A MESSAGE FROM THE PRESIDENT

by Mike West
ISBA President
mw@stat.duke.edu

Belated New Year greetings for 2009 to all ISBA members, and to any readers whose memberships are not yet paid up for 2009! In a time of substantial economic uncertainty for us all, I am pleased to report that ISBA is in good shape financially as well as scientifically. Thanks and kudos go to recent and past Board members, including especially my predecessor and now Past President Christian Robert, for thoughtful leadership and careful stewardship. Our revenues in 2008 were buoyed by the increased membership associated with the most successful World Meeting in Australia, and by the beginning impact of both the new lifetime ISBA membership and the promotion of ISBA membership through IMS.

In its 17 years since inception, I believe ISBA has done much to advance the appreciation of Bayesian statistical science, especially in terms of international and interdisciplinary outreach. Beyond the intellectual and socio-professional community ISBA represents, and that many of us value and promote, some of the more immediate, tangible activities are conference organisation and support, the Bayesian Analysis journal, and the administration of Bayesian awards. Success and maturation over the longer term requires planning and development to ensure the professional and financial vitality of these activities; this, in turn, *Continue in page* .

A MESSAGE FROM THE EDITOR

by Raphael Gottardo
raphael.gottardo@ircm.qc.ca

I would like to start this new issue of the bulletin by welcoming our new President, Mike West. As you will see in his message, our new President has a long list of great news to share with you. He also has many ideas on how to improve ISBA and make it even better for its members. Speaking of Presidents, I would also like to thank our Past President, Christian Robert, for all his great work and contributions to ISBA. Among writing many books, papers, being editor of JRSSB, Christian still finds time to maintain a very entertaining blog. So if you miss Christian's contribution to the bulletin make sure you check out his blog at http://xianblog. wordpress.com/ where you will find some interesting discussion. In this issue of the bulletin, you will find an interesting application section written by Peter Müller (ISBA President Elect), an interview of Sonia Petrone, a nice review on ABC methods and an update from our BA editor-in-chief, Brad Carlin. I hope you will enjoy this issue, and as usual, please send me you comments/concerns/contributions! ▲

WORDS FROM THE PRESIDENT, *Continued from page 1.* will rely on sustained efforts to ensure a robust membership and expanded appreciation of the benefits, and value, of ISBA membership. ISBA spends sparely, with most of the operating expenses associated with the Bayesian prizes and awards that ISBA co-sponsors and manages on behalf of the profession.

I am pleased to report that, during the first three months of 2009, the ISBA Board has been active – on your behalf – in discussions of these general questions, and in particular on matters of membership development and benefits/services. Here are some specifics on these discussions and initial outcomes, as well as news and information about ISBA meetings.

Membership: In January 2009, the Board approved the establishment of an ad-hoc *Committee on Membership* that is already up-and-running. I am pleased to report the immediate and enthusiastic agreement of each of the Board members invited to serve – and work! – on this important committee. They are David Dunson, Sylvia Frühwirth-Schnatter, Lurdes Inoue, Brunero Liseo and David van Dyk, along with an able and vigourous chairman – and long-term significant volunteer contributor to ISBA – Hedibert Lopes. Full details are available from the committee page at the ISBA web site. In brief, the committee will plan, propose and develop *strategies for consolidating and expanding the membership* and for *expanding the appreciation of "what ISBA does for the members"*. Some of the initial topics under discussion are international and disciplinary outreach, coordinated communication with academic departments, targeted PR at national and international meetings, short-courses, more systematic and membership-focused email and web communications, and others. ISBA members will receive summary reports from the committee in the Bulletin. Please feel free – and be encouraged – to email Hedibert or any of the committee members to help move these important discussions along, and consider ways in which you can contribute to the activities that are emerging.

New Award: I am very pleased to report to you that, by Board resolution, ISBA has just established a new *Lifetime Members Junior Researcher Award*. Beginning with the 2010 meeting in Spain, this award will be made to junior researchers and/or students to provide support for participation in ISBA World Meetings. We expect to announce the process for this award later this year when ISBA2010 is announced.

The *Lifetime Members Junior Researcher Award* expands the portfolio of ISBA awards – a key member service – and will be funded, in part, from revenue generated on Lifetime Member dues. The award thus recognises life members as significant contributors to, and supporters of, ISBA and our interests in assisting new/junior members. We hope and expect that this will contribute to growing the ISBA core membership by encouraging colleagues to consider life membership. In this regard, I am delighted to add that almost all of the *ISBA Past Presidents* and *current elected members of the ISBA Board* enthusiastically agreed to sign-up for life membership to support this initiative and award.

Meetings/2009: This is an "odd" year – we are between World Meetings. We have several ISBA co-sponsored or endorsed meetings in 2009 – the recent conference on Bayesian Biostatistics in Houston, and two meetings in Italy in June: the 6th Workshop on Bayesian Inference In Stochastic Processes (BISP6) in Bressanone, and the 7th Workshop on Bayesian Nonparametrics in Turino. There is no place like Italy in June for Bayesians! There are other meetings coming up that are or may be ISBA sponsored: make sure to visit the meetings page at the ISBA web site.

Meetings/2010: Mark your calendars for early June 2010! At the invitation of the Valencia Committee, ISBA has now agreed to hold ISBA10 – the 10th ISBA World Meeting – in conjunction with the 9th Valencia International Meeting on Bayesian Statistics, in Spain next year. The joint meeting will be a full 7-day conference during the first ten days of June 2010; we are currently finalising discussions about precise timing and location. ISBA10 will involve talks, posters, tutorials, the ISBA congress/meeting of members (and parties), with publication in our journal Bayesian Analysis. ISBA10 is in the hands of the organising committee and its subcommittees: members are Cathy Chen, Andrés Christen, Simon Godsill, Aparna Huzurbazar, Xiao-Li Meng, Kerrie Mengersen, Paul Mostert, Peter Müller, Sonia Petrone, Gareth Roberts, Alex Schmidt, led by our able and and vigourous Program Council Chairman, Herbie Lee. This committee will man-

age all details – logistics, finance and science – and may solicit your help as they proceed.

Relationships: Among other topics on the radar screen are relationships with other professional societies. Our new association with IMS has established co-membership sign-up opportunities and IMS provision of PR for the ISBA journal Bayesian Analysis (see also the news from BA Editor-in Chief, Brad Carlin, in this Bulletin). In the last couple of months, we have had discussions with the leadership of the ASA Section on Bayesian Statistical Science (SBSS) about common interests. You will know that we already collaborate with SBSS on co-sponsoring awards and meetings. As a result of these recent discussions, ISBA has now been invited to have a "presence" at the annual SBSS business meeting at the Joint Statistical Meeting this August in Washington DC, and SBSS has agreed to take an active hand in promoting ISBA membership among its members. We are also in initial discussions with the leadership of the International Chinese Statistical Association (ICSA) about opportunities for co-operative activities. I hope to be able to report on progress in these discussions in future Bulletins.

As a final note on Board matters, we are looking into questions of the ISBA "office" (i.e., the ISBA web site!) and IT matters of membership management, email services/communications, IT support for conference and workshop organisation, and so forth. As these discussions evolve in the near future, we will need your input!

2009 has kicked-off actively for the ISBA Board, and I hope to see some of the initiatives move along to fruition to the benefit of all current and future members. You can help: on all matters of ISBA societal business, become engaged! Board members – and the various ISBA members that volunteer time and effort to run the society, its publications, its awards and its meetings for the benefit of all – value input and suggestions from the membership. So, feel free and be encouraged to provide your thoughts and suggestions to help to grow ISBA – your Society – to a central, visible position within the broad field of statistical science. ▲

BAYESIAN ANALYSIS - A MESSAGE FROM THE EDITOR

Update from BA
by Brad Carlin
BA **Editor-in-Chief**
carli002@umn.edu

Hello everyone! ISBA President Mike West has asked me to write a few lines about what's happening with Bayesian Analysis, and while I do not have any particularly earth-shattering news at the moment I am happy to provide a brief update of what is going on.

Editorially, the journal remains strong, and is relying more and more on author-initiated submissions. However, we continue to recruit papers related to ISBA-sponsored conferences, in order to maintain a healthy overall number of submissions. I encourage submissions from all authors giving talks connected to any ISBA-sponsored meeting or session at the meeting of another society (JSM, RSS, IBS, etc.). Related to this, BA is looking forward to handling all contributed papers from the next Valencia conference (June 2010), as well as both the invited and contributed papers from the next Case Studies in Bayesian Statistics workshop at Carnegie Mellon University (October 2009).

The journal continues to be an all-volunteer effort, and I am extremely grateful to all the referees, AEs, editors, and staffers who continue to produce a journal of which the society can be proud. Persons joining the board in the past 12 months include Sudipto Banerjee (AE, University of Minnesota), Kate Cowles (editor, University of Iowa), Dennis Cox (AE, Rice University), Tim Hanson (AE, University of Minnesota), Christina Kendziorski (AE, University of Wisconsin), Thanasis Kottas (AE, University of California Santa Cruz), and Amy Racine (AE, Novartis Pharma Switzerland). Of the continuing editors and AEs, all are outstanding but worthy of special mention is System Managing Editor Pantelis Vlachos, who continues to provide crucial production and technical support at no cost to the society despite being on sabbatical in Europe this semester!

Speaking of costs, we do have a few, and we continue to look for ways to defray them. Our

joint arrangement with IMS to provide BA with a small revenue stream via print subscriptions and page charges has yet to produce any revenue, but we remain hopeful for 2009. I do hope all successful BA authors will say "yes" to paying the voluntary page charges; you will be doing ISBA a tremendous favor and helping a cause that I'm sure all readers of this column agree is worthy of support!

Regarding indexing, we have heard your cries for the journal to establish a formal impact factor, and as such have now completed our submission to Thomson-Reuters for inclusion in their extensive and popular database. The company required us to submit at least three consecutive current issues "in real time" to their evaluation, a step we have just completed (we submitted September and December 2008 as they came out, and have just now submitted March 2009). We hope for a favorable decision from them soon; I will keep you posted.

Finally (and again relatedly) I hope you will enjoy reading the March 2009 issue, which includes a featured discussion article entitled, "Hi-erarchical Model Building, Fitting, and Checking: A Behind-the-Scenes Look at a Bayesian Analysis of Arsenic Exposure Pathways," by Peter F. Craigmile, Catherine A. Calder, Hongfei Li, Rajib Paul, and Noel Cressie. The article offers a careful look at a Bayesian hierarchical analysis of pathways of exposure to arsenic (a toxic heavy metal) using the Phase I National Human Exposure Assessment Survey carried out in Arizona. The analysis combines individual-level personal exposure measurements (biomarker and environmental media) with water, soil, and air observations from the ambient environment. The article features invited discussions by four persons well-acquainted with Bayesian research on adverse environmental exposures and related health outcomes (Chris Barr, Francesca Dominici, David Dunson, and Alexandra Schmidt), and rejoinder by the authors.

The current issue (featuring this discussion paper and 5 others) is freely available at
http://ba.stat.cmu.edu/vol04is01.php
It continues to be an honor to serve as editor-in-chief; thanks again for the opportunity!

INTERVIEW

SONIA PETRONE

by Donatello Telesca
donatello.telesca@gmail.com

Sonia Petrone is associate professor of Statistics at Bocconi University (Milano), where she is also vice-director of the Ph. D. program in Statistics. Previously she was assistant professor of Statistics at the University of Pavia. Her current research interests focus on Bayesian nonparametric inference, mixtures and latent variables models, and dynamic regression models. She is also working on state-space models and about to publish a book on Bayesian time series analysis with dynamic linear models in R.

Sonia visited several statistical departments in the US, the Russian Academy of Science in Moscow and the ISI in Delhi. She was elected in the ISBA Board last year and previously served in the Board from 2002 to 2004. She organized and served in the scientific committees of several meetings, including the series of workshops on Bayesian Inference for Stochastic Processes (BISP) and Bayesian Nonparametrics.

I had the pleasure of meeting Sonia during one of her visits to M.D. Anderson and she agreed to answer some of our questions.

1. *When you name Sonia Petrone, the first things that come to mind are NP Bayes and Bernstein polynomials. In words understandable by a non statistician (or so that I can understand), what is a Bernstein polynomial and how is it useful in non parametric constructions?*

Proprio come prima domanda? *(As the first question?)* Bernstein polynomials are used in analysis to give a simple, constructive prove of Weirstrass theorem (any continuous function on [0,1] can be uniformly approximated by polynomials). You can think of them as polynomials whose terms are rearranged so that the coefficients are explicitly related to the function to be approximated. In BNP inference on a bounded function, a simple idea is to model the function as a polynomial, with potentially infinite order, and assign a prior to the coefficients of the polynomial expansion.

Bernstein polynomials are simple and their co-efficients are easily interpretable, so assigning a prior on them should be easier. When I studied them, years ago, I wanted a simple way to have a "nonparametric" prior on a continuous d.f., since I found that the discreteness of the Dirichlet process had some shortcoming, e.g. for partially exchangeable data.

2. *How did you get originally interested in NP Bayes problems?*

In my undergraduate thesis - many years ago! I studied empirical Bayes methods, and the aim of my thesis was to get to Bayes-empirical-Bayes: estimating a mixing distribution nonparametrically, using a Dirichlet process prior. I struggled on the paper of Antoniak (1974), but I found it very rich in content. In my thesis, I did a Bayes-empirical-Bayes regression model with random effects, but I could manage only moderate sample size: it was before Gibbs sampling!

3. *Is there a book or a particular paper that at some point made you think: "I like this. I want to work on this stuff"?*

Well: de Finetti's work! Again, when I was an undergraduate student in Economics: I couldn't stand all these models assuming perfect information and rational agents ... Then we studied Keynes: uncertainty, animal spirits, subjective probability! And we began to read de Finetti: and there was so much to learn! Btw: you know that de Finetti somehow had a Nobel prize in Economics ...

4. *What are you currently working on?*

I'm working on BNP for multivariate and functional data, and on multivariate urn schemes: there have been many extensions of the Dirichlet process recently, but I would like to have something that is fairly simple to compute. I am also working on state-space models: they have been used for Bayesian time series analysis but there are still unsolved problems, for instance with multivariate time series, and in control.

5. *Is it true that if you are Italian, then you are bound to be Bayesian? What do you think is the pulse on Bayesian Statistics in Italy?*

I think this is a misleading belief! In fact, Bayesian statistics in Italy was a quite small minority. Perhaps, small but active! But it's true that we had good schools - after de Finetti, of course - in Trieste, Roma, Milano, Pavia, Cagliari... We had absolutely great "maestri", like Eugenio Regazzini, and then young people grew up. I think that now Bayesian statistics in Italy is much more popular than just a few years ago - but after all, this seems true all over the world.

6. *As a young (at least at heart) Italian statistician, I sometimes look with apprehension at what is going on in our country. What is your perspective on the state and future of academic research in the bel Paese?*

... ohi ohi ohi! just sometimes? Academic life in Italy has never been easy, little money, but especially a lot of uncertainty. Any new government starts its own "reform" of the University System, and the current one ... We used to have excellent undergraduate programs, that however have been spoilt by the recent "3+2" reform. What we needed, instead, was to have more resources in the PhD programs; but we are still exporting brilliant minds. Sometimes, I feel it is difficult to encourage a young student to start an academic career in Italy. Nevertheless, there are excellent research groups in Italy – and this is still a wonderful job.

7. *Who are the people, you think had the most influence on your career?*

Most influence? Several people - ok, I just mention three. Eugenio Regazzini, a great maestro for me and many of us; Pietro Muliere, with his curiosity and intelligence; and I am grateful to Persi Diaconis: for his advise and his criticism, and for giving me enthusiasm for research.. for this amazing privilege we have to be part - even if very small, in my case! - of the great adventure of knowledge! I learned a lot from him: talking with him always opens my mind.

8. *You had the chance to visit a few departments around the world. Do you have any fun story from your travels?*

Well, travelling is another privilege we have! Let's see, instead of a fun story I will tell you a "traveller's rule": if you are in a new country, with short time to go around (you are there to work!), what would you visit to quickly get a idea of how people live there?... My rule: go

someplace where people get food!

Many thanks to Sonia for her contribution!

Following are are some references to the books and articles mentioned in Sonia's interview:

○ Petris, G., Petrone, S., Campagnoli, P. (2009). Dynamic linear models with R. (in press) Springer-Verlag, N.Y.

○ Petrone, S., Guindani, M. and Gelfand, A.E. (2009). Hybrid Dirichlet mixture models for functional data. Journal of the Royal Statistical Society, Ser. B, to appear.

○ Trippa, L., Bulla, P. and Petrone, S. (2009). Extended Bernstein Prior via Reinforced Urn Processes. Annals of the Institute of Statistical Mathematics, to appear.

○ Petrone, S. and Veronese, P. (2008). Feller operators and mixture priors in Bayesian non-parametrics. Statistica Sinica, to appear.

○ Gelfand, A.E., Guindani, M. and Petrone, S. (2007). Bayesian nonparametric modelling for spatial data using Dirichlet processes (with discussion). In Bayesian Statistics 8, J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West (Eds). Oxford: Oxford University Press, 175–200.

○ Hjort, N.L. and Petrone, S. (2007). Nonparametric quantile inference with Dirichlet processes. In Advances in Statistical Modeling and Inference. Essays in Honor of Kjell A Doksum, Vijay Nair Ed., 463–492.

ANNOTATED BIBLIOGRAPHY

APPROXIMATE BAYESIAN COMPUTATION

S. A. Sisson and Y. Fan
S.A.Sisson@unsw.edu.au

Approximate Bayesian computation (ABC) methods (also known as likelihood-free methods) have been gaining in popularity in recent years. They aim to implement Bayesian inference for models with computationally intractable likelihood functions, i.e., where the likelihood function of the desired model cannot be numerically evaluated. This naturally presents a number of obvious challenges in performing Bayesian analyses with such models, since tools for Bayesian computation, such as MCMC, are explicitly based on the evaluation of likelihoods. However, once these problems have been overcome, there are clear advantages to working with models which are not constrained by their mathematical tractability.

Though it was not immediately obvious when such methods first emerged, ABC posterior simulation methods involve sampling on an augmented parameter space in such a way that the need to evaluate the intractable likelihood function is avoided. More intuitively, consider drawing a parameter vector from the prior. Conditional upon this, draw a dataset from the model. As the likelihood function describes the probability of the data given a parameter vector, the probability of generating a dataset "close" to the observed data will be in direct proportion to the likelihood. In ABC methods, "closeness" is usually measured through vectors of low-dimensional summary statistics. Accordingly, if this parameter vector is given a weight which is higher the closer the simulated and observed data become, the collection of weighted draws from the prior will generate a sample from an approximation to the intractable posterior.

ABC methods primarily originated in the population genetics literature (note the journals in which many of the following papers have appeared), but have now propagated through an ever-increasing range of disciplines, given their obvious appeal to applied research in general.

It is difficult to present an exhaustive subject overview in this short bibliography, and there are many excellent articles which are not listed here. Instead we present lists of some of the major ABC-methods papers, some of the most interesting current research, and point towards some overview articles for expository reading.

Primary ABC methods papers

Each of the articles in this section either represented a major forward-step in ABC methods al-

gorithms, or established new perspectives that have enabled (or are expected to enable) substantially improved inferential methods. They are presented in chronological order.

- Tavaré S., D. J. Balding, R. C. Griffiths and P. Donnelly (1997). Inferring coalescence times from DNA sequence data. *Genetics*, **145**, 505-518. Generally considered the primary historical ABC reference, this paper developed a basic rejection sampling ABC algorithm in an applied setting. This approach was described in the introductory text. A number of variations of this approach have subsequently been developed.

- Beaumont M. A., W. Zhang and D. J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025-2035. A very novel idea which improved the ABC approximation to the true posterior while also improving efficiency. This paper ignited the subsequent flurry of ABC research proper, including much applied work. Following standard rejection sampling, the improvement is based on a regression-adjustment which aims to reduce the discrepancy between the simulated and observed data to almost zero.

- Marjoram P., J. Molitor, V. Plagnol and S. Tavaré (2003). Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA*, **100**, 15324-15328. As with standard Bayesian inference, adopting MCMC methods over rejection sampling afforded the opportunity to increase sampling efficiency. In the analyses presented, algorithm acceptance rates of 0.0008% under rejection sampling improved to 0.2% under MCMC. (Yes, ABC methods are very computationally intensive!) This paper brought ABC methods closer to the mainstream of Bayesian computational methods.

- Reeves R. W. and A. N. Pettitt (2005). A theoretical framework for approximate Bayesian computation. Proceedings of the International Workshop for Statistical Modelling, Sydney 2005: 393-396. This conference paper appears to have been the first to explicitly describe the exact nature of the ABC approximation to the true posterior distribution. As such, it maintains a strong and positive influence on subsequent ABC methods development

- Bortot P., S. G. Coles and S. A. Sisson (2007). Inference for stereological extremes. *J. Amer. Statist. Assoc.*, **102**, 284-92. The auxiliary parameter describing the distribution of simulated and observed data similarity was established as a model (tempering) parameter to be estimated. The paper developed this idea in the MCMC setting, permitting a basic *a posteriori* evaluation of model accuracy.

- Sisson S. A, Y. Fan and M. M. Tanaka (2007). Sequential Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA*, **104**, 1760-1765. Errata (2009). This article introduced sequential Monte Carlo algorithms as an efficient manner to perform ABC inference, in comparison to rejection sampling and MCMC methods. A mutation kernel based on partial rejection control was required to maintain particle diversity.

- Ratmann O., C. Andrieu, T. Hinkley, C. Wuif and S. Richardson (2009). Model criticism based on likelihood-free inference, with an example in protein network evolution. *Proc. Natl. Acad. Sci. USA*, to appear. Up to this point, most ABC methods were focused on fitting the ABC model approximation as accurately as possible. This paper introduced the insight that the degree to which the likelihood could reproduce each of the individual observed summary data could be estimated as part of the model-fitting process, and that this information provided a natural diagnostic for informed model-criticism.

Some current papers

Current ABC research is very active. As such there are a number of current articles either just published, or still in the tech report stage that offer interesting perspectives on developing ABC methods further. They are presented alphabetically in the first author.

- Blum M. G. B. and O. Francois (2009). Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing*, to appear. arXiv:0809.4178. Despite the huge popularity of the local-linear regression approach of Beaumont et al. (2002), it is surprising that it has taken so

long for this to be extended to non-linear and heteroscedastic regression to provide greater flexibility and efficiency.

- Del Moral P., A. Doucet and A. Jasra (2008). An adaptive sequential Monte Carlo method for Approximate Bayesian Computation. Preprint via Sequential Monte Carlo Methods Homepage. This article resents a sequential Monte Carlo sampler of a particular form with an automated distribution schedule. This work is an ABC-application of the more general "expected auxiliary variable" approach (Andrieu et al. 2008).

- Joyce P. and P. Marjoram (2008). Approximately sufficient statistics and Bayesian computation. *Stat. Appl. Genet. Molec. Biol.*, **7**, 23. The choice of summary statistics directly affects the quality of the ABC approximation to the true posterior. This article develops a scheme which determines whether including additional summary statistics to construct the ABC approximation will substantially improve the quality of inference.

- Peters G. W., Y. Fan and S. A. Sisson (2008). On sequential Monte Carlo, partial rejection control and approximate Bayesian computation. arXiv:0808.3466v1. This paper presents a generalised version of the sequential Monte Carlo ABC sampler, based on improved representations of the ABC approximation. The algorithm is one of the most efficient currently available.

- Wilkinson R. D. (2008). Approximate

Bayesian computation (ABC) gives exact results under the assumption of model error. arXiv:0811.3355v1. This article presents a more detailed discussion along the lines of Reeves and Pettitt (2005), but with a focus on the "model-error" arising from the mismatch of simulated and observed summary statistics.

Review papers

As ABC methods are still undergoing a rapid development, there has been little opportunity to provide subject overviews (such as this!) as these are likely to be out-of-date within a short period of time. However, a few review-type articles have appeared, listed below.

- Marjoram P. and S. Tavaré (2006). Modern computational approaches for analysing molecular genetic variation data. *Nat. Rev. Genet.*, **7**, 759-770. A nice review from a molecular genetics perspective, evaluating the state of the art in 2006.

- Sisson S. A. (2007). Genetics and stochastic simulation do mix! *The American Statistician*, **61**, 112-119. ABC methods (circa 2007) are discussed as one instance in which the interplay between genetics and stochastic simulation research has advanced methodological development in both disciplines.

- Sisson S. A., G. W. Peters, Y. Fan and M. Briers (2008). Likelihood-free samplers. Technical report, University of New South Wales. An overview and illustration of non-regression-based ABC methods.

APPLICATIONS

BAYESIAN NONPARAMETRICS

Peter Müller and Fernando Quintana
pmueller@mdanderson.org

Statistical models almost never reflect reality. All models involve certain parametric and structural assumptions. Nonparametric Bayesian inference is an increasingly widely used approach to mitigate the dependence on such assumptions. Technically, nonparametric Bayesian (NPB) models

can be defined as probability models on infinite dimensional parameter spaces, usually devised for random distributions or random mean functions. Typical examples are the Dirichlet process (DP) and Polya tree (PT) priors for random distributions, or Gaussian process priors for random functions. In this short discussion we will focus on a specific class of models that can be

characterized as probability models for families of related distributions. Recent more comprehensive reviews of BNP inference appear in Walker et al. (1999), Hjort (2003) or Müller and Quintana (2004).

Consider a family of random probability measures $\mathcal{F} = \{F_x, \ x \in X\}$ indexed with a covariate x. For example, F_x could be a random effects distribution for patients with baseline covariates x. Baseline covariates might include treatment, dose, initial severity of a disease etc. Or F_x could be the distribution of a set of biomarkers in the subpopulation of high risk patients ($x = 1$) and low risk patients ($x = 0$). Or F_x could be the distribution of suitably defined difference scores for genes that are differentially expressed ($x = 1$) versus not differentially expressed ($x = 0$). A parametric model might proceed by assuming a (multivariate) normal model for F_x. However, when the experimental units are patients, then the normal assumption can be criticized for ignoring the ubiquitous heterogeneity of patient populations.

Alternatively we can construct an NPB prior model $p(\mathcal{F})$. Many popular constructions are variations of non-parametric mixture of normal models. The construction starts with $F_x(\theta) = \int N(\theta; \ \mu, \tau) \, dG_x(\mu)$, i.e., a mixture of normal kernels with respect to a mixing measure G_x on the location parameter (including the scale parameter in the mixture is easily possible). The model for F_x is completed by specifying a prior for G_x. A prior on G_x is a probability model on the random distribution G_x. The most popular of all is a DP prior, $p(G_x) = \mathrm{DP}(G^\star, c)$ for a random probability measure G_x. The DP prior requires the specification of two parameters, the base measure $G^\star$, which defines the prior mean $E(G_x) = G^\star$, and the total mass parameter c, which can be interpreted as a precision parameter. The DP prior generates discrete random measures $G_x = \sum w_h \delta(m_{xh})$. The discrete nature of G is the main reason for using the additional convolution with the normal kernel in F_x. One of the many defining properties of the DP is that the point masses are an i.i.d. sample from the base measure, $m_{xh} \sim G^\star$, and the weights are scaled independent Beta draws, $w_h \sim \left(1 - \sum_{g<h} w_g\right) \mathrm{Be}(c, 1)$. In summary, we now have a constructive definition of $p(F_x)$ as

$$F_x(\theta) = \sum w_h N(\theta; \ m_{xh}, \tau) \qquad (1)$$

with a DP prior on $G_x = \sum w_h \delta(m_{xh})$. The def-

inition is marginally for each x. By definition of the DP prior the m_{xh} are independent across h. MacEachern (1999) extends the marginal model (1) to a probability model for $\{F_x, \ x \in X\}$ by assuming that m_{xh} are *dependent* across x. The model is known as the dependent DP (DDP) model.

Perhaps the easiest form of dependence is a linear model on the m_{xh}. If x are only categorical covariates, we get an ANOVA model for $\{m_{xh}, \ x \in X\}$, one model for each h. This ANOVA DDP is the construction in De Iorio et al. (2004) and in Müller et al. (2005).

Müller et al. (2004) consider a variation of the DDP mixture of normal model for the special case when $x \in \{1, 2, \ldots, k\}$ indexes k related studies. We define $p(\mathcal{F})$ by simply piecing together mixing measures with some probability mass that is common to all studies and some remaining part that is specific to each study. Formally, the mixing measure in the mixture of normals F_x is defined as $G_x = \epsilon H_0 + (1 - \epsilon) H_x$ and $H_j \sim \mathrm{DP}(H^\star, c)$, independently across $j = 0, 1, \ldots, k$. The construction has a natural interpretation when the F_x are distributions for patient-specific random effects in related studies $x = 1, \ldots, k$. The model reflects heterogeneity of patient populations, with H_0 representing a subpopulation that is common across studies and H_x representing patient subpopulations specific to each study.

An interesting application arises when x indexes different subpopulations of interest, and the aim of the study is to classify a new patient into one of these subpopulations. Without loss of generality assume $x \in \{0, 1\}$. Cruz-Mesía et al. (2007) construct a BNP model that allows such classification. Let y_i denote the response for the i-th subject, and let x_i denote the classification into the two subpopulations. Assume that x_i is known for $i = 1, \ldots, n$, and the unknown classification x_{n+1} is to be predicted on the basis of the observed response y_{n+1}. Cruz-Mesía et al. (2007) define a a sampling model $p(y_i \mid x_i, \mathcal{F}) = F_x$ and an ANOVA DDP on $\mathcal{F}$. The model is completed with $p(x_i = 1) = \pi$. The (marginal posterior) predictive probability $p(x_{n+1} = 1 \mid y_{n+1}, data)$ defines the desired inference. When $y_i = (y_{i1}, \ldots, y_{im_i})$ is a longitudinal response the model allows easy continuous updating of $p(x_{n+1} = 1 \mid y_{n+1,1}, \ldots, y_{n+1,m_i}, data)$ as m_i increases. A similar model is used in Do et al. (2005) to classify genes as differentially versus non-differentially expressed.

In recent literature, similar constructions are the order based DDP of Griffin and Steel (2006), the probit stick-breaking model (PSBP) of Chung and Dunson (2008) and the weighted mixture of DPs (WMDP) of Dunson et al. (2007). The order based DDP introduces the desired dependence across F_x by permuting the weights in a systematic fashion as x changes. The PSBP parametrization uses a representation like (1), but with covariate-dependent weights w_{xh} and common point masses m_h. The weights are explicitly parametrized as a regression on x. The WMDP assumes that the random distributions F_x are weighted mixtures of independent random probability distributions F_ℓ^o. The weights are functions of the covariates.

Example

We use a dataset from Klein and Moeschberger (2003, Section 1.11). The data report survival times for 80 patients with cancers of the mouth. Samples are recorded as aneuploid (abnormal number of chromosomes) versus diploid (two copies of each chromosome) tumors. We use the R function `ddpsurvival()` from the package *ddpanova* to estimate an ANOVA DDP model for survival. The only covariate is an indicator for aneuploid versus diploid tumors. The inference is shown in Figure 1 below.

```
require(KMsurv)     # from R CRAN
require(ddpanova)
## from http://odin.mdacc.tmc.edu/~pm

## tongue data from Section 1.11,
## Klein & Moeschberger (2003)
data(tongue)
attach(tongue)
Y <- cbind(time,delta,time)
D <-  cbind(1, ifelse(type==2,1,-1))
D0 <- cbind(1, c(-1,1))
ddpsurvival(Y,D,n.iter=3000,d0=D0,
   S.init=100,S.prior=0)

## posterior predictive
pp <- post.pred()
matplot(pp$ygrid,t(pp$Sy),type="l")
## add KM plot
fit <- survfit(Surv(time,delta)~type)
lines(fit,col=1:2,bty="l",lty=1:2)
```
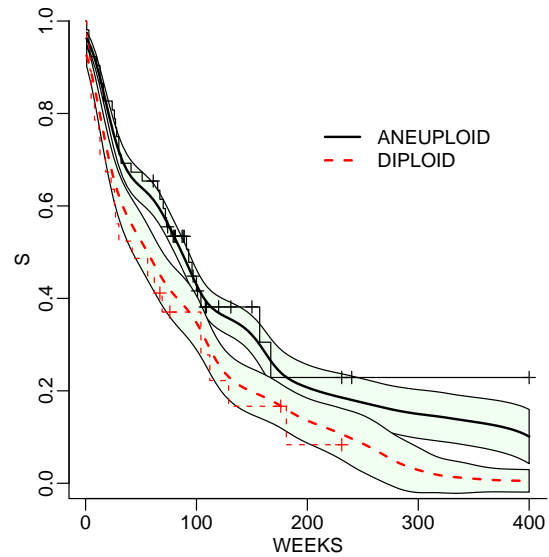

Figure 1. Estimated survival function by tumor type (solid black and dashed red curves). The grey shaded bands around the estimated survival functions show pointwise ± 1.0 posterior standard deviation bounds. The piecewise constant lines plot the Kaplan-Meier estimates.

References

Chung, Y. and Dunson, D. B. (2008), "Nonparametric Bayes Conditional Distribution Modeling with Variable Selection," Tech. rep., Department of Statistical Science, Duke University.

Cruz-Mesía, R. D. l., Quintana, F. A., and Müller, P. (2007), "Semiparametric Bayesian Classification with Longitudinal Markers," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 56, 119–137.

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205–215.

Do, K.-A., Müller, P., and Tang, F. (2005), "A Bayesian Mixture Model for Differential Gene Expression," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 54, 627–644.

Dunson, D. B., Pillai, N., and Park, J.-H. (2007), "Bayesian Density Regression," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 69, 163–183.

Griffin, J. E. and Steel, M. F. J. (2006), "Order-based Dependent Dirichlet Processes," *Journal of the American Statistical Association*, 101, 179–194.

Hjort, N. L. (2003), "Topics in Nonparametric Bayesian Statistics," in *Highly structured stochastic systems*, eds. Green, P., Hjort, N., and Richardson, S., Oxford University Press, pp. 455–487.

Klein, J. P. and Moeschberger, M. L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, Springer-Verlag Inc.

MacEachern, S. N. (1999), "Dependent Nonparametric Processes," in *ASA Proceedings of the Section on Bayesian Statistical Science*, American Statistical Association, pp. 50–55.

Müller, P., Quintana, F., and Rosner, G. (2004), "A Method for Combining Inference across Related Nonparametric Bayesian Models," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 66, 735–749.

Müller, P. and Quintana, F. A. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110.

Müller, P., Rosner, G. L., Iorio, M. D., and MacEachern, S. (2005), "A Nonparametric Bayesian Model for Inference in Related Longitudinal Studies," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 54, 611–626.

Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999), "Bayesian Nonparametric Inference for Random Distributions and Related Functions (Disc: P510-527)," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 61, 485–509.

STUDENTS' CORNER

Luke Bornn
l.bornn@stat.ubc.ca

This quarter's Students' Corner features the PhD abstract of University of Florida graduate Georgios Papageorgiou. As always, if you (or your student) have recently graduated and would like your dissertation abstract published, please contact me.

Dissertation Abstracts

MULTIVARIATE LIMITED TRANSLATION ESTIMATORS

by Georgios Papageorgiou
gpapageo@gmail.com
Department of Statistics,
University of Florida
PhD Supervisor: Dr. Malay Ghosh

Bayes estimators are used quite often in the theory and practice of statistics. Their popularity is attributed to the more efficient inference that they often lead to compared to classical frequentist procedures. However, the Bayes estimators can result in high Bayes risk when the prior distribution is misspecifed. Additionally, they can result in high frequentist risk when they are used for estimating parameters which depart widely from the assumed prior means.

We developed multivariate limited translation Bayes estimators of the normal mean vector which serve as a compromise between the Bayes and the maximum likelihood estimators. We demonstrated the superiority of the limited translation estimators over the usual Bayes estimators under misspecifed priors and often also in terms of the frequentist risks under the situation stated in the previous paragraph.

We also extended the above results and developed multivariate limited translation empirical Bayes estimators of the normal mean vector which serve as a compromise between the empirical Bayes estimators and the maximum likelihood estimators. These compromise estimators perform better than the regular empirical Bayes estimators, in a frequentist sense, when there is wide departure of an individual observation from the grand average.

NEWS FROM THE WORLD

Announcements

I would like to encourage those who have any announcements or would like to draw attention to an up-coming conference, to get in touch with me and I would be happy to place them here.

Events

O-Bayes09, University of Pennsylvania, Philadelphia, 6-9th Jun. 2009.

O-Bayes09, the 2009 International Workshop on Objective Bayes Methodology will take place at the Wharton School of the University of Pennsylvania, Philadelphia, PA, USA. The principal objectives of O-Bayes09 are to facilitate the exchange of recent research developments in objective Bayes methodology, to provide opportunities for new researchers to shine, and to establish new collaborations and partnerships that will channel efforts into pending problems and open new directions for further study. O-Bayes09 will also serve to further crystallize objective Bayes methodology as an established area for statistical research.

The workshop will consist of a series of invited talks followed by discussion and one or more sessions dedicated to contributed posters.For more information visit the website, http://stat.wharton.upenn.edu/statweb/ Conference/OBayes09/OBayes.html.

BISP6 - Bayesian Inference in Stochastic Processes, Bressanone/Brixen, Italy, 18-20th Jun. 2009.

In this workshop, we will bring together experts in the field to review, discuss and explore directions of development of Bayesian Inference in Stochastic Processes and in the use of Stochastic Processes for Bayesian Inference. There will be sessions on Markov processes, state-space models, spatial, empirical, birth-death and branching processes. Theoretical and applied contributions (for example queueing, population modelling, signal processing) are both welcome. The workshop will thus be of interest to workers in both Bayesian Inference and Stochastic Processes. For more information visit the website, http://www. mi.imati.cnr.it/conferences/bisp6.html.

7th Workshop on Bayesian Nonparametrics, Moncalieri (Turin), Italy, 21-25th Jun. 2009.

The aim of the 7th Workshop on Bayesian Nonparametrics is to highlight the latest developments in Bayesian Nonparametrics covering a wide variety of both theoretical and applied topics. The meeting will be held at the Collegio Carlo Alberto, a Research Institution housed in an historical building located in Moncalieri on the outskirts of Turin, Italy. For more information visit the website, http://bnpworkshop. carloalberto.org, or contact Pierpaolo De Blasi *bnp@carloalberto.org*.