Statistics & Decisions 15, 129 – 140 (1997)

EXPANSION OF BAYES RISK FOR ENTROPY LOSS AND REFERENCE PRIOR IN NONREGULAR CASES

Subhashis Ghosal[1] and Tapas Samanta

Received: Revised version: *June 27, 1996*

Abstract

Lindley's measure of information in a sample about a parameter is given by the average Kullback-Leibler distance between the posterior and the prior. This is also equal to the Bayes risk when one estimates the density using the entropy loss. In this paper, an asymptotic expansion of this measure is obtained for a one-parameter family of discontinuous densities. This expansion is then used to obtain the reference prior in the sense of Bernardo.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent observations each having a distribution P_θ with a density $f(x;\theta)$ with respect to a fixed dominating measure where $\theta \in \Theta$, an open subset of $\mathbb{R}$. Consider a prior on Θ having a density $\pi(\cdot)$ with respect to the Lebesgue measure. Lindley's measure of information (see Lindley [15]) $I(\pi; X^n)$ in $X^n = (X_1, \ldots, X_n)$ about θ is given by the average relative entropy or Kullback-Leibler distance between the posterior distribution of θ given X^n and the prior π (see Sec. 2). This measure is also equal to the average (with respect to π) relative entropy distance between the distribution of X^n given θ and the marginal distribution of X^n and indeed is the Bayes risk when one estimates the density of X^n given θ using the entropy loss (see Aitchison [1]).

In Section 2 of this paper, we obtain an asymptotic expansion of this Bayes risk (or measure of information) $I(\pi; X^n)$ for a family of nonregular cases. We restrict our attention to the cases which, by results of Ghosh *et al.* [11], are essentially the only cases where the posterior distributions converge. Our treatment is similar to that of Clarke and Barron [9] who obtained an expansion of the entropy risk for the regular cases. Results similar to those of Clarke and Barron [9] were obtained earlier by Ibragimov and Has'minskii [13]. For extensions to non

[1]Research supported by the National Board of Higher Mathematics, Department of Atomic Energy, Bombay, India.

AMS 1991 *subject classification:* Primary 62F25.

Key words and phrases: Asymptotic expansion, Bayes risk, discontinuous densities, Kullback-Leibler number, posterior distribution, reference prior.

i.i.d. regular cases, see, for example, Polson [16]. Certain related results are obtained by Rissanen [18, 19].

The reference prior method for development of noninformative priors was initiated by Bernardo [7] and developed further in a number of papers including Berger and Bernardo [2, 3, 4, 5], Berger, Bernardo and Mendoza [6], Ghosh and Mukerjee [12] and Chang and Eaves [8]. Bernardo [7] considered the measure $I(\pi; X^n)$ as a measure of information in X^n about θ and argued that the larger the measure, the less informative is the prior. The reference prior is defined as a π that maximizes this measure in an asymptotic sense. Since this measure is also the Bayes risk with respect to the entropy loss, maximizing $I(\pi; X^n)$ would lead to an asymptotically least favourable π and therefore under reasonable conditions, the corresponding Bayes estimate is asymptotically minimax. Also, the reference priors usually have the property that the corresponding procedures match with some standard frequentist procedures upto a certain order; see Ghosh and Mukerjee [12] and the references therein.

In Section 3, we use the asymptotic expansion obtained in Section 2 to find a prior that maximizes $I(\pi; X^n)$ in an asymptotic sense. Explicit forms are derived for some important examples.

2. Expansion of Bayes Risk for Entropy Loss

Let $X_1, X_2, \ldots$ be i.i.d. observations with a common distribution P_θ having a density $f(x;\theta)$, $\theta \in \Theta$, where Θ is an open interval in $\mathbb{R}$. We assume that for all $\theta \in \Theta, f(\cdot;\theta)$ is strictly positive on a closed interval (bounded or unbounded) $S(\theta) := [a_1(\theta), a_2(\theta)]$ depending on θ and is zero outside $S(\theta)$. It is permitted that one of the endpoints is free of θ and may be plus or minus infinity. In view of the results of Ghosh *et al.* [11] (Theorem 2.4 and Example 3.2), in order to have a limit of the posterior, it is necessary that the sets $S(\theta)$ are either increasing or decreasing in θ. The case where $S(\theta)$ increases with θ may be reduced to the case where $S(\theta)$ decreases, by the reparametrization $\theta \mapsto (-\theta)$. We, therefore, consider only the latter, namely the case where $a_1(\theta)$ is increasing and $a_2(\theta)$ is decreasing in θ. Moreover, we assume that these functions are strictly monotone unless they are infinite or free from θ.

We now make the following assumptions on the density $f(x;\theta)$.

(A1) For every $\varepsilon > 0$, $\inf\{r_2^2(\theta, \theta+h) : |h| > \varepsilon\}$ is bounded away from zero on compact subsets of Θ, where $r_2^2(\theta, \theta+h) := \int (f^{1/2}(x;\theta) - f^{1/2}(x;\theta+h))^2 dx$ is the squared Hellinger distance between $f(\cdot;\theta)$ and $f(\cdot;\theta+h)$.

(A2) If not infinite, $a_1(\theta)$ and $a_2(\theta)$ are continuously differentiable in θ.

(A3) The limits

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

exist $\forall \theta \in \Theta$, and the above convergences are uniform on compact subsets of Θ.

(A4) For each x, $\log f(x;\theta)$ is twice differentiable in θ on $\{\theta : a_1(\theta) < x < a_2(\theta)\}$ and for any θ, there exist a neighbourhood N_θ of θ and a function $H_\theta(x)$ such that

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

where $E_\theta H_\theta(X_1)$ is finite and continuous in θ.

(A5) The functions $p(\theta)$, $q(\theta)$ and $\int |(\partial/\partial\theta) f(x;\theta)| dx$ are (finite and) continuous and have polynomial majorants.

(A6) As $|u| \to \infty$, for some $\gamma > 0$,

$$\int f^{1/2}(x;\theta) f^{1/2}(x;\theta+u)dx \leq B(\theta)|u|^{-\gamma},$$

where $B(\theta)$ is bounded on compact subsets of Θ.

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics based on $X_1, X_2, \ldots, X_n$ and set $W_n = \min\{a_1^{-1}(X_{1:n}), a_2^{-1}(X_{n:n})\}$ (The first term in the minimum is ignored if a_1 is minus infinity or free of θ, and the second term is ignored if a_2 is infinity or free of θ). Note that $X_{1:n} \downarrow a_1(\theta)$ and $X_{n:n} \uparrow a_2(\theta)$ a.s. $[P_\theta]$, and hence for almost all samples, W_n is defined for all n, sufficiently large. Now the likelihood function $p^n(X^n;\theta) = \prod_{i=1}^n f(X_i;\theta)$ is positive if and only if $\theta \leq W_n$. Define $c(\theta) = E_\theta((\partial/\partial\theta)\log f(X_1;\theta))$ and observe that $c(\theta) = a_1'(\theta)p(\theta) - a_2'(\theta)q(\theta) > 0$. Let $\sigma_n = \sigma_n(\theta) = n(W_n - \theta)$. We make one more assumption:

(A7) For any compact subset K of Θ, $\sup_{\theta\in K}\sup_{n\geq 1} E_\theta\sigma_n < \infty$.

We shall give useful sufficient conditions for (A7) at the end of this section.

Let $\pi(\theta)$ be a (proper) prior density on Θ and $\pi(\theta|X^n)$ be the corresponding posterior. Lindley's measure of information $I(\pi;X^n)$ is defined to be the expected Kullback-Leibler distance between the posterior and the prior, i.e.,

$$I(\pi;X^n) = H(\pi) - H_{X^n}(\pi),$$

where $H_{X^n}(\pi) = EH(\theta|X^n)$, $H(\theta|X^n) = -\int \pi(\theta|X^n)\log\pi(\theta|X^n)d\theta$ and $H(\pi) = -\int \pi(\theta)\log\pi(\theta)d\theta$.

The following theorem is the main result of this section.

Theorem 2.1. *Assume Conditions* (A1)–(A7). *Let K be a compact subset of Θ and $\pi(\theta)$ be a proper prior which is positive, continuous and concentrated on K. Then as $n \to \infty$,*

$$I(\pi;X^n) = \log(n/e) + \int_K \pi(\theta)\log(c(\theta)/\pi(\theta))d\theta + o(1). \tag{2.1}$$

The proof of this theorem is long and we break it into several parts. It suffices to show that

$$\liminf_{n\to\infty}[I(\pi;X^n) - \log n] \geq -1 + \int_K \pi(\theta)\log(c(\theta)/\pi(\theta))d\theta \tag{2.2}$$

and

$$\limsup_{n\to\infty}[I(\pi;X^n) - \log n] \leq -1 + \int_K \pi(\theta)\log(c(\theta)/\pi(\theta))d\theta. \tag{2.3}$$

We first establish (2.2), the proof of which is much easier than that of (2.3). The following entropy maximization property of (the negative of) the exponential distribution (see, e.g., p. 217 of Rao [17]) is a straightforward consequence of the information inequality.

Lemma 2.1. *Let f be a density on $(-\infty, 0]$ having expectation $-\mu$, $\mu > 0$. Then*

$$-\int f(x)\log f(x)dx \leq 1 + \log\mu.$$

Let $m_n(x^n) = \int p^n(x^n;\theta)\pi(\theta)d\theta$ be the marginal density of X^n. The following result is an important step towards the proof of Theorem 2.1, and is also of interest in its own right.

Proposition 2.1. *Under Assumptions* (A1)–(A6),

(i) $\log(m_n(X^n)/p^n(X^n;\theta)) + \log n + \log(c(\theta)/\pi(\theta)) - c(\theta)\sigma_n \xrightarrow{P_\theta^n} 0$,

(ii) $n(E(t|X^n) - W_n) \xrightarrow{P_\theta^n} -1/c(\theta)$,

where t stands for a dummy variable for the parameter.

Proof: By the general results derived in Ibragimov and Has'minskii [14], Ch. V,

$$p^n(X^n;\theta+u/n)/p^n(X^n;\theta) = \exp[c(\theta)u]1\{u<\sigma_n\} + o_p(1), \tag{2.4}$$

where the convergences are uniform on $|u| \le H$ for any $H > 0$. Now

$$\begin{aligned}
&\int |\pi(\theta+u/n)\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)} - \pi(\theta)\exp[c(\theta)u]1\{u<\sigma_n\}|du \\
&\le \int_{|u|\le H} |\pi(\theta+u/n)-\pi(\theta)|\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)}du \\
&+ \int_{|u|\le H} \pi(\theta)\left|\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)} - \exp[c(\theta)u]1\{u<\sigma_n\}\right|du \\
&+ \int_{|u|>H} \pi(\theta+u/n)\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)}du \\
&+ \int_{|u|>H} \pi(\theta)\exp[c(\theta)u]1\{u<\sigma_n\}du.
\end{aligned} \tag{2.5}$$

Let $\varepsilon > 0$ and $\delta > 0$ be given. The variable σ_n is asymptotically exponential and so it is tight. Hence we can choose $H > 0$ such that $P\{\sigma_n > H\} < \delta/4$ for all n and $(\pi(\theta)/c(\theta))\exp[-c(\theta)H] < \varepsilon/4$. Thus the last term on the right hand side (RHS) of (2.5) is less than $\varepsilon/4$ with probability greater than $1-\delta/4$. The third term can also be made less than $\varepsilon/4$ with probability greater than $1-\delta/4$ by Lemma I.5.2 of Ibragimov and Has'minskii [14], provided H is chosen large enough. Now, for such an H, choose n large enough so that the first two terms are less than $\varepsilon/4$ with probability greater than $1-\delta/4$, by virtue of (2.4). Thus

$$\int |\pi(\theta+u/n)\frac{p^n(X^n;\theta+u/n)}{p^n(X^n;\theta)} - \pi(\theta)\exp[c(\theta)u]1\{u<\sigma_n\}|du \xrightarrow{P_\theta^n} 0 \tag{2.6}$$

which also leads to the posterior approximation:

$$\int\left|\frac{\pi(\theta+u/n)p^n(X^n;\theta+u/n)}{nm_n(X^n;\theta)} - c(\theta)\exp[c(\theta)(u-\sigma_n)]1\{u<\sigma_n\}\right|du \xrightarrow{P_\theta^n} 0. \tag{2.7}$$

From (2.6), we immediately have

$$n\,m_n(X^n)/p^n(X^n;\theta) - (\pi(\theta)/c(\theta))\exp[c(\theta)\sigma_n] \xrightarrow{P_\theta^n} 0 \tag{2.8}$$

which is equivalent to (i). Part (ii) can be proved in a similar manner. ▱

Remark 2.1. An almost sure version of Proposition 2.1, together with asymptotic expansions, can also be obtained under some further assumptions; see Ghosal and Samanta [10].

We now prove (2.2). Let $v = n(t - W_n)$ where t is a dummy variable for the parameter. Then the posterior density of v is $\pi_n^*(v) = n^{-1}\pi_n(W_n + v/n)$, where π_n is the posterior density of t. Note that $\pi_n^*(v)$ is concentrated on $(-\infty, 0]$. Thus

$$\begin{aligned} I(\pi; X^n) &= -\int_K \pi(\theta)\log\pi(\theta)d\theta + \int m_n(x^n)\int_K \pi_n(\theta)\log\pi_n(\theta)d\theta dx^n \\ &= -\int_K \pi(\theta)\log\pi(\theta)d\theta + \int m_n(x^n)\int_{-\infty}^0 \pi_n^*(v)[\log n + \log\pi_n^*(v)]dv dx^n \\ &= -\int_K \pi(\theta)\log\pi(\theta)d\theta + \log n + \int m_n(x^n)\int_{-\infty}^0 \pi_n^*(v)\log\pi_n^*(v)dv dx^n \\ &\geq -\int_K \pi(\theta)\log\pi(\theta)d\theta + \log n - \int m_n(x^n)[1 + \log E(-v|x^n)]dx^n \qquad (2.9) \end{aligned}$$

by Lemma 2.1. By Proposition 2.1 (ii), $E(-v|X^n) \xrightarrow{P_\theta^n} (c(\theta))^{-1}$. Further, $E(-v|X^n) = n(W_n - E(t|X^n)) = \sigma_n - n(E(t|X^n) - \theta)$ is uniformly integrable by (A7) and Theorem I.5.2 of Ibragimov and Has'minskii [14]. (Here the uniformity is also with respect to $n \geq 1$ and θ belonging to compacts.) Hence $\log E(-v|X^n)$ is uniformly integrable from above and so

$$\begin{aligned} &\liminf_{n\to\infty}[I(\pi; X^n) - \log(n/e) + \int_K \pi(\theta)\log\pi(\theta)d\theta] \\ &\geq -\limsup_{n\to\infty}\int_K \pi(\theta)\int \log E(v|x^n)p^n(x^n;\theta)dx^n d\theta \\ &\geq -\int_K \pi(\theta)\log(c(\theta))^{-1}d\theta = \int_K \pi(\theta)\log c(\theta)d\theta, \qquad (2.10) \end{aligned}$$

which is equivalent to (2.2). ▱

We now prove (2.3). Fix a $\theta \in \Theta$ and let

$$\begin{aligned} R_n &= -\log(m_n(x^n)/p^n(x^n;\theta)) - \log n, \\ \psi_n(\theta) &= K(P_\theta^n; m_n) - \log n = E_\theta R_n, \\ \psi(\theta) &= \log(c(\theta)/\pi(\theta)) - 1 \end{aligned}$$

where $K(P_\theta^n; m_n)$ stands for the Kullback-Leibler information number between the two probability measures P_θ^n and m_n. Note that

$$I(\pi; X^n) = \int_K \psi_n(\theta)\pi(\theta)d\theta. \qquad (2.11)$$

The proof of (2.3) follows from the following result.

Lemma 2.2. *Under Assumptions* (A1)–(A6), *we have*

(a) $\limsup_{n\to\infty}\psi_n(\theta) \leq \psi(\theta)$,

(b) $\psi_n(\theta)$ *is uniformly dominated from above by an integrable function on* K.

The following result will be used to prove Lemma 2.2.

Lemma 2.3. *Let $\theta \in \Theta$. Then under P_θ, R_n is uniformly integrable from above, i.e., there exists a uniformly integrable sequence of random variables R_n' such that $R_n \le R_n'$, $n \ge 1$.*

Definition 2.1. Let P and Q be two probability measures on a measurable space $(\Omega, \mathcal{A})$ and let h stand for the Radon-Nikodym derivative of the absolutely continuous part of P with respect to Q. The *modified Kullback-Leibler information number* $K^*(Q;P)$ of Q with respect to P is defined by

$$K^*(Q;P) = -\int_{\{h>0\}} \log h(\omega) Q(d\omega).$$

Unlike the usual Kullback-Leibler information number $K(Q;P)$, $K^*(Q;P)$ can be negative and hence cannot be viewed as a distance measure. The following property of K^* will be used in the sequel; the proof is fairly simple.

Let P^n and Q^n be n-fold products of two probabilities P and Q. Then

$$K^*(Q^n;P^n) = nK^*(Q;P)(Q\{h>0\})^{n-1}. \tag{2.12}$$

Proof of Lemma 2.3: For any $A > 0$,

$$\begin{aligned} R_n &= -\log[(1/p^n(x^n;\theta))\int_K p^n(x^n;t)\pi(t)dt] - \log n \\ &\le -\log[\int_{\{-A\le u\le \sigma_n(A)\}} \frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\pi(\theta+u/n)du] \end{aligned} \tag{2.13}$$

where $\sigma_n(A) = \min\{\sigma_n, A\}$.

Now applying Jensen's inequality, the RHS of (2.13) can be bounded by

$$\begin{aligned} &-(A+\sigma_n(A))^{-1}\int_{\{-A\le u\le\sigma_n(A)\}} \log(p^n(x^n;\theta+u/n)/p^n(x^n;\theta))du \\ &\quad - \log(A+\sigma_n(A)) - \log\inf\{\pi(\theta+u/n) : |u| \le A\}. \end{aligned} \tag{2.14}$$

The second term in (2.14) is uniformly bounded and the last term converges to $\log\pi(\theta) > -\infty$. Hence it is enough to show that the first term is bounded by an uniformly integrable function from above. Now

$$[\log(p^n(X^n;\theta+u/n)/p^n(X^n;\theta)) - c(\theta)u]1\{u<\sigma_n\} \xrightarrow{P_\theta^n} 0 \tag{2.15}$$

for any fixed u. Therefore, by Fubini's theorem, the above convergence also holds in the joint probability $(\nu \times P_\theta^n)$ where ν is the uniform distribution on $[-A, A]$. So it is now enough to show that

$$[\log(p^n(X^n;\theta+u/n)/p^n(X^n;\theta)) - c(\theta)u]1\{u<\sigma_n\}$$

is uniformly integrable with respect to $\nu \times P_\theta^n$. Indeed, we then have

$$\int\int_{[-A,A]} |\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)} - c(\theta)u|1\{u<\sigma_n\}du\, p^n(x^n;\theta)dx^n \to 0,$$

which implies that

$$\int_{[-A,A]} \log \frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)} 1\{u<\sigma_n\}du - \int_{[-A,A]} c(\theta)u\,1\{u<\sigma_n\}du$$

is uniformly integrable with respect to P_θ^n. Since the last term is always uniformly integrable by the tightness of $\{\sigma_n\}$, the result will then follow. Now

$$\begin{aligned}
&[-\log(p^n(x^n;\theta+u/n)/p^n(x^n;\theta)) + c(\theta)u]1\{u<\sigma_n\}\\
&= 2\left[-\log\left(\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\right)^{1/2} + \left(\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\right)^{1/2}\right]1\{u<\sigma_n\}\\
&\quad - \left[2\left(\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\right)^{1/2} + c(\theta)u\right]1\{u<\sigma_n\}. \qquad (2.16)
\end{aligned}$$

The second term on the RHS of (2.16) has a bounded second moment and hence is uniformly integrable. The first term is nonnegative and being a continuous function of the normalized likelihood ratio, is itself weakly convergent with respect to P_θ^n for each fixed u, and hence is also so with respect to $(\nu \times P_\theta^n)$-probability. The weak limit here is

$$(-c(\theta)u + 2\exp[c(\theta)u/2])1\{u<\sigma\},$$

where σ has an exponential distribution with parameter $c(\theta)$. The expectation of the limit is

$$\begin{aligned}
&(2A)^{-1}\{-\int_{[-A,A]}\int_0^\infty c(\theta)u\,1\{u<s\}c(\theta)\exp[-c(\theta)s]dsdu\\
&\qquad +2\int_{[-A,A]}\int_0^\infty \exp[c(\theta)u/2]1\{u<s\}c(\theta)\exp[-c(\theta)s]dsdu\}\\
&= (2A)^{-1}\{\int_{[-A,0]}(-c(\theta)u + 2\exp[c(\theta)u/2])du\\
&\qquad + \int_{[0,A]}(-c(\theta)u + 2\exp[c(\theta)u/2])\exp[-c(\theta)u]du\}. \qquad (2.17)
\end{aligned}$$

By a well known fact, the uniform integrability verification now boils down to showing that the limit of the expectations coincides with the expression in (2.17). Fix $u \geq 0$ and note that the expectation of the first term on the RHS of (2.16) given θ is equal to

$$\begin{aligned}
&\int_{p^n(x^n;\theta+u/n)>0} p^n(x^n;\theta)\log(p^n(x^n;\theta+u/n)/p^n(x^n;\theta))dx^n\\
&\qquad +2\int (p^n(x^n;\theta+u/n)p^n(x^n;\theta))^{1/2}dx^n\\
&= nK^*(\theta;\theta+u/n)\left[1-\int_{a_1(\theta)}^{a_1(\theta+u/n)} f(x;\theta)dx - \int_{a_2(\theta+u/n)}^{a_2(\theta)} f(x;\theta)dx\right]^{n-1}\\
&\qquad +2[1-r_2^2(\theta,\theta+u/n)/2]^n\\
&= n(-c(\theta)u/n + o(1/n))(1-c(\theta)u/n)^{n-1} + 2(1-c(\theta)u/(2n)+o(1/n))^n.
\end{aligned}$$

The last expression converges to

$$-c(\theta)u\exp[-c(\theta)u] + 2\exp[-c(\theta)u/2]$$

and the convergence is uniform in u belonging to compact subsets. Similarly for $u < 0$, the limit is

$$-c(\theta)u + 2\exp[c(\theta)u/2]$$

and the convergence is uniform on compacts again. Clearly, the last two statements are sufficient to imply the desired result. ▱

Proof of Lemma 2.2: We first prove Statement (a). As a consequence of Proposition 2.1,

$$R_n \xrightarrow{d} \log(c(\theta)/\pi(\theta)) - c(\theta)\sigma \tag{2.18}$$

where σ has an exponential distribution with parameter $c(\theta)$. By Lemma 2.3, we have

$$\limsup_{n\to\infty} E_\theta R_n \leq \log(c(\theta)/\pi(\theta)) - c(\theta)E\sigma = \log(c(\theta)/\pi(\theta)) - 1 \tag{2.19}$$

as desired.

It remains to prove Statement (b). From (2.14),

$$\begin{aligned} \psi_n(\theta) \leq\ & A^{-1}\int\int_{[-A,\sigma_n(A)]} p^n(x^n;\theta)|\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}|du\,dx^n \\ & -\log A - \log\inf\{\pi(t) : t\in K\}. \end{aligned} \tag{2.20}$$

If $p^n(x^n;\theta+u/n) > 0$,

$$\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)} = \frac{u}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(x_i;\theta) + \frac{u^2}{2n^2}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log f(x_i;\theta^*)$$

where θ^* lies between θ and $\theta + u/n$. Consequently by (A4), for all sufficiently large n,

$$\left|\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}\right| \leq \frac{|u|}{n}\sum_{i=1}^n\left|\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right| + \frac{u^2}{2n^2}\sum_{i=1}^n H_\theta(x_i),$$

which yields

$$\begin{aligned} \psi_n(\theta) \leq\ & \int_{[-A,A]}\int_{p^n(x^n;\theta+u/n)>0} p^n(x^n;\theta)|\log\frac{p^n(x^n;\theta+u/n)}{p^n(x^n;\theta)}|dx^n du \\ & -\log A - \log\inf\{\pi(t) : t\in K\} \\ \leq\ & \int_{[-A,A]}|u|\int\left|\frac{\partial}{\partial\theta}f(x;\theta)\right|dx\,du + \frac{1}{2}\int_{[-A,A]}u^2\int H_\theta(x)f(x;\theta)dxdu \\ & -\log A - \log\inf\{\pi(t) : t\in K\} \\ \leq\ & 2A\int\left|\frac{\partial}{\partial\theta}f(x;\theta)\right|dx + 2A^3\int H_\theta(x)f(x;\theta)dx \\ & -\log A - \log\inf\{\pi(t) : t\in K\}. \end{aligned} \tag{2.21}$$

By assumption, the terms are continuous in θ and hence bounded on compacts. Thus (b) is proved. ▱

By Proposition 2.1, the posterior converges in the L^1-distance:

$$\int_{-\infty}^{0} |\pi_n^*(v|X^n) - c(\theta)\exp[c(\theta)v]|dv \xrightarrow{P_\theta^n} 0,$$

where π_n^* is the posterior density of $v = n(t - W_n)$. This can be equivalently written as

$$\int_{-\infty}^{0} |\pi_n(t|X^n) - nc(\theta)\exp[nc(\theta)(t - W_n)]|dt \xrightarrow{P_\theta^n} 0. \tag{2.22}$$

We shall now present an information theoretic version of (2.22).

Theorem 2.2. *Assume* (A1)–(A6) *and the following condition:*
(A7)′ $\sigma_n(\theta)$ *is uniformly (in n and also in θ belonging to compacts) integrable. Then*

$$\int\int\int_{-\infty}^{W_n} (\pi_n(t|x^n)\log(\pi_n(t|x^n)/\rho_n(t;\theta,x^n))dt)\, p^n(x^n;\theta)dx^n\pi(\theta)d\theta \to 0, \tag{2.23}$$

where $\rho_n(t;\theta,x^n) = nc(\theta)\exp[nc(\theta)(t - W_n)]$.

Proof: The expression on the left hand side of (2.23) is equal to

$$\begin{aligned} I(\pi;X^n) &+ \int_K \pi(\theta)\log\pi(\theta)d\theta - \log n - \int_K \pi(\theta)\log c(\theta)d\theta \\ &- \int_K \int n(E(t|x^n) - W_n)c(\theta)p^n(x^n;\theta)dx^n\pi(\theta)d\theta. \end{aligned} \tag{2.24}$$

By Proposition 2.1 (ii) and (A7)′, the last term converges to one. Using Theorem 2.1, we now get the result. ▱

As we have promised, we now give sufficient conditions for (A7) to hold.

Proposition 2.2. *Let* $g_\theta(v) = P_\theta(S(\theta + v))$. *If*

$$\sup_{\theta\in K}\int_0^\infty g_\theta(v)dv < \infty, \tag{2.25}$$

then (A7) *holds. If for some* $\delta > 0$,

$$\sup_{\theta\in K}\int_0^\infty v^\delta g_\theta(v)dv < \infty, \tag{2.26}$$

then (A7)′ *holds.*

Proof: We shall prove only the first assertion; proof of the second one is similar. We have

$$E_\theta\sigma_n(\theta) = \int_0^\infty P_\theta^n\{\sigma_n(\theta) > v\}dv = \int_0^\infty n(g_\theta(v))^n dv. \tag{2.27}$$

Now $n(g_\theta(v))^n \le g_\theta(v)$ if and only if $g_\theta(v) \le n^{-1/(n-1)}$, which is true if and only if $v > c_n$, where $g_\theta(c_n) = n^{-1/(n-1)}$. Clearly, $c_n \to 0$ since $n^{-1/(n-1)} \to 1$. Thus

$$E_\theta\sigma_n(\theta) \le \int_0^{c_n} n(g_\theta(v))^n dv + \int_0^\infty g_\theta(v)dv. \tag{2.28}$$

By assumption, the last term on the RHS of (2.28) is finite whereas the first term is equal to

$$\int_0^{nc_n} (1 - \int_{a_1(\theta)}^{a_1(\theta+u/n)} f(x;\theta)dx)^n du. \tag{2.29}$$

Find $\varepsilon > 0$ and $\delta > 0$ such that

$$\inf_{\theta\in K} \inf_{a_1(\theta)<x<a_1(\theta)+\delta} f(x;\theta) \geq \varepsilon > 0,$$
$$\inf_{\theta\in K} \inf_{a_2(\theta)-\delta<x<a_2(\theta)} f(x;\theta) \geq \varepsilon > 0,$$
$$\inf_{\theta\in K} \inf_{0<u<\delta} a_1'(\theta+u) \geq \varepsilon > 0,$$
$$\inf_{\theta\in K} \inf_{0<u<\delta} (-a_2'(\theta)) \geq \varepsilon > 0;$$

this is possible by the assumed conditions in the setup. Then the term in (2.29) is less than

$$\int_0^{nc_n} (1 - 2\varepsilon^2 u/n)^n du \leq \int_0^{\infty} \exp[-2\varepsilon^2 u] du < \infty,$$

completing the proof. ▱

Remark 2.2. Condition (2.26) is clearly satisfied if Θ is bounded above. In the very important case when $a_1(\theta) = \theta$ and $a_2(\theta) = \infty$, if we have $\sup_{\theta\in K} E_\theta(X_1 - \theta)^{1+\delta} < \infty$ for some $\delta > 0$, then (2.26) is satisfied. This is simply because

$$g_\theta(v) = P_\theta\{X_1 - \theta > v\} \leq v^{-(1+\delta)} E_\theta(X_1 - \theta)^{1+\delta}. \tag{2.30}$$

3. Reference Prior

Reference priors are proposed by Bernardo [7] as noninformative priors which can be thought of as a reference point against which any particular subjective prior belief can be judged. These are obtained by maximizing the expected Kullback-Leibler divergence between the posterior and the prior in an asymptotic sense. However, Bernardo's original suggestion needs to be slightly modified for technical reasons and we shall follow essentially Ghosh and Mukerjee [12].

For any (possibly improper) prior $\pi(\theta)$ on Θ and $K \subset \Theta$ compact, we write $\pi|_K$ for the proper prior defined by

$$\pi|_K(A) = \pi(A)/\pi(K), \qquad A \subset K.$$

Definition 3.1. A prior density π^* on Θ is called a *reference prior* if

(i) π^* is positive and continuous on Θ;

(ii) for all compact $K \subset \Theta$ and for all prior density π positive and continuous on Θ, we have

$$J(\pi^*|_K) \geq J(\pi|_K)$$

where $J(q) = \int_K q(\theta) \log(c(\theta)/\pi(\theta)) d\theta$.

Motivation of Condition (ii) in Definition 3.1 comes from Bernardo [7] (the present form is suggested by Ghosh and Mukerjee [12]). The first one is included for technical reasons;

nevertheless, it is an impartiality requirement. One would hardly like to call a prior as a reference prior if (i) is not satisfied.

As is immediate, the prior $\pi^*(\theta) \propto c(\theta)$ is the reference prior. (In case $S(\theta)$ is increasing, $\pi^*(\theta) \propto |c(\theta)|$.) In the particular case of a location family, we have $f(x;\theta) = f(x-\theta)$, $\theta \in \mathbb{R}$, with $f(0+) > 0$, $f(0-) = 0$, and so $c(\theta) = f(0+)$. Consequently, $\pi^*(\theta)$ is the improper uniform prior, as expected. In the case of truncation model

$$f(x;\theta) = g(x)/\bar{G}(\theta), \qquad x \geq \theta$$

where g is a smooth positive density and $\bar{G}(x) = \int_x^\infty g(y)dy$, we have $c(\theta) = g(\theta)/\bar{G}(\theta)$. Consequently, the reference prior $\pi^*(\theta)$ is proportional to the hazard rate for the density g.

Acknowledgement. The authors are indebted to Professor J. K. Ghosh for several helpful suggestions and for providing important references.

References

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

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

[3] Berger, J. O. and Bernardo, J. M.: On the development of reference priors (with discussions). In *Bayesian Statistics* 4 (J. M. Bernardo et al., eds.), Oxford University Press (1992), 35–60.

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

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

[6] Berger, J. O., Bernardo, J. M. and Mendoza, M.: 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 (1989), 1–20.

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

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

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

[10] Ghosal, Ṡ. and Samanta, T.: Asymptotic expansions of posterior distributions in nonregular cases. Technical Report, Indian Statistical Institute (1994).

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

[12] Ghosh, J. K. and Mukerjee, R.: Non-informative priors (with discussion). In *Bayesian Statistics* 4 (J. M. Bernardo et al., eds.) (1992), 195–210.

[13] Ibragimov, I. A. and Has'minskii, R. Z.: On the information contained in a sample about a parameter. In *Second International Symposium on Information Theory.* Akademiai, Kiado, Budapest (1973), 295–309.

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

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

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

[17] Rao, C. R.: *Linear Statistical Inference and Its Applications. Second edition.* John Wiley, New York (1973).

[18] Rissanen, J.: Stochastic complexity and modelling. *Ann. Statist.* **14** (1986), 1080–1100.

[19] Rissanen, J.: Stochastic complexity. *J. Roy. Statist. Soc. Ser. B* **49** (1987), 223–239.

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

Tapas Samanta
Computer Science Unit
Indian Statistical Institute
203, B. T. Road, Calcutta 700035
India