THE ISBA BULLETIN

Vol. 8 No. 3 *September 2001*

The official bulletin of the International Society for Bayesian Analysis

A WORD FROM THE PRESIDENT

by Alicia Carriquiry
ISBA President
alicia@iastate.edu

Dear ISBA members:

I hope that this issue of the Bulletin finds everyone in good health and in good spirits. For our members in the United States, a special hope that the appalling tragedy of September 11 did not affect family or friends in any serious way.

First of all, I'd like to extend the warmest congratulations on behalf of ISBA to our new Savage Award winners: Peter Hoff, Tzee-Ming Huang, and Jeremy Oakley, and also to Tim Hanson who obtained an honorable mention. The committee ably chaired by Ehsan Soofi is already working to select next year's Savage Award winners.

And talking of awards, I am delighted to report that ISBA has agreed to administer the endowment of the Mitchell Prize, that was established in 1993 in memory of Toby J. Mitchell.

The final news about awards has to do with the DeGroot Prize, that will be awarded for the first time during the Valencia 7 meeting, and every second year after that. Steve Fienberg, from Carnegie Mellon University, has enthusiastically agreed to chair the first DeGroot Prize selection committee, and announcements will be forthcoming in the next few weeks.

The Executive Committee and the Board of Directors of ISBA have been actively discussing several issues of importance to ISBA and its membership. Underway are discussions, for example, on the always elusive problem of our stagnant membership numbers.

How should we try to increase the number of ISBA members? In the recent past, ISBA has made efforts on a variety of fronts to attract and then retain new members.

A very convenient and secure system to register and pay dues online has been in operation for some time now, thanks to the efforts of our Treasurer Val Johnson and of our Web Master Mike Evans. The ISBA Bulletin has been an essential channel of communication with our membership. Under the very capable hands of its Editor Fabrizio Ruggeri, the ISBA Bulletin has grown into a real asset for ISBA, and perhaps one of our most effective tools today to engage our members with ISBA.

The Program Committee, chaired by Tony O'Hagan, has been busy, and with exciting results; some of the meetings that ISBA will be sponsoring or co-sponsoring in the near future include the Latin American Conference on Statistics and Probability (Cuba, November 2001), the First Latin American Bayesian Meeting (Brazil, February 2002), and Valencia 7 (Spain, June 2002). Details can be found on our web site at `www.bayesian.org`. A very promising new venture for ISBA is an international meeting planned for 2003, that will be jointly sponsored by the IMS; stay tuned for more news in the coming months.

Contents

The valiant efforts described above have not yet resulted in increased membership, however, so it may be time to think about new services that the society can offer its members. Projects are in the works, and will be presented to the membership for feedback in the near future. On the nuts and bolts side, the Executive will propose to the Board of Directors that the Board authorize ISBA to establish a system for automatic deduction of annual membership fees, as is currently done by, for example, the Royal Statistical Society. A closer cooperation and increased communication with the Section on Bayesian Statistical Sciences of the ASA need to be established. Alan Gelfand, current Chair of SBSS (and ISBA member!), has been very helpful by raising awareness about ISBA among SBSS members.

We are always delighted to hear from you, and to get input and ideas on how to make our society more valuable to its members. Please feel free to write to any of the officers of ISBA (you will find their addresses in the ISBA web site), but in particular, please do not hesitate to contact me at alicia@iastate.edu.

ISBA ELECTIONS

by Philip Dawid
Chair, ISBA Nominating Committee
dawid@stats.ucl.ac.uk

The election for new ISBA Officers and Board members, to take up office on 1 January 2002, will shortly take place. The following positions are to be filled:

President-Elect (2002), to serve as President in 2003, and as Past President in 2004.

Treasurer (2002–4), to replace Valen Johnson.

Four Directors (2002–4), to replace those retiring from the Board.

The President-Elect and Treasurer are also *ex officio* members of both the Board and the Executive Committee. Details of the continuing and retiring members of the Executive and Board may be found on the ISBA website at www.bayesian.org.

This year's ISBA Nominating Committee consisted of *Philip Dawid*, University College London, UK, Past President ISBA (Chair); *Gabriel Huerta*, CIMAT, Mexico; *Katja Ickstadt*, University of Darmstadt, Germany; *Rosangela Loschi*, Universidad Federal de Minas Gerais, Brazil; *Kerrie Mengersen*, University of Newcastle, Australia; *Lawrence Pettit*, Goldsmiths' College, UK; and *Hal Stern*, Iowa State University, USA. In accordance with the Constitution and By-Laws, the Nominating Committee has nominated two candidates for each vacancy. The names of the candidates, and a short statement from each, appear elsewhere in this issue of the Bulletin. The By-Laws also provide for nominations by petition of 30 members of ISBA. This year no such petition was received by the deadline of 15 September 2001.

Ballots will be made available by 15 October 2001, and are to be returned by 15 November. The primary medium for transmission of ballots and other election information will be electronic. ISBA members who have not provided e-mail addresses will receive a paper ballot by regular mail. The ISBA Treasurer keeps the official record of postal and electronic mail addresses. Until December 2001 notification of changes to either address should be made by e-mail to isba@stat.duke.edu, or by regular mail to Professor Valen Johnson, ISBA Treasurer, Los Alamos National Laboratory, P.O. Box 1663, MS F600, Los Alamos, NM 87545, USA.

2002 MITCHELL PRIZE

The Mitchell Prize, sponsored jointly by the SBSS, the ISBA and the Founders Committee, is given annually in recognition of an outstanding paper that describes how a Bayesian analysis has solved an important applied problem. The Prize includes an award of $1000 and a commemorative plaque. It is given in memory of Toby Mitchell who spent the bulk of his professional career at the Oak Ridge National Laboratories on applied problems in biology and engineering, and who argued often and persuasively for a Bayesian approach to solving them.

The 2002 Committee (Rod Little, Tony O'Hagan and Henry Wynn, Chair), seeks nominations for the Prize that will be awarded at the ISBA Meeting in June of 2002. Complete details about the Prize and the nomination process are provided in the Mitchell Prize Charter (see www.bayesian.org under ISBA awards). Entries for the 2002 prize should be sent (before 1st February, 2002) to Professor Henry Wynn, Dept. of Statistics, University of Warwick, Coventry CV4 7AL, UK. H.P.Wynn@warwick.ac.uk

INFORMATION ON CANDIDATES

President Elect

Edward I. George

➤ *Affiliation and Current Status*: Professor of Statistics, Department of Statistics, The Wharton School of the University of Pennsylvania.

➤ *Web Page and e-mail Address*: `www-stat.wharton.upenn.edu/~edgeorge/`; `edgeorge@wharton.upenn.edu`

➤ *Areas of Interest*: Hierarchical Modeling, Model Uncertainty, Shrinkage Estimation, Treed Modeling, Variable Selection, Wavelet Regression.

➤ *Most Important Journals or Books*: Annals of Statistics; Biometrika; JASA; JRSSB; Machine Learning; *Aspects of Uncertainty* (A.F.M. Smith and P. Freeman, eds.); *Markov Chain Monte Carlo in Practice* (W.R. Gilks, S. Richardson and D.J. Spiegelhalter, eds.); *Bayesian Statistics and Econometrics* (D. Berry, K. Chaloner and J. Geweke, eds.).

➤ *Honors*: Elected Fellow, ASA, IMS; Elected Member, ISI.

➤ *Previous Service to ISBA*: Editor, *Bayesian Methods With Applications to Science, Policy, and Official Statistics. Selected Papers from ISBA 2000: The Sixth World Meeting of the International Society for Bayesian Analysis*, Eurostat; Program Council Chair 2000; Board Member 1996-1999; Program Committee for 5th World Meeting in Turkey; Program Chair for 3rd World Meeting in Mexico.

➤ *Services to other Societies*: Associate Editor - Biometrika, JASA; IMS - Committee on Fellows, Nomination Committee Chair; ASA- COS Representative for SBSS; Program Chair for B&E Section 1999; Nomination Committee for SBSS.

My view of ISBA

In existence for just 9 years, ISBA has already become a remarkable professional society. We have enjoyed six grand ISBA World Meetings, most recently, the tremendously successful Crete meeting attended by 297 participants from 30 different countries. Enveloped by a magical spirit of hard work and hard play, these meetings have been spurred on by a marvelous camaraderie that ensures a productive and fun time for all. Beyond these World Meetings, ISBA continues to sponsor numerous successful regional meetings, and is gearing up to play major role as a co-sponsor of the upcoming 7th Valencia Meeting. ISBA now co-sponsors four prestigious research awards, the DeGroot Prize, the Lindley Prize, Mitchell Prize, and the Savage Award. Both the ISBA Bulletin and the ISBA Website continue to reach new levels of excellence.

So where do we go from here? Well, I think we all want to see ISBA grow and become more influential. With roughly 325 members, ISBA is still small compared to other statistical societies. As Bayesian analysis continues to flourish and become widespread, ISBA is too often not involved. By increasing our visibility through joint efforts with other societies of statisticians and scientists, we can increase our membership and our impact. Preliminary plans for a joint ISBA-IMS meeting in 2003 are an excellent step in this direction. We may also want to consider ISBA sessions at the meetings of other societies with reciprocal sessions for them at our meetings. To further encourage the dissemination of our accomplishments, I believe it is time for a concerted effort to establish relations between ISBA and the popular media. ISBA publications, such as the upcoming ISBA 2000 volume, are yet another route to increased visibility. Although I am open to the idea of an ISBA journal, I can also see its drawbacks and look forward to an open and friendly debate of its merits across our membership.

While we contemplate such initiatives, I think it is crucial to make sure we maintain and build on what we already have - a tightly knit organization that serves as an extended family haven for the international community of Bayesians. In addition to our strong intellectual discipline, we must be careful to maintain the strong sense of good will, intimacy and trust that have become the heart of the ISBA environment. It is precisely these qualities that will attract other statisticians and scientists to ISBA, and will inevitably lead us to our ultimate goals.

President Elect

Joseph B. ("Jay") Kadane

➤ *Affiliation*: Carnegie Mellon University.

➤ *Current Status*: Leonard J. Savage University Professor of Statistics and Social Sciences.

➤ *Web Page and e-mail Address*: `www.stat.cmu.edu/~kadane/`; `kadane@stat.cmu.edu`

➤ *Areas of Interest*: Foundations of statistics; applications in social, behavioral, physical, biological, medical and environmental sciences; forensic statistics; statistical computing; econometrics.

➤ *Most Important Journals and Books*: Journal of the American Statistical Association; Biometrika; Annals of Statistics; Journal of the Royal Statistical Society; *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*, with D. A. Schum; *Statistics and the Law*, with M. H. DeGroot and S. E. Fienberg (Eds.); *Bayesian Methods and Ethics in a Clinical Trial Design* (Ed.); *Rethinking the Foundations of Statistics*, with M. Schervish and T. Seidenfeld (Eds.).

➤ *Honors*: Elected Fellow, ASA, IMS, AAAS, elected member ISI; Wilcoxon Award; Pittsburgh Statistician of the Year.

➤ *Previous Service to ISBA*: International Council of Advisors, 1994–1996; Chair, Constitution and Bylaws Committee, 1994–2000; Director, 1996–2000.

➤ *Services to other Societies*: *American Statistical Association*: Chair, Advisory Committee to the US Census Bureau, 1982; Board of Directors, 1983; Elected Chair, Section on Bayesian Statistical Science, 2002. *L. J. Savage Memorial Foundation*: Director, 1978–2000. *Journal of the American Statistical Association*: Associate Editor, 1968–1973; Acting Theory and Methods Editor, 1986–1987; Applications and Coordinating Editor, 1983–1985. Associate Editor, *Annals of Statistics*, 1974–1976, *Journal of Business and Economic Statistics*, 1987–1998.

My view of ISBA

These are very exciting times for Bayesians. I am delighted with the enormous successes we now enjoy. The issues faced by ISBA, and by Bayesians generally, are the consequences of our expanded influence, and call for rethinking our strategy in the light of new circumstances.
The combination of sound philosophical underpinning, hierarchical modeling and our recently enhanced ability to compute the quantities of interest has made the Bayesian perspective attractive to many. This strength has led me to believe that the place of Bayesianism within statistics is now secure, and that our attention should increasingly go to the various applied areas to which Bayesian ideas can contribute.
ISBA should first of all consolidate the advances it has already made: holding very successful meetings, providing an umbrella for various prizes (Savage, Mitchell, Lindley, DeGroot), and publishing its excellent Bulletin. Additionally, ISBA should continue to be responsive to the needs of geographical regions that have chosen to organize ISBA chapters. But all of this is continuation of what ISBA has been doing already.
How should ISBA take advantage of the new opportunities now open to us? One possibility is organizing smaller more specialized meetings grouped around a particular area of application, a specific computing technique, or comparative foundations. This would allow us to have more, and smaller meetings of Bayesians who share a specific interest. It would also allow us to be more active in reaching out to substantive experts with Bayesian interest.
A second opportunity we should pursue is to the creation of an on-line journal of Bayesian thought. A few years ago I opposed the creation of such a journal because I wanted Bayesians to continue to publish in the major statistical journals. I think we are now past the point of any danger of Bayesians becoming a ghettoized specialty in statistics. Thus, we now have more to gain than to lose by creating the world's first Bayesian journal.
It would be an honor to serve as President of ISBA for a year, if that is the wish of the membership.

Treasurer

Steve MacEachern

Steven MacEachern is a Professor in the Department of Statistics at The Ohio State University (www.stat.ohio-state.edu/~snm). Steve's main areas of interest are nonparametric Bayesian methods, computational methods, Bayesian data analysis, and dynamical systems and chaos. His publications have appeared in, among other places, Biometrika, the Journal of Computational and Graphical Statistics, the American Mathematical Monthly, the Scandinavian Journal of Statistics and the Canadian Journal of Statistics. He has served as Chapter Representative and President of the Columbus Chapter of the American Statistical Association, and is currently Associate Editor of the journals Technometrics and The American Statistician.

Peter Müller

Since I got my PhD from Purdue in 1991, I have been working at Duke University and, starting this fall, at U Texas M.D. Anderson Cancer Center where I am currently Professor at the Department of Biostatistics. My main areas of interest are Bayesian nonparametrics, MCMC simulation, optimal design, and longitudinal data models. I have published papers in Applied Statistics, Biometrics, Biometrika, JASA, JCGS, and JRSSB. Further information about me (including postscript versions of unpublished papers) can be obtained from my Web page address odin.mdacc.tmc.edu/~pm/. I was a member of the ISBA Board of Directors, 1995–8.

Board Members

Caitlin Buck

Caitlin Buck (PhD 1994, University of Nottingham) is a Lecturer in the Department of Probability and Statistics at the University of Sheffield (www.sheffield.ac.uk/st1ceb/). Her research focuses on the application of the Bayesian paradigm to archaeological data interpretation. She is best known for her work on the interpretation of chronological information (in particular radiocarbon data) and is the co-ordinator of an on-line, MCMC-based chronological data interpretation service known as BCal (http://bcal.sheffield.ac.uk/). In the last decade or so, she has also worked on the interpretation of archaeological surface survey data and architectural problems using change point analysis and on the interpretation of chemical compositional data from ancient ceramics using Bayesian clustering methods. She has published in Valencia 5, the Journal of Archaeological Science, Archaeometry, Applied Statistics, the Gilks et al 1996 book on MCMC in Practice, and (in 1996) she co-authored: The Bayesian Approach to Interpreting Archaeological Data, Wiley, Chichester. In 2000, she served on the nominations committee for ISBA.

Pilar Iglesias

Pilar Iglesias (Ph.D. 1993, Universidad de Sao Paulo) is Associate Professor and Head of the Department of Statistics, Pontificia Universidad Catolica de Chile. Her research interests include Bayesian inference in elliptical regression, model selection and diagnostics, representation theorems, and teaching of Bayesian Statistics. She has published articles in several books and journals, including Bayesian Statistics 6, Computational Statistics, Journal of Statistical Planning and Inference, Journal of the Chilean Statistical Society and Test. She has also written several Bayesian monographs for courses of the Chilean Statistical Society, of which she is currently President. She was a member of the ISBA nominating committee in 1998 and 2000, and has been President of the Chilean Chapter of ISBA since 1997.

Jun Liu

After getting my Ph.D. in 1991 from the University of Chicago, I worked as Assistant Professor at Harvard and Stanford Universities from 1991 to 2000. I am currently Professor of Statistics and Biostatistics at Harvard University. My research interests include bioinformatics, Bayesian modeling, genetics, Monte Carlo methods, signal processing, and time series. I have published papers in Annals of Statistics, Biometrika, JASA, JRSSB, Science, etc., and a book "Monte Carlo Strategies in Scientific Computing" with Springer-Verlag. I actively participated in the past ISBA conferences and was the 2000 Mitchell prize recipient. Further information about me and my research can be found in my webpage www.fas.harvard.edu/~junliu.

Helio Migon

I have been working for a long time at the Federal University of Rio de Janeiro, where I have a joint appointment between the Department of Statistics at the Mathematical Institute and the Operation Research Section at the Engineering Graduate School -COOPE. I am currently Graduate Director of the Mathematical Institute. I was President of the Associacao Brasileira de Estatistica for 1998-2000. My main areas of research interest include hierarchical models, applications of Bayesian statistics and decision theory in actuarial models, dynamic models and, more recently, stochastic production frontier models. I have published in a variety of journals including JASA, JRSS, Computational Statistics and Data Analysis, Computational Statistics. I have also written a book "Statistical Inference : An Integrated Approach" (Arnold, 1999), joint with Dani Gamerman.

Sonia Petrone

Sonia Petrone (Ph.D. 1989, Università di Trento) is Associate Professor of Statistics, Università Bocconi, Milano, Italy. She has been working on foundations of Statistics, along the approach of de Finetti, with applications in decision theory. Her recent areas of interest are Bayesian nonparametric inference, mixture models, predictive inference, dynamic models. She has published papers in several journals, including the Scandinavian Journal of Statistics, the Canadian Journal of Statistics, Statistics and Probability Letters, the Journal of the Italian Statistical Society, Metron, the Journal of the Royal Statistical Society, B (in press). `eco.uninsubria.it/Webdocenti/spetrone/homepage2_va.html`

Dale Poirier

I am a Professor of Economics at the University of California, Irvine, currently engaged in helping to create a new Department of Statistics on campus. My main area of interest is econometrics but all areas of theory and application of Bayesian statistics interest me. I publish in both econometrics and statistics journals (e.g., Journal of Econometrics, Journal of the American Statistical Association, and Journal of Business and Economic Statistics). My textbook, Intermediate Statistics and Econometrics: A Comparative Approach (MIT Press, 1995), is one of few econometrics texts covering the Bayesian viewpoint. I have been involved with ISBA since its inception. I was appointed to the initial Temporary International Advisory Board and I was elected to the first International Advisory Board. I served on the Local Arrangements committee for our very first meeting in San Francisco in 1993 and I organized our most recent meeting in Laguna Beach in April, 2001. `aris.ss.uci.edu/econ/personnel/poirier/poirier.html`.

Fabrizio Ruggeri

Fabrizio Ruggeri (Ph.D. Duke, M.Sc. Carnegie Mellon) is Senior Researcher at the Italian National Research Council in Milano and teaches at Politecnico di Milano and Università di Pavia. His main areas of interest are Bayesian robustness, wavelets, industrial applications (reliability and bidding) and inference in stochastic processes (queues, Poisson and self-similar ones). He is one of the editors of the most recent books on Bayesian robustness (IMS Lecture Notes, 1996, and Springer, 2000), and has published in a range of journals, including Journal of Statistical Planning and Inference, Journal of Computational and Graphical Statistics, Reliability Engineering and Systems Safety and Sankhya. ISBA member since its foundation, he is the ISBA Bulletin Editor for the term 1999-2001, has served in the 1998 and 1999 ISBA Nominations Committees and has been one of the organisers of three ISBA sponsored conferences (on Bayesian Robustness in Rimini, 1995, and Bayesian Inference in Stochastic Processes in Madrid, 1998, and Varenna, 2001). `www.iami.mi.cnr.it/~fabrizio`.

Robert Wolpert

Robert Wolpert (PhD Princeton Univ, mathematics) is Professor of Statistics in Duke University's Institute of Statistics and Decision Sciences (ISDS) and Nicholas School of the Environment and Earth Sciences (NSEES). Some of his current research interests include a random field approach to spatial statistical analysis and Bayesian nonparametric statistics, and likelihood-based statistical analysis in spatial environmental and epidemiological applications. He has published a book (The Likelihood Principle, joint with Jim Berger) and articles in a range of journals (e.g. JASA, Biometrika, Statistical Science) and volumes (e.g. CMU Case Studies in Bayesian Statistics, Valencia Bayesian Statistics). He has served ISBA as Program Chair (1999-00), as member of the Publications Committee (1998-00), and as editor of the Bayesian Abstracts Archive. `www.isds.duke.edu/~rlw`.

HOW DIFFERENT MIGHT THINGS HAD BEEN HAD BAYESIAN STATISTICAL SCIENCE BEEN AVAILABLE TO THE MOST FAMOUS SCIENTISTS IN HISTORY

by S. James Press
and Judith M. Tanur [1]

It might be of interest to conjecture a bit about what 12 of the famous scientists in history, from Aristotle to Einstein, might have done had Bayesian analysis been available to them. (A detailed rationale for how these 12 scientists were selected is provided in Press and Tanur, 2001.) Suppose that advanced mathematics, subjective probability, and Bayesian statistical methods of analyzing scientific data had already been developed and were readily available at the time some of these scientists lived and worked. How might that research methodology have been employed to strengthen the positions that those scientists held, thus perhaps settling more definitively some of the questions still raised, even today, perhaps hundreds of years later? In many cases we are still questioning their research methodologies, their judgments, their conclusions, and their decision making that was based upon the conclusions they reached. In the following summaries of the subjectivity of each of these scientists, we shall speculate in some cases as to how they might have used Bayesian analysis. In particular, we shall speculate on how they might have used such analysis to compare competing models of the phenomena under investigation.

Aristotle (384-322 B.C.) was a dedicated and careful observer of the natural world. Nevertheless, his belief in final causes and other overarching theories sometime led him to over-generalize or otherwise distort his observations. For example, he wrote that human beings are superior to other animals in purity of blood and softness of flesh. Because he believed that males are better equipped than females with offensive and defensive weapons, he concluded that worker bees are male (when in fact they are female). He over-generalized his observation that extremely light objects fall more slowly than do heavy ones to conclude that the speed of falling is generally related to the weight of an object. (We now understand that the difference in speed between extremely light and heavy objects only occurs because the effect of air resistance is more obvious for very light objects.) He described detailed observations that purport to show, mistakenly, that the heart is the first part of the embryo to develop. These mistaken observations lent support to his doctrine that the heart is the principle of life, the seat of the soul, of locomotion, and of what we now call higher mental functions. Since Aristotle was more inclined to speculate about scientific facts than to collect data to support his conclusions, we shall not conjecture in his case about how he might have handled probability.

Galileo (1564-1642) seems to have been convinced a priori of the law of motion that the distance fallen by a dropped object is proportional to the square of the time elapsed from dropping well before he conducted experiments to demonstrate the law. Indeed, he seems to have derived the law from a set of false premises. Further, there is some doubt as to whether the instrumentation of his day would have permitted Galileo to carry out the experiments he claims to have done, or at least to have achieved results of sufficient accuracy to prove his thesis, had he not already been convinced of its truth. Despite his position that seemed to challenge the Church's insistence on a geocentric universe, Galileo's own belief in cosmic order prevented him from accepting Kepler's substitution of elliptical planetary orbits for the more cosmically perfect circular ones. This rejection barred Galileo from full formulation of the inertial law, eventually formulated by Newton.

Galileo was forced to entertain two opposing theories of how Earth travels through the universe: does Earth travel around the Sun, or does the Sun travel around the Earth?

Copernicus and Kepler had already concluded that the heliocentric theory was to be preferred to Ptolomy's geocentric theory. Galileo's professional career was dominated by his life-threatening confrontation with the Catholic Church over which of these two theories should be accepted as truth. Putting political and religious considerations aside, Galileo might have calculated the (posterior) odds favoring each theory from a Bayesian viewpoint by comparing the (subjective) probabilities of each

[1] Adapted from Section 4.14 of, "The Subjectivity of Scientists and the Bayesian Approach", by S. James Press and Judith M. Tanur (2001), New York: John Wiley and Sons, Inc.

of the two theories to see how they compared. The powerful political position of the Church at that time, and its strong and long-held fixed belief in geocentrism, suggest that even if such a calculation had yielded results that showed the probability of the heliocentric theory overwhelming the probability of the geocentric theory, it still would have been difficult to convince the Church authorities. William Harvey (1578-1657) used what he called a "meditation" on the amount of blood issuing from the heart over a period of time to deduce that such a large quantity could not be manufactured or stored in the body; hence he was led to the notion of circulation. Having convinced himself of the reality of circulation and thus the necessity for blood to flow from arteries to veins, he was willing to believe (correctly) in the existence of a connecting network of capillaries without ever having physically observed such a network. In his work on the reproductive mechanisms of animals, Harvey was unable to identify a human ovum, but was sufficiently confident of the generality of his findings in lower animals that he subjectively (and correctly) generalized them to the entire animal kingdom.

Harvey confronted Galen's earlier theory about the heart, aorta, and all the arteries, that the blood pulsates through the body in an ebb-and-flow pattern, and certainly doesn't travel in a closed, one-directional path, by examining the implications of such a theory. He actually made calculations about what such a pulsation theory would imply about the buildup of blood in various places in the body. And he collected data that would bear on both the pulsation and circulation hypotheses. Had Bayesian methods of analysis been available he might have formed the posterior odds ratio for the probability of the "pulsation model" for blood compared with the probability of the "circulation" model he was considering. A ratio much less than 1 would have argued very strongly for the advocacy of the circulation model, which might have been very convincing to Harvey's colleagues. Harvey might also have applied such a probabilistic approach to comparing his belief in the old Aristotelian theory of "epigenesis" (the formation of a fetus by the addition of one part after another) with the more commonly-held thinking of his time, the theory of "pre-formation" (all parts of the animal fetus were present, if invisible, from the start). His data, obtained from numerous dissections, strongly supported his belief.

Subjectivity seems to have entered the work of Sir Isaac Newton (1642-1727) at a later point in the scientific process. Between the first and second editions of the *Principia*, Newton has been found to have made several changes to make the data he reported seem to fit more closely to the theories he was advocating. He reported a calculated value of the speed of sound that agreed too exactly with a value reported by another investigator, despite the fact that the earlier value was the average of many distinct measurements. Newton claimed that water vapor constitutes ten percent of air (in fact, it does not, but varies with temperature, pressure, and geographic location). He further claimed, without empirical support, that sound is not propagated through water vapor. To compensate, Newton arbitrarily increased the calculated speed of sound by 10%. Numbers purported to be data on the precession of the equinoxes were altered to better fit some corrected mathematics. Because Newton was sure of the truth of his theories, he saw his task as that of convincing his peers, and hence sometimes altered his experimental results to make them seem to support the theory more strongly. But excluding his work in alchemy and other metaphysical researches, he got his physical laws right (although they were somewhat modified later by Einstein). A major issue of model comparison concerned Newton during most of his career, the issue of the physical nature of light. Should light be thought of as corpuscular (made up of tiny particles), or as a wave motion? Newton was mostly persuaded during his early years that light should be thought of as a wave motion (and he studied the interference patterns now called "Newton's rings" which involve wave-motion thinking). Later in his career he switched positions and then thought about light as made up of particles whose behavior could be predicted by his theory of the laws of motion of material bodies. In this matter he was to be pitted against Christian Huygens, another famous scientist of the time, who strongly believed in the wave theory of light. The issue remains an arguable matter today. In what is called the dual nature of light (a theory partially attributable to Max Planck), light sometimes seems to act one way, and sometimes the other. Would a Bayesian model comparison have helped?

We believe it might have, and might still, but such a probability comparison does not yet seem to have been carried out.

For Antoine Lavoisier (1743-1794), there is a sense in which his entire corpus of work, consisting as it did primarily of a theoretical synthesis of the experimental work of others, constitutes an exercise in subjectivity. He spent a good part of his career refuting the phlogiston theory (the mistaken theory that all flammable substances contain something called phlogiston that is given off when they burn). Indeed, he was the originator of the antiphlogiston theory of combustion. Yet he was not above invoking explanations drawn from phlogiston when these seemed convenient, and he advocated the imponderable "caloric" to help explain the varying states of matter. Not only may Lavoisier have plagiarized the experiments of others, claiming to have repeated those experiments but perhaps not having actually done so, but experiments that he reported having done were sometimes carried out after the publication in which they were cited. Further, the accuracy of experimental results was sometimes exaggerated, and the number of replications achieved similarly inflated. Like Newton, Lavoisier was subjectively convinced of the accuracy of the scientific system he was proposing, the systematization of chemistry, and sometimes gave in to the temptation to embellish the empirical support for that system. But the system and most of its details, especially the understanding of combustion and other forms of oxidation, were sufficiently correct to earn Lavoisier the sobriquet, "father of modern chemistry."

The argument between the antiphlogistians and the advocates of the phlogiston theory of combustion might have been resolved more readily had the posterior probability calculations been presented to researchers caught up in Lavoisier's chemical revolution. Again, the issue was the selection of the correct model to use in order to come closer to truth.

Alexander von Humboldt's (1769-1859) subjectivity influenced him, early in his career, to interpret geological observations as evidence supporting the Neptunist theory of the origin of rocks in sedimentation. Later, more extensive observations and a change in his theoretical orientation led to a reversal of the interpretation of the same observations as evidence in favor of the volcanic origin of the rocks (the Plutonist theory). Early in his scientific career, von Humboldt set out to demonstrate the harmoniousness of nature; accordingly, his voluminous writings repeatedly make the point that nature is one great whole, in which plants, animals, and geological and meteorological phenomena as well as human beings and their culture fit coherently. Bayesian statistical methods of analysis, had they been available at the time, could readily have compared the Neptunist and Plutonist theories.

Michael Faraday (1791-1867), like von Humboldt, sought a unity in nature. Under the influence of his religion of Sandemanianism, Faraday believed in the simplicity and integration of nature. This led him to attempt to establish the unity of the forces of magnetism, electricity, and gravitation. He achieved the conversion of electrical to mechanical energy and established the equivalence of all types of electricity. His conviction led to his persistence through many years of attempts before he was successful in showing the conversion of magnetism to electricity. He continued to believe in the relationship between electricity and gravity, although he was unable to establish that relationship. Faraday even left a letter to be opened 60 years after his death that carried the idea of unity of forces further with speculation on the existence of electromagnetic waves analogous to those of the ocean.

Charles Darwin (1809-1882) believed theory must guide observations despite his professed belief in Baconian induction. His synthesis of his observations led him to a theory of evolution powered by a mechanism of natural selection, which would yield gradual changes in organisms. This subjective belief led him to explain gaps in the fossil record where intermediate forms should have been represented as being merely temporary gaps in data gathering, to be filled as paleontologists explored further. It also led him to disbelieve then current calculations of the age of Earth, since the accumulation of gradual changes would require more time than those calculations provided. Scientific opposition to Darwin's proposed mechanism of natural selection was coupled with the mistaken belief (which Darwin shared) that the material of heredity was "infinitely divisible." If that were the case, the substance in the human body that passed heredity characteristics, such as eye color, on to the next generation could pass on any degree of "blueness," for example, as opposed to just "blue" or "not blue." An implication of this mistaken belief was that a change (mutation) in an organism would be diluted through interbreeding. That implication caused Darwin to change his mind about natural selection. His later writings gave but scant importance to natural selection and stressed instead a purely subjective notion of pangenesis, which implied the inheritance of acquired characteristics. Neither Darwin nor any of his contemporaries or successors could find any empirical evidence for the inheritance of acquired characteristics, called the Lamarckian view. The lack of empirical support for Lamarckian mechanisms was coupled with the re-discovery of the work of Gregor Mendel, which showed that the material of heredity is "particulate," that is, not infinitely divisible but discrete and thus that mutations can be stored in the genes without being diluted. These findings eventually caused the scientific community to revive natural selection as the mechanism for evolution, despite the fact that Darwin had abandoned the idea.

These two models of inheritance, "pangenesis," implying the inheritance of acquired characteristics (the Lamarckian view), and "natural selection," were opposed to one another. There were considerable data available that bore on the two models. They could easily have been compared probabilistically, as part of the scientific analysis. Such an analysis, and related analyses about the model of evolution versus. the model of biblical genesis have not yet been carried out, and the arguments surrounding such competing models still rage. As in the case of Galileo and the heliocentric and geocentric theories, the arguments regarding natural selection and evolution go far beyond science and mathematics by pitting religious belief against the opinions of the scientific community. While it is likely that probability would have helped somewhat in both Galileo's time, and in Darwin's, the issue of convincing the public is one that pits science against religion.

Louis Pasteur (1822-1895) has been shown within recent years to have permitted his strong prior beliefs influence his clinical treatments and his reports of data he falsely claimed to have acquired through experimentation. On the basis of theory and its demonstrated effectiveness in the case of the microbe for chicken cholera, Pasteur believed that atmospheric oxygen would be effective in attenuating the virulence of other microbes. He was extremely enthusiastic about this theory of the usefulness of atmospheric oxygen. Hence in reporting a highly successful public test of the effectiveness of a vaccine against anthrax, Pasteur claimed that he had prepared the vaccine using atmospheric oxygen, when in fact it had been produced using potassium bichromate. Similarly, when he believed he had hit upon a process that would produce immunity to rabies, Pasteur felt himself justified in using that vaccine on two young boys who had been bitten by a rabid dog. He claimed that he had already successfully immunized 50 dogs with this same vaccine; his recently revealed laboratory notebooks make it clear that he had barely begun trials of this vaccine and had as yet no results at all. In his work to disprove the doctrine of the spontaneous generation of life, Pasteur is seen to have suppressed the results of experiments that had what he considered "errors," that is, data that might lend support to the theory he was trying to disprove. We now know that Pasteur's experimental procedure generated insufficient heat to kill all the microorganisms originally present in his experimental material, so that the "erroneous" results do not in fact support the idea of spontaneous generation, but Pasteur himself did not have this explanation available. Regardless, Pasteur's sterilization process was eventually justified, and resulted in the pasteurization process that has currency today.

One of the several model comparisons that Pasteur was involved with was the one that examined the theory of "spontaneous generation" and he confronted it with the theory that life can only derive from other life. He introduced the notion of sterilization (now called pasteurization) to prove his point. Would probabilistic analysis have helped his cause? Probably, but despite all of his brilliance and flashes of scientific insight, Pasteur's approach to data was to suppress them if they didn't agree with his pre-conceptions of what they ought to be. More objective Bayesian analyses by others of data analogous to Pasteur's would surely have helped Pasteur make his case more convincing.

Partly because Sigmund Freud (1856-1939) used subjectivity so pervasively in his work, some observers believe that his work is not science at all. He used data based on patients' free associations and recounting of dreams to interpret their subconscious thoughts, and thus derive what he considered universal laws indicating that the sexual urge and the libido are the driving forces behind all repression and inner conflict. Observers of Freud's work have stressed that his ideas evolved in a mind suffused with a sense of personal destiny and that maintained a lively interest in the occult. His attention to mysticism perhaps stemmed from his father's Chassidic mysticism and from his own heavy use of cocaine. Because they argue that Freud never scientifically tested his claims for the causal links he proposed between sexual repression and neurosis, and between the processes of psychoanalysis and cure, some critics feel Freud's work is pure subjectivity (rather than the required combination of subjectivity and objectivity found in all good science).

Freud's data are questionable in a scientific sense since he did not carry out controlled experiments and his observations were totally subjective assessments of what was happening to his patients. But that has been the nature of most of the field of psychoanalysis, although there are positive signs that the stance may be undergoing change in the appropriate direction.

Marie Curie (1967-1934) was long sustained in her investigation of radium by her belief that the new substance was an element. But her belief that elements could not be transmuted to other elements (even through radioactive decay) kept her from being able to explain the origin of radioactivity, although she (together with her husband, Pierre) originated the insight that radiation was an atomic property. The fact that the elements with which she was working (not only radium, but polonium as well) had long half lives made it extremely difficult for Curie to measure the loss of weight and energy due to radioactive decay. Thus, despite her toying in a 1900 publication with the idea of transmutation, Marie Curie's strong subjective prior belief in the established principles of physics, coupled with her strong belief in what her data told her, kept the next insight from her. A Bayesian model comparison of alternative theories purporting to explain the radioactivity phenomenon would probably have suggested new and exciting paths of discovery to Madam Curie.

Albert Einstein (1879-1955) was a *theoretical* physicist in the sense that he derived his beliefs about the physical world not only from a profound understanding of the way the physics of the world works, but also from the mathematical models that he developed of relationships that physical phenomena should obey. His equations governing the special and general theories of relativity, for example, were used to make corrections to Isaac Newton's laws of motion, and to generate predictions about what might be observed in appropriate hypothetical circumstances, as when traveling at speeds near the speed of light. But he never experimentally tested any of his theories; he merely stated them subjectively and mathematically, fully expecting them to correctly characterize what would be observed in the real world. He also believed that if data were collected to test his theories, and if the data contradicted theory, he would be inclined to reject the data as erroneous, because they contradicted the theory that he simply knew a priori to be true.

Einstein was fully capable of carrying out any mathematical analysis of his choosing. So to examine competing theories of physics using probability and posterior odds ratios would have been well within his grasp. He even used probability to develop the (predictive) law of Brownian motion, one of his famous papers of 1905. But he had difficulties with collecting his own data. He was not an experimentalist. So it was not his approach to decide about any physical laws on the basis of being convinced by a preponderance of scientific data–expressed probabilistically or otherwise. Einstein just intuitively and mathematically decided how things ought to be, and then enunciated what the general principles had to be. Probability considerations are very unlikely to have been much of a cogent force on his own beliefs, although they might have strengthened the scientific beliefs of others in his theories (such as in the theory of general relativity) at the earlier stages in his career.

We can discern several themes in this recital of the earlier uses of informal subjectivity in science, and the use of formal probability to examine competing scientific hypotheses.

First, we see generalizations from data (induction), precisely as Bacon had prescribed (and contrary to Popper's views involving falsification). But sometimes, these generalizations go well beyond the data at hand,

and can lead either to correct inferential leaps or mistaken over-generalizations. We have seen Harvey correctly assume that the origin of all life is in the ovum, even though he was unable to understand the significance of the ovaries in mammals. But we have also seen Aristotle over-generalize from watching extremely light and heavy objects fall. He concluded that heavy objects tend to fall toward the earth and light ones tend to rise and thus that the speed of fall is related to an object's weight.

Second, we have seen instances in which a scientist, overwhelmed by a theory, is apt to see what the theory prescribes (or is unlikely to see what the theory denies). Marie Curie, guided by a belief in the conservation of energy, failed to identify the mechanism that gives rise to radioactivity. And Albert Einstein insisted that any data that might seem to contradict his theory would prime facie be mistaken. Note how this statement echoes Galileo's expression of awe at learning that the adherents of the Copernican (heliocentric) system were able to ignore the evidence of their own senses (that the Sun appears to rotate around Earth) to conclude that Earth revolves around the Sun.

Third, we have seen instances in which a scientist becomes sufficiently convinced of the correctness of the line of investigation being followed that a single-minded stubbornness develops. We have seen Marie Curie insisting that her newly discovered substance, radium, is an element, and we see Michael Faraday laboring for a decade to demonstrate the induction of electricity from magnetism.

Finally there are instances of subjectivity that fall on the border of outright fraud - and instances that fall squarely within the domain of the fraudulent. To range outside of the 12 scientists we have been discussing for a moment, Johannes Kepler, Gregor Mendel, Robert Millikan and Cyril Burt reported data that were "too good to be true", to extend use the term employed by Fisher (see Fisher, R. A., 1936, "Has Mendel's Work Been Rediscovered?", Annals of Science, 1, 115-137) to describe his view of the work of Mendel, but that in each case fit the theory that the scientist was trying to prove or to demonstrate. The data doctoring in three of these cases was in the service of a theory now believed to be true; in Burt's case the jury is still out. Sir Isaac Newton altered data between editions of his masterpiece, *The Principia*, in order that his theoretical contentions appear in the best possible light. Both Antoine Lavoisier and Louis Pasteur claimed to have carried out experiments that were either incomplete or even not yet begun at the time they published the supposed results. We do not condone their bending, or even breaking, the accepted rules of honesty applicable no less in the practice of science than more broadly in normal human interaction. Nevertheless, we must note that Newton, Lavoisier, and Pasteur were most often correct in their scientific conclusions, and that they enormously advanced the cause of science with their thinking.

Subjectivity occurs, and should occur, in the work of scientists; it is not just a factor that plays a minor role that we need to ignore as a flaw that sometimes creeps into otherwise "objective" scientific analysis. Total objectivity in science is a myth. Good science inevitably involves a mixture of subjective and objective parts.

MORE ABOUT THOMAS BAYES

by Anthony O'Hagan
a.ohagan@sheffield.ac.uk

Since writing in the last Bulletin about my very amateur piece of historical investigation on Thomas Bayes and Richard Price, I have heard about two much more extensive and authoritative studies by colleagues in Canada and South Africa. I am sure that many ISBA members would like to know about these.

David Bellhouse has discovered some manuscripts that throw light on Bayes's status in contemporary mathematical circles, and has written a fascinating biography of Bayes. This and a technical article on the manuscripts themselves can be found on David's website at `www.stats.uwo.ca/faculty/bellhouse/bayespage.htm`.

Then you can look forward to a substantial book on Bayes written by Andrew Dale (`DALE@nu.ac.za`) and due to be published by Springer-Verlag next year. Andrew says, "It should run to about 600 pages, I think. There will be some chapters on the Bayes family (as much as I have been able to find) and on Thomas's time in Tunbridge Wells. Then chapters on the works, and finally something about Bunhill Fields, the Bayes family vault and wills of some of the Bayes family".

Bayesians in Agricultural Economics

by Garth Holloway, Jeff Dorfman and George Davis
g.holloway@cgiar.org, jdorfman@agecon.uga.edu, gdavis@tamu.edu

Growing interest in Bayesian methods among agricultural economists was evident in the recent, joint meetings of the American Agricultural Economics Association (AAEA) and the Canadian Agricultural Economics Society (CAES), Chicago, Illinois, August 5-8. There, three sessions were organized that focused on Bayesian methods, computational issues in Bayesian analysis and Bayesian comparisons with sampling theory approaches to model selection. The three sessions were an immense success, being some of the most well attended in the authors' memories. George Davis organized a session entitled *Econometric Methodologies for the Model Specification Problem: Addressing Old Problems in The New Century*. Its purpose was to compare and contrast the Bayesian approach to model selection with the frequentist approach (as represented by probabilistic reduction). Aris Spanos and Anya McGuirk contributed a paper on "A Probabilistic Reduction Approach to the Model Specification Problem" (discussed by George Davis) and John Geweke and Bill McCausland contributed one on "Bayesian Specification Analysis in Econometrics" (discussed by Wally Thurman). The papers and discussions will appear in the December 2001 Proceedings issue of the American Journal of Agricultural Economics. A "frontier" session on Monte Carlo methods was also organized by George Davis. Frontier sessions, a recent innovation at AAEA and CAES meetings, demonstrate cutting-edge techniques and their application to problems in agricultural and natural-resource economics. Bill McCausland, with John Geweke, demonstrated Bayesian estimation of the Seemingly Unrelated Regression model using their software Bayesian Analysis, Computation, and Communication (BACC). The session generated considerable interest, signaling to the many who attended that perceived entry costs associated with using Bayesian techniques are in significant decline. Finally, a session entitled *Markov Chain Monte Carlo Methods: A User's Guide For Agricultural Economists* was organized by Jeff Dorfman and Garth Holloway. The objectives were to make accessible to agricultural economists MCMC methods; demonstrate how routine application of MCMC solves problems of fundamental importance in agriculture and economic development; and, generally, advance the pace of adoption of Bayesian methods among agricultural economists. John Geweke presented an introduction to MCMC theory and recent developments in its applications, including projections about fruitful avenues for future research. Bill Griffiths used MCMC methods to impose both equality restrictions (homogeneity, adding up and symmetry) and inequality restrictions (monotonicity and concavity) on the parameters of the popular translog and almost-ideal demand systems and demonstrated how model choice problems are resolved easily using model averaging. Garth Holloway presented a non-zero censored Tobit regression applied to data on milk-market participation in the Ethiopian highlands and demonstrated how routine application of MCMC generates quantities of importance for development policy. Finally, although an unavoidable circumstance prevented Peter Rossi from delivering informative discussion comments (Jeff Dorfman moderated an enlivened post-presentation debate) participants (including recent ISBA inductees Wally Thurman and George Davis) heeded his culinary advice, adjourning for a delightful evening of Italian fare at a local restaurant and conversation on a wide range of topics, including Lindley's paradox and a list of potential venues for future ISBA meetings. In the great tradition of ISBA, we stayed until the restaurant was ready to close, with considerable wine drunk in the interim. Electronic versions of the papers in this last session ("Markov Chain Monte Carlo Methods: A Tutorial" by Bill McCausland and John Geweke, "Bayesian Model Averaging in Consumer Demand Systems with Inequality Constraints" by C.L. Chua, Bill Griffiths and Chris O'Donnell, "Tobit estimation With Unknown Point of Censoring With An application to Milk-Market Participation In The Ethiopian Highlands" by Garth Holloway, Jeff Dorfman and Simeon Ehui and "Discussion" by Peter Rossi) are available at `ftp://ftp.cgiar.org/ilri/LPAP`. The papers will appear in the November 2001 issue of the Canadian Journal of Agricultural Economics.

7TH VALENCIA INTERNATIONAL MEETING ON BAYESIAN STATISTICS

2002 ISBA INTERNATIONAL MEETING

Spain, June 2 - 6, 2002

Preparations for the Valencia 7 meetings are now well underway, and the deadline for submitting an abstract to the competition for Selected Contributed Papers is fast approaching. As a reminder, the Selection Committee chaired by Hal Stern, has established October 15, 2001, as the deadline for receiving abstracts to be considered for SCP. For additional information about the new SCP sessions, and to obtain instructions on how to submit an abstract for consideration, please check the ISBA web site at `www.bayesian.org` or the Valencia 7 web site at `www.uv.es/valencia7` or its mirror site at `www.stat.duke.edu/valencia7`.

For the first time, ISBA is using a fully electronic system for the submission of abstracts to the SCP competition. With the generous cooperation of Microsoft Corporation, the Selection Committee has now access to software that enables the members to receive, distribute, and keep track of abstracts, using only the web. Because the system has not been previously available to us, we hope to receive feedback from the ISBA members who had a chance to try it out themselves. So please write and let us know of your experiences with abstract submission!

Please plan to attend Valencia 7. The meetings will take place on a beautiful beach (Playa de las Americas) on the island of Tenerife, one of the Canary Islands. There are ample opportunities to present your work at the Valencia 7 meetings; while the invited portion of the program has already been set, you can still choose to submit an abstract to the SCP competition, or decide to present a poster in the always popular and well attended evening poster sessions. The deadline for submitting a paper for presentation in one of the poster sessions is April 10, 2002. For additional information on the poster presentations, please visit the ISBA web site or the two Valencia 7 web sites listed above.

ISBA members at the time of registration will be able to register to participate in the meeting at a reduced fee, so do not let your ISBA membership expire at the end of the year. You can renew your ISBA membership very conveniently on-line; the registration web page provided by ISBA is secure and thus payment can be submitted safely through the web.

We hope to see many of our ISBA friends in Tenerife, so mark your calendars, and start planning for a week of excellent talks, invigorating discussions, and of course sun, surf, and a great time.

CESAREO VILLEGAS

Dear Colleagues:

Cesareo Villegas, Professor Emeritus in the Department of Statistics and Actuarial Science at Simon Fraser University passed away on July 8/2001 at 80 years of age. He is survived in Canada by his wife Nellie, four children and four grandchildren, and in Uruguay by three brothers and two sisters.

Professor Villegas received the Ing Ind degree in Engineering from the U. de la Republica in Uruguay in 1953. After 20 years in faculty positions at U. de la Republica, he came to North America as a visiting Associate Professor at the University of Rochester (1968 to 1970). He joined Simon Fraser University in 1970 as an Associate Professor and was the founding statistician. From 1979 until his retirement in 1986, he served as Full Professor.

Cesareo Villegas was an expert in the foundations of Bayesian statistics, beginning his work in the days when Bayesian methods were not so fashionable. He was one of the original handful of pioneers who participated in the now wildly popular Valencia meetings that promote the Bayesian point of view.

His publications were theoretical and included amongst others, eight papers in the Annals and three in JASA. Some of his best known work involved the development of priors satisfying certain invariance properties.

Although his published work was characterized by mathematics, and in particular algebra and probability theory, Cesareo had an interest in applications. One topic which caught his fancy for a sustained period involved the possible relationship between river flows and sunspots. Professor Villegas was a scholar; he read widely, he thought long and deeply and he wrote quality papers. He was active in his retirement and maintained an NSERC grant up until the year of his death.

Cesareo was a gentle man who lived his life with dignity. Although quiet in nature, he could become animated when engaged in almost any topic, spiritual or scientific. He was generous with his time to young investigators and when it was clear that he was unable to spend his grant in 2001, he used the balance to support graduate students at SFU. He was a role model who demonstrated how to love and how to attend consistently to one's work without being overly distracted by the politics of academia. His priorities in life were firmly established, and in increasing order of importance, these included statistics, his family and his faith.

Cesareo had a slow growing prostate cancer for a number of years. The last three months he was hospitalized and was further diagnosed with a brain tumour. He lived his last months and days pain free.

He is deeply missed.

Respectfully submitted,

Tim Swartz and Michael Stephens
Department of Statistics and Actuarial Science, Simon Fraser University

---

Getting Epidemiology Students Started with Bayes

by Charles Poole
cpoole@unc.edu

I teach intermediate and advanced epidemiologic methods to epidemiology graduate students. In the statistical part of my teaching, I try to encourage future epidemiologists to become informed consumers of statistical methods. Informed consumers have goals they want to achieve and problems they want to solve. They aren't afraid to keep asking "stupid questions" until they get understandable, helpful answers. They make choices and don't just accept the first tool they're handed. They create demand, in the economic sense. In short, informed epidemiologic consumers of statistical methods can be a major pain to their friendly neighborhood biostatisticians!

I began introducing students to Bayesian methods out of necessity. I've found that after I explain what p-values and confidence intervals aren't, many students remain incredulous. Some of the remainder feel betrayed to greater or lesser degrees by what they've read in the medical and epidemiologic literature and by some of what they've been taught (or not taught) in "biostatistics for epidemiologists" courses. Almost all feel profoundly disappointed when they learn that p-values and confidence intervals refer to probabilities of data and not to probabilities of hypotheses.

I believe epidemiologists have every right and every good reason to ask the questions about probabilities of hypotheses, to which p-values and confidence intervals give such notoriously misleading answers. This belief obliges me to make my students aware, at a minimum, that Bayesian methods exist. This is a fairly bold step, given that many biostatistics and epidemiology textbooks don't even have index entries for Bayes's theorem, let alone Bayesian statistics. I try to teach "intro to Bayes" side-by-side with frequentist methods, and regularly prove the well known downsides of such an approach. But I don't seem to have a choice.

Aside from the incredulity, betrayal and disappointment many of my students feel upon learning that a helpful set of methods has been kept hidden from them as though it were a dirty little secret, the main challenges I've encountered thus far have been the usual fears of applied users: fears about the difficulty and subjectivity in specifying priors, fears about needing to use unfamiliar software, and fears about producing work that won't be understood and that therefore won't be publishable. The first of these challenges is an especially big one for me, as I don't countenance noninformative priors. I'm in the business of training informed epidemiologists. I hope none of my students would begin an analysis of real data by knowingly and explicitly stating that, say, a relative risk they're trying to estimate might just as well be zero, infinity or any value in between. Epidemiologists always have more information than that. I hope the epidemiologists I help to train will feel obliged to use that information, not free to ignore it.

So my students have to specify proper priors. At first, they don't like doing that any more than other newcomers to Bayesian methods. I've tried using a very small study in which a prior probability can be specified for every possible parameter value, but that is like pulling teeth. Now I use normal priors for log relative risks, but only as a rough and ready starter kit. Symmetrical priors are often not very realistic in my field, where we usually have pretty strong information favoring causation over prevention, or prevention over causation. The balance I'm seeking is to impart a view that priors are necessary, important, and subject to critical discussion, but without fostering too much obsession over specifying exactly "the right" prior.

I'm constantly searching for "hooks" into Bayes, by which I mean familiar contexts in which a Bayesian outlook seems not only unthreatening but obviously helpful or perhaps even necessary. For clinically oriented students, the interpretation of screening and diagnostic tests often serves nicely; clinicians and their patients reasonably want to know the probability of disease given a test result, but they can't arrive at this posterior probability without specifying a prior one. The "multiple comparisons problem" is a nice hook for some students, for many of whom a Bayesian outlook creates an affirmative way out of a perplexing dilemma of being forced to choose between the result and the ensemble as the reference frame. A growing interest in multilevel modeling, occasioned in part by a recent renaissance in social epidemiology, puts some students into a hierarchical frame of mind in which prior expectations enter naturally. The occasional student who is familiar with medical cartography, where Bayesian methods for smoothing rates have enjoyed a long tradition, may not find other applications quite as alien as some other students do. Lately I've considered illustrating the use of prior information on confounders, on the hunch that this application might rank relatively low on the "threatening scale" for many students. But perhaps the best hook for epidemiology students into Bayes at the present time, however, is meta-analysis. I often describe a well-done meta-analysis, especially a meta-regression analysis providing a summary estimate from studies "like this one", as a reasonably good source for a prior distribution.

BAYESIANS IN JAPAN

by Yasuhiro Omori and Hajime Wago
omori@bcomp.metro-u.ac.jp
wago@ism.ac.jp

In 1952, a group of students met at the Institute of Statistical Mathematics (ISM) every week to study *Statistical Decision Function* (Wald, 1950). It was a beginning of the Bayesian research in Japan. Some of them were students from Department of Mathematics at University of Tokyo and were attracted by the mathematical beauty of Wald (1950). Among them were Yukio Suzuki, Kenichi Inada, Hirofumi Uzawa, Kameo Matsuhita. Suzuki became one of the first Bayesians in Japan, while Inada and Uzawa moved their interest to the economics, and are now well-known economists. There is no Department of Statistics in Japan that deals with classical statistics and Bayesian statistics. Statistical education is performed in each department or institute. Most statisticians belong to the department of economics or engineering. In Japan, each fields has its own statistical association, related to their origin. Unfortunately, this situation hinders the chance of gathering a research team that studies Bayesian Statistics. We have not captured how many members of ISBA are there in Japan. A brief look at statistics community in Japan may show that most of the statisticians are non-Bayesian. However, it seems quite interesting that they accept the Bayesian approach as a useful method with a good theoretical base in statistical decision theory. Suzuki was hired by ISM in 1953, and continued reading the literature on Bayesian statistics with Matsushita and Hiroshi Fujimoto, such as *The Foundations of Statistics* (Savage, 1954), *Game Theory and Statistical Decision* (Blackwell and Girshick, 1954) and *Applied Statistical Decision Theory* (Raiffa and Schlaifer, 1961). In 1960, there was a conference held by International Statistical Institute in Tokyo, and Suzuki gave a talk on "Sequential Sampling Plan" using non-Bayesian approach. Lindley from London University suggested him to use Bayesian approach, and he published his first paper using Bayesian approach in *Annals of Institute of Statistical Mathematics*. Thus Bayesian research started to unfold gradually at ISM. Nozomu Matsubara, a researcher of ISM, became a student of Chernoff at Stanford University (1968-1972) and published Bayesian articles as well as some Japanese textbooks on the statistical decision theory (Fujimoto and Matsubara, 1976 and Matsubara, 1977). He is also one of pure Bayesians in Japan such as Suzuki and Fujimoto, and is now at University of Tokyo. Around that time, Yoshiko Nogami, returned from Michigan State University and became an empirical Bayesian at University of Tsukuba. There was another movement of Bayesian activities starting with Koichi Miyazawa at Department of Economics, University of Tokyo. He started his research on the statistical decision theory with Akio Kudo who wrote several papers on slippage problems at Kyusyu University. When Miyazawa first tried to raise a new generation of Bayesians, non-Bayesian approach was dominant at University of Tokyo. This is because Motosaburo Masuyama who was a non-Bayesian statistician at Department of Medicine received Asahi award (award from the newspaper) for his book on non-Bayesian statistics. To overcome the difficulties, Miyazawa called Suzuki to University of Tokyo in 1968, and thence it provided another stronghold of Bayesian research. They used textbooks such as Savage (1954), *Optimal Statistical Decisions* (DeGroot, 1970) , *Bayesian Inference in Statistical Analysis* (Box and Tiao, 1973), and *Statistical Decision Theory: Foundations, Concepts, and Methods* (Berger, 1980). Among their students were Shintaro Sono (Hokkaido University), Michikazu Aoi (Keio University), Akira Sekiya (Daito Bunka University), Michiko Hiradate (Kanazawa University of Economics), Kensei Araya (Fukushima College) and Yasuto Yoshizoe (Aoyama Gakuin University). In 1971, Miyazawa wrote some books on the statistical decision theory, but later moved his interest towards management science such as information and system theory. On the other hand, Suzuki continued his work on Bayesian research and also wrote some Bayesian textbooks in Japanese. Kazuo Shigemasu and Hiroshi Watanabe are also Bayesians from School of Education at University of Tokyo. Shigemasu obtained a Ph.D. degree under the supervision of Melvin Novick at University of Iowa in 1974 and now at Department of Life Sciences, University of Tokyo. He applied Bayesian approach in the field of psychology using multivariate analysis, structural equation model, and item response model. Shigemasu wrote a couple of Bayesian textbooks in Japanese. Similarly Watanabe obtained a Ph.D. degree under the supervision of Novick in 1977 and now at School of Education, University of Tokyo. On the other hand, Hiroki Tsurumi (Rutgers University) got a Ph.D. degree at University of Pennsylvania in 1967 under the supervision of Lawrence Klein, making econometric models of U.S. automobile industry and companies. When he attended the 2nd World Congress in Econometrics in England in 1970, there was an invited lecture by Harold Jeffreys chaired by Arnold Zellner. Jeffreys' lecture made him more curious about Bayesian statistics. Realizing that most of the Bayesian statistical papers and books were devoid of applications, he decided to learn Bayesian methods by applying them to some econometric models. He wrote papers on "Bayesian estimation of CES production functions", "A Bayesian test of the product life cycle hypothesis–application of a gradual switching regression", and "Bayesian test of a parameter shift". (These papers were later published in the *Journal of Econometrics* in 1976-1977). Tsurumi took a sabbatical leave from Queen's University in Kingston, Ontario, Canada and stayed at the Japan Economic Research Center for 1973-1974. At the suggestion of Hajime Wago and Kuriyama, he gave a series of seminars focusing on Bayesian applications, supported by the Fuyo Research Center. Wago is now a Bayesian econometrician at Nagoya University and works on cointegration analysis of time series, structural change in production functions, Bayesian estimation of business cycle model using such as an asymmetric GARCH model and Markov switching model. Let us move back to ISM. Hirotsugu Akaike is also at ISM since 1952 and well known for his Akaike Information Criterion. He was one year senior to Suzuki at University of Tokyo, and was originally non-Bayesian who wrote papers on state space modeling of time series. In 1976, he started Bayesian research when he visited Harvard University as Vinton Hayes Senior Fellow. He is not a pure Bayesian, but rather a practical Bayesian in the sense that he considered that a Bayesian model is just another type of statistical model for extracting the information provided by the data. His practical attitude toward statistical method was acquired by his visit to Tukey at Princeton University during 1966-1967. Akaike (1978) found that the use of AIC values for a prior weight of competing models improve the accuracy of prediction (which is related to Bayesian model averaging). He also applied Gaussian smoothness prior to the seasonal adjustment of time series and gave a talk at the first International Bayesian Meeting in Valencia, Spain in 1980. Akaike influenced many researchers to work on Bayesian statistics at ISM in the Eighties. Among them were Makio Ishiguro, Yoshihiko Ogata, Genshiro Kitagawa, Yoshiyuki Sakamoto, Giichiro Suzuki and Kunio Tanabe. Ishiguro worked on Bayesian time series modeling and seasonal adjustment using BAYSEA (Bayesian seasonal adjustment program). Kitagawa also conducted a research on seasonal adjustment of time series using Decomp (its on-line version, Web-Decomp, is now available by Seisho Sato at ISM). He is well known for the smoothness prior models of Kitagawa and Gersh (1984) for the state space modeling of time series with trend and seasonality. Ogata wrote papers on point process, spatial statistics, statistical seismology and Markov chain Monte Carlo method, while Sakamoto focused on Bayesian analysis of categorical data. Suzuki G. pursued empirical Bayesian analysis, and Tanabe worked on Bayesian nonparametric density estimation and inverse problem. In the late Eighties and the Nineties, Takashi Nakamura used Bayesian Cohort model to analyze data obtained using a multiple-choice question, and Nobuhisa Kashiwagi wrote papers on the spatial smoothing. Yukito Iba worked on the improvement of Markov chain Monte Carlo sampler in the computational physics, and Bayesian variable selection in multiple regression models and neural network models. In the field of Geophysics, Tomoyuki Higuchi developed Monte Carlo filter using the genetic algorithm operator and also designed Bayesian seasonal adjustment of the time series in the natural phenomena. In 1993 the first International Bayesian meeting was held in Fuji, Japan. This meeting was requested by Zellner and organized by Tanabe and Wago, supported by ISM. Many old and new friends attended including Arnold Zellner, Jim Berger, Jose Bernard, Dale Poirier, Herman van Dijk, Wolfgang Polasek, John Geweke, Hiroki Tsurumi, Seymour Geisser, Noel Cressie, Yu-ch-ho and others. This was quite a nice opportunity for graduate student majoring statistics and econometrics to attend such a stimulus meeting. After that several Bayesian sections in two international meetings were held by ISM. Some Japanese presented their papers in Valencia Meeting (from the first to the sixth: Akaike, Suzuki, Wago, Arai, Shigemasu, Terui, and Sekiya), ISBA (Wago), Riverboat Conference (Tsurumi, Wago, Kato, and Naniwa), ISI and other international professional societies meetings. Bayesian activities become widespread in the Nineties. Humiyasu Komaki (Department of Mathematical Engineering and Information Physics, University of Tokyo) uses Bayesian approach to state-space modeling of time series sampled from continuous process with pulses. Hirohisa Kishino (Graduate School of Agriculture and Life Sciences University of Tokyo) recently started to use Bayesian hierarchical model in the molecular phylogenetics. Yasuo Ohashi (Department of Medicine, University of Tokyo) also started to apply Bayesian approach in Biostatistics. Eiichiro Funo (Kanto Gakuin University) works on Bayesian estimation under sampling with missing, pooling and censoring. Katsuaki Iwaki (Asia University) visited Purdue University in 1996, and now works on the noninformative prior for model selection. Further, there has been an increasing number of Bayesian researchers at Department of Economics, working on Markov chain Monte Carlo method. Among them are Manabu Asai (Ristumeikan University), Hideo Kozumi (Hokkaido University), Teruo Nakatsuma (Keio University), Yasuhiro Omori and Toshiaki Watanabe (Tokyo Metropolitan University). Watanabe T. and Asai applied Bayesian estimation method in the field of financial econometrics such as stochastic volatility model. These applications in econometrics and financial econometrics are extending the boundaries of Bayesian activities to the financial engineers such as Toshinori Takayama (SG Yamaichi Asset Management) and Kasuya (Bank of Japan). Bayesian research activities are now very active in Japanese universities with strong interaction with universities abroad. We are planning a joint Bayesian research and meeting with Korea, China and the countries of Southeast Asia. We are also planning to hold a meeting on Bayesian application to financial econometrics using MCMC in the near future.

IMPLEMENTING MCMC

by Brad Carlin, Chris Holmes, Håvard Rue, Mike Smith and Simon Wilson

brad@biostat.umn.edu
c.holmes@ic.ac.uk
hrue@math.ntnu.no
mikes@econ.usyd.edu.au
simon.wilson@tcd.ie

➤ Editorial by Leo Knorr-Held

As an associate editor of the ISBA Bulletin I asked Mike Smith a few months ago if he could come up with an interesting topic on Bayesian software. Mike suggested to have a general discussion on how to implement MCMC these days. I liked the idea and so I asked a number of experts in the field to write a few lines about their views. Fortunately all five of them (including Mike) agreed and I am very glad to be able to present their experiences and recommendations in this issue. I like to thank all of the contributors as I believe these comments can be very helpful, especially for young PhD students, who are unsure which software to use to implement MCMC. It starts with Mike's comments on `Fortran`, other software discussed includes `C`, `MATLAB`, and `WinBUGS`.

➤ Mike Smith on `Fortran`

I use MCMC methods to estimate a range of time series and cross-sectional econometric models. These models typically involve a large number of parameters and are applied to datasets often numbering into the tens of thousands of observations. This gives rise to serious computational issues that are compounded by the fact that the sampling schemes usually have to be run for thousands (or tens of thousands) of iterates. Such computational demands push the limits of modern processors and equipment, and it is essential that the sampling schemes are coded in a thoughtful manner, and using a lower level language. Higher level languages, such as the otherwise excellent `MATLAB`, are simply not feasible options due to speed, storage and RAM requirements. However, my research focus is on methodology development, and I would very concerned to minimise code development time. Therefore, while it is common to turn to `C` or `C++` to tackle this task, I instead recommend my students implement their sampling schemes in `Fortran95` because I believe it represents the best balance between computational requirements and code development time.

The first thing to note about `Fortran95` (and the almost identical `Fortran90`) is that it is an almost unrecognisable improvement over `Fortran77`. Out are long lists of subroutine arguments, line indenting and endless loops for matrix calculations. In are basic object handling for defined types (such as banded matrices or various decompositions), more compact syntax, options to compile real numbers to quad precision (32 decimal places), a full suite of methods for efficient dynamic memory allocation and pointers. Excellent compilers can be found for all popular operating systems, and particularly useful is that the language is backward compatible, allowing easy access to the vast and stable `Fortran77` libraries. Moreover, `Fortran95` is well suited for use on parallel machines, with the source usually only requiring relatively minor alterations to compile to parallel. However, GUI and graphical support is weak, and it is much quicker to simply dump results to ASCII files, and read them into Splus or `MATLAB`, and use their vast plotting infrastructure to produce figures. However, when combined with IMSL, Lapack, Linpack and personal subroutine libraries, I have not yet come across a quicker way of efficiently coding sampling schemes at the more computationally intensive end of the spectrum.

➤ Håvard Rue on `C`

High-level computing software like `S-PLUS`, `R`, `MATLAB`, or `Octave` have become increasingly popular and provide a powerful framework to do exploration and testing of new ideas and algorithms. The benefits are obvious, a high-level environment allows you to code programs very efficiently as all sorts of utilities like operating on matrices and preparing plots are a natural part of the software. The cost is less efficient code, where "how much less efficient" depends roughly on how much time the code uses in its "environment" compare to its "core". As an example, consider

```
for j=1 to ncol(x)
  for i=1 to nrow(x)
    x(i,j) = 0.0
```

which is very inefficient compared to calling a built-in function (written in `C` or `Fortran`) `FillMatrix(x,0.0)`. A speed-up of 100 and more using the built-in function is not uncommon. For this particular example on a 100×100 matrix, I got a speed-up of 500 (!) compared to `Octave`, (a `MATLAB`-"clone" which is free software) on my laptop running Linux.

The majority of all MCMC programs use some kind of single-site updates, which naturally have the overall structure for models specified through its directed acyclic graph

```
for iter=1 to Niter+Nburnin
  for j=1 to Nvariables
    Update(j,variables,data)
```

where the function `Update` contains a loop like

```
for k in Neighbours(j)
  logdens +=
    lcd(k, variables, data),
```

here `lcd()` denotes a function which evaluates the corresponding log conditional density. Obviously MCMC programs are a natural candidate for running very inefficiently using a high-level programming environment, since they often consist of simple operations (i.e. single updates) inside nested loops. Hence, the speed-up writing the same code in C or C++ can be quite surprising. One may argue that coding using loops inside S-PLUS, say, is not very smart and there are better and more efficient ways to do it; Well, I agree, but most people "think in loops" and will implement the algorithm in the way they think.

I have experienced MCMC schemes for even quite simple models coded in S-PLUS, which ran to slow to get reasonably precise estimates. After re-coding it into C, we got precise estimates in less than a minute. I am sure the Splus-code could be written more efficient, but the person who wrote it could not and was not able to code it in C or C++. Perhaps this is the point; Those who are able to write (relatively) efficient S-PLUS-code are also those with enough interest in programming to learn C or C++, and they will even choose/combine the environment/language which is most efficient for each problem to optimize their own time! There is no reason why PhD-students in statistics using MCMC should not be able to learn C or C++ and thereby get some basic knowledge in computing, so the faster they master it the better!

➤ Simon Wilson on combining MATLAB and C

I am using more and more the combination of MATLAB and C. This is mainly because most of my research is with engineers, who are usually familiar with MATLAB and know nothing of S-PLUS or WinBUGS. But I also find that MATLAB is much easier than S-PLUS for students to pick up at first, especially if they have programmed in C or Java as an undergraduate. It also runs faster than S-PLUS in my experience. However, it is important that you have the MATLAB "Statistics Toolbox" available; this has a range of extra plots, does regression, ANOVA, etc., and has functions for simulating from many different distribution functions. Without it, MATLAB has a very limited range of statistical functions.

Once the students really get into the MCMC, they start to complain at how slow MATLAB is! Then, if they have not done it already, I start them programming in C. MATLAB can call C routines to do the heavy computation, leaving it to plot output, run diagnostics, etc. This combination combines speed with the ability of MATLAB to display results well.

Disadvantages of this combination are that MATLAB is not a statistics package, and so lacks the large library of statistical routines that S-PLUS has built up. For example, there is not, as far as I am aware, a suite of convergence diagnostic functions available for MATLAB. All the useful options in WinBUGS for plotting kernel density estimators, etc. must also be written.

➤ Chris Holmes on MATLAB

MATLAB provides a general purpose environment for scientific programming and in particular mathematical computing. It comes with a large number of built in functions designed to automate many standard tasks. Typical functions that I tend to use include, matrix inversion, optimisation, sort routines and random number generation from various distributions.

MATLAB has I feel a great deal to offer the user wishing to code up an MCMC algorithm for simulating Bayesian statistical models. Some attractive features of the MATLAB language are (in no particular order): (i) speed of development; (ii) ease of implementation and interpretation; (iii) run-time graphics facilities; (iv) numerical stability and quality of built in library functions; (v) debug editor. Its major disadvantage in comparison to a compiled language such as C or Fortran is its relative speed; at a rough guess around 10 times slower, though MATLAB would appear to be much faster than many other interpreted languages such as S-PLUS.

The points raised above have a number of practical implications for MCMC simulation. Features (i) and (ii) enable the rapid prototyping of ideas as working code can be written extremely quickly; the excellent debug facilities (v) also help with this. This can save days of development time over the same program written in say C. This facilitates the testing of more speculative ideas as you know that potentially you will only waste a couple of hours of programming if the method does not work! The second point (ii) is also significant if you are working on a joint project with co-workers. Trying to read even my own C code after a period away from it is usually a painful experience. Trying to read someone else's C code is in my experience practically impossible. MATLAB code is remarkably interpretable as the basic syntax is lifted from linear algebra and the function names are transparent. Moreover, I have not come across any difficulty running the same code on different operating systems and on different platforms such as Sun, DEC or PC. The run-time graphics mentioned in (iii) allow the display of summary statistics from the Markov chain during simulation. This is useful for monitoring convergence as well as highlighting any unusual behaviour that would suggest that the program is not doing what you expect it to do. Numerical stability (iv) can be important for instance if you are inverting large matrices with large condition numbers (ratio of smallest to largest eigenvalues). MATLAB's pseudo random number generators have excellent properties which is incredibly important when performing Monte Carlo simulation where we may implement millions of calls to a generator.

My only real gripe concerns the speed. Compiled languages remain much faster especially for routines that use nested loops for which MATLAB is notoriously poor. A compiler is available for MATLAB which will translate MATLAB code into C (or Fortran) code which can then be compiled. However, I have never managed to get this to work properly, which of course is not to say that it is not entirely straightforward.

In conclusion my recommendation would be to always implement first within MATLAB for rapid development and test of concept. If speed is subsequently an issue then redevelop in C, C++ or Fortran using the MATLAB code as a blue print. Finally check that the results concur.

➤ **Brad Carlin on** WinBUGS

Appendix C of my book with Tom Louis (Carlin and Louis, 2000, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall/CRC Press) contains a brief review of many commercial and noncommercial packages available for Bayesian analysis. While more are of course emerging every day, of those currently available, only the BUGS package www.mrc-bsu.cam.ac.uk/bugs seems to offer enough generality to be considered as an all-purpose Bayesian model-fitting tool. The original UNIX BUGS product was essentially only a sampling engine, and so needed the assistance of the CODA (and now BOA) functions for output analysis and convergence diagnosis (see the latter package's website at www.public-health.uiowa.edu/boa). But the Windows version of BUGS, WinBUGS, does offer enough extra plotting, diagnostic, and post-sampling summary features that it can be used on its own, at least for many standard models (e.g. hierarchical linear models using standard distributional families). The package is also now sufficiently bug-free that I do not hesitate recommending it to statisticians used to the high-quality output produced by SAS or S-PLUS. It is also general enough that I no longer teach anything else in my Masters-level courses or short courses, though in my PhD-level course here at Minnesota I still require students to work a couple of problems in Fortran or C, since their dissertation work may well involve models still beyond WinBUGS' reach.

Still, at least in my own area (spatio-temporal modeling), this reach continues to extend: the next release of GeoBUGS, the spatial statistics add-on to WinBUGS, will offer functions for both lattice and geostatistical (kriging) models, as well as the ability to read in geographic boundary files from either S-PLUS or ARC/INFO. This latter part is particularly exciting, since it means we can now use the maps library in S-PLUS to output county boundaries for any state in the US, read them into GeoBUGS, perform the required MCMC sampling, and draw summary maps of fitted county-specific rates suitable for publication all within the Bayesian sampling software; there is no need for the user to learn a GIS like ARC/INFO or ArcView at all. Coupled with language developments in other areas (e.g. the PKBUGS add-on for pharmacokinetic modeling), WinBUGS seems to remain the best choice for users seeking an entre into Bayesian analysis that is both relatively easy to learn, yet general enough to justify its modest learning curve.

ANNOTATED BIBLIOGRAPHY ON SOME RECENT CONTRIBUTIONS TO BAYESIAN MIXTURE MODELING

by Paul L. Mosquin
and Steven F. Arnold
mosquin@stat.psu.edu
sfa@stat.psu.edu

Research in Bayesian mixture models continues to be an active and diverse field. Here we consider some recent contributions, grouped roughly according to primary area of focus.

➤ *Label Switching*

The usual invariance of mixture likelihoods and priors to permutation of component labels leads to a similar invariance in the posterior distribution, and a tendency for component labels to relabel during simulation.

• G. CELEUX (1998) **Bayesian inference for mixtures: The label switching problem**. In: *COMPSTAT 98* R. Payne and P Green (Eds.) Physica-Verlag, 227-232.

A clustering method is developed for unraveling the component relabelings. Simulation of a two component exponential mixture reveals that the method performs favorably compared to either the use of raw simulation output or the imposition identifying of component parameter constraints.

• M. STEPHENS (2000) **Dealing with label switching in mixture models** *Journal of the Royal Statistical Society, Series B.* 62, 795-809.

Presents a Kullback-Leibler based loss function algorithm to post-processing of mixture simulation output so as to unravel mixture component label switching. Performance of the algorithm is examined and some modifications discussed.

• G. CELEUX, M. HURN, AND C.P. ROBERT (2000) **Computational and inferential difficulties with mixture posterior distributions** *Journal of the American Statistical Association*. 451, 957-970.

This paper considers the label switching problem in mixture simulation. A review of an earlier clustering based approach is provided, and its performance compared with that of some new approaches based on a loss function. The relabeling algorithms are found to perform in a largely similar manner, and provide an improvement over the use of imposed constraints.

➤ *Prior Specification and Model Parameterization*

The inability to directly specify improper priors for the component parameters of a finite mixture has motivated the development of priors and parameterizations which allow subjective input to be minimized.

• K.L. MENGERSEN AND C.P. ROBERT (1996) **Testing for mixtures: a Bayesian entropic approach**. In: J.O. Berger, J.M. Bernardo, A.P. Dawid, D.V. Lindley, and A.F.M. Smith (Eds.) *Bayesian Statistics, Vol. 5*, Oxford University Press, London, 255-266.

This paper considers the use of the entropy (Kullback-Leibler distance) between a non-mixing and a two-component mixing distribution to test whether a mixture structure is present. Univariate normal models are considered with a new parameterization that separates certain location and scale parameters from remaining parameters of interest, and allows for a partially proper prior.

• C.P. ROBERT AND K.L. MENGERSEN (1999) **Reparameterization issues in mixture modeling and their bearing on MCMC algorithms** *Computational Statistics and Data Analysis*. 29, 325-343.

This paper develops earlier work to extend partially proper parameterizations of the finite mixture model to arbitrary number of components. Two such parameterizations are considered, and one found to be superior according to convergence properties.

• K. ROEDER AND L. WASSERMAN (1997) **Practical Bayesian density estimation using mixtures of normals**. *Journal of the American Statistical Association*. 439, 894-902.

A partially proper prior is presented for the analysis of normal mixtures of fixed component number. The prior leads to a proper posterior, and does not require subjective input for either the location or the scale. Consistency results are given as the number of components grows with the sample size.

• L. WASSERMAN (2000) **Asymptotic Inference for mixture models using data-dependent priors** *Journal of the Royal Statistical Society, Series B*. 62, 159-180.

Data dependent priors are studied as a way to introduce vague priors into the mixture model. These priors are shown in some cases to be the only priors that produce intervals with second order correct frequentist coverage. In particular this is shown for mixtures of univariate normals. Other results are given as well as interpretations of the resulting posterior distributions.

➤ *Selecting the Number of Components*

These articles provide approaches for determining the number of components present in the mixture model.

• S. RICHARDSON AND P.J. GREEN (1997) **On Bayesian analysis of mixtures with an unknown number of components**. *Journal of the Royal Statistical Society, Series B.* 59, 731-792.

Reversible jump methods are applied to the modeling of finite mixtures allowing for a random component number and so a fully Bayesian approach to the problem. These methods allow the Markov chain simulation to reversibly move between model subspaces of differing dimension, as represented by mixtures of differing component number. The methods are applied to hierarchical univariate normal mixture model using a variety of datasets. An extensive, insightful discussion follows.

• M. STEPHENS (2000) **Bayesian analysis of mixtures with an unknown number of components - an alternative to reversible jump methods**. *Annals of Statistics* 28, 40-74.

A birth-death process approach to analyzing mixtures of a random number of components is given. The process is a marked point process, with the mixture component parameters being points in its sample space. The method is implemented on univariate and bivariate examples. Issues relating to prior selection and choice of birth distribution discussed, as well as extension of the point process approach to non-mixture problems are discussed.

• H. ISHWARAN, L.F. JAMES, AND J. SUN (2001) **Bayesian model selection in finite mixtures by marginal density decompositions**. *Journal of the American Statistical Association*. To appear.

In this paper a prior is placed over the space of mixing distributions of at most K components. Decomposition of the marginal component number density leads to a weighted Bayes factor method for consistently estimating the true component number. Simulation is conducted by use of the generalized weighted Chinese restaurant algorithm and the blocked Gibbs sampler. The rate of estimation is shown to be the frequentist optimal $O_p(n^{-1/4})$.

➤ *Dirichlet Process Mixtures*

Dirichlet process mixture models are so named since the prior for the mixing distribution is modeled as a Dirichlet process. The resulting mixtures of countably infinite component number provide a broad approach to the mixture problem.

• S.N. MACEACHERN AND P. MÜLLER (1998) **Estimating mixtures of Dirichlet process models**. *Journal of Computational and Graphical Statistics*. 7,223-238.

Gibbs sampler approaches to the simulation of Dirichlet process mixtures with non-conjugate priors are presented. The methods are applied to a bivariate normal dataset. Approaches for extension to incorporate hyperparameters and for prediction are described, and convergence issues discussed.

• R. NEAL (2000) **Markov Chain Sampling Methods for Dirichlet Process Mixture Models**. *Journal of Computational and Graphical Statistics*. 9, 249-265.

Methods for Dirichlet process mixtures are further developed for situations with non-conjugate priors. These methods are based on introduction of a Metropolis-Hastings or introduction of both a Metropolis-Hastings step and a Gibbs step. Simulations are conducted to evaluate relative performance of developed and existing approaches, with some of the developed approaches showing good performance. A review of previous work on Dirichlet process mixtures is provided.

➤ *Reviews*

Finally, the following review chapter provides an introduction to the area.

• G. MCLACHLAN AND D. PEEL (2000) **Bayesian approach to mixture analysis** Chapter 4 in *Finite Mixture Models*, Wiley, 117-134.

Gives a general overview of the developments in the field. Also covers some Bayesian methods found largely in the computer science literature such as a variational approach, and one based on minimum message length.

NEW CONTRIBUTIONS

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

This section includes three papers. The first one consists in an abstract of the thesis of Dr. Alessandra Nuccitelli. The other two are summaries of the investigation that two students are doing about applications in the financial field.

Alessandra Nuccitelli
Università di Roma Tre
nuccitel@istat.it
Exact matching techniques: a review and a Bayesian proposal.
Advisor: Dr. Brunero Liseo

Record linkage (or exact matching) refers to the use of an algorithmic technique to identify pairs of records (a, b), a from file A and b from file B, that correspond to the same statistical unit (e.g. individual, business, dwelling). The need for record linkage procedures is ubiquitous in official statistics; applications can be characterised as falling into two broad groups: 1) problems where it is desired to draw inferences about relationships between variables collected in different data files; 2) problems where interest focuses directly on the number of individuals represented in one or both data files.

Clerks can review pairs of records, but that is time consuming and costly. Given a decision rule, computers can quickly designate each pair as a link, a non-link or a possible link. The decision rule is based on the comparison between the common fields (or matching variables) of the two files of records. Unfortunately, matching variables are likely to be observed with errors and that makes the linkage process not trivial. From a classical viewpoint, the decision rule is chosen in order to minimise the expected number of possible links at specified error levels. As a measure of performance of the procedure, the False Match Rate (FMR), defined as the number of false declared matches divided by the total number of declared matches, can be adopted.

In this work a fully Bayesian approach to record linkage is proposed. The Bayesian framework is particularly suitable for the solution of the following problems: 1) exact computation of the probability that each pair is a match, conditionally on the observed data (the comparisons between the matching variables); 2) computation of conditional probabilities that several pairs are simultaneously matches. The second point represents an improvement on the classical methods, where decision rules establish separately for each pair if the records correspond to the same unit or not, without considering the compatibility constraints, unless additional procedures based on operational research techniques are used. In this approach a matrix, which indicates the pattern of matches in the two files, is considered as the quantity of interest. The combinatorial nature of the record linkage problem makes the analytic use of the marginal posterior distribution for the matrix practically impossible. So, a Metropolis-Hastings algorithm is used to extract the posterior distribution. Some possible inferential summaries are introduced: one is based on a decision theoretic approach, using a loss function which is based on FMR and False Non-Match Rate, defined as the number of false declared non-matches divided by the total number of declared non-matches; the other is a more eclectic approach which slightly deviates from a Bayesian road at the reward of a more flexible analysis. In this work, the use of the posterior mode(s) as point estimate(s) of the parameter of interest is discussed. The proposed methodology is applied to simulated data. For the application presented, the knowledge on truth and falsehood is used to evaluate the performance of the posterior mode, detected by a simulated annealing algorithm, compared to the classical decision rule implemented in the Record Linkage Software, developed by the Bureau of the Census.

Extreme value analysis and outlier detection: a Bayesian procedure

Stefano Cabras
Department of Statistics.
University of Florence.
cabras@ds.unifi.it

This work considers the problem of estimating the probability of an extreme event for the following independent random variables $X_1, X_2, ..., X_n$: $\Pr(X_i \leq x | X_i > u), i = 1, 2, ..., n$, where u is regarded as an *high* threshold. If the empirical distribution of X_i has heavy tails, then an adequate model, for describing the probability of extreme events, is the Generalized Pareto Distribution (GPD). The parameter u, of the GPD, has to be specified in order to identify those observations to be considered as extremes. Since the u parameter determines the tails beginning or the support of the

GPD, then a MLE of u is not available, so empirical frequentist methods are often used.

The aim of this work is to propose a Bayesian method for estimating u by defining as extremes, those observations that are outliers for a specified model f with unknown parameters, θ. By this way, the original set $X_1, X_2, ..., X_n$ is partitioned in two subsets: $\overline{E} = \{X_1, X_2, ..., X_j, \ i.i.d. \ f\}$ and $E = \{X_{j+1}, X_{j+2}, ..., X_n, \ i.i.d. \ GPD\}$. The Bayesian estimation of θ is conducted under the assumption that not all observations are informative for the parameters, in particular when θ has the meaning of location and scale parameters, the observed maximum is suspected to give no information. Under the assumption of $X_i \sim f$, the u parameter is the threshold by which all $\{X_i : X_i \in E, x_i > u\}$. This result is due to the Pickands's theorem which state that the tails of a wide class of models - f included - belong to Domain of Attraction of the GPD.

The partial posterior predictive $p-$value (p_{ppp}) has been taken as a measure for outlier detection; it produces the estimation of the threshold, $\widehat{u}$. The main problem, that arises in multiple outliers detection, is the presence of the well known *swamping* and *masking* effects, that grow weak every detection procedure. In order to partially overcome this problem, it has been used an expert elicitation on the amount of possible outliers in the sample. By this way, all the observations indicated by the expert are dropped out and the p_{ppp} calculated. The assumption of $X_1, X_2, ..., X_n$ independent leads to assume as outliers all observations larger then $\widehat{u}$ (included those indicated by the expert). However the robustness of $\widehat{u}$ has to be verified with respect to the expert elicitation.

The method has been applied to the Value at Risk analysis on a time series of ten years daily stock *returns*(defined as the log difference of two prices). For this series a normal model f has been considered. In order to show the robustness of the estimations, it has been observed that expert elicitations, varying in the range $[0.04, 0.08]$ of percentages of possible outliers, lead to $\widehat{u}$ in the range $[0.052, 0.062]$ and, in general, $\widehat{u}$ variability is an half of the expert elicitations' variability. All numerical results have been obtained by implementing MCMC methods.

Measuring the market risk

Gonzalo Garcia-Donato
Department of Economics.
University of Castilla-La Mancha.
gdonato@correo.ecem-ab.uclm.es

In a financial context it is important to obtain measures of the market risk of a portfolio of financial assets. One of that measures is Value at Risk (VaR). VaR usually is defined as the maximum loss that a portfolio can achieve during a certain planning horizon with a previously determined probability c. Assuming a planning horizon of one temporarily unit, VaR is proportional to the $1-c$th percentile of the distribution of R_{t+1}, where R_t represents the portfolio return at present instant t.

In the Bayesian way, k historical portfolio returns, r$= (r_t, \ldots, r_{t-k+1})$, and a priori information, I_t, can be used to obtain the predictive distribution $p(R_{t+1} \mid \text{r}, I_t)$. It is well known that financial returns exhibit high positive kurtosis, therefore the Student model has been proposed as more adequate for explaining the distribution of financial returns. We have applied that model to obtain a VaR with an appealing interpretation in terms of kurtosis and that it is very easy to calculate.

Aside from the model selection for the financial returns, there are two questions that are of special interest.

On the one hand we would rather include the problem of VaR in a decision problem, finding loss functions that exclude the (sometimes) arbitrary election of c. On the other hand, the frequentist characteristics of VaR are also interesting, since these ones can be used as management devices in the medium and long term. It is wanted that the proportion of periods where the change of the portfolio value is superior to VaR is close to $1-c$. The high dependence between consecutive VaR makes that the *probabilistic* VaR does not have a good frequentist behavior. Therefore, it is necessary to define in a different manner the VaR for those cases.

NEWS FROM THE WORLD

by Antonio Lijoi
lijoi@unipv.it

✻ *denotes an ISBA activity*

➤ Events

The Second Annual NIPS Unlabeled Data Competition and Workshop. *December 7, 2001, Whistler, Canada.*
This competition is a challenge to the machine learning community to develop and demonstrate methods to use unlabeled data to improve supervised learning. There is a web-site where participants can download and submit problem sets and compete head to head with other contestants in a series of challenging unlabeled-data, supervised-learning problems. Details at q.cis.uoguelph.ca/~skremer/NIPS2001/

Haifa Winter Workshop on Computer Science and Statistics. *December 17-20, 2001, Haifa, Israel.*
The purpose of the workshop is to bring together experts from the fields of computer science and statistics and to explore potential areas of research in order to stimulate collaborative work. Particular areas of interest are: data mining, simulation-based computation, expert systems, automated learning and robotics. For further information please contact libi@cs.haifa.ac.il.

***I COBAL.** *February 3-7, 2002, Sao Paulo, Brasil.*
The First Latin American Bayesian Meeting (I COBAL) is the first meeting by ISBRA (the Brazilian Chapter of ISBA) after its foundation last July during the XIV Brazilian Symposium of Probability and Statistics. The main objective of the I COBAL is to enhance scientific integration and exchange across Latin American Bayesian communities by promoting researches recently developed by latin american researchers. I COBAL will also have the sixth edition of the Brazilian Meeting of Bayesian Statistics (EBEB) as a satellite conference. The EBEB has been a biannual national meeting since 1991 with recognition and promotion by the Brazilian Statistical Society since 1997. The EBEB organization has been transferred to ISBRA since July 2000. Long talks, short talks and poster sessions are part of I COBAL's scientific program. The deadline for abstracts is November, 10, 2001. Detailed information are available at: www.est.ufmg.br/cobal/indexIng.htm

International Conference on Current Advances and Trends in Nonparametric Statistics. *July 15-19, 2002, Crete, Greece.*
The conference will highlight the major trends in several areas of nonparametric statistics. Among other themes, the conference will also focus on Bayesian methods in nonparametric statistics. Website: www.stat.psu.edu/~npconf/index.html

International Biometric Conference 2002. *July 21-26, 2002, University of Freiburg, Germany.*
Deadline for abstracts: January 15, 2002. Website: www.ibc2002.uni-freiburg.de

Symposium on Stochastics and its Applications – ICM 2002 satellite conference. *August 15-17, 2002, National University of Singapore.*
The symposium is a satellite conference of the International Congress of Mathematicians that will be held in August 20-28, 2002, Bejing, China. The main topics the symposium will deal with are: financial mathematics, Gaussian random fields, Markov chain Monte Carlo, probability approximations and random matrices. Website: www.math.nus.edu.sg/ssa

ENBIS 2002. *September 23-24, 2002, Rimini, Italy.*
The Second Annual Conference of the European Network for Business and Industrial Statistics aims to create a forum for users of statistics to get together, share ideas and network. Information at www.ibisuva.nl/ENBIS/ or contacting Soren Bisgaard (s.bisgaard@monitoring-intl.com) or Fabrizio Ruggeri (fabrizio@iami.mi.cnr.it).

Third International Conference on Data Mining Methods and Databases for Engineering, Finance and other Fields. *September 25-27, 2002, Bologna, Italy.*
Data Mining 2002 provides an international forum for researchers and application developers from many different areas, to share state-of-the-art research results and practical development experiences. Main topics of the conference are: data warehousing and databases, web mining, data analysis and data mining on large databases, data mining methodologies, knowledge discovery and data mining, etc. Papers are invited on the topics outlined and others falling within the scope of the meeting. Deadline for abstracts: January, 8, 2002. Website: www.wessex.ac.uk/conferences/2002/datamining02/index.html

➤ Internet Resources

Bayes Linear Methods home page.

The website `fourier.dur.ac.uk/stats/bayeslin/` contains useful information about Bayes linear methods. It contains a series of technical reports providing an introduction to Bayes linear methods. Moreover, one can find a list of researchers currently working on the subject and a tutorial on solving basic problems using a computer programming language named [B/D].

Open Bayes discussion group.

Richard Dybowski formed the OpenBayes discussion group/email list on January 17, 2001. The goal is to discuss the development of an open source library for probabilistic graphical models. It contains some useful links to some relevant articles, assembled (and mostly written) by Kevin Murphy. One can find instructions on how to subscribe to the list at the following URL: `http.cs.berkeley.edu/~murphyk/OpenBayes/index.html`

➤ Awards and Prizes

***2001 Mitchell Prize.**

The 2001 Mitchell Prize has been awarded to Keisuke Hirano, Guido Imbens, Donald Rubin and Xiao-Hua Zhou for the Bayesian analysis of a substantive and concrete problem that appeared in their 2000 paper “Assessing the effect of an influenza vaccine in an encouragement design”. The award presentation took place at the SBSS meeting in Atlanta during the JSM in August, with Don Rubin accepting the Prize on behalf of his co-authors.

Students Awards Program.

The American Statistical Association’s Section for Statistics in Marketing offers two awards, with honoraria, to students who best apply statistical thinking to marketing problems. One award is for the best theory and modeling entry and the other for the best applied case study entry. Submissions in the theory and modeling award must describe a novel contribution to the statistics (or statistical modeling) literature as applied to marketing. Submissions to the applied case study will illustrate able application of statistical methodology in a marketing setting but need not make a unique contribution to the statistics literature. Each award is accompanied by a 2,000 USD scholarship to support students in traveling to and attending the Joint Statistical Meetings (to be held in New York City, August 2002) where they must present their work. The deadline for next year’s entry is April 15, 2002. For further details, contact prof. Bradlow: `ebradlow@wharton.upenn.edu`

➤ Miscellanea

Bayesian Statistics in Greece.

The Greek Statistical Institute is the official association of statisticians in Greece. Among other activities, the Greek Statistical Institute organizes an annual conference. The 14th Conference was held in April 2001 at the island of Skiathos. The Bayesian presentation was stronger this year. There was a session named “Bayesian Statistics and Decision Sciences” and several Bayesian talks were presented, including the talk of the Lefkopoulou award. Some of the topics were : Bayesian model and variable selection using MCMC, Bayesian modelling for the prediction of outstanding automobile claim amounts, Bayesian estimation of distribution parameters of random sums, a Bayesian approach to quality control problems. During the Conference, the Greek Statistical Institute assigned the 2001 Lefkopoulou award. This year the prize was shared by Ioannis Ntzoufras, for his PhD thesis “Aspects of Bayesian model and variable selection using MCMC”, and Haritini Tsangari, for her thesis entitled “New models and methods for nonparametric analysis of covariance (ANCOVA)”. The components of the Award Committee were Professors Stavros Kourouklis, University of Patras and Takis Papaioannou, University of Piraeus.

The Proceedings of the Conference can be requested at `esi@ath.forthnet.gr`. (*Thanks to Prof. Takis Papaioannou*)

Call for Papers - New INFORMS Journal on Decision Analysis

The Decision Analysis Society and INFORMS are going to introduce *Decision Analysis*, a new journal that focuses on the development and study of operational decision-making methods, drawing on all aspects of decision theory and decision analysis, with the ultimate objective of providing practical guidance for decision makers. The anticipated publication date of the first issue is scheduled for 2003. Interested researchers are invited to submit their decision analysis manuscripts for publication consideration. Further details about the journal’s editorial objectives, intended audience, and review process can be found at the journal’s website: `da.pubs.informs.org`

INTERNATIONAL SOCIETY FOR BAYESIAN ANALYSIS

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*