THE ISBA BULLETIN

Vol. 16 No. 4 December 2009

The official bulletin of the International Society for Bayesian Analysis

A MESSAGE FROM THE PRESIDENT

- Mike West -

ISBA President, 2009

mw@stat.duke.edu

With end-of-year holiday greetings to all ISBA Members, my final *Bulletin* message as ISBA President will be a short and summary one. I began my year as ISBA President with a few specific goals related to increasing the stability and visibility of the Society and setting in place mechanisms to ensure growth and vitality for the longer term. Initiation, development and progress on key activities have been reported in detail in earlier 2009 issues of the *Bulletin*. Some recent updates are noted below, along with news of emerging activities.

ISBA IT & VIRTUAL OFFICE.
Our development of the virtual office is now complete. With the establishment in mid-2009 of a new membership data base and services system, a negotiated agreement with Duke University to partner with ISBA and contribute to the maintenance and management of the IT system, and the appointment of the amazingly effective and efficient *ISBA Membership Services Coordinator* Alicia Korenman, the Society has now transitioned to operations – email communications and membership records – completely within the system. Thanks are due to Lance Brown of Duke University for his continued systems support.

ISBA ELECTIONS.
You will have received the email announcement of results of the recent ISBA elections (posted at
. . . *Continued on page 2.*

A MESSAGE FROM THE EDITOR

by Raphael Gottardo

raphael.gottardo@ircm.qc.ca

It's December, and Christmas is only a week away! This should be a great and happy time of the year, but I am a bit sad as this is my last bulletin as editor. I have enjoyed my work very much thanks to a great team of AEs, many member contributions and three great ISBA presidents (Peter Green, Christian Robert, and Mike West). Many thanks to all of you! Starting in January, Manual Mendoza will take over my position as new editor of the ISBA bulletin. I am sure Manual will do a terrific job, and I will of course work with Manual to enable a smooth transition over the next few months. Welcome Manual!

I think our field and community, Bayesian statistics, is very strong and growing even more every day. *Continue on page 2.* ▲

In this issue

➤ A MESSAGE FROM THE BA EDITOR
☛*Page 5*

➤ INTERVIEW
☛*Page 6*

➤ ANNOTATED BIBLIOGRAPHY
☛*Page 8*

➤ BAYESIAN HISTORY
☛*Page 11*

➤ APPLICATIONS
☛*Page 13*

➤ STUDENTS' CORNER
☛*Page 17*

➤ NEWS FROM THE WORLD
☛*Page 18*

MESSAGE FROM THE PRESIDENT, *Continued from page 1.* ...http://bayesian.org/election/), namely:

- Michael I. Jordan, as President 2011,
- Merlise Clyde, as Executive Secretary 2010-2012,
- Siddhartha Chib, Arnaud Doucet, Peter Hoff and Raquel Prado, as members of the Board of Directors 2010-2012.

It remains for me to publicly congratulate those elected and thank them again, on your behalf, for their willingness to serve. I look forward to their leadership contributions to ISBA in coming years.

ISBA EDITORS.
You will also have received the email announcement of results of Board decisions following the recommendations of the ISBA Editor Search Committee. As of January 1st 2010, our three existing publications are to be edited and led by:

- Herbie Lee, as Editor in Chief of *Bayesian Analysis*,
- Manuel Mendoza, as Editor of the ISBA *Bulletin*,
- Chris Hans, as ISBA Web Editor.

Again, I am very pleased to publicly congratulate Herbie, Manuel and Chris, and thank them for their willingness to serve in these central leadership positions – for ISBA and the worldwide Bayesian community ISBA represents.

I am also pleased to pass on the good news, from BA Editor in Chief elect Herbie Lee, that Alyson Wilson (Iowa State University) has agreed to join Bayesian Analysis as Managing Editor from January 1st. This is a critical position in the operation and direction of our flagship journal, and I am sure ISBA will be ably served by Alyson taking up this challenging position.

NEW PUBLICATION AND A NEW ISBA EDITOR.
I also have news for the members on a new ISBA publication, to debut in 2010. As a result of discussions between several statistics societies and the Springer publisher, 2010 will see the establishment of a new online encyclopedia entitled *Probability and Statistics Online*. This inter-societal venture will be a free, open source, online encyclopedia hosted by Springer, run and developed via an Editorial Board comprised of editors representing several statistics societies. The initial core of societies participating will have the right to approve additional societies and the Editors will solicit, oversee and regulate the development of the content. This is not a commercial venture and the overall flavour is that of developing a major resource for the profession that is driven and managed by the profession through the collaborating societies. There will be more, detailed news in early 2010 prior to the official launch.

As of now, ISBA is a founding society, along with the American Statistical Association (ASA), the Bernoulli Society (BS), the Institute of Mathematical Statistics (IMS), the International Biometric Society (IBS), the International Chinese Statistical Association (ICSA), the International Statistical Institute (ISI), and the Statistical Society of Canada (SSC). As you will recall from previous 2009 issues of the *Bulletin*, we have active or initiated discussions with several of these organisations on ways and means of collaborating to the benefit of Bayesian statistics, and this new online publication represents one new dimension for inter-societal activism.

I am also pleased to report that the ISBA Board has selected Susie Bayarri (1998 ISBA President, currently Professor of Statistics, University of Valencia and Adjunct Professor of Statistical Science at Duke University) as our founding Editor of *Probability and Statistics Online*. Congratulations are due to Susie, as are our thanks for her willingness to assume this important position – a major new publication venture that has the potential to become a substantial resource for Bayesian communication, research and education.

MEMBERSHIP.
First, if you have a membership ending on December 31st this year and have not yet renewed, now is the time! Just login at the ISBA home page .....

The opportunities for 3-year and Life memberships are attractive for some of us, and the stability and longer-term planning of the Society is enhanced and eased by increased numbers of members in these categories. As some of you will know, membership notices were sent in early December and reminders will follow to only those that have not yet renewed.

Second, I am very pleased to report that membership is increasingly buoyant and growing. Based on a number of initiatives of the Membership Committee we established this year coupled with general activism to promote ISBA –

by broadening awareness of the roles and services ISBA provides to members, and by simply communicating more widely – we have seen an upsurge in membership. The core membership over the last decade or so has been *very* volatile, averaging in the low 400s and being boosted by one-year/once-only memberships at World Meetings and free memberships given through various workshops (a practice we discontinued in early 2009). As we approach the end of 2009, I am very pleased to report that the membership stands at over 675, an all-time high and with strong indications of continued core growth. We will see a boost to numbers based on the upcoming World Meeting in 2010, but the important trajectory is sustained growth of renewals and new members. With a firmer Societal focus on visibility and promotion of services, and an active ongoing Membership Committee, I believe ISBA is poised for longer-term growth to sustained membership levels that represent the thousands of active Bayesians worldwide. I look forward to seeing us reach those levels soon!

MORE THANKS!
Finally, thanks are due to several ISBA members in addition to those already named above.

First, to our current editors who will step down at the end of 2009: Brad Carlin, outgoing Editor in Chief of *Bayesian Analysis*, Raphaël Gottardo, outgoing Editor of the *ISBA Bulletin*, and Robert Wolpert, outgoing ISBA Web Editor. Each of these is ending a three year term on December 31st, and ISBA has been very well served by their efforts and contributions since 2007.

Second, to our current elected officials who will step down at the end of 2009: outgoing Past President Christian Robert, outgoing Executive Secretary Robert Wolpert, and outgoing Board members David Heckerman, Xiao-Li Meng, Gareth Roberts and Alexandra Schmidt. Each of these is also ending a three year term on December 31st, and we should thank them for generous and effective service to the profession. Alex's service to ISBA through other formal roles continues into 2010-2011, of course, as she takes over as chair of the ISBA Program Council on January 1st.

Third, but by no means least, I would like to recognize and thank Heidi Sestrich (on staff at Carnegie Mellon University) for her contributions in support of ISBA membership and email communications for several years. Heidi came on board as a part-time Administrative Secretary in 2004, and served ably and actively in membership management and communications until this year. Most recently, Heidi's work in assisting the transition of ISBA membership records to the new system and virtual office has been invaluable. On her retirement from ISBA service in 2009, Heidi has been formally recognized with an ISBA *Certificate of Appreciation* for her work and support over the last six years.

It has been a busy year for ISBA, and a productive one. We are in good shape in terms of membership growth, and very solid financially; we have a central focus on developing membership services more actively and visibly; we have begun to develop inter-societal interactions more aggressively; and we have a core focus on promoting membership growth linked to increasing visibility and roles of the Society in the professional life of Bayesians worldwide. I believe ISBA is well positioned for my successor, 2010 President Peter Müller, together with the new, incoming executive and Board members, to be able to substantially build momentum during 2010.

As always, I invite and encourage all ISBA members to contact any member of the elected Board or Executive committee on any societal matters, and to consider playing active roles to further grow and enrich the existing activities of ISBA on behalf of Bayesians worldwide.

In the meantime, very best wishes to all for a happy, peaceful and increasingly Bayesian world in 2010!

▲

WORDS FROM THE EDITOR, *Continued from page 1.* Even though all of our current editors (BA, Web, Bulletin) are stepping down, fresh blood is coming in and I am sure our new editors will help take Bayesian statistics to the next level. Speaking of new editors, I have included a short bio description for our three editors: Herbie Lee, Chris Hand and Manuel Mendoza. This will give you an opportunity to learn more about their background, research activities and goals for their respective positions. I'd like to welcome each of them and wish them the best of luck in their next position.

Last but not least, I'd like to thank all of my associate editors that have done a terrific job over the years. I should say that as editor my job

has always been very easy thanks to them. They have always worked very hard to obtain interesting contributions, and that under very rigid time constraints. I would like you to join me in thanking Sebastien Haneuse, Mayetri Gupta, Beatrix Jones, Tim Johnson, Luke Bornn, Donatello Telesca and Alex Lewin for their great work and numerous contributions to the bulletin. ▲

ISBA New Editors

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

Here is a short bio of our future ISBA editors.

Bayesian Analysis – Herbie Lee http://users.soe.ucsc.edu/~herbie/
Herbie received his PhD in statistics from Carnegie Mellon University in 1998. He completed a post-doc at Duke University before joining the faculty at the University of California, Santa Cruz, where he is now a full professor. He is a fellow of the American Statistical Association. He is a co-author of the book "Multiscale Modeling: A Bayesian Perspective", and author of the book "Bayesian Nonparametrics Via Neural Networks", and has published widely in journals such as the Journal of the American Statistical Association, Technometrics, and Bayesian Analysis. Awards include a UCSC excellence in teaching award, best JSM contributed papers in the Statistical Education Section and in the Statistical Computing Section, and an ASA-SIAM book series author award. His current research focuses on statistical issues of computer simulation (e.g., surrogate modeling, sequential design, optimization), spatial modeling, and connections between statistics and machine learning.

Herbie's past elected positions include board member of the Classification Society of North America and Publications Officer of the American Statistical Association, Section on Bayesian Statistical Science. He is the 2009 Program Chair for ISBA, and has previously served on the ISBA nominating committee.

His current and past editorial duties included service as an Editor for Chance, as an associate editor for the Journal of the American Statistical Associate and for Statistics and Computing, and as managing editor for Bayesian Analysis. Being managing editor of BA since 2005, Herbie has worked very closely with both past Editors in Chief – Rob Kass and Brad Carlin – and is intimately familiar with the history of the journal and most of the policy issues it currently faces.

Web – Chris Hans http://www.stat.osu.edu/~hans I am an Assistant Professor in the Department of Statistics at The Ohio State University, and have been at this position since receiving my PhD in statistics from Duke University in 2005. My research is focused on Bayesian methods and associated computation for regression models, especially in high-dimensional settings. Much of this work is computationally oriented, including parallel computing methods for model space exploration, and includes public domain software distribution. Recently I have been investigating Bayesian analogues of several regularized regression algorithms. My research appears in mainstream statistics journals such as Biometrics, Biometrika, JASA and Statistical Science, and is funded by NSF and other grants. I am currently an Associate Editor of JCGS.

I am a lifetime member of ISBA. I have contributed several articles to the ISBA Bulletin, and have contributed to the Bayesian community (though not specifically to ISBA) by organizing and chairing SBSS sessions at JSM and serving on the SBSS Student Paper Award Selection Committee. I have experience with HTML and maintain my own web page: http://www.stat.osu.edu/~hans

Bulletin – Manuel Mendoza http://allman.rhon.itam.mx/~mendoza/mendoza.html I got a Doctor in Sciences degree (Mathematics) in 1988 at the National University (UNAM) in Mexico. To this purpose, I spent two years in Valencia, Spain, where I wrote my thesis under the supervision of J.M. Bernardo. Prior to this, I got a Master in Science in Statistics and a Bachelor degree in Actuarial Sciences from the same university. I was director of the Statistics master program at the UNAM for some time and later I moved to the Instituto Tecnologico Autonomo de Mexico (ITAM) as Statistics Professor. Here, I am direc-

tor of the Risk Management master program. I was President of the Mexican Statistical Association (1998-1999). For some time I was also Associate Editor of the TEST journal in Spain and Theory and Methods Editor of ESTADISTICA, the journal of the Inter American Statistical Institute. These days I am mainly interested in Bayesian inference for risk management.

BAYESIAN ANALYSIS - A MESSAGE FROM THE EDITOR

Bayesian Analysis LAST WORDS

by Brad Carlin
***Bayesian Analysis* Editor-in-Chief**
brad@biostat.umn.edu

The current issue of Bayesian Analysis, Volume 4 Number 4, is the twelfth and final one for which I have the privilege of serving as editor-in-chief (EiC); my three-year term (2007-09) is drawing to a close. It's been an enormously gratifying and enlightening run, so I'd like to take just a few paragraphs to say thanks, describe a bit of what we've accomplished as a journal in the past year or so, and mention where the journal is headed and what challenges and opportunities it's likely to face there.

As I type this, it is near the end of Thanksgiving Day in the United States, and it's impossible not to reflect on how thankful I am for the chance to serve as EiC, and for the many dedicated men and women who do all the real work of the journal, completely without financial compensation and at a time when ever-increasing pressures to further improve productivity encourage one to forego often-thankless volunteer work like editing and refereeing whenever possible. Simply put, I am eternally grateful to all the editors, associate editors, referees, and production staff who make each issue possible. It is dangerous to begin naming names in situations like this since one is sure to miss someone important, but I do want to mention a few key persons who have been around BA since the very beginning some 5 years ago. These include System Managing Editor Pantelis Vlachos, Managing Editor Herbie Lee, Deputy Editor Marina Vannucci, and Editors Philip Dawid, David Heckerman, Michael Jordan, and Fabrizio Ruggeri. Philip and Marina have decided to step down as part of the transition, and let me thank them at this time for their years of outstanding service. I am hopeful that most if not all of the others are willing to continue, along with Production Editor Angelika van der Linde and Editors Kate Cowles, David Dunson, Antonietta Mira, and Bruno Sansó. All do wonderful work, as do all our AEs and referees. Thanks again.

2009 has been another good year for the journal. We will again publish about 850 pages, very similar to our page counts in 2007 and 2008. We submitted three consecutive issues of the journal to Thomson-Reuters as evidence that we merit inclusion in their indexing systems, and in October we received the good news that BA has been accepted into the Science Citation Index-Expanded (SCIE) including the Web of Science, the ISI Alerting Service, and Current Contents/Physical, Chemical and Earth Sciences (CC/PC&ES). We have been told that Thomson-Reuters will have sufficient source item and citation data to compute an impact factor for the next Journal Citation Reports (JCR), which is scheduled to be published in June 2010. Getting BA on the road to an impact factor (critically important these days, especially to our European contributors) was one of my primary goals as EiC, so I'm very pleased this got done before my term expired. My other primary goal was to keep the flow of interesting discussion papers going, and here again I've been very pleased by the results. I have striven to select a wide range of papers for this very visible quarterly slot, from foundations to applications and even an occasional thought piece (such as Andrew Gelman's "April Fools' Day blog" paper in Volume 3 Number 3). The current issue's discussion paper is on a subject near and dear to my heart (baseball), and as usual features some state of the art Bayesian methods followed by a spirited question-and-answer period in the discussions and rejoinder. We also get a nice stream of potential discussion papers from the Case Studies in Bayesian Statistics (Carnegie Mellon) and other ISBA-related meetings throughout the year. Some in the profession have suggested opening the papers on the BA website to discussion by anyone, rather than permit only a few papers to be discussed by a

few high-profile discussants selected by the EiC. I must confess I have been hesitant to change our current system, since I like the idea of one "special" paper per issue, plus selecting this paper and coordinating the discussions and rejoinders is perhaps the best part of the EiC job! But it's a good suggestion and one with which the next EiC and editorial board may choose to grapple.

Speaking of the new EiC, I am very pleased to announce to those who missed the email from ISBA President Mike West that it will be none other than Herbie Lee, the current (and founding) managing editor of BA. I am very pleased the search committee offered the position to Herbie, and that he accepted! It means an especially easy transition period since Herbie's long tenure with the journal means he needs essentially no "training" of any kind. Herbie will no doubt want to bring on some new editors and AEs, and I'm confident the journal will remain strong under his leadership. Of course, as I said there will be challenges to greet Herbie and his team. One involves the online review system, which was constructed specifically for us several years ago, but is now beginning to show its age somewhat. Other online review system products are out there that may offer advantages over ours in terms of flexibility and extendability in the long run. One of these is already used by our institutional partner IMS, for whom BA is already an IMS Supported Journal and with whom ISBA already has a joint membership agreement and a variety of well-attended jointly sponsored conferences (including the MCMSki series, the next of which will be January 5-7, 2011). Expanding this IMS partnership may be most natural. A second issue continues to be finding a reliable revenue stream to support the journal. We now offer on-demand printing of issues for a small fee, but BA's free online availability has essentially precluded any significant sales revenue. Other arrangements may include adding advertising, or even affiliating with the Berkeley Electronic Press, which has a fairly long history of profitably running journals like ours. I happen to know BEP is interested in seeing this happen, but whether we should surrender our independence in such a dramatic way for a modest revenue stream is again something Herbie and the ISBA Board will need to ponder. On that note, I close this editorial. Thanks again for the opportunity to serve as EiC, and for your support of the journal during my tenure. While my involvement with the journal will now shrink dramatically in order for me to free up the time necessary to chair my own biostatistics group at Minnesota, I will stay involved as a guest editor, helping the journal process the contributed papers from the upcoming Valencia 9/2010 ISBA World Meeting in Spain this June. I look forward to seeing many of you at that meeting, and as always, to your submissions at http://ba.stat.cmu.edu, and your more personal thoughts and reactions via email to brad@biostat.umn.edu.

INTERVIEW

MARINA VANNUCCI

by Donatello Telesca
donatello.telesca@gmail.com

Marina Vannucci is Full Professor in the Department of Statistics at Rice University, Houston, TX. She received her Ph.D. degree in Statistics in 1996 from the University of Florence, Italy. Prior to joining Rice in 2007, Marina was in the faculty at Texas A&M University (1998-2007) and Research Fellow at the University of Kent at Canterbury, UK (1996-1998). Marina's research focuses on the theory and practice of Bayesian variable selection techniques and on the development of wavelet-based statistical models and their application. She was the recipient of an NSF CAREER award in 2001 and won the Mitchell prize from the International Society for Bayesian Analysis in 2003. She is an elected Fellow of the American Statistical Association (ASA) and of the Institute of Mathematical Statistics (IMS) and an elected member of the International Statistical Institute (ISI). Marina has advised 9 Ph.D. students and 4 postdoctoral fellows. Her CV currently lists more than 60 technical papers.

I had the pleasure of meeting Marina as I was serving some time in Texas. She graciously agreed to answer some of our questions.

1. *First of all, congratulations on your recent election as a Fellow of the IMS! Among your many recognitions and awards, can you identify one in particular*

which makes you feel particularly proud?

Thanks! It is natural to feel a sense of pride when your achievements get acknowledged. One recognition that was particularly instrumental to my career is the SIS (Italian Statistical Society) prize "Best Doctoral Thesis in Statistics" that I won in 1996, for my Ph.D. thesis work. It came at a point in my life when I was trying to decide what to do next. It represented an expression of appreciation from the Italian statistical community and gave me confidence in myself and in my abilities as a researcher. Recognitions at early stages are key in stimulating young researchers to advance in their academic careers. ISBA certainly recognizes the importance of this, with the Savage awards for outstanding doctoral dissertations.

2. *As a Scientist that has co-authored more than 60 papers. What would be your advice to a Statistician at the beginning of her academic career?*

In order to succeed in the academic world you need a combination of talent, skills and determination. Clearly, there is a strong individual component that can only come from yourself. However, it is also important that you do not miss any opportunity to receive the necessary training and to work with people who can expose you to the process of doing research. As an advisor, I always find fascinating to see a diligent student growing into a fully developed researcher. Personally, I consider advising a very important part of my profession and have in fact advised several Ph.D. students and postdoctoral fellows throughout my career. In the past few years I have particularly enjoyed leading small interdisciplinary teams of students and young researchers. I find this a particularly effective way of training students.

3. *Who do you think most influenced your own career?*

I have been fortunate to have had the support of people who, early on in my career, helped me to achieve my goal of becoming an independent researcher. The most influential person has certainly been my Ph.D. advisor, Antonio Moro, at the University of Florence, in Italy. He was very supportive and understood my potential. He also encouraged me to look beyond the Italian academic world. He introduced me to wavelets and put me in contact with a couple of people in America who at that time were doing research on the topic. Later on I had the opportunity to work with talented researchers in the field of wavelets, both in statistics and engineering. Another turning point in my career was the postdoctoral experience I did in the UK, working with Phil Brown. It is then that I became a Bayesian!

4. *You mentioned that your first Statistical work involved Wavelets. What kind of problems got you first interested in the Subject?*

The topic was suggested to me by my Ph.D. advisor. He knew I had a background in Math and Computing and helped me navigating through the literature. At that time (mid 1990's) statisticians had just started to look at wavelets. The majority of books and articles available were written by mathematicians and engineers, mainly addressing the theoretical developments of wavelets, and were quite intimidating. Once I got through that learning stage then it was clear to me that wavelets had potential for application in a variety of statistical topics. My early work was on density estimation and on Bayesian wavelet shrinkage. Later on I focused on the development of Bayesian wavelet- based modeling of functional data, proposing original ways to use wavelet methods for dimension reduction when multiple curves are under study. I have also done work on wavelet-based methods that relate to the modeling of time series data, with various applications. Nowadays, wavelets are for me one of the many tools I can pull out of my statistical hat whenever appropriate. For example, recently I have used wavelets for the denoising of mass spectrometry data in Bioinformatics. Over the last 2-3 years I have worked at organizing the now huge literature on wavelets and statistical applications into what I hope will become a monograph (one day).

5. *On the Web of Science, the 1998 paper with J. P. Brown and Fearn T, is identified as one of your highest impact publications, with more than 50 citations. What do you think was the main contribution of that manuscript?*

Together with wavelets, the development of Bayesian methods for variable selection represents my other area of strong interest. The 1998 paper was my first contribution to the topic, with Brown and Fearn. There we addressed multi-

variate regression settings, extending stochastic variable selection methods that use prior distributions with a spike at zero. My later work has been on innovative ways of performing variable selection in model-based classification with probit models and in sample clustering, both in terms of finite mixture models and infinite mixture of distributions, via Dirichlet process mixtures.

6. *The Bayesian literature has seen increased popularity of model selection procedures, with different flavors and levels of sophistication. From a more general perspective, do you think there still are some interesting open problems in Bayesian Model Selection?*

It is true that the topic has received considerable attention. This is certainly an indication of the importance and usefulness of these methodologies. Indeed, Bayesian techniques for variable selection are relatively simple and natural to use, with their employment of latent binary indicators and mixture priors. More important, they can be applied to cases with a large number of explanatory variables, typically many more than the number of observations, and come with efficient stochastic search techniques that allow fast implementations. Certainly, thorough investigations have been done, at least in the case of linear settings. However, there may be still work do in extending methods and techniques to generalized and nonlinear/nonparametric modeling frameworks, where the necessary conditional and marginal densities may not be directly available, therefore requiring more sophisticated sampling techniques.

7. *What is the link between model selection and high throughput data in Bioinformatics?*
One of the reasons for the popularity of the Bayesian methods for model/variable selection is in their success in applications to important real-data problems. One example is the analysis of high-throughput data in Bioinformatics. The major challenge in analyzing high-throughput microarray and proteomics data is in their dimensionality, in that datasets typically consist of a large number of variables (genes or proteins) measured on a relatively small number of patients. In addition, novel methodological questions are now being generated and require the integration of different concepts, methods, tools and data types. This is an ideal scenario for Bayesian methodologies. In spite of the substantial amount of data that the scientific community has produced, the development of statistical tools for an integrated use of these data is still at an early stage.

8. *What does Marina Vannucci do when she is not busy writing grants and papers?*

Well, I must admit it: I am a workaholic! As I am sure you know, the academic profession can be very rewarding but it also very demanding. When I am not working I enjoy simple things that take my mind away: music, a dinner with friends, a bike ride and running around with my latest toy, a (fully) electric scooter!

Thanks to Marina, for the stimulating discussion and for her scooter-related bravery. (Electric scooters in Houston? What will come next a gay mayor?).

ANNOTATED BIBLIOGRAPHY

PRIOR DISTRIBUTIONS FOR COVARIANCE/PRECISION MATRICES, PART I

Helene Massam
Department of Statistics, York University,
Tornonto M3J1P3, Canada.
Support for the author provided by NSERC grant A8947.
massamh@mathstat.yorku.ca

An accurate estimation of the covariance or precision matrix of a multivariate normal distribution is a central problem in the analysis of high-dimensional data: it is one of the most common task in statistical analysis but also a difficult task due to the large number of parameters in a covariance matrix and the fact that it is constrained to belong to a cone. In this issue we consider priors for full matrices; priors for graphical models will be considered next quarter.

We see in several early papers the fundamental ideas that will motivate most of the sub-

sequent work on the estimation of covariance matrices: shrinking the estimate of the covariance or precision matrix towards a structure, given or revealed by the data, keeping the eigenvalues in the range of the eigenvalues of Σ and reducing the number of parameters. From the Bayesian point of view, shrinkage can be achieved by choosing prior distributions for various parametrization of the covariance matrix. Standard parametrizations include the polar decomposition ODO^t where O is an orthogonal matrix and D the diagonal matrix of eigenvalues, the variance-correlation decomposition DRD^t and the Cholesky decomposition TDT^t where T is lower (or upper) triangular with entries function of the regression coefficients and diagonal elements equal to 1, D is diagonal with positive entries. With few exceptions, these distributions are derived from the Wishart (or equivalently inverse Wishart) distribution.

The Wishart distribution has qualities and defects. Its main deficiency is the fact that it has only one shape parameter to model the form of the prior. The priors on the various parametrization will therefore strive to introduce more parameters in order to shape the prior according to needs. The choice of these hyperparameters has an impact on inference and reference priors have also been developped to avoid arbitrary or inadequate choices of the hyperparameters. The Wishart's main quality is that the inverse Wishart is conjugate to the normal as a prior on Σ and the estimate of Σ can thus be obtained analytically. As a consequence, even though conjugacy is viewed as a purely mathematical convenience which might prevent us from properly expressing our prior knowledge, many of the priors presented in the literature strive to come as close to conjugacy as possible and thus minimize computationally expensive Monte-Carlo methods.

The following list of papers is by no means complete. Rather it is meant to illustrate the ideas outlined above. We concentrate on priors with broad applicability. There are many interesting priors for specific problems that are not mentionned below.

1 Some of the early papers

- Jeffreys, H. (1961) Theory of probability, Oxford university press. Jeffreys prior for the covariance matrix Σ is given. It is the measure on the cone P of positive definite matrices invariant under the group action $x \in P \mapsto \rho(g)(x) = gxg^t$ for g in the linear group.

- Geisser, S. and Cornfield, J. (1963) Posterior Distributions for Multivariate Normal Parameters, *JRSS B*, **25**: 368-376. They use a prior similar to Jeffreys prior but with a different power on the dispersion matrix Σ^{-1}.

- Villegas, C. (1969) On the *apriori* distribution of the covariance matrix, *Ann. Math. Statist.*, **40**, 1098-1099. This is a short note to give fiducial support to the usage of Jeffreys prior.

- Chen, Chan-Fu (1979) Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis *JRSS B*, **41**: 235-248. One of the first papers to use the conjugate inverse Wishart as a prior for Σ and choose a scale and shape hyperparameters derived from the data. The Bayes estimate is given explicitly as the convex combination of the sample covariance matrix and the estimate of the scale parameter of the inverse Wishart prior.

- Diaconis, P. and Ylvisaker (1979) Conjugate priors for exponential families, *Ann. Statist.*, **7**, 269-281. This paper gives a characterization of the conjugate family for the canonical parameter of a natural exponential familiy. This is work with a more general scope.

- Dickey, J. M. and Lindley, D. V. and Press, S. J. (1985) Bayesian estimation of the dispersion matrix of a multivariate normal distribution, *Commun. Stat. Theory Methods*, **14**, 1019-1024. Though published in 1985, according to the authors, this work was done in 1975. It is one of the first papers to shrink towards a structure and to be concerned about the choice of this structure.

2 Priors for full matrices

- Leonard, T. and Hsu, J. S. J. (1992) Bayesian inference for a covariance matrix, *Ann. Statist.*, **20**, 1669-1696. This paper radically moves away from the Wishart. The precision matrix is written as $\Sigma^{-1} = \exp A$ so that the entries of the symmetric matrix A

are unconstrained. A multivariate normal prior distribution is then proposed for the vectorized version of A. Since the orthogonal matrices in the spectral decomposition of A and Σ^{-1} are the same, the transformation amounts to taking the logarithm of the eigenvalues of the precision matrix. Computation of the estimates is done through importance sampling.

- Yang, R. and Berger, J. O. (1994) Estimation of a covariance matrix using the reference prior, *Ann. Statist.*, **22**, 1195-1211. The proposed prior (which was previously derived in a work of Chang and Eaves, 1990) is the reference prior on the monotonically ordered eigenvalues of Σ. This prior puts more mass where the eigenvalues are close, thus attempting to correct the traditional distortion of the eigenvalues of the estimate. The Bayes estimates are derived and computed with the hit-and-run sampler. Risk levels compare favourably with those of decision theoretic estimates of Stein' and Haff's.

- Brown, P. J. and Le, N. D. and Zidek, J. M. (1994) Inference for a covariance matrix, *Aspects of Uncertainty: A Tribute to D. V. Lindley*, Wiley, Chichester. Suppose that Y is a Gaussian $N(0, \Sigma)$ vector which is split into blocks as $Y = (Y_{q_1}, \ldots, Y_{q_k})$. Suppose Σ is split accordingly and follows an inverse Wishart distribution. Then, writing $r_i = \sum_{j=i}^{k} q_j$, for $i = 1, \ldots, k$, Σ_{r_i} admits a Choleski decomposition $T_i^t D_i T_i$ where T_i is lower triangular with diagonal elements equal to 1 and D_i is diagonal. The elements of the decomposition are $(\Sigma_{q_i, r_{i+1}} \Sigma_{r_{i+1}}^{-1}, \Sigma_{q_i \cdot r_{i+1}}, \Sigma_{r_{i+1}})$ such that $\Sigma_{r_{i+1}}$ is independent of $(\Sigma_{q_i, r_{i+1}} \Sigma_{r_{i+1}}^{-1}, \Sigma_{q_i \cdot r_{i+1}})$ with both $\Sigma_{q_i \cdot r_{i+1}}$ and $\Sigma_{r_{i+1}}$ following Wishart distributions and $\Sigma_{q_i, r_{i+1}} \Sigma_{r_{i+1}}^{-1}$ given $\Sigma_{q_i \cdot r_{i+1}}$ following a Gaussian. The authors exploit and generalize this property of the inverse Wishart. They obtain a very flexible generalized inverse Wishart distribution by keeping the independence as for the inverse Wishart but taking different shape hyperparameters for $\Sigma_{q_i \cdot r_{i+1}}$, $i = 1, \ldots, k$ and allowing different scale hyperparameters for all the parameters while preserving conjugacy. The parameters are clearly the regression coefficient and covariance in the conditional distributions of Y_{q_i} given the succeeding $Y_{r_{i+1}}$. This paper can be seen as the seed paper for all generalized Wisharts or inverse Wisharts cited below.

- Daniels, M.J. and Kass, R.E. (1999) Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models, *J. Stat.Am. Ass.*, **94**, 1254–1263. The authors have the interesting idea to use the ODO^t parametrization and then further parametrize O with its Givens angles. A normal prior is then put on something like the z-transform of the angles and a flat prior on the eigenvalues. The induced prior on the covariance has the factor $\prod(\lambda_i - \lambda_j)$ in the denominator and thus brings the eigenvalues of the posterior mean closer together than the eigenvalues of the sample covariance matrix. Sampling from the posterior is computationally very heavy.

- Barnard, J. and McCulloch, R. and Meng, X. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage, *Statistica Sinica*, **10**, 1281-1311. Uses the variance-correlation decomposition DRD, chooses to put a normal prior on $\log(S)$ because the chi-square distributions derived from the Wishart only allows for one shape parameter and are too restrictive to express prior knowledge. The prior distribution of R given S aims to be noninformative. It is interesting to note that one of the proposed distributions for R is derived from the Wishart. They use the Gibbs sampler to sample from the posterior. One must make sure that the correlations generated are positive definite.

- Daniels, M.J. and Pourahmadi, M. (2002) Bayesian analysis of covariance matrices and dynamic models for longitudinal data, *Biometrika*, **86**, 677-690. The parameters are the elements of the Choleski decomposition $\Sigma = T^{-1} D T^{-1'}$ where T is lower triangular with diagonal elements equal to 1. The advantage of this Choleski decomposition is that the off-diagonal elements of T on line t are precisely minus the regression coefficients of Y_t on $(Y_1, \ldots, Y_{t-1})$. The t-th diagonal element of the diagonal matrix D is the conditional covariance of Y_t given the preceding components. The au-

thors put a Gaussian prior on ϕ, the set of off-diagonal elements of T, and independent inverse Gammas with various shape parameters on the diagonal elements of D. Moreover ϕ and D are assumed independent in the prior. This approach does not allow for conjugacy as in Brown et al. (1994). It leads to a posterior that has to be computed with a Gibbs sampler, alternatively drawing from a normal given D and from inverse Gammas given ϕ. The computational burden is much lighter than in Daniels and Kass (1999) but the conjugacy property of the inverse or generalized inverse Wishart is lost.

- Consonni, G. and Veronese, P. (2003) Enriched conjugate and reference priors for the Wishart family on symmetric cones, *Ann.Statist.*, 31, 1491-1516. Using the same notation as for Brown et al. (1994) and setting $a_i = \sum_{j=1}^{i} q_j$, the parameters used here are those of the conditional distribution of Y_{q_i} given $Y_{a_{i-1}}$ that is $(\Sigma_{q_i,a_{i-1}}\Sigma_{a_{i-1}}^{-1}, \Sigma_{q_i \cdot a_{i-1}})$, $i = 1, \ldots, k$ (that is the ϕ parametrization of Daniels and Pourahmadi, 2002) but they are expressed in terms of entries of Σ^{-1} rather than of Σ. Like Brown et al. (1994), they use the properties of the Wishart (equivalent to those of the inverse Wishart) and generalize them to have more hyperparameters. For real covariance matrices, the class of priors that are obtained is the same as that in Brown et al. (1994). However, these priors are given in the more general framework of symmetric cones and are thus valid for complex and quaternionic matrices. Reference priors are also obtained. The conjugacy property of these priors yields an analytic expression for the Bayes estimates of Σ, thus avoiding computationally expensive MC methods.

- Berger, J.O. and Sun, D. (2008) Objective priors for the bivariate normal model. *Ann. Statist.*, **36**, 963-982. A wide range of issues concerning objective inference is studied through the bivariate normal. This includes reference priors for various parameters constructed from the $(\mu_1, \mu_2, \rho, \sigma_1, \sigma_2)$. This is work with a more general scope.

BAYESIAN HISTORY

HIROTUGU AKAIKE, 1927–2009: A PERSONAL OBITUARY

by Wolfgang Polasek
Institut für höhere Studien, Wien
polasek@ihs.ac.at

Japanese friends coming to the JEuBES[1] and Econometric Society European meeting (ESEM) in Barcelona, August 2009 brought the news: Professor Akaike passed away August 4, 2009. It was in May in Spain 1979, just 30 years ago, at the first "Valencia" meetings of Bayesian statistics, where I first met Professor H. Akaike (call me "Hiro", he said). The unglobalized world was quite different then: Japan was an exotic country at the end of Asia, just starting to sell small cars in Europe. At Valencia I, Akaike was invited to present his ideas on AIC and BIC, average and Bayesian information criterion, as he called then. For those who know about time series, (t)his approach was considered to be a revolution. Now for the first time in statistics one could estimate the parameters and the complexity of a model in a simple way; in time series it was the lag order of the auto-Aregressive (AR) process up to a finite order. I was fascinated: how was it possible to come up with such an idea, I asked him. "It's the likelihood of a model you have to estimate", he replied. I knew the likelihood of the parameter of a model, but the whole model? And the BIC, what is that? "The prior is like a window you can view the world", he said, and you can estimate it. Estimating a prior leads to a criterion? I was totally confused.

Three years later I passed a tiny shop at the Japanese Information Center near the opera in Vienna. I wanted to get some information about Origami, since I wanted to teach it to my $6\frac{1}{2}$ year old son. On the way out I saw it: JSPS (Japanese Society for the Promotion of Science) was offering stipends to go to Japan. Why not, I reasoned, I still had not understood the derivation of AIC and BIC. And when I called the ministry, I found

[1] Japan-Europe Bayesian Econometrics and Statistics: www.ihs.ac.at/jeubes

that one last slot was left, but I needed to act quickly. So I wrote H. Akaike a letter (e-mail did not exist at that time), and then in February 1983 my biggest adventure started: sharing an office with Hiro for half a year. He even found my accommodation for the stay, which was within walking distance of the Research Institute of Statistical Mathematics[2] , located in the embassy district of Tokyo (next to the German embassy).

To get into the right mood I studied Japanese for six months. It was quite tough, since studying had to fit around my regular work, but it paid off later. Understanding the Japanese mentality starts with the language: Politeness, culture, history, distance, hierarchy and the gender differences, all is laid out there. Still, I remember my first walk to the institute; it was a walk on a different planet, despite my basic reading knowledge.

Japanese offices are spartan and so was Hiro Os office; small grey metal desks and book closets, full of books, obscuring the view, with the charm of a library storage. His office was simply two ordinary offices, where I occupied one part. Japanese like a flat hierarchy at offices, and prefer to express differences in their language when they have to talk. (Complicated, even for Japanese, so they prefer a neutral language like English for business.) Akaike could speak English perfectly, while most of the other researchers wrote English better than they spoke; a situation that has not changed much to this day.

Akaike had developed a program package in FORTRAN and everything had to be supplied by cards to the computer. The computer was located one floor down and Professor Akaike and his coworkers could frequently be met on the way there: carrying the punch cards or the printed output. Akaike had frequent visitors from industry, since he—as a real engineer—was very proud that his TIMSAC[3] programs could help business and the economy to save money, as his application to the rotary kiln showed or the steering system of big tanker ships (Akaike and Nakagawa, 1988) .

Akaike-sensei—that was the formal way to address him at these official occasions (not Hiro, of course). Sensei in Japanese means teacher, but the title expresses esteem and is a public tribute to someone from whom one can learn a lot. My origami teacher Yoshizawa was also called sensei. The lunch meetings at the secretary Os office at his department were unforgettable. You can order or bring the obento (lunch boxes) to such meetings held around a big grey metal table. Japanese have adapted metal equipment for their (and not Western) needs and use it in most (public) offices, for equality, fire and earthquake reasons. The constant "trembling earth" is rarely recognized by Japanese but lonely foreigners like me had to get accustomed to the trembling grey equipment. Akaike used these office lunches to get his views across, and all his disciples listened: because that was the reason to come here to the ISM and study with one of the most famous professors in Japan. Akaike often complained that his work was not always appreciated. Even if we had cheese and wine parties at the end of a week or went to the Japan style tatami first floor of "Edogawa", a former excellent old fashion Sushi restaurant across the "Hiroo" subway station that could be only reached by ladder without shoes, people listened and agreed, but did not express their own views. Except those wild impolite foreigners: "Please, please the boss", as I got reminded at times. A lesson, hard learned: Japanese politeness demands respect for hierarchies. Otherwise you lose face. Once I used such a meeting to find out the reading skills of my colleagues. From my language course I knew a Japanese compound word that even Japanese cannot decipher easily: Seamless gas pipeline pipes. Professor Akaike was the only one who could read it.

Mentioning the atomic bomb was taboo. Japanese want to be seen as a peace-loving nation, any face-losing talk about bad war-time news should be avoided. H. Akaike was stationed near Hiroshima as a student. He told me that they could see American airplanes high in the air and the Japanese aircrafts that could not reach them. He could see the fire of Hiroshima but mentioned that people did not know for a long time what was going on. At the end of my stay he sent me to Hiroshima and Nagasaki. At the Hiroshima peace park museum the front page of an English newspaper from August 1945 was displayed: "Allies won the battle of science: Atomic bomb dropped at Hiroshima". I am sure Akaike would agree that a battle of science is the last thing we want to see again.

The last time the scientific community saw

[2]ISM: Tookei suri kenkyu joo or short Too-su-ken in Japanese.

[3]There were four time series and control (TIMSAC) packages published over the years. A new R-package is available at: cran.r-project.org/web/packages/timsac/index.html. TIMSAC is now also available as a program package in R (CRAN).

him officially, was at his retirement conference at Mount Fuji in December 1992, close to his birth place Fujinomiya. A fine celebration after a scientific life devoted completely to the practical side of statistics: computing, modeling and simplification. If we have to come up with a list of people who have contributed to the view that statistics is an independent science, then Akaike should be on it. His paper on AIC was put into the hall of fame of the most cited papers in IEEE, see Akaike (1981). The interview in Statistical Science (Findley and Parzen, 1995) gives a nice summary of how he came to discover AIC. Akaike Os name is now attached to the term AIC and his outstanding contribution of extending estimation to the problem of model choice lives on. In 2006 Akaike was awarded the Kyoto[4] Prize for his major contributions to statistical science and modeling in the development of the "Akaike Information Criterion" (AIC).

Akaike means red pond. On my walks through the Asian mountains I just saw this red pond once in midst of red earth. And that stays in my memory: Professor Akaike was a quiet red Japanese pond in the landscaped garden of modern statistical science.

References

[1] Akaike, H. (1974). "A new look at the statistical model identification", *IEEE Transactions on Automatic Control* 19(6):716–723.

[2] Akaike, H. (1980). "Likelihood and the Bayes procedure", In *Bayesian Statistics*, Bernardo, J. .M., DeGroot, M. H., Lindley, D. V. and Smith, A. F. M., editors. 143–166. University Press. Valencia, Spain.

[3] Selected papers of Hirotugu Akaike (1998). Parzen, E., Tanabe, K., Kitagawa, G., editors.

[4] Akaike, H. and Nakagawa, T. (1988). *Statistical analysis and control of dynamic systems*, Tokyo, Dordrecht: KTK Scientific; London: Kluwer. (Translation of the 1972 book in Japanese.)

[5] Akaike, H. (1981). *This Week Os Citation Classic, December 21, 1981*, `http://www.garfield.library.upenn.edu/classics1981/A1981MS54100001.pdf`

[6] Findley, D. F. and Parzen, E. (1995). "A Conversation with Hirotugu Akaike", *Statistical Science* 10(1):104–117.

APPLICATIONS

Statistical Challenges in Spectral Analysis in High-Energy Astrophysics

Taeyoung Park
Department of Statistics
University of Pittsburgh `tpark@pitt.edu`

In recent years, tremendous advances in technology have enabled a revolution in the quantity and quality of X-ray data available to astronomers. As the most sophisticated X-ray telescope ever built, the *Chandra X-Ray Observatory* uses a highly sensitive high-resolution X-ray detector to observe much fainter stars and produce much sharper images than any previous X-ray telescope. While the resulting high-resolution data carry much information on stars in the universe, analysis of such data requires significant advances in data analysis methodology. In this article, I will address statistical challenges in high-energy spectral analysis and review some of the work that has tried to solve the challenges.

Spectral model

The first issue to discuss is modeling the energy spectrum of photons emanating from a star. The energy spectrum is the distribution of photon emissions as a function of energy, which gives insight into the composition, temperature, distance, and velocity of a star. Due to instrumental constraints, each high-energy photon arriving at the detector from a star is recorded in an energy bin that contains the energy value of the photon. The high resolution of *Chandra* means

[4]The Kyoto Prize (Kyōto-shō) has been awarded annually since 1985 by the Inamori Foundation. It is a Japanese equivalent of the Nobel Prize for outstanding works in the fields of philosophy, arts, music, science and technology. Rudolf Emil Kalman won the first prize in Electronics 1985, Karl Raimund Popper in 1992 and 2004 Jürgen Habermas in philosophy.

that the observed energy spectrum has a much finer binning, resulting in smaller counts of expected photons per energy bin. Such low-count data make traditional statistical methods based on Gaussian approximations unjustifiable– the arrival of photons is better modeled as an inhomogeneous Poisson process [12].

The energy spectrum is generally formulated as a finite mixture of the continuum term and several line profiles, where the continuum describes the general shape of the spectrum over the entire energy range of interest and a line profile is a local spectral feature added to the continuum. Specifically, the expected Poisson counts in energy bin j for $j = 1, \ldots, J$ are modeled by

$$\Lambda_j(\theta) = \Delta_j f(\theta^C, E_j) + \sum_{k=1}^{K} \lambda_k \pi_j(\mu_k, \nu_k), \qquad (1)$$

where θ represents the set of model parameters, Δ_j is the width of bin j, E_j is the mean energy of bin j, $f(\theta^C, E_j)$ is the expected number of photons per unit energy bin due to the continuum term at energy E_j, θ^C represents the set of model parameters in the continuum model, K is the number of line profiles, λ_k is the expected number of photons due to line profile k, and $\pi_j(\mu_k, \nu_k)$ is the proportion of line profile k with mean μ_k and standard deviation ν_k falling into bin j. While the continuum term is modeled by various smooth parametric functions in some bounded energy range, the line profile is modeled via the proportions of a Gaussian distribution, a Lorentzian distribution, or a delta function.

Statistical challenges

Here I briefly outline a number of statistical challenges that arise when analyzing spectral data from the X-ray detector on *Chandra* or any other high-resolution count-based detectors.

Data collection processes. The observed data are subject to non-ignorable stochastic processes involved in data collection and recording. First, a proportion of photons emitted from a star is absorbed by interstellar material between the star and the detector, and the probability of absorption varies with photon energy. Second, the effective area of the detector varies with the energy of a photon, so that photons with different energy values have different probabilities of being recorded. Figure 1 shows 1000 effective area curves in the calibration samples of [1], illustrating the nonconstant effective area. Third, due to instrument response, the energy of a photon is blurred across a region of the detector and the blurring probability depends on photon energy. Figure 2 presents an example of a blurring matrix $M = \{M_{lj}\}$, where M_{lj} is the probability of a photon arriving with energy corresponding to bin j being recorded in detector channel l and brighter pixels indicate higher probabilities. Fourth, observed photons are contaminated by background photons originating somewhere other than the star of interest. Last, photon pileup occurs when more than one photon arrives at the same region of the detector within the same time frame, thereby creating a single arrival with energy equal to the total energy of coincident photon arrivals.

Narrow line profiles. The physical nature of a line profile is sometimes considered to be narrow, so that it can be contained in a single energy bin with relatively low-resolution X-ray detectors. In this case, a delta function is typically used to model the narrow line profile. When it comes to the high-resolution X-ray detector on *Chandra*, such a narrow line profile may occupy a few energy bins, and thus it is modeled with a narrow Gaussian distribution. Fitting a narrow line profile modeled with a delta function or a narrow Gaussian distribution involves a multimodal likelihood which causes significant statistical challenges [8].

Model selection. Testing for an additional model component in a spectral model is of scientific interest. To do so, we may compare two models, where the null model is nested within the alternative model, based on a likelihood ratio test (LRT) statistic or the related F-test statistic. However, significance tests such as the LRT or F-test cannot be used to test whether the null value of any additional parameter is on the boundary of the parameter space. This is the case when testing for a line profile in a spectral model because the line intensity λ_k in (1) is nonnegative and its null value of zero is on the boundary of the nonnegative numbers. In this case, the LRT statistic is not even asymptotically distributed as a chi-square [9].

Calibration uncertainty. The effective area and blurring matrix of the X-ray detectors are subject to systematic, or calibration, uncertainty. Figure 1 shows 1000 nonconstant effective area curves evaluated by the instrument calibration of the X-ray detector on *Chandra*. In typical spectral analysis, one default effective area is randomly selected out of the calibration samples

and presumed known. Thus calibration uncertainty is usually ignored in practice or the calibration error is assumed to be uniform over the energy range of interest. Failing to account for the calibration uncertainty, however, can cause large systematic errors in statistical inference for model parameters.

Currently available solutions

The statistical challenges above have motivated the need for new statistical models and methods, and below I describe some of the currently proposed solutions.

Data collection processes. To account for instrumental effects and other aspects of the data collection procedure, [12] proposed a more realistic spectral model by modifying the model in (1) via

$$\Xi_l(\theta) = \sum_{j=1}^{J} M_{lj}\Lambda_j(\theta)d_j u(\theta^A, E_j) + \theta_l^B, \qquad (2)$$

where $\Xi_l(\theta)$ is the expected Poisson counts in detector channel l for $l = 1, \ldots, L$, M_{lj} is the blurring probability of channel l and bin j, d_j is the effective area of bin j, $u(\theta^A, E_j)$ is the probability that a photon with energy E_j is *not* absorbed, θ^A represents the set of model parameters in the absorption model, and θ_l^B is the expected Poisson counts of the background in channel l. Here pile-up is ignored, which can be justified for low-count data.

The data collection processes along with the finite mixture model in (1) can be addressed through a hierarchy of missing data by the method of data augmentation; pile-up can be addressed by adding another level in the hierarchy. To fit the resulting multilevel spectral model that includes a Gaussian line profile, [12] developed a Markov chain Monte Carlo (MCMC) sampler under a Bayesian perspective. Other applications of this model-based approach include the estimation of hardness ratios [5], Bayesian multiscale image analysis [2, 11], and reconstruction of the temperature distribution of matter in a stellar corona [11].

Narrow line profiles. Fitting the location of a narrow line profile requires new and more sophisticated computational techniques than those developed by [12]. For posterior optimization of model parameters in a spectral model that includes narrow line profiles, [10] devised new EM-type algorithms. The posterior simulation of the location (and width) of narrow line profiles is based on a nontrivial generalization of MCMC sampling known as partially collapsed Gibbs sampling [14, 7]. Partially collapsed Gibbs samplers tailored for fitting narrow line profiles were developed by [8].

Model selection. When testing the null model that is on the boundary of the parameter space, the LRT and F-test statistics do not adhere to their nominal chi-squared and F-distributions even asymptotically. As possible alternative methods to the LRT and F-statistics, [9] proposed the methods of Bayes factors, the Bayesian information criterion, and posterior predictive p-values. The method of posterior predictive p-values has been extended to test for the presence of broad line profiles [13] or narrow line profiles [8] in X-ray spectra.

Calibration uncertainty. Because uncertainties in instrument calibration can cause large systematic errors in statistical inference, some strategies have been developed in an attempt to incorporate calibration uncertainty into spectral analysis. The previously proposed methods are, however, based on Gaussian assumptions for systematic errors, which is not appropriate for high-resolution low-count data. To properly account for systematic errors in X-ray spectral modeling and imaging, general statistical strategies based on the multiple imputation and fully Bayesian methods were developed by [3] and [6].

Summary

Innovative advances in astronomical instrumentation have made it possible to collect high-resolution low-count X-ray data. The complexity of such data in terms of both the underlying spectral features and the data collection processes, however, requires not only new statistical models but also cutting-edge computational tools for data analysis. In this article I have outlined some statistical challenges arising from the advances in technology and currently available model-based Bayesian solutions to the challenges. The work introduced in this article belongs to the interdisciplinary area of astrostatistics that applies the tools of statistics to astronomical/astrophysical data. Readers interested in astrostatistics are encouraged to check out more articles in a recent special issue on statistics and astronomy in *The Annals of Applied Statistics* [4].

References

[1] J. Drake, P. Ratzlaff, V. Kashyap, R. Edgar, R. Izem, D. Jerius, A. Siemiginowska, and A. Vikhlinin. Monte Carlo processes for including Chandra instrument response uncertainties in parameter estimation studies. In *Proceedings of SPIE*, pages 6270–49, 2006.

[2] D. N. Esch, A. Connors, M. Karovska, and D. A. van Dyk. An image reconstruction technique with error estimates. *The Astrophysical Journal*, 610:1213–1227, 2004.

[3] V. Kashyap, H. Lee, A. Siemiginowska, J. McDowell, A. Rots, J. Drake, P. Ratzlaff, A. Zezas, R. Izem, A. Connors, D. A. van Dyk, and T. Park. How to handle calibration uncertainties in high-energy astrophysics. In *Proceedings of SPIE*, pages 7016–23, 2008.

[4] T. J. Loredo, J. Rice, and M. L. Stein. Introduction to papers on astrostatistics. *Annals of Applied Statistics*, 3:1–5, 2009.

[5] T. Park, V. Kashyap, A. Siemiginowska, D. A. van Dyk, A. Zezas, C. Heinke, and B. Wargelin. Bayesian estimation of hardness ratios: Modeling and computations. *The Astrophysical Journal*, 652:610–628, 2006.

[6] T. Park, V. Kashyap, A. Siemiginowska, D. A. van Dyk, A. Zezas, C. Heinke, and B. Wargelin. Bayesian estimation of hardness ratios: Modeling and computations. *CHASC Technical Report*, 2009.

[7] T. Park and D. A. van Dyk. Partially collapsed Gibbs samplers: Illustrations and applications. *Journal of Computational and Graphical Statistics*, 18:283–305, 2009.

[8] T. Park, D. A. van Dyk, and A. Siemiginowska. Searching for narrow emission lines in X-ray spectra: Computation and methods. *The Astrophysical Journal*, 688:807–825, 2008.

[9] R. Protassov, D. A. van Dyk, A. Connors, V. Kashyap, and A. Siemiginowska. Statistics, handle with care: Detecting multiple model components with the likelihood ratio test. *The Astrophysical Journal*, 571:545–559, 2002.

[10] D. van Dyk and T. Park. Efficient EM-type algorithms for fitting spectral lines in high-energy astrophysics. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (Editors: A. Gelman and X.-L. Meng), pages 285–296. Wiley & Sons, New York, 2004.

[11] D. A. van Dyk, A. Connors, D. N. Esch, P. Freeman, H. Kang, M. Karovska, V. Kashyap, A. Siemiginowska, and A. Zezas. Deconvolution in high-energy astrophysics: Science, instrumentation, and methods. *Bayesian Analysis*, 1:139–236, 2006.

[12] D. A. van Dyk, A. Connors, V. Kashyap, and A. Siemiginowska. Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal*, 548:224–243, 2001.

[13] D. A. van Dyk and H. Kang. Highly structured models for spectral analysis in high-energy astrophysics. *Statistical Science*, 19:275–293, 2004.

[14] D. A. van Dyk and T. Park. Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103:790–796, 2008.

Figure 1: Effective area curves evaluated by the instrument calibration of the X-ray detector on *Chandra*.

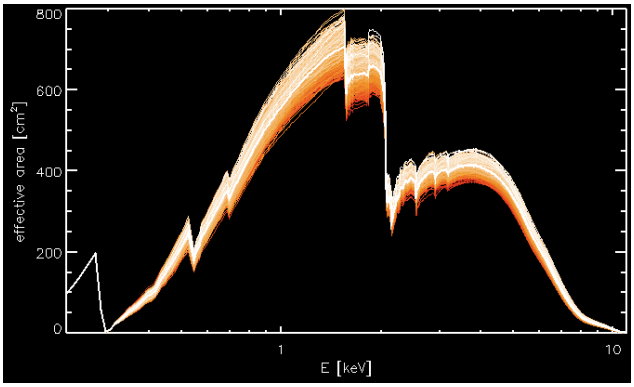
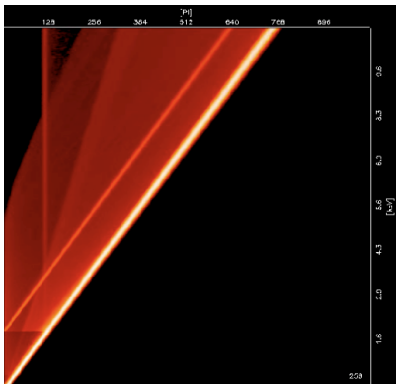


Figure 2: Blurring matrix of the X-ray detector on *Chandra*.



STUDENTS' CORNER

Luke Bornn
l.bornn@stat.ubc.ca

This Students' Corner features the dissertation abstract of recent University of British Columbia graduate Roman Holenstein. You can also see his doctoral work in an upcoming read paper in JRSSB. If you are newly graduated and would like to publish your thesis abstract, don't hesitate to contact me.

Dissertation Abstracts

PARTICLE MARKOV CHAIN MONTE CARLO

by Roman Holenstein
roman@holenstein.org
http://holenstein.org
Department of Computer Science
University of British Columbia
PhD Supervisor: Arnaud Doucet

MCMC and SMC methods have emerged as the two main tools to sample from high-dimensional probability distributions. Although asymptotic convergence of MCMC algorithms is ensured under weak assumptions, the performance of these latters is unreliable when the proposal distributions used to explore the space are poorly chosen and/or if highly correlated variables are updated independently. In this thesis

we propose a new Monte Carlo framework in which we build efficient high-dimensional proposal distributions using SMC methods. This allows us to design effective MCMC algorithms in complex scenarios where standard strategies fail. We demonstrate these algorithms on a number of example problems, including simulated tempering, non-linear non-Gaussian state-space model, and protein folding.

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.

ISBA 2010 World Meeting/Ninth Valencia International Meeting on Bayesian Statistics, Benidorm, Spain. 3-8th June, 2010.

Although, abstracts for contributed talks are no longer being accepted, submissions for the poster sessions will be accepted until February 20th, 2010 and potential participants are encouraged to do so. Submissions must follow a specific structure, available at http://www.bayesian.org/events/isba2010/abstructure.html, and should be sent by e-mail to isba10-poster@im.ufrj.br.

We hope to see you in Benidorm!

Events

2010 Bayesian Biostatistics Conference, Houston, Texas. 27-29th January, 2010.

Current and prospective users of Bayesian biostatistics are invited to join experts in the field for a three-day conference sponsored by the Department of Biostatistics at The University of Texas M. D. Anderson Cancer Center in Houston, Texas, USA. Attendees will have the opportunity to attend two courses on the first day of the conference (Wednesday): The Use of Bayesian Statistics in Clinical Trials, and Applications of Bayesian Methods to Drug and Medical Device Development. On Thursday and Friday, invited presentations will cover a variety of topics, possibly including comprehensive decision modeling; using predictive probabilities in clinical studies and drug development; roles for hierarchical modeling; how Bayesian methods can be used to augment traditional methods; Bayesian methods in epidemiology; the Bayesian approach and medical ethics; how to assure good quality and scientific rigor in taking a Bayesian approach; and guidelines for publishing Bayesian analyses. Registration fees will be modest. Program co-chairs: Donald A. Berry, Ph.D., The University of Texas M. D. Anderson Cancer Center, and Telba Z. Irony, Ph.D., Center for Devices and Radiological Health, U.S. Food and Drug Administration.

Information will be available at http://biostatistics.mdanderson.org/BBC2010/

Frontier of Statistical Decision Making and Bayesian Analysis, San Antonio, Texas. 17-20th March, 2010.

This conference consists of plenary, invited and poster sessions. Plenary speakers include Donald Berry, Lawrence Brown, Persi Diaconis, Stephen Fienberg, and Alan Gelfand. The conference will provide an overview of past, present and future developments of statistical decision making and Bayesian analysis. Prior to the conference, short courses on various statistical topics will be offered.

For more information visit the website, http://bergerconference2010.utsa.edu/.

10th Bayesian Statistics Brazilian Meeting, Green Coast, Brazil. 21-24th March, 2010.

The 10th EBEB will take place at Portogalo Suites Hotel, located in the pleasant Green Coast area of the State of Rio de Janeiro, Brazil. It is about 150km far from the city of Rio de Janeiro. In this 10th edition, we aim to discuss recent developments in the area both from the methodological and computational points of view. These developments will be presented and discussed by leading researchers in the world with a short course, 13 plenary talks, 8 oral presentations and 2 poster sessions. Since its 6th edition, EBEB is organized by ISBrA, the Brazilian chapter of ISBA, EBEB X is supported by ISBA.

Submissions and registration can be made at www.dme.ufrj.br/ebebx.

ISBA 2010 World Meeting/Ninth Valencia International Meeting on Bayesian Statistics, Benidorm, Spain. 3-8th June, 2010.

The ISBA 2010 World Meeting will be held in conjunction with the Ninth Valencia International Meeting on Bayesian Statistics in Benidorm, Spain. As already announced in Valencia 8, this will be the last Valencia meeting personally organized by José M. Bernardo (who will be 60 when the conference takes place). After Valencia 9, the Valencia meetings will become regular ISBA World Meetings (which will not necessarily take place in the State of Valencia). ISBA world meetings will therefore take place every two years. The meeting will be structured in tutorials on the first day, contributed talks in the late afternoons, and poster sessions in the evenings.

Additional information can be found at http://www.uv.es/valenciameeting.

2010 Summer Program on Semiparametric Bayesian Inference: Applications in Pharmacokinetics and Pharmacodynamics, Research Triangle Park, North Carolina. 12-23rd July, 2010.

The purpose of this program is to bring together a mix of experts in pharmacokinetics (PK) and pharmacodynamics (PD) modeling, nonparametric Bayesian inference, and computation. The aims of the program and workshop are (i) to identify the critical new developments of inference methods for PK and PD data; (ii) to determine open challenges; and (iii) to establish inference for PK and PD as an important motivating application area of non-parametric Bayes.

he program will begin with a week of tutorials and workshop activities. There will be extended, tutorial-style talks during morning sessions, and contributed and invited research talks during the afternoons. Afternoon talks will be selected to complement topics covered in the morning sessions. At the end of the first week workshop research working groups will be formed. The working groups will tackle particular research problems in the area. Working group activities can include workshop-style presentations by group members to stimulate discussion on specific issues

A detailed description of activities, along with application information is available at http://www.samsi.info/programs/2010bayes-summer-program.shtml.

Call for Pre-Proposals for Hosting the ISBA World Meeting in 2012

With the 10th World Meeting in Spain next year fast approaching, the ISBA Program Council is also beginning the early steps towards planning for our 11th World Meeting in 2012! We are now actively soliciting pre-proposals to host the 2012 meeting. Pre-proposals are due by February 1st 2010 to alex@im.ufrj.br and shall consist of:

- a paragraph on the proposed location and expected conference facilities;
- the names of members of a tentative local organizing committee, with the likely lead/chair local organizer identified;
- some mention of possible local/regional funding streams and tentative level of local funding expected/targeted.

Pre-proposals are likely to be quite tentative, and ISBA Program Council is ready to discuss and advise as groups begin to explore potential locations and opportunities. We hope to see several exciting potential venues proposed eventually, and will also work with local organizers in developing funding proposals and seeking funding internationally.
The Program Council will review these pre-proposals and invite full proposals from as many groups as the Council decides. The Program Council will comment on the full proposals and then submit them to the Board for their discussion. The Board and Council will decide an initial prioritization of proposals, weighing balance of geographical location (over a period of years, aiming to represent all active Bayesian regions worldwide), cost and accessibility. At the World Meeting in Spain in 2010, the potential sites will be discussed and represented at the Member Meeting, following which the Program Council will make a final recommendation for the site of ISBA11 to the Board, for Board approval.
Timeline:

February 1st 2010 - pre-proposals due
February 15th 2010 - full proposals invited
May 1st 2010 - full proposals due
May 15th 2010 - program council forwards comments to board
June 2010 - discussion at ISBA10 in Benidorm leading to final decision

For questions, please contact Alex Schmidt at alex@im.ufrj.br. Alex will be Program Chair in 2010 and responsible for moving the post-selection planning of ISBA11 along during the year.