-tailed Gamma autoregressive processes for stochastic volatility with jumps
- Joshua Chan (Australian National University, Australia)
  Efficient samplers for dynamic factor models with stochastic volatilities

---

16:30–17:00 Coffee break

17:00–18:30

Room A *Special topic session.* **Bayesian analysis of astronomical data**

Organizer: Joseph M. Hilbe (Arizona State University, USA)

Chair: David van Dyk (University of California, Irvine, USA)

Speakers:

- Roberto Trotta (Imperial College London, UK)
  Bayesian hierarchical modeling of cosmological supernovae type Ia data
- Ashish Mahabal (California Institute of Technology, USA)
  Using Bayesian networks for real-time classification of transients
- Ninan Philip (St. Thomas College, Kozhencheri, India)
  Bayesian analysis of sparse astronomical data matrices
- Radu S. Stoica (Université Lille 1, France), Vincent Martinez (University of Valencia, Spain), Enn Saar (Tartu Observatory, Estonia)
  Use of marked point processes in a Bayesian framework for detecting and characterizing cosmic galactic filaments

Room B *Special topic session.* **Recent advances in Bayesian variable selection**

Organizer: Joyee Ghosh (University of Iowa, USA)

Chair: Philip Brown (University of Kent, UK)

Speakers:

- Christopher Hans (Ohio State University, USA)
  Structuring dependence in regression: radius mixtures of spherically uniform priors
- Mahlet Tadesse (Georgetown University, USA)
  A stochastic partitioning method to associate high-dimensional datasets
- Scotland Leman (Virginia Tech, USA)
  The multiset model selector
- Joyee Ghosh (University of Iowa, USA)
  Data augmentation and sandwich algorithms for Bayesian model averaging

Room C *Invited session.* **Nonparametric Bayes applications in Biostatistics**

Organizer: Luis Nieto-Barajas (ITAM, Mexico)

Chair: Athanasios Kottas (University of California, Santa Cruz, USA)

Speakers:

- Michele Guindani (University of Texas MD Anderson Cancer Center, USA)
  Species Sampling prior for the analysis of array CGH data
- Lorenzo Trippa (Harvard University, USA)
  A nonparametric approach for validating, comparing and integrating predictive models
- Luis Nieto-Barajas (ITAM, Mexico)
  Bayesian analysis of functional proteomics profiles

Discussant: Sujit Ghosh (North Carolina State University, USA)

Room D *Special topic session.* **Advances in honest Monte Carlo**

Organizer: Mark Huber (Claremont McKenna College, USA)

Chair: Mark Huber (Claremont McKenna College, USA)

Speakers:

- James Flegal (University of California, Riverside, USA)
  Expectation and quantile estimation via Markov chain Monte Carlo
- Mark Huber (Claremont McKenna College, USA)
  Fast approximation algorithms for partition functions of Gibbs distributions
- Galin Jones (University of Minnesota, USA)
  Markov chain Monte Carlo: can we trust the third significant figure?
- Alexandros Beskos (University College London, UK)
  Sequential Monte Carlo methods in high dimensions

18:30–21:00 *Seminar Room No. 1–3:* **Poster session**

Thursday, June 28th

08:00–09:00 Registration

09:00–10:30

Room A *Invited session.* **Applications of particle filtering and sequential updating**

Organizer: Robert Kohn (University of New South Wales, Australia)

Chair: Jonathan Stroud (George Washington University, USA)

Speakers:

- Julien Cornebise (UK)
  Practicalities in implementing Adaptive Particle MCMC
- Scott Sisson (University of New South Wales, Australia)
  Regression density estimation for approximate Bayesian computation
- Robert Kohn (University of New South Wales, Australia)
  Bayesian Inference for Complex Time Series Models

Discussant: Hedibert Lopes (University of Chicago Booth School of Business, USA)

Room B *Special topic session.* **Bayesian graphical and factor models: structure, sparsity and dimension**

Organizer: Mike West (Duke University, USA)

Chair: Emily Fox (University of Pennsylvania, USA)

Speakers:

- Ryo Yoshida (Research Organization of Information and Systems, Institute of Statistical Mathematics, Japan)
  Bayesian sparse reconstruction: Latent factor analysis of gene regulatory programs
- Andrew Cron (Duke University, USA)
  Modeling sparse full-rank orthogonal matrices
- Hao Wang (University of South Carolina, USA)
  Efficient Gaussian graphical model determination under G-Wishart prior distributions
- P. Richard Hahn (University of Chicago Booth School of Business, USA)
  Sparse factor model approaches to the weak instrument problem

Room C *Special topic session.* **Bayesian methods in biological, environmental and ecological systems**

Organizer: Debashis Mondal (University of Chicago, USA)

Chair: David Dahl (Texas A&M University, USA)

Speakers:

- Donatello Telesca (University of California, Los Angeles, USA)
  Statistical issues in nanometerial toxicology
- Saman Muthukumarana (University of Manitoba, Canada)
  Modelling heterogeneity in mark-recapture data using the Dirichlet process
- Veronica J. Berrocal (University of Michigan, USA)
  Directional weights CAR models using Gaussian process mixing
- Debashis Mondal (University of Chicago, USA)
  Spatial analysis of environmental bioassays

Room D *Invited session.* **Bayesian semi-parametric analysis: theory**

Organizer: Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)

Chair: Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)

Speakers:

- Harry van Zanten (Eindhoven University of Technology, Netherlands)
  Asymptotic theory for empirical Bayes procedures
- Ismael Castillo (CNRS Paris, France)
  Semiparametric Bernstein-von Mises theorem, low regularity
- Pierpaolo De Blasi (University of Torino and Collegio Carlo Alberto, Italy)
  Bayesian estimation of the discrepancy with misspecified parametric models

Discussant: Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)

10:30–11:00 Coffee break

11:00–12:30

Room A *Special topic session.* **Approximate Bayesian computation (ABC): likelihood-free Bayesian inference II**

Organizers: Christian Robert (Université Paris Dauphine, France), Scott Sisson (University of New South Wales, Australia)

Chair: Marc Suchard (University of California, Los Angeles, USA)

Speakers:

- David Nott (National University of Singapore, Singapore)
  Approximate Bayesian computation and Bayes linear analysis: towards high-dimensional ABC
- Michael Blum (CNRS, Université Joseph Fourier, France)
  A comparison of dimension reduction methods in approximate Bayesian computation
- Jean-Michel Marin (Université Montpellier 2, France)
  Estimation of demo-genetic model probabilities with approximate Bayesian computation using linear discriminant analysis on summary statistics
- Oliver Ratman (Duke University, USA)
  Summary errors, conflict and evidence synthesis - exploiting the cutting edges of approximate Bayesian computation to analyze the phylodynamics of human pathogens

Room B *Invited session.* **Time Series analysis and Finance**

Organizer: Siddartha Chib (Washington University, St. Louis, USA)

Chair: Hedibert Lopes (University of Chicago Booth School of Business, USA)

Speakers:

- Antonietta Mira (University of Lugano, Switzerland)
  A Bayesian estimator of the multivariate covariance of noisy and asynchronous returns
- Jouchi Nakajima (Duke University, USA)
  Multivariate dynamic sparsity modelling in financial time series
- Yasuhiro Omori (University of Tokyo, Japan)
  A class of multivariate stochastic volatility models with leverage

Discussant: Teruo Nakatsuma (Keio University, Japan)

Room C *Special topic session.* **Bayesian model assessment**

Organizer: Joris Mulder (Tilburg University, Netherlands)

Chair: Brad Carlin (University of Minnesota, USA)

Speakers:

- Merlise Clyde (Duke University, USA)
  Perspectives on Bayesian model combination
- Laura Hatfield (Harvard Medical School, USA)
  Identifiability and learning in Bayesian joint longitudinal-survival models
- Gonzalo García-Donato (University of Castilla-La Mancha, Spain)
  On the development of formal criteria to determine objective model selection priors
- Joris Mulder (Tilburg University, Netherlands)
  Default Bayes factors for comparing (in)equality constrained models

Room D *Invited session.* **Problem-driven developments in Bayesian nonparametrics**

Organizer: Sonia Petrone (Bocconi University, Italy)

Chair: Ramses Mena (UNAM, Mexico)

Speakers:

- Yee Whye Teh (University College London, UK)
  A Bayesian nonparametric model for ranking data
- Sergio Bacallado (Stanford University, USA)
  A Bayesian nonparametric model for reversible Markov chains with applications in molecular dynamics
- Sonia Petrone (Bocconi University, Italy)
  Bayes and empirical Bayes: do they merge?

Discussant: Raffaele Argiento (CNR-IMATI, Italy)

12:30–14:00 Lunch

14:00–14:45 **Keynote Lecture**

Room A Chair: Ed George (University of Pennsylvania, USA)

Speaker: Stephen G. Walker (University of Kent, UK)
Bayesian nonparametrics and conditional distributions

15:00–16:30

Room A *Special topic session.* **Case studies of Bayesian success stories: babies, trials and ratings**

Organizer: Peter Mueller (University of Texas at Austin, USA)

Chair: Peter Mueller (University of Texas at Austin, USA)

Speakers:

- Suchi Saria (Harvard University and Johns Hopkins University, USA)
  Modeling at-risk infants: Bayesian nonparametrics to the rescue
- Brad Carlin (University of Minnesota, USA)
  Success stories in Bayesian adaptive methods for phase I-II clinical trials
- Yuan Ji (Northshore University Healthcare System, USA)
  Clustering, network, epigenetics
- Mark Glickman (Boston University, USA)
  Measuring competitor strength in games and sports through an approximate Bayesian filter: the Glicko system

Room B *Special topic session.* **Parallel processing in Bayesian computing**

Organizers: Christian Robert (Université Paris Dauphine, France), Marc Suchard (University of California, Los Angeles, USA)

Chair: Christian Robert (Université Paris Dauphine, France)

Speakers:

- Marc Suchard (University of California, Los Angeles, USA)
  Ridiculously parallel, serial Bayesian inference algorithms
- Steven Scott (Google, USA)
  Distributed hierarchical logistic regression
- Jarad Niemi (Iowa State University, USA)
  Rejection sampling on a graphical processing unit
- Kenichiro McAlinn (Keio University, Japan)
  GPGPU parallel computing for Bayesian portfolio selection with massive number of assets

Room C *Invited session.* **Adaptive Bayesian function estimation**

Organizer: Subhashis Ghosal (North Carolina State University, USA)

Chair: Subhashis Ghosal (North Carolina State University, USA)

Speakers:

- Eduard Belitser (Eindhoven University of Technology, Netherlands)
  Projection oracle convergence rate of posterior
- Surya Tokdar (Duke University, USA)
  Dimension adaptability of Gaussian process models with variable selection and projection
- Heng Lian (Nanyang Technological University, Singapore)
  Posterior convergence rates for wavelet regression

Discussant: Subhashis Ghosal (North Carolina State University, USA)

Room D *Special topic session.* **Bayesian Econometrics III**

Organizer: Tomohiro Ando (Keio University, Japan)

Chair: Tomohiro Ando (Keio University, Japan)

Speakers:

- Tomohiro Ando (Keio University, Japan)
  Factor augmented approach for predicting stock market behavior
- Toshiaki Watanabe (Hitotsubashi University, Japan)
  Bayesian analysis of identifying restrictions for the time-varying parameter vector autoregressive model
- Nalan Basturk (Erasmus University Rotterdam, Netherlands)
  Instrumental variables, errors in variables, and simultaneous equations models: applicability and limitations of direct Monte Carlo
- Taeryon Choi (Korea University, Korea)
  Bayesian analysis of partially linear regression models

16:30–17:00 Coffee break

17:00–18:30

Room A *Invited session.* **Spatial state-space models**

Organizer: Dani Gamerman (Federal Univeristy of Rio de Janeiro, Brazil)

Chair: Gabriel Huerta (Indiana University, USA)

Speakers:

- Bruno Sanso (University of California, Santa Cruz, USA)
  Blending and downscaling ensembles of climate model predictions
- Ioanna Manolopoulou (Duke University, USA)
  Diffusion modeling of motion trajectories under the influence of covariates
- Christopher Strickland (Queensland University of Technology, Australia)
  Change point detection in multivariate time series and space time data sets

Discussant: Sudipto Banerjee (University of Minnesota, USA)

Room B *Special topic session.* **Bayesian modeling and its applications in social science**

Organizer: Jun Lu (American University, USA)

Chair: Jeff Gill (Washington University, St. Louis, USA)

Speakers:

- Dongchu Sun (University of Missouri-Columbia, USA)
  Bayesian analysis of variance
- Geoff Nicholls (University of Oxford, UK)
  Phylogenetic models for the ancestry of a cultural trait with applications to the analysis of vocabulary trait data
- Chong He (University of Missouri-Columbia, USA)
  Adjusting nonresponse bias in small area estimation via a Bayesian hierarchical spatial model
- Jun Lu (American University, USA)
  A Bayesian analysis to explicit and implicit memory

Room C *Special topic session.* **High dimensional graphical models in genomics**

Organizers: Christopher Yau (University of Oxford, UK), Leonardo Bottolo (Imperial College London, UK)

Chair: Anthony Lee (University of Warwick, UK)

Speakers:

- Francesco Stingo (University of Texas MD Anderson Cancer Center, USA)
  Bayesian hierarchical models for data integration in genomics
- Silvia Liverani (Imperial College London, UK)
  Bayesian dependence model for regulatory networks
- Alex Lenkoski (Heidelberg University, Germany)
  Hierarchical Gaussian graphical models and the G-Wishart distribution
- Leonardo Bottolo (Imperial College London, UK)
  Stochastic recurrent heavy subgraphs of denoised weighted networks

Room D *Special topic session.* **Bayesian methods in reliability**

Organizer: Refik Soyer (George Washington University, USA)

Chair: Fabrizio Ruggeri (CNR-IMATI, Italy)

Speakers:

- Thomas Mazzuchi (George Washington University, USA)
  A Bayesian analysis of variable reliability growth
- Refik Soyer (George Washington University, USA)
  To survive or to fail: what is the question?
- J. Andrés Christen (CIMAT, Mexico)
  Fully sequential analysis of accelerated life testing
- Simon P. Wilson (Trinity College Dublin, Ireland)
  Inference on phase-type models via MCMC with application to networks of repairable redundant systems

---

18:30–21:00 *Seminar Room No. 1–3:* **Poster session**

Friday, June 29th

08:00–09:00 Registration

09:00–10:30

Room A *Invited session.* **Savage Award Session**

Organizer: ISBA

Chair: Bertrand Clarke (University of Miami, USA)

Speakers:

- Gun Ho Jang (University of Pennsylvania, USA)
  Invariant procedures in model checking, checking for prior-data conflict and Bayesian analysis
- Kaisey Mandel (Imperial College London, UK)
  Improving cosmological distances to illuminate dark energy: hierarchical Bayesian models for type Ia supernovae in the optical and near-infrared
- Juan Carlos Martínez-Ovando (Banco de México, Mexico)
  On stationary modelling for time-series data
- Fabian Scheipl (Ludwig-Maximilians-Universität München, Germany)
  Spike-and-slab priors for function selection in structured additive regression models

Room B *Invited session.* **Topics in Bayesian Statistics**

Organizer: Raquel Prado (University of California, Santa Cruz, USA)

Chair: Raquel Prado (University of California, Santa Cruz, USA)

Speakers:

- Valen E. Johnson (University of Texas MD Anderson Cancer Center, USA)
  Bayesian model selection in ultrahigh-dimensional settings
- Michael Daniels (University of Florida, USA)
  Proper Bayesian inference with missingness and auxiliary information
- Hans R. Kuensch (ETH Zurich, Switzerland)
  A new tool for Bayesian model diagnostics

Discussant: Steven MacEachern (Ohio State University, USA)

Room C *Special topic session.* **Bayesian spatio-temporal disease mapping: new frontiers**

Organizer: Susanna Cramb (Cancer Council Queensland and Queensland University of Technology, Australia)

Chair: Veronica J. Berrocal (University of Michigan, USA)

Speakers:

- Susanna Cramb (Cancer Council Queensland and Queensland University of Technology, Australia)
  Spatio-temporal cancer mapping: Bayesian dynamic factor models
- Nicole White (Queensland University of Technology, Australia)
  Spatial modelling of health service provision and utilisation
- Guangquan Li (Imperial College London, UK)
  BaySTDetect: detecting unusual temporal patterns in small area disease rates using Bayesian posterior model probabilities
- Marian Farah (MRC Biostatistics Unit, Cambridge, UK)
  Dynamic Bayesian modelling for emulation and calibration of epidemic models

Room D *Special topic session.* **Bayes modeling in marketing**

Organizer: Nobuhiko Terui (Tohoku University, Japan)

Chair: Nobuhiko Terui (Tohoku University, Japan)

Speakers:

- Greg Allenby (Ohio State University, USA)
  Modeling indivisible demand
- Makoto Abe (University of Tokyo, Japan)
  A brand purchase model of consumer goods incorporating
- Fumiyo Kondo (University of Tsukuba, Japan)
  Bayesian dynamic factor analysis and the corresponding clusters for changing needs
- Nobuhiko Terui (Tohoku University, Japan)
  Dynamic brand satiation

10:30–11:00 Coffee break

11:00–12:30

Room A *Special topic session.* **Bayesian models for high-dimensional complex-structured data**

Organizer: Veera Baladandayuthapani (University of Texas MD Anderson Cancer Center, USA)

Chair: Veera Baladandayuthapani (University of Texas MD Anderson Cancer Center, USA)

Speakers:

- Philip Brown (University of Kent, UK)
  Flexible Bayesian sparsity modelling for high dimensional data
- Marco Ferreira (University of Missouri, USA)
  Bayesian hierarchical multi-subject multiscale analysis of functional MRI data
- Jeffrey Morris (University of Texas MD Anderson Cancer Center, USA)
  Robust, adaptive, Bayesian functional response regression
- Sudipto Banerjee (University of Minnesota, USA)
  Statistical inference on temporal gradients in regionally aggregated data

Room B *Invited session.* **S'Bayes: constructing and using subjective priors for Bayesian modelling**

Organizer: Kerrie Mengersen (Queensland University of Technology, Australia)

Chair: Kerrie Mengersen (Queensland University of Technology, Australia)

Speakers:

- Michael Goldstein (University of Durham, UK)
  Subjective belief specification for complex physical models
- Samantha Low-Choy (Queensland University of Technology, Australia)
  Strategies for aggregating multiple sources of subjective prior information in Bayesian analysis
- Jeff Gill (Washington University, St. Louis, USA)
  Revealing latent clusters from Dirichlet process mixtures models using product partitions

Discussant: Kerrie Mengersen (Queensland University of Technology, Australia)

Room C *Invited session.* **Bayesian analysis of inverse problems**

Organizer: Colin Fox (University of Otago, New Zealand)

Chair: Geoff Nicholls (University of Oxford, UK)

Speakers:

- Heikki Haario (Lappeenranta University of Technology, Finland)
  State and parameter estimation of large models
- Radu Herbei (Ohio State University, USA)
  Bayesian inverse problems via a Bernoulli factory
- Colin Fox (University of Otago, New Zealand)
  Polynomial accelerated Gibbs sampling for conductivity imaging, and other inverse problems

Discussant: J. Andrés Christen (CIMAT, Mexico)

Room D *Special topic session.* **Applications of non- and semi-parametric Bayesian methods**

Organizer: Fernando Quintana (Pontificia Universidad Católica de Chile, Chile)

Chair: Antonio Lijoi (University of Pavia, Italy)

Speakers:

- Raffaele Argiento (CNR-IMATI, Italy)
  A "density-based" algorithm for cluster analysis using Dirichlet process Gaussian mixture models
- Jim Griffin (University of Kent, UK)
  A Bayesian semiparametric model for yield curves
- Athanasios Kottas (University of California, Santa Cruz, USA)
  Nonparametric Bayesian analysis of developmental toxicity experiments with clustered discrete-continuous outcomes
- Fernando Quintana (Pontificia Universidad Católica de Chile, Chile)
  Cluster-specific variable selection for product partition models

---

12:30–14:00 Lunch

14:00–15:30 **Keynote Lectures**

Room A Chair: Fabrizio Ruggeri (CNR-IMATI, Italy)

Speakers:

- Arnaud Doucet (University of Oxford, UK)
  Bayesian inference for intractable state-space models
- Alan Gelfand (Duke University, USA)
  Demographic analysis of forest dynamics using stochastic integral projection models

15:30–16:00 Coffee break

16:00–17:30 *Room A:* General meeting

18:00–22:00 Banquet and Cabaret at Rhiga Royal Hotel

22:30–01:00 Afterparty at "Top of Kyoto" Rhiga Royal Hotel

Poster presenters and titles

Tuesday, June 26th

- Colin Aitken (University of Edinburgh, UK)
  The evaluation of evidence in forensic science for discrete data
- Lorraine Allchin (University of Oxford, UK)
  Bayesian machine learning methods for genome-wide association data
- Pierre Alquier (Université Paris 7 and CREST, France)
  PAC-Bayesian bounds for high-dimensionnal estimation
- Clair Alston (Queensland University of Technology, Australia)
  PyMCMC — a new alternative in the implementation of MCMC
- Osvaldo Anacleto (The Open University, UK)
  Dynamic graphical models for real-time multivariate road traffic flow forecasting
- Ioannis Andrianakis (National Oceanography Centre, UK)
  A Bayesian hierarchical model for the reconstruction of the sea level in the Mediterranean basin for the late 20th century
- Isadora Antoniano Villalobos (University of Kent, UK)
  A nonparametric regression model with normalized weights
- Julyan Arbel (CREST, Université Paris Dauphine, France)
  Multidimensional covariate dependent Dirichlet processes
- Raffaele Argiento (CNR-IMATI, Italy)
  Semi-Markov modelling of electricity co-generation in residential applications with time-dependent covariates
- Richard Arnold (Victoria University of Wellington, New Zealand)
  Multicomponent systems with dependent failures
- Christian Asseburg (ESiOR Oy, Finland)
  Estimating Weibull parameters from Kaplan-Meier curves
- Paul Baines (University of California, Davis, USA)
  Interwoven EM algorithms
- Roderick Ball (Scion - New Zealand Forest Research Institute, New Zealand)
  Peeling and Bayesian QTL mapping for allo-polyploids
- Dipankar Bandyopadhyay (University of Minnesota, USA)
  Nonparametric spatial models for periodontal disease data with spatially-varying non-random missingness
- Anjishnu Banerjee (Duke University, USA)
  Infinite tensor factorization priors
- Lorna Barclay (University of Warwick, UK)
  Chain event graphs in Bayesian model selection for health studies
- Maria Asuncion Beamonte (Universidad de Zaragoza, Spain)
  A Bayesian geographically weighted regression model applied to real estate markets
- Candace Berrett (Brigham Young University, USA)
  Bayesian models for multicategory spatial data
- Fernando V. Bonassi (Duke University, USA)
  Mixture modeling strategies applied to approximate Bayesian computation
- Bjoern Bornkamp (Novartis Pharma AG, Switzerland)
  Approximating posterior densities by iterated Laplace approximations
- Luke Bornn (University of British Columbia, Canada)
  The product graphical model
- Lane Burgette (RAND Corporation, USA)
  Asset smoothing in the health and retirement study: a Bayesian nonparametric approach
- Ben Calderhead (University College London, UK)
  Bayesian modelling of ion channels
- Ewan Cameron (Queensland University of Technology, Australia)
  Approximate Bayesian computation for astronomical model analysis
- Antonio Canale (University of Torino, Italy)
  Bayesian multivariate mixed scale density estimation

- Francois Caron (Inria Bordeaux, France)
  Efficient Bayesian inference for (dynamic) generalized Bradley-Terry models
- Shu-Yu Chen (Feng Chia University, Taiwan)
  Bayesian testing unit roots versus threshold specifications
- Ting-Li Chen (Academia Sinica, Taiwan)
  Locally optimal MCMC sampling
- Delson Chikobvu (University of Free State, South Africa)
  Winter peak electricity load forecasting in South Africa using extreme value theory with a Bayesian flavour
- Sam Clifford (Queensland University of Technology, Australia)
  Bayesian spatio-temporal modelling of ultrafine particle number concentration from a panel design
- David Conesa (University of Valencia, Spain)
  Assessing the spatial distribution of species using Bayesian latent Gaussian models
- Forrest Crawford (University of California, Los Angeles, USA)
  Birth-death regression
- Madeleine Cule (University of Oxford, UK)
  Modelling the transmission of Clostridium difficile in hospitals
- Jonathan Cumming (Durham University, UK)
  Emulating expensive decision choices with application to computer models of complex physical systems
- Gabriela Cybis (University of California, Los Angeles, USA)
  Bayesian nonparametric clustering in phylogenetics: modeling antigenic evolution in influenza
- David Dahl (Texas A&M University, USA)
  Cluster analysis via random partition distributions
- Moumita Das (Indian Statistical Institute, India)
  Transformation based reversible jump MCMC
- Maria De Iorio (University College London, UK)
  Inference on population structure using random partition models
- Haydar Demirhan (Hacettepe University, Turkey)
  Forecasting election results via the sequential nature of Bayes' theorem and an application to the Turkish general elections
- Sophie Donnet (Université Paris Dauphine, France)
  Bayesian estimation for partially observed multiplicative intensity Poisson processes
- Oya Ekici (Duke University, USA)
  Estimation of short term private external debt of turkey within the DLM framework
- Tarek El Moselhy (Massachusetts Institute of Technology, USA)
  Bayesian inference with optimal maps
- Katie Evans (University of Rochester, USA)
  Outlier identification for model-based clustering: an application to the Seychelles child development and nutrition study
- Christopher Eves (University of Kent, UK)
  An industrial application of Bayesian statistics
- Mariel Finucane (Harvard University, USA)
  Semiparametric estimation of population-level distributions of nutritional markers using disparate data sources
- Chiara Franco (University of Essex, UK)
  Modelling the dynamics of CaCO3 budgets in a changing environment using a Bayesian belief network approach
- Kassie Fronczyk (University of Texas MD Anderson Cancer Center, USA)
  A nonparametric Bayesian approach to the analysis of bioassay experiments with ordinal responses
- Kenji Fukumizu (Institute of Statistical Mathematics, Japan)
  Kernel Bayes' rule and its applications
- Agnes Fussl (Johannes Kepler University of Linz, Austria)
  Efficient MCMC methods for estimating binomial logit models
- Chris Gamble (University of Oxford, UK)
  A Bayesian approach to detecting signals of recent natural selection in large scale phased haplotype data
- Souparno Ghosh (Duke University, USA)
  Ontogenetic niche shifts among plant species of eastern United States
- Robert Gramacy (University of Chicago Booth School of Business, USA)
  A regression-adjusted Hockey plus-minus Metric

- Carolina Granado (Venezuelan Foundation for Seismological Research, Venezuela)
  Application of Bayes' theorem to study seismic events and active faults in Venezuela
- Thiago Guerrera (Norwegian University of Science and Technology, Norway)
  Fast Bayesian approximate inference in near-Gaussian latent models
- Gilles Guillot (Technical University of Denmark, Denmark)
  Integrating phenotypic, genetic and geographic data in the analysis of population structure
- Jan Hannig (University of North Carolina at Chapel Hill, USA)
  Model penalized inference
- Yuning He (University of California, Santa Cruz, USA)
  Predicting variable-length functional outputs for emulation of a NASA flight simulator
- Charles Hogg (National Institute of Standards and Technology, USA)
  Geometry-based hyperpriors for nonstationary lengthscales
- Hwanhee Hong (University of Minnesota, USA)
  Hierarchical Bayesian methods for combining efficacy and safety in multiple treatment comparisons
- Ibuki Hoshina (Chuo University, Japan)
  Algorithm for constructing sparse regression models via Bayesian lasso
- Reza Hosseini (University of Southern California, USA)
  Modeling air pollution mixtures in Southern California
- Leanna House (Virginia Tech, USA)
  Revolutionizing the education of data analytics using Bayesian interactive visualizations
- Yung-Hsiang Huang (National Taiwan University, Taiwan)
  A marker-set approach using Bayesian regression model with regularization for family association studies
- John Hughes (University of Minnesota, USA)
  Dimension reduction and alleviation of confounding for spatial generalized linear mixed models
- Francesca Ieva (Politecnico di Milano, Italy)
  Semiparametric Bayesian modelling for the classification of patients with observed high survival probability
- Kaoru Irie (University of Tokyo, Japan)
  Nonparametric stochastic volatility: mixture approach
- Irina Irincheeva (Duke University, USA)
  Bayesian additive latent variable models for multivariate densities
- Tsukasa Ishigaki (Tohoku University, Japan)
  Toward large-scale customer analysis in the framework of random utility model
- Tsunehiro Ishihara (Hitotsubashi University, Japan)
  Multivariate realized stochastic volatility model with leverage
- Väinö Jääskinen (University of Helsinki, Finland)
  Learning of sequence data by compression with sparse Markov Chains
- Teresa Jacobson (University of California, Santa Cruz, USA)
  Gaussian process transfer function models for discretely sampled but continuous time series
- Shane Jensen (University of Pennsylvania, USA)
  A level-set hit-and-run sampler for quasi-concave distributions
- Seongil Jo (Seoul National University, Korea)
  Nonparametric Bayesian modeling for spatially smooth density estimation
- Alicia Johnson (Macalester College, USA)
  Convergence rates among component-wise MCMC algorithms for Bayesian inference
- Ian Johnston (Boston University, USA)
  Graph-regularized centroid estimation on a hierarchical Bayesian model for genome-wide association studies
- Edmund Jones (University of Bristol, UK)
  Efficient searching in tree and forest graphical models
- Anthony Quinn (Trinity College Dublin, Ireland)
  Bayesian inference with hard parameter constraints
- Erik Vanem (University of Oslo, Norway)
  Bayesian hierarchical space-time models for the North Atlantic ocean wave climate

Wednesday, June 27th

- Satoshi Kabe (University of Tsukuba, Japan)
  Estimating the Markov switching almost ideal demand system: a Bayesian approach
- Kengo Kamatani (Osaka University, Japan)
  Asymptotic properties of Monte Carlo strategies for cumulative link model
- Su Yun Kang (Queensland University of Technology, Australia)
  Bayesian hierarchical models for analyzing spatial point-based data at a grid level: a comparison of approaches
- George Karabatsos (University of Illinois-Chicago, USA)
  A Bayesian unimodal density regression model, with applications
- Gregor Kastner (Vienna University of Economics and Business, Austria)
  Efficient Bayesian inference for multivariate factor stochastic volatility (SV) models
- Kentaro Kato (Center for Research on Educational Testing, Japan)
  Assessing prior distributions for the item parameters in the two-parameter logistic IRT model
- Hiroaki Katsura (Keio University, Japan)
  Parallel particle learning for Bayesian asset price prediction
- Peter Kecskemethy (University of Oxford, UK)
  Genome-wide structural variation analysis
- Khaled Khatab (RWTH Aachen University, Germany)
  Application of Bayesian approach on the child's health problems
- Bartek Knapik (VU University Amsterdam, Netherlands)
  Semiparametric posterior limits under local asymptotic exponentiality
- Genya Kobayashi (Kobe University, Japan)
  Stable priors for linear regression model
- Daiki Koizumi (Cyber University, Japan)
  On the loss functions of Bayesian forecasting for WWW traffic based on time varying Poisson distribution
- Michalis Kolossiatis (Cyprus University of Technology, Cyprus)
  Comparing distributions using dependent normalized random measure mixtures
- Tsuyoshi Kunihama (Duke University, USA)
  Bayesian modeling of temporal dependence in large sparse contingency tables
- Yuta Kurose (University of Tokyo, Japan)
  Dynamic equicorrelation stochastic volatility
- Tristan Launay (Université de Nantes, France)
  Construction of an informative hierarchical prior distribution for electricity load forecasting
- Dominic Lee (University of Canterbury, New Zealand)
  Bayesian vertex nomination
- Kate Lee (Auckland University of Technology, New Zealand)
  Threshold selection method for modelling multivariate extremes using a Bayesian measure of surprise
- Min Lee (University of Southampton, UK)
  Multiply imputing missing values in data sets with mixed measurement scales using a sequence of generalized linear models
- Thomas Leininger (Duke University, USA)
  Quadratic scaling models for spatial compositional data with application to forest fragmentation and land use/land cover classification
- Maxime Lenormand (IRSTEA, France)
  Adaptive approximate Bayesian computation for complex models
- Feng Li (Stockholm University, Sweden)
  Efficient Bayesian multivariate surface regression
- Guangquan Li (Imperial College London, UK)
  Bayesian methods for subnational mortality forecasts
- Li-Jung Liang (University of California, Los Angeles, USA)
  Longitudinal actor-partner interdependence models for studies on HIV-impacted families
- Merrill Liechty (Drexel University, USA)
  Revealed preferences for portfolio selection — does skewness matter?
- Chien-Hua Lin (Providence University, Taiwan)
  A study of feedforward and feedback EWMA controller for multistage processes

- Edward M. H. Lin (Feng Chia University, Taiwan)
  Bayesian estimation of smoothly mixing time-varying parameter GARCH models
- Pei-Chun Lin (National Cheng Kung University, Taiwan)
  Ticket vending service and customer waiting: evidence from Taiwan high speed rail system
- Tanzy Love (University of Rochester, USA)
  Effect modification using latent mixture analysis
- Li Ma (Duke University, USA)
  Bayesian recursive model selection
- Steven MacEachern (Ohio State University, USA)
  Clustered Bayesian model averaging
- Ying MacNab (University of British Columbia, Canada)
  Bayesian disease mapping with sparse data and smoothing — Poisson or zero-inflated Poisson?
- Tiep Mai (Trinity College Dublin, Ireland)
  Short term traffic flow forecasting with spatial temporal model
- Himel Mallick (University of Alabama at Birmingham, USA)
  A new Bayesian LASSO
- Juan Carlos Martínez-Ovando (Banco de México, Mexico)
  On marked duration models and the study of price setting strategies
- Yuzo Maruyama (University of Tokyo, Japan)
  Robust Bayesian variable selection with sub-harmonic priors
- Samira Masoumi (University of Waterloo, Canada)
  A Bayesian sequential approach in model discrimination
- Kazuki Matsuda (Chuo University, Japan)
  Nonlinear statistical modeling via the relevance vector machine and its application to change point analysis
- Takeshi Matsuda (Cyber University, Japan)
  Predictive distribution of SQL injection attacks detection model
- Clare McGrory (University of Queensland, Australia)
  Fast approximate inference for Bayesian mixture models with application to volume estimation of tissue types from computed tomography
- Zairul Nor Deana Md Desa (University of Kansas and Universiti Teknologi Malaysia, USA)
  Gibss sampling estimation method for subscores estimation using bi-factor multidimensional item response theory
- Hsin-Chao Mi (National Tsing Hua University, Taiwan)
  Stability analysis of the EWMA controller for mixed-product processes
- Minna Miettinen (University of Helsinki, Finland)
  Bayesian methods for modeling phenotypic time-series and whole-genome sequence data for bacteria
- Masaki Misonou (Keio University, Japan)
  Redefining the momentum strategy: a Bayesian approach
- Shakir Mohamed (University of British Columbia, Canada)
  Large-scale approximate Bayesian inference for exponential family latent Gaussian models
- Silvia Montagna (Duke University, USA)
  Bayesian latent factor regression for functional and longitudinal data
- Paul J. Mostert (University of Stellenbosch, South Africa)
  A Bayes analysis of the censored Rayleigh model when using a generalised hypergeometric as a prior
- Alexis Muir Watt (University of Oxford, UK)
  Monte Carlo inference for partial orders
- Sabyasachi Mukhopadhyay (Indian Statistical Institute, India)
  An improved Bayesian semiparametric model for palaeoclimate reconstruction
- Jared Murray (Duke University, USA)
  Joint stochastic blockmodeling of attributed random graphs
- Sungmin Myoung (Jungwon University, South Korea)
  Mixture of component models for hidden stratification in repeated measures data
- Haruhisa Nagata (Tokyo Institute of Technology, Japan)
  Bayesian posterior probability computation of identified network communities by bootstrap resampling
- Kenji Nagata (University of Tokyo, Japan)
  Analysis of acceptance rate for Metropolis algorithm and relationship between Bayesian learning

- Yuji Nakayama (Osaka Prefecture University, Japan)
  A Bayesian econometric analysis of auction under variable reserve price and unknown number of potential bidders
- Consuelo R. Nava (University of Torino, Italy)
  Bayesian construction and estimation of GIG stationary models
- Maxim Nazarov (Bocconi University, Italy)
  Functional representation of Markov exchangeable sequences
- Abdoul Aziz Junior Ndoye (GREQAM, AMSE, France)
  Bayesian unconditional quantile regression: An analysis of recent expansions in wage structure and earnings inequality in the U.S. 1992-2009
- Ronald Neath (Hunter College, City University of New York, USA)
  Regenerative simulation for Metropolised Gibbs samplers
- Bernardo Nipoti (University of Texas MD Anderson Cancer Center, USA)
  A Bayesian semi-parametric model for covariate dependent survival data
- Haruhisa Nishino (Chiba University, Japan)
  Bayesian estimation of income distribution using grouped data
- Hidehisa Noguchi (Tokyo University of Science, Japan)
  Bayesian Lasso for screening experiments
- Hisashi Noma (Kyoto University School of Public Health, Japan)
  The optimal discovery procedure in multiple significance testing: an empirical Bayes approach
- Didit Budi Nugroho (Kwansei Gakuin University, Japan)
  Parameter estimation in LNSV models: griddy Gibbs versus Metropolis-Hastings
- Takashi Oga (Chiba University, Japan)
  Sampling EGARCH parameters
- Cheongeun Oh (New York University, USA)
  Bayesian inference of associations and SNP-environment interactions for GWAS-identified RA risk alleles with risk of infection via perfect sampling scheme
- Man-Suk Oh (Ewha Womans University, Korea)
  Bayesian multiple comparison of models for binary data with inequality constraints
- Kensuke Okada (Senshu University, Japan)
  Bayesian model averaging in factor analysis to estimate factor reliability
- Thais Paiva (Duke University, USA)
  Imputation of confidential datasets with spatial locations using point process models
- Julia Palacios (University of Washington, USA)
  Gaussian process-based estimation of population size trajectories from gene sequence data. A phylodynamic approach
- Anastasios Panagiotelis (Monash University, Australia)
  Bayesian estimation and model selection for high dimensional copula models
- Georgios Papageorgiou (Imperial College London, UK)
  Bayesian nonparametric spatial modeling with applications in environmental epidemiology
- Justinas Pelenis (Institute for Advanced Studies, Vienna, Austria)
  Bayesian semiparametric regression
- Stefano Peluso (Università della Svizzera Italiana, Switzerland)
  Estimation of the multivariate covariance of noisy and asynchronous returns
- Chien-Yu Peng (Academia Sinica, Taiwan)
  Optimal classification policy for highly reliable products
- Johan Pensar (Åbo Akademi University, Finland)
  Context specific graphical models
- Alberto Pessia (University of Helsinki, Finland)
  Bayesian bidirectional clustering method for detection of gene loss and acquisition processes
- Matti Pirinen (University of Oxford, UK)
  Modeling heterogeneity in genotype-phenotype associations
- Dale Poirier (University of California, Irvine, USA)
  Multivariate versus multinomial probit: when are binary decisions made separately also jointly optimal?
- Wolfgang Polasek (Institute for Advanced Studies, Vienna, Austria)
  MCMC estimation for the extended Hodrick-Prescott (eHP) filter

- Alexandra Posekany (Vienna University of Economics and Business, Austria)
  Robust microarray analysis with mixtures of Gaussians and heavy-tailed student's t distributions
- Benedict Powell (Durham University, UK)
  Calibrated smoothing: sharing structure between computer simulations and historical data

Thursday, June 28th

- Sarah Filippi (Imperial College London, UK)
  Rational threshold schemes for approximate Bayesian computation via unscented transforms
- Paul Kirk (University of Warwick, UK)
  MDI: a method for integrative modeling of multiple functional genomics datasets
- Aurore Lavigne (UMR 518 AgroParisTech/INRA, France)
  Spatio-temporal modeling of avalanche frequencies in the French Alps
- Leslie Pibouleau (INSERM UMR 717 and Université Paris 7, France)
  Bayesian hierarchical meta-analysis model for medical device evaluation
- Vinayak Rao (University College London, UK)
  MCMC for continuous-time discrete-state systems: beyond uniformization
- Lizanne Raubenheimer (Rhodes University, South Africa)
  Bayesian inference on nonlinear functions of Poisson rates
- Perla Reyes (University of California, Santa Cruz, USA)
  Analysis of collaboration networks: elucidating firms relationships
- Vincent Rivoirard (Université Paris Dauphine, France)
  Bernstein-von Mises theorem for linear functionals of the density
- Josemar Rodrigues (Federal University of São Carlos, Brazil)
  A Bayesian correlated cure rate model under a hybrid latent activation schemes
- Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)
  Asymptotic properties of the posterior distributions for HMMs with applications to the selection of the prior
- Vivekananda Roy (Iowa State University, USA)
  Estimating link function parameters in robust Bayesian binary regression
- Elizabeth Ryan (Queensland University of Technology, Australia)
  Optimal sampling times for high dimensional pharmacokinetic studies
- Robin Ryder (CEREMADE, Université Paris Dauphine, France)
  The Wang-Landau algorithm reaches the flat histogram criterion in finite time
- Gustaf Rydevik (University of York, UK)
  Can you hindcast the epidemic curve of an infectious disease
- Daniel Sabanés Bové (University of Zurich, Switzerland)
  Mixtures of g-priors for objective Bayes selection of generalized additive models
- Gen Sakurai (National Institute for Agro-Environmental Sciences, Japan)
  Application of Bayesian method for estimating past effect of climate change on crop production
- Tim Salimans (Erasmus University Rotterdam, Netherlands)
  Automated structured variational Bayes by stochastic natural gradient descent
- Jean Bernard Salomond (ENSAE-CREST and Université Paris Dauphine, France)
  Bayes test for monotonicity
- Christian Schäfer (ENSAE-CREST, CEREMADE and Université Paris Dauphine, France)
  Parallelized sequential Monte Carlo for large scale Bayesian variable selection
- Susanne Schmitz (Trinity College Dublin, Ireland)
  Bayesian MTC models to combine evidence from different sources of evidence. An application in rheumatoid arthritis
- Paulo Serra (Technical University of Eindhoven, Netherlands)
  Posterior contraction rates in estimating holder smooth Poisson intensities
- Hidetoshi Shimodaira (Tokyo Institute of Technology, Japan)
  An MCMC method for estimating the rate of preferential attachment in growing networks with missing timelines
- Shinichiro Shirota (University of Tokyo, Japan)
  Realized stochastic volatility with leverage and long memory

- Isabelle Smith (Laboratoire des Sciences du Climat et de l'Environnement, France)
  Generalization of the posterior distribution of the likelihood ratio to composite vs composite hypotheses testing
- Theresa Smith (University of Washington, USA)
  Bayesian inference for small area health data using the G-Wishart prior
- Reza Solgi (University of Lugano, Switzerland)
  A Bayesian semiparametric multiplicative error model
- Hiroko Solvang (Oslo University Hospital, Norway)
  Identification of Bayesian causal association models for investigation among gene expression, genotype variation and fatigue of breast cancer
- Michio Sonoda (Cyber University, Japan)
  Markov chain Monte Carlo method simulation of SQL injection attack detection
- Alberto Sorrentino (University of Warwick, UK)
  Particle filtering for estimation of stationary dipoles in magnetoencephalogaphy
- Polina Sporysheva (Bocconi University, Italy)
  Bivariate species sampling models
- Rebecca Steorts (University of Florida, USA)
  On estimation of mean squared errors of benchmarked empirical Bayes estimators
- Shinya Sugawara (University of Tokyo, Japan)
  Nonparametric analysis of a price mechanism of the Japanese private nursing home market via Polya tree mixture for a simultaneous model of demand
- Botond Szabo (Technical University of Eindhoven, Netherlands)
  Exploring the asymptotic behaviour of empirical Bayes procedures
- Ryota Takada (Keio University, Japan)
  Spatial interdependence of fiscal efficiency among Japanese local governments
- Fuyuhiko Tanaka (University of Tokyo, Japan)
  Minimax Bayesian predictive density operator
- Rusty Tchernis (Georgia State University, USA)
  Modeling area-level health rankings
- Satoshi Teramukai (Kyoto University Hospital, Japan)
  A Bayesian predictive sample size selection design for single-arm exploratory clinical trials
- Maria Terres (Duke University, USA)
  Analyzing first flowering event data using survival models with spatio-temporal covariates
- Thomas Thorne (Imperial College London, UK)
  Interaction networks changing with time
- Adrien Todeschini (Inria Bordeaux, France)
  BiiPS: A software for inference in Bayesian graphical models with sequential Monte Carlo methods
- Minh Ngoc Tran (University of New South Wales, Australia)
  Model selection for regression density estimation with mixtures of heteroscedastic experts
- Sofia Tsepletidou (Université Paris Dauphine, France)
  Computational Bayesian tools for modeling the aging process
- Hiroki Tsurumi (Rutgers University, USA)
  Comparing three MCMC algorithms: probability integral transformation, griddy Gibbs, and random walk draws
- Catalina Vallejos (University of Warwick, UK)
  Robust Bayesian methods for survival analysis
- Raquel Vasquez (Venezuelan Foundation for Seismological Research, Venezuela)
  Bayesian estimation of the spatial variation of the completeness magnitude for the Venezuelan seismic catalogue
- Ian Vernon (Durham University, UK)
  Emulation and efficient history matching of stochastic systems biology models
- Cristiano Villa (University of Kent, UK)
  An objective prior for the number of trials in a binomial distribution
- Sara Wade (Bocconi University, Italy)
  More informative density regression with multivariate covariates
- Charlotte Wang (National Taiwan University, Taiwan)
  SNP-set association test with Hamming distance

- Fangpo Wang (Duke University, USA)
  Analyzing spatial directional data using projected Gaussian processes
- Shi-Heng Wang (National Taiwan University, Taiwan)
  A Bayesian hierarchical model for family association studies with rare CNV variants
- Wenyi Wang (University of Texas MD Anderson Cancer Center, USA)
  Analysis of differential gene expression in heterogeneous tumor samples
- Xue Wang (University of Kent, UK)
  Bayesian nonparametric estimation of a copula
- Bertil Wegmann (Linköping University, Sweden)
  Bayesian inference in structural second-price auctions with both private-value and common-value bidders
- Chi-Chung Wen (Tamkang University, Taiwan)
  Bayesian analysis of current status data with missing covariates
- Nuttanan Wichitaksorn (University of Sydney, Australia)
  Efficient MCMC estimation of some elliptical copula regression models through scale mixture of normal
- Daniel Williamson (Durham University, UK)
  What is the value of longer AMOC observations?
- Amy Wilson (University of Edinburgh, UK)
  The evaluation of evidence relating to traces of drugs on banknotes
- Jesse Windle (University of Texas at Austin, USA)
  Forecasting high-dimensional, time-varying covariance matrices: factor stochastic volatility vs. realized covariances
- Tomasz Wozniak (University of Melbourne, Australia)
  Granger causal analysis of VARMA GARCH models
- Chieh-Hsi Wu (University of Auckland, New Zealand)
  Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov chain Monte Carlo
- Yuefeng Wu (University of California, Santa Cruz, USA)
  Consistency of stochastic blockmodels
- Yuh-Jenn Wu (Chung Yuan Christian University, Taiwan)
  Robustness of prior information to the predictive power
- Jie Xiong (University of Helsinki, Finland)
  Predictive sequential classification
- Masanao Yajima (University of California, Los Angeles, USA)
  Differential patterns of interaction and Gaussian graphical models
- Yohei Yamamoto (University of Alberta, Canada)
  Time instability of the U.S. monetary system: multiple break tests and reduced rank TVP VAR
- Jin Yan (University of Maryland, College Park, USA)
  Hidden Markov model and reversible jump algorithm for the analysis of areas of interest from eye tracking data
- Takemi Yanagimoto (Chuo University, Japan)
  Permissive boundary prior function as a virtually proper prior density
- Jae Ho Yoon (POSCO Research Institute, Korea)
  Instrumental variables estimation of a flexible nonlinear model
- Hideo Yunoue (Chiba University, Japan)
  Spatial patterns of flypaper effects for local expenditure by policy objective in Japan: a Bayesian approach
- Zamira Zamzuri (Macquarie University, Australia)
  Spatio temporal model for multivariate traffic accident count data
- Mahmoud Zarepour (University of Ottawa, Canada)
  The interplay of frequentist and Bayesian nonparametric Statistics
- Jing Zhou (University of North Carolina at Chapel Hill, USA)
  Nonparametric Bayes methods for estimation in high-dimensional contingency tables

ISBA Lectures on Bayesian Foundations

Monday, June 25th

13:00–14:15 **Confidence in nonparametric credible sets?**

Aad van der Vaart (University of Leiden, Netherlands)

In nonparametric statistics the posterior distribution is used in exactly the same way as in any Bayesian analysis. It supposedly gives us the likelihood of various parameter values given the data. A difference with parametric analysis is that it is often difficult to have an intuitive understanding of the prior, which affects the believability of the posterior distribution as a quantification of uncertainty. A second difference is that the posterior distribution is much more sensitive to the prior: its “fine properties” matter. This is true even in the asymptotic situation when the informativeness of the data increases indefinitely. In this talk we start by reviewing frequentist asymptotic results and insights on posterior distributions in the semi- and nonparametric setting obtained in the last decade. These results show that posterior distributions can be effective in recovering a true parameter provided some care is taken when choosing a prior. We next go on to ask whether posterior distributions are also capable in giving a correct idea of error in the reconstructions. Are credible sets in any way comparable to confidence regions? We shall not present an answer to this question, but show by example that it will be delicate.

Coauthors: Bartek Knapik, Botond Szabo and Harry van Zanten

⁂

14:15–15:30 **Bayesian dynamic modelling**

Mike West (Duke University, USA)

Since the 1970s, applications of Bayesian time series models and forecasting methods have represented major success stories for our discipline. Dynamic modelling is a very broad field, so this ISBA Lecture on Bayesian Foundations will rather selectively note key concepts and some core model contexts, leavened with extracts from a few time series analysis and forecasting examples from various application fields. The Lecture with then link into and briefly discuss a range of recent developments in exciting and challenging areas of Bayesian time series analysis.

⁂

16:00–17:15 **Approximate Bayesian computation (ABC): advances and questions**

Christian Robert (Université Paris Dauphine, France)

The lack of closed form likelihoods has been the bane of Bayesian computation for many years and, prior to the introduction of MCMC methods, a strong impediment to the propagation of the Bayesian paradigm. We are now facing models where an MCMC completion of the model towards closed-form likelihoods seems unachievable and where a further degree of approximation appears unavoidable. In this tutorial, I will present the motivation for approximative Bayesian computation (ABC) methods, the various implementations found in the current literature, as well as the inferential, rather than computational, challenges set by these methods.

⁂

17:15–18:30 Slowly but surely, Bayesian ideas revolutionize medical research

Donald A. Berry (University of Texas MD Anderson Cancer Center, USA)

Bayesian theory is elegant and intuitive. But elegance may have little value in practical settings. The "Bayesian Revolution" of the last half of the 20th century was irrelevant for biostatisticians. They were busy changing the world in another way, and they neither needed nor wanted more methodology than they already had. The randomized controlled trial (RCT) came into existence in the 1940s and it changed medical research from an art into a science, with biostatisticians guiding the process. To make matters worse for the reputation of Bayesians, we seemed to be anti-randomization, and medical researchers feared we wanted to return them to the dark ages.

The standard approach to clinical experimentation is frequentist, which has advantages and disadvantages. One disadvantage is that unit of statistical inference is the entire experiment. As a consequence, the RCT has remained largely unchanged. It is still the gold standard of medical research, but it can make research ponderously slow. And it is not ideally suited for the "personalized medicine" approach of today, identifying which types of patients benefit from which therapies.

In this presentation I'll chronicle the increased use of the Bayesian perspective in medical research over this period. An important niche regards adaptive design. I'll describe a variety of approaches, most of which employ randomization, and all employ Bayesian updating. Accumulating trial results are analyzed frequently with the possibility of modifying the trial's future course based on the overall theme of the trial. It is possible to have many treatment arms. Including combination therapies enables learning how treatments interact with each other as well as the way they interact with biomarkers of disease that are specific to individual patients. I will give an example (called I-SPY 2) of a Bayesian adaptive biomarker-driven trial in neoadjuvant breast cancer. The goal is to efficiently identify biomarker signatures for a variety of agents and combinations being considered simultaneously. Longitudinal modeling plays a vital role.

Although the Bayesian approach supplies important tools for designing informative and efficient clinical trials, I've learned to not try to change things too abruptly. In particular, we can stay rooted in the well established frequentist tradition by evaluating false-positive rates and statistical power using simulation. The most exciting aspect of this story is the potential for utilizing Bayesian ideas in the future to build ever more efficient study designs and associated processes for developing therapies, based on the existing solid foundation.

❉

Keynote lectures

Tuesday, June 26th

14:00–14:45 **Data assimilation and sequential Bayes filters: information fusion with numerical simulation**

Tomoyuki HIGUCHI (Institute of Statistical Mathematics, Japan)

Data assimilation (DA) is a synthesis technique based on the Bayesian filtering method by embedding observation/experiment data in a numerical simulation. It yields an accommodation ability to make a simulation real, and the better initial and boundary conditions can be automatically obtained. DA is an emerging area in earth sciences, particularly oceanography. In statistical methodology, DA can be formulated in the state space model that draws much interest of the researchers in various domains such as the time series analysis, signal processing, and control theory. There are two types of DA in terms of a methodology; sequential DA and variational (non-sequential) DA. An ensemble-based sequential DA (EnSDA) has an advantage in terms of less human resources which is achieved by plugging into the existing "forward" simulation codes. We briefly explain a recent advancement in EnSDA, and give a simple description on the relationship among the nonlinear non-Gaussian filters. A part of applications carried out by our DA research group (Research and Development Center for Data Assimilation: http://daweb.ism.ac.jp/en/) is demonstrated with an emphasis on what and how we develop the EnSDA methods for realizing DA on high performance computing (HPC), in particular, the world's fastest supercomputer "Kei".

❉

Wednesday, June 27th

14:00–14:45 **Bayesian methods in cancer genomics**

Chris C. HOLMES (University of Oxford, UK)

Cancer occurs through the sequential accumulation of genetic alterations in particular regions of the genome involved in cell growth and survival. One of the most important types of alteration is copy-number change whereby stretches of DNA are duplicated or deleted. Recent advances in DNA sequencing and array-based technologies now allow for the routine measurement of such events at 1,000,000s of positions across a genome. This talk will discuss research in general Bayesian methods for the robust detection of copy-number-aberrations in cancer. At one level of abstraction this can be thought of as a task in change-point modelling or genome segmentation. We will describe how Bayesian signal-processing methods using Hidden Markov models scaled to large-data are ideally suited to this task using informative priors and decision theoretic loss functions to provide predictions under the posterior model.

Analysis is complicated by the highly heterogeneous nature of tumours, which typically contain multiple sub-types and clinical samples are often contaminated with normal DNA. This introduces a deconvolution task where the data can be represented as a mixture of signals each with their own unknown copy-number profile. Again, Bayesian methods using mixtures of hidden Markov models and auxiliary variables work well here. The techniques are demonstrated on a number of on-going real world studies.

❉

Thursday, June 28th

14:00–14:45 **Bayesian nonparametrics and conditional distributions**

Stephen G. WALKER (University of Kent, UK)

In many statistical contexts, it is the modeling of a conditional distribution which is required. Key examples include time series and regression. Bayesian nonparametrics has clear ideas in the case of independent and identically distributed data, but are less clear in conditional non-identically distributed settings. This talk will discuss how Bayesian nonparametric thinking can be unified to cover both situations.

❉

Friday, June 29th

14:00–14:45 **Bayesian inference for intractable state-space models**

Arnaud DOUCET (University of Oxford, UK)

Carrying out Bayesian inference in presence of intractable likelihood functions is a computationally challenging problem arising in numerous applications. Standard Monte Carlo strategies are not directly applicable in this context as the Metropolis-Hastings ratio cannot be computed exactly. I will review various approaches which allow us to bypass the intractability of the likelihood function and will discuss novel strategies. The resulting methods will be demonstrated in a variety of applications.

* * *

14:45–15:30 **Demographic analysis of forest dynamics using stochastic integral projection models**

Alan GELFAND (Duke University, USA)

Demographic analysis for plant and animal populations is a prominent problem in studying ecological processes. Typical demography studies age, size, length, or mass distributions. Historical treatment employs matrix projection models, creating classification bins, to model bin to bin transitions using individual level data. Integral projection models (IPMs) offer a continuous version of this approach, providing evolving *mean* surfaces over time to explain the dynamics in such trait distributions.

The general form for such models is $\gamma_{t+1}(y) = \int_D K_t(y|x)\gamma_t(x)dx$ where k_t is a time varying redistribution kernel and D is a suitable interval. These models are a class of integro-differential equations and, as such, may be driven by a partial differential equation. However, for demography, it is more natural to specify the redistribution kernel mechanistically using demographic functions, i.e., parametric models for demographic processes such as survival, growth, and replenishment. Evidently, there is much opportunity for modeling with IPMs, incorporating process features and available environmental covariates but, in addition, there are various ways to introduce stochasticity.

With interest in scaling in space, we work with data in the form of point patterns rather than with individual level data (hopeless to scale) yielding intensities (which are easy to scale). As a result, we learn about the IPM through marginal (e.g., annual) patterns. We also clarify that estimating K_t through individual- level data fails to align estimated intensities with observed intensities, resulting in inadequate models.

Fitting IPMs in our setting is quite challenging and is most feasibly done either by working in the spectral domain or with a *pseudo*-IPM, in conjunction with Laplace approximation. We illustrate with an investigation of forest dynamics using data from Duke Forest as well as a national survey called the Forest Inventory Analysis (FIA), maintained by the US Forest Service.

Coauthors: S. Ghosh and J.S. Clark

* * *

Invited sessions

Tuesday, June 26th

09:00–10:30 Room A **Advances in Gaussian processes**

Organizer: Herbie Lee (University of California, Santa Cruz, USA)
Chair: Robert Gramacy (University of Chicago Booth School of Business, USA)
Discussant: Kate Calder (Ohio State University, USA)

Bayesian nonstationary Gaussian process models via treed process convolutions

Waley LIANG (University of California, Santa Cruz, USA)

Spatial modeling with stationary Gaussian processes (GPs) has been widely used, but the assumption that the correlation structure is independent of the spatial location is invalid in many applications. Various nonstationary GP models have been developed to solve this problem, however, many of them become impractical when the sample size is large. To tackle this problem, we develop a process convolutions-based GP model by convolving a smoothing kernel with a partitioned latent process. Nonstationarity in the GP is obtained by allowing the variability of the latent process and the kernel size to change across partitions. Partitioning is achieved using a method similar to that of Classification and Regression Trees, which results in a binary tree structure. A Bayesian approach is used to simultaneously guide the partitioning process and estimate the parameters of the treed model.

Coauthors: Herbert Lee

❊

Efficient emulators of computer experiments using compactly supported correlation functions

Cari KAUFMAN (University of California, Berkeley, USA)

Statistical emulators of computer simulators have proven to be useful in a variety of applications. The widely adopted model for emulator building, using a Gaussian process model with strictly positive correlation function, is computationally intractable when the number of simulator evaluations is large. We propose a new model that uses a combination of low-order regression terms and compactly supported correlation functions to recreate the desired predictive behavior of the emulator at a fraction of the computational cost. Following the usual approach of taking the correlation to be a product of correlations in each input dimension, we show how to impose restrictions on the ranges of the correlations, giving sparsity, while also allowing the ranges to trade off against one another, thereby giving good predictive performance. We illustrate the method using data from a computer simulator of photometric redshift with 20,000 simulator evaluations and 80,000 predictions.

Coauthors: Derek Bingham

❊

Gaussian process modeling of derivative curves

Herbie LEE (University of California, Santa Cruz, USA)

Gaussian process (GP) models provide nonparametric methods tofit continuous curves observed with noise. Here we develop a GP-based inverse method that allows for the direct estimation of the derivative of a curve, motivated by a cosmological problem on the nature of dark energy. In principle, a GP model may be fit to the data directly, with the derivatives obtained by means of differentiation of the correlation function. However, it is known that this approach can be inadequate due to loss of information when differentiating. Instead, we present a new method of obtaining the derivative process by viewing this procedure as an inverse problem, andwe use the properties of a GP to obtain a computationally efficient fit. We apply our methodology to help with our understanding of dark energy.

Coauthors: Tracy Holsclaw, Bruno Sanso, David Higdon, Katrin Heitmann, Ujjaini Alam and Salman Habib

❊

09:00–10:30 Room B **Bayesian methods in Biostatistics**

Organizer: Valen E. Johnson (University of Texas MD Anderson Cancer Center, USA)
Chair: David Rossell (IRB Barcelona, Spain)
Discussant: Donatello Telesca (University of California, Los Angeles, USA)

A novel empirical Bayes approach for profiling immunoprecipitation-based DNA methylation

Andrea RIEBLER (University of Zurich, Switzerland)

DNA methylation plays an important role in gene regulation and aberrations occur in many diseases, most notably cancer. For example, tumor suppressor genes are silenced through increased methylation at CpG-island-associated promoters and oncogenes become activated through demethylation. Genome-wide DNA methylation analyses are essential to identify genes commonly interrupted in disease, to discover markers related to disease progression and for understanding biological mechanisms. There are three major approaches to analyze DNA methylation, all of which have platform-specific biases and vary in cost and resolution. We focus on the statistical analysis of immunoprecipitation (IP) capture data, such as MBD-seq. The efficiency of IP capture varies widely across the genome, but is known to depend on sequence-specific properties, such as CpG density. We present a novel empirical Bayes model based on the Poisson distribution that takes advantage of fully methylated (SssI) control data in two ways: i) we model SssI IP data jointly with IP data from a sample of interest to preserve the linearity of the methylation estimation; ii) we get information about the region-specific read density as a function of CpG density. Furthermore, we can model copy number variation (CNV), which directly affects read densities, and specify context-specific priors on the methylation level. Notably, we derive an analytic expression for the posterior methylation level, thus avoiding the high computational cost of competing methods. Our algorithm performs better compared to state-of-the-art techniques using lung fibroblast MBD-seq data, where “true” methylation levels are available from whole genome bisulphite sequencing. For cancer cell line data, the estimation performance improves significantly when modelling CNV.

Coauthors: Mark D. Robinson

⁂

A Bayesian graphical model for ChIP-Seq data on histone modifications

Peter MUELLER (University of Texas at Austin, USA)

Histone modifications (HMs) are an important post-translational feature. Different types of HMs are believed to co-regulate biological processes such as gene expression, and therefore are intrinsically dependent on each other. We develop inference for this complex biological network of HMs based on a graphical model for the dependence structure across HMs. A critical computational hurdle in the inference for the proposed graphical model is the evaluation of a normalization constant in an autologistic model that builds on the graphical model. We tackle the problem by Monte Carlo evaluation of ratios of normalization constants. We carry out a set of simulations to validate the proposed approach and to compare it with a standard approach using Bayesian networks. We report inference on HM dependence in a case study with ChIP-Seq data from a next-generation sequencing experiment. An important feature of our approach is that we can report coherent probabilities and estimates related to any event or parameter of interest, including honest uncertainties. Posterior inference is obtained from a joint probability model on latent indicators for the recorded HMs. We illustrate this in the motivating case study.

Coauthors: Riten Mitra, Shoudan Liang, Lu Yue and Yuan Ji

⁂

Bayesian de novo characterization of alternative splicing using high-throughput sequencing

David ROSSELL (IRB Barcelona, Spain)

High-throughput sequencing offers an unprecedented resolution to study gene alternative splicing (AS). AS has important implications for human health, e.g. it is involved in multiple diseases and cellular malfunctions. De novo characterization of AS, i.e. not conditioning on a set of known variants, poses very serious statistical challenges. The main issue is that the problem, which can be seen as a particular case of model selection, has a supra-exponential number of possible models.

Our proposed Bayesian framework improves upon previous approaches by (i) strongly enforcing parsimony by specifying non-local priors (NLPs) and (ii) incorporating valuable, publicly available, prior knowledge. NLPs have been shown in previous work to have good theoretical properties for sparse model selection. Integrating the experimental data with the accumulated prior knowledge allows recovering the set of expressed variants at a lower sequencing depth. We explore large posterior probability models via a random-walk MCMC tailor-suited to this problem. The framework is computationally efficient and shows good performance.

Coauthors: Camille Stephan-Otto Attolini, Manuel Kroiss and Almond Stöcker

⁂

11:00–12:30 Room A **Model selection**

Organizer: Susie Bayarri (University of Valencia, Spain)
Chair: David Conesa (University of Valencia, Spain)
Discussant: Gonzalo García-Donato (University of Castilla-La Mancha, Spain)

Shrinkage adjustment for model selection

Ed GEORGE (University of Pennsylvania, USA)

Should Bayesians worry about model selection? Maybe no, because after conditioning on the data, the Bayesian paradigm will make all the proper adjustments. Maybe yes, because selection which only reports effects that are “large enough” tends to yield upward biases. In this talk I will argue that the Bayesian paradigm will indeed make the proper adjustments, but only when a selection adjusted likelihood is used to account for selection uncertainty. In specific cases, this is seen to lead to additional selection shrinkage regardless of the prior. A general adjusted approach to Bayesian model selection is proposed.

❊

A new look at logit likelihoods

James SCOTT (University of Texas at Austin, USA)

We propose a new data-augmentation scheme for estimating models with logit likelihoods. This talk will build up the relevant theory involving Polya-Gamma distributions, and demonstrate three useful features of the approach. First, it leads to simple EM and Gibbs-sampling algorithms, circumventing the need for analytic approximations, numerical integration, Metropolis-Hastings, or variational methods. Second, it allows modelers much more flexibility when choosing priors, which have usually come from the Dirichlet or logistic-normal family. Finally, our approach naturally suggests a default strategy for prior selection based on thelogistic-Z model, which is strongly related to Jeffreys’ prior for the binomial. To illustrate the method we focus primarily on two cases: multiway contingency tables with fixed margins, and high-dimensional logistic regression. But the general approachencompasses many other common situations, including topic models, network models, the Cox model, anddiscrete-choice models.

Coauthors: Nicholas Polson, Liang Sun and Jesse Windle

❊

TESS: the effective sample size for linear models

Susie BAYARRI (University of Valencia, Spain)

Objective Bayes criteria for model selection crucially depend on n, the “sample size.” It appears explicitly in the formulae of asymptotic Bayes factors (Laplace, BIC and generalizations, etc.). Similarly, it is often used when choosing of the appropriate scale for objective (but proper) priors, which are generally needed to compute Bayes factors or posterior odds, or, more generally, in Bayes analysis under model uncertainty. There are, however, some difficulties in defining “n,” some of which are explored in this talk. Although there have been some attempts towards appropriately defining “n,” the problem remains essentially open. We provide in this talk a new definition which seems to apply satisfactorily to challenging situations, including examples for which the previous definitions fail. We call it TESS (The Effective Sample Size).

Coauthors: Jim Berger and Luis Pericchi

❊

11:00–12:30 Room D **Bayesian Econometrics**

Organizer: John Geweke (University of Technology, Sydney, Australia)
Chair: Yasuhiro Omori (University of Tokyo, Japan)
Discussant: Robert Kohn (University of New South Wales, Australia)

Flexible econometric modelling based on sparse finite mixtures

Sylvia FRÜHWIRTH-SCHNATTER (Vienna University of Economics and Business, Austria)

There has been a tremendous increase in applied work both in statistics as well as econometric using infinite mixtures based in particular on Dirichlet process priors. However, going to infinity makes the mathematics much more complicated , in particular, for applied statisticians and econometricians.

This talk investigates the concept of sparse finite mixture modelling, which is based on a shrinkage prior on the weights that removes all redundant components automatically. The sparsity prior provides an automatic tool to select the number K of component and avoids the cumbersome compution of the marginal likelihood for each K. Futhermore, it is shown how the label switching problem could be solved using the framework of sparse finite mixtures. In contrast to infinite mixtures, this allow identification of component-specific parameters and classification. This approach is applied to various issues in econometric modelling such as choosing a flexible link function in binary data modelling , model-based clustering based on finite mixtures of normals, and to switching regression models. Furthermore, the sparse finite mixture approach is compared to infinite mixtures based on Dirichlet process priors. An effcient method for computing the marginal likelihood of a sparse finite mixture is developped which makes direct comparison with Dirichlet process priors possible.

❈

Bayesian regression with nonparametric heteroskedasticity

Andriy NORETS (Princeton University, USA)

This paper presents a large sample justification for a semiparametric Bayesian approach to inference in a linear regression model. The approach is to model the distribution of the error term by a normal distribution with the variance that is a flexible function of covariates. It is shown that even when the data generating distribution of the error term is not normal the posterior distribution of the linear coefficients converges to a normal distribution with the mean equal to the asymptotically efficient estimator and the variance given by the semiparametric efficiency bound. This implies that the estimation procedure is robust and conservative from the Bayesian standpoint and at the same time it can be used as an implementation of semiparametrically efficient frequentist inference.

❈

Massively parallel sequential Monte Carlo for Bayesian inference

John GEWEKE (University of Technology, Sydney, Australia)

This paper reconsiders sequential Monte Carlo approaches to Bayesian inference in the light of massively parallel desktop computing capabilities now well within the reach of individual academics. It first develops an algorithm that is well suited to parallel computing in general and for which convergence results have been established in the sequential Monte Carlo literature but that tends to require manual tuning in practical application. It then introduces endogenous adaptations in the algorithm that obviate the need for tuning, using a new approach based on the structure of parallel computing to show that convergence properties are preserved and to provide reliable assessment of simulation error in the approximation of posterior moments. The algorithm is generic, requiring only code for simulation from the prior distribution and evaluation of the prior and data densities, thereby shortening development cycles for new models. Through its use of data point tempering it is robust to irregular posteriors, including multimodal distributions. The sequential structure of the algorithm leads to reliable and generic computation of marginal likelihood as a by-product. The paper includes three detailed examples taken from state-of-the-art substantive research applications. These examples illustrate the many desirable properties of the algorithm. and demonstrate that it can reduce computing time by several orders of magnitude.

Coauthors: Garland Durham

❈

15:00–16:30 Room B **Networks and relational data**

Organizer: Peter Hoff (University of Washington, USA)
Chair: Perla Reyes (University of California, Santa Cruz, USA)
Discussant: Abel Rodriguez (University of California, Santa Cruz, USA)

A Bayesian framework for indirect sampling of relational data

Tyler McCORMICK (University of Washington, USA)

Indirect sampling (Lavall'ee, 2007) is an increasingly common design for producing samples from populations without an easily specified sampling frame. Indirect sampling begins with a probability sample of persons or units that are not the target population. It then links the target population to the sampled units, forming a type of relational data and allowing probabilities of selection of the target population to be computed. We develop a Bayesian framework for analyzing this sort of indirectly sampled data. This framework maintains many consistency properties of previous design-based approaches but enjoys greatly reduced variance due to hierarchical smoothing. We demonstrate this framework first using Aggregated Relational Data (questions of the form "how many X's do you know?") to estimate the sizes of traditionally hard-to-count populations and then using verbal autopsy reports to estimate mortality in areas without well-developed vital records systems.

❈

Marginal-additive models and processes for network-correlated outcomes

Andrew THOMAS (Carnegie Mellon University, USA)

A key promise of social networks is the ability to detect and model the correlation of personal attributes along the structure of the network, in either static or dynamic settings. The basis for most of these models, the Markov Random Field on a lattice, has several assumptions that may not be reflected in real network data, namely the assumptions that the process is stationary on the lattice, and that the ties in the model are correctly specified. Additionally, it is less than clear how correlation over longer distances on networks can be adequately specified under the lattice mechanism, given the assumption of a stationary process at work.

Based on concepts from generalized additive models and spatial/geostatistical methods, I introduce a class of models that is more robust to the failure of these assumptions, more flexible to different definitions of network distance, and more generally applicable to large-scale studies of network phenomena. I apply this method to outcomes from two large-scale social network studies to demonstrate its use and versatility.

❈

Bayesian inference from non-ignorable network sampling designs

Simon LUNAGOMEZ (Harvard University, USA)

The objective is to perform inferences on a population quantity (like the incidence of HIV on a high-risk subpopulation) via sampling mechanisms based on a social network (link-tracing designs, RDS). We phrase this problem in terms of the framework proposed by Rubin (1976). A new notion of ignorability (graph-ignorability) is proposed for this context and it is proved that RDS is not graph-ignorable. We develop a general framework for making Bayesian inference on the population quantity that: models the uncertainty in the underlying social network, incorporates dependence among the individual responses according to the social network via a Markov Random Field, models the uncertainty regarding the sampling on the social network, and deals with the non-ignorability of the sampling design. The proposed framework is general in the sense that it allows a wide range of different specifications for the components of the model we just mentioned. Samples from the posterior distribution are obtained via Bayesian model averaging. Our model is compared with state of the art methods in simulation studies to show its superiority. An important contribution of our method is that it provides credible intervals for the population quantity, while our competitors can only give a point estimate.

Coauthors: Edoardo Airoldi

❈

17:00–18:30 Room A **Predictive inference and Bayes methods**

Organizer: Fumiyasu Komaki (University of Tokyo, Japan)
Chair: Yuzo Maruyama (University of Tokyo, Japan)
Discussant: Xinyi Xu (Ohio State University, USA)

Bayesian clustering stability

Bertrand CLARKE (University of Miami, USA)

We propose a new Bayesian method for evaluating the stability of a clustering. It rests on finding the probability of the set of coefficients that preserve the distance relationships among the data points and the cluster centroids. We verify that this stability criterion matches intuition in the sense of assigning high stability to clusterings that should be stable. We verify that asymptotically the stability assessment converges to a limit and that under reasonable conditions can be used to estimate the number of clusters consistently. In a series of examples we compare our method to several existing methods on the basis of correctly determining the number of clusters.

Coauthors: Hoyt Koepke

❉

Proper local scoring rules

Matthew PARRY (University of Otago, New Zealand)

Scoring rules have long been used to assess the quality of probabilistic predictions. We report on a recently discovered class of scoring rules with the remarkable property that they do not require knowledge of the normalization constant of the predictive model. On both continuous and discrete outcome spaces, we show how such scoring rules can be constructed from the idea of a score being local. One interesting consequence is that Besag's pseudolikelihood is given a firm theoretical foundation as a result of being a proper local scoring rule. We discuss recent applications of local scoring rules and connections to the work of Ehm & Gneiting.

Coauthors: Philip Dawid and Steffen Lauritzen

❉

On the use of latent information priors

Fumiyasu KOMAKI (University of Tokyo, Japan)

Construction methods of objective priors by using conditional mutual information between future observables and unknown parameters given observations are discussed. The priors maximizing the conditional mutual information are called latent information priors. These priors can be regarded as a generalization of k-reference priors and are often quite different from the Jeffreys priors or other widely used objective priors. Several properties and applications of prediction and inference based on the latent information priors are discussed. The performance of prediction is evaluated by the Kullback-Leibler divergence.

❉

Wednesday, June 27th

09:00–10:30 Room A **Being simultaneously Bayesian and frequentist**

Organizer: Kenneth Rice (University of Washington, USA)
Chair: Fumiyasu Komaki (University of Tokyo, Japan)
Discussant: David Stephens (Mc Gill University, Canada)

Likelihood of the empirical CDF: a coarsening approach to complex samples

Thomas LUMLEY (University of Auckland, New Zealand)

Standard Bayesian analyses of data from complex samples rely on either exchangeability or on modelling the sampling process. These are often satisfactory, but not ideal: for national multistage surveys the exchangeability assumptions are typically not true, and neighbourhood-level design variables used in constructing the sampling are often not present in public-use data sets. I will present initial investigations into a third non-ideal approach, based on treating the data as a single sample from the asymptotic Gaussian distribution of the empirical CDF. This approach separates the survey methodology used to estimate the asymptotic distribution from the Bayesian modelling used for inference, a split corresponding to the typical separation in expertise of data analysts.

⁂

Prediction in high dimensional datasets: an ensemble of shrinkage approaches

Bhramar MUKHERJEE (University of Michigan, USA)

With advancement in genomic technologies, it is common that two high-dimensional datasets are available, both measuring the same underlying biological phenomenon with different techniques. We consider predicting a continuous outcome Y using X, a set of p markers which is the best measure of the underlying biological process. This same biological process may also be measured by W, coming from prior technology but correlated with X. On a moderately sized sample we have (Y,X,W), and on larger sample we have (Y,W). We utilize the data on W to boost prediction of Y by X. When p is large and the subsample containing X is small, this is a p>n situation. When p is small, this is akin to the classical measurement error problem; however, ours is not the typical goal of calibrating W for use in future studies. We propose to shrink the regression coefficients of Y on X towards different targets that use information derived from W in the larger dataset, comparing these with the classical ridge regression of Y on X, which does not use W. We also unify all of these methods as targeted ridge estimators. Finally, we propose a hybrid estimator which is a linear combination of multiple estimators of the regression coefficients and balances efficiency and robustness in a data-adaptive way to theoretically yield smaller prediction error than any of its constituents. The methods are evaluated via simulation studies. We also apply them to a gene-expression dataset. mRNA expression of 91 genes is measured by quantitative real-time polymerase chain reaction (qRT-PCR) and microarray technology on 47 lung cancer patients with microarray measurements available on an additional 392 patients. The goal is to predict survival time using qRT-PCR. The methods are evaluated on an independent sample of 101 patients.

Coauthors: Philip S. Boonstra and Jeremy M. G. Taylor

⁂

Interpreting tests which are both Bayesian and frequentist

Kenneth RICE (University of Washington, USA)

Statistical testing has a long history of controversy; the Fisher and Neyman-Pearson approaches have fundamental differences, and neither of them agree with standard Bayesian procedures. This talk describes recent research which aims to clarify the differences between testing frameworks, by describing them all within a single decision-theoretic framework, where they are motivated as trade-offs. Different forms of testing are seen to trade off different quantities, and at different rates. As well as describing Bayesian justifications of frequentist procedures, we also consider the large-sample frequentist properties of the testing procedures, and their connection to shrinkage methods.

⁂

09:00–10:30 Room D **Auxiliary variable and particle MCMC methods**

Organizer: Omiros Papaspiliopoulos (University Pompeu Fabra, Spain)
Chair: Arnaud Doucet (University of Oxford, UK)
Discussant: Arnaud Doucet (University of Oxford, UK)

ABC-EP: expectation propagation for likelihood-free Bayesian computation

Nicolas CHOPIN (CREST-ENSAE, France)

Many models of interest in the natural and social sciences have no closed-form likelihood function, which means that they cannot be treated using the usual techniques of statistical inference. In the case where such models can be efficiently simulated, Bayesian inference is still possible thanks to the Approximate Bayesian Computation (ABC) algorithm. Although many refinements have since been suggested, the technique suffers from three major shortcomings. First, it requires introducing a vector of “summary statistics”, the choice of which is arbitrary and may lead to strong biases. Second, ABC may be excruciatingly slow due to very low acceptance rates. Third, it cannot produce a reliable estimate of the marginal likelihood of the model. We introduce a technique that solves the first and the third issues, and considerably alleviates the second. We adapt to the likelihood-free context a variational approximation algorithm, Expectation Propagation (Minka, 2001). The resulting algorithm is shown to be faster by a few orders of magnitude than alternative algorithms, while producing an overall approximation error which is typically negligible. Comparisons are performed in three real-world applications which are typical of likelihood-free inference, including one application in neuroscience which is novel, and possibly too challenging for standard ABC techniques.

Coauthors: Simon Barthelmé

⁂

Auxiliary variables, active particles and locally adaptive Monte Carlo methods

Anthony LEE (University of Warwick, UK)

The use of auxiliary variables in various Monte Carlo methods has proliferated both explicitly and implicitly over the last two decades, as our understanding of how to devise effective algorithms has grown. In addition, massively parallel ‘many-core’ processors have become the focus of the high performance computing community for a variety of physical reasons, providing a strong incentive for algorithms in computational statistics to exhibit specific types of parallelism. Within the field of Monte Carlo methodology, population-based methods such as sequential Monte Carlo, parallel tempering and pseudo-marginal methods are promising auxiliary variable algorithms that can take advantage of available parallel resources while allowing the advantageous interaction of simulated random variables. We present a new perspective on auxiliary variables within reversible MCMC kernels that allow for the construction of novel population-based MCMC, SMC and particle MCMC algorithms. These algorithms can utilize the availability of many-core processors and/or adapt their behaviour to characteristics of the target density of interest.

Coauthors: Christophe Andrieu and Arnaud Doucet

⁂

SMC^2: an algorithm for exact sequential inference in state space models

Pierre JACOB (CEREMADE, Université Paris Dauphine, France)

We consider the generic problem of performing sequential Bayesian inference in a state-space model with observation process y, state process x and fixed parameter theta. An idealized approach would be to apply the iterated batch importance sampling (IBIS) algorithm of Chopin (2002). This is a sequential Monte Carlo algorithm in the theta-dimension, that samples values of theta, reweights iteratively these values using the likelihood increments $p(y_t|y_{1:t-1}, theta)$, and rejuvenates the theta-particles through a resample-move step. In state-space models these likelihood increments are intractable in most cases, but they may be unbiasedly estimated by a particle filter in the x-dimension, for any fixed theta. This motivates a sequential Monte Carlo algorithm, defined in the theta-dimension, which propagates and resamples many particle filters in the x-dimension to “replace” the likelihood increments by unbiased estimates. On the other hand, the particle Markov chain Monte Carlo (PMCMC) framework developed in Andrieu et al. (2010) allows to design appropriate MCMC rejuvenation steps. Thus, the algorithm targets the correct posterior distribution at each iteration t, despite the intractability of the likelihood increments. We illustrate the method on various non-linear non-gaussian state space models.

Coauthors: Nicolas Chopin and Omiros Papaspiliopoulos

⁂

11:00–12:30 Room B **Partial identification and causal inference: what can Bayes bring to the table?**

Organizer: Paul Gustafson (University of British Columbia, Canada)
Chair: Fan Li (Duke University, USA)
Discussant: David Stephens (Mc Gill University, Canada)

Estimation and inference for set-identified parameters using posterior lower probability

Toru KITAGAWA (University College London, UK)

This paper develops inferences and statistical decisions for set-identified parameters from a robust Bayes perspective. When a model is set-identified, prior knowledge of model parameters is decomposed into two parts: one that can be updated by data (revisable prior knowledge) and one that can never be updated (unrevisable prior knowledge). We introduce a class of prior distributions that share a single prior distribution for revisable prior knowledge, but allow for arbitrary prior distributions for unrevisable prior knowledge. A posterior inference procedure proposed in this paper operates on the resulting class of posteriors by focusing on the posterior lower probability. We analyze point estimation of the set-identified parameters by applying the gamma-minimax criterion. We propose a robustified posterior credible region for the set-identified parameters by focusing on a contour set of the posterior lower probability. Our framework offers a procedure for eliminating set-identified nuisance parameters, and yields inferences for the marginalized identified set. For the case of an interval-identified parameter, we establish the Bernstein-von Mises type asymptotic equivalence of the lower probability inference to frequentist inference for the identified set.

⁂

A Bayesian causal effect model with weak stochastic assumption for clinical trials with incomplete longitudinal outcomes

Chenguang WANG (Johns Hopkins University, USA)

In randomized clinical trials, the causal effect of treatment is generally not identifiable when outcome data are unobserved because of early study exit caused by protocol defined events (e.g. death). Since the unobserved responses after early study exit are often undefined, standard missing data analysis methods which impute these responses are not applicable. We propose a Bayesian approach to estimate the causal treatment effect among those who would not have had protocol defined events regardless of treatment assignment. In this approach, we consider longitudinal stochastic assumptions that are weaker than commonly applied deterministic assumptions. For inference, we compute the upper and lower bounds of the causal effects and characterize the uncertainty associated with the estimated intervals using Bayesian methods.

Coauthors: Mike Daniels and Daniel Scharfstein

⁂

Bayesian inference in partially identified models

Paul GUSTAFSON (University of British Columbia, Canada)

Intuitively, partial identification describes a situation where the true value of the target parameter cannot be recovered in the large-sample limit, but some values of the target can be ruled out. Causal inference models, and other models for coarsened data, often give rise to partially identified models. While there is no conceptual problem in turning the Bayesian crank in a partially identified setting, there is clearly a need to understand the utility of such inference. Special attention is paid to the fact a non-Bayesian approach yields an identification region for the target parameter, whereas a Bayesian approach yields this region plus a distribution over the region. So, the extent to which the data can upweight or downweight values within the region, and the utility of this weighting, is of interest.

⁂

11:00–12:30 Room C **Bayesian analysis of network data: from network determination to network modeling**

Organizer: Abel Rodriguez (University of California, Santa Cruz, USA)
Chair: Michele Guindani (University of Texas MD Anderson Cancer Center, USA)
Discussant: Andrew Thomas (Carnegie Mellon University, USA)

Bayesian sparse graphical models and their applications

Veera BALADANDAYUTHAPANI (University of Texas MD Anderson Cancer Center, USA)

We propose Bayesian methods for estimating the precision matrix in Gaussian graphical models. The methods lead to sparse and adaptively shrunk estimators of the precision matrix, and thus conduct model selection and estimation simultaneously. Our methods are based on selection priors leading to parsimonious parameterization of the precision (inverse covariance) matrix, which is essential in several applications involving learning relationships among the variables. We introduce a novel type of selection prior that develops a sparse structure of the precision matrix by making most of the elements exactly zero, in addition to ensuring positive definiteness. Motivated by a novel protein expression data set we extend these methods to classification problems. The novelty of our Bayesian model lies in the ability to draw information from the network data as well as from the associated categorical outcome in a unified hierarchical model for classification. In addition, our method allows for intuitive integration of a priori network information directly in the model and allows for posterior inference on the network topologies both within and between classes.

Coauthors: Bani Mallick, Rajesh Talluri, Yuan Ji and Kevin Coombes

* * *

On a class of shrinkage priors for covariance matrix estimation

Natesh PILLAI (Harvard University, USA)

We propose a flexible class of models based on scale mixture of uniform distributions to construct shrinkage priors for covariance matrix estimation. This new class of priors enjoys a number of advantages over the traditional scale mixture of normal priors, including its simplicity and flexibility in characterizing the prior density. We also exhibit a simple, easy to implement Gibbs sampler for posterior simulation which leads to efficient estimation in high dimensional problems. We first discuss the theory and computational details of this new approach and then extend the basic model to a new class of multivariate conditional autoregressive models for analyzing multivariate areal data. The proposed spatial model flexibly characterizes both the spatial and the outcome correlation structures at an appealing computational cost. Examples consisting of both synthetic and real-world data show the utility of this new framework in terms of robust estimation as well as improved predictive performance.

Coauthors: Hao Wang

* * *

Dynamic models for financial trading networks

Abel RODRIGUEZ (University of California, Santa Cruz, USA)

This talk describes a novel class of dynamic blockmodels for sequences of networks. More specifically, we introduce flexible hidden Markov models for network data that incorporate a generalization of the Dirichlet process prior to jointly exchangeable arrays. We discuss both Markov chain Monte Carlo and sequential Monte Carlo methods for computation. Our goal is to create models that can help us understand the effect of the introduction of an electronic trading platform in energy futures markets.

Coauthors: Brenda Betancourt and Naomi Boyd

* * *

15:00–16:30 Room C **On the uses of random probabilities in Bayesian inference**

Organizer: Antonio Lijoi (University of Pavia, Italy)
Chair: Fernando Quintana (Pontificia Universidad Católica de Chile, Chile)
Discussant: Julyan Arbel (CREST, Université Paris Dauphine, France)

Constructions of priors on probability measures

Peter ORBANZ (Columbia University, USA)

A pivotal problem in Bayesian nonparametrics is the construction of prior distributions on the space M(V) of probability measures on a given domain V. I will first focus on one of the two main approaches, the construction from finite-dimensional distributions, and present a representation theorem applicable to the construction of arbitrary priors on probability measures, including continuous ones. I will then contrast the result to the other main construction approach, stick-breaking representations. Roughly speaking, stick-breaking methods are applicable on a wider variety of spaces, whereas constructions from finite-dimensional distributions can represent a wider range of priors. I will illustrate the differences between the two approaches with examples and discuss in how far the construction of new priors using the representation result can be expected to be feasible in practice.

⁂

An EPPF from independent sequences of geometric random variables

Ramses MENA (UNAM, Mexico)

We consider generating exchangeable partitions probability functions from an independent and identically distributed sample from a geometric distribution. We show that the model is rich and while different from exchangeable random variables based on nonparametric models, such as the Dirichlet process, both are driven by a single parameter, and hence to some extent comparable.

Coauthors: Stephen Walker

⁂

Discrete random probabilities for Bayesian inference on species variety

Antonio LIJOI (University of Pavia, Italy)

Species sampling problems have a long history in ecological and biological studies and a number of issues, including the evaluation of species richness, the design of sampling experiments, the estimation of rare species variety, are to be addressed. Typical applications we will be interested in deal with very large populations containing a huge number of distinct species. Discrete random probabilities, acting as nonparametric priors, perfectly fit this framework and we rely on the flexible class of Gibbs-type priors. Given an observed sample of size n, focus will be on prediction of some key aspects of the outcome from an additional sample of size m. Indeed, we will address issues such as the estimation of: (i) a measure of species richness, (ii) the number of species that will be detected with a certain frequency in the enalrged sample of size n+m; (iii) the discovery probability, i.e. the probability of detecting, at the (n+m+1)-th observation, species that have been observed with any given frequency in the enlarged sample of size n+m; (iv) rare species variety. The results will then be illustrated by an application to real datasets.

Coauthors: Stefano Favaro, Ramses Mena and Igor Pruenster

⁂

17:00–18:30 Room C **Nonparametric Bayes applications in Biostatistics**

Organizer: Luis Nieto-Barajas (ITAM, Mexico)
Chair: Athanasios Kottas (University of California, Santa Cruz, USA)
Discussant: Sujit Ghosh (North Carolina State University, USA)

Species Sampling prior for the analysis of array CGH data

Michele GUINDANI (University of Texas MD Anderson Cancer Center, USA)

Many popular Bayesian Nonparametric priors can be characterized in terms of exchangeable species sampling sequences. However, in some applications, exchangeability may not be appropriate. We introduce non exchangeable generalized species sampling sequences characterized by a tractable predictive probability function with weights driven by a sequence of independent Beta random variables. We compare their clustering properties with those of the Dirichlet Process and the two parameters Poisson-Dirichlet process. We propose the use of such sequences as prior distributions in a hierarchical Bayes modeling framework. We detail on Markov Chain Monte Carlo posterior sampling and discuss the resulting inference in a simulation study, comparing their performance with that of popular Dirichlet Processes mixtures and Hidden Markov Models. Finally, we discuss an application to the detection of chromosomal aberrations in breast cancer using array CGH data.

Coauthors: Thiago Costa, Federico Bassetti, Fabrizio Leisen and Edoardo M. Airoldi

⁂

A nonparametric approach for validating, comparing and integrating predictive models

Lorenzo TRIPPA (Harvard University, USA)

In many scientific areas predictive models are defined and trained, and later on validated, using data from distinct independent studies. Prediction of long versus short survival for specific cancer types, for example, have been approached independently by multiple groups of scientists using similar technologies for measuring gene expressions and, more generally, relevant biomarkers. We consider two objectives in the integration of independent studies: the assessment of study-specific data quality, with the aim of distinguishing which datasets are likely to incorporate non negligible systematic errors, and the assessment of alternative methods for defining predictive models. A hierarchical Bayesian model for the integration of data and predictive models defined by independent institutions or groups of scientists is proposed. The proposed methodology is discussed and the main results from the analysis of data sets generated by recent cancer studies are presented.

⁂

Bayesian analysis of functional proteomics profiles

Luis NIETO-BARAJAS (ITAM, Mexico)

Using a new type of array technology, the reverse phase protein array (RPPA), we measure time-course protein expression for a set of selected markers that are known to co-regulate biological functions in a pathway structure. To accommodate the complex dependent nature of the data, including temporal correlation and pathway dependence for the protein markers, we propose a mixed effects model with temporal and protein-specific components. We develop a sequence of random probability measures (RPM) to account for the dependence in time of the protein expression measurements. We also acknowledge the pathway dependence among proteins via a conditionally autoregressive (CAR) model. Applying our model to the RPPA data, we reveal a pathway-dependent functional profile for the set of proteins as well as marginal expression profiles over time for individual markers.

Coauthors: Peter Mueller, Yuan Ji, Yiling Lu and Gordon Mills

⁂

Thursday, June 28th

09:00–10:30 Room A **Applications of particle filtering and sequential updating**

Organizer: Robert Kohn (University of New South Wales, Australia)
Chair: Jonathan Stroud (George Washington University, USA)
Discussant: Hedibert Lopes (University of Chicago Booth School of Business, USA)

Practicalities in implementing Adaptive Particle MCMC

Julien CORNEBISE (UK)

Particle Markov Chaine Monte Carlo methods are one of the most cutting edge algorithmic developments in Bayesian computational statistics. Being so recent, their implementation is not as familiar as more classical MCMC algorithms: computational bottlenecks can appear in a too-hastily implemented particle filter; MCMC adaptation can be overlooked or failing because “too close to the book”; and while Particle MCMC is a neat combination of existing algorithms, spagetthi code can quickly ensue – or worse, some algorthmic speed-ups can be cast aside to keep the code simple and short, losing on their benefits. We review here three practical tricks to avoid those typical pitfalls: * oft-overlooked linear-complexity resampling algorithm in the particle filter step, well known but not emphasized enough – with dramatic speed loss when forgotten, * online adaptation: waiting more than theoretically needed can pay off, * for the most code-enclined, investing in object-oriented implementation allows to mirror the mathematical structure, making it much easier to combine and reuse fine-tuned algorithmic components.

⁂

Regression density estimation for approximate Bayesian computation

Scott SISSON (University of New South Wales, Australia)

Approximate Bayesian computation (ABC) refers to a collection of Bayesian techniques that allow approximate posterior simulation in the presence of computationally intractable likelihood functions. Typically, these methods require the close matching of summary statistics generated under the model with those of the observed data. One obvious problem with this is that if the summary statistics are high dimensional in order to avoid information loss, the close matching of simulated and observed statistics becomes untenable. As such, in the majority of ABC applications, the dimension of the summary statistics is necessarily low. In this talk we propose a regression density ABC approach, which does not rely on the close matching of summary statistics in high dimensions. In effect, the method “matches” the observed summary statistics exactly thereby increasing accuracy, and scales well with the dimension of the summary statistics which can number in the hundreds. We make a number of practical illustrations.

Coauthors: Yanan Fan and David Nott

⁂

Bayesian Inference for Complex Time Series Models

Robert KOHN (University of New South Wales, Australia)

We develop new particle filter MCMC algorithms to deal with complex structural time series model that are of importance in both Economics and Finance.

Coauthors: Michael Pitt

⁂

09:00–10:30 Room D **Bayesian semi-parametric analysis: theory**

Organizer: Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)
Chair: Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)
Discussant: Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)

Asymptotic theory for empirical Bayes procedures

Harry VAN ZANTEN (Eindhoven University of Technology, Netherlands)

Empirical Bayes procedures are widely used in practice, but at the present time there is little theory available about them. It is not well understood how they perform in terms of contraction rates, for instance compared to an oracle, nor is it known whether credible sets constructed using empirical Bayes have good frequentist coverage probabilities. In this talk we present some recent results regarding these issues.

Coauthors: Botond Szabo and Aad van der Vaart

⁂

Semiparametric Bernstein-von Mises theorem, low regularity

Ismael CASTILLO (CNRS Paris, France)

In this work we propose a framework to derive semiparametric posterior concentration results. Sufficient conditions are given for the semiparametric Bernstein-von Mises theorem to hold and are studied for Gaussian process and truncated series priors. As an application we consider several examples of functionals, including cases with low regularity. This talk is based partly on joint work with Judith Rousseau and Vincent Rivoirard.

Coauthors: Vincent Rivoirard and Judith Rousseau

⁂

Bayesian estimation of the discrepancy with misspecified parametric models

Pierpaolo DE BLASI (University of Torino and Collegio Carlo Alberto, Italy)

We study a Bayesian semiparametric model where we have made specific requests about the parameter values to be estimated. The aim is to find the parameter of a parametric family which minimises a distance to the data generating density and then to find the discrepancy using nonparametric methods. We illustrate how coherent Bayesian updating can proceed given that we are out of the standard Bayes framework. Bayesian updating is performed using MCMC methods and in particular a novel method for dealing with intractable normalizing constants is required. Illustrations using synthetic data are provided.

Coauthors: Stephen Walker

⁂

11:00–12:30 Room B **Time Series analysis and Finance**

Organizer: Siddartha Chib (Washington University, St. Louis, USA)
Chair: Hedibert Lopes (University of Chicago Booth School of Business, USA)
Discussant: Teruo Nakatsuma (Keio University, Japan)

A Bayesian estimator of the multivariate covariance of noisy and asynchronous returns

Antonietta MIRA (University of Lugano, Switzerland)

A multivariate positive definite estimator of the covariance matrix of noisy and asynchronously observed asset returns is proposed. We adopt a Bayesian Dynamic Linear Model which allows us to interprete microstructure noise as measurement errors, and the asynchronous trading as missing observations in an otherwise synchronous series. These missing observations are treated as any other parameter of the problem as typically done in a Bayesian framework. We use an augmented Gibbs algorithm and thus sample the covariance matrix, the observational error variance matrix, the latent process and the missing observations of the noisy process from their full conditional distributions. Convergence issues and robustness of the Gibbs sampler are discussed. A simulation study compares our Bayesian estimator with recently proposed pair-wise QMLE-type and Multivariate Realized Kernel estimators, under different liquidity and microstructure noise conditions. The results suggest that our estimator is superior in terms of RMSE in both a two- and ten-dimensional settings, especially with dispersed and high missing percentages and with high noise. This suggests that our Bayesian estimator is more robust in severe conditions, such as portfolios of assets with heterogeneous liquidity profiles, or particularly illiquid, or when there is a high level of microstructure noise in the market.

Coauthors: Stefano Peluso and Fulvio Corsi

⁂

Multivariate dynamic sparsity modelling in financial time series

Jouchi NAKAJIMA (Duke University, USA)

We describe a general approach to dynamic sparsity modelling in time series and state-space models. Time-varying parameters are linked to latent processes that are thresholded to induce zero values adaptively, providing dynamic variable inclusion/selection. We discuss Bayesian model estimation and prediction in dynamic regressions, time-varying vector autoregressions, dynamic factor models and multivariate volatility models using latent thresholding. Substantive examples in dynamic factor modeling of financial FX and commodity time series show that interpretable, data-driven dynamic sparsity can reduce estimation uncertainties, improve predictions and Bayesian decisions about financial portfolio revision leading to improved returns.

Coauthors: Mike West and Jouchi Nakajima

⁂

A class of multivariate stochastic volatility models with leverage

Yasuhiro OMORI (University of Tokyo, Japan)

We discuss three multivariate stochastic volatility (SV) models with cross leverage for multivariate financial asset returns. First, we introduce a natural extension of the univariate SV models where we consider (constant) cross leverage. The factor SV models are also discussed to describe common dynamic factors among asset returns. Finally, the matrix exponential multivariate SV model is proposed to allow general dynamic correlations. An efficient MCMC algorithm is proposed based on the multi-move sampler.

⁂

11:00–12:30 Room D **Problem-driven developments in Bayesian nonparametrics**

Organizer: Sonia Petrone (Bocconi University, Italy)
Chair: Ramses Mena (UNAM, Mexico)
Discussant: Raffaele Argiento (CNR-IMATI, Italy)

A Bayesian nonparametric model for ranking data

Yee Whye TEH (University College London, UK)

Data in the form of partial rankings (for example in terms of an ordered list of the top-N items) arise in many contexts, e.g. skill modelling, choice modelling, marketing, and voting. In many situations the pool of available items can be very large, and a nonparametric approach, where the pool is assumed to be infinite, can be sensible. In this talk we propose such a Bayesian nonparametric model, where each item is assigned a positive "desirability" parameter and the model is a nonparametric generalization of the popular Plackett-Luce choice model. We use completely random measures as priors for the desirability parameters, and develop a simple posterior characterization in terms of auxiliary variables. We discuss Markov chain Monte Carlo algorithms for inference, as well as generalizations of our model to temporal and mixture models.

Coauthors: Francois Caron and Brendan Murphy

❉

A Bayesian nonparametric model for reversible Markov chains with applications in molecular dynamics

Sergio BACALLADO (Stanford University, USA)

Reversible Markov chains are a simple stochastic model for dynamical systems whose equations of motions are invariant to a sign change in time. In the field of molecular dynamics simulation, they have been applied with much success to model conformational transitions in proteins and nucleic acids.

We introduce a conjugate prior for reversible Markov chains on infinite spaces which allows us to address species sampling problems in the application and makes it possible to enforce reversibility in Bayesian nonparametric Hidden Markov Models. The prior is defined via a random walk with reinforcement with a notion of partial exchangeability, which is its predictive distribution. This process generalizes the linearly edge-reinforced random walk proposed as a Bayesian predictive scheme by Diaconis and Rolles [1]. Our process also generalizes two well-known exchangeable urn schemes whose de Finetti measures are the Dirichlet process and the Pitman-Yor process. The prior is tuned by three parameters, which we demonstrate provide flexibility needed in applications and facilitate sharing statistical strength between transitions from different states.

References [1] P. Diaconis and S. Rolles. Bayesian analysis for reversible Markov chains. Ann. Stat., 2006.

Coauthors: Lorenzo Trippa and Stefano Favaro

❉

Bayes and empirical Bayes: do they merge?

Sonia PETRONE (Bocconi University, Italy)

It is a common experience that being a honest Bayesian and eliciting a honest prior is often difficult, and this is possibly more true in the case of complex Bayesian nonparametric models. Indeed, in many applications one is tempted to take an "empirical Bayes" approach, plugging some empirical estimates of the prior hyperparameters into the posterior distribution. Even if awkward for a Bayesian, the underlying idea is that, if the sample size is large, the Bayesian and empirical Bayes posterior distributions will be close. This asymptotic equivalence seems to be commonly accepted; in fact, precise mathematical results appear to be missing. In this talk we try to give a more rigorous discussion, formalizing the problem in terms of merging of Bayes and empirical Bayes procedures. We first discuss Bayesian weak merging, and we show that it holds if and only if the empirical Bayes posterior is consistent in frequentist sense. Then we give general conditions for consistency of empirical Bayes. Examples with Dirichlet process mixtures models are provided. Finally, we give further results in terms of frequentist strong merging.

Coauthors: Judith Rousseau and Catia Scricciolo

❉

15:00–16:30 Room C **Adaptive Bayesian function estimation**

Organizer: Subhashis Ghosal (North Carolina State University, USA)
Chair: Subhashis Ghosal (North Carolina State University, USA)
Discussant: Subhashis Ghosal (North Carolina State University, USA)

Projection oracle convergence rate of posterior

Eduard BELITSER (Eindhoven University of Technology, Netherlands)

We apply the Bayes approach to the problem of projection estimation of a signal observed in the Gaussian white noise model and we study the rate at which the posterior distribution concentrates about the true signal, as the information in observations tends to infinity. A benchmark is the rate of a so called oracle projection risk, i.e., the smallest risk of an unknown true signal over all projection estimators. Under an appropriate hierarchical prior, we study the performance of the resulting (appropriately adjusted by the empirical Bayes approach) posterior distribution and establish that the posterior concentrates about the true signal with the oracle projection convergence rate. We also study implications for the model selection problem, namely we propose a Bayes model selector and assess its quality in terms of the so called false selection probability.

⁂

Dimension adaptability of Gaussian process models with variable selection and projection

Surya TOKDAR (Duke University, USA)

It is now known that an extended Gaussian process model equipped with rescaling can adapt to different smoothness levels of a function valued parameter in many nonparametric Bayesian analyses, offering a posterior convergence rate that is optimal (up to logarithmic factors) for the smoothness class the true function belongs to. This optimal rate also depends on the dimension of the function's domain and one could potentially obtain a faster rate of convergence by casting the analysis in a lower dimensional subspace that does not amount to any loss of information about the true function. In general such a subspace is not known a priori but can be explored by equipping the model with variable selection or linear projection. We demonstrate that for nonparametric regression, classification, density estimation and density regression, a rescaled Gaussian process model equipped with variable selection or linear projection offers a posterior convergence rate that is optimal (up to logarithmic factors) for the lowest dimension in which the analysis could be cast without any loss of information about the true function. Theoretical exploration of such dimension reduction features appears novel for Bayesian nonparametric models with or without Gaussian processes.

⁂

Posterior convergence rates for wavelet regression

Heng LIAN (Nanyang Technological University, Singapore)

We investigate the posterior rate of convergence for wavelet shrinkage in Besov spaces. Many previous works have focused on various Bayesian estimators such as posterior mean or posterior median. Instead, we study the posterior contraction rates from a nonparametric Bayesian asymptotics point of view. The results obtained are thus complementary to those works. Unlike the theoretical investigation of Bayesian estimator, adapting to unknown smoothness of the function is relatively easy in our case.

⁂

17:00–18:30 Room A **Spatial state-space models**

Organizer: Dani Gamerman (Federal Univeristy of Rio de Janeiro, Brazil)
Chair: Gabriel Huerta (Indiana University, USA)
Discussant: Sudipto Banerjee (University of Minnesota, USA)

Blending and downscaling ensembles of climate model predictions

Bruno SANSO (University of California, Santa Cruz, USA)

We consider a general framework for the analysis of ensembles of simulations produced by climate models. The methods are based on hierarchical Bayesian models with space and time components. Our models provide summaries of the simulated variables that allow for sensible comparisons between model simulations and historical records, and quantify possible discrepancies. Such discrepancies are used as the basis for the averaging of the simulation ensembles. We are particularly interested in obtaining blended results for time varying factors that can summarize the variability of large spatio-temporal fields. Then use process convolutions and high resolution data to downscale those results in order to obtain predictions at small spatial scales. We consider predictions of oceanic indexes in the North Pacific and predictions of sea surface temperature along the California coast, at scales that are meaningful for coastal ecosystems.

⁂

Diffusion modeling of motion trajectories under the influence of covariates

Ioanna MANOLOPOULOU (Duke University, USA)

We present dynamic spatial modeling and computational methods for the analysis of collections of objects moving in a spatially inhomogeneous force field under the influence of covariates. Core motivating examples come from movement ecology, where multiple animals are tracked moving in 2-D or 3-D largely driven by the external environmental characteristics. Interest lies in identifying the role of different covariates in guiding the motion, both in terms of the shape of their implied field, as well as their overall presence or absence of influence. Models are based on discrete-time, dynamic state-space models for locations and directional velocities of each of a set of animals, combined with a latent force-field over the temporal domain that drives changes in velocities. We extend models for the force fields using dynamic Bayesian radial basis function regression to define a potential surface varying in space but also in the space of covariates, with the force field given by the gradient of the potential in 3-D. Corresponding variable selection priors allow us to detect which covariates play a role in shaping the motion, and provide a basis for understanding their precise functional form. We exemplify the work with analysis of GPS tracking data from a set of toucans in central America, where primary interest lies in characterizing the "birds'" response and contribution to different temperature levels.

Coauthors: Mike West

⁂

Change point detection in multivariate time series and space time data sets

Christopher STRICKLAND (Queensland University of Technology, Australia)

We present a Bayesian approach that can be used for the detection of change in multivariate time series and space time data sets. In particular, efficient algorithms, which are linear in computational complexity, are proposed that enable the implementation of fast Markov chain Monte Carlo sampling schemes. Specifically, the algorithms are developed for the analysis of dynamic mixture models that have been cast into a conditionally Gaussian state space framework. The proposed approach is in an analysis of multitemporal remotely sensed image data.

Coauthors: Robert Burdett, Robert Kohn, Robert Denham and Kerrie Mengersen

⁂

Friday, June 29th

09:00–10:30 Room A **Savage Award Session**

Organizer: ISBA
Chair: Bertrand Clarke (University of Miami, USA)

Invariant procedures in model checking, checking for prior-data conflict and Bayesian analysis

Gun Ho JANG (University of Pennsylvania, USA)

We consider a statistical theory as being invariant when the results of two statisticians' independent data analyses, based upon the same statistical theory and using effectively the same statistical ingredients, are the same. We discuss three aspects of invariant statistical theories.

Both model checking and checking for prior-data conflict are assessments of single null hypothesis without any specific alternative hypothesis. Hence, we conduct these assessments using a measure of surprise based on a discrepancy statistic. For the discrete case, it is natural to use the probability of obtaining a data point that is less probable than the observed data. For the continuous case, the natural analog of this is not invariant under equivalent choices of discrepancies. A new method is developed to obtain an invariant assessment. This approach also allows several discrepancies to be combined into one discrepancy via a single P-value.

Second, Bayesians developed many noninformative priors that are supposed to contain no information concerning the true parameter value. Any of these are data dependent or improper which can lead to a variety of difficulties. Gelman (2006) introduced the notion of the weak informativity as a compromise between informative and noninformative priors without a precise definition. We give a precise definition of weak informativity using a measure of prior-data conflict that assesses whether or not a prior places its mass around the parameter values having relatively high likelihood. In particular, we say a prior Π_2 is weakly informative relative to another prior Π_1 whenever Π_2 leads to fewer prior-data conflicts a priori than Π_1. This leads to a precise quantitative measure of how much less informative a weakly informative prior is.

In Bayesian data analysis, highest posterior density inference is a commonly used method. This approach is not invariant to the choice of dominating measure or reparametrizations. We explore properties of relative surprise inferences suggested by Evans (1997). Relative surprise inferences which compare the belief changes from a priori to a posteriori are invariant under reparametrizations. We mainly focus on the connection of relative surprise inferences to classical Bayesian decision theory as well as important optimalities.

❉

Improving cosmological distances to illuminate dark energy: hierarchical Bayesian models for type Ia supernovae in the optical and near-infrared

Kaisey MANDEL (Imperial College London, UK)

Type Ia supernovae, brilliant stellar explosions, are observed in distant galaxies in the Universe. These supernovae can be used as "standard candles": if their luminosity is known, their distances can be estimated from their observed brightness. Observing optical light from faraway supernovae to determine their distances, astronomers found that the expansion of the Universe is accelerating, rather than slowing down under the attractive force of gravity, an astounding discovery that won the 2011 Nobel Prize in Physics. The cosmic acceleration is thought to be caused by a repulsive "dark energy", whose physical nature is the most perplexing mystery in modern cosmology. To shed light on the nature of dark energy, I seek to improve statistical inference of the history of cosmic expansion using supernovae. The major source of systematic error confounding current distance estimates is interstellar dust in the galaxies of the supernovae. Dust dims a supernova's optical light, making it appear farther away. However, in near-infrared (NIR) light, the dust is nearly transparent, and supernovae are even better standard candles. The combination of optical and NIR time series and spectroscopic data mitigates the pernicious effects of dust and can improve inferences in supernova cosmology. To optimize distance estimates, I constructed a principled, hierarchical Bayesian framework, described by a directed acyclic graph, to coherently model the multiple random and uncertain effects underlying supernova time series observations, including measurement error, dust, intrinsic supernova covariances across time and wavelength, galaxy motions and distances. I developed an MCMC code, BayeSN, using a Gibbs sampling structure to efficiently compute probabilistic inferences for the parameters of individual supernovae and the hyperparameters of their population, while dealing with incomplete data. Applying this to nearby supernova data, I demonstrate that the combination of optical and NIR data almost doubles the precision of cross-validated distance predictions, and is a more powerful method to measure the properties of dark energy.

❉

On stationary modelling for time-series data

Juan Carlos MARTÍNEZ-OVANDO (Banco de México, Mexico)

Stationary models provide us with a solid framework for analysing and predicting data observed sequentially in time. In this talk we shall explore a well-known intuitive idea to model construction in order to develop flexible stationary models capable of capturing particular features observed in real data, while preserving stationarity. Different model specifications will be explored within the Bayesian framework, considering: i) doubly nonparametric first-order stationary models, ii) stationary models admitting exogenous interventions in the transition dynamics while having a given (invariant) marginal distribution, and iii) marginalization coherent stationary models with arbitrary order of dependence. Illustrations of these models will be discussed, along with related research topics.

Coauthors: Stephen G. Walker

❉

Spike-and-slab priors for function selection in structured additive regression models

Fabian SCHEIPL (Ludwig-Maximilians-Universität München, Germany)

Structured additive regression provides a general framework for complex Gaussian and non-Gaussian regression models, with predictors comprising arbitrary combinations of nonlinear functions and surfaces, spatial effects, varying coefficients, random effects and further regression terms. The large flexibility of structured additive regression makes function selection a challenging and important task, aiming at (1) selecting the relevant covariates, (2) choosing an appropriate and parsimonious representation of the impact of covariates on the predictor and (3) determining the required interactions. We present a novel spike-and-slab prior structure for function selection that allows to include or exclude single coefficients as well as blocks of coefficients representing specific model terms. An innovative multiplicative parameter expansion is required to obtain good mixing and convergence properties in a Markov chain Monte Carlo simulation approach and is shown to induce desirable shrinkage properties. We investigate sensitivity to hyperparameter settings and compare performance to competitors in simulation studies and on real benchmark classification data. We also present results from geoadditive and additive mixed logit model applications.

Coauthors: Ludwig Fahrmeir and Thomas Kneib

❉

09:00–10:30 Room B **Topics in Bayesian Statistics**

Organizer: Raquel Prado (University of California, Santa Cruz, USA)
Chair: Raquel Prado (University of California, Santa Cruz, USA)
Discussant: Steven MacEachern (Ohio State University, USA)

Bayesian model selection in ultrahigh-dimensional settings

Valen E. JOHNSON (University of Texas MD Anderson Cancer Center, USA)

Standard assumptions incorporated into Bayesian model selection procedures result in model selection procedures that are not competitive with commonly used penalized likelihood methods. We propose modifications of these methods by imposing non-local prior densities on model parameters. We show that the resulting model selection procedures are consistent in linear model settings when the number of possible covariates p is bounded by the number of observations n, a property that has not been extended to other model selection procedures. In addition to consistently identifying the true model, the proposed procedures provide accurate estimates of the posterior probability that each identified model is correct. Through simulation studies, we demonstrate that these model selection procedures perform as well or better than commonly used penalized likelihood methods in a range of simulation settings, including ultrahigh-dimensional setting for which p » n. In particular, we show that model selection procedures based on non-local prior densities are better able to identify the correct model and provide equivalent or smaller estimation errors than SIS and ISIS-SCAD procedures.

Coauthors: Valen E. Johnson

❉

Proper Bayesian inference with missingness and auxiliary information

Michael DANIELS (University of Florida, USA)

In order to make a missing at random (MAR) or ignorability assumption realistic, auxiliary covariates are often required. However, the auxiliary covariates are usually not desired in the model for inference. Typical multiple imputation approaches do not assume that the imputation model marginalizes to the inference model. This has been termed ’uncongenial’ by Meng (Meng, 1994) and also as ’Bayesianly improper imputation’ (Schafer, 1997). The lack of marginalization is unsatisfactory on many levels. In order to make the two models congenial (or compatible), we do not want to assume a parametric model for the marginal distribution of the auxiliary covariates, but we typically do not have enough data to estimate the joint distribution well non-parametrically. In addition, when the imputation model uses a non-linear link function (e.g., the logistic link for a binary response), the marginalization over the auxiliary covariates to derive the inference model typically results in a difficult to interpret form for effect of covariates of interest in the inference model. In this talk, we propose an approach to ensure that the models are congenial and that also addresses the two complications described above. We evaluate the approach via simulations and illustrate it on a recent clinical trial.

Coauthors: Chenguang Wang

⁂

A new tool for Bayesian model diagnostics

Hans R. KUENSCH (ETH Zurich, Switzerland)

We introduce a new tool for assessing the adequacy of a Bayesian temporal or spatio-temporal model which we call the primary posterior predictive distribution (PPPD). Often, the fit of a model is explored by simulating from the posterior predictive distribution (PPD) and comparing characteristics of such simulated values with characteristics of observations. However, if the model contains, in addition to the primary parameters which do not change in time or space, also a latent process, the PPD often fails to dectect deficiencies in the spatio-temporal dependence structure. This phenomenon occurs when the latent process is informative for the data, because then the observations will impose the correct dependence on the latent process in the posterior. Hence observations simulated from the posterior will always have a dependence structure that is similar to the one in the actual data, even if the dependence in the model is inadequate. However, as soon as we use the model for extrapolations where no observations are available, the model deficiency will have drastic consequences. In order to circumvent this problem, we propose the PPPD which simulates from the model where the actual data provide information only on the primary parameters, but not on the latent process. The main difference between the PPD and the PPPD is thus whether we integrate over the prior or the posterior of the latent process. We illustrate the problem and our solution on a model for short term prediction of precipitation.

Coauthors: Fabio Sigrist and Werner A. Stahel

⁂

11:00–12:30 Room B **S’Bayes: constructing and using subjective priors for Bayesian modelling**

Organizer: Kerrie Mengersen (Queensland University of Technology, Australia)
Chair: Kerrie Mengersen (Queensland University of Technology, Australia)
Discussant: Kerrie Mengersen (Queensland University of Technology, Australia)

Subjective belief specification for complex physical models

Michael GOLDSTEIN (University of Durham, UK)

The familiar problems involved in specifying expert beliefs for Bayesian analysis become progressively more challenging as the phenomena in question increase in complexity. When we deal with complex physical systems modelled by computer simulators, such as climate, then there are so many sources of uncertainty (function uncertainty, input and condition uncertainty, structural discrepancy, observational uncertainty, multi-model uncertainty and so forth) that we need to develop global formulations which link all of the uncertainties within a coherent structure and novel elicitation tools to facilitate structured expert elicitation for the various elements of the uncertainty specification. In this talk we shall discuss the ingredients of this global structuring and give examples of the types of elicitation tools that we have found to be helpful and appropriate for expert elicitation for such problems.

Coauthors: Daniel Williamson

⁂

Strategies for aggregating multiple sources of subjective prior information in Bayesian analysis

Samantha LOW-CHOY (Queensland University of Technology, Australia)

Multiple sources of prior information may be incorporated either sequentially, via Bayesian updating, or simultaneously through aggregation into a single prior. The most common strategies fall into the latter category: averaging and pooling. Averaging highlights the consensus, and deviations from this. Linear pooling (an additive mixture with fixed weights), pools diversity across experts with individual uncertainty. Recent work has explored Bayesian hierarchical models for aggregating expert assessments, to reflect both the consensus and diversity across experts, as well as different sources of individual uncertainty. It builds on gradual emergence of Bayesian approaches for encoding subjective priors, from the perspective of a decision maker who updates their prior using elicited data.

This recent work applied a Bayesian hierarchical model to indirectly encode and aggregate parameters (such as the mean and variance), inferred from summary statistics about the observable response (such as cumulative probabilities and quantiles). Here we consider this approach for the more extreme case of indirect elicitation that occurs when eliciting summary statistics about means conditional on covariates in generalized linear models (Conditional Mean Priors).

We have identified five stages at which aggregation of expert inputs can occur: (A) on raw elicited summary statistics of observables, (B) on inferred distributions for these observables, (C) on encoded priors for regression coefficients, (D) sequential Bayesian updating, and (E) model-averaged posteriors. We present a case study that clearly demonstrates different outcomes depending on stage of aggregation. This leads to insights relevant to designing elicitation from multiple experts, including feedback to individuals or groups.

Coauthors: Kerrie Mengersen, Justine Murray and Judith Rousseau

❊

Revealing latent clusters from Dirichlet process mixtures models using product partitions

Jeff GILL (Washington University, St. Louis, USA)

In this work we adapt our Generalized Linear Mixed Dirichlet Process Model algorithms to provide substantive clusters of interest through posterior probabilities using covariate information. This is an algorithmic approach that directly includes variable selection within clusters into a general clustering model. The value lies in establishing a new paradigm for using Bayesian nonparametric approaches to produce clustering based on posterior probabilities. This approach promises to improve current designs since it inherently includes fit and overfit criteria, which is not possible without new Bayesian stochastic simulation tools.

Coauthors: George Casella

❊

11:00–12:30 Room C **Bayesian analysis of inverse problems**

Organizer: Colin Fox (University of Otago, New Zealand)
Chair: Geoff Nicholls (University of Oxford, UK)
Discussant: J. Andrés Christen (CIMAT, Mexico)

State and parameter estimation of large models

Heikki HAARIO (Lappeenranta University of Technology, Finland)

The core of a Bayesian solution to inverse problems might be defined as the construction of the posterior distribution of the unknown. With state of art MCMC methods, this can now routinely be done for large classes of models, such as various ODE or PDE systems. Models with high CPU demand or high dimension still provide a challenge. We discuss options of identifying parameters of large scale climate models. While MCMC appears doable, the construction of the likelihood in case of a chaotic model needs special attention. The use of filtering methods provides an approach here. We present various optimization-based approximative filtering algorithms, and how to use them both for state and parameter estimation.

Coauthors: Antti Solonen andAlexander Bibov

❊

Bayesian inverse problems via a Bernoulli factory

Radu HERBEI (Ohio State University, USA)

The posterior distribution resulting from a standard approach to Bayesian inverse problems is typically explored via Markov chain Monte Carlo methods. In this work I use a perfect sampling algorithm, following the work of Flegal and Herbei (2012). They show how it is possible to draw *exact* Bernoulli random variates with a success probability which cannot be calculated exactly, leading to a perfect sampling algorithm for general distributions. I analyze inverse problems for which the likelihood evaluation requires solving a partial differential equation. In this case I show how it is possible to reduce the computational burden by only evaluating the likelihood at accepted states. In addition, I provide a drift and minorization condition for the MH sampler. Using the perfect sampling algorithm described above, this leads to exact draws from the posterior distribution. This methodology is illustrated with examples from oceanography.

⁂

Polynomial accelerated Gibbs sampling for conductivity imaging, and other inverse problems

Colin FOX (University of Otago, New Zealand)

Gibbs samplers have many desirable theoretical properties, but also have the pesky requirement that conditional distributions be available. We give a new analysis that shows how conditional densities can be evaluated for the posterior distribution in conductivity imaging — for free in coordinate directions and very cheaply in other “special” directions. The analysis actually applies to a broad class of non-invasive imaging techniques that utilize strong scattering of energy. The resulting Gibbs sampler draws an independent conductivity image in only a little more compute time than required for optimization. One of the nice properties of Gibbs samplers is that they are essentially identical to stationary linear solvers, and so are amenable to polynomial acceleration. We demonstrate acceleration in the convergence of first and second moments of the basic Gibbs sampler for conductivity imaging.

Coauthors: Al Parker and Markus Neumayer

⁂

Special topic sessions

Tuesday, June 26th

09:00–10:30 Room C **Bayesian Econometrics IV**

Organizer: Cathy W. S. Chen (Feng Chia University, Taiwan)
Chair: Cathy W. S. Chen (Feng Chia University, Taiwan)

Bayesian semi-parametric forecasting of expected shortfall

Richard GERLACH (University of Sydney, Australia)

Bayesian semi-parametric estimation has proven effective for quantile estimation in general and specifically in financial Value at Risk forecasting. Expected shortfall is a competing tail risk measure, involving a conditional expectation beyond a quantile, that has recently been semi-parametrically estimated via asymmetric least squares and so-called expectiles. An asymmetric Gaussian density is proposed allowing a likelihood to be developed that leads to Bayesian semi-parametric estimation and forecasts of expectiles and expected shortfall. Further, the conditional autoregressive expectile class of model is generalised to two fully nonlinear families. Adaptive Markov chain Monte Carlo sampling schemes are employed for estimation in these families. The proposed models are clearly favoured in an empirical study forecasting eleven financial return series: clear evidence of more accurate expected shortfall forecasting, compared to a range of competing methods is found. Further, the most favoured models are those estimated by Bayesian methods.

Coauthors: Cathy Chen and Liou-Yan Lin

⁂

Bayesian risk assessment with threshold mixture extreme value models

Teruo NAKATSUMA (Keio University, Japan)

This study proposes a new Markov chain Monte Carlo (MCMC) method to estimate a threshold mixture model of financial losses based on the generalized Pareto distribution (GPD). In our model setting, small losses are supposed to follow a normal or gamma distribution, which we refer to as the base distribution, while large ones that exceed a threshold are supposed to follow the GPD. This model is designed to take into account a stylized fact that the loss distribution tends to be positively skewed and has a heavy right tail. It is also consistent with the extreme value theory on exceedances over a high threshold. Bayesian MCMC approach enables us to simultaneously estimate the threshold along with parameters in the base distribution and GPD. Our new method explicitly takes care of discontinuity in the posterior density of the threshold to obtain a workable proposal distribution for the Metropolis-Hastings algorithm. As a result, it takes less time in computation and can achieve a good mixing property of the sample path. As an application, we apply it to evaluation of risk measures for operational loss data and compare threshold mixture models by a Bayesian model selection procedure.

⁂

Efficient filtering of stochastic volatility models

Catherine FORBES (Monash University, Australia)

Understanding and predicting volatility is central to the management of financial risk, with stochastic volatility (SV) models considered a useful way to conceptualise and forecast risk associated with asset returns. Estimation of such models inevitably involves high dimensional simulation procedures associated with the joint specification of the latent random vector. Although there are many ways to undertake this simulation, particle filtering methods, in conjunction with other simulation based algorithms, have become increasingly popular. However, due to the large sample sizes in many empirical problems, the filtered probability samples resulting from existing particle filter methods at some point must confront a degeneracy problem that may compromise the quality of the inference. In this paper we consider Bayesian estimation of SV models using an alternative particle filtering method that completely avoids this degeneracy problem. The new filter, based on the transformation from the measurement error variable to the corresponding observation, enables direct sampling of the latent vector conditional upon the fixed parameters and observed data. The method is illustrated using with the discrete time stochastic volatility model, both with and without leverage effects.

Coauthors: Gael Martin, Brendan McCabe and Jason Ng

⁂

A Bayesian perspective on backtesting value-at-risk models

Cathy W. S. CHEN (Feng Chia University, Taiwan)

Many papers have proposed forecasting VaR (Value-at-Risk) measures in a Bayesian framework. However, when the authors deal with backtesting, most works jump outside the Bayeisan framework or use informal criteria to compare VaR models. An important issue that arises in this context is how to evaluate the performance of VaR models/methods. It is desirable to have formal testing procedures for comparison, which do not necessarily require knowledge of the underlying model, or if the model is known, do not restrict attention to a specific estimation procedure. The motivation of the study is to propose backtesting based on the idea of Gaglianone et al. (2011 JBES) which evaluates VaR models via quantile regression. It does not rely solely on binary variables. For the empirical study, we compare the performance of the proposed methods and other three non-Bayesian backtesting methods to figure out which backtests are the most reliable, and which tests are suitable for forthcoming model validation processes. The proposed Bayesian test provides a sound result to assess the finite sample performance of a quantile model.

Coauthors: Edward Lin

⁂

09:00–10:30 Room D **Bayesian empirical likelihood**

Organizer: Sanjay Chaudhuri (National University of Singapore, Singapore)
Chair: Sanjay Chaudhuri (National University of Singapore, Singapore)

Empirical likelihood weighting

Nicole LAZAR (University of Georgia, USA)

An important by-product of the empirical likelihood procedure are weights that are assigned to the individual data points. Although these weights are not often directly studied or utilized, they are nonetheless of some intrinsic interest. For example, they can provide an alternative to standard bootstrap and multiple imputation procedures. In this talk, I will discuss empirical likelihood weights and some of their uses, with an emphasis on connections to the bootstrap. A thorough exploration of the empirical likelihood weights also leads to some new thoughts on Bayesian empirical likelihood.

⁂

Bayesian empirical likelihood

Sanib BASU (Northern Illinois University, USA)

Empirical likelihood is a semi-parametric approach which requires minimal distributional assumptions. We consider the use of empirical likelihood as an alternative to traditional parametric likelihoods in Bayesian analysis and argue that empirical likelihood may provide several advantages. In many applications it provides comparable or better performance than parametric models. Furthermore, restrictions to certain classes of estimators, often assumed for convenience, can also be dispensed with. We discuss the formulation and computational methodologies for the proposed estimators. Performances of the estimators are illustrated through examples.

Coauthors: Sanjay Chaudhuri

⁂

Bayesian empirical likelihood for longitudinal data with small area application

Malay GHOSH (University of Florida, USA)

The paper will discuss some Bayesian empirical likelihood techniques in the context of small area estimation based on longitudinal data.

Coauthors: Sanjay Chaudhuri

⁂

Empirical likelihood based Bayesian methodology for complex survey datasets

Sanjay CHAUDHURI (National University of Singapore, Singapore)

In many applications, it is necessary that the data is collected through a complex design. In such datasets the distribution in the sample differs drastically from the distribution in the population. Moreover, the observations may not be independent. Modelling and analysing such datasets has been major interest in statistics. In frequentist paradigm, parameter estimates are usually obtained by solving Horvitz-Thompson estimate of the population sum of estimating equations obtained from the assumed model. These estimators are not likelihood based and thus cannot be directly used in Bayesian paradigm. In this presentation we discuss an alternative empirical likelihood based Bayesian approach to model and analyse such complex datasets. Our method is based on a *sample likelihood*, which together with empirical likelihood provides an easy way to include prior information in such analysis. Furthermore, because of the use of empirical likelihood, the procedure is semiparametric and enjoys many advantages over the traditional parametric likelihood. We, shall illustrate our methodology with real data examples.

⁂

11:00–12:30 Room B **Approximate Bayesian computation (ABC): likelihood-free Bayesian inference I**

Organizers: Christian Robert (Université Paris Dauphine, France)
Scott Sisson (University of New South Wales, Australia)
Chair: Scott Sisson (University of New South Wales, Australia)

Considerate approaches to ABC model selection

Michael STUMPF (Imperial College London, UK)

For nearly any challenging scientific problem evaluation of the likelihood is problematic if not impos- sible. Approximate Bayesian computation (ABC) allows us to employ the whole Bayesian formalism to problems where we can use simulations from a model, but cannot evaluate the likelihood directly. When summary statistics of real and simulated data are compared rather than the data directly infor- mation is lost, unless the summary statistics are sufficient. Here we employ an information-theoretical framework that can be used to construct (approximately) sufficient statistics by combining different statistics until the loss of information is minimized. Such sufficient sets of statistics are constructed for both parameter estimation and model selection problems. We apply our approach to a range of illustrative and real-world model selection problems. We also comment on the appropriate choice of summary statistics for model checking.

Coauthors: Christopher Barnes, Sarah Filippi, Thomas Thorne and Carsten Wiuf

⁂

Summary statistics for ABC model choice

Dennis PRANGLE (Lancaster University, UK)

Recently there has been much interest in using approximate Bayesian computation (ABC) methods for model choice. However, concerns have been raised that the results are often not robust to summary statistic choice. We propose a method of choosing useful summary statistics, extending the idea of semi-automatic ABC (Fearnhead and Prangle 2012). The method is illustrated by a substantive application in population genetics; detecting demographic change in a bacterial population.

Coauthors: Paul Fearnhead

⁂

Assessing the influence of the choice of summary statistics for ABC: application to model choice

Judith ROUSSEAU (ENSAE-CREST and Université Paris Dauphine, France)

Approximate Bayesian Computation relies on the choice of a reasonable set of summary statistics. This choice is very influencial on the outcome. We examine necessary and sufficient conditions on those statistics for a corresponding Bayes factor to be convergent, in a generic testing problem. The conditions thus obtained are then usable in ABC settings to determine which summary statistics are appropriate, following a standard Monte Carlo validation.

Coauthors: Jean-Michel Marin, Natesh Pilai and Christian Robert

⁂

Approximate Bayesian computation using indirect inference

Christopher DROVANDI (Queensland University of Technology, Australia)

For an increasing number of high-impact research problems from the biological, ecological and medical sciences, the primary models of interest are often sufficiently complex that the likelihood function is computationally intractable. Rather than discard these models in favour of inferior alternatives, a class of Bayesian “likelihood-free” techniques (often termed approximate Bayesian computation (ABC)) has emerged in the last few years, which avoids direct likelihood computation through repeated sampling of data from the model and comparing summary statistics of the simulated data with the summary statistics of the observed data. Determining application-specific summary statistics is a challenge and a current area of research. In this presentation I will discuss a technique for obtaining summary statistics in the presence of covariates. This approach, based on indirect inference, uses the estimates of the parameters of an auxiliary model with a tractable likelihood as the observed summary statistics. The ABC inference is performed via an efficient sequential Monte Carlo algorithm that requires very little user intervention. The motivating example for this work is a problem in macroparasite population evolution inside a host modelled using a trivariate Markov process. The objective is to determine which parameters are estimable from data, which are the numbers of adult parasites observed in hosts at autopsy. A good auxiliary model in this application is a beta-binomial model, and I will discuss sensitivities to the auxiliary model specification.

Coauthors: Anthony Pettitt andMalcolm Faddy

❇

11:00–12:30 Room C **Optimal Bayesian experimental design**

Organizers: Thomas J. Loredo (Cornell University, USA)
Youssef M. Marzouk (Massachusetts Institute of Technology, USA)
Chair: Youssef M. Marzouk (Massachusetts Institute of Technology, USA)

Optimal experimental design for model discrimination in cognitive science: a nonparametric extension

Woojae KIM (Ohio State University, USA)

Discriminating among competing theoretical explanations is a vital issue in the field of cognitive science. While it has been adequately demonstrated that cognitive modeling may benefit from the methods of statistical model comparison, the issue of finding an optimal experimental design for effective model discrimination is a relatively new topic in the field (Myung & Pitt, 2009). Quite typical in the process of experimentation and modeling of cognitive tasks is a complex decision-making task in which it is desirable to find not only optimal levels of an independent variable in an experiment but also an optimal number of such levels as well as suitable sample sizes in each condition. In the present study, a procedure is developed for optimizing all such design variables jointly through sequential experiments in an adaptive fashion. To that end, the Dirichlet process (Ferguson, 1973) is adopted as a prior on the design space and integrated into the sampling-based optimization algorithm for nonlinear models (Mueller, 1999; Amzal et al., 2006). The method’s effectiveness is demonstrated through a few modeling examples in cognitive science.

Coauthors: Jay Myung, Mark Pitt and Mark Steyvers

❇

Optimal sequential Bayesian experimental design via approximate dynamic programming

Xun HUAN (Massachusetts Institute of Technology, USA)

Experimental data play a crucial role in the development and refinement of models. Since experiments are often resource intensive, they must be chosen intelligently in order to maximize their value.

We will consider the optimal design of multiple experiments in systems with nonlinear parameter dependence. Open-loop (i.e., batch) design involves choosing all the experiments simultaneously. This may be necessary if, for example, experiments need to be performed in parallel. Closed-loop or sequential design, on the other hand, involves a sequence of decisions wherein intermediate experimental results can guide the design of subsequent experiments. The expected impact of all future experiments must be quantified in order to achieve optimality. This more general closed-loop problem is the focus of our study.

We will present a rigorous formulation and computational methodology for optimal sequential Bayesian experimental design. The closed-loop design process is formulated as a discrete-time finite-horizon dynamic programming problem with perfect state information. A natural choice of state is the parameter distribution, with the system dynamics given by Bayes' rule. The control space is the (continuous) experimental design space, the disturbance is the experimental measurement from a given likelihood function, and the reward function is the information gain from prior to posterior. Direct solution of this problem is computationally impractical. By using particles and functions to represent the state, polynomial chaos surrogates to successively approximate the cost-to-go functions, and gradient-based stochastic optimization to maximize the reward, we can identify sequences of experiments that are optimal for parameter inference in a computationally feasible manner.

Coauthors: Youssef Marzouk

⁂

Bayesian study designs for new drug development

Satoshi MORITA (Yokohama City University, Japan)

Recently, Bayesian study designs have been increasingly used in medical area, especially in clinical trials to develop new treatments. It may be owing to the flexibility of Bayesian approaches. This presentation focuses on Bayesian clinical study designs for new drug development, especially in oncology area. Despite the ever-increasing number of new agents, the number of patients available for clinical trials remains limited. Thus, clinical trials should be designed in an efficient, adaptive and ethical way to save resources, draw correct conclusions earlier, benefit more patients and result in less unnecessary toxicities. Given the limited resources and especially a small sample size, for example, phase I dose-finding trials aim to correctly find the maximum tolerated dose for an investigational drug. Therefore, it is critically important to conduct clinical trials efficiently and appropriately. These concerns have motivated the development of more useful and efficient study designs. While novel design and analytic approaches to clinical trials are nowadays more frequently developed and adopted, many issues have arisen in applications to clinical trials from practical perspectives. This presentation discusses these new developments and emerging issues including patient heterogeneity, genetic difference, multiple outcomes, etc. With respect to patient heterogeneity, we will focus on investigation of a predictive biomarker which can identify the sensitive subpopulation of patients who are most likely to benefit from a specific treatment.

⁂

Simulation-based optimal design using MCMC

Antti SOLONEN (Lappeenranta University of Technology, Finland)

In classical parameter estimation of nonlinear models, a Gaussian approximation of parameter uncertainty is usually obtained by linearizing the model around a point estimate. Classical optimal design methods are based on this approximation, and aim, for instance, at minimizing the covariance of the parameters. These approaches suffer from two flaws. First, they depend on the linearization point and on the validity of the Gaussian approximation. Second, classical design methods are often unavailable in ill-posed estimation situations, where previous data lacks the information needed to properly construct the design criteria.

With MCMC methods, nonlinear parameter estimation problems can be solved without using, e.g., Gaussian approximations, and MCMC has been intensively used for model fitting in many fields of science and engineering. Therefore, it is natural to study ways how the MCMC output — samples from the parameter posterior — can be used further in optimal design. This can be done via simulation based optimal design introduced by Müller et al. (2004). In this talk, we discuss several aspects of simulation based optimal design based on MCMC parameter estimation, and illustrate the benefits of the approach by numerical examples.

Coauthors: Heikki Haario and Marko Laine

⁂

15:00–16:30 Room A **Hierarchies of Bayesian nonparametric processes**

Organizer: Lancelot F. James (Hong Kong University of Science and Technology, Hong Kong)
Chair: Peter Orbanz (Columbia University, USA)

Multiresolution Gaussian process regression

Emily FOX (University of Pennsylvania, USA)

A key challenge in many time series applications is efficiently capturing potentially long-range, non-Markovian dependencies. One method of addressing this challenge is through employing a Gaussian process (GP) with an appropriate (non-band-limited) kernel function. However, the smoothness of the GP random functions can blur key elements of the signal if abrupt changes occur. Likewise, a changepoint model between smooth functions cannot capture long-range dependencies. Instead, we propose a multiresolution GP that hierarchically couples a collection of smooth GPs, each defined over an element of a random nested partition. Long-range dependencies are captured by the top-level GP while the partition points define the abrupt changes in the time series. Due to the inherent conjugacy of the GPs, one can analytically marginalize the GPs and compute the conditional probability of the observations given the partition tree. This allows for efficient inference of the partition itself, for which we employ graph-theoretic techniques. We analyze the theoretical properties of the multiresolution GP, as well as applying it to the analysis of Magnetoencephalography (MEG) recordings of brain activity.

Coauthors: David Dunson

⁂

Hierarchical random measure

Dohyun KIM (Seoul National University, Korea)

An Indian Buffet Process can be explained by counting process with independent increment models with random measures, which are specified by two parameter Beta processes, as intensity processes. In this paper, we extended it with more general Lévy densities and also proposed a Naïve Bayes classifier with a Lévy density as an application. To construct the hierarchy in random measure, defining random measures on non-smooth base measure is an important issue. A hierarchical two parameter Poisson Dirichlet process(HPD) is constructed by compositions of two parameter Poisson Dirichlet processes. In analogous to the hierarchical Dirichlet process and the hierarchical Beta process, we defined a hierarchical random measure with a hierarchical stick breaking process of HPD. As an application, we also proposed a hierarchical Naïve Bayes classifier with a HPD. It has several advantage. Firstly, our proposed models can apply the hierarchy of classes. Secondly, these can capture and model power-law behavior, which is increasing order of newly chosen feature, and sparsity that is sparseness of samples in a chosen feature. We described these properties by analyzing simulated data sets and real data sets.

Coauthors: Lancelot F. James

⁂

On the stick-breaking representation for Gibbs-type priors

Stefano FAVARO (University of Torino, Italy)

Random probability measures are the main tool for Bayesian nonparametric inference, given their law acts as a prior distribution. Many well-known priors used in practice admit different, though (in distribution) equivalent, representations. Some of these are convenient if one wishes to thoroughly analyze the theoretical properties of the priors being used, others are more useful in terms of modeling and computation As for the latter purpose, the so-called stick-breaking constructions certainly stand out. In this talk we focus on the recently introduced class of Gibbs-type priors and provide a stick-breaking representation for it.

Coauthors: Antonio Lijoi, Igor Pruenster and Stephen Walker

⁂

Infinite structured explicit duration hidden Markov models

Frank WOOD (Columbia University, USA)

We develop a class of models we call infinite structured explicit duration hidden Markov models (ISEDHMM)s. ISEDHMMs are HMMs that possess an unbounded number of states, encode state dwell-time distributions explicitly, and have constraints on what state transitions are allowed. The ISEDHMM framework is uses a hierarchical approach to generating dependent Dirichlet processes via selective renormalization of gamma processes. ISEDHMMs generalize explicit duration finite HMMs (hidden semi-Markov models), infinite HMMs, left-to-right HMMs, and more (all are recoverable by specific choices of ISEDHMM parameters). This suggests that ISEDHMMs should be applicable to data-analysis problems in a variety of settings.

Coauthors: Jonathan Huggins

❉

15:00–16:30 Room C **Recent advances in Bayesian causal inference**

Organizer: Fan Li (Duke University, USA)
Chair: Fan Li (Duke University, USA)

Using multivariate outcomes in Bayesian causal inference with noncompliance

Fabrizia MEALLI (University of Florence, Italy)

A Bayesian approach to causal inference in the presence of noncompliance to assigned randomized treatment is considered. It exploits multivariate outcomes for improving estimation of weakly identified models. Using artificial data sets, we analyze the properties of the posterior distribution of causal estimands to evaluate the potential gains of jointly modelling more than one outcome. The approach can be used to assess robustness with respect to deviations from structural identifying assumptions. It can also be extended to the analysis of observational studies with instrumental variables where exclusion restriction assumptions are usually questionable. An application using data from the Jobs II study illustrates the approach.

Coauthors: Fan Li and Alessandra Mattei

❉

Bayesian multivariate inference for a non-standard fuzzy regression discontinuity design

Fan LI (Duke University, USA)

Regression discontinuity designs (RDD) identify causal effects of interventions by exploiting treatment assignment mechanisms that are discontinuous functions of observed covariates. In standard RDDs, the probability of treatment changes discontinuously if a covariate exceeds a threshold. We consider a more complex RDD setup where the treatment is determined by both a covariate and an application status. In particular, we focus on a fuzzy RDD with this setup, where the causal estimand and estimation strategies are different from those in the standard instrumental variable approach to fuzzy RDDs. A Bayesian approach is developed for drawing inferences of the causal effect and multivariate outcomes are utilized to sharpen the analysis. The method is applied to evaluate the effects of Italian university grant on student dropout and academic performances.

Coauthors: Fabrizia Mealli

❉

Assessing the causal effect of treatment dosages in the presence of self-selection

Michael R. ELLIOTT (University of Michigan, USA)

To make drug therapy as effective as possible, patients are often put on an escalating dosing schedule. But patients may choose to take a lower dose because of side effects. Thus even in a randomized trial, the dose level received is a post-randomization variable, and comparison between groups based on the dosing level may no longer have a causal interpretation. Hence we use the potential outcomes framework to define pre-randomization “principal strata” from the joint distribution of doses selected under control and treatment arms, with the goal of estimating the effect of treatment within the subgroups of the population who will select a given dosing level under treatment. We utilize Bayesian framework with Markov chain Monte Carlo algorithm for analysis of a randomized clinical study on painful bladder syndrome, and compare the analysis results from our proposed model with traditional analysis method. Simulation results show that the estimates of interests in our proposed causal model have correct coverage.

Coauthors: Xin Gao

❉

A robust outcome-free procedure for interval estimation of causal effects

Roee Gutman (Brown University, USA)

The estimation of population/subpopulation average treatment effects has been a subject of extensive research. In a randomized experiment, common practice entails estimating the average treatment effect by calculating the difference between the average outcome in the treatment group and the average outcome in the control group. In cases where covariates that affect the outcome exist, estimation of the treatment effect is typically controlled by using a model that combines the treatment effect and the covariates in an additive manner. This modeling relies on the assumption that the response surfaces from the outcome given the covariates are parallel in the control and treatment groups. When this assumption is incorrect, the estimation of the treatment effect may be unreliable. In observational studies, the effect of this assumption is even more substantial, because the distributions of the covariates in the two groups differ.

This talk will describe a three-stage procedure based on Rubin's framework for causal inference. First, we create subclasses that include observations from each group based on the covariates. Next, we independently estimate the response surface in each group using flexible spline model. Lastly, multiple imputations of the missing potential outcomes are performed. A simulation analysis which resembles real life situations and compares this procedure to other common methods is carried out. In relation to other methods and in many of the experimental conditions examined, our proposed method is the only one that produced a valid statistical procedure while providing a relatively precise point estimate and a relatively short interval estimate.

Coauthors: Donald B. Rubin

⁂

15:00–16:30 Room D **Adaptive Monte Carlo**

Organizers:	Nando de Freitas (University of British Columbia, Canada)
	Christophe Andrieu (University of Bristol, UK)
	Arnaud Doucet (University of Oxford, UK)
Chair:	Nando de Freitas (University of British Columbia, Canada)

Adaptive equi-energy samplers

Gersende Fort (LTCI, CNRS and Telecom ParisTech, France)

Interacting Markov chain Monte Carlo (MCMC) algorithms have been introduced in order to have better mixing properties than classical MCMC when the target distribution is multimodal. The Equi-Energy sampler (Kou et al. 2006) is an example of such an interacting MCMC, which has been defined in order to favor the interactions between the parallel chains. This algorithm depends on many design parameters (such as the number of parallel chains, the temperatures of each chain, the definition of the energy rings, ...) and the role of these parameters on the convergence and the efficiency of the sampler is an open question.

In this talk, I will present new results for the convergence of the interacting MCMC samplers. New Equi-Energy algorithms will be proposed, based on adaptive procedure in order to tune on the fly some design parameters. We will show application of these algorithms to motif discovery in DNA sequences.

Coauthors: Eric Moulines, Pierre Priouret and Amandine Schreck

⁂

Robustness in adaptive MCMC

Matti Vihola (University of Jyväskylä, Finland)

The adaptive Metropolis (AM) algorithm of Haario, Saksman and Tamminen [Bernoulli 7 (2001) 223-242] employs the empirical covariance to tune the random-walk Metropolis algorithm. We consider a new robust adaptive Metropolis algorithm estimating the shape of the target distribution and simultaneously optimising the acceptance rate. The adaptation rule is computationally simple adding no extra cost compared with the AM algorithm. The adaptation strategy can be seen as a multidimensional extension of the previously proposed method adapting the scale of the proposal distribution in order to attain a given acceptance rate. The empirical results show promising behaviour of the new algorithm in an example with a Student target distribution having no finite second moment, where the AM covariance estimate is unstable. In the examples with finite second moments, the performance of the new approach seems to be competitive with the AM algorithm combined with scale adaptation. We also consider some techniques to alleviate the slow mixing of the random-walk Metropolis in multimodal scenarios.

Coauthors: Blazej Miasojedow and Eric Moulines

⁂

Learning high-dimensional deep Boltzmann machines using adaptive MCMC

Russ SALAKHUTDINOV (University of Toronto, Canada)

Extracting meaningful representations from high-dimensional data lies at the core of solving many AI related tasks, including visual object recognition, language and speech perception, anomaly detection, and time series analysis.

In this talk, I will first introduce a broad class of hierarchical probabilistic models called Deep Boltzmann Machines (DBMs). When modeling high-dimensional data, the distribution defined by the Deep Boltzmann Machine has a rough energy landscape with many local minima separated by high energy barriers. The commonly used Gibbs sampler tends to get trapped in one local mode, which results in unstable learning dynamics and leads to poor parameter estimates. I will therefore focus on learning DBMs using adaptive MCMC algorithms. In particular, I will introduce a Coupled Adaptive Simulated Tempering (CAST) algorithm that can be used to better explore a highly multimodal energy landscape. I will then demonstrate that the proposed algorithm considerably improves parameter estimates, particularly when learning large-scale DBMs. Finally, I will show that CAST allows DBMs to learn useful hierarchical representations that can be successfully applied in many application domains, including information retrieval, 3-D visual object recognition, speech perception, and nonlinear dimensionality reduction.

⁂

Adaptive MCMC with bandits

Ziyu WANG (University of British Columbia, Canada)

The talk introduces novel ways of adapting MCMC algorithms using parametric and nonparametric bandits. The first part of the talk covers the use of Gaussian process bandits to adapt the parameters of a sophisticated algorithm for sampling from discrete distributions: SARDONICS. This nonparametric adaptation strategy is also known as Bayesian optimization, and has been used routinely in experimental design. SARDONICS allows for large moves in the discrete state space. This is achieved by constructing self-avoiding walks (SAWs) in the state space. As a consequence, many bits are flipped in a single MCMC step. The algorithm has several free parameters, which require careful adaptation for it to work in practice. The adaptive SARDONICS algorithm performs remarkably well in a broad number of sampling tasks: toroidal ferromagnetic and frustrated Ising models, 3D Ising models, restricted Boltzmann machines and chimera graphs arising in the design of quantum computers. The second part of the talk introduces a strategy for adapting the Hybrid Monte Carlo (HMC) algorithm using parametric bandits with nonlinear features. HMC is a powerful Markov chain Monte Carlo (MCMC) method, but it requires careful tuning of its hyper-parameters. A Bayesian parametric bandit approach is introduced to carry out the adaptation of the hyper-parameters while the Markov chain progresses. We propose the use of cross-validation error measures for adaptation, which we believe are more pragmatic than many existing adaptation objectives. The new measures take the intended statistical use of the model, whose parameters are estimated by HMC, into consideration. We apply these two innovations to the adaptation of HMC for prediction and feature selection with multi-layer feed-forward neural networks. The experiments with synthetic and real data show that the proposed adaptive scheme is not only automatic, but also does better tuning than human experts.

Coauthors: Firas Hamze, Nimalan Mahendran and Nando de Freitas

⁂

17:00–18:30 Room B **Beyond MCMC methods in Bayesian inference**

Organizer: Paul Damien (University of Texas at Austin, USA)
Chair: Yee Whye Teh (University College London, UK)

Generalized direct sampling for Bayesian hierarchical models

Michael BRAUN (Massachusetts Institute of Technology, USA)

We develop a new method to sample from posterior distributions in hierarchical models without using Markov chain Monte Carlo. This method is generally applicable to high-dimensional models involving large data sets. Illustrative analysis exemplifies the ease with which one could implement our method, which results in independent samples from the posterior distributions of interest.

Coauthors: Paul Damien

⁂

The pseudo-marginal approach and exact approximations for efficient MC

Christophe ANDRIEU (University of Bristol, UK)

We present and review exact approximations to standard Monte Carlo algorithms and describe some of their known theoretical properties.

❉

Sequential estimation in dynamic spatio-temporal models

Jonathan STROUD (George Washington University, USA)

Sequential Monte Carlo methods for state-space models have been sucessfully applied in many problems. However, standard methods like the SIR particle filter and auxiliary particle filter are known to suffer from particle degeneracy in the presence of outliers, high dimensionality and unknown static parameters. In this talk, we propose a new method for sequential estimation of dynamic spatio-temporal models. The method allows for online Bayesian estimation of states and parameters and works well in nonlinear, high-dimensional models. It is based on a combination of ensemble Kalman filter and Gaussian mixture updates. We illustrate the approach using the Lorenz (1996) 40-variable system, and on a space-time pollution monitoring problem in Mexico City.

❉

Understanding the slice sampler

Maria KALLI (University of Kent, UK)

Inference in Bayesian nonparametric mixtures has traditionally used the Polya urn scheme representation of the non parametric prior . Recently there has been interest in conditional methods which include the unknown distribution in the sampler. One such method is the slice sampler Walker (2007). The method introduces a latent variable which avoids the full conditional for the allocations having an infinite number of atoms. Kalli et al (2011) describe different ways of introducing this latent variable. This talk will consider the effect of this choice on the behaviour of the Markov Chain.

❉

17:00–18:30 Room C **Bayesian inference in science: the pursuit of a synergy**

Organizer: Fabio Rigat (University of Warwick, UK)
Chair: Fabio Rigat (University of Warwick, UK)

Assessing the impact of a movement network on the spatiotemporal spread of infectious diseases

Leonhard HELD (University of Zurich, Switzerland)

Linking information on a movement network with space-time data on disease incidence is one of the key challenges in infectious disease epidemiology. In this article, we propose and compare two statistical frameworks for this purpose, namely, parameter-driven and observation-driven models. Bayesian inference in PD models is done using integrated nested Laplace approximations, while OD models can be easily fitted with existing software using maximum likelihood. The predictive performance of both formulations is assessed using proper scoring rules. As a case study, the impact of cattle trade on the spatiotemporal spread of Coxiellosis in Swiss cows, 2004-2009, is finally investigated.

Coauthors: Birgit Schrödle and Havard Rue

❉

Mortality projection incorporating model uncertainty

Jonathan FORSTER (University of Southampton, UK)

Researchers interested in longevity and mortality projection have available a wide variety of mortality projection models from which to choose. Having been chosen, the favoured model is typically fitted against a suitable dataset and projected forward in time to produce estimated future mortality rates together with a prediction interval. In this paper, we describe a Bayesian approach to the quantification of mortality projection uncertainty that incorporates model uncertainty.

Initially, we focus on individual models, and develop Bayesian methodology for computing the predictive (forecast) distributions for various models. We demonstrate how to effectively compute probabilistic projections across mortality projection models. We use a computational approach which involves separate MCMC generation of parameter values for each model, and then uses the MCMC output to estimate the posterior model probabilities.

Our approach is illustrated on data from England and Wales. We compute forecasts which incorporate a diverse selection of models. The integrated projection uncertainty provides a coherent and more realistic assessment of uncertainty than any corresponding analysis based upon a single model.

Coauthors: Xiaoling Ou

⁂

Getting the facts right on the climate change debate

Ricardo LEMOS (University of Queensland, Australia)

Two contentious aspects in the climate change debate are discussed: observed 20th century trends and the usefulness of Global Climate Models (GCMs). Correct detection of significant trends from measurements precedes attribution to human or natural causes, thus it is important to develop statistical models capable of tackling spatial and temporal sparsity of data, location-dependent autocorrelation and anisotropy. GCMs should be able to reproduce those trends, when supplied with 20th century forcing, and should produce a cone of uncertainty around mean forecasts, for every emissions scenario considered. Since state-of-the-art GCMs are deterministic, the procedure to obtain such cones is an open issue. Two Bayesian models, purposely built to address these problems, are described. The first blends Dynamic Linear Models and Discrete Process Convolutions, in a hierarchical fashion. The second weights the output from 17 GCMs, based on their ability to reproduce the observed dominant pattern of temperature variability in the North Pacific Ocean. Novel techniques are introduced so that models can be applied to global, long-term problems, where massive data sets of various origins are available. Bottlenecks in computation and methodological dissemination conclude this talk.

Coauthors: Bruno Sansó, Francisco Beltrán and Roy Mendelssohn

⁂

On the construction of parametric hierarchical predictive distributions

Fabio RIGAT (University of Warwick, UK)

Applied statistics often relies on postulated theoretical models to fit experimental data, allowing for a data-driven assessment of such embedded theories. Although this learning process provides the necessary feedback to improve an understanding of the data generating process, it may not provide a satisfactory predictive tool for short-term decision making. This potential trade-off between a model's explanatory power and its predictive power is indeed well-known, motivating most non-parametric and predictivistic approaches to applied statistical modeling. A glaring example is provided by non-parametric Bayesian hierarchical models, where a finite-dimensional centering distribution only partially defines a predictive stochastic process. In the first part of this talk we illustrate a novel methodology aiming at maximising the predictive power of parametric statistical models using a hierarchical but finite-dimensional construction of the predictive distribution. While sharing the same pragmatic emphasis on prediction placed by non-parametric hierarchical models, this methodology provides a more parsimonious avenue to detecting and correcting intrinsic limitations of the underlying explanatory model. Illustrative applications will be provided in the second part of the talk.

Coauthors: Luca Moraschini

⁂

17:00–18:30 Room D **Bayesian approaches to design and model comparison**

Organizer: Marina Savelieva (Novartis Pharma AG, Switzerland)
Chair: Luis Nieto-Barajas (ITAM, Mexico)

A particle filter for Bayesian sequential design

James McGREE (Queensland University of Technology, Australia)

A particle filter algorithm will be presented that can be used for Bayesian sequential design in the presence of model and parameter uncertainty. Our focus is on adaptive design for model discrimination but the methodology is applicable if one has a different design objective. A particle filter is run in parallel for each model and the algorithm relies on a convenient estimator of the marginal likelihood of each model. Approximating posterior model probabilities in this way allows us to use model discrimination utilities derived from information theory that were previously difficult to compute for all but the simplest of models. The algorithm requires very little problem specific tuning. Our work is motivated by an application to phase I clinical trials.

Coauthors: Christopher Drovandi and Anthony Pettitt

⁂

Design of text mining experiments

Matt TADDY (University of Chicago Booth School of Business, USA)

Sequential design of experiments is an interest area for many in both engineering and statistics. Previous work by the authors has looked at this problem in the context of robust search optimization, wherein we were tasked with augmenting a local pattern search algorithm with search locations that had a high probability of improvement. In particular, parallel execution required optimal sets of multiple new search points rather than a single ‘best’ new location. This talk will re-visit these ideas, and look to problems in text-sentiment analysis as a new application area. Here, the goal is to predict variables that motivated language use (e.g., the author’s political beliefs in a news article). Typically, huge amounts of text are available, but obtaining sentiment-scored text samples for model training is very expensive. Hence we discuss methods for choosing optimal subsamples from the available conversation. The technology will be illustrated in scoring of various sentiment indices for text data from the streaming twitter feed.

⁂

Functional uniform prior distributions for nonlinear regression

Bjoern BORNKAMP (Novartis Pharma AG, Switzerland)

In this talk I will consider the topic of finding prior distributions in nonlinear modelling situations, that is, when a major component of the statistical model depends on a non-linear function. Making use of a functional change of variables theorem, one can derive a distribution that is uniform in the space of functional shapes of the underlying nonlinear function and then back-transform to obtain a prior distribution for the original model parameters. The primary application considered in this talk is non-linear regression in the context of pharmaceutical dose-finding trials. Here the so constructed priors have the advantage that they are parametrization invariant as opposed to uniform priors on parameter scale and can be calculated prior to data collection as opposed to the Jeffrey’s prior. I will investigate the priors for a real data example and for calculation of Bayesian optimal designs, which require the prior distribution to be available before data collection has started.

⁂

The calibrated Bayes factor for model comparison

Steven MACEACHERN (Ohio State University, USA)

The traditional Bayesian approach to model comparison is based on the Bayes factor—often said to be the relative support of the data for one model in comparison to another. In fact, the Bayes factor can be heavily influenced by the prior distributions under the two models. This causes Bayesian model comparison to break down when a prior is improper; it creates difficulties when prior information is weak and an arbitrarily overdispersed prior distribution is used; the problems compound when one of the models is of high or infinite dimension, where the prior is, in many ways, arbitrary. To reduce the arbitrariness of the Bayes factor, we propose a novel criterion called the calibrated Bayes factor. The new criterion relies on training samples to yield a partial-posterior distribution that is calibrated to a fixed level of concentration. The calibrated Bayes factor is then computed as the Bayes factor over the remaining data. We motivate the method, explain it, illustrate its benefits through simulation in simple (one-sample) and more complex (regression) settings, and use it on obesity data from the Ohio Family Health Survey.

Coauthors: Xinyi Xu, Pingbo Lu and Ruoxi Xu

⁂

Wednesday, June 27th

09:00–10:30 Room B **Bayesian applications**

Organizer: David Draper (University of California, Santa Cruz, USA)
Chair: David Draper (University of California, Santa Cruz, USA)

Bayesian testing of factor asset pricing models

Carlos CARVALHO (University of Texas at Austin, USA)

The Capital Asset Pricing Model (CAPM) and its various multi-factor extensions are generally viewed as the central paradigm of asset pricing research. In this paper we describe a formal statistical methodology to evaluate the validity of these models. Our approach is couched in the context of the “Size” and “Value” anomalies and thus is based on a well specified alternative hypothesis. We explore the finite sample properties of our approach by working with a fully Bayesian linear model and conclude that most of the “fail to reject” results that recently appear in the finance literature are a consequence of the lack of power of the statistical procedures commonly used.

⁂

Power-intrinsic Bayesian variable selection in Gaussian linear models

David DRAPER (University of California, Santa Cruz, USA)

Bayes factors are at the heart of one leading method for Bayesian variable selection in regression models, but this approach has long been plaguedby a fundamental difficulty when scientific context suggests diffuseness in the prior distributions on the parameters in the models under comparison. Intrinsic prior distributions, which use improper baseline priors as a starting point, attempt to solve this problem by extracting training samples from the full data set and averaging over the resulting intrinsic Bayes factors, but (a) the number of possible training samples is typically too large to average over all of them and (b) choosing a random subset of the set of all possible training samples adds an extraneous layer of Monte-Carlo noise to the model-comparison process. In this work we combine ideas from power-prior and unit-information-prior methods to greatly diminish the effect of training samples on the intrinsic-prior approach: we raise the likelihood involved in the intrinsic prior distribution to a power that produces aprior information content equivalent to one data point. The result is that in practice our power-intrinsic methodology is sufficiently insensitive tothe size n* of the training sample that one may take n* equal to the full-data sample size and dispense with training samples altogether; this promotes stability of the resulting Bayes factors, removes the arbitrariness arising from individual training-sample selections, and can greatly increase computational speed, allowing many more models to be compared within a fixed CPU budget.

Coauthors: Dimitris Fouskakis and Ioannis Ntzoufras

⁂

Hierarchically related regression structures for integrative genomics analysis

Sylvia Richardson (Imperial College London, UK)

This talk will consider the task of building efficient regression models for sparse multivariate analysis of high dimensional data sets, focussing on cases where the numbers q of outcomes Y and p of predictors X to analyse jointly are both large with respect to the sample size n, a challenging bi-directional situation of 'large p, small n' paradigm. The analysis of such data sets arise commonly in genetical genomics, with X linked to the DNA characteristics and Y corresponding to measurements of fundamental biological processes such as transcription, protein or metabolite production. In this context, a particular interest is the detection of 'hot spots', i.e. genetic variants that are associated with several responses, as these could correspond to candidate master regulator of a co-expression network. Building on the Bayesian variable selection set-up for the linear model and associated efficient MCMC algorithms, we propose a generic set-up of hierarchical related sparse regressions, where parallel regressions of Y on the set of covariates X are linked in a hierarchical fashion, in particular through the prior model of the variable selection indicators. This hierarchical set-up is flexible as it can accommodate different probability models underlying of the distribution of the joint indicators of associations between X and Y, including in particular models that will enhance the detection of hot spots by facilitating the borrowing of information between the responses. The applicability of the model for uncovering functional eQTL hot-spots will be illustrated on case studies.

Coauthors: Leonardo Bottolo

❉

Bayesian models for multiple physiological time series in multi-subject studies

Raquel Prado (University of California, Santa Cruz, USA)

We present time-domain approaches for describing the relationships across multiple physiological signals. We consider state-space models with sparsity priors for inferring effective brain connectivity from fMRI data at multiple brain regions. The sparsity priors are placed on the coefficients that model region-specific activations as linear functions of the blood oxygenation level-dependent (BOLD) signals. Such prior structure allows us to determine the nature of the coupling relationship between any two brain regions. We also consider hierarchical dynamic factor models for the analysis of brain signals collected in multi-subject and multi-group studies. Models with subject and group-specific connectivity parameters as well as models with subject and group-specific latent structure are considered. We illustrate the use of these models in the analysis of fMRI and EEG data.

Coauthors: Jairo Fuquene

❉

09:00–10:30 Room C **Beta processes: extensions and applications**

Organizer: Nils Lid Hjort (University of Oslo, Norway)
Chair: Sonia Petrone (Bocconi University, Italy)

Analysis of credit risk history data using a mixture of beta-Dirichlet processes prior

Yongdai Kim (Seoul National University, Korea)

We propose a Bayesian model for analyzing credit risk history data. We model the data using a finite state time-continuous inhomogeneous Markov process and derive the posterior distribution of the cumulative intensity functions with a mixture of beta-Dirichlet processes prior. A mixture of beta-Dirichlet processes prior is an extension of a beta-Dirichlet process prior proposed by Kim et al. (2012). Advantages of a mixture of beta-Dirichlet processes prior over a standard beta-Dirichlet process prior is similar to those for mixture of Dirichlet processes (Doss, 1994) and mixture of beta processes (Kim, 2001). That is, we can incorporate information of the parametric form of the cumulative intensity functions into the prior. We explain why incorporating such information is important. We apply the proposed model to a real data set of credit risk history collected by WestLB AG (a bank in Germany).

Coauthors: Rafael Weissbach

❉

Inducing dependency between beta processes using a Poisson process representation

Sinead Williamson (Carnegie Mellon University, USA)

As completely random measures, beta processes can be described in terms of Poisson processes. Such representations allow us to make use of well-known properties of the Poisson process in developing extensions to the beta process. In this talk, I will discuss ways of extending the beta process to model collections of data located in some covariate space. Such models are appropriate for latent feature models where the feature probabilities vary with covariates such as time and space, and can also be applied to categorical covariates.

Coauthors: Peter Orbanz, Zoubin Ghaharamani and Nick Foti

❈

The beta process, exchangeable feature models, and the feature paintbox

Tamara Broderick (University of California, Berkeley, USA)

While the beta process has opened up a realm of practical Bayesian nonparametric feature models much as the Dirichlet process did for Bayesian nonparametric clustering models, we have yet to see a treatment of the full class of exchangeable feature models. To that end, we examine such models as the beta Bernoulli process and beta negative binomial process as examples of this larger class. Just as the exchangeable partition probability function (EPPF) can be used to describe the distribution of cluster membership under an exchangeable clustering model, we examine an analogous “exchangeable feature probability function” for certain types of feature models. Moreover, recalling Kingman’s paintbox theorem as a characterization of the class of exchangeable clustering models, we study a similar construction in the feature model case.

Coauthors: Jim Pitman and Michael I. Jordan

❈

A brief history of beta processes

Nils Lid Hjort (University of Oslo, Norway)

I attempt to present a brief history and overview of the Beta processes, from their introduction in papers of mine from 1985 and 1990 up to present extensions and new application areas. The first papers aimed at providing nonparametric Bayesian methods for the modelling, analysis and interpretation of survival and event history data, which along with various generalisations and advances regarding computation, simulation and theoretical performance issues is still an active and rich field of research. There are also various bigger and more complex model constructions where the Beta processes serve as one of several building blocks. The Beta processes are also enjoying an upsurge in interest from the perspectives of machine learning, in particular via the Indian Buffet Processes and their relatives.

❈

11:00–12:30 Room A **Scaling Bayesian computation to handle big data: methods and feasibility**

Organizer: Dawn Woodard (Cornell University, USA)
Chair: Galin Jones (University of Minnesota, USA)

Massive data and the bootstrap

Michael Jordan (University of California, Berkeley, USA)

Datasets in the terabyte and petabyte range are becoming increasingly common in science and technology and it is essential that statisticians begin to address the inferential and computational issues that arise in the setting of such “massive data.” In this talk we (provocatively)take a non-Bayesian perspective and consider the classical bootstrap,which is readily parallelizable and thus seems to be a good match to modern distributed computing architectures. For a terabyte dataset, however, each bootstrap sample is several hundred gigabytes in size and thus a naive implementation of the bootstrap for such massive datasetsis simply not feasible. There is an elegant solution to this dilemma, however, that we refer to as the “Bag of Little Bootstraps” (BLB). After discussing this solution we will pose (and perhaps not resolve) the question as to whether there is a Bayesian counterpart of BLB.

Coauthors: Ariel Kleiner, Ameet Talwalkar and Purna Sarkar

❈

Multi-resolution inference of stochastic models from partially observed data

Samuel KOU (Harvard University, USA)

Stochastic models, diffusion models in particular, are widely used in science, engineering and economics. Inferring the parameter values from data is often complicated by the fact that the underlying stochastic processes are only partially observed. Examples include inference of discretely observed diffusion processes, stochastic volatility models, and double stochastic Poisson (Cox) processes. Likelihood based inference faces the difficulty that the likelihood is usually not available even numerically. Conventional approach discretizes the stochastic model to approximate the likelihood. In order to have desirable accuracy, one has to use highly dense discretization. However, dense discretization usually imposes unbearable computation burden. In this talk we will introduce the framework of Bayesian multi-resolution inference to address this difficulty. By working on different resolution (discretization) levels simultaneously and by letting the resolutions talk to each other, we substantially improve not only the computational efficiency, but also the estimation accuracy. We will illustrate the strength of the multi-resolution approach by examples.

Coauthors: Benjamin Olding and Jun S. Liu

⁂

Potts-modulated Gaussian processes for high-dimensional output computer models

Yves ATCHADE (University of Michigan, USA)

We combine Markov random fields (from a Potts distribution) and Gaussian processes to obtain 2/3-dimensional random functions with non-smooth sample paths. The model is potentially useful for calibrating computer models with high-dimensional, non-smooth outputs. The talk will discuss a number of approaches for fitting this class of models in a Bayesian framework.

⁂

Scalability of Markov chain methods for genomic motif discovery

Dawn WOODARD (Cornell University, USA)

We analyze the efficiency and scalability of a popular Gibbs sampling method used for statistical discovery of gene regulatory binding motifs in DNA sequences. We bound its convergence rate, and show that, due to multimodality of the posterior distribution, the rate of convergence often decreases exponentially as a function of the length of the DNA sequence. This implies that the run time of the algorithm grows exponentially in the sequence length, to attain a fixed accuracy. Specifically, we show that this occurs whenever there is more than one true repeating pattern in the data. In practice there are typically multiple such patterns in biological data, the goal being to detect the best-conserved and most frequently-occurring of these. Our findings match empirical results, in which the motif-discovery Gibbs sampler has exhibited such slow convergence that it is used for finding modes of the posterior distribution (candidate motifs) rather than for obtaining samples from that distribution. Ours are some of the first meaningful bounds on the convergence rate of a Markov chain method for sampling from a multimodal posterior distribution, as a function of statistical quantities like the number of observations.

Coauthors: Jeffrey Rosenthal

⁂

11:00–12:30 Room D **Applied Bayesian Econometrics**

Organizer: Sylvia Frühwirth-Schnatter (Vienna University of Economics and Business, Austria)
Chair: Sylvia Frühwirth-Schnatter (Vienna University of Economics and Business, Austria)

On the long run volatility of stocks

Hedibert LOPES (University of Chicago Booth School of Business, USA)

In this paper we address the question "what is the variance of long horizons portfolios?" The work is directly motivated by "Are Stocks Really Less Volatile in the Long Run" (2011) by Pastor and Stambaugh. In that work, they use their "Predictive Systems" framework to challenge the conventional wisdom that stocks are less volatile over long horizons when compared to short horizons. Their conclusion is reached by the incorporation of parameter uncertainty and "imperfect" predictors. Preserving the economic motivation of their approach, we develop parsimonious alternatives to "Predictive Systems" and show that, when compared to the correct benchmark, stocks can still be appealing for long run portfolios. Central to our results is a careful assessment of the priors investors have about key quantities underlying market behavior. In particular we explore the impact of priors that enforce the correlation between expected and realized returns to be negative in a time-varying volatility context.

Coauthors: Carlos Carvalho and Robert McCulloch

⁂

Estimation of copula models with discrete margins via Bayesian data augmentation

Michael SMITH (Melbourne Business School, Australia)

Estimation of copula models with discrete margins can be difficult beyond the bivariate case. We show how this can be achieved by augmenting the likelihood with latent variables, and computing inference using the resulting augmented posterior. Our method applies to all parametric copulas where the conditional copula functions can be evaluated, not just elliptical copulas as in much previous work. Moreover, the copula parameters can be estimated joint with any marginal parameters, and Bayesian selection ideas employed. To demonstrate the potential in higher dimensions we estimate 16 dimensional D-vine copulas for a longitudinal model of usage of a bicycle path in the city of Melbourne, Australia.

Coauthors: Mohamad Khaled

⁂

Bayesian mixture models for large microeconomic panels

Mattias VILLANI (Linköping University, Sweden)

We explore the potential of Bayesian mixture models for analyzing large microeconomic panel data sets. The size of many of the currently available data sets is a serious computational challenge for mixture models, and we explore techniques to speed up the MCMC. As a working example, we analyze the bankruptcy risk of Swedish corporate firms in a large unbalanced panel with roughly 200,000 firms during the time period 1991-2008. The bankruptcy risk is modelled flexibly using a smooth mixture of survival models with financial ratios from the firms' annual balance sheets as time-varying covariates.

Coauthors: Matias Quiroz

⁂

Bayesian treatment effects models for panel outcomes

Helga WAGNER (University of Linz, Austria)

We consider a Bayesian treatment effects model for continuous outcomes observed over subsequent time periods. A joint model for the categorical treatment and the longitudinal outcome data is specified which includes level-specific and heterogeneous treatment effects on the outcome. To control for unobserved subject specific factors a random intercept is included in the model for the longitudinally observed outcomes. The treatment variable and the panel outcomes are assumed to be correlated, due to unobserved or unmeasured confounders. Estimation and inference is accomplished by MCMC methods. Incorporation of Bayesian variable selection methods allows to determine which covariates should be included in the model and to test for the existence common and level-specific effects of the treatment and the covariates. In the application the effects of incentives for maternal leave are of interest. We exploit a recent change in the parental leave policy in Austria that extended maternal benefits from 18 months since birth of the child to 30 months but left the period of job protection unchanged at 24 months. For our analysis we use data from the Austrian Social Security Register which contains complete individual employment histories since 1972 and also reports number of births and maternity and parental leave spells for all Austrian employees.

Coauthors: Sylvia Fruehwirth-Schnatter and Liana Jacobi

⁂

15:00–16:30 Room A **Bayesian methods for Spatial Statistics**

Organizer: Mikyoung Jun (Texas A&M University, USA)
Chair: Bruno Sanso (University of California, Santa Cruz, USA)

Computer model calibration with high dimensional multivariate space-time observations

Murali HARAN (Pennsylvania State University, USA)

Complex computer models are frequently used to make projections about climate. Calibrating unknown parameters of these models increasingly involves high dimensional multivariate space-time observations and computer model output. Extracting information from these data sources while accounting for various uncertainties appropriately poses considerable computational and inferential challenges. I will describe Gaussian process-based modeling approaches that allow for flexible dependence structures while at the same time retaining computational tractability by the use of dimension-reduction methods. The computational efficiency of these methods allow our scientific collaborators in climate science to learn about key climate model parameters, and to study important features related to the spatial scale of complex climate processes.

Coauthors: Sham Bhat, Klaus Keller, Roman Olson and Won Chang

⁂

Covariance approximation for large multivariate spatial datasets

Huiyan SANG (Texas A&M University, USA)

Multivariate spatial processes models for the analysis of multivariate spatial data are receiving increased attention these days. In many applications, the presence of large data sets poses major computational challenge for the fitting and prediction of these models. Here, we propose a class of multivariate cross-covariance models based upon the full-scale cross-covariance approximation, that yields a new cross-covariance function to approximate the original cross-covariance function at both large and small scales. This cross-covariance approximation method greatly facilitates computation in the application of the multivariate spatial processes for large datasets. We also offer several modifications of the full-scale covariance approximation for various popular multivariate cross-covariance models, utilizing their specific cross-covariance structures. We illustrate with both a simulated dataset as well as a real dataset.

Coauthors: Jianhua Huang

⁂

Estimating parametric uncertainties in computationally enabled strategies for climate model development

Gabriel HUERTA (Indiana University, USA)

The scientific, statistical, and computational strategies that are used for uncertainty quantification are key to the future of climate model development. The objective of this talk is to show how Bayesian statistical methodologies can leverage on HPC resources to help reduce some of the biases in future versions of the NCAR Community Atmosphere Model (CAM). In particular we use ideas from Gaussian Markov random fields to define a metric that takes into account spatial and field dependencies that avoids many of the limitations imposed by more traditional strategies based on empirical orthogonal functions. In combination to this metric, we also show how parametric uncertainties in climate models are more effectively estimated with Adaptive/Delayed rejection Metropolis Hastings methods in contrast to approaches based on stochastic optimization methods.

Coauthors: Charles Jackson and Alejandro Villagran

⁂

Adaptive sampling for Bayesian geospatial models

Hongxia YANG (IBM Watson Research, USA)

Bayesian hierarchical modeling with Gaussian process random effects provides a popular approach for analyzing point-referenced spatial data. For large spatial data sets, however, generic posterior sampling is infeasible due to the extremely high computational burden in decomposing the spatial correlation matrix. In this paper, we propose an efficient algorithm, called the adaptive griddy Gibbs (AGG) algorithm, to address the computational issues with large spatial data sets. The proposed algorithm dramatically reduces the computational complexity. We show theoretically that the proposed method can approximate the real posterior distribution accurately. The sufficient number of grid points for a required accuracy will also be derived. We compare the performance of the proposed method with that of some existing methods in simulation studies. Finally, we apply the proposed method to spatially indexed data concerning building energy consumption.

Coauthors: Fei Liu, Chunlin Ji and David Dunson

⁂

15:00–16:30 Room B **Bayesian analysis of protein structure and evolution**

Organizer: Scott Schmidler (Duke University, USA)
Chair: Scott Schmidler (Duke University, USA)

Bayesian evolutionary modeling of protein structures

Christopher CHALLIS (Duke University, USA)

We present a stochastic process model for the joint evolution of protein primary and tertiary structure, suitable for use in alignment and estimation of phylogeny. Indels arise from a classic Links model and mutations follow a standard substitution matrix, while backbone atoms diffuse in three-dimensional space according to an Ornstein-Uhlenbeck process. The model allows for simultaneous estimation of evolutionary distances, indel rates, structural drift rates, and alignments, while fully accounting for uncertainty. The inclusion of structural information enables phylogenetic inference on time scales not previously attainable with sequence evolution models. The model also provides a tool for testing evolutionary hypotheses and improving our understanding of protein structural evolution.

Coauthors: Scott Schmidler

⁂

Revealing the across site heterogeneity of nucleotide substitution patterns using Dirichlet process mixture model and Bayesian model selection

Chieh-Hsi WU (University of Auckland, New Zealand)

When inferring a phylogenetic tree from a nucleotide sequence alignment, a nucleotide substitution model is used to specify the relative exchange frequencies between two nucleotides i.e. the substitution pattern. An appropriate model is most commonly selected by using likelihood ratio-based tests conditioned on a fixed tree topology (Posada and Crandall, 1998), which assumes that the substitution pattern is homogeneous across all sites of the alignment. However, several papers have presented evidence for heterogeneity in substitution pattern across sites (Huelsenbeck and Nielsen, 1999; Shapiro et al. 2006). Here, the problem of interest is to decide which substitution models should be used at which sites of the alignment, in order to accommodate across-site heterogeneity in substitution pattern. This is solved by using a Dirichlet process prior to partition the alignment into categories, and within each category we use Bayesian model selection over a set of standard nucleotide substitution models. Two Dirichlet process mixture (DPM) models are investigated. The first assumes that the substitution pattern and site rate share the same site-to-category assignment. This assumption is relaxed in the second model where substitution pattern and site rate have independent Dirichlet process priors. These two models have been applied to three RNA virus alignments and one mammalian gene alignment. The analyses with DPM models provide substantial evidence for across-site heterogeneity in substitution pattern. Moreover, for some data sets, the phylogenetic parameters estimates from analyses with the DPM models are significantly different to those produced by the analyses assuming homogeneity in substitution pattern.

Coauthors: Marc A. Suchard and Alexei J. Drummond

⁂

Bayesian alignment of unlabeled marked point sets using random fields - molecular shape analysis

Irina CZOGIEL (Max Planck Institute for Molecular Genetics, Germany)

In structural bioinformatics it is of great interest to align molecules for structural comparison. In many cases the molecular data are thereby given in the form of atomic coordinates and some values of molecular properties (e.g. partial charge values) which have been observed at these coordinates. As there usually are no clear one-to-one correspondences between atoms of different molecules, the task of comparing molecular structures from a methodological point of view is therefore that of comparing unlabeled marked point sets.

We propose a statistical approach to the above problem. In order to define a suitable measure of similarity between two molecules, methods from statistical shape analysis are combined with techniques for predicting random fields in spatial statistics. Bayesian modeling of the predicted field overlap between pairs of molecules is proposed, and posterior inferenceof the alignment is carried out using MCMC. By representing the fields in reproducing kernel Hilbert spaces, the degree of molecule overlap can be computed without expensive numerical integration, and superimposing continuous molecular fields also avoids the problem of lacking one-to-one atom correspondences between the molecules. In addition, mask parameters are introduced in the model so that partial matching of molecules can be carried out. A generalization for the simultaneous alignment of multiple molecules is also proposed.

Our methodology is illustrated with a simulation study and then appliedto a dataset of 31 steroid molecules, where the relationship between shape and binding activity to the corticosteroid binding globulin receptor is explored.

Coauthors: Ian L. Dryden and Christopher J. Brignell

⁂

Some theoretical bounds on Bayesian estimation of evolutionary distance

Scott SCHMIDLER (Duke University, USA)

We provide bounds on the estimability of evolutionary distances under Markov evolution processes, relevant to Bayesian phylogenetic reconstruction. Based on the observation that independent-site processes trivially exhibit the Markov chain “cutoff phenomenon” of Diaconis, we show that these bounds provide a quantitative version of the phase transition results of Mossel and colleagues that can be used in practical analyses.

⁂

15:00–16:30 Room D **Bayesian Econometrics II**

Organizer: Roberto Leon-Gonzalez (National Graduate Institute for Policy Studies, Japan)
Chair: Rodney Strachan (Australian National University, Australia)

Consumption decisions in markets with limited accessibility: the case of cannabis

Liana JACOBI (University of Melbourne, Australia)

Among illicit drugs marijuana is the most common, where the US government spends upwards of $7.7 billion per year in enforcement of the laws for marijuana sales (Miron, 2005). Over the past 30 years there has been a debate regarding whether marijuana should be legalized. While previous literature has examined the impact of decriminalization on marijuana use, studies to date have not disentangled the impact of limited accessibility from consumption decisions based solely on preferences. However, this distinction is particularly important. If most individuals do not use because they do not know where to buy it, but would otherwise use, we would see a large increase in consumption ceteris paribus. On the other hand, if accessibility plays little role in consumption decisions, then making drugs more readily available would impact the supply more. In order to access the impact of legalization on use, it is necessary to explicitly consider the role played by accessibility in use, the impact of illegal actions in utility, as well as the impact on the supply side. In this paper, we develop and estimate a model of buyer behavior that explicitly considers the impact of illegal behavior on utility as well as the impact of limited accessibility (either knowing where to buy or being offered) an illicit drug on using the drug. We use the demand side estimates to conduct counterfactuals on how use would change under a policy of legalization under different assumptions, such as tax policies on the price of cannabis.

Coauthors: Michelle Sovinsky

❉

Time varying dimension models

Rodney STRACHAN (Australian National University, Australia)

Time varying parameter (TVP) models have enjoyed an increasing popularity in empirical macroeconomics. However, TVP models are parameter-rich and risk over-fitting unless the dimension of the model is small. Motivated by this worry, this paper proposes several Time Varying Dimension (TVD) models where the dimension of the model can change overtime, allowing for the model to automatically choose a more parsimonious TVP representation, or to switch between different parsimonious representations. Our TVD models all fall in the category of dynamic mixture models. We discuss the properties of these models and present methods for Bayesian inference. An application involving US inflation forecasting illustrates and compares the different TVD models. We find our TVD approaches exhibit better forecasting performance than many standard benchmarks and shrink towards parsimonious specifications.

Coauthors: Gary Koop and Roberto Leon-Gonzalez

❉

Fat-tailed Gamma autoregressive processes for stochastic volatility with jumps

Roberto LEON-GONZALEZ (National Graduate Institute for Policy Studies, Japan)

In times of macroeconomic or financial turmoil volatility typically jumps suddenly to extreme values. We develop a distribution for volatility that allows for fat tails and that generalizes the Gamma Autoregressive Process. We show that this specification allows for a simple and efficient MCMC algorithm. By conditioning on some auxiliary variables, all volatilities can be sampled jointly from the joint conditional posterior. Because of its simplicity and efficiency, the MCMC algorithm can be used to tackle long time series. In applications to real data we show that the extension to fat tails is empirically relevant.

❉

Efficient samplers for dynamic factor models with stochastic volatilities

Joshua CHAN (Australian National University, Australia)

In recent years, dynamic factor models have become increasingly popular for modeling high-dimensional macroeconomic and financial data. Research in this area has found that allowing for stochastic volatility in these models improves in-sample fitting and forecasting performance. Estimating these high-dimensional latent variable models, however, proves to be challenging. Conventional MCMC sampling schemes typically construct Markov chains that exhibit slow mixing properties and high autocorrelation. In view of this, we consider various efficient samplers based on recent advances in precision-based methods for estimating state space models. We illustrate the efficiency of the proposed algorithms with a finance application involving equity returns.

❉

17:00–18:30 Room A **Bayesian analysis of astronomical data**

Organizer: Joseph M. Hilbe (Arizona State University, USA)
Chair: David van Dyk (University of California, Irvine, USA)

Bayesian hierarchical modeling of cosmological supernovae type Ia data

Roberto Trotta (Imperial College London, UK)

Supernovae type Ia (SNIa) are standardazible candles that can be used to determine distances in the Universe and hence to learn about the expansion history of the cosmos, as well as to detect and characterize its properties, for example dark energy.

In this talk I will present a new method based on a Bayesian hierarchical model to extract cosmological parameters from SALT-II lightcurve fits of SNIa data. I will demonstrate that this new, fully Bayesian method outperforms the standard approach 90% of the time, delivering better cosmological constraints and smaller statistical bias. I will present the ensuing cosmological parameters constraints and discuss applications of the method to learn about the characteristics and possible evolution of the SNIa population.

Coauthors: Marisa March, Pietro Berkes, Glenn Starkman and Pascal Vaudrevange

⁂

Using Bayesian networks for real-time classification of transients

Ashish Mahabal (California Institute of Technology, USA)

With an ever growing number of transients being detected per night, the various ongoing and forthcoming synoptic surveys provide a unique opportunity to explore the variable sky. A majority of these belong to fairly well understood classes on which one need not waste the scarce follow-up resources. As a result selecting which transients to follow becomes more and more critical for understanding newer and/or rarer classes. The inputs are diverse and not easy to make a cohesive sense of especially when one is interested in the classification in as close to real-time as possible. We will present the use of priors and Bayesian networks using them to do quick characterization which can be gradually refined as new data become available. We will demonstrate the performance using Catalina Realtime Transient Survey (CRTS) transients and auxiliary data from other sources.

Coauthors: Ciro Donalek, George Djorgovski and CRTS Team

⁂

Bayesian analysis of sparse astronomical data matrices

Ninan Philip (St. Thomas College, Kozhencheri, India)

It is often desired to make decisions based on whatever subset is available from a large set of features. This is one step better than the mere ability to handle missing data. The possible number of features may be large, dynamic and there might be redundancies and noise in the feature estimation. If the input features are conditionally independent, the Bayesian framework offers the simplest update rule for dynamic addition of information to estimate the posterior. Though conditional independence is uncommon, it is possible to impose conditional independence on the features if the feature values are binned and constrains are applied on the allowed values of other features when a given feature is having a certain bin value. The application of the method to the automated classification of transients detected by the Catalina Real-Time Transient Survey (CRTS) using an implementation of the concepts in a Difference Boosting Neural Network (DBNN) is presented.

Coauthors: Sheelu Abraham, Ashish Mahabal and CRTS Team

⁂

Use of marked point processes in a Bayesian framework for detecting and characterizing cosmic galactic filaments

Radu S. STOICA (Université Lille 1, France)
Vincent MARTINEZ (University of Valencia, Spain)
Enn SAAR (Tartu Observatory, Estonia)

The main feature of the spatial large-scale galaxy distribution is an intricate network of galaxy filaments. Although many attempts have been made to quantify this network, there is no unique and satisfactory recipe for that yet.

The present work uses a Bayesian framework to compare the filaments in the real data and in the numerical models, to see if our best models reproduce statistically the filamentary network of galaxies.

The Bisous model - a marked point process with interactions - is applied to trace and describe the filamentary network both in the observed samples (the 2dFGRS catalogue) and in the numerical models that have been prepared to mimic the data. We compare the networks.

The filamentary network properties are caught by the sufficient statistics of the model. We find that the properties of filaments in numerical models (mock samples) have a large variance. Some mock samples display filaments that resemble the observed filaments, but frequently the model filaments are much shorter and do not form an extended network.

We conclude that although we can build numerical models that are similar to observations in many respects, they fail yet to explain the global structure seen in the data. The Bisous-built filaments are a good test for such a structure.

Coauthors: Vicent J. Martinez and Enn Saar

⁂

17:00–18:30 Room B **Recent advances in Bayesian variable selection**

Organizer: Joyee Ghosh (University of Iowa, USA)
Chair: Philip Brown (University of Kent, UK)

Structuring dependence in regression: radius mixtures of spherically uniform priors

Christopher HANS (Ohio State University, USA)

We investigate prior distributions that are designed to incorporate information about the strength of a regression relationship. The most commonly-used prior distributions for regression models typically assume that coefficients are a priori independent or induce dependence via the empirical design matrix. While these standard priors (and recently-refined versions of them) may exhibit desirable behavior with respect to targeted inferential goals, we should not expect them to distribute probability throughout the entire parameter space in a way that is consistent with all of our prior beliefs. Examination reveals that when we focus on the strength of the regression relationship, standard priors place nearly all of their mass in regions of the parameter space that are not only inconsistent with reasonable prior belief but are nearly certain to clash so greatly with the likelihood that we should question the validity of particular inferences.

We describe a new class of priors that allows one to directly incorporate information about the strength of the regression relationship. We compare the Bayesian model uncertainty properties of our priors with those of standard priors, highlighting the consequences of inappropriately ignoring prior information when it is indeed available, and highlighting the consequences of unintentionally incorporating strong prior information when it does not exist. We describe MCMC algorithms that scale well with model size and require minimal storage by using a fixed-dimensional parameterization across models of different sizes. We discuss several strategies for improving MCMC output-based estimation using the structure of the posterior.

Coauthors: Steven MacEachern and Agniva Som

⁂

A stochastic partitioning method to associate high-dimensional datasets

Mahlet TADESSE (Georgetown University, USA)

In recent years, there has been a growing interest in relating data sets in which both the number of regressors and response variables are substantially larger than the sample size. For example, in the context of genomic studies, a common goal is to identify groups of correlated gene expression levels that are modulated by sets of DNA sequence variations. This may give insights into molecular processes underlying various phenotypes. We propose a Bayesian stochastic partitioning method that combines ideas of mixtures of regression models and variable selection methods to identify cluster structures and relationships across high-dimensional data sets. We illustrate the method with applications to genomic studies.

Coauthors: Stefano Monni

⁂

The multiset model selector

Scotland LEMAN (Virginia Tech, USA)

Multiset Sampling (Leman et. al., 2009) has previously been developed for efficient sampling from complex stochastic processes. We extend the sampler and surrounding theory to high dimensional model selection problems. In such problems, efficient exploration of the model space is oftentimes a challenge since independent and ad-hoc proposals might not be able to jointly propose parameter sets which correctly explain a new proposed model. In order to overcome this issue, we apply multiset sampling to the model space in order to efficiently explore complex model structures. The Multiset Model Selector (MSMS) is general and can be applied to a multitude of modeling frameworks. Under this method we do not directly obtain typical Bayesian model averaged estimates for the parameters; however, we show that the multiset parameter estimates are a mixture distribution from which the true Bayesian model probabilities and model averaged parameter estimates can be obtained. We explore the effects of model dimensionality, correlation in the predictor space, and size of the multiset on the computational complexity and efficiency of the model search.

Coauthors: Dipayan Maiti

* * *

Data augmentation and sandwich algorithms for Bayesian model averaging

Joyee GHOSH (University of Iowa, USA)

Several variants of Markov chain Monte Carlo algorithms have been proposed in the literature for sampling models from the posterior distribution in Bayesian model averaging. Our focus is on the data augmentation algorithm of Ghosh and Clyde, which augments the parameter space to exploit the properties of orthogonal design matrices in Bayesian variable selection and Bayesian model averaging. This augmented data approach leads to a Gibbs sampler with a fast updating scheme and the availability of Rao-Blackwellized estimates of quantities of interest. We propose a sandwich algorithm based on permutation groups which is shown to improve upon the original data augmentation algorithm. We illustrate via simulation studies that the sandwich algorithm can lead to substantial gains in the presence of high multicollinearity, with a small increase in computational cost. This is joint work with Aixin Tan and James Hobert.

Coauthors: Aixin Tan and James Hobert

* * *

17:00–18:30 Room D **Advances in honest Monte Carlo**

Organizer: Mark Huber (Claremont McKenna College, USA)
Chair: Mark Huber (Claremont McKenna College, USA)

Expectation and quantile estimation via Markov chain Monte Carlo

James FLEGAL (University of California, Riverside, USA)

Calculating a Monte Carlo standard error (MCSE) is an important step in the statistical analysis of the simulation output obtained from a Markov chain Monte Carlo experiment. For example, it can be used to provide a rigorous method for terminating the simulation. An MCSE is usually based on an estimate of the variance of the asymptotic normal distribution. We consider spectral, batch means, and subsampling bootstrap methods for estimating this variance. Finally, we investigate the finite sample properties through two examples and provide recommendations to practitioners.

Coauthors: Galin Jones

* * *

Fast approximation algorithms for partition functions of Gibbs distributions

Mark HUBER (Claremont McKenna College, USA)

A new algorithm for estimating the partition function of a Gibbs distribution is presented. The method combines a well balanced cooling schedule created through TPA and a product importance sampler. One advantage of the algorithm over existing methods is the standard deviation of the estimate can be bounded theoretically. The number of samples necessary to build a close estimate grows almost linearly in the logarithm of the partition function, making the approach suitable for high dimensional problems. The samples needed for the estimate can be generated rapidly by methods such as parallel tempering.

* * *

Markov chain Monte Carlo: can we trust the third significant figure?

Galin JONES (University of Minnesota, USA)

Calculating a Monte Carlo standard error (MCSE) is an important step in the statistical analysis of the simulation output obtained from a Markov chain Monte Carlo experiment. An MCSE is usually based on an estimate of the variance of the asymptotic normal distribution. We show that batch means, spectral methods and regenerative simulation yield consistent estimators under much weaker conditions than previously obtained. We focus on the practical consequences of these theoretical results.

Coauthors: Krzysztof Łatuszyński

❉

Sequential Monte Carlo methods in high dimensions

Alexandros BESKOS (University College London, UK)

Sequential Monte Carlo (SMC) methods are nowadays routinely applied in a variety of complex applications: hidden Markov models, dynamical systems, target tracking, control problems, just to name a few. Whereas SMC methods have been dramatically improved and refined in the last decades, they are still known to suffer from the curse of the dimensionality: algorithms can sometimes break down exponentially fast with the dimension of the state space. The talk will concentrate on a particular version of SMC and will look at methods that can reduce the asymptotic cost of the algorithms from exponential to quadratic in the dimension of the state space. Novel explicit asymptotic results will clarify the effect of the dimension at the properties of the algorithm and could provide a platform for algorithmic optimisation in high dimensions.

Coauthors: Dan Crisan and Ajay Jasra

❉

Thursday, June 28th

09:00–10:30 Room B **Bayesian graphical and factor models: structure, sparsity and dimension**

Organizer: Mike West (Duke University, USA)
Chair: Emily Fox (University of Pennsylvania, USA)

Bayesian sparse reconstruction: Latent factor analysis of gene regulatory programs

Ryo YOSHIDA (Research Organization of Information and Systems, Institute of Statistical Mathematics, Japan)

Cells can sense and respond to many different cellular signals by switching activation or inactivation of specific biochemical molecules via several types of regulatory circuits. Robustness of cellular functions to environmental changes, damages, reaction failures and kinetic perturbations is an essential property of biological systems. Understanding the network structures, motifs and design principles underlying biological robustness is necessary for an in-depth understanding of cells at the system level. This study aims to reconstruct, in systematic ways, such robust regulatory circuits, i.e., graphical structures, from time series measurements on molecular abundances or concentrations. The method involves a hierarchical latent factor models based on ordinary differential equations, and a newly-derived evolutionary MCMC algorithms for the stochastic graph search problem.

❉

Modeling sparse full-rank orthogonal matrices

Andrew CRON (Duke University, USA)

We present a novel method for modeling square orthogonal matrices. By decomposing the orthogonal matrix using Givens rotations, we can incorporate sparsity and parsimony via model selection while maintaining full rank and orthogonality. Furthermore in the context of Gaussian covariance estimation, our model also has some interesting properties on the implied graphical model. Model fitting is demonstrated in an approximate exploratory algorithm and a fully Bayesian reversible jump Markov chain Monte Carlo algorithm. The effectiveness of our model is shown in a gene expression for breast cancer classification example where we incorporate our model into a Gaussian mixture model.

Coauthors: Mike West

❉

Efficient Gaussian graphical model determination under G-Wishart prior distributions

Hao WANG (University of South Carolina, USA)

This paper proposes a new algorithm for Bayesian model determination in Gaussian graphical models under G-Wishart prior distributions. We first review recent development in sampling from G-Wishart distributions for given graphs, with a particular interest in the efficiency of the block Gibbs samplers and other competing methods. We generalize the maximum clique block Gibbs samplers to a class of flexible block Gibbs samplers and prove its convergence. This class of block Gibbs samplers substantially outperforms its competitors along a variety of dimensions. We next develop the theory and computational details of a novel Markov chain Monte Carlo sampling scheme for Gaussian graphical model determination. Our method relies on the partial analytic structure of G-Wishart distributions integrated with the exchange algorithm. Unlike existing methods, the new method requires neither proposal tuning nor evaluation of normalizing constants of G-Wishart distributions.

Coauthors: Sophia Zhengzi Li

❉

Sparse factor model approaches to the weak instrument problem

P. Richard HAHN (University of Chicago Booth School of Business, USA)

We adapt sparse factor models to the causal inference setting, specifically the instrumental variables linear model. Sparse factor priors prove beneficial in the common but challenging situation where the instruments are only weakly correlated with the focal predictor variable. We discuss the relative benefits of various parametrizations of the model in terms of regularization properties and computational feasibility.

Coauthors: Hedibert Lopes

❉

09:00–10:30 Room C **Bayesian methods in biological, environmental and ecological systems**

Organizer: Debashis Mondal (University of Chicago, USA)
Chair: David Dahl (Texas A&M University, USA)

Statistical issues in nanometerial toxicology

Donatello TELESCA (University of California, Los Angeles, USA)

The development of high throughput screening (HTS) assays in the field of nanotoxicology provides new opportunities for the hazard assessment of engineered nanomaterials (ENM). The statistical challenge associated with hazard assessment from HTS data lies in its richness and heterogeneity as multi-dimensional measurements are often taken over a small number of replicates with relatively low signal. We discuss how multivariate Bayesian analysis can efficiently address some of the statistical challenges associated with nanomaterial toxicology and provide a unified framework for toxicity profiling, hazard ranking and the analysis of ENM quantitative structure activity relationships.

Coauthors: Cecile Low-Kam and Trina Patel

❉

Modelling heterogeneity in mark-recapture data using the Dirichlet process

Saman MUTHUKUMARANA (University of Manitoba, Canada)

A main assumption in modelling mark-recapture data is that survival probabilities are homogeneous. We treat survival probabilities as a function of two parameters that explain variations due to unknown biological and environmental reasons. The heterogeneity in travel times and survival probabilities is accounted using the Dirichlet process. The approach is highlighted using actual data arising from the Pacific Ocean Shelf Tracking (POST) project. Results from the data analyses and log-pseudo marginal likelihood (LPML) model selection procedure indicate that DP model perform better over alternative methods.

❉

Directional weights CAR models using Gaussian process mixing

Veronica J. Berrocal (University of Michigan, USA)

Areal or lattice data is usually analyzed by specifying a likelihood model that introduces spatial random effects provided with a conditionally autoregressive (CAR) prior. In a CAR specification the weighting scheme used to derive the full conditionals is constant over space and fixed a priori to be either binary or inversely proportional to the distance among subregions/cells. In this paper, we propose a more general class of CAR models where the weights are non-constant and are random variables obtained by appropriately transforming a latent Gaussian process. The resulting class of CAR models is flexible and generalizes the class of fixed-weight CAR models which arises as a special case. Distributional properties of the weights and of the spatial random effects can be derived. As an illustration, we have applied this general class of CAR models in the context of image restoration and disease mapping.

Coauthors: Alan E. Gelfand

❈

Spatial analysis of environmental bioassays

Debashis Mondal (University of Chicago, USA)

Recent years have seen an increased research focus on environmental toxicity sparked by concerns related to watershed protection in lakes, rivers, estuaries and coastal areas. Environmental bioassays are now routinely conducted in large numbers in order to determine adverse biological effects of contaminants present in soil, water and sediments across a geographic region under study. However, till date, statistical analyses of these environmental bioassays have been naive and limited, and, thus, require further attention. For example, the result of Ampeisca abdita toxicity test is often summarized by percentage of ratios of survivals in cases and controls, and toxicity is determined by whether or nor this percentage is below eighty. In this talk, I present some critical statistical evaluations of such bioassays, and propose spatial modeling and Bayesian computations to borrow strength from nearby bioassays values. I show how spatial models allow us to obtain maps and predictions of toxicity levels across the watershed by making explicit use of the geographic distribution of sampling locations.

❈

11:00–12:30 Room A **Approximate Bayesian computation (ABC): likelihood-free Bayesian inference II**

Organizers: Christian Robert (Université Paris Dauphine, France)
Scott Sisson (University of New South Wales, Australia)
Chair: Marc Suchard (University of California, Los Angeles, USA)

Approximate Bayesian computation and Bayes linear analysis: towards high-dimensional ABC

David Nott (National University of Singapore, Singapore)

Bayes linear analysis and approximate Bayesian computation (ABC) are techniques commonly used in the Bayesian analysis of complex models. It is possible to connect these ideas in a very simple way by demonstrating that regression-adjustment ABC algorithms produce samples for which first and second order moment summaries approximate adjusted expectation and variance for a Bayes linear analysis. This gives regression-adjustment methods a useful interpretation and role in exploratory analysis in high-dimensional problems. As a result, we propose a new method for combining high-dimensional, regression-adjustment ABC with lower-dimensional approaches (such as using MCMC for ABC). This method first obtains a rough estimate of the joint posterior via regression-adjustment ABC, and then estimates each univariate marginal posterior distribution separately in a lower-dimensional analysis. The marginal distributions of the initial estimate are then modified to equal the separately estimated marginals, thereby providing an improved estimate of the joint posterior. We illustrate this method with several examples.

Coauthors: Yanan Fan, Lucy Marshall and Scott Sisson

❈

A comparison of dimension reduction methods in approximate Bayesian computation

Michael Blum (CNRS, Université Joseph Fourier, France)

Approximate Bayesian computation relies on comparisons between observed summary statistics and simulated ones. As typical in statistics, there is a tradeoff between having too many and too few summary statistics. We provide in this talk a comparative review of the main ABC methods that address the problem of dimension reduction for ABC inference. In addition, we propose two original approaches for dimension reduction. The first approach considers statistical information criteria to find an optimal subset of statistics. The second approach relies on a ridge regression layer that penalizes uninformative summary statistics.

Coauthors: M. A. Nunes, D. Prangle and S. A. Sisson

⁂

Estimation of demo-genetic model probabilities with approximate Bayesian computation using linear discriminant analysis on summary statistics

Jean-Michel Marin (Université Montpellier 2, France)

Comparison of demo-genetic models using Approximate Bayesian Computation (ABC) is an active research field. Although large numbers of populations and models (i.e. scenarios) can be analysed with ABC using data obtained from various marker types, methodological and computational issues arise when such numbers becomes too large. Moreover, Robert et al. (2011) have shown that the conclusions drawn on ABC model comparison cannot be trusted per se and required further simulation analyses. However, when the number of summary statistics (Ss) to be manipulated is large, inferential techniques used to compute and validate the ABC estimates are very time consuming. We here describe a methodological innovation to process efficient ABC scenario probability computation using linear discriminant analysis (LDA) on Ss before computing logistic regression. We used controlled simulated pseudo-observed data sets (pods) to assess the main features of the method (precision and computation time) in comparison to more traditional probability estimation using raw (i.e. not LDA-transformed) Ss. We also illustrate the method on real microsatellite data sets produced to make inferences about the invasion routes of the coccinelid Harmonia axyridis. We found that scenario probabilities computed from LDA-transformed and raw Ss were strongly correlated. Type I and II errors were similar for both methods. The faster probability computation that we observed (speed gain around a factor 100 for LDA-transformed Ss) substantially increases the ability of ABC practitioners to analyze large numbers of pods and hence provides a decent way to empirically evaluate the power to discriminate among a set of complex scenarios.

Coauthors: A. Estoup, E. Lombaert, T. Guillemaud, P. Pudlo, C. P. Robert and J.-M. Cornuet

⁂

Summary errors, conflict and evidence synthesis - exploiting the cutting edges of approximate Bayesian computation to analyze the phylodynamics of human pathogens

Oliver Ratman (Duke University, USA)

The infectious disease dynamics of many human pathogens like influenza, norovirus and coronavirus are inextricably tied to their evolution. Large data sets on disease incidence and the molecular genetic evolution of the virus are readily available, but difficult to combine with existing likelihood-based methods. Taking interpandemic human influenza subtype A(H3N2) as an example, we discuss how ABC enables us to interface complex phylodynamic models that simulate pathogen evolution and ecology with both types of data. Summary errors can be used at no extra computational cost to diagnose specifically in which direction models on the interaction of disease ecology and evolution are inconsistent with the data taken together. Whenever the evolution and ecology of rapidly evolving pathogens are linked, case report and phylogenetic summaries are co-dependent. We exploit these co-dependencies to test for conflict between summaries, thereby using the data synergistically. Phylodynamic inference and hypothesis testing has long been contemplated, and the nature of ABC offers unique opportunities to address unresolved phylodynamic questions in a Bayesian spirit.

Coauthors: Gé Donker, Adam Meijer, Christophe Fraser and Katia Koelle

⁂

11:00–12:30 Room C **Bayesian model assessment**

Organizer: Joris Mulder (Tilburg University, Netherlands)
Chair: Brad Carlin (University of Minnesota, USA)

Perspectives on Bayesian model combination

Merlise CLYDE (Duke University, USA)

Consideration of multiple models is routine in statistical practice. With computational advances over the past decade, there has been increased interest in methods for making inferences based on combining models. Examples include boosting, bagging, stacking, and Bayesian Model Averaging (BMA), which often lead to improved performance over methods based on selecting a single model. Bernardo and Smith have described two Bayesian frameworks for model selection known as the M-closed and M-open perspectives. The standard formulation of Bayesian Model Averaging arises as an optimal solution for combining models in the M-closed perspective where one believes that the “true” model is included in the list of models under consideration. In the M-open perspectives the “true” model is outside the space of models to be combined, so that model averaging using posterior model probabilities is no longer applicable. Using a decision theoretic approach, we present optimal Bayesian solutions for combining models in both frameworks and contrast this forumation with frequentists methods and Bayesian Ensemble methods. Finally we discuss asymptotic properties of the M-open solutions.

❈

Identifiability and learning in Bayesian joint longitudinal-survival models

Laura HATFIELD (Harvard Medical School, USA)

In studying biological processes, investigators may repeatedly measure features of the process (longitudinal data) and also measure the time until some event (survival data). For example, a clinical trial may measure symptom severity and time until death. A broad class of joint models for simultaneously analyzing longitudinal and time-to-event data has been developed. The most popular such models use latent variables to link longitudinal and survival submodels. These approaches have expanded to accommodate many data complexities, yet little attention has been paid to these approaches’ properties and performance.

To quantify the benefit of joint versus separate modeling, we derive closed-form expressions for posterior quantities in a simplified normal-lognormal joint model. We show that as the prior variance on fixed effects increases, the resulting posterior means and variances from this joint model and either of its submodels alone converge to the same limit. We use a single latent variable to link the two submodels and thus requirea coefficient to establish the magnitude and direction of its contribution to one of the submodels. The posterior for this scaling parameter may exhibit two modes symmetric about the origin when the sample size is small, indicating weak identification of its sign. We argue that such counterintuitive results in a very simple joint model imply great care should be taken when using complex joint models in applied settings.

Coauthors: James Hodges and Bradley Carlin

❈

On the development of formal criteria to determine objective model selection priors

Gonzalo GARCÍA-DONATO (University of Castilla-La Mancha, Spain)

In a non-informative a priori setting, Bayes factors are very sensitive to the prior distribution (here called model selection prior). Further, it is well known that neither objective estimation priors (e.g reference or Jeffreys) nor proper vague priors can in general be used. In this context, many different mathematical rules have been proposed to derive convincing objective priors. Unfortunately, the resulting Bayes factor can substantially differ among the different priors and the Bayesian community has adopted none of them as a standard approach. Nevertheless, a usual practice is to evaluate the goodness of the priors by broadly agreed criteria (like consistency, predictive matching, etc). Here, we explore the potential of such criteria as a mechanism to determine objective model selection priors with high degree of consensus. We first formalize the most general and compelling criteria that have been suggested, together with a new criterion. Then we illustrate the methodology in a number of testing problems.

Coauthors: Jim Berger, Susie Bayarri and Anabel Forte

❈

Default Bayes factors for comparing (in)equality constrained models

Joris MULDER (Tilburg University, Netherlands)

In practice, researchers often formulate their expectations using equality and/or inequality constraints on the parameters of interest. This results in a set of competing (in)equality constrained models (or hypotheses). The goal is then to determine which model is most supported by the data. Default Bayes factors, such as the intrinsic Bayes factor (IBF) or the fractional Bayes factor (FBF), are not recommended for this problem because they are not consistent when evaluating nested inequality constrained models. This is a result of the fact that the corresponding default priors of the IBF and FBF are centered around the likelihood of the data. A sufficient condition for the prior to obtain consistent Bayes factors is that it is centered on the boundary of the constrained parameter space. For this reason, a default Bayes factor is proposed that is based on the FBF of O'Hagan (1995) with the modification that the default prior is centered on the boundary of the constrained parameter space under investigation. The proposed criterion is referred to as the prior adjusted FBF. It is also effective for small data sets; it satisfies several principles for coherence; and it converges fastest to the true inequality constrained model.

⁂

15:00–16:30 Room A **Case studies of Bayesian success stories: babies, trials and ratings**

Organizer: Peter Mueller (University of Texas at Austin, USA)
Chair: Peter Mueller (University of Texas at Austin, USA)

Modeling at-risk infants: Bayesian nonparametrics to the rescue

Suchi SARIA (Harvard University and Johns Hopkins University, USA)

Physiological data are routinely recorded in intensive care, but their use for rapid assessment of illness severity has been limited. The data is high-dimensional, noisy, and changes rapidly; moreover, small changes that occur in a patient's physiology over long periods of time are difficult to detect, yet can lead to catastrophic outcomes. A physician's ability to recognize complex patterns across these high-dimensional measurements is limited.

We propose a hierarchical Bayesian model that employs Bayesian nonparametrics for discovering informative representations in such continuous time series. This model aids both exploratory data analysis and feature construction. When applied to data from premature infants in the neonatal ICU (NICU), our model obtains novel clinical insights. Based on these insights, we devised the Physiscore, a novel risk prediction score that combines patterns from continuous physiological signals to predict infants at risk for developing major complications in the NICU. Using only 3 hours of non-invasive data from birth, Physiscore very successfully predicts morbidity in preterm infants. Physiscore performed consistently better than other neonatal scoring systems, including the Apgar, which is the current standard of care, and SNAP, a machine learning based score that requires multiple invasive tests. This work was published on the cover of Science Translational Medicine (Science's new journal aimed at translational medicine work), and was covered by numerous press sources.

Coauthors: Daphne Koller and Anna Penn

⁂

Success stories in Bayesian adaptive methods for phase I-II clinical trials

Brad Carlin (University of Minnesota, USA)

Bayesian methods have a long history of success in clinical trial settings where patients and other resources are scarce, where good reliable external information is available, or both. In this talk we review several real-life settings where Bayesian methods have paid such dividends. First, we describe an adaptive Bayesian design for ensuring 95medical device at 5 years, where interim posterior predictive distributions are used to decide whether to stop patient accrual. Next, in the area of Phase I oncology trials, we describe a trivariate binary model for the probabilities of toxicity, efficacy, and surrogate efficacy given a particular dose. An asymmetric loss function permits differential penalties for over- and under-dosing, and we also consider semiparametric extensions to the usual logistic dose-response curve, which may be too smooth and simple for our data. Next, we describe a novel Bayesian adaptive trial design that incorporates multi-trial historical information on the relationship between a surrogate and a clinical endpoint. If the joint outcomes in the new trial seem plausible given similar historical trials, we proceed with the surrogate endpoint (perhaps stopping the trial early for treatment success, inferiority, or futility); otherwise, we discard the surrogate and switch back to the original primary endpoint. Finally, we describe recent Bayesian advances in the adaptive incorporation of historical information in clinical trials, showing connections with traditional meta-analytic methods and illustrating their potential for improved power and ethical patient treatment (e.g., by permitting randomization of fewer patients to the inferior treatment). While like Nebraska's Platte River this talk will be "a mile wide and an inch deep," all of our success stories will be illustrated in the context of real trial settings arising in both academics and industry.

Coauthors: Thomas Murray, Wei Zhong, Joseph, Lindsay Renfro, Daniel Sargent and Brian Hobbs

⁂

Clustering, network, epigenetics

Yuan Ji (Northshore University Healthcare System, USA)

In this talk, I will describe how Bayesian models are successfully applied to the field of epigenetics, which is concerned about regulatory mechanism of gene expression. Epigenetics, one of the most heavily researched and challenging field in biology, increasingly draws attention from statistic ans due to breakthroughs in bioengineer and biotechnology that allow large-scale and high-throughput experiments to be routinely conducted with affordable resources. A central topic of epigenetics is to understand the chromatin state – modifications to histones and other proteins that package the DNA. A complex mechanism called "histone code" is believed to dictate the dynamics of DNA expression. As a step towards deciphering the histone code, we develop Bayesian parametric and nonparametric models based on genome-wide mapping of histone modification data. Such models are only initial attempts to decipher the complex histone code but highlight the need of Bayesian inference in the research of gene regulations, receiving relatively small amount of attention from statisticians. I will summarize our recent work and results using a comprehensive ChIP-Seq data set.

Coauthors: Juhee Lee, Shoudan Liang, Riten Mitra and Peter Mueller

⁂

Measuring competitor strength in games and sports through an approximate Bayesian filter: the Glicko system

Mark Glickman (Boston University, USA)

For many years, the only serious algorithm for tracking player strength over time in organized competitive games such as tournament chess was a system developed by Arpad Elo in the early 1960s. This algorithm was adopted with minor modifications by various national and international gaming organizations throughout and 1960s and 1970s, and is still in widespread use today. Despite its popularity, one major drawback of the Elo system is that its connections to probability models is ad hoc at best, so the resulting ratings from the system are difficult to interpret. In the mid-1990s, I developed an approximate Bayesian algorithm that was a linearization of a reasonably standard Bayesian time series model for binary outcomes. This new algorithm, conventionally known as the Glicko system, overcomes many of the difficulties with the Elo system. In fact, the Elo system can be seen as a special case of the Glicko system in which parameter estimates have no posterior uncertainty. The Glicko system is now used by many gaming organizations, such as the International Wargames Federation, the Italian Othello Federation, and the Australian Chess Federation, along with many online gaming organizations. In this talk, I describe the basic development of the Glicko system, its advantages over the Elo system, and its adoption by various gaming organizations over the years.

⁂

15:00–16:30 Room B **Parallel processing in Bayesian computing**

Organizers: Christian Robert (Université Paris Dauphine, France)
Marc Suchard (University of California, Los Angeles, USA)
Chair: Christian Robert (Université Paris Dauphine, France)

Ridiculously parallel, serial Bayesian inference algorithms

Marc SUCHARD (University of California, Los Angeles, USA)

Following a series of high-profile drug safety disasters in recent years, many countries are redoubling their efforts to ensure the safety of licensed medical products. Large-scale observational databases such as claims databases or electronic health record systems are attracting particular attention in this regard, but present significant methodological and computational concerns. Likewise, fusion of real-time satellite data with in situ sea surface temperature measurements remains taxing for probabilistic spatial-temporal models on a global scale. In this talk, I discuss how high-performance statistical computation, including graphics processing units, can enable complex Bayesian methods in these large datasets. I focus on algorithm restructuring through techniques like block relaxation (Gibbs, cyclic coordinate descent, MM) to exploit increased data/parameter conditional independence within traditional serial algorithms. I find orders-of-magnitude improvement in overall run-time fitting models involving tens of millions of observations. These approaches are ubiquitous in Bayesian statistics and I conclude with a seemingly (but not) unrelated example using an infinite hidden Markov model to explore the genomic evolution of influenza.

Coauthors: Subha Guha, Ricardo Lemos, David Madigan, Bruno Sanso and Steve Scott

⁂

Distributed hierarchical logistic regression

Steven SCOTT (Google, USA)

Markov chain Monte Carlo methods are inherently sequential, but they are often applied to models with strong conditional independence relationships. The most computationally intensive sub-steps in an MCMC chain are often locally embarrassingly parallel in the sense that the work involved can be carried out by separate machines, but the machines must communicate before the next MCMC step is executed. The communication is a performance bottleneck that limits the overall speed of the sampling algorithm. This talk describes an implementation of hierarchical logistic regression in a distributed cloud environment and illustrates the size of the network performance bottleneck as the number of machines grows.

⁂

Rejection sampling on a graphical processing unit

Jarad NIEMI (Iowa State University, USA)

Rejection sampling is a basic tool of Monte Carlo sampling, easily implemented in many situations but often not useful due to extremely low acceptance rates. Parallelizing obviates this practical issue by simultaneously generating a large number of candidate samples. Recent advances in graphical processing unit (GPU) technology has made massive parallelization available to many end users at little cost. A unique aspect of rejection sampling on a GPU is that the number of required samples is random. We utilize dynamic creation of a Mersenne-Twister family of generators to create independent streams of pseudorandom numbers avoiding jump-ahead techniques. We illustrate the methodology on a data augmentation Markov chain Monte Carlo algorithm for inference in stochastic chemical kinetic system based on Michaelis-Menton kinetics. We show that the efficiency of the GPU vs a single-core CPU increases as the acceptance probability decreases and has a maximum around 200-fold speedup in this application.

Coauthors: Matthew Wheeler

⁂

GPGPU parallel computing for Bayesian portfolio selection with massive number of assets

Kenichiro McALINN (Keio University, Japan)

Bayesian methods that utilize subjective views of investors have proven to be successful in solving a range of problems in finance, from portfolio selection to program trading. Regardless of the mounting supportive evidence for Bayesian methods, they have yet to suffice for practical use in everyday decision making by investors because of their time consuming nature. With the recent development of fast and inexpensive devices for parallel computing such as general purpose graphic processing units (GPGPU), however, formerly impractical computations that take hours or days can be completed in minutes or even seconds. With this new paradigm of parallel computing in mind, we investigate whether the use of GPGPU can ease the burden of computationally intensive Bayesian methods, in particular Bayesian portfolio selection with massive number of assets. In our parallel computing approach, we simultaneously perform multiple runs of Gibbs sampling for the factor model of asset returns and compute posterior statistics necessary for Bayesian portfolio selection. In numerical experiments, we find that parallel computing by GPGPU substantively reduces computing time for portfolio selection, compared with the traditional CPU, making the process realistic for any standard. We also demonstrate that Bayesian methods far surpass other more conventional frequentist methods; when dealing with massive asset portfolios.

Coauthors: Teruo Nakatsuma

⁂

15:00–16:30 Room D **Bayesian Econometrics III**

Organizer: Tomohiro Ando (Keio University, Japan)
Chair: Tomohiro Ando (Keio University, Japan)

Factor augmented approach for predicting stock market behavior

Tomohiro ANDO (Keio University, Japan)

Vast amount of variables that represent dynamics of stock market and real economic activity provides useful information for predicting future stock market returns. To implement this idea, we apply a factor augmented predictive regression system approach to incorporate information of thousands of variables. The method intends to capture and predict the behavior of future stock market in terms of expected return, and also tail parts, including 5% and 95% quantile points. We apply this method to the stock return data from Tokyo Stock Exchange and tries to predictor the TOPIX returns. The found evidence for short-horizon predictability was stronger than for log-term horizons.

Coauthors: Saburo Ohno and Ruey Tsay

⁂

Bayesian analysis of identifying restrictions for the time-varying parameter vector autoregressive model

Toshiaki WATANABE (Hitotsubashi University, Japan)

The time-varying parameter vector autoregressive (TVP-VAR) model proposed by Primiceri (2005) assumes that the both parameters and volatilities are time-varying. Since this model assumes the recursive structure for identification, the ordering of variables matters. This article develops a Bayesian method for the ordering of variables for the TVP-VAR model using the reversible jump Markov chain Monte Carlo. The simulation study reveals that the performance of this method depends on how variable the parameters are and does not depend on how variable the volatilities are. The empirical result using the macroeconomic data in Japan during the period from 1981 to 2008 provides evidence that fixing the order as the one which has the highest posterior probability is favored over taking account of order uncertainty and that the introduction of zero interest rate policy may have changed the order of variables. Some other identifying restrictions are also compared with the recursive restriction.

Coauthors: Jouchi Nakajima

⁂

Instrumental variables, errors in variables, and simultaneous equations models: applicability and limitations of direct Monte Carlo

Nalan Basturk (Erasmus University Rotterdam, Netherlands)

A Direct Monte Carlo (DMC) approach is introduced for posterior simulation in the Instrumental Variables (IV) model with one possibly endogenous regressor, multiple instruments and Gaussian errors under a flat prior. This DMC method can also be applied in an IV model (with one or multiple instruments) under an informative prior for the endogenous regressor's effect. This DMC approach can not be applied to more complex IV models or Simultaneous Equations Models with multiple endogenous regressors. An Approximate DMC (ADMC) approach is introduced that makes use of the proposed Hybrid Mixture Sampling (HMS) method, which facilitates Metropolis-Hastings (MH) or Importance Sampling from a proper marginal posterior density with highly non-elliptical shapes that tend to infinity for a point of singularity. After one has simulated from the irregularly shaped marginal distribution using the HMS method, one easily samples the other parameters from their conditional Student-t and Inverse-Wishart posteriors. An example illustrates the close approximation and high MH acceptance rate. While using a simple candidate distribution such as the Student-t may lead to an infinite variance of Importance Sampling weights. The choice betweenthe IV model and a simple linear model under the restriction of exogeneity may be based on predictive likelihoods, for which the efficient simulation of all model parameters may be quite useful. In future work the ADMC approach may be extended to more extensive IV models such as IV with non-Gaussian errors, panel IV, or probit/logit IV.

Coauthors: Arnold Zellner, Tomohiro Ando, Lennart Hoogerheide and Herman K. van Dijk

* * *

Instrumental variables, errors in variables, and simultaneous equations models: applicability and limitations of direct Monte Carlo

Nalan Basturk (Erasmus University Rotterdam, Netherlands)

A Direct Monte Carlo (DMC) approach is introduced for posterior simulation in the Instrumental Variables (IV) model with one possibly endogenous regressor, multiple instruments and Gaussian errors under a flat prior. This DMC method can also be applied in an IV model (with one or multiple instruments) under an informative prior for the endogenous regressor's effect. This DMC approach can not be applied to more complex IV models or Simultaneous Equations Models with multiple endogenous regressors. An Approximate DMC (ADMC) approach is introduced that makes use of the proposed Hybrid Mixture Sampling (HMS) method, which facilitates Metropolis-Hastings (MH) or Importance Sampling from a proper marginal posterior density with highly non-elliptical shapes that tend to infinity for a point of singularity. After one has simulated from the irregularly shaped marginal distribution using the HMS method, one easily samples the other parameters from their conditional Student-t and Inverse-Wishart posteriors. An example illustrates the close approximation and high MH acceptance rate. While using a simple candidate distribution such as the Student-t may lead to an infinite variance of Importance Sampling weights. The choice between the IV model and a simple linear model under the restriction of exogeneity may be based on predictive likelihoods, for which the efficient simulation of all model parameters may be quite useful. In future work the ADMC approach may be extended to more extensive IV models such as IV with non-Gaussian errors, panel IV, or probit/logit IV.

Coauthors: Arnold Zellner, Tomohiro Ando, Lennart Hoogerheide and Herman K. van Dijk

* * *

Bayesian analysis of partially linear regression models

Taeryon Choi (Korea University, Korea)

A partially linear regression model is a semiparametric regression model that consists of parametric and nonparametric regression components in an additive form. In this paper, we consider statistical inference for partially linear regression models using Gaussian process priors. Specifically, a Gaussian process prior with a zero mean function and a suitably chosen covariance function is put on the nonparametric regression component in the partially linear regression model. The posterior inference is performed for estimation as well as hypotheses testing, i.e. selecting appropriate regression models based on Bayes factors. For estimation, posterior distributions are derived for unknown parameters, and numerical schemes are discussed to generate posterior samples from them. For hypotheses testing, we deal with model comparisons between parametric representation and semi/nonparametric ones. We illustrate empirical performance of the proposed model based on synthetic data and real data applications, and investigate the asymptotic issue on consistency.

* * *

17:00–18:30 Room B **Bayesian modeling and its applications in social science**

Organizer: Jun Lu (American University, USA)
Chair: Jeff Gill (Washington University, St. Louis, USA)

Bayesian analysis of variance

Dongchu SUN (University of Missouri-Columbia, USA)

The ANOVA models have been applied in analyzing data from a wide range of areas. We consider two versions of the ANOVA models in literature: the fixed effect model and the random effect model. The main difference of these two types of models lies in whether or not the effects of the factor levels are treated as random. In practice, however, there are many cases where it is not clear whether or not the factor effects should be treated as fixed or random. From the Bayesian point of view, all parameters are considered as random variables, making the distinction between the fixed effect model and the random effect model rather obscure. The primary goal of this talk is to seek a unified Bayesian approach to deal with ANOVA models with fixed effects and random effects. We consider a Normal linear regression models under a class of Zellner's (1986) g-priors. The closed form expression of the marginal likelihood function is derived assuming the commutativity of the projection matrices from the design matrices. As illustrations, the marginal likelihood functions of the balanced m-way ANOVA models with either main effects only or with all interaction effects are calculated using the closed form expression. The commutativity condition is discussed using the tool of orthogonal arrays. In case of one-way ANOVA models, we show that the proposed prior will result in good-consistency properties in terms of model selection and posterior distributions, in the settings of either the fixed effect model or the random effect model.

Coauthors: Xiaoyi Min and Paul Speckman

⁂

Phylogenetic models for the ancestry of a cultural trait with applications to the analysis of vocabulary trait data

Geoff NICHOLLS (University of Oxford, UK)

Nicholls and Gray (JRSSB, 2008) describe a phylogenetic model for binary trait data. They use their model to estimate branching times on a Indo-European language tree from the lexical data of Ringe et al. (Trans. Philological Soc., 2002), dropping seven languages with missing data. We summarise this model and extend it in three stages. First, we give a model for rate heterogeneity in the trait evolution by adding a point process of evolutionary bursts to the background birth-death process. Second, we show how to compute the likelihood when there are trait data missing at random. Third, we consider a model for lateral transfer of traits, in which the tree becomes a network. This is a word borrowing model. We fit the model to the full dataset, and check the fit using Bayes factors to measure support for known constraints on the phylogenetic branch lengths and subtree structures. We reject three of thirty historically attested constraints. Some of the datasets of interest are very large, and it is impractical to load the entire dataset at any given time. We briefly describe MCMC scheme which is new to Statistics, called the Penalty Method. The method is due to Ceperley et al. (J. Chem Phys, 1999) and it allows us to fit some large data sets approximately, using only random subsets of the

Coauthors: Robin Ryder and Alexis Muir Watt

⁂

Adjusting nonresponse bias in small area estimation via a Bayesian hierarchical spatial model

Chong HE (University of Missouri-Columbia, USA)

A survey sample is drawn from a large area although the estimate of interest is at a smaller subdomain level. This may cause a small sample problem. The estimation problem is further complicated by survey nonresponse. We build a Bayesian hierarchical spatial model that deals with the small sample problem and nonresponse. This Bayesian model is used to simultaneously estimate response rates and conditional probability of success given response status. This model also allows us to calculate the marginal probability of success for each subdomain. The method is illustrated using the 2001 Missouri Deer Hunter Attitude Survey. The success is defined as satisfaction regarding to Missouri's Department of Conservation deer management program. The estimated satisfaction rates are lower after adjusting nonresponse than the estimates based on respondents only.

⁂

A Bayesian analysis to explicit and implicit memory

Jun LU (American University, USA)

Modern theories of human memory postulate separate systems for explicit memory (i.e., conscious recollection) and implicit memory (i.e. unconscious, automatic recall). Consequently, psychologists have proposed nonlinear measurement models to separate contributions from these systems to overall memory. Conventional analysis has ignored important sources of variability, which in principle can lead to biased estimates and erroneous conclusions. In this article, a hierarchical model is proposed along with Bayesian analysis. The model employs nested probit components and priors with possibly correlated random effects in situations where these correlations have theoretical interest. Three types of priors are examined, and their efficacies are compared in a simulation study. The methods are also illustrated with data from a memory experiment.

Coauthors: Paul Speckman, Dongchu Sun and Jeffrey Rouder

⁂

17:00–18:30 Room C **High dimensional graphical models in genomics**

Organizers: Christopher Yau (University of Oxford, UK)
Leonardo Bottolo (Imperial College London, UK)
Chair: Anthony Lee (University of Warwick, UK)

Bayesian hierarchical models for data integration in genomics

Francesco STINGO (University of Texas MD Anderson Cancer Center, USA)

In this talk I will focus on recent prior constructions that can be used for the analysis of cancer related high-throughput data. I will describe Bayesian approaches that allow for the integration of different sources of biological information into the analysis of experimental data. I will then explain how a graphical model approach can account for the biological relations among the variables of interest even when the number of observed variables is much greater than the sample size. This approach can be successfully applied to several types of biological networks like gene networks, networks based on the amount of linkage disequilibrium (LD), and regulatory networks. The value of these methods is that they can provide new insights into the analysis of experimental data, test new types of hypotheses, take into account complex dependence structures among several variables, and in general, provide a better understanding of the biological processes involved in the development of cancer.

⁂

Bayesian dependence model for regulatory networks

Silvia LIVERANI (Imperial College London, UK)

The usual assumption for clustering is that observations in the same cluster share the same expression profile – accounting for measurement error – and are expressed independently of observations outside their own cluster. However, when searching over certain types of data, it is unrealistic to assume independence between clusters as, for instance, in the context of microarray experiments, it is well known that genes belong to regulatory pathways and activate and inhibit each other.

In this talk we present a formal methodology for developing the semantics of new Gaussian graphical models so that these are customised to be faithful to a given scientific domain. This new class of models is designed to utilize as much as possible the convenient separation properties of a Bayes cluster model whilst continuing to respect the types of dependence that scientists might conjecture exist. We demonstrate how this class of models can be generalized to encompass more credible deterministic functional dependence structures associated with regulation. We show that it is possible to search over a large number of different hypothesized regulation models, just as in Bayes cluster models, and we interpret the results in a causal framework that provides the scientist with a new hypothesis-generating tool.

We demonstrate the efficacy of these methods on a class of circadian models where interest focuses on the regulatory relationships between different genes. Each member in the class can be evocatively labelled by a regulation graph closely resembling in its semantics the graphical representations of regulation familiar to biologists.

Coauthors: Jim Q. Smith

⁂

Hierarchical Gaussian graphical models and the G-Wishart distribution

Alex LENKOSKI (Heidelberg University, Germany)

The Gaussian Graphical Model (GGM) has proven a useful tool for imposing sparsity when data are sampled from a multivariate normal distribution. However, in many practical situations observations violate either independence, normality or identically distributed assumptions. We review recent work focused on embedding the GGM in hierarchical Bayesian frameworks that incorporates uncertainty in both the precision matrix and the structure of the graphical model. Our work combines theoretical results related to the G-Wishart distribution with computational algorithms for manipulating sparse, positive definite matrices. We conclude with several examples including Gaussian copulas, multivariate stochastic volatility models and matrix and array-variate data. In all cases we show that predictive performance is improved when using the G-Wishart distribution, both in comparison to fully specified Wishart models as well as simpler, but equally sparse conditional autoregressive models.

Coauthors: Adrian Dobra, Abel Rodriguez and Yuan Cheng

❉

Stochastic recurrent heavy subgraphs of denoised weighted networks

Leonardo BOTTOLO (Imperial College London, UK)

There is a mounting interest both from a theoretical and applied point of view about graphical models that explore dependence relationships between random variables across different conditions. In this framework, a separate model for each condition is not able to identify the common structure of the data, while pooling together all conditions will mask the underlying heterogeneity. Inspired by the recent work of Li etal. (2011), we propose a hierarchical Bayesian model to identify stochastic subgraphs that possess two important features: they are recurrent in many conditions and they are highly interconnected. These graphs are particularly important in genetics/genomics since, for instance, they can be used to identify highly coherent transcriptional programs conserved across species and/or tissues. Besides using population-based Monte Carlo methods to search for Recurrent Heavy Subgraphs in the space of all possible subgraphs, we explicitly model the data (weighted networks) using normal matrix variates. This allows us to capture systematic variations (for instance batch effects) present in the data and directly model missing values. We will demonstrate the applicability of the proposed method and the benefits with respect to other mining algorithms in different real-case data sets, including weighted networks generated in animal models and human genetic settings. This is a joint work with Adrian Dobra (University of Washington), Marc Chadeau-Hyam, Mario Falchi and Enrico Petretto (Imperial College, London). [Li W. etal. (2011) PLoS Comput Biol, 7:e1001106]

❉

17:00–18:30 Room D **Bayesian methods in reliability**

Organizer: Refik Soyer (George Washington University, USA)
Chair: Fabrizio Ruggeri (CNR-IMATI, Italy)

A Bayesian analysis of variable reliability growth

Thomas MAZZUCHI (George Washington University, USA)

In this paper we present a Bayesian analysis of the variable reliability growth problem by defining a very flexible prior distribution over the TAAF stages. The prior can be used to define the initial assumption of constant reliability growth, average reliability growth and stochastic reliability growth. In addition, the prior can easily incorporate expert information concerning the magnitude of growth over the TAAF stages. Posterior analysis is conducted using MCMC methods. We will illustrate the incorporation of expert judgment and the reliability growth analysis fo both actual and simulated data sets.

Coauthors: Refik Soyer and Selda Kapan-Ulusoy

❉

To survive or to fail: what is the question?

Refik Soyer (George Washington University, USA)

During the conduct of a life test, does one prefer to observe more failures than survivals? Intuitively, because of the definitiveness of failures, the answer seems to be the former. However,it has been shown that the above is not always true and that the answer depends on the parameters about which inferences are desired. The aim of this paper is to generalize this previous work and more importantly, to address this question when interest centers around the predictive distribution of a lifetime. It is shown that for a certain class of distributions failures can be more informative than survivals and vice versa for some others. The results of this paper question a routine use of accelerated life testing for predictive inference.

Coauthors: Nader Ebrahimi and Ehsan Soofi

⁂

Fully sequential analysis of accelerated life testing

J. Andrés Christen (CIMAT, Mexico)

Accelarated life testing is used in industrial reliability studies were the natural working conditions do not permit standard data gathering that may result in extremely long life testing experiments. Items are subjected to certain stressed working conditions (high voltage, high temperatures, etc.) were life spans are shorter and thus data gathering is faster, and a model is used to infer the normal working conditions reliability (time to failure). However, design issues are crucial regarding the number of items and stress levels to use in the time to failure experiments. We use a fully sequential, backward induction approach, to establish optimal designs using a discretized version of the time to failure data. An intensive, parallel computing approach is used to numerically solve the backward induction problem, obtaining the optimal number of items to sequentially place at each stress level, in order to correctly predict the item's survival function at normal working conditions. We study a synthetic data example as well as a real data case study from an industrial application in Mexico.

Coauthors: Fabrizio Ruggeri and Enrique Villa

⁂

Inference on phase-type models via MCMC with application to networks of repairable redundant systems

Simon P. Wilson (Trinity College Dublin, Ireland)

Continuous-time Markov chains are widely-used models for repairable systems with redundancy, with an absorbing state representing failure. The failure time is then the first passage time to the absorbing state, which is described by a Phase-type distribution.

We consider Bayesian inference via MCMC for Phase-type distributions when data consist only of absorption times. Extensions to the existing methodology are presented which accommodate censored data, enable specific structure to be imposed on the underlying continuous time Markov process and expand computational tractability to a wider class of situations. An example is presented of a simple application to a repairable redundant electronic system when ultimate system failure time (as opposed to individual component failure time) comprise the data. This example is then expanded to illustrate approaches to inference using this methodology for networks of such systems in the very reduced information setting of overall network failure time. This provides one example of a class of reliability problems for which the extended methodology improves both parameter estimation via MCMC and vastly improves computational speed (by several orders of magnitude) over existing approaches to Bayesian inference.

Coauthors: Simon Wilson

⁂

Friday, June 29th

09:00–10:30 Room C **Bayesian spatio-temporal disease mapping: new frontiers**

Organizer: Susanna Cramb (Cancer Council Queensland and Queensland University of Technology, Australia)
Chair: Veronica J. Berrocal (University of Michigan, USA)

Spatio-temporal cancer mapping: Bayesian dynamic factor models

Susanna CRAMB (Cancer Council Queensland and Queensland University of Technology, Australia)

Understanding disease patterns across space and time is important to detect inequalities. Identifying the unobserved risk patterns underlying the disease may provide greater insight.

Bayesian dynamic factor models aim to describe the underlying latent processes of multivariate datasets and how these change across time. These models, incorporating spatial components, have been widely applied to high-dimensional genetic and image analysis data. However, their application in disease mapping is a recent occurrence. Advantages include the ability to model multiple diseases simultaneously, resulting in decreased uncertainty in estimates, as well as enabling the investigation of similar risk patterns.

In Queensland, Australia, cancer is a notifiable disease. Many cancers share similar risk factors, such as smoking, alcohol, diet and physical inactivity, although some cancers have poorly defined risk factors. Even when risk factors are well understood, only limited data are available for risk factors.

We applied Bayesian dynamic factor models to the incidence of selected smoking-related cancers across 478 regions in Queensland to obtain the underlying risk patterns. We also compared disease incidence results with those obtained by the more commonly used Bayesian hierarchical spatio-temporal models of Waller et al (1997) and Knorr-Held (2000). The challenges in applying these models to an environment of high data sparseness will be discussed.

References

Waller, LA, Carlin, BP, Xia, H and Gelfand, AE, 1997. Hierarchical spatio-temporal mapping of disease rates. Journal of the American Statistical Association, 92(438): 607-617.

Knorr-Held, L, 2000. Bayesian modelling of inseparable space-time variation in disease risk. Statistics in Medicine, 19: 2555-2567.

Coauthors: Peter Baade and Kerrie Mengersen

⁂

Spatial modelling of health service provision and utilisation

Nicole WHITE (Queensland University of Technology, Australia)

The provision of publicly funded health services aims to promote health awareness, provide equity in access and maximise positive health outcomes in the community. Fulfillment of these objectives demands continual evaluation of resource allocation and patterns of utilisation, as well as to ensure these resources are made available in a cost-effective manner.

To conduct this surveillance, we develop a Bayesian model for describing patterns in health service utilisation over space and time, using the example of a government funded, breast cancer screening program in Queensland, Australia. As with other studies in public health, data is available only in the form of aggregated counts, for example by Statistical Local Area (SLA), to protect patient confidentiality. In our study, data is in the form of number of screening episodes, cross-tabulated by residential SLA and individual screening facility. To model these data, we propose a generalised linear mixed model framework that allows for the modelling of utilisation rates for each residential SLA and resource. Issues considered in model development include the nature of the relationship between each residential SLA and resource access, the identification of risk factors, small area estimation and zero inflation in observed counts.

Coauthors: Kerrie Mengersen

⁂

BaySTDetect: detecting unusual temporal patterns in small area disease rates using Bayesian posterior model probabilities

Guangquan LI (Imperial College London, UK)

Space-time modelling of small area data is often used in epidemiology for mapping chronic disease rates and by government statistical agencies for producing local estimates of, for example, unemployment or crime rates. Although there is typically a general temporal trend that affects all areas similarly, abrupt changes may occur in a particular area, e.g., due to emergence of localized predictors/risk factor(s) or impact of a new policy. Detection of areas with "unusual" temporal patterns is therefore important as a screening tool for further investigations.

In this talk, we describe BaySTDetect, a novel detection method for short time series of small area data using Bayesian model choice between two competing space-time models. The first model is a multiplicative decomposition of the area effect and the temporal effect, assuming one common temporal pattern across the whole study region. The second model estimates the time trends independently for each area. For each area, the posterior probability of belonging to the common trend model is calculated, which is then used to classify the local time trend as "unusual" or not. An important feature of this method is that the classification rule can be fine-tuned to control for the false detection rate, crucial to any detection method. We will demonstrate the use of BaySTDetect in an application related to policy assessment together with a second application on retrospective surveillance using space-time cancer incidence data from South East England covering the period of 1981-2008.

Coauthors: Sylvia Richardson, Lea Fortunato, Ismail Ahmed, Anna Hansell, Mireille Toledano and Nicky Best

⁂

Dynamic Bayesian modelling for emulation and calibration of epidemic models

Marian FARAH (MRC Biostatistics Unit, Cambridge, UK)

Tracking and predicting the behaviour of emerging epidemics in real-time is important for determining a prompt and appropriate response by the health-care system. Modelling the dynamics of an epidemic is typically done by combining a model for the transmission of infection with a model for the reported data. We consider a model developed to study the dynamics of the 2009 influenza A/H1N1pdm epidemic in London. Transmission is modelled using a system of differential equations that depend on time-varying parameters. Then, a deterministic function of these parameters is used in the data reporting model for doctor consultations reported as time series over the lifetime of the epidemic. The transmission-reporting model is then embedded in a Bayesian framework to obtain real-time posterior inference for the model parameters. However, the deterministic function linking the transmission and reporting components is computationally expensive, making real-time Bayesian inference practically infeasible. We tackle this computational inefficiency by treating the slow-to-compute deterministic function as a dynamic computer model output, and we estimate it with a dynamic emulator, which is a computationally efficient statistical approximation. The emulator is constructed by combining a Bayesian multivariate dynamic linear model that captures dependence over time, with a Gaussian process prior that captures correlation across series over the input space. The dynamic emulator is then combined with the field data and the priors for the model parameters in order to obtain real-time Bayesian inference for the epidemic. The proposed method is illustrated with data from the 2009 influenza A/H1N1pdm epidemic in London.

Coauthors: Paul Birrell, Stefano Conti and Daniela De Angelis

⁂

09:00–10:30 Room D **Bayes modeling in marketing**

Organizer: Nobuhiko Terui (Tohoku University, Japan)
Chair: Nobuhiko Terui (Tohoku University, Japan)

Modeling indivisible demand

Greg ALLENBY (Ohio State University, USA)

Disaggregate demand in the marketplace exists on a grid determined by the package sizes offered by manufacturers and retailers. While consumers may want to purchase a continuous-valued amount of a product, realized purchases are constrained by available package sizes. This constraint might not be problematic for high-volume demand, but it is potentially troubling when demand is small. Despite the prevalence of packaging constraints on choice, economic models of choice have been slow to deal with their effects on parameter estimates and policy implications. In this paper we propose a general framework for dealing with indivisible demand in economic models of choice, and show how to estimate model parameters using Bayesian methods. An analysis of simulated data and a scanner-panel dataset of yogurt purchases indicates that packaging constraints can severely bias parameter estimates and measures of model fit. The effect of introducing a smaller package size is investigated in a counter-factual analysis.

Coauthors: Sanghak Lee

⁂

A brand purchase model of consumer goods incorporating

Makoto ABE (University of Tokyo, Japan)

In this research, we construct a stochastic brand choice model of consumer goods, in which consumers repeatedly choose a brand from "many" alternatives, using scanner panel data. We reexamine consumers' repetitive purchase behavior from the perspective of information processing theory. In particular, we explicitly incorporate the construct of internal search, external search, and learning, which are proposed in behavioral study, into the brand choice model. The existing research on brand choice from many alternatives suggests the existence of choice subsets, such as "awareness set" and "consideration set," in the mind of consumers when making the decision. These subsets cannot be observed directly from purchase data, however, and their identification requires either direct questioning or inference through behavioral modeling. Instead, we introduce a concept of the "experiential set," which is observable from the purchase records, as a means for consumers to process information and decide on brand choice. Because choice subsets are constructed from observable data, this concept helps in building brand choice models that incorporate more elaborate information search and learning by consumers. This, in turn, results in the high predictive validity of the model.

Coauthors: Sotaro Katsumata

⁂

Bayesian dynamic factor analysis and the corresponding clusters for changing needs

Fumiyo KONDO (University of Tsukuba, Japan)

The aim of this research is to construct a dynamic factor analysis model that makes it possible to do accurate comparisons of factors between the data of 2 time points in a dynamically changing market, and to do segmentation by using the estimated factor scores. For conventional factor analyses, accurate comparisons between two time points are not possible and the authenticity and the preciseness for the obtained results are lacking. This research provides a solution, a dynamic factor analysis model in a hierarchical Bayesian framework. The model was applied to use intention survey data on mobile information services in 2008 and those in 2009, which shows dynamic changes in its market. Hypotheses on dynamics are established on the part of factor loadings and/or that of factor scores. The analyses based on the best model gave interesting results in reducing the number of factors that have meaningful results from four to three. Those are factors of communication, information, and entertainment. The last factor yielded no meaningful insight. Furthermore, the factor loadings between 2 time points showed mild changes and a cluster analysis based on the estimated factor scores resulted in a change in the number of segments, from 3 in 2008 to 5 in 2009.

Coauthors: Takayuki Kawamoto

⁂

Dynamic brand satiation

Nobuhiko Terui (Tohoku University, Japan)

A dynamic factor model of brand satiation is developed to explain longitudinal variation in consumer purchases. Factor loadings are associated with a brand's position along a satiation dimension, and factorscores are associated with a household's sensitivity to satiation effects. Dynamics are introduced by allowing the factor scores to evolve through time, reflecting variation in household satiation sensitivity. The factor model is imbedded within a direct utility model that allows for both corner and interior solutions, and is shown to fit the data better than alternative specifications. Analysis of a panel dataset of Doritos corn chips indicates that respondent satiation exhibits faster temporal effects than temporal variation in baseline preferences for brands, and it is, in general, temporally accelerating. Implications for product line assortment are explored.

Coauthors: Shoehi Hasegawa and Greg Allenby

⁂

11:00–12:30 Room A **Bayesian models for high-dimensional complex-structured data**

Organizer: Veera Baladandayuthapani (University of Texas MD Anderson Cancer Center, USA)
Chair: Veera Baladandayuthapani (University of Texas MD Anderson Cancer Center, USA)

Flexible Bayesian sparsity modelling for high dimensional data

Philip Brown (University of Kent, UK)

A number of variants of the lasso have arisen in the literature, from the grouped to the fused lasso and their Bayesian clones. We seek to give a much more general Bayesian prior which simultaneously shrinks both marginal effects and differences of effects to zero and so allows clustering of effects within a group. This leads to classes of spiked long tailed multivariate prior distributions which provide flexible structured alternatives for model choice in high dimensional data. We highlight the advantage of a single multivariate continuous prior over the more traditional univariate 'slab and spike' solution, especially for sparse problems.

The form of shrinkage and selection with robust likelihoods is also considered.

Coauthors: James Griffin

⁂

Bayesian hierarchical multi-subject multiscale analysis of functional MRI data

Marco Ferreira (University of Missouri, USA)

We develop methodology for Bayesian hierarchical multi-subject multiscale analysis of functional Magnetic Resonance Imaging (fMRI) data. We begin by modeling the brain images temporally with a standard general linear model. After that, we transform the resulting estimated standardized regression coefficient maps through a discrete wavelet transformation to obtain a sparse representation in the wavelet space. Subsequently, we assign to the wavelet coefficients a prior that is a mixture of a point mass at zero and a Gaussian white noise. In this mixture prior for the wavelet coefficients, the mixture probabilities are related to the pattern of brain activity across different resolutions. To incorporate this information, we assume that the mixture probabilities for wavelet coefficients at same location and level are common across subjects. Furthermore, we assign for the mixture probabilities a prior that depends on few hyperparameters. We develop empirical Bayes methodology to estimate the hyperparameters and, as these hyperparameters are shared by all subjects, we obtain precise estimated values. Then we carry out inference in the wavelet space and obtain smoothed images of the regression coefficients by applying the inverse wavelet transform to the posterior means of the wavelet coefficients. An application to computer simulated synthetic data has shown that, when compared to single-subject analysis, our multi-subject methodology performs better in terms of mean squared error. Finally, we illustrate the utility and flexibility of our multi-subject methodology with an application to an event-related fMRI dataset generated by Postle (2005) through a multi-subject fMRI study of working memory related brain activation.

⁂

Robust, adaptive, Bayesian functional response regression

Jeffrey MORRIS (University of Texas MD Anderson Cancer Center, USA)

New methods have been developed in recent years to analyze functional and image data, many of which involve extensions of linear regression such as functional regression and functional mixed models. Existing methods, however, tend to be sensitive to outliers, as no analogs to robust linear regression have been developed for the functional setting. Here, we discuss a unified Bayesian method for robust functional regression, whereby a functional response of unspecified form is regressed on a set of linear predictors. The method is developed within the general functional mixed model framework, which can simultaneously model multiple factors and accommodate between-function correlation induced by the experimental design. We demonstrate outstanding robustness properties, doing an excellent job estimating functional regression coefficients even in the presence of Cauchy errors and random effects, and yet not trading off much efficiency when the true likelihood is Gaussian. We also observed remarkable adaptive smoothing properties in our estimates of the fixed and random effect functions, which arise from an interaction of the robust likelihood and adaptive sparsity priors.

Coauthors: Hongxiao Zhu and Philip J. Brown

❈

Statistical inference on temporal gradients in regionally aggregated data

Sudipto BANERJEE (University of Minnesota, USA)

Advances in Geographical Information Systems (GIS) have led to a burgeoning of spatial-temporal databases. Accounting for associations across space and time in analysing such data sets is routinely called for. Here we depart from the rather rich literature in space-time modelling by considering the setting where space is discrete (e.g. aggregated data over regions) and time is continuous. This can be envisioned when, for instance, we have a collection of functions of time over regions, but these functions are spatially associated. A major objective of our current project is to carry out inference on gradients of the temporal process, while at the same time accounting for spatial similarities across neighbouring regions. Rather than use parametric forms to model time, we opt for a more flexible stochastic process for time embedded within a dynamic Markov Random Field framework. Through the cross-covariance function we can ensure that the temporal process realizations are mean square differentiable and we subsequently carry out inference on temporal gradients in a posterior predictive fashion. We use this approach to evaluate temporal gradients in a dataset comprised of monthly county level asthma hospitalization rates in the state of California where we are concerned with temporal changes in the residual and fitted rate curves after accounting for seasonality, spatial-temporal ozone levels, and several spatially-resolved important socio-demographic covariates.

Coauthors: Harrison Quick and Bradley P. Carlin

❈

11:00–12:30 Room D **Applications of non- and semi-parametric Bayesian methods**

Organizer: Fernando Quintana (Pontificia Universidad Católica de Chile, Chile)
Chair: Antonio Lijoi (University of Pavia, Italy)

A "density-based" algorithm for cluster analysis using Dirichlet process Gaussian mixture models

Raffaele ARGIENTO (CNR-IMATI, Italy)

In this talk we propose a Bayesian nonparametric model-based cluster procedure. In particular, we assume a Dirichlet process mixture (DPM) model, with Gaussian kernels, for univariate and multivariate data. Most of the approaches proposed in literature are based on a similarity matrix built estimating the posterior probabilities that observations i and j share exactly the same kernel component. In these cases, cluster estimation turns out to be very sensible to the choice of hyperparameters, since the estimated number of the components of the mixture is strongly influenced by the shape of the "true" distribution of the data. As an example, we consider a simulated dataset from a mixture of bivariate densities with non-convex contour lines. Here, in order to "reconstruct" the shape of the target distribution, the DPM model estimates a number of components in the mixture that is greater than the real one. Clearly, in this case the estimated number of components cannot be directly used to cluster the data. To overcome this problem, we will consider a similarity matrix built by taking into account the posterior probability that some distance between the kernel densities associated with observations i and j is smaller than a threshold. An equivalence relationship among the kernel densities is constructed via the DBSCAN ("density-based spatial clustering of application with noise", Ester et al.,1996) empirical procedure applied to the posterior sample of a Polya urn Gibbs sampler algorithm for the DPM model. We point out that "density-based" refers to points crowded closely together.

Coauthors: Andrea Cremaschi and Alessandra Guglielmi

❈

A Bayesian semiparametric model for yield curves

Jim GRIFFIN (University of Kent, UK)

The yield curve describes how the yield of a bond changes with its maturity. Predicting and modelling yield curves over time is an important topic in finance and have proved to be challenging problems. In this talk, a Bayesian nonparametric version of the popular dynamic Nelson-Siegel model for yield curves will be described which uses the autoregressive Dirichlet Process (DPAR) constructed in Griffin and Steel (2011). The availability of the predictive distribution of the DPAR allows us to construct sequential Monte Carlo methods for inference. The method will be illustrate on yield curves of US Treasury bonds.

Coauthors: Mark Steel Pulak Ghosh

⁂

Nonparametric Bayesian analysis of developmental toxicity experiments with clustered discrete-continuous outcomes

Athanasios KOTTAS (University of California, Santa Cruz, USA)

We present a Bayesian nonparametric mixture modeling approach to inference and risk assessment for developmental toxicity studies. The primary objective of such studies is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical response. We consider the general data setting which involves clustered categorical responses on the number of prenatal deaths, the number of live pups, and the number of live malformed pups from each laboratory animal, as well as continuous outcomes (e.g., body weight) on each of the live pups. To provide flexibility in the functional form of both the response distribution and the various dose-response curves of interest, we propose a mixture model built from a dependent Dirichlet process prior, with the dependence of the mixing distributions governed by the dose level. Particular emphasis is placed on the structure of the mixture kernel and the formulation of the nonparametric prior. The nonparametric mixture modeling framework enables flexible shapes for the implied dose-response relationships and correlations between the different endpoints that can change with dose, features which provide practical advances relative to traditional parametric models for developmental toxicology. The methodology is illustrated with data from a toxicity experiment that investigated the toxic effects of ethylene glycol, a widely used industrial chemical.

Coauthors: Kassandra Fronczyk

⁂

Cluster-specific variable selection for product partition models

Fernando QUINTANA (Pontificia Universidad Católica de Chile, Chile)

Motivated by data on mortality collected in an intensive careunit in Portugal, we study a covariate dependent product partition model with the ability of selecting variables by clusters. This is achieved by a modification of the usual product distribution, by means of the introduction of cluster-specific covariate inclusion parameters. We study the effectof various prior partition choices to the resulting clusters, and discuss how the resulting clusters are influenced by the available covariates.

Coauthors: Peter Mueller and Ana-Luisa Papoila

⁂

Posters

Tuesday, June 26th

The evaluation of evidence in forensic science for discrete data

Colin Aitken (University of Edinburgh, UK)

In forensic science the value of evidence is determined with the likelihood ratio. This compares the likelihood of the evidence if the prosecution proposition is true with the likelihood of the evidence if the defence proposition is true. When the evidence is in the form of measurements, methods are well-developed for multivariate, hierarchical Bayesian multivariate random effects models. Methods are not so well developed for discrete data.

Data are available from a project in forensic phonetics at the University of York in which the number of clicks per minute are recorded for each of 100 speakers, over a period of time ranging from four to six minutes. The evidence may be considered to be the number of clicks from a piece of speech from an unknown source and the number of clicks from a piece of speech from a known source, such as a suspect. The prosecution proposition would be that these two pieces of speech were made by the same speaker and the defence proposition would be that they were made by different speakers. Using these data as an exemplar, possible models for such data are currently under investigation. They will be presented for discussion along with preliminary results. These models include a beta-binomial model and its generalisation to a Dirichlet-multinomial model, a Poisson-gamma model, an empirical model based on relative frequencies and one allowing for correlated discrete data.

Coauthors: Erica Gold

⁂

Bayesian machine learning methods for genome-wide association data

Lorraine Allchin (University of Oxford, UK)

Over the last few years genome wide associations studies have risen to the forefront of research. In these studies the analysis of the data tries to find specific regions in the genome which have a causal link to the disease status or quantitive value of a trait, such as cholesterol levels. However most of these studies only look at the markers in the data, whether they be single nucleotide polymorphisms (SNPs) or copy number variations, one at a time. It may be better to look at multiple SNPs as analysing multiple SNPs at once can make a weak effect more noticeable and reduce the effect of a false positive. I have developed a Bayesian method using Markov Chain Monte Carlo that will analyse data using multivariate marker relationships. It is applicable to both case control and quantitive data. I present simulation for case control based on a human colon cancer data set from a region of chromosome 18q, mouse quantitative data for crlCFW-ALP from outbred mice and simulations comparing my algorithm to both frequentist and other Bayesian methods.

Coauthors: Chris Holmes and Richard Mott

⁂

PAC-Bayesian bounds for high-dimensionnal estimation

Pierre Alquier (Université Paris 7 and CREST, France)

A central question in statistical learning theory is to learn to predict as well as the best predictor in a given family, sometimes referred as “set of experts”. The rate of convergence highly depends on the complexity of this family. PAC-Bayesian bounds were introduced by McAllester (1998). The idea is to control the complexity of the set of experts thanks to a prior distribution, like in Bayesian statistics. In Catoni (2004) a link with statistics is done, as the author shows that this method allow to obtain optimal estimation in classical parametric and non-parametric models. In this poster, we will present recent attempts to apply PAC-Bayesian tools to various problems of high-dimensionnal estimation: sparse linear regression, sparse single-index model, autoregression. In each case, a sparsity inducing prior is given. We then show that the optimal learning rate is reached for a Bayesian type estimator, under very general assumption. In the case of sparse regression estimation, this is particularly satisfying as, for example, a strong assumption on the design is needed to prove the same results for the LASSO estimator. We also show that our estimator can be computed by a MCMC type algorithm, with good performances on simulations, and on real datasets.

Coauthors: Gérard Biau, Olivier Wintenberger and Karim Lounici

⁂

PyMCMC — a new alternative in the implementation of MCMC

Clair Alston (Queensland University of Technology, Australia)

The use of Bayesian statistics for data analysis is now abundant in the research community. Markov chain Monte Carlo (MCMC) is the most common approach to Bayesian estimation in this literature. Packages such as WinBugs and user contributed R libraries are useful and freely available. However, when data size is large or models are non-standard, there is a real need for a powerful and flexible package.

Strickland et al (2012) are developing open source code called PyMCMC. It is a Python module which aids in the construction of efficient MCMC samplers and greatly aids in reducing the likelihood of coding errors by minimising repetitive code. To date, PyMCMC contains classes for Gibbs, Metropolis Hastings, orientational bias MC and slice samplers.

PyMCMC Standard, is a library in which specific modules were developed for models such as linear regression, generalised linear models, mixed models and mixture models.

In this poster I will outline PyMCMC and demonstrate its use in a series of models, providing a feel for the level of expertise required by users and its extensibility.

Coauthors: Christopher Strickland and Kerrie Mengersen

❈

Dynamic graphical models for real-time multivariate road traffic flow forecasting

Osvaldo Anacleto (The Open University, UK)

Dynamic graphical models can combine graphical representations of multivariate time series with a dynamic modelling approach. A particular class of these models, the linear multiregression dynamic model (LMDM), has been shown to be promising for real-time multivariate traffic flow forecasting. However, its performance can be improved further by addressing some issues that still remain as a challenge in on-line traffic modelling.

Data on occupancy, headway and speed are routinely collected in traffic networks and usually not considered in forecasting flow models. Since these variables have a non-linear relationship with flow, cubic splines can be used to incorporate these extra traffic variables into the LMDM. Cubic splines can also be used to accommodate the daily flow cycle, thus providing a more parsimonious method for modelling systematic daily patterns. This approach can be shown to provide better flow forecasts when compared to usual seasonal representations, such as seasonal factor or Fourier forms.

The current LMDM formulation imposes zero covariance between the flows at the entrances to the network. However, the LMDM can be extended by combining different structures of graphical models, which allows the accommodation of possible symmetries in the network and leads to a new class of dynamic graphical models. It will be shown how this class of models can be derived, as well as some of their theoretical and practical implications in the context of traffic modelling. Minute-by-minute data collected at the intersection of three motorways in the UK will be used to illustrate the methodology.

Coauthors: Catriona Queen and Casper Albers

❈

A Bayesian hierarchical model for the reconstruction of the sea level in the Mediterranean basin for the late 20th century

Ioannis Andrianakis (National Oceanography Centre, UK)

The aim of this work is the reconstruction of the sea surface level in the Mediterranean basin for the second half of the 20th century. We use data from two distinct sources: the first data set comes from tide gauge observations, which span the second half of the century but are sparse and discontinuous. The second data set comes from satellite altimetry, which is very dense, but covers only the last two decades. The reconstruction is based on a recent method that expresses a Gaussian field as a solution to a stochastic partial differential equation (Lindgren et al., 2011) and solves for the posterior using Integrated Nested Laplace Approximations (INLA). This approach allows us to handle large data sets, such as the altimetry data set in our case, can define non-stationary fields or fields bounded in a spherical manifold area (important when we extend from the Mediterranean to the globe). The standard method for this problem (Church et al. 2004) is based on Principal Components calculated on the altimeter data and used to fit the tide gauges. Our method tackles the problem within a Bayesian framework. We do not separate the data into two sets but learn simultaneously from both about the spatio-temporal structure of sea level. Furthermore, as we are building a Bayesian hierarchical model, we can include covariates based on climate indices, such as the El Niño Southern Oscillation (ENSO), as well as directly estimating the rate (and acceleration) of sea level rise.

Coauthors: Peter G. Challenor

❈

A nonparametric regression model with normalized weights

Isadora ANTONIANO VILLALOBOS (University of Kent, UK)

In recent years, Bayesian nonparametric mixture models for regression have become a subject of intense research activity. This is due to the flexibility that models of this type can achieve, while retaining useful statistical properties.

For these models to be truly flexible, it is necessary to construct covariate dependent weights which can be guaranteed to add up to one at each point of the covariate space. This can be easily done by defining normalized weights. While in doing so the verification of theoretical properties, such as complete support, becomes accessible, practical difficulties arise from the intractability of the normalizing constant. To solve this problem, alternative models have been proposed, within a family of dependent stick breaking priors. Effectively, posterior simulation is made possible by imposing restrictions on the structure of the dependent weights, that may not have a clear interpretation.

We present a method that allows posterior inference for a nonparametric mixture model with normalized dependent weights, via MCMC simulation. Therefore, we propose the use of a Bayesian model for regression which is flexible and has an intuitive interpretation.

Coauthors: Sara Wade and Stephen G. Walker

❉

Multidimensional covariate dependent Dirichlet processes

Julyan ARBEL (CREST, Université Paris Dauphine, France)

We present a new class of dependent Dirichlet processes, which are an extension of Dirichlet processes that make use of covariates. It allows nonparametric density regression, i.e. regression varying with a covariate index. The dependent Dirichlet process is defined with varying weights and clusters (two infinite dimensional parameters of a Dirichlet process). It is novel in the way we define the weights, and it allows a sensible tuning of the dependence of the resulting predictor-dependent probability distributions. This construction results marginally in a Dirichlet process prior for the random measure at any specific predictor value. A blocked Gibbs sampler is developed for posterior computation, and is illustrated on simulated and real data.

❉

Semi-Markov modelling of electricity co-generation in residential applications with time-dependent covariates

Raffaele ARGIENTO (CNR-IMATI, Italy)

In this poster we analyze data coming from a particular machines called micro co-generators or micro combined Heat and power generators. These are machines for domestic use, capable of the simultaneous production of heat and electricity, controlled by heat requirements. In particular, in this work we will try to characterize the stochastic process describing the profile of the electricity produced in a "typical" winter day by a system of cogenerators. Firstly, we describe the stochastic process used to model the energy produced by one co-generator, the induced likelihood on the data and the related statistical inference: we end up with an Accelereted Failure Time (AFT) model with time-varying covariates. Secondly, the main part of the poster will be devoted to analyze how the estimated model can be used to make inference on the function describing the electricity fed into the electrical grid by a large number of CHP systems by studying the superposition of many realizations of the estimated process. Our approach will be computational. We will use a Markov chain to simulate the stationary trajectories of an alternating renewal process with time varying covariates. The method will be applied to a real dataset. The data were collected at the Institute of Construction Technology of the Italian National Research Council where a micro CHP system is installed to an experimental building.

Coauthors: Antonio Pievatolo and Stefano Galli

❉

Multicomponent systems with dependent failures

Richard ARNOLD (Victoria University of Wellington, New Zealand)

Models of failures in multicomponent systems often make unjustifiable simplifying assumptions about the dependence among the components. In this poster we present a Bayesian treatment of a multi-component system accumulating damage due to a series of fatal and non-fatal shocks. The model introduces dependence amongst the system components in a physically realistic manner by associating individual shock processes with potentially overlapping subsystems made up of groupings of components. The model has the Marshall-Olkin multivariate exponential model as a special case. We present a Bayesian approach to the estimation of the parameters of the model, given system survival data for various model configurations (series, parallel, and other arrangements) of up to 5 dependent components.

Coauthors: Stefanka Chukova and Yu Hayakawa

❉

Estimating Weibull parameters from Kaplan-Meier curves

Christian ASSEBURG (ESiOR Oy, Finland)

Background: The Kaplan-Meier estimator is used to present time-to-event data in graphical form in the presence of censoring. For example, in a clinical trial on cancer, the Kaplan-Meier estimator may be used to illustrate the probability of progression-free, corrected for patients who are lost to follow-up. Treatment effect is usually published numerically as a hazard ratio, but for health-economic models an absolute measure of difference is required together with the baseline survival function. The Kaplan-Meier estimator itself is not a sufficient statistic of the survival dataset, but because the patient-level clinical data are rarely published, it is often the only available source of data for health-economic models. The statistical bias in different methods for extracting the survival function from a published Kaplan-Meier diagram has not been described.
Methods: Here, hypothetical clinical trials are generated and three statistical models are compared regarding their ability to estimate the underlying survival function parameters and treatment effect. (1) Complementary log-log transformed linear model, using numbers-at-risk as sampling weights. (2) Same but weighted using Greenwood's estimate of standard error. (3) Model 1 with additional explicit modelling of numbers lost to follow-up. Models were fitted in OpenBUGS 3.2.1 using vague priors for all model parameters.
Results: Mean squared error in the parameter estimates was lowest with model 2. The 95%-credibility intervals of models 1 and 3 are too narrow. Explicitly modelling censoring does not improve accuracy of model estimates.
Conclusion: Standard errors of the Kaplan-Meier estimator should be published to facilitate accurate modelling of survival functions.

❈

Interwoven EM algorithms

Paul BAINES (University of California, Davis, USA)

Since its first appearance in the statistical literature, the EM algorithm has spawned a large number of variants, generalizations and specializations. In this talk we introduce a new class of EM algorithm to accelerate the rate of convergence: the interwoven EM (IEM) algorithm. The fundamental idea of interweaving is to combine two or more different data augmentation schemes into a single efficient algorithm. We prove the optimality of one particular pairing, the ancillary and sufficient augmentations, within a large class of problems.

Our algorithm is the optimization counterpart of the recently proposed interweaving approach for data augmentation. We establish formal connections between both the structure and performance of the sampling and optimization algorithms. IEM is also intimately connected with many other flavors of EM such as parameter-expanded EM and variational EM. By formalizing the connections between IEM and a selection of alternative optimization algorithms, we will attempt to convey some intuition behind interweaving. Examples will be given describing the construction, implementation and performance of the IEM algorithm for both mixed effects models and dynamic linear models. We conclude with ideas for extensions of IEM suggested by this intuition and a general framework for multi-model statistical computation.

Coauthors: Xiao-Li Meng

❈

Peeling and Bayesian QTL mapping for allo-polyploids

Roderick BALL (Scion - New Zealand Forest Research Institute, New Zealand)

We develop a Bayesian multi-locus quantitative trait locus (QTL) mapping method for allo-polyploids, incorporating model uncertainty in the number and location of QTL. Allo-polyploids (e.g. clover, kiwifruit) contain multiple genomes (e.g. of the form AABBCC) inherited from distinct ancestral species. For QTL mapping, polyploids present an additional challenge due to missing information. Multiple copies of an allele cannot be distinguished from single copies and few markers are fully informative. Previous approaches have addressed special cases, work best only for special single dose markers, and base inference on a likelihood ratio conditional on one or more flanking markers, that are rarely fully informative in polyploids.

Our algorithm uses peeling (Elston and Stewart 1971; the original Bayesian graphical modelling method for evaluating probability distributions on pedigrees) to sample from the posterior distribution of fully informative 'virtual markers'. The virtual markers are then used as input to a Bayesian QTL mapping method (the BIC method; Ball Genetics 2001; R package BayesQTLBIC; http://cran.r-project.org/web/packages/BayesQTLBIC/index.html), a non-MCMC method based on using the BIC criterion to obtain good approximations to posterior probabilities for linear models representing possible QTL configurations, with multiple peeling samples analysed jointly using multiple imputation. Single marker peeling is used together with an incremental HMM peeling strategy to sample from the posterior distribution of virtual marker genotypes. By eliminating genotype errors and eliminating low probability combinations at each stage we ensure the computation is linear in the number of loci and progeny.

The method is applied to allo-tetraploid clover data.

Coauthors: Gail Timmerman-Vaughan

❈

Nonparametric spatial models for periodontal disease data with spatially-varying non-random missingness

Dipankar BANDYOPADHYAY (University of Minnesota, USA)

Periodontal disease process is often quantified by clinical attachment level (CAL) defined as the distance down a tooth's root that is detached from the surrounding bone. Measured at 6 locations per tooth throughout the mouth (excluding the molars), it gives rise to a typical clustered data set-up; however routinely subjected to a one-number summary such as the whole mouth average, or the number of observations (sites) greater than a threshold, or some regression of these summaries that identifies important co-variates determining the current state of a subject' periodontal health. Rather than a simple one number summary, we set forward to analyze all available CAL data for each subject exploiting the presence of spatial referencing, non-stationarity, non-normality, etc. Also, the subjects have a considerable proportion of absent (missing) teeth, and cannot be considered 'missing at random', because periodontal disease is the leading cause of adult tooth loss, and the nature of this missingness can vary in different locations within the mouth. Under a Bayesian paradigm, we propose a semi-parametric flexible spatial (joint) model of observed CAL and the location of missing tooth via. kernel convolution methods, incorporating the aforementioned features of CAL data under a unified framework. Application of this methodology to a real data recording periodontal health of an interesting African-American population, as well as simulation studies reveal the gain in model fit and inference, and provides a new perspective into unraveling covariate-response relationships in presence of complexities posed by these data

Coauthors: Brian J. Reich and Howard Bondell

⁂

Infinite tensor factorization priors

Anjishnu BANERJEE (Duke University, USA)

There is increasing interest in broad application areas in defining flexible joint models for data having a variety of measurement scales, while also allowing data of complex types, such as functions, images and documents. We consider a general framework for nonparametric Bayes joint modeling through mixture models that incorporate dependence across data types through a joint mixing measure. The mixing measure is assigned a novel infinite tensor factorization (ITF) prior that allows flexible dependence in cluster allocation across data types. The ITF prior is formulated as a tensor product of stick-breaking processes. Focusing on a convenient special case corresponding to a Parafac factorization, we provide basic theory justifying the flexibility of the proposed prior and resulting asymptotic properties. Focusing on ITF mixtures of product kernels, we develop a Gibbs sampling algorithm for routine implementation relying on slice sampling. The methods are compared with alternative joint mixture models based on Dirichlet processes and related approaches through simulations and real data applications.

Coauthors: David Dunson

⁂

Chain event graphs in Bayesian model selection for health studies

Lorna BARCLAY (University of Warwick, UK)

Bayesian Networks (BN) are one of the most widely used graphical models and strategies for searching useful explanatory models are well established. However, such methods, in their simplest form, do not search over context-specific symmetries within the system. We introduce a new and more general class of graphical models, the Chain Event Graphs (CEG). These capture the discrete BN as a particular subclass and further allow for asymmetries within the dependence structure of the variables. CEGs are derived from a probability tree by merging the nodes of the tree whose associated conditional probabilities are the same and hence retain the paths of the tree in a more compact graph. Therefore they share with the BN the property of providing an expressive graphical framework through which conclusions can be read back to the client. We illustrate the improvement via the CEG by applying it to a subset of the Christchurch Health and Development Study, which looks at the effect of social and family factors on children's health in a New Zealand birth cohort. We show how searching the much larger space of graphically explanatory models of the CEG can be of great advantage. Scoring methods are set up analogously to the standard Bayesian Dirichlet metric for BNs and hence a direct comparison between the models is feasible. Our results demonstrate that the CEG search leads to a significantly higher scoring model than the MAP BN model and further that it lets us draw plausible additional conclusions from the health study.

Coauthors: Jim Q. Smith and Jane L. Hutton

⁂

A Bayesian geographically weighted regression model applied to real estate markets

Maria Asuncion BEAMONTE (Universidad de Zaragoza, Spain)

In this paper we propose a Bayesian approach to analyse the Geographically Weighted Regression model (BGWR) that combines the local character of the estimation of the regression coefficients with the relationships of global character over those coefficients through the prior distribution. We aim, in the first place, to evaluate the ability of this procedure to discriminate between a spatially constant process and one with spatially varying relationships, and subsequently to accurately retrieve spatially varying relationships. The methodology is based on the use of a geographically weighted likelihood and MCMC methods are utilized to obtain the posterior distribution of the model coefficients. Moreover, in order to analyze the spatial variation of the coefficients, we carry out a model comparison process with Bayesian tools. For illustrative purposes we show an application to the Real Estate Market in the city of Zaragoza (Spain).

Coauthors: Pilar Gargallo and Manuel Salvador

⁂

Bayesian models for multicategory spatial data

Candace BERRETT (Brigham Young University, USA)

Albert and Chib (1993)'s latent variable representation of the Bayesian probit regression model for categorical outcomes is widely recognized to facilitate model fitting. This representation has also been used in various settings to incorporate residual dependence among the discrete response variables. In this poster, we further extend this latent variable strategy to specify models for multicategory spatially-dependent outcomes. In particular, we discuss parameter identifiability issues in the latent mean specification and introduce covariance structures for describing the cross spatial/category residual dependence. We also consider data augmentation MCMC strategies for improving the efficiency of model fitting algorithms. Finally, we illustrate the proposed modeling framework through an analysis of land-cover/land-use observations taken over mainland Southeast Asia.

Coauthors: Catherine A. Calder

⁂

Mixture modeling strategies applied to approximate Bayesian computation

Fernando V. BONASSI (Duke University, USA)

Approximate Bayesian Computation (ABC) is a class of methods applied to inference problems in which the likelihood function of the experiment is intractable. Such methods rely on the basic idea of generating massive prior:model simulation and accepting only the cases for which the generated synthetic data set is close to the actual observation. One well-known problem about ABC is its high inefficiency in terms of low acceptance rate. Two computational strategies that are commonly used to deal with this problem are: (i) smart sampling schemes such as SMC or MCMC; and (ii) post-processing regression adjustment of the accepted values. In this work, we propose two new strategies that follow the ideas of (i) and (ii), respectively. The first strategy consists of an automatic and adaptive sampling scheme based on SMC, and the second strategy consists of a post-processing adjustment based on nonlinear regression. In both strategies a mixture modeling approach is incorporated. As a motivating application, we address the problem of statistical inference from marginal data in bionetwork models, using a toggle switch example integrating simulated and real data as context. Finally, in order to show the better efficiency of the SMC strategy proposed, we applied our method to other examples that are commonly used in the ABC literature.

Coauthors: Mike West

⁂

Approximating posterior densities by iterated Laplace approximations

Bjoern BORNKAMP (Novartis Pharma AG, Switzerland)

The Laplace approximation is an old, but frequently used method to approximate integrals for Bayesian calculations. In this poster I present an extension of the Laplace approximation, by applying it iteratively to the residual, i.e., the difference between the current approximation and the true function. The final approximation is thus a linear combination of multivariate normal densities, where the coefficients are chosen to achieve a good fit to the target distribution. We illustrate on real and artificial examples that the proposed procedure is a computationally efficient alternative to current approaches for approximation of multivariate probability densities.

Reference: Bornkamp, B. (2011), Approximating Probability Densities by Iterated Laplace Approximations, Journal of Computational and Graphical Statistics, 20, 656-669

⁂

The product graphical model

Luke BORNN (University of British Columbia, Canada)

In this paper we propose a class of prior distributions on decomposable graphs, allowing for improved modeling flexibility. While existing methods solely penalize the number of edges, the proposed work empowers practitioners to control clustering, level of separation, and other features of the graph. Emphasis is placed on a particular prior distribution which derives its motivation from the class of product partition models; the properties of this prior relative to existing priors are examined through theory and simulation. We then demonstrate the use of graphical models in the field of agriculture, showing how the proposed prior distribution alleviates the inflexibility of previous approaches in properly modeling the interactions between the yield of different crop varieties. Lastly, we explore American voting data, comparing the voting patterns amongst the states over the last century.

Coauthors: Francois Caron

❈

Asset smoothing in the health and retirement study: a Bayesian nonparametric approach

Lane BURGETTE (RAND Corporation, USA)

The Health and Retirement Study (HRS) is a large, longitudinal survey of American households. Self-reported financial variables (such as total assets) are a key component of the HRS. Data-users have complained that household-level asset trajectories contain many “spikes” and “notches” where reported assets may change by an order of magnitude from one wave of data collection to the next, and return nearly to the old level in the following wave. It is believed that these are the product of misreports, and some data users have asked the HRS to provide smoothed asset reports that more closely align with patterns that social scientists feel are reasonable.

We propose a Bayesian nonparametric approach to such smoothing. Our model assumes that a Dirichlet process mixture of Gaussian processes describes the actual asset trajectories. On top of this, we propose a contamination model that assumes that the reported assets may – with small probability – have been drawn from some other distribution. A by-product of the MCMC is the posterior probability that each household had a misreport (conditional on the model). We find that the distribution of such probabilities is strongly bimodal: the model assigns very low probability of any misreports to the majority of households, while assigning a very high probability of at least one misreport to 5-10% of the households.

Coauthors: Michael Hurd and Erik Meijer

❈

Bayesian modelling of ion channels

Ben CALDERHEAD (University College London, UK)

This project is concerned with Bayesian statistical modelling of ion channels. Ion channels are transmembrane proteins present in cells, which are crucial for a range of important physiological processes, such as modulating the synaptic connection between neurones, or between nerves and muscle. The conformational changes that describe how ion channels open and close are statistically well modelled as aggregated Markov processes. Postulated Markov models are fitted from single channel recordings using different agonists and agonist concentrations.

For any hypothesised mechanism, the transition rates of states and time distributions of openings and closings of the channel must be inferred from the observed (open/close) experimental recording. A maximum likelihood approach has been fruitful thus far in order to infer rate parameters and distributions for simple mechanisms. However, new experimental data suggests a larger number of conformational states than previously thought. As the complexity of models grows, a Bayesian approach to rate parameter inference and model selection is likely to be vital for making accurate inferences from noisy experimental data and uncertain structures.

We employ recent advances made in differential geometric MCMC methods to perform Bayesian inference and model selection over proposed models, which allows rational and systematic analysis of competing Markov models. This approach, allied with structural data from X-ray crystallography, aims to provide a better understanding of the relationship between structure and function of ion-channels. In the longer term it is expected that this approach will help inform the process of rational drug design for specific ion channel targets.

Coauthors: Michael Epstein and Mark Girolami

❈

Approximate Bayesian computation for astronomical model analysis

Ewan CAMERON (Queensland University of Technology, Australia)

"Approximate Bayesian Computation" (ABC) represents a powerful methodology for the analysis of complex stochastic systems for which the likelihood of the observed data under an arbitrary set of input parameters may be entirely intractable—the latter condition rendering useless the standard machinery of tractable likelihood-based, Bayesian statistical inference (e.g. conventional Markov Chain Monte Carlo simulation; MCMC). In this poster we highlight the potential of the ABC technique for the analysis of astronomical models with a specific case study in the morphological transformation of high redshift galaxies. In particular, we (1) outline a novel stochastic model for the birth, merging, and secular evolution of massive galaxies in the early Universe, (2) constrain the free parameters of this model against a detailed demographic dataset (compiled from the latest CANDELS HST/WFC3 imaging) using a "Sequential Monte Carlo" (SMC) implementation of ABC, and (3) describe the physical implications of the resulting posteriors for contemporary theories of galaxy formation. Finally, we conclude with a discussion of the many exciting possibilities afforded by the further application of ABC in this field. [cf. Cameron & Pettitt 2012: http://arxiv.org/abs/1202.1426]

Coauthors: Tony Pettitt

❈

Bayesian multivariate mixed scale density estimation

Antonio CANALE (University of Torino, Italy)

Although univariate continuous density estimation has received abundant attention in the Bayesian nonparametrics literature, there is essentially no theory on multivariate mixed scale density estimation. In this article, we consider a general framework to jointly model continuous, count and categorical variables under a nonparametric prior, which is induced through rounding latent variables having an unknown density with respect to Lesbesgue measure. For the proposed class of priors, we provide sufficient conditions for large support, strong consistency and rates of posterior contraction. These conditions, which primarily relate to the prior on the latent variable density and heaviness of the tails for the observed continuous variables, allow one to convert sufficient conditions obtained in the setting of multivariate continuous density estimation to the mixed scale case. We provide new results in the multivariate continuous density estimation case, showing the Kullback-Leibler property and strong consistency for different mixture priors including priors that parsimoniously model the covariance in a multivariate Gaussian mixture via a sparse factor model. In particular, the results hold for Dirichlet process location and location-scale mixtures of multivariate Gaussians with various prior specifications on the covariance matrix.

Coauthors: David Dunson and Pierpaolo de Blasi

❈

Efficient Bayesian inference for (dynamic) generalized Bradley-Terry models

Francois CARON (Inria Bordeaux, France)

The Bradley-Terry model is a popular approach to describe probabilities of the possible outcomes when elements of a set are repeatedly compared with one another in pairs. It has found many applications including animal behaviour, chess/sport ranking and multiclass classification. Numerous extensions of the basic model have also been proposed in the literature including models with ties, multiple comparisons, group comparisons and graphs. Using a suitable set of latent variables, we develop Expectation-Maximization algorithms to obtain MAP estimates and simple Gibbs samplers to approximate the posterior distribution for these models. We also describe a simple latent variable construction that allows to consider dynamic generalized Bradley-Terry, and show how efficient Bayesian inference can still be performed with this class of models. We demonstrate experimentally the efficiency of the algorithms on a variety of applications.

Coauthors: Luke Bornn and Arnaud Doucet

❈

Bayesian testing unit roots versus threshold specifications

Shu-Yu CHEN (Feng Chia University, Taiwan)

Testing the unit root null hypothesis in a threshold autoregressive model has received much attention recently. This paper aims to develop a test of the unit root null hypothesis against a stationary threshold process. This testing problem is nonstandard and complicated because a parameter is unidentified and the process is nonstationary under the null hypothesis. A new posterior odds analysis is proposed to test for a unit root in the context of threshold autoregressive models with volatility dynamics. A mixture prior is proposed that alleviates the usual identifiability problem as the AR(1) coefficient parameter tends to one that allows for reliable inference and a proper posterior, despite the non-integrability of the likelihood function. The posterior odds ratio is employed to compare the nonstationary versus stationary threshold heteroskedastic process. We illustrate the proposed methodology in simulation study and empirical example.

Coauthors: Shu-Yu Chen and Cathy W.S. Chen

❉

Locally optimal MCMC sampling

Ting-Li CHEN (Academia Sinica, Taiwan)

Let X be a finite space and p be an underlying probability on X. For any real-valued function f defined on X, we are interested in calculating the expectation of f under p. Let X0, X1,... ,Xn,... be a Markov chain generated by some transition matrix P with invariant distribution p. The time average, the sum of f(Xk), is a reasonable approximation to the expectation, E[f(X)]. In this paper, we propose an MCMC algorithm using a locally optimal transition matrix. From our simulation studies, our proposed method outperformed two famous MCMC algorithms, the Gibbs Sampling and the Metropolis-Hastings algorithm.

❉

Winter peak electricity load forecasting in South Africa using extreme value theory with a Bayesian flavour

Delson CHIKOBVU (University of Free State, South Africa)

The paper explores the use of extreme value theory in calculating the probability of exceedance (POE) levels and posterior predictive densities of peak electricity demand in South Africa for the period 2000 to 20011. In this paper the exceedances above a sufficiently large threshold is modelled through a Generalized Pareto-type (GP-type) distribution. The GP-type distribution is derived as an approximation of the tail of Generalized Burr-Gamma (GBG) distribution (a convenient extreme value distribution fitted to the whole data set) by Verster and De Waal (2011). The main advantage of this modelling approach is that the GP-type distribution has only one parameter to estimate and not two as is the case with the well-known Generalized Pareto distribution (GPD). An application of a Generalized Pareto-type (GP-type) in modelling daily peak electricity demand is presented. The optimum threshold is determined using a Generalized Pareto quantile plot. Empirical results show that the GP-type distribution is appropriate in modelling of extreme daily peak electricity demand above a sufficiently high threshold. A comparative analysis was done by fitting a Generalized Extreme Value (GEV) distribution. Empirical results show that the Weibull class of distributions can be used to model the daily peak electricity demand data. The results from the GP-type distribution are better than those from fitting a GEV distribution.

Coauthors: Caston Sigauke

❉

Bayesian spatio-temporal modelling of ultrafine particle number concentration from a panel design

Sam CLIFFORD (Queensland University of Technology, Australia)

The UPTECH project aims to link health measurements to air pollution in order to quantify the impact of air quality on the health of primary school children in South-East Queensland, Australia. Due to the costs and time constraints of health and air quality measurements, a panel design with random school selection has been employed, with measurements being taken for two weeks at each of 25 of the 188 state primary schools in the Metropolitan School District. The school data will be augmented by continuous monitoring data recorded by three Environmental Protection Agency monitoring sites.

One component of the UPTECH project is the development of a spatio-temporal model of ultrafine particle number concentration (PNC). This semi-parametric regression model will be used to quantify the effects of meteorology, temporal trends and background levels at each site and across all sites so that predictions of ultrafine PNC can be made outside the two week measurement period at each school. These predictions will be used to approximate the exposure to ultrafine particles of the children whose health outcomes have been measured as part of UPTECH.

Coauthors: Sama Low Choy, Kerrie Mengersen and Lidia Morawska

❉

Assessing the spatial distribution of species using Bayesian latent Gaussian models

David CONESA (University of Valencia, Spain)

Methodological approaches for modelling species occurrence patterns have been proposed during the last years. The presence/absence of species is usually modelled with a hierarchical Bayesian model using the geographical and environmental characteristics of each sampling locations. Maps of predicted probabilities of presence are generated using Bayesian kriging. Our interest here is to describe how to use the integrated nested Laplace approximation jointly with the Stochastic Partial Differential Equation approach to perform fast inference and prediction in such models. Two particular examples are presented: the first one is related to the presence of a fish species (Mediterranean mackerel), while the second deals with the presence of a species of parasites in cows of Galicia (NW Spain). Moreover, our interest is also to show that our approach can be adapted to be used in situations where sampling is likely to be preferential. Preferential sampling arises when the process that determines the data locations and the process being modelled are stochastically dependent. Commercial fishery is a clear example of this situation: fishermen go to fish in the area where they presume that can take the target species. To show how to perform preferential sampling we also present an example about the abundance of a target species (hake) in the Western Mediterranean.

Coauthors: Antonio López-Quílez, Silvia Lladosa, Facundo Muñoz, Maria G. Pennino, Jose M. Bellido, Janine. B. Illian, Daniel Simpson, Marta González-Warleta and Mercedes Mezo

⁂

Birth-death regression

Forrest CRAWFORD (University of California, Los Angeles, USA)

Statistical inference for stochastic counting processes is a vital part of modern biostatistical and epidemiological research. The general birth-death process provides a flexible generalization of Markov counting processes on the non-negative integers. The Poisson, linear birth-death, susceptible-infected-susceptible (SIS), Moran, and Ehrenfest processes are widely-used examples of general birth-death processes. Traditionally, regression analyses for counting processes have utilized maximum likelihood, least-squares, or Martingale approaches, which require unrealistic large-sample assumptions and are not flexible enough to accommodate more complicated population processes such as those often encountered. A formal framework for Bayesian hierarchical regression in Markov population/counting processes has remained elusive, in part because of the difficulty in computing the likelihood for discretely observed or interval censored data. In this paper, we propose a novel Bayesian framework for hierarchical regression for arbitrary general birth-death processes. We employ an innovative method for computing the likelihood using a continued fraction representation of the Laplace transform of birth-death transition probabilities. For some models, we derive exact representations of full conditional distributions as exponential family densities, enabling efficient Gibbs sampling approaches to inference. We also provide a computationally efficient method for fast maximum a posteriori estimation via the EM algorithm. We validate our methods using simulated data and apply our technique to the growth of cancer cells under different chemotherapeutic treatments, revealing the effect of treatment on cell birth and death rates.

Coauthors: Marc A. Suchard

⁂

Modelling the transmission of Clostridium difficile in hospitals

Madeleine CULE (University of Oxford, UK)

Clostridium difficile is a major cause of healthcare-associated diarrhoea, and controlling its spread within a healthcare setting is an ongoing focus of significant public health effort. However, the dynamics of the disease are still relatively poorly understood, and the complex, dynamic contact network of a large hospital system makes it difficult to apply standard techniques for analysing disease outbreaks. Focussing on data collected within the Oxford Radcliffe Hospitals NHS Trust between 2007 and 2009 (containing a total of 250,000 hospital admissions and 1,000 confirmed cases of Clostridium difficile), we describe a compartmental model for the transmission of Clostridium difficile within the hospital system. A Bayesian approach allows us to estimate key epidemiological parameters, as well as to assess the impact of assumptions made about the dynamics of disease. An advantage of this framework is that it enables us to introduce additional genetic typing information about the bacteria as it becomes available.

Coauthors: Peter Donnelly

⁂

Emulating expensive decision choices with application to computer models of complex physical systems

Jonathan CUMMING (Durham University, UK)

Most experiments on physical systems require the specification of various decision parameters. When such systems are modelled by computer simulators, then we must optimise our choice of decision inputs for the computer simulator while recognising all of the other sources of uncertainty which are characteristic of problems in computer modelling. One such example is found in well testing in hydrocarbon reservoirs, where the parameters of the well test (such as its duration in time) must be chosen before the experiment is performed to learn about the geological configuration of the reservoir. While the parameters of the well test are controllable, the configuration of the reservoir is unknown and to be discovered. Given such a problem, we consider the goal of identifying good choices of decision parameters with respect to the uncertainty in the reservoir's configuration. We approach this problem by using a simple computer model of reservoir behaviour which simulates the results of any well test given its geology. We can then assess the utility of a single choice of decision parameters by simulating many well tests under those decisions and investigating differing geologies. Investigating many different choices of decision parameters allows us to build a picture of the behaviour of the utility function over the decision space. However, as this function is both uncertain and expensive to evaluate, we use emulation methods to represent our expected utility surface. Applying history matching techniques to the emulated expected utility then enables us to identify good regions of the decision space, and hence good test configurations. This study illustrates a general methodology for practical use of computer simulators to support decision problems of realistic size and complexity.

Coauthors: Michael Goldstein

❉

Bayesian nonparametric clustering in phylogenetics: modeling antigenic evolution in influenza

Gabriela CYBIS (University of California, Los Angeles, USA)

Influenza is alone responsible for 250,000 to 500,000 deaths every year, and antigenic variability represents much of its epidemiological burden. To visualize antigenic differences across many viral strains, antigenic cartography methods use binding assay data to map influenza antigenicity onto a 2 or 3 dimensional space. In these assays, the influenza strains naturally form clusters of similar antigenicity that correlate with sequence evolution. To understand the dynamics of these antigenic groups, we jointly model genetic and antigenic evolution using Bayesian phylogenetic machinery and nonparametric clustering methods. We use a modified version of the Chinese Restaurant Process to incorporate the phylogenetic dependency structure between strains in the modeling of antigenic clusters. Our goal is to use the genetic information to better understand the evolution of antigenicity and potentially make predictions about the emergence of new antigenic groups. This information could be particularly relevant for the design of new vaccines. We present an application of this model to a dataset of seasonal influenza.

Coauthors: Marc Suchard and Janet Sinsheimer

❉

Cluster analysis via random partition distributions

David DAHL (Texas A&M University, USA)

We present a suite of clustering algorithms based on random partition distributions indexed by pairwise distance information. Having their roots from the Bayesian nonparametric literature, these proposals are interesting in that they use the same input data as does hierarchical clustering yet are based on formal probability distribution. Hence, the resulting clusterings can be justified by theory and statements about clustering uncertainty can be readily addressed. We find empirically that our proposals are competitive with traditional clustering methods across a variety of clustering scenarios and are superior in select cases.

Coauthors: Bryce Little and Mahlet G. Tadesse

❉

Transformation based reversible jump MCMC

Moumita DAS (Indian Statistical Institute, India)

Since its introduction by Green (1995), reversible jump MCMC (RJMCMC) has been recognised as a powerful tool for making posterior inference about a wide range of statistical problems where the model dimension is unknown. But despite its well-deserved recognition, a major criticism of the method is that efficient "move-types" are difficult to design.Typically,dimension jumping moves in reversible jump samplers yield much lower acceptance rates than the other move types, The problem is many times aggravated in the case of multivariate parameters. or the multimodal target distributions, also for large dimesional distribution give particularly low acceptence rates. lead to poor mixing, slow convergence. Moreover, the computational complexities involved in RJMCMC makes it error-prone and difficult to implement in practice. In this article we propose a novel methodology for variable-dimensional problems based on deterministic transformations of low-dimensional random variables. The idea is originally inspired by the method of transformation based MCMC (TMCMC) proposed by Dutta and Bhattacharya (2011) in the case of fixed-dimensional distributions. One of the remarkable advantages of TMCMC is that underlying target distribution is very high dimensional, deterministic transformation of a one dimensional random variable is sufficient to generate an appropriate Markov chain that is guaranteed to converge to the high-dimensional target distribution. This idea of transforming a single random variable leads to excellent acceptance rates. Since it is well-known that joint updating of many random variables using Metropolis-Hastings (MH) algorithm generally leads to poor acceptance rates, TMCMC, in this regard, seems to provide a significant advance. Using the key idea based on deterministic transformations of low-dimensional random variables we propose efficient and easy-to-implement methodology for trans-dimensional problems as an alternative to RJMCMC. We refer to this new methodology as transformation-based RJMCMC (TRJMCMC). We validate our proposal theoretically, establishing convergence properties. We apply our methodology to the variable-dimensional problems for comparison purpose. Moreover in a particular challenging example, where it is required to update a random series consisting of unknown, but large, number of random summands, our methodology ensured significant computational savings over RJMCMC and far better acceptance rates and mixing properties.

Coauthors: Sourabh Bhattacharya

⁂

Inference on population structure using random partition models

Maria DE IORIO (University College London, UK)

We propose a class of Bayesian random partition models to infer population structure from genetic data. Given multilocus genotype data from a sample of individuals collected from a population with unknown structure, the model clusters subjects into subpopulations, whose number is unknown. We develop methods that allow for correlation between loci due to linkage disequilibrium by exploiting recent work on covariate-dependent random partition models. We demonstrate the proposed model on simulated and real data and discuss methods to summarise the results of the MCMC output for the analysis of population structure.

Coauthors: Stefano Favaro and Yee Whye Teh

⁂

Forecasting election results via the sequential nature of Bayes' theorem and an application to the Turkish general elections

Haydar DEMIRHAN (Hacettepe University, Turkey)

Prediction of election results is a challenging issue. Especially, when an election is approaching, numerous public opinion research firms compete with each other in the prediction of election results. They spend a considerable amount of effort and budget in order to forecast results of elections with the best accuracy. In general, sample sizes used for this purpose are limited. Hence, accuracy of their predictions is questionable. However, when we aim to forecast results of an election, there are several data sources. Results of past elections are based on the whole population, so they are the main source of data. Although being limited in sample size, pre-election polls are also useful data sources about the current political situation. We need an approach to handle and combine information coming from these data sources, effectively. Our main point is to combine information coming from past elections and available pre-election polls by using the sequential nature of Bayes' theorem in order to predict the results of the forthcoming election. Election results are generally presented in histograms. We simply estimate a histogram constructed for results of an election in the Bayesian way. We accumulate information obtained from the past elections and polls by the sequential use of Bayes' theorem. We apply our approach in order to predict results of Turkish General Elections between 1987-2011. As the result, we are able to forecast results of the June 2011 Turkish General Elections with an acceptable amount of error, and provide predictions for the forthcoming election.

Coauthors: Kamil Demirhan

⁂

Bayesian estimation for partially observed multiplicative intensity Poisson processes

Sophie DONNET (Université Paris Dauphine, France)

Poisson processes are used in various application fields applications (public health biology, reliability, ...). In their homogeneous version, the intensity process is a deterministic constant. In their inhomogeneous version, it depends on time. To allow for an endogenous evolution of the intensity process we consider multiplicative intensity processes. Inference methods have been developed when the trajectories are fully observed.

We deal with the case of a partially observed process. As a motivating example, consider the analysis of an electrical network through time. This network is composed of cables and accessories (joints). When a cable fails, the cable is replaced by a new cable connected to the network by two new accessories. When an accessory fails, the same kind of reparation is done leading to the addition of only one accessory. The failure rate depends on the stochastically evolving number of accessories. We only observe the times events; the initial number of accessories and the cause of the incident (cable or accessory) are only partially observed.

The aim is to estimate the different failure rates or to make predictions. The inference is strongly influenced by the initial number of accessories. We deduce a sensible prior on this quantity from an asymptotic study of the process. We illustrate the performances of our methodology on a large simulation study.

Coauthors: Judith Rousseau

❈

Estimation of short term private external debt of turkey within the DLM framework

Oya EKICI (Duke University, USA)

Considering the economic stability or crisis, short-term private external debt stock of a country is one of the key figures that are needed to be followed. In economics literature, the convention is to monitor the ratio of the debt stock to Central Bank's international reserves. However the debt stock series itself can be main focus as an indicator. So we examine Turkey's short-term external debt stock data. We propose usage of linear growth model to model short-term external debt stock and Bayesian method for estimation and forecasting strategy of the model. The empirical findings and comparisons illustrate the performance of our predictions in capturing instable terms of the economy.

Coauthors: Karun Nemlioglu

❈

Bayesian inference with optimal maps

Tarek EL MOSELHY (Massachusetts Institute of Technology, USA)

We present a new approach to Bayesian inference that entirely avoids Markov chain simulation, by constructing a map that pushes forward the prior measure to the posterior measure. Existence and uniqueness of a suitable measure-preserving map is established by formulating the problem in the context of optimal transport theory. We discuss various means of explicitly parameterizing the map and computing it efficiently through solution of an optimization problem, exploiting gradient information from the target density when possible. The resulting algorithm overcomes many of the computational bottlenecks associated with Markov chain Monte Carlo.

Advantages of a map-based representation of the posterior include analytical expressions for posterior moments and the ability to generate arbitrary numbers of independent posterior samples without additional likelihood evaluations or "forward solves." The optimization approach also provides clear convergence criteria for posterior approximation and facilitates model selection through automatic evaluation of the marginal likelihood. We demonstrate the accuracy and efficiency of the approach via inference in several statistical models (e.g., Bayesian logistic regression, log-Gaussian Cox processes), as well as high-dimensional parameter inference in ordinary and partial differential equations.

Coauthors: Youssef Marzouk

❈

Outlier identification for model-based clustering: an application to the Seychelles child development and nutrition study

Katie EVANS (University of Rochester, USA)

This work was motivated by the Seychelles Child Development Nutrition Study (SCDNS), an international research project to study whether nutrients from fish consumption modify the effects of prenatal mercury exposure on childhood neurodevelopment. To capture information on their entire nutrition profile, mothers were asked to keep a food-diary over the course of four consecutive days. The data set includes ordinal, zero-dominated, and correlated variables, all of which can be problematic for model-based clustering methods based on normal mixture models.

One implementation of such a method is the contributed R package MCLUST, a powerful clustering tool, sometimes criticized for its instabilities. When parametrized correctly, adding a prior can avoid degeneracies and singularities; we propose a modified parametrization for the variance prior with better properties. Additionally, clustering on all available nutrition variables may be difficult to interpret or influenced by outliers; thus, a clustering method which simultaneously performs variable selection is used. Once important clustering variables are identified, the presence of outliers should be considered and appropriately handled. We have developed a method to identify outliers based on cluster-specific variances estimated with and without removal of each observation separately. Results from a simulation study demonstrate the ability of our method to detect true outliers without falsely identifying many non-outliers.

Clustering on fish-related variables important to investigators resulted in no outliers being identified. Clustering on variables chosen from variable selection on all nutrition variables, two outliers were identified. While not overly influential on clustering results, their presence and differentiating features are noted.

Coauthors: Tanzy Love and Sally Thurston

⁂

An industrial application of Bayesian statistics

Christopher Eves (University of Kent, UK)

Frequentist Statistics are commonly used in industry. The applications vary from quality control to direct production. They are the default choice when companies take a statistical approach to problems. Bayesian Statistics could provide a better alternative to problems. However, due to limited understanding these methods are rarely developed. One thing that may deter companies from using such methods is the ambiguity surrounding the selection of a prior. This is a recurring argument amongst academics and one that there is no simple answers for. This is a result of the choices available, none of which can be defined as clearly the best. Such uncertainty could somewhat compromise the reputation of the subject. It is perhaps this uncertainty which stops companies investing in the development of Bayesian methods, despite the potential for large savings or the generation of revenue. A company specialising in measurement instrumentation has decided to implement Bayesian methods. They currently employ classical OLS methods to correct for errors arising from the changes of process conditions. This method requires a lengthy data collection process, which leaves the data vulnerable to external influences. A simple analysis suggested parameters for thousands of previously calibrated meters demonstrated near normality. This was supported by histograms and ECDF plots. This resulted in a clear, undisputed choice for the Prior information. This poster identifies the loss of ambiguity with industrial prior selection given the existence of substantial amounts of data. We will also see the positive financial implications as a result of the Bayesian method.

Coauthors: Xue Wang and Stephen Walker

⁂

Semiparametric estimation of population-level distributions of nutritional markers using disparate data sources

Mariel Finucane (Harvard University, USA)

Restricted growth, as a result of inadequate nutrition and infections, is an important cause of morbidity and mortality in infants and children. Furthermore, the hazardous effects of undernutrition occur along a continuum, and hence all levels of mild, moderate and severe undernutrition are important for their potential effects on mortality. We present a semiparametric Bayesian model that makes annual estimates of population-level distributions of height-for-age z-scores for 0-5 year old children in the world's 141 low- and middle-income countries between 1985 and 2011. Since tail probabilities are of clinical and public health importance for these markers, estimating the distributions' shapes is of particular substantive importance.

We use a hierarchical probit stick-breaking model on the weights of a finite normal mixture. The model allows for nonlinear changes in time, and it borrows strength in time, covariates, and within and across regional country clusters to make estimates where data are sparse or missing. Importantly, we incorporate both individual-level data when available, as well as aggregated summary statistics from studies whose individual-level data could not be obtained, accounting for dependence when multiple summaries are reported for a single study sample.

Our analysis includes data from 563 country-years, comprising anthropometric measurements of 4.3 million children. Our cross-validated model has good external validity, substantiating our predictions for country-years without data. Although we find that height-for-age z-scores in developing countries have improved on average between 1985 and 2011, worldwide progress has been uneven.

Coauthors: Christopher Paciorek, Gretchen Stevens and Majid Ezzati

⁂

Modelling the dynamics of CaCO3 budgets in a changing environment using a Bayesian belief network approach

Chiara FRANCO (University of Essex, UK)

A probabilistic approach based on Bayesian Belief Network (BBN) was used to address the difficulties associated with the quantitative evaluation of complex systems such as coral reef ecosystems. Coral reefs are experiencing continuous deterioration due to increasing anthropogenic pressures. Carbonate budget assessments have been largely used to evaluate coral reefs health and the status of their physical structure, named framework. The intimate connection between coral reef framework and human pressures associated with the exploitation of this ecosystem, were investigated using a holistic approach based on BBN. A BBN parametrized on empirical data collected in south Sulawesi (Indonesia), was developed to quantitatively assess rates of changes in reef framework carbonate production and removal across environmental gradients. Bayesian Belief Networks simplify the effects of the events involved in maintaining coral reef framework under a positive or negative carbonate balance by marginalizing the irrelevant variables and incorporating uncertainty. To reduce the bias introduced by inappropriate elicitation the variables and the relationships between variables constituting the model structure were selected based upon a systematic review of carbonate budget studies. The network was populated with discretized data and the reef sites status information were graphically visualised through a direct acyclic graph. The model was tested for portability using a second dataset from Jamaica, with data extrapolated from published literature or environmental monitoring programs. By encapsulating the complexity of coral reef ecosystem in a BBN frame, we intended to quantify the uncertainty associated with some of the variables thus to improve management and conservation actions.

Coauthors: David Smith, Allan Tucker and Leanne J Hepburn

❉

A nonparametric Bayesian approach to the analysis of bioassay experiments with ordinal responses

Kassie FRONCZYK (University of Texas MD Anderson Cancer Center, USA)

Motivated by the area of cytogenetic dosimetry, where interest lies in the relationship between the dosage of exposure to radiation and some measure of genetic aberration, we aim to model the general case of bioassay experiments with ordinal outcomes. We adopt a structured nonparametric mixture model built upon modeling dose-dependent response distributions. The framework capitalizes on a continuation-ratio type construction for the mixture kernel to promote flexibility in the dose-response curves and to maintain relative ease of implementation. A portion of an experimental data set is used to illustrate the method with respect to the prime inferential objectives.

Coauthors: Athanasios Kottas

❉

Kernel Bayes' rule and its applications

Kenji FUKUMIZU (Institute of Statistical Mathematics, Japan)

A general nonparametric method for realizing Bayes' rule is proposed, based on representations of probabilities in reproducing kernel Hilbert spaces (RKHSs) associated with positive definite kernels. The prior and conditional probabilities are expressed as empirical means and covariance operators defined by embedding random variables into the RKHSs with kernels, and the kernel mean of the posterior distribution is computed in the form of a weighted sample. The kernel Bayes' rule can be applied to a wide variety of Bayesian inference problems, where exact density form may not be available. As applications of this method, we mainly discuss Bayesian computation without likelihood in population genetics, for which Approximate Bayesian Computation (ABC) is popularly used, and show promising results in comparison with the standard ABC approach.

Coauthors: Shigeki Nakagome, Le Song, Arthur Gretton and Shuhei Mano

❉

Efficient MCMC methods for estimating binomial logit models

Agnes Fussl (Johannes Kepler University of Linz, Austria)

Our work considers efficient Bayesian methods for data originating from experiments where binary outcome variables are aggregated in terms of binomial outcome variables. Such data arise e.g. if repeated measurements are taken for each covariate pattern in the design matrix of a planned experiment or when frequencies in the form of two-way or three-way contingency tables are observed. For any of these examples, the data are modeled by a binomial regression model. We present estimation methods based on various MCMC sampling procedures for fitting binary logit models with data augmentation, where the regression coefficients are estimated by rewriting the logit model as random utility model (RUM) or difference RUM (dRUM). Using the reconstructed binary observations to estimate the parameters in the binomial model leads to a high-dimensional latent variable. To reduce the dimension of this individual representation, we introduce an aggregated RUM version of the binomial model like in Frühwirth-Schnatter, Frühwirth, Held and Rue (2009). To improve the sampler further, we suggest a new aggregated dRUM representation of the binomial logistic regression model. The parameters are estimated by using three different MCMC algorithms: a data-augmented MH-sampler, an auxiliary mixture sampler and a novel hybrid auxiliary mixture (HAM) sampler. To show that the modifications lead to a considerable reduction of computing time and a noticeable gain in efficiency, we present a comparison of their performances within a comparative case study on various data sets.

Coauthors: Sylvia Frühwirth-Schnatter and Rudolf Frühwirth

❉

A Bayesian approach to detecting signals of recent natural selection in large scale phased haplotype data

Chris Gamble (University of Oxford, UK)

Identifying regions of natural selection is a natural way to learn about recent adaptation, and may help establish candidates for loci that harbour disease susceptibility or drug resistance variants. We approach the problem of detecting signals of selection by developing a novel Bayesian approach which searches for genomic regions that have specific departures from the genome-wide population structure. This structure is estimated by exploiting the computationally efficient Hidden Markov Model (HMM) framework to characterise the relationships between the studied haplotypes. We demonstrate the utility of our approach by applying it to both simulated and real whole genome data from diverse human populations and the malaria parasite Plasmodium Falciparum.

Coauthors: Chris Spencer and Chris Holmes

❉

Ontogenetic niche shifts among plant species of eastern United States

Souparno Ghosh (Duke University, USA)

A climate niche can be defined as a set of climate conditions required for a species to develop. There are evidences that the niche changes during a species' lifespan indicating organisms' changes in sensitivity towards climatological factors during their lifecycle. Such changes of niche during a species' lifespan are known as ontogenetic niche shifts. An expansion of niche occurs when species' requirements are stricter in earlier stages and a niche contraction occurs when the scenario is reversed. There has been no attempt to study the niche shifts among plant species by jointly modeling different life stages and investigating both occurrence and abundance simultaneously. The contribution here is to develop a Bayesian hierarchical model to jointly analyse the occurrences and abundances at different life stages of plant species using the climate information as the covariate. The climate niche is obtained as a response surface corresponding to different life stages of the concerned species. Further we develop an inferential framework that enables us to statistically examine if there is sufficient evidence of ontogenetic niche shift. The methodology is subsequently applied on several plant species of the eastern United States.

Coauthors: Alan E. Gelfand, Kai Zhu and James S. Clark

❉

A regression-adjusted Hockey plus-minus Metric

Robert GRAMACY (University of Chicago Booth School of Business, USA)

We present a logistic regression model for backing out player contributions when goals are scored in hockey. The traditional plus-minus statistic, which gives each player on the ice one unit of credit (for or against), has long been regarded as a poor (and noisy) metric for ranking players. Our approach takes a longer view by considering the aggregate of all matchups involved in each goal, as well as overall team contributions and other factors. The large-p large-n nature of the data requires the use of shrinkage priors. Inference in this context has, until recently, been cumbersome. Therefore, this paper acts as a fitting showcase of two new methods: a Gibbs sampler for fully Bayesian inference; and an EM-style algorithm for variable selection and MAP inference. The result of the analysis is that most players don't stand out as strong (positive or negative) contributors. This allows the stars to really shine, and brings to light a few surprising underperformers overlooked in earlier analyses.

Coauthors: Shane Jensen and Matthew A. Taddy

* * *

Application of Bayes' theorem to study seismic events and active faults in Venezuela

Carolina GRANADO (Venezuelan Foundation for Seismological Research, Venezuela)

We used Bayesian Techniques to establish quantitatively the connection between the hypocenters location of earthquakes occuring in active fault system segments in Venezuela. We analysed probability density functions and prior probabilities based on the application of the the Bayes theorem to establish a correlation between a seismic event and a specific fault combined with the frecuency-magnitude analysis of the seismicity. The idea of the process is to contribute with valuable information in the estimation of seismic hazard in seismic regions with crustal seismicity.

Coauthors: Raquel Vasquez

* * *

Fast Bayesian approximate inference in near-Gaussian latent models

Thiago GUERRERA (Norwegian University of Science and Technology, Norway)

This work extends the Integrated Nested Laplace Approximation (INLA) method to latent models outside the scope of latent Gaussian models, where independent components of the latent field can have a non-Gaussian distribution. Two important class of models that can be addressed with our proposed method are non-Gaussian random effects models and dynamic models with non-Gaussian error term for the observation and/or system equation. Our approach is applied to two examples of interest to the applied community and the results are compared with that obtained by Markov Chain Monte Carlo (MCMC), showing similar accuracy with only a small fraction of computational time.

Coauthors: Håvard Rue

* * *

Integrating phenotypic, genetic and geographic data in the analysis of population structure

Gilles GUILLOT (Technical University of Denmark, Denmark)

We propose a statistical model to detect population structure from a combination of phenotypic, genetic and geographic data. The method models structure as a set of homogeneous clusters. Its main outputs are estimates of the number of such clusters and of their spatial location. We develop a Markov chain Monte Carlo algorithm to perform inference. We show from simulated data and data belonging to the classic statistical clustering literature that our method is accurate and improves on alternative approaches in many situations. We also implement the method on an original dataset of georeferenced genetic and morphometric markers of the bank vole Myodes glareolus in Sweden.

* * *

Model penalized inference

Jan HANNIG (University of North Carolina at Chapel Hill, USA)

In this poster we discuss a connection between approximate Bayesian computation and generalized fiducial inference. This connection is done by defining a model penalized distribution to which Bayes Posterior and generalized fiducial distribution are a special case. This model penalized distribution than allows for incorporation LASSO like effect into the resulting "posterior" distribution.

* * *

Predicting variable-length functional outputs for emulation of a NASA flight simulator

Yuning HE (University of California, Santa Cruz, USA)

We address the problem of predicting the output of a function of multiple real variables in which the output is itself a function of a real variable, e.g., time. We propose a new statistical method for such emulation of computer models with a multi-dimensional input space and variable-length time series output space.

When output curves can be grouped into a small set of clusters of similar curve length, shape, and frequency content, we can improve the prediction by fitting distinct models for different classes of output curves. Therefore, our complete model is a multi-stage model built on top of statistical models for the single output functions from input to class, input to output curve length, and input to each of the coefficients in a reduced basis curve representation.

The motivating application for our prediction method is statistical emulation of a NASA flight controller simulator with 11 input variables and 12 output variables, where each output variable is a function over time indicating some aircraft configuration measurement. Our statistical methods help to assist in the NASA simulator verification and validation process.

Coauthors: Herbert Lee

⁂

Geometry-based hyperpriors for nonstationary lengthscales

Charles HOGG (National Institute of Standards and Technology, USA)

Gaussian Processes are flexible nonparametric tools for quantifying uncertainty in continuous functions $f(X)$. They achieve smoothness in $f(X)$ by specifying the covariance among its values, rather than by assuming its functional form. For simple covariance functions like the Squared Exponential (SE), this amounts to choosing characteristic horizontal and vertical lengthscales ℓ and σ for features in the function. Nonstationary Gaussian Processes allow these lengthscales to vary, generalizing ℓ to $\ell(X)$. This creates the need for priors on $\ell(X)$ ("hyperpriors") which guarantee smoothness, especially since kinks and discontinuities in $\ell(X)$ propagate to the functions $f(X)$ which it governs.

The present work argues that $\ell(X)$ curves have a hidden intrinsic geometry which can help construct these hyperpriors. Two key observations motivate this geometry. First, ℓ and X necessarily have the same units, since ℓ describes distances in X-space. Second, it does not matter what those units are, since $\ell(X)$ defines the local units. These principles define equivalence classes for segments of $\ell(X)$ under simultaneous scaling of ℓ and X by the same factor. Hyperpriors which exploit this geometry do not discriminate among curves of the same class.

A case study (simulated X-ray scattering data with varying peakwidth) illustrates this idea using one particular geometry-based hyperprior. Many others are possible, and their relative merits remain to be explored. The main contribution of the present work is to recognize that lengthscale curves carry intrinsic geometry, which forms a natural foundation for expressing prior knowledge about them.

⁂

Hierarchical Bayesian methods for combining efficacy and safety in multiple treatment comparisons

Hwanhee HONG (University of Minnesota, USA)

Biomedical decision makers confronted with questions about the comparative effectiveness and safety of interventions often wish to combine all sources of data. Such multiple treatment comparisons (MTCs) may or may not include head-to-head randomized controlled trials of the treatments of primary interest, instead relying largely on indirect comparisons (say, trials that separately compare each treatment to placebo). In such settings, hierarchical Bayes-MCMC meta-analytic methods offer a natural approach (e.g., by enabling full posterior inference on the probability that a given treatment is best). In this paper, we summarize the current state of such methods in the binary response setting, and consider extension to the case of multiple outcomes (say, on both efficacy and safety) where we account for correlation and use clinically-informed weights to arrive at an overall decision regarding the best treatment. We offer several simulations to support the use of our methods over more standard approaches that ignore cross-endpoint correlation. We also illustrate our methods with data from a recent MTC comparing pharmacological treatments for female urinary incontinence. We close with a discussion of our results and a few avenues for future methodological development.

Coauthors: Haitao Chu, Bradley Carlin, Tatyana Shamliyan and Robert Kane

⁂

Algorithm for constructing sparse regression models via Bayesian lasso

Ibuki Hoshina (Chuo University, Japan)

The Lasso proposed by Tibshirani (1996, JRSS Ser.B 58) has sparked much interest in the use of regularization for simultaneous variable selection and estimation. In linear regression models, the lasso has been extended by imposing different penalties for regression coefficients, and various L1 type of regularization procedures such as adaptive lasso (Zou, 2006, JASA, 101), elastic net (Zou and Hastie, 2005, JRSS Ser.B 67), SCAD (Fan and Li, 2001, JASA, 96) have been proposed. In Bayesian contexts the lasso estimates can be interpreted as posterior mode when independent and identical Laplace prior distributions are placed on the regression coefficients. Park and Casella (2008, JASA, 103) introduced the Bayesian lasso linear regression model, using a conditional Laplace prior distribution represented as a scale mixture of normal distributions with an exponential mixing density (Andrews and Mallows, 1974, JRSS Ser.B 36). The lasso shrinks regression coefficients towards zero, and can also set some coefficients to be exactly zero. However, in the context of variable selection, the Bayesian lasso loses this property, not setting any of the coefficients to zero. Hence the Bayesian lasso is not simultaneous variable selection and parameter estimation procedure. In the present paper, we propose an efficient algorithm which modifies the Bayesian lasso estimates so as to be sparse. Monte Carlo simulations are conducted to investigate the efficiency of the proposed algorithm for sparse Bayesian regression models.

Coauthors: Sadanori Konishi

❊

Modeling air pollution mixtures in Southern California

Reza Hosseini (University of Southern California, USA)

Air quality has consistently been shown to be an important determinant of public health. However, characterizing the complex behavior of the multiple pollutant mixture inherent in ambient air is a challenging statistical problem. This work develops a multivariate model that accounts for correlation between pollutants across time and space by employing latent spatial processes. A Bayesian hierarchical approach is used to estimate the parameters and predict pollution levels at unobserved times and locations.

Application of the model incorporates air pollution concentrations (NO2, NOx and O3) gathered at several locations within 12 Southern California communities during three periods in 2005-2006. We study the change of the between-pollutant correlation over time.

Coauthors: Kiros Berhane, Duncan Thomas and Meredith Franklin

❊

Revolutionizing the education of data analytics using Bayesian interactive visualizations

Leanna House (Virginia Tech, USA)

Data (or Discovery) Analytics (DA) is a field that merges quantitative methods in statistics, data mining, and computer science to discover and summarize information in data. In practice, learning from data via DA requires comprehensive critical thinking skills which 1) extend beyond the application of quantitative statistical or computational methods and 2) include qualitative forms of thought, such as formalizing potential biases, communicating personal judgement, exploring multiple solutions, assimilating new information with old, and assessing implications of discoveries. Unfortunately, current methods in teaching DA focus primarily on its quantitative aspects; only after students master quantitative theory and methods, do students have an opportunity to think critically about applications. Students who fail to complete current DA courses (for whatever reason), do not develop comprehensive DA skills nor experience how complex quantitative summaries may enhance critical thinking and advance knowledge. We use novel interactive data visualizations that rely on Bayesian Visual Analytics (House, Leman, and Han; 2011) to design student-centric DA courses that will synchronize DA with critical thinking and make DA accessible to diverse learners.

Coauthors: Scotland Leman

❊

A marker-set approach using Bayesian regression model with regularization for family association studies

Yung-Hsiang HUANG (National Taiwan University, Taiwan)

Missing heritability has been an important issue in genetic association studies under the assumption of common disease/common variants (CD/CV). One approach is then to aim at rare variants and developed statistical methodology based on the assumption of common disease/rare variants (CD/RV). Most of the existing methods, however, handled only one type of the variants, not both at the same time. They overlook the possibility that the susceptible variants can contain both common and rare ones. Moreover, these methods were applied to unrelated populations, not pedigree data. Compared with the population-based study designs, family-based studies may have a better chance to detect rare variants, because the susceptible genetic factors were usually transmitted from ancestral generation to offspring.

To include both common and rare variants in a single model, we utilize the uncertainty coding matrix in a Bayesian generalized linear mixed-effect regression model for the marker-set data of family studies. Although marker-set data contain more information than single SNP association tests, it raises another problem of dimensionality. Here we use regularization method such as LASSO and ridge regression to estimate the marker-set effect, and use the posterior probability of the variance parameter to evaluate the evidence of association. Simulation studies will be conducted to demonstrate the validity of the proposed test and an application will be included for illustration.

Coauthors: Chuhsing Kate Hsiao

❉

Dimension reduction and alleviation of confounding for spatial generalized linear mixed models

John HUGHES (University of Minnesota, USA)

Non-Gaussian spatial data are very common in many disciplines. For instance, count data are common in disease mapping, and binary data are common in ecology. When fitting spatial regressions for such data, one needs to account for dependence to ensure reliable inference for the regression coefficients. The spatial generalized linear mixed model (SGLMM) offers a very popular and flexible approach to modeling such data, but the SGLMM suffers from two major shortcomings: (1) variance inflation due to spatial confounding, and (2) high-dimensional spatial random effects that make fully Bayesian inference for such models computationally challenging. We propose a new parameterization of the SGLMM that alleviates spatial confounding and speeds computation by greatly reducing the dimension of the spatial random effects. We illustrate the application of our approach to simulated binary, count, and Gaussian spatial datasets, and to a large infant mortality dataset.

Coauthors: Murali Haran

❉

Semiparametric Bayesian modelling for the classification of patients with observed high survival probability

Francesca IEVA (Politecnico di Milano, Italy)

In this work, two different Bayesian semiparametric logit models are fitted to grouped data related to the in-hospital survival outcome of patients hospitalized with STEMI (ST segment Elevation Myocardial Infarction) diagnosis. Both Dirichlet (DP) and Dependent Dirichlet Process (DDP) priors are considered for modelling the random effects distribution of the grouping factor (the hospital of admission), with the aim of studying the clustering induced on them by the random partition. The study is within Strategic Program of Regione Lombardia, and is aimed at supporting decisions in healthcare policies. Two are the main goals of the work: providing methods for clustering providers with similar effect on patients' outcome, enabling the evaluation of providers' performances with respect to clinical gold standards, and making predictions on outcomes at patient's level. According to the first issue, the flexibility of semiparametric DP and DDP models suits well the overdispersed nature of data and the in-built clustering they provide is exploited to point out groups among random effect estimates. This is achieved by minimizing suitable loss functions on misclassification errors through linear integer programming, assigning different weights to the wrong association of hospitals belonging to different clusters as well as to the missed association of hospitals whose effects are similar, within the random partition. Concerning the second issue we propose a new method for classifying patients, starting from Bayesian posterior interval estimates of survival probabilities. In terms of predictive power, it provides better results than other criteria usually adopted in literature, based on pointwise estimates.

Coauthors: Alessandra Guglielmi, Anna Maria Paganoni, Fabrizio Ruggeri and Jacopo Soriano

❉

Nonparametric stochastic volatility: mixture approach

Kaoru Irie (University of Tokyo, Japan)

This paper proposes the new nonparametric models of the stochastic volatility in order to consider both the negative skewness and high kurtosis of asset returns and the leverage effect. Our models assume the distribution of logarithm of squared rates of asset returns to be Dirichlet process mixture of normals. To employ MCMC, we found that mixture sampler and slice sampler are available when we sample SV parameters and mixture components, respectively. Simulation studies and empirical research are also conducted by using TOPIX daily prices.

Coauthors: Yasuhiro Omori

⁂

Bayesian additive latent variable models for multivariate densities

Irina Irincheeva (Duke University, USA)

There is an increasing need for flexible statistical models to sparsely characterize high-dimensional multivariate distributions without making restrictive assumptions, such as normality and linearity of associations. We propose a novel class of Bayesian additive latent variable models, which induces nonparametric priors for multivariate densities that have fully flexible marginals while accommodating non-linear dependencies among variables. By allowing a one-dimensional unknown function to characterize dependence between each pair of variables, we address the curse of dimensionality which arises in unconstrained multivariate density estimation. Using uniform latent variables with Bayesian splines for the unknown nonlinear functions, we obtain an efficient Gibbs sampler for posterior computation. The methodology is applied to simulated and real data applications.

Coauthors: David B. Dunson

⁂

Toward large-scale customer analysis in the framework of random utility model

Tsukasa Ishigaki (Tohoku University, Japan)

In this presentation, we describe a customer analysis model using a large-scale purchase history including several hundred of different type items. Today, we are easily able to gather a large scale purchased history by point of sales system. The feedback of information extracted from such big dataset to practical business and management become increasingly important. In practical, it is natural that service providers and managers hope to care more and various kinds of items for customer relationship management or personalization. However, conventional studies of random utility model that aims to strict deduction model based on the utility maximization principle and the consumer behavior theory mainly focused on few kinds of items in a category due to the limitation of computational cost. Here, we propose a customer analysis model that can deal with several hundred of different type items. The model is realized by a fusion of the hierarchical Bayes binary logit model and the latent class model called as topic model in the field of machine learning research. We show the model structure and the estimation results in the poster presentation.

Coauthors: Nobuhiko Terui and Tadahiko Sato

⁂

Multivariate realized stochastic volatility model with leverage

Tsunehiro Ishihara (Hitotsubashi University, Japan)

A joint model of multivariate returns and realized measures of covariance is proposed. The model of returns is described by a multivariate stochastic volatility model with leverage. The matrix exponential transformation is used to keep the time varying covariance matrices positive definite. The measurement equation of the multivariate realized measure is formulated as a matrix log-linear form, which is a matrix-variate extentsion of Takahashi, Omori, and Watanabe (2009). An efficient Bayesian estimation method using Markov chain Monte Carlo is discussed. The proposed model and estimation method are applied to stock return data.

Coauthors: Yasuhiro Omori

⁂

Learning of sequence data by compression with sparse Markov Chains

Väinö JÄÄSKINEN (University of Helsinki, Finland)

Finite memory sources and variable length Markov chains have recently gained popularity in data compression and mining, in particular for applications in bioinformatics. Here we consider denser data compression with a family of sparse Bayesian predictive models for Markov chains in finite state spaces. Our approach lumps transition probabilities into classes comprised of invariant probabilities, such that the resulting models need not have a hierarchical structure as in context-tree based approaches. This can lead to a substantially higher rate of data compression, and such non-hierarchical sparse models can be motivated for instance by data dependence structures existing in the bioinformatics context. We describe an exact Bayesian inference algorithm for learning sparse Markov models through clustering of transition probabilities, and illustrate it by an application to DNA sequence data.

Coauthors: Jie Xiong, Jukka Corander and Timo Koski

⁂

Gaussian process transfer function models for discretely sampled but continuous time series

Teresa JACOBSON (University of California, Santa Cruz, USA)

Gaussian process models provide flexible modeling approaches for a variety of applications involving curve-fitting. We present an application in time series analysis, using Gaussian process based models to fit continuous-time transfer functions to real-life data. Advantages and extensions include the ability to easily accommodate differently-spaced time series and unevenly-spaced data collection, as well as easing extensions to multivariate and spatial data. We also demonstrate a few techniques to improve computation and MCMC mixing with this class of model.

Coauthors: Bruno Sansó

⁂

A level-set hit-and-run sampler for quasi-concave distributions

Shane JENSEN (University of Pennsylvania, USA)

We develop a new sampling strategy that uses the hit-and-run algorithm within level sets of the target density. Our method can be applied to any quasi-concave density, which covers a broad class of models. Our sampler performs well in high-dimensional settings, which we illustrate with a comparison to Gibbs sampling on a spike-and-slab mixture model. We also extend our method to exponentially-tilted quasi-concave densities, which arise often in Bayesian models consisting of a log-concave likelihood and quasi-concave prior density. Within this class of models, our method is effective at sampling from posterior distributions with high dependence between parameters, which we illustrate with a simple multivariate normal example. We also implement our level-set sampler on a Cauchy-normal model where we demonstrate the ability of our level set sampler to handle multi-modal posterior distributions.

Coauthors: Dean Foster

⁂

Nonparametric Bayesian modeling for spatially smooth density estimation

Seongil JO (Seoul National University, Korea)

In this study, we consider Bayesian models for spatially varying density estimation. In particular, we utilize mixtures of species sampling model which allows intuitive modeling of spatial variation. We use a straightforward numerical approach to calculate the predictive probability function of the species sampling model. This approach is based on the size-biased permutation of the sequence of weights in the stick-breaking representation of the species sampling model. The method is illustrated using climate data.

Coauthors: Jaeyong Lee

⁂

Convergence rates among component-wise MCMC algorithms for Bayesian inference

Alicia Johnson (Macalester College, USA)

In Bayesian settings requiring Markov chain Monte Carlo (MCMC) simulations for inference, applying the fundamental Metropolis-Hastings algorithm is often prohibitively difficult. Specifically, producing the required full-dimensional simulation updates can be a challenge. In such cases, a component-wise algorithm which updates the simulation one variable or sub-block of variables at a time might be more effective. Common examples include the Gibbs and Metropolis-within-Gibbs algorithms, within which component-wise updates can be performed using different strategies. For example, a "fixed scan" strategy updates variables in a fixed, pre-determined order whereas a "random scan" strategy updates a single, randomly selected variable in each iteration. We explore Markov chain convergence properties between and within these component-wise strategies, paying special attention to conditions under which this convergence occurs at a geometric rate. We illustrate our results for a Bayesian version of a linear mixed model.

⁂

Graph-regularized centroid estimation on a hierarchical Bayesian model for genome-wide association studies

Ian Johnston (Boston University, USA)

Genome-wide association studies (GWAS) attempt to determine which genomic markers (SNPs) are predictors of genetic traits, most commonly human diseases. In practice, despite the extreme imbalance of having millions of markers recorded for only a few thousand individuals, it is of great interest to glean as much information as possible from this type of data. To this end, we propose a novel Bayesian statistical model that exploits a hierarchical structure between markers and genes to leverage information between levels and alleviate the "large p small n"; regimen while still attaining a reasonably complex and realistic model. To obtain posterior samples we describe a collapsed Gibbs sampler that takes advantage of particular features of the resulting graphical model to obtain an efficient sampling procedure. We conduct inference on which SNPs and genes are associated with the studied trait using graph-regularized centroid estimation. Finally, we illustrate the proposed model and estimation procedure on simulated data and discuss preliminary results on real-world data.

Coauthors: Luis Carvalho

⁂

Efficient searching in tree and forest graphical models

Edmund Jones (University of Bristol, UK)

Gaussian graphical models are used to model gene regulation networks, and learning the graph structure from data can be useful in suggesting regulatory relationships between genes. The Bayesian approach to structure-learning enables prior knowledge about the graph structure to be used and allows for uncertainty in the posterior distribution.

However, Bayesian learning of Gaussian graphical model structure is highly computer-intensive, because of the large number of possible graphs and the need for time-consuming approximations on non-decomposable graphs. One possibility is to restrict attention to forests, which are graphs that have no cycles, or trees, which are connected forests. It is believed that most real-world graphs are sparse, so the restriction to forests can be justified by the notion that sparse graphs are locally tree-like. The poster will show how this notion can be made rigorous using random graph theory.

Even with the restriction to trees and only 15 or 20 nodes, checking all the graphs is still too time-consuming. MCMC and other methods can be used to approximate the posterior distribution by moving through the space of possible graphs. For forests the simplest move is to add or remove an edge, and for trees the simplest move is to reposition an edge. The poster will show how forests and trees can be stored so that these moves can easily be chosen uniformly at random and after each move the information that is stored can be updated locally, and the results of experiments to assess different ways of storing trees.

Coauthors: Vanessa Didelez

⁂

Bayesian inference with hard parameter constraints

Anthony Quinn (Trinity College Dublin, Ireland)

The paper addresses the Bayesian inference of parameters with hard constraints on their range. In particular, the design of conjugate priors in this case is studied formally, confirming that the finite sufficient statistics, if available, do not depend on the geometry of the constraint subset. This insight has important general consequences for prior design in constrained parameter problems. Two cases of constraint subset are studied: (i) proper subsets of measure greater than zero, and (ii) manifold subsets. As an example of the former, the constraint on the parameters of a stable autoregressive model is handled. Methods for efficient posterior simulation and computation are derived in this case, and predictive performance is assessed. In the latter case, we examine a simplex constraint for the parameters of the linear regression model, such as arises in the linear mixing model (LMM) for geo-sensing and functional medical imaging applications. The relative merits of a constrained Gaussian prior and a Dirichlet prior are examined in this case. Once again, an efficient posterior simulation strategy is derived, and an application in terrestrial remote sensing is presented.

⁂

Bayesian hierarchical space-time models for the North Atlantic ocean wave climate

Erik Vanem (University of Oslo, Norway)

This poster will present Bayesian hierarchical space-time models developed to analyse the spatial and temporal variability of significant wave height over an area in the North Atlantic ocean. This modelling framework allows modelling of complex dependence structures in space and time and incorporation of physical features and prior knowledge, yet at the same time remains intuitive and easily interpreted. Bayesian inference is performed by way of Markov chain Monte Carlo methods, i.e. a Gibbs sampler with additional Metropolis Hastings steps.

The models are hierarhical and contain an observation model and several state models. The different model components will be presented and the results obtained by applying the model to significant wave height data for an area in the North Atlantic ocean will be discussed. Of special interest is a purely temporal component included to model any long-term trends and possible effects of climate change. Different model alternatives and extentions have been tried out with respect to this long-term contribution, i.a. a CO2 regresion component for the long-term trends. For this extended model, projections of future ocean wave climate have been made based on different emission scenarios.

Overall, the model and its various components are found to perform reasonably well and to yield interesting results. Hence, it is believed that the results presented herein, estimating future trends in the North Atlantic wave climate, is an important contribution to the scientific discussion on climate change, with important implications for maritime, offshore and coastal safety.

⁂

Wednesday, June 27th

Estimating the Markov switching almost ideal demand system: a Bayesian approach

Satoshi Kabe (University of Tsukuba, Japan)

Allais and Nichele (2007) proposed a Markov-switching almost ideal demand system (MS-AIDS) model by extending the idea of Hamilton (1989). This model enables us to determine when the regime shifts occurred and to estimate parameters characterized across the different regimes. Moreover, degree of belongingness to each of the regimes and transitions between regimes are quantified by the probabilities. In this paper, we propose the Bayesian estimation for MS-AIDS model and illustrate the applicability of our proposed method. The Bayesian estimation has some important advantages. First, Bayesian estimation enables us to avoid the singularity problem suggested in Hamilton (1990, 1991). Second, our proposed Bayesian estimation ensures that transition probabilities have to lie between zero and one. Third, Bayesian estimation is able to avoid the messy calculations entailed in the score functions of log-likelihood. In the empirical study on the Japanese meat market, we found that our Bayesian estimation improves the mean squared errors for all meat products over the maximum likelihood estimation, while successfully capturing the regime shifts of meat demand coinciding with the timing of Bovine Spongiform Encephalopathy (BSE) cases in Japan and U.S.

Coauthors: Yuichiro Kanazawa

⁂

Asymptotic properties of Monte Carlo strategies for cumulative link model

Kengo KAMATANI (Osaka University, Japan)

For cumulative link model in Bayesian context, the posterior distribution can not be obtained in closed form and we have to resort to approximation method. Simple data augmentation strategy, widely used for that purpose, is known to work poorly. Marginal augmentation method (MA) is considered to be a remedy for it. However, according to Kamatani (Arxiv 2011), MA is also not free from poor convergence except for small number of categories. In this paper we propose two methods, two step marginal augmentation and multiple marginal augmentation. The former is free from Metropolis update and it works well (locally consistent) up to 3-categories. The latter is always locally consistent. Numerical simulation is provided to describe efficiency. We propose a practical strategy to implement these methods.

⁂

Bayesian hierarchical models for analyzing spatial point-based data at a grid level: a comparison of approaches

Su Yun KANG (Queensland University of Technology, Australia)

Spatial data are prevalent in a wide range of fields including environmental and health science. This has led to the development of a range of approaches for analysing patterns in these data. There is now a growing interest in modelling spatial point-based data at grid level, by discretizing the study region into regular grid cells which are far smaller than the common areal regions (e.g., administrative districts) and thus more geographically accurate. Spatial data may also be collected directly at grid level with the aid of geographical information systems and related software, in the form of pixel or raster data. Grid-based spatial data have the advantages of allowing the formation of a generalized linear model, and the approximation of the covariance structure by a Markov random field which makes the computation more feasible. Given these identified advantages in analysing point-based data at grid level, it is of interest to evaluate and compare models that have proposed for this purpose. In this study, we select three popular approaches, review the model and computational methodology for each approach, and then compare the approaches in terms of estimation of the underlying intensity and goodness of fit under different spatial structures. We consider dense, sparse, clustered and non-clustered point-based data, and conduct the comparison using simulated data at similar scale of discretization of the study region. We select three popular Bayesian approaches: two parametric models (log Gaussian Cox process and gamma moving average model) and a semiparametric model (adaptive Gaussian Markov random field model).

Coauthors: James McGree and Kerrie Mengersen

⁂

A Bayesian unimodal density regression model, with applications

George KARABATSOS (University of Illinois-Chicago, USA)

We investigate a Bayesian nonparametric, infinite-mixture regression model. In the model, the mixture weights are specified by an infinite ordered-category probit regression, with covariate dependence in both the mean and variance of the probit function. Each kernel of the mixture is a general unimodal density, i.e., an infinite-mixture of uniform densities, with mixture distribution assigned an independent stick-breaking prior (for e.g., a Dirichlet process, or more general stick-breaking prior). If the dependent response variable is discrete, then the unimodal kernel density is specified for the underlying latent response. In all, the regression model is a covariate-dependent infinite-mixture of unimodal infinite-mixture densities. In the regression model, the probit variance reflects the explanatory power of a covariate vector. On the one hand, the model predicts a density that becomes increasingly unimodal as the explanatory power of the covariate (vector) increases. On the other hand, the model predicts a density that becomes increasingly multimodal as the explanatory power decreases. Moreover, the model allows the explanatory power (probit variance) to vary from one covariate (vector) value to another. The regression model is illustrated through the analysis of more than 30 real data sets, and the analysis of many simulated data sets. The real data applications reflect a wide range of substantive areas, including the analysis of student performance, and psychometric analysis of test item responses. For all data sets, the new regression model is shown to have better predictive performance than many other regression models, including the Bayesian Additive Regression Trees model.

Coauthors: Stephen G. Walker

⁂

Efficient Bayesian inference for multivariate factor stochastic volatility (SV) models

Gregor KASTNER (Vienna University of Economics and Business, Austria)

Multivariate factor SV models are increasingly used for the analysis of multivariate financial and economic time series because they can capture the volatility dynamics by a small number of latent factors. The main advantage of such a model is its parsimony, where all variances and covariances of a time series vector are governed by a low-dimensional common factor with the components following independent SV models.

These approaches have recently been applied to important problems in financial econometrics such as asset allocation and asset pricing. They extend standard factor pricing models such as the arbitrage pricing theory and the capital asset pricing model. As opposed to SV factor models, standard factor pricing models do not attempt to model the dynamics of the volatilities of the asset returns and usually assume that the covariance matrix is constant. Empirical evidence suggests that multivariate factor SV models are a promising approach for capturing multivariate time-varying volatility, explaining excess asset returns, and generating optimal portfolio strategies.

For high dimensional problems of this kind, Bayesian MCMC estimation is a very efficient estimation method, however, it is associated with a considerable computational burden when the number of assets is moderate to large. To overcome this, we avoid the usual forward-filtering backward-sampling (FFBS) algorithm by sampling "all without a loop" (AWOL), consider various reparameterizations such as (partial) non-centering, and apply an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation at an univariate level, which can be applied directly to heteroscedasticity estimation for latent variables such as factors.

Coauthors: Sylvia Fruehwirth-Schnatter and Hedibert Freitas Lopes

⁂

Assessing prior distributions for the item parameters in the two-parameter logistic IRT model

Kentaro Kato (Center for Research on Educational Testing, Japan)

Item response theory (IRT) is a form of psychometric modeling. Basic IRT models represent the probability of a correct response to a test item as the probability conditional on the latent "ability" parameter of the test taker. An advantage of IRT is that it separates the item-specific property such as discrimination and difficulty, which are represented by item parameters, from the person-specific ability which is intended to be measured by the test items. One of the downsides of IRT is that it requires a large sample to obtain stable estimates of the item parameters. Bayesian estimation methods with a variety of prior distributions are available and can be used when the sample size is not adequate. In most cases neutral or mild prior distributions are sought for, but incorporating expert prior opinions could rather improve the estimation. Based on the assumption that experienced item writers or content experts could be trained to express their belief about item properties in probability, this study proposes a method to derive informative prior distributions for the item parameters under the two-parameter logistic IRT model. The method repeatedly asks an expert the probability of a correct response to each item at a certain ability level, and then transforms these correct response probabilities onto the space of item parameters to estimate their prior distributions of a certain class. A few similar methods have been proposed in the literature; strengths and weaknesses compared to the existing methods will be discussed.

⁂

Parallel particle learning for Bayesian asset price prediction

Hiroaki Katsura (Keio University, Japan)

Posterior simulation for Bayesian inference using particle filters and particle learning algorithms have proven to be successful in potently predicting asset prices and other financial measures like volatility. However, these particle based methods for posterior simulation are, by nature, computationally strenuous and more so as the model becomes more complicated. This fact has stymied its success in the practical world and is yet to replace MCMC algorithms for many researchers. However, with the recent development of fast and inexpensive devices for parallel computing, such as general purpose graphic processing units (GPGPU), formerly impractical computations, that would take hours or even days, can be completed in minutes or even seconds. With this new paradigm in mind, we have developed a new algorithm for particle learning that is fully parallelazied and thus making it suitable for GPGPU computing. Utilizing this new algorithm, we will show, using numerous financial data, how accurate and effective posterior simulation for Bayesian inference can be achieved with great alacrity and efficiency.

Coauthors: Kenichiro McAlinn and Teruo Nakatsuma

⁂

Genome-wide structural variation analysis

Peter Kecskemethy (University of Oxford, UK)

Imputation has been widely used as a tool for incorporating reference single nucleotide-polymorphism (SNP) information into association study data to improve the efficiency of association studies. As the focus of research shifts from SNPs to a wider range of genetic variation it is important that information on different polymorphism classes can be integrated and imputed jointly. The ideal situation of joint imputation would be to have a reference panel including all classes alike. Achieving such a combined reference panel raises multiple issues, such as integrating information for analysis such as genome-wide association studies.

We found that linkage disequilibrium (LD) between SNPs and short insertions and deletions (INDELs) is not very different from that observed between SNPs, and there is little reason to believe that it will be very different in the case of most other types of variation.

We propose a single locus Bayesian model for relating different types of genetic variation with the aim of integrating new genetic information into SNP haplotypes. We present results of various methods for phasing and imputation based on this model using SNP and INDEL data from the 1000 Genomes Project.

We show that acceptable accuracy can be achieved by phasing INDELs into SNP haplotypes using the LD structure between SNPs and INDELs - genotype predictions made by our phasing algorithms improved predictions compared to those based on genotype likelihood modes in the original data. We also found that our method is comparable to widely used methods both in accuracy and speed.

Coauthors: Chris Holmes and Jonathan Marchini

⁂

Application of Bayesian approach on the child's health problems

Khaled KHATAB (RWTH Aachen University, Germany)

Childhood malnutrition is the most serious health issue facing developing countries. Major progress has been made over the last 30 years in reducing the prevalence of malnutrition amongst children less than 5 years of age in developing countries (UNICEF 2008). Previous analyses of children's nutritional status have often been based on Demographic and Health Surveys (DHS). The aim of this work focuses on childhood malnutrition in Nigeria. We apply the geoadditive Gaussian latent variable model for continuous indicators in which the 3 measurements of the malnutrition status of children are assumed as indicators for the latent variable "nutritional status". In this work, a latent variable model (LVM) for the nutritional status based on continuous indicators is applied. This model permits modeling of covariate effects on the latent variables through a flexible geoadditive predictor. It gives us the opportunity to study the association or interrelationship between the 3 types of malnutrition as indicators for nutritional status. The factor loadings describe the association between these indicators and their impact on the nutritional status of a child. The results of the LVM have shown that children living in some provinces in the southeast regions and some regions in the southern part of the country are more likely to suffer from undernutrition. More attention needs to be given to some areas which have high rates of poverty, such as the southern part of Nigeria. These areas are more likely to have a higher proportion of undernutrition compared to other areas, due to poor health facilities and complications during childbirth or even careless and misdiagnosis during hospital care.

⁂

Semiparametric posterior limits under local asymptotic exponentiality

Bartek KNAPIK (VU University Amsterdam, Netherlands)

In this poster we investigate the asymptotic behavior of the marginal posterior for the parameter of interest in nonregular semiparametric estimation problems. It is known in the literature that there are few nonregular cases for which the posterior distributions converge. A crucial tool in studying the asymptotic behavior of the posterior distribution is likelihood expansion. In this work we focus on nonregular models that exhibit a so-called local asymptotic exponentiality (LAE). This setting covers important examples such as densities with jumps, uniform distributions, etc.

We show that under certain straightforward and interpretable conditions, the marginal posterior for the parameter of interest converges in total variation to a negative exponential distribution, with parameters present also in the LAE expansion. In addition to the local asymptotic exponentiality, the model is required to satisfy metric entropy conditions, while the nuisance prior must assign non-zero mass to certain Kullback-Leibler neighborhoods, like in Ghosal et al. (2000). In addition, the marginal posterior is required to converge at parametric rate (i.e. when the nuisance is fixed).

The results are applied to estimation of location in a model, where underlying densities possess a jump at zero, with a Gaussian prior for the nuisance.

Coauthors: Bas Kleijn

⁂

Stable priors for linear regression model

Genya Kobayashi (Kobe University, Japan)

Symmetric alpha stable (SAS) prior has been used as wavelet shrinker in the image processing field. SAS prior includes normal and Cauchy priors as special cases and has some desirable properties for the shrinkage problem such as spike at zero and thick tail. However, the use of SAS prior is currently quite limited, because it does not allow a closed form density. The non-existence of a closed form density leads to difficulty in estimating the hyperparameter, which plays an important role in the shrinkage problem. Furthermore, it is little known about the shrinkage rule of SAS prior. In this paper, we consider SAS prior and its extension, namely geometric stable (GS) prior for the linear regression model. GS prior is an interesting shrinkage prior, since it permits a similar scale mixture of normal representation with SAS prior and includes an important special case, double exponential prior. By looking at the shrinkage coefficients of those priors, we confirm that they achieve much more flexible shrinkage rules than Cauchy and double exponential priors. The hyperparameters of SAS and GS priors are estimated using approximate Bayesian computational method. The prior distributions are demonstrated using the simulated and real datasets.

Coauthors: Hideo Kozumi

⁂

On the loss functions of Bayesian forecasting for WWW traffic based on time varying Poisson distribution

Daiki Koizumi (Cyber University, Japan)

Traffic forecasting from past observed traffic data with small calculation complexity is one of important problems for planning servers and networks. For this problem, the author previously proposed the Bayesian forecasting algorithm from the view point of statistical decision theory. This forecasting algorithm assumes certain class of time-varying parameter to save calculation costs as much as possible. In fact, the hyperparameter estimation for his algorithm requires numerical calculation with interval between 0 and 1. Furthermore, the traffic forecasting only requires arithmetic operations if the hyper parameter is known. On the other hand, this algorithm have been assumed the quadratic loss function. In general, the both overestimation and underestimation are equally weighted on the quadratic loss function since this function is analytically symmetric. However, it is not necessary the only assumption for the traffic forecasting problem. Therefore, some asymmetric functions can be assumed for the loss function for this problem. Typically, it is possible that the overestimation is more seriously considered than its underestimation. This presentation explores the relationship between assumptions of loss functions and traffic forecasting performance with real observed data.

⁂

Comparing distributions using dependent normalized random measure mixtures

Michalis Kolossiatis (Cyprus University of Technology, Cyprus)

An area of Bayesian statistics that has attracted a lot of attention in the last years is nonparametric modelling. One important reason for this is the advancement of various simulation methods, and especially Markov Chain Monte Carlo methods. These models are particularly useful for flexible modelling of various data, for example grouped data. Here we describe a methodology for the simultaneous Bayesian nonparametric modelling of several distributions, which are assumed to be dependent. Our approach makes use of normalised random measures with independent increments and builds dependence through the superposition of shared processes. We present the properties of these priors, as well as the modelling possibilities in this framework. Regarding posterior analysis, we develop efficient slice sampling simulation methods. We also introduce a new graphical method for comparing pairs of distributions, which allows better understanding of the differences between the dependent distributions. The proposed models are finally applied to a stochastic frontier setting and used to analyse the efficiency of some hospital firms, with interesting results.

Coauthors: Jim Griffin and Mark Steel

⁂

Bayesian modeling of temporal dependence in large sparse contingency tables

Tsuyoshi KUNIHAMA (Duke University, USA)

In many applications, it is of interest to study trends over time in relationships among categorical variables, such as age group, ethnicity, religious affiliation, political party and preference for particular policies. At each time point, a sample of individuals provide responses to a set of questions, with different individuals sampled at each time. In such settings, there tends to be abundant missing data and the variables being measured may change over time. At each time point, one obtains a large sparse contingency table, with the number of cells often much larger than the number of individuals being surveyed. To borrow information across time in modeling large sparse contingency tables, we propose a Bayesian autoregressive tensor factorization approach. The proposed model relies on a probabilistic Parafac factorization of the joint pmf characterizing the categorical data distribution at each time point, with autocorrelation included across times. Efficient computational methods are developed relying on MCMC. The methods are evaluated through simulation examples and applied to social survey data.

Coauthors: David B. Dunson

⁂

Dynamic equicorrelation stochastic volatility

Yuta KUROSE (University of Tokyo, Japan)

A multivariate stochastic volatility model with dynamic equicorrelation and cross leverage effect is proposed and estimated. Using a Bayesian approach, an efficient Markov chain Monte Carlo algorithm is described where we use the multi-move sampler, which generates multiple latent variables simultaneously. Numerical examples are provided to show its sampling efficiency in comparison with the simple algorithm that generates one latent variable at a time given other latent variables. Furthermore, the proposed model is applied to the multivariate daily stock price index data.

Coauthors: Yasuhiro Omori

⁂

Construction of an informative hierarchical prior distribution for electricity load forecasting

Tristan LAUNAY (Université de Nantes, France)

The EVENTAIL model is a non linear Gaussian model currently used to estimate and forecast the electricity load in France. The load can be decomposed as a sum of two components : the seasonalities and the weather dependant part. Due to its high dimensionality, the model usually requires 4 to 5 years worth of data to compute reasonable predictions.

We are interested in the estimation of this basic model on a subset of the French population, the precise monitoring of which started only recently : the history of data available for estimation is thus very short, typically around a year worth of data.

To overcome the difficulty, we use the estimation of the model on another population, who has been monitored for a longer time. We assume both datasets (the short and the long) are somehow similar and introduce hyperparameters into a hierarchical prior to model and estimate this similarity. We first apply this methodology on simulated datasets to ensure its validity and then show that using it upon real datasets leads to a substantial improvement of the predictions' quality.

Coauthors: Anne Philippe and Sophie Lamarche

⁂

Bayesian vertex nomination

Dominic LEE (University of Canterbury, New Zealand)

Consider an attributed graph G with n vertices, each of which is colored green or red, but only k vertices are observed to be red. The color of the other vertices is unobserved. Typically, the unknown total number, m, of red vertices is small, satisfying $k < m \ll n$. The vertex nomination problem is to nominate one of the unobserved vertices as being red. The edge set of G is a subset of the set of unordered pairs of vertices. Suppose that each edge is also colored green or red and this is observed for all edges. For a vertex, v, define its context statistic as the number of observed red vertices connected to v; and its content statistic as the number of red edges incident to v. Assuming that these statistics are independent between vertices and that red edges are more likely between red vertices, a likelihood model is formulated, with prior distributions chosen for the unknown parameters and unobserved vertex colors. From the resulting posterior distribution, the nominated vertex is the one with the highest posterior probability of being red. Inference with the model is conducted using a Metropolis-within-Gibbs algorithm, and performance is illustrated by Monte Carlo simulation experiments. An application example is provided using the Enron email corpus, where vertices represent Enron employees, observed red vertices are known fraudsters, red edges represent email communications perceived as fraudulent, and we wish to identify one of the latent vertices as most likely to be a fraudster.

Coauthors: Carey Priebe

⁂

Threshold selection method for modelling multivariate extremes using a Bayesian measure of surprise

Kate LEE (Auckland University of Technology, New Zealand)

In modelling exceedances over the threshold, the threshold selection plays a key role in fitting the tail of extremal data. For multivariate extremes, the spectral density function estimates the dependency of spatial extremes and the joint threshold determines the set of extremes whilst the spatial structure is still hold. In this poster, we adopt a Bayesian measure of surprise (MS) for the threshold selection problem and the simulation studies using real data are provided. Semiparametric and parametric models are considered for the spectral density, and the posterior predictive p-value is used for the MS estimate.

Coauthors: Scott Sisson and Yana Fan

⁂

Multiply imputing missing values in data sets with mixed measurement scales using a sequence of generalized linear models

Min LEE (University of Southampton, UK)

Missing data are a common problem in many fields which can complicate statistical analyses of interest. Multiple Imputation is a commonly used approach to deal with missing values. In this approach, an imputer repeatedly imputes the missing values from their posterior predictive distribution to generate multiple completed data sets; these data sets are then released to the analysts. Analysts can now perform the same standard analysis that would have been performed on fully observed data, on each of these completed data sets. The analyst can then combine the analysis from the different data sets using simple combining rules and make inferences that take the uncertainty due to missing values into account. It is crucial that imputations are drawn from an appropriate distribution to preserve relationships present in the data and allow valid conclusions to be made from any analysis. In data sets with different types of variables, e.g. some categorical and some continuous variables, Multivariate Imputation by Chained Equations (MICE) is a commonly used multiple imputation method; however, imputations from such an approach are not necessarily drawn from a proper posterior predictive distribution. We propose a method to multiply impute missing values in such data sets by modelling the joint distribution of the variables in the data through a sequence of generalised linear models. We use data augmentation methods to draw imputations from a proper posterior distribution using Markov Chain Monte Carlo (MCMC). We compare the performance of our method with MICE using simulations and a breast-feeding study.

Coauthors: Robin Mitra

⁂

Quadratic scaling models for spatial compositional data with application to forest fragmentation and land use/land cover classification

Thomas Leininger (Duke University, USA)

We introduce two spatial regression models for compositional data with zeros in the components. Current methods either impute small values for the zero components or use transformations with complicated interpretations of model and regression parameters. For data with a large proportion of zeros, we propose a model which allows zero inflation using a latent process which takes on negative values when the corresponding observed components are zero. For data with a smaller proportion of zeros, we propose a model based on the projected normal distribution which admits zeros but does not place any probability mass at zero. Both methods result in a tractable likelihood which easily incorporates typical regression and spatial methods and provides meaningful interpretations of regression parameters. We illustrate the performance of the models on land use/land cover and forest fragmentation data in the Northeastern U.S. and compare the results to existing methods.

Coauthors: Alan E. Gelfand, Jenica M. Allen and John A. Silander, Jr.

⁂

Adaptive approximate Bayesian computation for complex models

Maxime Lenormand (IRSTEA, France)

Approximate Bayesian computation (ABC) is a family of computational techniques in Bayesian statistics. These techniques allow to fit a model to data without relying on the computation of the model likelihood. They instead require to simulate a large number of times the model to be fitted. A number of refinements to the original rejection-based ABC scheme have been proposed, including the sequential improvement of posterior distributions. This technique allows to decrease the number of model simulations required, but it still presents several shortcomings which are particularly problematic for costly to simulate complex models.

We here provide a new algorithm to perform adaptive approximate Bayesian computation. We present a modification of the PMC-ABC algorithm proposed by Beaumont et al. (2009). We compare this new algorithm with the population Monte Carlo ABC algorithm of Beaumont et al. (2009), the replenishment SMC ABC algorithm of Drovandi and Pettitt (2011) and the adaptive SMC ABC algorithm of Del Moral et al. (2011) on a toy example. Finally, we apply our new algorithm to a complex individual-based social model, the PRIMA model. We show that our algorithm outperforms the three other algorithms in the two applications by requiring less simulations to reach the same posterior density quality.

References

Beaumont, M. A., Cornuet, J., Marin, J., Robert, C. P.: Adaptive approximate Bayesian computation. Biometrika. 96(4),983–990 (2009)

Drovandi, C. C., Pettitt, A. N.: Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics. 67(1),225–233 (2011)

Del Moral, P., Doucet,A., Jasra,A.: An adaptive sequential Monte Carlo method for approximate Bayesian computation. Statistics and Computing. (2011)

Coauthors: Franck Jabot and Guillaume Deffuant

⁂

Efficient Bayesian multivariate surface regression

Feng Li (Stockholm University, Sweden)

Methods for choosing a fixed set of knot locations in additive spline models are fairly well established in the statistical literature. While most of these methods are in principle directly extendable to non-additive surface models, they are likely to be less successful in that setting because of the curse of dimensionality, especially when there are more than a couple of covariates. We propose a regression model for a multivariate Gaussian response that combines both additive splines and interactive splines, and a highly efficient MCMC algorithm that updates all the multivariate knot locations jointly. We use shrinkage priors to avoid overfitting with different estimated shrinkage factors for the additive and surface part of the model, and also different shrinkage parameters for the different response variables. This makes it possible for the model to adapt to varying degrees of nonlinearity in different parts of the data in a parsimonious way. Simulated data and an application to firm leverage data show that the approach is computationally efficient, and that allowing for freely estimated knot locations can offer a substantial improvement in out-of-sample predictive performance.

Coauthors: Mattias Villani

⁂

Bayesian methods for subnational mortality forecasts

Guangquan Li (Imperial College London, UK)

Mortality forecasts are used extensively by governments for health spending allocation and social security planning and by financial institutes for pricing life insurance/annuity. While national forecasts are produced routinely, work on subnational mortality forecasts is limited, partly due to data availability and various methodological challenges. Data sparsity, amongst others, is a key issue to deal with when analysing small area data. Bayesian hierarchical models are well suited for this situation as they allow for borrowing information across various strata defined by, for example, age and area. In particular, we extend the Age-Period-Cohort modelling framework (a) to encompass the spatial dimension and (b) to allow for flexible temporal smoothing functions (e.g., spline and random walk models), which are the main driving force for forecasting. While these temporal smoothing functions are constructed to be stratum-specific, the associated parameters (the spline coefficients for example) are modelled hierarchically; in addition to yielding more reliable estimates and hence forecasts, this setting follows the spirit of coherent forecasting for different populations (Li and Lee, 2005). Under the Bayesian framework, forecast uncertainty can be better estimated than the conventional principal component methods, the uncertainty intervals from which tend to be too narrow. Using the all-cause mortality data in England and Wales at the Local Authority District level (376 in total), we present comparisons of the forecasting performance of various Bayesian models as well as of some conventional methods through a cross validation study with different lengths for the validation period (5-/10-years ahead). We also explore the use of INLA/PyMC for implementing these Bayesian models in practise.

Coauthors: Nicky Best and Majid Ezzati

⁂

Longitudinal actor-partner interdependence models for studies on HIV-impacted families

Li-Jung Liang (University of California, Los Angeles, USA)

More and more studies focus on HIV-impacted families, including persons living with HIV (PLH) and also other family members. It is important to develop an analytical approach to properly analyze such complex family-system data. Several challenges have been identified: (1) longitudinal observations per person within families and covariates that differ by group in nested designs are encountered, (2) longitudinal data, including outcomes and predictors of interest, collected from multiple members of a family are likely to be correlated, and (3) one person's predictor variable may influence not only that person's outcome variable, but also that person's family members' outcome variables. We propose an analytical approach that uses a Bayesian framework to combine a bivariate growth-curve model with an actor-partner interdependence model. Both models have been used in some research fields (e.g., psychology), but have not been applied in analyzing longitudinal data from the HIV-impacted family studies. The combined approach will allow researchers to assess the trajectories of outcomes and predictors for PLH and family members jointly in a single hierarchical model, to incorporate covariates that differ by PLH and family member, to account for clustering within subjects and within families, and to investigate whether the PLH or family members' predictors influence their own or/and partners' trajectories of outcomes. Our approach will be applied to a longitudinal study on HIV-impacted families in China.

Coauthors: Li Li

⁂

Revealed preferences for portfolio selection — does skewness matter?

Merrill Liechty (Drexel University, USA)

We take a critical look at the paradigm of mean/variance efficient portfolios. Posing the portfolio selection problem as a decision problem we hope to show how reasonable assumptions on utility and the probability model lead to asset allocation rules that depend on mean, variance and skewness of future returns. Our main goal is the empirical validation of this argument by comparing it with traditional mean variance efficient portfolios. We use a fair benchmark for comparison, i.e. we set up a revealed preference study where we develop a framework to explain observed investor preferences by the two alternative utility functions. Minimizing the discrepancy between the optimal decision under the considered utility functions and the observed data formalizes the comparison.

In other words, we implement the inverse problem of expected utility maximization. Given observed decisions we back out inference about the underlying probability model and utility function.

⁂

A study of feedforward and feedback EWMA controller for multistage processes

Chien-Hua LIN (Providence University, Taiwan)

Run-to-run (R2R) process control has well developed in semiconductor manufacturing communities. The exponentially weighted moving average (EWMA) feedback controller is a popular model-based controller which primarily uses data from past process run to adjust the input recipe for the next run and it has received great attention in literatures. However, IC manufacturing is a hybrid and multistage process that includes cleaning, layering, patterning, etching, deposition, planarization, etc. Each process is highly correlated with others such as the deposition time is deeply affected by the depth and width of etching process. Focusing on the known information (such as the depth of etching) in previous process, and assuming that the process disturbance follows a general ARIMA (p, d, q) series, we propose a systematic approach that combines feedforward controller and double EWMA feedback controller to adjust IC processes. First, the long-term stability and short-term output performance are addressed. Furthermore, under the optimal discount factor of double EWMA controller, a simulation study is presented to illustrate that the proposed controller has the capability of reducing the total mean square error (TMSE) of process response variable by comparing with double EWMA controller. Finally, a sensitivity analysis is addressed to the impacts of parameters on the process TMSE.

Coauthors: Bo-Yuan Huang

⁂

Bayesian estimation of smoothly mixing time-varying parameter GARCH models

Edward M. H. LIN (Feng Chia University, Taiwan)

In this research, asymmetric time-varying (TV) GARCH models via the asymmetric logistic functions mechanism are proposed, which are incorporated into the conditional volatility equation for capturing smooth volatility asymmetries in the financial time series. The proposed models allow smooth transitions for different break points. A Bayesian computational method is employed to identify the locations of structural breaks, and for estimation and inference, simultaneously accounting for heteroskedasticity. The proposed methods are illustrated using simulated data, and an empirical study also provides evidence for smooth asymmetric time-varying volatilities in two international stock market return series. In addition, marginal likelihood is used to choose the break number in the TV-GARCH model. We find significant improvements in fitness in asymmetric smooth transition TV-GARCH models.

Coauthors: Cathy W. S. Chen, Richard Gerlach and Edward M.H. Lin

⁂

Ticket vending service and customer waiting: evidence from Taiwan high speed rail system

Pei-Chun LIN (National Cheng Kung University, Taiwan)

Waiting-line experience affects the overall perceptions of the quality of service provided. This study aims to investigate the appropriate timing of adding extra ticket vending service window to shorten the queue length of waiting passengers. Therefore, this study contributes to propose an analytical optimization model to deal with a waiting line being an integral part of the critical skill of time management of High Speed Rail system operator. The threshold of queuing length in the waiting line for deciding whether additional ticket vending window are open and closed could be estimated. Several managerial implications are incorporated as a useful reference for the transportation system operator.

Coauthors: Yu-Shian Cheng

⁂

Effect modification using latent mixture analysis

Tanzy LOVE (University of Rochester, USA)

The Seychelles Child Development Study (SCDS) is examining associations between prenatal exposure to low doses of methylmercury (MeHg) from maternal fish consumption and children's developmental outcomes. Whether MeHg has neurotoxic effects at low doses remains unclear and recommendations for pregnant women and children to reduce fish intake may prevent a substantial number of people from receiving sufficient nutrients that are abundant in fish. The primary findings of the SCDS are inconsistent with adverse associations between MeHg from fish consumption and neurodevelopmental outcomes. However, whether there are subpopulations of children who are particularly sensitive to this diet is an open question. Secondary analysis from this study found significant interactions between prenatal MeHg levels and both caregiver IQ and income on 19 month IQ. These results are dependent on the categories chosen for these covariates and are difficult to interpret collectively.

Here, we estimate effect modification of the association between prenatal MeHg exposure and 19 month IQ using a general formulation of mixture regression. Our model creates a latent categorical group membership variable which interacts with MeHg in predicting the outcome. We also fit the same outcome model when in addition the latent variable is assumed to be a parametric function of three distinct socioeconomic measures. Bayesian MCMC methods allow group membership and the regression coefficients to be estimated simultaneously and our approach yields a principled choice of the number of distinct subpopulations. The results show three different response patterns between prenatal MeHg exposure and 19 month IQ in this population.

Coauthors: Sally W. Thurston and Philip W. Davidson

⁂

Bayesian recursive model selection

Li Ma (Duke University, USA)

In this work we introduce a new model space prior for the Bayesian variable selection problem. This prior is designed based on a recursive constructive procedure that randomly generates models by including variables in a stagewise fashion. While the procedure is nominally similar to the classical forward-stagewise variable selection approach, it constitutes a coherent probability model on the space of regression models. Consequently, inference based on this prior avoids difficulties such as overfitting that are commonly faced by greedy methods. We provide a recipe for carrying out Bayesian variable selection using this prior, and show that it possesses several desirable features for the problem of variable selection. In particular, it is conjugate—the posterior is the same stepwise procedure. Moreover, the posterior parameters can be computed analytically through a sequence of recursive computation. This property is particularly desirable in high dimensional settings as it provides an alternative approach for exploring the posterior model space without resorting to Markov Chain Monte Carlo (MCMC), whose convergence behavior is often hard to guarantee and difficult to evaluate in such cases. In addition, the prior also allows flexible ways to incorporate structural features of the model space such as the dependence structure among the predictors. In particular, we illuatrate how the prior can be specified to take into account model space redundancy arising from strong correlation among potential predictors.

⁂

Clustered Bayesian model averaging

Steven MacEachern (Ohio State University, USA)

It is sometimes preferable to conduct statistical analyses based on the combination of several models rather than on the selection of a single model, thus taking into account the uncertainty about the true model. Models are usually combined using constant weights that do not distinguish between different regions of the covariate space. However, a procedure that performs well in a given situation may not do so in another situation. In this paper, we propose the concept of local Bayes factors, where we calculate the Bayes factors by restricting the models to subregions of the covariate space. The covariate space is split in such a way that the relative model efficiencies of the various Bayesian models are about the same in the same region while differing in different regions. An algorithm for clustered Bayes averaging is then proposed for model combination, where local Bayes factors are used to guide the weighting of the Bayesian models. Simulations and real data studies show that clustered Bayesian averaging results in better predictive performance compared to a single Bayesian model or Bayesian model averaging where models are combined using the same weights over the entire covariate space.

Coauthors: Steven N. MacEachern and Mario Peruggia

⁂

Bayesian disease mapping with sparse data and smoothing — Poisson or zero-inflated Poisson?

Ying MacNab (University of British Columbia, Canada)

Spatial and spatiotemporal disease mapping data are often spatially sparse and structurally multidimensional. In this paper, the characteristics of “sparse data” (e.g. small counts, noisy rates, and excessive number of areas with zero case-observation) and “multidimensionality” (e.g. response(s) of spatial, temporal, and multivariate in nature) are discussed in the context of Bayesian disease mapping. Multidimensional smoothing for mapping rare disease or health outcomes, such as autoregressive spatial smoothing and cross-component smoothing for multivariate disease mapping as well as autoregressive spatial smoothing and spline temporal smoothing for spatiotemporal disease mapping, are discussed for Poisson and zero-inflated Poisson models. Several real-life examples of small area disease risks prediction and inference under Bayesian hierarchical spatial and spatiotemporal Poisson and zero-inflated Poisson models are presented.

⁂

Short term traffic flow forecasting with spatial temporal model

Tiep Mai (Trinity College Dublin, Ireland)

Short-term Traffic Flow Forecasting (STFF), the process of predicting future traffic conditions based on historical and real-time observations, is an essential aspect of Intelligent Transportation Systems (ITS). Even though ARIMA has been widely used as one of the most precise methods in such a problem, it has several weaknesses: the exponential decaying in multi-step prediction; The slow computation of its multivariate version (VARIMA) due to intractable MA part. Furthermore, the implicit trend constituted by the difference or seasonal difference operator is not adaptive to dynamic scenarios such as blocking incident, traffic accident. Hence, in this work, a spatial temporal model is proposed and the joint forecast is performed on traffic flow and vehicle number of the link, another important network indicator. Using the additional variable, vehicle number, allows better forecast and adaption to the dynamic traffic scenarios. Applying this model on an expanded traffic network enables the prediction of spatial effect from one source to another, which is not accomplished by previous models. Real time prediction is obtained by parallel and independent inference on separate links and junctions.

Coauthors: Simon Wilson and Bidisha Ghosh

❈

A new Bayesian LASSO

Himel Mallick (University of Alabama at Birmingham, USA)

Regularization methods for co-efficient estimation and variable selection in linear regression, especially those based on the LASSO of Tibshirani (1996), have enjoyed a great deal of applicability in recent years. The LASSO estimate corresponds to the posterior mode when independent Laplace priors are assigned on the regression co-efficients. Park and Casella (2008) provided the Bayesian LASSO using scale mixtures of normal priors for the parameters and independent exponential priors on their variances. In this paper, we propose a new hierarchical representation of Bayesian LASSO. The new representation is based on the characterization of a Laplace distribution as a scale mixture of uniform, the mixing distribution being a particular gamma distribution. Our method borrows all the major advantages of hierarchical models. Firstly, as a Bayesian method, the estimates can be easily interpreted, resulting in easier statistical inference. Secondly, it provides valid standard errors as well as interval estimates (Bayesian credible intervals). We consider a fully Bayesian treatment which leads to a new efficient Gibbs sampler in the MCMC estimation with tractable full conditional distributions. Furthermore, we develop a fast expectation-maximization (EM) algorithm to estimate the co-efficients, extending the latent weighted least squares (LWLS) estimator (Walker et al, 1997) to penalized regression setting. Finally, we show how our approach can be extended to generalized linear models and Bayesian versions of other regularization methods e.g. Bridge estimator, Adaptive LASSO etc. We compare the performance of our method to the existing one as well as to its frequentist counterpart using simulations and real data.

Coauthors: Nengjun Yi

❈

On marked duration models and the study of price setting strategies

Juan Carlos Martínez-Ovando (Banco de México, Mexico)

Motivated by the study of the price setting behavior of Mexican firms, we introduce a new stationary dependent marked duration model and develop some of its properties. The model accounts to capturing some stylized facts observed in micro price data, such as duration heterogeneity and asymmetric price changes across firms. Model construction and inference follows recent Bayesian ideas on latent processes modelling. We illustrate this model with the study of micro data underlying the consumer and producer price Mexican indices and contrast the results with existing time-dependent models for price setting under heterogeneity. We conclude with a discussion of possible model extensions and some open statistical problems related to the macroeconomic implications for the understanding of inflation dynamics.

❈

Robust Bayesian variable selection with sub-harmonic priors

Yuzo MARUYAMA (University of Tokyo, Japan)

We study Bayesian variable selection in linear models with general spherically symmetric error distributions. We propose sub-harmonic priors which arise as a class of mixtures of Zellner's g-priors for which the Bayes factors are independent of the underlying error distribution, as long as it is spherically symmetric. Because of this invariance to spherically symmetric error distribution, we refer to our method as a robust Bayesian variable selection method. We demonstrate that our Bayes factors have model selection consistency and are coherent. We also develop Laplace approximations to Bayes factors for a number of recently studied mixtures of g-priors that have appeared in the literature (including our own) for Gaussian errors. These approximations, in each case, are given by the Gaussian Bayes factor based on BIC times a simple rational function of the prior's hyper-parameters and the R^2's for the respective models. We also extend model selection consistency for several g-prior based Bayes factor methods for Gaussian errors to the entire class of spherically symmetric error distributions.

Coauthors: William E. Strawderman

⁂

A Bayesian sequential approach in model discrimination

Samira MASOUMI (University of Waterloo, Canada)

Model discrimination methods can be used to determine the underlying mechanism of a physical system when different mechanistic models are considered as rival models. Candidate models may be derived based on different assumptions about the mechanism of the real system. Obtaining the data needed from the real system is one of the challenges particularly in applications where experiments are expensive or time consuming. Therefore, it is beneficial to get the maximum information possible from the real system using the least possible number of experiments. Several model selection methods have been presented before. The advantage of the current method is its ability to handle nonlinear systems and the combination of an experimental design procedure with a model selection step. In any iteration, the Bayesian sequential model discrimination method designs the next experiment which provides information which highlights the differences among the rival models. The model selection analysis then combines the new experiment with previous data and updates the model probabilities using the Bayesian approach. The model selection procedure used in this current method is based on sampling from the posterior probability of parameters by adaptive random walk metropolis method. The promising results of case studies demonstrate the benefits of using this method.

Coauthors: Thomas A. Duever and Park M. Reilly

⁂

Nonlinear statistical modeling via the relevance vector machine and its application to change point analysis

Kazuki MATSUDA (Chuo University, Japan)

The relevance vector machine introduced by Tipping (2001; Journal of Machine Learning Research 1) has emerged as a useful tool for constructing sparse regression models and classifiers, which sets many weights to zero in the estimated models. The relevance vector machine achieves sparsity by specifying a Gaussian prior over the weights in the nonlinear regression model expressed as a weighted sum of basis functions. This attractive feature relies on appropriate choice of basis functions and tuning parameters, which essentially control the model complexity. Choosing the optimal model from a predictive point of view by applying cross-validation can result in computational difficulties and often yields unstable model estimates. In order to construct more stable and flexible models via the relevance vector machine, we propose nonlinear regression models with multi-overlapping basis functions. We also introduce nonlinear modeling procedures based on discontinuous basis functions for determining whether a change has taken place. The proposed procedure is applied to analyze the time series data of radioactivity released into the atmosphere at the Fukushima Daiichi nuclear disaster. Monte Carlo simulations are conducted to examine the properties of our sparse Bayesian regression modeling.

Coauthors: Sadanori Konishi

⁂

Predictive distribution of SQL injection attacks detection model

Takeshi MATSUDA (Cyber University, Japan)

Recently, SQL Injection attacks have been observed frequently, and a lot of prevention and detection methods against SQL Injection attacks have been developed to prevent Web applications driven database. At the same time, attackers continue to make new evasion techniques. As it stands now, there are many cases that web application firewall has been introduced to prevent database of web applications. However, Web application firewall detects SQL injection attacks using the black list based on the observed attack samples. Therefore, the black list of web application firewall must be updated when a newly-developed attacks. Because of that, the black list balloons, and the time cost involved with detecting attacks may increase. So, it is very important problem to develop the new detection method independent of the black list. In this study, we proposed the SQL injection attack detection model based on a binomial distribution. To extract an attack feature, we focused attention on some symbols including in an input strings. In our previous study, we had proposed the selection algorithm of symbols may be used as attack features. Using the selection algorithm of symbols, we computed the Bayesian predictive distribution of our proposed model. To obtain predictive distribution, we use the beta distribution as the natural conjugate prior distribution of the binomial distribution. We will discuss on the effectiveness of our proposed model by using test SQL injection attack samples, and compare with our previous study results .

Coauthors: Michio Sonoda, Daiki Koizumi and Shigeishi Hirasawa

⁂

Fast approximate inference for Bayesian mixture models with application to volume estimation of tissue types from computed tomography

Clare MCGRORY (University of Queensland, Australia)

The use of Bayesian Mixture Models for analysing computed tomography (CT) scan data has resulted in significant improvements in the analysis and interpretation of experimental data in relation to volume calculation of tissue and longitudinal responses to various treatment regimes in terms of sheep nutrition. The use of Markov Chain Monte Carlo (MCMC) techniques to estimate tissue component membership via Bayesian mixture models provides a reliable and accurate estimation method. However, in order to enable this modelling to be carried out on a regular basis, which is essential in order for it to be of practical use to the Australian sheep meat industry, a fast to implement approach is needed. Exploring techniques for increasing the implementation speed of an MCMC algorithm is one possibility, but another option is to explore the use of efficient approximate inference techniques as we do here. We explore the use of a fast variational Bayes approach for volume estimation of tissue types from CT scans of live sheep and compare results with those obtained from using an MCMC-based approach.

Coauthors: Clair Alston and Christopher Strickland

⁂

Gibss sampling estimation method for subscores estimation using bi-factor multidimensional item response theory

Zairul Nor Deana MD DESA (University of Kansas and Universiti Teknologi Malaysia, USA)

One of the widely used probabilistic models for large-scale educational testing data is item response theory (IRT). IRT models maximize the information obtained from items in a test regarding the ability of the examinees. This paper empirically evaluates a random sample of 1,337 examinees' item responses from the mathematics assessment in the Trends in International Mathematics and Science Study (TIMSS 2007). Subscores are estimated from the pairing of multidimensional IRT and bi-factor models. Both compensatory and partially compensatory multidimensional IRT (Reckase, 1985, 1997, 2009) are defined. A Monte Carlo simulation study with 1,500 examinees is carried out to examine different test lengths and different levels of discriminations between secondary and primary abilities. The Markov Chain Monte Carlo (MCMC) with the Gibbs sampling method is used for parameter estimations using OpenBUGS. Bayesian convergence diagnostics are studied and measurement of the accuracy and precision of the parameters estimations (e.g. absolute bias, root mean squared error and standard error of estimation) will be reported. Finally, the estimated subscores will be quantified for evaluation of reliability and validity.

Coauthors: William P. Skorupski

⁂

Stability analysis of the EWMA controller for mixed-product processes

Hsin-Chao MI (National Tsing Hua University, Taiwan)

The exponentially weighted moving average (EWMA) feedback controller is a popular model-based run-to-run (R2R) controller that primarily uses data from previous process runs to adjust the settings of the next run. The long-term stability conditions and the transient performance of the EWMA controller have received considerable attention in literature. Most studies on this topic assume that there is only one single product in the manufacturing line. Generally, there are many different products produced in the manufacturing line, which ensures the production idle time as little as possible. Even though a few studies adopted EWMA controller to address this control issue, they simply assumed that the process disturbance came from some specific time series model. Therefore, aiming to analyze the long-term stability conditions of the commonly used EWMA controller under the case that the process disturbance follows a generalized time series model, this study derives an analytical expression for the process output and addresses the problem of long-term stability conditions successfully.

❉

Bayesian methods for modeling phenotypic time-series and whole-genome sequence data for bacteria

Minna MIETTINEN (University of Helsinki, Finland)

Accelerating advances in measurement technology have enabled an explosion of information available on genetic and phenotypic characteristics of bacteria. Inexpensive whole-genome sequencing has just started to shape our detailed understanding about transmission patterns and how rapidly bacterial populations can respond to environmental pressures, e.g., by vaccines and antibiotics. Recently introduced cell-based phenotypic microarrays can be used to simultaneously assess genotype-phenotype relationships of bacterial, fungal, and animal cells for almost 2,000 different metabolic substrates. These arrays are being used in a wide range of applications including basic research in cancer, metabolic and mitochondrial disorders, nutrition, toxicology, drug discovery, and bioprocess development. The success of PM technology is based on simple and efficient testing of cell properties. To enable appropriate statistical analyses of factorial experimental designs producing such data, we have developed an analysis pipeline where mixture-mixed models are fitted to the phenotypic time-series data using a combination of different computational approaches including an EM-algorithm and rapid numerical curve fitting. Results from such an analysis can be linked to various properties of the fully sequenced genomes of the bacteria included in the phenotypic experiments. To make our statistical tools widely available for biologists, an open source R package implementing the pipeline will be made available in the near future.

Coauthors: Shubin Mikhail, Corander Jukka

❉

Redefining the momentum strategy: a Bayesian approach

Masaki MISONOU (Keio University, Japan)

Momentum strategies have been a persistent and effective factor in practical and academic financial models. In fact, for a long time, the de facto standard in finance has been the four factor model (Fama-French's three factors and the momentum factor); otherwise called the Carhart four factor model. This model is often viewed as the benchmark model for comparing new models and the one to beat. Although the effectiveness of the momentum strategy has been proven through numerous academic research and years of enduring the financial market, there has yet to be a coherent concept and model to formulate this phenomenon other than by rule of thumb (e.g. buy/sell based on 12 month winners/losers). In this research, we intend to formulate the momentum effect, through the Bayesian approach, by mixing data with different frequencies. In particular, we define the momentum effect (especially short-term momentum) as a prior based on high frequency data to explain the momentum effect of more low frequency data. In our numerical example, we examine how the inclusion of the prior momentum can improve the predictive ability of other frequentist models.

Coauthors: Kenichiro McAlinn and Teruo Nakatsuma

❉

Large-scale approximate Bayesian inference for exponential family latent Gaussian models

Shakir MOHAMED (University of British Columbia, Canada)

Latent Gaussian models (LGMs) such as factor analysis, principal components analysis, and discrete-choice models, form the core of statistical tools for solving problems ranging from search and recommendation to genome-wide association studies. Many real world datasets contain noisy and sparse observations, e.g., movie rating datasets have users with a different number of observed ratings, making the uncertainty associated with them important. In these situations, conducting Bayesian analysis is important since the posterior distribution over latent variables allows for improved predictions and the marginal likelihood provides a principled approach to model selection. However, this is challenging since the data size may be extremely large, making fully Bayesian inference using MCMC prohibitive.

In this work, we make the following contributions. (1) We develop a unified variational EM approach for exponential family LGMs that learns approximate posterior distributions by maximizing concave lower bounds to the log-marginal likelihood. We demonstrate cases where efficient numerical optimization can be used to speed up convergence. Our approach gives comparable performance to MCMC but takes much less time. (2) For data vectors with many missing entries, we derive an efficient version of our EM algorithm with computational complexity that depends only on the number of observed entries, but is independent of the length of the data vectors. This allows us to apply our methods to many large datasets, such as Netflix. (3) Finally, we embed our algorithm in an online learning framework that allows our method to scale well with a large number of data vectors.

Coauthors: Emtiyaz Khan and Kevin Murphy

⁂

Bayesian latent factor regression for functional and longitudinal data

Silvia MONTAGNA (Duke University, USA)

In studies involving functional data, it is commonly of interest to model the impact of predictors on the distribution of the curves, allowing flexible effects on not only the mean curve but also the distribution about the mean. Characterizing the curve for each subject as a linear combination of a high-dimensional set of potential basis functions, we place a sparse latent factor regression model on the basis coefficients. We induce basis selection by choosing a shrinkage prior that allows many of the loadings to be close to zero. The number of latent factors is treated as unknown through a highly-efficient, adaptive-blocked Gibbs sampler. Predictors are included on the latent variables level, while allowing different predictors to impact different latent factors. This model induces a framework for functional response regression in which the distribution of the curves is allowed to change flexibly with predictors. The performance is assessed through simulation studies and the methods are applied to data on blood pressure trajectories during pregnancy.

Coauthors: Silvia Montagna, Surya T. Tokdar, Brian Neelon and David B. Dunson

⁂

A Bayes analysis of the censored Rayleigh model when using a generalised hypergeometric as a prior

Paul J. MOSTERT (University of Stellenbosch, South Africa)

Based on a type II censored sample, Bayesian estimation for the scale parameter of the Rayleigh model is carried out under the assumption of the squared error loss function. A generalised hypergeometric distribution is introduced as a prior, and beta special cases are examined. Numerical examples are used to illustrate the effect of this parameter-rich family of priors. A simulation study was carried out to investigate the sensitivity of this prior distribution and to show the effect on robustness. Prediction bounds are derived for the lifetimes of unused components (two-sample prediction).

Coauthors: Andriette Bekker and Kotie Roux

⁂

Monte Carlo inference for partial orders

Alexis MUIR WATT (University of Oxford, UK)

We introduce a new time-series model for an evolving hierarchy parameterized by partial orders. A partial order on a set P corresponds to a transitively closed, directed acyclic graph or DAG h(P) with vertices in P. Such orders generalize hierarchies defined by partitioning the elements of P and ranking the elements of the partition. In the following h(P) represents a social order on individuals P and the data is a list of random linear extensions of suborders of an unobserved partial order h(P), and we wish to make inference for the partial order.

Following (Mogapi et al. 2010), we specify a model for random partial orders with base measure the random k-dimensional orders (Winkler 1985), and a parameter controlling the typical height of a random partial order. We extend the static model to a Hidden Markov Model in which the hierarchy and its members evolve in time. The partial order h(P) evolves according to a hidden Markov process which has the static model as its equilibrium: singleton events re-order individual nodes while change-point events re-sample the entire partial order. The process is observed by taking random linear extensions from suborders of h(P) at a sequence of sampling times. The sampling times are uncertain (up to an interval). The posterior distribution for the unobserved process and parameters, which is determined by the HMM, is doubly-intractable. We use the new hybrid MCMC/particle-filter approach of (Andrieu et al. 2010) in order to carry out Monte Carlo based inference.

Coauthors: Geoff Nicholls

⁂

An improved Bayesian semiparametric model for palaeoclimate reconstruction

Sabyasachi MUKHOPADHYAY (Indian Statistical Institute, India)

Fossil-based palaeoclimate reconstruction is an important area of ecological science that has gained momentum in the backdrop of the global climate change debate. Vasko, Toivonen and Korhola (2000) were arguably the first to propose a Bayesian model for palaeoclimate reconstruction, on which the full-fledged attempt to reconstruct past Irish climate from fossil pollen was based by Haslett et al. (2006). The shortcomings of Vasko, Toivonen and Korhola (2000) and Haslett et al. (2006) motivated Bhattacharya (2006) to modify the species response function using a mixture of unknown number of Gaussian functions. But even though the latter seemed to rectify several of the drawbacks of the model of Haslett et al. (2006) and showed very significant improvement of the results yielded by that of Vasko, Toivonen and Korhola (2000), further improvement was desirable. To clarify, the results of cross-validation with respect to Vasko, Toivonen and Korhola (2000) showed only 43data set consisting of chironomid counts apart from observed temperatures. For the same data, Bhattacharya (2006) obtained 83coverage is desired for serious consideration as a candidate for palaeoclimate reconstruction. In this paper, we propose a simple, albeit very important modification of the model of Bhattacharya (2006). Even though the the zero-inflated multinomial model in place of the ordinary mutinomial model is the only issue discriminating our current model and that of Vasko, Toivonen and Korhola (2000), still this new model now yields the much-desired 100% coverage of the observed temperatures of the chironomid data, evidently providing the green signal for prehistoric climate reconstruction.

Coauthors: Sourabh Bhattacharya

❊

Joint stochastic blockmodeling of attributed random graphs

Jared MURRAY (Duke University, USA)

Data from a range of application areas are concerned with how pairs of units interact - examples include relationships between individuals or nation-states, links between webpages, or interactions between genes or proteins. Such relational data can be summarized in a graph, where distinct observational units are nodes and edges describe the (possibly valued and/or directed) relationships between the units. Often there are additional covariates or responses (“attributes”) observed for each node, which could be of mixed continuous, count, ordinal and/or unordered categorical types. We may be interested in using using the relational structure to predict or infer node-level attributes, leveraging the attributes to uncover latent structure in the graph and predict links, or both. Additionally, the attributes and the graph may be subject to measurement error or missingness. For these reasons we develop a novel Bayesian nonparametric joint model for graphs and their node-level attributes. Our model is related to existing stochastic blockmodels and other models for networks and relational data, but unlike these methods we model the attributes and allow them to impact the network model flexibly. Fully Bayesian inference via MCMC is straightforward and efficient. Applications to some network datasets are considered.

Coauthors: Jerome Reiter

❊

Mixture of component models for hidden stratification in repeated measures data

Sungmin MYOUNG (Jungwon University, South Korea)

The Mixture of experts (ME) is modular neural network architecture for supervised learning among a number of existed methods. This model can be considered as a mixture model that is consisted of the mixed distributions and weights in input variable x. In this research, we propose and develop a novel class of the mixture of component model for hidden stratification(the cluster-specific effect) in repeated measure data. The proposed model can be considered to joint modeling of the mixture of expert(ME) model via the linear mixed linear model. To resolve this, firstly, we considered the construction of an expert, which has linear mixed-effect model, in ME. Afterward, the finding estimates were obtained for gating network and expert via EM-algorithm. The proposed model is more flexible than classical linear mixed-effect model in identifying several hidden stratification with different patterns. The simulated data and real data examples will be used to illustrate the feasibility of the proposed method. We expect this model to apply to a wide range of a variety of clinical studies.

Coauthors: Jinnam Jo

❊

Bayesian posterior probability computation of identified network communities by bootstrap resampling

Haruhisa NAGATA (Tokyo Institute of Technology, Japan)

Identifying communities in a given network is an important subject for understanding diverse systems such as social networks and protein-protein interaction networks. Communities (or clusters) are groups of related nodes of a specified graph. There have been several methods proposed in literature for identifying communities. Among them, Ahn et al. (2010) proposed recently one particularly promising method by applying hierarchical clustering to edges in a graph, instead of nodes, so that identified communities allow overlap. These methods, however, may produce too many garbage communities, making the interpretation very difficult. In this presentation, we develop methods to get bayesian posterior probability of identified communities. The communities may be ranked by the bootstrap probability (bp) and only those with high bp are left for interpretation. Our method is to generate replicate graphs from a given graph. A key idea is to bootstrap edges of the graph by allowing multiple edges between a pair of nodes. Then the bp of each community is computed by counting how many times the same community appears in the replicate graphs (allowing some discrepancy), which means its bayesian posterior. In addition to this graph bootstrapping, we also apply the multiscale bootstrap method of Shimodaira (2008) for adjusting the bias of bp. We will show numerical examples of the link structure of Wikipedia.

Coauthors: Hidetoshi Shimodaira

❉

Analysis of acceptance rate for Metropolis algorithm and relationship between Bayesian learning

Kenji NAGATA (University of Tokyo, Japan)

Markov chain Monte Carlo (MCMC) method is the algorithm to sample from arbitrary probability distribution, and is widely used in various areas such as Bayesian estimation. While the algorithm has the advantage that it can sample the probability density not normalized, it has the disadvantages of huge computational cost and the difficulty of convergence judgement. Hence, the optimal settings of tuning parameters such as Monte Carlo step and step size for the Metropolis algorithm are very important for the efficient MCMC method. In this study, we analytically calculate the average of acceptance rate, which is the criterion for the step size for the Metropolis algorithm. In particular, our analysis focuses on the situation for Bayesian estimation of singular learning machines such as the neural networks and the mixture of Gaussian. Moreover, we verify the efficiency of our analysis by comparing the analytical result with that of the computer simulation.

Coauthors: Sumio Watanabe and Masato Okada

❉

A Bayesian econometric analysis of auction under variable reserve price and unknown number of potential bidders

Yuji NAKAYAMA (Osaka Prefecture University, Japan)

In this paper, we conduct a Bayesian econometric analysis of auction data. We assume that (1) reserve price varies across auctions and (2) the number of potential bidders is unknown, which are usual in Internet auctions. Under Independent Private Value Paradigm (IPVP), we show that a modified version of Bayesian Tobit model with data augmentation by Chib (1992) can be applied to online auction data with the above two assumptions. Using simulated data, we examine the validity of our method. Various extensions of our method are also discussed.

Coauthors: Tomonori Ishigaki and Nagateru Araki

❉

Bayesian construction and estimation of GIG stationary models

Consuelo R. NAVA (University of Torino, Italy)

We construct a Markovian stationary model via the Esscher transform of the Generalized Inverse Gaussian (GIG) moment distribution. Our approach allows full control of the corresponding transition probabilities. This turns out to be highly flexible to model both discrete and continuous time random phenomena. We study some of its properties and propose a Bayesian estimation procedure for financial data. Some connections to SDEs and dependent nonparametric models are also highlighted.

Coauthors: Igor Prünster and Ramsés H. Mena

❉

Functional representation of Markov exchangeable sequences

Maxim NAZAROV (Bocconi University, Italy)

Our work is motivated by the problem of modeling dynamics in network models. Network models are increasingly popular in representing relational data, arising in social sciences and other areas of life (interaction of individuals, groups of people, organizations; international trade; biological applications). An interesting direction of research lies in modeling dynamics of a network as it evolves, e.g. over time or space. It is customary to assume exchangeability in the data. Functional representations of two- and higher dimensional exchangeable arrays were obtained by Aldous (1981) and Hoover (1982). These representations are widely used in relational data modeling. At the same time, when considering the evolution of a network, the assumption of exchangeability is undesirable. So, it seems interesting to combine notions of row-column exchangeability for arrays with less restrictive Markov exchangeability along the third/time dimension. We give a first result in this direction by providing a functional representation for recurrent Markov exchangeable sequences. The result exploits the characterization of Markov exchangeable sequences in terms of (de Finetti) partial exchangeability of the matrix of successor states and a functional representation for a partially exchangeable family of random variables.

References

D. J. Aldous, "Representations for partially exchangeable arrays of random variables", Journal of Multivariate Analysis, Volume 11, Issue 4, December 1981, pages 581–598

D. Hoover, "Row-columns exchangeability and a generalized model for exchangeability", in G.Koch and F.Spizzichino (Editors), Exchangeability in probability and statistics (Rome, 1981), North-Holland, Amsterdam, 1982, pages 281–291

Coauthors: Sonia Petrone

⁂

Bayesian unconditional quantile regression: An analysis of recent expansions in wage structure and earnings inequality in the U.S. 1992-2009

Abdoul Aziz Junior NDOYE (GREQAM, AMSE, France)

We develop a reliable Bayesian inference for the RIF-regression model of Firpo, Fortin and Lemieux (Econometrica, 2009) in which we first estimate the log wage distribution by a mixture of normal densities. This approach is pursued so as to provide better estimates in the upper tail of the wage distribution as well as valid confidence intervals for the Oaxaca-Blinder decomposition. We apply our method to a Mincer equation for analysing the recent changes in the U.S. wage structure and in earnings inequality. Our analysis uses data from the CPS Outgoing Rotation Group (ORG) from 1992 to 2009. We find first that the largest part (around 77% on average) of the recent changes in the U.S. wage inequality is explained by the wage structure effect and second that the earnings inequality is rising more at the top end of the wage distribution, even in the most recent years. The decline in the unionisation rate has a small impact on total wage inequality while differences in returns to education and gender discrimination are the dominant factors accounting for these recent changes.

Coauthors: Michel Lubrano

⁂

Regenerative simulation for Metropolised Gibbs samplers

Ronald NEATH (Hunter College, City University of New York, USA)

Regenerative simulation is a technique for identifying times at which a Markov chain "probabilistically restarts itself", so that the paths taken between regeneration times are independent and identically distributed. Regeneration was introduced by Mykland et al. (1995) as a method for Markov chain Monte Carlo (MCMC) output analysis, and by Gilks et al. (1998) as a method for constructing adaptive MCMC algorithms. In the years since those papers, regeneration has been employed to great effect for approximating posterior moments from Gibbs samplers and block-updated Metropolis-Hastings algorithms, but its use in variable-at-a-time Metropolis-Hastings, or "Metropolised Gibbs", remains unexplored. In this poster we develop a framework for conducting regenerative simulation in a general component-wise Metropolis-Hastings sampler. We illustrate our results in a simple but practically relevant problem in Bayesian inference.

Coauthors: James Flegal and Galin Jones

⁂

A Bayesian semi-parametric model for covariate dependent survival data

Bernardo NIPOTI (University of Texas MD Anderson Cancer Center, USA)

We propose a Bayesian semi-parametric model for the analysis of covariate dependent survival data that accommodates situations where the proportional hazard assumption may fail. More specifically, we resort to the construction of a vector of gamma completely random measures in order to define (possibly dependent) Cox proportional hazard models where the baseline hazard functions are distributed as a vector of extended gamma processes. Further dependence is introduced in the prior distributions of the regression coefficients of each Cox model. A characterization of the posterior distribution of the baseline hazard functions leads us to the implementation of a Gibbs sampling type algorithm to estimate correlated survival functions. Borrowing information across different models turns out to be useful when carrying out inference on both the baseline hazards and the regression coefficients. We apply the proposed model to a dataset of possibly censored survival times of patients under different treatment regimens.

Coauthors: Michele Guindani

*

Bayesian estimation of income distribution using grouped data

Haruhisa NISHINO (Chiba University, Japan)

Grouped data are often used in estimation for income distribution because it is sometimes difficult to access micro individual data. We propose three Bayesian estimations for group data using lognormal distribution. The first is a Gibbs sampling applied to a likelihood for grouped data, the second is a Gibbs sampling applied to a likelihood for individual data with data augmentation, and the third is an acceptance-rejection Metropolis-Hastings algorithm method. We present Monte Carlo simulation results of the three Bayesian estimation and empirical results for Japanese grouped income data.

Coauthors: Kazuhiko Kakamu

*

Bayesian Lasso for screening experiments

Hidehisa NOGUCHI (Tokyo University of Science, Japan)

In a screening experiment, eliminating the unimportant predictors from data sets and identifying a good subset of predictors are important. Those screening factors are supposed to have active effect. When many potential factors are examined, variable selection plays an important part in screening experiments. There are many variable selection methods however most of them have a problem of not accounting for any relationships between interaction and its main effect. Because of this, the subsets selected by variable selection can be hard to interpret. Some statisticians adopt the convention that a model containing an interaction should also contain the corresponding main effects. In practice, this relationship is defined as the principle; Effect Heredity Principle. In our study, we focus on the Lasso method which is useful variable selection method in the situation where there are more parameters than sample size. The Lasso estimate can be interpreted as a Bayesian posterior mode estimate when the regression parameters have independent Laplace priors. However this Bayesian Lasso method also doesn't consider the relationship between the predictors. So, we extend the Bayesian Lasso for identifying important factors which are obeying Strong and Weak Heredity Principle. And we consider the relationship between the main effect posterior and its corresponding interaction posterior. Methods are illustrated on simulation and real data.

Coauthors: Yoshikazu Ojima and Seiichi Yasui

*

The optimal discovery procedure in multiple significance testing: an empirical Bayes approach

Hisashi NOMA (Kyoto University School of Public Health, Japan)

Multiple testing has been widely adopted for genome-wide studies such as microarray experiments. To improve the power of multiple testing, Storey (2007, J. Royal Statist Soc B 69: 347-368) recently developed the optimal discovery procedure (ODP) which maximizes the number of expected true positives for each fixed number of expected false positives. However, in applying the ODP, we must estimate the true status of each significance test (null or alternative) and the true probability distribution corresponding to each test. In this study, we derive the ODP under hierarchical, random effects models and develop an empirical Bayes estimation method for the derived ODP. Our methods can effectively circumvent the estimation problems in applying the ODP presented by Storey (2007). Simulations and applications to clinical studies of leukemia and breast cancer demonstrated that our empirical Bayes method achieved theoretical optimality and performed well in comparison with existing multiple testing procedures.

Coauthors: Shigeyuki Matsui

*

Parameter estimation in LNSV models: griddy Gibbs versus Metropolis-Hastings

Didit Budi Nugroho (Kwansei Gakuin University, Japan)

We compare performances between Gibbs sampler-griddy Gibbs (GS-GG) and Gibbs sampler-Metropolis-Hastings (GS-MH) sampling algorithm to estimate the log-normal stochastic volatility (LNSV) models: LNSVn (standard), LNSVt (with fat-tails), LNSVL (with correlated errors) and LNSVLt (with fat-tails and correlated errors). The models and algorithms are applied to the daily returns on TOPIX and three stocks of the TOPIX Core 30: Hitachi, Nissan Motor and Panasonic, from January 2004 to December 2011. The comparison results for the model parameters and latent volatility are as follows. Indicated by persistence and variance of volatility, the volatility by GS-MH algorithm is more persistent and less variable than those by GS-GG algorithm. In the case of LNSV involving fat-tails, GS-GG algorithm exhibits the higher degrees of freedom than those by GS-MH algorithm. For the LNSV models involving correlated errors, the posterior means of the leverage effect parameter in GS-GG algorithm suggest a weaker negative correlation between returns and volatility process. Finally, after comparing the estimation performance of latent volatility, where the daily Realized Volatility is used as a proxy, it was found that the GG sampler is superior according to all six loss functions for TOPIX in the LNSVLt model, while the GS-MH algorithm is superior according to five loss functions for Hitachi in the LNSVL model and for Nissan in the LNSVLt model and three loss functions for Panasonic in the LNSVt model. Furthermore, GMLE (Gaussian quasi-maximum likelihood function) loss function is only minimized on all stocks in the same model and sampler: LNSVLt model by GS-GG algorithm.

Coauthors: Takayuki Morimoto

⁂

Sampling EGARCH parameters

Takashi Oga (Chiba University, Japan)

In this paper, we present a new and powerful sampling tool for Nelson's EGARCH (Exponential Generalized Auto-Regressive Conditional Heteroskedasticity) model. Our new sampling methods for non-conjugate parameters are based on independent MH algorithm, utilizing discrete approximations. We can find good Gaussian candidate distributions using this method. In discretizing unknown continuous distributions, which have $(-\infty,\infty)$ support, the problems are how the upper and lower point are determined and how the number of gridding points are selected. We show the best setting is that the points should be $[-7.5\sigma, 7.5\sigma]$ and grid interval should be 1σ, that is; we only need 16 discretizing point. In MCMC loop, we need to evaluate full conditional density 16 times and it's a little time consuming, but the efficiency is higher than random-walk MH algorithms, namely, the adaptive MCMC. This paper presents the methods and theoretical background, and empirical studies using EGARCH models.

⁂

Bayesian inference of associations and SNP-environment interactions for GWAS-identified RA risk alleles with risk of infection via perfect sampling scheme

Cheongeun Oh (New York University, USA)

Bayesian variable selection sampling via Markov chain Monte-Carlo (MCMC) has been extensively applied to various biomedical applications, noticeably in handling high-dimensional data such as genome wide association studies (GWAS). However, these methods have been limited in practice, partly due to the uncertainty of convergence and sensitivity to the initial starting points. Besides, the applications and the interpretation to GWAS have still presented great statistical and computational challenges especially in testing interactive effects. Here we address the drawbacks of the current Bayesian variable selection approaches and propose remedies by utilizing the perfect sampling scheme, so-called "Coupling From The Past" (CFTP). Our methods ensure unbiased samples from the posterior distributions, but also are able to handle a large number of variables by controlling the size of the model via a prior distribution over the dimension of the model. We also suggest various prior choices in order to handle highly correlated SNPs with strong linkage disequilibrium (LD) and model sparsity in searching interactive effects. The effectiveness of our method is illustrated to identifying multiple single nucleotide polymorphisms (SNPs) associated with infections, epistases, and their interactive effects with environmental factors among rheumatoid arthritis patients. Our method can provide useful inferences in these large scale problems, and produces better power and predictive performance compared with standard single-SNP analyses or current Bayesian variable selection approaches.

Coauthors: Jeffrey Greenberg

⁂

Bayesian multiple comparison of models for binary data with inequality constraints

Man-Suk OH (Ewha Womans University, Korea)

In this paper we consider generalized linear models for binary data subject to inequality constraints on the regression coefficients, and propose a simple and efficient Bayesian method for parameter estimation and model selection by using Markov chain Monte Carlo (MCMC). In implementing MCMC, we introduce appropriate latent variables and use a simple approximation of a link function, to resolve computational difficulties and obtain convenient forms for full conditional posterior densities of elements of parameters. Bayes factors are computed via the Savage-Dickey density ratios and the method of Oh(1999), for which posterior samples from the full model with no degenerate parameter and the full conditional posterior densities of elements are needed. Since it uses one set of posterior samples from the full model for any model in consideration, it performs simultaneous comparison of all possible models and is very efficient compared with other model selection methods which require one to fit all candidate models.

A simulation study shows that significant improvements can be made by taking the constraints into account. Real data on purchase intention of a product subject to order constraints is analyzed by using the proposed method. The analysis results show that there exist some price changes which significantly affect the consumer behavior. The results also show the importance of simultaneous comparison of models rather than separate pairwise comparisons of models since the latter may yield misleading results from ignoring possible correlations between parameters.

⁂

Bayesian model averaging in factor analysis to estimate factor reliability

Kensuke OKADA (Senshu University, Japan)

The estimation of reliability coefficient is one of the central issues in the application of factor analysis. One of the most commonly used reliability index is Cronbach's alpha, but there are also many alternative methods proposed over years. However, most methods currently available do not consider the ambiguity included in the model structure. Therefore, we propose a Bayesian model averaging (BMA) approach to estimate the reliability under model uncertainty. In the proposed method, the number of factors in the model is taken to be unknown. The posterior model probabilities of models having different number of factors are calculated by using the default prior distribution and path sampling methodology proposed by Ghosh and Dunson (2009). For a single model, the parameter expanded (PX) Gibbs sampler, which introduces an overparameterized working model, is used to enhance the mixing in Gibbs sampling. The performance of the proposed method is evaluated by a Monte Carlo simulation study. Compared to reliability coefficients and SEM approach, the proposed method achieved robust performance under most of the experimental conditions.

⁂

Imputation of confidential datasets with spatial locations using point process models

Thais PAIVA (Duke University, USA)

We suggest a method to generate synthetic data sets with imputed spatial locations, respecting individuals' confidentiality without losing statistical utility. Our primary interest is to impute the spatial coordinates conditional on the response and explanatory variables. We generate the imputed data sets using spatial point models. The underlying spatial intensities are modeled allowing flexible relationships among the variables and the spatial locations. Using a Bayesian framework, we obtain posterior samples of the intensities, and use them to generate imputed data sets for public release. We verify the quality of the synthetic data, along with the level of confidentiality.

Coauthors: Jerome Reiter

⁂

Gaussian process-based estimation of population size trajectories from gene sequence data. A phylodynamic approach

Julia PALACIOS (University of Washington, USA)

Bayesian estimation of population size trajectories from gene sequence data usually involves two levels of hierarchical modeling. Given gene sequence data and a mutational model, one can estimate a genealogical tree. Given a genealogical tree and the coalescent model, one can estimate the effective population size. In order to add as much flexibility as possible in terms of the shape of the population trajectory, a Gaussian process-based (GP) Bayesian nonparametric method has been recently developed for a fixed genealogy. We extend the GP method to sample genealogies and estimate population trajectories directly from gene sequence data. Additionally, we incorporate the flexibility to correlate effective population size trajectories with other temporal processes. In nearly all molecular data-based demographic analyses, the analysis does not end at estimating population dynamics. Often, one is interested in demonstrating that some other temporal process correlates with the population size fluctuations. Our GP-based framework enable practitioners to put their hypotheses about relationships between other temporal processes and the population size trajectory into a rigorous quantitative framework.

Coauthors: Vladimir Minin

⁂

Bayesian estimation and model selection for high dimensional copula models

Anastasios PANAGIOTELIS (Monash University, Australia)

Recent years, have seen notable growth in the development of high dimensional copulas.These can be used to flexibly model features that are commonly observed in multivariate data in finance and econometrics, such as non-linear dependence and tail dependence. Alongside these developments in copula modelling, a rich literature on Bayesian estimation of these models has also emerged.In a recent paper, Smith and Khaled (2011) develop a general efficient Markov chain Monte Carlo (MCMC) scheme that can be used to estimate the parameters of almost any copula model, including in cases where margins are continuous, discrete or some combination of the two.

This poster presents work that incorporates principles of Bayesian model selection into the modern literature on Bayesian estimation of copulas models.Margins are assumed to follow generalised additive models which include a potentially large number of covariates.Binary indicator variables are introduced to explicitly parameterise whether each covariate should be included.A number of alternative sampling schemes are developed for sampling these binary indicators and are compared in a simulated setting.The overall modelling framework can be interpreted as a more general, robust and flexible version of the seemingly unrelated regression model that is prevalent in econometric modelling.

⁂

Bayesian nonparametric spatial modeling with applications in environmental epidemiology

Georgios PAPAGEORGIOU (Imperial College London, UK)

We present Bayesian nonparametric mixture models suitable for the analysis of spatially indexed data. Recently, there have been several proposals for nonparametric modeling of the mechanism that produces spatial dependence. Here we focus our attention on models that induced spatial dependence by allowing the mixture weights to vary between locations. The weights are obtained as probit transformations (Rodriguez & Dunson 2011) of latent variables that arise from a Gaussian Markov random field (Rue & Held 2005). As a result, the correlation function of the Gaussian random field induces correlations among the distributions on the geographical space of interest.

Our specific focus is on spatially distributed, highly correlated, covariates and their link with disease data. In such circumstances classical methods such as regression via generalized linear models become problematic due to issues related to multicollinearity. We circumvent this problem by avoiding regression and instead adopting a multivariate mixture model for both covariates and health outcomes at each location (Molitor et al. 2010). Further tools are developed for quantifying the effects that covariates have on health outcomes.

The flexibility of the model is illustrated by simulation studies and an application to a real data set. Specifically, we compare clusters of geographical areas created by environmental covariates and health outcomes with and without the use of the latent Markov random field in the model. In addition, we obtain spatial predictions of disease rates for areas for which only covariates or no data at all have been observed.

Coauthors: Sylvia Richardson and John Molitor

⁂

Bayesian semiparametric regression

Justinas PELENIS (Institute for Advanced Studies, Vienna, Austria)

The standard practice in the Bayesian literature for linear regression modeling is to use flexible families of distributions for the errors and to assume that the errors are independent from regressors. However, a model with flexible predictor dependent error distributions should be preferred for the following reasons. First, assuming that the error distribution is independent of predictors might lead to inconsistent estimation when errors and regressors are dependent. Second, the prediction intervals obtained from a model with predictor dependent error distributions are likely to be superior to the ones obtained assuming a constant error distribution. Third, modeling conditional error density might allow one to obtain a more efficient estimator of the regression coefficients under heteroscedasticity. To address these issues, I develop a Bayesian semiparametric regression model with flexible predictor dependent error densities and with mean restricted by a conditional moment condition. I consider a flexible prior for conditional residual densities using a finite and an infinite mixture of mixtures of two normal distributions constrained to have mean zero. I provide sufficient conditions for posterior consistency of the regression coefficients and weak posterior consistency of the conditional error densities. This approach allows for estimation of both regression coefficients and residual density that varies nonparametrically with the covariates. In experiments, the proposed method compares favorably with classical and alternative Bayesian estimation methods for regression coefficients.

⁂

Estimation of the multivariate covariance of noisy and asynchronous returns

Stefano PELUSO (Università della Svizzera Italiana, Switzerland)

Motivated by the need of an unbiased and positive-definite estimator of multivariate realized covariance matrices, we model noisy and asynchronous ultra-high-frequency asset prices in a state-space framework with missing data. A frequentist and a Bayesian multivariate positive definite estimators are proposed. In the frequentist setting, we estimate the covariance matrix of the latent states through a Kalman smoother and Expectation Maximization (KEM) algorithm. In particular, in the expectation step, by means of the Kalman filter with missing data, we reconstruct the smoothed and synchronized series of the latent price processes. In the maximization step, we search for covariance matrices that maximize the expected likelihood obtained with the reconstructed price series. Iterating between the two EM steps we obtain a KEM-improved covariance matrix estimate which is robust to both asynchronicity and microstructure noise, and positive definite by construction. In the Bayesian approach, Gibbs sampler and Metropolis-Hastings algorithms with different degrees of augmentation are implemented to sample the covariance matrix, the observational error variance matrix, the latent process and the missing observations of the noisy process. In a Monte Carlo simulation and empirical settings, we compare our KEM and Bayesian estimators with alternative covariance matrix estimates and under different liquidity and microstructure noise conditions.

Coauthors: Antonietta Mira and Fulvio Corsi

⁂

Optimal classification policy for highly reliable products

Chien-Yu PENG (Academia Sinica, Taiwan)

Nowadays in the competitive marketplace, manufacturers need to classify products in a short time according to market demand. Hence, it is a challenge for a manufacturer to implement a classification test that can distinguish the different levels of products quickly and efficiently. For highly reliable products, if quality characteristics do exist whose degradation over time can be related to the lifetime of the product, the degradation model can then be constructed based on the degradation data. In this study, we propose a non-linear degradation model that simultaneously considers unit-to-unit variation with time-dependent error structure and measurement error. Then, by adopting the concept of linear discriminant analysis, we propose a three-step classification policy to determine optimal vector of coefficients, optimal cut-off points and optimal testing time subject to cost. In addition, we also use an analytic approach to compare the efficiency of our proposed procedure with the methods that are previously reported in the literature. Finally, we use several data sets to illustrate the proposed classification procedure.

Coauthors: Shih-Chi Hsu

⁂

Context specific graphical models

Johan PENSAR (Åbo Akademi University, Finland)

We introduce a novel class of context specific graphical models (CS-models) for finite sets of discrete variables, which allow for context specific independencies to be captured within a graph structure. While conventional graphical models are restricted to conveying global conditional independence statements, CS-models offer the possibility of representing conditional independencies spanning only a subset of the outcome space. Decomposable CS models are defined, such that the marginal likelihood can be calculated analytically for them under a specific conjugate prior enjoying appropriate Markov properties. For non-decomposable models where the joint distribution of variables cannot be fully factorized, we utilize cyclic projections to derive a BIC based approximation to the marginal likelihood. A non-reversible MCMC algorithm is developed for model learning and we will illustrate the advantages of CS-models by both synthetic and real datasets.

Coauthors: Henrik Nyman, Jukka Corander and Timo Koski

*

Bayesian bidirectional clustering method for detection of gene loss and acquisition processes

Alberto PESSIA (University of Helsinki, Finland)

The rapidly increasing amount of whole genome sequences of bacteria that are present in public databases has opened up new possibilities to investigate processes of gene acquisition and loss over genus, family and order levels. When no prior knowledge of the structure of the data is assumed, the problem can be formalized as a bidirectional unsupervised cluster analysis, both in the units space and the features space, the objective being to separate relevant features from uninteresting features. We introduce a parametric Bayesian clustering method, based on a product partition model (PPM), to simultaneously identify a structure of mutually exclusive groups for the units and the features characterizing such clusters. In addition, a novel sequential learning algorithm has been developed for identifying global posterior modes in certain subspaces of the total model space. Both synthetic and real-world data are used to evaluate the performance of our proposed method.

Coauthors: Jukka Corander

*

Modeling heterogeneity in genotype-phenotype associations

Matti PIRINEN (University of Oxford, UK)

Contemporary genome-wide association studies (GWAS) consider genotype data of thousands of individuals at millions of polymorphic loci. A central question is to quantify the statistical evidence for associations between the genetic variants and the phenotypes of interest. In a past few years GWAS have identified many loci that show convincing statistical association with complex human diseases and the expectation is that these signals will point to the actual biological mechanisms behind the phenotypes.

We introduce a joint model for genotype data from multiple collections of individuals using hierarchical Bayesian modeling. We use the model to quantify the level of genetic heterogeneity between different subtypes of a disease at a single locus, and to identify genetic effects that are heterogeneous across different populations.

Computationally, we consider approximations to the model that allow its application to large genome-wide data sets, and compare the accuracy of these approximations to more complex sampling algorithms.

*

Multivariate versus multinomial probit: when are binary decisions made separately also jointly optimal?

Dale POIRIER (University of California, Irvine, USA)

We provide an analysis of the question in the title in terms of a bivariate probit framework representing two (possibly correlated) separate decisions, and a multinomial probit framework representing the four possible outcomes viewed as one joint decision. We offer a Bayesian treatment that builds on Weeks and Orme (1998) and Di Tommaso and Weeks (2000) who showed that the bivariate probit corresponds to a singular four-dimensional multinomial probit under testable restrictions. We also discuss extensions to trivariate and quadrivariate probit.

Coauthors: Deven Kapadia

*

MCMC estimation for the extended Hodrick-Prescott (eHP) filter

Wolfgang POLASEK (Institute for Advanced Studies, Vienna, Austria)

The Hodrick-Prescott (HP) method is a popular smoothing method for economic time series to get a smooth or long-term component of stationary series like growth rates. We show that the HP smoother can be viewed as a Bayesian linear model with a strong prior using differencing matrices for the smoothness component. The HP smoothing approach requires a linear regression model with a Bayesian conjugate multi-normal-gamma distribution. The Bayesian approach also allows to make predictions of the HP smoother on both ends of the time series. Furthermore, we show how Bayes tests can determine the order of smoothness in the HP smoothing model. The extended HP smoothing approach is demonstrated for the non-stationary (textbook) airline passenger time series. Thus, the Bayesian extension of the HP model defines a new class of model-based smoothers for (non-stationary) time series and spatial models.

⁂

Robust microarray analysis with mixtures of Gaussians and heavy-tailed student's t distributions

Alexandra POSEKANY (Vienna University of Economics and Business, Austria)

Frequently, data violate normality assumptions, required for analysis of variance (ANOVA) and other tests and models. Thus, we robustified Bayesian inference by means of mixture models, without fitting solely complex non-Gaussian distributions. We cover scenarios where the data is partially over-dispersed, while the rest follows a normal distribution. As the Gaussian model provides several advantages for normally distributed data, we model this part with the well-fitting normal distribution. Non-Gaussian distributions model the over-dispersed, outlying values, possibly originating from measurement errors or an underlying data generating process. Combining both normal and t distributions in a mixture model allows modelling both types of noise behaviour. For inference and dealing with label switching we defined a measure of non-Gaussianity, based on peakedness. Practically, we conducted a case study on microarray data, well-known for their complicated, usually over-dispersed noise behaviour. Matching previous findings, we showed that the noise behaves very heavy- tailed, while only a minor part is normally distributed, and showed that noisy behaviour originates at least partially from biological processes. In summary, our method constitutes a high degree of robustness and handles data with complex, yet symmetric structure, as appropriate for linear regression settings with noisy data. While providing a general and flexible way of modelling residuals, our model avoids fitting more parameters than necessary by not always using non-Gaussian distribution. Our model qualifies for data with too few sample to conduct non-parametric analyses, while remaining close to the ANOVA model.

Coauthors: Peter Sykacek and Klaus Felsenstein

⁂

Calibrated smoothing: sharing structure between computer simulations and historical data

Benedict POWELL (Durham University, UK)

We assess the connection between the output of a simulator and the observable quantities in the system it simulates, with our attention turned to time series of data. There is no hope nor ambition to match the two exactly. In fact such a match would be met with high suspicion because our scientific knowledge tells us that sustained synchronization would be impossible, due to structural differences between simulator and system, or even just impossible to find, due to chaotic behaviour imprecisely observed. What then ought we see as the common element between the two? We propose that a smooth trend, representing a hypothesised low frequency signal like the 'climate' as opposed to 'weather', is the object that we can hope to calibrate and simulate successfully.

In pursuing this strategy the usual difficulties involved in model emulation are encountered but often with interesting differences. Smoothness measures may serve as suitable metrics for data, and properties of those metrics may be used to reduce effectively the high dimensional space in which they reside. Even the likelihood functions, that form a part of calibration exercises, inherit smoothness properties that can be used to aid their approximation.

We will show, by example, how a time series can be smoothed using the information from a calibrated simulator, and discuss the conceptual basis for doing so.

⁂

Thursday, June 28th

Rational threshold schemes for approximate Bayesian computation via unscented transforms

Sarah FILIPPI (Imperial College London, UK)

The choice of acceptance thresholds remains an important challenge for sequential Approximate Bayesian Computation (ABC) algorithms. Such algorithms proceed by constructing a succession of probability distributions over the parameter space conditional upon the simulated data lying in an epsilon ball around the observed data, for decreasing values of the threshold epsilon. While in theory, the distributions (starting from a suitably defined prior) will converge towards the unknown posterior as epsilon tends to zero, the exact sequence of epsilons can impact upon the computational efficiency and success of a particular application. Currently, epsilons are typically chosen as a pre-determined quantile of the distances between simulated and observed data from the previous population. Here we present an automated and adaptive method for epsilon choice that is based upon predictions of the future distributions using the unscented transform (UT). Generally known for its use in extending the Kalman filter to non-linear problems, the UT allows the statistics of a Gaussian random variable that has undergone a non-linear transform to be estimated. By fitting Gaussian mixture models to the prior distributions, we can use the UT to predict the ABC acceptance rate for any value of epsilon, and subsequently choose the acceptance thresholds that optimally balance the need to minimise epsilon with computational efficiency. Further, global (approximate) knowledge of the epsilon / acceptance rate curve can enable us to avoid local minima - a problem that has plagued previous threshold schemes.

Coauthors: Daniel Silk and Michael Stumpf

❈

MDI: a method for integrative modeling of multiple functional genomics datasets

Paul KIRK (University of Warwick, UK)

High-throughput technologies used in the modern biomedical and life sciences generate a broad array of different data types, providing distinct - but often complementary - information. As a result of dependencies between the biological processes probed by these technologies (e.g. gene expression and protein-DNA binding), the clustering structures within the various datasets may be similar. We consider modeling each dataset using Dirichlet-multinomial allocation (DMA) mixture models, but - rather than modeling each dataset independently - we modify the approach so that the mixture component to which gene i is allocated in Dataset k also depends upon the component allocations for gene i in the other datasets. We infer parameters that control the strength of these dependencies, and which thereby describe the level of similarity in clustering structure between the datasets. We refer to our approach as MDI, simply as shorthand for “multiple dataset integration”. We consider a variety of different types of functional genomic datasets (discrete/continuous, static/time series, ...), and use our approach to identify genes that tend to cluster together within several of the datasets. MDI allows us to identify more biologically homogeneous gene groups, and provides insights into the relationships between the various datasets.

Coauthors: Jim Griffin, Richard Savage, Zoubin Ghahramani and David Wild

❈

Spatio-temporal modeling of avalanche frequencies in the French Alps

Aurore LAVIGNE (UMR 518 AgroParisTech/INRA, France)

Over a large area as the whole French Alps, we relax the common assumption of separability between space and time of avalanche counts. Based on climate studies which exhibit zones characterized by distinct climatic behaviors, we assume that our 63-years series of avalanche counts can be divided into several clusters presenting a spatial pattern. On the latent layer of a hierarchical negative binomial-lognormal structure, we assume the existence of clusters characterized by independent temporal evolutions, modeled using smoothing splines. The clusters arise as the result of a multinomial probit regression, whith spatial coordinates as regressors. Bayesian inference of our case study is performed thanks to a Gibbs sampler. The number of clusters is assessed using cross validation, and is evaluated to 2 or 3. Details results for the 3-cluster model show that one cluster is located in the east of the region, whereas the 2 other clusters are located together in the west. We highlight three contrasted temporal trends for the three clusters. They result from regional climate change effect interacting with a strong altitudinal control. Some data with poor quality may somehow blur the results. For further interpretation, additional climate or topographic data could be brought into the analysis.

Coauthors: Eric Parent, Nicolas Eckert and Liliane Bel

❈

Bayesian hierarchical meta-analysis model for medical device evaluation

Leslie PIBOULEAU (INSERM UMR 717 and Université Paris 7, France)

Rationale: Implantable medical devices (IMDs) display specific characteristics that lead to some important issues regarding their clinical evaluation. Indeed, contrary to innovative drugs, IMDs are usually updated continuously. Although clinical data on previous versions are available, how to integrate these data when evaluating an IMD is an open issue. To combine the evidence, Bayesian meta-analysis models provide an opportunity to learn about variations in treatment effectiveness.
Objective: to study Bayesian three-level hierarchical meta-analysis models, allowing for between-version heterogeneity, comparing them to standard random effects for IMD evaluation when the effects of the IMDs varied.
Methods: A simulation study was conducted with models compared on the Deviance Information Criterion (DIC), bias and standard deviation of the mean success rates under repeated sampling. The practical use of these methods was illustrated in the evaluation of intracranial stents for the treatment of wide-necked aneurysms.
Results: When there was no heterogeneity, the two-level random-effects model was the best-fitting model. However when there was a difference in efficacy between versions, the three-level hierarchical model allowed to reach a better DIC and minimised the bias and standard deviation in its estimates.
Conclusion: The comparison of the DIC between the models appears to be a useful way of detecting some version effects and discriminating among all of the versions those that should be considered equivalent in efficacy. When there is a detected version effect, hierarchical model may provide precise estimates of the success rates of the currently available versions of IMD.

⁂

MCMC for continuous-time discrete-state systems: beyond uniformization

Vinayak RAO (University College London, UK)

In this work, we consider a novel class of MCMC algorithms for posterior simulation in continuous-time discrete-state models. Examples of such models include Markov jump processes, renewal processes and semi-Markov jump processes. Our work builds on recent work on Markov jump processes [1] where the idea of uniformization [2] was used to construct an auxiliary variable blocked Gibbs sampler. Uniformization involves determining a single rate that bounds all event rates in the system. This is not possible for systems with unbounded rates and can be inefficient for systems where states have widely varying rates. In this work, we show how to generalize uniformization to allow different bounding rates for different states. We exploit this generalization to construct a flexible and efficient MCMC sampler without the limitations listed previously. We apply our sampler to the Lotka-Volterra model and a semi-Markov jump process with bursty states. Both these systems have unbounded rates, and we demonstrate the usefulness of our sampler on these problems.

References

[1] Fast MCMC sampling for Markov jump processes and continuous time Bayesian networks. V. Rao and Y. W. Teh, UAI 2011.

[2] Markoff chains as an aid in the study of Markoff processes. A. Jensen, Skand. Aktuarietiedskr, 1953.

Coauthors: Yee Whye Teh

⁂

Bayesian inference on nonlinear functions of Poisson rates

Lizanne RAUBENHEIMER (Rhodes University, South Africa)

In this paper the probability matching prior for the product of different powers of k Poisson rates is derived. This is achieved by using the differential equation procedure of Datta and Ghosh (1995). Using the method of Berger and Bernardo (1992), the reference prior for the ratio of two Poisson rates is also obtained. Simulation studies are done to compare different methods for constructing Bayesian confidence intervals. It seems that if one is interested in making Bayesian inferences on the product of different powers of k Poisson rates, the probability matching prior is the best. If on the other hand we want to obtain point estimates, credibility intervals or do hypothesis testing about the ratio of two Poisson rates, the uniform prior must be used.

Coauthors: Abrie Van der Merwe

⁂

Analysis of collaboration networks: elucidating firms relationships

Perla Reyes (University of California, Santa Cruz, USA)

In the last few years, there has been a growing interest into studying how inter-firm networks affect innovation and industry evolution. In particular, the effect of a collaboration network on a firm's propensity for innovation, measured on number of patents has been analyzed. However, thus far only descriptive statistics (degree distribution, clustering and centering coefficients) or graphical techniques have been used to analyze firm's alliances. We propose using Bayesian nonparametric mixture models to systematically analyze the network formation. The resulting models provide insides into how companies make alliances and may be used to asses the influence of the network in other aspects of the companies activities. As a case study we analyze information on alliances for 647 worldwide companies.

Coauthors: Abel Rodriguez, Zouhaier M'Chirgui and Enrique ter Horst

⁂

Bernstein-von Mises theorem for linear functionals of the density

Vincent Rivoirard (Université Paris Dauphine, France)

We study the asymptotic posterior distribution of linear functionals of the density by deriving general conditions to obtain a semiparametric version of the Bernstein-von Mises theorem. The special case of the cumulative distributive function evaluated at a specific point is widely considered. In particular, we show that for infinite dimensional exponential families, under quite general assumptions, the asymptotic posterior distribution of the functional can be either Gaussian or a mixture of Gaussian distributions with different centering points. This illustrates the positive but also the negative phenomena that can occur for the study of Bernstein-von Mises results.

Coauthors: Judith Rousseau

⁂

A Bayesian correlated cure rate model under a hybrid latent activation schemes

Josemar Rodrigues (Federal University of São Carlos, Brazil)

In this paper, we develop a Bayesian flexible cure rate survival model under the unobservable stages (initiation, promotion and progression) of the carcinogenesis process with dependent structure between the initiated cells. In order to formulate this dependence, we use an extension of the generalized power series distribution by including an additional parameter (the inflated-parameter generalized power series (IGPS) distribution, studied by Kolev et al. (2000), which has a natural interpretation in terms of "zero-inflated" proportion. The model is formulated under the hybrid or minimax latent activation schemes. In addtion, it is assumed that a tumor becomes detectable when its size reaches some threshold random level which means the proliferation of tumorais cells (or descendants) generated by the malignant cells. The number of initiated and tumorais cells are supposed to follow the IGPS probability distributions. The main advantage of the proposed model is to take in account the proportion of malignant cells that die before tumor induction in a competitive scenario. An example with a real data set is presented from the Bayesian paradigm.

Coauthors: Patrick Borges and Francisco Louzada-Neto

⁂

Asymptotic properties of the posterior distributions for HMMs with applications to the selection of the prior

Judith Rousseau (ENSAE-CREST and Université Paris Dauphine, France)

In this paper we study the asymptotic behaviour of the posterior distribution of HMMs. The emphasize is on finite state space HMMs although a general theorem is proposed. For finite state space HMMs, we prove that if the prior does not favour small transitions in the matrix then the posterior distribution has a good concentration behaviour and in the case of miss-specified HMMs, i.e. when the number of states in the model is greater than the actual number of states, under some reasonable choices on the types of priors on the transition matrix, the posterior distribution concentrates on the configuration which tends to merge the extra states to true states. Contrary-wise to the case of static mixture models no results are provided when the prior favours small probability transitions, because it may also correspond to small mixing hidden Markov chains.

Coauthors: Elisabeth Gassiat

⁂

Estimating link function parameters in robust Bayesian binary regression

Vivekananda ROY (Iowa State University, USA)

The logistic and probit regression models are most commonly used to analyze binary response data, but it is well known that their maximum likelihood estimators are not robust to outliers. Liu(2004) proposed a robust regression model, called the robit model, which replaces the normal (logistic) distribution in the probit (logit) regression model with the Student's $t-$distribution. Unlike probit and logistic model, the robit model has an extra degrees of freedom parameter. In this paper, we propose an empirical Bayes approach for estimating the degrees of freedom parameter along with estimation of the regression coefficients. We show that a combination of importance sampling based on a fast mixing Markov chain and an application of control variates can be used to efficiently estimate a large class of Bayes factors for selecting the degrees of freedom parameter of the robit model.

❈

Optimal sampling times for high dimensional pharmacokinetic studies

Elizabeth RYAN (Queensland University of Technology, Australia)

Here we present a sequential, simulation-based approach that can be used to determine optimal sampling times for population pharmacokinetic-pharmacodynamic studies, within a Bayesian decision theoretic framework. We consider the problem of choosing the placement of a fixed number of sampling times for racing horses that have been treated with the phenothiazine tranquiliser, Acepromazine (ACP), and enter the study sequentially. Design criteria that involve the time at which the drug concentration falls below a certain detection value and sampling cost are utilised. We also consider the effect of the deletion of sampling times from the fixed set of design points on the efficiency of prediction from the pharmacokinetic model.

Coauthors: Helen Thompson, Christopher Drovandi and Tony Pettitt

❈

The Wang-Landau algorithm reaches the flat histogram criterion in finite time

Robin RYDER (CEREMADE, Université Paris Dauphine, France)

The Wang-Landau algorithm aims at sampling from a probability distribution, while penalizing some regions of the state space and favouring others. It is widely used, but its convergence properties are still unknown. We show that for some variations of the algorithm, the Wang-Landau algorithm reaches the so-called Flat Histogram criterion in finite time, and that this criterion can be never reached for other variations. The arguments are shown on an simple context – compact spaces, density functions bounded from both sides– for the sake of clarity, and could be extended to more general contexts.

Coauthors: Pierre Jacob

❈

Can you hindcast the epidemic curve of an infectious disease

Gustaf RYDEVIK (University of York, UK)

When an epidemic outbreak is discovered, it has often been developing unseen for some time. Understanding its origin is then valuable. We present a method to retrospectively find out the development of an infectious disease from a cross-sectional data set collected at a single time point. By exploiting the differing time dependencies of different diagnostic tests, we can triangulate the time of infection for an individual. An MCMC approach is used to set up the posterior distribution of test results given infection, and estimating the distribution of time points given test results. We investigate the properties of this procedure by applying it to artificial data generated from several combinations of diagnostic test responses and epidemic shapes. The estimation of the epidemiological curve is proven to be both robust and precise. The method may be applied if two requirements are fulfilled: that each individual has been tested for the disease with two diagnostic methods, each giving a quantitative measure (e.g. ELISA and real-time PCR for testing immune response presence of pathogen); and having knowledge of test values as a function of time since infection. Our results indicate that hindcasting an epidemic curve without access to longitudinal data is theoretically possible and may be a useful tool in the epidemiology toolbox. They also highlight the additional information gained by utilising a variety of different methods when testing for disease.

Coauthors: Glenn Marion, Piran White and Mike Hutchings

❈

Mixtures of g-priors for objective Bayes selection of generalized additive models

Daniel SABANÉS BOVÉ (University of Zurich, Switzerland)

Recently, there has been increasing interest in mixtures of g-priors for objective Bayesian variable selection in linear models. These priors can be extended naturally to additive models, where smooth transformations of continuous covariates can be incorporated. We achieve this by using fixed-dimensional penalised splines in mixed model representation, and considering the implied marginal linear model with general covariance structure. The prior variances on the spline coefficients, defining the smoothness of the spline transformations, are taken from covariate-tailored finite sets of values and are part of the model index. Since the marginal likelihood has either a closed form or can be numerically approximated, efficient MCMC exploration of the discrete posterior model distribution is possible. We extend the approach to generalized additive models by deriving a generalized g-prior for non-normal mixed models. This directly extends the hyper-g priors for generalized linear models (Sabanes Bove and Held, 2011), from which the integrated Laplace approximation of the marginal likelihood is adopted (an efficient implementation is available in the R-package “hypergsplines”). Finally, we show how the methodology can be used for objective Bayes selection of proportional hazards models. We illustrate our approach with applications from nutritional epidemiology, where a large number of potentially influential dietary characteristics is available, and contrast it with competing methods from the literature.

Coauthors: Leonhard Held and Göran Kauermann

❈

Application of Bayesian method for estimating past effect of climate change on crop production

Gen SAKURAI (National Institute for Agro-Environmental Sciences, Japan)

How can we apply Bayesian methods to the evaluation of past and future climate change? The enhancement of photosynthetic efficiency predicted to occur under elevated [CO2], known as the CO2 fertilization effect, is a key source of uncertainty in predicting the effects of past and future climate change on crop production. Crop growth simulation models are one of the most important instruments to investigate how crop growth will respond to the possible abiotic environment of the future. However, many experimental studies have reported a decrease of photosynthesis and/or biomass growth compared with the theoretical increase in response to increased [CO2] (photosynthetic down-regulation). In this study, we inversely estimated the posterior distributions of the parameters of a soybean growth model (SPM-Soybean) together with the effect of down-regulation using a hierarchical Bayesian method (using replica exchange method and model selection by BPIC) by using an in-situ experimental data of crop growth. Furthermore, in order to expand the model to larger scale inference, we re-estimated the posterior distributions of the parameters of the model using yield data of soybean during 1980 to 2006 in three major production countries (U.S.A, China, and Brazil) using DiffeRential Evolution Adaptive method, and we estimated past CO2 fertilization effect. We estimated that the countries obtained quite large benefit from past CO2 increase: the yield was increased by from 8 to 30% during past 26 years.

Coauthors: Toshichika Iizumi and Masayuki Yokozawa

❈

Automated structured variational Bayes by stochastic natural gradient descent

Tim SALIMANS (Erasmus University Rotterdam, Netherlands)

Approximating an intractable posterior distribution by an approximating distribution of more convenient form is a central problem in Bayesian inference. A popular method of solving this problem is structured or fixed-form Variational Bayes, which works by numerically minimizing the Kullback-Leibler divergence of an approximating distribution in the exponential family to the intractable target distribution. An efficient way of performing this numerical optimization is by the use of a quasi-Newton method called natural gradient descent. We analyze the functional form of these natural gradients in a structured Variational Bayes setting and we present some results that facilitate their calculation. We then show how we can stochastically approximate these natural gradients in very general settings, which allows us to automate the optimization procedure and to extend the structured Variational Bayes approach to problems where it was previously not applicable. Our proposed algorithm makes approximation of posterior densities as easy and user friendly as gradient based optimization: The user supplies as input a function to calculate a (unnormalized) log posterior density and/or its gradient for given parameters, and the algorithm outputs a structured variational approximation to the corresponding posterior distribution. The type of approximating distribution is chosen by the user and can be any distribution in the exponential family or any mixture of such distributions, which means that our approximations can in principle be made arbitrarily precise. Despite its generality, we find our approach to be very competitive with Monte Carlo approximation, both in terms of speed as well as accuracy.

❈

Bayes test for monotonicity

Jean Bernard SALOMOND (ENSAE-CREST and Université Paris Dauphine, France)

We study the asymptotic behaviour of a Bayesian nonparametric test of qualitative hypotheses. More precisely, we focus on the problem of testing the monotonicity of a regression function, following an on going work of McVinish and Rousseau [2011]. Even if some results are known in the frequentist framework, no Bayesian testing procedure has been proposed, at least none has been studied theoretically. The test presented in this work is straightforward to implement, which is a great advantage compared to the frequentist tests proposed in the literature. We prove consistency of our testing procedure and also derive an upper bound on its separation rate. The separation rate is the rate at which the minimal distance between the null and a given alternative for which the test procedure is consistent, decreases with the number of observations. In our case, these alternative classes of regression functions, will be defined in terms of a Hölder smoothness and in terms of their distance to the set of monotone regression functions. We also investigate the possibility to design a testing procedure which would be adaptive, in the sense that it would not depend on the smoothness of the alternative class of functions and would lead to the optimal separation rate, for all classes of H"older functions with smoothness bounded by 1. Our approach can be generalized in order to test for other qualitative hypotheses such as positivity, unimodality or convexity.

⁂

Parallelized sequential Monte Carlo for large scale Bayesian variable selection

Christian SCHÄFER (ENSAE-CREST, CEREMADE and Université Paris Dauphine, France)

We present a novel Sequential Monte Carlo approach for estimating the posterior mean in Bayesian variable selection on normal linear regression models. In extensive numerical experiments, we show the proposed algorithm to yield significantly better results than standard Markov chain methods. Further, we demonstrate how a parallelized version of our algorithm can provide a precise estimate of the posterior mean within a few hours for real-world problems from plant genetics having thousands of covariates.

⁂

Bayesian MTC models to combine evidence from different sources of evidence. An application in rheumatoid arthritis

Susanne SCHMITZ (Trinity College Dublin, Ireland)

Bayesian mixed treatment comparison (MTC) models are an extension of Meta analysis and provide a flexible methodology to estimating the relative efficacy between treatments where direct evidence is not available or insufficient. Such estimates are essential for health care decision making to ensure an optimal distribution of funds. While randomized controlled trials (RCTs) are the gold standard of evidence in health care, other sources of information, such as observational data or registry data often exist and provide additional information. This additional data should also be included in the analysis; there is an argument that real life data gives a more realistic estimate of efficacy than controlled trials and the Bayesian approach advocates the inclusion of all available evidence. In this work we have extended a MTC model to include information from different trial designs. We have applied three alternative methodologies to data in rheumatoid arthritis (RA). The aim was the estimation of relative efficacy between five anti-TNF alpha antagonists, which are currently licensed for the treatment of RA in Europe. The simplest method, naïve pooling, does not differentiate between the trial designs, we cannot adjust for bias or measure consistency between trial designs. Using additional data to inform the prior distribution gives us the freedom to adjust for possible overestimation or variance inflation. The most flexible approach is a three level hierarchical model. It is possible to adjust for potential bias and estimates on trial design level and overall level allow to measure consistency between trial designs.

Coauthors: Cathal Walsh and Roisin Adams

⁂

Posterior contraction rates in estimating holder smooth Poisson intensities

Paulo SERRA (Technical University of Eindhoven, Netherlands)

Non-homogeneous Poisson processes are among the most widely used stochastic processes with applications in medicine, modeling of natural phenomenon such as cyclones and earthquakes, and finance, to mention but a few.

We observe increments of such a process with intensity $n\lambda$ where the function λ is a member of space Λ of bounded functions with Hölder smoothness α. We propose a prior measure supported on increasingly denser subspaces of Λ which are spanned by B-splines; the prior on the intensity function are induced by putting a prior on the coefficients of the B-splines. The corresponding posterior measure is studied from a frequentist point of view, in that we obtain a contraction rate – a Bayesian analog of the frequentist convergence rate – for the posterior measure in terms of the scaling parameter n.

The contraction rate of the posterior measure can then be determined based on entropy conditions on the support of the prior and on how the prior mass decays in comparison to the prior mass in a shrinking Kullback–Leibler ball around the true measure of the data. We show that the contraction rate $n^{-\alpha/(2\alpha+1)}$ holds for several (e.g. Normal, Uniform) prior measures on the coefficients of the B-splines.

We implement an MCMC algorithm to sample intensity functions from the posterior measure for simulated data.

Our approach combines the power of the Bayesian approach with the parsimony of orthonormal series based priors to provide posterior measures whose proximity to the true measure of the data can be quantified.

❉

An MCMC method for estimating the rate of preferential attachment in growing networks with missing timelines

Hidetoshi SHIMODAIRA (Tokyo Institute of Technology, Japan)

Preferential attachment is a simple mechanism that drives the evolution of growing networks by governing the formation of node-to-node connections, and measuring the extent that it occurs during the growth of a network is important for explaining its overall topology. Conventional methods require the timeline of a growing network is known, that is, the order in which the nodes of the network arrived in time is known. But growing network datasets are commonly accompanied by missing-timelines, meaning the order of the nodes in time is unknown. To address this shortcoming, we introduce an MCMC method for estimating the rate of preferential attachment in growing networks with missing-timelines. Key to our approach is that any growing network model gives rise to a probability distribution over the space of networks. In this manner a growing network model may be fitted to growing network data with missing-timeline, allowing not only for the rate of preferential attachment to be estimated as a model parameter, but the timeline also. Parameter estimation is achieved by implementing a novel Metropolis-Hastings sampling scheme for updating both the preferential attachment parameter and timeline. In a simulation study, we show that our method is able to correctly estimate preferential attachment in networks generated according to the underlying model. We also illustrate our method on a sub-network of the United States patent citation network. In this example the timeline is known, so we are able to validate our approach against the conventional methods, showing that they give consistent estimates.

Coauthors: Paul Sheridan

❉

Realized stochastic volatility with leverage and long memory

Shinichiro SHIROTA (University of Tokyo, Japan)

The daily return and the realized volatility are simultaneously modeled in the stochastic volatility model with leverage and long memory. In addition to the stochastic volatility model with leverage for the daily returns, ARFIMA process is jointly considered for the realized volatilities. Using a state space representation of the model, we estimate parameters by Markov chain Monte Carlo methods. Model comparison with similar realized stochastic volatility models with short memory is conducted by computing marginal likelihood and volatility forecast.

Coauthors: Takayuki Hizu and Yasuhiro Omori

❉

Generalization of the posterior distribution of the likelihood ratio to composite vs composite hypotheses testing

Isabelle SMITH (Laboratoire des Sciences du Climat et de l'Environnement, France)

The Posterior distribution of the Likelihood Ratio (PLR) concept is proposed by Dempster in 1974 for significance testing in the simple vs composite hypotheses case. The PLR (with inner threshold 1) is shown to be equal to 1 minus the frequentist p-value in a simple case. In 1997, Aitkin extends Dempster's result by adding a nuisance parameter and showing its asymptotic validity under more general distributions. In 2010, Smith & Ferrari extended the framework of reconciliation between the PLR and a frequentist p-value to a framework analogous to the one in which a credible (Bayesian) domain is equal to a confidence (frequentist) domain. By contrast, frequentist and classical Bayesian hypotheses testing seem to be irreconciliable (see Berger & Selke 1987 among others).

This general reconciliation result seems important but only concerns simple vs composite hypotheses testing. We propose now an extension of the PLR concept to the general composite vs composite hypothesis test. The PLR concept is also extended to the posterior distribution of an event related to a discrepancy variable different to the mere likelihood. Indeed, like in the posterior predictive p-value proposed by Meng in 1994, the choice of such a discrepancy variable could be discussed. But actually, we prove by a lemma analogous to the Neyman-Pearson lemma that under some reasonable (but necessarily debatable) optimization paradigm the likelihood is the most suited discrepancy variable. The (composite) PLR is also a p-value and a contributory term which is natural in other respects that will be discussed.

Coauthors: André Ferrari

❉

Bayesian inference for small area health data using the G-Wishart prior

Theresa SMITH (University of Washington, USA)

Hierarchical models are the most common way to incorporate spatial dependence in a fully Bayesian analysis of small area health data. The Gaussian intrinsic autoregression, or ICAR, prior is commonly used in applications as prior for spatially-dependent random effects in log-linear models [Besag et al., 1991]. Subsequent authors [Green and Richardson [2002]; Knorr-Held and Rasser [2000]] have developed semiparametric random effects models where clusters or collections of small areas share the same random effect. A priori, contiguous areas are more likely to be allocated to the same cluster or collection. We propose a Bayesian model for spatial data that uses Gaussian random effects with a flexible G-Wishart prior on the variance-covariance matrix [Dobra et al., 2011]. Despite its relevance in the Gaussian graphical models literature, the G-Wishart distribution has not been previously employed in spatial epidemiology. Our methodological research bridges this gap and brings recent state-of-the-art theoretical developments from the Gaussian graphical models literature to spatial epidemiology. The resulting Bayesian hierarchical models represent a novel, flexible class of models for spatial epidemiological applications. We illustrate Markov chain Monte Carlo algorithms for using the G-Wishart prior for disease mapping and compare our approach to Bayesian hierarchical models based on ICAR or semiparametric models.

Coauthors: Adrian Dobra

❉

A Bayesian semiparametric multiplicative error model

Reza SOLGI (University of Lugano, Switzerland)

In this paper we propose a semi parametric multiplicative error model (MEM) for daily realized volatility. In traditional MEM models, the innovations are typically assumed to be Gamma distributed (with one free parameter that ensures unit mean of the innovations and identifiability of the model), however empirical investigations unveils the inappropriateness of this choice. In the proposed model, the conditional mean of the time series is modeled parametrically, while we model its conditional distribution nonparametrically. In particular, in order to model the innovations nonparametrically, we employ a Dirichlet process mixture model with the Gamma distribution as its kernel. The inference is performed using Markov chain Monte Carlo simulation. This model is applied to the time series of daily realized volatility of some indices, and it has been compared to similar parametric models. A comparison in terms of predictive performance of the proposed model relative to competing parametric models available in the literature has been conducted, showing better predictive performance, flexibility and robustness to mis-specification of our Bayesian semi parametric approach.

Coauthors: Antonietta Mira

❉

Identification of Bayesian causal association models for investigation among gene expression, genotype variation and fatigue of breast cancer

Hiroko Solvang (Oslo University Hospital, Norway)

Fatigue is known as the most common side effect of cancer therapy. To elucidate the etiological mechanisms behind fatigue, we focus on developing an approach integrating gene expression, genotype variation and fatigue. To such approach, Lee et al (2009) developed model-based statistical tests applying likelihood-based model selection (LMS), which originally proposed by Schadt et al (2005). LMS involves three Bayesian models to causal associations among gene expression, SNP and disease, called 'causal', 'reactive' and 'independent' models. Although Schadt et al. applied LCM after detecting significant features by some statistical procedures, Lee et al., first applied LCM and secondly applied specific statistical test for each group to identify significant feature. Lee et al. called their approach two-step procedure. They pointed out that two-step procedures could avoid false negative to identify significant feature. However, our target for whole genome involves over 10K genes for expression and 100 K loci for genotype variation. Even if representative genes could be selected from whole genomic region, we should take into account all combinations for in-trans relationship. In such case, p-values obtained by the specific statistical test become much larger than FDR (false discovery rate) corrected p-values and it is hard to find significant features. Lee et al. also indicated that FDR did not find significant features to all combinations for some datasets. To such challenge, we consider LMS approach taking account Bayesian model selection in the first step of two-step procedure to select less number of features than the original approach.

⁂

Markov chain Monte Carlo method simulation of SQL injection attack detection

Michio Sonoda (Cyber University, Japan)

The SQL injection attack causes very serious problem to web applications which have database including personal data. To detect the SQL injection attack, the parsing and the black list based on the existed attack have been widely used. However, it is getting difficult to detect SQL injection attack because new attacks have been developed now. We have proposed the detection algorithm and mathematical model of SQL injection attack that do not depend on black lists. These methods extract an attack character by counting the content rate of some specific symbols included in the string of SQL injection attacks. But in the mathematical model that we have proposed, we used the maximum likelihood estimation of hyper parameter. But, we found that there exists weak point for detecting normal input in the method above. In this study, we computed the Bayesian predictive distribution of our proposed detection model by using Markov Chain Monte Carlo method. From our previous study results, we have already known the information on the parameter of our proposed model as the prior distribution. We will compare the Bayesian predictive distribution based on the result above with the predictive distribution using non informative prior distribution, and discuss the effectiveness of them.

Coauthors: Takeshi Matsuda, Daiki Koizumi and Shigeichi Hirasawa

⁂

Particle filtering for estimation of stationary dipoles in magnetoencephalogaphy

Alberto Sorrentino (University of Warwick, UK)

In MagnetoEncephaloGraphy (MEG) one looks for dynamic estimates of brain activity from non-invasive, millisecond-by-millisecond recordings of the magnetic fields produced by large neural populations inside the brain and measured outside the scalp. Under the dipole approximation, a single neural population is approximated as a point source (the dipole), characterized by a location, an orientation and a strength; in this case brain activity amounts to a time-varying set of point sources. The non-linear dependence of the data on the dipole location and the dynamic nature of the problem encourage application of particle filtering techniques. Particle filters are sequential Monte Carlo algorithms that produce samples approximately distributed according to the posterior distribution of an evolving unobserved system. In previous work we proposed a Sequential Importance Resampling (SIR) particle filter to estimate dynamically the number of sources and the source parameters from MEG data, achieving performances comparable to state-of-the-art methods, but with a notably higher level of automation. However, current dipoles were modelled as moving randomly within the brain, a key assumption for the SIR particle filter to work, but in conflict with the physiological interpretation of the current dipole as the activity of a particular neural population. Here we explicitly assume that the current dipoles do not move, and construct both a more realistic model and a more effective algorithm — based on the Resample Move idea of Gilks and Berzuini - that approximates reasonably the posterior distribution of a time-varying dipole configuration under the assumption of stationary dipole locations.

Coauthors: Adam M Johansen, John AD Aston, Tom E Nichols and Wilfrid S Kendall

⁂

Bivariate species sampling models

Polina SPORYSHEVA (Bocconi University, Italy)

Species Sampling Models were developed by Pitman [3, 4] and are based on the partition structure introduced by Kingman [2] and Aldous [1]. The corresponding sample is a Species Sampling Sequence if it is exchangeable and has a given prediction rule. We would like to extend these notions to problems of species sampling from two (or more) populations, in such a way that the resulting Species Sampling Models are connected and the overall sample is partially exchangeable (in the sense of de Finetti) with a given prediction scheme. In order to achieve this aim, we define a bivariate random partition. We present related prediction schemes and provide conditions under which samples from these schemes are partially exchangeable. The key role is played by a notion of partially exchangeable bivariate partition probability function.

References

1. D. J. Aldous. Exchangeability and related topics. In École d'été de probabilités de Saint-Flour, XIII—1983, volume 1117 of Lecture Notes in Math., pages 1–98. Springer, Berlin, 1985.
2. J. F. Kingman. The representation of partition structures. J. London Math. Soc, 18:374-380, 1978.
3. J. Pitman. Exchangeable and partially exchangeable random partitions. Probability Theory and Related Fields, 102:145-158, 1995.
4. J. Pitman. Some developments of the Blackwell-MacQueen urn scheme. In T.S. Ferguson et al., editor, Statistics, Probability and Game Theory; Papers in honor of David Blackwell, volume 30 of Lecture Notes-Monograph Series, pages 245–267. Institute of Mathematical Statistics, Hayward, California, 1996.

Coauthors: Sonia Petrone

❊

On estimation of mean squared errors of benchmarked empirical Bayes estimators

Rebecca STEORTS (University of Florida, USA)

We consider benchmarked empirical Bayes (EB) estimators under the basic area-level model of Fay and Herriot while requiring the standard benchmarking constraint. In this paper we determine how much mean squared error (MSE) is lost by constraining the estimates through benchmarking. We show that the increase due to benchmarking is $O(m^{-1})$, where m is the number of small areas. Furthermore, we find an asymptotically unbiased estimator of this MSE and compare it to the second-order approximation of the MSE of the EB estimator or equivalently of the MSE of the empirical best linear unbiased predictor (EBLUP), which was derived by Prasad and Rao(1990). Morever, using methods similar to those of Butar and Lahiri (2003), we compute a parametric bootstrap estimate of the MSE of the benchmarked EB estimate under the Fay-Herriot model and compare it to the MSE of the benchmarked EB estimate found by a second-order approximation. Finally, we illustrate our methods using SAIPE data from the U.S. Census Bureau.

Coauthors: Malay Ghosh

❊

Nonparametric analysis of a price mechanism of the Japanese private nursing home market via Polya tree mixture for a simultaneous model of demand

Shinya SUGAWARA (University of Tokyo, Japan)

This paper provides an econometric analysis for the Japanese private nursing home market. In this market, information asymmetry for consumers' longevity risks yields a complicated price mechanism which might worsen payoffs of many residents. We construct an economic model of simultaneous demand and supply and conduct a prediction analysis to simulate a governmental intervention which forces to vanish the mechanism. Since the conventional econometric methodology based on the generalized method of moments is not suitable for the prediction analysis, we adopt a nonparametric Bayesian methodology called the Polya tree mixture. Our empirical analysis via the Markov chain Monte Monte finds that the intervention increases the welfare of many consumers through a reduction of their payment.

Coauthors: Yasuhiro Omori

❊

Exploring the asymptotic behaviour of empirical Bayes procedures

Botond SZABO (Technical University of Eindhoven, Netherlands)

In recent years there has been a huge increase in the use of Bayesian methods in nonparametric statistical problems. One very popular adaptive Bayesian technique is the empirical Bayes method. Although it has a wide range of application, the technique itself does not have a full theoretical underpinning. In our work we aim to contribute to the fundamental understanding of this method. In Bayesian nonparametrics it is well known that the performance of a statistical procedure depends crucially on the choice of the prior distribution. A common approach that helps to avoid this problem is to work with a whole family of prior distributions, indexed by a scaling parameter. Popular adaptive methods for choosing the appropriate value of this hyper parameter are full (or hierarchical) and empirical Bayes procedures. In the hierarchical Bayes method the hyper parameter is endowed with a prior distribution and in the Bayesian framework this two level prior distribution is applied, while in the empirical Bayes method the hyper parameter is set at a data-dependent value, for instance by maximizing the marginal likelihood for the parameter. We study the latter approach in a standard nonparametric framework, the Gaussian sequence model. We show that the empirical Bayes method behaves optimally over a scale of regularity classes. Somewhat surprisingly the scale is limited, depending on the regularity of the prior distribution. Applying too rough a prior the method fails to work optimally and gives a sub-optimal posterior contraction rate. We investigate this phenomena in the hierarchical Bayes approach as well and establish that the performance of the posterior distribution for a large variety of hyper prior distributions is very similar to the empirical Bayes posterior distribution.

Coauthors: Aad van der Vaart and Harry van Zanten

❉

Spatial interdependence of fiscal efficiency among Japanese local governments

Ryota TAKADA (Keio University, Japan)

In the last decade, municipal mergers have been promoted in Japan as national policy which is based on the premises that (i) many municipalities in Japan are too small to be efficient; (ii) thus those smaller municipalities should be merged into bigger and more efficient ones. The policy to promote “bigger and better” local governments have prompted numerous studies on the validity of the premises it relies on, especially relationship between the size of a municipality and fiscal efficiency of the local government. Many previous studies have scrutinized fiscal efficiency of local governments by estimating their cost functions, though most of them do not give a full consideration on spatial policy interdependence among neighboring municipalities nor possible existence of unobservable inefficiency in the local government. This paper proposes a new Bayesian approach to evaluate fiscal efficiency of Japanese local governments with spatial autoregressive stochastic frontier models in which spatial autoregressive structure is introduced in order to incorporate spatial interdependence among municipalities and unobservable inefficiency is explained by a parametric stochastic frontier model with a half-normal or exponential municipality-specific error term. The results indicate that neglected spatial interdependence would have a significant impact on evaluation of fiscal efficiency.

Coauthors: Teruo Nakatsuma

❉

Minimax Bayesian predictive density operator

Fuyuhiko TANAKA (University of Tokyo, Japan)

In the quantum state tomography, which is the task of estimating the unknown quantum state on a Hilbert space from measurement data, the standard Bayesian technique is often used. In finite-dimensional Hilbert spaces, a quantum state is described by a positive semidefinite complex matrix of trace one, which is called a density operator. Since a density operator is considered as a quantum analogue of a probability distribution, Bayesian predictive density is naturally extended to the quantum setting, which is called a Bayesian predictive density operator. Although a Bayesian predictive density operator is shown to be optimal with a prior distribution, practically we need to specify an objective prior as in usual Bayesian statistics. Recently Komaki clarifies the relation between minimax prediction based on observed data and a least favorable prior for the nonasymptotic setting in a parametric submodel of a multinomial model. His main criterion is maximizing the conditional mutual information between the unknown parameter and future data. In this presentation, we consider a quantum analogue of this result. Since future data depend on how we select a measurement over the quantum system, it seems difficult to extend the definition of the mutual information between the unknown parameter and future data. Nevertheless, it is possible to formulate the problem of the estimation of the future density operator to be measured. Under some regularity conditions, we show that there exists a least favorable prior by which we can construct a minimax estimation of the unknown quantum density operator.

❉

Modeling area-level health rankings

Rusty TCHERNIS (Georgia State University, USA)

Objectives: To propose a model for estimation of area level health rankings and to compare it to the existing measures.
Methods: The area-level measures of health are estimated using factor analysis model with spatial correlation. The contribution of each variable to the resulting rank is derived from the data and population sizes are incorporated into a measure of uncertainty. The model is applied to county level data from Texas and Wisconsin and the results are compared with ranks from County Health Rankings.
Results: We summarize the posterior distribution of county health rankings in each state and compare them to the single ranks from County Health Rankings. In the case of Wisconsin we find few disagreements between the two rankings, but our model results in a coherent measure of uncertainty in the ranks. In the case of Texas we report vast differences in the resulting rankings. We also observe a strong suburban-urban health ranking gradient within metropolitan statistical areas.
Conclusions: Rankings that include health, social, and environmental variables and incorporate statistical uncertainty are useful for governmental agencies. Different methods may produce similar or dissimilar rankings. These differences should be reconciled before policy decisions are made.

Coauthors: Samir Soneji

⁂

A Bayesian predictive sample size selection design for single-arm exploratory clinical trials

Satoshi TERAMUKAI (Kyoto University Hospital, Japan)

The aim of an exploratory clinical trial is to determine if a new intervention is promising for further testing in confirmatory clinical trials. Most exploratory clinical trials are designed as single-arm trials using a binary outcome with or without interim monitoring for early stopping. In this context, we propose a Bayesian adaptive design based on the two-priors approach and predictive probabilities. This approach incorporates sample size determination and sample size selection during multiple interim analyses. First, the sample size is determined using the method of Sambucini (Statist. Med. 2008; 27:1199-1224), which adopts a predictive probability criterion with two kinds of prior distributions, that is, an ‘analysis prior’ used to compute posterior probabilities and a ‘design prior’ used to obtain prior predictive distributions. In the sample size determination, two sample sizes, i.e., N and Nmax, are prepared using two types of design priors, ‘enthusiastic’ and ‘sceptical’, respectively. Second, a choice between the two sample sizes is allowed at the time of any pre-specified interim analysis. At each interim analysis, the predictive probability is calculated, and the trial is stopped in case of low or high efficacy; otherwise, an optimal sample size, i.e., either N or Nmax, is selected. The proposed design, termed predictive sample size selection design, is an adaptive design that combines sample size determination, incorporating the frequentist features as a special case, together with sample size selection during multiple interim analyses. The properties of the design are investigated using simulation studies.

Coauthors: Takashi Daimon and Sarah Zohar

⁂

Analyzing first flowering event data using survival models with spatio-temporal covariates

Maria TERRES (Duke University, USA)

First flowering events in cherry trees are believed to be closely related to temperature patterns during the winter and spring months. Earlier work has incorporated the idea of temperature thresholds, defining chill and heat functions based on these thresholds. However, selection of the thresholds is often arbitrary and shared across species and locations. We propose a survival analysis framework with time-varying covariates having functional forms representing chill and heat accumulation leading up to first flowering events. The models are fit with a variety of possible chill and heat thresholds in order to select the most appropriate threshold pair. We utilize Cumulative Rank Probability Scores for model selection to choose a “best” threshold pair.

The model is first applied to 29 years of flowering data for four cherry species grown at a single location in Japan. This allows us to investigate the idea that relationship with temperature may differ across species. Next, the model is applied to 54 years of flowering data for 74 trees of a single species grown across Japan. By exploring flowering dates across locations we can investigate whether the relationship between temperature and first flowering event varies through space.

Coauthors: Alan Gelfand

⁂

Interaction networks changing with time

Thomas THORNE (Imperial College London, UK)

Modelling of gene regulatory network structures based on microarray time series data in the field of systems biology often assumes a rigid network structure that does not change over the course of the experiment. In reality many factors may influence the regulatory interactions at work, so that the network structure does not stay fixed, but varies depending on external influences. Previous methods have addressed this through the use of changepoint models, dividing the time series into discrete sections and allowing the network structure to vary within each segment. Doing so in a Bayesian framework requires us to make some prior assumptions about the number and distribution of changepoints within the time series. Here we present a method that instead applies Bayesian nonparametrics to infer a hidden state sequence governing the form of the interaction network, through the use of the Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM). This allows the number of parameters of the model to adapt as necessary to match the complexity of the observed data, and unlike other methods employing Bayesian nonparametrics takes into account the sequential nature of the data. We combine the HDP-HMM with a graphical model of the network structure describing the regulatory interactions at work. We apply our methodology to both simulated data and gene expression microarray data.

Coauthors: Michael P.H Stumpf

❈

BiiPS: A software for inference in Bayesian graphical models with sequential Monte Carlo methods

Adrien TODESCHINI (Inria Bordeaux, France)

The main factor in the success of Markov Chain Monte Carlo Methods is that they can be implemented with little efforts in a large variety of settings. Many softwares have been developped such as BUGS and JAGS, that helped to popularize Bayesian methods. These softwares allow the user to define his statistical model in a so-called BUGS language, then runs MCMC algorithms as a black box.

Although sequential Monte Carlo methods have become a very popular class of numerical methods over the last 15 years, there is no such “black box software” for this class of methods. The BiiPS software, which stands for Bayesian Inference with Interacting Particle Systems, aims at bridging this gap. From a graphical model defined in BUGS language, it automatically implements sequential Monte Carlo algorithms and provides summaries of the posterior distributions. In this poster, we will highlight some of the features of the BiiPS software, its R interface, and illustrations of this software applied to various models in target tracking, financial econometrics, chemical engineering or systems biology.

Coauthors: François Caron, Pierrick Legrand and Pierre Del Moral

❈

Model selection for regression density estimation with mixtures of heteroscedastic experts

Minh Ngoc TRAN (University of New South Wales, Australia)

This paper is concerned with the flexible regression density estimation problem which models the density of a response as a mixture of heteroscedastic normals with the means, variances and mixing probabilities all varying smoothly as functions of covariates. We use the variational Bayes approach and propose a novel fast algorithm for simultaneous covariate selection, component selection and parameter estimation. Our method is able to deal with the local maxima problem inherent in mixture model fitting, and is applicable to high-dimensional settings where the number of covariates can be larger than the sample size. In the special case of the classical regression model, the proposed algorithm is similar to currently used greedy algorithms while having many attractive properties and working efficiently in high-dimensional problems. The methodology is demonstrated through simulated and real examples.

Coauthors: David Nott and Robert Kohn

❈

Computational Bayesian tools for modeling the aging process

Sofia TSEPLETIDOU (Université Paris Dauphine, France)

Whereas aging is obvious in macroscopic organisms, it is not in single-celled ones, such as in E.Coli. However, the quantification of aging in E.Coli becomes possible due to an asymmetry observed during the inheritance of cellular membranes, poles. For this purpose, the complete lineage and some physiological characteristics of an exponentially growing colony have been determined experimentally. Devising a hidden Markov model where the age of the old pole cell would be modeled as a hidden variable given the observed physiological measurements is the key development of this project.

The non-linear dynamics of the hidden variable are reconstructed under the "Mass Conservation" principle and independently from the observed cell growth. Additionally, a mixture of two particles parameterizes the asymmetry's unknown directionality which has to be maintained at each step. Estimation becomes possible in form of exploration of the posterior distribution for the likelihood parameters: the growth rate of the hidden quantity, the asymmetry's parameters, and the trajectory of the hidden quantity along the tree. A popular solution for such problems is provided by sequential Monte-Carlo methods (SMC). Here, the tree structure introduces an additional difficulty: the particle system representing the posterior distribution of the hidden variable in the lineage tree has to be updated in two different manners, one for each child cell. Hence, the SMC framework is extended by introducing a branching SMC sampler, the particle system growing with the lineage tree.

Coauthors: Christian P. Robert

⁂

Comparing three MCMC algorithms: probability integral transformation, griddy Gibbs, and random walk draws

Hiroki TSURUMI (Rutgers University , USA)

Probability integral transformation, griddy Gibbs and random walk draws are compared for their efficiency of convergence using censored and noncensored models. Probability integral transformation performs best.

Coauthors: Cheng Gao and Xiangjin Shen

⁂

Robust Bayesian methods for survival analysis

Catalina VALLEJOS (University of Warwick, UK)

Classical survival distributions (exponential, gamma, Weibull, log-normal among others) produce inference that is not robust to the presence of outliers and does not represent the heterogeneity present between the observations in most real situations. This work presents the use of infinite mixture families of lifetime distributions as a solution for these two issues. This can be interpreted as the introduction of a random effect on the survival distribution. One example of these families are the Shape Mixture of Log-Normal (SMLN) distributions which can be interpreted either as the exponential of a random variable distributed as a Scale Mixture of Normals or as an infinite mixture of log-normal distributions with random shape parameter. This new class covers a wide range of shapes, in particular cases in which the log-normal cannot capture the tail behaviour of the data. It includes the already studied log-Laplace, log-logistic, log exponential power and log Student-t distributions among others. Bayesian inference under non-subjective priors based on the Jeffreys' rule is examined. In particular, it is highlighted that the existence of the posterior distribution of the parameters on the basis of a sample of point observations is not always guaranteed and a solution that considers set observations is implemented. Finally, the advantages of the use of the SMLN family are illustrated using a real dataset.

Coauthors: Mark F. J. Steel

⁂

Bayesian estimation of the spatial variation of the completeness magnitude for the Venezuelan seismic catalogue

Raquel VASQUEZ (Venezuelan Foundation for Seismological Research, Venezuela)

We apply Bayesian inference in order to assess the completeness magnitude Mc of an instrumental seismic catalogue compiled by the Venezuelan Foundation for Seismological Research during 2000-2010. This approach is known as the Bayesian magnitude of completeness (BMC) method. The events registered in this catalogue have been detected by the Venezuelan Seismological Network. In particular, we explore the spatial variation of Mc depending on the minimum number of stations where the seismic event has been detected and the distribution of seismic stations. We explore this technique to characterize the quality of the catalogue, to perform high resolution Mc mapping and to minimize errors in the Ms estimates due to spatial heterogeneities.

Coauthors: Lelys Bravo and Carolina Granado

⁂

Emulation and efficient history matching of stochastic systems biology models

Ian VERNON (Durham University, UK)

Systems Biology is a rapidly expanding area within the Biological Sciences. Typically it has involved the modelling of large chemical reaction networks through the use of ODE's. However, for many intracellular networks, especially those concerning gene transcription, the discrete number of molecules involved and the inherently stochastic behaviour of the network become important.

These networks can be accurately modelled as stochastic processes, that possess many unknown reaction rate constants representing all the various reactions involved. Systems Biologists want to learn about these rate parameters by comparing the model output to observed data on molecule counts.

By treating these networks as stochastic computer models, we have shown how to apply computer model methodology in order to learn about the reaction rate parameters. We have developed novel Bayes Linear emulators of stochastic models and shown how they can be used to perform history matching in order to efficiently identify regions of rate parameter space that are consistent with observed data. The history matching process is a powerful technique and is either a complete alternative to fully probabilistic Bayesian calibration or an extremely useful precursor to it. It proceeds using an iterative reduction of the rate or input parameter space using a series of increasingly more accurate emulators. This process raises several interesting questions in both the deterministic and stochastic model cases which we will discuss, including when is it best to change to the next iteration or wave, and how do we apportion an initial budget of runs between future waves.

Coauthors: Michael Goldstein

⁂

An objective prior for the number of trials in a binomial distribution

Cristiano VILLA (University of Kent, UK)

We propose an objective prior for the discrete parameter n (number of trials) of a binomial distribution. The prior assigns a mass to each value of the parameter space, not by means of an assessment of an objective probability directly, but by defining a utility expressing the “worth” of each n. We can assess the worth of n by describing and evaluating what is lost if n is removed from the parameter space. For if n is removed, and it is the true value, the Bayesian model will asymptotically accumulate at $n+1$. Hence we lose the relative information between the model with n and $n+1$ (measured by Kullback–Leibler divergence). We show that, even though the prior is improper, one observation of the data is sufficient to produce a proper posterior. Furthermore, we show that the posterior is consistent.

Coauthors: Stephen G. Walker

⁂

More informative density regression with multivariate covariates

Sara WADE (Bocconi University, Italy)

Density regression models obtained from Dirichlet Process mixtures of the joint distribution of the response and covariates allow the density of the response to change flexibly with the covariates by randomly partitioning subjects into clusters with similar behavior in x and similar relationship between y and x. Similarity is measured by the kernels of the mixture model where the joint kernels are parameterized as the marginal of x and the conditional of y. For multivariate covariates, independence across x is typically assumed for the marginal kernels due to computational and statistical reasons and known asymptotic properties. However, as the number of covariates, p, increases, a larger number of kernels is typically needed to approximate the marginal of x, while the relationship between y and x may be more stable. Furthermore, even for moderate p, the likelihood for x tends to dominate the posterior of the random partition so that the partition is solely determined by similarity in the covariate space. To address this issue, we propose to replace the Dirichlet Process with the Enriched Dirichlet Process, which has a richer parameterization and allows for a different number of kernels for the marginal and conditionals. This leads to more informative use of the data resulting in more reliable estimates, smaller credible intervals, and less influence of the prior due to larger sample sizes within cluster. Moreover, our proposal maintains a simple allocation rule. We demonstrate the advantages of our approach through simulated examples and an application to predict Alzheimer’s Disease.

Coauthors: Sonia Petrone

⁂

SNP-set association test with Hamming distance

Charlotte WANG (National Taiwan University, Taiwan)

With the advance in biotechnology and the completion of the international HapMap project, researchers are able to carry out association studies to investigate the association between diseases and genetic factors. For complex diseases, however, the genetic variants identified through genome-wide association studies (GWAS) provide only small to moderate increase in disease risk and contribute little to the overall heritability of the diseases. Many explanations for this missing heritability have been suggested. For instance, variants with small effects have not been found yet; rare variants with possibly larger effects have not been detected; gene-gene or gene-environment interactions are hard to be investigated. Another problem is, statistical methods for GWAS focusing on single-marker association tests are accompanied by multiple testing problems and low powers. These issues lead to several methods, such as multiple-maker tests, haplotype association tests, SNP-set analysis, gene set analysis, and pathway analysis. The Hamming distance, a measure of dis-similarity between two strings of equal length, is a well-known metric in information theory. This distance is also widely used in areas, including numerical taxonomy, cluster analysis, and construction of phylogenetic trees. In this proposal, we develop a SNP-set association test via the distribution of Hamming distance. This test has broad applicability in the sense that the SNP-set does not necessarily represent a haplotype and can stand for a gene, genes in a known pathway, or even longer regions in a chromosome. We will demonstrate its operational characteristics and compare its performance with other existing tests.

Coauthors: Chuhsing Kate Hsiao

❉

Analyzing spatial directional data using projected Gaussian processes

Fangpo WANG (Duke University, USA)

Directional data naturally arise in many scientific fields, such as oceanography (wave direction), meteorology (wind direction) and biology (animal migration direction). Our contribution is to develop a fully model-based approach to capture structured spatial dependence for modeling directional data at different spatial locations. We build a projected Gaussian process, induced from an inline multivariate Gaussian process. We illustrate the properties of the projected Gaussian process and show how to fit this model using suitable latent variable and Markov chain Monte Carlo methods. We also show how to implement spatial interpolation and conduct model comparison in this setting. Simulated and real data examples are provided for illustration

Coauthors: Alan E. Gelfand

❉

A Bayesian hierarchical model for family association studies with rare CNV variants

Shi-Heng WANG (National Taiwan University, Taiwan)

Many statistical methods for association studies have been developed based on the assumption of common disease common variant. For genotypic variation that is caused collectively by multiple rare variants (common disease rare variant), however, statistical methodology reamins limited. Existing methods compared the pooled variant allele distribution in cases and controls, yet such methods cannot be easily applied to family-based studies. This paper aimed to explore the association between the collapsing CNV variants and disease status with a Bayesian hierarchical model to account for the dependence within family members and among the repeated measurements of CNVs data. This Bayesian hierarchical model assigns integers in a probabilistic sense to the quantitatively measured copy numbers, and is able to test the association simultaneously between all variants under study. This integrated model can account for the uncertainly of copy number assignment and differentiate if the variation was de novo or inherited on the basis of posterior probabilities. In addition, this Bayesian model also considers the case of mutation by allowing the insertion and deletion to occur in offspring. This Bayesian hierarchical model will be applied to a real application, schizophrenia family study, and to simulations containing families with different distributions for phenotype and CNV.

Coauthors: Yu-Chin Tsai, Wei J. Chen, Hai-Gwo Hwu and Chuhsing K. Hsiao

❉

Analysis of differential gene expression in heterogeneous tumor samples

Wenyi WANG (University of Texas MD Anderson Cancer Center, USA)

Clinically derived tumor tissues are often times made of both cancer and normal stromal cells. The expression measures of these samples are therefore partially derived from normal cells. This may explain why some previous studies have identified only a fraction of differentially expressed genes between tumor and normal samples. What makes the in silico estimation of mixture components more difficult is that the percentage of normal cells varies from one tissue sample to another. Until recently, there has been limited work on statistical methods development that accounts for tumor heterogeneity in gene expression data. To this end, we have built a Bayesian hierarchical model and developed a Markov chain Monte Carlo approach to simultaneously estimate, in each tumor sample, the normal cell fractions (i.e. level of stromal contamination), as well as cancer cell-specific gene expressions. We illustrate the performance of our model in simulated microarray and RNAseq data as well as in real data.

Coauthors: Jaeil Ahn and Ying Yuan

❉

Bayesian nonparametric estimation of a copula

Xue WANG (University of Kent, UK)

Copulas have recently become popular as a modeling tool for accounting for the dependence structure of multivariate data. Unlike the marginal and the joint distributions, a copula is a hidden dependence structure that couples a joint distribution with its marginals. This makes the task of proposing a parametric copula model non-trivial and more attention is placed on using nonparametric methods.

We present here a Bayesian nonparametric methodology to estimate arbitrary bivariate copula density and prove that any bivariate copula density can be approximated by an infinite mixture of the Gaussian copula densities, the dependence structures of the pairs with standard normal marginals. A slice sampling idea is introduced for this infinite structure that makes up for the impossible via full Bayesian posterior analysis. The estimation procedure is proposed by the Gibbs sampling algorithm.

The proposed approach extends the range of estimation since there is no need to any assumption of the copula family models. If a copula function is known, our approach can reconstruct the key features of a copula that are symmetry, skewness or linear relationship. It is also true on the phenomenon when the copula is unknown.

Coauthors: Juan Wu and Stephen G. Walker

❉

Bayesian inference in structural second-price auctions with both private-value and common-value bidders

Bertil WEGMANN (Linköping University, Sweden)

Auctions with asymmetric bidders have been actively studied in recent years. Tan and Xing (2011) show the existence of monotone pure-strategy equilibrium in auctions with both private-value and common-value bidders. The equilibrium bid function is given as the solution to an ordinary differential equation (ODE). We approximate the ODE and obtain a very accurate, approximate inverse bid as an explicit function of a given bid. This results in fast and numerically stable likelihood evaluations, which is an extremely valuable property for inference. We propose a model where the valuations of both common-value and private-value bidders are functions of covariates. The probability of being a common- value bidder is modeled by a logistic regression model with Bayesian variable selection. The model is estimated on a dataset of eBay coin auctions. We analyze the model using Bayesian methods implemented via a Metropolis-within-Gibbs algorithm. The posterior inference of the common-value part of the model is similar to the ones obtained from a model with only common-value bidders, whereas the private-value part of the model is more affected by the introduction of common-value bidders. There is on average a slightly larger probability for a bidder to be a common-value bidder, but this probability depends very little on the auction- specific covariates.

❉

Bayesian analysis of current status data with missing covariates

Chi-Chung WEN (Tamkang University, Taiwan)

In current status data, the failure time of interest is only known whether it exceeds a random survey time or not. Covariate missing problems have been extensively studies for right-censored data, but less so for current status data. Motivated by the fracture data from 2005 National Health Survey of Taiwan, where the occurrence of fractures was current status censored and the covariate osteoporosis was reported only for 28% respondents, we proposed a Bayesian inverse probability weighted (IPW) method to analyze current status data with missing covariates under the Cox proportional hazards model. The general complete-case method is inconsistent when the missingness depends on outcome variable and the frequentist IPW method has difficulty in variance estimation due to the information matrix does not have a closed form. By contrast, the proposed method is consistent and can easily provide interval estimate for regression parameter. Simulations and a real data analysis are included.

⁂

Efficient MCMC estimation of some elliptical copula regression models through scale mixture of normal

Nuttanan WICHITAKSORN (University of Sydney, Australia)

In this study, we simplify the implementation of some elliptical copula regression models by expressing the copula and its marginals as the scale mixture of normal and some other distributions. This allows us to obtain the parameter estimates through the efficient Markov Chain Monte Carlo estimation. We conduct a simulation study to assess the performance of our proposed method. We also apply our model and method to some real data applications.

Coauthors: S.T. Boris Choy and Richard Gerlach

⁂

What is the value of longer AMOC observations?

Daniel WILLIAMSON (Durham University, UK)

NERC-RAPID has currently taken 10 years of observations of the Atlantic Meridional Overturning Circulation (AMOC) at latitude 26 degrees north. How many more years are required in order to refine long term forecasts for future AMOC change under anthropogenic forcing? We consider the value of AMOC data in terms of its ability to constrain the parameter space of coupled climate models and thus reduce uncertainty in current AMOC projections. Using part of an unprecedented 10,000 member ensemble of runs of the global coupled climate model HadCM3 that we have designed and managed through climateprediction.net, we build emulators using constrained Bayesian dynamic linear models to help us answer this question via History Matching. Using these emulators we can explore the answer to this question over a range of anthropogenic forcing scenarios and for a range of model discrepancies.

Coauthors: Michael Goldstein and Adam Blaker

⁂

The evaluation of evidence relating to traces of drugs on banknotes

Amy WILSON (University of Edinburgh, UK)

Banknotes can be seized from crime scenes as evidence of illicit drug use or dealing. Mass Spec Analytical Ltd., an analytical chemistry company, have developed a technique to analyse the quantities of drugs on banknotes. Data are available from banknotes seized in criminal investigations, as well as from banknotes from the general circulation. For each sample tested, the analytical response over time is recorded for five different drugs. A peak detection algorithm used to convert these data into a measurement of the quantity of drug on each banknote will be presented.

Two questions are considered. The first focuses on the likelihood of the data under each of two propositions: that a set of seized banknotes is associated with drug crime, and that these banknotes are from the general circulation. The aim is to evaluate the associated likelihood ratio. There is evidence of autocorrelation between adjacent banknotes in samples. Two models have been developed to take this into account: an autoregressive process of order one and a hidden Markov model. Non-parametric models using kernel regression are being developed. These models will be described, with preliminary results presented.

The second question involves the calculation of a likelihood ratio where data are available from bundles within samples, thus the within sample variation may be measured. The propositions are that two samples of banknotes have originated from the same source, and that they have originated from different sources. The use of the above models in evaluating this likelihood ratio will be described.

Coauthors: Colin Aitken, Richard Sleeman and Jim Carter

⁂

Forecasting high-dimensional, time-varying covariance matrices: factor stochastic volatility vs. realized covariances

Jesse Windle (University of Texas at Austin, USA)

Factor stochastic volatility (FSVol) models are the work horse of Bayesian multivariate volatility estimation and forecasting. An alternative approach to variance-covariance estimation relies upon high-frequency intraday data to construct approximations to the daily variance-covariance matrix. In one dimension these high-frequency estimates are superior to the estimates produced by stochastic volatility. Consequently, one may treat these high-frequency statistics as data to model and arrive at superior forecasts of daily volatility as well.

Following an analogous path of exploration, we compare high-dimensional estimates and forecasts found using FSVol models and high-frequency data. We methodically interpolate between procedures that make no use of intraday data and procedures that exclusively use intraday data to examine the important directions of variation found in high-frequency variance-covariance estimates. Under an economically meaningful measure of loss, the empirical standard deviation of one-day ahead minimum variance portfolios, we find that simple estimation and forecasting procedures that use high-frequency data are superior to FSVol, even in high-dimensional settings. Thus one should model the high-frequency estimates of the daily variance-covariance matrix directly.

Models of high-frequency statistics have been implemented in low dimensional settings. These models assume that the daily covariance matrix comes from a stationary, matrix-valued time series. We suggest employing simple Bayesian forecasting techniques when working in high-dimensions. Such simple techniques allow one to efficiently and robustly forecast large covariance matrices while accommodating intervention for occasional extreme events.

Coauthors: Carlos Carvalho

*

Granger causal analysis of VARMA GARCH models

Tomasz Wozniak (University of Melbourne, Australia)

Recent economic developments have shown the importance of spillover and contagion effects in financial markets. Such effects are not limited to relations between the levels of financial variables but also impact on the volatility. I investigate Granger causality in conditional mean and conditional variances of time series. For this purpose a VARMA GARCH model is used. I derive parametric restrictions for the hypothesis of noncausality in conditional variances between two groups of variables, when there are other variables in the system as well. These novel conditions are convenient for the analysis of potentially big systems of economic variables. Such systems should be considered in order to avoid the problem of omitted variable bias. Further, I apply a Bayesian procedure in order to test the restrictions. The test avoids the singularity problem that may appear in the classical Wald test. Also, it relaxes the assumption of existence of higher order moments of the residuals.

*

Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov chain Monte Carlo

Chieh-Hsi Wu (University of Auckland, New Zealand)

Microsatellites are popular genetic markers for making inferences on molecular evolution and population genetics, as they are abundant, widely distributed in the genome, and highly polymorphic. We provide a framework for Bayesian coalescent inference from microsatellite data that enables inference of population history parameters averaged over microsatellite mutation models. We introduce a new model that facilitates Bayesian model selection over a rich family of microsatellite mutation models defined in existing literature. These microsatellite models are based on the replication slippage mechanism and focus on three properties of microsatellite mutation: length dependency of mutation rate, mutational bias towards expansion or contraction, and number of repeat units changed in a single mutation event. Using our new model in a coalescent inference, the posterior distributions of population history parameters are integrated across a set of microsatellite models and thus account for model uncertainty. Simulated data are used to evaluate our method in terms of accuracy and precision of θ (product of mutation rate and population size) estimation and also identification of the true mutation model.

Coauthors: Alexei J. Drummond

*

Consistency of stochastic blockmodels

Yuefeng WU (University of California, Santa Cruz, USA)

Stochastic blockmodels are used to analyze the community structure of biological, social or computer networks. Nonparametric Bayesian hierarchical models have been used to model the latent clusters by researchers in different fields. However, the important issue of consistency was left open. We give sufficient conditions, under which these nonparametric priors possess consistency under partial exchangeability.

❊

Robustness of prior information to the predictive power

Yuh-Jenn WU (Chung Yuan Christian University, Taiwan)

At some interim point during the course of a clinical trial, the decision making group may wish to use the data on the initial series of patients to judge the likely consequences of further patient accrual. Many trials have used the concept of conditional power developed by Halperin et al. (1982) as a guideline for interim data analyses. Spiegelhalter, Freedman, and Blackburn (1986) argue that simply conditioning on selected hypotheses that were described at the beginning of the trial would ignore the noteworthy knowledge concerning the treatment effects that has accumulated by the time of data monitoring. Hence they suggest averaging the conditional power function with respect to the current opinion about the treatment effect for consideration of early termination. Doing so leads to a Bayesian approach and the unconditional probability is considered as predictive power. However the choice of prior is often criticized for being subjective and nonrobust. In the study, we compare the conditional power with predictive power. In addition, we also explore the robustness of predictive power with respect to changes in prior distributions.

Coauthors: Chin-Fu Hsiao

❊

Predictive sequential classification

Jie XIONG (University of Helsinki, Finland)

We develop inductive rules for sequential probabilistic supervised and semi-supervised classification of data arising from multiple finite alphabets. The rules are based on combining a decision-theoretic formulation with predictive representation of data derived under random urn models and hidden Markov models, as well as generalized exchangeability. Our classification rules have attractive theoretical properties that conditional on the training data predict the labels of sequentially observed items based on their features and the previously observed items whose true labels are unknown. Optimal classification rules are defined under different utility functions and we examine the behavior of the supervised rules as a function of the amount of sequential testing data to show the advantages of the simultaneous classifiers compared to a standard predictive classifier which assigns all test items independently. Asymptotic properties of the predictive sequential classifiers are investigated and we illustrate that the advantages of considering data from the previous items reaches a saturation when the amount of query data increases. Multiple possible stochastic learning algorithms are considered for implementing the sequential classifiers.

Coauthors: Jukka Corander, Yaqiong Cui and Timo Koski

❊

Differential patterns of interaction and Gaussian graphical models

Masanao YAJIMA (University of California, Los Angeles, USA)

We propose a methodological framework to assess heterogeneous patterns of association amongst components of a random vector expressed as a Gaussian directed acyclic graph. The proposed framework is likely to be useful when primary interest focuses on potential contrasts characterizing the association structure between known subgroups of a given sample. We provide inferential frameworks as well as an efficient computational algorithm to fit such a model and illustrate its validity through a simulation. We apply the model to a Reverse Phase Protein Array data on Acute Myeloid Leukemia patients to show the contrast of association structure between the refractory patients and the relapsed patients.

Coauthors: Donatello Telesca, Yuan Ji and Peter Muller

❊

Time instability of the U.S. monetary system: multiple break tests and reduced rank TVP VAR

Yohei Yamamoto (University of Alberta, Canada)

Earlier attempts to find evidence of time varying coefficients in the U.S. monetary vector autoregression have been only partially successful. Structural break tests applied to typical data sets often fail to reject the null hypothesis of no break. Bayesian inferences using time varying parameter vector autoregressions provide posterior median values that capture some important movements over time, but the associated confidence intervals are often very wide and make the entire results less conclusive. We apply recently developed multiple structural break tests and find statistically significant evidence of time varying coefficients. We also develop a reduced rank time varying parameter vector autoregression with multivariate stochastic volatility. Our model has a smaller number of free parameters and thereby yields tighter confidence intervals than previously employed unrestricted time varying parameter models.

Coauthors: Dukpa Kim

⁂

Hidden Markov model and reversible jump algorithm for the analysis of areas of interest from eye tracking data

Jin Yan (University of Maryland, College Park, USA)

When looking at an advertisement, consumers' attention jumps from one point to another, until enough information is collected. Previous studies have pointed out that only part of the ad-image is frequently viewed by consumers, while other areas are rarely visited. By analyzing of eye tracking data of consumers on 17 different pages of a magazine, we want to find out what areas on the pages attract consumers more, and how consumers' attention jumps among these areas. We assume that there are unknown number of areas of interest (AOIs) with unknown centers and sizes, and that consumers' eye-fixations are located on one of these AOIs at a time. The sequence of AOIs visited by each consumer, which is unobservable, is modeled as a Markov chain. Conditional on these hidden chains, we assume that each consumer's fixations follow a spatial Poison distribution, whose intensity is proportional to a truncated bivariate normal density. The means of the densities equal the centers of these AOIs, and the covariance matrices represent the sizes. When the number of AOIs is fixed, we can estimate the means, the covariance matrices and the transition matrices using Markov chain Monte Carlo method and the forward-backward algorithm. When the number of AOIs is unknown, we use a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm to estimate the number of AOIs in addition. The RJMCMC algorithm is known to have slow convergence and small acceptance rates, so we investigate modifications to speed up the convergence and initiate more mixing of the chain.

Coauthors: Michel Wedel and Paul J. Smith

⁂

Permissive boundary prior function as a virtually proper prior density

Takemi Yanagimoto (Chuo University, Japan)

Researches on the validity of the use of an improper prior function (density) are still surprisingly sparse in spite of its familiarity in practical applications. The strong reservation on the use of an improper function comes from the use of the marginal likelihood, which is the key in the empirical Bayes method and the model selection. Because of the primary role of the posterior density in Bayesian theory, we attempt to justify the use of an improper function by regarding it as a virtually proper prior density. An improper prior function is called a permissive boundary function, when it satisfies the following two conditions; 1) it is a limit of a sequence of prior densities and 2) the sequence of the induced posterior densities converges smoothly under the dual logarithmic divergences to the corresponding posterior density. We observe that these properties allow an improper function to treat it virtually as a proper density, and give examples where familiar improper functions do not satisfy the latter condition. Implications of this notion to the empirical Bayes method are discussed.

Coauthors: Toshio Ohnishi

⁂

Instrumental variables estimation of a flexible nonlinear model

Jae Ho Yoon (POSCO Research Institute, Korea)

The applicability of Hamilton's (2001) flexible nonlinear model for estimating simultaneous equations model or errors in variables model is placed under restraint due to the existence of endogenous explanatory variables. This paper proposes IVFM (instrumental variables estimation of a flexible nonlinear model) for solving the case of endogenous explanatory variables using a standard estimation method. The findings of this paper are as follows: this paper theoretically solves a flexible nonlinear model with the endogenous explanatory variables by using instrumental variables; and also empirically proves the applicability of IVFM for simultaneous equations model or error in variables model. As we applied the proposed model to Campbell and Mankiw's (1989) consumption function, we found that the relationship is linear between the log difference of per-capita disposable income and the log difference of per-capita consumption on non-durable goods and services.

Coauthors: Young-Wan Goo

* * *

Spatial patterns of flypaper effects for local expenditure by policy objective in Japan: a Bayesian approach

Hideo Yunoue (Chiba University, Japan)

The empirical work on local public finance has found that the marginal effect of lump-sum grants on expenditure is larger than that of income, thereby providing evidence of the "flypaper effect". However, most existing studies only employ single equation models to test the flypaper effect. In this paper, we specify a seemingly unrelated regression (SUR) model to examine the flypaper effect in Japan, primarily because other categories of expenditure influence the expenditure on particular policy objectives. We also include spatial interaction in our estimation model and employ a Bayesian approach in estimating our model. Our results show that SUR with a spatial error model is better for this purpose than several other specifications. Using this approach, we observe evidence of the flypaper effect in land development, police, education, and debt expenditure, and spatial interaction in sanitation, police, education, and disaster recovery expenditure.

Coauthors: Kazuhiko Kakamu and Takashi Kuramoto

* * *

Spatio temporal model for multivariate traffic accident count data

Zamira Zamzuri (Macquarie University, Australia)

This paper introduces a new spatio-temporal model to accommodate multivariate accident count data. Available models in the accident literature to date can either only cope with one factor; spatial or temporal, or are being developed under a univariate framework. Based on the multivariate Poisson — lognormal model, we further develop this model by introducing a linear combinations of random impulses to capture spatial correlation and adding lagged observations for day to day effects. The estimation of this model is then carried out using the Markov Chain Monte Carlo simulation method. Simulated data sets are used in assessing the performance of this model. Then, this model is applied to the New South Wales traffic accident count data to test its practicality. An advantage of this new model is that it not only copes with three sources of variations; time, space and multivariate data variations, but also provides information on time and space dependency that has been often ignored in past models. Finally, this model will benefit practitioners by providing an improvement in the accuracy of traffic accident modelling.

Coauthors: Graham Wood, Ross Sparks and Gillian Heller

* * *

The interplay of frequentist and Bayesian nonparametric Statistics

Mahmoud Zarepour (University of Ottawa, Canada)

The Dirichlet Process and its generalization, the two-parameter Poisson-Dirichlet process, has been used in Bayesian inference frequently. When the concentration parameter is large, the two-parameter Poisson-Dirichlet process and its corresponding quantile process share many asymptotic properties with the frequentist empirical and quantile processes. Specifically, let $P_{H,\theta,a}$ be the two- parameter Poisson-Dirichlet random probability measure on $\mathbb{R}$ with a base probability measure H, a discount parameter $\theta \in (0,1]$ and a concentration parameter $a > -\theta$. For the two important cases when $\theta \in \{0, 1/2\}$, we show that the process $\sqrt{a(1-\theta)^{-1}}\left(P_{H,\theta,a}((-\infty,t]) - H((-\infty,t])\right)$ converge s to a Brownian bridge as $a \to \infty$. The weak convergence of the corresponding quantile process is also obtained. In addition, a Glivenko-Cantelli theorem is derived for any $\theta \ in[0,1)$ as $a \to \infty$.

Coauthors: Luai Al Labadi

* * *

Nonparametric Bayes methods for estimation in high-dimensional contingency tables

Jing ZHOU (University of North Carolina at Chapel Hill, USA)

A fundamental barrier limiting more rapid discovery of adverse risk factors and development of accurate predictive models is the multiplicity problem, particularly given that it is crucial in many settings to account for interactions. We consider data from the National Birth Defects Prevention Study (NBDPS), the largest population-based study ever conducted in the United States on the etiology of birth defects. The study was designed to evaluate environmental, behavioral, biomedical, sociodemographic, and genetic factors associated with the occurrence of congenital malformations. If one considers placing individual pregnant women in a multi-way table defined by levels of these potentially important variables, only a few cells in the table will have more than a few women, with most cells having no women. Hence, it is not possible simultaneously to estimate an odds ratio for each cell of the table without borrowing information strongly. Potentially, one can preselect variables that seem likely to be important and ignore interactions to reduce dimensionality. Alternatively, one can include variables one at a time as predictors in a low-dimensional logistic regression model, with interactions only added for predictors with main effects, and significance thresholds on p-values adjusted for false discovery rate control. However, such approaches can lead to overlooking important risk factors and lack a probabilistic characterization of uncertainty, which is important in making inferences and representing uncertainty in predictions. We apply the nonparametric Bayesian approach of Dunson and Xing (2009, JASA) to NBDPS data, introducing two measures of association for interpreting results obtained using their approach.

Coauthors: Amy H. Herring, Arthur S. Aylsworth and Andrew F. Olshan

❊

List of participants

Abe, Makoto • *University of Tokyo, Japan* • `abe@e.u-tokyo.ac.jp`
Abe, Takayuki • *Keio University, Japan* • `tabe@z5.keio.jp`
Aida, Toshiaki • *Okayama University, Japan* • `aida@cs.okayama-u.ac.jp`
Aitken, Colin • *University of Edinburgh, UK* • `c.g.g.aitken@ed.ac.uk`
Allchin, Lorraine • *University of Oxford, UK* • `lorraine.allchin@lincoln.ox.ac.uk`
Allenby, Greg • *Ohio State University, USA* • `allenby.1@osu.edu`
Alquier, Pierre • *Université Paris 7 and CREST, France* • `alquier@math.univ-paris-diderot.fr`
Alston, Clair • *Queensland University of Technology, Australia* • `c.alston@qut.edu.au`
Amano, Tomoyuki • *Wakayama University, Japan* • `tomchami@center.wakayama-u.ac.jp`
Anacleto, Osvaldo • *The Open University, UK* • `o.anacleto-junior@open.ac.uk`
Ando, Tomohiro • *Keio University, Japan* • `andoh@kbs.keio.ac.jp`
Andrianakis, Ioannis • *National Oceanography Centre, UK* • `i.andrianakis@noc.ac.uk`
Andrieu, Christophe • *University of Bristol, UK* • `c.andrieu@bristol.ac.uk`
Antoniano Villalobos, Isadora • *University of Kent, UK* • `ia57@kent.ac.uk`
Araki, Nagateru • *Osaka Prefecture University, Japan* • `araki@eco.osakafu-u.ac.jp`
Arbel, Julyan • *CREST, Université Paris Dauphine, France* • `julyan.arbel@polytechnique.org`
Argiento, Raffaele • *CNR-IMATI, Italy* • `raffaele@mi.imati.cnr.it`
Arnold, Richard • *Victoria University of Wellington, New Zealand* • `richard.arnold@msor.vuw.ac.nz`
Asahina, Seitaro • *Astellas Pharma Inc., Japan* • `seitaro.asahina@astellas.com`
Aslett, Louis • *Trinity College Dublin, Ireland* • `louis@maths.tcd.ie`
Asseburg, Christian • *ESiOR Oy, Finland* • `christian.asseburg@esior.fi`
Atchade, Yves • *University of Michigan, USA* • `yvesa@umich.edu`
Bacallado, Sergio • *Stanford University, USA* • `sergiobacallado@gmail.com`
Baines, Paul • *University of California, Davis, USA* • `pdbaines@ucdavis.edu`
Baladandayuthapani, Veera • *University of Texas MD Anderson Cancer Center, USA* • `veera@mdanderson.org`
Ball, Roderick • *Scion - New Zealand Forest Research Institute, New Zealand* • `rod.ball@scionresearch.com`
Bandyopadhyay, Dipankar • *University of Minnesota, USA* • `dbandyop@umn.edu`
Banerjee, Anjishnu • *Duke University, USA* • `ab229@stat.duke.edu`
Banerjee, Sudipto • *University of Minnesota, USA* • `baner009@umn.edu`
Banks, David • *Duke University, United States* • `banks@stat.duke.edu`
Barbillon, Pierre • *AgroParisTech, France* • `pierre.barbillon@agroparistech.fr`
Barclay, Lorna • *University of Warwick, UK* • `L.M.Barclay@warwick.ac.uk`
Basturk, Nalan • *Erasmus University Rotterdam, Netherlands* • `basturk@ese.eur.nl`
Basu, Sanib • *Northern Illinois University, USA* • `basu@math.niu.edu`
Bayarri, Susie • *University of Valencia, Spain* • `susie.bayarri@uv.es`
Beamonte, Maria Asuncion • *Universidad de Zaragoza, Spain* • `asunbea@unizar.es`
Belitser, Eduard • *Eindhoven University of Technology, Netherlands* • `e.n.belitser@tue.nl`
Berihuete, Angel • *University of Cadiz, Spain* • `angel.berihuete@uca.es`
Berrett, Candace • *Brigham Young University, USA* • `cberrett@stat.byu.edu`
Berrocal, Veronica J. • *University of Michigan, USA* • `berrocal@umich.edu`
Berry, Donald A. • *University of Texas MD Anderson Cancer Center, USA* • `dberry@mdanderson.org`
Beskos, Alexandros • *University College London, UK* • `a.beskos@ucl.ac.uk`
Blum, Michael • *CNRS, Université Joseph Fourier, France* • `michael.blum@imag.fr`
Bonassi, Fernando V. • *Duke University, USA* • `fernando@stat.duke.edu`
Bornkamp, Bjoern • *Novartis Pharma AG, Switzerland* • `bjoern.bornkamp@novartis.com`
Bornn, Luke • *University of British Columbia, Canada* • `l.bornn@stat.ubc.ca`
Bottolo, Leonardo • *Imperial College London, UK* • `lbottolo@imperial.ac.uk`
Braun, Michael • *Massachusetts Institute of Technology, USA* • `braunm@mit.edu`
Broderick, Tamara • *University of California, Berkeley, USA* • `tab@stat.berkeley.edu`
Brown, Philip • *University of Kent, UK* • `pjb8@kent.ac.uk`
Burgette, Lane • *RAND Corporation, USA* • `burgette@rand.org`
Calder, Kate • *Ohio State University, USA* • `calder@stat.osu.edu`

Calderhead, Ben • *University College London, UK* • b.calderhead@ucl.ac.uk
Cameron, Ewan • *Queensland University of Technology, Australia* • dr.ewan.cameron@gmail.com
Canale, Antonio • *University of Torino, Italy* • antonio.canale@unito.it
Carlin, Brad • *University of Minnesota, USA* • brad@biostat.umn.edu
Caron, Francois • *Inria Bordeaux, France* • Francois.Caron@inria.fr
Carvalho, Carlos • *University of Texas at Austin, USA* • carlos.carvalho@mccombs.utexas.edu
Carvalho, Luis • *Boston University, United States* • lecarval@math.bu.edu
Castillo, Ismael • *CNRS Paris, France* • ismael.castillo@upmc.fr
Chae, Minwoo • *Seoul National University, Korea* • demian87@snu.ac.kr
Challis, Christopher • *Duke University, USA* • christopher.challis@stat.duke.edu
Chan, Joshua • *Australian National University, Australia* • joshuacc.chan@gmail.com
Chang, Sheng-Kai • *National Taiwan University, Taiwan* • schang@ntu.edu.tw
Chaudhuri, Sanjay • *National University of Singapore, Singapore* • sanjay@stat.nus.edu.sg
Chen, Cathy W. S. • *Feng Chia University, Taiwan* • chenws@mail.fcu.edu.tw
Chen, Fang • *SAS Institute Inc., United States* • fangk.chen@sas.com
Chen, Shu-Yu • *Feng Chia University, Taiwan* • shu-yu@mail.fcu.edu.tw
Chen, Ting-Li • *Academia Sinica, Taiwan* • tlchen@stat.sinica.edu.tw
Chen, Xi • *Zhejiang University, China* • chenxi199008@gmail.com
Chib, Siddartha • *Washington University, St. Louis, USA* • chib@wustl.edu
Chikobvu, Delson • *University of Free State, South Africa* • chikobvu@ufs.ac.za
Choi, Taeryon • *Korea University, Korea* • trchoi@korea.ac.kr
Chopin, Nicolas • *CREST-ENSAE, France* • nicolas.chopin@ensae.fr
Choy, Samantha • *QUT, Australia* • dow.jaemjamrat@qui.edu.au
Christen, J. Andrés • *CIMAT, Mexico* • jac@cimat.mx
Clarke, Bertrand • *University of Miami, USA* • bclarke2@med.miami.edu
Clifford, Sam • *Queensland University of Technology, Australia* • sj.clifford@student.qut.edu.au
Clyde, Merlise • *Duke University, USA* • clyde@stat.duke.edu
Conesa, David • *University of Valencia, Spain* • david.v.conesa@uv.es
Corander, Jukka • *University of Helsinki, Finland* • jukka.corander@helsinki.fi
Cornebise, Julien • *UK* • julien@cornebise.com
Cramb, Susanna • *Cancer Council Queensland and Queensland University of Technology, Australia* • susannacramb@cancerqld.org.au
Crawford, Forrest • *University of California, Los Angeles, USA* • fcrawford@ucla.edu
Cron, Andrew • *Duke University, USA* • ajc40@stat.duke.edu
Cule, Madeleine • *University of Oxford, UK* • cule@stats.ox.ac.uk
Cumming, Jonathan • *Durham University, UK* • j.a.cumming@durham.ac.uk
Cybis, Gabriela • *University of California, Los Angeles, USA* • gcybis@uclas.edu
Czogiel, Irina • *Max Planck Institute for Molecular Genetics, Germany* • czogiel@molgen.mpg.de
Dahl, David • *Texas A&M University, USA* • dahl@stat.tamu.edu
Daniels, Michael • *University of Florida, USA* • mdaniels@stat.ufl.edu
Das, Moumita • *Indian Statistical Institute, India* • moumita.statistics@gmail.com
De Blasi, Pierpaolo • *University of Torino and Collegio Carlo Alberto, Italy* • pierpaolo.deblasi@unito.it
De Iorio, Maria • *University College London, UK* • m.deiorio@ucl.ac.uk
Demirhan, Haydar • *Hacettepe University, Turkey* • haydarde@hacettepe.edu.tr
dan Der Merwe, Abraham • *University of the Free State, South Africa* • matheeme@ufs.ac.za
Dietz, Zach • *University of Rochester, United States* • tanzy_love@urmc.rochester.edu
Donnet, Sophie • *Université Paris Dauphine, France* • donnet@ceremade.dauphine.fr
Doucet, Arnaud • *University of Oxford, UK* • doucet@stats.ox.ac.uk
Draper, David • *University of California, Santa Cruz, USA* • draper@ams.ucsc.edu
Drovandi, Christopher • *Queensland University of Technology, Australia* • c.drovandi@qut.edu.au
van Dyk, David • *University of California, Irvine, USA* • dvd@uci.edu
Ekici, Oya • *Duke University, USA* • oe8@stat.duke.edu
El Moselhy, Tarek • *Massachusetts Institute of Technology, USA* • tmoselhy@mit.edu
Elliott, Michael R. • *University of Michigan, USA* • mrelliot@isr.umich.edu
Evans, Katie • *University of Rochester, USA* • katie_evans@urmc.rochester.edu
Eves, Christopher • *University of Kent, UK* • christopher.eves4@gmail.com
Fanie, Samuel • *Richtessy trading, South Africa* • banbofabunmi@yahoo.co.uk

Farah, Marian • *MRC Biostatistics Unit, Cambridge, UK* • `marian.farah@mrc-bsu.cam.ac.uk`
Favaro, Stefano • *University of Torino, Italy* • `stefano.favaro@unito.it`
Ferkingstad, Egil • *Norwegian Computing Center, Norway* • `egil@nr.no`
Ferreira, Marco • *University of Missouri, USA* • `ferreiram@missouri.edu`
Filippi, Sarah • *Imperial College London, UK* • `s.filippi@imperial.ac.uk`
Finucane, Mariel • *Harvard University, USA* • `mfinucan@hsph.harvard.edu`
Flegal, James • *University of California, Riverside, USA* • `jflegal@ucr.edu`
Forbes, Catherine • *Monash University, Australia* • `catherine.forbes@monash.edu`
Forster, Jonathan • *University of Southampton, UK* • `J.J.Forster@soton.ac.uk`
Fort, Gersende • *LTCI, CNRS and Telecom ParisTech, France* • `gersende.fort@telecom-paristech.fr`
Fox, Colin • *University of Otago, New Zealand* • `fox@physics.otago.ac.nz`
Fox, Emily • *University of Pennsylvania, USA* • `ebfox@wharton.upenn.edu`
Franco, Chiara • *University of Essex, UK* • `cfrancb@essex.ac.uk`
de Freitas, Nando • *University of British Columbia, Canada* • `nando@cs.ubc.ca`
Fronczyk, Kassie • *University of Texas MD Anderson Cancer Center, USA* • `kassandra.fronczyk@gmail.com`
Frühwirth-Schnatter, Sylvia • *Vienna University of Economics and Business, Austria* • `sfruehwi@wu.ac.at`
Fukumizu, Kenji • *Institute of Statistical Mathematics, Japan* • `fukumizu@ism.ac.jp`
Fukushige, Mototsugu • *Osaka University, Japan* • `mfuku@econ.osaka-u.ac.jp`
Fussl, Agnes • *Johannes Kepler University of Linz, Austria* • `agnes.fussl@jku.at`
Gamble, Chris • *University of Oxford, UK* • `chris.gamble@well.ox.ac.uk`
García-Donato, Gonzalo • *University of Castilla-La Mancha, Spain* • `gonzalo.garciadonato@uclm.es`
Gelfand, Alan • *Duke University, USA* • `alan@stat.duke.edu`
George, Ed • *University of Pennsylvania, USA* • `edgeorge@wharton.upenn.edu`
Gerlach, Richard • *University of Sydney, Australia* • `richard.gerlach@sydney.edu.au`
Geweke, John • *University of Technology, Sydney, Australia* • `John.Geweke@uts.edu.au`
Ghosal, Subhashis • *North Carolina State University, USA* • `subhashis_ghoshal@ncsu.edu`
Ghosh, Joyee • *University of Iowa, USA* • `joyee-ghosh@uiowa.edu`
Ghosh, Malay • *University of Florida, USA* • `ghoshm@stat.ufl.edu`
Ghosh, Souparno • *Duke University, USA* • `sg147@stat.duke.edu`
Ghosh, Sujit • *North Carolina State University, USA* • `sujit_ghosh@ncsu.edu`
Gill, Jeff • *Washington University, St. Louis, USA* • `jgill@wustl.edu`
Glickman, Mark • *Boston University, USA* • `mg@bu.edu`
Goldstein, Michael • *University of Durham, UK* • `michael.goldstein@durham.ac.uk`
Gramacy, Robert • *University of Chicago Booth School of Business, USA* • `rbgramacy@chicagobooth.edu`
Granado, Carolina • *Venezuelan Foundation for Seismological Research, Venezuela* • `cgranado@funvisis.gob.ve`
Griffin, Jim • *University of Kent, UK* • `J.E.Griffin-28@kent.ac.uk`
Guerrera, Thiago • *Norwegian University of Science and Technology, Norway* • `thigm85@gmail.com`
Guestrin, Carlos • *University of Pennsylvania, United States* • `ebfox@wharton.upenn.edu`
Guillot, Gilles • *Technical University of Denmark, Denmark* • `gilles.b.guillot@gmail.com`
Guimond, Timothy • *University of Toronto, Canada* • `tim.guimond@utoronto.ca`
Guindani, Michele • *University of Texas MD Anderson Cancer Center, USA* • `mguindani@mdanderson.org`
Gustafson, Paul • *University of British Columbia, Canada* • `gustaf@stat.ubc.ca`
Gutman, Roee • *Brown University, USA* • `Roee_Gutman@brown.edu`
Haario, Heikki • *Lappeenranta University of Technology, Finland* • `heikki.haario@lut.fi`
Hahn, P. Richard • *University of Chicago Booth School of Business, USA* • `richard.hahn@chicagobooth.edu`
Haitovsky, Yoel • *The Hebrew University of Jerusalem, Israel* • `msyoel@mscc.huji.ac.il`
Han, Minkyu • *SNU, Korea* • `crysnal2@snu.ac.kr`
Hannig, Jan • *University of North Carolina at Chapel Hill, USA* • `jan.hannig@unc.edu`
Hans, Christopher • *Ohio State University, USA* • `hans@stat.osu.edu`
Haran, Murali • *Pennsylvania State University, USA* • `mharan@stat.psu.edu`
Hardt, Nino • *The Ohio State University, United States* • `hardt.8@osu.edu`
Hasegawa, Hikaru • *Hokkaido University, Japan* • `hasegawa@econ.hokudai.ac.jp`
Hasegawa, Shohei • *Tohoku University, Japan* • `s.hasegawa@s.tohoku.ac.jp`
Hashiguchi, Yoshihiro • *Institute of Developing Economies, Japan* • `yha4guchi@gmail.com`
Hatfield, Laura • *Harvard Medical School, USA* • `hatfield@hcp.med.harvard.edu`
Hayakawa, Yu • *Waseda University, Japan* • `yu.hayakawa@waseda.jp`

Hayashi, Kentaro • *University of Hawaii at Manoa, United States* • `hayashik@hawaii.edu`
He, Chong • *University of Missouri-Columbia, USA* • `hezh@missouri.edu`
He, Yuning • *University of California, Santa Cruz, USA* • `yuning@soe.ucsc.edu`
Healy, Liam • *Naval Research Laboratory, United States* • `liam.healy@nrl.navy.mil`
Held, Leonhard • *University of Zurich, Switzerland* • `held@ifspm.uzh.ch`
Herbei, Radu • *Ohio State University, USA* • `herbei@stat.osu.edu`
Higuchi, Tomoyuki • *Institute of Statistical Mathematics, Japan* • `higuchi@ism.ac.jp`
Hjort, Nils Lid • *University of Oslo, Norway* • `nils@math.uio.no`
Hogg, Charles • *National Institute of Standards and Technology, USA* • `charles.hogg@nist.gov`
Holmes, Chris C. • *University of Oxford, UK* • `c.holmes@stats.ox.ac.uk`
Hong, Hwanhee • *University of Minnesota, USA* • `hong0362@umn.edu`
Hori, Kazuki • *Benesse Corporation, Japan* • `horik@mail.benesse.co.jp`
Hoshina, Ibuki • *Chuo University, Japan* • `i.hoshina@gmail.com`
Hosseini, Reza • *University of Southern California, USA* • `reza1317@gmail.com`
House, Leanna • *Virginia Tech, USA* • `lhouse@vt.edu`
Hsiao, Chuhsing • *National Taiwan University, Taiwan* • `ckhsiao@ntu.edu.tw`
Huan, Xun • *Massachusetts Institute of Technology, USA* • `xunhuan@mit.edu`
Huang, Yung-Hsiang • *National Taiwan University, Taiwan* • `d95842002@ntu.edu.tw`
Huber, Mark • *Claremont McKenna College, USA* • `autotomic@gmail.com`
Huerta, Gabriel • *Indiana University, USA* • `gahuerta@indiana.edu`
Hughes, John • *University of Minnesota, USA* • `hughesj@umn.edu`
Hyodo, Masashi • *University of Tokyo, Japan* • `hyodoh_h@yahoo.co.jp`
I, Peerasak • *Thammasat University, Thailand* • `intarapaiboon@gmail.com`
Iba, Yukito • *Institute of Statistical Mathematics, Japan* • `iba@ism.ac.jp`
Ieva, Francesca • *Politecnico di Milano, Italy* • `francesca.ieva@mail.polimi.it`
Irie, Kaoru • *University of Tokyo, Japan* • `iriekao@gmail.com`
Irincheeva, Irina • *Duke University, USA* • `Irina.Irincheeva@stat.duke.edu`
Ishigaki, Tomonori • *Nanzan University, Japan* • `ishigaki@nanzan-u.ac.jp`
Ishigaki, Tsukasa • *Tohoku University, Japan* • `isgk@econ.tohoku.ac.jp`
Ishihara, Tsunehiro • *Hitotsubashi University, Japan* • `tishr@e.u-tokyo.ac.jp`
Isogawa, Naoki • *Pfizer Japan Inc., Japan* • `Naoki.Isogawa@pfizer.com`
Jääskinen, Väinö • *University of Helsinki, Finland* • `vaino.jaaskinen@helsinki.fi`
Jacob, Pierre • *CEREMADE, Université Paris Dauphine, France* • `jacob@ceremade.dauphine.fr`
Jacobi, Liana • *University of Melbourne, Australia* • `ljacobi@unimelb.edu.au`
Jacobson, Teresa • *University of California, Santa Cruz, USA* • `jacobson@soe.ucsc.edu`
James, Lancelot F. • *Hong Kong University of Science and Technology, Hong Kong* • `lancelot@ust.hk`
Jang, Gun Ho • *University of Pennsylvania, USA* • `gunjang@upenn.edu`
Jensen, Shane • *University of Pennsylvania, USA* • `stjensen@wharton.upenn.edu`
Ji, Yuan • *Northshore University Healthcare System, USA* • `koaeraser@gmail.com`
Jo, Seongil • *Seoul National University, Korea* • `joseongil@gmail.com`
Johnson, Alicia • *Macalester College, USA* • `ajohns24@macalester.edu`
Johnson, Valen E. • *University of Texas MD Anderson Cancer Center, USA* • `vejohnson@mdanderson.org`
Johnston, Ian • *Boston University, USA* • `ianj@math.bu.edu`
Jones, Edmund • *University of Bristol, UK* • `edmund.jones@bristol.ac.uk`
Jones, Galin • *University of Minnesota, USA* • `galin@umn.edu`
Jordan, Michael • *University of California, Berkeley, USA* • `jordan@cs.berkeley.edu`
Jun, Mikyoung • *Texas A&M University, USA* • `mjun@stat.tamu.edu`
Kabe, Satoshi • *University of Tsukuba, Japan* • `k0420214@sk.tsukuba.ac.jp`
Kakamu, Kazuhiko • *Chiba University, Japan* • `kakamu@le.chiba-u.ac.jp`
Kalli, Maria • *University of Kent, UK* • `M.Kalli@kent.ac.uk`
Kamatani, Kengo • *Osaka University, Japan* • `kamatani@sigmath.es.osaka-u.ac.jp`
Kanazawa, Yuichiro • *University of Tsukuba, Japan* • `kanazawa@flamenco.plala.or.jp`
Kang, Su Yun • *Queensland University of Technology, Australia* • `s7.kang@qut.edu.au`
Karabatsos, George • *University of Illinois-Chicago, USA* • `georgek@uic.edu`
Kastner, Gregor • *Vienna University of Economics and Business, Austria* • `gregor.kastner@wu.ac.at`
Kato, Kengo • *Hiroshima University, Japan* • `kkato@hiroshima-u.ac.jp`

Kato, Kentaro • *Center for Research on Educational Testing, Japan* • katoken@cret.or.jp
Katsumata, Sotaro • *Nagasaki University, Japan* • katsumat@nagasaki-u.ac.jp
Katsura, Hiroaki • *Keio University, Japan* • katsura-0225@z3.keio.jp
Kaufman, Cari • *University of California, Berkeley, USA* • cgk@stat.berkeley.edu
Kawashima, Yuichi • *Keio University, Japan* • pawn.promotes@gmail.com
Kecskemethy, Peter • *University of Oxford, UK* • peter.kecskemethy@merton.ox.ac.uk
Khatab, Khaled • *RWTH Aachen University, Germany* • khaledkhatab314@yahoo.com
Kim, Dohyun • *Seoul National University, Korea* • dhkim0211@gmail.com
Kim, Woojae • *Ohio State University, USA* • kim.1124@osu.edu
Kim, Yongdai • *Seoul National University, Korea* • ydkim0903@gmail.com
Kirk, Paul • *University of Warwick, UK* • paul.kirk06@imperial.ac.uk
Kitagawa, Toru • *University College London, UK* • t.kitagawa@ucl.ac.uk
Knapik, Bartek • *VU University Amsterdam, Netherlands* • b.t.knapik@vu.nl
Kobayashi, Genya • *Kobe University, Japan* • 081b007b@stu.kobe-u.ac.jp
Kohn, Robert • *University of New South Wales, Australia* • r.kohn@unsw.edu.au
Koizumi, Daiki • *Cyber University, Japan* • dkoizumi@m.ieice.org
Kolossiatis, Michalis • *Cyprus University of Technology, Cyprus* • Michalis.Kolossiatis@cut.ac.cy
Komaki, Fumiyasu • *University of Tokyo, Japan* • komaki@mist.i.u-tokyo.ac.jp
Kondo, Fumiyo • *University of Tsukuba, Japan* • kondo@sk.tsukuba.ac.jp
Kottas, Athanasios • *University of California, Santa Cruz, USA* • thanos@ams.ucsc.edu
Kou, Samuel • *Harvard University, USA* • kou@stat.harvard.edu
Kubota, Sho • *Keio University, Japan* • sho_kubota@a7.keio.jp
Kunihama, Tsuyoshi • *Duke University, USA* • tsuyoshi.kunihama@stat.duke.edu
Kuensch, Hans R. • *ETH Zurich, Switzerland* • kuensch@stat.math.ethz.ch
Kurose, Yuta • *University of Tokyo, Japan* • 8385906877@mail.ecc.u-tokyo.ac.jp
Lassere, Marissa • *St George Hospital, Australia* • marissa.lassere@sesiahs.health.nsw.gov.au
Launay, Tristan • *Université de Nantes, France* • tristan.launay@gmail.com
Lavigne, Aurore • *INRA, France* • aurore.lavigne@agroparistech.fr
Lazar, Nicole • *University of Georgia, USA* • nlazar@stat.uga.edu
Lee, Anthony • *University of Warwick, UK* • anthony.lee@warwick.ac.uk
Lee, Dominic • *University of Canterbury, New Zealand* • dominic.lee@canterbury.ac.nz
Lee, Herbie • *University of California, Santa Cruz, USA* • herbie@ams.ucsc.edu
Lee, Jaeyong • *Seoul National University, Korea* • leejyc@gmail.com
Lee, Ju Hee • *UT M.D. Anderson Cancer Center, United States* • jlee14@mdanderson.org
Lee, Kate • *Auckland University of Technology, New Zealand* • jelee@aut.ac.nz
Lee, Min • *University of Southampton, UK* • mcl206@soton.ac.uk
Lee, Taewook • *Hankuk University of Foreign Studies, Korea* • twlee@hufs.ac.kr
Leininger, Thomas • *Duke University, USA* • tjl13@duke.edu
Leman, Scotland • *Virginia Tech, USA* • leman@vt.edu
Lemos, Ricardo • *University of Queensland, Australia* • rtl@net.sapo.pt
Lenkoski, Alex • *Heidelberg University, Germany* • alex.lenkoski@uni-heidelberg.de
Lenormand, Maxime • *IRSTEA, France* • maxime.lenormand@irstea.fr
Leon-Gonzalez, Roberto • *National Graduate Institute for Policy Studies, Japan* • rlg@grips.ac.jp
Li, Fan • *Duke University, USA* • fli@stat.duke.edu
Li, Feng • *Stockholm University, Sweden* • feng.li@stat.su.se
Li, Guangquan • *Imperial College London, UK* • guang.li@imperial.ac.uk
Lian, Heng • *Nanyang Technological University, Singapore* • henglian@ntu.edu.sg
Liang, Li-Jung • *University of California, Los Angeles, USA* • liangl@ucla.edu
Liang, Waley • *University of California, Santa Cruz, USA* • wliang@ams.ucsc.edu
Liechty, Merrill • *Drexel University, USA* • merrill@drexel.edu
Lijoi, Antonio • *University of Pavia, Italy* • lijoi@unipv.it
Lin, Chien-Hua • *Providence University, Taiwan* • chlin@pu.edu.tw
Lin, Edward M. H. • *Feng Chia University, Taiwan* • p9431404@mail.fcu.edu.tw
Lin, Pei-Chun • *National Cheng Kung University, Taiwan* • peichunl@gmail.com
Liverani, Silvia • *Imperial College London, UK* • s.liverani@imperial.ac.uk
Lopes, Hedibert • *University of Chicago Booth School of Business, USA* • hlopes@chicagobooth.edu

Love, Tanzy • *University of Rochester, USA* • `tanzy_love@urmc.rochester.edu`
Low-Choy, Samantha • *Queensland University of Technology, Australia* • `s.lowchoy@qut.edu.au`
Lu, Jun • *American University, USA* • `lu@american.edu`
Lumley, Thomas • *University of Auckland, New Zealand* • `t.lumley@auckland.ac.nz`
Lunagomez, Simon • *Harvard University, USA* • `simon.lgz@gmail.com`
Ma, Li • *Duke University, USA* • `li.ma@duke.edu`
Macaro, Christian • *SAS Institute Inc, United States* • `Christian.Macaro@sas.com`
MacEachern, Steven • *Ohio State University, USA* • `snm@stat.osu.edu`
MacNab, Ying • *University of British Columbia, Canada* • `ying.macnab@ubc.ca`
Mahabal, Ashish • *California Institute of Technology, USA* • `aam@astro.caltech.edu`
Mai, Tiep • *Trinity College Dublin, Ireland* • `maik@tcd.ie`
Mallick, Himel • *University of Alabama at Birmingham, USA* • `himel@uab.edu`
Mandel, Kaisey • *Imperial College London, UK* • `k.mandel@imperial.ac.uk`
Manolopoulou, Ioanna • *Duke University, USA* • `im30@duke.edu`
Marin, Jean-Michel • *Université Montpellier 2, France* • `Jean-Michel.Marin@univ-montp2.fr`
Martinez, Vincent • *University of Valencia, Spain* • `vicent.martinez@uv.es`
Martínez-Ovando, Juan Carlos • *Banco de México, Mexico* • `juan.martinez@banxico.org.mx`
Martins, Thiago • *Norwegian University of Science and Technology, Norway* • `thigm85@gmail.com`
Maruyama, Yuzo • *University of Tokyo, Japan* • `maruyama@csis.u-tokyo.ac.jp`
Marzouk, Youssef M. • *Massachusetts Institute of Technology, USA* • `ymarz@mit.edu`
Masoumi, Samira • *University of Waterloo, Canada* • `smasoumi@uwaterloo.ca`
Matsuda, Kazuki • *Chuo University, Japan* • `kazuki.matsuda.chuo@gmail.com`
Matsuda, Takeshi • *Cyber University, Japan* • `takeshi_matsuda@cyber-u.ac.jp`
Mazzuchi, Thomas • *George Washington University, USA* • `mazzu@gwu.edu`
McAlinn, Kenichiro • *Keio University, Japan* • `kenmcalinn@gmail.com`
McCormick, Tyler • *University of Washington, USA* • `tylermc@u.washington.edu`
McGree, James • *Queensland University of Technology, Australia* • `james.mcgree@qut.edu.au`
McGrory, Clare • *University of Queensland, Australia* • `c.mcgrory@uq.edu.au`
Md Desa, Zairul Nor Deana • *University of Kansas and Universiti Teknologi Malaysia, USA* • `zndeana@ku.edu`
Mealli, Fabrizia • *University of Florence, Italy* • `mealli@ds.unifi.it`
Mena, Ramses • *UNAM, Mexico* • `ramses@sigma.iimas.unam.mx`
Mengersen, Kerrie • *Queensland University of Technology, Australia* • `k.mengersen@qut.edu.au`
Mi, Hsin-Chao • *National Tsing Hua University, Taiwan* • `d9624802@oz.nthu.edu.tw`
Miettinen, Minna • *University of Helsinki, Finland* • `minna.miettinen@helsinki.fi`
Mira, Antonietta • *University of Lugano, Switzerland* • `antonietta.mira@usi.ch`
Misonou, Masaki • *Keio University, Japan* • `masaki6413@gmail.com`
Miyata, Yoichi • *Takasaki City University of Economics, Japan* • `ymiyatagbt@tcue.ac.jp`
Miyawaki, Koji • *Kwansei Gakuin University, Japan* • `koji.miyawaki@09.alumni.u-tokyo.ac.jp`
Mohamed, Shakir • *University of British Columbia, Canada* • `emtiyaz@cs.ubc.ca`
Mondal, Debashis • *University of Chicago, USA* • `mondal@galton.uchicago.edu`
Montagna, Silvia • *Duke University, USA* • `silvia.montagna@duke.edu`
Mori, Joji • *Osaka University, Japan* • `j.mori@ono.co.jp`
Morimoto, Takayuki • *Kwansei Gakuin University, Japan* • `morimot@kwansei.ac.jp`
Morita, Satoshi • *Yokohama City University, Japan* • `smorita@yokohama-cu.ac.jp`
Morris, Jeffrey • *University of Texas MD Anderson Cancer Center, USA* • `jefmorris@mdanderson.org`
Mostert, Paul J. • *University of Stellenbosch, South Africa* • `pjmos@sun.ac.za`
Mueller, Peter • *University of Texas at Austin, USA* • `pmueller@math.utexas.edu`
Muir Watt, Alexis • *University of Oxford, UK* • `alexis.muirwatt@stats.ox.ac.uk`
Mukherjee, Bhramar • *University of Michigan, USA* • `bhramar@umich.edu`
Mukhopadhyay, Sabyasachi • *Indian Statistical Institute, India* • `sabstat123@gmail.com`
Mulder, Joris • *Tilburg University, Netherlands* • `j.mulder3@uvt.nl`
Murray, Jared • *Duke University, USA* • `jared.murray@stat.duke.edu`
Murray, Lawrence • *CSIRO, Australia* • `lawrence.murray@csiro.au`
Muthukumarana, Saman • *University of Manitoba, Canada* • `muthukum@cc.umanitoba.ca`
Myoung, Sungmin • *Jungwon University, South Korea* • `smmyoung@jwu.ac.kr`
Myung, Chong • *Ohio State University, United States* • `myung.1@osu.edu`

Myung, Jay • *Ohio State University, United States* • ijmyung@gmail.com
Nagata, Haruhisa • *Tokyo Institute of Technology, Japan* • nagata.h.aa@m.titech.ac.jp
Nagata, Kenji • *University of Tokyo, Japan* • nagata@mns.k.u-tokyo.ac.jp
Nagata, Shuichi • *Kwansei Gakuin University, Japan* • nagatas@kwansei.ac.jp
Nakajima, Jouchi • *Duke University, USA* • jouchi.nakajima@stat.duke.edu
Nakatsuma, Teruo • *Keio University, Japan* • nakatuma@econ.keio.ac.jp
Nakayama, Yuji • *Osaka Prefecture University, Japan* • nakayama@eco.osakafu-u.ac.jp
Nava, Consuelo R. • *University of Torino, Italy* • consuelo.nava@carloalberto.org
Nazarov, Maxim • *Bocconi University, Italy* • maxim.nazarov@phd.unibocconi.it
Ndoye, Abdoul Aziz Junior • *GREQAM, AMSE, France* • abdoul-aziz-junior.NDOYE@univ-amu.fr
Neath, Ronald • *Hunter College, City University of New York, USA* • rneath@hunter.cuny.edu
Nicholls, Geoff • *University of Oxford, UK* • nicholls@stats.ox.ac.uk
Niemi, Jarad • *Iowa State University, USA* • niemi@iastate.edu
Nieto-Barajas, Luis • *ITAM, Mexico* • lnieto@itam.mx
Nipoti, Bernardo • *University of Texas MD Anderson Cancer Center, USA* • bernardo.nipoti@carloalberto.org
Nishino, Haruhisa • *Chiba University, Japan* • nishino@le.chiba-u.ac.jp
Nishiyama, Takahiro • *Tokyo University of Science, Japan* • nishiyam@rs.kagu.tus.ac.jp
Noguchi, Hidehisa • *Tokyo University of Science, Japan* • j7410702@ed.tus.ac.jp
Noma, Hisashi • *Kyoto University School of Public Health, Japan* • nomahi-bstat@umin.ac.jp
Nomura, Shogo • *National Cancer Center Hospital East, Japan* • shnomura@east.ncc.go.jp
Norets, Andriy • *Princeton University, USA* • anorets@princeton.edu
Nott, David • *National University of Singapore, Singapore* • standj@nus.edu.sg
Nugroho, Didit Budi • *Kwansei Gakuin University, Japan* • bke76145@kwansei.ac.jp
Nyman, Henrik • *Abo Akademi University, Finland* • hennyman@abo.fi
Odani, Motoi • *Kyoto University, Japan* • m.odani@ono.co.jp
Oga, Takashi • *Chiba University, Japan* • ohgatakashi@me.com
Oh, Cheongeun • *New York University, USA* • cceohh@gmail.com
Oh, Man-Suk • *Ewha Womans University, Korea* • msoh@ewha.ac.kr
Ohashi, Yasuo • *University of Tokyo, Japan* • ohashi@epistat.m.u-tokyo.ac.jp
Ohwada, Shoichi • *Daiichi Sankyo Co., Ltd.* • Japan
Okada, Kensuke • *Senshu University, Japan* • ken@psy.senshu-u.ac.jp
Omori, Yasuhiro • *University of Tokyo, Japan* • omori@e.u-tokyo.ac.jp
Orbanz, Peter • *Columbia University, USA* • po245@cam.ac.uk
Oya, Kosuke • *Osaka University, Japan* • oya@econ.osaka-u.ac.jp
Paiva, Thais • *Duke University, USA* • tvp@stat.duke.edu
Palacios, Julia • *University of Washington, USA* • jpalacio@u.washington.edu
Panagiotelis, Anastasios • *Monash University, Australia* • anastasios.panagiotelis@monash.edu
Papageorgiou, Georgios • *Imperial College London, UK* • g.papageorgiou@imperial.ac.uk
Parent, Eric • *Agroparistech, France* • eric.parent@agroparistech.fr
Parry, Matthew • *University of Otago, New Zealand* • mparry@maths.otago.ac.nz
Pelenis, Justinas • *Institute for Advanced Studies, Vienna, Austria* • pelenis@ihs.ac.at
Peluso, Stefano • *Università della Svizzera Italiana, Switzerland* • stefano.peluso@usi.ch
Peng, Chien-Yu • *Academia Sinica, Taiwan* • chienyu@stat.sinica.edu.tw
Pensar, Johan • *Åbo Akademi University, Finland* • jopensar@abo.fi
Perron, Francois • *University of Montreal, Canada* • perronf@dms.umontreal.ca
Pessia, Alberto • *University of Helsinki, Finland* • alberto.pessia@helsinki.fi
Petrone, Sonia • *Bocconi University, Italy* • sonia.petrone@unibocconi.it
Pettitt, Anthony • *Queensland University of Technology, Australia* • a.pettitt@qut.edu.au
Pham, Thong • *Tokyo Institute of Technology, Japan* • thongphamthe@gmail.com
Philip, Ninan • *St. Thomas College, Kozhencheri, India* • sajeethphilip@gmail.com
Philippe, Anne • *UniversitÂŐ de Nantes, France* • anne.philippe@univ-nantes.fr
Pibouleau, Leslie • *INSERM UMR 717 and Université Paris 7, France* • leslie.pibouleau@univ-paris-diderot.fr
Pillai, Natesh • *Harvard University, USA* • pillai@fas.harvard.edu
Pirinen, Matti • *University of Oxford, UK* • matti.pirinen@well.ox.ac.uk
Poirier, Dale • *University of California, Irvine, USA* • dpoirier@uci.edu
Polasek, Wolfgang • *Institute for Advanced Studies, Vienna, Austria* • polasek@ihs.ac.at

Posekany, Alexandra • *Vienna University of Economics and Business, Austria* • alexandra.posekany@gmail.com
Powell, Benedict • *Durham University, UK* • benedict.powell@durham.ac.uk
Prado, Raquel • *University of California, Santa Cruz, USA* • raquel@ams.ucsc.edu
Prangle, Dennis • *Lancaster University, UK* • d.prangle@lancaster.ac.uk
Pruenster, Igor • *University of Torino & Collegio Carlo Alberto, Italy* • igor@econ.unito.it
Quinn, Anthony • *Trinity College Dublin, Ireland* • aquinn@tcd.ie
Quintana, Fernando • *Pontificia Universidad Católica de Chile, Chile* • quintana@mat.puc.cl
Quiroz, Matias • *Sveriges Riksbank, Sweden* • quiroz.matias@gmail.com
Rao, Vinayak • *University College London, UK* • vrao@gatsby.ucl.ac.uk
Ratman, Oliver • *Duke University, USA* • oliver.ratmann@duke.edu
Raubenheimer, Lizanne • *Rhodes University, South Africa* • L.Raubenheimer@ru.ac.za
Reyes, Perla • *University of California, Santa Cruz, USA* • perla@soe.ucsc.edu
Rice, Kenneth • *University of Washington, USA* • kenrice@u.washington.edu
Richardson, Sylvia • *Imperial College London, UK* • sylvia.richardson@imperial.ac.uk
Riebler, Andrea • *University of Zurich, Switzerland* • andrea.riebler@ifspm.uzh.ch
Rigat, Fabio • *University of Warwick, UK* • f.rigat@warwick.ac.uk
Rivoirard, Vincent • *Université Paris Dauphine, France* • Vincent.Rivoirard@dauphine.fr
Robert, Christian • *Université Paris Dauphine, France* • xian@ceremade.dauphine.fr
Rodrigues, Josemar • *Federal University of São Carlos, Brazil* • vjosemar@ufscar.br
Rodriguez, Abel • *University of California, Santa Cruz, USA* • abel@soe.ucsc.edu
Rossell, David • *IRB Barcelona, Spain* • david.rossell@irbbarcelona.org
Rousseau, Judith • *ENSAE-CREST and Université Paris Dauphine, France* • rousseau@ceremade.dauphine.fr
Roy, Vivekananda • *Iowa State University, USA* • vroy@iastate.edu
Ruggeri, Fabrizio • *CNR-IMATI, Italy* • fabrizio@mi.imati.cnr.it
Ryan, Elizabeth • *Queensland University of Technology, Australia* • elizabeth.ryan@student.qut.edu.au
Ryder, Robin • *CEREMADE, Université Paris Dauphine, France* • ryder@ceremade.dauphine.fr
Rydevik, Gustaf • *University of York, UK* • gustaf.rydevik@sac.ac.uk
Saar, Enn • *Tartu Observatory, Estonia* • saar@aai.ee
Sabanés Bové, Daniel • *University of Zurich, Switzerland* • daniel.sabanesbove@ifspm.uzh.ch
Saine, Isatu • *Global Prevention Alliance, Gabon* • preventionalliance@aol.com
Sakurai, Gen • *National Institute for Agro-Environmental Sciences, Japan* • sakuraigen@gmail.com
Sakurai, Hirohito • *National Center for University Entrance Examinations, Japan* • sakurai@rd.dnc.ac.jp
Salakhutdinov, Russ • *University of Toronto, Canada* • rsalakhu@utstat.toronto.edu
Salimans, Tim • *Erasmus University Rotterdam, Netherlands* • salimans@ese.eur.nl
Salomond, Jean Bernard • *ENSAE-CREST and Université Paris Dauphine, France* • jean.bernard.salomond@ensae.fr
Sang, Huiyan • *Texas A&M University, USA* • huiyan.sang@gmail.com
Sanso, Bruno • *University of California, Santa Cruz, USA* • bruno@ams.ucsc.edu
Saria, Suchi • *Harvard University and Johns Hopkins University, USA* • suchi.saria@gmail.com
Sato, Seisho • *Institute of Statistical Mathematics, Japan* • sato@ism.ac.jp
Savelieva, Marina • *Novartis Pharma AG, Switzerland* • marina.savelieva@novartis.com
Schäfer, Christian • *ENSAE-CREST, CEREMADE and Université Paris Dauphine, France* • christian.schafer@ensae.fr
Scheipl, Fabian • *Ludwig-Maximilians-Universität München, Germany* • fabian.scheipl@stat.uni-muenchen.de
Schmidler, Scott • *Duke University, USA* • schmidler@stat.duke.edu
Schmitz, Susanne • *Trinity College Dublin, Ireland* • schmitzs@tcd.ie
Scott, James • *University of Texas at Austin, USA* • james.scott@mccombs.utexas.edu
Scott, Steven • *Google, USA* • stevescott@google.com
Seker, Sadi • *Istanbul University, Turkey* • isba@shedai.net
Serra, Paulo • *Technical University of Eindhoven, Netherlands* • serra@eurandom.tue.nl
Shigemasu, Kazuo • *Teikyo University, Japan* • kshige@ca2.so-net.ne.jp
Shimodaira, Hidetoshi • *Tokyo Institute of Technology, Japan* • shimo@is.titech.ac.jp
Shirota, Shinichiro • *University of Tokyo, Japan* • ee116010@mail.ecc.u-tokyo.ac.jp
Shubin, Mikhail • *University of Helsinki, Finland* • 2pi360@gmail.com
Sigauke, Caston • *University of Limpopo, South Africa* • csigauke@gmail.com
Sisson, Scott • *University of New South Wales, Australia* • Scott.Sisson@unsw.edu.au
Smith, Isabelle • *Laboratoire des Sciences du Climat et de l'Environnement, France* • isabelle.smith@lsce.ipsl.fr
Smith, Michael • *Melbourne Business School, Australia* • mike.smith@mbs.edu

Smith, Theresa • *University of Washington, USA* • tsmith7@uw.edu
Solgi, Reza • *University of Lugano, Switzerland* • reza.solgi@usi.ch
Solonen, Antti • *Lappeenranta University of Technology, Finland* • antti.solonen@gmail.com
Solvang, Hiroko • *Oslo University Hospital, Norway* • hsolvang@rr-research.no
Sono, Shintaro • *Hokkaido University, Japan* • sono@econ.hokudai.ac.jp
Sonoda, Michio • *Cyber University, Japan* • michio_sonoda@cyber-u.ac.jp
Sorrentino, Alberto • *University of Warwick, UK* • A.Sorrentino@warwick.ac.uk
Soyer, Refik • *George Washington University, USA* • soyer@gwu.edu
Sporysheva, Polina • *Bocconi University, Italy* • polina.sporysheva@phd.unibocconi.it
Steorts, Rebecca • *University of Florida, USA* • rsteorts@stat.ufl.edu
Stephens, David • *Mc Gill University, Canada* • d.stephens@math.mcgill.ca
Stingo, Francesco • *University of Texas MD Anderson Cancer Center, USA* • fra.stingo@gmail.com
Stoica, Radu S. • *Université Lille 1, France* • stoica@math.univ-lille1.fr
Strachan, Rodney • *Australian National University, Australia* • rodney.strachan@anu.edu.au
Strickland, Christopher • *Queensland University of Technology, Australia* • christopher.strickland@qut.edu.au
Stroud, Jonathan • *George Washington University, USA* • stroud@gwu.edu
Stumpf, Michael • *Imperial College London, UK* • m.stumpf@imperial.ac.uk
Suchard, Marc • *University of California, Los Angeles, USA* • msuchard@ucla.edu
Sugawara, Shinya • *University of Tokyo, Japan* • sugahra@gmail.com
Sun, Dongchu • *University of Missouri-Columbia, USA* • sund@missouri.edu
Sunada, Hiroshi • *Yamagata University, Japan* • sunada@human.kj.yamagata-u.ac.jp
Szabo, Botond • *Technical University of Eindhoven, Netherlands* • b.szabo@tue.nl
Taddy, Matt • *University of Chicago Booth School of Business, USA* • taddy@chicagobooth.edu
Tadesse, Mahlet • *Georgetown University, USA* • mgt26@georgetown.edu
Takada, Ryota • *Keio University, Japan* • ryota1987@gmail.com
Takanashi, Kosaku • *Keio University, Japan* • kousaku.takanashi@gmail.com
Takeda, Kentaro • *Astellas Pharma Inc., Japan* • kentaro.takeda@astellas.com
Tanaka, Fuyuhiko • *University of Tokyo, Japan* • ftanaka@stat.t.u-tokyo.ac.jp
Tchernis, Rusty • *Georgia State University, USA* • rtchernis@gsu.edu
Teh, Yee Whye • *University College London, UK* • ywteh@gatsby.ucl.ac.uk
Telesca, Donatello • *University of California, Los Angeles, USA* • dtelesca@ucla.edu
Teramukai, Satoshi • *Kyoto University Hospital, Japan* • steramu@kuhp.kyoto-u.ac.jp
Terres, Maria • *Duke University, USA* • mat40@stat.duke.edu
Terui, Nobuhiko • *Tohoku University, Japan* • terui@econ.tohoku.ac.jp
Thomas, Andrew • *Carnegie Mellon University, USA* • act@acthomas.ca
Thorne, Thomas • *Imperial College London, UK* • thomas.thorne@imperial.ac.uk
Todeschini, Adrien • *Inria Bordeaux, France* • adrien.todeschini@inria.fr
Todo, Naoya • *University of Tokyo, Japan* • 5949488744@mail.ecc.u-tokyo.ac.jp
Tokdar, Surya • *Duke University, USA* • tokdar@stat.duke.edu
Tran, Minh Ngoc • *University of New South Wales, Australia* • ngocmath@gmail.com
Trippa, Lorenzo • *Harvard University, USA* • ltrippa@jimmy.harvard.edu
Trotta, Roberto • *Imperial College London, UK* • r.trotta@imperial.ac.uk
Tsepletidou, Sofia • *Université Paris Dauphine, France* • tsepletidou@ceremade.dauphine.fr
Tsurumi, Hiroki • *Rutgers University , USA* • tsurumi@econ.rutgers.edu
van der Vaart, Aad • *University of Leiden, Netherlands* • avdvaart@math.leidenuniv.nl
Vallejos, Catalina • *University of Warwick, UK* • C.A.Vallejos-Meneses@warwick.ac.uk
Vanem, Erik • *University of Oslo, Norway* • erikvan@math.uio.no
Vasquez, Raquel • *Venezuelan Foundation for Seismological Research, Venezuela* • rvasquez@funvisis.gob.ve
Vernon, Ian • *Durham University, UK* • i.r.vernon@durham.ac.uk
Vihola, Matti • *University of Jyväskylä, Finland* • matti.s.vihola@jyu.fi
Villa, Cristiano • *University of Kent, UK* • cv60@kent.ac.uk
Villani, Mattias • *Linköping University, Sweden* • mattias.villani@liu.se
Vukcevic, Damjan • *Murdoch Childrens Research Institute, Australia* • damjan.vukcevic@st-hughs.oxon.org
Wade, Sara • *Bocconi University, Italy* • sara.wade@phd.unibocconi.it
Wagner, Helga • *University of Linz, Austria* • helga.wagner@jku.at
Wago, Hajime • *Kyoto Sangyo University, Japan* • wago.hajime@gmail.com

Walker, Stephen G. • *University of Kent, UK* • S.G.Walker@kent.ac.uk
Walsh, Cathal • *Trinity College Dublin, Ireland* • walshc@tcd.ie
Wang, Charlotte • *National Taiwan University, Taiwan* • d99849002@ntu.edu.tw
Wang, Chenguang • *Johns Hopkins University, USA* • cwang68@jhmi.edu
Wang, Fangpo • *Duke University, USA* • fw19@stat.duke.edu
Wang, Hao • *University of South Carolina, USA* • haowang@sc.edu
Wang, Shi-Heng • *National Taiwan University, Taiwan* • d96842004@ntu.edu.tw
Wang, Wenyi • *University of Texas MD Anderson Cancer Center, USA* • wwang7@mdanderson
Wang, Xue • *University of Kent, UK* • Xue.Wang@kent.ac.uk
Wang, Ziyu • *University of British Columbia, Canada* • ziyuw@cs.ubc.ca
Watanabe, Hiroshi • *Benesse Corporation, Japan* • horik@mail.benesse.co.jp
Watanabe, Toshiaki • *Hitotsubashi University, Japan* • watanabe@ier.hit-u.ac.jp
Wegmann, Bertil • *Linköping University, Sweden* • bertil.wegmann@ida.liu.se
Weissbach, Rafael • *University of Rostock, Germany* • rafael.weissbach@uni-rostock.de
Wen, Chi-Chung • *Tamkang University, Taiwan* • ccwen@mail.tku.edu.tw
West, Mike • *Duke University, USA* • mw@stat.duke.edu
White, Nicole • *Queensland University of Technology, Australia* • nm.white@qut.edu.au
Wichitaksorn, Nuttanan • *University of Sydney, Australia* • nwic9927@uni.sydney.edu.au
Williamson, Daniel • *Durham University, UK* • daniel.williamson@durham.ac.uk
Williamson, Sinead • *Carnegie Mellon University, USA* • sinead@cs.cmu.edu
Wilson, Amy • *University of Edinburgh, UK* • A.L.Wilson-4@sms.ed.ac.uk
Wilson, Simon P. • *Trinity College Dublin, Ireland* • swilson@tcd.ie
Windle, Jesse • *University of Texas at Austin, USA* • jwindle@ices.utexas.edu
Wood, Frank • *Columbia University, USA* • fwood@stat.columbia.edu
Woodard, Dawn • *Cornell University, USA* • woodard@cornell.edu
Wozniak, Tomasz • *University of Melbourne, Australia* • wozniak.tom@gmail.com
Wu, Chieh-Hsi • *University of Auckland, New Zealand* • cwu080@aucklanduni.ac.nz
Wu, Yuefeng • *University of California, Santa Cruz, USA* • yuefeng@soe.ucsc.edu
Wu, Yuh-Jenn • *Chung Yuan Christian University, Taiwan* • yuhjenn@cycu.edu.tw
Xiong, Jie • *University of Helsinki, Finland* • jie.xiong@helsinki.fi
Xu, Xinyi • *Ohio State University, USA* • xinyi@stat.osu.edu
Yajima, Masanao • *University of California, Los Angeles, USA* • yajima@ucla.edu
Yamamoto, Michio • *Osaka University, Japan* • myamamoto@sigmath.es.osaka-u.ac.jp
Yamamoto, Yohei • *University of Alberta, Canada* • yohei.yamamoto@ualberta.ca
Yan, Guohua • *University of New Brunswick, Canada* • gyan@unb.ca
Yan, Jin • *University of Maryland, College Park, USA* • yjconnie@umd.edu
Yanagimoto, Takemi • *Chuo University, Japan* • yanagmt@indsys.chuo-u.ac.jp
Yang, Hongxia • *IBM Watson Research, USA* • yangho@us.ibm.com
Yi, Seongbaek • *Pukyong National University, Korea* • sbyi0108@gmail.com
Yonemoto, Naohiro • *National Center of Neurology and Psychiatry, Japan* • nyonemoto@gmail.com
Yoon, Jae Ho • *POSCO Research Institute, Korea* • jhyoon@posri.re.kr
Yoshida, Ryo • *Research Organization of Information and Systems, Institute of Statistical Mathematics, Japan* • yoshidar@ism.ac.jp
Yoshida, Yasushi • *Chiba University of Commerce, Japan* • yyoshida@mug.biglobe.ne.jp
Yoshizoe, Yasuto • *Aoyama Gakuin University, Japan* • yasuto_yoshizoe@post.harvard.edu
Yunoue, Hideo • *Chiba University, Japan* • yunoue@cuc.ac.jp
Zamzuri, Zamira • *Macquarie University, Australia* • nalagnayas@yahoo.com
van Zanten, Harry • *Eindhoven University of Technology, Netherlands* • j.h.v.zanten@tue.nl
Zarepour, Mahmoud • *University of Ottawa, Canada* • zarepour@uottawa.ca
Zhou, Jing • *University of North Carolina at Chapel Hill, USA* • jingzhou@live.unc.edu