re, please contact Michele Guindani, Program Chair of the BNP section of ISBA, at michele.guindani@me.com.

- The June issue of Bayesian Analysis will feature a captivating paper by Peter Mueller and Riten Mitra, titled "*Bayesian Nonparametric Inference - Why and How*". The editor in chief of Bayesian Analysis has organized invited discussions by three masterminds of anything Bayes, as well as a set of contributed discussions by other many prominent authors.

- Last but not least, Judith Rousseau has organized a one-day satellite workshop to MCMCski, on "Bayesian nonparametrics, modelling and computations (BNPski)". The workshop will be held in Chamonix on January 9th, 2014 and is free for any person registered at MCMCski. For more information, see the MCMCski website (`http://www.pages.drexel.edu/~mwl25/mcmski`) , or contact directly Judith at rousseau@ceremade.dauphine.fr.

For any more information on BNP related events (or propose your own), stay tuned on our Section website at `http://bayesian.org/sections/BNP`. ▲

INDUSTRIAL STATISTICS SECTION

- Refik Soyer -
Section Chair
`soyer@gwu.edu`

Upcoming ISBA IS (Industrial Statistics) activities:

- An ISBA/IS session is organized at the ISBA South Africa Chapter meeting in Grahamstown during June 26-28, 2013.

- ISBA/IS sponsored session on "Advances in Bayesian Reliability Analysisïs organized at the, Mathematical Methods in Reliability Conference at Stellenbosch, South Africa during July 1-4, 2013.

- The ISBA/IS executive committee is actively involved in organization of the Third Symposium on Games and Decisions in Reliability and Risk in Kinsale, County Cork, Ireland during July 8th - 10th, 2013. There will be an ISBA/IS sponsored section at the symposium.

- At the INFORMS 2013 annual conference in Minneapolis, October 6-9, 2013, ISBA-IS is organizing a session titled on"Bayesian Methods for Industrial Statistics". The session is cosponsored by ISBA/IS and Quality, Statistics and Reliability (QSR) section of INFORMS.

- An ISBA/IS sponsored invited section is organized at the European Network Business and Industrial Statistics (ENBIS) meeting in Ankara, Turkey, September 15-19, 2013. ▲

INVITED CONTRIBUTION

IN PRAISE OF THE REFEREE

- Nicolas Chopin, Andrew Gelman, Kerrie L. Mengersen & Christian P. Robert -
`nicolas.chopin@ensae.fr`
`gelman@stat.columbia.edu`
`k.mengersen@qut.edu.au`
`xian@ceremade.dauphine.fr`

A peer-review system in flux

Scientific and scholarly publishing has for years been centred on peer-reviewed journals, where the authors of published articles are responsible for their correctness, while editors and referees vouch for this correctness to some extent, but mostly for the novelty and importance of the work.

Widely-acknowledged problems with the current refereeing system include inefficiency for authors (e.g., waiting time for referees, referee reports of different quality), waste of reviewers' efforts (e.g., resubmissions of papers to other journals without cross-reference to previous reviews, long referee reports that are read by at most two people—the author and the journal editor), a proliferation of journals (so that it is no longer sufficient for scholars to keep up with a field by reading a few top journals), and, most importantly, a profusion of unreplicated or unreplicable claims even in the highest-prestige outlets.

For instance, Wasserman's (2012) remarks that "we are using a refereeing system that is almost 350 years old. If we used the same printing methods as we did in 1665 it would be considered laughable." He describes the refereeing process as"'noisy, time consuming and arbitrary," that it "limits dissemination" and that provides an "illusion" of quality control. He likens the process to a "priesthood" or "guild" and advocates its replacement by a "marketplace" of ideas.

Proposals for reform typically vary among the following options: (1) replacing the formal referee process with a communal process by bypassing the journals altogether and posting articles freely on the web, (2) formalising a post-publication peer-review process so that referee reports are open and available for all to read, and (3) putting more of the burden of proof of replicability on published work by requiring data-based articles to come with full replication materials.

Each of these steps has been taken already, to some extent. Personal websites and servers such as arXiv (physics and mathematics) and SSRN (social science) are widely used for posting unreviewed preprints. While arXiv or SSRN is not completely open, it is not difficult for a researcher to establish the connections necessary to post there. Post-publication peer review exists in some journals and, more effectively, in an informal network of scientific blogs. The goal of ensuring replicability is tougher, but some journals (for example, the *Quarterly Journal of Political Science*) do require a full suite of replication materials before allowing any empirical article to be published.

Thus, proposed reforms typically involve taking some aspect of the current system and pushing them further. Here are three illustrations:

1. Theoretical statistician Larry Wasserman (2012) calls for "a world without referees":

> "We should be disseminating our research as widely as possible. Instead, we let two or three referees stand in between our work and the rest of our field (...) We should think about our field like a marketplace of ideas. Everyone should be free to put their ideas out there. There is no need for referees. Good ideas will get recognised, used and cited. Bad ideas will be ignored."

2. Cognitive psychologist Nikolaus Kriegeskorte (2009, 2011) proposes "open post-publication peer-review":

> "Any scientist can instantly publish a peer review on any published paper. The scientist will submit the review to a public repository (...) The repository will link each paper to all its reviews, such that that readers are automatically presented with the evaluative meta-information. In addition, the repository allows anyone to rank papers according to a personal objective function computed on the basis of the public reviews and their numerical quality ratings."

3. Political scientist Brendan Nyhan (2012), following ideas that have become popular in medical research, recommends that data-collection protocols be published ahead of time, with the commitment to publish the eventual results:

> "In the case of experimental data, a better practice would be for journals to accept articles before the study was conducted. The article should be written up to the point of the results section, which would then be populated using a pre-specified analysis plan submitted by the author. The journal would then allow for post-hoc analysis and interpretation by the author that would be (...) distinguished from the previously submitted material. By offering such an option, journals would create a positive incentive for preregistration that would avoid file drawer bias."

All three of these proposals are appealing, compelling, and radical—and go in different directions, with the statistician wanting to eliminate referees, the psychologist recommending reviews but in a different structure, and the political scientist proposing a more stringent system of pre-publication quality control.

Our goal is not to evaluate these particular proposals but rather (a) to consider the relevance of these ideas for the field and (b) to emphasise the value of the referee system and to focus attention on how to not lose its benefits in this time of change. As statisticians, it would be most appropriate for us to evaluate reform proposals by analysing existing data, gathering new information, or at the very least proposing a plan for sampling, measurement, and causal inference. Unfortunately, as in much of our professional lives, we do not practice what we preach.

This note attempts to find a middle ground between what we have now and various proposed reforms. In our opinion, the debate is as much about ethics as it is about science, namely how to work out a system of dissemination in which papers are evaluated on the basis of their scientific worth, rather than on the paper's conformity with existing norms (a problem with the traditional system of peer review), its potential popular impact (an issue with proposed open alternatives), the author's reputation or networks, or the reviewer's own long-term plans. Based on our own experiences, we argue that in this era of data explosion, the referee system remains preferable to the frightening morass of an uncontrolled accumulation of self-published documents.

Background

Each field brings its own perspective on publishing. For a mathematician or theoretical statistician such as Wasserman, what is important in a publication is the idea. Mathematical ideas can be evaluated openly and, in principle, by anyone. From the other direction, Nyhan focuses on the *difficulty* of replicating empirical results, especially given the selection problem that positive rather then negative findings tend to get published. As applied statisticians, we see the merits of both approaches, depending on our focus.

At the same time that mathematicians are moving to deregulate academic publications, many experimental scientists are pushing toward more formal registries. Beyond their direct benefit in replicability, such reforms involve incentives for better behaviour of researchers. If you know ahead of time that you will have to supply details of your design, methods, data and computer code, you will be motivated to keep better records and clearer codes from the start, which in turn leads to a positive feedback in which later analyses are improved by iterating on existing material, as argued by O'Rourke and Detsky (1989).

Publication patterns also vary among academic fields. Some of the best mathematicians and economists work alone or in small collaborations and publish papers after they have been honed by workshop and seminar presentations, while, at the other extreme, leading physicists, biologists, and electrical engineers supervise laboratories producing dozens of publications a year. In the first case, one could argue against an extra refereeing stage; however biases in the workshop process also need to be ironed out by this anonymous refereeing step.

We should ask the same of sport and of scientific referees: assurance of quality—in terms of the merit, originality and substantive contribution of the scientific content; fairness—in terms of equitable treatment for all authors; consistency—in terms of reasonable, useful feedback to authors; and timeliness—a fast turnaround of reviews. These are the very qualities that Wasserman laments are lacking in the current process.

Horror stories with happy endings

The previous section seems to proceed along the line that refereeing is a necessary evil. We believe on the contrary that it is a necessary good. Yes, certain referees are annoying, or even aggressive or too dismissive about one's work. Of course, like others, we can tell horror stories about referees completely missing the point or even being outright dishonest. As authors of many peer-reviewed publications, we have however benefited immensely from the unpaid labour of referees (and repaid this by serving as referees, associate editors, and sometimes editors).

At times, we've been annoyed at having to jump through hoops but more often than not the suggestions are helpful. For example, Gelman's (2006) most successful article of the past decade was his paper on prior distributions for hierarchical variance parameters. Originally an example in (*Bayesian Data Analysis*), it was solicited as an article by the editor of *Bayesian Analysis*. The referees were brutal and the paper could only be published in the journal as discussion of another accepted article. However, as a consequence of this revision process, Gelman was motivated to add a whole new section that made the research much more general and interesting. It is thanks to the referees that the author put in the work to make the paper what it was.

Another extreme example experienced by Chopin is that of a referee who was adamant about rejection on grounds that the authors believed unreasonable, but who in the third revision exposed a mistake in a sampling algorithm. Since publishing a wrong paper is much more damaging in the long run than being rejected by a given journal, this turned out to be most useful.

At the other end of the spectrum are sloppy referees who form a strong opinion based on a cursory read, along with their particular priors about the topic in question. The result, especially for competitive journals, is often a rejection based on unconstructive comments, which also contributes to an incentive structure that favours incremental and conventional work. Alternatively, an "accept" decision based on shallow refereeing can allow a poor paper to appear. Often, however, the system corrects itself, the discrepancy with the other reports or the lack of substance in the review being spotted by associate editors or editors.

We also believe that our papers are preemptively improved by refereeing, in that we mostly write better papers because we know they will be critically evaluated by colleagues prior to publication. We go the extra mile, chase typos, think more carefully about real examples, and so on, before submitting, because we do not want to give a negative referee this additional and objective leverage we can ourselves perceive.

Wheat from chaff

While scientific review processes have been evolving forever, the current paradigm is that editors send submitted manuscripts to selected reviewers for comments, and then make a decision based on these comments and their own judgement. The issues of concern in such a simple system arise from the arbitrary and often narrow selection of reviewers, the generous, even unreasonable time allowed for response, the mostly unhelpful guidelines for comments, the opaque manner in which the final decision occurs, and the huge and often wasted investment in time by all actors. In particular, junior scholars can take their refereeing duties very seriously, writing long and careful reports even on papers that are not worth the effort.

We agree with Kriegeskorte and Deca (2012) that a better use of reviewers' time and effort would be to have many reviews of important papers and only zero or one review for the sorts of minor contributions that fill up our journals. Conversely, a very specialised result can sometimes be useful; in this case it might well merit a post-publication review thread by its user community, in a *Tripadvisor* manner.

When faced with these issues, some journals have evolved from the traditional model. For example, some have databases of reviewers from which to more objectively draw subject-specific referees; others demand short review times; others have formalised the referee process by instituting a detailed checklist or providing careful guidelines about the type of review required; and a small number have adopted the post-modern (or pre-traditional) practice of an editorial board making decisions at regular team meetings.

A strong argument against doing away with referees is the problem of sifting through the chaff. The daily volume of published research documents is overwhelming and accelerating, perhaps not so much in statistics but certainly in biomedical research and engineering. There is a maximum amount of time one can dedicate to looking at websites, blogs, twitter accounts, and such. And blog comments have certainly not delivered the post-publication quality control some had hoped. Commenting on a blog is not a well-respected use of time, while commenting on a busy blog might not get noticed amidst all the chaff. Right now there seem to be very few blogs providing a useful communal review function (and none of these focus on statistics).

Even keeping track of new arXiv postings may gets overwhelming. Wasserman writes, "if you don't check arXiv for new papers every day, then you are really missing out," but our own experience is that it is almost impossible *not* to miss out. Checking arxiv.org/list/stat/new indeed takes less than a minute, checking potentially interesting papers takes much longer!

Without an organised system of reviews, why should anyone bother to comment on poor or wrong, but not newsworthy, papers? The result could well be a clutter of mediocre and uncommented results making it difficult for researchers who are not well-connected to navigate the field. We, the authors of the present article, know enough experts in our research areas that we can often get a quick evaluation of unpublished work. But a student whose advisor is not an expert on statistical computation or a researcher in biology (say) who wants to use the latest computational methods, will not generally have the resources provided by our social network. The review process does not completely level the playing field—nothing could, given institutional disparities of resources—but it comes closer to equalising the information available to differently-equipped teams.

Given the amount of chaff and the connected tendency to choke on it, filtering will be done—somehow or another. Getting rid of referees and journals in favour of repositories like (the great) arXiv would force us to rely on other and less well-defined sources for ranking, selecting, and eliminating papers. Again this would be subject to arbitrariness, subjectivity, bias, variation, randomness, peer pressure, and so on. In addition, having no prior quality control makes reading a new paper a tremendous chore as one would have to check the references as well, leading to a type of infinite regress, or forcing one to rely on reputation and peer opinions.

In fact, one may wonder if it is really possible to go that far in reducing the impact of peer reviewing. For many of us, so much depends on our publication record (including jobs, promotions, grants, and eventually salaries) that very few would be bold enough to stop sending papers to peer-reviewed journals from their own initiative. Getting rid of peer-reviewed publications would make sense only if the vast majority of scientists in a given field would agree to do all at once. And, since it is not only individuals but also scientific fields that compete for grant money, one could argue that a simultaneous move from *all fields* would be required to ditch peer reviewing, which is of course even less likely.

Thus, despite the appeal of chucking the journals and starting over, we think an uncontrolled system would be even more unethical than what we currently have and may be exactly what we would like to avoid. If our profession did start from scratch, that new institutions would certainly arise to serve the filtering and reviewing functions, but we would prefer to see a smooth switch. In the next sections, we make two proposals that constitute a middle ground between *statu quo* now and Wasserman's suggestion. The first is a further evolutionary step in the review process, while the second is more radical.

Proposal 1: Post-publication peer review

In a world where (nearly) everything is published, how can the scientific community sift through the mass of results? It should be possible to use more efficiently the effort that is currently going into peer review. While writing dozens of careful referee reports per year, we realize the futility of creating mini-articles for such a tiny audience (the author and the journal editor). It makes much more sense to switch to blogging about important papers, as to reach a much wider audience. And to keep reviews short and to the point (and available to the readers of the article in question at some point; see below). This notion is met with reluctance by many, for whom the secrecy of the reviewing process and the anonymity of the reviewer appear like sacrosanct principles.

Post-publication peer review could be done in different ways, most simply by adding a comment thread to each arXiv article (with the caveat of being possibly unread), but more formal approaches are possible. Kriegeskorte (2009, 2011) recommends "peer-to-peer editing: authors ask a senior scientist to edit the paper; editor chooses 3 reviewers and asks them to openly review the paper; editor is named on the paper."

Another, perhaps complementary, approach would be for groups of scholars and academic societies to manage a filtering service. For example, instead of the ASA running *JASA*, *JABES*, *JCGS*, etc., and maintaining a separate editorial staff for each of those journals (representing a huge amount of possibly overlapping and hence redundant volunteer service), it could support filtering services. The editors of each filter would be expected to scan the literature and handle submissions (which in this case would point to articles already published on the web). Editorial boards would have the responsibility to come up with monthly (say) recommended reading material. This would require some work, but less than the existing job of producing a journal. The main concern we see would be to keep the editors focused on solid research rather than getting tabloid-like, but the latter seems less likely if the process involves simply flagging articles rather than formally and exclusively publishing them. The flagging could even be multidimensional, with some papers tagged as potentially exciting but speculative, and others labelled as solid contributions within an existing paradigm.

Instead of a simple thumbs-up or down, reviewers would have the task of situating each new paper within the literature. As journal editors and frequent referees ourselves, we would appreciate the opportunity to prepare reviews that are directed outward to the potential users of the published articles rather than inward.

We suspect that a key step in getting post-publication peer review to work is to transfer the efforts that *would have gone into refereeing* into filtering. It would be difficult to start up a filter all on its own without the free labour that comes from referees (who are in turn motivated by a sense of obligation and scientific community). Time being a limited resource, we foresee a challenge in instilling the same sense of duty for filtering and post-publication review as is now present in the journal review process.

Proposal 2: A reviewer commons

Just as it is useful to ask why sport referees do not always get it right, we could ask the same here. What is broken in our system? There is a constant proliferation of new arenas of training and competition and an exponential growth in the community of participants, which has great potential benefits for science but is daunting for reviewers. However, instead of the open scrutiny to which we subject our sports referees, scientific reviews are conducted behind closed doors. Perhaps it is time we came out.

We are thus suggesting a dramatic move in the creation of reviewer commons, namely a (virtual) repository for the placement of scientific reviews, open to all. The advantages of such a commons are many. It would encourage high quality, fair and useful reviews. It would facilitate acknowledgement of reviewer contributions, benefiting both the journals and authors (since reviews could be referenced in the manuscript) and the reviewers (since reviews could be accessed by peers). Reviewers would then write not only for the authors but also for the readers, turning their comments and suggestions into a valuable discussion at the end of the reviewing process, to be added to their publication list as well. Furthermore, as well as improving quality, this notion of a commons might also help to reduce the workload of reviewers and editors. For example, until the current practice of not requiring authors to declare prior submissions of articles is revised, access to previous reviews might help to mitigate replication of effort by reviewers in dealing with manuscripts doing the rounds of journals.

We are not the first to argue that revealing the names of referees, not only to the authors, but also to the public, would deter referees from being complacent or un-constructively negative. Indeed, it may bring more explicit recognition in the scientific sense to referees and to their role in publishing better research, possibly all the way to referees' reports becoming a valued part of their own publication record, as is already the case for referees for *Hydrology and Earth System Sciences*.

A related concern is the increasing focus of some journals on headline-grabbing articles. This can lead to evaluating articles on the basis of their popularity rather than their science. As discussed above, this is against the principles that we laid down for good refereeing practice. Psychologist Sanjay Srivastava (2011) identifies the problem: "As long as a journal pursues a strategy of publishing 'wow' studies, it will inevitably contain more unreplicable findings and unsupportable conclusions than equally rigorous but more 'boring' journals. Ground-breaking will always be higher-risk. And definitive will be the territory of journals that publish meta-analyses and reviews."

Looking forward

Three cheers for the referee: One cheer for quality, two for fairness, three for excellence. Just as backyard players aspire to higher levels of play, (true) scientists *want* to be reviewed. We want our work to be high quality and accepted by our peers, and we accept refereeing—and journals—as part of this evaluation excellence. This does not mean that we must accept poor practice, in terms of quality or ethics, among referees or publishers. Nor does it mean that having found such faults, we should abolish the system. Indeed, for the self-same reasons of ethics and quality, it is likely that even if we did away with scientific refereeing, if we opted instead for a web-free-for-all, a system for identifying excellence and equity would soon emerge. So instead of evicting, let us try evolving. Like any good complex system, improvements such as the establishment of a commons or of society supported post-publication peer review might exhibit similar self-organisation whereby a more satisfactory process of scientific review evolves of its own accord—or then again, it might equally implode.

Finally, we have not addressed the problems of non-replicability in social science, medicine, and applied and computational statistics. Just as biomedical journals are moving toward registration of protocols and data, statistics researchers might soon be expected to produce replicable papers with code, data, and even random seeds. This would in turn impact further the refereeing process.

Acknowledgements

The first and last authors' research is partly supported by the Agence Nationale de la Recherche through the grants ANR-008-BLAN-0218 "BigMCänd ANR-11-BS01-0010 "Calibration". The second author (AG) thanks the U.S. National Science Foundation for partial support of this work. The third author (KM) acknowledges the generosity of CEREMADE and CREST as hosts while this article was written.

References

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models *Bayesian Analysis* 1(3): 515–533

Kriegeskorte, N. 2009. The future of scientific publishing: Open post-publication peer review. http://futureofscipub.wordpress.com/

Kriegeskorte, N. 2012. Open evaluation (OE): post-publication peer review and rating. Parliamentary Session 2010-2011. Written Evidence (PR 14) 28 February 2011 http://tinyurl.com/cfc437z

Kriegeskorte, N., and Deca, D. 2011. Beyond open access: visions for open evaluation of scientific papers by post-publication peer review *Frontiers in Computational Neuroscience*, Research Topic.

Nyhan, B. 2012. Academic reforms: A four-part proposal. http://tinyurl.com/cwllabs

O'Rourke, K., and Detsky, A. S. 1989. Meta-analysis in medical research: strong encouragement for higher quality in individual research efforts. *Journal of Clinical Epidemiology* 42(10): 1021–1024.

Srivastava, S. 2011. Groundbreaking or definitive? Journals need to pick one. http://tinyurl.com/d874oey

Wasserman, L. 2012. A world without referees. *ISBA Bulletin* 19(1): 7–8.

STUDENTS' CORNER

Isadora Antoniano and Antonio Ortiz
isadora.antoniano@unibocconi.it
aao33@kent.ac.uk

Dear Student's Corner audience. With this issue, the ISBA bulletin has reached a turning point. We welcome our new editor, Feng Liang. And, to mark the new beginning, we announce the introduction of a new format for this section.

The first and most important change is that we have decided to give a rest to the Q & A scheme and, particularly, to the experienced members of the Bayesian community who made such section possible. Before we tell you about our new format, we pause to make a bow to those who deserve it so much. We hope you're still following the section, and reading this:

> "Dear Q & A Panel Members, in the name of the previous editors, present co-editors and readers, we wish to express our most sincere gratitude for sharing your points of view, experiences, worries, feelings, and opinions during all these time. We appreciate every single of your collaborations in this Corner and wish you the best."

Now, what were we thinking when we decided to suspend the Q & A scheme? To give an answer, we need to explain the vision we have for the section, which benefited from many helpful suggestions from the same panel members we are so grateful to. The Q & A scheme, after an initial gathering of very interesting questions from the Bayesian Student community, was starting to become one-way communication from