Ann Inst Stat Math (2011) 63:43–58
DOI 10.1007/s10463-009-0226-4

A general divergence criterion for prior selection

Malay Ghosh · Victor Mergel · Ruitao Liu

Received: 14 September 2006 / Revised: 26 February 2008 / Published online: 10 March 2009

Abstract The paper revisits the problem of selection of priors for regular one-parameter family of distributions. The goal is to find some "objective" or "default" prior by approximate maximization of the distance between the prior and the posterior under a general divergence criterion as introduced by Amari (Ann Stat 10:357–387, 1982) and Cressie and Read (J R Stat Soc Ser B 46:440–464, 1984). The maximization is based on an asymptotic expansion of this distance. The Kullback–Leibler, Bhattacharyya–Hellinger and Chi-square divergence are special cases of this general divergence criterion. It is shown that with the exception of one particular case, namely the Chi-square divergence, the general divergence criterion yields Jeffreys' prior. For the Chi-square divergence, we obtain a prior different from that of Jeffreys and also from that of Clarke and Sun (Sankhya Ser A 59:215–231, 1997).

Keywords Chi-square distance · Fisher information number · Jeffreys' prior

1 Introduction

The selection of priors has always been a controversial and much debated issue in the Bayesian literature. Prior elicitation is often possible with sufficient background information. However, this is not the case for most real situations, and one needs

This research was partially supported by NSF Grants SES-0317589, SES-0631426 and NSA Grant MSPF-076-097.

M. Ghosh (✉) · R. Liu
Department of Statistics, University of Florida, Gainesville, FL 32611-8545, USA
e-mail: ghoshm@stat.ufl.edu

V. Mergel
Bristol Myers Squibb, Princeton, NJ, USA

to use "objective" or "default" prior. Laplace (1812), one of the earliest proponents of such a prior, advocated the use of a uniform or a flat prior on the entire parameter space. But his prior was criticized due to its lack of invariance under one-to-one transformation. This led Jeffreys (1961) to the discovery of a prior proportional to the positive square root of the determinant of the Fisher information matrix, which has now become universally known as Jeffreys' prior. This prior is invariant under smooth one-to-one transformation.

There are various optimality results associated with Jeffreys' prior (e.g., Kass and Wasserman 1996; Bernardo 1979). Indeed, in the absence of nuisance parameters, Bernardo's reference prior is identical with Jeffreys' prior. Roughly speaking, Bernardo's reference prior is obtained by maximizing the Kullback–Leibler (KL) distance between the prior and the corresponding posterior for a given likelihood. A formal and very rigorous justification of reference priors is given in Clarke and Barron (1990, 1994). Clarke and Sun (1997, 1999) had an interesting result for the one-parameter exponential family of distributions with the canonical parameter (or any one-to-one function of it which includes the mean) as the parameter of interest. They showed in this situation that maximization of the Chi-square distance instead of the KL distance between the prior and the posterior led to a prior different from Jeffreys' prior (even in the absence of nuisance parameters).

The objective of the present paper is to extend the previous results for the regular one-parameter family of distributions under a general divergence criterion which includes not only the KL divergence, but also the Bhattacharyya–Hellinger (Bhattacharyya 1943; Hellinger 1909) divergence as well as the Chi-square divergence. An asymptotic expansion of the divergence between the prior and the posterior is provided. The prior is selected based on first or second order approximation of this divergence as needed.

The general prior is derived in Sect. 2. Some final remarks are made in Sect. 3. The proof of the main result, due to its technical nature, is deferred to the Appendix.

It turns out that with the exception of the Chi-square divergence, Jeffreys' prior turns out to be the unique optimizing prior. This is not so for the Chi-square divergence. The reason for this is whereas in other cases, a first order approximation suffices, and yields Jeffreys' prior, for the Chi-square divergence, the first order approximation does not yield any prior, and one needs to go for a second order approximation in finding the desired prior. The resulting prior turns out to be the positive fourth root rather than the positive square root of the Fisher information number. As a consequence, this leads to the Beta(3/4, 3/4) prior rather than Jeffreys' Beta(1/2, 1/2) prior for the binomial parameter θ and the prior $\pi(\theta) = \theta^{-1/4}$ rather than Jeffreys' prior $\pi(\theta) = \theta^{-1/2}$ for the Poisson parameter θ. Our prior also differs from Hartigan's (1998) maximum likelihood (ML) prior. In particular, for the one-parameter exponential family of distributions, the ML prior of Hartigan (1998) is the Fisher information number which is different from either Jeffreys' prior or the Chi-square divergence prior. Finally, our Chi-square divergence prior differs also from that of Clarke and Sun (1997) who considered the special case of the one-parameter exponential family of distributions. They used a technique different from ours in their asymptotic expansion, and obtained the reciprocal of Jeffreys' prior as their solution. It seems that there was a minor algebraic oversight in their expansion, which resulted in a prior different from ours. In particular, it appears that the third term within the integrand in Eq. 2.3 of Clarke and Sun (1997) should involve the multiplier $1/2$ instead of $1/8$.

2 Derivation of priors

Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)$, where the X_i are independent and identically distributed with common pdf (with respect to some σ-finite measure μ) $p(x \mid \theta)$. Consider a prior $\pi(\theta)$ which puts all its mass on a compact set (closed bounded interval) on the real line. One passes on to the limit eventually in many of actual examples considered in the literature.

Write

$$L_n(\theta) = \prod_{i=1}^{n} p(x_i \mid \theta)$$

and $\boldsymbol{x}_n = (x_1, \ldots, x_n)$. The posterior $\pi(\theta \mid \boldsymbol{x}_n)$ is then given by

$$\pi(\theta \mid \boldsymbol{x}_n) \propto L_n(\theta)\pi(\theta). \tag{1}$$

Also, let $m(\boldsymbol{x}_n)$ denote the marginal pdf of $\boldsymbol{x}_n$.

The general expected divergence between the prior and the posterior is given by

$$R^\beta(\pi) = \frac{1 - \int \left[\int \pi^\beta(\theta)\pi^{1-\beta}(\theta|\, \boldsymbol{x}_n)\, \mathrm{d}\theta\right] m(\boldsymbol{x}_n)\mu(\mathrm{d}\boldsymbol{x}_n)}{\beta(1-\beta)}. \tag{2}$$

where $\mu(\mathrm{d}\boldsymbol{x}_n) = (\mu(\mathrm{d}x_1), \ldots, \mu(\mathrm{d}x_n))$.

For $\beta = 0$ or 1, we need to interpret $R^\beta(\pi)$ as its limiting value (when it exists). In particular,

$$R^0(\pi) = \iint \left\{\log \frac{\pi(\theta \mid \boldsymbol{x}_n)}{\pi(\theta)}\right\} \pi(\theta \mid \boldsymbol{x}_n) m(\boldsymbol{x}_n)\, \mathrm{d}\theta\, \mu(\mathrm{d}\boldsymbol{x}_n), \tag{3}$$

which is the KL divergence between the prior and the posterior considered for example in Lindley (1956), Bernardo (1979), Clarke and Barron (1990, 1994) and Ghosh and Mukerjee (1992).

From the relation $L_n(\theta)\pi(\theta) = \pi(\theta \mid \boldsymbol{x}_n)m(\boldsymbol{x}_n)$, one can reexpress $R^\beta(\pi)$ given in (2) as

$$\begin{aligned} R^\beta(\pi) &= \frac{1 - \iint \pi^{\beta+1}(\theta)\pi^{-\beta}(\theta \mid \boldsymbol{x}_n) L_n(\theta)\, \mu(\mathrm{d}\boldsymbol{x}_n)\, \mathrm{d}\theta}{\beta(1-\beta)} \\ &= \frac{1 - \int \pi^{\beta+1}(\theta) \left[E_\theta \left[\pi^{-\beta}(\theta \mid \boldsymbol{X}_n)\right]\right] \mathrm{d}\theta}{\beta(1-\beta)}, \end{aligned} \tag{4}$$

where E_θ denotes the conditional expectation given θ.

Let $I(\theta)$ denote the per observation Fisher information number, that is $I(\theta) = E_\theta\left[-l_n''(\theta)\right]$, where $l_n(\theta) = n^{-1}\log L_n(\theta)$. Also, let $g_3(\theta) = E_\theta\left[-l_n'''(\theta)\right]$, and $g_4(\theta) = E_\theta\left[-l_n''''(\theta)\right]$, where the appropriate derivatives are asumed to exist. Then one has the following theorem.

Theorem 1 *Assume*

(A_1) *For each* x, $\log p(x|\theta)$ *is* 5 *times continuously differentiable in* θ. *Further, there exist a neighborhood* $N_\theta(\delta) = (\theta - \delta, \theta + \delta)$ *of* θ *and* P_θ *integrable function* $H_{k,\theta,\delta}(x)$, $k = 2, 3, 4, 5$ *such that*

$$\sup_{\theta' \in N_\theta(\delta)} \left| \frac{\partial k}{\partial \theta^k} \log p(x|\theta) \right|_{\theta=\theta'} \le H_{k,\theta,\delta}(x)$$

(A_2) *For all sufficiently large* $\lambda > 0$

$$E\left[\sup_{|\theta'-\theta|>\lambda} \log \frac{p(x|\theta')}{p(x|\theta)}\right] < 0$$

(A_3)

$$E\left[\sup_{\theta' \in (\theta-\rho,\theta+\rho)} \log p(x|\theta') \middle| \theta\right] \to E[\log p(x|\theta)]$$

as $\rho \to 0$

(A_4) *The prior density* $\pi(\theta)$ *is* 3 *times continuously differentiable in a neighborhood of* θ *and* $\pi(\theta) > 0$.

Then for $\beta < 1$,

$$E_\theta\left[\pi_\pi^{-\beta}(\theta \mid \boldsymbol{x}_n)\right] = \left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}} \frac{1}{(1-\beta)^{1/2}} \left[1 + \frac{\beta}{2n}\left\{\frac{\beta g_3(\theta) - (2+\beta)I'(\theta)}{(1-\beta)I^2(\theta)}\frac{\pi'(\theta)}{\pi(\theta)} - \frac{\beta}{I(\theta)}\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2 + \frac{\beta(2-\beta)}{(1-\beta)I(\theta)}\frac{\pi''(\theta)}{\pi(\theta)} + k(\theta)\right\}\right] + o(n^{-1}), \tag{5}$$

where

$$k(\theta) = \frac{5\beta(2-\beta)\,(g_3(\theta))^2}{12(1-\beta)I^3(\theta)} - \frac{\beta(2-\beta)g_4(\theta)}{4(1-\beta)I^2(\theta)} + \frac{\beta}{1-\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{g_3(\theta)}{I^2(\theta)} + \frac{\beta(2+\beta)(4+\beta)}{4(1-\beta)}\frac{I'(\theta)}{I^2(\theta)}\left[\frac{I'(\theta)}{I(\theta)} + I''(\theta)\right]$$

and does not involve $\pi(\theta)$ *or its derivatives.*

The proof of the result is very technical and is included in the Appendix. The proof uses a technique proposed by Ghosh and is discussed in details in Datta and Mukerjee (2004).

Remark 1 The theorem does not hold for $\beta \geq 1$ as is evident from the right-hand-side expression in (5).

In view of Theorem 1, for $\beta < 1$ and $\beta \neq 0$ or -1, one has

$$R^{\beta}(\pi) = \frac{1 - (1/2)\left(\frac{2\pi}{n}\right)^{\beta/2}(1-\beta)^{-\frac{1}{2}} \int \left\{\frac{\pi(\theta)}{I^{\frac{1}{2}}(\theta)}\right\}^{\beta} \pi(\theta)\,\mathrm{d}\theta}{\beta(1-\beta)} + o(n^{-\frac{\beta}{2}-1}). \tag{6}$$

Thus the first order approximation to $R^{\beta}(\pi)$ is given by

$$\frac{1 - (1/2)\left(\frac{2\pi}{n}\right)^{\beta/2}(1-\beta)^{-\frac{1}{2}} \int \left\{\frac{\pi(\theta)}{I^{\frac{1}{2}}(\theta)}\right\}^{\beta} \pi(\theta)\,\mathrm{d}\theta}{\beta(1-\beta)}. \tag{7}$$

We now consider maximization of (7) in several different cases noting $\int \pi(\theta)\mathrm{d}\theta = 1$. It turns out that there is no maximizer when $\beta < -1$.

Case I $0 < \beta < 1$.

Here it suffices to minimize

$$\int \left\{\pi(\theta)/I^{1/2}(\theta)\right\}^{\beta} \pi(\theta)\,\mathrm{d}\theta$$

with respect to $\pi(\cdot)$ subject to $\int \pi(\theta)\,\mathrm{d}\theta = 1$. To this end, we first recall Hölders inequality, namely

$$\int |f_1(\theta) f_2(\theta)|\,\mathrm{d}\theta \leq \left(\int |f_1(\theta)|^p\,\mathrm{d}\theta\right)^{\frac{1}{p}} \left(\int |f_2(\theta)|^q\,\mathrm{d}\theta\right)^{\frac{1}{q}} \tag{8}$$

where $p > 1, q > 1$ and $p^{-1} + q^{-1} = 1$.

Now putting $f_1(\theta) = \pi(\theta) I^{-\frac{\beta}{2(1+\beta)}}(\theta)$, $f_2(\theta) = I^{\frac{\beta}{2(1+\beta)}}(\theta)$, $p = 1+\beta$ and $q = (1+\beta)/\beta$, one gets

$$\left(\int \pi^{1+\beta}(\theta) I^{-\frac{\beta}{2}}(\theta)\,\mathrm{d}\theta\right)^{\frac{1}{1+\beta}} \left(\int I^{\frac{1}{2}}(\theta)\,\mathrm{d}\theta\right)^{\frac{\beta}{1+\beta}} \geq \int \pi(\theta)\,\mathrm{d}\theta = 1$$

or equivalently

$$\int \pi^{1+\beta}(\theta) I^{-\frac{\beta}{2}}(\theta)\,\mathrm{d}\theta \geq \left(\int I^{\frac{1}{2}}(\theta)\,\mathrm{d}\theta\right)^{-\beta}$$

with equality if and only if

$$\pi(\theta) \propto I^{\frac{1}{2}}(\theta),$$

that is $\pi(\theta)$ is Jeffreys' prior.

Case II $-1 < \beta < 0$.

One needs to maximize

$$\int \left\{\pi(\theta)/I^{1/2}(\theta)\right\}^{\beta} \pi(\theta)\, \mathrm{d}\theta$$

subject to $\int \pi(\theta)\mathrm{d}\theta = 1$.

Here let $f_1(\theta) = \pi^{1+\beta}(\theta)$, $f_2(\theta) = I^{-\frac{\beta}{2}}(\theta)$, $p = (1+\beta)^{-1}$ and $q = -1/\beta$. Then, again by Hölders inequality

$$\int \pi^{1+\beta}(\theta) I^{-\frac{\beta}{2}}(\theta)\, \mathrm{d}\theta \leq \left(\int I^{\frac{1}{2}}(\theta)\, \mathrm{d}\theta\right)^{-\beta}$$

with equality if and only if

$$\pi(\theta) \propto I^{\frac{1}{2}}(\theta),$$

which is once again Jeffreys' prior.

Case III $\beta < -1$.

Here putting $\beta = -\lambda$, we rewrite

$$R^{\beta}(\pi) = \frac{(1/2)\left(\frac{2\pi}{n}\right)^{-\lambda/2}(1+\lambda)^{-\frac{1}{2}} \int \left(\frac{I^{1/2}(\theta)}{\pi(\theta)}\right)^{\lambda} \pi(\theta)\, \mathrm{d}\theta - 1}{\lambda(1+\lambda)}, \quad \lambda > 1 \tag{9}$$

Hence it suffices to maximize

$$\int \left\{I^{1/2}(\theta)/\pi(\theta)\right\}^{\lambda} \pi(\theta)\, \mathrm{d}\theta$$

subject to $\int \pi(\theta)\, \mathrm{d}\theta = 1$.

Writing $\int \left\{I^{1/2}(\theta)/\pi(\theta)\right\}^{\lambda} \pi(\theta)\, \mathrm{d}\theta = E\left[I^{1/2}(\theta)/\pi(\theta)\right]^{\lambda}$, expectation being taken with respect to the prior $\pi(\cdot)$, one gets

$$\int \left\{I^{1/2}(\theta)/\pi(\theta)\right\}^{\lambda} \pi(\theta)\, \mathrm{d}\theta \geq \left[\int \left\{I^{1/2}(\theta)/\pi(\theta)\right\} \pi(\theta)\, \mathrm{d}\theta\right]^{\lambda} = \left(\int I^{1/2}(\theta)\, \mathrm{d}\theta\right)^{-\beta}$$

since $\lambda > 1$. Equality holds if and only if

$$\pi(\theta) \propto I^{\frac{1}{2}}(\theta).$$

Thus in this case Jeffreys' prior is the minimizer rather then the maximizer of $R^{\beta}(\pi)$. Also there is no maximizing prior in this case. To see this, let

$$r(\pi(\theta)) = \left\{I^{1/2}(\theta)/\pi(\theta)\right\}^{\lambda} \pi(\theta) - \eta\pi(\theta),$$

where η is the Lagrangian multiplier. Hence,

$$r'(\pi(\theta)) = (1-\lambda)I^{\lambda/2}(\theta)\pi^{-\lambda}(\theta) - \eta$$

so that equating $r'(\pi(\theta))$ to zero, $\pi(\theta) \propto I^{\frac{1}{2}}(\theta)$. Accordingly, the only inflection prior is Jeffreys' prior which is the minimizer and not the maximizer.

Case IV $\beta \to 0$, from (7) one obtains an expression due to Clarke and Barron (1990, 1994), namely,

$$R^0(\pi) = \frac{1}{2}\log\left(\frac{n}{2\pi e}\right) - \int \pi(\theta)\log\frac{\pi(\theta)}{I^{1/2}(\theta)}\,\mathrm{d}\theta + o(1)$$

which is maximized up to first order approximation by $\int \pi(\theta)\log\frac{\pi(\theta)}{I^{1/2}(\theta)}\,\mathrm{d}\theta = 0$ i.e. $\pi(\theta) = I^{1/2}(\theta)$.

Case V The only remaining case is $\beta = -1$, the Chi-square distance as considered by Clarke and Sun (1997) for the one parameter exponential family. We consider the general framework in this paper. Here $\pi^{1+\beta}(\theta) = 1$ so that the first order term as obtained from Theorem 1 is a constant, and one needs to consider the second order term. Leaving out all the terms which do not involve $\pi(\theta)$ or its derivatives and putting $\beta = -1$ in (5), and noting that $\beta(1-\beta) = -2 < 0$, it suffices to maximize

$$\int \left[\frac{I'(\theta) + g_3(\theta)}{2I^{3/2}(\theta)}\frac{\pi'(\theta)}{\pi(\theta)} + \frac{1}{I^{1/2}(\theta)}\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2 - \frac{3}{2I^{1/2}(\theta)}\frac{\pi''(\theta)}{\pi(\theta)}\right]\mathrm{d}\theta. \tag{10}$$

Putting

$$y(\theta) = \frac{\pi'(\theta)}{\pi(\theta)},$$

the integral in (10) is rewritten as

$$\begin{aligned}
&\int \left[\frac{I'(\theta) + g_3(\theta)}{2I^{3/2}(\theta)}y(\theta) + \frac{y^2(\theta)}{I^{1/2}(\theta)} - \frac{3}{2I^{1/2}(\theta)}[y'(\theta) + y^2(\theta)]\right]\mathrm{d}\theta \\
&= \int \left[\frac{I'(\theta) + g_3(\theta)}{2I^{3/2}(\theta)}y(\theta) - \frac{y^2(\theta)}{2I^{1/2}(\theta)} - \frac{3y'(\theta)}{2I^{1/2}(\theta)}\right]\mathrm{d}\theta. \qquad (11)
\end{aligned}$$

Writing the integrand as $s(y(\theta), y'(\theta))$, the solution is found by solving the Euler–Lagrange equation (Hewitt and Stromberg 1969)

$$\frac{\partial s}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\partial s}{\partial y'}\right) = 0$$

or equivalently

$$\frac{I'(\theta) + g_3(\theta)}{2I^{3/2}(\theta)} - \frac{y(\theta)}{I^{1/2}(\theta)} + \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{3}{2I^{1/2}(\theta)}\right) = 0.$$

This leads to equation

$$\frac{I'(\theta) + g_3(\theta)}{2I^{3/2}(\theta)} - \frac{3I'(\theta)}{4I^{3/2}(\theta)} = \frac{y(\theta)}{I^{1/2}(\theta)},$$

that is

$$\frac{\pi'(\theta)}{\pi(\theta)} = y(\theta) = \frac{2g_3(\theta) - I'(\theta)}{4I(\theta)}.$$

Accordingly, the solution is given by

$$\pi(\theta) \propto \exp\left[\int \frac{2g_3(\theta) - I'(\theta)}{4I(\theta)}\,\mathrm{d}\theta\right].$$

It is interesting to consider some examples where one gets priors different from Jeffreys' prior under the above divergence criterion.

Example 1 Consider the one-parameter exponential family of distributions with $p(x|\theta) = \exp[\theta x - \psi(\theta) + h(x)]$. Then $g_3(\theta) = I'(\theta)$ so that $\pi(\theta) \propto \exp\left[\frac{1}{4}\int \frac{I'(\theta)}{I(\theta)}\mathrm{d}\theta\right]$ $= I^{\frac{1}{4}}(\theta)$. Thus, in particular, for the Binomial(n, θ) problem, one gets $\pi(\theta) \propto \theta^{-\frac{1}{4}}(1-\theta)^{-\frac{1}{4}}$ which is a Beta$(\frac{3}{4}, \frac{3}{4})$ distribution quite different from Jeffreys' Beta$(\frac{1}{2}, \frac{1}{2})$ prior or Laplace's Beta(1, 1) prior or Haldane's improper Beta(0, 0) prior. Similarly, for the Poisson(θ) case, one gets $\pi(\theta) \propto \theta^{-\frac{1}{4}}$, again different from Jeffreys' $\pi_J(\theta) \propto \theta^{-\frac{1}{2}}$ prior. However, for the $N(\theta, 1)$ situation, since $I(\theta) = 1$ and $g_3(\theta) = I'(\theta) = 0$, $\pi(\theta) = c(> 0)$, a constant, which is the same as Jeffreys' prior. We may point out also that for the one-parameter exponential family, our prior differs from Hartigan's (1998) maximum likelihood prior $\pi_H(\theta) = I(\theta)$.

Example 2 For the one-parameter location family of distributions with $p(x|\theta) = f(x - \theta)$, where f is a pdf, both $g_3(\theta)$ and $I(\theta)$ are constants implying $I'(\theta) = 0$ so that $\pi(\theta)$ is of the form $\pi(\theta) = \exp(\mu\theta)$ for some constant μ. However, for the special case of a symmetric f, i.e. $f(x) = f(-x)$ for all x, $g_3(\theta) = 0$, and then $\pi(\theta)$ reduces once again to $\pi(\theta) = c$ which is the same as Jeffreys' prior.

Example 3 For the general scale family of distributions with $p(x|\theta) = \theta^{-1} f(\frac{x}{\theta})$, $\theta > 0$, where f is a pdf, $I(\theta) = \frac{c_1}{\theta^2}$ for some constant $c_1 (> 0)$, where $g_3(\theta) = \frac{c_2}{\theta^3}$ for some constant c_2. Then $\pi(\theta) \propto \exp(c \log \theta) = \theta^c$ for some constant c. In particular, when $p(x|\theta) = \theta^{-1} \exp(-\frac{x}{\theta})$, $\pi(\theta) \propto \theta^{-\frac{3}{2}}$, different from Jeffreys' $\pi(\theta) \propto \theta^{-1}$ for the general scale family of distributions.

3 Summary

The paper considers a general divergence measure between the prior and the posterior for regular one-parameter distributions to develop objective priors. The one-parameter exponential family appears as a special case. It follows that with one exception, Jeffreys' prior is the desired objective prior. The exception, the Chi-squared divergence yields the positive fourth root rather than the positive square root of the Fisher information number.

There is an enormous scope for future research. The first one to look at is a regular multiparameter family of distributions without any nuisance parameter, and examine whether a similar result holds. The second and possibly the more important investigation is to develop priors under this general divergence criterion in the presence of nuisance parameters.

Acknowledgments The paper has benefitted from the comments of two anonymous referees. The research was partially supported by a Natural Security Agency Grant MSPF-07G-097.

Appendix

Proof of Theorem 1 Let $\hat\theta$ denote the MLE of θ, $a_2 = l_n''(\hat\theta)$, $c = -a_2$, $a_3 = l_n'''(\hat\theta)$, $a_4 = l_n''''(\hat\theta)$ and $h = \sqrt{n}(\theta - \hat\theta)$. Then from (2.2.19) of Datta and Mukerjee (2004, p. 13) one gets

$$\pi(h|\boldsymbol{x}_n) = \hat\pi(h|\boldsymbol{x}_n) + o(n^{-1}), \tag{12}$$

where

$$\begin{aligned}\hat\pi(h|\boldsymbol{x}_n) = \sqrt{\frac{c}{2\pi}} \exp\left\{-\frac{ch^2}{2}\right\} \Bigg[1 + \frac{1}{\sqrt{n}} \left\{ \frac{\pi'(\hat\theta)}{\pi(\hat\theta)} h + \frac{1}{6} a_3 h^3 \right\} \\ + \frac{1}{n} \Bigg\{ \frac{1}{2} \left(\frac{\pi''(\hat\theta)}{\pi(\hat\theta)} h^2 - \frac{\pi''(\hat\theta)}{\pi(\hat\theta) c} \right) + \frac{1}{6} \left(a_3 \frac{\pi'(\hat\theta)}{\pi(\hat\theta)} h^4 - \frac{3a_3}{c^2} \frac{\pi'(\hat\theta)}{\pi(\hat\theta)} \right) \\ + \frac{1}{24} \left(a_4 h^4 - \frac{3a_4}{c^2} \right) + \frac{1}{72} \left(a_3^2 h^6 - \frac{15 a_3^2}{c^3} \right) \Bigg\} \Bigg].\end{aligned} \tag{13}$$

With the general expansion

$$\left(\frac{1}{b_1 + \frac{b_2}{\sqrt{n}} + \frac{b_3}{n} + o(n^{-1})} \right)^\beta = b_1^{-\beta} \left(1 - \beta \frac{b_2}{b_1 \sqrt{n}} + \frac{\beta}{n} \left(\frac{\beta+1}{2} \frac{b_2^2}{b_1^2} - \frac{b_3}{b_1} \right) \right) + o(n^{-1}),$$

one gets

$$\pi^{-\beta}(h|\boldsymbol{x}_n) = \hat{\pi}_\beta(h|\boldsymbol{x}_n) + o(n^{-1}), \tag{14}$$

where

$$\begin{aligned}\hat{\pi}_\beta(h|\boldsymbol{x}_n) = \left(\frac{c}{2\pi}\right)^{-\frac{\beta}{2}} \exp\left\{\frac{c\beta h^2}{2}\right\} \Bigg[1 - \frac{\beta}{\sqrt{n}} \left\{ \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} h + \frac{1}{6} a_3 h^3 \right\} \\ + \frac{\beta}{n} \Bigg\{ \frac{1+\beta}{2} \left(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} h + \frac{1}{6} a_3 h^3 \right)^2 - \frac{1}{2} \left(\frac{\pi''(\hat{\theta})}{\pi(\hat{\theta})} h^2 - \frac{\pi''(\hat{\theta})}{\pi(\hat{\theta}) c} \right) \\ - \frac{1}{6} \left(a_3 \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} h^4 - \frac{3a_3}{c^2} \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} \right) \\ - \frac{1}{24} \left(a_4 h^4 - \frac{3a_4}{c^2} \right) - \frac{1}{72} \left(a_3^2 h^6 - \frac{15 a_3^2}{c^3} \right) \Bigg\} \Bigg]. \end{aligned} \tag{15}$$

By (12) and (14), for a prior $\bar{\pi}$ (not the same as π) twice differentiable in its argument, we get

$$\begin{aligned} \pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n) = \left(\frac{c}{2\pi}\right)^{\frac{1-\beta}{2}} \exp\left\{-\frac{c(1-\beta)h^2}{2}\right\} \\ \times \Bigg[1 + \frac{1}{\sqrt{n}} \left\{ \frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} h - \beta \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} h + \frac{1-\beta}{6} a_3 h^3 \right\} \\ + \frac{1}{n} \Bigg\{ -\beta \left(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} \frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} h^2 + \frac{1}{6} a_3 h^4 \left(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} + \frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} \right) + \frac{1}{36} a_3^2 h^6 \right) \\ + \frac{\beta(1+\beta)}{2} \left(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} h + \frac{1}{6} a_3 h^3 \right)^2 - \frac{\beta}{2} \left(\frac{\pi''(\hat{\theta})}{\pi(\hat{\theta})} h^2 - \frac{\pi''(\hat{\theta})}{\pi(\hat{\theta}) c} \right) \\ + \frac{1}{2} \left(\frac{\bar{\pi}''(\hat{\theta})}{\bar{\pi}(\hat{\theta})} h^2 - \frac{\bar{\pi}''(\hat{\theta})}{\bar{\pi}(\hat{\theta}) c} \right) + \frac{1}{6} \Bigg(a_3 \frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} h^4 - \frac{3a_3}{c^2} \frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} \\ - \beta \left(a_3 \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} h^4 - \frac{3a_3}{c^2} \frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} \right) \Bigg) + \frac{1-\beta}{24} \left(a_4 h^4 - \frac{3a_4}{c^2} \right) \\ + \frac{1-\beta}{72} \left(a_3^2 h^6 - \frac{15 a_3^2}{c^3} \right) \Bigg\} \Bigg] + o(n^{-1}). \end{aligned} \tag{16}$$

The right-hand-side of (16) can be expressed as

$$\hat{\pi}_\beta(h|\boldsymbol{x}_n)\hat{\bar{\pi}}(h|\boldsymbol{x}_n) + n^{-\frac{3}{2}} \left(\frac{c}{2\pi}\right)^{\frac{1-\beta}{2}} \exp\left\{-\frac{c(1-\beta)h^2}{2}\right\} R(h) + o(n^{-1}), \tag{17}$$

where $R(h)$ is a polynomial of h. This leads to

$$\Big|\pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\hat{\bar{\pi}}(h|\boldsymbol{x}_n) - n^{-\frac{3}{2}}\left(\frac{c}{2\pi}\right)^{\frac{1-\beta}{2}} \exp\left\{-\frac{c(1-\beta)h^2}{2}\right\} R(h)\Big| = o(n^{-1}). \tag{18}$$

However, we need to show

$$\int \Big|\pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\hat{\bar{\pi}}(h|\boldsymbol{x}_n) - n^{-\frac{3}{2}}\left(\frac{c}{2\pi}\right)^{\frac{1-\beta}{2}} \exp\left\{-\frac{c(1-\beta)h^2}{2}\right\} R(h)\Big| \, dh = o(n^{-1}).$$

To this end, it suffices to show that

$$\int \left|\pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\hat{\bar{\pi}}(h|\boldsymbol{x}_n)\right| dh = o(n^{-1}). \tag{19}$$

To prove (19), we need the following lemma:

Lemma 2 *There exist $\delta > 0$ and $\epsilon > 0$ (ϵ depends on δ), such that with probability one, and for all large n*

1. $$\log Z_n(h) \le -M\frac{h^2}{2}, \quad -n^{\frac{1}{2}}\delta \le h \le n^{\frac{1}{2}}\delta \quad \text{some } M > 0$$

2. $$\log Z_n(h) \le -n\varepsilon, \quad \forall h < -n^{\frac{1}{2}}\delta$$

3. $$\log Z_n(h) \le -n\varepsilon, \quad \forall h > n^{\frac{1}{2}}\delta$$

where $Z_n(h) = \prod_{i=1}^n \frac{p(x_i|\hat{\theta}_n + h/\sqrt{n})}{p(x_i|\hat{\theta}_n)}$, $\hat{\theta}_n$ is MLE. Based on assumptions (A_1), (A_2), (A_3), applying the argument similar to that used by Ghosal and Samanta (1997a,b) and Ghosal (1997), one can prove the Lemma 2.

Notice that

$$\begin{aligned}
&\int \left|\pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\hat{\bar{\pi}}(h|\boldsymbol{x}_n)\right| \mathrm{d}h \\
&\le \int_{-\infty}^{-n^{\frac{1}{2}}\delta} \left|\pi^{-\beta}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\right| \bar{\pi}(h|\boldsymbol{x}_n)\mathrm{d}h \\
&\quad + \int_{-n^{\frac{1}{2}}\delta}^{n^{\frac{1}{2}}\delta} \left|\pi^{-\beta}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\right| \bar{\pi}(h|\boldsymbol{x}_n)\mathrm{d}h \\
&\quad + \int_{n^{\frac{1}{2}}\delta}^{+\infty} \left|\pi^{-\beta}(h|\boldsymbol{x}_n) - \hat{\pi}_\beta(h|\boldsymbol{x}_n)\right| \bar{\pi}(h|\boldsymbol{x}_n)\mathrm{d}h \\
&\quad + \int_{-\infty}^{-n^{\frac{1}{2}}\delta} \left|\hat{\pi}_\beta(h|\boldsymbol{x}_n)\right| \left|\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\bar{\pi}}(h|\boldsymbol{x}_n)\right| \mathrm{d}h \\
&\quad + \int_{-n^{\frac{1}{2}}\delta}^{n^{\frac{1}{2}}\delta} \left|\hat{\pi}_\beta(h|\boldsymbol{x}_n)\right| \left|\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\bar{\pi}}(h|\boldsymbol{x}_n)\right| \mathrm{d}h \\
&\quad + \int_{n^{\frac{1}{2}}\delta}^{+\infty} \left|\hat{\pi}_\beta(h|\boldsymbol{x}_n)\right| \left|\bar{\pi}(h|\boldsymbol{x}_n) - \hat{\bar{\pi}}(h|\boldsymbol{x}_n)\right| \mathrm{d}h
\end{aligned}$$

Based on assumption (A_4), by suitably choosing δ and using Lemma 2, one can show each integration on the right side of the inequality is $o(n^{-1})$. Now (19) holds.

So, by (18), integration of (16) with respect to h will lead to

$$\begin{aligned}
E^{\bar{\pi}}\left[\pi^{-\beta}(h|\boldsymbol{x}_n)\,|\,\boldsymbol{x}_n\right] &= \int \pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n)\,\mathrm{d}h = \left(\frac{c}{2\pi}\right)^{-\frac{\beta}{2}} \frac{1}{\sqrt{1-\beta}} \\
&\times \Bigg[1 + \frac{1}{n}\Bigg\{-\frac{\beta}{c(1-\beta)}\Bigg(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})}\frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} + \frac{a_3}{2c(1-\beta)} \\
&\times\left(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} + \frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})}\right) + \frac{5a_3^2}{12c^2(1-\beta)^2}\Bigg) \\
&+ \frac{\beta(1+\beta)}{2c(1-\beta)}\left(\left(\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})}\right)^2 + \frac{a_3}{c(1-\beta)}\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} + \frac{15a_3^2}{36c^2(1-\beta)^2}\right) \\
&+ \frac{\beta}{2c(1-\beta)}\frac{\bar{\pi}''(\hat{\theta})}{\bar{\pi}(\hat{\theta})} - \frac{\beta^2}{2c(1-\beta)}\frac{\pi''(\hat{\theta})}{\pi(\hat{\theta})} + \frac{a_3\beta(2-\beta)}{2c^2(1-\beta)^2}\frac{\bar{\pi}'(\hat{\theta})}{\bar{\pi}(\hat{\theta})} \\
&- \frac{a_3\beta^2(2-\beta)}{2c^2(1-\beta)^2}\frac{\pi'(\hat{\theta})}{\pi(\hat{\theta})} + \frac{a_4\beta(2-\beta)}{8c^2(1-\beta)} \\
&+ \frac{5a_3^2(\beta^3 - 3\beta^2 + 3\beta)}{24c^3(1-\beta)^2}\Bigg\}\Bigg] + o(n^{-1}). \qquad (20)
\end{aligned}$$

From the relation $\theta = h/\sqrt{n} + \hat{\theta}$ we get

$$E^{\bar{\pi}}\left[\pi^{-\beta}(\theta|\boldsymbol{x}_n)|\boldsymbol{x}_n\right] = n^{-\frac{\beta}{2}}\int \pi^{-\beta}(h|\boldsymbol{x}_n)\bar{\pi}(h|\boldsymbol{x}_n)\,\mathrm{d}h. \tag{21}$$

Now taking expectation over the conditional distribution of $\boldsymbol{x}_n$ given θ from (20) and (21), we have

$$\begin{aligned}
\lambda(\theta) = E_\theta\left(E^{\bar{\pi}}\left[\pi^{-\beta}(\theta|\boldsymbol{x}_n)|\boldsymbol{x}_n\right]\right) &= \left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\\
&\times\Bigg[1+\frac{1}{n}\Bigg\{-\frac{\beta}{(1-\beta)I(\theta)}\left(\frac{\pi'(\theta)}{\pi(\theta)}\frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}-\frac{g_3(\theta)}{2(1-\beta)I(\theta)}\right.\\
&\times\left.\left(\frac{\pi'(\theta)}{\pi(\theta)}+\frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}\right)+\frac{5(g_3(\theta))^2}{12(1-\beta)^2I^2(\theta)}\right)\\
&+\frac{\beta(1+\beta)}{2(1-\beta)I(\theta)}\left(\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2-\frac{g_3(\theta)}{(1-\beta)I(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}+\frac{5(g_3(\theta))^2}{12(1-\beta)^2I^2(\theta)}\right)\\
&+\frac{\beta}{2(1-\beta)I(\theta)}\left(\frac{\bar{\pi}''(\theta)}{\bar{\pi}(\theta)}-\beta\frac{\pi''(\theta)}{\pi(\theta)}\right)-\frac{g_3(\theta)\beta(2-\beta)}{2(1-\beta)^2I^2(\theta)}\left(\frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta)}-\beta\frac{\pi'(\theta)}{\pi(\theta)}\right)\\
&-\frac{g_4(\theta)\beta(2-\beta)}{8(1-\beta)I^2(\theta)}+\frac{5(g_3(\theta))^2\beta(\beta^2-3\beta+3)}{24(1-\beta)^2I^3(\theta)}\Bigg\}\Bigg]+o(n^{-1-\beta/2}).
\end{aligned} \tag{22}$$

In the next step, we find

$$\begin{aligned}
\int\lambda(\theta)\bar{\pi}(\theta)\,\mathrm{d}\theta &= \int\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\\
&\times\Bigg[1+\frac{1}{n}\Bigg\{\frac{g_3(\theta)\beta}{2(1-\beta)^2I^2(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}-\frac{5\beta(g_3(\theta))^2}{12(1-\beta)^3I^3(\theta)}+\frac{\beta(1+\beta)}{2(1-\beta)I(\theta)}\\
&\times\left(\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2-\frac{g_3(\theta)}{(1-\beta)I(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}+\frac{5(g_3(\theta))^2}{12(1-\beta)^2I^2(\theta)}\right)-\frac{\beta^2}{2(1-\beta)I(\theta)}\frac{\pi''(\theta)}{\pi(\theta)}\\
&+\frac{g_3(\theta)\beta^2(2-\beta)}{2(1-\beta)^2I^2(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}-\frac{g_4(\theta)\beta(2-\beta)}{8(1-\beta)I^2(\theta)}+\frac{5(g_3(\theta))^2\beta(\beta^2-3\beta+3)}{24(1-\beta)^2I^3(\theta)}\Bigg\}\Bigg]\bar{\pi}(\theta)\,\mathrm{d}\theta\\
&+\int\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\left[\frac{1}{n}\left(-\frac{\beta}{(1-\beta)I(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}+\frac{g_3(\theta)\beta}{2(1-\beta)^2I^2(\theta)}\right.\right.\\
&\left.\left.-\frac{g_3(\theta)\beta(2-\beta)}{2(1-\beta)^2I^2(\theta)}\right)\bar{\pi}'(\theta)\right]\mathrm{d}\theta\\
&+\int\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{\beta}{2nI(\theta)(1-\beta)^{3/2}}\bar{\pi}''(\theta)\,\mathrm{d}\theta+o(n^{-1-\beta/2}).
\end{aligned} \tag{23}$$

The last step will give an expression for $E_\theta\left[\pi^{-\beta}(\theta|\boldsymbol{x}_n)\right]$. We consider $\bar{\pi}(\theta)$ to converge weakly to the degenerate prior at true θ and have chosen $\bar{\pi}(\theta)$ in such a way that we could integrate the last two integrals in (23) by parts and have the first term equal to zero every time we use integration by parts.

Thus we have

$$
\begin{aligned}
E_\theta\left[\pi^{-\beta}(\theta|\boldsymbol{x}_n)\right] = &\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}} \frac{1}{\sqrt{1-\beta}}\left[1+\frac{1}{n}\left(\frac{g_3(\theta)\beta^2}{2(1-\beta)I^2(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}\right.\right.\\
&+\frac{5(g_3(\theta))^2\beta(2-\beta)}{24(1-\beta)I^3(\theta)}+\frac{\beta(1+\beta)}{2(1-\beta)I(\theta)}\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2\\
&\left.\left.-\frac{\beta^2}{2(1-\beta)I(\theta)}\frac{\pi''(\theta)}{\pi(\theta)}-\frac{g_4(\theta)\beta(2-\beta)}{8(1-\beta)I^2(\theta)}\right)\right]\\
&+\frac{1}{n}\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\left(\frac{\beta}{(1-\beta)I(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}\right.\right.\\
&\left.\left.+\frac{g_3(\theta)\beta}{2(1-\beta)I^2(\theta)}\right)\right]\\
&+\frac{1}{n}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left[\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\frac{\beta}{2(1-\beta)I(\theta)}\right]+o(n^{-1-\beta/2}).
\end{aligned}
\tag{24}
$$

Finally, we have

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\left(\frac{\beta}{(1-\beta)I(\theta)}\frac{\pi'(\theta)}{\pi(\theta)}+\frac{g_3(\theta)\beta}{2(1-\beta)I^2(\theta)}\right)\right]\\
&=\left(\frac{2\pi}{n}\right)^{\frac{\beta}{2}}\frac{\beta}{(1-\beta)^{3/2}}\times\left[-(1+\frac{\beta}{2})\frac{\pi'(\theta)}{\pi(\theta)}I^{-2-\beta/2}(\theta)I'(\theta)\right.\\
&\quad\left.+\left\{\frac{\pi''(\theta)}{\pi(\theta)}-\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2\right\}I^{-1-\beta/2}(\theta)+\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{g_3(\theta}{I^2(\theta)}\right)\right],
\end{aligned}
\tag{25}
$$

and

$$
\begin{aligned}
&\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left[\left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}}\frac{1}{\sqrt{1-\beta}}\frac{\beta}{2(1-\beta)I(\theta)}\right]\\
&=\left(\frac{2\pi}{n}\right)^{\frac{\beta}{2}}\frac{\beta}{2(1-\beta)^{3/2}}\left(1+\frac{1}{2}\beta\right)\left(2+\frac{1}{2}\beta\right)\\
&\quad\times I^{-2-\beta/2}(\theta)I'(\theta)\left[I^{-1}(\theta)I'(\theta)+I''(\theta)\right].
\end{aligned}
\tag{26}
$$

From (24)–(26), we get

$$E_\theta\left[\pi^{-\beta}(\theta|\boldsymbol{x}_n)\right] = \left(\frac{2\pi}{nI(\theta)}\right)^{\frac{\beta}{2}} \frac{1}{\sqrt{1-\beta}} \left[1 + \frac{1}{n}\left(\frac{g_3(\theta)\beta^2 - \beta(2+\beta)I'(\theta)}{2(1-\beta)I^2(\theta)} \frac{\pi'(\theta)}{\pi(\theta)} \right.\right.$$
$$\left.\left. + \frac{\beta(2-\beta)}{2(1-\beta)I(\theta)} \frac{\pi''(\theta)}{\pi(\theta)} - \frac{\beta}{2I(\theta)}\left(\frac{\pi'(\theta)}{\pi(\theta)}\right)^2 + \frac{1}{2}k(\theta)\right)\right]$$
$$+ o(n^{-1-\beta/2}). \tag{27}$$

with

$$k(\theta) = \frac{5\beta(2-\beta)\,(g_3(\theta))^2}{12(1-\beta)I^3(\theta)} - \frac{\beta(2-\beta)g_4(\theta)}{4(1-\beta)I^2(\theta)} + \frac{\beta}{1-\beta}\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{g_3(\theta)}{I^2(\theta)}$$
$$+ \frac{\beta(2+\beta)(4+\beta)}{4(1-\beta)} \frac{I'(\theta)}{I^2(\theta)}\left[\frac{I'(\theta)}{I(\theta)} + I''(\theta)\right]$$

This completes the proof. □

References

Amari, S. (1982). Differential geometry of curved exponential families—curvatures and information loss. *The Annals of Statistics*, *10*, 357–387.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Series B*, *41*, 113–147.

Bhattacharyya, A. K. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, *35*, 99–109.

Clarke, B., Barron, A. (1990). Information-theoretic asymptotics of Bayes methods. *IEEE Transactions on Information Theory*, *36*, 453–471.

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

Clarke, B., Sun, D. (1997). Reference priors under the chi-square distance. *Sankhya, Series A*, *59*, 215–231.

Clarke, B., Sun, D. (1999). Asymptotics of the expected posterior. *Annals of the Institute of Statistical Mathematics*, *51*, 163–185.

Cressie, N., Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, *46*, 440–464.

Datta, G. S., Mukerjee, R. (2004). *Probability matching priors: higher order asymptotics*. New York: Springer.

Ghosal, S. (1997). Reference prior in multiparameter nonregular cases. *Test*, *6*(1), 159–186.

Ghosal, S., Samanta, T. (1997a). Asymptotic expansion of posterior distributions in nonregular cases. *Annals of the Institute of Statistical Mathematics*, *49*(1), 181–197.

Ghosal, S., Samanta, T. (1997b). Expansion of Bayes risk for entropy loss and refernce prior in nonregular cases. *Statistics and Decisions*, *15*, 129–140.

Ghosh, J. K., Mukerjee, R. (1992). Non-informative priors (with discussion). In J. M. Bernardo, J. O. Berger, A. P. David, A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 195–210). London: Oxford University Press.

Hartigan, J. A. (1998). The maximum likelihood prior. *The Annals of Statistics*, *26*(6), 2083–2103.

Hellinger, E. (1909). Neue begründung der theorie quadratischen formen von unendlichen vielen veränderlichen. *Journal Für Reine und Angewandte Mathematik*, *136*, 210–271.

Hewitt, E., Stromberg, K. (1969). *Real and abstract analysis. A modern treatment of the theory of functions of a real variable* (second printing corrected). Berlin: Springer.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). London: Oxford University Press.

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

Laplace, P. S. (1812). *Theorie analytique des probabilities*. Paris: Courceir.

Lindley, D. V. (1956). On the measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, *27*, 986–1005.