THE ISBA BULLETIN

Vol. 14 No. 3 September 2007

The official bulletin of the International Society for Bayesian Analysis

A MESSAGE FROM THE PRESIDENT

by Peter Green
ISBA President
P.J.Green@bristol.ac.uk

Greetings to all ISBA members, and welcome to the first Bulletin of our new Editor, Raphael Gottardo. I am really grateful to Raph for taking this on, and I am confident he will do a great job. He has already recruited an exciting new team of associate editors, and I am looking forward to seeing and reading the results of their efforts in this and the following issues.

In the June issue, we reported that the new By-laws for awards and prizes were now in place, and I am pleased to tell you that the new overarching Prize committee has been appointed. The idea is to bring the four main awards administered by ISBA (DeGroot, Lindley, Mitchell and Savage) under a single umbrella, to provide more consistency and to help avoid conflicts-of-interest. The committee is chaired by Susie Bayarri, and the other members are Chris Carter, Phil Dawid, Ed George, Fernando Quintana and Marina Vannucci. Between them they have very broad interests, and a lot of experience of all the awards: thanks to all of them for agreeing to serve ISBA in this way. Nominations are currently open for the 2007 Mitchell Prize and Savage awards. *Continue in page 2.*

A MESSAGE FROM THE EDITOR

by Raphael Gottardo
raph@stat.ubc.ca

I will begin my first message by saying that I am extremely excited to be the new editor of the ISBA bulletin. As the new editor, not only will I do my best to continue the good quality of the bulletin as a source of information and exchange of ideas among the ISBA members, but I will try to make it even better. Note, however, that this will only be possible with great contributions from ISBA members and more generally, interested readers. So, if you have anything you would like to publish in the bulletin or any ideas on how to improve the bulletin, you should contact me or any of the associate editors. Contact information are available on the last page. Speaking of AE's, I would like to welcome and thank our new AE's: Donatello Telesca, Mayetri Gupta, Beatrix Jones, Alex Lewin, Tim Johnson, Luke Bornn, and Sebastien Haneuse. *Continue on page 2.* ▲

In this issue

WORDS FROM THE PRESIDENT, *Continued from page 1.*

You can read about how to nominate for awards, and many, many other things about ISBA on the website. Almost all back issues of the Bulletin can now be found there. Fresh content is being added nearly every week by the new Web Editor (and Executive Secretary) Robert Wolpert. He is enjoying this job so much that he would love to do more of it! If you have ideas for additional content for the site, please let him know, especially if you are willing to write or collate it yourself, of course. I would especially like to see more introductory material, to help to motivate and draw people in, and more pages of links useful to Bayesians.

Finally, the elections for officers and board members take place from 15 October to 15 November. Help to ensure that ISBA goes in the direction you want: read the election addresses later in this Bulletin, and don't forget to vote!▲

WORDS FROM THE EDITOR, *Continued from page 1.*

As my first, and perhaps most apparent contribution, I have tried to improve the presentation of the bulletin by adding colors and enhancing url links to allow for a better reading experience and easier navigation. Now you can jump directly to the article that most interests you, though I still hope you will read the whole bulletin. Note that many of the url links were already active in previous issues but were not highlighted in colors, which made it hard to know where urls were. In addition, we are currently working on other new improved features, which I am sure many of you will enjoy. You should hear more about these new features very soon ...

If you have any comments and/or suggestions about the presentation of the bulletin, please feel free to contact me.▲

ISBA ELECTIONS

2007 ISBA ELECTIONS

by Robert Wolpert
wolpert@stat.duke.edu

Biographical information for each of the candidates appears below. The candidates for president have also included statements about what they intent to accomplish. This information is also currently accessible on the ISBA web-site. The 2007 elections of future ISBA officers will take place electronically at the ISBA web-site from 15 October through 15 November. Instructions for voting will be emailed to all current ISBA members prior to the election.

President 2009 (President Elect 2008, Past President 2010)

Tony O'Hagan (U Sheffield, UK)
Statement: It's a great honour to be nominated for President of ISBA. The list of past presidents is imposing, and if elected I will do my best to uphold the high standard they have set. I have been a passionate advocate of Bayesian analysis since 1970, and here are the challenges where I hope to bring that passion to bear in the role of President. I want to see "Bayesian Analysis" become one of the top journals in Statistics; it currently doesn't have that feel for me. I want to see the ISBA world meetings working better as a mechanism to bring all our members together; that means getting bigger as well as more accessible - tricky! And I want to bring more members in from major application areas; despite our aspirations, we are perhaps not yet inclusive enough.

Mike West (Duke U, US)
Statement: As a member of the founding committee that established ISBA over fifteen years ago, I have been delighted with the development of the Society and its increasing role as a hub of professional activities for the Bayesian communities. ISBA has done much to advance the appreciation of Bayesian statistical science, especially in terms of international and interdisci-

plinary outreach. But, as an early teenage society, we have only just begun. Let us now look ahead to the next fifteen years... where do we imagine ISBA will be at its thirtieth birthday? What does and what should the membership want and expect from ISBA in the coming years? What paths should we be planning now in order to move the Society to a central, visible position within the broad field of statistical science rather than being regarded by some as representing only a small "sub-field" or "niche" area?

Beyond the intellectual and socio-professional community ISBA represents, its tangible activities are conference organisation, the new 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. If elected, I will focus leadership attention on:
1. Membership: Current paid-up membership is currently under 450. Active membership has been much higher (around world meeting times) and the current figure is woeful in the context of the expansion of Bayesian analysis over the last couple of decades. (The Bayesian section of ASA has over 1200 members, the curated Valencia email list over 1700). Systematising membership drives, developing connections with other professional societies, and improving recruitment of students and new researchers via university liaison are efforts to promote.
2. Connections to other societies: Visibility and membership will be enhanced by improved inter-connections with several of the leading statistical societies. This might involve increased endorsement and co-sponsorship of conferences and workshops, and initiating discussions about co-listings on membership renewals.
3. Organisation and funding: As a wholly volunteer organisation, ISBA is fragile in terms of institutional memory and long-term organisational stability. Ongoing financial organisation, including rolling fund-raising and grant generation for conference sponsorship, and especially support for participation of junior researchers at international meetings, will eventually require a longer-term dedicated strategy. With a much expanded continuing membership, ISBA will need planning to move towards a hybrid volunteer/permanent office model, either alone or via connections with other organisations.

Some of Mike's past contributions to ISBA: Member of the ad-hoc Founding committee that established ISBA; Past member of the International Advisory Board; Chair of ISBA 2000 Scientific Committee; Led the fund-raising campaigns to establish the Lindley and DeGroot Prize foundations, to expand the Mitchell Prize foundation, and to establish the three as ISBA administered awards.

Treasurer 2008-2010

Gabriel Huerta (U New Mexico, US)
Statement: Gabriel Huerta is currently Associate Professor and Regents Lecturer in the Department of Mathematics and Statistics at the University of New Mexico. His research interests include Bayesian time series, space-time models with environmental applications, extreme value modeling and parameter uncertainty estimation for climate models. He has published papers in JRSS(B), Applied Statistics, Journal of Time Series Analysis, Statistica Sinica, JCGS and JSPI. He has served as associate editor of the ISBA Bulletin and as a member of the nominations committee of ISBA. He served on the Program Committees for JSM 2006 and Cobal II. He has been a board member of the Albuquerque Chapter of the ASA.

Athanasios Kottas (UC Santa Cruz, US)
Statement: Athanasios Kottas (PhD, University of Connecticut, 2000). I am currently Assistant Professor of Applied Mathematics and Statistics at University of California, Santa Cruz. I am interested in the methodology and applications of Bayesian nonparametrics, including analysis of computer model experiments, population dynamics modeling, regression models and survival analysis. I have published papers in Scandinavian J. Stat., JSPI, JCGS, JASA and Biometrics.

Board of Directors 2008-2010 (4 openings)

Hedibert Lopes (U Chicago GSB, US)
Statement: I am Associate Professor of Econometrics and Statistics at the Chicago Business School. After graduating from Duke's ISDS in 2000, I returned to the Institute of Mathematics, Federal University of Rio de Janeiro, as Assistant Professor of Statistics. Since then I have taught several PhD-level courses on Bayesian statistics, ministered dozens of scientific talks and advised several graduate students.

My research interests includes spatial dynamic

factor models, nonlinear time series models, vector autoregressive models, multivariate mixture models, extreme value theory. I have published papers in the Biometrics, Statistica Sinica, Journal of the Time Series Analysis, Journal of Statistical Planning and Inference, Computational Statistics and Data Analysis. I co-authored the book MCMC: Stochastic Simulation for Bayesian Inference (2/e).

I served as ISBA Bulletin Editor (2002-2004). I co-founded Brazilian Chapter of ISBA (ISBrA) and launched its Bulletin. I served on the Savage Committee in 2006 and 2007. I humbly look forward to the chance of contributing to our growing community in such an honorable and important role.

Lurdes Inoue (U Washington SPHCM, US)
Statement: Lurdes Inoue obtained her PhD degree from Duke University in 1999. In the same year she joined the Department of Biostatistics at MD Anderson Cancer Center as a post-doctoral research associate. She joined the department of Biostatistics at the University of Washington in 2002 as an Assistant Professor. Her research interests are on Bayesian methods for biostatistics, more specifically, the design and analysis of clinical trials; models for disease progression; decision theory and cancer research. She has published papers in JASA, Biometrics, Biostatistics and TAS. She is also co-authoring a book on decision theory.

Caitlin Buck (U Sheffield, UK)
Statement: I am a professor in the Department of Probability and Statistics at the University of Sheffield with research interests in applied Bayesian statistics. I work mostly on applications in archaeology and palaeoenvironmental reconstruction, but am also interested in issues that impact on applied Bayesian work more generally including prior elicitation. My current research projects relate to developing models for: chronology construction for ice cores, estimating radiocarbon calibration curves, and the spread of domesticated cereals during the early neolithic in Europe. Work on the radiocarbon calibration curves forms the focus of an invited talk at the Ninth Case Studies in Bayesian Statistics workshop at Carnegie Mellon in October 2007.

I have published in a wide range of journals including: Applied Statistics, The Statistician, Bayesian Analysis, Antiquity, Quaternary Science Reviews and the Holocene. I have been an Associate Editor for Bayesian Analysis since its launch in 2006.

Havard Rue (Norwegian U Science & Technology, NO)
Statement: Havard Rue (PhD, NTNU, 1993) is currently Professor in Statistics at the Department of Mathematical Sciences, Norwegian University of Science and Technology, in Trondheim, Norway. His main research interests are computational and spatial statistics, but he also does (very) applied engineering type of research in ocean wave statistics. He has been an associate editor for JRSS(B), and is currently associate editor for Annals of Statistics, Scandinavian Journal of Statistics and Statistics Surveys. He has written a book on the "green book series" of Chapman & Hall with Leonhard Held, about Gaussian Markov random fields.

Marc Suchard (UC Los Angeles SM, US)
Statement: Marc Suchard (PhD, UCLA, 2002; MD, UCLA, 2004). I am currently an Assistant Professor in the Departments of Biomathematics, Biostatistics and Human Genetics at UCLA. My research interests cover stochastic modeling in biology and evolution, bioinformatics/molecular sequence analysis and biomedical data analysis, all of which I approach from a Bayesian perspective. My published papers have appeared in such journals as JASA, Biometrics, PNAS, Systematic Biology, British Medical Journal and Bioinformatics. In my work, I constantly strive to bridge the chasm between statistician and biologist and, to this end, serve as an Associate Editor for both the Annals of Applied Statistics and Systematic Biology, the top research journal in the field of evolutionary biology. ISBA has met my work with enthusiasm – my dissertation claimed the 2002 Savage Award and a recent paper garnered the 2006 Mitchell Prize – and I look forward to repaying this support.

Tony Pettitt (Lancaster U, UK & Queensland U Technology, AU)
Statement: I currently hold a position in applied statistics at Lancaster University UK and additionally at Queensland University of Technology, Brisbane, Australia. I have long had research interests in Bayesian statistics (a paper of mine has the words "posterior probabilitie" in the 1981 volume of Biometrika). From 1989 I helped establish a research profile in statistics at QUT with a strong emphasis on Bayesian statistics in a coun-

try, Australia, then not known for research in this area. My contributions in the last few years have been in applied Bayesian statistics in the areas of spatial statistics, infectious diseases (especially hospital pathogen transmission) and neurology. A jointly authored paper, published in Applied Statistics, with neurologists and other statisticians on motor unit number estimation using RJMCMC was read to the Royal Statistical Society in November 2006 whilst my other recent Bayesian work has been published in JRSSB, Biometrics, Biostatistics, Biometrika, J Theoretical Biology, and J Royal Society Interface. I was one of three co-editors of Biometrics for two years, 1999-2001. I am on the organising committee for ISBA2008 and I organised a Bayesian invited session at IBC2006.

Sylvia Frühwirth-Schnatter (Johannes Kepler U, AT)
Statement: Since 2003 I have been Professor of Applied Statistics and Econometrics at the Johannes Kepler University in Linz, Austria. After obtaining my PhD (Mathematics, TU Vienna, 1988) I held various research positions in Vienna at the Technical University and the Vienna University of Economics. During these 20 years I have introduced collaborators from such diverse areas as economics, finance, hydrology, marketing and road safety research to the Bayesian approach.

My research interests include MCMC methods, mixture modelling, Bayesian econometrics, and times series analysis using Markov switching and state space models. I have published papers in Biometrika, JASA, JBES, JRSS(B), Journal of Applied Econometrics, and Journal of Time Series Analysis. In 2006, I finished a book on Finite Mixture and Markov Switching Models which appeared in Springer Series in Statistics. Currently I am Editor of Statistical Papers, AE of Journal of Econometrics and a member of the Program Committee for the ISBA 2008 world meeting.

Sonia Petrone (U Bocconi, IT)
Statement: I am currently associate Professor of Statistics at Bocconi University (Milano, Italy), with a qualification (idoneitÃ) as Full Professor since 2002. My interest and enthusiasm for Bayesian Statistics arose from studying the work of de Finetti (as an undergraduate and in my PhD (1989)). My main research areas are now in Bayesian nonparametrics, mixtures and latent variables models, dynamic models. I have published papers in JRSS(B), Scandinavian J. Stat., Canadian J. Stat., Stat. Prob. Letters, Metron. I am co-authoring a book on dynamic linear models with R. I was member of the ISBA Board in 2002-2004. I have been in the scientific and organizing committee of several international conferences, including the series of workshops on Bayesian Nonparametrics and on Bayesian Inference for Stochastic Processes (BISP). I would be pleased and honoured if my work experience could be again a useful service for ISBA as a member of the Board.

INTERVIEW

DON BERRY

by Donatello Telesca
telesd@u.washington.edu

Donald Berry (better known as Don) is head Division of Quantitative Sciences, and Chairman or the Department of Biostatistics at the University of Texas, M.D. Anderson Cancer Center. He previously served on the faculty at the University of Minnesota and at Duke University, where he held the Edger Thompson Professorship in the College of Arts and Sciences. Dr. Berry's research has dealt with the theory and applications of statistics, especially Bayesian statistics, and particularly that dealing with the sequential design of experiments. His recent research has focused on the design and analysis of clinical trials and developing models in statistical genetics, modeling the relevant benefits of interventions and treatment, and medical decision making.

1. *While preparing an interview with Don Berry, It is almost impossible not to notice your extensive involvement as a proponent of Bayesian clinical trials. What do you think is the main advantage of the Bayesian choice in a design framework?*

That's easy. The Bayesian approach enables on-

line learning. It's the very model of the way humans think. It's why we have evolved into the Earth's dominant creatures. The people alive today are those whose ancestors' thought processes were closer to that of Bayes rule. Make an observation and update what you know on the basis of that observation. Decide what observation you want to make next with a clear understanding of the associated uncertainties. An important aspect of this decision is assessing the consequences of the various possible futures when following any particular sampling strategy. Regarding clinical trial design, predictive probabilities based on one's current information are of critical importance, and there is no satisfactory way to find predictive probabilities outside of the Bayesian paradigm.

Just to follow-up on the evolutionary aspects of the Bayesian approach. You'd think from what I said that "survival of the fittest" dictates that Bayesians will take over statistical thought. Not necessarily. On average, Bayesians don't understand the appropriate use of Bayes rule any better than do non-Bayesians! Sure, Bayesians condition on "the data". The problem is that they don't really understand what is "the data". The simplest example may be that part of "the data" is the fact that they are looking at the data. Frequentists are not wonderfully adept at handling this problem, but their statistical philosophy gives them a bit of protection from what might be called "the curse of silent multiplicities."

2. *When you arrived at M.D. Anderson, how did the medical community react to the "cultural revolution" you brought to the world of clinical trials design?*

I was lucky. My predecessors at M.D. Anderson were highly respected statisticians, so I too was treated with respect. Regarding the Bayesian approach, Peter Thall was one of the statisticians at M.D. Anderson when I arrived and he had introduced some Bayesian ideas into the language and practice of clinical research. So the seeds had been sown before I arrived. And, of course, clinicians are naturally Bayesian.

In retrospect, it was important in my first year that I accepted appointment to M.D. Anderson's Institutional Review Board. I learned the culture and I was able to work effectively in modifying it. The IRB members came to believe, rightly or wrongly, that I knew what I was doing. I had a compelling story to tell, but I had learned the hard way that having a compelling story is not enough. (A statistician at the FDA once told me that every time he heard me give a talk he became a Bayesian for 10 minutes!) What mattered was more than a compelling story and more than intuitive appeal. What mattered was trust, and gaining the confidence of the clinicians in the institution, from the President on down. "The rest is history," as the saying goes. If you ask an MD at our institution about our statistical approach, you'll no doubt get the answer, "We're Bayesians here".

3. *Going back to the origins, I will ask you a popular question. How did you decide to become a statistician? (I know you are going to have an interesting story here).*

Oh goodness, it's not very interesting. To tell it I have to go way back. My family was poor and I went to a small high school. I aspired to be a farmer-maybe I missed my calling! (Actually, earlier on I had wanted to be a priest, but that was before I discovered girls!) Then I read that PhD mathematicians had annual salaries of 12,000. (At the time, Coke from vending machines cost a nickel and first-class postage was 3 cents.) I figured that was for me. I went to Dartmouth on a full scholarship. My Freshman advisor was John Kemeny, clearly one in a billion on the scale of brilliance. (He was a mathematician/probabilist par excellance but is best known to some as a co-inventor (with Tom Kurtz) of the BASIC programming language, to others as head of the Three Mile Island Commission, and to others as a President of Dartmouth.) I was a hick. In high school I thought I was smarter than my teachers. But college was different. I had never been in the presence of an intellect like Kemeny. It was a head-wrenching experience. Further, even my fellow students seemed a lot smarter than I was, and they were infinitely more sophisticated. Anyway, I learned how to play bridge and poker, which was much more fun than studying and attending classes. I proceeded to flunk out. A stint in the Army built back my confidence and my resolve. I returned to Dartmouth, having saved up enough money (from untaken Army leaves) for one quarter's tuition. I managed to win back my scholarship after the first quarter. By then I had 3 children to support and so I worked full time as a bartender in addition to carrying a full load of courses. I did well, majoring in math. I was not turned

on by Lie algebras and non-Euclidean geometries. But I liked probability. One of my advisors suggested graduate school in statistics. I said, "What's that?" He said he wasn't sure himself, but knew that it had something to do with probability. Having nothing better to do, I took my first statistics course as a senior, and after I had applied to graduate schools. So you see, the short answer to your question is that my choice of vocation was random.

4. *During your studies at Yale, you had the chance to work with Leaonard Savage. Is there a particular story you would like to share with us, about this extraordinary mind? How was it working with him?*

Amazing! Jimmie Savage was brilliant beyond all imagination. A towering, towering mind. I was young and naïve, but by now I was used to being intimidated. I treasured opportunities for repartee with Savage. The atmosphere tingled in his presence. When you were wrong he would gently walk you down your path until you realized that you had fallen over a cliff!

Stories? Lots of stories. They're all neat. But no particular one conveys the total persona. Here's a story that speaks to the breadth of Savage's knowledge, as well as to his ability to teach, and it involves another statistical and intellectual giant: Frank Anscombe, who was the department chairman at Yale. Remember the 3 children? Well, when we showed up at Yale, Donna was pregnant with our fourth. After 3 boys she wanted a girl. Her obstetrician told her that she was "due" for a girl. Now I was statistically naïve, but I wasn't that naïve! At the fall department picnic I queried Professor Anscombe about the issue-pun intended! I asked him what I should tell Donna was the probability of a boy. I said I realized the MLE was 1, but this didn't make sense. He explained that if one assumes a uniform prior distribution then Laplace's rule of succession applies to give $4/5$. Since this prior distribution is probably too conservative, this is a rough upper bound. I had no idea what he was talking about! A few weeks later I was chatting with Professor Savage and recounted my discussion with Anscombe. Savage showed me how to use data to come up with a credible prior distribution. But the amazing part of this story is that he went to a shelf and pulled out a book by Corrado Gini (of Gini coefficient fame). The book had been published in the early 1900s and contained an incredible compilation from different countries of sex distributions in hundreds of thousands of families of size 1, 2, on up to 20 or so. Especially in the larger families there was an obviously greater variability than in any binomial distribution. Savage explained that an appropriate (empirical) prior distribution was the mixing distribution over the binomial parameters that gave the best fit to the data.

Bringing Professor Anscombe back into the picture, a few months later when we were getting close to observing the result of the experiment at issue(!). Anscombe asked me how I was going to pay for the baby. My attitude has been to not worry about such mundane matters! I had a generous fellowship, but what with our family's size we had to stretch to make ends meet. I told him I didn't know. He said, "When you get the bills, bring them to me". I did. And he paid them. The rest of the story is that the baby was a boy (Scott, whom you mention in your next question), but our fifth and sixth children were girls.

5. *If I am correct after the death of Leaonard Savage, you continued your work with Jay Kadane. I was once told that by the end of my dissertation I would learn to hate my advisor. Then I noticed that your son Scott Berry graduated with Kadane as well. Was that tough love?*

No, I love Jay. Always have. But you're not quite correct. Savage and Kadane were my co-advisors, with Savage being senior and primary. I had finished my dissertation before Savage died. I've never had anything but the fondest of thoughts for both of them. Regarding Savage, my admiration and affection for him never diminished. However, he made me work hard, putting me through 5 full dissertation drafts. The first one was 10 pages long. I thought it was neat, and so did he, or so he said. He indicated that there weren't many 10-page dissertations, but what I had done sufficed. "But," he said "let's try to do more". And more. And still more. I remember some frustration at the time! The final version was 70 pages long. So polished was it in the end that I simply cut out a couple of examples and submitted it as is to The Annals of Mathematical Statistics, where at a published 27 pages it became one of the longest of Annals articles. (The Associate Editor was still another giant: Tom Ferguson.) I learned a lot from both Savage and Kadane. And I've imitated them in my advising. I hope my students regard me with a fraction of the respect with which I hold them

both.

6. *Coming back to the present and speaking of Scott Berry, is he also a Bayesian? Do you collaborate with him?*

Both yeses. Scott and I have a company, Berry Consultants. He's full time and is the main thinker and doer in the company. We specialize in Bayesian approaches to clinical trial design and analysis, especially the former. We work with virtually all the major drug and medical device companies and with many small biotechs as well. We exploit the on-line learning aspect of the Bayesian approach to build efficient clinical trials-we call them "adaptive". Scott has designed more adaptive Bayesian clinical trials than anyone else on Earth. In part because of our activities, this headlined the lead article in a recent Japanese pharmaceutical industry newsletter (in Japanese, of course): "The Coming Bayesian Tsunami in Clinical Drug Development."

7. *M. D. Anderson has been rated the number one Cancer Center for a number of years and recently number one place for a post-doc. What do you think should be emphasized in the training of the new generation of Statisticians and Biostatisticians?*

Thanks for asking this, Donatello. It is one of my favorite subjects. Newly minted PhDs are green-present company excluded, I'm sure! I admit to being green long after I got my PhD. In the area of clinical trials I remember being told by biostatisticians of the day that I was naïve and did not understand clinical research. I bristled at the suggestion. My retort (not always voiced!) was that traditional clinical research was in a rut, that they had only one way to do things and that way happened to be terribly inefficient, ineffective, and it resulted in delivering inferior treatment to many patients, those in the trial as well as those who followed. Well, it turns out that we were both right. To be an effective innovator one must understand the culture and the rationale of what one is trying to change. I did not have the background to really understand. Since no one was listening to what I had to say, I reversed fields and immersed myself into the culture of clinical research as it existed-"If you can't beat 'em, join 'em!" I was able to achieve a high level of credibility in the traditional world of medicine. That made changing it a lot easier. I was no longer regarded to be from the lunatic fringe.

Now to your question: one size does not fit all. One's training should depend on one's goals. I'll address the young statistician who wants to have a real impact, let's say, in medicine. Research MDs specialize, and so too should statisticians. We must learn the substantive field of our application. My main area of application is breast cancer. I've worked hard to learn its biology and its treatment. (I am proud to have once been introduced to an audience of several hundred breast oncologists as "one of two statisticians in the world who could walk into a breast clinic and actually treat patients.") If I put an array of data in front of you and you say you can analyze it, you're wrong. As much as I hate the statistician-as-mechanic analogy, I'll use it here: this would be like turning a bolt on a car engine without knowing its location on the engine or the make of the car. Frequentists have an excuse for making this kind of error, Bayesians do not. But Bayesian make it at least as often as do frequentists.

Like so many of my attitudes, this one comes directly from Jimmie Savage. He promoted mastering the science before doing any statistical analysis. Once when we were involved in a project I was explaining to him some of the science. He asked where I had learned so much about the subject. I told him that my source was the scientific encyclopedia in the department library. He was pleased as punch because he had lobbied for the purchase of said encyclopedia, and he used this example as a justification for having done so.

Once you've learned the science of your specialty, start worrying about big things. Worry about things that will really make a difference, say, in the treatment of patients. Make a reputation for yourself in the substantive field. Emulate R.A. Fisher. We think of him as a statistician who dabbled in genetics. Geneticists think of him as a geneticist who dabbled in statistics!

8. *In your career you always managed to put together theoretical research with an impressive involvement in scientific panels, which brought you very often in the public eye. What are the pros and cons of this talent?*

I'm frequently asked related questions by statisticians. For example, I've been asked to address this topic in the ENAR President's Invited Address at the 2008 meetings. How is it that The

New York Times quotes me so often, including in a feature called Quotation of the Day? The same for many other newspapers and interviews on major TV news programs? The subjects range from cancer screening to Mad Cow Disease to the validity of death counts in Iraq. I'd like to say that this requires talent, but it doesn't. The interviews started when I became involved in the so-called mammography wars. Breast cancer is a highly political disease in which otherwise rational views become major points of contention. I was one of a few people who questioned the evidence concerning the value of screening mammography. I testified at a U.S. Senate hearing regarding this question, which further contributed to putting me in the public eye.

The principal pro of being in the public eye is that it makes life interesting, as does controversy more generally, at least for me. It's fun! Another pro is that it's good for the ego when people actually listen to what you say. The only con is that talking to the press is time consuming. Yes, I've received death threats because of my views, but I take this as more positive than negative: it suggests that my existence matters because some people wish I didn't have it! Oh, and I've experienced ad hominem attacks. Having a tough skin is essential. In one example someone e-mailed an especially vitriolic and personal attack to hundreds of respected researchers. I sent a simple reply to all, thanking the author for not including my mother in the "to" list!

As it turns out, my long-held views concerning screening mammography (and screening more generally) are becoming almost mainstream. The status quo has been for caregivers to tell women over 40 years old that mammography is a sine qua non for good health. I have pushed instead for the medical establishment to encourage caregivers to inform women of the risks and benefits of screening mammography, helping them make their own personal decision. Some major medical organizations are now saying exactly the same thing.

9. *One final question. What are your views about ISBA and what if any changes you would like?*

It's interesting that you ask this question. Readers may surmise that it is a set-up. But I had no idea you were going to ask it. I was opposed to forming ISBA. My arguments to Arnold Zellner, Jim Press, and others were that the Bayesian philosophy was but one statistical attitude and not a separate discipline. I worried that ISBA would isolate us from the mainstream of statistical thought. I worried that we'd have Bayesian publications that would be regarded as second-rate statistics journals by our peers. I'll let others decide whether any of this has come to pass. I do know that Bayesian biostatistics has been largely immune to any of the effects of ISBA, whether positive or negative. ISBA has ignored us and we have ignored it. Our papers are published in traditional biostatistics journals such as Biometrics and Biometrika. Personally speaking, publishing in Bayesian Analysis is not something I would consider doing. Incidentally, I've never liked the restriction "Bayesian Analysis" in the name ISBA. Similarly for the journal. In biostatistics the benefits of Bayesian design are at least as important as the benefits of Bayesian analysis. Bayesian Analysis and Design may sound bad but this is a case in which ISBAD is good!

Thanks to Don for his stimulating and entertaining answers.

BAYESIAN HISTORY

A Call for Contributors

by Timothy D. Johnson
tdjtdj@umich.edu

The editor of the ISBA Bulletin and I would like to make a general call for contributors to the Bayesian History section of this bulletin. We would like to propose that this section be devoted to the historical development of Bayesian theory and methods in particular countries—yours, for example.

This one to two page article could consist of an overview of the development of Bayesian methods in your particular country, a synopsis of the contributions to a particular aspect of Bayesian methods/theory that have taken place in your country or by fellow countrymen or an interview with a prominent statistician.

We are also open to suggestions and articles pertaining to the general history of Bayesian statistics. So, if you have an idea, or wish to contribute to this section of the ISBA Bulletin, please contact me.

APPLICATIONS

A Call for Contributors

by Mayetri Gupta
gupta@bios.unc.edu

The editor of the ISBA Bulletin and I, the Applications section AE, would like to make a general call for contributors to the Applications section of this bulletin. We hope that this section continues to be devoted to new and exciting application areas of Bayesian statistical methods in the sciences, engineering and social sciences.

The one to two page article could consist of an overview of the development of Bayesian methods in a particular application area, a synopsis of contributions to a particular aspect of Bayesian methods in this area, or to introduce a new area where Bayesian statistics has made a recent impact. Areas that have been highlighted in the past have a diverse range, just to give a few examples, there have been articles on oceanography and climate change, fossil records, market forecasting, medical imaging, clinical trials, high-dimensional problems in genomics, as well as online gaming! The idea is to give our readers a general idea about the topic and point them to research articles where they can learn more.

We are also open to other suggestions and articles pertaining to applications of Bayesian statistics. So, if you have an idea, or wish to contribute to this section of the ISBA Bulletin, please contact me.

ANNOTATED BIBLIOGRAPHY

Bayesian Parentage Analysis

Beatrix Jones
m.b.jones@massey.ac.nz

I am the new associate editor for the annotated bibliography section of the ISBA bulletin. For this issue I've taken the opportunity to write a bibliography about one of my research areas. However, before launching into that I'd like to mention my hopes and dreams for the annotated bibliography section. One of these, suggested by Jim Pitman, is to include clickable links in the electronic version of the bibliography, which appears at http://www.bayesian.org/bulletin/bulletin.html. We will link to the abstract, to the full text whenever possible, and provide bib-

TeX entries. Jim has generously offered to lend his expertise to the project, so I hope to start including some of these features with the next issue. My second dream is that many Bayesian statisticians who are keen to introduce their area of research for to the wider Bayesian public will write to me at the email address above and offer to provide annotated bibliographies; but being a realist I also welcome suggestions of the form "topic X is all the rage these days and I would like to learn more about it; statistician Y is an expert in topic X, perhaps you could get her to write an annotated bibliography."

On to Bayesian Parentage Analysis! Parentage analysis is the process of using genetic information to match offspring with candidate parents, or, more broadly, reconstruct family structures (including reconstruction of sibling groups, and the genotypes of parent individuals). This is a niche area to be sure, but also an opportunity to observe how one group of subject matter scientists interact with Bayesian statistics. I restrict myself to consideration of parentage analysis in natural populations of plants and animals. In parentage analysis among humans (e.g. paternity testing), the identity of the true parents is the key issue, there are usually a limited number of offspring and candidate parents, and a "spare no expense" approach can be taken when collecting genetic information. When studying natural populations, typically one is trying to match tens or hundreds of offspring with tens or hundreds of parents, and the genetic structure and demography of the population is of interest rather than the specific identity of parents. For instance, one may be interested in the variance across males in the number of offspring fathered, the heritability of a certain quantitative trait, or the distance sapling trees end up from their seed and pollen parents. Thus, uncertainty in the assignment of parents is tolerable as long as a reasonably clear picture of the demographics emerges.

There are some fully Bayesian approaches: both the parent assignments (which can be thought of as nuisance parameters or latent variables) and the demographic parameters are treated in a Bayesian framework. There are also methods that use a Bayesian approach primarily to deal with the parent assignments. For each offspring the posterior probability of belonging to candidate parent i is computed. These are called "fractional assignment methods," as parents are then assigned a fraction of each offspring in proportion to this posterior probability. A group of likelihood methods have also come to be known as fractional methods. Finally, a third section includes a brief selection of papers outlining other major inference methods used with natural populations. Omitted is a large body of work on design: the number of individuals that must be genotyped, and at how many genetic markers, to attain a given level of precision.

Fractional Assignment Methods

- Devlin B, Roeder K, Ellstrand NC (1988) Fractional paternity assignment: theoretical development and comparison to other methods. *Theoretical and Applied Genetics*, **76**: 369–380. Introduces the fractional parentage approach. While this paper does not refer to the method as Bayesian, the key quantity used is the posterior probability of a particular offspring j belonging to parent i, using a prior of equal probability over the candidate parents. The number of offspring parented by a individual i is then estimated as the sum of these probabilities over all offspring.

- Roeder K, Devlin B, Lindsay BG (1989) Application of maximum likelihood methods to population genetic data for the estimation of individual fertilities. *Biometrics* **45**: 363-379. The first paper that refers to Devlin et al (1988) as a Bayesian approach. Here, the probabilities of the observed offspring genotypes are looked at as a function of individual fertilities; these parameters are inferred using by maximum likelihood using the EM algorithm. The paper is notable in treating the assignments of offspring to parents as latent variables to be dealt with via standard methods.

- Smouse PE, Meagher TR (1994) Genetic Analysis of Male Reproductive Contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). *Genetics* **136**: 313–322. Refers to the method in Roeder et al (1989) as an "iterative fractional allocation" method because the expected value of the number of offspring belonging to each parent, computed in the expectation step of the EM algorithm, is typically a a fraction rather than a whole number. Develops likelihood ratio tests for whether the individual fertilities are all equal.

- Neff BD, Repka J, Gross MR (2001) A Bayesian framework for parentage analysis: the value of genetic and other biological data. *Theoretical Population Biology*: **59**: 315-331. This extends the approach of Devlin et al (1988) for use with informative priors derived from additional biological data (for example, behavioral observations). The proportion of offspring parented by a particular individual is treated in a fully Bayesian way, eg with credible intervals created from the posterior (although the authors call them confidence intervals).

- Nielsen R, Mattila DF, Clapham PJ, Palsboll PJ (2001) Statistical approaches to Paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* 157: 1673–1682. Extends the computation of the posterior probability of offspring i belonging to parent j to the case where some proportion of the parent individuals are not observed. Also considers inferring the relative fertility α of two groups of males. The method for fitting α is a 'fractional-likelihood' method in the sense of Smouse and Meagher (1994); however the maximum is found by a quasi-Newton method rather than EM. Unknown population sizes are also coped with by integrating over a prior distribution for the population size.

- Signorovitch J, Nielsen R (2002) PATRI–paternity inference using genetic data. *Bioinformatics*:**18** 341–342. Describes the program PATRI, which implements the methods described in Nielsen et al (2001).

- Neff BD, Pitcher TE, Repka J (2002) A Bayesian model for assessing the frequency of multiple mating in nature. *Journal of Heredity* **92**: 406–414. This is not a fractional parentage approach, but I have placed it in this section as it is Bayesian without fully embracing the Bayesian machinery. Inference is for the frequency of multiple mating f_{MM}; the data is reduced to the proportion of broods P_M with more than three paternal alleles (indicating more than one father contributing). A prior is placed on the frequency of multiple mating, and the likelihood $\Pr(P_M|f_{MM})$ is assessed for a grid of values of f_{MM} by simulation, using a fixed values for the number of contributing fathers and paternal skew (proportion of offspring produced by each father). Papers in the following section approach similar problems using all the genotype data, and doing joint inference for the number of fathers contributing to an offspring brood and the reproductive skew.

Fully Bayesian Methods

- Painter I. (1997) "Sibship reconstruction without parental information." *Journal of Agricultural, Biological and Environmental Statistics* **2**: 212-229. Considers a group of individuals that consists of several full sibships; Markov chain Monte Carlo is used sample from the posterior partitions into full sibships. This is easily transformed into a posterior for the number of parent pairs contributing to the group under study.

- Emery AM, Wilson IJ, Craig S, Boyle PR, Noble LR (2001) Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. *Molecular Ecology* **10**: 1265-1278. Reconstructs the family structure and parental genotypes of a single brood, nest, or (in the case of squid) egg string. It allows for both multiple paternity and maternity, using mild to moderately informative priors on the number of parents of each sex.

- Wilson, IJ (2001) *Parentage* software. http://www.mas.ncl.ac.uk/~nijw/ Implementation of the methods from Emery et al, with several additional options for modeling the number of parents and the proportion of offspring allocated to each. Allows the use of known or candidate parents where available; heated chains are used to improve mixing.

- Jones B, Clark AG (2003). Bayesian sperm competition estimates. *Genetics* **163**: 1193-1199. Inference for the distribution of the number of mates females take, using a sample of females and their offspring. The proportion of offspring fathered by each mate is governed by a model for sperm displacement, in which later mating males father more offspring; the parameter governing this process is also a target for inference.

- Jones B, Grossman GD, Walsh DCI, Porter BA, Avise JC, Fiumera AC (2007) Estimating differential reproductive success from nests of related individuals, with application to a study of the mottled sculpin, *Cottus bairdi*.*Genetics* **176**: 2427-2439.
  Inference for the distribution of mothers and fathers contributing to a nest where both multiple maternity and paternity is possible, using many nests and a partial sample of candidate parents. The use of candidate parents (who have been measured and aged as well as genotyped) allows model of the effect of age class on nest participation and the fraction of eggs produced by each individual.

Selected Additional Papers

- Thompson EA, Meagher TR (1987) Parental and sib likelihoods in genealogy reconstruction. *Biometrics* 43: 585-600. One of many papers by these authors that considers the strategy of categorically assigning offspring to the parents that result in the maximum probability for the offspring genotypes.

- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**: 639-655. Develops a simulation based methodology for assessing the confidence level of parentage assignments based on maximum likelihood. Uses the log likelihood ratio of the two most likely parents as its key statistic. The program implementing this method, CERVUS, is one of the most well developed and widely used parentage programs.

- Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in *Eucalyptus regnans* seed orchard using allozyme markers and the neighborhood model. *Molecular Ecology* **11**: 2379-2391. An excellent example of using parentage information to get at demographic parameters of interest–in this case selfing rate, local outcrossing, and long distance outcrossing–rather than parent assignments per se. One of many papers by these authors adapting maximum likelihood methods for different plant systems.

- Jones A, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* **12**:2511-2523. A comprehensive review.

- Butler K, Field C, Herbinger CM, Smith BR (2004) Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. *Molecular Ecology* **13**: 1589-1600. Sibship reconstruction typically shares the same goals as parentage analysis. This paper reviews sibship reconstruction methods, with numerical experiments comparing four methods.

SOFTWARE HIGHLIGHT

BGX: Bayesian Hierarchical analysis of 3′ GeneChip data

by Ernest Turro
ernest.turro@ic.ac.uk

GeneChips

High-density oligonucleotide microarrays allow biomedical researchers to estimate the expression of thousands of genes simultaneously through their messenger RNA transcripts. A microarray is a very small array to which thousands of strands of DNA with known sequence are attached. A sample of mRNA consisting of strands with unknown sequence is labelled and hybridised to the array. The non-hybridised mRNA is then washed away and the array is dyed and scanned in order to determine which of the sequences were present in the sample. Affymetrix 3′ GeneChip arrays represent genes by sets of probe pairs, each of which consists of a "perfect match" sequence, which matches a corresponding RNA subsequence perfectly, and an identical probe with an inverted base in the

middle position that is intended to measure non-specific hybridisation (binding by mRNA from other genes). Microarray data are notorious for their low replicability and high levels of background noise. Variability in the data is introduced at various steps, including array manufacture, sample preparation, array hybridisation and the scanning process.

BGX

BGX is a new Bioconductor R package for the analysis of GeneChip data. It is an implementation of a Bayesian hierarchical model, outlined in [2], that takes into account additive and multiplicative error, non-specific hybridisation and replicate summarisation at different levels in the hierarchy. This approach is in contrast with the majority of alternative algorithms, which extract the signal from GeneChip experiments in a sequence of separate steps, thereby inhibiting the simultaneous use of all available information. The software provides a full posterior distribution for the expression of each gene, even when only one array per condition is available [1].

Probe affinity effects

The propensity of probes to hybridise to mRNA has recently been shown to be affected by the composition of their DNA sequence. BGX accounts for this in an extension to the core model that incorporates affinity effects in the modelling of non-specific hybridisation. The probes on the arrays are, prior to analysis, grouped into a number of probe affinity categories according to their oligonucleotide sequences. Subsequently, the affinity-specific parameters are estimated from the data, simultaneously with all other parameters. BGX can also estimate the probe affinity categories by treating the categorisation mapping as a random variable with prior probability equal to the observed frequency of the categories.

Stratifying the non-specific hybridisation parameter according to probe affinity categories improves BGX's estimates of non-specific hybridisation and thus its gene expression estimates. For instance, Figure 1 shows an increased capacity to detect RNA concentration changes when probe affinity effects are incorporated.

Adaptive MCMC

The full conditional distributions of several parameters in the model are non-standard and are therefore estimated using a Metropolis-Hastings MCMC algorithm. Due to the high-dimensionality of gene expression data sets, each component of a given parameter has a different support and consequently a different optimal Random Walk proposal variance. Using a fixed variance for all components results in excessively low or high acceptance ratios for a large proportion of components, leading to highly autocorrelated chains.

BGX tackles this problem with an implementation of the novel Adaptive Metropolis-Within-Gibbs algorithm recently proposed by Roberts and Rosenthal [4]. It uses a unique proposal variance for each component of the model, which adapts to its optimal value after successive batches of 50 iterations. The aim is to achieve an acceptance ratio of around 0.44, which has been shown to be optimal for one-dimensional proposals in certain settings [3], and is commonly accepted as being a sensible benchmark. An acceptance rate that is close to zero implies inefficient mixing, while an acceptance rate that is close to one implies the probability space is not efficiently explored. The algorithm preserves ergodicity as long as each kernel has the right stationary distribution; the total variation distance between successive kernels tends to zero in probability; and the convergence time of each kernel is bounded in probability [4].

Figure 2 shows a dramatic reduction in the autocorrelation of samples from the full posterior distributions of the non-specific hybridisation parameter (H). This reduction allows the MCMC process to explore the target distributions more efficiently, resulting in shorter MCMC runs for a given quality of parameter estimates.

Conclusion

BGX performs well relative to other widely used methods at estimating expression levels and fold changes while having the advantage of providing a measure of uncertainty for its estimates. Figure 3 shows receiver operating characteristic curves between BGX and other methods for the Golden Spike data set. The package includes various analysis functions to visualise and exploit the rich output that is produced by the Bayesian model. BGX is fully described in [5] and may be

downloaded freely from the Bioconductor web site: http://bioconductor.org.

References

[1] Anne-Mette K Hein and Sylvia Richardson. A powerful method for detecting differentially expressed genes from genechip arrays that does not require replicates. *BMC Bioinformatics*, 7:353, 2006.

[2] Anne-Mette K Hein, Sylvia Richardson, Helen C Causton, Graeme K Ambler, and Peter J Green. BGX: a fully Bayesian integrated approach to the analysis of Affymetrix GeneChip data. *Biostatistics*, 6(3):349–373, 2005.

[3] Gareth O. Roberts and Jeffrey S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. *Statistical Science*, 16(4):351–367, nov 2001.

[4] Gareth O. Roberts and Jeffrey S. Rosenthal. Coupling and ergodicity of Adaptive MCMC. *Journal of Applied Probability*, 2005. To appear.

[5] Ernest Turro, Natalia Bochkina, Anne-Mette K Hein, and Sylvia Richardson. BGX: a Bioconductor package for the Bayesian integrated analysis of Affymetrix GeneChips, 2007. Under review for BMC Bioinformatics.

Fig. 1: Concentration plots – HGU95a

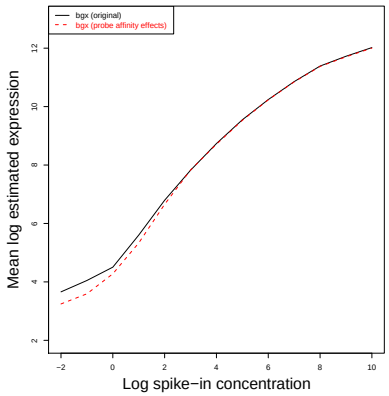


Fig. 2: IACT of log(H+1) for subsample of all genes

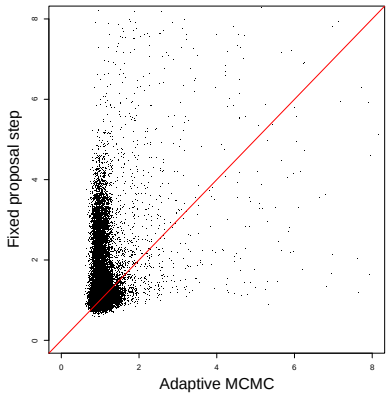
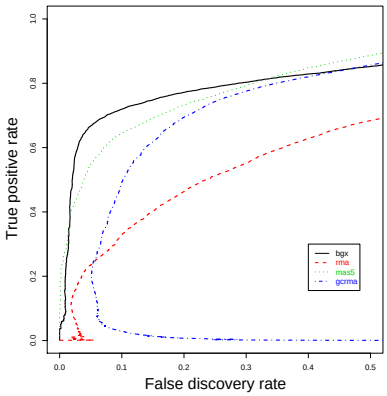


Fig. 3: Golden Spike average single array comparisons



STUDENTS' CORNER

NEW SEMESTER, NEW FACES

by Luke Bornn
l.bornn@stat.ubc.ca

As the new associate editor of the Student Corner, let me start with an introduction. I am a graduate student at the University of British Columbia, and have lived in the Vancouver area for the majority of my life. When not doing school-related work I enjoy tennis, hockey, hiking, and when I can find time, playing the guitar. I'd like to thank Raphael Gottardo for this opportunity and all of the previous editors for their contributions to this section. I would like to continue their work showcasing recent PhD thesis abstracts and presenting relevant and practical articles. If you have recently finished or are near finishing your degree, I invite you to send me your thesis abstract for publication here. Additionally, if you would like to see an article on a certain topic or think that you may have something to contribute, give me a shout at l.bornn@stat.ubc.ca.

With new students pouring onto campus and a summer research hangover, it's easy to forget deadlines for conferences happening in the winter. I want to highlight one such deadline here. MCMSki will be held in Bormio, Italy from Wednesday, January 9 to Friday, January 11, 2008. The NSF will likely provide funding for students from US institutions who will be presenting a paper at the conference.

The current deadline for this funding is October 20th (Early-bird conference registration ends November 1st). For more information visit `http://musing.unipv.it/IMS-ISBA-08`

This issue we have an article by Lawrence McCandless highlighting some opportunities and pitfalls in cross-disciplinary research for statistics students. Following this we have 3 dissertation abstracts from recent graduates in the field of Bayesian statistics.

EVER THOUGHT OF BRANCHING OUT FROM STATISTICS?

by Lawrence McCandless
l.mccandless@imperial.ac.uk
http://www1.imperial.ac.uk/medicine/people/l.mccandless
Department of Epidemiology and Public Health
Faculty of Medicine, Imperial College

Opportunities to participate in cross-disciplinary work are increasingly common for students studying Bayesian statistics. The emergence of complex datasets in areas like computer science and genomics presents students with numerous opportunities for collaboration with scientists in fields outside of statistics. I recently completed a PhD in statistics and work at the interface between Bayesian statistics and epidemiology. As a junior scientist starting an academic career, I would like to share some of my impressions on the advantages and challenges of cross-disciplinary work.

Early on in my graduate studies I realized that cross-disciplinary work can provide access to funding and jobs that would be otherwise unavailable to statistics students. For instance, there are many scholarships and grants available for health research. Graduate students in statistics may be eligible to compete for funding from multiple sources, and this can increase the odds that a proposal will be accepted. As a student, I applied for many scholarships and most of my applications were unsuccessful. But I was fortunate enough to be awarded a scholarship in healthcare research, which provided me with a steady income and relieved me of teaching responsibilities so that I could concentrate on finishing my PhD. Similarly, statistics students who are pursuing an academic career have the option of applying for positions in departments outside of statistics. A quick perusal of recent statistics job postings reveals numerous tenure track positions in computer science, finance and medicine.

Cross-disciplinary work can also provide access to areas for statistical innovation. Again speaking from my own experience, my thesis work involved developing Bayesian methods for reducing confounding bias in observational studies. Concern over confounding is ubiquitous in epidemiology, but the topic has historically received modest attention from statisticians. In recent years there has been renewed interest in Bayesian methods for causal inference, but there is still much uncharted territory and competing methodological approaches. Much of the attention in causal inference has been driven by scientists in economics, computer science and epidemiology. Thus my sense is that by working with scientists from other disciplines, it may be possible to get on the leading edge of emerging statistical problems.

A challenge of cross-disciplinary research is that it requires subject area knowledge in addition to the usual training in statistics. This knowledge is essential for understanding and communicating research ideas, and it often has to be acquired through independent study. As a graduate student, I had to make frequent tradeoffs between learning about epidemiological concepts and topics in statistics. Furthermore by studying different fields, students risk becoming a 'jack of all trades and master of none'. James Berger recently visited the statistics department at the University of British Columbia, where I did my PhD. In giving career advice to the graduate students, he emphasized the importance of postdoctoral research for statistics students doing cross-disciplinary work. My interpretation of his comments was that such students may require additional research training in order to keep up with their fellow peers in statistics.

If you are a student with interests in fields outside of statistics, I would encourage you to consider branching out. Although it can be daunting to study a new field while completing a statistics degree, there can be many rewards and opportunities. If you are interested in chatting more, please feel free to contact me.

Dissertation Abstracts

IMPROVING CLASSIFICATION MODELS WHEN A CLASS HIERARCHY IS AVAILABLE

by Babak Shahbaba
babaks@stanford.edu
http://www.stanford.edu/~babaks
Department of Statistics, Stanford University
PhD Supervisor: Ranford Neal (Toronto)

We introduce a new method for modeling hierarchical classes, when we have prior knowledge of how these classes can be arranged in a hierarchy. The application of this approach is discussed for linear models, as well as nonlinear models based on Dirichlet process mixtures. Our method uses a Bayesian form of the multinomial logit (MNL) model, with a prior that introduces correlations between the parameters for classes that are nearby in the hierarchy. Using simulated data, we compare the performance of the new method with the results from the ordinary MNL model, and a hierarchical model based on a set of nested MNL models. We find that when classes have a hierarchical structure, models that acknowledge such existing structure in data can perform better than a model that ignores such information (i.e., MNL). We also show that our model is more robust against missspecification of class structure compared to the alternative hierarchical model. Moreover, we test the new method on page layout analysis and document classification problems, and find that it performs better than the other methods. Our original motivation for conducting this research was classification of gene functions. Here, we investigate whether functional annotation of genes can be improved using the hierarchical structure of functional classes. We also introduce a new nonlinear model for classification, in which we model the joint distribution of response variable, y, and covariates, x, non-parametrically using Dirichlet process mixtures. In this approach, we keep the relationship between y and x linear within each component of the mixture. The overall relationship becomes nonlinear if the mixture contains more than one component. We extend this method to classification problems where a class hierarchy is available. We use our model to predict protein folding classes, which can be arranged in a hierarchy. We find that our model provides substantial improvement over previous methods, which were based on Neural Networks (NN) and Support Vector Machines (SVM). Together, the results presented in this thesis show that higher predictive accuracy can be obtained using Bayesian models that incorporate suitable prior information.

SPATIAL PROCESS MODELS FOR SOCIAL NETWORK ANALYSIS

by Crystal Linkletter
cdlinkle@stat.brown.edu
http://www.sfu.ca/~cdlinkle
Department of Community Health, Brown University
PhD Supervisor: Randy Sitter (Simon Fraser)

There has been a recent increase in the use of network models for representing interactions and structure in many complex systems. In this thesis we introduce the use of spatial process models for the statistical analysis of networks, emphasizing applications to social networks.

The first methodology we propose is the latent socio-spatial process model. In the spirit of a random effects model, pairwise connections are assumed to be conditionally independent given a latent spatial process evaluated at observed points in a covariate space. This smooths the relationship between connections and covariates in a sample network using relatively few parameters, so the probabilities of connection for a population can be inferred. The second model that is proposed is the meta-distance model. Here, a random function is used to represent the logistic relationship between covariates and binary relations. A spatial covariance structure is assumed for the random function, where the points in space are distances between attribute pairs. A Bayesian framework is used for estimation and prediction.

While spatial process models can be very flexible and provide reasonable fit and predictions in many contexts, interpretation of these models can be complicated. To aid in the identification of important covariates, we propose a reference distribution variable selection procedure. An inert variable is appended to the data for analysis, and the posterior distribution of an "activity" parameter associated with toe covariates is used as a reference distribution against which the true variables can be assessed. The approach is Bayesian, but the variable selection has a frequentist flavor.

Finally, we illustrate one important application of the proposed methodology. Local network topology can have a significant impact on contact-based processes, such as epidemics. This is demonstrated by looking at susceptible-infected-susceptible and susceptible-infected-removed epidemic models. We explore how using a predictive network model, such as the latent socio-spatial process mode, can help in predicting how a disease might spread in a population.

Some Approaches to Bayesian Design of Experiments and Microarray Data

by David Rossell
rosselldavid@gmail.com
www.stat.rice.edu/~rusi
Department of Biostatistics, M.D. Anderson Cancer Center
PhD Supervisor: Peter Mueller (M.D. Anderson)

This thesis consists of three projects. The first project introduces methodology to design drug development studies in an optimal fashion. Optimality is defined in a decision-theoretic framework where the goal is expected utility maximization. We show how our approach outperforms some other conventional designs. The second project generalizes the hierarchical Gamma/Gamma model for microarray data analysis. We illustrate how our generalization improves the fit without increasing the model complexity, and how one can use it to find differentially expressed genes and to build a classifier. When the sample size is small our method finds more genes and classifies samples better than several standard methods. Only as the number of microarrays grows large competing methods detect more genes. The last project explores the use of L2E partial density estimation as an exploratory technique in the context of microarray data analysis. We propose a heuristic that combines frequentist and Bayesian ideas. Our approach outperforms other competing methods when its assumptions hold, but it presents increased false positive rates when the assumptions do not hold.

NEWS FROM THE WORLD

Announcements

Call for 2007 Savage Awards

ISBA is pleased to announce two Savage Awards for outstanding Bayesian PhD dissertations in Theory and Methods and Applied Methodology, as well as two Honorable Mentions. Submissions are accepted between September 10 and October 10, 2007. For details on how to submit a PhD thesis for the 2007 Savage Award please visit http://www.stat.duke.edu/research/isba-sbss/SavageAward/. For descriptions of the award, please visit http://www.isds.duke.edu/research/isba-sbss/SavageAward

Call for 2007 Mitchell Prize

Nominations are now being accepted for the 2007 Mitchell Prize. The 2007 Mitchell Prize is awarded in recognition of an outstanding paper where a Bayesian analysis has been used to solve an important applied problem. The Prize includes a commemorative plaque an award of $1,000. Eligible papers for the 2007 Mitchell Prize must be published or accepted in a refereed journal or conference proceedings during 2005 or 2006. Deadline for submissions is 31 December 2007. For details on nomination for the 2007 Mitchell Prize please visit http://www.stat.duke.edu/apps/MitchellPrize.

Call for Papers for special IJF issue

Special Issue of the International Journal of Forecasting (IJF) on Applied Bayesian Forecasting in Economics
Guest Editors: Kajal Lahiri and Gael Martin.

The editors of this special issue of the International Journal of Forecasting (IJF) invite submissions on recent advances in the use of the Bayesian forecasting method in economics and allied social sciences.

The primary aim of the issue is to showcase the applicability of the Bayesian forecasting paradigm to a broad range of economic models

and empirical problems. We are not dictating the nature of the contribution, but are looking for papers with a reasonably substantive applied component. We are also looking for methodological contributions that highlight recent developments in Bayesian forecasting, including, but by no means restricted to, computational advances and issues related to forecast evaluation.

We request that authors initially submit a brief abstract to the editors of the special issue. A quick response as to the suitability of the proposed topic will be given. Final papers should be submitted electronically to the editorial office at ijf@forecasters.org, with a note to indicate that the paper is intended for the special issue. All contributions will be refereed and held to the usual IJF standards. Please refer to the guidelines for preparing papers for submission, at http://www.forecasters.org/pdfs/Guideforauthors.pdf.

The deadline for submission of papers is June 1, 2008, but earlier submissions are welcome. We are aiming for publication of the issue by mid-2009. Please submit your initial abstract electronically to both:

Kajal Lahiri klahiri@albany.edu and Gael Martin Gael.Martin@buseco.monash.edu.au.

Fund Raising in Honor of Pilar Iglesias

Contributions to the ISBA award in honor of Pilar Iglesias are still actively sought. The award will benefit students and young researchers from developing countries by providing travel grants to attend the Valencia Meetings or the ISBA World Meetings. So that the award can be given out in perpetuity, we are hoping to raise an endowment of US$20,000. We have already reached (either through contributions or commitments) almost two third of the sum, but an extra effort is needed.
Please consider making a contribution to the fund, even if small. To do so, please go to the ISBA web site at www.bayesian.org, and then to the News link. If you wish to learn more about Pilar Iglesias and her contributions to Bayesian statistics in Latin America and elsewhere, please see the March issue of the ISBA Bulletin, at http://www.bayesian.org/bulletin/0703.pdf.

New at Duke statistics

Effective July 1, 2007, The Institute of Statistics and Decision Sciences at Duke University becomes the Department of Statistical Science. This name change reflects expansion of the program to include a new undergraduate major and minor that complements our already outstanding graduate program and makes us a full-function department. The new name also signifies that statistical research and interdisciplinary science remain as primary missions. Alan Gelfand will serve as Head of the Department, following upon Dalene Stangl's five years as director of the Institute. Professor Stangl's service was exemplary and enabled the changes underway.

The new department will begin in growth mode. At least six new permanent positions (open rank) will be filled over the next three years. This will provide an exciting opportunity to shape the future for statistical science here at Duke and will make the department an even more attractive place for visiting researchers. Formal hiring advertisements will appear soon.

Events

Ninth Workshop on Case Studies of Bayesian Statistics, Carnegie Mellon University, Pittsburgh, PA, 19-20 Oct. 2007.

The Workshop will feature in-depth presentations and discussions of substantial applications of Bayesian statistics to problems in science and technology, poster presentations of contributed papers on applied Bayesian work, and contributed presentations by young researchers. In conjunction with the workshop, the Department of Statistics' Tenth Morris H. DeGroot memorial lecture will be delivered by Professor Larry Brown, University of Pennsylvania.

Posters and contributed presentations abstracts are due by October 5th. For more information visit the website, http://lib.stat.cmu.edu/bayesworkshop/2007/.

MCMCSki: Markov Chain Monte Carlo in Theory and Practice, Bormio, Italy, 9-11 Jan. 2008.

The unifying theme of the third joint international meeting of the IMS and ISBA is MCMC and its impact on the theory and practice of statistics, but invited sessions and poster presentations will cover a broad range of statisti-

cal topics. Plenary speakers are Peter Green of the University of Bristol, Kerrie Mengersen of the Queensland University of Technology and Xiao-Li Meng of Harvard University.

There will also be a pre-conference "satellite" meeting, from 7-8 January, intended to provide a snapshot of the methodological, practical and theoretical aspects of an emerging group of methods (adaptive MCMC, adaptive population Monte Carlo, and various breeds of adaptive importance sampling amongst others) that attempt to automatically optimize their performance for a given task.

Abstract submission for poster presentations is now open. Limited financial support for the travel of junior ($<$ 5 years since PhD) is anticipated for those presenting in a poster session. For more information visit the website, http://musing.unipv.it/IMS-ISBA-08/, or contact Brad Carlin brad@biostat.umn.edu.

9th Brazilian Meeting on Bayesian Statistics (EBEB), Maresias Beach Hotel, Maresias, Sao Paulo, Brazil, 24-27 Feb. 2008.

The 9th EBEB will have two special sessions, one dedicated to Professor Carlos A. de Braganca Pereira, to thank him for his many contributions in the development of the Bayesian Statistics in Brazil and, more broadly, in Latin America. Another will pay a tribute to Professor Pilar Iglesias, who passed away in March 2007 and had a close relationship with the Bayesian Brazilian community. You are also invited to visit the website to get more information about 9th EBEB and to appreciate the Atlantic Forest's beauties. We are looking forward to seeing you in Brazil.

Oral presentation abstracts are due by October 30th, and poster abstracts are due by December 8th. For more information visit the website, http://www.ime.usp.br/~isbra/ebeb/9ebeb/.

World Meeting of the International Society for Bayesian Analysis, Hamilton Island, Australia, 21-25 Jul. 2008.

ISBA 2008 will combine an excellent scientific program - including five keynote speakers, 90 oral presentations, three parallel sessions and two poster evenings - with an active social schedule.

Abstract submission is now open. For additional details go to the website, http://www.isba2008.sci.qut.edu.au/, or e-mail isba08@qut.edu.au.

NIPS Workshop on Statistical Network Models, December 8th 2007, Whistler, BC.

Organizers: Kevin Murphy, Lise Getoor, Eric Xing, Raphael Gottardo

The purpose of the workshop is to bring together people from different disciplines - computer science, statistics, biology, physics, social science, etc - to discuss foundational issues in the modeling of network and relational data. In particular, we hope to discuss various open research issues, such as (1) How to represent graphs at varying levels of abstraction, whose topology is potentially condition-specific and time-varying (2) How to combine techniques from the graphical model structure learning community with techniques from the statistical network modeling community (3) How to integrate relational data with other kinds of data (e.g., gene expression or text data). For more details please visit: http://www.cs.ubc.ca/~murphyk/nips07NetworkWorkshop/.