for an *acceptable* hypothesis. In my elementary statistics classes we can and sometimes do get into extended discussions about sample size, and the arbitrariness of any cutoff value, as well as the virtues of estimation rather than testing. However, statistics students—like everyone else—need to have clear and concise guidelines they can remember. I tell students to interpret $p > .05$ as "no evidence against H_0" and $p < .05$ as "some evidence against H_0" but I frequently emphasize that a p near .05 would be pretty shaky. In doing this I think I am offering a bit more conservative advice than is traditional, but it seems to me to be good advice that is more or less consistent with its Gosset/Fisher origins while being enlightened by subsequent analysis. So, my last question is this: How has any of this reconsideration of the foundations affected their thinking, and what scale of evidence do they tell their students to use?

Let me now, with great pleasure, thank the authors for a stimulating article.

ADDITIONAL REFERENCES

Edwards, W., Lindman, H., and Savage, L.J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.* **70**, 193-242.

Efron, B. (1986). Why isn't everyone a Bayesian? *Amer. Statist.* **40**, 1-11.

George, E.I. and Foster, D.P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87**, 731-748.

Pearson, E.S. (1938) "Student" as statistician, in William Sealy Gosset, 1876-1937. *Biometrika* **30**, 210-250.

G. S. Datta and P.Lahiri

University of Georgia and University of Nebraska-Lincoln

Starting from a simple testing hypothesis problem, the authors have very successfully demonstrated how one can possibly reconcile two apparently different approaches (Fisher's and Jeffreys') to model selection. In the process, they have proposed a new frequentist Bayes factor for model selection.

G. S. Datta is Professor, Department of Statisitcs, University of Georgia, Athens, GA 30602, U.S.A; email: gauri@stat.uga.edu. P. Lahiri is Milton Mohr Distinguished Professor of Statisitcs, Department of Mathematics and Statistics, University of Nebraska-Lincoln, Lincoln, NE 68588-0323, U.S.A; email: plahiri@math.unl.edu. Datta was partially supported by NSF grants SBR-9705145 and DMS-0071642 and by NSA. Lahiri was supported in part by NSF grant SBR-9978145. The authors thank Mr. Shijie Chen for computing help.

In this discussion, we will consider the important problem of selection between a random effect model and a fixed effect model. Such a problem is encountered in many applications including animal breeding and small-area estimation. For the simplicity of exposition, let us consider the following one-way balanced random effects model:

$$y_{ij} = \mu + a_i + e_{ij}, \qquad i = 1, ..., m; \qquad j = 1, ..., n_0,$$

where a_i's and e_{ij}'s are independent with $a_i \overset{\text{i.i.d}}{\sim} N(0, \sigma_a^2)$ and $e_{ij} \overset{\text{i.i.d}}{\sim} N(0, \sigma_e^2)$. Let us call this model M.

Note that this model belongs to the exponential family considered by the authors. The model selection in question can be equivalently viewed as the following one-sided testing hypothesis problem:

$$H_0 : \sigma_a^2 = 0 \qquad \text{vs} \qquad H_a : \sigma_a^2 > 0.$$

Under this null hypothesis model M reduces to model M_0: $y_{ij} \overset{\text{i.i.d}}{\sim} N(\mu, \sigma_e^2)$. We consider this model selection problem for two reasons. First, unlike the testing of μ considered by the authors, our example distinguishes among different types of noninformative priors available in the literature. Secondly, the parameter value specified by the null hypothesis falls on the boundary of the parameter space, a case not covered by the authors. Is the approximation formula given in Lemma (section 2.2) valid in this situation? We try to investigate this question by numerical examples. Do the authors have any comments here?

Define $S = \Sigma\Sigma(y_{ij} - \bar{y}_i)^2$, $SSB = n_0\Sigma(\bar{y}_i - \bar{y})^2$ and $SST = \sum_i \sum_j (y_{ij} - \bar{y})^2$. We also define $\eta = \sigma_a^2/\sigma_e^2$. To calculate the Bayes factor in favor of M against M_0, we need to calculate the marginal densities $f_M(y)$ and $f_{M_0}(y)$, where

$$f_M(y) = \int L(\eta, \sigma_e^2, \mu)\pi(\eta, \sigma_e^2, \mu) d\mu d\eta d\sigma_e^2, \tag{1}$$

and

$$f_{M_0}(y) = \int L(0, \sigma_e^2, \mu)\pi_0(\sigma_e^2, \mu) d\mu d\sigma_e^2, \tag{2}$$

where the unrestricted likelihood $L(\eta, \sigma_e^2, \mu)$ and the likelihood under the null hypothesis $L(0, \sigma_e^2, \mu)$ are given by

$$\begin{aligned} L(\eta, \sigma_e^2, \mu) &= (2\pi)^{-\frac{n_0 m}{2}} (\sigma_e^2)^{-\frac{mn_0}{2}} (1 + n_0\eta)^{-\frac{m}{2}} \times \exp[-\frac{1}{2\sigma_e^2}\{s + \frac{n_0}{1 + n_0\eta}\sum_{i=1}^{m}(\bar{y}_i - \mu)^2\}], \\ L(0, \sigma_e^2, \mu) &= (2\pi)^{-\frac{mn_0}{2}} (\sigma_e^2)^{-\frac{mn_0}{2}} \exp[-\frac{s + n_0\sum_{i=1}^{m}(\bar{y}_i - \mu)^2}{2\sigma_e^2}]. \end{aligned} \tag{3}$$

To consider authors' suggestion in calculating Bayes factor based on Jeffreys' priors (under M and M_0), we use a general class of priors. Under the model M, we use

$$\pi(\eta, \sigma_e^2, \mu) = (\sigma_e^2)^{\{a_1+a_2+a_3+2a_4+1\}}\eta^{a_1}(1+n_0\eta)^{a_3}[(n_0-1)(1+n_0\eta)^2+1]^{a_4}. \quad (4)$$

Under $H_0 : \sigma_a^2 = 0$, the class of priors to be considered is

$$\pi_0(\mu, \sigma_e^2) = (\sigma_e^2)^{a_2^*}. \quad (5)$$

It can be shown (see Datta and Lahiri 2000) that the Bayes factor B in favor of M against M_0 is given by

$$B = \frac{f_M(y)}{f_{M_0}(y)} = 2^{-(a_1+a_2+a_3+2a_4+1-a_2^*)}[SSB+s]^{1/2(mn_0-2a_2^*-3)}$$
$$\times\Gamma(\frac{mn_0-2a_1-2a_2-2a_3-4a_4-5}{2})\{\Gamma(\frac{mn_0-2a_2^*-3}{2})\}^{-1}$$
$$\times \int_0^\infty \eta^{a_1}(1+n_0\eta)^{1/2\{m(n_0-1)-2a_1-2a_2-4a_4-4\}}[SSB+s(1+n_0\eta)]^{-1/2(mn_0-2a_1-2a_2-2a_3-4a_4-5)}d\eta. \quad (6)$$

In order that B in (6) remains free from the unit of measurement we must have

$$a_1+a_2+a_3+2a_4-a_2^*+1=0. \quad (7)$$

Thus, we must be careful in choosing the noninformative priors for the model M and M_0 so that the resulting Bayes factor is unit-free. We will consider the following pairs of priors from the classes (4) and (5) described above in calculating the Bayes factor B.

(a) Jeffreys' priors: Here $a_2^* = a_3 = -3/2$.
Prior under $M = \pi_{JM}(\mu, \sigma_a^2, \sigma_e^2) = (\sigma_e^2)^{-1}(\sigma_e^2+n_0\sigma_a^2)^{-3/2}$,
Prior under $M_0 = \pi_{JM_0}(\mu, \sigma_e^2) = (\sigma_e^2)^{-3/2}$.

(b) Matching prior for η vs. Jeffrey's recommended prior: Here $a_2^* = a_3 = -1$.
Prior under $M = \pi_{\eta M}(\mu, \sigma_a^2, \sigma_e^2) = \{\sigma_e^2(\sigma_e^2+n_0\sigma_a^2)\}^{-1}$
Prior under $M_0 = \pi_{M_0}(\mu, \sigma_e^2) = (\sigma_e^2)^{-1}$.

(c) Matching prior for σ_a^2 in the full model: Here $a_2^* = a_3 = -2$.
Prior under $M = \pi_{\sigma_a^2 M}(\mu, \sigma_a^2, \sigma_e^2) = \pi_M^{(2)}(\mu, \sigma_a^2, \sigma_e^2) = \frac{[(n_0-1)(\sigma_e^2+n_0\sigma_a^2)^2+\sigma_e^4]^{1/2}}{[\sigma_e^2(\sigma_e^2+n_0\sigma_a^2)]^2}$.
Prior under $M_0 = \pi(\mu, \sigma_e^2) = (\sigma_e^2)^{-2}$.

Datta and Lahiri (2000) showed that for the prior-pair (a)-(c), B further simplifies to

$$B = \frac{1}{n_0} \frac{\int_b^1 w^{\frac{1}{2}\{m(n_0-1)\}-1}(1-w)^{\frac{1}{2}(m-2a_3-3)-1} dw}{b^{\frac{1}{2}(n_0-1)m}(1-b)^{\frac{1}{2}(m-2a_3-3)}} = B(a_3) \;\; (\text{say}), \tag{8}$$

where $b = S/SST$. Note that for $b < w < 1$ and for $a_3'' < a_3'$ it is easy to see that $(1-w)^{-a_3'} > (1-w)^{-a_3''}(1-b)^{a_3''-a_3'}$. Using this it can be shown from (8) that $B(a_3') > B(a_3'')$. Thus, B is increasing in a_3. Hence, B for prior pair $(b) > B$ for prior pair $(a) > B$ for prior pair (c). In other words, the matching prior for η is the least favorable for M_0, the matching prior for σ_a^2 is the most favorable for M_0, and Jeffreys' prior (the recommended prior of the authors) falls in between.

The frequentist Bayes factor as given by the authors involves the log-likelihood ratio statistic

$$\begin{aligned} \hat{B}_l(y) &= 2\{\sup_{\mu,\sigma_a^2,\sigma_e^2} \log\; L(\sigma_a^2, \sigma_e^2, \mu) - \sup_{\mu,\sigma_e^2} \log\; L(0, \sigma_e^2, \mu)\} \\ &= 2\{\sup_{\sigma_a^2,\sigma_e^2} \log\; L(\sigma_a^2, \sigma_e^2, \bar{y}) - \sup_{\sigma_e^2} \log\; L(0, \sigma_e^2, \bar{y})\}. \end{aligned} \tag{9}$$

It can be shown that

$$\hat{B}_l(y) \equiv \hat{B}_l(b) = m\{h(\tilde{b}) - h(b)\}I(b < \tilde{b}),$$

where $\tilde{b} = (n_0-1)/n_0$ and $h(x) = (n_0-1)\log\; x + \log(1-x), \quad 0 < x < 1$.

Then the authors' approximation to frequentist Bayes factor $\hat{B}_{Freq}(b)$ is given by

$$\hat{B}_{Freq}(b) = \exp\{\frac{1}{2}\hat{B}_l(b)\}/2.2687, \tag{10}$$

where 2.2687 is based on the cut-off point at level 0.10 of $\hat{B}_l(b)$. Since asymptotically (see Chernoff 1954) $P_{M_0}[\hat{B}_l(b) > x] \approx \frac{1}{2}P[\chi_1^2 > x]$, we get at level 0.10, $x = (1.28)^2$ and $\exp\{x/2\} = 2.2687$.

Consider the dyestuff data given in Box and Tiao (1973, Sec. 5.1.2). Five samples (n_0=5) from each of the six (m=6) randomly chosen batches of raw materials were taken and yield of dyestuff of standard color for each sample was determined. Table 1 reports the values of Bayes factors for three different choices of the noninformative priors and the frequentist Bayes factor. It is interesting to note that the frequentist Bayes factor and the Bayes factors under the noninformative priors (a) - (c) all provide "positive" support for random batch effect.

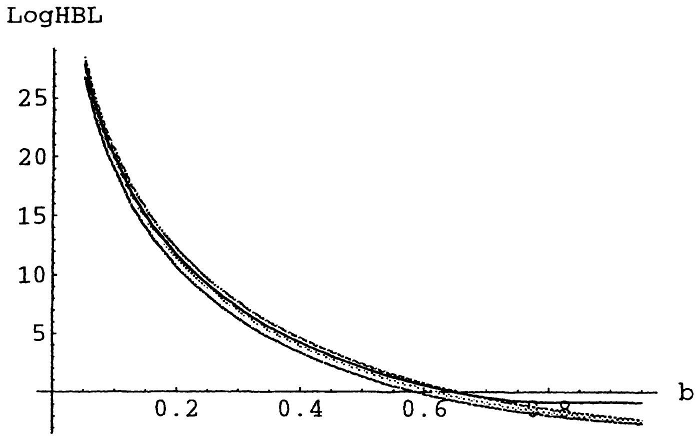

Figure 1: *Logarithm of frequentist Bayes factor and various noninformative Bayes factors for dyestuff data (six batches) of Box and Tiao.*

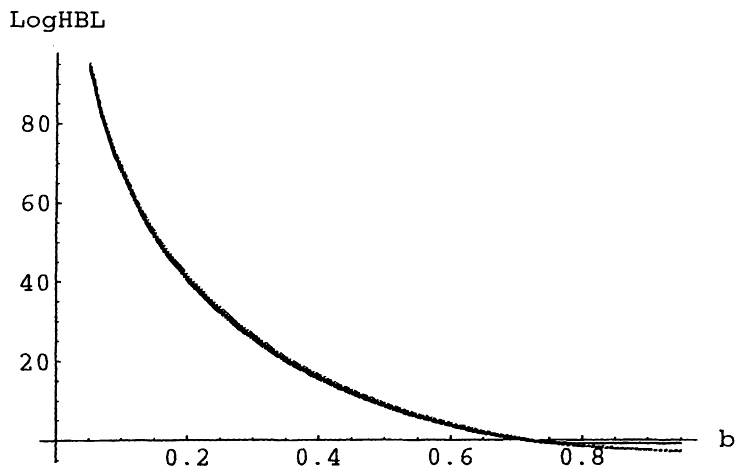

Figure 2: *Logarithm of frequentist Bayes factor and various noninformative Bayes factors for dyestuff data (20 batches, simulated) of Box and Tiao*

Table 1. Frequentist Bayes factor and the Bayes factors under priors (a)-(c)for dyestuff data.

Frequentist	(a)	(b)	(c)
6.568	4.92424	8.67469	3.02869

Note that all the three Bayes factors constructed using noninformative priors (a)-(c) and the frequentist Bayes factor is a function of b. Figures 1 and 2 plot logarithm of Bayes factors against b for $m = 6$ and $m = 20$ (in each case $n_0 = 5$). It is clear that there is very good reconciliation of the Bayes factors under noninformative priors (a)-(c).

ADDITIONAL REFERENCES

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Wiley, New York.

Chernoff, H. (1954). On the distribution of the likelihood ratio, *J. Amer. Statist. Asso.* **25**, 573-578.

Datta, G.S. and Lahiri, P. (2000). A comparison of the Bayes factors and frequentist Bayes factor for a balanced one-way random effects model, *unpublished manuscript.*

REJOINDER

Bradley Efron and Alan Gous

This article was written under the following rule of thumb: no method that's been heavily used in serious statistical practice can be entirely wrong. The rule certainly applies to Fisherian hypothesis testing, but it also applies to Jeffreys and the BIC, leaving us to worry about Figure 1. The two scales of evidence seem to be giving radically different answers, even for sample sizes as small as $n = 100$.

Our paper localizes the disagreement to coherency, in this case sample size coherency, the key distinguishing feature of modern Bayesian philosophy. The BIC, along with any other methodology that acts coherently across different sample sizes, must share Figure 1's behavior, treating the smaller hypothesis M_o ever more favorably as n increases. Fisher's theory, which is usually presented with the sample size fixed, eschews sample size coherency in favor of a more aggressive demand for statistical power.