Reference priors via α-divergence for a certain non-regular model in the presence of a nuisance parameter

Shintaro Hashimoto[1]

Department of Mathematics, Hiroshima University, Hiroshima, Japan

Abstract

This paper presents reference priors for non-regular model whose support depends on an unknown parameter. A multi-parameter family which includes both regular and non-regular structures is considered. The resulting priors are obtained by asymptotically maximizing the expected α-divergence between the prior and the corresponding posterior distribution. Some examples of reference priors for typical multi-parameter non-regular distributions such as the location-scale family of distributions and the truncated Weibull distribution are also given.

Keywords: α-divergence, Bayesian inference, Location-scale model, Non-regular, Reference priors

1. Introduction

In Bayesian inference, when there is no prior information, we often begin inference by using objective priors such as non-informative or default priors. Then we are often faced with a problem of the selection of objective prior in a given context. One of the most widely used objective priors is the Jeffreys prior proposed by Jeffreys (1961). The Jeffreys prior is proportional to the positive square root of the Fisher information function in one-dimensional case. On the other hand, the reference prior was proposed by Bernardo (1979) and was extended by Berger and Bernardo (1989) in the presence of nuisance parameters. The reference prior is defined by maximizing the Kullback-Leibler (KL) divergence between the prior and the posterior under some regularity conditions. This prior maximizes the expected posterior information to the prior, i.e., the prior is the 'least informative' prior in some aspects. In the context of the reference priors, Ghosh et al. (2011) derives the priors which asymptotically maximize a more general divergence measure (called the α-divergence) between the prior and the corresponding posterior under some regularity conditions. We note that the α-divergence smoothly connected the KL divergence ($\alpha \to 0$), the reverse KL divergence ($\alpha \to 1$), the squared Bhattacharyya-Hellinger divergence ($\alpha = 1/2$) and the chi-square divergence ($\alpha = -1$) (see e.g. Amari (1982) and Cressie and Read (1984)). Recently, Liu et al. (2014) extends the result of Ghosh et al. (2011) to a multi-parameter model with or without nuisance parameters for regular parametric family. Beside the prior selection problem, statistical inference and prediction based on the α-divergence have been also developed in recent years (see

[1]Address correspondence to Shintaro Hashimoto, Department of Mathematics, Hiroshima University, Higashi-Hiroshima, Hiroshima, 739–8521 Japan
E-mail: s-hashimoto@hiroshima-u.ac.jp

e.g. (Corcuera and Giummolè, 1999; Ghosh et al., 2008; Ghosh and Mergel, 2009; Maruyama et al., 2019).

However, Ghosh et al. (2011) and Liu et al. (2014) deal with the regular parametric models and these results are not applied for non-regular cases whose supports of the density depend on unknown parameter. For example, the uniform and shifted exponential distributions have the parameter dependent supports and such non-regular distributions are also important in applications. For examples, the auction and search models in structural econometric models have a jump in the density and the jump is very informative about the parameters (e.g. Chernozhukov and Hong (2004)). In such non-regular cases, for example, the asymptotic normality of the posterior distribution does not hold (Ghosal and Samanta (1995)). Ghosal and Samanta (1997) shows the prior which maximizes the KL divergence for a non-regular one-parameter family of distributions. In non-regular case, the prior which is different from the Jeffreys prior is derived. For a multi-parameter case, Ghosal (1997) gives the reference prior based on the KL divergence for multi-parameter non-regular model from the perspective of information theory. As related results, Ghosal (1999) also derives probability matching priors and Hashimoto (2019) derives moment matching priors for the same non-regular model as that of Ghosal (1997).

In this paper, we consider a certain multi-parameter family of distributions $\mathcal{P} = \{f(\cdot;\theta,\varphi) \mid \theta \in \Theta, \varphi \in \Phi\}$ which includes both regular and non-regular structures. In other words, this model is regular with respect to φ for fixed θ, and is non-regular with respect to θ for fixed φ. In this paper, we call θ and φ, respectively. For example, the shifted exponential distribution with the density function $f(x;\theta,\varphi) = \varphi^{-1}e^{-(x-\theta)/\varphi}$ $(x > \theta,\ \varphi > 0)$ belongs to this family of distributions. For such model, the reference priors based on the expected α-divergence for $\alpha < 1$ are derived by using the higher order asymptotic expansion for the posterior distribution. The results in this paper are a kind of generalizations of the result in Ghosal (1997) which is used the expected KL divergence. The resulting reference priors are different forms from Ghosal (1997) except for $\alpha = 0$. However, in the location-scale family (see Example 1), if θ is the parameter of interest, the reference prior for (θ,φ) is the same as Ghosal (1997)'s one which is the right invariant Haar measure when $-1 < \alpha < 0$ and $0 < \alpha < 1$. In other words, in this case, our prior has loss-robustness for $-1 < \alpha < 1$. On the other hand, if φ is the parameter of interest, the resulting prior for (θ,φ) is not same as that of Ghosal (1997), that is, our prior does not have loss-robustness for $-1 < \alpha < 1$ in such case. This is very interesting phenomenon. Furthermore, for $\alpha = -1$, that is, the chi-square divergence, we also derive a new reference prior when φ is the parameter of interest.

This paper organizes as follows: in Section 2, we introduce the higher order asymptotic representations for posterior density in non-regular case and the definition of the maximum α-divergence prior in the presence of a nuisance parameter. In Section 3, we derive the marginal reference priors for the non-regular parameter θ in the presence of the regular nuisance parameter φ for $-1 < \alpha < 1$. It is also shown that there is generally no reference prior for $\alpha \leq -1$. In a similar way, we derive the marginal reference prior for the regular parameter φ in the presence of the non-regular nuisance parameter θ for $-1 < \alpha < 1$. It is also shown that there is generally no reference prior for $\alpha < -1$. Further, we give the explicit form of the marginal reference prior for φ in the case of $\alpha = -1$. Overall reference priors for (θ,φ) are calculated by using Berger and Bernardo (1989)'s algorithm (for details, see Subsection 2.2). As examples, we show the reference priors in the case of the (non-regular) location-scale family of distributions and the truncated Weibull distribution with known shape parameter.

2. Assumptions and formulations

2.1. Setting

In this paper, we consider the same family of non-regular distributions as that of Ghosal (1997) and Ghosal (1999). Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) observations from a density $f(x; \theta, \varphi)$ $(\theta \in \Theta \subset \mathbb{R}, \varphi \in \Phi \subset \mathbb{R})$ with respect to the Lebesgue measure, where Θ and Φ are parameter spaces of θ and φ, respectively. For simplicity, we consider a scalar θ and φ, respectively. When φ is vector-value, we may also consider in the same manner. We assume that for all $\theta \in \Theta$ and $\varphi \in \Phi$, $f(x; \theta, \varphi)$ is strictly positive and forth times differentiable in θ and φ on a closed interval $S(\theta) := [a_1(\theta), a_2(\theta)]$ depending only on unknown parameter θ and is zero outside $S(\theta)$. Namely, θ is a non-regular parameter and φ is a regular parameter. For example, the two-parameter shifted exponential distribution which has a truncation parameter θ and a scale parameter φ belongs to this family. Note that for a given θ, the family of distributions $\mathcal{P} = \{f(\cdot; \theta, \varphi) \mid \theta \in \Theta, \varphi \in \Phi\}$ is regular with respect to φ. It is permitted that one of the endpoints is free from θ and may be plus or minus infinity. We assume that the endpoints $a_1(\theta)$ and $a_2(\theta)$ of the support are continuously differentiable functions of θ. Let $\pi(\theta, \varphi)$ be the joint prior density of (θ, φ), and $\pi(\theta)$ and $\pi(\varphi)$ be marginal prior densities of θ and φ, respectively. We assume that the prior density $\pi(\theta, \varphi)$ is three times continuously differentiable in a neighborhood of (θ, φ). Further, we assume the conditions which ensure the validity of the second order asymptotic expansion of the posterior distribution such as Ghosal (1999).

In order to have a limit of the posterior distributions, Ghosh et al. (1994) show that it is necessary that the set $S(\theta)$ is either increasing or decreasing in θ, that is, $S(\theta)$ satisfies either $S(\theta) \subseteq S(\theta + \varepsilon)$ for $\varepsilon > 0$ or $S(\theta) \subseteq S(\theta + \varepsilon)$ for $\varepsilon < 0$, respectively. For this reason, we may assume $S(\theta)$ is decreasing without loss of generality. Indeed, the case where $S(\theta)$ increases with θ may be reduced to the case where $S(\theta)$ decreases by the reparametrization $\theta \mapsto (-\theta)$. As an example of family with non-monotone support, one directed family of distribution is discussed by Akahira and Takeuchi (1987). For such family of distributions, the reference priors are discussed by Berger et al. (2009) and Wang and Sun (2012). However, we do not deal with such distributions in this paper. When $S(\theta)$ is decreasing, the set $\{a_1(\theta) \leq X_i \leq a_2(\theta), i = 1, 2, \ldots, n\}$ can be expressed as $\{\hat{\theta}_n(X_1, \ldots, X_n) \geq \theta\}$ where $\hat{\theta}_n := \min\{a_1^{-1}(X_{(1)}), a_2^{-1}(X_{(n)})\}$, $X_{(1)} := \min_{1 \leq i \leq n} X_i$ and $X_{(n)} := \max_{1 \leq i \leq n} X_i$. If a_1 does not depend on θ, then we interpret the above $\hat{\theta}_n$ as $a_2^{-1}(X_{(n)})$ while it is interpreted as $a_1^{-1}(X_{(1)})$ if a_2 does not depend on θ. We note that $\hat{\theta}_n - \theta = O_p(n^{-1})$ $(n \to \infty)$. Hereafter, we omit "$n \to \infty$" for simplicity. Let $\hat{\varphi}_n$ be a solution of the the modified likelihood equation

$$\sum_{i=1}^{n} \frac{\partial}{\partial \varphi} \log f(X_i; \hat{\theta}_n, \hat{\varphi}_n) = 0.$$

Smith (1985) showed the consistency for the special case when θ is a location parameter, but the argument can easily be generalized. Hence, we may assume that $(\hat{\theta}_n, \hat{\varphi}_n)$ is consistent estimator of

(θ, φ). We put

$$\sigma := \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i;\hat{\theta}_n,\hat{\varphi}_n), \quad b^2 := -\sum_{i=1}^{n}\frac{\partial^2}{\partial\varphi^2}\log f(X_i;\hat{\theta}_n,\hat{\varphi}_n)$$

and we note that $\sigma \to c(\theta,\varphi)$ and $b^2 \to \lambda^2(\theta,\varphi)$ almost surely, where

$$c(\theta,\varphi) := \mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta,\varphi)\right], \quad \lambda^2(\theta,\varphi) := \mathrm{E}\left[-\frac{\partial^2}{\partial\varphi^2}\log f(X_1;\theta,\varphi)\right].$$

When $S(\theta)$ is monotone decreasing, we can show that $c(\theta,\varphi) > 0$. Hereafter, we may assume that $c(\theta,\varphi) > 0$. Let $u := n\sigma(\theta - \hat{\theta}_n)$ and $v := \sqrt{n}b(\varphi - \hat{\varphi}_n)$ be normalized random variables of θ and φ, respectively. From Appendix in Ghosal (1999) the joint posterior density of (u, v) given $X = (X_1, \ldots, X_n)$ has the asymptotic expansion up to the order $O(n^{-3/2})$

$$\pi(u,v|X) = \frac{1}{\sqrt{2\pi}}e^{u-(v^2/2)}\left\{1 + \frac{1}{\sqrt{n}}D_1(u,v) + \frac{1}{n}D_2(u,v) + O\left(n^{-3/2}\right)\right\} \tag{1}$$

for $u < 0$, where

$$\begin{aligned}
D_1(u,v) =& \frac{\hat{\pi}_{01}}{\hat{\pi}_{00}b}v + \frac{2a_{11}}{\sigma b}uv + \frac{a_{03}}{b^3}v^3,\\
D_2(u,v) =& \frac{\hat{\pi}_{10}}{\hat{\pi}_{00}\sigma}(u+1) + \frac{\hat{\pi}_{02}}{2\hat{\pi}_{00}b^2}(v^2-1) + \frac{a_{20}}{\sigma^2}(u^2-2)\\
&+ \frac{2(\hat{\pi}_{01}/\hat{\pi}_{00})a_{11} + 3a_{12}}{\sigma b^2}(uv^2+1) + \frac{\hat{\pi}_{01}a_{03}}{\hat{\pi}_{00}b^4}(v^4-3)\\
&+ \frac{2a_{11}^2}{\sigma^2 b^2}(u^2v^2-2) + \frac{2a_{11}a_{03}}{\sigma b^4}(uv^4+3) + \frac{a_{03}^2}{b^6}(v^6-15)
\end{aligned}$$

with

$$\hat{\pi}_{rs} = \frac{\partial^{r+s}}{\partial\theta^r\partial\varphi^s}\pi(\hat{\theta}_n,\hat{\varphi}_n), \quad a_{rs} = \frac{1}{(r+s)!\,n}\sum_{i=1}^{n}\frac{\partial^{r+s}}{\partial\theta^r\partial\varphi^s}\log f(X_i;\hat{\theta}_n,\hat{\varphi}_n)$$

for $r, s = 0, 1, 2, \ldots$, and note that $a_{rs} \to A_{rs}(\theta,\varphi)$ almost surely, where

$$A_{rs}(\theta,\varphi) = \frac{1}{(r+s)!\,n}\mathrm{E}\left[\frac{\partial^{r+s}}{\partial\theta^r\partial\varphi^s}\log f(X_1;\theta,\varphi)\right]$$

for $r, s = 0, 1, 2, \ldots$. Note that $\sigma = a_{10}$ and $b^2 = -2a_{02}$. From (1) we can find that the random variables u and v are the first order asymptotic independent and their first order asymptotic marginal posterior distributions are the exponential and the normal distributions, respectively (see also Ghosal and Samanta (1995)). From (1) we can obtain the second order asymptotic marginal posterior densities $\pi(u|X)$ and $\pi(v|X)$. The second order asymptotic marginal posterior density of

u is given by

$$
\begin{aligned}
\pi(u|X) =& \int \pi(u,v|X)\mathrm{d}v \\
=& e^u\Bigg[1+\frac{1}{n}\Bigg\{\left(\frac{\hat{\pi}_{10}}{\hat{\pi}_{00}\sigma}+\frac{2(\hat{\pi}_{01}/\hat{\pi}_{00})a_{11}+3a_{12}}{\sigma b^2}+\frac{6a_{11}a_{03}}{\sigma b^4}\right)(u+1) \\
&+\left(\frac{a_{20}}{\sigma^2}+\frac{2a_{11}^2}{\sigma^2 b^2}\right)(u^2-2)\Bigg\}+O\left(n^{-2}\right)\Bigg]
\end{aligned}
\tag{2}
$$

for $u<0$, while that of v is given by

$$
\begin{aligned}
\pi(v|X) =& \int \pi(u,v|X)\mathrm{d}u \\
=& \frac{1}{\sqrt{2\pi}}e^{-v^2/2}\Bigg[1+\frac{1}{\sqrt{n}}\Bigg\{\left(\frac{\hat{\pi}_{01}}{\hat{\pi}_{00}b}-\frac{2a_{11}}{\sigma b}\right)v+\frac{a_{03}}{b^3}v^3\Bigg\} \\
&+\frac{1}{n}\Bigg\{\left(\frac{\hat{\pi}_{02}}{2\hat{\pi}_{00}b^2}-\frac{2(\hat{\pi}_{01}/\hat{\pi}_{00})a_{11}+3a_{12}}{\sigma b^2}+\frac{4a_{11}^2}{\sigma^2 b^2}\right)(v^2-1) \\
&+\left(\frac{\hat{\pi}_{01}a_{03}}{\hat{\pi}_{00}b^4}-\frac{2a_{11}a_{03}}{\sigma b^4}\right)(v^4-3)+\frac{a_{03}^2}{b^6}(v^6-15)\Bigg\}+O\left(n^{-3/2}\right)\Bigg].
\end{aligned}
\tag{3}
$$

2.2. Reference prior as a maximizer of the expected divergence

As we mentioned in Section 1, the reference prior was firstly proposed by Bernardo (1979). The reference prior is defined by maximizing the expected KL divergence between the prior and the corresponding posterior under some regularity conditions. Clarke and Barron (1994) showed a rigorous proof of the derivation of reference priors from the perspective of information theory (see also Ghosal and Samanta (1997)). Now, we define the reference prior in the sense of Berger and Bernardo (1989) under a more general divergence measure.

For simplicity, we consider the two-parameter model $f(x;\theta)$, where $\theta=(\theta_1,\theta_2)\in\Theta_1\times\Theta_2\subset\mathbb{R}^2$. We assume that θ_1 is a parameter of interest and θ_2 is a nuisance parameter. Let $\pi(\theta)=\pi(\theta_1,\theta_2)=\pi(\theta_2|\theta_1)\pi(\theta_1)$ be a joint prior density of $\theta=(\theta_1,\theta_2)$. Then the reference prior with a general divergence is defined by the following (see also Liu et al. (2014)).

Definition 1. *When the conditional prior $\pi(\theta_2|\theta_1)$ is chosen as a reasonable prior on a compact subset of Θ_2, the marginal reference prior for θ_1 with a general divergence is define by*

$$
\pi(\theta_1)=\arg\max_{\pi(\theta_1)}\int D\left(\pi(\theta_1),\pi(\theta_1|x)\right)m(x)\mathrm{d}x, \tag{4}
$$

where $D(\pi(\cdot),\pi(\cdot|x))$ is a divergence measure between the prior $\pi(\theta_1)$ and the corresponding posterior $\pi(\theta_1|x)$, and $m(x)$ is the marginal density of $X=(X_1,\ldots,X_n)$ with respect to the joint prior $\pi(\theta)=\pi(\theta_1,\theta_2)$.

The reference prior for (θ,φ) is given by the following algorithm by Berger and Bernardo (1989).

1. Choose the reference prior for θ_2 given θ_1 as $\pi^*(\theta_2|\theta_1)$.

2. Choose a sequence $\Theta_{2,1} \subset \Theta_{2,2} \subset \cdots$ of compact subsets of Θ_2 such that $\cup_{l=1}^{\infty}\Theta_{2,l} = \Theta_2$.
3. Set $K_l(\theta_1) = \left(\int_{\Theta_{2,l}} \pi^*(\theta_2|\theta_1)\mathrm{d}\theta_2\right)^{-1}$ and $p_l(\theta_2|\theta_1) = K_l(\theta_1)\pi^*(\theta_2|\theta_1)\mathbb{1}\{\theta_2 \in \Theta_{2,l}\}$.
4. The marginal reference prior for θ_1 at stage l is calculated by using (4):

$$\pi_l^*(\theta_1) = \arg\max_{\pi(\theta_1)} \int D\left(\pi(\theta_1), \pi_l(\theta_1|x)\right) m_l(x)\mathrm{d}x,$$

where $m_l(x) = \int_{\Theta_1\times\Theta_{2,l}} f(x|\theta_1,\theta_2)\pi(\theta_1)p_l(\theta_2|\theta_1)\mathrm{d}\theta_1\mathrm{d}\theta_2$ and

$$\pi_l(\theta_1|x) = \int_{\Theta_{2,l}} \pi_l(\theta_1,\theta_2|x)\mathrm{d}\theta_2 = \int_{\Theta_{2,l}} \frac{f(x|\theta_1,\theta_2)\pi(\theta_1)p_l(\theta_2|\theta_1)}{m_l(x)}\mathrm{d}\theta_2.$$

5. Let $\theta_{0,1}$ be a fixed point in Θ_1. The reference prior for (θ_1, θ_2) with a nuisance θ_2 is obtained by

$$\pi^*(\theta_1,\theta_2) = \lim_{l\to\infty}\left(\frac{K_l(\theta_1)\pi_l^*(\theta_1)}{K_l(\theta_{0,1})\pi_l^*(\theta_{0,1})}\right)\pi^*(\theta_2|\theta_1),$$

provided the limit exists.

We note that the bigger the divergence between prior and posterior is, the less information in a prior is. In this paper, we consider the α-divergence which includes the KL divergence (Amari, 1982; Cressie and Read, 1984). The α-divergence between the prior and the posterior is defined by

$$D^{\alpha}(\pi(\cdot),\pi(\cdot|X)) := \frac{1-\int \pi^{\alpha}(\theta_1)\pi^{1-\alpha}(\theta_1|x)\mathrm{d}\theta_1}{\alpha(1-\alpha)} \tag{5}$$

for $\alpha \in \mathbb{R}\setminus\{0,1\}$ and $\theta_1 \in \Theta_1 \subset \mathbb{R}$. Then the expected α-divergence between the prior and the posterior is defined by the following functional

$$R^{\alpha}(\pi) := \int D^{\alpha}(\pi(\cdot),\pi(\cdot|X))m(x)\mathrm{d}x = \frac{1-\int[\int \pi^{\alpha}(\theta_1)\pi^{1-\alpha}(\theta_1|x)\mathrm{d}\theta_1]m(x)\mathrm{d}x}{\alpha(1-\alpha)}, \tag{6}$$

where $m(x)$ is the marginal density of $X = (X_1,\ldots,X_n)$ with respect to the joint prior $\pi(\theta) = \pi(\theta_1,\theta_2)$ (see Ghosh et al. (2011), Liu et al. (2014)). Note that the α-divergence smoothly connects the KL divergence ($\alpha \to 0$), the reverse KL divergence ($\alpha \to 1$), the squared Bhattacharyya-Hellinger divergence ($\alpha = 1/2$), and the chi-square divergence ($\alpha = -1$). Let $L_n(\theta_1)$ be the likelihood function of the parameter θ_1. From the relation $\pi(\theta_1|x)m(x) = L_n(\theta_1)\pi(\theta_1)$ we can express as

$$\begin{aligned} R^{\alpha}(\pi) &= \frac{1-\iint \pi^{\alpha+1}(\theta_1)\pi^{-\alpha}(\theta_1|x)L_n(\theta_1)\mathrm{d}x\mathrm{d}\theta_1}{\alpha(1-\alpha)} \\ &= \frac{1-\int \pi^{\alpha+1}(\theta_1)\mathrm{E}_{\theta_1}[\pi^{-\alpha}(\theta_1|X)]\mathrm{d}\theta_1}{\alpha(1-\alpha)}, \end{aligned} \tag{7}$$

where E_{θ_1} denotes the conditional expectation of X given θ_1. In Section 3 and 3.3, we derive the reference priors which maximize (7) for the multi-parameter non-regular family of distributions $\mathcal{P}$. Hereafter, we assume $\alpha < 1$ for some technical reasons described later.

3. Reference priors via α-divergence

We consider the two-parameter model $f(x;\theta,\varphi)$ which includes both regular parameter φ and non-regular parameter θ defined in the previous section.

3.1. Reference priors for the non-regular parameter in the presence of the regular nuisance parameter

In this subsection, we assume that θ is the parameter of interest and φ is the nuisance parameter. We derive the marginal reference prior for θ under the α-divergence in the sense of Berger and Bernardo (1989). First of all, we note that the joint prior density $\pi(\theta,\varphi)$ is rewritten by $\pi(\theta,\varphi)=\pi(\varphi|\theta)\pi(\theta)$. In the first step, we assign the conditional Jeffreys prior $\pi(\varphi|\theta)\propto\sqrt{\lambda^2(\theta,\varphi)}$ on a compact subset of Φ to the parameter φ given θ (Ghosh et al. (2011)). Then we may maximize the following functional with respect to π

$$R^{\alpha}(\pi)=\frac{1-\int\pi^{\alpha+1}(\theta)\mathrm{E}_{\theta}[\pi^{-\alpha}(\theta|X)]\mathrm{d}\theta}{\alpha(1-\alpha)}, \tag{8}$$

where E_θ denotes the conditional expectation of $X=(X_1,\dots,X_n)$ given θ. In order to derive the prior which maximizes the expected α-divergence, we may calculate the expectation $\mathrm{E}_\theta[\pi^{-\alpha}(\theta|X)]$. Since the exact calculation of this expectation is not easy, we consider the asymptotic approximation of $\mathrm{E}_\theta[\pi^{-\alpha}(\theta|X)]$ by using the first order asymptotic representation of the marginal posterior density of θ in (1). Here, we use the computation method called the shrinkage argument which is a Bayesian approach for frequentist computations (see Ghosh (1994), Datta and Mukerjee (2004)). Then we have the following lemma.

Lemma 1. *For $\alpha<1$, the second order asymptotic approximation of $\mathrm{E}_\theta[\pi^{-\alpha}(\theta|X)]$ is given by*

$$\begin{aligned}\mathrm{E}_{\theta}\left[\pi^{-\alpha}(\theta|X)\right]=&n^{-\alpha}\int\frac{c(\theta,\varphi)^{-\alpha}}{1-\alpha}\Bigg[1+\frac{1}{n}\Bigg\{\frac{\alpha^2}{1-\alpha}\frac{(\partial/\partial\theta)\pi(\theta,\varphi)}{\pi(\theta,\varphi)}\frac{1}{c(\theta,\varphi)}\\&+\frac{2\alpha^2}{1-\alpha}\frac{A_{11}(\theta,\varphi)}{c(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{(\partial/\partial\varphi)\pi(\theta,\varphi)}{\pi(\theta,\varphi)}+S(\theta,\varphi)\Bigg\}+O(n^{-2})\Bigg]\pi(\varphi|\theta)d\varphi,\end{aligned} \tag{9}$$

where $S(\theta,\varphi)$ is some continuous function which does not depend on the prior density $\pi(\theta)$, and $\pi(\varphi|\theta)\propto\sqrt{\lambda^2(\theta,\varphi)}$ is the conditional Jeffreys prior for φ given θ on a compact subset of Φ.

The proof of Lemma 1 is given in Section 4. We note that the equation (9) does not hold for $\alpha\geq 1$ as is evident from the right-hand-side expression in (9). For $\alpha<1$, we have the following theorem.

Theorem 1. *The marginal reference priors for θ in the presence of the nuisance parameter φ are given by*

$$\pi(\theta)\propto\begin{cases}\left(\int c^{-\alpha}(\theta,\varphi)\pi(\varphi|\theta)\mathrm{d}\varphi\right)^{-1/\alpha} & (-1<\alpha<0,\ 0<\alpha<1),\\ \exp\left(\int\log c(\theta,\varphi)\pi(\varphi|\theta)\mathrm{d}\varphi\right) & (\alpha=0),\end{cases} \tag{10}$$

while the marginal reference priors for θ generally do not exist for $\alpha\leq -1$.

The proof of Theorem 1 is given in Section 4.

Remark 1. When $\alpha = 0$ in Theorem 1, we may interpret $\alpha \to 0$ as $\alpha = 0$ because the expectation (9) is not defined for $\alpha = 0$. In this case, the α-divergence corresponds to the KL divergence. Note that the marginal reference prior (10) for $\alpha = 0$ is the same as the marginal reference prior based on the expected KL divergence in Ghosal (1997). As a related result, Ghosal (1999) derive the probability matching prior with θ as the parameter of interest is given by

$$\pi(\theta) \propto \left(\int c^{-1}(\theta, \varphi)\pi(\varphi|\theta)\mathrm{d}\varphi \right)^{-1}. \tag{11}$$

We note that the prior (11) is slightly different from the maximum KL divergence prior (10) for $\alpha = 0$.

3.2. Reference priors for the regular parameter in the presence of the non-regular nuisance parameter

Next, we consider the case where the regular parameter φ is assumed to be more interest than the non-regular parameter θ, that is, we assume that φ is the parameter of interest and θ is the nuisance parameter. In a similar way to Subsection 3.1, we derive marginal reference priors for φ based on the maximization of the expected α-divergence. The joint prior density can be written by $\pi(\theta, \varphi) = \pi(\theta|\varphi)\pi(\varphi)$. Since the conditional prior density $\pi(\theta|\varphi)$ is a function of θ for fixed φ, we use the conditional prior density $\pi(\theta|\varphi) \propto c(\theta, \varphi)$ on a compact subset of Θ (see Ghosal and Samanta (1997)). Note that this prior is known as a non-informative prior for non-regular case. If $S(\theta)$ is monotone increasing, then we may put $\pi(\theta|\varphi) \propto |c(\theta, \varphi)|$ because $c(\theta, \varphi) < 0$ in such case. Having fixed this conditional prior, marginal reference priors for φ is given by maximizing the following expected α-divergence

$$R^{\alpha}(\pi) = \frac{1 - \int \pi^{\alpha+1}(\varphi)\mathrm{E}_{\varphi}[\pi^{-\alpha}(\varphi|X)]\mathrm{d}\varphi}{\alpha(1-\alpha)}, \tag{12}$$

where E_{φ} denotes the conditional expectation of $X = (X_1, \ldots, X_n)$ given φ. In order to derive the prior which maximizes (12), we give the second order asymptotic approximation of $\mathrm{E}_{\varphi}[\pi^{-\alpha}(\varphi|X)]$ by using the shrinkage argument in a similar way to Subsection 3.1.

Lemma 2. *For $\alpha < 1$, the second order asymptotic approximation of $\mathrm{E}_\varphi[\pi^{-\alpha}(\varphi|X)]$ is given by*

$$
\begin{aligned}
\mathrm{E}_\varphi\left[\pi^{-\alpha}(\varphi|X)\right] =& n^{-\alpha/2}\Bigg[\int B_\alpha(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \\
&+\frac{1}{n}\Bigg\{-\frac{\alpha^2}{1-\alpha}\left(\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}H_1^{(\alpha)}(\varphi)+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\right) \\
&-\frac{3\alpha^2(2-\alpha)}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&+\frac{\alpha(\alpha+1)}{(1-\alpha)}\left(\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2H_1^{(\alpha)}(\varphi)+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\right) \\
&+\frac{3\alpha(\alpha+1)}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&-\frac{\alpha}{1-\alpha}\Bigg(-2\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-2\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_{2,\varphi}^{(\alpha)}(\varphi)+2\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2H_2^{(\alpha)}(\varphi) \\
&\qquad\qquad+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\Bigg) \\
&-\frac{3\alpha}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right)+S(\varphi)\Bigg\}+O(n^{-2})\Bigg],
\end{aligned}
\tag{13}
$$

where $S(\varphi)$ is a continuous function, not involving $\pi(\varphi)$, $B_\alpha(\theta,\varphi) := ((2\pi)^{\alpha/2}\lambda^{-\alpha}(\theta,\varphi))/\sqrt{1-\alpha}$,

$$
\begin{aligned}
H_1^{(\alpha)}(\varphi) &:= \int \frac{B_\alpha(\theta,\varphi)}{2\lambda^2(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta, \\
H_2^{(\alpha)}(\varphi) &:= \int \frac{B_\alpha(\theta,\varphi)}{\lambda^2(\theta,\varphi)}\frac{\pi_\varphi(\theta|\varphi)}{\pi(\theta|\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta, \\
H_3^{(\alpha)}(\varphi) &:= \int \frac{2A_{11}(\theta,\varphi)B_\alpha(\theta,\varphi)}{c(\theta,\varphi)\lambda^2(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta, \\
H_4^{(\alpha)}(\varphi) &:= \int \frac{A_{03}(\theta,\varphi)B_\alpha(\theta,\varphi)}{\lambda^4(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta,
\end{aligned}
$$

and $H_{2,\varphi}^{(\alpha)}(\varphi) = (\partial/\partial\varphi)H_2^{(\alpha)}(\varphi)$, and $\pi(\theta|\varphi) \propto c(\theta,\varphi)$ is the conditional prior for θ given φ on a compact subset of Θ.

The proof of Lemma 2 is given in Section 4. As we mentioned in Subsection 3.1, we note that the equation (2) does not hold for $\alpha \geq 1$ as is evident from the right-hand-side expression in (2). For $\alpha < 1$ and $\alpha \neq -1$, we have the following theorem.

Theorem 2. *The marginal reference priors for φ in the presence of the nuisance parameter θ are* *given by*

$$\pi(\varphi) \propto \begin{cases} \left(\int \lambda^{-\alpha}(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \right)^{-1/\alpha} & (-1<\alpha<0,\ 0<\alpha<1), \\ \exp\left(\int \log \lambda(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \right) & (\alpha=0), \end{cases} \tag{14}$$

while the marginal reference priors for φ *generally does not exist for* $\alpha < -1$.

The proof of Theorem 2 is omitted since it is similar to that of Theorem 1 in Subsection 3.1.

Remark 2. Note that the marginal reference prior (14) for $\alpha = 0$ is the same as the marginal reference prior under the expected KL divergence in Ghosal (1997). As a related result, Ghosal (1999) derive the probability matching prior with φ as the parameter of interest, given by

$$\pi(\varphi) \propto \left(\int \lambda^{-1}(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \right)^{-1}. \tag{15}$$

Next, we consider the case $\alpha = -1$. In this case, the α-divergence corresponds to the chi-square divergence. We note that the reference priors which maximizes the expected chi-square divergence are also discussed by Clarke and Sun (1997), Ghosh et al. (2011) and Liu et al. (2014) for regular parametric family.

Putting $\alpha = -1$ in (13), we have

$$\begin{aligned} &\mathrm{E}_{\varphi}\left[\pi(\varphi|X)\right] \\ =&n^{1/2}\left[\int B_{-1}(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \right. \\ &+\frac{1}{n}\left\{ \frac{\pi_{\varphi}(\varphi)}{\pi(\varphi)}\left(-H_{2,\varphi}^{(-1)}(\varphi) - \frac{3}{2}H_4^{(-1)}(\varphi) \right) \right. \\ &\left.\left. + \left(\frac{\pi_{\varphi}(\varphi)}{\pi(\varphi)}\right)^2 H_2^{(-1)}(\varphi) + \frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}\left(-\frac{1}{2}H_1^{(-1)}(\varphi) - H_2^{(-1)}(\varphi) \right) + S(\varphi) \right\} + O(n^{-2}) \right]. \end{aligned}$$

Further, we put

$$\begin{aligned} M_1(\varphi) :=& -H_{2,\varphi}^{(-1)}(\varphi) - \frac{3}{2}H_4^{(-1)}(\varphi), \\ M_2(\varphi) :=& H_2^{(-1)}(\varphi), \\ M_3(\varphi) :=& -\frac{1}{2}H_1^{(-1)}(\varphi) - H_2^{(-1)}(\varphi). \end{aligned} \tag{16}$$

By using (16), we can rewrite

$$\mathrm{E}_\varphi\left[\pi(\varphi|X)\right] = n^{1/2}\Bigg[\int B_{-1}(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta + \frac{1}{n}\left\{\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}M_1(\varphi) + \left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2 M_2(\varphi) + \frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}M_3(\varphi) + S(\varphi)\right\} + O(n^{-2})\Bigg]. \tag{17}$$

Then we have the following theorem concerning with the marginal reference prior for φ under the expected chi-square divergence.

Theorem 3. *For $\alpha = -1$, the marginal reference prior for φ in the presence of nuisance parameter θ is given by*

$$\pi(\varphi) \propto \exp\left(\int \frac{(\partial/\partial\varphi)M_3(\varphi) - M_1(\varphi)}{2(M_2(\varphi) + M_3(\varphi))}\mathrm{d}\varphi\right), \tag{18}$$

where the integral in (18) *is the indefinite integral, and $M_1(\varphi)$, $M_2(\varphi)$ and $M_3(\varphi)$ are functions of φ defined by* (16).

The proof of Theorem 3 is given in Section 4. From Theorem 3 we can find that it appears a new prior distribution which is different from (14). From theorems 1 and 3, the results of the case $\alpha = -1$ change depending on whether we are interested in θ or φ. This interesting phenomenon has beed also pointed out by Ghosh et al. (2011) and Liu et al. (2014) in a regular parametric family.

3.3. Some examples

In this subsection, we show some examples concerned with the reference priors given in theorems 1, 2 and 3. To compute reference priors, we use Berger and Bernardo (1989)'s algorithm which is mentioned in Subsection 2.2. We also discussed the differences of reference priors between α-divergence and KL divergence.

Example 1 (Location-scale family)**.** Let f_0 be a strictly positive density on $[0,\infty)$ and consider the family $f(x;\theta,\varphi) = \varphi^{-1}f_0\{(x-\theta)/\varphi\}$ $(x > \theta)$, where θ is a location parameter and φ is a scale parameter. We note that the support of the density depends on θ. Further, we assume that the right-hand limit of $f_0(x)$ at $x = 0$ exists. For example, the shifted exponential distribution with the density function $f(x;\theta,\varphi) = \varphi^{-1}e^{-(x-\theta)/\varphi}$ $(x > \theta,\ \varphi > 0)$ belongs to this (non-regular) location-scale family. In this case, we have

$$c(\theta,\varphi) = f_0(0+)/\varphi, \quad \lambda^2(\theta,\varphi) = c_1/\varphi^2,$$

where c_1 is the constant number defined by $c_1 = \int\{1 + xf_0'(t)/f_0(t)\}^2 f_0(t)\mathrm{d}t$. If θ is the parameter of interest and φ is the nuisance parameter, we adapt the conditional reference prior $\pi(\varphi|\theta) = \sqrt{\lambda^2(\theta,\varphi)} \propto \varphi^{-1}$ on the sequence $\Phi_1 \subset \Phi_2 \subset \cdots$ of compact set of Φ such that $\cup_{l=1}^\infty \Phi_l = \Phi$. Then the marginal reference prior (10) for $-1 < \alpha < 0$ and $0 < \alpha < 1$ is the improper uniform distribution $\pi(\theta) \propto 1$, and by using Berger and Bernardo (1989)'s algorithm in Section 2, the resulting reference prior is given by $\pi(\theta,\varphi) \propto \varphi^{-1}$ for $-1<\alpha<0$ and $0<\alpha<1$. For $\alpha=0$, by using the second equation of (10), the reference prior for (θ,φ) is also given by $\pi(\theta,\varphi) \propto \varphi^{-1}$ which is the same as the result of Ghosal (1997). In this case, for $-1<\alpha<1$, the reference prior is given by $\pi(\theta,\varphi)$ which is the right invariant Haar measure, and it is known that the right Haar measure has a very attractive properties (see Chang and Eaves (1990)).

In a similar way to the above, if φ is the parameter of interest and θ is the nuisance parameter, we may consider the conditional reference prior $\pi(\theta|\varphi)=c(\theta,\varphi)$ on the sequence $\Theta_1 \subset \Theta_2 \subset \cdots$ of compact set of Θ such that $\cup_{l=1}^{\infty}\Theta_l=\Theta$. Then the marginal reference prior (14) is given by $\pi(\varphi) \propto \varphi^{(1-\alpha)/\alpha}$ for $-1<\alpha<0$ and $0<\alpha<1$, and the resulting reference prior is $\pi(\theta,\varphi) \propto \varphi^{(1-2\alpha)/\alpha}$ for $-1<\alpha<0$ and $0<\alpha<1$. For $\alpha=0$, by using the second equation of (14), the reference prior for (θ,φ) is given by $\pi(\theta,\varphi) \propto \varphi^{-1}$ which is the same as the result of Ghosal (1997), but is different from the case for $-1<\alpha<0$ and $0<\alpha<1$. For $\alpha=-1$, we have the marginal reference prior (18) is given by $\pi(\varphi) \propto \varphi^{-2}$, and resulting reference prior for (θ,φ) is given by $\pi(\theta,\varphi) \propto \varphi^{-3}$. This prior is neither the right invariant Haar measure nor left invariant Haar measure.

In both cases, resulting reference priors are improper. So, we now check the posterior propriety. Since, it is not easy to show the posterior propriety for general location-scale family, in particular, we consider the shifted exponential distribution which belongs to the location-scale family. Let $X_1,\dots,X_n$ be a sequence of random variables from the density $f(x;\theta,\varphi)=\varphi^{-1}e^{-(x-\theta)/\varphi}$ ($x>\theta$, $\varphi>0$). and $x=(x_1,\dots,x_n)$ be the observation from this model. To show the posterior propriety, it is enough to show that the normalized constant in the posterior density is finite under the priors $\pi(\theta,\varphi) \propto \varphi^{-1}$, $\pi(\theta,\varphi) \propto \varphi^{(1-2\alpha)/\alpha}$ for $-1<\alpha<0$ and $0<\alpha<1$, and $\pi(\theta,\varphi) \propto \varphi^{-3}$. The normalized constant in the posterior density is defined by

$$m(x)=\int_0^{\infty}\int_{-\infty}^{x_{(1)}}\prod_{i=1}\frac{1}{\varphi}e^{-(1/\varphi)\sum_{i=1}^n(x_i-\theta)}\pi(\theta,\varphi)\mathrm{d}\theta\mathrm{d}\varphi,$$

where $x_{(1)}:=\max_{1\le i\le 1}x_i$. Under the prior $\pi(\theta,\varphi) \propto \varphi^{-1}$, we have $m(x)=\Gamma(n-1)/\{n(\sum_{i=1}^n(x_i-x_{(1)}))^{n-1}\}<\infty$ for $n\ge 2$, where $\Gamma(k)$ is the gamma function defined by $\Gamma(k)=\int_0^{\infty}x^{k-1}e^{-x}\mathrm{d}x$. Further, under the prior $\pi(\theta,\varphi) \propto \varphi^{(1-2\alpha)/\alpha}$ for $-1<\alpha<0$ and $0<\alpha<1$, we have $m(x)=\Gamma(n-(1/\alpha))/\{n(\sum_{i=1}^n(x_i-x_{(1)}))^{n-(1/\alpha))}\}<\infty$ for $n\ge\max(1/\alpha,2)$ $(-1<\alpha<0,\ 0<\alpha<1)$. Finally, under the prior $\pi(\theta,\varphi) \propto \varphi^{-3}$, we have $m(x)=\Gamma(n+1)/\{n(\sum_{i=1}^n(x_i-x_{(1)}))^{n+1}\}<\infty$ for $n\ge 2$. Hence, resulting improper reference priors for (θ,φ) lead to proper posteriors.

Example 2 (Truncated Weibull distribution)**.** Consider the truncated Weibull distribution with the truncation parameter θ, the scale parameter $\varphi>0$ and the shape parameter $k>0$ with the density function $f(x;\theta,\varphi)=k\varphi^k x^{k-1}\exp\{-\varphi^k(x^k-\theta^k)\}$ $(x>\theta)$. We assume that the shape parameter $k>0$ is known and $\theta>0$ in this example, and consider reference priors for (θ,φ). In this case, we have

$$c(\theta,\varphi)=k\varphi^k\theta^{k-1},\quad \lambda^2(\theta,\varphi)=k^2/\varphi^2.$$

We now derive reference priors for (θ,φ) by using the same operation as that of Example 1. If θ is the parameter of interest and φ is the nuisance parameter, then the reference prior (10) for $-1<\alpha<0$ and $0<\alpha<1$ is given by $\pi(\theta,\varphi) \propto \theta^{k-1}\varphi^{-1}$. For $\alpha=0$, the reference prior (10) is also given by $\pi(\theta,\varphi)\propto\theta^{k-1}\varphi^{-1}$ which is the same prior as that of Ghosal (1997) under KL divergence.

On the other hand, If φ is the parameter of interest and θ is the nuisance parameter, then the reference prior (14) is given by $\pi(\varphi)\propto\theta^{k-1}\varphi^{-(k/\alpha)+k-1}$ for $-1<\alpha<0$ and $0<\alpha<1$. For $\alpha=0$, the reference prior (14) is given by $\pi(\theta,\varphi)\propto\theta^{k-1}\varphi^{-1}$ which is the same prior as that of Ghosal (1997) under KL divergence. For $\alpha=-1$, the calculation of the prior (18) may be messy. Hence, we omit it here.

Check for the posterior propriety is also omitted here, but we may be able to obtain the sufficient condition for the finiteness of the normalized constant in the posterior density.

Remark 3. Ghosal (1997) and Ghosal (1999) discussed the important special case where the factorizations

$$c(\theta,\varphi)=c_1(\theta)c_2(\varphi),\quad \sqrt{\lambda^2(\theta,\varphi)}=\lambda_1(\theta)\lambda_2(\varphi) \tag{19}$$

hold. If such factorizations hold, they argued that under KL divergence, the Berger and Bernardo (1989)'s algorithm yields the reference prior $c_1(\theta)\lambda_2(\varphi)$ which does not depend on the order of importance and the choice of compact sets. However, it does not always hold under α-divergence as we seen in Examples 1 and 2 even though the factorizations (19) hold. Further, Ghosal (1997) argued that the invariant property of reference priors under KL divergence for non-regular when the factorizations (19) hold, while reference priors under α-divergence do not generally hold invariant property except for $\alpha=0$.

4. Proofs

We give proofs of lemmas and theorems in Section 3.

Proof of Lemma 1. We show (9) by using the shrinkage argument which consists following three steps (see Datta and Mukerjee (2004)). We put $u=n\sigma(\theta-\hat{\theta}_n)$ in (2). Then we have

$$\begin{aligned}\pi(\theta|X)=&|n\sigma|e^{n\sigma(\theta-\hat{\theta}_n)}\\ &\cdot\left[1+\frac{1}{n}\left\{\hat{R}_1(n\sigma(\theta-\hat{\theta}_n)+1)+\hat{R}_2((n\sigma)^2(\theta-\hat{\theta}_n)^2-2)+O\left(n^{-2}\right)\right\}\right]\end{aligned} \tag{20}$$

for $\theta<\hat{\theta}_n$, where

$$\hat{R}_1=\frac{\hat{\pi}_{10}}{\hat{\pi}_{00}\sigma}+\frac{2(\hat{\pi}_{01}/\hat{\pi}_{00})a_{11}+3a_{12}}{\sigma b^2}+\frac{6a_{11}a_{03}}{\sigma b^4},\quad \hat{R}_2=\frac{a_{20}}{\sigma^2}+\frac{2a_{11}^2}{\sigma^2b^2}.$$

Step 1. We consider a proper prior density $\bar{p}(\theta)$, such that the support of $\bar{p}(\theta)$ is compact in the parameter space and $\bar{p}(\theta)$ vanishes outside of the support while remaining positive in the interior. Next, we compute the expectation $\mathrm{E}^{\bar{\pi}}[\pi^{-\alpha}(u|X)|X]$, where $\mathrm{E}^{\bar{\pi}}[\cdot|X]$ denotes the expectation with respect to the posterior density $\bar{\pi}(\cdot|X)$ under the prior $\bar{\pi}(\theta,\varphi)=\bar{p}(\theta)\pi(\varphi|\theta)$. First, we compute the following product

$$\pi^{-\alpha}(\theta|X)\bar{\pi}(\theta|X)=(n|\sigma|)^{1-\alpha}e^{(1-\alpha)n\sigma(\theta-\hat{\theta}_n)}$$

$$\cdot\left[1+\frac{1}{n}\left\{-\alpha\hat{R}_1(n\sigma(\theta-\hat{\theta}_n)+1)+\hat{\bar{R}}_1(n\sigma(\theta-\hat{\theta}_n)+1)+k(\theta)\right\}+O(n^{-2})\right]$$

for $\theta < \hat{\theta}_n$, where $k(\theta)$ is a continuous parametric function, not involving $\pi(\theta)$ and

$$\hat{\bar{R}}_1=\frac{\hat{\bar{\pi}}_{10}}{\hat{\bar{\pi}}_{00}\sigma}+\frac{2(\hat{\bar{\pi}}_{01}/\hat{\bar{\pi}}_{00})a_{11}+3a_{12}}{\sigma b^2}+\frac{6a_{11}a_{03}}{\sigma b^4},\quad \hat{\bar{\pi}}_{rs}=\frac{\partial^{r+s}}{\partial\theta^r\partial\varphi^s}\bar{\pi}(\hat{\theta}_n,\hat{\varphi}_n)\quad (r,s=0,1,\dots).$$

Then the expectation of $\pi^{-\alpha}(\theta|X)$ with respect to $\bar{\pi}(\cdot|X)$ is given by

$$\begin{aligned}\mathrm{E}^{\bar{\pi}}[\pi^{-\alpha}(\theta|X)|X] &= \int_{-\infty}^{\hat{\theta}_n}\pi^{-\alpha}(\theta|X)\bar{\pi}(\theta|X)\mathrm{d}\theta\\ &= \frac{(n|\sigma|)^{-\alpha}}{1-\alpha}\left[1+\frac{1}{n(1-\alpha)}\left\{\alpha^2\hat{R}_1-\alpha\hat{\bar{R}}_1+\alpha(2-\alpha)\hat{R}_2+C\right\}+O(n^{-2})\right]\\ &= G(X)\quad\text{(say)}\end{aligned}$$

where C is a constant number. Note that in order to compute above integration, we put $n\sigma(\theta-\hat{\theta}_n)=-t$ and regard the integration as the expectation of the exponential distribution with mean parameter $(1-\alpha)$.

Step 2. For θ in the interior point of the support of $\bar{p}(\theta)$, we calculate the following expectation

$$\lambda(\theta):=\int G(x)f_n(x;\theta)\mathrm{d}x=\int G(x)\left\{\int\prod_{i=1}^n f(x_i;\theta,\varphi)\pi(\varphi|\theta)\mathrm{d}\varphi\right\}\mathrm{d}x=\int\lambda_0(\theta,\varphi)\pi(\varphi|\theta)\mathrm{d}x$$

where $\lambda_0(\theta,\varphi)=\int G(x)\prod_{i=1}^n f(x_i;\theta,\varphi)\mathrm{d}x$. Since

$$b^2=\lambda^2(\theta,\varphi)+o(1),\quad \sigma=c(\theta,\varphi)+o(1),\quad a_{rs}=A_{rs}(\theta,\varphi)+o(1),\quad (r,s=0,1,\dots),$$

by using Taylor's expansion, we have

$$\lambda_0(\theta,\varphi)=\frac{(n|\sigma|)^{-\alpha}}{1-\alpha}\left[1+\frac{1}{n(1-\alpha)}\left\{\alpha^2R_1-\alpha\bar{R}_1+S_0(\theta,\varphi)\right\}+O(n^{-2})\right]$$

where $S_0(\theta,\varphi)$ is a continuous parametric function, not involving $\pi(\theta)$ and

$$\begin{aligned}R_1&=\frac{\pi_{10}}{\pi_{00}c}+\frac{2(\pi_{10}/\pi_{00})A_{11}+3A_{12}}{c\lambda^2}+\frac{6A_{11}A_{03}}{c\lambda^4},\\ \bar{R}_1&=\frac{\bar{\pi}_{10}}{\bar{\pi}_{00}c}+\frac{2(\bar{\pi}_{10}/\bar{\pi}_{00})A_{11}+3A_{12}}{c\lambda^2}+\frac{6A_{11}A_{03}}{c\lambda^4}\end{aligned}$$

with

$$\begin{aligned}c:=&c(\theta,\varphi)=\mathrm{E}\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta,\varphi)\right],\quad \lambda^2:=\lambda^2(\theta,\varphi)=\mathrm{E}\left[-\frac{\partial^2}{\partial\varphi^2}\log f(X_1;\theta,\varphi)\right],\\ \pi_{rs}:=&\pi_{rs}(\theta,\varphi)=\frac{\partial^{r+s}}{\partial\theta^r\partial\varphi^s}\pi(\theta,\varphi)\quad(r,s=0,1,2,\dots),\end{aligned}$$

$$\bar{\pi}_{rs} := \bar{\pi}_{rs}(\theta,\varphi) = \frac{\partial^{r+s}}{\partial\theta^r\partial\varphi^s}\bar{\pi}(\theta,\varphi) \quad (r,s=0,1,2,\dots),$$
$$A_{rs} := A_{rs}(\theta,\varphi) = \frac{1}{(r+s)!}\mathrm{E}\left[\frac{\partial^{rs}}{\partial\theta^r\partial\varphi^s}\log f(X_1;\theta,\varphi)\right] \quad (r,s=0,1,2,\dots).$$

Step 3. The final step of this argument involves integrating $\lambda(\theta)$ with respect to $\bar{p}(\theta)$ and then making $\bar{p}(\theta)$ degenerate at θ. We have

$$\int \lambda(\theta)\bar{p}(\theta)\mathrm{d}\theta$$
$$= \iint \frac{\{nc(\theta,\varphi)\}^{-\alpha}}{1-\alpha}\left[1+\frac{1}{n(1-\alpha)}\left\{\alpha^2 R_1 - \alpha\bar{R}_1 + S(\theta,\varphi)\right\} + O(n^{-2})\right]\pi(\varphi|\theta)\bar{p}(\theta)\mathrm{d}\varphi\mathrm{d}\theta.$$

We note that the following identities hold

$$\frac{\bar{\pi}_{10}}{\bar{\pi}_{00}} = \frac{(\partial/\partial\theta)(\bar{p}(\theta)\pi(\varphi|\theta))}{\bar{p}(\theta)\pi(\varphi|\theta)} = \frac{(\partial/\partial\theta)\bar{p}(\theta)}{\bar{p}(\theta)} + \frac{(\partial/\partial\theta)\pi(\varphi|\theta)}{\pi(\varphi|\theta)},$$
$$\frac{\bar{\pi}_{01}}{\bar{\pi}_{00}} = \frac{(\partial/\partial\varphi)(\bar{p}(\theta)\pi(\varphi|\theta))}{\bar{p}(\theta)\pi(\varphi|\theta)} = \frac{(\partial/\partial\varphi)\pi(\varphi|\theta)}{\pi(\varphi|\theta)}.$$

So, we have

$$\iint \frac{1}{c^{1+\alpha}(\theta,\varphi)}\frac{\bar{\pi}_{10}}{\bar{\pi}_{00}}\pi(\varphi|\theta)\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta$$
$$= \iint \frac{1}{c^{1+\alpha}(\theta,\varphi)}\left(\frac{(\partial/\partial\theta)\bar{p}(\theta)}{\bar{p}(\theta)} + \frac{(\partial/\partial\theta)\pi(\varphi|\theta)}{\pi(\varphi|\theta)}\right)\pi(\varphi|\theta)\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta$$
$$= \iint \frac{1}{c^{1+\alpha}(\theta,\varphi)}\pi(\varphi|\theta)\mathrm{d}\varphi\frac{\partial}{\partial\theta}\bar{p}(\theta)\mathrm{d}\theta + \iint \frac{(\partial/\partial\theta)\pi(\varphi|\theta)}{c^{1+\alpha}(\theta,\varphi)}\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta,$$
$$\iint \frac{A_{11}(\theta,\varphi)}{c^{1+\alpha}(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{\bar{\pi}_{01}}{\bar{\pi}_{00}}\pi(\varphi|\theta)\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta$$
$$= \iint \frac{A_{11}(\theta,\varphi)}{c^{1+\alpha}(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{(\partial/\partial\varphi)\pi(\varphi|\theta)}{\pi(\varphi|\theta)}\pi(\varphi|\theta)\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta.$$

Hence, we have

$$\iint \frac{1}{c^{1+\alpha}(\theta,\varphi)}\frac{\bar{\pi}_{10}}{\bar{\pi}_{00}}\pi(\varphi|\theta)\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta \to -\frac{\partial}{\partial\theta}\left(\int \frac{1}{c^{1+\alpha}(\theta,\varphi)}\pi(\varphi|\theta)\mathrm{d}\varphi\right) + \int \frac{(\partial/\partial\theta)\pi(\varphi|\theta)}{c^{1+\alpha}(\theta,\varphi)}\mathrm{d}\varphi,$$
$$\iint \frac{A_{11}(\theta,\varphi)}{c^{1+\alpha}(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{\bar{\pi}_{01}}{\bar{\pi}_{00}}\pi(\varphi|\theta)\mathrm{d}\varphi\bar{p}(\theta)\mathrm{d}\theta \to \int \frac{A_{11}(\theta,\varphi)}{c^{1+\alpha}(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{(\partial/\partial\varphi)\pi(\varphi|\theta)}{\pi(\varphi|\theta)}\pi(\varphi|\theta)\mathrm{d}\varphi.$$

Bu using the shrinkage argument, the second order asymptotic approximation of $\mathrm{E}_\theta[\pi^{-\alpha}(\theta|X)]$ is given by

$$\mathrm{E}_\theta\left[\pi^{-\alpha}(\theta|X)\right] = n^{-\alpha}\int \frac{c(\theta,\varphi)^{-\alpha}}{1-\alpha}\left[1+\frac{1}{n}\left\{\frac{\alpha^2}{1-\alpha}\frac{(\partial/\partial\theta)\pi(\theta,\varphi)}{\pi(\theta,\varphi)}\frac{1}{c(\theta,\varphi)}\right.\right.$$

$$+\frac{2\alpha^2}{1-\alpha}\frac{A_{11}(\theta,\varphi)}{c(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{(\partial/\partial\varphi)\pi(\theta,\varphi)}{\pi(\theta,\varphi)}+S(\theta,\varphi)\Bigg\}+O(n^{-2})\Bigg]\pi(\varphi|\theta)d\varphi,$$

where $S(\theta,\varphi)$ is a continuous function not involving π. This completes the proof. □

Proof of Theorem 1. From Lemma 1, the first order asymptotic approximation of (8) is given by

$$R^\alpha(\pi)\approx\frac{1}{\alpha(1-\alpha)}\left\{1-\frac{n^{-\alpha}}{1-\alpha}\int\left(\frac{\xi(\theta)}{\pi(\theta)}\right)^{-\alpha}\pi(\theta)\mathrm{d}\theta\right\}, \tag{21}$$

where $\xi(\theta)=\{\int c^{-\alpha}(\theta,\varphi)\pi(\varphi|\theta)\mathrm{d}\varphi\}^{-1/\alpha}$. We consider the following five cases separately: (i) $0<\alpha<1$, (ii) $-1<\alpha<0$, (iii) $\alpha=0$, (iv) $\alpha<-1$ and (v) $\alpha=-1$.

(i) First, we consider the case $0<\alpha<1$. Since $\alpha(1-\alpha)>0$, it suffuses to minimize the following

$$\int\left(\frac{\xi(\theta)}{\pi(\theta)}\right)^{-\alpha}\pi(\theta)\mathrm{d}\theta=\int g\left(\frac{\xi(\theta)}{\pi(\theta)}\right)\pi(\theta)\mathrm{d}\theta,$$

where $g(t)=t^{-\alpha}$ $(t>0)$. Noting that $g(t)$ is a convex function of t for $0<\alpha<1$, by Jensen's inequality, we have

$$\int g\left(\frac{\xi(\theta)}{\pi(\theta)}\right)\pi(\theta)\mathrm{d}\theta\geq g\left(\int\frac{\xi(\theta)}{\pi(\theta)}\pi(\theta)\mathrm{d}\theta\right)=\left\{\int\xi(\theta)\mathrm{d}\theta\right\}^{-\alpha}.$$

The equality holds if and only if $\pi(\theta)\propto\xi(\theta)$ which is the marginal reference prior with maximum α-divergence for $0<\alpha<1$.

(ii) Next, we consider the case $-1<\alpha<0$. Since $\alpha(1-\alpha)<0$, it suffuses to maximize the following

$$\int\left(\frac{\xi(\theta)}{\pi(\theta)}\right)^{-\alpha}\pi(\theta)\mathrm{d}\theta=\int g\left(\frac{\xi(\theta)}{\pi(\theta)}\right)\pi(\theta)\mathrm{d}\theta.$$

Noting that $g(t)=t^{-\alpha}$ $(t>0)$ is a concave function of t for $-1<\alpha<0$, by Jensen's inequality, we have

$$\int g\left(\frac{\xi(\theta)}{\pi(\theta)}\right)\pi(\theta)\mathrm{d}\theta\leq g\left(\int\frac{\xi(\theta)}{\pi(\theta)}\pi(\theta)\mathrm{d}\theta\right)=\left\{\int\xi(\theta)\mathrm{d}\theta\right\}^{-\alpha}.$$

The equality holds if and only if $\pi(\theta)\propto\xi(\theta)$ which is the marginal reference prior with maximum α-divergence for $-1<\alpha<0$.

(iii) For $\alpha=0$, we need to interpret (8) as its limiting value (when it exists). In this case, the α-divergence corresponds to the KL divergence. By using the L'Hôpital's rule, it suffices to maximize the following

$$R^0(\pi)=\log n-1+\iint\log\left(\frac{c(\theta,\varphi)}{\pi(\theta)}\right)\pi(\varphi|\theta)\pi(\theta)\mathrm{d}\varphi\mathrm{d}\theta.$$

Putting $\log \zeta(\theta) = \int (\log c(\theta, \varphi))\pi(\varphi|\theta)\mathrm{d}\varphi$, we may maximize the following

$$\int \log\left(\frac{\zeta(\theta)}{\pi(\theta)}\right)\pi(\theta)\mathrm{d}\theta. \tag{22}$$

From the property of the KL divergence, (22) is maximized by $\pi(\theta) \propto \zeta(\theta)$. Note that as $\alpha \to 0$ in (10), we can get $\zeta(\theta)$.

(iv) Consider the case $\alpha < -1$. Putting $\alpha = -\beta\,(\beta > 1)$, we can rewrite (21)

$$R^\alpha(\pi) \approx \frac{1}{\beta(\beta+1)}\left\{\frac{n^\beta}{\beta+1}\int\left(\frac{\xi(\theta)}{\pi(\theta)}\right)^\beta \pi(\theta)\mathrm{d}\theta - 1\right\}$$

for $\beta > 1$. Hence it suffices to maximize the following

$$\int\left(\frac{\xi(\theta)}{\pi(\theta)}\right)^\beta \pi(\theta)\mathrm{d}\theta.$$

By using the Lyapounov inequality (e.g. DasGupta (2008)), we have for $\beta > 1$

$$\int\left(\frac{\xi(\theta)}{\pi(\theta)}\right)^\beta \pi(\theta)\mathrm{d}\theta \geq \left\{\int \xi(\theta)\mathrm{d}\theta\right\}^\beta = \left\{\int \xi(\theta)\mathrm{d}\theta\right\}^{-\alpha}. \tag{23}$$

The equality holds if and only if $\pi(\theta) \propto \xi(\theta)$. However, this prior is the minimizer rather than the maximizer of (22) from (23). We can show that there is no maximizing prior in this case. It suffuses to show that

$$\sup_\pi \int \xi^\beta(\theta)\pi^{1-\beta}(\theta)\mathrm{d}\theta = +\infty. \tag{24}$$

In order to prove (24), we consider a compact set $A \subset \mathbb{R}$. Then there exists $c > 0$ such that $\xi(\theta) \geq c$ for all $\theta \in A$. For any $M > 0$, we can make a prior $\pi(\theta) = \{M/(\mu(A_M)c)\}^{1/(1-\beta)}\,(\theta \in A_M)$, where $A_M \subseteq A$ satisfying $\int_{A_M}\pi(\theta)\mathrm{d}\theta < 1$ and $\mu(\cdot)$ is the Lebesgue measure on $\mathbb{R}$. If θ is not in A_M, we can assign some suitable value to $\pi(\theta)$ to make $\pi(\theta)$ a probability density. Then, we have

$$\int \xi^\beta(\theta)\pi^{1-\beta}(\theta)\mu(\mathrm{d}\theta) \geq \int_A \xi^\beta(\theta)\pi^{1-\beta}(\theta)\mu(\mathrm{d}\theta) \geq \int_{A_M} c\frac{M}{\mu(A_M)c}\mu(\mathrm{d}\theta) = M.$$

Hence, for any $M > 0$, we can find $\pi(\theta)$ such that $\int \xi^\beta(\theta)\pi^{1-\beta}(\theta)\mu(\mathrm{d}\theta) \geq M$. Therefore, it holds $\sup_\pi \int \xi^\beta(\theta)\pi^{1-\beta}(\theta)\mathrm{d}\theta = +\infty$.

(v) Finally we consider the case $\alpha = -1$. For $\alpha = -1$, the first order term in (8) is a constant because of $\pi^{\alpha+1}(\varphi) = 1$. We need to consider the second order term. From Lemma 1 we have

$$\begin{aligned}\mathrm{E}_\theta\left[\pi(\theta|X)\right] =& n\int \frac{c(\theta,\varphi)}{2}\Bigg[1 + \frac{1}{n}\Bigg\{\frac{\alpha^2}{2}\frac{(\partial/\partial\theta)\pi(\theta,\varphi)}{\pi(\theta,\varphi)}\frac{1}{c(\theta,\varphi)} \\ &+ \alpha^2\frac{A_{11}(\theta,\varphi)}{c(\theta,\varphi)\lambda^2(\theta,\varphi)}\frac{(\partial/\partial\varphi)\pi(\theta,\varphi)}{\pi(\theta,\varphi)} + S(\theta,\varphi)\Bigg\} + O(n^{-2})\Bigg]\pi(\varphi|\theta)d\varphi.\end{aligned}$$

For $\alpha = -1$, the expected α-divergence is expressed by

$$R^{-1}(\pi) \approx \frac{\int \mathrm{E}_\theta[\pi(\theta|X)]\mathrm{d}\theta - 1}{2}.$$

Hence, we consider the maximization problem:

$$\begin{aligned}
&\max_{\pi(\theta)} \int \left\{ \int \left(\frac{1}{2} \frac{(\partial/\partial\theta)\pi(\theta,\varphi)}{\pi(\theta,\varphi)} + \frac{A_{11}(\theta,\varphi)}{\lambda^2(\theta,\varphi)} \frac{(\partial/\partial\varphi)\pi(\theta,\varphi)}{\pi(\theta,\varphi)} \right) \pi(\varphi|\theta)\mathrm{d}\varphi \right\} \mathrm{d}\theta \\
&= \max_{\pi(\theta)} \int \frac{(\partial/\partial\theta)\pi(\theta)}{\pi(\theta)} \mathrm{d}\theta.
\end{aligned}$$

However, we can not find such $\pi(\theta)$ in general. For example, putting $\pi(\theta) = \sin\theta (0 \le \theta \le \pi/2)$, we have $\int \pi'(\theta)/\pi(\theta)\mathrm{d}\theta = \int_0^{\pi/2} (\tan\theta)^{-1} d\theta = \infty$.

□

Proof of Lemma 2. From (3) the asymptotic marginal posterior density of v is given by

$$\begin{aligned}
\pi(v|X) =& \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left[1 + \frac{1}{\sqrt{n}} (\hat{S}_1 v + \hat{S}_2 v^3) \right. \\
&\left. + \frac{1}{n} \left\{ \hat{S}_3(v^2 - 1) + \hat{S}_4(v^4 - 3) + \hat{S}_5(v^6 - 15) \right\} + O\left(n^{-3/2}\right) \right],
\end{aligned} \tag{25}$$

where

$$\begin{aligned}
&\hat{S}_1 = \frac{\hat{\pi}_{01}}{\hat{\pi}_{00} b} - \frac{2a_{11}}{\sigma b}, \quad \hat{S}_2 = \frac{a_{03}}{b^3}, \quad \hat{S}_3 = \frac{\hat{\pi}_{02}}{2\hat{\pi}_{00} b^2} - \frac{2(\hat{\pi}_{01}/\hat{\pi}_{00})a_{11} + 3a_{12}}{\sigma b^2} + \frac{4a_{11}^2}{\sigma^2 b^2}, \\
&\hat{S}_4 = \frac{\hat{\pi}_{01} a_{03}}{\hat{\pi}_{00} b^4} - \frac{2a_{11}a_{03}}{\sigma b^4}, \quad \hat{S}_5 = \frac{a_{03}^2}{b^6}.
\end{aligned}$$

Putting $v = \sqrt{n} b(\varphi - \hat{\varphi})$ in (25), we have

$$\begin{aligned}
&\pi(\varphi|X) \\
=& \frac{\sqrt{n} b}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} n b^2 (\varphi - \hat{\varphi})^2 \right\} \\
&\cdot \left[1 + \frac{1}{\sqrt{n}} \left\{ \hat{S}_1 \sqrt{n} b(\varphi - \hat{\varphi}) + \hat{S}_2 n\sqrt{n} b^3 (\varphi - \hat{\varphi})^3 \right\} \right. \\
&\left. + \frac{1}{n} \left\{ \hat{S}_3(nb^2(\varphi - \hat{\varphi})^2 - 1) + \hat{S}_4(n^2 b^4 (\varphi - \hat{\varphi})^4 - 3) + \hat{S}_5(n^3 b^6 (\varphi - \hat{\varphi})^6 - 15) \right\} + O\left(n^{-3/2}\right) \right].
\end{aligned}$$

We now put

$$\begin{aligned}
A_n &= \hat{S}_1 \sqrt{n} b(\varphi - \hat{\varphi}) + \hat{S}_2 n\sqrt{n} b^3 (\varphi - \hat{\varphi})^3, \\
B_n &= \hat{S}_3(nb^2(\varphi - \hat{\varphi})^2 - 1) + \hat{S}_4(n^2 b^4(\varphi - \hat{\varphi})^4 - 3) + \hat{S}_5(n^3 b^6 (\varphi - \hat{\varphi})^6 - 15).
\end{aligned}$$

Step one. We consider a proper prior density $\bar{p}(\varphi)$ such that the support of $\bar{p}(\varphi)$ is compact in the parameter space and $\bar{p}(\varphi)$ vanishes outside of the support while remaining positive in the interior. Let $\bar{\pi}(\cdot|X)$ be the posterior density under the prior $\bar{\pi}(\theta, \varphi) = \bar{p}(\varphi)\pi(\theta|\varphi)$. First we compute the product

$$
\begin{aligned}
&\pi^{-\alpha}(\varphi|X)\bar{\pi}(\varphi|X) \\
=&(\sqrt{n}b)^{1-\alpha}(2\pi)^{-(1-\alpha)/2}\exp\left\{-\frac{1}{2}(1-\alpha)nb^2(\varphi-\hat{\varphi})^2\right\} \\
&\cdot\left\{1+\frac{A_n}{\sqrt{n}}+\frac{B_n}{n}+O\left(n^{-3/2}\right)\right\}^{-\alpha}\left\{1+\frac{\bar{A}_n}{\sqrt{n}}+\frac{\bar{B}_n}{n}+O\left(n^{-3/2}\right)\right\} \\
=&(\sqrt{n}b)^{1-\alpha}(2\pi)^{-(1-\alpha)/2}\exp\left\{-\frac{1}{2}(1-\alpha)nb^2(\varphi-\hat{\varphi})^2\right\} \\
&\cdot\left\{1-\alpha\left(\frac{A_n}{\sqrt{n}}+\frac{B_n}{n}\right)+\frac{\alpha(\alpha+1)}{2}\left(\frac{A_n}{\sqrt{n}}+\frac{B_n}{n}\right)^2+O\left(n^{-5/2}\right)\right\}\left\{1+\frac{\bar{A}_n}{\sqrt{n}}+\frac{\bar{B}_n}{n}+O\left(n^{-3/2}\right)\right\} \\
=&(\sqrt{n}b)^{1-\alpha}(2\pi)^{-(1-\alpha)/2}\exp\left\{-\frac{1}{2}(1-\alpha)nb^2(\varphi-\hat{\varphi})^2\right\} \\
&\cdot\left\{1+\frac{1}{\sqrt{n}}\left(-\alpha A_n+\bar{A}_n\right)+\frac{1}{n}\left(-\alpha B_n+\bar{B}_n-\alpha A_n\bar{A}_n+\frac{\alpha(\alpha+1)}{2}A_n^2\right)+O\left(n^{-3/2}\right)\right\},
\end{aligned}
$$

where $\bar{A}_n$ and $\bar{B}_n$ are the same forms as A_n and B_n under the prior $\bar{\pi}$, respectively. The expectation of $\pi^{-\alpha}(\varphi|X)$ under the density $\bar{\pi}(\cdot|X)$ is

$$
\begin{aligned}
&\mathrm{E}^{\bar{\pi}}\left[\pi^{-\alpha}(\varphi|X)\right] \\
=&\int_{-\infty}^{\infty}(\sqrt{n}b)^{1-\alpha}\left\{1+\frac{1}{\sqrt{n}}\left(-\alpha A_n+\bar{A}_n\right)+\frac{1}{n}\left(-\alpha B_n+\bar{B}_n-\alpha A_n\bar{A}_n+\frac{\alpha(\alpha+1)}{2}A_n^2\right)+O\left(n^{-3/2}\right)\right\} \\
&\cdot(2\pi)^{-(1-\alpha)/2}\exp\left\{-\frac{1}{2}(1-\alpha)nb^2(\varphi-\hat{\varphi})^2\right\}\mathrm{d}\varphi \\
=&\int_{-\infty}^{\infty}\left(\frac{2\pi}{nb^2}\right)^{\alpha/2}\left[1+\frac{1}{\sqrt{n}}\left\{-\alpha(\hat{S}_1t+\hat{S}_2t^3)+(\hat{\bar{S}}_1t+\hat{\bar{S}}_2t^3)\right\}\right. \\
&+\frac{1}{n}\left\{-\alpha(\hat{S}_3(t^2-1)+\hat{S}_4(t^4-3)+\hat{S}_5(t^6-15))+(\hat{\bar{S}}_3(t^2-1)+\hat{\bar{S}}_4(t^4-3)+\hat{\bar{S}}_5(t^6-15))\right. \\
&\left.\left.+\frac{\alpha(\alpha+1)}{2}(\hat{S}_1t+\hat{S}_2t^3)^2-\alpha(\hat{S}_1t+\hat{S}_2t^3)(\hat{\bar{S}}_1t+\hat{\bar{S}}_2t^3)\right\}+O\left(n^{-3/2}\right)\right] \\
&\cdot\sqrt{\frac{1-\alpha}{2\pi}}\exp\left\{-\frac{(1-\alpha)t^2}{2}\right\}\mathrm{d}t \\
=&\left(\frac{2\pi}{nb^2}\right)^{\alpha/2}\frac{1}{\sqrt{1-\alpha}}\left[1+\frac{1}{n}\left\{-\alpha\left(\frac{\alpha}{1-\alpha}\hat{S}_3+\frac{3\alpha(2-\alpha)}{(1-\alpha)^2}\hat{S}_4\right)+\left(\frac{\alpha}{1-\alpha}\hat{\bar{S}}_3+\frac{3\alpha(2-\alpha)}{(1-\alpha)^2}\hat{\bar{S}}_4\right)\right.\right. \\
&+\frac{\alpha(\alpha+1)}{2}\left(\frac{1}{1-\alpha}\hat{S}_1^2+\frac{6}{(1-\alpha)^2}\hat{S}_1\hat{S}_2\right)
\end{aligned}
$$

$$- \alpha\left(\frac{1}{1-\alpha}\hat{S}_1\hat{\bar{S}}_1 + \frac{3}{(1-\alpha)^2}\hat{S}_1\hat{\bar{S}}_2 + \frac{3}{(1-\alpha)^2}\hat{S}_2\hat{\bar{S}}_1\right) + \hat{K}_1\Bigg\} + O\left(n^{-2}\right)\Bigg] = Q(X) \quad \text{(say)},$$

where $\hat{\bar{S}}_i$ $(i = 1, \ldots, 5)$ are the same forms as $\hat{S}_i$ $(i = 1, \ldots, 5)$ under the prior $\bar{\pi}$, and $\hat{K}_1 := K(\hat{\theta}, \hat{\varphi})$ is a random variable not involving the prior π.

Step 2. For φ in the interior point of the support of $\bar{p}(\varphi)$, we calculate the following expectation

$$\lambda(\varphi) := \int Q(x) f_n(x;\varphi)\mathrm{d}x = \int Q(x)\left\{\int \prod_{i=1}^n f(x_i;\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta\right\}\mathrm{d}x = \int \lambda_0(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}x$$

where $\lambda_0(\theta,\varphi) = \int Q(x)\prod_{i=1}^n f(x_i;\theta,\varphi)\mathrm{d}x$. Note that

$$b^2 = \lambda^2(\theta,\varphi) + o(1), \quad \sigma = c(\theta,\varphi) + o(1), \quad \hat{K}_1 = K_1 + o(1),$$

where $K_1 = K_1(\theta,\varphi)$ and

$$a_{rs} = A_{rs}(\theta,\varphi) + o(1) \quad (r, s = 0, 1, \ldots)$$

By using Taylor's expansion, we have

$$\begin{aligned}
\lambda_0(\theta,\varphi) =& \frac{1}{\sqrt{1-\alpha}}\left(\frac{2\pi}{n\lambda^2(\theta,\varphi)}\right)^{\alpha/2}\Bigg[1 + \frac{1}{n}\Bigg\{ - \alpha\left(\frac{\alpha}{1-\alpha}S_3 + \frac{3\alpha(2-\alpha)}{(1-\alpha)^2}S_4\right) \\
&+ \left(\frac{\alpha}{1-\alpha}\bar{S}_3 + \frac{3\alpha(2-\alpha)}{(1-\alpha)^2}\bar{S}_4\right) + \frac{\alpha(\alpha+1)}{2}\left(\frac{1}{1-\alpha}S_1^2 + \frac{6}{(1-\alpha)^2}S_1S_2\right) \\
&- \alpha\left(\frac{1}{1-\alpha}S_1\bar{S}_1 + \frac{3}{(1-\alpha)^2}S_1\bar{S}_2 + \frac{3}{(1-\alpha)^2}\bar{S}_1S_2\right) + K_2(\theta,\varphi)\Bigg\} + O\left(n^{-2}\right)\Bigg],
\end{aligned}$$

where $K_2(\theta,\varphi)$ is a continuous function not involving π and

$$\begin{aligned}
S_1 &= \frac{\pi_{01}}{\pi_{00}\lambda} - \frac{2A_{11}}{c\lambda}, \quad S_2 = \frac{A_{03}}{\lambda^3}, \\
S_3 &= \frac{\pi_{02}}{2\pi_{00}\lambda^2} - \frac{2(\pi_{01}/\pi_{00})A_{11} + 3A_{12}}{c\lambda^2} + \frac{4A_{11}^2}{c^2\lambda^2}, \quad S_4 = \frac{\pi_{01}A_{03}}{\pi_{00}\lambda^4} - \frac{2A_{11}A_{03}}{c\lambda^4},
\end{aligned}$$

and $\bar{S}_i$ $(i = 1, \ldots, 4)$ are the same forms as S_i $(i = 1, \ldots, 4)$ under the prior density $\bar{\pi}(\varphi,\theta) = \bar{p}(\varphi)\pi(\theta|\varphi)$.

Step 3. The final step of this argument involves integrating $\lambda(\varphi)$ with respect to $\bar{p}(\varphi)$ and then making $\bar{p}(\varphi)$ degenerate at φ. We consider the integral

$$
\begin{aligned}
\int \lambda(\varphi)\bar{p}(\varphi)\mathrm{d}\varphi =& n^{-\alpha/2} \iint B_\alpha(\theta,\varphi)\Bigg[1+\frac{1}{n}\Bigg\{-\alpha\left(\frac{\alpha}{1-\alpha}S_3+\frac{3\alpha(2-\alpha)}{(1-\alpha)^2}S_4\right) \\
&+\left(\frac{\alpha}{1-\alpha}\bar{S}_3+\frac{3\alpha(2-\alpha)}{(1-\alpha)^2}\bar{S}_4\right)+\frac{\alpha(\alpha+1)}{2}\left(\frac{1}{1-\alpha}S_1^2+\frac{6}{(1-\alpha)^2}S_1S_2\right) \\
&-\alpha\left(\frac{1}{1-\alpha}S_1\bar{S}_1+\frac{3}{(1-\alpha)^2}S_1\bar{S}_2+\frac{3}{(1-\alpha)^2}\bar{S}_1S_2\right)+K_2(\theta,\varphi)\Bigg\} \\
&+O\left(n^{-2}\right)\Bigg]\pi(\theta|\varphi)\bar{p}(\varphi)\mathrm{d}\theta\mathrm{d}\varphi,
\end{aligned}
\tag{26}
$$

where $B_\alpha(\theta,\varphi)=\frac{(2\pi)^{\alpha/2}\lambda^{-\alpha}(\theta,\varphi)}{\sqrt{1-\alpha}}$. From $\bar{\pi}=\bar{\pi}(\theta,\varphi)=\bar{p}(\varphi)\pi(\theta|\varphi)$ we note that

$$
\begin{aligned}
\frac{\bar{\pi}_{01}}{\bar{\pi}_{00}} &= \frac{(\partial/\partial\varphi)\{\bar{p}(\varphi)\pi(\theta|\varphi)\}}{\bar{p}(\varphi)\pi(\theta|\varphi)} = \frac{(\partial/\partial\varphi)\bar{p}(\varphi)}{\bar{p}(\varphi)}+\frac{(\partial/\partial\varphi)\pi(\theta|\varphi)}{\pi(\theta|\varphi)}, \\
\frac{\bar{\pi}_{02}}{\bar{\pi}_{00}} &= \frac{(\partial^2/\partial\varphi^2)\bar{p}(\varphi)}{\bar{p}(\varphi)}+2\frac{(\partial/\partial\varphi)\bar{p}(\varphi)(\partial/\partial\varphi)\pi(\theta|\varphi)}{\bar{p}(\varphi)\pi(\theta|\varphi)}+\frac{(\partial^2/\partial\varphi^2)\pi(\theta|\varphi)}{\pi(\theta|\varphi)}.
\end{aligned}
\tag{27}
$$

Similarly, since $\pi=\pi(\theta,\varphi)=\pi(\varphi)\pi(\theta|\varphi)$, we have the same formulae as (27) for π_{01}/π_{00} and π_{02}/π_{00}, respectively. Because we are only interested in the terms depending $\pi(\varphi)$ and its derivatives, we divide terms in (26) which involve π and its derivatives into two parts. For example,

$$
\begin{aligned}
&\int B_\alpha(\theta,\varphi)\bar{S}_4\pi(\theta|\varphi)\mathrm{d}\theta \\
=&\int B_\alpha(\theta,\varphi)\left(\frac{\bar{\pi}_{01}A_{03}}{\bar{\pi}_{00}\lambda^4}-\frac{2A_{11}A_{03}}{c\lambda^4}\right)\pi(\theta|\varphi)\mathrm{d}\theta \\
=&\int \frac{\bar{\pi}_{01}}{\bar{\pi}_{00}}\frac{A_{03}(\theta,\varphi)B_\alpha(\theta,\varphi)}{\lambda^4(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta+(\text{terms not involving }\bar{p}(\varphi)) \\
=&\int \frac{(\partial/\partial\varphi)\bar{p}(\varphi)}{\bar{p}(\varphi)}\frac{A_{03}(\theta,\varphi)B_\alpha(\theta,\varphi)}{\lambda^4(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta \\
&+\int \frac{(\partial/\partial\varphi)\pi(\theta|\varphi)}{\pi(\theta|\varphi)}\frac{A_{03}(\theta,\varphi)B_\alpha(\theta,\varphi)}{\lambda^4(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta+(\text{terms not involving }\bar{p}(\varphi)) \\
=&\int \frac{(\partial/\partial\varphi)\bar{p}(\varphi)}{\bar{p}(\varphi)}\frac{A_{03}(\theta,\varphi)B_\alpha(\theta,\varphi)}{\lambda^4(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta+(\text{terms not involving }\bar{p}(\varphi)).
\end{aligned}
$$

Here, we are not interested in the terms not involving $\bar{p}(\varphi)$. Later, in our final asymptotic expansion, we put all of these terms into one term $K_3(\varphi)$. We can rewrite (26) as

$$
\begin{aligned}
\int \lambda(\varphi)\bar{p}(\varphi)\mathrm{d}\varphi =& n^{-\alpha/2}\int\Bigg[\int B_\alpha(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \\
&+\frac{1}{n}\Bigg\{-\frac{\alpha^2}{1-\alpha}\left(\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}H_1^{(\alpha)}(\varphi)+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{3\alpha^2(2-\alpha)}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&+ \frac{\alpha}{1-\alpha}\left(\frac{\bar{p}_{\varphi\varphi}(\varphi)}{\bar{p}(\varphi)}H_1^{(\alpha)}(\varphi) + \frac{\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_2^{(\alpha)}(\varphi) - \frac{\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_3^{(\alpha)}(\varphi)\right) \\
&+ \frac{3\alpha(2-\alpha)}{(1-\alpha)^2}\left(\frac{\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&+ \frac{\alpha(\alpha+1)}{(1-\alpha)}\left(\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2 H_1^{(\alpha)}(\varphi) + \frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi) - \frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\right) \\
&+ \frac{3\alpha(\alpha+1)}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&- \frac{\alpha}{1-\alpha}\Bigg(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\frac{2\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_2^{(\alpha)}(\varphi) + \frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi) + \frac{\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_2^{(\alpha)}(\varphi) \\
&\qquad\qquad - \frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi) - \frac{\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_3^{(\alpha)}(\varphi)\Bigg) \\
&- \frac{3\alpha}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&- \frac{3\alpha}{(1-\alpha)^2}\left(\frac{\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_4^{(\alpha)}(\varphi)\right) + K_3(\varphi)\Bigg\} + O(n^{-2})\Bigg]\bar{p}(\varphi)\mathrm{d}\varphi,
\end{aligned}
$$

where

$$
\begin{aligned}
H_1^{(\alpha)}(\varphi) &:= \int \frac{B_\alpha(\theta,\varphi)}{2\lambda^2(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta \\
H_2^{(\alpha)}(\varphi) &:= \int \frac{B_\alpha(\theta,\varphi)}{\lambda^2(\theta,\varphi)}\frac{\pi_\varphi(\theta|\varphi)}{\pi(\theta|\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta, \\
H_3^{(\alpha)}(\varphi) &:= \int \frac{2A_{11}(\theta,\varphi)B_\alpha(\theta,\varphi)}{c(\theta,\varphi)\lambda^2(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta, \\
H_4^{(\alpha)}(\varphi) &:= \int \frac{A_{03}(\theta,\varphi)B_\alpha(\theta,\varphi)}{\lambda^4(\theta,\varphi)}\pi(\theta|\varphi)\mathrm{d}\theta
\end{aligned}
$$

and $K_3(\varphi)$ is a continuous function, not involving $\pi(\varphi)$ and $\bar{p}(\varphi)$. Also, with the choice of $\bar{p}(\varphi)$ which values on the boundary of parameter space is zero, one can prove that for any twice differentiable function of φ, say $H(\varphi)$,

$$
\begin{aligned}
\int H(\varphi)\frac{\partial}{\partial\varphi}\bar{p}(\varphi)\mathrm{d}\varphi &= -\int \frac{\partial H(\varphi)}{\partial\varphi}\bar{p}(\varphi)\mathrm{d}\varphi \\
\int H(\varphi)\frac{\partial^2}{\partial\varphi^2}\bar{p}(\varphi)\mathrm{d}\varphi &= \int \frac{\partial^2 H(\varphi)}{\partial\varphi^2}\bar{p}(\varphi)\mathrm{d}\varphi
\end{aligned} \tag{28}
$$

by using the integration by parts. Now suppose that the support of $\bar{p}(\varphi)$ contains the true φ as an interior point. Then arrowing $\bar{p}(\varphi)$ weakly converge to the degenerate density of true φ, we obtain the second order asymptotic approximation of $\mathrm{E}_\varphi[\pi^{-\alpha}(\varphi|X)]$. By using the equations in (28), we have

$$
\begin{aligned}
\mathrm{E}_\varphi\left[\pi^{-\alpha}(\varphi|X)\right] =& n^{-\alpha/2}\Bigg[\int B_\alpha(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \\
&+\frac{1}{n}\Bigg\{-\frac{\alpha^2}{1-\alpha}\left(\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}H_1^{(\alpha)}(\varphi)+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\right) \\
&-\frac{3\alpha^2(2-\alpha)}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&+\frac{\alpha(\alpha+1)}{(1-\alpha)}\left(\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2H_1^{(\alpha)}(\varphi)+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\right) \\
&+\frac{3\alpha(\alpha+1)}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right) \\
&-\frac{\alpha}{1-\alpha}\Bigg(-2\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-2\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_{2,\varphi}^{(\alpha)}(\varphi)+2\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2H_2^{(\alpha)}(\varphi) \\
&\qquad\qquad+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)-\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_3^{(\alpha)}(\varphi)\Bigg) \\
&-\frac{3\alpha}{(1-\alpha)^2}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_4^{(\alpha)}(\varphi)\right)+S(\varphi)\Bigg\}+O(n^{-2})\Bigg],
\end{aligned}
$$

where $S(\varphi)$ is a continuous function, not involving $\pi(\varphi)$ and $H_{2,\varphi}^{(\alpha)}(\varphi)=(\partial/\partial\varphi)H_2^{(\alpha)}(\varphi)$. This completes the proof.

□

In the proof of Lemma 2, we note that

$$
\begin{aligned}
\int \frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\frac{2\bar{p}_\varphi(\varphi)}{\bar{p}(\varphi)}H_2^{(\alpha)}(\varphi)\bar{p}(\varphi)\mathrm{d}\varphi &= 2\int\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)\bar{p}_\varphi(\varphi)\mathrm{d}\varphi \\
&= -2\int\frac{\partial}{\partial\varphi}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)\right)\bar{p}(\varphi)\mathrm{d}\varphi \\
&\to -2\frac{\partial}{\partial\varphi}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)\right),
\end{aligned}
$$

and

$$
\frac{\partial}{\partial\varphi}\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)\right)=\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}H_2^{(\alpha)}(\varphi)+\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}H_{2,\varphi}^{(\alpha)}(\varphi)-\left(\frac{\pi_\varphi(\varphi)}{\pi(\varphi)}\right)^2H_2^{(\alpha)}(\varphi),
$$

where $H_{2,\varphi}^{(\alpha)}(\varphi)=(\partial/\partial\varphi)H_2^{(\alpha)}(\varphi)$.

Proof of Theorem 3. Putting $\alpha=-1$, the first order term in (12) is a constant because of $\pi^{\alpha+1}(\varphi)=$

1. We need to consider the second order term in (13). From (17) we have

$$\begin{aligned}\mathrm{E}_{\varphi}\left[\pi(\varphi|X)\right] =& n^{1/2}\Bigg[\int B_{-1}(\theta,\varphi)\pi(\theta|\varphi)\mathrm{d}\theta \\ &+\frac{1}{n}\left\{\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}M_1(\varphi)+\frac{\pi_{\varphi\varphi}(\varphi)}{\pi(\varphi)}M_2(\varphi)+\left(\frac{\pi_{\varphi}(\varphi)}{\pi(\varphi)}\right)^2 M_3(\varphi)+S(\varphi)\right\}+O(n^{-2})\Bigg],\end{aligned}$$

where $B_{-1}(\theta,\varphi)$, $M_1(\varphi)$, $M_2(\varphi)$ and $M_3(\varphi)$ are defined by Lemma 2 and equation (16).

It suffuces to maximize the following with respect to $\pi(\cdot)$

$$R^{-1}(\pi)\approx\frac{1}{2}\left\{\int \mathrm{E}_{\varphi}[\pi(\varphi|X)]\mathrm{d}\varphi-1\right\}$$

or equivalently

$$\int\left\{M_1(\varphi)\frac{\pi'(\varphi)}{\pi(\varphi)}+M_2(\varphi)\left(\frac{\pi'(\varphi)}{\pi(\varphi)}\right)^2+M_3(\varphi)\frac{\pi''(\varphi)}{\pi(\varphi)}\right\}\mathrm{d}\varphi, \tag{29}$$

where $\pi'(\varphi)=\pi_{\varphi}(\varphi)=(\partial/\partial\varphi)\pi(\varphi)$ and $\pi''(\varphi)=\pi_{\varphi\varphi}(\varphi)=(\partial^2/\partial\varphi^2)\pi(\varphi)$ Putting $y(\varphi)=\pi'(\varphi)/\pi(\varphi)$, the integral in (29) is rewritten as

$$\begin{aligned}J(y):=&\int\left\{M_1(\varphi)y(\varphi)+M_2(\varphi)y^2(\varphi)+M_3(\varphi)(y'(\varphi)+y^2(\varphi))\right\}\mathrm{d}\varphi\\ =&\int\left\{M_1(\varphi)y(\varphi)+(M_2(\varphi)+M_3(\varphi))y^2(\varphi)+M_3(\varphi)y'(\varphi)\right\}\mathrm{d}\varphi\\ =&\int F(\varphi,y(\varphi),y'(\varphi))\mathrm{d}\varphi\quad\text{(say)}.\end{aligned}$$

A candidate of local extremum is found by solving the Euler-Lagrange equation

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

(see e.g. Giaquinta (1983)) or equivalently

$$M_1(\varphi)+2(M_2(\varphi)+M_3(\varphi))y(\varphi)-\frac{\mathrm{d}}{\mathrm{d}\varphi}M_3(\varphi)=0.$$

By solving this equation, we have the following

$$y(\varphi)=\frac{\pi'(\varphi)}{\pi(\varphi)}=\frac{(\partial/\partial\varphi)M_3(\varphi)-M_1(\varphi)}{2(M_2(\varphi)+M_3(\varphi))}=:y^*(\varphi). \tag{30}$$

Hence y^* is a candidate which may be a local extremum. We need to consider the second variation of a functional J. In fact, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}J(y+\varepsilon\eta)=\int\Big[F_{yy}(\varphi,y(\varphi)+\varepsilon\eta(\varphi),y'(\varphi)+\varepsilon\eta'(\varphi))\cdot\eta^2(\varphi)$$

$$+ 2F_{yy'}(\varphi, y(\varphi) + \varepsilon\eta(\varphi), y'(\varphi) + \varepsilon\eta'(\varphi)) \cdot \eta(\varphi)\eta'(\varphi)$$
$$+ F_{y'y'}(\varphi, y(\varphi) + \varepsilon\eta(\varphi), y'(\varphi) + \varepsilon\eta'(\varphi)) \cdot (\eta'(\varphi))^2\Big] \mathrm{d}\varphi$$

for any $\eta(\varphi)$ and small number ε. As $\varepsilon \to 0$, the second variation $\delta^2 J(y;\eta)$ is given by

$$\delta^2 J(y;\eta) = \int \left\{2(M_2(\varphi) + M_3(\varphi))\eta^2(\varphi)\right\} \mathrm{d}\varphi < 0 \tag{31}$$

for any $\eta(\varphi)$ because of $M_2(\varphi) + M_3(\varphi) < 0$. Since it holds $\delta^2 J(y;\eta) < 0$ for any η, the marginal reference prior under the expected chi-square divergence is given by

$$\pi(\varphi) \propto \exp\left(\int \frac{(\partial/\partial\varphi)M_3(\varphi) - M_1(\varphi)}{2(M_2(\varphi) + M_3(\varphi))} \mathrm{d}\varphi\right), \tag{32}$$

where the integral in (32) is the indefinite integral. Therefore, we have the desired result.

□

5. Concluding remarks

Reference priors which maximize the expected α-divergence for multi-parameter non-regular model in the presence of nuisance parameter were given. By using the second order asymptotic approximation for the marginal posterior density of the parameter of interest, we considered the maximization of the expected α-divergence for $\alpha < 1$ with respect to the prior density function. Some examples were also given, and we discuss the differences between the α-divergence ($-1 < \alpha < 0$ and $0 < \alpha < 1$) and KL divergence ($\alpha = 0$).

Further, considering the reference priors for multi-parameter non-regular model in other settings is also interesting problem. For example, Kuboki (1998) discussed the reference priors for Bayesian prediction for regular parametric family of distributions. Although we consider the i.i.d. setting in this paper, Smith (1994) presents the non-regular regression which is the regression model under the error distribution with positive support such as Weibull and exponential distributions. For such non-regular regression model, it is important to derive objective prior for regression coefficient vector. Ghosal (1997) and our results should be extended to such non i.i.d. setting.

Acknowledgments

The author would like to thank two anonymous referees for their valuable comments and suggestions to improve the quality of this article. This work was supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 17K14233.

References

Akahira, M. and Takeuchi, K. (1987). The lower bound for the variance of unbiased estimators for one-directional family of distributions. *Annals of the Institute of Statistical Mathematics*, **39**, Part A, 593–610.

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

Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *Journal of American Statistical Association*, **84**, 200–207.

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *The Annals of Statistics*, **37**, 905–938.

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

Chang, T. and Eaves, D. (1990). Reference priors for the orbit in a group model. *The Annals of Statistics*, **18**(4), 1595–1614.

Chernozhukov, V. and Hong, H. (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica*, **72**(5), 1445–1480.

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

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

Corcuera, J. M. and Giummolè, F. (1999). A generalized Bayes rule for prediction. *Scandinavian Journal of Statistics*, **26**, 265–279.

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

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer, New York.

Datta, G. S. and Mukarjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. Springer, New York.

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

Ghosal, S. (1999). Probability matching priors for non-regular cases. *Biometrika*, **86**, 956–964.

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

Ghosal, S. and Samanta, T. (1997a) Expansion of Bayes risk for entropy loss and reference prior in non-regular cases. *Statistics & Decisions*, **15**, 129–140.

Ghosh, J. K. (1994). *Higher Order Asymptotics*. NSF–CBMS Regional Conference Series in Probability and Statistics, Vol.4, Inst. of Math. Statist., Hayward.

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* (eds. S. S. Gupta and J. O. Berger), 183–199, Springer, New York.

Ghosh, M. (2011). Objective priors: an introduction for frequentists. *Statistical Science*, **26**, 187–202.

Ghosh, M. and Mergel, V. (2009). On the Stein phenomenon under divergence loss and an unknown variance-covariance matrix. *Journal of Multivariate Analysis*, **100**, 2331–2336.

Ghosh, M., Mergel, V. and Datta, G. (2008). Estimation, prediction and the Stein phenomenon under divergence loss. *Journal of Multivariate Analysis*, **99**, 1941–1961.

Ghosh, M., Mergel, V. and Liu, R. (2011). A general divergence criterion for prior selection. *Annals of the Institute of Statistical Mathematics*, **63** (1), 49–58.

Giaquinta, M. (1983). *Multiple integrals in the calculus of variations and nonlinear elliptic systems.* Princeton University Press, Princeton, New Jersey.

Hashimoto, S. (2019). Moment matching priors for non-regular models. *Journal of Statistical Planning and Inference*, **203**, 169–177.

Jeffreys, H. (1961). *Theory of Probability, 3rd Ed.* Oxford University Press, London.

Kuboki, H. (1998). Reference priors for prediction. *Journal of Statistical Planning and Inference*, **69**, 295–317.

Liu, R., Chakrabarti, A., Samanta, T., Ghosh, J. K. and Ghosh, M. (2014). On divergence measures leading to Jeffreys and other reference priors. *Bayesian Analysis*, **9** (2), 331–370.

Maruyama, Y., Matsuda, T. and Ohnishi, T. (2019). Harmonic Bayesian prediction under α-divergence. *IEEE Translations on Information Theory*, **65**, 5352–5366.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67–90.

Smith, R. L. (1994). Nonregular regression. *Biometrika*, **81**(1), 173–83.

Wang, H. and Sun, D. (2012). Objective Bayesian analysis for a truncated model. *Statistics and Probability Letters*, **82**, 2125–2135.