A new approach to default priors and robust Bayes methodology

Mohan DELAMPADY, Anirban DasGUPTA,
George CASELLA, Herman RUBIN and William E. STRAWDERMAN

Key words and phrases: Bayesian estimator; correlation; Fisher information; Gaussian tilting; maximum likelihood estimator; robust estimation; squared error loss.

Abstract: Following the developments in DasGupta *et al.* (2000), the authors propose and explore a new method for constructing proper default priors and a method for selecting a Bayes estimate from a family. Their results are based on asymptotic expansions of certain marginal correlations. For ease of exposition, most results are presented for location families and squared error loss only. The default prior methodology amounts, ultimately, to the minimization of Fisher information, and hence, Bickel's prior works out as the default prior if the location parameter is bounded. As for the selected Bayes estimate, it corresponds to "Gaussian tilting" of an initial reference prior.

Une nouvelle méthode de sélection automatique de lois a priori et d'estimateurs bayésiens robustes

Résumé : Dans la foulée des travaux de DasGupta *et al.* (2000), les auteurs proposent et explorent une nouvelle méthode de construction de lois a priori génériques intégrables, ainsi qu'une méthode de sélection d'estimateur bayésien dans une classe donnée. Leurs résultats sont déduits de l'expansion asymptotique de certaines corrélations marginales. Pour simplifier la présentation, la plupart de ces résultats ne sont exposés que dans le cadre des modèles de localisation et pour la fonction de perte quadratique. La loi a priori à laquelle les auteurs sont conduits minimise l'information au sens de Fisher et se réduit donc à la loi a priori de Bickel si le paramètre de localisation est borné. Pour sa part, l'estimateur de Bayes choisi correspond à une "inclinaison gaussienne" de la loi a priori référentielle de départ.

1. INTRODUCTION

A radically different way of looking at Pearson's correlation coefficient is detailed in DasGupta *et al.* (2000) where it is used as a binding theme to connect together various approaches to statistical inference. It is also shown there that some of the properties of Pearson's correlation can lead to useful and substantial developments in mathematical statistics, particularly Bayesian statistics. This last theme is further developed in this article.

We begin with an observable X which is distributed as $f(x|\theta)$. The parameter θ is assumed to be distributed according to a prior π, and hence there is then a joint probability distribution for X and θ which we shall call P. As in the above mentioned article, all the developments here follow from the consideration of Pearson's correlation coefficient between two functions $g_1(X,\theta)$ and $g_2(X,\theta)$ under the probability distribution P. However, most of our derivations involve asymptotic arguments which can be rather intractable in general, and hence we shall restrict our attention mostly to random samples from location families. The exact set-up required is outlined in Section 2.

In Section 3, we use the correlation coefficient to address an important problem in robust Bayesian analysis. The overwhelming bulk of the work in robust Bayesian analysis has dealt with sensitivity with respect to the choice of the prior. However, the problem of selecting a specific Bayes estimate is also an important one. Indeed, classical robustness flourished and was taken seriously because specific procedures such as M-estimates were developed. We first select a reference prior π; this reference prior need not belong to the family Γ of priors in consideration. However, usually one will choose π from inside Γ. For a general prior $\nu \in \Gamma$, we then consider the correlation $\rho_P(\theta,\delta_\nu)$ as a criterion for selecting a specific estimate δ_ν; and then ρ_P is calculated under the chosen reference prior π. We maximize an accurate approximation $\hat{\rho}_\nu(\theta, \delta_\nu)$ over δ_ν. The specific chosen estimate δ_ν is Bayes with respect to a prior density of the form

$$\nu(\theta) = \text{constant} \times \pi(\theta) \times \exp\left\{-\frac{1}{2\tau^2}(\theta - \mu)^2\right\}.$$

There are several interesting things about this. First, the generality of the form: one always gets a Gaussian factor. Second, $\nu(\theta)$ has the following interpretation: presumably one will start with a flat reference density $\pi(\theta)$ due to robustness concerns. The final prior $\nu(\theta)$ is formally just the posterior density of θ when a Gaussian observation has been obtained and θ has the prior $\pi(\theta)$. By starting with a flat prior π and ultimately settling for a "formal posterior" as ν, one will pull in the tails but it will still be a more conservative choice than a straight Gaussian prior. However, δ_ν will provide greater shrinkage than δ_π. We have examples illustrating these results. Note that in other contexts, Bayesians have been talking about such a "tilting" of an initial prior by collecting a pilot sample; see Perez (1998), for example. It is interesting that we see this tilting emerge in a purely theoretical way in our results.

Next, in Section 4, we apply the correlation criterion to outline a new method for construction of default priors. Default prior Bayesian analysis has been a very active area of research for a considerable time. After the initial classic contributions of Laplace and Jeffreys, the recent renewed interest has much to do with objective Bayesian inference and the realization that default prior Bayes methods often provide satisfactory frequentist properties. See Berger (1986), Efron (1986), Stein (1982), among many. Conventional default priors in use tend to be improper; thus, nice frequentist properties such as admissibility often have to be established case by case. We develop here an outline for construction of proper default priors. The method suggested is general, although we have worked it out here in detail only for a location parameter.

The method we suggest is as follows. Many Bayesians take the view that post-data opinion about a parameter should be reported simply in terms of a posterior density. On the other hand, there is another clear candidate for such a summary, namely the likelihood function. Just as one can try to minimize an appropriate distance between the two summaries, we suggest maximizing the correlation between them under the joint probability measure P.

Now, the exact correlation, of course, is not something that one can work with. So we provide an appropriate expansion for the correlation, and maximize the appropriate term of this expansion. The expansion is very technical and is presented in the Appendix. It is remarkable that in the end, the maximization based on this expansion corresponds to minimization of the Fisher information of the prior in the chosen family of proper priors. Minimization of Fisher information is a well known variational problem that has arisen in other statistical problems; see Bickel (1981), Bickel & Collins (1983), Huber (1964, 1974), Levit (1979, 1980), Kagan, Linnik & Rao (1973), and Brown (1971). We find this ultimate reduction of our approach to the minimization of Fisher information quite interesting. As a result of this reduction, the Bickel prior (1981) is now seen to have the asymptotic correlation maximization property; compare this with the asymptotic entropy maximization by Jeffreys priors (Clarke & Barron 1994).

The asymptotic expansions we needed are substantially more involved than what is necessary in other problems (e.g., Ghosh, Sinha & Joshi 1982) because we need expansions to more terms for our results. The derivations thus require more smoothness assumptions on the likelihood function and the prior. Exact finite-sample implementation of our approach was not pursued in this article. We remark, however, that our formulation leads to the uniform prior in the binomial case for every finite sample size.

2. THE SET-UP

The following general notation will be used in the sequel: π and ν will denote prior densities for the parameter θ, m the marginal of X (actually m_π or m_ν if the context needs implying dependence of m on the prior used), and E_θ will denote conditional expectation given θ; cov_P, var_P, and ρ_P will, respectively, denote the covariance, variance and correlation under the joint distribution P, whereas cov_π and var_π will denote the covariance and variance under the distribution π on θ.

We consider i.i.d. observations $X_1, X_2, \ldots$ from a location parameter density

$$f(x|\theta) = \exp\{-h(x-\theta)\}.$$

We assume that h is seven times continuously differentiable and that both $h^{(6)}$ and $h^{(7)}$ are bounded. This form of density is used mainly for the convenience it yields while imposing conditions on its logarithm; the loss of generality is minimal, once such conditions are accepted.

The following notation will be used:

$$\left.\begin{array}{lll} \mathcal{L}(\theta, x) = \sum_{i=1}^{n} \log f(x_i|\theta), & \mathcal{L}^{(i)} = \frac{\partial^i}{\partial \theta^i}\mathcal{L}(\theta, x)|_{\theta=\hat{\theta}}, & \hat{\sigma}^2 = -\left(\mathcal{L}^{(2)}/n\right)^{-1}, \\ \ell_i = \mathrm{E}_\theta\left\{\frac{\partial^i}{\partial\theta^i}\log f(X_1|\theta)\right\}, & w_2 = \mathrm{var}_P\{h''(X)\}, & \sigma^2 = \ell_2^{-1}. \end{array}\right\} \quad (1)$$

Here, $\hat{\theta}$ denotes the maximum likelihood estimator (MLE) of θ. We assume that $\sigma^2 > 0$, $\ell_i = 0$ for odd i, and $\mathcal{L}^{(1)} = 0$, i.e., the MLE $\hat{\theta}$ solves the likelihood equation.

By elementary calculations, one can see that $\mathcal{L}^{(3)} = O_p(n^{\frac{1}{2}})$, and that if $W_n = \sqrt{n}(\mathcal{L}^{(2)}/n + 1/\sigma^2)$, then $\mathrm{E}_P(W_n) = O(n^{-1})$ and $\mathrm{E}_P(W_n^2) = w_2 + O(n^{-1})$. The normal location model with known variance trivially fits into our set-up. It can be checked that some other standard location models such as Student's t and the logistic also fit into our set-up. Regarding the prior distributions under consideration in this article, we make the following assumption.

ASSUMPTION A. The reference prior density π and every density ν in Γ is five times continuously differentiable a.e., with bounded fourth and fifth derivatives, and $\mathrm{E}_\nu(\theta) = 0$, $\mathrm{E}_\nu(\theta^2) < \infty$.

3. SELECTING A BAYES ESTIMATE

Robust Bayesian analysis has almost exclusively concentrated on sensitivity of the Bayes estimate and other posterior quantities to the choice of the prior. There is an extensive literature on this now; see the review article by Berger (1994). Far less has been done in the direction of presenting methods for choosing a Bayes estimate from a collection specified by a family of priors; see Zen & DasGupta (1993) for some results on this question. We present below a method for selecting a particular Bayes estimate from a collection for the location parameter case by using the correlation criterion. The notation and the final result are as follows: let Γ be a specified family of priors. Let π be a special prior, a reference. Let ν be a generic element of Γ with δ_ν as the corresponding Bayes estimate. The criterion for selecting a special δ_ν is to maximize over all ν the correlation $\rho_P(\theta, \delta_\nu)$, under the reference prior π. This criterion has an intuitive appeal similar to that of maximizing the expected utility in a decision theoretic set-up.

3.1. A useful approximation to $\rho_P(\theta, \delta_\nu)$.

The problem we wish to address cannot be solved in closed form (and possibly not in any form) if we work with the exact correlation $\rho_P(\theta, \delta_\nu)$. Instead, we present an approximation $\hat{\rho}_P(\theta, \delta_\nu)$. It is an asymptotic approximation, but it can be highly accurate even for $n = 3$. The derivation of the approximation is intensely technical, so we shall break it up into small steps at a time and we will present only the gist.

We now present the approximation $\hat{\rho}_P(\theta, \delta_\nu)$ which we shall maximize over ν. For this, first, we need an expansion for the Bayes estimate $\delta_\nu(x)$ as a function of the MLE $\hat{\theta}$ and $x - \hat{\theta}\cdot 1$, where 1 stands for a vector of ones.

THEOREM 1. *The Bayes estimate $\delta_\nu(x)$ satisfies*

$$\begin{aligned}\delta_\nu(x) &= \hat\theta + \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left(\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)}W_n\right)\\ &+ \frac{\sigma^4}{2n^2}\left\{\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)} + \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2 + \frac{\mathcal{L}^{(4)}}{n}\right) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\right\}\\ &+ O(n^{-5/2}),\end{aligned}$$

uniformly in x.

Proof. The essential details of this long derivation are outlined in the Appendix. We would like to note here that expansions for the posterior mean of this nature can be found in Ghosh (1994), Ghosh, Sinha & Joshi (1982), Johnson (1970) and Lindley (1961). However, they are only accurate up to $O(n^{-3/2})$, whereas we need an approximation which is good to $O(n^{-5/2})$ (and, this of course requires a lot more work). □

Using Theorem 1, we have the following asymptotic expressions for the covariance of δ_ν with θ and the variance of δ_ν. Proofs of Propositions 1 and 2 are again algebraically very involved and will not be given here. However, these can be constructed along the lines of those of Theorem 1 and the complete details can be found in Delampady *et al.* (1999).

For notational convenience, the following notation is used:

$$a = \int \theta\pi(\theta)\tfrac{\nu'(\theta)}{\nu(\theta)}d\theta, \qquad b = \int \theta\pi(\theta)\tfrac{\nu^{(3)}(\theta)}{\nu(\theta)}\,d\theta,$$

$$c = \int \theta\pi(\theta)\tfrac{\nu'(\theta)\nu''(\theta)}{\nu^2(\theta)}\,d\theta, \quad d = \int \theta\pi''(\theta)\tfrac{\nu'(\theta)}{\nu(\theta)}\,d\theta,$$

$$f = \int \pi'(\theta)\tfrac{\nu'(\theta)}{\nu(\theta)}\,d\theta, \qquad g = \int \pi(\theta)\left\{\tfrac{\nu'(\theta)}{\nu(\theta)}\right\}^2\,d\theta - \left\{\int \pi(\theta)\tfrac{\nu'(\theta)}{\nu(\theta)}d\theta\right\}^2.$$

Also, recall that $w_2 = \mathrm{var}_P\{h''(X)\}$ and $\ell_4 = \mathrm{E}_P\{h^{(4)}(X)\}$. In addition, $k(\pi)$, $k_1(\pi)$ and $K(\pi)$ used below are constants which can depend on π but not on ν.

PROPOSITION 1. *One has*

$$\begin{aligned}\mathrm{cov}_P(\theta,\delta_\nu) &= \mathrm{var}_\pi(\theta) + \frac{\sigma^2}{n}a + \frac{\sigma^4}{8n}(3w_2+\ell_4) + \frac{\sigma^4}{2n^2}(b-c+d+2f)\\ &\quad + \frac{5\sigma^6}{8n^2}(3w_2+\ell_4)a + \frac{\sigma^6}{4n^2}k(\pi) + O(n^{-5/2}).\end{aligned} \tag{2}$$

PROPOSITION 2. *One has*

$$\begin{aligned}\mathrm{var}_P(\delta_\nu) &= \mathrm{var}_\pi(\theta) + \frac{\sigma^2}{n}(1+2a) + \frac{\sigma^4}{8n}\mathrm{var}_\pi(\theta)(3w_2+\ell_4) + \frac{\sigma^4}{n^2}(b-c+d+g)\\ &\quad + \frac{5\sigma^6}{4n^2}(3w_2+\ell_4)a + \frac{\sigma^6}{4n^2}k_1(\pi) + O(n^{-5/2}).\end{aligned} \tag{3}$$

These two results now lead to the desired approximation to $\rho_P(\theta,\delta_\nu)$ stated below.

PROPOSITION 3. $\rho_P(\theta, \delta_\nu) = \hat{\rho}_P(\theta, \delta_\nu) + O(n^{-5/2})$, *where*

$$\begin{aligned}\hat{\rho}_P(\theta, \delta_\nu) &= 1 - \frac{\sigma^2}{2n}\left\{\frac{\sigma^2}{8}(3w_2 + \ell_4) - \frac{1}{\mathrm{var}_\pi(\theta)}\right\} + \frac{\sigma^4}{n^2}\left\{\frac{f - \frac{1}{2}g}{\mathrm{var}_\pi(\theta)} + \frac{a + \frac{1}{2}a^2}{V^2(\pi)}\right\} \\ &+ \frac{\sigma^4}{n^2 V^2(\pi)}\left[\frac{3}{8} + \frac{\sigma^2}{8}(3w_2 + \ell_4)\left\{2 - \frac{\sigma^2}{8}(3w_2 + \ell_4)\right\} + K(\pi)\right]. \quad (4)\end{aligned}$$

Proof. Use the definition of $\rho_P(\theta, \delta_\nu)$ and substitute the expressions (2) and (3) given above. Equation (4) will follow then based on simple algebra. Details are given in the Appendix.

We can now state the result describing the particular selected Bayes estimate $\delta_\nu(X)$.

THEOREM 2. *The estimate $\delta_\nu(X)$ maximizing $\hat{\rho}_P(\theta, \delta_\nu)$ is Bayes with respect to the prior density*

$$\nu(\theta) = c\pi(\theta)\exp\left\{-\frac{1}{2\tau^2}(\theta - \mu)^2\right\}, \quad (5)$$

where μ, τ^2 are arbitrary and c is a normalizing constant.

Proof. We will give the proof for the case $\mathrm{var}_\pi(\theta) = 1$. A minor modification works for $\mathrm{var}_\pi(\theta) \neq 1$. First, note that from (4) we would like to maximize $f - g/2 + a + a^2/2$ and hence minimize

$$g - 2f - 2a - a^2 = \mathrm{var}_\pi\left\{\frac{\nu'(\theta)}{\nu(\theta)}\right\} - 2\mathrm{cov}_\pi\left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}, \frac{\nu'(\theta)}{\nu(\theta)}\right\} - \mathrm{cov}_\pi^2\left\{\theta, \frac{\nu'(\theta)}{\nu(\theta)}\right\}.$$

Now, using the fact that $\mathrm{var}_\pi(\theta) = 1$, observe that

$$\mathrm{cov}_\pi\left\{\theta, \theta + \frac{\pi'(\theta)}{\pi(\theta)}\right\} = 0.$$

Next, write

$$\begin{aligned}&\mathrm{var}_\pi\left\{\frac{\nu'(\theta)}{\nu(\theta)}\right\} - 2\mathrm{cov}_\pi\left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}, \frac{\nu'(\theta)}{\nu(\theta)}\right\} - \mathrm{cov}_\pi^2\left\{\theta, \frac{\nu'(\theta)}{\nu(\theta)}\right\} \\ &= \mathrm{var}_\pi\left[\frac{\nu'(\theta)}{\nu(\theta)} - \left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right\}\right] - \mathrm{cov}_\pi^2\left[\theta, \frac{\nu'(\theta)}{\nu(\theta)} - \left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right\}\right] - \mathrm{var}_\pi\left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right\}\end{aligned}$$

At this point, make the important observation that since $\mathrm{var}_\pi(\theta) = 1$, by Schwartz's inequality,

$$\mathrm{var}_\pi\left[\frac{\nu'(\theta)}{\nu(\theta)} - \left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right\}\right] - \mathrm{cov}_\pi^2\left[\theta, \frac{\nu'(\theta)}{\nu(\theta)} - \left\{\theta + \frac{\pi'(\theta)}{\pi(\theta)}\right\}\right] \geq 0,$$

with equality if

$$\frac{\nu'(\theta)}{\nu(\theta)} - \theta - \frac{\pi'(\theta)}{\pi(\theta)} = a + b\theta. \quad (6)$$

The final step is to note that the solutions of the differential equation (6) are of the form (5). □

3.2. *Investigation of the selected Bayes estimate and examples.*

Suppose the reference prior density is $\pi(\theta) = e^{-|\theta|}/2$, a good middle ground between sharp and flat priors. Further suppose that the family Γ under consideration contains only symmetric priors and so $\nu(\theta)$ is of the form

$$\nu(\theta) = c\exp\left(-|\theta| - \frac{\theta^2}{2\tau^2}\right).$$

There is a value of τ (approximately 7.52) that gives the largest variance among all priors of this form. For our first illustration, this is the specific $\nu(\theta)$ we use.

Example 1. Let us take $X \sim N(\theta, 1)$. Under both $\pi(\theta)$ and $\nu(\theta)$, the marginal density of X can be found in closed form and hence the Bayes estimates $\delta_\pi(X)$ and $\delta_\nu(X)$ can also be found in closed form by using the familiar identity (Brown 1985):

$$\delta_\nu(x) = x + \frac{\sigma^2}{n}\frac{m'_\nu(x)}{m_\nu(x)},$$

where m_ν is the marginal density. Some selected values are reported in Table 1; as we have already noted, δ_ν results in a bit more shrinkage than δ_π.

TABLE 1: Values of $\delta_\pi(X)$ and $\delta_\nu(X)$ for various choices of $X = x$ in Example 1.

x	0	.5	1	1.5	2	3	5	8	10	15
$\delta_\pi(x)$	0	.241	.503	.806	1.161	2.026	4	7	9	14
$\delta_\nu(x)$	0	.238	.497	.795	1.144	1.992	3.931	6.878	8.844	13.758

Example 2. Consider $X \sim \mathrm{Logistic}(\theta, \sigma)$, with known σ and having density

$$f(x|\theta) = \frac{1}{\sigma}\exp\left(-\frac{x-\theta}{\sigma}\right)\left\{1 + \exp\left(-\frac{x-\theta}{\sigma}\right)\right\}^{-2}.$$

In this case it is not possible to obtain in closed form either the marginals or the Bayes estimates, but both $\delta_\pi(x)$ and $\delta_\nu(x)$ can be easily computed for any given x.

Again, some selected values are reported in Table 2 for two different values of σ, 0.5 and 1. As before, δ_ν receives a little more shrinkage than δ_π. Further, $\sigma = 1$ results in much more shrinkage than $\sigma = 0.5$, with the values for $\sigma = 0.5$ being closer to those in Example 1.

TABLE 2: Values of $\delta_\pi(X)$ and $\delta_\nu(X)$ for two values of σ and various choices of $X = x$ in Example 2.

σ	x	0	.5	1	1.5	2	3	5	8	10	15
0.5	$\delta_\pi(x)$	0	.276	.580	.922	1.301	2.147	4.030	7.002	9	14
	$\delta_\nu(x)$	0	.274	.574	.912	1.285	2.113	3.948	6.834	8.771	13.607
1.0	$\delta_\pi(x)$	0	.166	.336	.513	.698	1.095	1.970	3.386	4.357	6.820
	$\delta_\nu(x)$	0	.163	.329	.501	.680	1.059	1.864	3.038	3.722	4.938

From the numerical tables above, it seems that in cases where the reference prior is a double exponential, δ_π and δ_ν behave similarly. For another choice of the reference prior, this need not be the case.

To see this, consider the reference prior density $\pi(\theta) \propto (1+\theta^2/3)^{-2}$, density of the Student's t_3 prior which is a flat prior. Suppose again that the family Γ under consideration contains only symmetric priors and so $\nu(\theta)$ is of the form $\nu(\theta) = c(1+\theta^2/3)^2 \exp\{-\theta^2/(2\tau^2)\}$.

Example 3. Now consider $X \sim \text{Cauchy}(\theta, \sigma)$, with known σ and having density

$$f(x|\theta) = \frac{1}{\sigma\pi\left\{1 + \left(\frac{x-\theta}{\sigma}\right)^2\right\}}.$$

Some selected values are reported in Table 3 for $\sigma = 0.2$.

TABLE 3: Values of $\delta_\pi(X)$ and $\delta_\nu(X)$ for various choices of $X = x$ in Example 3.

x	0	.5	1	1.5	2	3	5	8	10	15
$\delta_\pi(x)$	0	.349	.687	1.002	1.284	1.735	2.197	2.216	2.065	1.645
$\delta_\nu(x)$	0	.348	.683	.993	1.267	1.685	1.976	1.60	1.15	.481

Note that, for small and moderate values of x, δ_π and δ_ν behave similarly, whereas for large values, δ_ν results in much more shrinkage than δ_π. But we would expect this because the penultimate ν has normal tails, whereas the reference π has very flat tails.

4. FURTHER POTENTIAL FOR PRACTICAL USES

4.1. *Selecting a default prior.*

An extensive literature exists on default prior Bayesian analysis; the literature includes much general theory and methods and applications of these to specific problems. There seem to be widely different opinions regarding appropriate definitions of default priors. We will say that a prior chosen from a specified class by a specified (and hopefully reasonable) selection rule is a default prior. Assessment of such a default prior is a separate issue and we will not address that here. Our intention is to show a potential use of the correlation approach in the default prior construction problem. The priors that result from our analysis here are constructed from an automatic and principled method, and in that sense, they are default priors. In addition, as we shall show, our proposal has a very distinct connection to the Fisher information. These are interesting consequences of our general approach and the Bickel prior arises as special from this development. Among the literature on default priors, particularly pertinent to our discussion are Cifarelli & Regazzini (1987), Clarke & Wasserman (1993), Ghosh & Mukerjee (1992), Datta & Ghosh (1995), and Kass &Wasserman (1996).

The approach we take is the following. A likelihood-based method will summarize the post-data opinion about θ by the likelihood function $f(x|\theta)$; a Bayesian method based on a given prior $\pi(\theta)$ will use the posterior density $\pi(\theta|x)$. Minimizing a suitable distance between these two summaries is a well accepted approach for the construction of default priors. We are proposing, instead, maximization of the correlation between $f(x|\theta)$ and $\pi(\theta|x)$ in the joint probability space.

Before we derive the results of this section for the location parameter case, let us look at an important case as an illustrative example for our suggested approach. This example will show that the general approach we are suggesting has the potential for producing standard default priors.

Example 4. Suppose $X \sim$ Bin (n, θ) and we wish to estimate θ. In the literature, various priors have been suggested as default priors for θ, the uniform and the Jeffreys prior included; see Berger (1986). It is well known that if $\theta \sim \mathcal{U}(0, 1)$, then, curiously, X has a marginal uniform distribution, too. Thus, $\pi(\theta|x) = (n+1)f(x|\theta)$, for all x and for all θ. We therefore have the curious result that the correlation between $f(x|\theta)$ and $\pi(\theta|x)$ in the joint probability space is 1 if θ has a uniform prior. *A fortiori*, the uniform prior is the default prior for θ according to our criterion just as long as the class of priors entertained includes the uniform prior. Thus, in the important binomial case, our general approach leads to a credible default prior. This is encouraging.

For ease of exposition, we shall only state the main results here. Major steps involved in the proofs will be outlined in the Appendix. However, many of the details are similar to those in the proof of Theorem 1 and are skipped. The complete details can again be found in Delampady *et al.* (1999). The set-up required in this section is similar, but somewhat weaker than that stated in Section 2. Specifically, it is enough to assume that the likelihood function is continuously differentiable five times with a bounded fifth derivative. With regard to the prior densities also, we can weaken Assumption A, and work with the following Assumption B.

ASSUMPTION B. The class Γ of prior densities under consideration consists of prior densities π which are three times continuously differentiable a.e., with bounded third derivative, and $\mathrm{E}_\pi(\theta) = 0$, $\mathrm{E}_\pi(\theta^2) < \infty$.

4.2. An expansion for correlation.

It is not possible to derive any analytical results by working with the exact marginal correlation $\rho_P\{f(x|\theta), \pi(\theta|x)\}$. We will present an expansion for the marginal correlation; in this expansion, the leading term is 1, and the second term is $-a_1\sigma^2 \cdot I(\pi)/n + a_2/n$, where $I(\pi)$ is the Fisher information of the prior π, while $a_1 > 0$ is an absolute constant, and a_2 depends only on f. Therefore, according to our correlation criterion, we propose to maximize this second term as a rule for selecting a default prior. This is formally similar to certain results in Clarke & Barron (1994) and Clarke & Wasserman (1993). As indicated in the Introduction, we thus end up minimizing the Fisher information of the prior, a well-known approach which has been adopted by other authors for different reasons altogether. This connection of our default prior methodology to minimization of Fisher information is interesting.

THEOREM 3. *Under Assumption B,*

$$\rho_P\{f(x|\theta), \pi(\theta|x)\} = 1 - \frac{\sigma^2 c_2 I(\pi)}{2n(c_1 - c_3^2)} - \frac{\sigma^4}{8n(c_1 - c_3^2)}\{c_1 w_2 + (c_1 - c_6)C_4\} + o(n^{-1}). \quad (7)$$

where $I(\pi)$ is the Fisher information functional, and c_1 to c_6 are the following constants:

$$c_1 = \int_{-\infty}^{\infty} \phi^3(z)dz = \frac{1}{2\pi\sqrt{3}}, \quad c_2 = \int_{-\infty}^{\infty} z^2\phi^3(z)dz = \frac{1}{6\pi\sqrt{3}}, \quad c_3 = \int_{-\infty}^{\infty} \phi^2(z)dz = \frac{1}{2\sqrt{\pi}},$$

$$c_4 = \int_{-\infty}^{\infty} z^2\phi^2(z)dz = \frac{1}{4\sqrt{\pi}}, \quad c_5 = \int_{-\infty}^{\infty} z^4\phi^2(z)dz = \frac{3}{8\sqrt{\pi}}, \quad c_6 = \int_{-\infty}^{\infty} z^4\phi^3(z)dz = \frac{1}{6\sqrt{3}\pi}.$$

Remark. The constants c_4 and c_5 appear in the derivation but ultimately disappear. Since $c_1 - c_3^2 = 0.01231 > 0$, we would want to minimize $I(\pi)$ in appropriate families Γ. A formal derivation of the expansion is outlined in the Appendix.

4.3. Illustrations.

Consider a general location parameter model (with parameter θ) which fits into our set-up. Then we have the following results.

Illustration 1. Suppose $|\theta| \leq 1$. Then from Bickel (1981) or Huber (1974), the following prior density achieves the minimum Fisher information in the class of *all* priors (i.e., now compactly supported on $[-1, 1]$ since this is the parameter space):

$$\pi(\theta) = \begin{cases} \cos^2(\pi\theta/2), & \text{if } |\theta| \leq 1; \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the Bickel prior is the default prior under our correlation criterion.

Illustration 2. Fix $\tau > 0$, and consider

$$\Gamma = \left\{ \pi : \pi \text{ is symmetric about 0 and } \int_{-\infty}^{\infty} \theta^2 \pi(\theta)\, d\theta = \tau^2 \right\}.$$

The prior which achieves the minimum Fisher information in this class is $N(0, \tau^2)$, (see Kagan, Linnik & Rao 1973). This class, however, is somewhat restricted since it excludes very heavy tailed priors such as the Cauchy and Student's t distribution with two degrees of freedom.

Illustration 3. Now take

$$\Gamma = \left\{ \pi : \pi(\theta) = \int \phi(\theta\rho^{-\gamma})\mu(d\gamma),\ \mu \text{ an arbitrary probability measure on } [-\infty, \infty] \right\}.$$

This is a useful collection of priors in robustness since it is the class of scale mixtures of normal priors, thus including heavy tailed distributions such as Student's t (Cauchy being a special case), and double exponential. Unfortunately, it is the case that (see Bickel & Collins 1983) the infimum Fisher information for this class is 0.

Illustration 4. Bickel & Collins (1983) modify the above class and consider instead the ϵ-contamination class (in some other context),

$$\Gamma = \left\{ \pi : \pi(\theta) = \epsilon\phi(\theta) + (1-\epsilon) \int_{[-\infty, \infty]} \phi(\theta\epsilon^{-\gamma})\mu(d\gamma),\ \mu \text{ arbitrary} \right\}.$$

They show that the prior π which minimizes the Fisher information in this class is given by π^*, where

$$\pi^*(\theta) = \epsilon\phi(\theta) + (1-\epsilon) \sum_{i=1}^{\infty} p_i \sigma_i^{-1} \phi(\sigma_i^{-1}\theta),$$

with $0 < \sigma_i < \infty, 0 < p_i < 1, \sum_{i=1}^{\infty} p_i = 1$. Identifying the p_i and σ_i, however, is a numerically challenging problem.

APPENDIX

Proof of Theorem 1. If $\nu(\theta)$ is any prior density, then the Bayes estimator of θ with respect to this prior is

$$\delta_\nu(x) = \frac{\int \theta \exp\{\mathcal{L}(\theta, x)\}\nu(\theta)\, d\theta}{\int \exp\{\mathcal{L}(\theta, x)\}\nu(\theta)\, d\theta} = \hat\theta + \frac{\int (\theta - \hat\theta) \exp\{\mathcal{L}(\theta, x)\}\nu(\theta)\, d\theta}{\int \exp\{\mathcal{L}(\theta, x)\}\nu(\theta)\, d\theta} = \hat\theta + R_n(x).$$

say. An asymptotic expansion is derived for R_n now. The proof sketched below can be modified and used in Propositions 1 and 2 and Theorem 3.

Expand ν in a Taylor Series around $\hat{\theta}$, noting that it is continuously differentiable four times, and that the 4th derivative is bounded. Also expand $\mathcal{L}$ in a Taylor Series around $\hat{\theta}$, and note that $\mathcal{L}^{(6)}(\theta, x)/n$ is uniformly bounded. Let ϕ denote the standard normal density and let $z = \sqrt{n}(\theta - \hat{\theta})/\sigma$. Then, using (1), we obtain, on Taylor expansions, and some tedious algebra,

$$R_n(x) = \frac{\sigma}{\sqrt{n}} \frac{\mathrm{num}(x)}{\mathrm{denom}(x)},$$

where

$$\begin{aligned}\mathrm{num}(x) = \int z\phi(z) &\left\{1 + \frac{\sigma^2}{2\sqrt{n}} W_n z^3 + \frac{\sigma^4}{8n} W_n^2 z^4 + \frac{\sigma^6}{48n^{3/2}} W_n^3 z^6 + O(n^{-2})\right\} \\ \times &\left\{1 + \frac{\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n} z^4 \mathcal{L}^{(4)} + \frac{\sigma^5}{120n^{3/2}} z^5 \frac{\mathcal{L}^{(5)}}{n} + O(n^{-2})\right\} \\ \times &\left\{\nu(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z\nu'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \nu''(\hat{\theta}) + \frac{\sigma^3}{6n^{3/2}} z^3 \nu^{(3)}(\hat{\theta}) + O(n^{-2})\right\} dz,\end{aligned}$$

and

$$\begin{aligned}\mathrm{denom}(x) = \int \phi(z) &\left\{1 + \frac{\sigma^2}{2\sqrt{n}} W_n z^3 + \frac{\sigma^4}{8n} W_n^2 z^4 + \frac{\sigma^6}{48n^{3/2}} W_n^3 z^6 + O(n^{-2})\right\} \\ \times &\left\{1 + \frac{\sigma^3}{6n} z^3 \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{\sigma^4}{24n^2} z^4 \mathcal{L}^{(4)} + \frac{\sigma^5}{120n^{3/2}} z^5 \frac{\mathcal{L}^{(5)}}{n} + O(n^{-2})\right\} \\ \times &\left\{\nu(\hat{\theta}) + \frac{\sigma}{\sqrt{n}} z\nu'(\hat{\theta}) + \frac{\sigma^2}{2n} z^2 \nu''(\hat{\theta}) \frac{\sigma^3}{6n^{3/2}} z^3 \nu^{(3)}(\hat{\theta}) + O(n^{-2})\right\} dz.\end{aligned}$$

Using the moments of the standard normal distribution and some more algebra, these expressions reduce to

$$\begin{aligned}\mathrm{num}(x) = \frac{\sigma}{\sqrt{n}} &\left[\nu'(\hat{\theta}) + \frac{\sigma^2}{2\sqrt{n}}\left\{\nu(\theta) \frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 3\nu'(\hat{\theta}) W_n\right\}\right. \\ &\left.+ \frac{\sigma^2}{n}\left\{\frac{\nu'''(\hat{\theta})}{2} + \frac{15\sigma^2}{8}\nu'(\theta)\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) + \frac{\sigma^2}{4}\nu(\theta)\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}} W_n\right)\right\}\right] + O(n^{-3/2}).\end{aligned}$$

and

$$\mathrm{denom}(x) = \nu(\hat{\theta}) + \frac{\sigma^2}{2\sqrt{n}} \nu(\theta) W_n + \frac{\sigma^2}{n}\left\{\frac{\nu''(\hat{\theta})}{2} + \frac{3\sigma^2 \nu(\hat{\theta})}{8}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\right\} + O(n^{-3/2}).$$

Therefore,

$$\begin{aligned}R_n(x) \;=\; &\frac{\sigma^2}{n}\left[\nu'(\hat{\theta}) + \frac{\sigma^2}{\sqrt{n}}\left\{\frac{\nu(\hat{\theta})}{2}\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + \frac{3\nu'(\hat{\theta})W_n}{2}\right\} + \frac{\sigma^2}{n}\left\{\frac{\nu'''(\hat{\theta})}{2}\right.\right. \\ &\left.\left.+\frac{15\nu'(\hat{\theta})\sigma^2}{8}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right) + \frac{\nu(\hat{\theta})\sigma^2}{4}\left(\frac{\mathcal{L}^{(5)}}{2n} + 5\frac{\mathcal{L}^{(3)}}{\sqrt{n}}W_n\right)\right\} + O(n^{-3/2})\right] \\ &\times \left[\nu(\hat{\theta}) + \frac{\sigma^2}{2\sqrt{n}} \nu(\hat{\theta}) W_n + \frac{\sigma^2}{2n}\left\{\nu''(\hat{\theta}) + \frac{3\sigma^2 \nu(\hat{\theta})}{4}\left(W_n^2 + \frac{\mathcal{L}^{(4)}}{3n}\right)\right\} + O(n^{-3/2})\right]^{-1}\end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{\sigma^2}{n}\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} + \frac{\sigma^4}{2n^{3/2}}\left\{\frac{\mathcal{L}^{(3)}}{\sqrt{n}} + 2\frac{\nu'(\hat\theta)}{\nu(\hat\theta)} W_n\right\} + \frac{\sigma^4}{2n^2}\left\{\frac{\nu'''(\hat\theta)}{\nu(\hat\theta)} - \frac{\nu'(\hat\theta)\nu''(\hat\theta)}{\nu^2(\hat\theta)}\right. \\
& \left. + \frac{\nu'(\hat\theta)}{\nu(\hat\theta)}\sigma^2\left(2W_n^2 - \frac{\mathcal{L}^{(4)}}{n}\right) + \sigma^2\left(\frac{\mathcal{L}^{(5)}}{4n} + 2\frac{\mathcal{L}^{(3)}}{\sqrt{n}} W_n\right)\right\} + O(n^{-5/2}).
\end{aligned}
$$

This yields the desired approximation for δ_ν.

Proof of Proposition 3. One has

$$
\begin{aligned}
\rho_P(\theta, \delta_\nu) \quad = \quad & \Big[\mathrm{var}_\pi(\theta) + \frac{\sigma^2}{n}\, a + \frac{\sigma^4}{8n}\,\mathrm{var}_\pi(\theta)(3w_2 + \ell_4) + \frac{\sigma^4}{2n^2}\,(b - c + d + 2f) \\
& + \frac{5\sigma^6}{8n^2}\,(3w_2 + \ell_4)a + \frac{\sigma^6}{4n^2}\,k(\pi)\Big] \\
& \times \Big[\mathrm{var}_\pi(\theta)\Big\{\mathrm{var}_\pi(\theta) + 2\frac{\sigma^2}{n}\, a + \frac{\sigma^2}{n} + \frac{\sigma^4}{8n}\,\mathrm{var}_\pi(\theta)(3w_2 + \ell_4) \\
& + \frac{\sigma^4}{n^2}\,(b - c + d + g) + \frac{5\sigma^6}{4n^2}\,(3w_2 + \ell_4)a + \frac{\sigma^6}{4n^2}k_1(\pi)\Big\}\Big]^{-1/2} + O(n^{-5/2}) \\
= \quad & \Big[1 + \frac{1}{\mathrm{var}_\pi(\theta)}\Big[\frac{\sigma^2}{n}\Big\{a + \frac{\sigma^2}{8}\,\mathrm{var}_\pi(\theta)(3w_2 + \ell_4)\Big\} \\
& + \frac{\sigma^4}{2n^2}\Big\{b - c + d + 2f + \frac{5\sigma^2}{4}(3w_2 + \ell_4)a\Big\} + \frac{\sigma^6}{4n^2}k(\pi)\Big]\Big] \\
& \times \Big[1 - \frac{1}{2\mathrm{var}_\pi(\theta)}\Big[\frac{\sigma^2}{n}\Big\{1 + 2a + \frac{\sigma^2}{8}\,\mathrm{var}_\pi(\theta)(3w_2 + \ell_4)\Big\} \\
& + \frac{\sigma^4}{n^2}\Big\{b - c + d + g + \frac{5\sigma^2}{4}(3w_2 + \ell_4)a\Big\} + \frac{\sigma^6}{4n^2}\,k_1(\pi)\Big] \\
& + \frac{3}{8V^2(\pi)}\Big[\frac{\sigma^2}{n}\,\Big\{1 + 2a + \frac{\sigma^4}{8}\mathrm{var}_\pi(\theta)(3w_2 + \ell_4)\Big\} \\
& + \frac{\sigma^4}{n^2}\Big\{b - c + d + g + \frac{5\sigma^2}{4}(3w_2 + \ell_4)a\Big\} + \frac{\sigma^6}{4n^2}k_1(\pi)\Big]\Big]^2 + O(n^{-5/2}) \\
= \quad & 1 + \frac{\sigma^2}{2n}\Big\{\frac{\sigma^2}{8}(3w_2 + \ell_4) - \frac{1}{\mathrm{var}_\pi(\theta)}\Big\} + \frac{\sigma^4}{n^2\mathrm{var}_\pi(\theta)}\Big\{f - \frac{g}{2} + \frac{a + \frac{1}{2}a^2}{\mathrm{var}_\pi(\theta)}\Big\} \\
& + \frac{\sigma^4}{n^2V^2(\pi)}\Big[\frac{3}{8} + \frac{\sigma^2(3w_2 - \ell_4)}{8}\Big\{2 - \frac{\sigma^2}{8}\,(3w_2 + \ell_4)\Big\} + K(\pi)\Big] + O(n^{-5/2}).
\end{aligned}
$$

Outline of the derivation of expansion (7) of Theorem 3. By definition, one has

$$
\rho_P\{f(x|\theta), \pi(\theta|x)\} = \frac{\mathrm{E}_P\{f(x|\theta)\pi(\theta|x)\} - \mathrm{E}_P\{f(x|\theta)\}\mathrm{E}_P\{\pi(\theta|x)\}}{\sqrt{[\mathrm{E}_P\{f^2(x|\theta)\} - \mathrm{E}_P^2\{f(x|\theta)\}][\mathrm{E}_P\{\pi^2(\theta|x)\} - E_P^2\{\pi(\theta|x)\}]}}\,.
$$

Proceeding as in the proof of Theorem 1, it can be shown that

$$
\begin{aligned}
\mathrm{E}_P\{f(x|\theta)\} \quad &= \quad \int\!\!\int f^2(x|\theta)\pi(\theta)d\theta dx = \int\!\!\int \exp\{2\mathcal{L}(\theta, x)\}\pi(\theta)d\theta dx \\
&= \quad \frac{\sqrt{n}}{\sigma}\left\{c_3 + \frac{\sigma^4}{2n}c_5\left(w_2 + \frac{\ell_4}{6}\right) + o(n^{-1})\right\}. \qquad (8)
\end{aligned}
$$

Similarly, the marginal density $m(x)$ admits the expansion

$$\frac{\sigma}{\sqrt{n}} \exp\{\mathcal{L}(\theta, x)\}\sqrt{2\pi}\pi(\theta) \left[1 - \frac{\sigma^3 W_n}{2\sqrt{n}} + \frac{\sigma^2}{2n} \left\{\frac{\pi''(\theta)}{\pi(\theta)} + \frac{\sigma^2}{4}\left(\frac{\mathcal{L}^{(4)}}{n} + 3W_n^2\right)\right\} + O(n^{-\frac{3}{2}})\right].$$

Consequently,

$$\mathrm{E}_P\{\pi(\theta|x)\} = \int\int \pi(\theta|x)f(x|\theta)\pi(\theta)d\theta dx = \int\int \frac{\exp\{2\mathcal{L}(\theta, x)\}\pi^2(\theta)}{m(x)}\, d\theta dx$$

$$= \frac{\sqrt{n}}{\sigma}\left[c_3 + \frac{\sigma^2}{n} c_4 I(\pi) + \frac{\sigma^4}{2n}\left\{(w_2\left(c_5 - c_4 - \frac{c_3}{4}\right) + \ell_4\left(\frac{c_5}{6} - \frac{c_3}{4}\right)\right\} + o(n^{-1})\right], \qquad (9)$$

where $I(\pi)$ is the Fisher information of π.

Analogously,

$$\mathrm{E}_P\{f(x|\theta)\pi(\theta|x)\}$$

$$= \frac{n}{\sigma^2}\left[c_1 + c_2\frac{\sigma^2}{n} I(\pi) + \frac{\sigma^4}{4n}\left\{w_2\left(\frac{9c_6}{2} - 3c_2 - \frac{c_1}{2}\right) + \ell_4\left(\frac{c_6}{2} - \frac{c_1}{2}\right)\right\} + o(n^{-1})\right],$$

and so

$$\mathrm{cov}_P\{f(x|\theta, \pi(\theta|x)\} = \frac{n}{\sigma^2}(c_1 - c_3^2) + (c_2 - c_3c_4)I(\pi)$$

$$+\frac{\sigma^2}{2}\left\{w_2\left(\frac{9}{4}c_6 - \frac{3}{2}c_2 - \frac{c_1}{4} - 2c_3c_5 + c_3c_4 + \frac{c_3^2}{4}\right) + \ell_4\left(\frac{c_6}{2} - \frac{c_1}{2} - \frac{c_3c_5}{3} + \frac{c_3^2}{4}\right)\right\} + o(1). \qquad (10)$$

One can also show that

$$\mathrm{E}_P\{f^2(x|\theta)\} = \frac{n}{\sigma^2}\left\{c_1 + \frac{\sigma^4}{8n}c_6(9w_2 + \ell_4) + o(n^{-1})\right\}. \qquad (11)$$

and that

$$\mathrm{E}_P\{\pi^2(\theta|x)\} = \int\int \pi^2(\theta|x)f(x|\theta)\pi(\theta)\, d\theta dx = \int\int \frac{f^3(x|\theta)\pi^3(\theta)}{m^2(x)}\, d\theta dx$$

$$= \frac{n}{\sigma^2}\left[c_1 + 3c_2\frac{\sigma^2}{n} I(\pi) + \frac{\sigma^4}{2n}\left\{w_2\left(\frac{9}{4}c_6 - 3c_2\right) + \ell_4\left(\frac{c_6}{4} - \frac{c_1}{2}\right)\right\} + o(n^{-1})\right]. \qquad (12)$$

In view of equations (8) and (11), we thus have

$$\mathrm{var}_P\{f(x|\theta)\} = \frac{n}{\sigma^2}(c_1 - c_3^2) + \sigma^2\left\{w_2\left(\frac{9}{8}c_6 - c_3c_5\right) + \ell_4\left(\frac{c_6}{8} - \frac{c_3c_5}{6}\right)\right\} + o(1), \qquad (13)$$

and by equations (9) and (12),

$$\mathrm{var}_P\{\pi(\theta|x)\} = \frac{n}{\sigma^2}(c_1 - c_3^2) + (3c_2 - 2c_3c_4)I(\pi)$$

$$+\frac{\sigma^2}{2}\left\{w_2\left(\frac{9}{4}c_6 - 3c_2 - 2c_3c_5 + 2c_3c_4 + \frac{c_3^2}{2}\right) + \ell_4\left(\frac{c_6}{4} - \frac{c_1}{2} - \frac{c_3c_5}{3} + \frac{c_3^2}{2}\right)\right\} + o(1). \qquad (14)$$

Combining relations (10), (13), and (14), after several steps,

$$\rho_P(f(x|\theta), \pi(\theta|x)) = 1 - \frac{\sigma^2}{2n(c_1 - c_3^2)}c_3 I(\pi) - \frac{\sigma^4}{8n(c_1 - c_3^2)}\{c_1w_2 + (c_1 - c_6)\ell_4\} + o(n^{-1}),$$

which concludes the description.

ACKNOWLEDGEMENTS

The authors are grateful to the former Editor, an Associate Editor and the referees for their comments on previous versions of this work. This research was supported in part by the National Science Foundation and the National Security Agency.

REFERENCES

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

J. O. Berger (1994). An overview of robust Bayesian analysis. *Test*, 3, 5–59.

P. J. Bickel (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *The Annals of Statistics*, 9, 1301–1309.

P. J. Bickel & J. R. Collins (1983). Minimizing Fisher information over mixtures of distributions. *Sankhyā Series A*, 45, 1–19.

L. D. Brown (1971). Admissibility, recurrent diffusions, and insoluble boundary value problems. *The Annals of Mathematical Statistics*, 42, 855–903. (Correction *The Annals of Statistics*, 1, 594.)

L. D. Brown (1985). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. IMS Lecture Notes-Monograph Series, vol. 9, Institute of Mathematical Statistics, Hayward, CA.

D. M. Cifarelli & E. Regazzini (1987). Priors for exponential families which maximize the association between past and future observations. In *Probability and Bayesian Statistics: Papers from the International Symposium held in Innsbruck, Austria, September 23–26, 1986* (R. Viertl, ed.), Plenum, London, pp. 83–95.

B. S. Clarke & A. R. Barron (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference*, 41, 37–60.

B. S. Clarke & L. A. Wasserman (1993). Noninformative priors and nuisance parameters. *Journal of the American Statistical Association*, 88, 1427–1432.

A. DasGupta, G. Casella, M. Delampady, C. Genest, H. Rubin & W. E. Strawderman (2000). Correlation in a Bayesian framework. *The Canadian Journal of Statistics*, 28, 675–687.

G. S. Datta & J. K. Ghosh (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*, 82, 37–45.

M. Delampady, A. DasGupta, G. Casella, H. Rubin & W. E. Strawderman (1999). *A New Approach to Default Priors and Robust Bayes Methodology*. Technical Report Number 99–27, Department of Statistics, Purdue University, West Lafayette, IN.

B. Efron (1986). Why isn't everyone a Bayesian (with discussion). *The American Statistician*, 40, 1–11.

J. K. Ghosh (1994). *Higher Order Asymptotics*. IMS Lecture Notes-Monograph Series, vol. 4, Institute of Mathematical Statistics, Hayward, CA.

J. K. Ghosh & R. Mukerjee (1992). Non-informative priors. In *Bayesian Statistics 4: Proceedings of the 4th Valencia International Meeting held in Peniscola, Spain, April 15–20, 1991: Dedicated to the Memory of Morris H. DeGroot, 1931–1989* (J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds.), Oxford University Press, pp. 195–210.

J. K. Ghosh, B. K. Sinha & S. N. Joshi (1982). Expansions for posterior probability and integrated Bayes risk. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta & J. O. Berger, eds.), Academic Press, New York, pp. 403–456.

P. J. Huber (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35, 73–101.

P. J. Huber (1974). Fisher information and spline interpolation. *The Annals of Statistics*, 2, 1029–1034.

R. A. Johnson (1970). Asymptotic expansions associated with posterior distributions. *The Annals of Mathematical Statistics*, 41, 851–864.

A. M. Kagan, Yu. V. Linnik & C. R. Rao (1973). *Characterization Problems in Mathematical Statistics*. Wiley, New York.

R. E. Kass & L. A. Wasserman (1996). Selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91, 1343–1370.

B. Ja. Levit (1979). On the theory of the asymptotic minimax property of second order. *Theory of Probability and its Applications*, 24, 435–437.

B. Ja. Levit (1981). On asymptotic minimax estimates of the second order. *Theory of Probability and its Applications*, 25, 552–568.

D. V. Lindley (1961). The use of prior probability distributions in statistical inference and decisions. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, vol. 1, pp. 453–468.

J. Perez (1998). *Development of Expected Posterior Prior Distributions for Model Comparison*. Doctoral dissertation, Purdue University, West Lafayette, IN.

C. Stein (1982). *The Coverage Probabilities of Confidence Sets Based on a Prior Distribution*. Mimeograph.

M.-M. Zen & A. DasGupta (1993). Estimating a binomial parameter: is robust Bayes real Bayes? *Statistics & Decisions*, 11, 37–60.

Mohan DELAMPADY: mohan@isibang.ac.in
Indian Statistical Institute, 8th Mile, Mysore Road
R. V. College Post Office, Bangalore 560 059, India

Anirban DasGUPTA: dasgupta@stat.purdue.edu
Department of Statistics, Purdue University
West Lafayette, IN 47907-1399, USA

George CASELLA: casella@stat.ufl.edu
Department of Statistics, University of Florida
PO Box 118545, Gainesville, FL 32611-8545, USA

Herman RUBIN: hrubin@stat.purdue.edu
Department of Statistics, Purdue University
West Lafayette, IN 47907-1399, USA

William E. STRAWDERMAN: straw@stat.rutgers.edu
Statistics Center, Rutgers University
110 Frelinghuysen Road, Piscataway, NJ 08854, USA