Test (1997)
Vol. 6, *No.* 1, *pp.* 159–186

Reference Priors in Multiparameter Nonregular Cases

S. GHOSAL
Division of Theoretical Statistics and Mathematics
Indian Statistical Institute, 203 B. T. Road, Calcutta 700035, India

SUMMARY

The reference prior in the sense of Bernardo is derived in some multiparameter nonregular cases. The family of densities we consider have discontinuities at some points which depend on one component of the parameter (say, θ) while, for fixed values of θ, the family is regular with respect to the other components (say, φ). We obtain the reference prior through an asymptotic expansion of Lindley's measure of information. The expansion is in itself of some importance. The results are illustrated using examples.

Keywords: ASYMPTOTIC EXPANSION; BAYES RISK, DISCONTINUOUS DENSITIES; KULLBACK-LEIBLER NUMBER; MULTIDIMENSIONAL PARAMETER; POSTERIOR DISTRIBUTION; REFERENCE PRIOR.

1. INTRODUCTION

In a pioneering paper, Bernardo (1979) introduced the concept of a reference prior which, to date, seems to represent most sutably the idea of a noninformative prior. The reference prior can be thought of as a reference point so that anyone with a subjective prior belief can judge where he or she stands. From an objective Bayesian viewpoint, the Bayes procedure corresponding to this (often improper) prior can be used for inferential purposes. These priors usually have the additional appealing property that their Bayes procedures match with some standard frequentist procedures up to a certain order; for example, see Ghosh

Received October 1995; Revised July 1996.

(1994). Since Bernardo's (1979) original paper, there has been much further development of this idea in the literature, notably by Berger and Bernardo (1989, 1992a, 1992b, 1992c), Berger, Bernardo and Mendoza (1989), Ghosh and Mukerjee (1992), Ye and Berger (1991), Yang and Berger (1994) and Sun and Ye (1995).

The reference prior can be described as follows. Lindley's (1956) measure of information about the parameter in the sample is defined as the expected Kullback-Leibler divergence between the posterior and the prior. This is the same as the Bayes risk when estimating the density with entropy as the loss function, see Aitchison (1975). Thus, the larger the measure, the more informative the data and, hence, the less informative, the prior. Under this interpretation, the prior which maximizes this measure can be taken as a reasonable choice of noninformative prior. However, direct maximization is difficult as well as uninteresting since the maximizer involves the sample size. Bernardo (1979) proposed maximization of this measure in an asymptotic sense. Based on essentially heuristic arguments, Bernardo (1979) concluded that Jeffreys' prior is the reference prior for smooth families when the whole parameter is the parameter of interest.

To make Bernardo's (1979) arguments rigorous, we need to obtain an asymptotic expansion of Lindley's measure. The leading term in this expansion turns out to be of the order of the logarithm of the sample size and is independent of the choice of the prior. The next term is a functional of the prior and is independent of the sample size. Maximization of this term with respect to the prior yields the reference prior. The remaining term converges to zero as the sample size increases. For the case of smooth densities involving independent and identically distributed (i.i.d.) observations, such a rigorous expansion has been obtained by Clarke and Barron (1994) and earlier by Ibragimov and Khas'minskii (1973). Some extensions to the non-i.i.d. case have been investigated by Polson (1988). Expansion of the Bayes risk for the unsmooth case of a one-parameter family of discontinuous densities has been considered by Ghosal and Samanta (1997). As in the regular cases, this expansion naturally leads to the reference prior.

In this paper, an asymptotic expansion of the Bayes risk under the entropy loss is obtained for a class of nonregular cases indexed by a multidimensional parameter. The family of densities exhibit nonregularity in the form of discontinuities at locations depending on one parameter (θ say) while otherwise being a regular family with respect to the remaining component of the parameter (φ say) if θ is held fixed; here ϕ can be vector-valued but θ is always a scalar. A simple example of such a family is the shifted exponential scale family $f(x;\theta,\varphi) = \varphi^{-1}\exp[-(x-\theta)/\varphi]$, $x \geq \theta$, $\varphi > 0$. Our treatment proceeds in a similar way to Clarke and Barron (1994) and Ghosal and Samanta (1997), although our case is more complicated due to the presence of two types of parameters.

The organization of the paper is as follows. In Section 2, the problem is defined and the regularity conditions stated. In Section 3, the main result relating to the asymptotic expansion of Lindley's measure is proved (Theorem 3.1). The proof is somewhat long and is divided into several lemmas for reasons of clarity of presentation. Proof of the auxiliary lemmas are given in the Appendix. In Theorem 3.2, an information theoretic version of the convergence of the posterior distribution is obtained. The reference prior is obtained in Section 4 and is illustrated using several examples.

We now introduce some notation which will be adopted. The indicator of a set is denoted by **I**. We use bold face to denote vectors and matrices except in the subscripts. However, if in some special situation, a vector or a matrix turns out to be a scalar, then the same letter in non-bold form is used. The transpose of a vector $\boldsymbol{v}$ is denoted by $\boldsymbol{v}^T$. We denote the d-variate normal density with mean $\boldsymbol{\mu}$ and dispersion matrix Σ by $\phi_d(\cdot;\boldsymbol{\mu},\Sigma)$. Convergence in probability and in law are indicated by $\to_p$ and $\to_d$ respectively.

2. REGULARITY CONDITIONS

Consider i.i.d. real-valued observations with a distribution $P_{\theta,\varphi}$ indexed by two parameters $\theta \in \Theta$ and $\phi \in \Phi$, where $\Theta \subset \mathbb{R}$ is an open interval and Φ is an open subset of $\mathbb{R}^d$. We assume that $P_{\theta,\varphi}$ admits a density $f(\cdot;\theta,\phi)$ with respect to the Lebesgue measure on $\mathbb{R}$. The family $\{f(\cdot;\theta,\phi) : \theta \in \Theta, \phi \in \Phi\}$ is regular with respect to ϕ and nonregular with respect to θ in the sense that $x \mapsto f(x;\theta,\phi)$ is discontinuous at some points which depend on θ only, whereas for fixed θ, the family $\{f(\cdot;\theta,\phi) : \phi \in \Phi\}$ is regular. Asymptotic properties relating to estimation theory for such families have been studied in Ghosal and Samanta

(1995) (henceforth, abbreviated as GS) using the theory developed by Ibragimov and Has'minskii (1981) (hereafter IH). See also Smith (1985) in this context. For the expansion of the Bayes risk under entropy loss and to obtain the reference prior, we consider a special (but important) subclass described by the assumptions (A1)–(A10) below. The reason for considering this subclass will be given later.

(A1) Uniformly on compact subsets of $\Theta \times \Phi$,

$$\inf_{\|(u,\boldsymbol{v})\|>\varepsilon} \int |f^{1/2}(x;\theta+u,\phi+\boldsymbol{v}) - f^{1/2}(x;\theta,\phi)|^2 dx > 0, \quad \varepsilon > 0.$$

(A2) The density $f(x;\theta,\phi)$ is supported on an interval $S(\theta) := [a_1(\theta), a_2(\theta)]$. Situations where of $a_1(\theta) \equiv -\infty$ or $a_2(\theta) \equiv \infty$ are allowed although are no permitted simultaneously. If an endpoint is not infinite, the function $x \mapsto f(x;\theta,\phi)$ has a discontinuity at that point. If x belongs to the interior of $S(\theta)$, then $f(x;\theta,\phi)$ is continuously differentiable in (θ,ϕ).

(A3) The support $S(\theta)$ is either increasing or decreasing. Further, the endpoints $a_1(\theta)$ and $a_2(\theta)$ are continuously differentiable in θ, provided they are finite.

(A4) For every $\theta \in \Theta$ and $\phi \in \Phi$, the following limits exist:

$$p(\theta,\phi) = \lim_{x \downarrow a_1(\theta)} f(x;\theta,\phi), \quad q(\theta,\phi) = \lim_{x \uparrow a_2(\theta)} f(x;\theta,\phi).$$

Moreover, convergence is uniform over compact subsets of $\Theta \times \Phi$ and the functions $p(\theta,\phi)$ and $q(\theta,\phi)$ are continuous in (θ,ϕ).

(A5) The functions defined by

$$J(\theta,\phi) = \int (g_\theta(x;\theta,\phi))^2 dx,$$
$$J_j(\theta,\phi) = \int g_\theta(x;\theta,\phi) g_{\varphi_j}(x;\theta,\phi) dx, \quad j = 1,\ldots,d,$$
$$J_{jk}(\theta,\phi) = \int g_{\varphi_j}(x;\theta,\phi) g_{\varphi_k}(x;\theta,\phi) dx, \quad j,k = 1,\ldots,d,$$

are (finite and) continuous in (θ,ϕ), where $g_\theta = \partial g/\partial\theta$, $g_{\varphi_j} = \partial g/\partial\varphi_j$ and $g = f^{1/2}$. Moreover, the matrix $((J_{jk}(\theta,\phi)))$ is positive definite.

(A6) The functions $a_1(\theta), a_2(\theta), a_1'(\theta), a_2'(\theta)$ (where prime denotes derivative), $J(\theta, \phi)$, $J_j(\theta, \phi)$, $J_{jk}(\theta, \phi)$, $j, k = 1, \ldots, d$, have a majorant which is the product of a polynomial in $|\theta|$ and an exponential function in $\|\phi\|$.

Further, if $F(\theta, \phi, R) = \sup\{f(x; \theta, \phi) : x \leq a_1(\theta) + R$ or $x \geq a_2(\theta) - R\}$, then $F(\theta, \phi, R)$ has a majorant which is the product of a polynomial in θ and R and an exponential function in $\|\phi\|$.

(A7) For any fixed (θ, ϕ), $\max\{(\partial/\partial\varphi_j) \log f(x; \theta, \phi) : j = 1, \ldots, d\}$ remains bounded as $x \downarrow a_1(\theta)$ and $x \uparrow a_2(\theta)$.

(A8) For some $C, \gamma > 0$,

$$\int f^{1/2}(x; \theta, \phi) f^{1/2}(x; \theta + u, \phi + v)\, dx \leq C|u|^{-\gamma} \exp[-\gamma\|v\|].$$

(A9) For $x \in (a_1(\theta), a_2(\theta))$, $\log f(x; \theta, \phi)$ is thrice continuously differentiable in (θ, ϕ). For any (θ, φ), there is a neighbourhood $N_{\theta,\varphi}$ and a $P_{\theta,\varphi}$-integrable function $H_{\theta,\varphi}$ such that

$$\sup_{(t, s) \in N_{\theta,\varphi}} \left| \frac{\partial^2}{\partial t^2} \log f(x; t, s) \right| \leq H_{\theta,\varphi}(x),$$

$$\max_{1 \leq j \leq d} \sup_{(t, s) \in N_{\theta,\varphi}} \left| \frac{\partial^2}{\partial t \partial s_j} \log f(x; t, s) \right| \leq H_{\theta,\varphi}(x),$$

$$\max_{1 \leq j, k \leq d} \sup_{(t, s) \in N_{\theta,\varphi}} \left| \frac{\partial^2}{\partial s_j \partial s_k} \log f(x; t, s) \right| \leq H_{\theta,\varphi}(x).$$

Furthermore, $E_{\theta,\varphi} H_{\theta,\varphi}(X_1)$ is continuous in (θ, φ).

Treatment of the two cases $S(\theta)$ increasing and $S(\theta)$ decreasing with θ, are similar and hence we consider the latter case only. Put $W_n = \min\{a_1^{-1}(X_{1:n}), a_2^{-1}(X_{n:n})\}$ with the convention that the corresponding term is ignored if any of the functions $a_1(\cdot)$ or $a_2(\cdot)$ are infinite. Note that W_n is $P_{\theta,\varphi}$-a.s. well-defined for large sample sizes, $W_n \uparrow \theta$ a.s. $[P_{\theta,\varphi}]$ and the likelihood at θ is positive if and only if $\theta \leq W_n$. Set $c(\theta, \varphi) = E_{\theta,\varphi}[(\partial/\partial\theta) \log f(X_1; \theta, \varphi)]$, i.e., $c(\theta, \varphi) = a_1'(\theta)p(\theta, \varphi) - a_2'(\theta)q(\theta, \varphi) > 0$.

Consider the variable $\sigma_n(\theta) = n(W_n - \theta)$. The following assumption will be used in one part of the main result.

(A10) For any compact subset $K \subset \Theta \times \Phi$, we have

$$\sup_{(\theta,\varphi)\in K} \sup_{n\geq 1} E_{\theta,\varphi}\sigma_n(\theta) < \infty.$$

The assumptions (A1)–(A8) ensure, by the results obtained in GS, that the conditions of Theorem I.10.2 of IH are satisfied with the limiting likelihood ratio (i.e., the weak limit of

$$Z_n(u,\boldsymbol{v}) = \prod_{i=1}^{n} \frac{f(X_i;\theta + u/n, \varphi + n^{-1/2}\boldsymbol{v})}{f(X_i;\theta,\varphi))}$$

given by

$$Z(u,\boldsymbol{v}) = \exp[uc(\theta,\varphi) + \boldsymbol{v}^T\Delta(\theta,\varphi) - \frac{1}{2}\boldsymbol{v}^T\boldsymbol{F}(\theta,\varphi)\boldsymbol{v}]\mathbb{I}\{u < \sigma(\theta)\}, \quad (2.1)$$

where $\sigma(\theta)$ has an exponential distribution with mean $1/c(\theta,\varphi)$, $\Delta(\theta,\varphi)$ is a d-variate normal variable with mean zero and dispersion matrix $\boldsymbol{F}(\theta,\varphi)$ defined by $\boldsymbol{F}(\theta,\varphi) = ((4J_{jk}(\theta,\varphi)))$ (see (A5)).

The above conditions are satisfied by many important families such as $f(x;\theta,\varphi) = \varphi\exp[-\varphi(x-\theta)]\mathbb{I}\{x > \theta\}$ (location-scale family corresponding to the exponential distribution) or $f(x;\theta,\varphi) = x^{\alpha-1}\exp[-\varphi x]\ \mathbb{I}\{x > \theta\}/\int_\theta^\infty y^{\alpha-1}\exp[-\varphi y]dy$ (a truncation of the gamma scale family) if φ lies in a compact subset of $(0,\infty)$. Otherwise, a logarithmic transformation of $\sigma = \log\varphi$ may be used. For examples of more general classes along with a discussion of these conditions, see GS. Assumption (A9) is a mild one and is satisfied in most of the examples. It is not difficult to see that

$$\sup_{(\theta,\varphi)\in K} \int_0^\infty P_{\theta,\varphi}(X_1 \in S(\theta + u))du < \infty \quad (2.2)$$

is sufficient to imply (A10) (see Proposition 2.2 of Ghosal and Samanta (1997)). The same arguments show that if (2.2) is modified to

$$\sup_{(\theta,\varphi)\in K} \int_0^\infty u^\delta P_{\theta,\varphi}\{X_1 \in S(\theta + u)\}du < \infty \quad \text{for some } \delta > 0, \quad (2.3)$$

then the following stronger version of (A10) holds.

(A10)′ The variable $\sigma_n(\theta)$ is uniformly $P_{\theta,\varphi}$-integrable, where the uniformity is in $n \geq 1$ and (θ, φ) belonging to compact sets.

The condition (A10)′ will be used only in Theorem 2.2. If Θ is bounded above, then (2.2) and (2.3) are trivially satisfied. In some important special cases, such as $a_1(\theta) \equiv \theta$ and $a_2(\theta) \equiv \infty$, the moment condition $\sup_{(\theta,\varphi)\in K} E_{\theta,\varphi}(X_1 - \theta)^{1+\delta} < \infty$ for some $\delta > 0$, implies (2.2) and (2.3) by Chebyshev's inequality.

Remark 2.1. Except for Proposition 3.1, in all following results (including the main result of Theorem 3.1), we restrict attention to compact subsets of the parameter space, although we will allow the compact subsets to grow eventually. Thus, for the purpose of obtaining expansions relevant to the derivation of the reference priors, the parameter space may be assumed compact. As a result conditions (A6), (A8) and (2.3) (and hence (A10) and (A10)′) are trivially satisfied. For the same reason, when we have a scale or scale-like parameter, the logarithmic transformation is unnecessary. This allows us to work with the original parametrization in the examples of Section 4.

3. EXPANSION OF BAYES RISK FOR ENTROPY LOSS

Lindley's measure of information contained in the data is defined as the expected Kullback-Leibler divergence between the posterior and the prior, where it is assumed that the prior is proper and admits a density. We write this as $I(\pi; \boldsymbol{X}^n) = H(\pi) - H_{X^n}(\pi)$, where

$$H(\pi) = -\int \pi(\theta, \varphi) \log \pi(\theta, \varphi) d\theta\, d\phi$$

is the prior entropy and

$$H_{X^n}(\pi) = -E\left(\int \pi_n((\theta, \varphi)|\boldsymbol{X}^n) \log \pi_n((\theta, \varphi)|\boldsymbol{X}^n) d\theta\, d\phi\right)$$

is the expected posterior entropy; here and below $\pi_n(\cdot|\boldsymbol{X}^n)$ is the posterior obtained from the prior $\pi(\cdot)$.

The following Theorem is the main result of this paper.

Theorem 3.1. *Under conditions (A1)-(A10), for any prior* $\pi(\cdot)$ *which is positive, continuous and concentrated on a compact subset*

$K \subset \Theta \times \Phi$, we have, as $n \to \infty$,

$$I(\pi; \boldsymbol{X}^n) = \log \frac{n}{e} + \frac{d}{2} \log \frac{n}{2\pi e} + \int_K \pi(\theta, \varphi) \log \left(\frac{c(\theta, \varphi) l(\theta, \varphi)}{\pi(\theta, \varphi)} \right) d\theta \, d\phi + o(1),$$

where $l(\theta, \varphi) = (\det \boldsymbol{F}(\theta, \varphi))^{1/2}$.

The theorem will be proved by establishing the following two complementary results.

$$\liminf_{n \to \infty} \left[I(\pi; \boldsymbol{X}^n) - \log \frac{n}{e} - \frac{d}{2} \log \frac{n}{2\pi e} \right] \geq \int_K \pi(\theta, \varphi) \log \left(\frac{c(\theta, \varphi) l(\theta, \varphi)}{\pi(\theta, \varphi)} \right) d\theta \, d\phi, \tag{2.4}$$

and

$$\limsup_{n \to \infty} \left[I(\pi; \boldsymbol{X}^n) - \log \frac{n}{e} - \frac{d}{2} \log \frac{n}{2\pi e} \right] \leq \int_K \pi(\theta, \varphi) \log \left(\frac{c(\theta, \varphi) l(\theta, \varphi)}{\pi(\theta, \varphi)} \right) d\theta \, d\phi. \tag{2.5}$$

The former result is the simplest to establish and we address it first. The following lemma concerning the maximization of entropy is easy to prove from the information inequality.

Lemma 3.1. *Let f be a density on $(-\infty, 0] \times \mathbb{R}^d$ satisfying*

$$\int x f(x, \boldsymbol{y}) dx \, d\boldsymbol{y} = -\mu, \qquad \int \boldsymbol{y} f(x, \boldsymbol{y}) dx \, d\boldsymbol{y} = 0$$

and

$$\int \boldsymbol{y}\boldsymbol{y}^T f(x, \boldsymbol{y}) dx \, d\boldsymbol{y} = \Sigma.$$

Then,

$$-\int f(x, \boldsymbol{y}) \log f(x, \boldsymbol{y}) dx \, d\boldsymbol{y} \leq 1 + \log \mu + (d/2)(1 + \log(2\pi)) + (1/2) \log \det \Sigma.$$

Hereafter, $p^n(\boldsymbol{x}^n;\theta,\boldsymbol{\phi})$ will denote $\prod_{i=1}^n f(x_i;\theta,\varphi)$ and $m_n(\boldsymbol{x}^n)$ will denote the marginal density of $\boldsymbol{X}^n$, i.e.,

$$m_n(\boldsymbol{x}^n) = \int \prod_{i=1}^{n} p^n(\boldsymbol{x}^n; t, \boldsymbol{s})\pi(t, \boldsymbol{s})dt\, d\boldsymbol{s}.$$

Furthermore, we let $\boldsymbol{g}_\varphi = (g_{\varphi_1},\dots,g_{\varphi_d})^T$ and

$$\Delta_n(\theta,\varphi) = 2n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_\varphi(X_i;\theta,\varphi).$$

Note that $\Delta_n(\theta,\varphi)$ is asymptotically $N_d(\boldsymbol{0}, \boldsymbol{F}(\theta,\varphi))$.

Proposition 3.1. *Fix (θ,φ) and let $(t,\boldsymbol{s})$ denote the random parameter. If* (A1)–(A8) *are satisfied, then*

(i) $\log\left(\frac{m_n(\boldsymbol{X}^n)}{p^n(\boldsymbol{X}^n;\theta,\boldsymbol{\phi})}\right) + (1 + d/2)\log n - \frac{d}{2}\log(2\pi) + \log\left(\frac{c(\theta,\varphi)l(\theta,\varphi)}{\pi(\theta,\varphi)}\right) - c(\theta,\varphi)\sigma_n(\theta) - \frac{1}{2}\Delta_n^T(\theta,\varphi)$ $\boldsymbol{F}^{-1}(\theta,\varphi)\Delta_n(\theta,\varphi) = O_p(1)$.

(ii) $n(E(t|\boldsymbol{X}^n) - W_n) \to_p -1/c(\theta,\varphi)$, $n\,\mathrm{cov}(\boldsymbol{s}|\boldsymbol{X}^n) \to_p$ $\boldsymbol{F}^{-1}(\theta,\varphi)$.

Proof of (3.2). Put $u^* = n(t - W_n)$ and $\boldsymbol{v}^* = n^{1/2}(\boldsymbol{s} - \tilde{\boldsymbol{\phi}}_n)$ where, as before, $(t,\boldsymbol{s})$ denotes the random parameter and $\tilde{\boldsymbol{\phi}}_n = E(\boldsymbol{s}|\boldsymbol{X}^n)$. The posterior density of $(u^*,\boldsymbol{v}^*)$ is given by $\pi_n^*(u^*,\boldsymbol{v}^*) = n^{-(1+d/2)}\pi_n(W_n - u^*/n, \tilde{\boldsymbol{\phi}}_n + n^{-1/2}\boldsymbol{v}^*)$ and has support in $(-\infty,0]\times\mathbb{R}^d$. Thus by Lemma 3.1, we have

$$\begin{aligned}
I(\pi,\boldsymbol{X}^n) &= -\int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi)d\theta\, d\boldsymbol{\phi} \\
&\quad + \int m_n(\boldsymbol{x}^n)\int_K \pi_n(\theta,\varphi)\log\pi_n(\theta,\varphi)d\theta\, d\boldsymbol{\phi}\, d\boldsymbol{x}^n \\
&= -\int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi)d\theta\, d\boldsymbol{\phi} \\
&\quad + \int m_n(\boldsymbol{x}^n)\int_{(-\infty,0]\times\mathbb{R}^d} \pi_n(u^*,\boldsymbol{v}^*)[(1 + d/2)\log n \\
&\quad + \log\pi_n(u^*,\boldsymbol{v}^*)]du^*\, d\boldsymbol{v}^*\, d\boldsymbol{x}^n
\end{aligned}$$

and, therefore,

$$I(\pi, \boldsymbol{X}^n) \geq -\int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi)d\theta\, d\phi + (1+d/2)\log n$$
$$-\int m_n(\boldsymbol{x}^n)[1+\log E(-u^*|\boldsymbol{x}^n)$$
$$+\frac{d}{2}(1+\log(2\pi)) + \frac{1}{2}\log\det\operatorname{cov}(\boldsymbol{v}^*|\boldsymbol{x}^n)]d\boldsymbol{x}^n$$

so that

$$I(\pi, \boldsymbol{X}^n) \geq -\int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi)d\theta\, d\phi + (1+d/2)(\log n - 1$$
$$-\frac{d}{2}\log(2\pi) - \int_K \pi(\theta,\varphi)E_{\theta,\varphi}[\log E(-u^*|\boldsymbol{X}^n)]d\theta\, d\phi$$
$$-\frac{1}{2}\int_K \pi(\theta,\varphi)E_{\theta,\varphi}[\log\det\operatorname{cov}(\boldsymbol{v}^*|\boldsymbol{X}^n)]d\theta\, d\phi. \quad (3.3$$

Note that $E(-u^*|\boldsymbol{X}^n) = \sigma_n(\theta) - n(E(t|\boldsymbol{X}^n) - \theta)$. By the gener results of IH (Ch. I, Sec. 5), $n(E(t|\boldsymbol{X}^n) - \theta)$ is uniformly integrab with respect to $P_{\theta,\varphi}$ with (θ,φ) varying over a compact set. So by (A1 we have $\sup_{(\theta,\varphi)\in K}\sup_{n\geq 1} E_{\theta,\varphi}E(-u^*|\boldsymbol{X}^n) < \infty$. This implies th $\log E(-u^*|\boldsymbol{X}^n)$ is uniformly (in n and (θ,φ) on compacts) integrab from above.

By Hadamard's inequality, we obtain $\log\det\operatorname{cov}(\boldsymbol{v}^*|\boldsymbol{x}^n) \leq \sum$ $\log(n\operatorname{var}(s_j|\boldsymbol{x}^n))$, where s_j is the jth component of $\boldsymbol{s}$. Now $E_{\theta,\varphi}(n$ v $(s_j|\boldsymbol{X}^n)) \leq E_{\theta,\varphi}E[(n^{1/2}(s_j-\theta_j))^2|\boldsymbol{X}^n] = E_{\theta,\varphi}(\int v_j^2\pi_n(u,\boldsymbol{v})du\, dv$ $j = 1,\ldots,d$, which is uniformly (in n and (θ,φ) on compacts) bounde again this is a consequence of the general results of IH (Ch. I, Sec. 5 Thus $\log\det(n\operatorname{cov}(\boldsymbol{v}^*|\boldsymbol{X}^n))$ is uniformly integrable from above.

These two observations and (3.3), together with part (ii) of Propo sition 3.1, imply that

$$\liminf_{n\to\infty}[I(\pi;\boldsymbol{X}^n) - (1+d/2)(\log n - 1)]$$
$$\geq -\int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi)d\theta\, d\phi - \frac{d}{2}\log(2\pi) \quad (3.4$$
$$+\int_K \pi(\theta,\varphi)\log[c(\theta,\varphi)I(\theta,\varphi)]d\theta\, d\phi,$$

which is equivalent to (3.1).

We now prove (3.2).

For a fixed (θ, φ), we set

$$R_n(\theta, \varphi) = -\log(m_n(\boldsymbol{X}^n)/p^n(\boldsymbol{X}^n; \theta, \varphi)) - (1 + d/2)\log n,$$
$$\psi_n(\theta, \varphi) = K(P^n_{\theta,\varphi}, m_n) - (1 + d/2)\log n$$
$$\psi(\theta, \varphi) = \log(c(\theta, \varphi)l(\theta, \varphi)/\pi(\theta, \varphi)) - (1 + d/2) - \frac{d}{2}\log(2\pi).$$

where $K(P^n_{\theta,\varphi}, m_n)$ is the Kullback-Leibler measure of divergence between $P^n_{\theta,\varphi}$ and m_n. Clearly

$$\psi_n(\theta, \varphi) = E_{\theta,\varphi}R_n(\theta, \varphi)$$

and

$$I(\pi; \boldsymbol{X}^n) = (1 + d/2)\log n + \int_K \psi_n(\theta, \varphi)\pi(\theta, \varphi)d\theta\, d\phi.$$

To prove (3.2), we therefore need to show that the following result holds, the proof of which we defer to the Appendix.

Lemma 3.2. *Under* (A1)–(A9), *we have*

(*a*) $\limsup_{n\to\infty} \psi_n(\theta, \varphi) \le \psi(\theta, \varphi)$.

(*b*) *On the compact set* K, $\psi_n(\theta, \varphi)$ *is uniformly dominated above by an integrable function.*

We then have

$$\begin{aligned}\limsup_{n\to\infty}&[I(\pi; \boldsymbol{X}^n) - (1 + d/2)\log n]\\ &= \limsup_{n\to\infty} \int_K \psi_n(\theta, \varphi)\pi(\theta, \varphi)d\theta\, d\phi\\ &\le \int_K \log(c(\theta, \varphi)l(\theta, \varphi)/\pi(\theta, \varphi))(\theta, \varphi)\pi(\theta, \varphi)d\theta\, d\phi\\ &\qquad - (1 + d/2) - \frac{d}{2}\log(2\pi),\end{aligned} \tag{3.5}$$

which is a re-statement of (3.2).

We conclude this section with an information theoretic version of the convergence of the posterior distribution. Proposition 3.1 implies that the posterior distribution of $(u^*, \boldsymbol{v}^*) = (n(t - W_n), n^{1/2}(\boldsymbol{s} - \tilde{\boldsymbol{\phi}}_n))$ converges in an L^1-sense, i.e.,

$$\begin{aligned}\int_{(-\infty,0]\times\mathbb{R}^d} |\pi_n(u^*, \boldsymbol{v}^*) - \exp[cu^*] \\ \phi_d(\boldsymbol{v}^*; 0, (\boldsymbol{F}(\theta,\varphi))^{-1})| du^*\, d\boldsymbol{v}^* \to_p 0,\end{aligned} \tag{3.6}$$

where $\tilde{\boldsymbol{\phi}}_n$ is the posterior mean of $\boldsymbol{\phi}$. From (3.6), it follows that

$$\begin{aligned}\lim_n \int_K \int \int_{(-\infty,W_n]\times\mathbb{R}^d} |\pi_n(t, \boldsymbol{s}) - nc(\theta,\varphi)\exp[nc(\theta,\varphi)(t - W_n)] \\ \times \phi_d(\boldsymbol{s}; \tilde{\boldsymbol{\phi}}_n, n^{-1}(\boldsymbol{F}(\theta,\varphi))^{-1})| dt\, d\boldsymbol{s}\, p^n(\boldsymbol{x}^n;\theta,\varphi) d\boldsymbol{x}^n\, \pi(\theta,\varphi) d\theta\end{aligned} \tag{3.7}$$

equals cero. We can make this result stronger as follows.

Theorem 3.2. *Under Assumptions* (A1)–(A9) *and* (A10)$'$, *we have*

$$\begin{aligned}\lim_n \int_K \int \int_{(-\infty,W_n]\times\mathbb{R}^d} \pi_n(t, \boldsymbol{s}) \log\left(\frac{\pi_n(t,\boldsymbol{s})}{\tilde{\pi}_n(t,\boldsymbol{s})}\right) dt\, d\boldsymbol{s} \\ \times p^n(\boldsymbol{x}^n;\theta,\varphi) d\boldsymbol{x}^n\, \pi(\theta,\varphi) d\theta\, d\boldsymbol{\phi} = 0,\end{aligned} \tag{3.8}$$

where

$$\tilde{\pi}_n(t, \boldsymbol{s}) = nc(\theta,\varphi) e^{nc(\theta,\varphi)(t-W_n)]} \phi_d(\boldsymbol{s}; \tilde{\boldsymbol{\phi}}_n, n^{-1}\boldsymbol{F}^{-1}(\theta,\varphi)).$$

Proof. The term inside the limit is

$$\begin{aligned}I(\pi; \boldsymbol{X}^n) + \int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi) d\theta\, d\boldsymbol{\phi} \\ - (1 + \frac{d}{2})\log n + \frac{d}{2}\log(2\pi) + A + B + C\end{aligned} \tag{3.9}$$

where

$$A = -\int_K c(\theta,\varphi)(E_{\theta,\varphi}\sigma_n(\theta))\pi(\theta,\varphi) d\theta\, d\boldsymbol{\phi}$$

$$B = \frac{1}{2}\int_K E_{\theta,\varphi}((\phi - \tilde{\phi}_n)^T (\boldsymbol{F}(\theta,\varphi))^{-1}(\phi - \tilde{\phi}_n))\pi(\theta,\varphi)d\theta\, d\phi$$

$$C = -\int_K \pi(\theta,\varphi)\log c(\theta,\varphi)d\theta\, d\phi - \int_K \pi(\theta,\varphi)\log l(\theta,\varphi)d\theta\, d\phi.$$

By (A10)$'$, $E_{\theta,\varphi}\sigma_n(\theta) = 1/c(\theta,\varphi) + o_p(1)$ and $E_{\theta,\varphi}((\phi - \tilde{\phi}_n)^T (\boldsymbol{F}(\theta,\varphi))^{-1}(\phi - \tilde{\phi}_n)) = d + o(1)$ uniformly in (θ,φ) belonging to compacts. Thus (3.9) can be written as

$$I(\pi;\boldsymbol{X}^n) - \log\frac{n}{e} - \frac{d}{2}\log\frac{n}{2\pi e} - \int_K \pi(\theta,\varphi) \log\left(\frac{c(\theta,\varphi)l(\theta,\varphi)}{\pi(\theta,\varphi)}\right)d\theta\, d\phi + o(1). \qquad (3.10)$$

By Theorem 3.1, the expression in (3.10) is $o(1)$, which completes the proof. $\triangleleft$

We now explain the reason for considering the subclass of densities studied by GS. Since the expansion leads to the convergence of the posterior distribution of a centered and normalized parameter, by the results of Ghosh *et al.* (1994), the limiting likelihood ratio must satisfy the necessary condition of their Theorem 2.4. The class of densities considered here is essentially the only one, among the class of all multiparameter nonregular families studied by GS, where the limiting likelihood ratio satisfies the afore mentioned condition.

4. REFERENCE PRIOR

Bernardo (1979) introduced the novel idea of a reference prior as a means of generating non-informative priors. The reference prior is perhaps the most suitable representation of a non-informative prior since it avoids some of the problems associated with those previously proposed; see Bernardo (1979) for a discussion of this. Non-informative priors are very appealing to both nonsubjective Bayesians and frequentists as the corresponding procedures match with some standard frequentist ones up to a certain order; see Ghosh (1994), Ghosh and Mukerjee (1992) and references therein. Furthermore, reference priors, and non-informative priors in general, can also be useful to subjective Bayesians as a particular subjective opinion may be elicited by thinking of the non-informative prior as a reference point.

Roughly speaking, Bernardo's (1979) reference prior is obtained by maximizing the expected Kullback-Leibler distance between the posterior and the prior (i.e., Lindley's measure of information) with respect to the prior. However, this maximization is done in an asymptotic sense as the sample size increases to infinity. To do this rigorously, one needs to obtain a one-term expansion of Lindley's measure as was done in Section 2. The dominating term of order $\log n$ is independent of the prior, and it is the second (constant) term which leads to the choice of a reference prior. More formally, the reference prior is defined as the (possibly improper) prior density π^* which is continuous and positive everywhere and whose restriction to any compact set K, after normalization, maximizes the functional

$$\int_K \pi(\theta, \varphi) \log \left(\frac{c(\theta, \varphi)(\det \boldsymbol{F}(\theta, \varphi))^{1/2}}{\pi(\theta, \varphi)} \right) d\theta\, d\phi. \tag{4.1}$$

It follows immediately that under our setup that the prior density

$$\pi^*(\theta, \varphi) = c(\theta, \varphi)(\det \boldsymbol{F}(\theta, \varphi))^{1/2} \tag{4.2}$$

is the unique reference prior for (θ, φ) modulo an irrelevant normalizing constant. In many cases, both $c(\theta, \varphi)$ and $\det \boldsymbol{F}(\theta, \varphi)$ factorize into functions of θ alone and ϕ alone. In those cases, θ and ϕ are a priori independently distributed according to the reference prior.

We note the following invariance property of the reference prior. Let $(\delta, \boldsymbol{\eta})$ be a reparametrization of (θ, φ), where $\delta = \delta(\theta)$ and $\boldsymbol{\eta} = \boldsymbol{\eta}(\varphi)$ are one-to-one smooth functions. Then the reference prior $\tilde{\pi}^*(\delta, \boldsymbol{\eta})$ of $(\delta, \boldsymbol{\eta})$ is related to the reference prior $\pi^*(\theta, \varphi)$ by

$$\tilde{\pi}^*(\delta, \boldsymbol{\eta}) = \left|\frac{d\theta}{d\delta}\right| \left|\det \left(\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\eta}}\right)\right| \pi^*(\theta(\delta), \boldsymbol{\phi}(\boldsymbol{\eta})), \tag{4.3}$$

where $\theta(\delta)$ and $\boldsymbol{\phi}(\boldsymbol{\eta})$ are the inverse transformations and $\frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\eta}}$ is the Jacobian matrix of the transformation $\boldsymbol{\phi} \mapsto \boldsymbol{\eta}$.

Often the parameters θ and ϕ are not of equal importance. For example, θ may be the parameter of interest while ϕ is the nuisance parameter. In such cases, the reference prior for (θ, φ) defined in (3.2) may not be appropriate. As argued in Bernardo, in this case one should look at the quantity

$$\begin{aligned} I(\pi(\theta), \boldsymbol{X}^n) &= E\left[\log\left(\frac{\pi_n(\theta|\boldsymbol{X}^n)}{\pi(\theta)}\right)\right] \\ &= I(\pi(\theta, \phi), \boldsymbol{X}^n) - \int \pi(\theta) I(\pi(\phi|\theta), \boldsymbol{X}^n) d\theta; \end{aligned} \tag{4.4}$$

here and below $\pi(\theta)$ is the marginal prior for θ, $\pi(\phi|\theta)$ is the conditional prior for ϕ given θ and so on. We restrict the parameter space of (θ, φ) to a compact subset of the form $K_1 \times K_2$. Note that for a given θ, the family is regular so that Lindley's measure inside the integral of the second term can be expanded as

$$\begin{aligned} I(\pi(\phi|\theta), \boldsymbol{X}^n) = \frac{d}{2} \log \frac{n}{2\pi e} + \int_{K_2} \pi(\phi|\theta) \log\left(\frac{l(\theta, \varphi)}{\pi(\phi|\theta)}\right) \\ + o(1); \end{aligned} \tag{4.5}$$

this follows from Clarke and Barron (1994). The $o(1)$ term in (4.5) depends on θ, but it follows from our assumptions that the $o(1)$ term is uniformly bounded as long as θ lies in a compact set. By a dominated convergence argument, it then follows that the integral of the $o(1)$ term inside is again $o(1)$. Thus, on simplification, it follows from (4.4) and (4.5) that

$$I(\pi(\theta), \boldsymbol{X}^n) = \log \frac{n}{e} + \int_{K_1} \pi(\theta) \log\left(\frac{\psi(\theta)}{\pi(\theta)}\right) + o(1), \tag{4.6}$$

where

$$\psi(\theta) = \exp\left[\int \pi(\phi|\theta) \log c(\theta, \varphi) d\phi\right]. \tag{4.7}$$

Hence, if we confine ϕ to a compact subset and $\pi(\phi|\theta)$ is given, then the (possibly improper) prior proportional to $\psi(\theta)$ should be taken as the marginal reference prior for θ. More precisely, this leads us to the following algorithm for obtaining the reference prior.

Procedure 1. This is in an adaptation of the method suggested by Berger and Bernardo (1989). Choose the reference prior for ϕ given

θ as $\pi^*(\phi|\theta) = (\det \boldsymbol{F}(\theta,\varphi))^{1/2}$. Choose a sequence $\Phi_1 \subset \Phi_2 \subset \cdots$ of compact subsets of Φ such that $\cup_{l=1}^{\infty}\Phi_l = \Phi$. Set $K_l(\theta) = (\int_{\Phi_l}(\det \boldsymbol{F}(\theta,\varphi))^{1/2}d\phi)^{-1}$ and $p_l(\phi|\theta) = K_l(\theta)\pi^*(\phi|\theta)\mathbb{I}\{\phi \in \Phi_l\}$. The marginal reference prior for θ at stage l is calculated as $\pi_l^*(\theta) = \exp\{\int_{\Phi_l} p_l(\phi|\theta)\log c(\theta,\varphi)d\phi\}$. Let θ_0 be a fixed point in Θ. Finally, the reference prior for (θ,φ), when ϕ is a nuisance parameters, is finally obtained as

$$\pi^*(\theta,\varphi) = \lim_{l\to\infty}\left[\frac{K_l(\theta)\pi_l^*(\theta)}{K_l(\theta_0)\pi_l^*(\theta_0)}\right]\pi^*(\phi|\theta), \tag{4.8}$$

provided the limit exists.

Instead of taking $\pi(\phi|\theta)$ as given, we may try to maximize the functional in (4.7) with respect to both $\pi(\theta)$ and $\pi(\phi|\theta)$, using an appropriate penalty term as suggested by Ghosh and Mukerjee (1992). We can then modify the procedure of Ghosh and Mukerjee (1992) to obtain the following algorithm.

Procedure 2. Maximize the functional

$$\int_K \pi(\theta,\varphi)\log c(\theta,\varphi)d\theta\, d\phi - l\int_K \pi(\theta,\varphi)\log\pi(\theta,\varphi)d\theta\, d\phi, \tag{4.9}$$

where l is a constant. For $l = 1$, this leads to the choice $\pi^*(\theta,\varphi) = c(\theta,\varphi)$.

If ϕ is the parameter of interest and θ is the nuisance parameter, both the above procedures are modified in the obvious manner, reversing the roles played by $c(\theta,\varphi)$ and $(\det \boldsymbol{F}(\theta,\varphi))^{1/2}$. The justification for this procedure is essentially the same except that we must use the result of Ghosal and Samanta (1997) for one-parameter nonregular cases (formally, $d = 0$) in place of the results of Clarke and Barron (1994). If the different components of ϕ are of varying importance (when φ is multidimensional), appropriate modifications may be suggested in analogy with the methods presented by Berger and Bernardo (1992a, 1992b, 1992c). The details are omitted here.

In the important special case where the factorizations

$$c(\theta,\phi) = c_1(\theta)c_2(\phi), \quad l(\theta,\varphi) = (\det \boldsymbol{F}(\theta,\varphi))^{1/2} = l_1(\theta)l_2(\phi),$$

hold, the Berger-Bernardo algorithm yields the prior $c_1(\theta)l_2(\phi)$, irrespective of the order of importance and the choice of compact sets. The prior then has the parametrization invariance property described in (4.3). Procedure 2, however, lacks this property. In Examples 4.1 and 4.2 below, $c(\theta, \varphi)$ and $l(\theta, \varphi)$ factorize in the above manner.

We illustrate the procedure for obtaining reference priors using three important examples.

Example 4.1. (Location-Scale Family). Let

$$f(x; \theta, \varphi) = \varphi^{-1} f_0\left(\frac{x-\theta}{\varphi}\right),$$

where $f_0(\cdot)$ is a density on $(0, \infty)$ with $f(0+) > 0$,

$$0 < \int \frac{(f_0'(x))^2}{f_0(x)}\, dx < \infty, \quad 0 < \int \frac{(1 + x f_0'(x))^2}{f_0(x)}\, dx < \infty.$$

In this case,

$$c(\theta, \varphi) = \varphi^{-1} f(0+), \quad F(\theta, \varphi) = \varphi^{-2} \int \frac{(1 + x f_0'(x))^2}{f_0(x)}\, dx.$$

Thus, the reference prior for (θ, φ) when both θ and φ are parameters of interest is $\pi^*(\theta, \varphi) = \varphi^{-2}$. Moreover, if either θ is the parameter of interest and φ is the nuisance parameter, or if φ is the parameter of interest and θ is the nuisance parameter, then both procedures yield the prior φ^{-1}.

Note that if X has density f_0, then $(\varphi x + \theta)$ has density $f(x; \theta, \varphi)$. Thus, the parameter space may be identified with the group of affine transformations $x \mapsto (\varphi x + \theta)$. The measure $\varphi^{-2} d\theta d\varphi$ is the left invariant Haar measure while the right invariant Haar measure is given by $\varphi^{-1} d\theta d\varphi$. It is well known that the right Haar measure is a very attractive prior with many desirable properties; see Chang and Eaves (1990). That the procedures lead to the choice of the right Haar measure is, therefore an attractive feature.

Example 4.2. Consider the family of Weibull distributions with known shape parameter α, unknown scale parameter φ and truncated at the left at some unknown point θ. Thus, we have observations from the density

$$f(x;\theta,\varphi) = \frac{x^{\alpha-1}\exp[-\varphi^{\alpha}x^{\alpha}]}{\int_{\theta}^{\infty} y^{\alpha-1}\exp[-\varphi^{\alpha}y^{\alpha}]dy}, \quad x > \theta \tag{4.10}$$
$$= \alpha\varphi^{\alpha}x^{\alpha-1}\exp[-\varphi^{\alpha}(x^{\alpha}-\theta^{\alpha})], \quad x > \theta.$$

In this case, $c(\theta,\varphi) = \alpha\theta^{\alpha-1}\varphi^{\alpha}$ and $F(\theta,\varphi) = \alpha^2\varphi^{-2}$. Thus, the reference prior for (θ,φ) is $\theta^{\alpha-1}\varphi^{\alpha-1}$.

When θ is the parameter of interest, the reference prior is $\varphi^{-1}\theta^{\alpha-1}$ if Procedure 1 is followed. Procedure 2 leads to the prior $\theta^{\alpha-1}\varphi^{\alpha}$.

If φ is the parameter of interest, Procedure 1 again proposes using $\varphi^{-1}\theta^{\alpha-1}$ while Procedure 2 suggests use of the prior φ^{-1}.

Example 4.3. Suppose we have observations from a gamma population with unknown scale and shape parameters σ and α respectively. However, due to some limitations of the measuring instrument, observations greater than an unknown threshold value θ (which is assumed to be much larger than the mean α/σ) cannot be detected and are completely lost. Thus, we have observations from the truncated gamma family

$$f(x;\theta,\sigma,\alpha) = \frac{e^{-\sigma x}x^{\alpha-1}}{\int_0^{\theta} e^{-\sigma y}y^{\alpha-1}dy}, \quad 0 < x < \theta, \tag{4.11}$$

and our aim is to make inference about σ and α, with θ as a nuisance parameter. Let $\Gamma(;z)$ denote the incomplete gamma integral $\int_0^z e^{-x}x^{-1}dx$ and set

$$\Phi_k(,z) = \int_0^z e^{-x}x^{-1}(\log x)^k dx, \quad k = 1,2.$$

It then follows that

$$c = c(\theta,\sigma,\alpha) = -\sigma^{\alpha}e^{-\sigma\theta}\theta^{\alpha-1}/\Gamma(\alpha;\sigma\theta) \tag{4.12}$$

and $\boldsymbol{F} = \boldsymbol{F}(\theta,\sigma,\alpha) = ((F_{jk}))_{1\le j,k\le 2}$, where

$$F_{11} = \sigma^{-2}\{\frac{\Gamma(\alpha+2;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)} - (\frac{\Gamma(\alpha+1;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)})^2\},$$
$$F_{12} = F_{21} = \sigma^{-1}\{\frac{\Phi_1(\alpha+1;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)} - \frac{\Gamma(\alpha+1;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)}\frac{\Phi_1(\alpha;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)}\},$$
$$F_{22} = \frac{\Phi_2(\alpha;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)} - \left(\frac{\Phi_1(\alpha;\sigma\theta)}{\Gamma(\alpha;\sigma\theta)}\right)^2. \tag{4.13}$$

These expressions are very complicated and it seems impossible to implement the Berger-Bernardo algorithm described by Procedure 1. Procedure 2 the use of the prior $(\det \boldsymbol{F}(\theta, \sigma, \alpha))^{1/2}$, but this is still complicated and intractable. In what follows, we suggest an approximation to the reference prior. Since $\sigma\theta$ is large, we replace all the incomplete integrals with the corresponding complete integrals to obtain the approximation c^* for c given by

$$c^* = -\sigma^{\alpha} e^{-\sigma\theta} \theta^{\alpha-1} / \Gamma(\alpha) \tag{4.14}$$

and $\boldsymbol{F}^*$ for $\boldsymbol{F}$ given by

$$\boldsymbol{F}^* = \begin{pmatrix} \sigma^{-2}\alpha & \sigma^{-1}\alpha(\Psi(\alpha+1) - \Psi(\alpha)) \\ \sigma^{-1}\alpha(\Psi(\alpha+1) - \Psi(\alpha)) & \Psi'(\alpha) \end{pmatrix}. \tag{4.15}$$

where $\Psi() = \frac{d}{d\beta} \log \Gamma(\beta)$ is the digamma function and Ψ' is its derivative.

To implement the Berger-Bernardo algorithm, we first put

$$\pi^*(\theta|\sigma, \alpha) = \frac{\sigma^{\alpha}}{\Gamma(\alpha)} e^{-\sigma\theta} \theta^{\alpha-1}, \quad \theta > 0. \tag{4.16}$$

For a given θ, we find a "reasonably non-informative" prior for (σ, α) based on $\boldsymbol{F}^*$ and then "average it" with respect to $\pi(\theta|\sigma, \alpha)$ in the sense described in Procedure 1. Finally, the overall reference prior is obtained by multiplication. However, as $\boldsymbol{F}^*$ is independent of θ, the prior for (σ, α) given θ will not depend on θ, and hence the "averaging" is superfluous. The prior for (σ, α) may be taken as the "Jeffreys' prior" $(\det \boldsymbol{F}^*(\sigma, \alpha))^{1/2}$ if both σ and α are of similar importance or a further Berger-Bernardo algorithm (of the regular case) may be applied to the "information matrix" $\boldsymbol{F}^*(\sigma, \alpha)$. In the first case, we obtain the reference prior

$$\begin{aligned} \pi^*(\theta, \sigma, \alpha) &= \sigma^{-1}\sqrt{\alpha}\sqrt{\Psi'(\alpha) - \alpha(\Psi(\alpha+1) - \Psi(\alpha))^2} \\ &\quad \frac{\sigma^{\alpha}}{\Gamma(\alpha)} e^{-\sigma\theta} \theta^{\alpha-1}, \quad \theta, \sigma, \alpha > 0. \end{aligned} \tag{4.17}$$

When α is teh parameter of interest, the reference prior is

$$\begin{aligned} \pi^*(\theta, \sigma, \alpha) &= \sigma^{-1}\sqrt{\Psi'(\alpha) - \alpha(\Psi(\alpha+1) - \Psi(\alpha))^2} \\ &\quad \frac{\sigma^{\alpha}}{\Gamma(\alpha)} e^{-\sigma\theta} \theta^{\alpha-1}, \quad \theta, \sigma, \alpha > 0. \end{aligned} \tag{4.18}$$

while, when the parameter of interest is σ, the reference prior is

$$\pi^{*}(\theta,\sigma,\alpha)=\sigma^{-1}\sqrt{\Psi'(\alpha)}\frac{\sigma^{\alpha}}{\Gamma(\alpha)}e^{-\sigma\theta}\theta^{\alpha-1},\quad \theta,\sigma,\alpha>0. \tag{4.19}$$

In both cases, the choice of the compact sets does not matter. We note that, although the conditional reference prior for θ given (σ,α) is proper (in fact, gamma), the "marginal" is an improper mixture of gammas and, hence, is improper. In addition, the prior for θ is supported on $(0,\infty)$ even though we have assumed in the derivation that θ is large. We may impose a lower bound for θ, but this will not have much effect on the inference as the posterior for θ given $X_1,\ldots,X_n$ is supported on $[X_{\max},\infty)$. The "marginal" for σ is σ^{-1} in each case and this is a reflection of the fact that σ is a scale parameter in the absence of a truncation. The prior for α is complicated and, in practice, we may prefer to make a suitable subjective choice.

APPENDIX

Proof of Proposition 3.1. By the results of GS,

$$\begin{aligned}Z_n(u,\boldsymbol{v})&=\frac{p^n(\boldsymbol{X}^n;\theta+u/n,\phi+n^{-1/2}\boldsymbol{v})}{p^n(\boldsymbol{X}^n;\theta,\varphi)}\\&=\exp[cu+\boldsymbol{v}^T\Delta_n-\frac{1}{2}\boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}]\mathbb{I}\{u<\sigma_n\}+o_p(1),\end{aligned} \tag{A.1}$$

uniformly on $(u,\boldsymbol{v})$ belonging to compacts, where $c=c(\theta,\varphi)$, $\boldsymbol{F}=\boldsymbol{F}(\theta,\varphi)$, $\sigma_n=\sigma_n(\theta)$ and $\Delta_n=\Delta_n(\theta,\varphi)$. Furthermore, convergence is uniform over $(u,\boldsymbol{v})$ belonging to compact sets. Now, for any $H>0$,

$$\begin{aligned}\int &|\pi(\theta+u/n,\phi+n^{-1/2}\boldsymbol{v})Z_n(u,\boldsymbol{v})\\&-\pi(\theta,\varphi)\exp[cu+\boldsymbol{v}^T\Delta_n-\frac{1}{2}\boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}]\mathbb{I}\{u<\sigma_n\}|\,du\,d\boldsymbol{v}\end{aligned}$$

is smaller or equal than

$$\begin{aligned}
&\int_{\|(u,\boldsymbol{v})\|\leq H} |\pi(\theta + u/n, \boldsymbol{\phi} + n^{-1/2}\boldsymbol{v}) - \pi(\theta,\varphi)| Z_n(u,\boldsymbol{v}) du\, d\boldsymbol{v} \\
&+ \int_{\|(u,\boldsymbol{v})\|\leq H} \pi(\theta,\varphi) \\
&|Z_n(u,\boldsymbol{v}) - \exp[cu + \boldsymbol{v}^T\Delta_n - \frac{1}{2}\boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}]\mathbb{I}\{u < \sigma_n\}| du\, d\boldsymbol{v} \\
&+ \int_{\|(u,\boldsymbol{v})\|>H} \pi(\theta + u/n, \boldsymbol{\phi} + n^{-1/2}\boldsymbol{v}) Z_n(u,\boldsymbol{v}) du\, d\boldsymbol{v} \\
&+ \int_{\|(u,\boldsymbol{v})\|>H} \pi(\theta,\varphi) e^{[cu+\boldsymbol{v}^T\Delta_n - \frac{1}{2}\boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}]}\mathbb{I}\{u < \sigma_n\} du\, d\boldsymbol{v}
\end{aligned} \qquad (A.2)$$

By choosing H large enough, we can make the third and fourth terms small with arbitrarily large probability. For the third term, this follows from Lemma I.5.2 of IH while for the fourth, it is a consequence of the stochastic boundedness of σ_n and Δ_n. Furthermore, for any $H > 0$, the first two terms converge to zero in probability by virtue of the continuity of $\pi(\cdot)$ and (A.1). Hence, the right hand side (RHS) of (A.2) goes to zero in probability and, thus,

$$\frac{n^{1+d/2} m_\pi(\boldsymbol{X}^n)}{p^n(\boldsymbol{X}^n;\theta,\varphi)} - \frac{\pi(\theta,\varphi)(2\pi)^{d/2}}{c(\det \boldsymbol{F})^{1/2}} e^{c\sigma_n + \frac{1}{2}\Delta_n^T \boldsymbol{F}^{-1}\Delta_n} \rightarrow_p 0. \quad (A.3)$$

Since σ_n and Δ_n are stochastically bounded, can be seen that (i) follows from (A.3).

The fact that the left hand side (LHS) of (A.2) goes to zero implies that the posterior density $\bar{\pi}_n(u,\boldsymbol{v})$ of $(u,\boldsymbol{v})$ admits the approximation

$$\int |\bar{\pi}_n(u,\boldsymbol{v}) - c\exp(cu)\mathbb{I}\{u < \sigma_n\}\phi_d(\boldsymbol{v};\boldsymbol{F}^{-1}\Delta_n,\boldsymbol{F}^{-1})| du\, d\boldsymbol{v} \rightarrow_p 0. \quad (A.4)$$

Similar arguments give

$$\int |u| |\bar{\pi}_n(u,\boldsymbol{v}) - c\exp(cu)\mathbb{I}\{u < \sigma_n\}\phi_d(\boldsymbol{v};\boldsymbol{F}^{-1}\Delta_n,\boldsymbol{F}^{-1})| du\, d\boldsymbol{v} \rightarrow_p 0 \quad (A.5)$$

and

$$\int \|\boldsymbol{v}\|^2 |\bar{\pi}_n(u, \boldsymbol{v}) - c\exp[cu]\mathbb{I}\{u < \sigma_n\}\phi_d(\boldsymbol{v}; \boldsymbol{F}^{-1}\Delta_n, \boldsymbol{F}^{-1}) |du\, d\boldsymbol{v} \to_p 0. \tag{A.6}$$

It then follows immediately that (A.5) and (A.6) imply part (ii).

Proof of part (*a*) *of Lemma 3.2.* It follows from Proposition 3.1 (i) that

$$R_n(\theta, \varphi) \to_d \log\left(\frac{c(\theta, \varphi) l(\theta, \varphi)}{\pi(\theta, \varphi)}\right) - c(\theta, \varphi)\sigma - \frac{1}{2}\Delta^T \boldsymbol{F}^{-1}(\theta, \varphi)\Delta - \frac{d}{2}\log(2\pi), \tag{A.7}$$

where σ is an exponential random variable with mean $1/c(\theta, \varphi)$ and Δ is distributed as $N_d(0, \boldsymbol{F}(\theta, \varphi))$. Hence, by Lemma A.1 proved below, we obtain

$$\begin{aligned}\limsup_{n\to\infty} E_{\theta,\varphi} R_n(\theta, \varphi) &\leq \log\left(\frac{c(\theta, \varphi) l(\theta, \varphi)}{\pi(\theta, \varphi)}\right) \\ &\quad - c(\theta, \varphi) E\sigma - \frac{1}{2} E(\Delta^T \boldsymbol{F}^{-1}(\theta, \varphi)\Delta) \\ &\quad - \frac{d}{2}\log(2\pi) \\ &= \log(c(\theta, \varphi) l(\theta, \varphi)\pi(\theta, \varphi)) - (1 + d/2) - \frac{d}{2}\log(2\pi),\end{aligned} \tag{A.8}$$

which completes the proof.

Lemma A.1. *The sequence $R_n(\theta, \varphi)$ is uniformly $P_{\theta,\varphi}$-integrable from above.*

To prove Lemma A.1, we need the following definition.

Definition A.1. *Let two probability measures P and Q be defined on a sample space $(\Omega, \mathcal{A})$. Let h denote the density of the absolutely continuous part of Q with respect to P. The modified Kullback-Leibler information number $K^*(Q, P)$ of Q with respect to P is then defined by $K^*(Q, P) = -\int \log h(\omega)\mathbb{I}\{h(\omega) > 0\}Q(d\omega)$.*

It is easy to see that K^* of the n-fold products P^n and Q^n is related to $K^*(Q, P)$ by $K^*(Q^n, P^n) = nK^*(Q, P)(Q\{h > 0\})^{n-1}$.

Proof of Lemma A.1. We fix (θ, φ) and suppress it whenever there is no chance of confusion. Let $A > 0$ and let n be sufficiently large. Then

$$R_n = -\log\left(\int_K \frac{p^n(\boldsymbol{X}^n; t, \boldsymbol{s})}{p^n(\boldsymbol{X}^n; \theta, \boldsymbol{\phi})}\pi(t, \boldsymbol{s})dt\, d\boldsymbol{s}\right)$$
$$\leq -\log\left(\int_{-A\leq u\leq\sigma_n(A), \|\boldsymbol{v}\|\leq A} Z_n(u, \boldsymbol{v}) \quad (A.9)\right.$$
$$\left.\pi(\theta + u/n, \varphi + n^{-1/2}\boldsymbol{v})du\, d\boldsymbol{v}\right).$$

where $\sigma_n(A) = \min(\sigma_n, A)$. By application of Jensen's inequality, we can bound the RHS of (2.16) by

$$-(A + \sigma_n(A))^{-1}A^{-d}\alpha^{-1}\int_{-A\leq u\leq\sigma_n(A), \|\boldsymbol{v}\|\leq A} \log Z_n(u, \boldsymbol{v})du\, d\boldsymbol{v}$$
$$- \log(A + \sigma_n(A)) - d\log A - \log\alpha$$
$$- \log\inf\{\pi(\theta + u/n, \varphi + n^{-1/2}\boldsymbol{v}) : |u| \leq A, \|\boldsymbol{v}\| \leq A\}, \quad (A.10)$$

where α is the volume of the unit ball in $\mathbb{R}^d$. The final term converges to $\log\pi(\theta, \varphi) > -\infty$ as $n \to \infty$, whereas the intermediate terms are clearly bounded as n varies. It therefore suffices to show that

$$\int_{-A\leq\sigma_n(A)\leq A, \|\boldsymbol{v}\|\leq A} \log Z_n(u, \boldsymbol{v})du\, d\boldsymbol{v}$$

is uniformly integrable.

Now, for any fixed $(u, \boldsymbol{v})$, we have

$$|\log Z_n(u, \boldsymbol{v}) - (cu + \boldsymbol{v}^T\Delta_n - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2)|\mathbb{I}\{u < \sigma_n\} \to_p 0. \quad (A.11)$$

If we vary $(u, \boldsymbol{v})$ according to the uniform distribution ν on $[-A, A] \times\{\boldsymbol{v} : \|\boldsymbol{v}\| \leq A\}$, then by Fubini's theorem, the LHS of (2.18) converges to zero in $(\nu \times P^n_{\theta,\varphi})$-probability. We claim that $|\log Z_n(u, \boldsymbol{v}) - (cu + \boldsymbol{v}^T\Delta_n - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2)|\mathbb{I}\{u < \sigma_n\}$ is uniformly integrable with respect to the joint distribution. If this is true, we then have

$$\int\int |\log Z_n(u, \boldsymbol{v}) - (cu + \boldsymbol{v}^T\Delta_n - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2)|$$
$$\mathbb{I}\{u < \sigma_n\}\nu(du, d\boldsymbol{v})p^n(\boldsymbol{x}^n; \theta, \varphi)d\boldsymbol{x}^n \to 0,$$

which in turn implies that

$$\int \log Z_n(u,\boldsymbol{v})\mathbb{I}\{u<\sigma_n\}\nu(du,d\boldsymbol{v}) - \int (cu+\boldsymbol{v}^T\Delta_n - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2)$$
$$\mathbb{I}\{u<\sigma_n\}\nu(du,d\boldsymbol{v})$$

is uniformly $P^n_{\theta,\varphi}$-integrable. The second term is, however, itself uniformly integrable. Hence, it follows that the first term must also be uniformly integrable which proves the lemma.

It remains for us to prove the claim that was made above. Note that

$$\begin{aligned}
&[-\log Z_n(u,\boldsymbol{v}) + (cu+\boldsymbol{v}^T\Delta_n - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2)]\mathbb{I}\{u<\sigma_n\}\\
&\quad = 2[-\log Z_n^{1/2}(u,\boldsymbol{v}) + Z_n^{1/2}(u,\boldsymbol{v})]\mathbb{I}\{u<\sigma_n\}\\
&\qquad + [-2Z_n^{1/2}(u,\boldsymbol{v}) + cu + \boldsymbol{v}^T\Delta_n - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2]\mathbb{I}\{u<\sigma_n\}.
\end{aligned} \qquad (A.12)$$

The second term has a uniformly bounded second moment and so is uniformly integrable. The first term is nonnegative and converges weakly to

$$\begin{aligned}\{-(cu+\boldsymbol{v}^T\Delta - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2) + 2\exp[cu/2 + \boldsymbol{v}^T\Delta/2\\ - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/4]\}\mathbb{I}\{u<\sigma\};\end{aligned} \qquad (A.13)$$

(see (2.1)). To complete the proof, it is enough to show that the limit of the expectation (with respect to the joint distribution) of the first term of the RHS of (A.12) is equal to the expectation of the expression in (A.13). It is therefore sufficient to show that the $P^n_{\theta,\varphi}$-expectation of the first term on the RHS of (A.12) converges to the expectation of the expression in (A.13) uniformly in $(u,\boldsymbol{v})$ belonging to compacts. We separate the two cases $u \geq 0$ and $u < 0$. In the former case, the expectation of the expression in (A.13) is

$$\{-(cu - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2) + 2\exp(cu/2 - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/8)\}\exp(-cu), \quad (A.14)$$

while in the latter case, this expectation is

$$\{-(cu - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2) + 2\exp(cu/2 - \boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/8)\}. \qquad (A.15)$$

Now, if $u \geq 0$, the expectation of the first term on the RHS of (2.20) is

$$\int_{p^n(\boldsymbol{x}^n;\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v})>0} p^n(\boldsymbol{x}^n;\theta,\varphi)$$
$$\log\left(\frac{p^n(\boldsymbol{x}^n;\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v})}{p^n(\boldsymbol{x}^n;\theta,\varphi)}\right) d\boldsymbol{x}^n$$
$$+2\int (p^n(\boldsymbol{x}^n;\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v})$$
$$p^n(\boldsymbol{x}^n;\theta,\varphi))^{1/2} d\boldsymbol{x}^n - 2P^n_{\theta,\varphi}\{\sigma_n > u\}$$

which may be expressed as

$$nK^*((\theta,\varphi),(\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v}))$$
$$\left[1-\int_{a_1(\theta)}^{a_1(\theta+u/n)} f(x;\theta,\varphi)dx - \int_{a_2(\theta+u/n)}^{a_2(\theta)} f(x;\theta,\varphi)dx\right]^{n-1}$$
$$+2\left(1-\frac{1}{2}\int (f^{1/2}(x;\theta+u/n,\phi+\boldsymbol{v}/n) - f^{1/2}(x;\theta,\phi))^2 dx\right)^n. \tag{A.16}$$

A Taylor series expansion yields

$$K^*((\theta,\varphi),(\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v}))$$
$$= n^{-1}(-cu+\boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/2)+o(n^{-1}),$$
$$1-\int_{a_1(\theta)}^{a_1(\theta+u/n)} f(x;\theta,\varphi)dx - \int_{a_2(\theta+u/n)}^{a_2(\theta)} f(x;\theta,\varphi)dx$$
$$= 1-cu/n+o(n^{-1}),\ \int (f^{1/2}(x;\theta+u/n,\phi+\boldsymbol{v}/n)$$
$$- f^{1/2}(x;\theta,\phi))^2 dx = n^{-1}(cu+\boldsymbol{v}^T\boldsymbol{F}\boldsymbol{v}/4)+o(n^{-1}),$$

where the asymptotic expansions are valid uniformly in $(u,\boldsymbol{v})$ belonging to compacts. Substituting in (2.24), the claim follows by taking the limit. The case $u < 0$ follows similarly. ◁

Proof of part (*b*) *of Lemma 3.2.* Taking $P_{\theta,\varphi}$-expectation in (2.17), we obtain

$$\begin{aligned}\psi_n(\theta,\varphi) &\le \alpha^{-1}A^{-(1+d)}\int\int_{-A\le u\le\sigma_n(A),\|\boldsymbol{v}\|\le A} p^n(\boldsymbol{x}^n;\theta,\varphi)\\ &\quad\times\left|\log\left(\frac{p^n(\boldsymbol{x}^n;\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v})}{p^n(\boldsymbol{x}^n;\theta,\varphi)}\right)\right| du\,d\boldsymbol{v}\,d\boldsymbol{x}^n\\ &\quad-(1+d)\log A-\log\alpha-\log\inf\{\pi(t,\boldsymbol{s}):(t,\boldsymbol{s})\in K\}\\ &= B_1+B_2\int\int_{-A\le u\le\sigma_n(A),\|\boldsymbol{v}\|\le A} p^n(\boldsymbol{x}^n;\theta,\varphi)\\ &\quad\times\left|\log\left(\frac{p^n(\boldsymbol{x}^n;\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v})}{p^n(\boldsymbol{x}^n;\theta,\varphi)}\right)\right| du\,d\boldsymbol{v}d\boldsymbol{x}^n,\end{aligned} \tag{A.17}$$

where B_1 and B_2 are constants. If $u<\sigma_n$, we can write

$$\begin{aligned}&\log Z_n(u,\boldsymbol{v})\\ &= \frac{u}{n}\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(X_i;\theta,\varphi)+n^{1/2}\Delta_n\\ &\quad+\frac{u^2}{2n^2}\sum_{i=1}^n\frac{\partial^2}{\partial\theta^2}\log f(X_i;\theta^*,\boldsymbol{\phi}^*)\\ &\quad+\frac{u}{2n^{3/2}}\sum_{j=1}^d v_j\sum_{i=1}^n\frac{\partial^2}{\partial\theta\partial\varphi_j}\log f(X_i;\theta^*,\boldsymbol{\phi}^*)\\ &\quad+\frac{1}{2n}\sum_{j=1}^d\sum_{k=1}^d v_jv_k\sum_{i=1}^n\frac{\partial^2}{\partial\varphi_j\partial\varphi_k}\log f(X_i;\theta^*,\boldsymbol{\phi}^*),\end{aligned} \tag{A.18}$$

where $(\theta^*,\boldsymbol{\phi}^*)$ lies between (θ,φ) and $(\theta+u/n,\varphi+n^{-1/2}\boldsymbol{v})$; here and below v_j and φ_j denote the jth components of $\boldsymbol{v}$ and $\boldsymbol{\phi}$ respectively.

Thus, by (A9), we may write

$$\begin{aligned}\log Z_n(u,\boldsymbol{v}) &= \frac{u}{n}\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(X_i;\theta,\varphi)+n^{1/2}\Delta_n\\ &\quad+\frac{1}{2n}(u+\sum_{j=1}^d v_j)^2\sum_{i=1}^n H_{\theta,\varphi}(X_i).\end{aligned} \tag{A.19}$$

Note that $E|\boldsymbol{v}^T\Delta_n| \leq (E(\boldsymbol{v}^T\Delta_n)^2)^{1/2} = (\boldsymbol{v}^T\boldsymbol{F}(\theta,\varphi)\boldsymbol{v})^{1/2}$. Hence

$$\psi_n(\theta,\varphi) \leq B_1 + B_2\left[\int_{|u|\leq A,\|\boldsymbol{v}\|\leq A}\left\{|u|\int|\frac{\partial}{\partial\theta}f(x;\theta,\varphi)|dx + (\boldsymbol{v}^T\boldsymbol{F}(\theta,\varphi)\boldsymbol{v})^{1/2} + \frac{1}{2}(u+\sum_{j=1}^{d}v_j)^2E(H_{\theta,\varphi}(X_1))\right\}\right]du\,d\boldsymbol{v}$$
$$\leq B_1' + B_2'\left\{\int|\frac{\partial}{\partial\theta}f(x;\theta,\varphi)|dx + \max_{1\leq j\leq d}J_{jj}(\theta,\varphi) + E(H_{\theta,\varphi}(X_1))\right\}, \qquad (A.20)$$

where B_1' and B_2' are constants. As a function of (θ,φ), the RHS of (2.28) is continuous, and so is bounded on compact sets.

ACKNOWLEDGEMENTS

The author is grateful to the referee for a very careful reading of the manuscript and suggestion which improved the quality of the paper and its presentation. He would also like to thank Professor J. K. Ghosh for a fruitful discussion. Research was supported by the National Board of Higher Mathematics, Department of Atomic Energy, Bombay, India

REFERENCES

Aitchison, J. (1975). Goodness of prediction fit. *Biometrika* **62**, 547–554.

Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84**, 200–207.

Berger, J. O. and Bernardo, J. M. (1992a). On the development of reference priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Berger, J. O. and Bernardo, J. M. (1992b). Ordered group reference priors with application to a multinomial problem. *Biometrika* 79, 25–37.

Berger, J. O. and Bernardo, J. M. (1992c). Reference priors in a variance component problem. In *Proceedings of the Indo–USA Workshop on Bayesian Analysis in Statistics and Econometrics* (P. Goel, ed.) New York: Springer, 323–340.

Berger, J. O., Bernardo, J. M. and Mendoza, M. (1989). On priors that maximize expected information. In *Recent Developments in Statistics and Their Applications* (J. P. Klein and J. C. Lee, eds.) Freedom Academy Publishing, Seoul, 1–20.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. B* **41**, 113–147 (with discussion).

Chang T. and Eaves, D. (1990). Reference priors for the orbit in a group model. *Ann. Statist.* **18**, 1070–1090.

Clarke, B. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Plan. Inf.* **41**, 37–60.

Ghosal, S. and Samanta, T. (1995). Asymptotic behaviour of Bayes estimates and posterior distributions in multiparameter nonregular cases. *Math. Methods Statist.* **4**, 361–388.

Ghosal, S. and Samanta, T. (1997). Expansion of Bayes risk for entropy loss and reference prior in nonregular cases. *Statistics and Decisions*, (to appear).

Ghosh, J. K. (1994). *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series in Probability and Statistics **4**, IMS, Hayward, CA.

Ghosh, J. K., Ghosal, S. and Samanta, T. (1994). Stability and convergence of posterior in non-regular problems. *Statistical Decision Theory and Related Topics V* (S. S. Gupta and J. O. Berger, eds.) New York: Springer-Verlag, 183–199.

Ghosh, J. K. and Mukerjee, R. (1992). Non-informative priors. *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 195–210 (with discussion).

Ibragimov, I. A. and Khas'minskii, R. Z. (1973). On the information contained in a sample about a parameter. *Second International Symposium on Information Theory* (B. N. Petrov and F. Csaki, eds.). Akademiaikiado, Budapest, 295–309.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* New York: Springer-Verlag.

Lindley, D. V. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist.* **27**, 986–1005.

Polson, N. G. (1988). On the expected amount of information from a nonlinear model. *J. Roy. Statist. Soc. Ser. B.* **54**, 889–895.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 69–90.

Sun, D. and Ye, K. (1995). Reference prior Bayesian analysis for normal mean products. *J. Amer. Statist. Assoc.* **90**, 589–597.

Yang, R. and Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22**, 1195–1211.

Ye, K. Y. and Berger, J. O. (1991). Noninformative priors for inference in exponential regression models. *Biometrika* **78**, 645–656.