Bayesian Analysis (2008) **3**, Number 3, pp. 459–462

Comment on Article by Gelman

Stephen Senn*

> 'I spent the decade of the 1950s preaching the gospel of likelihood'. (Barnard 1996) p266

The late and great George Barnard, through his promotion of the likelihood principle, probably did as much as any statistician in the second half of the last century to undermine the foundations of the then dominant Neyman-Pearson framework and hence prepare the way for the complete acceptance of Bayesian ideas that has been predicted will be achieved by the De Finetti-Lindley limit of 2020. (See Dennis Lindley's preface to (de Finetti 1974))

Nevertheless, George Barnard never joined the Bayesian camp and, indeed, in a thoughtful article towards the end of his life suggested that all statisticians should be familiar with what he considered were the four great systems of inference 'the Fisherian, the Jeffreys, the Neyman-Pearsonian, and the Bayesian'(Barnard 1996). By the Bayesian he meant the subjective Bayesian approach particular associated with Ramsey, DeFinetti, Savage, Lindley and Good(Senn 2003), although, of course, there are considerable differences between many of these authors (see, for example Lindley's comments on (Good 1967)) and indeed many who call themselves Bayesian today disagree with each other on many fundamentals, as Gelman's thought-provoking (but also tongue-in-cheek) piece illustrates.

What is interesting about Barnard's quartet is that Jeffreys is awarded a system that is regarded as being distinct from that which is 'Bayesian'. Anybody who looks at the practice of Bayesian statistics today might find that rather strange since, even where improper Jeffreys prior distributions are not used, in an attempt to produce similar inferences, vague prior probabilities are frequently used by those calling themselves Bayesian, although they form no necessary part of the subjective Bayesian approach. These prior distributions are in practice conjugate and thus not only *convenient* as noted by Gelman but implausibly vague. What is almost never used, however, is the Jeffreys significance test. However, adopting Jeffreys's approach to parameter estimation without adopting his approach to significance testing is a recipe for disaster and I would say as one who tries to follow Barnard's recommendation (within the limitations of my lesser abilities), that the Jeffreys-subjective synthesis now common betrays a much more dangerous confusion than the Neyman-Pearson-Fisher synthesis as regards significance/hypothesis tests, so frequently criticised by Bayesians(Goodman 1992; Senn 2001; Senn 2002; Senn 2003). There is thus some linguistic virtue in Barnard's system, since if adopted more widely, it would encourage statisticians to realise that using prior

*Department of Statistics, University of Glasgow, Glasgow, Scotland, mailto:stephen@stats.gla.ac.uk

 DOI:10.1214/08-BA318C

distributions, let alone Bayes theorem, as part of your approach to inference is not at all the same as being coherent in the sense meant by Savage and DeFinetti.

So, Gelman's comment 'Bayesian inference is a coherent mathematical theory but I don't trust it in scientific applications' is reasonable if it is translated to mean, 'the pure subjective Bayesian approach is coherent but difficult to implement and the mongrel surrogate used in practice has many weaknesses'. Of Barnard's four systems the subjective Bayesian one strikes me as being the most impressive in the sense of satisfying the dicta of being self-contained and logical but it also strikes me as being impossible to apply. In fact I would say it provides a very reasonable theory of how to remain perfect but it does not really explain how you can become good. This is because if you are currently coherent it provides a perfectly logical account of how you should update probability distributions to remain so. If, however, you are not currently coherent, you can only become so by striking out or replacing probability distributions to leave a coherent set. But if this is permitted it is not clear why it is not always permitted at any time, in which case you can renege on bets whenever you like and the Dutch book argument loses much of its force.

However, that is not the same as saying that a good approach to applied statistics should never rely on Bayesian methods and in particular it is not the same as saying that one should eschew Bayesian thinking. I think Ehrenberg's point c) quoted by Gelman is put too strongly. Bayesianism *as practiced* frequently ducks the issue. In particular, putting vague prior distributions on nuisance parameters can lead to very peculiar inferences but these are precisely under circumstance where frequentists would have difficulties in doing analyses.

Gelman mentions the attractions of unbiasedness but if Bayesians have taught us anything it is that there are two sorts of unbiasedness, *direct*, which is the standard frequentist characteristic that the estimate should on average be equal to the parameter and *inverse* , which is the Bayesian requirement that the parameter should on average be equal to the estimate. I think that Bayesians have missed a trick by not using a vocabulary that stresses a distinction between the two and hence allowing frequentists sole use of a word with many positive connotations for a property that is less than compelling. Confusion between the two sorts of bias is rife. For discussion of an example concerning errors in variables see (Senn 1995).

I am in perfect agreement with Gelman's strictures against using the data to construct the prior distribution. There is only one word for this and it is 'cheating'. But I would go further and draw attention to another point that George Barnard stressed and that is that posterior distributions are of little use to anybody else when reporting analyses. Thus if you believe (as I do) that far more use of Bayesian decision analysis should be made in clinical research it does not follow that Bayesian analyses should be used in reporting the results. In fact the gloomy conclusion to which I am drawn on reading (de Finetti 1974) is that ultimately the Bayesian theory is destructive of any form of public statistics. To the degree that we agreed on the model (and the nuisance parameters) we could communicate likelihoods but in practice all that may be left is the communication of data. I am thus tempted to adapt Niels Bohr's famous quip about quantum theory to say that, 'anyone who is not shocked by the Bayesian theory of inference has not understood it'.

Andrew Gelman, of course, cannot be accused of not understanding Bayesian theory and I take it that his tongue-in-cheek piece was written to shock the complacent into a deeper understanding. In that, I wish him every success.

References

[1] Barnard, G. A. (1996). "Fragments of a statistical autobiography." *Student*, 1: 257-268.

[2] de Finetti, B. D. (1974). Theory of Probability (Volume 1). Chichester, Wiley.

[3] Good, I. J. (1967). "A Bayesian significance test for multinomial distributions (with discussion)." *Journal of the Royal Statistical Society Series B*, 29: 399-431.

[4] Goodman, S. N. (1992). "A comment on replication, p-values and evidence." *Statistics in Medicine*, 11: 875-9.

[5] Senn, S. J. (1995). "In defence of analysis of covariance: a reply to Chambless and Roeback [letter; comment]." *Statistics in Medicine*, 14: 2283-5.

[6] Senn, S. J. (2001). "Two cheers for P-values." *Journal of Epidemiology and Biostatistics*, 6: 193-204.

[7] Senn, S. J. (2002). "A comment on replication, p-values and evidence S.N.Goodman, Statistics in Medicine 1992; 11:875-879." *Statistics in Medicine* 21: 2437-44.

[8] Senn, S. J. (2003). Dicing with Death. Cambridge, Cambridge University Press.

[9] Senn, S. J. (2003). P-Values. Encylopedia of Biopharmaceutical Statistics. S. C. Chow, Marcel Dekker: 685-695.