# Comments on "Mixtures of g-priors for Bayesian Variable Selection"

by

F. Liang, R. Paulo, G. Molina, M.A. Clyde and J.O. Berger

by

Arnold Zellner
Graduate School of Business
University of Chicago

I congratulate the authors of this very interesting paper on their work in which they implement my suggestion in Zellner (1986) to put an informative prior on the g parameter. The authors have introduced hierarchical priors on $g$ and shown how they can be used to solve model comparison and selection problems. Thanks to them for this valuable extension of my earlier results and for comparing their results to Zellner and Siow's (1980) use of multivariate Cauchy priors in computing Bayes' factors for model comparison and selection problems. However, there are some points about "undesirable consistency issues" and "paradoxes", to use the authors' terms, and $g-$priors that require comment in order to acquire an appropriate understanding of important underlying testing methodological issues. For example, does a perfect model fit, $R^2 = 1,$ require that the posterior odds for this model relative to others under consideration go to infinity? And if so, is it the true model?

Let us start with what the authors call the "Bartlett Paradox", considered in Section 2.3 of their paper, that supposedly is a negative aspect of the use of $g$-priors with a fixed value for the prior parameter $g$. They state that when one attempts to be "non-informative" by letting the value of $g$ approach + infinity, the Bayes' factor for comparing two models will go to 0 and provide the following interpretation: "That is, the large spread of the prior induced by the non-informative choice of $g$ has the unintended consequence of forcing the Bayes factor to favor the null model, the smallest model, regardless of the information in the data." (p.412) What the authors fail to appreciate is the fact that this will happen not only for an improper g-prior but also for any improper prior, as is well known. And this is not a paradox, since for an improper prior relating to a parameter or vector of parameters, say $b$, it implies that $b$ has a zero probability of falling in any finite region; see Zellner (1971, p. 42) for a discussion of this point in connection with Jeffreys' use of improper priors to represent ignorance. Thus when improper priors are used in connection with the formulation of posterior odds, they dogmatically introduce the information that $b$ has a zero probability of falling in any finite interval and thus it is no paradox that the posterior odds favor the null hypothesis $b = 0$ which is assigned a non-zero prior probability in the formulation of the prior odds. The posterior odds in this case reflects the dogmatic prior input information, or misinformation, and thus there is no paradox specially relating to the use of $g-$priors with $g =$ infinity.

Second, the authors introduce an "Information Paradox" that supposedly afflicts the use of $g-$priors with a fixed value of $g$. In this case they consider a model, say a regression model, $M$ that fits the data perfectly, $R^2 = 1$ and they write that "… one would expect that [this model $M$ ] should receive high posterior probability and that the Bayes Factor [relating it to the null model] would go to infinity as the information against the [null model] accumulates." [p. 413]. Instead as they point out "…the Bayes factor (6) with a fixed choice of $g$ tends to a constant $(1+g)^{(n-p-1)/2}$, as $R^2$ approaches $1$ (Zellner 1986, Berger and Pericchi 2001." [p.413]. Note that if $n$ is large, this constant can be very large but not infinite for finite $n$, $g$, and $p$.

In my view that the posterior odds does not become infinite as $R^2$ approaches $1$ for a particular model, is not "paradoxical" but very reasonable in view of the fact that Jeffreys (1961, p.3) has pointed out that for any given set of data, there are an infinity of models that will fit the given data set exactly, i.e. $R^2 = 1$, not just the assumed model under consideration. That is, if $y(t) = a + bx(t) + cz(t)$ fits exactly for $t = 1, 2, \ldots, T$, then $y(t) = a + bx(t) + cz(t) + f(t)(t-1)(t-2)\ldots(t-T)$ will do so also for any $f(t)$ that has a finite value for each value of $t$. This implies that the odds should not have an infinite value for one particular model with a perfect fit to the data.

Further, as John Geweke has reminded me recently, the eminent statistician George Box has been widely quoted as saying that "All models are false." Note too that in physics, many 19th century physicists believed that Newton's Laws were "absolutely true" and later were shocked to have Einstein's Laws replace Newton's Laws. And now there is empirical evidence contradicting an important prediction of Einstein's Laws, namely that the universe will expand at a decreasing rate. And so on and so on in the iterative, learning process called science that apparently has not as yet converged on "ultimate truth" or infinite posterior odds for any model. Apparently, there is always room for model improvement, not only with respect to automobile models but also statistical models in the spirit of Deming's continuous quality improvement approach.

However, if it is considered "reasonable" to have the posterior odds possibly equal to infinity, "ultimate truth" for an alternative model relative to a specific null model, this can be done using a fixed $g$ in a normal $g-$prior for the alternative model's extra parameters, by permitting the prior odds to be random with a range from 0 to +infinity, as I have suggested in recent papers. To illustrate, for two mutually exclusive models, say $M1$ and $M2$, usually the prior odds, $p/(1-p)$ is taken equal to $1$ in order to be “fair.” However, this involves the very strong assumption that $p = 1/2$ for each model, a rather high value. To still be “fair” and free up this dogmatic value of $1/2$, assume $Ep/(1-p) = 1$; that is, the prior expectation of the prior odds $=1$, a “fair” value. Then the posterior odds $(PO)$ becomes subjectively random and $EPO = Ep/(1-p)\,x\;BF = BF$, where $E$ denotes the subjective expectation operator and $BF$ designates the given Bayes’ factor. In this formulation, the random $PO$ can have a range of possible values from 0 to infinity and does not dogmatically preclude a possible value of infinity even when the $BF$ has a finite value that could possibly increase as more observations are obtained, i.e. as $n$ increases.

Last, as regards reasons for employing a Cauchy prior in computing posterior odds for a null hypothesis versus an alternative hypothesis rather than a normal prior, I believe that the authors in the first sentence of Section 3.1 misrepresent Jeffreys’ reasons for using a Cauchy prior. There was nothing related to the authors’ “paradoxes” involved. See Jeffreys (1961, pp. 268-270) for his rationale for use of a univariate Cauchy prior in testing that a normal mean $= 0$ versus that it has a non-zero mean that motivated Siow and me to employ a special multivariate Cauchy prior in our 1980 paper that dealt with deriving posterior odds for hypotheses about coefficient vectors.

I hope that these brief comments on the authors’ paper do not detract from its valuable methodological contribution that those working on model selection and comparison problems will, in my opinion, find very useful.

References

Liang, F., Paulo, R., Molina, G., Clyde, C.A. and Berger, J.O. (2008), “Mixtures of *g* Priors for Bayesian Variable Selection,” Journal of the American Statistical Association, Vol. 103, No. 481, 410-423.

Jeffreys, H. (1961), Theory of Probability, 3rd ed., in Oxford Classic Texts in the Physical Sciences, Oxford: Oxford University Press.

Zellner, A. (1971), An Introduction to Bayesian Inference in Econometrics, New York: Wiley.

_________ (1986), “On Assessing Prior Distributions and Bayesian Regression Analysis with g-prior Distributions,” in Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, eds., P.K. Goel and A. Zellner, Amsterdam: North-Holland/Elsevier, 233-343.

_________ and Siow, A. (1980), “Posterior Odds Ratios for Selected Regression Hypotheses, in Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia, eds., J.M. Bernardo, M.H. Degroot, D.V. Lindley and A.F.M. Smith, Valencia, Spain: University of Valencia Press, 585-603.