INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-21

"Not only Defended but also Applied" : A Look Back at Feller's take on Bayesian Inference

A. GELMAN[1]
C. P. ROBERT[2]

[1] Department of Statistics and Department of Political Science, Columbia University, USA. Email : gelman@stat.columbia.edu

[2] Université Paris-Dauphine, CEREMADE and CREST, Paris. Email : xian@ceremade.dauphine.fr

"Not only defended but also applied": A look back at Feller's take on Bayesian inference

Andrew Gelman
Department of Statistics and Department of Political Science, Columbia University
gelman@stat.columbia.edu
Christian P. Robert
Université Paris-Dauphine, CEREMADE, and CREST, Paris
xian@ceremade.dauphine.fr

27 June 2010

Abstract. William Feller has a *Note on Bayes' rule* in his classic probability book in which he expresses doubts about the Bayesian approach to statistics and decries it as a method of the past. We analyze in this note the motivations for Feller's attitude, without aiming at a complete historical coverage of the reasons for this dismissal.

Keywords: Foundations, frequentist, Bayesian, Laplace law of succession, doomsdsay argument, bogosity.

Unfortunately Bayes' rule has been somewhat discredited by metaphysical applications of the type described above. In routine practice, this kind of argument can be dangerous. A quality control engineer is concerned with one particular machine and not with an infinite population of machines from which one was chosen at random. He has been advised to use Bayes' rule on the grounds that it is logically acceptable and corresponds to our way of thinking. Plato used this type of argument to prove the existence of Atlantis, and philosophers used it to prove the absurdity of Newton's mechanics. In our case it overlooks the circumstance that the engineer desires success and that he will do better by estimating and minimizing the sources of various types of errors in predicting and guessing. The modern method of statistical tests and estimation is less intuitive but more realistic. It may be not only defended but also applied." — W. Feller, 1950 (pp. 124-125 of the 1970 edition).

"This type of estimation with $u(p) = 1$ was used by Bayes. Within the framework of our model (that is, if we are really concerned with a mixed population of coins with known density u) there can be no objection to the procedure. The trouble is that it is used indiscriminately to judge "probabilities of causes" when there is no randomization in sight." — W. Feller, 1971 (volume 2, p. 56).

"When, as a student in 1946, I decided that I ought to learn some probability theory, it was pure chance which led me to take the book Theory of Probability

by Jeffreys, from the library shelf. In reading it, I was puzzled by something which, I am afraid, will also puzzle many who read the present book. Why was he so much on the defensive? It seemed to me that Jeffreys' viewpoint and—most of his statements were the most obvious common sense—I could not imagine any sane person disputing them. Why, then, did he feel it necessary to insert so many interludes of argumentation vigorously defending his viewpoint? Wasn't he belaboring a straw man? This suspicion disappeared quickly a few years later when I consulted another well-known book on probability (Feller, 1950) and began to realize what a fantastic situation exists in this field. The whole approach of Jeffreys was summarily rejected as metaphysical nonsense, without even a description. The author assured us that Jeffreys' methods of estimation, which seemed to me so simple and satisfactory, were completely erroneous, and wrote in glowing terms about the success of a 'modern theory,' which had abolished all these mistakes. Naturally, I was eager to learn what was wrong with Jeffreys' methods, why such glaring errors had escaped me, and what the new, improved methods were. But when I tried to find the new methods for handling estimation problems (which Jeffreys could formulate in two or three lines of the most elementary mathematics), I found that the new book did not contain them." — E. T. Jaynes, 1974.

1 Introduction

Watching old movies can be a good way to learn about other times and places. Often we learn more from the scenery in the background than from the action in the foreground. More generally, we can sometimes learn much about an era by studying its unexamined assumptions.

It is in that spirit that we consider Feller's notorious dismissal of Bayesian statistics, which is exceptional not in its recommendation—after all, as of 1950 (when the first edition of his wonderful book came out) or even 1970 (the year of his death), Bayesian methods were indeed out of the mainstream of American statistics, both in theory and in application—but rather in its intensity. Feller combined a perhaps-understandable skepticism of the wilder claims of Bayesians with a naïve (in retrospect) faith in the classical Neyman-Pearson theory to solve practical problems in statistics.

To say this again: Feller's real error was not his anti-Bayesianism (an excusable position, given that many researchers at that time were apparently unaware of modern applied Bayesian work)[1] but rather his casual, implicit, unthinking belief that classical statistical methods were essentially complete. In short, he was defining Bayesian statistics by its limitations while crediting the Neyman-Pearson theory with the 1950 equivalent of vaporware: the unstated conviction that, having solved problems such as

[1] One might argue that, whatever the merits of Feller's statement today, it might have been true back in 1950. Such a claim, though, would seem to be contradicted by Jaynes's statement—after all, Jeffreys's *Theory of Probability* came out in 1939—as well as, for example, the success of Bayesian methods by Turing and others in codebreaking in the Second World War, followed up by expositions such as Good (1950). It would be more accurate, we believe, to refer to Bayesian inference as being an undeveloped subfield in statistics at that time, with Feller being one of many academics who were aware of some of the weaker Bayesian ideas but not of the good stuff. This goes even without mentioning Wald's complete class results of the 1940s (Wald's *Statistical Decision Functions* got published in 1950).

inference from the Gaussian, Poisson, binomial, etc., distributions, that it would be no problem to solve all sorts of applied problems in the future. In retrospect, Feller was wildly optimistic that the principle of "estimating and minimizing the sources of various types of errors" would continue to be the best approach to solving engineering problems. (Feller's appreciation of what a statistical problem is seems rather moderate: the two examples Feller concedes to the Bayesian team are (b) finding the probability a family has one child given that it has no girl and (d) urn models for stratification/spurious contagion, problems that are purely probabilistic, no statistics being involved.)

Where was this coming from, historically? With Stephen Stigler out of the room, we are reduced to speculation (or, maybe we should say, we are free to speculate). We doubt that Feller came to his own considered judgment about the relevance of Bayesian inference to the goals of quality control engineers. Rather, we suspect that it was from discussions with one or more statistician colleagues that he drew his strong opinions about the relative merits of different statistical philosophies. In that sense, Feller is interesting case in that he was a leading mathematician of his area, a person who one might have expected would be well informed about statistics,[2] and the quotation reveals the unexamined assumptions of his colleagues. It is doubtful that even the most rabid anti-Bayesian of 2010 would claim that Bayesian inference can be "defended" but not "applied." (We would further argue that the "modern methods of statistics" Feller refers to have to be understood in an historical context as eliminating older approaches by Bayes, Laplace and other 19th century authors, in a spirit akin to Keynes (1921). Modernity starts with the great anti-Bayesian Ronald Fisher who, along with Richard von Mises,[3] is mentioned on page 6 by Feller as the originator of "the statistical attitude towards probability.")

2 The link between Bayes and bogosity

Non-Bayesians still occasionally dredge up Feller's quotation as a pithy reminder of the perils of philosophy unchained by empiricism (see, for example, Ryder, 1976, and DiNardo, 2008). In a recent probability text, Stirzaker (1999) reviews some familiar probability paradoxes (e.g., the Monty Hall problem) and draws the following lesson:

> In any experiment, the procedures and rules that define the sample space and all the probabilities must be explicit and fixed before you begin. This predetermined structure is called a protocol. Embarking on experiments without a complete protocol has proved to be an extremely convenient method of faking results over the years. And will no doubt continue to be so.

[2] However, it may be that Feller indeed had a diminished opinion of statisticians. In the previous page, he criticises "statisticians [who] speak of contagion in a vague and misleading manner." There are 13 mentions made of "statistician" in volume 1 (using Amazon's Look Inside), some of them dismissive.

[3] Von Mises may have been strong in mathematics, but when it came to a simple comparison of binomial variances, he didn't know how to check for statistical significance; see Gelman (2010).

Strirzaker follows up with a portion of the Feller quote and writes, "despite all this experience, the popular press and even, sometimes, learned journals continue to print a variety of these bogus arguments in one form or another." We are not quite sure why he attributes these problems to Bayes, rather than, say, to Kolmogorov—after all, these error-ridden arguments can be viewed as misapplications of probability theory that might never have been made if people were to work with absolute frequencies rather than fractional probabilities (Gigerenzer, 2002).

In any case, no serious scientist can be interested in bogus arguments (except, perhaps, as a teaching tool or as a way to understand how intelligent and well-informed people can make evident mistakes, as discussed in chapter 3 of Gelman et al. (2008). What is perhaps more interesting is the presumed association between Bayes and bogosity. We suspect that it is Bayesians' openness to making assumptions that makes their work a particular target, along with (some) Bayesians' intemperate rhetoric about optimality. Somehow classical terms such as "uniformly most powerful test" do not seem so upsetting. Perhaps what has bothered mathematicians such as Feller and Stirzaker is that the Bayesians actually seem to believe their assumptions rather than merely treating them as counters in a mathematical game. In the first quote, the interpretation of the prior distribution as a reasonning based on an "infinite population of machines" certainly indicates that Feller takes the prior at face value! As shown by the recent foray of Burdzy (2009) into the philosophy of Bayesian foundations and in particular of deFinetti's, this interpretation may be common among probabilists.

But in Bayesian data analysis (which is not the same thing as subjective Bayesian inference!) we do not actually believe our assumptions. Rather, we make strong assumptions in order to make strong inferences and predictions, which can then be tested by comparing to observed and new data (comparable to the "severe testing" of Mayo, 1996; see Gelman and Shalizi, 2010). Unfortunately, we doubt Stirzaker is aware of this perspective—and certainly neither was Feller, writing years before either of the present authors were born.

Recall the following principle, to which we (admitted Bayesians) subscribe:

> Everyone uses Bayesian inference when it is appropriate. A Bayesian is someone who uses Bayesian inference even when it is inappropriate.

What does this mean? Mathematical modelers from R. A. Fisher on down have used and will use probability to model processes that are clearly random, from the scattering of atomic particles to mixing of genes in a cell to random-digit dialing. To be honest, most statisticians are pretty comfortable with probability models even for processes that are not so clearly probabilistic, for example fitting logistic regressions to purchasing decisions or survey responses or connections in a social network. (As discussed in Robert, 2009, Keynes' *Treatise on Probability* is an exception in that Keynes even questions the sampling models.) Bayesians will go the next step and assign a probability distribution to a parameter that one could not possibly imagine to have been generated by a random process, parameters such such as the coefficient of party identification in a regression on vote choice, or the overdispersion in a network model, or Hubble's constant in cosmology.

As noted above, it is our impression that the assumptions of the likelihood are generally more crucial—and often less carefully examined—than the assumptions in the prior. Still, we recognize that Bayeians take this extra step of mathematical modeling. In some ways, the role of Bayesians compared to other statisticians is similar to the position of economists compared to other social scientists, in both cases making additional assumptions that are clearly wrong (in the economists' case, models of rational behavior) in order to get stronger predictions. With great power comes great responsibility, and Bayesians and economists alike have the corresponding duty to check their predictions and abandon or extend their models as necessary.

To return briefly to Stirzaker's quote, we believe he is wrong—or, at least, does not give any good evidence—in his claim that "in any experiment, the procedures and rules that define the sample space and all the probabilities must be explicit and fixed before you begin." Setting a protocol is fine if it is practical, but as discussed by Rubin (1976), what is really important from a statistical perspective is that all the information used in the procedure be based on known and measured variables. This is similar to the idea in survey sampling that clean inference can be obtained from probability sampling—that is, rules under which all items have nonzero probabilities of being selected, with these probabilities being known (or, realistically, modeled in a reasonable way).

It is unfortunate that certain Bayesians have published misleading and oversimplified expositions of the Monty Hall problem; nonetheless, this should not be a reason for statisticians to abandon decades of successful theory and practice on adaptive designs of experiments and surveys, not to mention the use of probability models for non-experimental data (for which there is no "protocol" at all).

3 The sun'll come out tomorrow

The prequel to Feller's quotation above is the notorious argument, attributed to Laplace, that uses a flat prior distribution on a binomial probability to estimate the probability the sun will rise tomorrow. The idea is that the sun has risen n out of n successive days in the past, implying a the posterior mean of $(n+1)/(n+2)$ of the probability p of the sun rising on any future day.

To his credit, Feller immediately recognized the silliness of that argument. For one thing, we don't have direct information on the sun having risen on any particular day, thousands of years ago. So the analysis is conditioning on data that don't exist.

More than that, though, the big, big problem with the $\Pr(\text{sunrise tomorrow} \mid \text{sunrise in the past})$ argument is not in the prior but in the likelihood, which assumes a constant probability and independent events. Why should anyone believe that? Why does it make sense to model a series of astronomical events as though they were spins of a roulette wheel in Vegas? That's not frequentist, it isn't Bayesian, it's just dumb. Or, to put it more charitably, it's a plain vanilla default model that we should use only if we are ready to abandon it on the slightest pretext.[4]

[4]The Laplace law of succession has been discussed in relation to the Humean debate about inference

It is no surprise that when this model fails, it is the likelihood rather than the prior that is causing the problem. After all, the prior comes into the posterior distribution only once, and the likelihood comes in n times. It is perhaps merely an accident of history that skeptics and subjectivists alike strain on the gnat of the prior distribution while swallowing the camel that is the likelihood. In any case, it is instructive that Feller saw this example as an indictment of Bayes (or at least of the uniform prior as a prior for "no advance knowledge") rather than of the binomial distribution.

4 The "doomsday argument" and confusion between frequentist and Bayesian ideas

Bayesian inference has such a hegemonic position in philosophical discussions that, at this point, statistical arguments get interpreted as Bayesian even when they are not.

An example is the so-called doomsday argument (Carter, 1983), which holds that there is a high probability that humanity will be extinct (or drastically reduce in population) soon, because if this were not true—if, for example, humanity were to continue with 10 billion people or so for the next few thousand years—then each of us would be among the first people to exist, and that's highly unlikely.

For our purposes here, the (sociologically) interesting thing about this argument is that it's been presented as Bayesian (see, for example, Dieks, 1992) but it isn't a Bayesian analysis at all! The "doomsday argument" is actually a classical frequentist confidence interval. Averaging over all members of the group under consideration, 95% of these confidence intervals will contain the true value. Thus, if we go back and apply the doomsday argument to thousands of past data sets, its 95% intervals should indeed have 95% coverage. This is the essence of classical statistical theory, that it makes claims about averages, not about particular cases.

However, this does not mean that there is a 95% chance that any particular interval will contain the true value. Especially not in this situation, where we have additional subject-matter knowledge. That's where Bayesian statistics (or, short of that, some humility about applying classical confidence statements to particular cases) comes in. The doomsday argument is pretty silly and also, it's fundamentally not Bayesian.[5]

The doomsday argument sounds Bayesian, though, having three familiar features of traditional Bayesian reasoning:

- It sounds more like philosophy than science.

(see, e.g., Sober, 2008). Berger et al. (2009) discuss other prior distributions for the model. Here, however, we are focusing on the likelihood function, which, despite its extreme inappropriateness for this problem, is typically accepted without question.

[5] Bayesian versions of the doomsday argument have been constructed, but from our perspective these are just unsuccessful attempts to take what is fundamentally a frequentist idea and adapt it to make statements about particular cases. See Dieks (1992) and Neal (2008) for detailed critiques of the assumptions underlying Bayesian formulations of the doomsday argument.

- It's a probabilistic statement about a particular future event.
- It's wacky, in an overconfident, "you gotta believe this counterintuitive finding, it's supported by airtight logical reasoning," sort of way.

Really, though, it's a classical confidence interval, tricked up with enough philosophical mystery and invocation of Bayes that people think that the 95% interval applies to every individual case. Or, to put it another way, the doomsday argument is the ultimate triumph of the idea, beloved among Bayesian educators, that our students and clients don't really understand Neyman-Pearson confidence intervals and inevitably give them the intuitive Bayesian interpretation.

Misunderstandings of the unconditional nature of frequentist probability statements are hardly new. Consider Feller's statement, "A quality control engineer is concerned with one particular machine and not with an infinite population of machines from which one was chosen at random." It sounds as if Feller is objecting to the prior distribution or "infinite population," $p(\theta)$, and saying that he only wants inference for a particular value of θ. This misunderstanding is rather surprising when issued by a probabilist but it shows a confusion between data and parameter: as mentioned above, the engineer wants to condition upon the data at hand (with obviously a specific if unknown value of θ lurking in the background). It does not help that many Bayesians over the years have muddied the waters by describing parameters as random rather than fixed. Actually, for Bayesians as much as any other statisticians, parameters are fixed but unknown.

In any case, we suspect that many quality control engineers do care about multiple machines, maybe even populations of machines, but to us Feller's sentence noted above has the interesting feature that it is actually the opposite of the usual demarcation: typically it is the Bayesian who makes the claim for inference in a particular instance and the frequentist who restricts claims to infinite populations of replications.

5 Conclusions

Why write an article picking on a 60-year-old quotation? We are not seeking to malign the reputation of Feller, a brilliant mathematician and author of arguably the most innovative and intellectually stimulating book ever written on probability theory. Rather, it is Feller's brilliance and eminence that makes the quotation that much more interesting: that this centrally-located figure in probability theory could make a statement that could seem so silly in retrospect (and even not so long in retrospect, as indicated by the memoir of Jaynes quoted above).

Misunderstandings of Bayesian statistics can have practical consequences in the present era as well. We could well imagine a reader of Stirzaker's generally excellent probability text and taking from it the message that all probabilities "must be explicit and fixed before you begin," thus missing out on some of the most exciting and important work being done in statistics today.

Bayesians have the reputation (perhaps deserved) as philosophers who are all too willing to make broad claims about rationality, with optimality theorems that are ultimately built upon the house of cards that is subjective probability, in a denial of the garbage-in-garbage-out principle that defies all common sense. In place of this, Feller (and others of his time) placed the rigorous Neyman-Pearson theory, which "may be not only defended but also applied." And, indeed, if the classical theory of hypothesis testing had lived up to the promise it seemed to have in 1950 (fresh after solving important operations-research problems in the Second World War), then indeed maybe we could have stopped right there.

But, as the recent history of statistics makes so clear, no single paradigm—Bayesian or otherwise—comes close to solving all our statistical problems, and there are huge limitations to the type-1, type-2 error framework which seemed so definitive to Feller's colleagues at the time. At the very least, we hope Feller's example will make us wary of relying on the advice of colleagues to criticize ideas we do not fully understand. New ideas by their nature are often expressed awkwardly and with mistakes—but finding such mistakes can be an occasion for modifying and improving these ideas rather than rejecting them,

Acknowledgements

The first author (AG) thanks the Institute of Education Sciences, Department of Energy, National Science Foundation, and National Security Agency for partial support of this work. He remembers reading with pleasure much of Feller's first volume in college, after taking probability but before taking any statistics courses. The second author's (CPR) research is partly supported by the Agence Nationale de la Recherche (ANR, 212, rue de Bercy 75012 Paris) through the 2007–2010 grant ANR-07-BLAN-0237 "SPBayes." He also remembers buying Feller's first volume in a bookstore in Ann Arbor during a Bayesian econometrics conference where he was kindly supported by Jim Berger.

References

Berger, J., J. Bernardo, and S. D. 2009. Natural induction: An objective Bayesian approach. *Rev. Acad. Sci. Madrid* A 103: 125–159. (With discussion).

Burdzy, K. 2009. *The Search for Certainty*. Singapore: World Scientific.

Dieks, D. 1992. Doomsday–or: the dangers of statistics. *Philosophical Quarterly* 42.

DiNardo, J. 2008. Introductory Remarks on Metastatistics for The Practically Minded Non-Bayesian Regression Runner. In *Palgrave Handbook of Econometrics: Vol. 2 Applied Econometrics*, eds. T. C. Mills and K. Patterson. Basingstoke: Macmillan. URL `http://www-personal.umich.edu/~jdinardo/draftfinal4.pdf`

Feller, W. 1950. *An Introduction to Probability Theory and its Applications*. New York: John Wiley.

—. 1970. *An Introduction to Probability Theory and its Applications*, vol. 1. New York: John Wiley.

—. 1971. *An Introduction to Probability Theory and its Applications*, vol. 2. New York: John Wiley.

Gelman, A. 2010. Going beyond the book: Toward critical reading in statistics teaching. *Teaching Statistics* (To appear). `http://www.stat.columbia.edu/~gelman/research/published/dontbelieve3.pdf`.

Gelman, A., D. Park, B. Shor, J. Bafumi, and J. Cortina. 2008. *Red State, Blue State, Rich State, Poor State: Why Americans Vote the Way They Do*. Princeton University Press.

Gelman, A. and C. R. Shalizi. 2010. Philosophy and the practice of Bayesian statistics. Tech. rep., Department of Statistics, Columbia University. `http://www.stat.columbia.edu/~gelman/research/unpublished/philosophy.pdf`.

Gigerenzer, G. 2002. *Calculated Risks: How to Know When Numbers Deceive You*. New York: Simon and Schuster.

Good, I. 1950. *Probability and the Weighting of Evidence*. London: Charles Griffin.

Jaynes, E. 1974. Lectures on Probability Theory.

Jeffreys, H. 1939. *Theory of Probability*. 1st ed. Oxford: The Clarendon Press.

Keynes, J. 1920. *A Treatise on Probability*. London: Macmillan and Co.

Mayo, D. G. 1996. *Error and the Growth of Experimental Knowledge*. University of Chicago Press.

Robert, C. 2009. An attempt at reading Keynes' Treatise on Probability. Tech. Rep. 1003.5544, arXiv.

Ryder, J. M. 1976. Subjectivism—a reply in defense of classical actuarial methods (with discussion). *Journal of the Institute of Actuaries* 103: 59–112.

Sober, E. 2008. *Evidence and Evolution*. Cambridge: Cambridge University Press.

Stirzaker, D. 1999. *Probability and Random Variables: A Beginners Guide*. Cambridge University Press.

Wald, A. 1950. *Statistical Decision Functions*. New York: John Wiley.