THE ISBA BULLETIN

Vol. 8 No. 4 December 2001

The official bulletin of the International Society for Bayesian Analysis

AND WHAT A PRIVILEGE IT HAS BEEN

by Alicia Carriquiry
ISBA President
alicia@iastate.edu

Dear Friends:

In keeping up with what is by now a tradition for outgoing ISBA Presidents, I wish to bid all a fond farewell. This is the time for the inevitable inventory of great successes and better-forgotten disappointments; I am happy to report that in the first year of the millenium, the successes for ISBA greatly outweighed the disappointments (and this bodes well for the next thousand years!).

I would like to first extend my warm thanks to the ISBA Executive Committee and to the Board of Directors; the dedication of ISBA elected officers provides the institutional backbone that we need to dream up and implement new initiatives, organize meetings and publish a bulletin, and in general to continue promoting the development and use of Bayesian methodology among our colleagues in statistics and in other disciplines. Several ISBA officers complete their term this December. Among those are Phil Dawid, Val Johnson, and Tony O'Hagan, who end their terms as Past President, Treasurer, and Chair of the Program Council, respectively. With this December issue, Fabrizio Ruggeri completes a brilliant term as Editor of the ISBA Bulletin. Fabrizio has done a spectacular job, and has turned the Bulletin into a publication that ISBA can be really proud of. Mike Evans has functioned as our Webmaster for many years now, and has provided the leadership needed to update procedures such as voting, which we can do accurately and efficiently online. (A large international society that shall remain nameless is thinking of modeling their on-line voting system after ours!). Mark Berliner, Petros Dellaportas, Jayanta Ghosh, and Sylvia Richardson rotate out of the Board of Directors. On behalf of the ISBA membership, I thank all of them.

And in the spirit of recognizing the extraordinary contributions of many of our members, I now revisit some of the ongoing activities and new initiatives that were carried out this year. The Savage Award committee, chaired by Ehsan Soofi, pored over a large number of entries and produced a slate of three winners and one honorable mention. Congratulations to all of them! The committee to select the first DeGroot Prize winner is already functioning under the capable direction of Steve Fienberg. Finally, the Bylaws committee, chaired by Jay Kadane, reviewed the existing ISBA bylaws and modified or added bylaws as needed, to accommodate the increasing number and range of ISBA business. Dale Poirier hosted us in Laguna Beach in April, the site of the ISBA Regional Meeting held in 2001, and Hal

Contents

Stern and the committee he chairs are in the process of producing the first Selected Contributed papers program, that will be part of the larger program at the Valencia VII meetings. As usual, our chapters have been busy, and meetings were held (or are in the planning stages) for Chile, South Africa, and Brazil. The organization of the first joint ISBA/IMS international meeting, to be held in San Juan, Puerto Rico, in 2003 is already underway, thanks to the efforts of Tony O'Hagan and Luis Raul Pericchi in representation of ISBA, and Susan Murphy in representation of the IMS. Many of us have benefitted from the administrative and secretarial support that was generously provided by Nicole Scott from Duke University, Heidi Sestrich from Carnegie Mellon University, and Ruth Birch, from Iowa State University. A big round of applause to all.

Perhaps the most significant and promising activity undertaken by ISBA this year was the publication of the volume *Bayesian Methods with Applications to Science, Policy, and Official Statistics*, which includes selected papers from ISBA 2000. The volume, all wonderful 592 pages of it, was assembled and edited by an editorial board led by Ed George, who kept the submission, editorial, and publishing process running like clockwork. A printed copy of the volume will be sent to all who attended ISBA 2000, but the exact same volume can be browsed, downloaded, and printed out on-line at `www.stat.cmu.edu/ISBA`. Completely free of charge, for members and non-members alike.

Even after rigorous refereeing, the ISBA 2000 volume is significantly larger and broader in scope than was initially planned. A striking illustration of ISBA's mission to promote the development and practice of Bayesian analysis, the volume includes work on Bayesian methodology and applications on areas as diverse as astronomy, ecology, genetics, engineering, and epidemiology.

Would the ISBA membership benefit from a new venue for the publication of articles of interest to Bayesian analysts? A journal that serves as a vehicle for the exchange of ideas of interest to ISBA members, and that is unique in structure, character, and mission can be a powerful tool for ISBA to foster the development and application of Bayesian analysis, and an incentive for more of our colleagues in statistics and other areas to join ISBA and participate in our activities. A committee co-chaired by Rob Kass and myself has revisited the "journal question" and has written a proposal for the membership to mull over and comment. I urge you to read the committee's proposal which is printed later in this bulletin, and to participate in the discussion that I hope will ensue. Should ISBA launch a new journal? For the many reasons given in the proposal, and for a few more that I can think of myself, I would personally welcome this initiative. But we wish to hear the opinion of the membership before embarking on this major project for ISBA. Please check the ISBA web site at `www.bayesian.org`; during the next few days, we will list instructions on how to access an electronic bulletin board where you can post comments and communicate with other ISBA members. You can always contact Rob Kass or me directly (`kass@stat.cmu.edu`, `alicia@iastate.edu`).

Enough said! I'll ride now quietly into the sunset, wishing David Draper a productive, fruitful year, with twice as many ISBA members to think about. It has been an honor and a privilege (and a lot of fun too!) serving you. I am deeply grateful for the opportunity.

GOOD-BYE FROM THE EDITOR

by Fabrizio Ruggeri
ISBA Bulletin Editor
`fabrizio@iami.mi.cnr.it`

My job is over: a new Editor will be announced soon and a new Editorial Board will be appointed. I am getting ready for a different involvement in ISBA activities (thanks to those who contributed to my election in the ISBA Board of Directors). New exciting adventures are expected in the Bayesian world: some are described in this issue. The one I am looking forward the most is about the Bulletin. The new Editor has the opportunity of starting from what we (see page 9) have built in the last three years so that a new, better, Bulletin will be provided to ISBA members (possibly by email). (S)he will need all your support and suggestions, as we got during our term.

Today, I received a message from Dennis Lindley. He made some comments on the Bulletin, and he raised an important point I subscribe: never forget those who suffer and die all over the world!

Arrivederci!

ISBA ELECTIONS

by Cindy L. Christiansen
ISBA Executive Secretary
cindyLc@bu.edu

This year ISBA held an election for the positions of President-Elect, Treasurer and four new Board members.

One hundred and forty eight ISBA members cast votes in the election. This participation rate is about the same as in previous elections. Many thanks to all who participated and especially to those who stood as candidates. The following individuals were elected.

➤ **President-Elect**
Ed George
➤ **Treasurer**
Peter Müller
➤ **Board Members**
Pilar Iglesias
Sonia Petrone
Fabrizio Ruggeri
Robert Wolpert

VALENCIA 7 SELECTED CONTRIBUTED PAPERS UPDATE

by Hal Stern
SCP Selection Committee Chair
hstern@iastate.edu

This year for the first time the Valencia meeting program will include oral presentation of approximately 50 papers from among those submitted to the Selected Contributed Papers (SCP) competition. A committee of 14, chaired by Hal Stern, Iowa State University, is currently reviewing the extended abstracts (3 pages or less) that were submitted.

There were a total of 123 submissions from all over the world. There were 39 submissions with lead author from the United States, 27 submissions with lead author from the United Kingdom, and 11 with lead author from the Netherlands. The breakdown by region is as follows:

US/Canada	46
Europe (not UK)	29
UK	27
Central/South America	7
Asia	6
Australia/NZ	5
Africa	3

Authors were invited to identify up to 3 topic areas into which their contribution might be placed (from a list of 17 topics). The data here are a little messy as not everyone did so and some of the last minute submissions are not included in my tabulations. The most popular topics were:

Computational methods, algorithms, convergence, sampling, software:	28
Stochastic processes, time series, spatial and spatio-temporal models:	27
Linear models and regression, surveys, hierarchical models, etc.:	24
Economics, social science, public policy, law:	23
Inference, optimality, distribution theory, causality:	22

We look forward to adding the best of these submissions to the Valencia program.

1st ISBASA WORKSHOP

by Paul J. Mostert
Local Organizer
pjmos@akad.sun.ac.za

The 1st ISBASA Workshop will take place 3, 4 and 5 April 2002 at the University of Stellenbosch (50 km from Cape Town). It is an official activity of the Southern Africa ISBA Chapter and the Department of Statistics and Actuarial Science of Stellenbosch University will host it. The Executive Committee has invited Jim Berger and Peter Mueller to participate in the Workshop. Anders Madsen (Hugin, Denmark) is hosting a session in Bayesian Networks.

The workshop will be attended by a limited number of statisticians from Southern African and abroad. Amongst the delegates, will be young researchers in the Southern African region. The aim of this Workshop is not only to promote Bayesian activities in Southern Africa, but also to expose young statisticians to the Bayesian approach. Therefore, all sessions will start with basic theory and applications. One of the sessions will involve Business in South Africa. An open debate will take place in one of the sessions on the relevance of modern Bayesian analysis in practice. ISBASA intends to invite young top postgraduate students from universities all over South Africa. It is important for their education to get the opportunity to learn from international as well as local professionals. This is an opportunity for Business to meet these students and for students to get exposure to a possible career environment.

For further details regarding the programme, contact Paul Mostert (pjmos@akad.sun.ac.za) or Alta de Waal (adewaal@csir.co.za).

PROPOSAL FOR A NEW ELECTRONIC JOURNAL

➤ Background

The idea of founding a Bayesian journal has come up repeatedly over the past 10 years or so. There was a formal vote taken in ISBA in 1994, and by a very narrow margin (68 to 64) the proposal was defeated. A lot has changed since then.

Bayesian methods are now very widely used in applications, and all major statistics journals publish many Bayesian articles. In 1994 there were fears that a Bayesian journal would be divisive or isolationist, in the sense that it would serve to divert Bayesian work away from the mainstream of statistics. Now, the great success of Bayesian methodology has made many more ISBA members feel they have gained a secure foothold in the discipline of Statistics, and acceptance in a variety of application areas.

An electronic journal of the type we describe below would provide a lively forum for exchange of ideas. It would increase ISBA's leadership within the field of Statistics as a whole by demonstrating how a modern statistical journal can operate efficiently and effectively, while reaching out to other disciplines. It would also provide a valuable service to ISBA members, and would further spread the social network ISBA provides, increasing the society's ability to promote the development and practice of Bayesian analysis.

➤ Purpose and Scope

The new journal *Bayesian Analysis* would be unlike any of the mainstream statistics journals, and would be one of the first fully electronic journals in the statistical sciences.

Bayesian Analysis would be very broad in scope, and inclusive in outlook. It would publish any good article of interest to the ISBA membership. It would seek authors both within Statistics and from other domains, such as computer science, law, health policy, bioinformatics, marketing, astronomy, and neuroscience, who might not only bring fresh problems, but also differing paradigms for solving them.

A special feature of *Bayesian Analysis* would be rapid turnaround. We would expect most articles to be reviewed very quickly (median time-to-decision less than 10 weeks); revisions to be requested only when they could be accomplished in a timely fashion; and publication to occur a short time after acceptance.

We anticipate that *Bayesian Analysis* would be published monthly. This new journal would have the freedom to publish articles with substantial appendices devoted to philosophical arguments, extensive literature review, or mathematical details, which are typically omitted from our top maintstream outlets. Authors could also provide links to datasets, additional analyses or demonstrations, references, and even figures, video, and audio. We envision a journal that remains readable while providing many useful resources. When appropriate, articles could be published with commentary, which enlivens and deepens the presentation.

Bayesian Analysis would welcome case studies and, in particular, the Case Studies in Bayesian Statistics series would be published there. Other workshop and seminar papers would also fit nicely into the journal. The advantage of publishing in *Bayesian Analysis* would be greater accessibility, greater flexibility, and absence of space constraints except on editorial grounds. Papers sponsored by the Section on Bayesian Statistical Sciences (SBSS) and presented at the Joint Statistical Meetings could, after refereeing, also appear in the ISBA electronic journal rather than in the non-refereed JSM Proceedings. This would elevate their stature, another real benefit for ISBA members. In addition, *Bayesian Analysis* would publish reports on software development, teaching, and professional matters.

➤ Production

The ISBA Publication Committee will be charged with determining a specific plan for publication and bringing it to the Board for approval.

The Committee has already investigated several possibilities, ranging from independent publication to commercial ventures, and collected information on each.

The journal would be available for free to ISBA members. Abstracts of published papers would be available publicly on the journal's website. Because one of the objectives is to attract scientists from other disciplines who might not be interested in the entire journal, non-members would be allowed to down-load a limited number of articles as a kind of "introductory offer".

The Committee would also propose to the Board a plan regarding co-sponsorship of the journal by other societies including, we hope, not only statistical societies but potentially others as well.

➤ A Prototypical List of Papers

To illustrate what we have in mind, here are some papers we might like to see in *Bayesian Analysis*. Each potential author was contacted and asked to submit a title for this list.

M.J. BAYARRI, *Is it "safe" to use Vague Proper Priors?*

J.O. BERGER, *The Impossibility of Model Selection, But We Have to Try*

E.N. BROWN, R.E. KASS and V. VENTURA, *Bayesian Methods in Neurophysiology*

A. CARRIQUIRY and M. DANIELS, *Analyzing Nutrient Intakes and Requirements: Are Those Chips Good for You?*

C. CHRISTIANSEN, *Using a Bayesian Framework to Understand Implicit Assumptions of Standard Statistical Tests*

A.P. DAWID, *Bayesian thinking: A guide to life*

B. EFRON, *Is There Such a Thing as Objective Bayesianism?*

P. IGLESIAS and R. ARELLANO-VALLE, *Bayesian Inference in Elliptical Models: A Review*

V. JOHNSON, *Differential Grading Policies and Student Course Enrollments*

M. JORDAN, *Variational Methods and Dirichlet process priors*

J. KADANE, M. SCHERVISH and T. SEIDENFELD, *Measuring the Extent of Incoherence*

M. LEWICKI, *Multiscale Priors for Learning Efficient Codes in Natural Images*

J.S. LIU, M. GUPTA, X. LIU, L. MAYERHOFERE and C.E. LAWRENCE, *Statistical Models for Motif Discovery*

T. LOREDO AND I.M. WASSERMAN, *Guilt by Association: Bayesian Assessment of Spatio-Temporal Coincidences in Astronomical Surveys*

D.K. PAULER, *Bayesian Joint Analyses of Longitudinal and Failure Time Data for Optimal Prediction*

C. ROBERT, *MCMC: Current State and Future Iterations*

H. STERN, *A Mixed Marriage: Teaching Traditional Methods and Bayesian Concepts*

P.F. THALL, H.-G. SING and E.H. ESTEY, *Multi-Course Treatment Strategies for Rapidly Fatal Diseases*

Y. ZHAN, *Bayesian Data Analysis in S-PLUS*

➤ Some Frequently-Voiced Concerns

- Wouldn't *Bayesian Analysis* draw papers away from mainstream journals such as *Biometrika* or *JASA*?

Bayesian papers now make up a substantial percentage of the papers published in the top statistical journals. This is unlikely to change: authors will continue to submit to those journals most of their best papers.

- But wouldn't *Bayesian Analysis* then become a second-rate journal?

We would certainly expect to see in this journal papers that would not be considered appropriate for outlets like *Biometrika* or *JASA*. However, this would be a reflection of different emphasis and scope. Furthermore, in many ways *Bayesian Analysis* would compete favorably with existing journals: it would offer flexibility of topic and format, and it would provide rapid turnaround. Authors who feel a particular paper might not fit the constrained expectations of the leading mainstream journals, or who don't want to hassle the long and intrusive refereeing process, would choose to submit excellent work to *Bayesian Analysis*.

In addition, papers that can not be published in the top existing journals because of their specialized nature end up being scattered in many places, including lower-tier mainstream journals and proceedings volumes. Usually these are very expensive publications. Many potential readers are unaware of those papers, and authors are unable to have large numbers of their colleagues see their work.

If *Bayesian Analysis* rejects papers of low quality, while publishing some first-rate papers on diverse topics, its reputation will be very strong.

- At a time when Statistics is gaining strength in many other disciplines, wouldn't *Bayesian Analysis* encourage an inward-looking, parochial attitude among Bayesians?

Bayesian Analysis would, from the beginning, encourage and actively solicit participation from people in other disciplines. The success of this venture depends, in part, on being open-minded about contributions using differing approaches. Bayesian analysts would be encouraged to communicate their cross-disciplinary successes, not as a substitute for publishing in substantive scientific journals but, in line with the fundamental purpose of our field, to spread techniques across many application areas. The new journal would also encourage a wide range of theoretical and methodological contributions from the many workers in Bayesian analysis who now come from a non-traditional background, thus ensuring that relevant work anywhere in the broad field of Bayesian analysis becomes widely disseminated.

DANIEL PEÑA

by Michael Wiper
mwiper@est-econ.uc3m.es

Professor Peña is one of the best known statisticians in Spain, having written over 100 published papers and a number of textbooks on both Bayesian and classical statistics. Since 1991, Professor Peña has been working in the Statistics and Econometrics Department at the Universidad Carlos III de Madrid. You can find out more information from the Departmental homepage at:

http://halweb.uc3m.es/

We e–mailed Professor Peña a number of questions about his career and the Bayesian world in general. Here are his responses.

`1. When and why did you first get interested in statistics, and in particular, Bayesian statistics.`

The first time I heard about Bayesian Statistics was in 1972, when I was starting my Ph.D. at the University of Madrid (UPM). I had a degree in industrial engineering (Ingeniero Industrial) and I was interested in a PhD in Operational Research. I had already had a mathematical statistic course and after it I concluded that statistics was a pretty boring subject, hard to understand and not very useful for real problems. During the first year of the PhD I took a course on Decision Theory and the text-book recommended was Decision Analysis by Howard Raiffa. I started reading this book and I was completely fascinated by it. I was very much impressed by the beauty of the concepts and the strong logic of the ideas in the book. After this course, I read Schlaiffer's book and Pratt, Raiffa and Schlaiffer, and I was getting more and more interested in Bayesian Statistics. At this time I did not know anybody in Spain who was interested in this topic, and finally I wrote a thesis, pretty much by myself, on Bayesian Decision Analysis applied to medical diagnosis and treatment. At that time I had a rather left wing point of view, and I did not want to work on anything related to business or economics, so that I chose a medical application.

`2. Tell us about some of the people who have influenced your career.`

After my PhD I started teaching decision analysis in an Operational Research department. I wanted to integrate dynamics and time series data in the decision process and in 1978 I met Arthur Treadway, an economist from Chicago, who came as a visitor to the University of Madrid. He told me about a new methodology for time series that has just appeared, and we started meeting once a week to study together the Box and Jenkins book. Again, I was fascinated by this book and I was very much attracted by the iterative statistical learning process advocated. In particular, I was very much attracted by the steps of identification and diagnosis of the model, that are very important in time series but also elsewhere. At that time I did not know how to integrate these ideas into the Bayesian framework, but I was convinced that they were very useful and important. I invited George Box to come to Spain to teach a short course on time series and he came with George Tiao. I learned a lot from them and I was very attracted for their approach to time series. After their visit I decided to spend a year in Wisconsin. George Box was always very nice and he helped me to get support to spend the 83-84 academic year in Madison. It was a great year from all points of view. I addition to working with George Box on factor analysis on time series, I visited George Tiao that has just moved to Chicago and we started working together. I also met Irwin Guttman and Dennis Cook, who were visiting Wisconsin this year, and learned a lot about Bayesian Statistics working with Irwin and about influence analysis working with Dennis. Later on I was very much impressed and influenced by the work of Tukey and Efron, among others, but, altogether, I think that the three persons who have had most influence in the way I look at statistics are George Box, George Tiao and Irwin Guttman.

`3. You have done a lot of work in time series but mainly using classical statistics. (I don't know if your new book on time series (`*A Course in Time Series Analysis*`, Daniel Peña, George Tiao and Ruey Tsay eds., Wiley) contains anything on Bayesian methods). Do you think that classical techniques are better suited to this field than Bayesian methods? If so, why?`

Yes my book with George Tiao and Ruey Tsay on time series has a chapter on Bayesian Time series written by Ruey, but most of it is from the classical point of view. In many time series applications the sample information is much larger than the a priori information about the parameters, and thus maximum likelihood estimation is roughly similar to Bayesian estimation. I have not had real experience with short time series in which the prior information can really make a difference. Also, a key part of the time series model building process is the identification and diagnostics of the model, and, for these steps, I believe that the so called classical statistics are better suited than traditional Bayesian Statistics. This point has been stressed by Box (1980) and I fully agree with him that we need Bayes theorem for estimation but to build statistical models we also need many other tools that has been developed in the so called classical statistics. Bayesian time series has sometimes been identified with the structural approach using the Kalman filter, whereas ARIMA models have been considered mostly from the classical statistics viewpoint. I think that both approaches are complementary, and both can be estimated by Bayesian or maximum likelihood techniques. I usually prefer the reduced form, (ARMA type models) because we have better tools available for identification and diagnosis but there are many cases, for instance dynamic factor models, for which I believe that Bayesian recursive estimation using the Kalman filter is more useful. So, I do not think it is so relevant if we use ML (maximum likelihood) or BT (Bayes Theorem) for estimation, because I do not see classical and Bayesian statistics as rival approaches but rather as complementary, and we will be able to solve real problems better if we can use both. The idea of a unique best method to obtain the truth has disappeared in many scientific areas and it is surprising that this dogmatic point of view has such strong roots in our scientific community.

`4. Conversely, in your work on outliers, influence and robustness, you have used both Bayesian and classical techniques.`

Yes, and I have found both very useful for different things, but again I think it is not true that we can do everything better from the Bayesian point of view. For instance, many Bayesians do not understand the concept of masking very well. This is an idea that has been developed mostly in the classical robustness literature and there are many so called robust Bayesian procedures published in the last 10 years, that fail completely as soon as we have a small group of high leverage outlier observations. On the other hand I think that we have some classical procedures to deal with outliers in multivariate problems and in regression that are far ahead of the Bayesian alternatives. I find it surprising that some people seem to be more concerned about whether a procedure is truly Bayesian or not than whether or not the procedure is useful to solve the problem it tries to solve. I think that research in Bayesian statistics should concentrate more on solving problems that classical statistics is not well suited for, such as working with small samples, using subjective information in a better way or combining in a robust way different sources of information.

`5. You have written fairly extensively on education, and quality improvement in the university sector. What comments do you have on how to improve the teaching of (Bayesian) statistics ?`

I hope that Bayesian Statistics will be more used in all scientific areas in the future. I think that we should concentrate our teaching in presenting simple and flexible procedures that people can use in practice to solve the problems they will face in their professions. Sometimes a classical tool could be a convenient and fast approximation and then we should recommend using it. We should teach ALL statistics, that implies how to use subjective information, how to combine information from different sources, and how to incorporate all sources of uncertainty in the problem and all these problems can be better solved using Bayesian Statistics. However, we should also stress exploratory data analysis and model diagnosis, problems in which probability plays a small role and in which the most useful tools have been developed within classical statistics.

`6. Also, you have written statistical texts and research papers designed for social scientists, engineers, medics, and have collaborated on research projects with people from many fields. What are the major differences you have found in such diverse areas.`

As Tukey has said, the great thing about statistics is that you can play in someone else backyard. I have had a lot of fun working with people from many different fields in understanding their data. I believe that real problems are the most stimulating source for new statistical developments. I think that we will be better scientists and more useful for society if we concentrate our efforts in solving the difficult problems that are all around in the real world instead of concentrating our efforts in generalizing methods and ideas that are of very limited usefulness in practise. Many of my research interests have come from practical applications. For instance, my interest in outliers came from noticing the important effects they may have, even in large data sets, in our conclusions on public welfare policies in a joint work with the economist Javier Ruiz Castillo. My interest in dynamic factor models came from my collaboration with a historian, Nicolas Sánchez Albornoz, in understanding the effect of political events in wheat prices in Spain in the XIX century. My interest in cluster analysis and data heterogeneity came from building quality indexes for the railroad system in Spain.

`7. Continuing along the same lines as questions 3 and 4, what do you think Bayesians can learn from classical statisticians and vice versa.`

I think Bayesians can learn methods for exploratory analysis and model diagnostics from Classical statistics. Classical statistics can learn flexible methods for estimation and testing using several sources of information from Bayesian statistics.

`8. Also, looking into the future, Bayesian statistics seems to be gaining more converts every year. Do you think that classical statistics will eventually be replaced by Bayesian as the dominant method or do you think other techniques (neural nets, data mining etc.) will start to replace standard statistical analysis?`

I think that we will move towards a more balanced teaching of statistics. It is clear to me that today a competent statistician cannot ignore Bayesian statistics, as was unfortunately very common in the past. In the same way, I think that the Bayesian statistical community is more mature now and is more interested in solving new problems and finding new procedures than in competing against classical statisticians. But our changing world will force us to develop new tools and new paradigms. For instance, in the last century the standard paradigm for both classical and Bayesian statistics, is that our raw material was a sample from some statistical model. We can simplify this situation saying that our basic assumption was some kind of data homogeneity, may be with some small proportion of outliers. This paradigm is not appropriate today for the analyses of the available large data sets that include hundreds of variables and many thousands of observations. In this situations we do not have a central model, rather we expect that different models will explain the data in different regions of the sample space. This multi-model situation can be called the data heterogeneity situation. The tools to be used in these cases are more complex than the ones considered by cluster analysis or robust methods, and we need new statistical tools to extract the information in these data set. I believe than in order to solve these complicated problems that we have ahead, both Bayesian and classical statistics will be useful. We also need to develop more automatic procedures for data analysis and for this purpose Neural networks and Data mining take advantage of the available computer power. Neural networks are fast, and sometimes not very efficient, ways to build regression or time series models in which the response is a non linear function of linear combinations of the explanatory variables. Thus they are fast procedures for non linear factor models. Data mining includes fast multivariate exploratory methods that can be very appropriate in many situations. These two procedures are useful for gathering information from a given set of data, but if we want to generate knowledge, that is to understand not only the sample data but also similar samples not yet observed, and to be able to generate useful forecasts we need statistical models.

`9. Looking back, what are the things you are proudest of in your statistical career?`

I am very proud of having had the opportunity of working with such great statisticians as George Box, George Tiao, Irwin Guttman, Dennis Cook and Victor Yohai. I am also very proud of helping to develop the Department of Statistics and Econometrics at the Universidad Carlos III of Madrid, I have excellent colleagues there and the atmosphere is very stimulating. Also I am very proud of contributing a bit to developing the research potential of my Ph.D. students. I have been very lucky to have excellent Ph.D. students and, to be honest, I feel that I have learned from them more than they have learned from me.

`And looking ahead, what about your future plans in statistics?`

I have been working for many years now with George Tiao in the problem of data heterogeneity and this is one of my first priorities. Also I am interested in many other problems: diagnostic tests for time series (with J. Rodriguez), new methods for bootstrap in time series (with A. Alonso and J. Romo), outliers in Garch processes (with A. Carnero and E. Ruiz), Bayesian Model Averaging (with I. Guttman and D. Redondas), Robust Bayesian estimation (with R. Zamar), Projection Pursuit methods for multivariate time series (with P. Galeano and R. Tsay), Dynamic factor models (with P. Poncela), random coefficients models for quality (with V. Yohai), Forecasting Multivariate time series (with I. Sánchez) and image analysis (with M. Benito). These problems are going to keep me busy for a while!

`10. And finally, what are you looking forward to seeing next year in Valencia 7, or should it be Tenerife 1?`

New practical, flexible and iterative methods for getting knowledge from the large and heterogeneous data set that as statisticians we are going more and more to face. These methods should incorporate several dimensions : multivariate, dynamic, robust (in a very broad way) and computationally efficient. I am really looking forward to new advances in this field.

Thanks to Daniel for an interesting interview.

THANKS

TO ALL THOSE WHO SERVED IN THE EDITORIAL BOARD IN 1999 - 2001

Associate Editors: Jim Albert, Sudipto Banerjee, Maria Eugenia Castellanos, Kate Cowles, Maria De Iorio, Duncan Fong, Sujit Ghosh, Gabriel Huerta, Leonhard Knorr-Held, Antonio Lijoi, Javier Morales, Antonio Pievatolo, Raquel Prado, David Rios Insua, Siva Sivaganesan, Dalene Stangl, Isabella Verdinelli, Michael Wiper

Corresponding Editors: Christophe Abraham, Antonia Amaral Turkman, Carmen Armero, Marilena Barbieri, Bill Bolstad, Agata Boratynska, Cinzia Carota, Christopher Carter, Andres Christen, Roger M. Cooke, Petros Dellaportas, Gül Ergun, Josep Ginebra, Paul Gustafson, Eduardo Gutierrez Peña, Katja Ickstadt, Leonhard Knorr-Held, Robert Kohn, Jack C. Lee, Shu-Ing Liu, Hedibert Freitas Lopes, Udi Makov, Marek Meczarski, Renate Meyer, Ioannis Ntzoufras, Yasuhiro Omori, Suleyman Ozekici, Susan Paddock, Carlos Daniel Paulino, Jose Miguel Perez, Raquel Prado, Josemar Rodrigues, Abel Rodriguez, Alfredo Russo, Aldo Sacerdoti, Mike Smith, Daniel Thorburn, Brani Vidakovic, Mattias Villani, Hajime Wago, Cathal D. Walsh, Simon Wilson, Lara Wolfson, Karen Young

David Heckerman

by David Rios Insua
`d.rios@escet.urjc.es`

Our last interview will be with David Heckerman (`research.microsoft.com/~heckerman`).

David has contributed widely to the areas of statistics and data analysis, machine learning, decision theory, decision analysis, and artificial intelligence. After an early start in Physics, he completed his Ph.D. in Medical Information Science ('90) and an MD ('92) at Stanford. He then briefly moved to UCLA, and then to Microsoft Research, where he is currently Manager of the Machine Learning and Applied Statistics (MLAS) Group. We conducted our interview through the net, with (of course!) several Microsoft Word Attachments enclosed.

`1. David, tell us first about your interest in Statistics? You were initially into Physics and then Medicine. And why your interest in Bayesian analysis? I recall some of your papers pointing out issues concerning 'competing' formalisms like Certainty Factors and Fuzzy Sets.`

Since high school, I have been seeking to answer two important questions: "What is the nature of the universe?" and "What is the nature of intelligence". Entering college, I concentrated on the first one by studying Physics. When confronted with the results of the Stern-Gerlach experiments and its dissatisfying explanations such as the Heisenberg interpretation, I turned to the second question. Being rather naive, I went to Medical School to study the hardware of the brain. I quickly realized I was on the wrong track, but fortunately was studying at Stanford University where the field of AI was just getting started. This field clearly had it sights on understanding the nature of intelligence, so I jumped in.

Before too long, it became clear that the representation and manipulation of uncertainty was central to AI and its applications such as expert systems. At the time, Bayesian reasoning was dismissed as a useful tool for uncertainty management in favor of approaches such as Certainty Factors, Fuzzy Set Theory, and the theory of Dempster and Shafer. At first, I followed the trend, applying these methods to one of my projects, an expert system for automated medical diagnosis. The system performed rather poorly, and so I turned to and old friend from physics: probability. The performance of the system improved dramatically. I still can hear the expert with whom I was working exclaim "What did you change, the system is fantastic!" All of us working on expert systems were quite surprised. This surprise led me to a theoretical investigation of Bayesian reasoning and I quickly found the arguments of DeFinetti and Cox. After that, I was hooked. Looking back, it still amuses me that I "discovered" Bayesian probability through experiment.

My investigations also led me to graphical models–in particular directed acyclic graphical models or Bayes nets, on which work was just getting started. I developed an extension of this graphical model–the similarity network–which was suited to the construction of expert systems for diagnosis. The method worked well and formed the basis of my Ph.D. dissertation as well as two expert-system companies that I co-founded. But then I moved to Microsoft, where experts were more difficult to come by and data was abundant. My work naturally (and quickly) gravitated to the application of Bayesian reasoning for statistical inference. I've been enjoying work in this area ever since.

`2. How was your move to Microsoft? Had you planned an academic career?`

It came completely out of the blue. I had just taken a position as an Assistant Professor at UCLA and was looking forward to continuing my research in Bayesian reasoning. Less than a month after starting the job, I got a call from Nathan Mhyrvold, who had just created a research lab at a (relatively small) company, Microsoft. I couldn't imagine why the company that made mice and DOS would be interested in my work, but Nathan told me about his and Bill Gates' vision of the future of computing. We talked about how this future would abound with AI applications including natural-language dialog, speech recognition, and handwriting recognition, and how research in Bayesian reasoning was an essential step in realizing this future. In addition, Nathan made sure I knew that Microsoft sold software to over a million people, and that the results of my work had the potential to help all of them. I was sold.

`3. How is work at Microsoft Research? Are there suggestions from` `above on what directions your research should follow? Do you miss academic life?`

I honestly don't miss it. Ironically, I think that going to Microsoft increased my productivity as a researcher. With no teaching load, no need to write grant proposals, and almost no committee meetings, I've been able to devote the majority of each day at work to research. And equally important, I've been able to spend much of the time I've gained interacting with product groups–listening to the problems they want to be solved and getting immediate and good feedback on the solutions I propose. These interactions have greatly influenced my work and the work my colleagues (for the better, I think). For example, for years we had been assuming that the Bayesian network was an ideal tool for visualizing relationships among variables. But repeatedly the product teams came back with criticisms. The resolution of these criticisms led to a new and useful graphical model called a dependency network, which now can be found in Microsoft's data-mining and e-commerce products.

`4. I recall an LA Times article in which Bill Gates endorsed heavily Bayesian methods. Is this still the case?`

Yes. And it's important to note that this is not a casual endorsement. Bill is interested and extremely well versed in the technical aspects of my work and, in fact, an extremely wide range of research topics. Bill frequently meets with members of Microsoft Research to discuss their work and to suggest directions for application. On numerous occasions, I've watched him drill down into technical details at a level comparable to experts in the field.

`5. For those around who might not still be aware, could you briefly recall the story behind Office paperclip?`

I'm glad you asked me this question because I'm happy to report that Bayesian methods were not responsible for the annoying nature of the Paperclip, and that the Bayesian component of the Paperclip has survived its demise. The precursor of the paperclip was the Answer Wizard, which I designed for Office 95. This application was based on a simple (naive Bayes) model that contained expert assessments relating words in a help query to the help topic sought by a user. For example, the model included the assessment that the typed phrase "turn sideways" is likely when the user is seeking the help topic pertaining to landscape/portrait printing. The Answer Wizard was a big hit in Office 95 and was one of its top two selling features. But then came Office 97. In this version of the product, the Answer Wizard was combined with a character-based UI and a rule-based system that would pop up tips such as "I see you are writing a letter, let me help you....". It was this UI and its rule-based behavior that many of us came to dislike (to put it nicely). The Paperclip survived Office 2000, but was finally "retired" in Office XP. At the same time, the Answer Wizard was restored to a stand-alone component with a simple and well-exposed UI. If you are using Office XP, just look at the upper-left-hand corner of any application screen. You'll see a box that says "Type a question for help". This is the Answer Wizard.

`6. What other Bayesian developments have ended up within Microsoft technology?`

In Windows 98, Windows ME, and Windows XP you'll find a series of troubleshooters that help you fix problems such as the inability to print and the inability to install a particular application. These troubleshooters are based on expert assessed Bayesian networks. One interesting point about our approach is that we were careful to construct *causal* Bayesian networks, as these networks are used to model the consequences of (repair) actions.

Both SQL Server 2000 and Commerce Server 2000 contain data-analysis tools that use Bayesian statistical techniques. For example, we use Bayesian model selection to learn the structure of dependency networks found in both these products. Although you won't find much documentation talking about the Bayesian nature of these tools, we did put an Easter egg into both products to pay homage to the Bayesian approach. An Easter egg is a hidden credit screen–most products have them, but they are next-to-impossible to discover unless you know how to find them. To see the Easter egg in either of these products, go to the dependency-network viewer, click on the find-node button, type in "Rev. Bayes!", and click the "cancel" button.

And there will be plenty more applications in the future. For example, Microsoft's upcoming Japanese handwriting recognizer will contain a Bayesian component.

`7. What are the current interests and efforts at your group? Do you strictly adhere to the Bayesian postulates? For our youngest readers, what is the way to enter your group (or other statistically related Groups) at MR?`

I can't imagine an approach better than the Bayesian one for representing and managing uncertainty in an expert system or for combining prior knowledge with relatively small amounts of data. Furthermore, Bayesian model-selection techniques seem to be ideal for the analysis of massive data sets, because they are extremely computationally efficient. In other situations, however, I think there is room for deviations from strict Bayesian practice. One difficulty with the Bayesian approach, for example, is the often-made assumption that "truth" lies within at least one of the models under consideration. (This problem is discussed, for example, in Chapter 6 of Bernardo and Smith, 1994.) In addition, there is empirical evidence that non-Bayesian methods such as boosting outperform their Bayesian counterparts. Of course, I would not be surprised if Bayesian versions of such methods are found once they are better understood.

In my group at Microsoft Research, we have a collection of great people (see `research.microsoft.com/~mlas`) working on a variety of theory and applications. More theoretical interests include variational techniques for approximating Bayesian inference in complex models, characterizations of priors based on graphical-model properties, the study of algebraic properties of graphical models, and computationally efficient approximations to Bayesian model selection, Bayesian model selection, and Bayesian mixture modeling. Application interests include data mining and customer relationship management, user personalization and targeted advertising on the internet, speech recognition, handwriting recognition, and statistical natural-language processing.

`8. In a former interview, Jim Berger mentioned that (academic) statisticians were at risk of losing emerging fields (like data mining) to other disciplines like IT. What is your view?`

I can see this happening only if these emerging areas are actively avoided by statisticians. Take your example of data mining. In my opinion, data mining is ten percent computer science and ninety percent statistics. I do not mean to disparage computer science, but there are hundreds of years of research behind statistical methods relevant to data mining. I believe that data-mining practitioners eventually will come to appreciate this work and embrace statistics.

`9. What emerging directions do you see for Bayesian Statistics, in the Internet era?`

There are many. Areas that I've worked on include user personalization (for example, recommender systems), visualization and exploration of site activity, categorization of web pages, online advertising and marketing, and intelligent searching and filtering. Other promising areas include the modeling of web structure, automatic extraction of information from web pages, and security management.

`10. Finally, any new things for the Valencia 7 conference?`

Yes indeed. For the first time, the conference will be held on the beautiful island of Tenerife in the Spanish Canary Islands and will be jointly sponsored by the University of Valencia and ISBA. Also for the first time, the scientific program will include selected contributed papers, to be presented orally. Finally, we've added two exciting themes to the program: bioinformatics and the computer-science/Bayes interface.

Thanks David for such informative answers.

As we mentioned, this is our last interview for the ISBA bulletin. We are very grateful to our interviewees for their enlightening responses and Fabrizio Ruggeri for his patience with missed deadlines. See you all soon!!!

David Rios Insua and Mike Wiper

BMCMC: Mel Novick's CADA Project

by George G. Woodworth
george-woodworth@uiowa.edu

In the post-Sputnik summer of 1961, the NSF though it was important to recruit undergraduates into careers in science and mathematics. I had just finished my junior year at Carleton College with a statistics course taught by the gifted mathematical educator Frank Wolf under my belt (www.carleton.edu/campus/news/pr/wolf.html). Frank called my attention to Richard Savage's NSF-funded undergraduate research participation program at the University of Minnesota, to which I applied and was accepted. That was the summer I became a Bayesian.

A year later, I entered the graduate mathematical statistics program at Minnesota; however, apart from a seminar talk or two, I don't recall any Bayesian content in the course offerings, and my dissertation was entirely frequentist.

As an assistant professor at Stanford, I took my turn teaching the large undergraduate statistics service course using Chernoff and Moses' excellent and entirely Bayesian *Elementary Decision Theory*. That experience, along with reading Raiffa's text, *Analysis of Decisions Under Uncertainty*, revived my dormant Bayesianism, although I continued to work out the frequentist ideas in my dissertation.

I took an appointment at the University of Iowa in 1971 and shortly thereafter met Melvin R. Novick. Mel was a Bayesian's Bayesian. Dennis Lindley (1987) put it this way, "All Novick's statistical work was within the Bayesian paradigm. His object was to ally the vigorous, coherent theory with sensible data analysis." Mel's 1963 Ph.D. dissertation at the University of North Carolina at Chapel Hill was on indifference priors. His last publication in 1986 was an application of m-group (exchangeable) regression (Lindley, 1987).

After 7 years at Educational Testing Service in Princeton with visiting lectureships at the University College of Wales, Aberystwyth, and University College, London, Mel accepted a permanent academic position at the University of Iowa in 1970. He quickly developed a course sequence in Bayesian statistical methods and, according to Feldt and Hogg (1987), "When it became apparent that the application of Bayesian methods would be inhibited in many settings by computational complexities, he initiated development of an interactive computer package called Computer Assisted Data Analysis (CADA)."

I first encountered Mel when he spoke at our Departmental seminar. I made some sort of remark about preposterior analysis and I think that gave him the idea of inviting me to join the CADA project - Bayesians being rather thin on the ground at that time. With a growing family and a liberal arts salary, I was happy to accept the offer of summer support.

Twenty-first century students of Bayesian inference will find it difficult to imagine what it was like to try to do practical Bayesian analysis in the 70's and early 80's. At that time interactive computing was done on a "mini" computer. "Mini" meant about as big as a refrigerator. At the University of Iowa, we had access to Hewlett-Packard 2000 minicomputers, initially via Teletype terminals, later via "dumb" terminals with CRT displays. The HP's native language was BASIC and the user had access to 32 kilobytes of RAM. That allocation had to hold everything - program and data. Longer programs had to be chained – broken into functionally independent parts that were loaded into memory in sequence. The adaptation of Markov Chain Monte Carlo (MCMC) to statistical computation was about a decade in the future so numerical integration in CADA was done by classical methods such as Gauss-Hermite quadrature, or by approximations.

Despite these obstacles, Mel had a clear vision of what was needed to facilitate practical Bayesian statistical analysis and was determined to implement it. He insisted that computations had to be interactive (his term was "conversational"), menu-driven, and flexible. According to Lindley (1986), "Novick, earlier than most of us, had recognized that Bayesian statistics required its own computer packages... The Bayesian - being required to think in probabilistic terms about the data - cannot be content with routines that merely perform a calculation. The Bayesian has to have an interactive package in which the user thinks and the computer calculates showing the user the coherent consequences of his thought."

By its 1983 release, the CADA package included a sophisticated data management facility, including transformations. The package provided analyses of simple parametric models including interactive elicitation of informative priors (beta-binomial, beta-mixture-binomial, comparison of two beta-distributed variates, Dirichlet-multinomial, conjugate and non-conjugate two-parameter normal models, comparison of two t-distributed variates, and simple linear regression). The multiple regression component offered interactive prior elicitation (Kadane, et al 1980). I wrote the multivariate general linear model (MGLM) component (Woodworth 1979), which allowed only non-informative priors (Box and Tiao, 1973). Factorial ANOVA and MANOVA were shells that called the MGLM component and hence allowed only non-informative priors as well.

A "simultaneous estimation" component implemented hierarchical models with exchangeable components: exchangeable proportions, exchangeable normal means (with equal precisions), and simple regressions (with exchangeable intercepts and equal slopes); this component included interactive elicitation of informative priors. Apart from these fairly general statistical analyses, CADA also included fully Bayesian subjective expected utility (SEU) solutions to educational or employment selection and assignment, with interactive elicitation of priors and utility functions.

In addition, CADA offered non-Bayesian exploratory data analysis (EDA), psychometric methods, and actuarial calculations. Although the list of Bayesian analyses implemented in CADA seems short by modern standards it did have a feature which contemporary general-purpose Bayesian packages such as WinBUGS lack: interactive elicitation of informative priors.

The 1983 version of CADA was available for DEC PDP-11, VAX-11, HP 2000, HP 3000, PR1ME, RT-11, and IBM VS/370 dialects of BASIC. It was installed at over 50 sites. Mel and his son Raymond were working on a PC version at the time of Mel's untimely death in 1986.

Mel Novick was an inspiring colleague. He showed me that Bayesian methodology was not just another option in the statistician's toolkit - it is in fact the only available coherent, internally consistent epistemology. He knew that practical Bayesian analysis needed interactive computer-intensive support and he had the courage and energy to try to provide it with the tools at hand.

➤ References

BOX, G.E.P. and TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

FELDT, L.S. and HOGG, R.V. (1987). Melvin R. Novick: Recollections of Two Iowa Colleagues. *Journal of Educational Statistics*, **12**, 7-10.

KADANE, J.B., DICKEY, J.M., WINKLER, R.L., SMITH, W.S. and PETERS, S.C. (1980). Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association*, **75**, 845-854.

LINDLEY, D.V. (1987). Melvin R. Novick: His Work in Bayesian Statistics. *Journal of Educational Statistics*, **12**, 21-26.

WOODWORTH, G.G. (1979). Bayesian Full Rank MANOVA /MANCOVA: An Intermediate Exposition with Interactive Computer Examples. *Journal of Educational Statistics*, **4**, 357-404.

SPECIFYING AND DOCUMENTING PRIORS IN HIERARCHICAL MODELS

by Kate Cowles
kcowles@stat.uiowa.edu

Hierarchical models have become ubiquitous in Bayesian statistical practice for both scientific and pragmatic reasons. They make it possible to model some complex real-world phenomena in terms of sequences of fairly simple components. Furthermore, the Gibbs sampler and its extensions are ideally suited to fitting hierarchical models.

In order to obtain a proper joint posterior distribution, hierarchical models generally require proper priors on some model parameters (e.g. variances of unobservable subpopulation-specific random effects). Gelfand and Sahu (1999) discuss two conflicting issues involved in specifying these priors. (1) Propriety is not enough if MCMC sampling is used to fit the model, as overly-vague priors can cause excruciatingly slow or uncertain sampler convergence. (2) However, strongly informative priors on these parameters can drive posterior inference on other model parameters that may be of scientific interest.

Natarajan and McCulloch (1998) demonstrate point (1), showing that in the context of hierarchical probit models, "the use of the Gibbs sampler with diffuse priors can give inaccurate posterior estimates." Alternative non-conjugate priors that may mitigate these issues have been proposed but as yet have not come into widespread usage, due at least in part to the interpretability and computational convenience of more standard priors (such as inverse gamma for variances and inverse Wishart for covariance matrices).

The preceding discussion highlights the crucial importance of reporting how proper priors were specified, what they mean, and what impact they have on posterior inference, whenever hierarchical models are used to address real-world questions. The present article features an applied Bayesian paper in which the documentation of these issues is done in an exemplary way.

Wakefield, Aarons, and Racine-Poon (WARP) (1999) present complex hierarchical models for population pharmacokinetic (PK) and pharmacodynamic (PD) analysis. PK studies model the concentration of a drug in body fluids or tissues as a function of dosage and sampling times. Typical PK parameters of interest are the absorption, distribution, and elimination of the drug. PD studies relate the concentration of the drug to its action on the body. Population PK studies the variability in concentration-time profiles between individuals when standard dosage regimens are used. They may use very sparse data from individual patients.

We focus on the part of WARP's paper that concerns fitting a population PK model to data from Phase II clinical trials of REVASC^{TM}, a drug that inhibits clotting.

➤ **First stage**

The first stage models measurements y_{ij} of plasma concentration of REVASC^{TM} for each patient i as a function of the patient's PK parameters $\theta_i = logCl_i, logV_i, logk_{a,i}$ and the dose size D_i:

$$logy_{ij} = logf_1(D_i, \theta_i, t_{ij}) + \epsilon^y_{ij}$$

where y_{ij} is the drug concentration of individual i at time t_{ij} and errors ϵ^y_{ij} are assumed i.i.d $N(0, \sigma^2_y)$.

Due to sparsity of the data (only two concentration measurements per patient), a simple one-compartment model was used at the first stage, i.e.:

$$f_1(D_i, \theta_i, t_{ij}) = \frac{D_i k_{a,i}}{V_i(k_{a,i} - \frac{Cl_i}{V_i})} \sum_{l=1}^{2} \left(\frac{exp(-\frac{Cl_i}{V_i}(t_{ij} - t_l))}{1 - exp(-Cl_i \Delta_l / V_i)} - \frac{exp(-k_{a,i}(t_{ij} - t_l))}{1 - exp(-k_{a,i}\Delta_l)} \right)$$

Here t_l is the time since does l was given, Δ_l is the dosing interval, and the individual PK parameters for patient i are Cl_i (clearance rate), V_i (volume of distribution), and $k_{a,i}$ (absorption rate).

➤ **Second stage**

The second stage of the kinetic model involved covariates and distributional assumptions for the individual level PK-parameters. Biological considerations suggested that a patient's estimated creatinine clearance and weight would be strong predictors of the Cl_i and V_i respectively, yielding expressions:

$$logCl_i = \mu_{\theta 0} + \mu_{\theta 1} X_{1i} + \delta^\theta_{1i}$$

$$logV_i = \mu_{\theta 2} + \mu_{\theta 3} X_{2i} + \delta^\theta_{2i}$$

The intended second stage prior for $logk_{a,i}$ was

$$logk_{a,i} = \mu_{\theta 4} + \delta^\theta_{3i}$$

The authors found, however, that there was very little information in the data about the absorption rates, because there were no early sampling times — i.e. no measurements were taken while the drug was being absorbed. In the following paragraph (p. 239), they describe their resulting efforts to find a workable second-stage prior for the $k_{a,i}$:

`[The above model] was assumed with very tight priors being placed on` $\mu_{\theta 4}$ `and the variance of the` δ^θ_{3i}`. The variance on the prior for` $\mu_{\theta 4}$ `corresponded to a change in` k_a `of` $\pm 5\%$`. The posterior for` $\mu_{\theta 4}$ `from this analysis was located at an unreasonably high value, however, because even with a strong prior the data were too sparse to discount the intravenous model which corresponds to an infinite` k_a`. The same behavior occurred when` k_a `was treated as a fixed effect, again with a tight prior. Two final analyses were carried out. In the first of these` k_a `was allowed to take the single value exp(-1.2). This latter was chosen from [three previously described Phase I studies of the same drug] where the posterior mean of the population` $logk_a$ `was -1.2. In the second analysis` k_a `was allowed to take one of five discrete values, centred on exp(-1.2) and with a spread of` $\pm 5\%$`.`

To complete the second stage, the pairs $(\delta^{\theta}_{1i}, \delta^{\theta}_{2i})$ were given a bivariate normal prior with mean zero and covariance matrix Σ^{θ}.

➤ Third stage

In an earlier section of the paper (p. 223), describing the joint PK/PD models to be used, WARP stated the following with respect to requirements on the third-stage priors. (References to parameters from the PD part of the model have been deleted.)

`At the third stage of the hierarchy priors are specified for the population parameters` μ_θ, ... Σ_θ, σ^2_y .... `We first note that [for] those elements of` μ_θ ... `that correspond to nonlinear parameters at the first stage, proper priors are required in order to guarantee propriety of the posterior distribution. For the same reason we also need proper priors for` Σ_θ .... `For computational convenience ... [a] normal prior distribution` $N(c_\theta, C_\theta)$ `...[is] assumed for` μ_θ ...; `[a] Wishart distribution` $W(r_\theta, (r_\theta R_\theta^{-1})$ ... `[is] assumed for` Σ_θ^{-1} ...; `and [a] gamma distribution` $Ga(a^2_\theta/A_\theta, a_\theta/A_\theta)$ ... `for` σ_y^{-2} ....

Detailing how appropriate parameters of these proper priors were specified, the authors state (p. 240):

`On the basis of [five previous] studies using the three compartment model we can obtain prior distributions for` $\mu_{\theta 0}$ `(clearance intercept) and` $\mu_{\theta 2}$ `(volume intercept) based on predictions from the Phase I studies. There were also three additional studies which we have not described in which weight and creatinine clearance were measured. Hence prior estimates of the regressors describing these relationships [with REVASC`^{TM} `clearance rate and volume of distribution] were also obtained. This was done by first simulating a large number of individuals from the posterior distribution of the population parameters. For each of these individuals we calculate the clearance and the volume at steady-state, and the the mean and the variance/covariance matrices of these quantities were then evaluated to give` $c_\theta$ = `(2.41, 0.0101, 3.56, 0.095). The variance of the prior` $C_\theta$ `was taken to be a diagonal matrix with the diagonal elements 0.09. The prior estimate for` $\mu_{\theta 0}$ `corresponds to a clearance value of 11 litres/hour for the patient population. The prior estimate of` $\Sigma_\theta$, $R_\theta$ `was taken to be a diagonal matrix with diagonal elements 0.04. These values correspond to a coefficient of variation on clearance and volume of 20% which is typical for studies such as these. We take` $r_\theta = 2$, `which is the smallest value that gives a proper prior, and choose` $A_\theta = a_\theta = 0$, `[as] with 602 observations there is sufficient information in the data to estimate` σ^2_y.

WARP then present results of their analysis, followed by brief mention of sensitivity analyses of the effects of altering the fixed value assigned to k_a and specifying a different prior on $\mu_{\theta 0}$.

In short, the WARP paper is an example of how to carry out and report a serious substantive study using hierarchical models. The need for proper informative priors on parameters that are not identified in the likelihood is explicitly addressed, and the procedures used to obtain appropriate priors from biological understanding and previous studies are clearly described. The interpretations of the resulting priors are expressed in terms understandable by a layman. Finally, sensitivity analyses of the influence of these priors were carried out and the results described.

➤ References

GELFAND, A.E. and SAHU, S.K. (1999). Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models. *Journal of the American Statistical Association*, **94**, 247-253.

NATARAJAN, R. and MCCULLOCH, C.E. (1998). Gibbs Sampling With Diffuse Proper Priors: A Valid Approach to Data-Driven Inference? *Journal of Computational and Graphical Statistics*, **7**, 267-277.

WAKEFIELD, J., AARONS, L. and RACINE-POON, A. (1999). The Bayesian Approach to Population Pharmacokinetic/Pharmacodynamic Modeling. In *Case Studies in Bayesian Statistics, Vol. 4* (C. Gatsonis, R.E. Kass, B. P. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, M. West eds.), Springer-Verlag.

RICHARD LEWIS TWEEDIE

Richard Tweedie, Professor and Head of the Division of Biostatistics at the University of Minnesota, died suddenly on June 7, 2001. His wife Catherine, their daughter Marianne, his mother Nel, sister Joan and brother Bruce survive him.

His contribution to the fields of applied probability modeling and statistical science was outstanding and was tragically cut short with his death at age 53. He was an international scholar who also successfully worked in the private sector as well as being an outstanding university administrator. He was a wonderful teacher and mentor to students. His ability to inspire students and colleagues to do their best work was widely recognized and will be remembered by those of us who were fortunate enough to have worked or studied with him.

Richard was born in Leeton, an agricultural community in the Australian state of New South Wales. He obtained a BA in statistics with First Class Honours and a MA degree by research from the Australian National University. In 1972 he was awarded his PhD at Cambridge under the supervision of David Kendall on the topic of Markov Chains. In 1986 he was awarded the Doctor of Science degree from ANU based on his major research achievements.

Richard's professional career was rich and surprisingly varied. He had held the following positions: Postdoctoral Fellow in the Statistics Department in the Institute of Advanced Studies at the Australian National University from 1972 to 1974; varying positions in Canberra and Melbourne with the Division of Mathematics and Statistics at the Commonwealth Scientific and Industrial Research Organisation from 1974 until 1981 where he finished as Senior Regional Officer managing the Victorian region; Associate Professor in the Department of Mathematics at the University of Western Australia in 1978; General Manager and Managing Director of Siromath, a mathematical and statistical consulting company from 1981 to 1987; Foundation Dean and Professor of Information Sciences at Bond University, Australia's first private university from 1987 to 1991; Professor and Chair in the Department of Statistics at Colorado State University from 1991 to 1999; Professor and Head of the Division of Biostatistics at the University of Minnesota from 1999 until his death.

Although for much of his career Richard managed a heavy administrative load he continued his extraordinary research output throughout. Beginning with his PhD work and continuing through his whole career he contributed fundamental results in the theory of Markov Chains on general state spaces. A major focus of this research was development of stability theory for such processes that could be readily applied in modeling a range of complex phenomena arising in diverse fields such as queuing theory, systems theory, statistical modeling including time series and decision theory. A particular achievement, of which Richard was justifiably proud, was publication of his book with Sean Meyn entitled Markov Chains and Stochastic Stability in 1993 by Springer-Verlag and given the 1994 Operations Research Society of America's award for the best research publication in Applied Probability.

From about 1994 onwards Richard's developing interests in statistical methods in the epidemiology of the health effects of exposure to environmental and occupational pollutants led to research in Bayesian modeling for meta-analysis and publication bias. This also stimulated his interest in the question of convergence rates for MCMC sampling schemes, a major focus of his most recent research and publications.

Richard is known by many as a brilliant theoretician. However throughout his career he also worked with numerous collaborators on a range of interesting and challenging practical problems. The seeds of his interests in scientific collaboration and commercial consulting were sown during his time at CSIRO and he pursued these throughout his career. He was a wonderful consultant and communicator with a knack of explaining complicated statistical ideas to practical people. His consulting and collaborative work led to a substantial number of publications in a diverse range of topics and journals.

Richard was also devoted to the statistical profession and, no matter how busy he was, he found time to involve himself seriously in the organization, promotion and scholarship of professional societies. His substantial work with the Statistical Society of Australia was rewarded with Honorary Life Membership in 1998. He was elected to the International Statistical Institute in 1980 and was a member of its Bernoulli Society since that year also. He was elected as Fellow of the Institute of Mathematical Statistics in 1989 and as Fellow of the American Statistical Association in 1997. He served as Editor of the Annals of Applied Probability and of Statistical Science and was an Associate Editor for a number of other leading journals.

Richard was a warm and generous friend and colleague. A great conversationalist and an outstanding humor and quick wit made being in his company both memorable and stimulating. His numerous friends throughout the world will sadly miss him.

William T.M. Dunsmuir
Division of Biostatistics, University of Minnesota

An Overview of the Teaching of Bayesian Statistics in Chile

by Pilar L. Iglesias
and Victor H. Salinas
pliz@mat.puc.cl
vsalinas@fermat.usach.cl

➤ Introduction

This work summarizes the main conclusions obtained during the recent forum on the teaching of Bayesian Statistics in Chile as part of the Third Chilean Seminar on Bayesian Statistics (Valdivia, Chile, January 1998), supported by the Chilean Chapter of ISBA. The main motivation for this forum was the absence of courses of this type in in the academic curricula and the international discussion of this issues. Comments are included based upon the experience gained by the authors, who have taught courses on Bayesian Analysis in the last five years, both at the graduate and the undergraduate level. Our intention is to add some new arguments to the discussion initiated by Moore (1997) and Albert (1997a). Finally, we give some ideas about the teaching of Bayesian methods in High School.

➤ Teaching of Bayesian Statistics in an Undergraduate Level

The inclusion of courses in Bayesian Statistics in undergraduate programs in Statistics is quite recent in Chile. Furthermore, only a few universities offer courses in this area and they are elective. There are about eight universities offering a professional degree in Statistics, many under the name of Statistical Engineering (Iglesias, 2001). This is tied to the existence of research groups in this area, as is the case of Universidad Católica de Chile and Universidad de Santiago de Chile. Depending to the space available in the curricula, two types of courses have been proposed. The first falls in the context of Statistical Inference, following a Probability Calculus. This alternative was the one supported by the participants at the forum, particularly when the only alternative to teach Bayesian Methods is within the only course in Statistical Inference.

Nevertheless, our experience shows that this alternative has been successful, even when the curriculum contemplates specific courses in the field. The idea is to take into account both classical and Bayesian ideas, giving to the students elements to make there own judgements. A course with these characteristics includes estimation and hypothesis testing under an integrated approach, as presented in De Groot (1988) and Migon and Gamerman (1999).

The second type of course, which does not preclude teaching the first, is conceived for those programs which have at least one course in the area, under the perspective of Decision Theory. The course includes simple applications to Engineering, Economics, and Medicine. The softwares *First Bayes* (O'Hagan, 2000), Minitab (Albert, 1997b) and WinBUGS (Spiegelharter, et al., 1995) have been used as a first approximation to solving problems with real data.

On the other hand, the forum discussed the inclusion of Bayesian methods in service courses. The contents of these courses are often suggested from specialists in the subject matter, with the side problem of covering too many topics and placing more emphasis in methods and their use, rather than on the fundamental concepts. Incidentally, some participants mentioned to have encountered students who cannot distinguish between a parameter and its estimate.

➤ Teaching of Bayesian Statistics in Graduate Programs

There are four Chilean universities offering an M.Sc. in Statistics and one that offers an M.Sc. in Biostatistics. The only doctoral program is that of the Pontificia Universidad Católica, which started in 1998, which was built upon the M.Sc. program, offered since 1987. At this university, the first course on Bayesian statistics is common to the M.Sc. and the Ph.D. programs, and it does not assume any prior knowledge, since the M.Sc. is open to applicants from different areas, which have taken the traditional sequence in Probability and Inference. The aim is to provide the student with a general vision of Bayesian Statistics and to give basic tools to understand the literature at the methodological level. During the course, classical and nonstandard solutions are contrasted. The references used are rather varied. For instance, the introduction to the fundamentals of Statistics and Decision Theory follows Bernardo (1981). Next, with the aim of illustrating the application of Bayes formula and of introducing the inference problem for proportions in finite populations, using the concepts of exchangeability, as presented in Albert (1997b, chapter 9). Results on conditional independence and De Finetti representation theorems are described, in the perspective of Dawid (1979) and Heath and Sudderth (1976). The course continues with an analysis of parametric models, following Lee (1989), Berry (1997), Gelman et al. (1995) and Albert (1997b). An elementary exposition of these topics helps the student to understand better the origin of models and parameters. MCMC methods are included in a general Computational Statistics course. Among the references used is Robert (1994, Chapter 9), Robert and Casella, (1999) and Quintana (2000).

At the doctoral level an Advanced Bayesian Statistics course is offered, whose contents changes according to the interests of the students.

➤ Bayesian Statistics in High School

In Chile, the incorporation of statistical topics in the Mathematics curriculum at High School only starts with the educational reform gave an impulse by the Secretary of Education in 1997. The Chilean Statistical Society was invited to express its views, which are summarized in del Pino et al. (1996). Various programs addressed to teachers and students were developed, with the support of this organism government and scientific organizations. The contents include probability calculus, exploratory data analysis, and an introduction to statistical inference, with the traditional approach. These contents are scattered within the last three years of High School.

Our experience, obtained from various projects related to the teaching of Statistics (Aravena et al. 2001 and Iglesias et al., 2001) lead us to the conclusion that we are still very far from introducing Bayesian ideas at this level. On the one hand, the instructors of these courses are teachers of Mathematics, with at most a very light knowledge of Statistics. This causes most statistical contents to be omit. Thus Statistics is not presented as an independent discipline.

The inclusion of Bayesian methods in High School is a debatable issue (Ito, 2000; Moore, 1997). Some topics in this level has been tried by the authors with groups of students from the last year of High School, who have some notions of probability. When this topic has been treated in terms of an everyday decision problem, the students have responded quite satisfactorily. They do not appear to have problems with the assignment of probabilities to the unknown quantities involved in the decision problem. Furthermore, the computation of expected utility (obviously not using this name) was found to be a natural criterion for decision making by a substantial part of the students.

➤ Conclusions

The teaching of Bayesian Statistics at the undergraduate and graduate levels has generated a relevant activity, which ha experimented a sustained growth in our scientific community. This has translated into an important increment of the number of published articles in this area. Articles with a Bayesian content have reached about 50% of the global production in the last two years. An increase in publications considering both classical and nonstandard solutions in different areas of application, as Environment, Finance, and Medicine. There is also an increase in the number of theses at different levels, discussing classical and Bayesian solutions of theoretical and applied problems, with an emphasis in the computational implementation.

The continuity of our activity, will undoubtedly allow a significant increase in our critical national research mass. As a natural consequence, the teaching of Bayesian statistics, will have a natural link to research, and it will have an impact on teaching at all levels, even as part of quantitative literacy.

Acknowledgements: The authors thank Professor Guido E. del Pino from Pontificia Universidad Católica de Chile for reading a first version of this manuscript and for his useful comments.

➤ References

See `www.mat.puc.cl/~pliz`

BRCAPRO

A MODEL AND SOFTWARE FOR GENETIC COUNSELING OF WOMEN AT HIGH RISK OF HEREDITARY BREAST AND OVARIAN CANCER

by William Browne
w.browne@ioe.ac.uk

In the last issue of the ISBA Bulletin a review was given of the software packages that Bayesian statisticians use to perform MCMC analyses. Bayesian methods are now becoming far more widely used by practitioners from a wide varied of research backgrounds. Software for Bayesian statistical analysis can be classified on a sliding scale that mirrors the variety of statistical and computing abilities of these practitioners.

At the technically proficient end of the scale people will use standard computer languages such as C, C++, Fortran and Java to program their own software for their problems, perhaps with the use of some random number generator libraries or programs such as the BACC suite of functions. Next down this scale are the general purpose packages such as WinBUGS that aim to be hugely flexible in the number of models they can fit but to accommodate this aim have a 'pseudo' programming language type interface and often rely on generic single site updating algorithms. There are then specific purpose packages such as BayesX and MLwiN which aim to fit limited families of statistical models. Due to these limitations they can often then use more efficient, faster algorithms or have an easier 'user-friendly' interface to cater for less statistical users. Finally there are the single purpose pieces of code written by individuals for their own use or to perform the analysis for a journal article.

BRCAPRO stems from the latter category and is a piece of C code written by the authors to solve problems described in Parmigiani et al. (1998) and Berry et al. (1997). The code performs a Bayesian analysis that calculates the probabilities of carrying mutations of the breast cancer-susceptibility genes BRCA1 and BRCA2 for an individual based on their family history and other factors. Over the last ten years or so major advances in the understanding of (inherited) susceptibility to both breast and ovarian cancers have been made, primarily with the identification of the BRCA1 and BRCA2 genes. Mutations in one or both of these genes will give the carrier a much greater risk of developing both breast and ovarian cancer. Tests are available for mutations in both these genes but are expensive and a positive outcome can affect eligibility to health insurance and possibly result in employment discrimination.

The BRCAPRO program calculates for an individual (the counselee) the probabilities of carrying the two genes BRCA1 and BRCA2 based on many factors. These include the age of the counselee and their cancer status, the age or age at death, sex and cancer status for relatives of the counselee and the nature of their relation. For any cancer cases in the counselee's pedigree the type of the cancer and the age of the individual at occurrence of the cancer are also used as predictors. Finally the results of any tests for mutations in the two genes that have been carried are also included in the model. From the probabilities of gene mutation the program can also calculate probabilities of the counselee developing both breast and ovarian cancer in the future and gives predicted probabilities at five year intervals.

It is possible to run BRCAPRO as stand-alone C code with suitable input files. (Note the web site gives a manual including input file information for UNIX users.) However BRCAPRO is included as part of the (freeware) package CancerGene (version 3.3 used here), which is developed and maintained by David Euhus at the University of Texas Southwestern medical center and it is through this program that I have experimented with BRCAPRO. CancerGene is a package that allows the user to input and store a family history in the form of a pedigree. It can then calculate many probabilities from this data include age-specific breast cancer probabilities and BRCA mutation probabilities using either BRCAPRO or other (non-Bayesian) methods. Although the package does not come with a user manual it contains an information option that gives the user much background information on both the package and the models it fits.

The user can construct the pedigree for the counselee by simply filling in boxes giving details of the individual's age, sex and cancer status including when the cancer first occurred. Then the pedigree is constructed via an 'Add relative' button which gives the user a list of possible relationships for example parent, sibling. Then similar details for this new individual must be added. If a mistake is made it is easy to either modify details for an individual or remove them from the family tree.

Once we are happy with our pedigree we can click on a 'Done' button, which firstly saves the data for future use and then calculates the probabilities. This in my limited experience never takes more than a few seconds. The computations involve calculating the posterior distribution using Bayes theorem after constructing the marginal likelihood with respect to the genotypes of all the relatives of the counselee. The results are then given in well-presented tables and bar graphs plotting probabilities of getting breast and ovarian cancers against age.

The software is a good example of how to allow non-statisticians to use Bayesian techniques. I have a couple of suggested potential improvements based on my experience as a Bayesian software developer. Firstly it would be useful for the naive user to have an example pedigree in the information system that they could try and replicate. I tried to replicate the results given in the `UNIX` manual for an example pedigree using `CancerGene` and got slightly different answers.

This it turns out is due to the developers (as good Bayesians) periodically updating the inputs to the program as the literature in this field grows. Iversen Jr. et al. (2000) gives the latest updates used by the program. In fact the authors are currently involved in a multicentre study, sponsored by the Cancer Genetics Network to validate their model on over 3,000 pedigrees from 10 medical centres in the US. However an example analysis in the documentation (which could be updated as the package is updated) would give the user more confidence in their ability to use the package.

Secondly all probabilities give no measure of uncertainty (unlike the analysis in the article in Parmigiani et al., 1998), and are often very different to those given by other methods. Some documentation on why this occurs would also be useful and the authors intend to remedy this in the future.

All in all, `BRCAPRO` through `CancerGene` gives an interesting introduction into Bayesian human genetics modelling.

➤ Website

`biosun01.biostat.jhsph.edu/~gparmigi/brcapro.html`

➤ References

BERRY, D.A., PARMIGIANI, G., SANCHEZ, J., SCHILDKRAUT, J. and WINER, E. (1997). Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *Journal of the National Cancer Institute*, **89**(3), 227-238.

IVERSEN JR., E.S., PARMIGIANI, G., BERRY, D.A. and SCHILDKRAUT, J. (2000). Genetic susceptibility and survival: Application to breast cancer. *Journal of the American Statistical Association*. **95**, 28-42.

PARMIGIANI, G., BERRY, D.A. and AGUILAR, O. (1998). Determining carrier probabilities for breast cancer susceptibility genes BRCA1 and BRCA2. *American Journal of Human Genetics*, **62**, 145-158.

NONPARAMETRIC BAYESIAN FUNCTION ESTIMATION

by Jean-François Angers
`angers@dms.umontreal.ca`

Nonparametric Bayesian estimation of a function is an interesting problem, which has been studied by several authors. In this annotated bibliography, I will review some classic papers in this area along with more recent ones. Related problems such as density estimation and intensity function estimation are not covered. (Sorry if I offended anyone by omitting his/her work related to this problem.)

The following model is assumed:

$$y(t_i) = g(t_i) + \varepsilon_i, \qquad (1)$$

where $i = 1, 2, \ldots, n$, g is the unknown function to be estimated, t_i belongs to some set $\mathcal{T}$ and ε_i represents the error term. The unknown function will be expressed in the form $g(x) = \sum_j \theta_j b_j(x)$, where the functions b_j form a basis. Different bases are considered here: polynomials, splines and wavelets. An "other" category is also considered.

➤ Polynomial basis

A polynomial basis is obtained by expanding the unknown function g using Taylor's series around some point-of-interest.

• S. WEERAHANDI AND J.V. ZIDEK (1988) **Bayesian nonparametric smoothers for regular processes**, *The Canadian Journal of Statistics*, **16**, 61–74.

In this paper, the function is observed at n time points and g is assumed to be a "smooth" Taylor-expandable function around t_{n+1}, that is g is at least p times differentiable at $t = t_{n+1}$. Hence, equation (**??**) can be written as

$$\vec{y} = \mathbf{X}\vec{\beta} + \vec{\varepsilon},$$

where $\beta_i = \frac{\partial}{\partial t} g(t)|_{t=t_{n+1}}$. In their set up, the ε_i's are stochastically dependant on $\vec{\beta}$. However, a transformation on the data can be made so that $\vec{\beta}$ and $\vec{\varepsilon^*}$ are uncorrelated. Using standard normal model with the appropriate covariance structure on $\vec{\beta}$ and $\vec{\varepsilon^*}$, the Bayesian estimator of $\vec{\beta}$ is obtained. In order to specify the hyperparameters, an empirical Bayes paradigm is used. Some sensitivity analysis is done for the order of the Taylor's expansion.

• J.-F. ANGERS AND M. DELAMPADY (1992) **Hierarchical Bayesian Curve Fitting and Smoothing**, *The Canadian Journal of Statistics*, **20**, 35–49.

The authors use a similar model as in Weerahandi and Zidek (1988), however, a hierarchical Bayes paradigm is used and a vague prior is assumed for the hyperparameters. A procedure to choose the order of the Taylor's expansion is presented and a more extensive sensitivity study has been done on the choice of the prior on the hyperparameters.

➤ **Spline basis**

This basis is useful especially when one is not sure that the true function can be expressed as polynomials.

• G. WAHBA (1978) **Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression**, *Journal of the Royal Statistical Society, Series B*, **40**, 364-372.

In this paper, $\mathcal{T} = [0, 1]$ and Wahba considers the problem of finding g which minimizes

$$n^{-1} \sum_{i=1}^{n} (g(t_i) - y_i)^2$$

$$+\lambda \int_0^1 \left(g^{(m)}(u)\right)^2 du, \quad (2)$$

where λ controls the smoothness of the solution, g is assumed to be $(m-1)$ times differentiable and $g^{(m)} \in \mathcal{L}_2[0, 1]$. It is shown that, for a fixed value of λ, the solution of equation (**??**) corresponds to a Bayesian estimator of g under a "partially diffuse" prior and the squared error loss. To choose the hyperparameter λ, generalized cross-validation is proposed.

• A. VAN DER LINDE (1993) **A Note on Smoothing Splines as Bayesian Estimates**, *Statistics and Decisions*, **11**, 61-67.

This note addresses the conceptual difficulties of the Bayesian interpretation of smoothing splines as Bayes estimators. It illustrates the difference between the prior model introduced in Wahba (1978) and the one given in Silverman (1985) [*JRSS-B*, **47**, 1–52]. Even though both models lead to the same Bayes estimator, the posterior variances differ by the interpolation error.

➤ **Wavelet basis**

The wavelet basis is the latest one used in nonparametric function estimation. Since its introduction into statistics in the late nineteen eighty, a number of papers have been written on this subject.

• H.A. CHIPMAN, E.D. KOLACZIK AND R.E. MCCULLOCH (1997) **Adaptative Bayesian Wavelet Shrinkage**, *Journal of the American Statistical Association*, **92**, 1413–1421.

The set up of this paper is similar to equation (**??**) except that $t_i = i/n$ and the error terms are normally distributed here. Also, g can be a potentially complex and spatially inhomogeneous function. The prior is a mixture of two normal densities, one being concentrated at 0 while the other more spread out. Empirical Bayes estimators of the hyperparameters are proposed. Simulations with the SureShrink and VisuShrink estimators are done for the standard test functions.

• D.L. DONOHO AND I.M. JOHNSTONE (1998) **Minimax Estimation via Wavelet Shrinkage**, *The Annals of Statistics*, **26**, 879-921.

In this paper, it is shown that wavelet thresholding can achieve the minimax rate. The spatial adaptivity of the wavelet estimator is discussed and the least favorable prior is given.

• F. ABRAMOVICH AND T. SAPATINAS (1999) **Bayesian Approach to Wavelet Decomposition and Shrinkage**, in P. Müller and B. Vidakovic (Eds.) *Bayesian Inference in Wavelet-Based Models*, Lecture Notes in Statistics **141**, 33–50.

It is assumed that the observations in equation (**??**) are sampled at $t_i = i/n$ and n is a power of 2. This assumption allows easy transformation from the observation space into the wavelet domain. The prior is a mixture of normal density and a point mass at zero. This model is used to build in thresholding of the wavelet coefficients. Hyperparameters of the prior are estimated using the log marginal density of the important coefficients. (Those coefficients are obtained after applying the universal threshold of Donoho and Johnson (1994) [Biometrika, **81** 425–455].)

• P. YAU AND R. KOHN (1999) **Wavelet Nonparametric Regression Using Basis Averaging**, in P. Müller and B. Vidakovic (Eds.)

Bayesian Inference in Wavelet-Based Models, Lecture Notes in Statistics **141**, 95–108.

It is assumed that t_i are equally spaced and that n is a power of 2. The prior is similar to that in Chipman *et al.* (1997) and four possible bases are considered: Haar, Daubechies 4, Symmlet 8 and the Fourier basis. In this empirical Bayes paper, model selection and model averaging are compared for several test functions.

• P. MÜLLER AND B. VIDAKOVIC (1999) **MCMC Methods in Wavelet Shrinkage: Non-Equally Spaced Regression, Density and Spectral Density Estimation**, in P. Müller and B. Vidakovic (Eds.) *Bayesian Inference in Wavelet-Based Models*, Lecture Notes in Statistics **141**, 187–202.

The prior used in this paper is similar to the one used in the previous articles. However, in this paper, a hierarchical approach is used instead of an empirical Bayes one. An MCMC simulation scheme is proposed. Unlike most of the contributions using wavelet basis, the procedure proposed here allows non-equally spaced values of t_i.

• M.A. CLYDE AND E.I. GEORGE (1999) **Empirical Bayes Estimation in Wavelet Nonparametric Regression**, in P. Müller and B. Vidakovic (Eds.) *Bayesian Inference in Wavelet-Based Models*, Lecture Notes in Statistics **141**, 309–322.

The main result of this paper is the use of E-M algorithm to estimate the hyperparameters among other empirical Bayes methods. Using simulations, the authors compared their empirical Bayes estimators with other well-known wavelet estimators (Hard thresholding, SureShrink and Risk Inflation Criteria) for standard test functions.

• J.-F. ANGERS AND M. DELAMPADY (2001) **Bayesian Nonparametric Regression using Wavelet**, *Sankhyä Series B*. To appear.

Contrary to other contributions in this area, estimation is done directly with the observations here rather than with the wavelet domain. This approach removes the usual restriction of having equally spaced t_i and n being a power of 2. Using a hierarchical approach, a Bayes estimator is proposed and compared with several other estimators (SureShrink, VisuShrink, and the one proposed in Chipman *et al.* (1997)) through simulation.

➤ Other work

Here I review some contributions that do not fall into any of three discussed classes.

• A. O'HAGAN (1978) **Curve Fitting and Optimal Design for Prediction**, *Journal of the Royal Statistical Society, Series B*, **40**, 1-42.

The model used in this paper is

$$\mathbf{E}(Y \mid x, \vec{\beta}) = \vec{f}(x)^t \vec{\beta}(x), \qquad (3)$$

where $\vec{f}(x)$ is a vector of known functions of x and the goal is to estimate $\vec{\beta}$, which may depend on x. Using a normal model, an estimator of $\vec{\beta}$ is proposed. However, some hyperparameters and the covariance kernel of the $\beta_i(x)$ need to be specified. Suggestions on how to elicit them are given in the Examples section.

• M. SMITH AND R. KOHN (1998) **Nonparametric Estimation of Irregular Function with Independent or Autocorrelated Error**, in *Practical Nonparametric and Semiparametric Bayesian Statistics*, D. Dey, P. Müller, D. Sinha (Eds.), Lecture Notes in Statistics **133**, 157 – 179.

In this paper, g is assumed to be an irregular function, that is, it is discontinuous or has discontinuous first or higher order derivatives. Instead of considering only one basis, the authors proposed four bases: reproducing kernel, cubic regression spline, linear natural spline and mixed radial basis. Using a hierarchical model, the proposed estimator does model averaging over the four considered bases.

• L.H. ZHAO (2000) **Bayesian aspects of some nonparametric Problems**, *The Annals of Statistics*, **28**, 532–552.

In this paper, asymptotic properties of Bayes estimators in nonparametric regression are investigated. It is shown that all Bayes estimators based on Gaussian priors have zero asymptotic efficiency (in the minimax sense). This lack of efficiency is due to the fact that Bayesian nonparametric regression is an infinite dimensional problem. However, using compound priors (infinite mixture of Gaussian priors), it is shown that the minimax rate can be achieved by a Bayes estimator.

BAYESIANS IN MEXICO

by Manuel Mendoza
mendoza@gauss.rhon.itam.mx

The history of Bayesian Statistics as well as that of Statistics as an independent scientific discipline in Mexico is rather brief. I shall refer to the beginnings just by paraphrasing Mendoza and Mendez (1991).

"The first PhD in Statistics obtained by a Mexican citizen was awarded to Basilio Rojas in 1959 at the Iowa State University. The first formal master program on Statistics was created at the Centro de Estadistica y Calculo (CEC) of the Colegio de Posgraduados de Chapingo in 1964 and one year later there were only three Mexicans with a doctoral degree in Statistics. In 1969, a master program in Statistics and Computing was created at the CEC.

During the following years some other master programs were created, most of them with a specific orientation to some field of application. Thus, in 1966 a program in Statistics was created at the Colegio de Mexico (an institution oriented to economical and sociological studies). In 1973, a Master of Science program in Mathematical Statistics was created at the Instituto de Investigaciones en Matematicas Aplicadas y en Sistemas (IIMAS) of the National University."

As for the Bayesian approach to Statistics, the subject appeared as part of the courses taught by Prof. Basilio Rojas at the CEC since 1964. However, it was only in 1973, at the IIMAS, that the first graduate course on Bayesian Statistics was included in a master program (as an elective course).

According to the available information, I believe that the first Bayesian doctoral dissertation by a Mexican was that of Enrique de Alba who in 1974 got his PhD in Statistics from the University of Wisconsin at Madison. In his thesis he proposed a procedure to deal with outliers using a Empirical Bayes approach.

The interest for the Bayesianism in Mexico has been stimulated by som important events. Some colleagues remember that in 1974, a seminar was organised by the Colegio de Posgraduados to which several prominent statisticians were invited. Among these, Prof. G.E.P. Box was very enthusiastic about the Bayesian methods and the audience was really impressed by his talk. Another definitive influence can be attributed to the long-term relationship that Prof. Jose Miguel Bernardo, from Valencia University, established with some Mexican statisticians.

The first time that Jose visited Mexico was in 1979. He offered a one-month intensive course on the foundations of Bayesian Statistics at the Facultad de Ciencias of the National University. Jose is well known to be one of the most radical Bayesians all over the world and one month of this Bernardian influence could not be ignored. As one of the many results of that visit, a second Bayesian doctoral thesis was produced. In this case, it was a thesis submitted by Gustavo Valencia to obtain a PhD in Mathematics from the National University. To this purpose, he spent one year in Valencia in 1983 working under the supervision of Jose on the problem of regression analysis with incomplete observations.

In 1984, Bernardo repeated the dose. Another one-month intensive course was organised, again at the Facultad de Ciencias, now on Bayesian methods and some specific applications. As a particular consequence of this second visit, Manuel Mendoza who was interested in the analysis of bioassays, asked Jose to be his PhD supervisor at the National University. Manuel spent two years (1985-1987) in Valencia where he completed an investigation on the inference for the ratio of linear combinations of the coefficients of a regression model. This thesis was presented in 1988 and Jose was a member of the examination committee. This was his third visit to the National University although on that occasion he was hosted by the IIMAS.

Since then, Bernardo has continued his visits to Mexico; he came back in 1992 invited by Manuel Mendoza on behalf of the Instituto Tecnologico Autonomo de Mexico (ITAM). He visited the IIMAS again in 1997 and, in addition, he has been invited as speaker to a number of events in Mexico. The most important are the II Congreso Iberoamericano de Estadistica in 1995, the XII Foro Nacional de Estadistica in 1997, the Taller Mexicano de Estadistica Bayesiana in 1998 and the III International Workshop on Objective Bayesian Methodology in 2000. The relation between Jose and us has been fruitful in many ways. Some of us maintain active research projects with him and right now another Mexican, Miguel Angel Juarez from ITAM, is currently

Bernardo's graduate student at Valencia.

Just to summarise, up to 1988 three Mexican statisticians had obtained a doctoral degree with a thesis on Bayesian Statistics. Fortunately, the situation has been evolving very rapidly. In the 90's, an important number of students decided to work for their Master degree or their PhD in Bayesian Statistics at some prominent universities. E. de Alba and M. Mendoza at ITAM encouraged many of these students and some others were benefited from the influence of R. Rueda at IIMAS.

The universities where those colleagues studied are Imperial College of Science, Technology and Medicine, Warwick University, Nottingham University, Sheffield University, Essex University and Oxford University all of them in the United Kingdom. In the U.S. the favourite university has been, without any doubt, Duke followed by Chicago.

At least ten Mexicans have recently obtained a PhD in Bayesian Statistics (or some other related fields). They are: Andres Christen (1994), Eduardo Gutierrez-Peña (1995), Raul Rueda (1995), Ruben Haro (1997), Gabriel Huerta (1998), Juan Jose Fernandez (1998), Omar Aguilar (1998), Rafael Perera (1999), Viridiana Lourdes de Leon (2000) and Luis Enrique Nieto (2001).

In addition, around ten students from ITAM have obtained a master degree in Statistics at Warwick University and nowadays, we have some students working for their PhD at Sheffield, Warwick, Chicago and Valencia universities.

The community of Mexican Bayesians has been affiliated, for the most part, with the universities but some of them work at the government and the corporate world. The larger research groups are now at the Statistics Department of ITAM and the Probability and Statistics Department at IIMAS and some other colleagues are very active at the Instituto de Matematicas (IMATE) of the National University and the Centro de Investigacion en Matematicas (CIMAT).

The range of topics that they work on includes: Model selection, reference analysis, dynamic linear models, general inference for exponential families, sample size determination, time series, hypotheses testing, nonparametric analysis, linear models, classification procedures, survival analysis, branching processes as well as applications to finance, image restoration, actuarial sciences, archaeology, bioassays and election forecasting.

Another two facts might be of some interest. Firstly, the number and the nature of the events that Mexican Bayesians have organised. In 1986 a NSF-NBER Seminar on Bayesian Inference in Econometrics took place at ITAM with the strong support of Prof. Arnold Zellner. Later, in 1995, the World Meeting of ISBA was held in Oaxaca immediately after the II Congreso Iberoamericano de Estadistica. Two workshops, under the name of Taller Mexicano de Estadistica, were organised in 1998 and 1999. The invited speakers for these workshops were Susie Bayarri, James Berger, Dani Gamerman, Jose Bernardo and Andrew Gelman. Although the annual national statistical meeting has never been a Bayesian conference, several times they have included Bayesian invited speakers as, for example, Jose Bernardo, James Berger, Daniel Peña and Javier Giron among others. More recently, in 2000 the III International Workshop on Objective Bayesian Methodology was held in Ixtapa.

The other relevant aspect is the fact that the rather small Mexican Bayesians group has published papers in some of the most important statistical journals. Just to mention a few, papers have appeared in the following journals: Royal Statistical Society, series B, C and D, Journal of Applied Statistics, Biometrics, Journal of the American Statistical Association, Journal of Business and Economic Statistics, Biometrical Journal, TEST, Journal of Statistical Planning and Inference, Journal of Statistics and the North American Actuarial Journal.

➤ Reference

MENDOZA, M. and MENDEZ, I. (1991). Graduate Statistical Training in Mexico. *Estadistica*, **43**, 101-113.

ADVICE TO CURRENT STUDENTS

by M. Eugenia Castellanos and Javier Morales
me.castella@umh.es
j.morales@umh.es

In this issue we present papers giving advices to current students. The authors are Susan Paddock and Alyson Wilson, former Ph.D students at ISDS, Duke University. Both jobs are very interesting so we hope you will like it, and this will be useful for you.

Susan Paddock
RAND Corporation
paddock@smmail2.rand.org

I want to thank the Associate Editors of the Student's Corner for asking me to describe my experience as a statistician who is working in an "interesting and non-academic job" and to give advice to current students who might be considering such work following completion of graduate school. My first piece of advice is to congratulate yourself for making a terrific career choice – as statisticians we are lucky to have skills that are in high demand and that allow us to immerse ourselves in a variety of compelling subject matter areas; there are interesting professional opportunities both inside and outside of academia.

First, let me tell you a bit about RAND. RAND is a nonprofit institution whose mission is to improve policy and decisionmaking through research and analysis. Approximately half of RAND covers a wide range of topics such as health, drug policy, criminal justice, education, civil justice, labor and population, and science and technology, and the other half pertains to U.S. national security research issues. Much of RAND's work is based on the U.S., but RAND does have a few European offices.

Most of the funding for the work I do at RAND (health and drug policy) comes from grants and contracts from governmental agencies and private foundations. These grants and contracts are almost always awarded in a competitive manner, which means that my colleagues and I write and submit proposals which are reviewed by funding agencies, alongside applications from other non-academic agencies and/or universities, and the funding agency decides which proposals it will fund. Thus, participating in grant proposals is an important part of my job; this is similar to the situation in many biostatistics departments in the U.S., where professors are responsible for bringing in their own funding.

Now, I will tell you about the type of research I do at RAND. Since joining RAND two years ago after completing my Ph.D. at Duke University, I have worked on a variety of projects in the areas of health and drug policy. My largest project to date is one which is funded the U.S. governmental agency which is responsible for providing health care coverage to senior citizens (Medicare) and low-income persons (Medicaid). The project involves designing, implementing, and monitoring a payment system for inpatient rehabilitation care for Medicare recipients; the U.S. government spends 4 billion per year on such inpatient rehabilitation care, and the program is expected to grow in the future. The system will be implemented January 1, 2002. During the first phase of the project, my role was to determine a baseline payment for all patients which would ensure that the government would not pay more than it had budgeted for the payment system. Currently, we are beginning the next phase of the project: we will further refine and improve the elements of the payment system, and we will monitor changes in the way inpatient rehabilitation care is provided once the system is implemented. It will be important to identify ways in which the way that care is given is impacted by the new system, especially if those impacts are unforeseen. In particular, we are interested in whether hospitals will change the way they report to the government the information about the types of patients they treat, and whether this is indicative of hospitals "gaming" the system (i.e., hospitals unfairly taking advantage of the system in order to maximize profit) or other factors.

I also have worked on a variety of smaller projects during the past two years. As part of an interdisciplinary team of psychologists and medical doctors, I examined the connection between risky sexual behavior (i.e., unsafe sex) and relationship violence in national (U.S.) probability sample of HIV-positive persons who received medical care in the late 1990s. A new phase of the study will begin soon, in which we will examine mortality and health outcomes as well. I have also participated in several, smaller projects in the area of drug policy, two of which involved building mathematical models to assess the impact of school-based drug and alcohol prevention programs and to demonstrate that the argument which is often used to support the "gateway effect" hypothesis (marijuana use leads to the use of "harder" drugs such as heroin) is questionable.

In addition to project work, I spend about 15-20% of my time on purely statistical methodological research; currently, that work is funded by RAND, but as I become more experienced I will be expected to bring in grant money to support that research. I have spent most of my methodological research time finishing up research and writing papers related to my dissertation, though I have also been exploring diagnostic methods for hierarchical models and developing a method for a particular problem that grew out of my HIV project work which involves properly accounting for variance due to design features when analyzing clustered data.

Much of what I do at RAND is similar to what an academic might do; I am expected to publish papers and am encouraged to apply for grant money. In fact, two of my colleagues and I are waiting to hear whether we will receive funding for a statistical methodology grant we submitted earlier this year. However, unlike academics, I can also succeed at RAND by publishing in non-statistical journals, briefing clients, or successfully obtaining and managing large research projects. The environment at RAND is unique in this way; it is possible for statisticians to be very successful at RAND but to achieve success in very different ways. The most obvious difference between RAND and academics is that we are not required to teach; because RAND has a graduate school it is possible to teach, but this is optional since it is a duty which does not replace regular project work. Also, many of my colleagues have also taught at local universities because they enjoy teaching and want that experience.

The climate at RAND is similar to that of an academic department in many ways; the Statistics Group (`www.rand.org/methodology/stat`) is very collegial, and we have meetings several times a month for various purposes: we have a Reading Group, in which we read and discuss a statistics paper; we have meetings during which we discuss methodological research or statistical problems coming directly from project work, and we have a monthly seminar series in which we invite non-RAND statisticians to speak.

As a new Ph.D. statistician in a non-academic environment, the biggest challenges for me have been to achieve a balance between doing statistical methodological research and applied research and consulting (as I said before, we statisticians are in high demand!) and becoming a "subject matter" expert in areas of health and drug policy (it is critical for an applied statistician to be "more than just a statistician" and to participate fully in the research). An additional challenge as a Bayesian at RAND is that Bayes is not (yet) the paradigm of choice, but I believe that this will change; I was hired in part because I am a Bayesian and RAND recognizes the value of Bayesian methods (empirical Bayes methods are widely used and understood, I should add). Also, since my arrival at RAND just two years ago, I have been joined by a couple of new Bayesian colleagues, so our numbers (and therefore the possibilities for Bayesian methods) are increasing. I would like to bring Bayesian methods to my project research at RAND, and I have an opportunity to do so in the near future on one of my projects. So, a piece of advice for a new Bayesian Ph.D. in a non-academic position is to be flexible and be willing to learn a lot from your non-Bayesian colleagues, as this will be critical if you ever hope to teach them about Bayes!

More advice for graduate students who might be considering taking a non-academic position following graduation: Take as many courses as possible during graduate school in order to develop a solid base of knowledge. If possible, having a summer internship is a very good experience as well for getting a flavor of non-academic research; otherwise, doing some form of applied research, either as part of the thesis or as a research assistant while in graduate school, is helpful, too. Also, there are a wide variety of non-academic statistical positions, not all of which are research-oriented; it is up to you to decide what best suits your interests. In addition, you may have a lot of statistical colleagues in non-academic positions, or you might be isolated; you should decide which best suits your interests (though I will emphasize that there are a lot of advantages to having good statistical colleagues). Ask a lot of questions of faculty members and other statisticians about career options – find out the pros and cons. If you are considering a non-academic position, you must be flexible about your statistical interests; just because your thesis is in nonparametric Bayes (as was mine....) does not mean that you will be doing that kind of work in your job (...and it can be very stimulating to do something new). Nevertheless, there is something to be said for specializing early on; while it is not necessary, you can get a big jump in your career as an applied statistician if you bring an in-demand skill to your job out of your thesis work (e.g., missing data imputation to the social science setting; spatial statistics to an environmental organization, etc.). However, the most important skill you should learn as a Ph.D. student is how to think about uncertainty and how to learn new things, so even if taking a non-academic position means you will have to switch gears, and if this is something you truly want to do following graduation, then go for it!

Finally, there is a very nice article in the *American Statistician* (Vol. 55, Nr. 1, February 2001), "How to Get a Job in Academics", by Elizabeth Stasny; while the article focuses on academic job searches, the advice can be applied to non-academic searches as well.

Alyson Wilson
Statistician and Technical Lead
for DOD Programs D-1
Statistical Sciences
Los Alamos National Laboratory
agw@lanl.gov

I completed my Ph.D. in statistics at Duke University in 1995. My dissertation focused on developing prior distributions for capturing features in medical images. After graduation, I moved to Las Cruces, New Mexico, where my husband was just starting graduate school at New Mexico State University.

The employment opportunities for statisticians are slim in southern New Mexico; most of the technical jobs are with the defense industry. I took a job with a small company in El Paso, Texas, called Cowboy Programming Resources (CPR). There were seven people who worked for CPR, and I was the only statistician. Our company planned, executed, and analyzed large tests of air defense systems like the Stinger and PATRIOT missiles for the U.S. Army. These tests often took two years to plan, cost several million dollars, and required collecting data about a few hundred soldiers out in the field using real systems for up to six weeks. During a test, we would collect and reduce tens of megabytes of data every day to produce summaries that would help guide the testing on the following day. I worked at CPR for four years. The skills I drew on were very applied: programming, experimental design, exploratory data analysis, sample size calculation, graphical summaries of large amounts of data.

In 1999, I took a job in the Statistical Sciences group at Los Alamos National Laboratory (LANL). I am in a group of 18 statisticians (14 with Ph.D.s) and 4 knowledge modelers (Ph.D.s in cultural anthropology, rhetoric of science, and computer science). Los Alamos itself is a fascinating place. It was founded in 1942, during World War II, to host the Manhattan Project, which was the U.S.'s program for the development of atomic weapons (see www.atomicmuseum.com). The lab itself is geographically isolated on a mesa in northern New Mexico. Today, the city of Los Alamos has about 18,000 residents, and about 9000 people work at LANL.

The mission of the lab has broadened considerably since its founding; currently LANL has a three part mission: (1) To ensure the safety and reliability of the U.S. nuclear deterrent, (2) To reduce the global threat of weapons of mass destruction, and (3) To solve national problems in energy, environment, infrastructure, and health security. Our group has research focuses in biological sciences, computational statistics, information integration, Monte Carlo methods, reliability, industrial statistics and quality control, statistical population bounding, and computer model evaluation. Our work is project-driven, but has a very strong research component. As is evident from LANL's mission, the lab focuses on "big science." Some of our projects include working on how to assess the performance and safety of the nuclear stockpile in the absence of full-system testing, developing ecological risk assessments for environmental restoration, and collaborating on the development of simulation-based models for the U.S. infrastructure (gas and electric power, communications, transportation). Our projects always involve multi-disciplinary teams of scientists working together.

Personally, I am still involved in many conventional weapons projects, which build on the expertise I gained during my years working with the Army. In particular, I work in the areas of information integration and reliability, with emphasis on Bayesian methods. For example, when a new weapons system is under development, there are many sources of information about its performance and reliability (similar systems, expert judgment, computer models, subcomponent tests, full system tests, etc.): how can all of this information be put together to make "continuous and comprehensive" estimates about the system?

The Statistical Sciences group at Los Alamos has a very active student internship program, both during the academic year and in the summers. See www.lanl.gov/orgs/d/d1/intern.shtml.

NEWS FROM THE WORLD

by Antonio Lijoi
lijoi@unipv.it

➤ Events

DEINDE (DEsign of INDustrial Experiments) 2002. *February 21-22, 2002, Torino, Italy.*
The sixth workshop is targeted at experts and users of Voice of the Customer and Customer Satisfaction techniques, although topics and applications will not be limited to them. Contributions are invited from business, industry, manufacturing, administration, marketing, sales, logistics and service showcasing. People interested in presenting a paper should send a brief abstract (approx. 250 words) by January 15, 2002, to the Chair of the Workshop, Grazia Vicario (vicario@calvino.polito.it)

2002 Spring Research Conference (SRC) on Statistics in Industry and Technology. *May 20-22, 2002, University of Michigan, Ann Arbor, MI, U.S.A.*
The SRC is an annual conference jointly sponsored by ASA/SPES and the IMS to promote research and applications on statistics in engineering, technology, industry, information and physical sciences. The 2002 SRC will feature keynote address by David Cox, plenary talks by distinguished researchers and leaders in industry, invited sessions on key areas, and contributed paper sessions. There will also be pre- and post-conference workshops. Deadline for abstracts: February 1, 2002. Web site: www.stat.lsa.umich.edu/src/

Ecological Inference Conference. *June 17-18 2002, Harvard University, Mass., U.S.A.*
The conference is sponsored by, and will take place at, the Center for Basic Research in the Social Sciences at Harvard University. Contributed presentations are invited for research related to the theme of ecological inference. Deadline for extended abstracts: February 15, 2002. Web: www.cbrss.harvard.edu/events/ecological.htm

2002 Conference of The International Environmetrics Society – TIES. *June 18-22, 2002, Genoa, Italy.*
The Conference aims to provide a forum of discussion to scientists working for the environment in different fields and types of applications. Among other topics, sessions will focus on: Chemometrics, Environmental human health statistics, Environmental risk assessment, Modeling environmental systems, National environmental statistics, Statistical modeling of spatial data in ecology, Training in environmental statistics. The conference will include plenary papers, invited paper sessions, contributed paper sessions, and poster sessions. Two special lectures will be given: the J. Stuart Hunter Lecture and the TIES President Invited Lecture. The conference will include a short course on the estimation of human impact on the environment. Deadline for abstracts: April 30, 2002. More information at www2.stat.unibo.it/ties2002

The 2002 Taipei International Statistical Symposium and Bernoulli Society EAPR Conference *July, 7-10, Taipei, Taiwan.*
Contributed papers may be submitted before March 31, 2002. Web page: www.stat.sinica.edu.tw/2002symp/

17th International Workshop on Statistical Modelling. *July, 8-12, 2002, Chania, Crete, Greece.*
The Workshop aims to bring together researchers and all those interested in the development and applications of generalized linear models and, moreover, statistical modeling in its widest sense. Papers motivated by real practical problems are desirable, but theoretical contributions addressing problems of practical importance or related to software developments are also welcome. The scientific programme is characterized by having invited lectures and tutorials, contributed papers, posters and software demonstrations. Deadline for abstracts: February 21, 2002. Web page: tara.unl.ac.uk/~11stasinopou/indexIWSM.html

Urbino2002 Biomathematics Euro Summer School. *July 8-19, 2002, Urbino, Italy.*
The school, organised by the The European Society for Mathematical and Theoretical Biology (ESMTB), will offer five courses centered on the biological background and on the mathematical modelling of relevant biomedical phenomena: the spread of cardiac electrical excitation, with the possible study of arhytmias; the physiology of blood flow in the pulmonary circulation; the system controlling glucose blood levels by means of the hormone insulin; the activity and synchronisation of neurones; the mechanism of production of blood cells with attendant regulations and possible derangements. The School is addressed to doctoral students in mathematics or allied disciplines (engineering, physics, statistics) looking for

exposure to medical problems, and to young biologists and physicians intending to employ mathematical tools in their research. Financial support is available for participants. Web: www.biomatematica.it/urbino2002

Fifth International Conference on Forensic Statistics. *August 30 - September 2, 2002, Venezia, Italy.*

The Conference is intended to bring together forensic scientists, lawyers, statisticians and those from related disciplines to discuss the many and varied uses of statistics in legislative, administrative and judicial proceedings. The program will include one stream of invited talks, a second stream of contributed talks and contributed papers to be presented in the poster sessions. Electronic submission of a text-only abstract of 300 words long must be sent to icfs5@eco.uniroma3.it before December 31, 2001. The deadline for poster submissions is July 1, 2002. Web page: icfs5.eco.uniroma3.it.

➤ Internet Resources

Bayes Linear Methods home page.

The new web page is maths.dur.ac.uk/stats/bayeslin/home.html. *(Thanks to Kurtay Ogunc who spotted the mistake in the September issue)*

Bayesian Network Repository.

The authors of the website aim at constructing a repository that will allow empirical research within their community by facilitating (1) better reproducibility of results, and (2) better comparisons among competing approach. Both of these are required to measure progress on problems that are commonly agreed upon, such as inference and learning. See www.cs.huji.ac.il/labs/compbio/Repository/

Bayesian Analysis, Computation and Communication.

The objective of the Bayesian Analysis, Computation, and Communication (BACC) project is to make Bayesian software and related resources available to users at all levels. The latest version of the software developed in connection with such a project is BACC 2001. It has involved Siddhartha Chib and John Geweke as main investigators and it currently offers Console (command line), Matlab, and Gauss versions for both Unix and Windows NT/98/95, as well as Windows versions for S-Plus and R. See www.econ.umn.edu/~bacc/bacc2001/

Bayesian Knowledge Discovery.

The Information Society uses information and generates data. These data are stored in large and fast-growing databases and the task of exploiting these data to enhance planning, prediction, and decision making represents a remarkable challenge. The Bayesian Knowledge Discovery Project aims at developing methods and tools, based on sound statistical theories, to take up this challenge. The project is a result of a joint effort of the Knowledge Media Institute and of the Department of Statistics at the Open University. Beside a series of publications on the subject, one can also download two softwares: (BKD) Bayesian Knowledge Discoverer and (RoC) Robust Bayesian Classifier. BKD is a computer program able to learn Bayesian Belief Networks from (possibly incomplete) databases. On the other hand, RoC allows performing supervised Bayesian classification from incomplete databases, with no assumption about the pattern of missing data. The website containing the material just described is: http://kmi.open.ac.uk/projects/bkd/.

Materials for the History of Statistics.

The Department of Mathematics at the University of York maintains an interesting web page with some materials about the history of Statistics. Portraits of Statisticians (including some well known Bayesians!) are available. The web page is www.york.ac.uk/depts/maths/histstat/

➤ Awards and Prizes.

Recipient of the 2001 Aranda-Ordaz Award

Gabriel Huerta, former Associate Editor of the ISBA newsletter, is the recipient of the 2001 Aranda-Ordaz Award for the best PhD thesis in any area of Probability and Statistics written by a Latin American in the last three years. The award is sponsored by the Bernoulli Society through its Latin America Regional Committee.

John M. Chambers Statistical Software Award

An entrant must have designed and implemented a piece of statistical software. The applicant must have begun the development while a student, and must either currently be a student, or have completed all requirements for the last degree earned after January 1, 1999. All application materials must be received by February 28, 2002. Information is available at the Statistical Computing Section (sponsor of the award) web site, reachable from the ASA web site www.amstat.org.

JOINING AND REJOINING ISBA

As laid down in Section G of the by-laws (viewable on the ISBA web site at `http://www.bayesian.org`), the ISBA membership subscription for 2002 are due on 1 January 2002. The fee can be paid through the web site, or by completing this form and returning it to:

Professor Peter Mueller, ISBA Treasurer
The University of Texas M. D. Anderson Cancer Center
1515 Holcombe Boulevard, Box 447
Department of Biostatistics
Houston, TX USA 77030-40091

voice: +1 713 745-4937
fax: +1 713 745-4940
e-mail: `pm@odin.mdacc.tmc.edu`
web: `www.bayesian.org (for membership)`
`odin.mdacc.tmc.edu/~pm/(personal homepage)`

I wish to become a member of ISBA

◯ Regular rate (U.S. $ 25.00)

◯ Reduced rate (U.S. $ 10.00) (*)

I wish to renew my ISBA membership

◯ Regular rate (U.S. $ 25.00)

◯ Reduced rate (U.S. $ 10.00) (*)

Name ____________________

Institution/Company ____________________

Department ____________________

Street Address ____________________

City, State/Province ____________________

Country, ZIP/Postal Code ____________________

Phone ____________________ Fax ____________________

E-mail ____________________

The above information may be made public: Yes ◯ No ◯

◯ I enclose a cheque in U.S. $ payable to International Society for Bayesian Analysis

◯ Credit card payment

American Express ◯ MasterCard ◯ VISA ◯

Card # ____________________ Exp. ____________________

Date ____________________ Signature ____________________

(*) The reduced rate of $10 for 2002 is payable by full-time students (up to 4 consecutive years, proof of status required) and permanent residents of countries whose GNP per capita, as given by World Bank data for 1996, is no greater than $6000. This includes those countries (Brazil, Chile, India and South Africa) currently having ISBA local chapters. Local chapter members must pay the ISBA subscription in addition to any local fees (see ISBA Provisional Policy for Chapters on web site).

Mailing address: *ISBA Bulletin - CNR IAMI - Via Ampère 56 - 20131 Milano (Italy)*
E-mail: *isba@iami.mi.cnr.it* **Phone:** *+39 0270643206* **Fax:** *+39 0270643212*
Web page: *www.iami.mi.cnr.it/isba*