A MESSAGE FROM THE PRESIDENT

- Sonia Petrone -
ISBA President, 2014
sonia.petrone@unibocconi.it

Last summer, at a conference, I was discussing with Wilfrid Kendall, the President of Bernoulli, about possible joint initiatives between ISBA and the Bernoulli Society. He asked me for how long one serves as President at ISBA. I said it's one year. Smile and reply: "I know, that's the life of a butterfly! so short!" Short? I can assure you, the life of a butterfly is quite demanding! I cannot deny a certain relief now that my year of Presidency is coming to an end! Many initiatives have been going on this year at ISBA, many things to take care of. But speaking seriously, this is great, a sign of the liveliness of the Society! Bayesian Statistics is more and more relevant in many fields, and ISBA wonderfully reflects this spread of interests, applications, research, in the lively growing Bayesian community. My sincere thanks to all the people that generously contributed to ISBA activities this year, to the ISBA officers, the editorial Boards of our publications, the Committees, to all of you for supporting ISBA!

But before I keep on thanking you, let me take you on a butterfly-flight over some ongoing initiatives and objectives.

- We are proud that the ISBA official journal, *Bayesian Analysis*, continues to grow in prestige and diffusion, thanks to your contributions and the work of the editorial Board and of the Editor in Chief, Marina Vannucci. I'd like to underline here that we continue to offer online open-access of *Bayesian Analysis*, through ProjectEuclid, despite the costs. This is something we all should be proud to support - starting by joining ISBA!

- ISBA sponsors, co-sponsors and endorses many meetings around the world. As you know, we usually focus on the ISBA World Meeting in even years, and on different workshops on odd years. The World Meeting 2014 in Cancun was a great success, for which I would like to thank again Andres Christen and the Local Organizing Committee, and the Scientific Committee, chaired by Raquel Prado. Next year will bring many ISBA endorsed workshops, including the official meetings of the ISBA sections on Objective Bayes and Bayesian Nonparametrics, ISBA Chapters' meetings, and the 9th Workshop on Bayesian Inference in Stochastic Processes, BISP9.

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

More information at http://bayesian.org/meetings/ISBA-meetings.

With the growth of the Society, the ISBA World Meeting has become quite a big conference. The latest World Meeting, in Cancun, last July, had more than 400 participants. Sustaining this growth, while keeping something of the spirit of the Valencia meetings, the passion for sharing deep ideas and discussion, the sense of community, is an important challenge. We discussed about encouraging some special plenary moments at the World Meeting, and new initiatives will be reported in the next issue of the Bulletin. I will anticipate only one thing. This has been a sad autumn. We had lost Susie Bayarri, and last October we lost another great Bayesian statistician, Kathryn Chaloner. We will present, at the ISBA World Meeting 2016, initiatives to remember and honor Susie and Kathryn.

- Support to young researchers is a core objective for ISBA. ISBA initiatives for young researchers include prizes and numerous travel awards. The prestigious Savage award, co-sponsored by ISBA and by the ASA Section on Bayesian Statistical Science, SBSS, is assigned every year for outstanding doctoral dissertations in Bayesian theory and applications. The high quality of the applications is an exciting sign of the growth of Bayesian Statistics. I thank the 2014 Savage Committee, chaired by Judith Rousseau, for their work on the selections for the 2014 Savage Award.

  ISBA funds a generous number of Travel Awards for the World Meeting, and ISBA co-sponsorship of endorsed workshops is usually finalized to travel support for students and new researchers. A current objective is to give more visibility to ISBA Travel Awards, starting from proper announcement of the recipients in the Bulletin and at the meetings. Moreover, for a better planning, we are fixing a yearly deadline for applying for ISBA endorsement and travel support. Details will be promptly provided on the website, http://bayesian.org/meetings/planning, and in the Bulletin.

We unfortunately cannot afford yet to give free student membership. However, a first step has been taken, by acknowledging one free student membership for each ISBA member in countries with reduced subscription fees, starting from 2015.

A challenge that students and new researchers are facing, possibly more than others, is the impact of new technologies and big data - including their impact on the recruitment policies of many Statistics departments. As a Society, caring for such opportunities and challenges is unavoidable; and I believe that being open is good. The ISBA at NIPS initiative this year - reported in the last issue of the Bulletin - has been a first step. I also would like to thank Max Welling and Yee-Whye Teh for having taken on my call for a discussion in the Bulletin (with possible developments on *Bayesian Analysis*?), contributing a note – in this issue – about the First Workshop on Bayesian Inference with Big Data, that took place in Amsterdam last August.

- Enhancing collaborations with other Societies and supporting our profession are ongoing objectives. ISBA is a member of the Friends of COPSS. We encourage interaction of the junior section of ISBA, J-ISBA, with other young research communities and initiatives, such as the Bernoulli European Young Statisticians Meeting and the IMS Meeting of New Researchers in Statistics and Probability. Last year, Kerrie Mengersen was elected as the ISBA representative in the World of Statistics initiative. The latest news can be found at http://www.worldofstatistics.org/files/2014/12/December-9-2014.pdf. Let's all step in, communicate initiatives to Kerrie, and make Bayesian events more and more visible in the World of Statistics!

- A Continuing Education committee was created in 2012, to promote and coordinate short courses, webinars and other activities. These include the ISBA Latin American School of Bayesian Statistics, which aims at the dissemination of Bayesian methods and the promotion of an interaction between young statisticians in Latin America. Work is still needed to improve and finalize web-based Continuing Education initiatives, but I forecast a lot of energy co-

ming in from our new President, Alexandra Schmidt!

Many people have been doing fundamental work for the ISBA activities, and I thank them all.

The first, very special thanks go to our Past President, Merlise Clyde. Merlise has given to ISBA the most generous and extraordinary service, as the ISBA Executive Secretary in 2009-2011, and as an ISBA President in 2012-2014. Merlise is simply amazing. As the ISBA Executive Secretary, she has greatly contributed to significant improvements of the ISBA website, which now provides a wide range of crucial and ever-growing services to the Bayesian community. Merlise is "an amazing talent – wrote Mike Jordan in his presidential message in December 2011 - someone who is at home in the Web 2.0 world as she is in Bayesian theory and practice. After putting the kids to bed she appears to enjoy unwinding with a bit of Drupal and CiviCRM programming." I can tell that Merlise has been an even more wonderful President! She has a vision of ISBA life as nobody else has. Her advise in the ISBA Executive Board has been fundamental. And she has generously continued to offer her expertise, to be our website-Drupal-civiCRM magician, not only after putting the kids to bed, but also while being the Head of the Statistics Department at Duke University. We will never thank you enough for your support and advise, Merlise!

It has been a busy year, but a pleasure to work with a wonderful Board of Directors. I thank Steve Scott for his efficient, wise, prompt collaboration - and hilarious emails - as the ISBA Executive Secretary. Murali Haran, for starting his role as the ISBA Treasurer in the busy year of the World Meeting, and for surviving the difficulties - including the 1.45 hours driving to initially set the bank accounts - with kindness and calm, leading ISBA finances in excellent shape. A special thank you to Mike Daniels, who kindly accepted to serve as

the ISBA "Past Treasurer" in the Finance Committee, after being the ISBA Treasurer for 3 years. Having the " Past Treasurer" overlapping one year with the Treasurer has been greatly helpful and we plan to formalize such rotation in the future.

I am grateful to the ISBA Program Council, Ramses Mena, Michele Guindani and Raquel Prado - the scientific leaders of ISBA! I thank the members of the ISBA Board of Directors that will complete their mandate at the end of 2014: Kate Calder, Maria de Iorio, Igor Pruenster and Wes Johnson. A double thank you to Wes, who also wisely and efficiently served as the Chair of the ISBA Prize Committee in 2014. Thank you to Jim Berger for chairing the Committee on Fellows on its first year of activity, and to all the components of ISBA Committees for their precious work.

Last but not least, my sincere and grateful thank you goes to Gabriella Bonfanti, who started to collaborate with ISBA as Administrative Manager this year, and is giving precious support to ISBA officers. I could eventually steel and pass to Feng a picture of her to be published in the Bulletin, so you can all get to know her. Thank you Gabriella for your collaboration.

I am delighted to welcome the ISBA President Elect, Steve McEachern, and the new elected members of the Board, Sudipto Banerjee, Carlos M. Carvalho, Vanja Dukic and Alessandra Guglielmi, and all the new elected Sections' officers. I leave ISBA in good hands: to the Brazilian energy and German efficiency of the ISBA President Alexandra Schmidt.

My warmest wishes of a splendid 2015 to all of you! ▲

A Message from the Editor

- Feng Liang -
liangf@illinois.edu

Another great and fascinating year has passed away. As being summarized in the *last* message from our president, many activities and many in-

itiatives have been going on in our society this year. As the Editor of the Bulletin, I would like to thank Sonia for all her support and guidance; as a member of ISBA, I would like to thank her for taking us on such an exciting yet graceful butterfly-ride in the year of 2014!

I also want to take this opportunity to express my sincere gratitude to Isadora Antoniano and

Antonio Ortiz (the Associate Co-Editors of Students' Corner), Jarad Niemi (who's in charge of the web publishing of the Bulletin), and contributors like members from our Program Council, Section Chairs, Editor-in-Chief of BA, and many others. Thank you for all your help and support!

Big Data has started to affect our everyday lives in many ways, including the life of Mickey Mouse. I just got back from a trip to Disney World at Orlando. This is the first year Disney's FastPass goes into a paperless online system, and all the planning needs to be done at "My Disney Experience," an online vacation-planning system provided by Walt Disney. Every day hundreds and thousands bytes of data ranging from personal information to waiting time for a popular ride have been collected by this system from millions of Disney visitors, waiting to be analyzed by data scientists. On the one hand, it is good to know that opportunities for us statisticians in the era of Big Data are enormous, but on the other hand, we should be aware of the challenges that come with Big Data. That is what you will find out in several articles in this issue, including a special report on the first workshop on Bayesian Inference with Big Data.

In this December issue, you'll also find an update from our Editor-in-Chief of BA, announcements on ISBA World Meetings in 2016 and 2018 from our Program Council, reports on ISBA endorsed workshops, and other news in our society. In the Students' Corner Section, Julien Cornebise, a former Savage Award winner, is sharing with us his post-graduation experience, and a group of bloggers is discussing the pros and cons on blogging as young researchers.

Before closing my message, I want to sadly bring your attention to the Obituary for Kathryn Chaloner, another great Bayesian statistician we lost in 2014.

Last, but not least, a word of thanks to YOU, dear readers of the Bulletin! In order for the Bulletin to become better, we need the help from YOU, the readers, and the best way to help us is to email us your suggestions/contributions in 2015!   ▲

ISBA AT NIPS

- Raquel Prado, Ramsés H. Mena and Michele Guindani -
ramses@sigma.iimas.unam.mx

The ISBA Program Council is pleased to report on the **ISBA@NIPS Travel Awards**. These special Travel Awards were created as part of an ISBA initiative aimed at highlighting the importance and impact of Bayesian methods in the new era of data science. As part of such an initiative four Bayesian workshops at NIPS were granted with the ISBA endorsement (see the previous number of this Bulletin). The ISBA@NIPS Travel Awards were especially designed to recognize the work of two young participants of ISBA-endorsed workshops at NIPS, one in the category of **Invited Paper** and one in the category of **Contributed Paper**.

After carefully reviewing all applications the ISBA Program Council designated David Knowles as the awardee in the category of Invited Paper (see report in the previous number of this Bulletin) and Dennis Prangle, Richard Rado postdoctoral fellow in Statistics at the University of Reading, as the winner in the Contributed Paper Category. Dr. Prangle has done important interdisciplinary contributions in the development of advanced approximate Bayesian computation (ABC) methods. The ISBA Program Council greatly acknowledges his work and congratulates him with this award, as part of the ISBA@NIPS initiative.

As it can be seen from the report by Tamara Broderick [see page 6], co-organizer of the ISBA endorsed workshop at NIPS –Advances in Variational Inference–, the endorsed workshops were a success within NIPS. ISBA is happy to see that these first actions of the ISBA@NIPS initiative have been of the interest of both of our academic communities.   ▲

WORLD MEETINGS 2016 AND 2018

- Michele Guindani -
on behalf of the Program Council of ISBA
micheleguindani@gmail.com

The ISBA World Meetings are the continuation of the traditional Valencia/ISBA Meetings held regularly since 1979. They represent a unique event where the Bayesian community gathers together to discuss recent advances and the future of our profession, at the same time looking back to our roots and traditions, following the footsteps of those who laid the foundations of where we are now. The ISBA World Meetings are an opportunity for young, senior and "giant" statisticians to meet and talk to each other in an environment that breaks up commonly held barriers and promotes the free flow of ideas, respectful confrontations, and ever-lasting collaborations. Someone in marketing would say that the ISBA meeting is the place to be with the people who matter! Surely enough, many of us think that the Bayesian community would not be the same today without first the Valencia meetings and now the ISBA World meetings. This is why the next two editions of the ISBA World Meetings will be held in locations where this continuous link between past and future will be most tangible.

The ISBA 2016 World Meeting will be held in Italy, from June 13th to June 17th, 2016. This choice will definitely present an opportunity to acknowledge the role of Bruno de Finetti in modern Bayesian thought - the 30th anniversary of his death will be widely celebrated in 2015 by the Italian mathematical, probability, actuarial and, of course, statistical communities. Thanks to de Finetti and his disciples, Bayesian ideas and methods have been lively encouraged and disseminated in Italy during the last century. The meeting will be held in the beautiful region of Sardinia, the second largest island in the Mediterranean Sea, and one of the most geologically ancient bodies of land in Europe. Due to its position at the heart of the Mediterranean Sea, Sardinia boasts old traditions and a wild and pure landscape. Hundreds of nuraghe (stone buildings) are scattered through the landscape, testifying to an ancient culture that endured there from the 16th to 15th Centuries B.C. The mountains are home to a wide variety of rare or uncommon animals and plants. The seaside landscapes are considered among the most beautiful in the world. The Local Organizing Committee of the conference is currently finalizing the negotiations with a leading conference venue and hotel seaside resort in the south of Sardinia, to secure the best prices and conditions for the attendees. The Local Organizing Committee is composed of Stefano Cabras (chair), Walter Racugno, Monica Musio, Brunero Liseo, Igor Pruenster, Antonio Lijoi and Laura Ventura. Plans are also well under way to define the scientific program of the meeting, which will conform to the usual high standards of the ISBA World meetings. For example, the traditional feature poster presentation will be the well-attended center point of the nights of Tuesday, Wednesday and Thursday nights. The scientific committee has been recently formed and includes ISBA members who are representative of the various "soulsänd the diverse geography of the society. Short courses are planned on June 12th on the premises of the nearby University of Cagliari. Updates on the program, registration fees, accommodation, and VISA requirements, will be provided soon, as usual, on the bayesian.org website, the ISBA mailing lists and on a dedicated website for the conference.

The ISBA 2018 World Meeting will be held on the premises of University of Edinburgh, Scotland, UK. This will represent a return to the origins for the Bayesian community, as the very same rev. Thomas Bayes was a student of the University of Edinburgh from 1719 to c. 1722. Edinburgh is the capital of Scotland, and is rich of historical and cultural attractions. For example, Edinburgh's Old Town and New Town are jointly listed as a UNESCO World Heritage Site. According to the plan of the local organizing committee, the conference should be held in central Edinburgh, at the Appleton Tower (and Informatics Forum) which are located at the central campus of the University of Edinburgh. The local organizing committee is composed of Chris Holmes (chair), Natalia Boch-

kina, Mark Girolami, Judith Rousseau and Chris Williams. Of course, more details will be provided as we approach the conference.

Finally, a special thanks to those members of our community who have contributed proposals to organize the World Meetings in the past, and an encouragement to contribute proposals for future ISBA meetings, according to the schedule and calls available on the ISBA website. It is thanks to the generosity of our members that ISBA is prospering as a community and as a society. ▲

BAYESIAN ANALYSIS - A MESSAGE FROM THE EDITOR

UPDATE FROM BA

- Marina Vannucci -
Editor-in-Chief
marina@rice.edu

The December issue of Bayesian Analysis features a discussion paper by Jesse Windle and Carlos M. Carvalho titled "A Tractable State-Space Model for Symmetric Positive-Definite Matrices". In this paper the authors present a state-space model whose observations and latent states take values on the manifold of symmetric positive-definite matrices and for which one may easily compute the posterior distribution of the latent states and the systems parameters as well as filtered distributions and one-step ahead predictions. Employing the model within the context of finance, they show how one can use realized covariance matrices as data to predict latent time-varying covariance matrices. This approach out-performs factor stochastic volatility. The manuscript is published together with three invited discussions, followed by a rejoinder. The issue also contains other fine articles on various topics of Bayesian statistics.

We remind readers that at BA it is now possible for individual authors to submit manuscripts for consideration as discussion papers. Such submissions will first go through our regular review process and, if accepted, the Editor in charge and the EiC will make a decision as to whether the manuscript can make a good discussion paper. If you wish to submit your work for consideration, please select "Article with Discussion" when submitting at EJMS. ▲

ISBA@NIPS WORKSHOP

NIPS 2014 WORKSHOP ON ADVANCES IN VARIATIONAL INFERENCE

- Tamara Broderick -
tab@stat.berkeley.edu

I am excited to be able to report on the great success of the ISBA@NIPS initiative this year as a co-organizer of one of the ISBA-endorsed workshops at the 2014 Neural Information Processing Systems conference. In particular, I was privileged to organize the NIPS 2014 Workshop on Advances in Variational Inference together with Shakir Mohamed, Charles Blundell, Matt Hoffman, David Blei, and Michael I. Jordan. We had an amazing turnout, with over 150 attendees packed in to a large room at the Palais des congrès de Montréal. Throughout most of the day-long workshop on Saturday, December 13, even the standing room in the back was full as we discussed challenges and advances in variational inference.

Variational inference has been enjoying a recent resurgence in interest. The ever-increasing size of data sets—e.g., in large-scale scientific and industrial applications—calls for a focus on scalability in Bayesian inference. Traditional schemes for posterior approximation such as Markov

chain Monte Carlo (MCMC) can be slow to run and difficult to evaluate in finite time. By contrast, variational inference allows for competitive run times and more reliable convergence diagnostics on large-scale and streaming data. The recent focus on such data sets has led to a booming interest in variational methods. But there remain significant trade-offs in speed, accuracy, simplicity, applicability, and learned-model complexity between variational inference and other approximative schemes such as MCMC and point estimation. Our workshop brought together researchers and practitioners to highlight recent advances in variational inference, to identify areas for improvement in current methodologies and software, and to discuss future directions and solutions to these issues.

Our workshop captured the groundswell of interest and current research in variational methods; the event included 6 invited talks, 4 contributed talks, 6 spotlight talks, and 34 total accepted posters. The presentations covered a great diversity of advances within variational inference, including but not limited to: the location of global vs. local optima in the variational optimization problem, non-conjugate models, nonparametric models, streaming data, accurate uncertainty estimation, stochastic optimization, theoretical guarantees, integration of variational methods with MCMC, and a variety of applications. We finished the day with a rousing panel discussion featuring a line-up of leading experts in the field: David Blei, Zoubin Ghahramani, Neil Lawrence, Shinichi Nakajima, and Matthias Seeger. The full workshop schedule, talk slides, and poster PDFs are available at our workshop website:

http://variationalinference.org.

Abbildung 1: The workshop in progress.

Abbildung 2: The workshop panel. Left to right: Shinichi Nakajima, Zoubin Ghahramani, David Blei, Neil Lawrence, Matthias Seeger.

We are indebted to ISBA not only for its endorsement and for forming the ISBA@NIPS initiative, but also for awarding the ISBA@NIPS Special Travel Award in the category of invited speaker to one of our speakers: David Knowles. We are further grateful to Google DeepMind and Adobe Research for additional workshop funding, to our large program committee for their thoughtful reviews, and to all of the workshop participants. ▲

OBITUARY: KATHRYN M. CHALONER (1954-2014)

The world lost a great Bayesian statistician and an amazingly gracious and wonderful person on October 19, 2014. Kathryn Chaloner died after a long and courageous fight against cancer. She was 60 years old and much, much too young to leave us. Kathryn is survived by her husband, Luke Tierney who is also a world-class statistician, and her two sons, Graham and Patrick.

At the time of her death Kathryn was Professor and Head of the Department of Biostatistics at the University of Iowa in the College of Public Health. The department was established in 1999 and she became its Head in 2002 when she and Luke moved to Iowa from Minnesota. She proceeded to build a first-class department working closely with Iowa's Department of Statistics and Actuarial Science.

Kathryn's PhD degree in 1982 was from Carnegie-Mellon where her advisor was Morris DeGroot. The focus of much of her research was Bayesian. She made important contributions in Bayesian inference, decision theory, and designing experiments, including clinical trials. But Kathryn was different from most Bayesians in that she applied her work and that of others to real problems, especially in research involving HIV and other infectious diseases. For example, she was a widely acknowledged expert in the elicitation of subjective probabilities. But as opposed to experts in how it should be done, she actually did it!

I was at the University of Minnesota in 1982 when Kathryn joined the faculty. We had similar backgrounds and interests. I had then and I continued to have the greatest respect for her thinking and her research. She was brilliant. But she was also nice to be around. She and I worked together in editing (with John Geweke) a festschrift honoring Arnold Zellner. This volume owes much to her good judgment. She had a marvelous intuition for what research is relevant in this world and what will wither on the vine.

When Telba Irony and I set up an annual conference on Bayesian Biostatistics in Houston we naturally asked Kathryn to be on the Program Committee. Again, we benefited and the conference benefited greatly from her good judgment.

The professional aspects of the Kathryn Chaloner story center on teachingÑor rather, on students. The distinction is important. She was a great mentor and role model. She respected students and they respected her. Her students attest to this with tributes to her on the internet. For example, https://faculty.biostat.ucla.edu/robweiss/taxonomy/term/105.

As a final acknowledgment of her great work with students, especially women students, she was awarded the 2014 Elizabeth L. Scott Award by the Committee of Presidents of Statistical Societies (COPSS) just two months before her death. Her citation reads:"For her commitment and success in developing programs to encourage and facilitate women to undertake careers in statistics; for extensive mentoring of women students and young faculty; for work to identify and remove inequities in employment for under-represented components of the profession; and for serving as a role model, balancing work and family while excelling as a teacher, researcher and academic administrator."

A memorial fund has been established in honor of Kathryn through the University of Iowa Foundation to support underrepresented minority students in the College of Public Health Department of Biostatistics. Contributions should be sent to: Kathryn Chaloner Student Scholarship Fund, UI Foundation, P.O. Box 4550, Iowa City, IA 52244-4550.

Donald Berry
Houston, Texas

INVITED CONTRIBUTION

BAYESIAN INFERENCE & BIG DATA: A SNAPSHOT FROM A WORKSHOP

- Max Welling, Yee Whye Teh, Christophe Andrieu, Jakub Kominiarczuk, Ted Meeds,

Babak Shahbaba, & Sebastian Vollmer -

welling.max@gmail.com

Bayesian Inference in the Age of Big Data

Over the last half century, and particularly since the advent of Markov Chain Monte Carlo methods, Bayesian inference has enjoyed tremendous successes, so much so that some quarters have proclaimed the age-old philosophical arguments between Frequentism and Bayesianism resolved, and that the twenty-first century is the Bayesian century. Whether such proclamations are controversial or not, the twenty first century has brought a new challenge to Bayesian inference that may render moot such philosophical arguments without a serious discussion of computation. This is the challenge of Big Data.

Indeed, the growth in volume and variety of data in the twenty first century has been staggering, and one could ask the question whether Bayesian inference is still a relevant statistical framework in this context. Does one still have to worry about model uncertainty and overfitting if there is so much data that uncertainties are negligible for all practical purposes? We believe there are a number of responses to this critique that rather point to the opposite direction.

Firstly, big data often means large p with $p \gg N$. For instance, modern MRI scanners record millions of voxels, possibly over time, for a handful of patients only. For another example, the cost for sequencing entire genomes is falling faster than Moore's law, implying that within a few years a doctor could routinely sequence every patient's genome: this is four billion basepairs per patient. Clearly, model uncertainty remains an important problem. Secondly, most of the data may not be relevant to certain model components. For instance, in recommendation systems with billions of user-product interactions, almost all of this data is irrelevant for any new user or product entering the system: this is the so-called cold start problem. Lastly, with the recent advances in deep learning, the number of parameters of modern neural network models have been growing at a faster pace than data volumes, with tens of billions parameters been reported by large internet firms such as Google, Yahoo!, Facebook and Baidu. Add to this the fact that the amount of "useful" or predictive information in data usually grows sublinearly in N. We thus observe that model capacity can in fact be growing faster than the amount of in-

formation we would like to store in it. That overfitting can still be a serious issue in this domain is evidenced by the fact that a new, rather extreme form of regularization (dropout) was invented to prevent it in neural networks [9]. One could argue that any model where regularization is not a consideration because the model's capacity is small compared to the data volume is in fact underfitting and therefore not optimally predictive (unless the data was generated from the model, which is never the case in reality).

What then are the challenges that a Bayesian approach faces in a Big Data context? Not surprisingly, a major challenge is computation. Traditionally two approaches exist to Bayesian inference: MCMC sampling and variational optimization. The former is asymptotically unbiased but often suffers from a significant variance in its estimates while the second is deterministic and has no variance but often exhibits rather strong biases. Until recently, standard coordinate ascent implementations of both methods were unsuited for very large datasets because every update (sampling or variational) depended on all data items[1].

So what criteria does an inference method need to satisfy in order to be suitable for Big Data? For cases where big data means large N, we propose the following desiderata: Every update needs to only access a small subset of the whole dataset. Moreover, the size of this subset should not depend on the size of the whole dataset, in other words, the algorithm should still make progress if the dataset is infinitely large (i.e. streaming data). Further, the algorithms should be amenable to execution on distributed systems using, say, the MapReduce framework.

Fortunately, important progress is being made on both desiderata. Stochastic variational updates were first reported for Latent Dirichlet Allocation in [12] which have since then been extended to a wide class of models ([6]; [13]; [20]; [19]; [11]; [10]; [24]). Similarly, the first stochastic subset MCMC algorithm was reported in [26], and since then also so extended in multiple directions ([1]; [14]; [4]; [5]). Good progress has also been made in terms of distributed implementations of both variational and MCMC methods ([21]; [25]; [18]; [3]; [2]).

---

[1] Efficient implementations of variational inference algorithms did however exist in the form of message passing algorithms ([27]; [17]) while efficient MCMC updates can also be achieved in some models by collapsing over the parameters ([7]). (We thank Tom Minka for pointing this out.)

First Workshop on Bayesian Inference with Big Data (BIBiD)

On August 28 and 29, 2014, the authors, participants in a NSF/EPSRC joint funded project, convened the First Workshop on Bayesian Inference with Big Data (BIBiD'14) at the University of Amsterdam to discuss these issues and to present the initial results of our current research.

Sebastian Vollmer presented a convergence proof for the stochastic gradient Langevin dynamics (SGLD) algorithm first presented in (26), along with showing that its rate of convergence is rather slow at $O(N^{-1/3})$ (23). New variants and alternative analyses were also discussed which shed light on the non-asymptotic behaviour of the algorithms.

Babak Shahbaba presented ideas for improving Big Data MCMC by exploiting the Riemannian geometry of the posterior. One idea is to split the problem into "easy" and "hard" parts, and using Hamiltonian Monte Carlo (HMC) with varying step-sizes to reduce computation on the "hard" parts (22). Another idea, for models over a constrained space, is to map the space onto a sphere and use Riemannian HMC on the sphere for efficient sampling (15).

Ted Meeds presented a more efficient ABC MCMC sampler for simulation based (likelihood free) models (16). In approximate Bayesian Computation (ABC) the typical scenario is that the likelihood can only be probed through a complex and expensive simulation. The proposed method learns a surrogate functions of the likelihood during runtime, significantly reducing the number necessary simulations.

Jakub Kominiarczuk discussed a Gibbs sampler that depends only on stochastically chosen mini-batches. This led to a challenging problem, due to the need to remove estimation bias from the naive estimates of conditional probabilities.

Yee Whye Teh addressed the problem of distributed MCMC sampling (28). While parallel schemes can be easily implemented, they do either not exchange valuable information among the chains (embarassingly parallel) or use too much bandwidth exchanging information. The proposed method involves each machine running a Markov chain, and asynchronously exchanging a limited number of sufficient statistics among the machines. The algorithm can be interpreted as expectation propagation where the moments are computed using MCMC sampling.

Finally, Max Welling discussed variational Bayesian autoencoders (13) (VBA). VBAs are similar to Helmholtz machines (8) in their architecture but use a more principled and scalable variational Bayesian learning algorithm to train both generative and discriminative networks.

In conclusion, the BIBiD'14 attendees agreed that Bayesian methods are both necessary and possible in the context of very large datasets. Research on the fertile interface among statistics, machine learning and computation are necessary to make Bayesian inference for Big Data a reality. Two main classes of Big Data Bayesian inference methods have emerged in recent years: low variance but biased variational methods and unbiased but high variance MCMC methods. However, the gap between these methods seems to be closing fast since the expressive power of variational methods is rapidly improving (reducing bias) while the efficiency of MCMC sampling is also rapidly increasing (reducing variance by the fact that more samples can be drawn). Of course, much work remains to be done in order to further improve the efficiency and accuracy of both variational methods and MCMC samplers.

We felt that the workshop has been a great venue for informally exchanging our ideas on Bayesian inference for Big Data. A future workshop welcoming a larger slice of the Bayesian community may help to crystallize further work on making the Bayesian framework relevant in the time of Big Data.

References

[1] S. Ahn, A. Korattikara, and M. Welling. Bayesian posterior sampling via stochastic gradient Fisher scoring. In *Proceedings of the International Conference on Machine Learning*, 2012.

[2] S. Ahn, B. Shahbaba, and M. Welling. Distributed stochastic gradient MCMC. In *Proceedings of the International Conference on Machine Learning*, 2014.

[3] E. Angelino, E. Kohler, A. Waterland, M. Seltzer, and R. P. Adams. Accelerating MCMC via parallel predictive prefetching. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2014.

[4] R. Bardenet, A. Doucet, and C. Holmes. Towards scaling up MCMC: An adaptive sub-

sampling approach. In *Proceedings of the International Conference on Machine Learning*, 2014.

[5] T. Chen, E. B. Fox, and C. Guestrin. Stochastic gradient Hamiltonian Monte Carlo. arxiv:1402.4102v1, 2014.

[6] A. Graves. Practical variational inference for neural networks. In *Advances in Neural Information Processing Systems*, pp. 2348-2356, 2011.

[7] T.L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America 101.Suppl. 1 5228-5235*, 2004.

[8] The wake-sleep algorithm for unsupervised Neural Networks. Science, 268, pp. 1158-1161, 1995.

[9] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.

[10] M. Hoffman. Stochastic Structured Mean-Field Variational Inference. arXiv preprint arXiv:1404.4114, 2014.

[11] M. Hoffman, D. Blei, J. Paisley, and C. Wang. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–1347, 2013.

[12] M. D. Hoffman, D. M. Blei, and F. Bach. Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems*, 2010.

[13] D.P. Kingma and Max Welling. Auto-encoding variational bayes. International Conference on Learning Representations, 2013. arXiv preprint arXiv:1312.6114.

[14] A. Korattikara, Y. Chen, and M. Welling. Austerity in MCMC land: Cutting the Metropolis-Hastings budget. In *Proceedings of the International Conference on Machine Learning*, 2014.

[15] S. Lan, B. Zhou, and B. Shahbaba. Spherical Hamiltonian Monte Carlo for constrained target distributions. In *Proceedings of the International Conference on Machine Learning*, 2014.

[16] T. Meeds and M. Welling. GPS-ABC: Gaussian process surrogate approximate Bayesian

computation. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2014.

[17] T. Minka. Expectation propagation for approximate Bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, 2001.

[18] W. Neiswanger, C. Wang, and E. P. Xing. Asymptotically exact, embarrassingly parallel MCMC. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2014.

[19] R. Ranganath, S. Gerrish, and D.M. Blei. Black box variational inference. arXiv preprint arXiv:1401.0118, 2013.

[20] T. Salimans and D.A. Knowles. Fixed-form variational posterior approximation through stochastic linear regression. Bayesian Analysis 8, 837-882, 2013.

[21] S. L. Scott, A. W. Blocker, F. V. Bonassi, H. Chipman, E. George, and R. McCulloch. Bayes and big data: the consensus Monte Carlo algorithm. In *Bayes 250*, 2013.

[22] B. Shahbaba, S. Lan, W. O. Johnson, and R. M. Neal. Split Hamiltonian Monte Carlo. *Statistics and Computing*, 2014.

[23] Y. W. Teh, A. H. Thiéry, and S. J. Vollmer. Consistency and fluctuations for stochastic gradient Langevin dynamics. *submitted*, 2014.

[24] M. Titsias and M. LÃizaro-Gredilla. Doubly Stochastic Variational Bayes for non-Conjugate Inference. Proceedings of the 31st International Conference on Machine Learning (ICML-14), 2014.

[25] X. Wang and D. B. Dunson. Parallelizing MCMC via Weierstrass sampler, 2013.

[26] M. Welling and Y. W. Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the International Conference on Machine Learning*, 2011.

[27] J.M. Winn and C.M. Bishop. Variational message passing. *Journal of Machine Learning Research*, 2005.

[28] M. Xu, Y. W. Teh, J. Zhu, and B. Zhang. Distributed context-aware bayesian posterior sampling via expectation propagation. In *Advances in Neural Information Processing Systems*, 2014. ▲

STUDENTS' CORNER

Isadora Antoniano

isadora.antoniano@unibocconi.it

The Holiday season is upon us and, whatever else that may mean for you (haters and lovers alike), I hope it provides a peaceful period to catch up on your reading, for example. And if you're running short on reading material, this issue of the Students' Corner may be just what you need.

We open with a kind contribution from a former Savage Award winner, Julien Cornebise, and his views four years after. I won't say more, because I find his piece quite speaks for itself. Thank you, Julien, I would have loved to read something like this when I weas a Ph.D. student and I'm sure some of our readers will appreciate it.

The second piece for this number's Corner tells us about blogging. The ubiquitous nature of the activity makes for a perfect opportunity to involve different people from many places and so this is a joint collaboration from an interesting group of bloggers: Rasmus Bååth (Cognitive Science, Lund University, Sweden), Boris Hejblum (Ph.D student at ISPED; INSERM, Centre INSERM U897, University of Bordeaux, and INRIA SISTM Team, F-33000 Bordeaux, France), Thiago G. Martins (Data Scientist at Yahoo!), Julyan Arbel (postdoc at Collegio Carlo Alberto, Moncalieri, Italy), Pierre Jacob (postdoc at the University of Oxford, UK), Robin Ryder (Université Paris-Dauphine, France), Ewan Cameron (Computational Statistician in the Spatial Ecology and Epidemiology Group, University of Oxford, UK) and Gregory Gandenberger (Department of History and Philosophy of Science, University of Pittsburgh, USA). My special thanks to my dear friend Pierre for putting the tea together. Rather diverse, wouldn't you agree? Don't forget to check out their bloggs, you might find something you like.

Finally, an invitation, sent to us by Francesca Ieva, of a series of statistical events, specially designed for the "young" and taking place in Italy!

Don't forget, this is your corner, suggestions, announcements, questions and contributions are always welcome.

Happy read and happy Holidays!

On the Savage award, advices to Ph.D. candidates

by Julien Cornebise

http://www.cornebise.com/julien/

This article is an occasion to pay forward ever so slightly, by encouraging current Ph.D. candidates on their path, the support ISBA gave me. Four years ago, I was honored and humbled to receive the ISBA 2010 Savage Award, category Theory and Methods, for my Ph.D. dissertation defended in 2009. Looking back, I can now testify how much this brought to me both inside and outside of Academia.

Inside Academia: confirming and mitigating the widely-shared post-graduate's impostor syndrome

Upon hearing of the great news, a brilliant multi-awarded senior researcher in my lab very kindly wrote to me that such awards meant never having to prove one's worth again. Although genuinely touched by her congratulations, being far less accomplished and more junior than her, I felt all the more responsible to prove myself worth of this show of confidence from ISBA. It would be rather awkward to receive such an award only to fail miserably shortly after.

This resonated deeply with the shared secret of recent PhDs, discovered during my year at SAMSI, a vibrant institution where half a dozen new postdocs arrive each year: each and every one of us, fresh Ph.D.s from some of the best institutions (Cambridge, Duke, Waterloo, Paris...) secretly suffered the very same impostor syndrome. We were looking at each other's CV/website and thinking "jeez! this guy/girl across the door is an expert of his field, look at all he/she has done, whereas I just barely scrape by on my own research!" –all the while putting up a convincing facade of self-assurance in front of audiences and whiteboards, to the point of apparent cockiness. Only after candid exchanges in SAMSI's very open environment did we all discover being in the very same mindset.

In hindsight the explanation is simple: each young researcher in his/her own domain has the very expertise to measure how much he/she still

does not know and has yet to learn, while he/she hears other young researchers, experts in their own other field, present results not as familiar to him/her, thus sounding so much more advanced. This take-away from SAMSI was perfectly confirmed by theis Savage Award: yes, maybe indeed, I, just like my other colleagues, might actually know something relatively valuable, and my scraping by might just be not so bad – as is also the case for so many of my young colleagues.

Of course, impostor syndrome is a clingy beast and, healthily, I hope to never get entirely over it – merely overcoming it enough to say "Do not worry, thee young candidate, thy doubts pave a path well trodden".

A similar message is also part of the little-known yet gem of a guide, How to do Research at MIT AI Lab – Emotional Factors, relevant far beyond its original lab. I recommend it to any Ph.D. student; the feedback from readers is unanimous.

Outside Academia: incredibly increased readability

After two post-docs, and curious to see what was out there in atypical paths, I took a turn out of purely academic research, first as an independent consultant, then recruited out of the blue by a start-up's recruiter, and eventually doing my small share to help convince investors. I discovered there another facet of ISBA's Savage Award: tremendous readability.

In Academia, the dominating metric of quality is the length of the publication list – a debate for another day. Outside of Academia, however, not all interlocutors know how remarkable is a JRSSB Read Paper, or an oral presentation at NIPS, or a publication in Nature.

This is where international learned societies, like ISBA, come into play: the awards they bestow can serve as headline-grabbing material in a biography, easily spotted. The interlocutors do not need to be familiar with the subtleties of Bayesian Analysis. All they see is a stamp of approval from an official association of this researcher's peers. That, in itself, is enough of a quality metric to pass the first round of contact, raise interest, and get the chance to further the conversation.

First concrete example: the recruiter who contacted me for the start-up I joined in 2011 was tasked to find profiles for an Applied position. The Savage Award on the CV grabbed his attention, even though he had no inkling what Adaptive Sequential Monte Carlo Methods were, nor if they were immediately relevant to the start-up. Passing it to the start-up's managers, they immediately changed focus and interviewed me for their Research track instead: a profile that was not what they were looking for originally, yet stood out enough to interest them for a position they had not thought of filling via a recruiter – and indeed a unique position that I would never have thought to find this way either!

Second concrete example, years later, hard at work in this start-up's amazing team: investors were coming for a round of technical due diligence. Venture capitals sent their best scientists-in-residence to dive deeply into the technical details of our research. Of course what matters in the end is, and forever will be, the work that is done and presented. Yet, the Savage Award was mentioned in the first line of the biography that was sent ahead of time, as a salient point to give a strong first impression of our research team.

Advices to Ph.D. Candidates: apply, you are the world best expert on your topic

That may sound trivial, but the first advice: apply. Discuss with your advisor the possibility to put your dissertation up for consideration. This might sound obvious to North-American students, whose educative system is rife with awards for high-performing students. Not so much in France, where those would be at odds with the sometimes over-present culture of égalité in the younger-age public education system. As a cultural consequence, few French Ph.D. students, even the most brilliant, would consider putting up their dissertation for consideration. I have been very lucky in that regard to benefit from the advice of a long-term Bayesian, who offered to send it for me – thanks again Xi'an! Not all students, regardless how brilliant their work, are made aware of this possibility.

The second advice, closely linked: do not underestimate the quality of your work. *You are the foremost expert in the entire world on your Ph.D. topic*. As discussed above, it is all too easy to see how advanced are the maths wielded by your office-mate, yet oversee the as-much-advanced maths you are juggling on a day-to-day basis, more familiar to you, and whose limitations you know better than anyone else. Actually, knowing these very limitations is what proves you are an expert.

A word of thanks and final advice

Finally, a word of thanks. I have been incredibly

lucky, throughout my career so far, to meet great people. My dissertation already had four pages of acknowledgements: I doubt the Bulletin's editor would appreciate me renewing (and extending!) them here. They are just as heartfelt today as they were then. I must, of course, add ISBA and the Savage Award committee for their support, as well as all those who, by their generous donations, allow the Savage Fund to stay alive throughout the years.

Of interest to Ph.D. candidates, though, one special mention of a dual tutelage system, that I have seen successfully at work many times. The most senior, a professor with the deep knowledge necessary to steer the project brings his endless fonts of knowledge collected over decades, wrapped in hardened tough-love. The youngest, a postdoc or fresh assistant professor, brings virtuosity, emulation and day-to-day patience. In my case they were Pr. Eric Moulines and Dr. Jimmy Olsson. That might be the final advice to a student: if you ever stumble, as many do, as I most surely did, because Ph.D. studies can be a hell of a roller-coaster to go through, reach out to the people around you and the joint set of skills they want to offer you. In combination, they can be amazing, and help you open doors that, in retrospect, can be worth all the efforts.

Blogging as young researchers

by Rasmus Bååth from Publishable Stuff; Boris Hejblum from Research side effects; Thiago G. Martins from tgmstat@wordpress; Julyan Arbel, Pierre Jacob and Robin Ryder from Statisfaction; Ewan Cameron from Another Astrostatistics Blog; Gregory Gandenberger from gandenberger.org

Inspired by established blogs, such as the popular Statistical Modeling, Causal Inference, and Social Science or Xi'an's Og, each of us began blogging as a way to diarize our learning adventures, to share bits of R code or LaTeX tips, and to advertise our own papers and projects. Along the way we've come to a new appreciation of the world of academic blogging: a never-ending international seminar, attended by renowned scientists and anonymous users alike. Here we share our experiences by weighing the pros and cons of blogging from the point of view of young researchers.

```
[To blog...]
```

At least at face value blogging has some notable advantages over traditional academic communication: publication is instantaneous and thus proves efficient in sparking discussions and debates; it allows all sorts of technological sorcery (hyperlinks, animations, applications), while many journals are still adapting to grayscale plots; and it allows for humorous and colourful writing styles, freeing the writer from the constraints of the impersonal academic prose. Last but not least, it is acceptable to blog about almost any topic, from office politics to funding bodies, from complaints about the absurdity of p-values to debates on the net profits of publishing companies, not to mention quarrels about the term "data science".

For young researchers, some aspects are particularly appealing. By putting academics directly in touch with one another through comments and replies, young researchers are given the opportunity to "talk" directly on technical subjects to some of the most renowned experts in their fields–and indeed a surprising number of senior researchers are avid blog readers! This often proves much more efficient than trying to awkwardly stalk the same professors at conferences. Through such interactions, young academics can show off their many interests and skills, which can do much to fill out the picture painted by their academic CV.

Beyond those low and careerist considerations, we see blogging as a good tool to learn and to share scientific ideas. According to popular belief, only a third of all started research projects end up in a publication; but all of them can at least end up on a blog. So if you indulge in a bit of off-topic study or burn a few hours playing around with a new methodology it need not fuel your performance anxiety: a blog post explaining it will still feel like a delivered product. And you will very likely get some interesting feedback–though rarely to the depth given in journal reviews.

Finally, using blogs to advertise articles and packages seems particularly useful at the early stage of a career, where you might not be invited to that many conferences, or might only be given some dark corner of a giant poster session to talk about your work.

```
[... or not to blog ?]
```
Some cautionary notes now, blogging can be risky! As the adage goes, "better to keep your mouth shut and appear a fool than to open it and remove all doubt". Beyond the quality of the content being shared, blogs are also sometimes disre-

garded by academics as a frivolous medium; there is a risk that your colleagues will see your blogging hobby as a pure waste of time.

A second risk is to disclose too much information about promising research leads. There should be some balance between ideas shared and ideas kept secret, so that blogging does not jeopardize publication. Other platforms that formally establish precedence (such as arXiv) might be better suited for the initial presentation of new and exciting work. For this reason it seems wisest to blog a posteriori, though the interest of these blogs will be less than their potential to function as real time research diaries.

A third risk is genuine time-wasting. For those who have never tried, it can be surprising to discover how many hours are needed to write each post. It can be frustrating in the beginning when reader statistics indicate an audience of just one or two spam-bots and some curious relatives. On the other hand there are still a limited number of academic blogs on statistics so far, so the market is far from saturation: any new blog can quickly garner a decent amount of attention. Of course it can be hard to keep a regular posting schedule, which is necessary to maintain a stable reading base.

To conclude, blogging can be a clever way to bypass the hierarchical structure of academia. It gives everyone a direct and fast access to everyone else. In some respects it helps to alleviate key problems affecting young researchers, such as the lengthy reviewing process of top journals and the lack of communication space.

Invitation:
StaTalk on Biostatistics

Department of Mathematics "F. Enrique",

Università degli Studi di Milano, via Saldini 50, Milano (Italy)

We would like to announce the first StaTalk promoted by the young section of Italian Statistical Society (y-SIS), to be held in Milan, on April 24th, 2015.

All the information, including the tentative program and the sign-up link for the conference, can be found on the conference website.

StaTalks are a series of one-day workshops aimed at early career statisticians, i.e., master degree students, people carrying out a PhD, postdocs and young researchers. StaTalks are a fantastic opportunity to share knowledge on advanced topics of statistics in a relaxed and friendly atmosphere. For each talk, a hosting research group will propose tutorials, seminars and interactive round tables related with its main fields of research and application. Participants can contribute bringing their own experience through presentations, ideas, posters and material to be shared during the debate. A StaTalk will be held every 6 months in a different city in Italy. StaTalks on Bayesian Nonparametrics and on Computational Tools for Statistical Modelling are already planned in Turin (autumn 2015) and Rome (spring 2016), respectively.

Registrations to the first StaTalk are now open. No conference fees are due, but registration is mandatory, using the form available on the website.

For further information, doubts or questions, please visit the event website and do not hesitate to contact the y-SIS officers: Francesca Ieva, Serena Arima, Antonio Canale, Clara Grazian and Mariangela Sciandra, at the address e-mail youngsis.2013@gmail.com.

NEWS FROM THE WORLD

Announcements

BNP10: BNP10 contributed talks and posters deadline extended

The 10th conference on Bayesian Nonparametrics (BNP10) will be held in June 22-26, 2015 in Raleigh, NC. BNP10 will feature four special invited lectures, eight invited sessions, contributed talks and poster sessions. A proposal for a contributed talk or poster can be made through the conference website http://www.stat.duke.edu/bnp10/ and should consist of a title and an abstract. Slots for contributed talks are limited.

Deadline for submission of a contributed talk has been extended to January 15, 2015, and decisions about acceptance for contributed talk is expected by February 15, 2015. Deadline for poster submission is May 1, 2015.

If you have any questions, please contact the BNP10 Local Organizing Committee at 10thbnpmeeting@gmail.com.

Member News

Mike West is recipient of award from ASA NC Chapter

Sunday, October 12, 2014 - 4:30pm

Mike West, the Duke University distinguished Arts & Sciences Professor of Statistics and Decision Sciences, has been honored with a distinguished award from the NC Chapter of the American Statistical Association. The award was announced and presented at the October 10-12, 2014 International Conference on Advances in Interdisciplinary Statistics and Combinatorics. West, a plenary speaker at the conference, was recognized for his outstanding contributions to the theory and practice of statistics, and particularly for his role in the development of the subject in North Carolina.

Jerry Reiter, current president of the NC Chapter of the American Statistical Association, announced and presented the award to Mike West at the conference award ceremony.

▲