A discussion on "Is Bayes posterior just quick and dirty confidence?" by D.A.S. Fraser,

Christian P. Robert, Université Paris Dauphine, IuF, and CREST

While Don's paper does shed new insight on the evaluation of Bayesian bounds in a frequentist light, the main point of the paper seems to be a radical reexamination of the relevance of the whole Bayesian approach to condence regions. This is surprising given that the disagreement between classical and frequentist perspectives is usually quite limited (in contrast with tests) in that the coverage statements agree to orders between $n^{-1/2}$ and n^{-1}, following older results by Welch and Peers (1963).

First, the paper seems to contain a lot of apocryphal deeds attributed to Thomas Bayes. My understanding of the 1763 posthumous paper of Thomas Bayes is one of a derivation of the posterior distribution of a probability parameter θ driving a binomial observation $x \sim \mathcal{B}(n, \theta)$. It thus fails to contain anything about confidence statements or location parameters, in relation with the "translation invariance" mentioned in the Introduction or in Section 7. As noted by Fienberg (2006), Thomas Bayes does not either introduce explicitly the constant prior as a rule, even in his limited perspective, this had to wait for Pierre Simon de Laplace twenty to thirty years later. Thanks to Don's paper, however, I re-read Bayes' essay (in Edward Demings 1940 reprint) and found in both RULE 2 (page 400 and further) and RULE 3 (page 403 and further) that Bayes approximated the (posterior) probability that the parameter θ is between $x/n - z$ and $x/n + z$. However, a closer examination revealed that this part (starting on page 399) had actually been written by Richard Price (even though Price mentions "Mr Bayes's manuscript").

My second and more important point of contention is that the Bayesian perspective on confidence (or credible) regions and statements does not claim "correct coverage" from a frequentist viewpoint since it is articulated in terms of the parameters. Probability calculus remains probability calculus whether it applies to the parameter space or to the observation space, making the comment about *the term probability [being] less appropriate in the Bayesian weighted likelihood* quite debatable. Following Jaynes (2003), "there is only one kind of probability". The title of the paper is thus in complete contradiction with the purpose of Bayesian inference and the chance identity occurring for location parameters is a coincidence on which one should not build sandcastles.

I find Bayesian analysis neither quick (although it is logical) nor, obviously, dirty (on the opposite, it proposes a more complete and more elegant inferential framework!). Looking at a probability evaluation on the parameter space being "correct" (Section 3) is also strange in that the referential for a Bayesian analysis is the prior endowed space, not the consequences on observable values that have not been observed to paraphrase Harold Jeffreys. A Bayesian credible interval is therefore correct in terms of the posterior distribution it is derived from and it does not address the completely different target of finding a frequency-valid interval. (The distinction made in the Bayesian literature, as e.g. Berger (1985) between confidence and credible intervals is significant for those different purposes.) That a β quantile Bayesian confidence bound does not exclude the true value of the parameter in $100\beta\%$ of the observations is not a cause for worry when considering only the observed y_0 and the example of Section 4 is perfectly illustrating this perspective. When I see on Figure 4 (c) that the Bayesian coverage starts at 1 when $\theta = \theta_0$ I am indeed quite happy with the fact that this coherent procedure accounts for the fact that θ cannot be lesser than θ_0. I thus strongly object to the dire conclusion of *Bayes approach [being] viewed as a long history of misdirection*! I also fail to understand what is the meaning of "reality" in Section 7. When running Bayesian inference, the parameter θ driving the observed data is fixed but unknown. Having a prior attached to it has nothing to do with "reality", it is a reference measure that is necessary for making probability statements (or, quoting again from Jaynes, 2003, extending the logics framework). Thus the apparently logical concern in Section 7 on *how probabilities can reasonably be attached to a constant* has no raison d'être. The debate about *where the prior comes from* (Section 9) neither. If the matter is about improper versus proper priors (as hinted at by the comments about marginalisation paradoxes), this has been extensively discussed in the literature and the difference seems to me less important than the difference between Bayes and generalised Bayes estimators.

While this is directly related to the above, the discussion in the Paradigm section also confuses me. Introducing a temporal order between y_1 and y_2 does not make sense from a probabilistic viewpoint. Both representations $f(y_1)f(y_2^0|y_1)$ and $f(y_2^0)f(y_1|y_2^0)$ are equally valid. I note as a side remark that the derivation of $f(y_1|y_2^0)$ as the collection of simulated y_1's for which $y_2 = y_2^0$ is exactly the starting point of the ABC (Approximate Bayesian calculation) algorithm (Rubin, 1984; Pritchard et al., 1999).

I further object to the debate about optimality (and the subsequent relevance of Bayes procedures) as I do believe that decision theory brings a useful if formal representation of statistical inference. The *choice of the criterion* that I understand as the choice of the loss function is clearly important; however, it helps in putting a meaning to notions like "real" or "true" or "correct" found in the paper. Changing the criterion does change the outcome for the "optimal" interval but the underlying relevance of Bayesian procedures does not go away. For instance, we proposed in Robert and Casella (1994) several of such losses for evaluating confidence sets. The criticism found at the end of this Section 9 is inappropriate in that the posterior quantile is neither derived from a loss function nor evaluated under a specific loss function, since the "non-zero" curvature drawback stems from a frequentist perspective. Let me also add that, even from a frequentist perspective, strange and counter-intuitive phenomena can occur, like the domination of the classical confidence region by regions that are equal to the empty set with positive probability (Hwang and Chen, 1986).

In conclusion, I am quite sorry about the negative (and possibly strident) tone of this discussion. However, I do not see a convincing reason for opening afresh the Pandora box about the (lack of) justifications for the Bayesian approach, the true nature of probability and the philosophical relevance of priors: The last section is a nice and provocative enough collection of aphorisms, although I doubt it will make a dent in the convictions of Bayesian readers. Bayesian credible intervals *are not* frequentist confidence intervals and thus do not derive their optimality from providing an exact frequentist coverage.

References

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis.* Springer Verlag, New York.

Fienberg, S. (2006) When did Bayesian inference become "Bayesian"? *Bayesian Analysis* **1**(1), 1–40.

Hwang, J.T. and Chen, J. (1986) Improved Confidence Sets for the Coefficients of a Linear Model with Spherically Symmetric Errors *Annals Statist.* **14**(2), 444–460.

Jaynes, E.D. (2003) *Probability Theory.* Cambridge University Press.

Pritchard, J.K., Seielstad, M.T., Perez-Lezaun, A. and Feldman, M.W. (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16**, 1791–1798

Robert, C.P. and Casella, G. (1994) Distance penalized losses for testing and confidence set evaluation. *Test* **3** (1), 163-182.

Rubin, D.B. (1984) Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals Statist.*, **12**, 1151–1172.

Welch, B.L. and Peers, H.W. (1963) On formulae for confidence points based on integrals of weighted likelihoods. *J. Royal Statistical Society* Series B,

25, 318-329.