

**THE FORMAL DEFINITION OF REFERENCE PRIORS
UNDER A GENERAL CLASS OF DIVERGENCE**

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

TRI MINH LE

Dr. Dongchu Sun, Dissertation Supervisor

MAY 2014

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE FORMAL DEFINITION OF REFERENCE PRIORS UNDER A
GENERAL CLASS OF DIVERGENCE

presented by Tri Minh Le,
a candidate for the degree of Doctor of Philosophy,
and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Dongchu Sun

Professor Paul Speckman

Professor Chong He

Professor Subharup Guha

Professor Shawn Ni

ACKNOWLEDGEMENTS

This dissertation would not have been done if I had not received a lot of help, support, and guidance from many people.

First, I would like to express my sincere gratitude to my advisor, Dr. Dongchu Sun, for his guidance and his time supervising this dissertation. His guidance not only provides me the foundation to become a good statistician, but also inspires me the love of doing research.

I would also like to express my special thanks to Dr. Paul L. Speckman, Dr. Chong He, Dr. Subharup Guha, and Dr. Shawn Ni for being my committee members and for their valuable comments on my dissertation. I would also thank Dr. Athanasios Micheas for serving on my committee earlier.

I have got a lot of useful knowledge during my course works given by many great professors of the faculty of the Department of Statistics. I am particularly grateful to Dr. Dongchu Sun, Dr. Paul L. Speckman, Dr. Chong He, Dr. Athanasios Micheas, Dr. Nancy Flournoy, Dr. Jianguo Sun, Dr. Christopher K. Wikle, Dr. Lori Thombs, Dr. Subharup Guha, Dr. Sounak Chakraborty, and Dr. Jing Qiu for all that they have taught me.

To my parents, I am especially grateful for their faith in me, for their support, everything they could do for me to help me pursuing my dreams.

Last but not least, my special thanks go to my wife, Vi Nguyen, and my priceless treasure, the two little boys, Harry Le and Benny Le. Without you, I cannot do anything.

TABLE OF CONTENTS

Acknowledgements	ii
Abstract	iv
1 Introduction	1
1.1 Background	1
1.1.1 Invariant priors	2
1.1.1.1 Jeffreys' prior	2
1.1.1.2 Left and Right invariant prior	3
1.1.2 Probability matching prior	4
1.1.3 Reference prior	5
1.2 Goals	7
2 Reference prior	9
2.1 A general class of divergence measures	9
2.2 Reference prior definition	14
3 Construction of reference prior	27
3.1 Reference prior for $\beta = 0$	27
3.2 Reference prior for $\beta < -1$	28
3.3 Reference prior for $0 < \beta < 1$	28
3.3.1 The explicit form of the reference Prior	29
3.3.2 Numerical computation of the reference prior	50
3.4 Examples	51
4 Reference prior under the Chi-squared divergence	89
4.1 Examples without reference priors when $\beta = -1$	90
4.2 Examples with reference priors when $\beta = -1$	93
5 Summary and Comments	114
References	125
VITA	128

THE FORMAL DEFINITION OF REFERENCE PRIORS
UNDER A GENERAL CLASS OF DIVERGENCE

Tri Minh Le

Dongchu Sun, Dissertation Supervisor

ABSTRACT

Bayesian analysis is widely used recently in both theory and application of statistics. The choice of priors plays a key role in any Bayesian analysis. There are two types of priors: subjective priors and objective priors. In practice, however, the difficulties of subjective elicitation and time restrictions frequently limit us to use the objective priors constructed by some formal rules. In this dissertation, our methodology is using reference analysis to derive objective priors.

Objective Bayesian inference makes inference depending only on the assumed model and the available data. The prior distribution used to make an inference is least informative in a certain information-theoretic sense. Recently, Berger, Bernardo and Sun (2009) derived reference priors rigorously in the contexts under Kullback-Leibler divergence. In special cases with common support and other regularity conditions, Ghosh, Mergel and Liu (2011) derived a general f -divergence criterion for prior selection. We generalize Ghosh, Mergel and Liu's (2011) results to the case without common support and show how an explicit expression for the reference prior can be obtained under posterior consistency. The explicit expression can be used to derive new reference priors both analytically and numerically.

Chapter 1

Introduction

1.1 Background

Bayesian analysis is widely used recently in both theory and application of statistics. Given the data \mathbf{x} , a Bayesian analysis first assigns a prior distribution $\pi(\theta)$ to the unknown parameter θ , and then uses the well-known Bayes theorem to get the corresponding posterior,

$$\pi(\theta | \mathbf{x}) = \frac{\pi(\theta)p(\mathbf{x} | \theta)}{m(\mathbf{x})},$$

where $p(\mathbf{x} | \theta)$ is the likelihood of \mathbf{x} given θ and $m(\mathbf{x}) = \int \pi(\theta)p(\mathbf{x} | \theta)d\theta$ is the marginal distribution of \mathbf{x} . Then, one can use this posterior to make inferential statements about the parameter θ . So, the meaning of the Bayes theorem is that beginning with a prior guess about the parameter θ (this guess may be good or

bad), combining with the information about the parameter θ contained in the data, one could update the prior to have a better guess for describing the parameter θ . Obviously, the selection of priors plays an important role in any Bayesian analysis. It has always been a controversial and much debated issue in the Bayesian context.

Ideally, if there were enough information about θ , one could use this knowledge to elicit an appropriate subjective prior. In practice, however, the difficulties of subjective elicitation and time restrictions frequently limit us to use the so-called "objective" priors which are also known as "non-informative priors" or "default priors". Objective priors are those priors constructed by some formal rules, depending only on the model or equivalently the likelihood function and the available data. There is a lot of conceptual and theoretical literature devoted to find appropriate procedures for the expression of objective priors. Consequently, some methods for selecting objective priors have been proposed and we will introduce three main approaches to determine objective priors as below.

1.1.1 Invariant priors

1.1.1.1 Jeffreys' prior

Laplace (1812), one of the earliest proponents of objective priors, used a uniform or a flat prior on the entire parameter space based on the principle that if one knows nothing about θ , then there is no reason to put more weight on one point than the others. Uniform prior is simple and convenient, but it was criticized due to its lack of invariance under one-to-one transformation. This makes people

confused when they are not sure which parameter is preferred, the original one or the reparameterized one.

Due to the above reason, Jeffreys (1946) proposed a prior proportional to the square root of the determinant of the Fisher information matrix $I(\theta)$, which is known as Jeffreys' prior. It could be shown that this prior is invariant under one-to-one transformation. Therefore, one does not need to choose any specific parameterization.

1.1.1.2 Left and Right invariant prior

One has another type of objective priors based on a different invariance principle. Instead of being invariant under one-to-one transformation, the so-called left and right invariant priors are invariant under a certain group of transformations.

According to the general theory of locally compact groups, there exist two measures μ_L and μ_R on $\bar{\mathcal{G}}$, the group of transformations on the parameter space Θ , such that, for any $\bar{g} \in \bar{\mathcal{G}}$ and $B \subset \bar{\mathcal{G}}$,

$$\begin{aligned}\mu_L(\bar{g} \circ B) &= \mu_L(B), \\ \mu_R(B \circ \bar{g}) &= \mu_R(B).\end{aligned}$$

μ_L and μ_R are called the left invariant Haar and the right invariant Haar measures, respectively. Then, based on these measures, one can get the corresponding left and right invariant priors for θ .

1.1.2 Probability matching prior

Welch and Peers (1963) and Peers (1965) proposed another method to find objective prior. The early literature on matching priors centered around those which ensure approximate frequentist validity of the posterior quantiles of a one-dimensional interest parameter. Consider priors $\pi(\cdot)$ for which the relation

$$P_{\theta}[\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, \mathbf{x})] = 1 - \alpha + o(n^{-r/2})$$

holds for $r = 1$ or 2 and for each $\alpha \in (0, 1)$. Here n is the sample size, $\theta = (\theta_1, \dots, \theta_p)'$ is an unknown parameter vector, θ_1 is the one-dimensional interest parameter, $P_{\theta}(\cdot)$ is the frequentist probability measure under θ , and $\theta_1^{(1-\alpha)}(\pi, \mathbf{x})$ is the $(1 - \alpha)th$ posterior quantile of θ_1 under $\pi(\cdot)$. Priors satisfying the above equation for $r = 1$ or 2 are called first or second order matching priors, respectively.

As shown by Peers (1965), writing $I^{-1} = (I^{jk})$, a prior $\pi(\cdot)$ is first order probability matching if and only if it satisfies

$$\sum_{j=1}^p \frac{\partial}{\partial \theta_j} \{ \pi(\theta) I^{j1} (I^{11})^{-1/2} \} = 0.$$

In addition, proved by Mukerjee and Ghosh (1997), if it satisfies

$$\frac{1}{3} \sum_{j,r,s,u=1}^p \frac{\partial}{\partial \theta_u} \{ \pi(\theta) \tau^{jr} L_{jrs} (3\sigma^{su} + \tau^{su}) \} - \sum_{j,r=1}^p \frac{\partial^2}{\partial \theta_j \partial \theta_r} \{ \pi(\theta) \tau^{jr} \} = 0,$$

then $\pi(\cdot)$ is second order probability matching, where

$$L_{jrs} = E \left\{ \frac{\partial^3}{\partial \theta_j \partial \theta_r \partial \theta_s} \log p(x | \theta) \right\},$$

$$\tau^{jr} = I^{j1} I^{r1} / I^{11}, \sigma^{su} = I^{su} - \tau^{su}.$$

1.1.3 Reference prior

Reference analysis, originally introduced by Bernardo (1979) and further developed by Berger and Bernardo (1989, 1992), Clarke and Barron (1990, 1994), Clarke and Sun (1997, 1999), Berger, Bernardo and Sun (2009), and Ghosh (2011), among others, is another approach to derive objective priors.

Reference analysis uses information-theoretical concepts to make precise the idea of an objective prior which should be maximally dominated by the data, in the sense of maximizing a suitable divergence measure between the prior and the posterior. Intuitively, the posterior contains extra information coming from the data. Therefore, the divergence between the prior and the posterior could be viewed as a measure of the amount of the information contained in the data. In other words, the divergence between the prior and the posterior is the measure of the missing information about θ . Another way to justify for maximizing a divergence between the prior and the posterior may be given as follows. If the prior is a point mass, the most informative case, then the posterior will be also a point mass no matter what the data are, and the divergence is zero. So we could

say the bigger the divergence, the lower the information in the prior or smaller its influence on the posterior.

There are two interesting results for the continuous one-parameter problems. First, Berger, Bernardo and Sun (2009) produced a rigorous general definition for the reference prior $\pi(\theta)$ based on the use of the expected Kullback-Leibler divergence between the prior and the corresponding posterior as follows,

$$\pi(\theta) = \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)},$$

$$f_k(\theta) = \exp \left\{ \int p(\mathbf{x}^{(k)} | \theta) \log[\pi^*(\theta | \mathbf{x}^{(k)})] d\mathbf{x}^{(k)} \right\},$$

where θ_0 is an interior point of the parameter space Θ , $\mathbf{x}^{(k)} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ stands for k replications of \mathbf{x} , and $\pi^*(\theta | \mathbf{x}^{(k)})$ is the posterior distribution corresponding to some fixed arbitrary prior $\pi^*(\theta)$.

Second, for the regular one-parameter family of distributions, where there exists the Fisher information, Ghosh, Mergel and Liu (2011) found that, under a general divergence measures which include not only the Kullback-Leibler divergence, but also the Hellinger divergence as well as the Chi-squared divergence, in most of the case, with the exception of the Chi-squared divergence, reference prior is identical with Jeffreys' prior. For the Chi-squared divergence, the reference prior is the fourth root of the Fisher information number.

1.2 Goals

In this dissertation, we revisit the problem of using reference analysis to derive objective priors, generalize the result from Berger, Bernardo and Sun (2009), and make some new contributions based on the use of a very general class of divergence measures, known as β -divergence measures (Amari 1982, Cressie and Read 1984). This class of β -divergence measures includes the Kullback-Leibler divergence in a limiting sense ($\beta \rightarrow 0$), the Hellinger divergence ($\beta = 1/2$), and the Chi-squared divergence ($\beta = -1$).

In Chapter 2, we make precise the definition of the reference prior. Roughly speaking, a reference prior needs to have two properties, permissible and maximizing missing information properties.

In Chapter 3, we show how an explicit expression for the reference prior can be obtained under very weak regularity conditions in the case $0 < |\beta| < 1$ and it can be used to derive new reference prior both analytically and numerically. Indeed, for a model described by density $p(\mathbf{x}^{(k)} | \theta)$ as mentioned before, the formula for the reference prior in this case, $\pi(\theta)$, will be shown to be

$$\pi(\theta) = \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)},$$

$$f_k(\theta) = \left\{ \int p(\mathbf{x}^{(k)} | \theta) [\pi^*(\theta | \mathbf{x}^{(k)})]^{-\beta} d\mathbf{x}^{(k)} \right\}^{-1/\beta},$$

where θ_0 is an interior point of the parameter space Θ , and $\pi^*(\theta | \mathbf{x}^{(k)})$ is the posterior distribution corresponding to some fixed arbitrary prior $\pi^*(\theta)$. A major

interesting property is that this formula holds for any type of continuous parameter models, regardless of the asymptotic nature of the posterior. Various examples to illustrate this formula are also given.

In Chapter 4, we study the reference prior under the Chi-squared divergence, the boundary case, when $\beta = -1$. For regular models, as mentioned above, Ghosh, Mergel and Liu (2011) found that the reference prior is the fourth root of the Fisher information number. For non-regular models where there do not exist the Fisher information number, it turns out that there is no general formula for the reference priors in this case. For some models there exist reference priors, for some other models there do not. This will be illustrated through a couple examples.

Some final comments are made in Chapter 5.

Chapter 2

Reference prior

2.1 A general class of divergence measures

In this present section, we will introduce a general class of divergence measures, known as β -divergence measures including the Kullback-Leibler divergence, the Hellinger divergence, and the Chi-squared divergence. This general class of divergence measures will be used to derive the reference prior later in this dissertation.

Consider a statistical model described by the likelihood $p(\mathbf{x} | \theta)$ as follows,

$$\mathcal{M} = \{p(\mathbf{x} | \theta), \mathbf{x} \in \mathcal{X}, \theta \in \Theta \subset \mathbb{R}\}, \quad (2.1)$$

where \mathbf{x} is the complete data vector and θ is a continuous unknown parameter. Note that \mathbf{x} can have any dependency structure (say, it could consist of n normal random variables with mean 0, variance 1 and correlation ρ). For any prior density

of θ , $\pi(\theta)$, denote the corresponding posterior of θ given \mathbf{x} by $\pi(\theta | \mathbf{x})$,

$$\pi(\theta | \mathbf{x}) = \frac{p(\mathbf{x} | \theta)\pi(\theta)}{m(\mathbf{x})}, \quad (2.2)$$

where $m(\mathbf{x})$ is the marginal density of \mathbf{x} , given by

$$m(\mathbf{x}) = \int_{\Theta} \pi(\theta)p(\mathbf{x} | \theta)d\theta. \quad (2.3)$$

Throughout this dissertation, we will consider the general expected divergence measures between the prior $\pi(\theta)$ and the corresponding posterior $\pi(\theta | \mathbf{x})$ given in Ghosh et al. (2009) as follows

$$R^\beta(\pi) = \frac{1}{\beta(1-\beta)} \left\{ 1 - \int_{\mathcal{X}} \left[\int_{\Theta} \pi^\beta(\theta)\pi^{1-\beta}(\theta | \mathbf{x})d\theta \right] m(\mathbf{x})d\mathbf{x} \right\}. \quad (2.4)$$

Here the value of β is fixed, and the range of β is $(-\infty, 1) \setminus \{0\}$. The most popular choice for β is in the interval $(0, 1)$, and the value when $\beta \rightarrow 0$ needs to be defined by the l'hospital rule. Define

$$R^0(\pi) = \lim_{\beta \rightarrow 0} R^\beta(\pi). \quad (2.5)$$

This general divergence was also considered by Amari (1982) and Cressie and Read (1984) in other contexts. The expected Kullback-Leibler, Hellinger and Chi-squared divergences are special cases of the general expected divergence in (2.4)

as we consider the following three different particular values of $\beta : 0, 1/2$, and -1 , respectively.

1. For $\beta = 0$, we have

$$R^0(\pi) = \lim_{\beta \rightarrow 0} \frac{1 - \int_{\mathcal{X}} \left[\int_{\Theta} \pi^\beta(\theta) \pi^{1-\beta}(\theta | \mathbf{x}) d\theta \right] m(\mathbf{x}) d\mathbf{x}}{\beta(1-\beta)}.$$

By the l'hopital rule, we get

$$\begin{aligned} R^0(\pi) &= \lim_{\beta \rightarrow 0} \frac{- \int \int \left\{ \pi^\beta(\theta) \pi^{1-\beta}(\theta | \mathbf{x}) \log[\pi(\theta)] - \pi^\beta(\theta) \pi^{1-\beta}(\theta | \mathbf{x}) \log[\pi(\theta | \mathbf{x})] \right\} d\theta m(\mathbf{x}) d\mathbf{x}}{1 - 2\beta} \\ &= - \int_{\mathcal{X}} \int_{\Theta} \left\{ \pi(\theta | \mathbf{x}) \log[\pi(\theta)] - \pi(\theta | \mathbf{x}) \log[\pi(\theta | \mathbf{x})] \right\} d\theta m(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{X}} \int_{\Theta} \pi(\theta | \mathbf{x}) \log \left[\frac{\pi(\theta | \mathbf{x})}{\pi(\theta)} \right] d\theta m(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

which is the expected Kullback-Leibler divergence between the prior $\pi(\theta)$ and the corresponding posterior $\pi(\theta | \mathbf{x})$.

2. For $\beta = 1/2$. Assume the prior π is proper, then we have

$$\begin{aligned} R^{\frac{1}{2}}(\pi) &= 4 \left\{ 1 - \int_{\mathcal{X}} \left[\int_{\Theta} \pi^{\frac{1}{2}}(\theta) \pi^{\frac{1}{2}}(\theta | \mathbf{x}) d\theta \right] m(\mathbf{x}) d\mathbf{x} \right\} \\ &= 2 \int_{\mathcal{X}} \left\{ 2 - \left[\int_{\Theta} 2\pi^{\frac{1}{2}}(\theta) \pi^{\frac{1}{2}}(\theta | \mathbf{x}) d\theta \right] \right\} m(\mathbf{x}) d\mathbf{x} \\ &= 2 \int_{\mathcal{X}} \int_{\Theta} \left[\pi(\theta) + \pi(\theta | \mathbf{x}) - 2\pi^{\frac{1}{2}}(\theta) \pi^{\frac{1}{2}}(\theta | \mathbf{x}) \right] d\theta m(\mathbf{x}) d\mathbf{x} \\ &= 2 \int_{\mathcal{X}} m(\mathbf{x}) \left\{ \int_{\Theta} \left[\sqrt{\pi(\theta | \mathbf{x})} - \sqrt{\pi(\theta)} \right]^2 d\theta \right\} d\mathbf{x}, \end{aligned}$$

which is twice of the expected Hellinger divergence between the prior $\pi(\theta)$ and the corresponding posterior $\pi(\theta | \mathbf{x})$.

3. For $\beta = -1$. Assume the prior π is proper, then we have

$$\begin{aligned}
 R^{-1}(\pi) &= -\frac{1}{2} \left\{ 1 - \int_{\mathcal{X}} \left[\int_{\Theta} \pi^{-1}(\theta) \pi^2(\theta | \mathbf{x}) d\theta \right] m(\mathbf{x}) d\mathbf{x} \right\} \\
 &= \frac{1}{2} \int_{\mathcal{X}} \left[\int_{\Theta} \frac{\pi^2(\theta | \mathbf{x})}{\pi(\theta)} d\theta - 1 \right] m(\mathbf{x}) d\mathbf{x} \\
 &= \frac{1}{2} \int_{\mathcal{X}} \int_{\Theta} \left[\frac{\pi^2(\theta | \mathbf{x})}{\pi(\theta)} - 2\pi(\theta | \mathbf{x}) + \pi(\theta) \right] d\theta m(\mathbf{x}) d\mathbf{x} \\
 &= \frac{1}{2} \int_{\mathcal{X}} \left\{ \int_{\Theta} \frac{[\pi(\theta | \mathbf{x}) - \pi(\theta)]^2}{\pi(\theta)} d\theta \right\} m(\mathbf{x}) d\mathbf{x},
 \end{aligned}$$

which is one-half of the expected Chi-square divergence between the prior $\pi(\theta)$ and the corresponding posterior $\pi(\theta | \mathbf{x})$.

In fact, the general expected divergence (2.4) between a proper prior and its corresponding posterior is always non-negative because of the following theorem.

Theorem 2.1. *Let $\pi(\cdot)$ be any proper prior, so that $\int_{\Theta} \pi(\theta) d\theta = 1$. Then,*

$$R^{\beta}(\pi) \geq 0 \text{ for every } \beta \in (-\infty, 1).$$

Proof. When $\beta = 0$, as mentioned above, this is the case of the expected Kullback-Leibler divergence. So it is well known that $R^0(\pi) \geq 0$.

For $\beta \neq 0$, we write

$$R^{\beta}(\pi) = \frac{1}{\beta(1-\beta)} \left[1 - S^{\beta}(\pi) \right], \quad (2.6)$$

where

$$\begin{aligned} S^\beta(\pi) &= \int_{\mathcal{X}} \left[\int_{\Theta} \pi^\beta(\theta) \pi^{1-\beta}(\theta | \mathbf{x}) d\theta \right] m(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathcal{X}} \int_{\Theta} \left[\frac{\pi(\theta)}{\pi(\theta | \mathbf{x})} \right]^\beta \pi(\theta | \mathbf{x}) d\theta m(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (2.7)$$

Case I. $0 < \beta < 1$. In this case, u^β is a concave function for $u > 0$. It follows from the Jensen's inequality that,

$$\begin{aligned} \int_{\Theta} \left[\frac{\pi(\theta)}{\pi(\theta | \mathbf{x})} \right]^\beta \pi(\theta | \mathbf{x}) d\theta &\leq \left[\int_{\Theta} \frac{\pi(\theta)}{\pi(\theta | \mathbf{x})} \pi(\theta | \mathbf{x}) d\theta \right]^\beta \\ &\leq \left[\int_{\Theta} \pi(\theta) d\theta \right]^\beta \\ &\leq 1. \end{aligned} \quad (2.8)$$

So, (2.7) and (2.8) imply that

$$0 \leq S^\beta(\pi) \leq 1. \quad (2.9)$$

From (2.6) and (2.9) we have

$$0 \leq R^\beta(\pi) \leq \frac{1}{\beta(1-\beta)}. \quad (2.10)$$

This completes the proof of the theorem for *Case I*.

Case II. $\beta < 0$. In this case, u^β is a convex function for $u > 0$.

Again, using the Jensen's inequality, we have

$$\begin{aligned} \int_{\Theta} \left[\frac{\pi(\theta)}{\pi(\theta | \mathbf{x})} \right]^{\beta} \pi(\theta | \mathbf{x}) d\theta &\geq \left[\int_{\Theta} \frac{\pi(\theta)}{\pi(\theta | \mathbf{x})} \pi(\theta | \mathbf{x}) d\theta \right]^{\beta} \\ &\geq \left[\int_{\Theta} \pi(\theta) d\theta \right]^{\beta} \\ &\geq 1. \end{aligned} \tag{2.11}$$

So, (2.7) and (2.11) imply that

$$S^{\beta}(\pi) \geq 1. \tag{2.12}$$

Since $\beta(1 - \beta) < 0$, (2.6) and (2.12) imply that

$$R^{\beta}(\pi) \geq 0. \tag{2.13}$$

This completes the proof of the theorem for *Case II*. □

2.2 Reference prior definition

The selection of priors has always been a controversial and much debated problem in the Bayesian context. Once we do not have any information about priors, we need to use "objective" priors. Berger, Bernardo and Sun (2009) produced a rigorous general definition for the reference prior based on the use of the expected Kullback-Leibler divergence between the prior and the corresponding posterior.

In the case there exists the Fisher information, Ghosh, Mergel and Liu (2011) found an interesting result that, under the general expected divergence (2.4), with one exception (namely the expected Chi-squared divergence), reference prior is identical with Jeffreys' prior.

The goal of this dissertation is to extend the previous results to give a rigorous general definition of the reference prior for a given likelihood in the sense of maximizing the missing information (to be defined later in the definition 2.4) under the general expected divergence (2.4) between the prior and the corresponding posterior.

Consider the model $\mathcal{M} = \{p(\mathbf{x} | \theta), \mathbf{x} \in \mathcal{X}, \theta \in \Theta \subset \mathbb{R}\}$ with a strictly positive continuous prior $\pi(\theta)$ (this strictly positive restriction will eliminate many technical details). Then, if $\pi(\theta)$ is improper the Bayes theorem no longer applies and the use of the formal posterior $\pi(\theta | \mathbf{x})$ should be justified. The posterior $\pi(\theta | \mathbf{x})$ should be a suitable limit of a sequence of posteriors obtained from proper priors. If we restrict the prior to an increasing sequence of compact sets, we will have this design. Moreover, reference priors are often viewed as priors that yield a good approximation to the analyses on the unbounded parameter space.

Definition 2.2. (Approximating Compact Sequence) Consider the model \mathcal{M} with a strictly positive continuous prior $\pi(\theta)$ as above. An approximating compact sequence of the parameter space Θ is an increasing sequence of compact subsets of Θ , $\{\Theta_i\}_{i=1}^{\infty}$, converging to Θ . The corresponding sequence of posteriors with support on Θ_i , defined as $\{\pi_i(\theta | \mathbf{x})\}_{i=1}^{\infty}$, is called the approximating sequence of

posteriors to the posterior $\pi(\theta | \mathbf{x})$, where

$$\pi_i(\theta) = \frac{\pi(\theta)}{\int_{\Theta_i} \pi(\theta) d\theta} 1_{\Theta_i}(\theta), \quad (2.14)$$

$$m_i(\mathbf{x}) = \int_{\Theta_i} \pi_i(\theta) p(\mathbf{x} | \theta) d\theta, \quad (2.15)$$

$$\pi_i(\theta | \mathbf{x}) = \frac{p(\mathbf{x} | \theta) \pi_i(\theta)}{m_i(\mathbf{x})}. \quad (2.16)$$

Then, for some approximating compact sequence $\{\Theta_i\}_{i=1}^{\infty}$ of the parameter space Θ , the general expected divergence between the formal posterior $\pi(\theta | \mathbf{x})$ and the posterior $\pi_i(\theta | \mathbf{x})$ is

$$R^\beta[\pi(\cdot | \mathbf{x}), \pi_i(\cdot | \mathbf{x})] = \frac{1 - \int_{\mathcal{X}} \left[\int_{\Theta_i} \pi^\beta(\theta | \mathbf{x}) \pi_i^{1-\beta}(\theta | \mathbf{x}) d\theta \right] m_i(\mathbf{x}) d\mathbf{x}}{\beta(1-\beta)}. \quad (2.17)$$

We restrict our possibly reference priors to those that satisfy the following condition,

$$\lim_{i \rightarrow \infty} R^\beta[\pi(\cdot | \mathbf{x}), \pi_i(\cdot | \mathbf{x})] = 0. \quad (2.18)$$

This condition means that a prior which satisfies this condition will yield a posterior that, on the average over \mathbf{x} , is a good approximation to the proper posterior resulting from restriction to a large compact subset Θ_i of the parameter space Θ . We define these candidate priors as follows.

Definition 2.3. (Permissible Prior). A strictly positive continuous function $\pi(\theta)$ is called a permissible prior for the model $\mathcal{M} = \{p(\mathbf{x} | \theta), \mathbf{x} \in \mathcal{X}, \theta \in \Theta \subset \mathbb{R}\}$ if

1. $\pi(\theta | \mathbf{x})$ is proper i.e. $\int_{\Theta} p(\mathbf{x} | \theta)\pi(\theta)d\theta < \infty$ for all $\mathbf{x} \in \mathcal{X}$;
2. for some approximating compact sequence $\{\Theta_i\}_{i=1}^{\infty}$ of the parameter space Θ , the corresponding posterior sequence is convergent to the formal posterior $\pi(\theta | \mathbf{x})$, i.e. the limit in (2.18) is satisfied.

Notice that a prior could not be permissible for a small sample size, but it could turn to be permissible for a larger sample size.

In the following, we assume that for a simple random sample of size k , $\mathbf{x}^{(k)} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ (or the k -replicate of the model \mathcal{M}), there exists a sufficient statistic $\mathbf{t}_k = \mathbf{t}_k(x_1, \dots, x_k) \in \mathcal{T}_k$. While \mathbf{t}_k could be just $\mathbf{x}^{(k)}$ itself, it is computationally convenient to do with sufficient statistics if they are available. We denote

$$\mathcal{P} = \left\{ p(\theta) > 0 : \int_{\Theta} p(\mathbf{t}_k | \theta)p(\theta)d\theta < \infty \right\}.$$

For any compact subset $\Theta_0 \subset \Theta$ and any $p \in \mathcal{P}$, we write the renormalized restrictions of $p(\theta)$ to Θ_0 , the corresponding marginal likelihood and posterior as follows

$$p_0(\theta) = \frac{p(\theta)}{\int_{\Theta_0} p(\theta) d\theta} 1_{\Theta_0}(\theta), \quad (2.19)$$

$$m_0(\mathbf{t}_k) = \int_{\Theta_0} p_0(\theta) p(\mathbf{t}_k | \theta) d\theta, \quad (2.20)$$

$$p_0(\theta | \mathbf{t}_k) = \frac{p(\mathbf{t}_k | \theta) p_0(\theta)}{m_0(\mathbf{t}_k)}. \quad (2.21)$$

The corresponding general expected divergence between prior p_0 and its posterior $p_0(\theta | \mathbf{t}_k)$ is

$$R_{0k}^\beta(p_0) = \frac{1}{\beta(1-\beta)} \left\{ 1 - \int_{\mathcal{T}_k} \left[\int_{\Theta_0} p_0^\beta(\theta) p_0^{1-\beta}(\theta | \mathbf{t}_k) d\theta \right] m_0(\mathbf{t}_k) d\mathbf{t}_k \right\}. \quad (2.22)$$

Here the value of β is fixed, and again the range of β is $(-\infty, 1)$.

We could view the general expected divergence between the prior $p(\theta)$ and its posterior $p(\theta | \mathbf{x})$, $R^\beta(p)$, as being the information to be expected over the data $\mathbf{x}^{(k)}$ from the model $\mathcal{M} = \{p(\mathbf{x} | \theta), \mathbf{x} \in \mathcal{X}, \theta \in \Theta \subset \mathbb{R}\}$. As the sample size $k \rightarrow \infty$, the sequence of realizations $(\mathbf{x}_1, \dots, \mathbf{x}_k)$ would eventually provide any missing information about the value of the parameter θ . In other words, as $k \rightarrow \infty$, $R^\beta(p)$ would provide a measure of the missing information about the parameter θ associated to the prior $p(\theta)$. Intuitively, a reference prior would be a permissible prior which maximizes the missing information about the parameter θ . Furthermore, with a continuous parameter space, since an infinite amount of information would be required to learn about the parameter θ , the missing information $R^\beta(p)$ will typically diverge as $k \rightarrow \infty$. Similarly, the missing information

$R^\beta(p)$ is also typically not defined on an unbounded parameter space. Thus, we restrict ourselves to the compact subsets Θ_0 of the parameter space Θ and have the following definition.

Definition 2.4. (Maximizing Missing Information (MMI) Property). A prior $\pi \in \mathcal{P}$ is said to have the MMI property for the model \mathcal{M} given \mathcal{P} if, for any compact set $\Theta_0 \subset \Theta$ and any $p \in \mathcal{P}$, we have:

$$\lim_{k \rightarrow \infty} \left\{ R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right\} \geq 0. \quad (2.23)$$

Now, from the previous considerations we have the definition of the reference prior as follows.

Definition 2.5. (Reference Prior). A function $\pi(\theta) \in \mathcal{P}$ is called a reference prior for the model \mathcal{M} given \mathcal{P} if it is permissible and has the MMI property.

The condition (2.23) means that, as $k \rightarrow \infty$, the missing information for the reference prior $\pi(\theta)$ is larger than the missing information for any other candidate priors $p(\theta)$. Although we feel that a reference prior needs to have both permissible and MMI properties, the MMI property is considerably more important. Thus, others have defined reference priors only in relation to this property.

We see that the definition 2.5 of a reference prior is useful if $R_{0k}^\beta(\pi_0)$ and $R_{0k}^\beta(p_0)$ are finite. Therefore, we consider the conditions under which this will be so.

Definition 2.6. (Standard Prior). Let \mathcal{P}_s be the class of strictly positive and continuous prior functions on Θ which have proper posteriors. We call these the standard prior functions.

From now on, \mathcal{P}_s will be the class of candidate priors we use to define the reference priors. Any prior not satisfying the conditions in the previous definition would not be accepted as being a reasonable candidate for a reference prior.

Definition 2.7. (Standard Model). The model \mathcal{M} is said to be standard if, for any prior $p \in \mathcal{P}_s$ and any compact subset Θ_0 of Θ we have

$$R_{0k}^\beta(p_0) < \infty. \quad (2.24)$$

Notice that in the case $0 < \beta < 1$, from (2.10), satisfaction of (2.24) is always true. When $\beta = 0$, the case of Kullback-Leibler divergence, there are some conditions under which (2.24) is satisfied given in Lemma 1 and Lemma 2 of Berger, Bernardo and Sun (2009). When $\beta < 0$, we also have the following condition to check (2.24).

Theorem 2.8. When $\beta < 0$, for $p(\theta) \in \mathcal{P}_s$ and any compact subset Θ_0 of Θ , (2.24) is satisfied if, for any $\theta \in \Theta_0$ and $\theta' \in \Theta_0$,

$$\int_{\mathcal{T}_k} p^\beta(\mathbf{t}_k | \theta') p^{1-\beta}(\mathbf{t}_k | \theta) d\mathbf{t}_k < \infty. \quad (2.25)$$

Proof. We have

$$R_{0k}^\beta(p_0) = \frac{1}{\beta(1-\beta)} \left[1 - S_k^\beta(p_0) \right], \quad (2.26)$$

where

$$S_k^\beta(p_0) = \int_{\mathcal{T}_k} \left[\int_{\Theta_0} p_0^\beta(\theta) p_0^{1-\beta}(\theta | \mathbf{t}_k) d\theta \right] m_0(\mathbf{t}_k) d\mathbf{t}_k \quad (2.27)$$

$$= \int_{\mathcal{T}_k} \int_{\Theta_0} \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta p_0(\theta | \mathbf{t}_k) d\theta m_0(\mathbf{t}_k) d\mathbf{t}_k. \quad (2.28)$$

From the Fubini Theorem,

$$S_k^\beta(p_0) = \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \quad (2.29)$$

$$= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{m_0(\theta)}{p(\mathbf{t}_k | \theta)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \quad (2.30)$$

Since $m_0(\mathbf{t}_k) = \int_{\Theta_0} p_0(\theta') p(\mathbf{t}_k | \theta') d\theta'$ and u^β is a convex function when $\beta < 0$ for $u > 0$, using the Jensen's inequality, we have

$$\begin{aligned} & \int_{\mathcal{T}_k} \left[\frac{m_0(\theta)}{p(\mathbf{t}_k | \theta)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k \\ &= \int_{\mathcal{T}_k} \left[\int_{\Theta_0} \frac{p(\mathbf{t}_k | \theta')}{p(\mathbf{t}_k | \theta)} p_0(\theta') d\theta' \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k \\ &\leq \int_{\mathcal{T}_k} \int_{\Theta_0} \left[\frac{p(\mathbf{t}_k | \theta')}{p(\mathbf{t}_k | \theta)} \right]^\beta p_0(\theta') d\theta' p(\mathbf{t}_k | \theta) d\mathbf{t}_k \\ &\leq \int_{\Theta_0} \int_{\mathcal{T}_k} p^\beta(\mathbf{t}_k | \theta') p^{1-\beta}(\mathbf{t}_k | \theta) d\mathbf{t}_k p_0(\theta') d\theta'. \end{aligned} \quad (2.31)$$

From the assumption (2.25), if

$$\int_{\mathcal{T}_k} p^\beta(\mathbf{t}_k | \theta') p^{1-\beta}(\mathbf{t}_k | \theta) d\mathbf{t}_k \leq M,$$

then, from (2.31), we have

$$\int_{\mathcal{T}_k} \left[\frac{m_0(\theta)}{p(\mathbf{t}_k | \theta)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k \leq M \int_{\Theta_0} p_0(\theta') d\theta' = M. \quad (2.32)$$

This and (2.30), (2.12) imply

$$1 \leq S_k^\beta(p_0) \leq M \int_{\Theta_0} p_0(\theta) d\theta = M. \quad (2.33)$$

So, by $\beta < 0$, (2.26) and (2.33) imply

$$0 \leq R_{0k}^\beta(p_0) \leq \frac{1-M}{\beta(1-\beta)}. \quad (2.34)$$

This means $R_{0k}^\beta(p_0) < \infty$.

□

The following theorem will show that the general expected divergence between p_0 and the posterior $p_0(\theta | \mathbf{t}_k)$ is monotonically nondecreasing in sample size.

Theorem 2.9. $R_{0,k+1}^\beta(p_0) \geq R_{0k}^\beta(p_0)$ for every $\beta \in (-\infty, 1)$.

Proof. Case I. $\beta = 0$. This is the case of Kullback-Leibler divergence. We need to show

$$\begin{aligned} & \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0(\theta | \mathbf{x}_1, \mathbf{x}_2) \log \left[\frac{p_0(\theta | \mathbf{x}_1, \mathbf{x}_2)}{p_0(\theta)} \right] d\theta m_0(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ & \geq \int_{\mathcal{X}_1} \int_{\Theta_0} p_0(\theta | \mathbf{x}_1) \log \left[\frac{p_0(\theta | \mathbf{x}_1)}{p_0(\theta)} \right] d\theta m_0(\mathbf{x}_1) d\mathbf{x}_1. \end{aligned} \quad (2.35)$$

First, since the relations

$$p_0(\theta | \mathbf{x}_1, \mathbf{x}_2) m_0(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1, \mathbf{x}_2 | \theta) p_0(\theta), \quad (2.36)$$

$$p_0(\theta | \mathbf{x}_1) m_0(\mathbf{x}_1) = p(\mathbf{x}_1 | \theta) p_0(\theta), \quad (2.37)$$

$$p(\mathbf{x}_1, \mathbf{x}_2 | \theta) = p(\mathbf{x}_1 | \theta) p(\mathbf{x}_2 | \mathbf{x}_1, \theta), \quad (2.38)$$

$$m_0(\mathbf{x}_1, \mathbf{x}_2) = m_0(\mathbf{x}_1) m_0(\mathbf{x}_2 | \mathbf{x}_1), \quad (2.39)$$

we have

$$\begin{aligned} & \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0(\theta | \mathbf{x}_1, \mathbf{x}_2) \log \left[\frac{p_0(\theta | \mathbf{x}_1, \mathbf{x}_2)}{p_0(\theta)} \right] d\theta m_0(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ & = \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p(\mathbf{x}_1, \mathbf{x}_2 | \theta) p_0(\theta) \log \left[\frac{p(\mathbf{x}_1, \mathbf{x}_2 | \theta)}{m_0(\mathbf{x}_1, \mathbf{x}_2)} \right] d\theta d\mathbf{x}_1 d\mathbf{x}_2 \\ & = \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p(\mathbf{x}_1 | \theta) p(\mathbf{x}_2 | \mathbf{x}_1, \theta) p_0(\theta) \log \left[\frac{p(\mathbf{x}_1 | \theta) p(\mathbf{x}_2 | \mathbf{x}_1, \theta)}{m_0(\mathbf{x}_1) m_0(\mathbf{x}_2 | \mathbf{x}_1)} \right] d\theta d\mathbf{x}_1 d\mathbf{x}_2 \\ & = \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0(\theta | \mathbf{x}_1) m_0(\mathbf{x}_1) p(\mathbf{x}_2 | \mathbf{x}_1, \theta) \log \left[\frac{p_0(\theta | \mathbf{x}_1) p(\mathbf{x}_2 | \mathbf{x}_1, \theta)}{p_0(\theta) m_0(\mathbf{x}_2 | \mathbf{x}_1)} \right] d\theta d\mathbf{x}_1 d\mathbf{x}_2 \\ & = J_1 - J_2, \end{aligned} \quad (2.40)$$

where

$$\begin{aligned}
J_1 &= \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0(\theta | \mathbf{x}_1) m_0(\mathbf{x}_1) p(\mathbf{x}_2 | \mathbf{x}_1, \theta) \log \left[\frac{p_0(\theta | \mathbf{x}_1)}{p_0(\theta)} \right] d\theta d\mathbf{x}_1 d\mathbf{x}_2, \\
J_2 &= \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p(\mathbf{x}_1 | \theta) p_0(\theta) p(\mathbf{x}_2 | \mathbf{x}_1, \theta) \log \left[\frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} \right] d\theta d\mathbf{x}_1 d\mathbf{x}_2.
\end{aligned}$$

Clearly,

$$J_1 = \int_{\mathcal{X}_1} \int_{\Theta_0} p_0(\theta | \mathbf{x}_1) \log \left[\frac{p_0(\theta | \mathbf{x}_1)}{p_0(\theta)} \right] d\theta m_0(\mathbf{x}_1) d\mathbf{x}_1. \quad (2.41)$$

Because $\log(u)$ is a concave function for $u > 0$, using the Jensen's inequality, we have

$$\begin{aligned}
J_2 &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{X}_1} p(\mathbf{x}_1 | \theta) \int_{\mathcal{X}_2} p(\mathbf{x}_2 | \mathbf{x}_1, \theta) \log \left[\frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} \right] d\mathbf{x}_2 d\mathbf{x}_1 d\theta \\
&\leq \int_{\Theta_0} p_0(\theta) \int_{\mathcal{X}_1} p(\mathbf{x}_1 | \theta) \log \left[\int_{\mathcal{X}_2} p(\mathbf{x}_2 | \mathbf{x}_1, \theta) \frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} d\mathbf{x}_2 \right] d\mathbf{x}_1 d\theta \\
&\leq \int_{\Theta_0} p_0(\theta) \int_{\mathcal{X}_1} p(\mathbf{x}_1 | \theta) \log(1) d\mathbf{x}_1 d\theta \\
&\leq 0.
\end{aligned} \quad (2.42)$$

So (2.40), (2.41) and (2.42) imply (2.35).

For $\beta \neq 0$, we need to show the following inequality

$$\begin{aligned}
& \frac{1}{\beta(1-\beta)} \left[1 - \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0^\beta(\theta) p_0^{1-\beta}(\theta | \mathbf{x}_1, \mathbf{x}_2) d\theta m_0(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \right] \\
\geq & \frac{1}{\beta(1-\beta)} \left[1 - \int_{\mathcal{X}_1} \int_{\Theta_0} p_0^\beta(\theta) p_0^{1-\beta}(\theta | \mathbf{x}_1) d\theta m_0(\mathbf{x}_1) d\mathbf{x}_1 \right]. \tag{2.43}
\end{aligned}$$

Similarly as in *Case I*,

$$\begin{aligned}
& \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0^\beta(\theta) p_0^{1-\beta}(\theta | \mathbf{x}_1, \mathbf{x}_2) d\theta m_0(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
= & \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0^\beta(\theta) \left[\frac{p(\mathbf{x}_1, \mathbf{x}_2 | \theta) p_0(\theta)}{m_0(\mathbf{x}_1, \mathbf{x}_2)} \right]^{1-\beta} d\theta m_0(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
= & \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0^\beta(\theta) \left[\frac{p(\mathbf{x}_1 | \theta) p(\mathbf{x}_2 | \mathbf{x}_1, \theta) p_0(\theta)}{m_0(\mathbf{x}_1, \mathbf{x}_2)} \right]^{1-\beta} d\theta m_0(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
= & \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0^\beta(\theta) \left[\frac{p(\mathbf{x}_1 | \theta) p(\mathbf{x}_2 | \mathbf{x}_1, \theta) p_0(\theta)}{m_0(\mathbf{x}_1) m_0(\mathbf{x}_2 | \mathbf{x}_1)} \right]^{1-\beta} d\theta m_0(\mathbf{x}_1) m_0(\mathbf{x}_2 | \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\
= & \int_{\mathcal{X}_1 \times \mathcal{X}_2} \int_{\Theta_0} p_0^\beta(\theta) \left[\frac{p(\mathbf{x}_2 | \mathbf{x}_1, \theta) p_0(\theta | \mathbf{x}_1)}{m_0(\mathbf{x}_2 | \mathbf{x}_1)} \right]^{1-\beta} d\theta m_0(\mathbf{x}_1) m_0(\mathbf{x}_2 | \mathbf{x}_1) d\mathbf{x}_1 d\mathbf{x}_2 \\
= & \int_{\mathcal{X}_1} \int_{\Theta_0} p_0^\beta(\theta) p_0^{1-\beta}(\theta | \mathbf{x}_1) d\theta m_0(\mathbf{x}_1) d\mathbf{x}_1 \int_{\mathcal{X}_2} \left[\frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} \right]^\beta p(\mathbf{x}_2 | \mathbf{x}_1, \theta) d\mathbf{x}_2. \tag{2.44}
\end{aligned}$$

Case II. $0 < \beta < 1$. In this case, u^β is a concave function for $u > 0$. It follows from the Jensen's inequality that,

$$\begin{aligned}
\int_{\mathcal{X}_2} \left[\frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} \right]^\beta p(\mathbf{x}_2 | \mathbf{x}_1, \theta) d\mathbf{x}_2 & \leq \left[\int_{\mathcal{X}_2} \frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} p(\mathbf{x}_2 | \mathbf{x}_1, \theta) d\mathbf{x}_2 \right]^\beta \\
& \leq \left[\int_{\mathcal{X}_2} m_0(\mathbf{x}_2 | \mathbf{x}_1) d\mathbf{x}_2 \right]^\beta \\
& \leq 1. \tag{2.45}
\end{aligned}$$

So (2.44) and (2.45) imply (2.43).

Case III. $\beta < 0$. In this case, u^β is a convex function for $u > 0$. Again, using the Jensen's inequality, we have

$$\begin{aligned} \int_{\mathcal{X}_2} \left[\frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} \right]^\beta p(\mathbf{x}_2 | \mathbf{x}_1, \theta) d\mathbf{x}_2 &\geq \left[\int_{\mathcal{X}_2} \frac{m_0(\mathbf{x}_2 | \mathbf{x}_1)}{p(\mathbf{x}_2 | \mathbf{x}_1, \theta)} p(\mathbf{x}_2 | \mathbf{x}_1, \theta) d\mathbf{x}_2 \right]^\beta \\ &\geq \left[\int_{\mathcal{X}_2} m_0(\mathbf{x}_2 | \mathbf{x}_1) d\mathbf{x}_2 \right]^\beta \\ &\geq 1. \end{aligned} \tag{2.46}$$

Since $\beta(1 - \beta) < 0$, (2.44) and (2.46) imply (2.43). This completes the proof of the theorem. \square

Chapter 3

Construction of reference prior

3.1 Reference prior for $\beta = 0$

We know that when $\beta = 0$ this is the case of the expected Kullback-Leibler divergence. Under the expected Kullback-Leibler divergence Berger, Bernardo, and Sun (2009) have derived a simple constructive formula for a reference prior as follows. For a model described by density $p(\mathbf{x} | \theta)$, where \mathbf{x} is the complete data vector and θ is a continuous unknown parameter, the formula for the reference prior, $\pi(\theta)$, is

$$\begin{aligned}\pi(\theta) &= \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)}, \\ f_k(\theta) &= \exp \left\{ \int p(\mathbf{x}^{(k)} | \theta) \log[\pi^*(\theta | \mathbf{x}^{(k)})] d\mathbf{x}^{(k)} \right\},\end{aligned}$$

where θ_0 is an interior point of the parameter space Θ , $\mathbf{x}^{(k)} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ stands for k replications of \mathbf{x} , and $\pi^*(\theta | \mathbf{x}^{(k)})$ is the posterior distribution corresponding to some fixed arbitrary prior $\pi^*(\theta)$.

3.2 Reference prior for $\beta < -1$

In the case of $\beta < -1$ and if there exists the Fisher information, Ghosh, Mergel and Liu (2009) proved that there is no reference prior in this case. Therefore, we do not have reference prior when $\beta < -1$.

3.3 Reference prior for $0 < |\beta| < 1$

We have some new contributions for this case. The reference prior definition 2.5 does not give an explicit expression to derive a reference prior. Under some mild conditions, the following theorem will give us an explicit expression for the reference prior. Recall that \mathbf{x} refers to the entire vector of observations from the model, while $\mathbf{x}^{(k)} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ refers to a vector of independent replicates of these vector observations from the model. Finally, let $\mathbf{t}_k = \mathbf{t}_k(x_1, \dots, x_k) \in \mathcal{T}_k$ be any sufficient statistic for the replicated observations. While \mathbf{t}_k could be just $\mathbf{x}^{(k)}$ itself, it is computationally convenient to do with sufficient statistics if they are available.

3.3.1 The explicit form of the reference Prior

Theorem 3.1. (*The Explicit Form of the Reference Prior.*) Consider a standard model $\mathcal{M} = \{p(\mathbf{x} | \theta), \mathbf{x} \in \mathcal{X}, \theta \in \Theta \subset \mathbb{R}\}$, and the standard class \mathcal{P}_s of candidate priors. Let $\pi^*(\theta)$ be a function in \mathcal{P}_s such that the corresponding posterior $\pi^*(\theta | \mathbf{t}_k)$ is consistent. Define,

$$f_k(\theta) = \left\{ \int_{\mathcal{T}_k} p(\mathbf{t}_k | \theta) [\pi^*(\theta | \mathbf{t}_k)]^{-\beta} d\mathbf{t}_k \right\}^{-1/\beta}. \quad (3.1)$$

For any interior point $\theta_0 \in \Theta$, define

$$\pi(\theta) = \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)}. \quad (3.2)$$

Assume that

- (i) each $f_k(\theta)$ is continuous and, for any fixed θ and sufficiently large k , $\{f_k(\theta)/f_k(\theta_0)\}$ is either monotonic increasing in k or is bounded above by some function $h(\theta)$ which is integrable on any compact set.

- (ii) $\pi(\theta)$ is a permissible prior function.

Then $\pi(\theta)$ is a reference prior for the model \mathcal{M} given the class \mathcal{P}_s .

Proof. For any fixed compact subset Θ_0 of Θ , and any $p(\theta) \in \mathcal{P}_s$, denote the posterior corresponding to p_0 (the restriction of p to the compact set Θ_0) by $p_0(\theta | \mathbf{t}_k)$.

Due to the Theorem 2.9, $R_{0k}^\beta(p_0)$ is a non-decreasing sequence in k , and by the assumption that the model \mathcal{M} is standard we have $R_{0k}^\beta(p_0) < \infty$. Therefore, the limit of the sequence $R_{0k}^\beta(p_0)$ is finite as $k \rightarrow \infty$. So, by the relation (2.26), the limit of the sequence $S_k^\beta(p_0)$ is also finite as $k \rightarrow \infty$.

Step 1. From (2.29), we show that

$$\begin{aligned}
\lim_{k \rightarrow \infty} S_k^\beta(p_0) &\equiv \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta) \pi_0^*(\theta | \mathbf{t}_k)}{\pi_0^*(\theta | \mathbf{t}_k) p_0(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \tag{3.3}
\end{aligned}$$

Let

$$\begin{aligned}
X_k &= \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta, \\
Y_k &= \left[\frac{\pi_0^*(\theta | \mathbf{t}_k)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta.
\end{aligned}$$

Then, to show (3.3) is equivalent to show

$$\begin{aligned}
\lim_{k \rightarrow \infty} S_k^\beta(p_0) &\equiv \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \tag{3.4}
\end{aligned}$$

It follows from Formula (F.13) of Berger, Bernardo and Sun (2009),

$$\frac{\pi_0^*(\theta | \mathbf{t}_k)}{p_0(\theta | \mathbf{t}_k)} \xrightarrow{P} 1 \text{ in probability } p(\mathbf{t}_k | \theta) \text{ as } k \rightarrow \infty. \quad (3.5)$$

Since p and π^* belong to the standard class \mathcal{P}_s , $p_0(\theta)/\pi_0^*(\theta)$ is bounded above and below on the compact set Θ_0 ,

$$0 < L \leq \frac{p_0(\theta)}{\pi_0^*(\theta)} \leq N \text{ for every } \theta \in \Theta_0. \quad (3.6)$$

Then, we have

$$\begin{aligned} 0 < \frac{\pi_0^*(\theta | \mathbf{t}_k)}{p_0(\theta | \mathbf{t}_k)} &= \frac{p(\mathbf{t}_k | \theta)\pi_0^*(\theta) \int_{\Theta_0} p(\mathbf{t}_k | \theta)p_0(\theta)d\theta}{p(\mathbf{t}_k | \theta)p_0(\theta) \int_{\Theta_0} p(\mathbf{t}_k | \theta)\pi_0^*(\theta)d\theta} \\ &= \frac{\pi_0^*(\theta) \int_{\Theta_0} [p_0(\theta)/\pi_0^*(\theta)]p(\mathbf{t}_k | \theta)\pi_0^*(\theta)d\theta}{p_0(\theta) \int_{\Theta_0} p(\mathbf{t}_k | \theta)\pi_0^*(\theta)d\theta} \\ &\leq \frac{N}{L}. \end{aligned} \quad (3.7)$$

From (3.5), we know that

$$Y_k \xrightarrow{P} 1 \text{ in probability } p(\mathbf{t}_k | \theta) \text{ as } k \rightarrow \infty.$$

This means that

$$\begin{aligned} &\forall \epsilon > 0, \lim_{k \rightarrow \infty} P(|Y_k - 1| < \epsilon) = 1 \\ \Leftrightarrow &\forall \epsilon > 0, \forall \delta > 0, \text{ as } k \rightarrow \infty : |P(|Y_k - 1| < \epsilon) - 1| < \delta \end{aligned} \quad (3.8)$$

$$\begin{aligned}
&\Leftrightarrow 1 - \delta < P(|Y_k - 1| < \epsilon) < 1 + \delta \text{ as } k \rightarrow \infty \\
&\Leftrightarrow P(|Y_k - 1| < \epsilon) > 1 - \delta, \forall \epsilon > 0, \forall \delta > 0, \text{ as } k \rightarrow \infty. \tag{3.9}
\end{aligned}$$

Therefore, we have $1 - \epsilon < Y_k < 1 + \epsilon$ with probability greater than $1 - \delta$ and $Y_k \notin (1 - \epsilon, 1 + \epsilon)$ with probability less than δ as $k \rightarrow \infty$.

From (3.7), we also proved that Y_k is bounded,

$$0 < Y_k < M, \text{ where } M > 1. \tag{3.10}$$

So, if $0 < 1 - \epsilon < 1 + \epsilon < M$, we denote

$$\begin{aligned}
\mathcal{T}_{k,0,1-\epsilon} &= \{\mathbf{t}_k : 0 < Y_k \leq 1 - \epsilon\}, \\
\mathcal{T}_{k,1-\epsilon,1+\epsilon} &= \{\mathbf{t}_k : 1 - \epsilon < Y_k < 1 + \epsilon\}, \\
\mathcal{T}_{k,1+\epsilon,M} &= \{\mathbf{t}_k : 1 + \epsilon \leq Y_k < M\}.
\end{aligned}$$

Then,

$$\mathcal{T}_k = \mathcal{T}_{k,0,1-\epsilon} \cup \mathcal{T}_{k,1-\epsilon,1+\epsilon} \cup \mathcal{T}_{k,1+\epsilon,M}.$$

From this we have the decomposition for $S_k^\beta(p_0)$ as follows

$$\begin{aligned}
S_k^\beta(p_0) &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&= S_{k1} + S_{k2} + S_{k3}, \tag{3.11}
\end{aligned}$$

where

$$\begin{aligned} S_{k1} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\ S_{k2} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\ S_{k3} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1+\epsilon,M}} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned}$$

For S_{k1} :

From (3.9), as $k \rightarrow \infty$, we have

$$\int_{\mathcal{T}_{k,0,1-\epsilon}} p(\mathbf{t}_k | \theta) d\mathbf{t}_k < \delta. \quad (3.12)$$

We now want to show $[p_0(\theta | \mathbf{t}_k)/p_0(\theta)]^\beta$ is bounded almost everywhere. Indeed, we have two cases:

Case I: If $p(\mathbf{t}_k | \theta)$ is bounded, then $m_0(\mathbf{t}_k) = \int_{\Theta_0} p_0(\theta) p(\mathbf{t}_k | \theta) d\theta$ is also bounded. Therefore, the relation $p_0(\theta | \mathbf{t}_k)/p_0(\theta) = p(\mathbf{t}_k | \theta)/m_0(\mathbf{t}_k)$ yields $[p_0(\theta | \mathbf{t}_k)/p_0(\theta)]^\beta$ is bounded.

Case II: If $p(\mathbf{t}_k | \theta)$ is not bounded. As mentioned before, the limit of the sequence $S_k^\beta(p_0)$ is finite as $k \rightarrow \infty$. So $S_k^\beta(p_0)$ is bounded, that means

$$S_k^\beta(p_0) = \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta < A,$$

where A is some constant.

Therefore,

$$p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k < B \text{ almost everywhere, where } B \text{ is some constant.}$$

Since $p_0(\theta)$ is bounded, we then have

$$\left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) < C \text{ almost everywhere, where } C \text{ is some constant.}$$

So, if $p(\mathbf{t}_k | \theta)$ is not bounded, then

$$\left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta < D \text{ almost everywhere, where } D \text{ is some constant.}$$

Therefore, both of these cases give $[p_0(\theta | \mathbf{t}_k)/p_0(\theta)]^\beta$ is bounded almost everywhere. This and (3.10) imply that $X_k Y_k$ and X_k are bounded almost everywhere,

$$X_k Y_k = \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta < D, \quad (3.13)$$

$$X_k = \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta < E. \quad (3.14)$$

Then, by (3.12), as $k \rightarrow \infty$, we have

$$\begin{aligned} S_{k1} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &< D \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \end{aligned}$$

$$\begin{aligned}
&< D\delta \int_{\Theta_0} p_0(\theta) d\theta \\
&< D\delta.
\end{aligned}$$

Let $\delta \rightarrow 0$, we get

$$\lim_{k \rightarrow \infty} S_{k1} = 0. \quad (3.15)$$

For S_{k3} : similarly as for S_{k1} , we also have

$$\lim_{k \rightarrow \infty} S_{k3} = 0. \quad (3.16)$$

For S_{k2} : we have

$$\begin{aligned}
&\int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} (1-\epsilon) X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&< S_{k2} = \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X_k Y_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&< \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} (1+\epsilon) X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta.
\end{aligned}$$

Let $\epsilon \rightarrow 0$, then

$$\lim_{k \rightarrow \infty} S_{k2} = \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \quad (3.17)$$

So, (3.11), (3.15), (3.16), and (3.17) give

$$\begin{aligned}
\lim_{k \rightarrow \infty} S_k^\beta(p_0) &= \lim_{k \rightarrow \infty} S_{k1} + \lim_{k \rightarrow \infty} S_{k2} + \lim_{k \rightarrow \infty} S_{k3} \\
&= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \quad (3.18)
\end{aligned}$$

Now we also have the decomposition

$$\int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta = G_{k1} + G_{k2} + G_{k3},$$

where

$$\begin{aligned}
G_{k1} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\
G_{k2} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\
G_{k3} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1+\epsilon,M}} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta.
\end{aligned}$$

From (3.12) and (3.14), similarly as for S_{k1} and S_{k3} , we have

$$\lim_{k \rightarrow \infty} G_{k1} = \lim_{k \rightarrow \infty} G_{k3} = 0.$$

So,

$$\lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta = \lim_{k \rightarrow \infty} G_{k2}. \quad (3.19)$$

Therefore, (3.18) and (3.19) give the equation (3.3)

$$\begin{aligned} \lim_{k \rightarrow \infty} S_k^\beta(p_0) &= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned}$$

Step 2. From (3.3), we show that

$$\begin{aligned} \lim_{k \rightarrow \infty} S_k^\beta(p_0) &= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi^*(\theta | \mathbf{t}_k)} \frac{\pi^*(\theta | \mathbf{t}_k)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned} \quad (3.20)$$

In fact, for any measurable set $A \subset \mathbb{R}$, we define

$$P^*(A | \mathbf{t}_k) = \int_A \pi^*(\theta | \mathbf{t}_k) d\theta. \quad (3.21)$$

It follows from Berger, Bernardo and Sun (2009) that, for $\theta \in \Theta_0$,

$$\frac{\pi^*(\theta | \mathbf{t}_k)}{\pi_0^*(\theta | \mathbf{t}_k)} = P^*(\Theta_0 | \mathbf{t}_k). \quad (3.22)$$

Then

$$\lim_{k \rightarrow \infty} S_k^\beta(p_0) = \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi^*(\theta | \mathbf{t}_k)} \right]^\beta [P^*(\Theta_0 | \mathbf{t}_k)]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \quad (3.23)$$

From the posterior consistency of π^* ,

$$P^*(\Theta_0 \mid \mathbf{t}_k) \xrightarrow{P} 1 \text{ in probability } p(\mathbf{t}_k \mid \theta) \text{ as } k \rightarrow \infty. \quad (3.24)$$

When $0 < \beta < 1$, $[P^*(\Theta_0 \mid \mathbf{t}_k)]^\beta$ is bounded above and below on Θ_0 ,

$$0 < [P^*(\Theta_0 \mid \mathbf{t}_k)]^\beta \leq 1.$$

So (3.20) holds from (3.24) and the dominated convergence theorem.

When $-1 < \beta < 0$, our goal here is to make the reference prior on the unbounded parameter space Θ by using analyses on any very large compact subset Θ_0 of Θ . The analyses on the subspace Θ_0 would yield the same answer as the analyses on the space Θ . Therefore, we just need to do the proof for any very large compact subset Θ_0 of Θ .

Let a be any fixed constant in the open interval $(0, 1)$. We show that there exists a compact subset Θ_0 in $\{\Theta_i\}_{i=1}^\infty$, the approximating compact sequence of the parameter space Θ , such that for all positive integer k and for all $\mathbf{t}_k \in \mathcal{T}_k$,

$$P^*(\Theta_0 \mid \mathbf{t}_k) = \int_{\Theta_0} \pi^*(\theta \mid \mathbf{t}_k) d\theta \geq a. \quad (3.25)$$

We prove this statement by contradiction. Suppose we do not have the assertion in (3.25), i.e. for all compact subsets $\Theta_i \subseteq \Theta$, there exists \mathbf{t}_{k_i} such that

$$\int_{\Theta_i} \pi^*(\theta | \mathbf{t}_{k_i}) d\theta < a.$$

This implies

$$\lim_{i \rightarrow \infty} \int_{\Theta_i} \pi^*(\theta | \mathbf{t}_{k_i}) d\theta < a < 1. \quad (3.26)$$

On the other hand, by the dominated convergence theorem, we have

$$\begin{aligned} \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi^*(\theta | \mathbf{t}_{k_i}) d\theta &= \lim_{i \rightarrow \infty} \int_{\Theta} \pi^*(\theta | \mathbf{t}_{k_i}) 1_{\Theta_i}(\theta) d\theta \\ &= \lim_{i \rightarrow \infty} \int_{\Theta} \pi^*(\theta | \mathbf{t}_{k_i}) d\theta = 1 \end{aligned}$$

: a contradiction to (3.26). Therefore, we have the statement (3.25).

So, $P^*(\Theta_0 | \mathbf{t}_k)$ is bounded above and below on the large compact subsets $\Theta_0 \subseteq \Theta$,

$$0 < a \leq P^*(\Theta_0 | \mathbf{t}_k) \leq 1.$$

Therefore, when $-1 < \beta < 0$, $[P^*(\Theta_0 | \mathbf{t}_k)]^\beta$ is also bounded above and below on Θ_0 ,

$$1 \leq [P^*(\Theta_0 | \mathbf{t}_k)]^\beta \leq a^\beta, \quad a \in (0, 1).$$

Let

$$\begin{aligned} X'_k &= \left[\frac{p_0(\theta)}{\pi^*(\theta | \mathbf{t}_k)} \right]^\beta, \\ Y'_k &= [P^*(\Theta_0 | \mathbf{t}_k)]^\beta. \end{aligned}$$

From now on we will do the proof as we did in the *Step 2* as follows.

From (3.24), we know that

$$Y'_k \xrightarrow{P} 1 \text{ in probability } p(\mathbf{t}_k | \theta) \text{ as } k \rightarrow \infty.$$

This means that

$$\begin{aligned} &\forall \epsilon > 0, \lim_{k \rightarrow \infty} P(|Y'_k - 1| < \epsilon) = 1 \\ \Leftrightarrow &\forall \epsilon > 0, \forall \delta > 0, \text{ as } k \rightarrow \infty : |P(|Y'_k - 1| < \epsilon) - 1| < \delta \\ \Leftrightarrow &1 - \delta < P(|Y'_k - 1| < \epsilon) < 1 + \delta \text{ as } k \rightarrow \infty \\ \Leftrightarrow &P(|Y'_k - 1| < \epsilon) > 1 - \delta, \forall \epsilon > 0, \forall \delta > 0, \text{ as } k \rightarrow \infty. \end{aligned} \quad (3.27)$$

Therefore, we have $1 - \epsilon < Y'_k < 1 + \epsilon$ with probability greater than $1 - \delta$ and $Y'_k \notin (1 - \epsilon, 1 + \epsilon)$ with probability less than δ as $k \rightarrow \infty$.

We also proved that Y'_k is bounded,

$$0 < Y'_k < M', \text{ where } M' > 1. \quad (3.28)$$

So, if $0 < 1 - \epsilon < 1 + \epsilon < M'$, we denote

$$\begin{aligned}\mathcal{T}_{k,0,1-\epsilon} &= \{\mathbf{t}_k : 0 < Y'_k \leq 1 - \epsilon\}, \\ \mathcal{T}_{k,1-\epsilon,1+\epsilon} &= \{\mathbf{t}_k : 1 - \epsilon < Y'_k < 1 + \epsilon\}, \\ \mathcal{T}_{k,1+\epsilon,M'} &= \{\mathbf{t}_k : 1 + \epsilon \leq Y'_k < M'\}.\end{aligned}$$

Then,

$$\mathcal{T}_k = \mathcal{T}_{k,0,1-\epsilon} \cup \mathcal{T}_{k,1-\epsilon,1+\epsilon} \cup \mathcal{T}_{k,1+\epsilon,M'}.$$

From this we have the decomposition for $S'_k{}^\beta(p_0)$ as follows

$$\begin{aligned}S'_k{}^\beta(p_0) &\equiv \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X'_k Y'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &= S'_{k1} + S'_{k2} + S'_{k3},\end{aligned}\tag{3.29}$$

where

$$\begin{aligned}S'_{k1} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} X'_k Y'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\ S'_{k2} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X'_k Y'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\ S'_{k3} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1+\epsilon,M'}} X'_k Y'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta.\end{aligned}$$

For S'_{k1} :

From (3.27), as $k \rightarrow \infty$, we have

$$\int_{\mathcal{T}_{k,0,1-\epsilon}} p(\mathbf{t}_k | \theta) d\mathbf{t}_k < \delta. \quad (3.30)$$

We now want to show $[\pi_0^*(\theta | \mathbf{t}_k)/p_0(\theta)]^\beta$ is bounded almost everywhere.

Indeed, we have two cases:

Case I: If $p(\mathbf{t}_k | \theta)$ is bounded, then $m_0(\mathbf{t}_k) = \int_{\Theta_0} p_0(\theta) p(\mathbf{t}_k | \theta) d\theta$ is also bounded. Therefore, the relation $p_0(\theta | \mathbf{t}_k)/p_0(\theta) = p(\mathbf{t}_k | \theta)/m_0(\mathbf{t}_k)$ yields $p_0(\theta | \mathbf{t}_k)/p_0(\theta)$ is bounded.

So, by (3.7), we have

$$\left[\frac{\pi_0^*(\theta | \mathbf{t}_k)}{p_0(\theta)} \right]^\beta = \left[\frac{p_0(\theta | \mathbf{t}_k)}{p_0(\theta)} \right]^\beta \left[\frac{\pi_0^*(\theta | \mathbf{t}_k)}{p_0(\theta | \mathbf{t}_k)} \right]^\beta \text{ is bounded.}$$

Case II: If $p(\mathbf{t}_k | \theta)$ is not bounded. Since the limit of the sequence $S_k^{\prime\beta}(p_0)$ is finite as $k \rightarrow \infty$, $S_k^{\prime\beta}(p_0)$ is bounded, that means

$$S_k^{\prime\beta}(p_0) = \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta < A',$$

where A' is some constant. Therefore,

$$p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k < B' \text{ almost everywhere, where } B' \text{ is some constant.}$$

Since $p_0(\theta)$ is bounded, we then have

$\left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) < C'$ almost everywhere, where C' is some constant.

So, if $p(\mathbf{t}_k | \theta)$ is not bounded, then

$\left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta < D'$ almost everywhere, where D' is some constant.

Therefore, both of these cases give $[\pi_0^*(\theta | \mathbf{t}_k)/p_0(\theta)]^\beta$ is bounded almost everywhere. This and (3.28) imply that $X'_k Y'_k$ and X'_k are bounded almost everywhere,

$$X'_k Y'_k = \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta < D', \quad (3.31)$$

$$X'_k = \left[\frac{p_0(\theta)}{\pi_0^*(\theta | \mathbf{t}_k)} \right]^\beta < E'. \quad (3.32)$$

Then, by (3.30), as $k \rightarrow \infty$, we have

$$\begin{aligned} S'_{k1} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} X'_k Y'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &< D' \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &< D' \delta \int_{\Theta_0} p_0(\theta) d\theta \\ &< D' \delta. \end{aligned}$$

Let $\delta \rightarrow 0$, we get

$$\lim_{k \rightarrow \infty} S'_{k1} = 0. \quad (3.33)$$

For S'_{k3} : similarly as for S'_{k1} , we also have

$$\lim_{k \rightarrow \infty} S'_{k3} = 0. \quad (3.34)$$

For S'_{k2} : we have

$$\begin{aligned} & \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} (1-\epsilon) X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ < S'_{k2} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X'_k Y'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ < \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} (1+\epsilon) X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned}$$

Let $\epsilon \rightarrow 0$, then

$$\lim_{k \rightarrow \infty} S'_{k2} = \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \quad (3.35)$$

So, (3.29), (3.33), (3.34), and (3.35) give

$$\begin{aligned} \lim_{k \rightarrow \infty} S_k'^\beta(p_0) &= \lim_{k \rightarrow \infty} S'_{k1} + \lim_{k \rightarrow \infty} S'_{k2} + \lim_{k \rightarrow \infty} S'_{k3} \\ &= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned} \quad (3.36)$$

Now we also have the decomposition

$$\int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta = G'_{k1} + G'_{k2} + G'_{k3},$$

where

$$\begin{aligned} G'_{k1} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,0,1-\epsilon}} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\ G'_{k2} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1-\epsilon,1+\epsilon}} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta, \\ G'_{k3} &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_{k,1+\epsilon,M'}} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned}$$

From (3.30) and (3.32), similarly as for S'_{k1} and S'_{k3} , we have

$$\lim_{k \rightarrow \infty} G'_{k1} = \lim_{k \rightarrow \infty} G'_{k3} = 0.$$

So,

$$\lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta = \lim_{k \rightarrow \infty} G'_{k2}. \quad (3.37)$$

Therefore, (3.36) and (3.37) give the equation (3.20)

$$\begin{aligned} \lim_{k \rightarrow \infty} S_k^\beta(p_0) &= \lim_{k \rightarrow \infty} S_k^{\prime\beta}(p_0) = \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} X'_k p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\ &= \lim_{k \rightarrow \infty} \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \end{aligned}$$

Step 3. Define

$$c_0(f_k) = \int_{\Theta_0} f_k(\theta) d\theta, \quad (3.38)$$

$$\pi_{0k}^*(\theta) = \frac{f_k(\theta)}{c_0(f_k)} 1_{\Theta_0}(\theta), \quad (3.39)$$

$$\tilde{S}_k^\beta(p_0) = \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta | \mathbf{t}_k)} \right]^\beta p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta. \quad (3.40)$$

We show that

$$\tilde{S}_k^\beta(p_0) = \frac{1}{[c_0(f_k)]^\beta} \int_{\Theta_0} \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta p_0(\theta) d\theta. \quad (3.41)$$

In fact,

$$\begin{aligned} \tilde{S}_k^\beta(p_0) &= \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} p(\mathbf{t}_k | \theta) \left[\frac{p_0(\theta) \pi_{0k}^*(\theta)}{\pi_{0k}^*(\theta) \pi^*(\theta | \mathbf{t}_k)} \right]^\beta d\mathbf{t}_k d\theta \\ &= \int_{\Theta_0} \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta [\pi_{0k}^*(\theta)]^\beta \left\{ \int_{\mathcal{T}_k} p(\mathbf{t}_k | \theta) [\pi^*(\theta | \mathbf{t}_k)]^{-\beta} d\mathbf{t}_k \right\} p_0(\theta) d\theta. \end{aligned}$$

From the definition of f_k in (3.1),

$$\begin{aligned} \tilde{S}_k^\beta(p_0) &= \int_{\Theta_0} \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta [\pi_{0k}^*(\theta)]^\beta [f_k(\theta)]^{-\beta} p_0(\theta) d\theta \\ &= \int_{\Theta_0} \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta \left[\frac{\pi_{0k}^*(\theta)}{f_k(\theta)} \right]^\beta p_0(\theta) d\theta \\ &= \int_{\Theta_0} \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta \left[\frac{1}{c_0(f_k)} \right]^\beta p_0(\theta) d\theta. \end{aligned}$$

The last line implies (3.41).

Step 4. It follows from *Steps 2 and 3* that for the prior π defined in (3.2), and any prior p in the standard class \mathcal{P}_s ,

$$\begin{aligned} & \lim_{k \rightarrow \infty} \left[R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right] \\ = & \frac{1}{\beta(1-\beta)} \lim_{k \rightarrow \infty} \frac{1}{[c_0(f_k)]^\beta} \\ & \times \lim_{k \rightarrow \infty} \left\{ \int_{\Theta_0} p_0(\theta) \left[\frac{p_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta d\theta - \int_{\Theta_0} \pi_0(\theta) \left[\frac{\pi_0(\theta)}{\pi_{0k}^*(\theta)} \right]^\beta d\theta \right\}. \end{aligned} \quad (3.42)$$

By the condition (i) of the theorem 3.1, it follows from the monotone or the dominated convergence theorem that

$$\begin{aligned} \lim_{k \rightarrow \infty} \pi_{0k}^*(\theta) &= \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{\int_{\Theta_0} f_k(\theta) d\theta} 1_{\Theta_0}(\theta) \\ &= \lim_{k \rightarrow \infty} \frac{f_k(\theta)/f_k(\theta_0)}{\int_{\Theta_0} f_k(\theta)/f_k(\theta_0) d\theta} 1_{\Theta_0}(\theta) \\ &= \frac{\lim_{k \rightarrow \infty} [f_k(\theta)/f_k(\theta_0)]}{\int_{\Theta_0} \lim_{k \rightarrow \infty} [f_k(\theta)/f_k(\theta_0)] d\theta} 1_{\Theta_0}(\theta) \\ &= \frac{\pi(\theta)}{c_0(\pi)} 1_{\Theta_0}(\theta) = \pi_0(\theta). \end{aligned} \quad (3.43)$$

This implies that $\pi_{0k}^*(\theta)$ is bounded above and below on Θ_0 . Therefore, $p_0(\theta)/\pi_{0k}^*(\theta)$ and $\pi_0(\theta)/\pi_{0k}^*(\theta)$ are also bounded above and below on Θ_0 .

So, from (3.42), (3.43) and the dominated convergence theorem, we have

$$\begin{aligned}
& \lim_{k \rightarrow \infty} \left[R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right] \\
&= \frac{1}{\beta(1-\beta)} \lim_{k \rightarrow \infty} \frac{1}{[c_0(f_k)]^\beta} \left\{ \int_{\Theta_0} p_0(\theta) \left[\frac{p_0(\theta)}{\pi_0(\theta)} \right]^\beta d\theta - \int_{\Theta_0} \pi_0(\theta) d\theta \right\} \\
&= \frac{1}{\beta(1-\beta)} \lim_{k \rightarrow \infty} \frac{1}{[c_0(f_k)]^\beta} \left\{ \int_{\Theta_0} p_0(\theta) \left[\frac{p_0(\theta)}{\pi_0(\theta)} \right]^\beta d\theta - 1 \right\} \\
&= \frac{1}{\beta(1-\beta)} \lim_{k \rightarrow \infty} \frac{1}{[c_0(f_k)]^\beta} \left\{ \int_{\Theta_0} p_0(\theta) \left[\frac{\pi_0(\theta)}{p_0(\theta)} \right]^{-\beta} d\theta - 1 \right\}. \tag{3.44}
\end{aligned}$$

Case I. $0 < \beta < 1$. In this case, $u^{-\beta}$ is a convex function for $u > 0$. It follows from the Jensen's inequality that,

$$\begin{aligned}
\int_{\Theta_0} p_0(\theta) \left[\frac{\pi_0(\theta)}{p_0(\theta)} \right]^{-\beta} d\theta &\geq \left[\int_{\Theta_0} p_0(\theta) \frac{\pi_0(\theta)}{p_0(\theta)} d\theta \right]^{-\beta} \\
&\geq \left[\int_{\Theta_0} \pi_0(\theta) d\theta \right]^{-\beta} \\
&\geq 1. \tag{3.45}
\end{aligned}$$

So, (3.44) and (3.45) imply that

$$\lim_{k \rightarrow \infty} \left[R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right] \geq 0. \tag{3.46}$$

This completes the proof of the theorem for *Case I*, $0 < \beta < 1$.

Case II. $-1 < \beta < 0$. In this case, $u^{-\beta}$ is a concave function for $u > 0$.

Again, using the Jensen's inequality, we have

$$\begin{aligned}
\int_{\Theta_0} p_0(\theta) \left[\frac{\pi_0(\theta)}{p_0(\theta)} \right]^{-\beta} d\theta &\leq \left[\int_{\Theta_0} p_0(\theta) \frac{\pi_0(\theta)}{p_0(\theta)} d\theta \right]^{-\beta} \\
&\leq \left[\int_{\Theta_0} \pi_0(\theta) d\theta \right]^{-\beta} \\
&\leq 1.
\end{aligned} \tag{3.47}$$

Since $\beta(1 - \beta) < 0$, (3.44) and (3.47) imply that

$$\lim_{k \rightarrow \infty} \left[R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right] \geq 0. \tag{3.48}$$

This completes the proof of the theorem for *Case II*, $-1 < \beta < 0$.

Case III. $\beta = -1$. In this case, we have

$$\begin{aligned}
\int_{\Theta_0} p_0(\theta) \left[\frac{\pi_0(\theta)}{p_0(\theta)} \right]^{-\beta} d\theta &= \int_{\Theta_0} p_0(\theta) \frac{\pi_0(\theta)}{p_0(\theta)} d\theta \\
&= \int_{\Theta_0} \pi_0(\theta) d\theta \\
&= 1.
\end{aligned} \tag{3.49}$$

So, from (3.44),

$$\lim_{k \rightarrow \infty} \left[R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right] = 0, \tag{3.50}$$

this means that, as $k \rightarrow \infty$, $R_{0k}^\beta(p_0)$ does not depend on the prior $p(\theta)$ when

$\beta = -1$. □

Notice that the choice of the prior $\pi^*(\theta)$ is arbitrary and can be chosen for computational convenience. Also, the choice of θ_0 is not important.

3.3.2 Numerical computation of the reference prior

In the situations where analytical derivation of reference priors is not feasible, we can use the Theorem 3.1 in the previous section to obtain an approximation to the reference prior through numerical evaluation of the equation (3.1). Moderate values of k will often yield a good approximation to the reference prior. We have the following algorithm to find the reference prior numerically as follows.

ALGORITHM.

Step 1. Choose a moderate value for k ;

Choose an arbitrary positive function $\pi^*(\theta)$, say $\pi^*(\theta) = 1$;

Choose the number of m of samples to be simulated.

Step 2. For any given θ value, repeat, for $j = 1, \dots, m$:

Simulate a random sample $\{x_{1j}, \dots, x_{kj}\}$ of size k from $p(x | \theta)$;

Compute numerically the integral $c_j = \int_{\Theta} \prod_{i=1}^k p(x_{ij} | \theta) \pi^*(\theta) d\theta$;

Evaluate $r_j(\theta) = \left[\prod_{i=1}^k p(x_{ij} | \theta) \pi^*(\theta) / c_j \right]^{-\beta}$.

Step 3. Compute $\pi(\theta) = \left[m^{-1} \sum_{j=1}^m r_j(\theta) \right]^{-1/\beta}$ and store the pair $\{\theta, \pi(\theta)\}$.

Step 4. Repeat *Step 2* and *Step 3* for all θ values for which the pair $\{\theta, \pi(\theta)\}$ is required.

If desired, a continuous approximation to the reference prior $\pi(\theta)$ could be obtained from the computed points by using the interpolation techniques. We will see how to use this algorithm in the next section.

3.4 Examples

In the present section, we will give some examples, including regular and non-regular models, to see how our formula for the reference prior works in the case $0 < |\beta| < 1$.

Example 3.1. *The normal $N(\theta, 1)$ model.*

This is a regular model.

Let $\mathbf{x} = (x_1, \dots, x_k)$ be a random sample from the distribution $N(\theta, 1)$. We know that the sufficient statistic of θ is $t_k = \bar{x}$, where $\bar{x} = (x_1 + \dots + x_k)/k$, and $\bar{x} | \theta \sim p(t_k | \theta) = N(\theta, 1/k)$.

Now, choosing the prior π^* being the constant prior, $\pi^*(\theta) = 1$, we have $\pi^*(\theta | t_k) \sim N(t_k, 1/k)$.

Then,

$$\begin{aligned} & \int_{-\infty}^{\infty} p(t_k | \theta) [\pi^*(\theta | t_k)]^{-\beta} dt_k \\ &= \int_{-\infty}^{\infty} \frac{\sqrt{k}}{\sqrt{2\pi}} \exp\left[-\frac{k}{2}(t_k - \theta)^2\right] \left\{ \frac{\sqrt{k}}{\sqrt{2\pi}} \exp\left[-\frac{k}{2}(\theta - t_k)^2\right] \right\}^{-\beta} dt_k \end{aligned}$$

$$\begin{aligned}
&= \left(\frac{\sqrt{k}}{\sqrt{2\pi}} \right)^{1-\beta} \int_{-\infty}^{\infty} \exp \left[-\frac{k}{2}(1-\beta)(t_k - \theta)^2 \right] dt_k \\
&= \left(\frac{\sqrt{k}}{\sqrt{2\pi}} \right)^{1-\beta} \frac{\sqrt{2\pi}}{\sqrt{k(1-\beta)}} = \left(\frac{\sqrt{k}}{\sqrt{2\pi}} \right)^{-\beta} \frac{1}{\sqrt{1-\beta}}.
\end{aligned}$$

This implies

$$\begin{aligned}
f_k(\theta) &= \left\{ \int_{-\infty}^{\infty} p(t_k | \theta) [\pi^*(\theta | t_k)]^{-\beta} dt_k \right\}^{-1/\beta} \\
&= \frac{\sqrt{k}}{\sqrt{2\pi}} (1-\beta)^{1/(2\beta)}.
\end{aligned} \tag{3.51}$$

Note that $f_k(\theta)$ is a constant of θ . Therefore,

$$\pi(\theta) = \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)} = 1. \tag{3.52}$$

Next we need to check whether the conditions in the Theorem 3.1 are satisfied or not. As mentioned before, since $0 < \beta < 1$, the model here is standard. It is also easy to see that $\pi^*(\theta) = 1$ belongs to \mathcal{P}_s , the class of standard priors.

For the consistency of the posterior $\pi^*(\theta | t_k)$, for any $\theta \in \mathbb{R}$, we have

$$\begin{aligned}
E(\tau - \theta | t_k)^2 &= \text{Var}(\tau | t_k) + [E(\tau - \theta | t_k)]^2 \\
&= \frac{1}{k} + (t_k - \theta)^2.
\end{aligned} \tag{3.53}$$

Using the fact that $t_k = \bar{x} \xrightarrow{P} \theta$ in probability $p(t_k | \theta)$ as $k \rightarrow \infty$, from (3.53) and the Slutsky's theorem, we have

$$E(\tau - \theta | t_k)^2 \xrightarrow{P} 0 \quad (3.54)$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$.

By the Chebyshev's inequality, for any $\epsilon > 0$,

$$P^*(|\tau - \theta| > \epsilon | t_k) \leq \frac{E(\tau - \theta | t_k)^2}{\epsilon^2}.$$

From (3.54) it follows that

$$P^*(|\tau - \theta| > \epsilon | t_k) \xrightarrow{P} 0$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$. In other words, the posterior $\pi^*(\theta | t_k)$ is consistent.

It is easy to verify the function $f_k(\theta)$ in (3.51) satisfies the condition (i) in the Theorem 3.1.

Finally, we verify the condition (ii) in the Theorem 3.1, whether the prior $\pi(\theta) = 1$ is permissible or not.

We first notice that the corresponding posterior $\pi(\theta | t_k) \sim N(t_k, 1/k)$ is proper and by choosing $\Theta_i = [-i, i], i = 1, 2, \dots$, the approximating compact sequence Θ_i will converge to $\Theta = \mathbb{R}$.

We then consider the following limit

$$\lim_{i \rightarrow \infty} R^\beta[\pi(\theta | t_k), \pi_i(\theta | t_k)] \equiv \lim_{i \rightarrow \infty} \frac{1 - \int_{-\infty}^{\infty} \left[\int_{\Theta_i} \pi^\beta(\theta | t_k) \pi_i^{1-\beta}(\theta | t_k) d\theta \right] m_i(t_k) dt_k}{\beta(1 - \beta)}, \quad (3.55)$$

where

$$\begin{aligned} \pi_i(\theta) &= \frac{\pi(\theta)}{\int_{\Theta_i} \pi(\theta) d\theta} 1_{\Theta_i}(\theta) = \frac{1}{2i} 1_{\Theta_i}(\theta), \\ m_i(t_k) &= \int_{\Theta_i} \pi_i(\theta) p(t_k | \theta) d\theta = \frac{1}{2i} \left\{ \Phi \left[(i - t_k) \sqrt{k} \right] - \Phi \left[(-i - t_k) \sqrt{k} \right] \right\}, \\ &\quad (\Phi \text{ is the CDF of the distribution } N(0, 1)) \\ \pi_i(\theta | t_k) &= \frac{p(t_k | \theta) \pi_i(\theta)}{m_i(t_k)} = \frac{\frac{\sqrt{k}}{\sqrt{2\pi}} \exp \left[-\frac{k}{2} (t_k - \theta)^2 \right]}{\left\{ \Phi \left[(i - t_k) \sqrt{k} \right] - \Phi \left[(-i - t_k) \sqrt{k} \right] \right\}} 1_{\Theta_i}(\theta), \\ \pi(\theta | t_k) &= \frac{\sqrt{k}}{\sqrt{2\pi}} \exp \left[-\frac{k}{2} (\theta - t_k)^2 \right]. \end{aligned}$$

We show the limit in (3.55) is 0. In fact,

$$\begin{aligned} &\lim_{i \rightarrow \infty} \int_{-\infty}^{\infty} \left[\int_{\Theta_i} \pi^\beta(\theta | t_k) \pi_i^{1-\beta}(\theta | t_k) d\theta \right] m_i(t_k) dt_k \\ &= \lim_{i \rightarrow \infty} \int_{-\infty}^{\infty} \int_{\Theta_i} \left[\frac{\pi(\theta | t_k)}{\pi_i(\theta | t_k)} \right]^\beta \pi_i(\theta | t_k) m_i(t_k) d\theta dt_k \\ &= \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi_i(\theta) \int_{-\infty}^{\infty} \left[\frac{\pi(\theta | t_k)}{\pi_i(\theta | t_k)} \right]^\beta p(t_k | \theta) dt_k d\theta \\ &= \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi_i(\theta) \int_{-\infty}^{\infty} \left\{ \Phi \left[(i - t_k) \sqrt{k} \right] - \Phi \left[(-i - t_k) \sqrt{k} \right] \right\}^\beta p(t_k | \theta) dt_k d\theta. \end{aligned} \quad (3.56)$$

It is clear that,

$$0 < a < \Phi \left[(i - t_k)\sqrt{k} \right] - \Phi \left[(-i - t_k)\sqrt{k} \right] < 1, \text{ for some } a > 0,$$

and

$$\lim_{i \rightarrow \infty} \left\{ \Phi \left[(i - t_k)\sqrt{k} \right] - \Phi \left[(-i - t_k)\sqrt{k} \right] \right\}^\beta = 1.$$

So, as $0 < |\beta| < 1$, $\left\{ \Phi \left[(i - t_k)\sqrt{k} \right] - \Phi \left[(-i - t_k)\sqrt{k} \right] \right\}^\beta$ is bounded.

Hence, by the dominated convergence theorem, from (3.56),

$$\begin{aligned} & \lim_{i \rightarrow \infty} \int_{-\infty}^{\infty} \left[\int_{\Theta_i} \pi^\beta(\theta | t_k) \pi_i^{1-\beta}(\theta | t_k) d\theta \right] m_i(t_k) dt_k \\ &= \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi_i(\theta) \int_{-\infty}^{\infty} p(t_k | \theta) dt_k d\theta = 1. \end{aligned} \quad (3.57)$$

(3.55) and (3.57) imply

$$\lim_{i \rightarrow \infty} R^\beta[\pi(\theta | t_k), \pi_i(\theta | t_k)] = 0.$$

This means that the prior $\pi(\theta) = 1$ is permissible.

We have already verified all the conditions in the Theorem 3.1. Therefore, the reference prior for the normal $N(\theta, 1)$ model is $\pi(\theta) = 1$, which is consistent with well-known results for reference priors.

Example 3.2. *Uniform*(0, θ) model, $\theta > 0$.

This is a non-regular model where there does not exist the Fisher information and it belongs to the Ghosal-Samanta non-regular class.

Let $\mathbf{x} = (x_1, \dots, x_k)$ be a random sample from *unif*(0, θ), $\theta > 0, k \geq 2$.

The sufficient statistic of θ is $t_k = x_{(k)} = \max\{x_1, \dots, x_k\}$, and

$$t_k | \theta \sim p(t_k | \theta) = \frac{k}{\theta^k} t_k^{k-1}, 0 < t_k < \theta.$$

We also choose $\pi^*(\theta) = 1$, then $\pi^*(\theta | t_k) = \frac{(k-1)t_k^{k-1}}{\theta^k}, \theta > t_k$.

So,

$$\begin{aligned} \int_0^\theta p(t_k | \theta) [\pi^*(\theta | t_k)]^{-\beta} dt_k &= \int_0^\theta \frac{k}{\theta^k} t_k^{k-1} \left[\frac{(k-1)t_k^{k-1}}{\theta^k} \right]^{-\beta} dt_k \\ &= \frac{k(k-1)^{-\beta}}{\theta^{k(1-\beta)}} \int_0^\theta t_k^{(k-1)(1-\beta)} dt_k \\ &= \frac{k(k-1)^{-\beta}}{\theta^{k(1-\beta)}} \left[\frac{t_k^{(k-1)(1-\beta)+1}}{(k-1)(1-\beta)+1} \right]_0^\theta \\ &= \frac{k(k-1)^{-\beta}}{(k-1)(1-\beta)+1} \theta^\beta. \end{aligned}$$

This implies

$$\begin{aligned} f_k(\theta) &= \left\{ \int_{-\infty}^\infty p(t_k | \theta) [\pi^*(\theta | t_k)]^{-\beta} dt_k \right\}^{-1/\beta} \\ &= \frac{k^{-1/\beta} (k-1)}{[(k-1)(1-\beta)+1]^{-1/\beta} \theta}. \end{aligned} \tag{3.58}$$

Therefore,

$$\pi(\theta) = \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)} = \lim_{k \rightarrow \infty} \frac{\theta_0}{\theta} = \frac{\theta_0}{\theta} \propto \frac{1}{\theta}. \quad (3.59)$$

Now we need to check whether the conditions in the Theorem 3.1 are satisfied or not. As mentioned before, since $0 < \beta < 1$, the model here is standard. It is also easy to see that $\pi^*(\theta) = 1$ belongs to \mathcal{P}_s , the class of standard priors.

We will check the consistency of the posterior $\pi^*(\theta | t_k)$ directly by definition.

For any $\theta > 0$, for any $\epsilon > 0$, we have

$$\begin{aligned} \int_{\{\tau: |\tau - \theta| \leq \epsilon\}} \pi^*(\tau | t_k) d\tau &= \int_{\max\{\theta - \epsilon, t_k\}}^{\theta + \epsilon} \frac{(k-1)t_k^{k-1}}{\tau^k} d\tau \\ &= (k-1)t_k^{k-1} \left[\frac{\tau^{-k+1}}{-k+1} \right]_{\max\{\theta - \epsilon, t_k\}}^{\theta + \epsilon} \\ &= \left(\frac{t_k}{\max\{\theta - \epsilon, t_k\}} \right)^{k-1} - \left(\frac{t_k}{\theta + \epsilon} \right)^{k-1}. \end{aligned} \quad (3.60)$$

Since $t_k < \theta + \epsilon$,

$$\left(\frac{t_k}{\theta + \epsilon} \right)^{k-1} \xrightarrow{P} 0 \quad (3.61)$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$.

On the other hand, we notice that

$$P(|t_k - \theta| > \epsilon) = P(\theta - t_k > \epsilon) = P(t_k < \theta - \epsilon)$$

$$\begin{aligned}
&= \int_0^{\theta-\epsilon} \frac{k}{\theta^k} t_k^{k-1} dt_k \\
&= \frac{(\theta - \epsilon)^k}{\theta^k} \\
&= \left(1 - \frac{\epsilon}{\theta}\right)^k.
\end{aligned} \tag{3.62}$$

So $\lim_{k \rightarrow \infty} P(|t_k - \theta| > \epsilon) = 0$, or

$$t_k \xrightarrow{P} \theta \tag{3.63}$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$.

Furthermore, since

$$\begin{aligned}
P(|\max\{\theta - \epsilon, t_k\} - \theta| > \epsilon) &= P(\theta - \max\{\theta - \epsilon, t_k\} > \epsilon) \\
&= P(\max\{\theta - \epsilon, t_k\} < \theta - \epsilon) \\
&= 0,
\end{aligned} \tag{3.64}$$

we also have

$$\max\{\theta - \epsilon, t_k\} \xrightarrow{P} \theta \tag{3.65}$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$.

From (3.63), (3.65) and the Slutsky's theorem we have $t_k / \max\{\theta - \epsilon, t_k\} \xrightarrow{P} 1$,

or

$$\lim_{k \rightarrow \infty} P \left(1 - \frac{t_k}{\max\{\theta - \epsilon, t_k\}} < \epsilon \right) = 1. \quad (3.66)$$

Consider the following probability,

$$\begin{aligned} P \left[1 - \left(\frac{t_k}{\max\{\theta - \epsilon, t_k\}} \right)^k < \epsilon \right] &= P \left(\frac{t_k}{\max\{\theta - \epsilon, t_k\}} > \sqrt[k]{1 - \epsilon} \right) \\ &= P \left(1 - \frac{t_k}{\max\{\theta - \epsilon, t_k\}} < 1 - \sqrt[k]{1 - \epsilon} \right) \end{aligned} \quad (3.67)$$

Since $\sqrt[k]{1 - \epsilon} < 1$ and $\lim_{k \rightarrow \infty} \sqrt[k]{1 - \epsilon} = 1$, from (3.66) and (3.67),

$$\begin{aligned} \lim_{k \rightarrow \infty} P \left[1 - \left(\frac{t_k}{\max\{\theta - \epsilon, t_k\}} \right)^k < \epsilon \right] &= \lim_{k \rightarrow \infty} P \left(1 - \frac{t_k}{\max\{\theta - \epsilon, t_k\}} < 1 - \sqrt[k]{1 - \epsilon} \right) \\ &= 1. \end{aligned}$$

This implies that

$$\left(\frac{t_k}{\max\{\theta - \epsilon, t_k\}} \right)^k \xrightarrow{P} 1 \quad (3.68)$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$.

So (3.60), (3.61) and (3.68) give

$$\int_{\{\tau: |\tau - \theta| \leq \epsilon\}} \pi^*(\tau | t_k) d\tau \xrightarrow{P} 1 \quad (3.69)$$

in probability $p(t_k | \theta)$ as $k \rightarrow \infty$. In other words, the posterior $\pi^*(\theta | t_k)$ is consistent.

It is easy to verify the function $f_k(\theta)$ in (3.58) satisfies the condition (i) in the Theorem 3.1.

Finally, we will verify the condition (ii) in the Theorem 3.1, whether the prior $\pi(\theta) = 1/\theta$ is permissible or not.

The corresponding posterior,

$$\pi(\theta | t_k) = \frac{kt_k^k}{\theta^{k+1}} \quad (3.70)$$

is proper. If we choose $\Theta_i = [1/i, i], i = 1, 2, \dots$, then Θ_i converges to $\Theta = (0, \infty)$.

Consider the following limit

$$\lim_{i \rightarrow \infty} R^\beta[\pi(\theta | t_k), \pi_i(\theta | t_k)] \equiv \lim_{i \rightarrow \infty} \frac{1 - \int_0^\theta \left[\int_{\Theta_i} \pi^\beta(\theta | t_k) \pi_i^{1-\beta}(\theta | t_k) d\theta \right] m_i(t_k) dt_k}{\beta(1 - \beta)}, \quad (3.71)$$

where

$$\begin{aligned} \pi_i(\theta) &= \frac{\pi(\theta)}{\int_{\Theta_i} \pi(\theta) d\theta} 1_{\Theta_i}(\theta) = \frac{1}{2 \log(i)\theta} 1_{\Theta_i}(\theta), \\ m_i(t_k) &= \int_{t_k}^i \pi_i(\theta) p(t_k | \theta) d\theta = \frac{t_k^{k-1}}{2 \log(i)} \left(\frac{1}{t_k^k} - \frac{1}{i^k} \right) \quad (\text{as } i \text{ is sufficiently large}), \\ \pi_i(\theta | t_k) &= \frac{p(t_k | \theta) \pi_i(\theta)}{m_i(t_k)} = \frac{k(it_k)^k}{i^k - t_k^k} \frac{1}{\theta^{k+1}} 1_{(t_k, i)}(\theta). \end{aligned}$$

We show the limit in (3.71) is 0. In fact,

$$\begin{aligned}
& \lim_{i \rightarrow \infty} \int_0^\theta \left[\int_{\Theta_i} \pi^\beta(\theta | t_k) \pi_i^{1-\beta}(\theta | t_k) d\theta \right] m_i(t_k) dt_k \\
&= \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi_i(\theta) \int_0^\theta \left[\frac{\pi(\theta | t_k)}{\pi_i(\theta | t_k)} \right]^\beta p(t_k | \theta) dt_k d\theta \\
&= \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi_i(\theta) \int_0^\theta \left[1 - \left(\frac{t_k}{i} \right)^k \right]^\beta p(t_k | \theta) dt_k d\theta. \tag{3.72}
\end{aligned}$$

It is clear that, as $t_k < i$,

$$0 < a < 1 - \left(\frac{t_k}{i} \right)^k < 1, \text{ for some } a > 0,$$

and

$$\lim_{i \rightarrow \infty} \left[1 - \left(\frac{t_k}{i} \right)^k \right]^\beta = 1.$$

So, as $0 < |\beta| < 1$, $\left[1 - \left(\frac{t_k}{i} \right)^k \right]^\beta$ is bounded.

Hence, by the dominated convergence theorem, from (3.72),

$$\begin{aligned}
& \lim_{i \rightarrow \infty} \int_0^\theta \left[\int_{\Theta_i} \pi^\beta(\theta | t_k) \pi_i^{1-\beta}(\theta | t_k) d\theta \right] m_i(t_k) dt_k \\
&= \lim_{i \rightarrow \infty} \int_{\Theta_i} \pi_i(\theta) \int_0^\theta p(t_k | \theta) dt_k d\theta = 1. \tag{3.73}
\end{aligned}$$

(3.71) and (3.73) imply

$$\lim_{i \rightarrow \infty} R^\beta[\pi(\theta | t_k), \pi_i(\theta | t_k)] = 0.$$

This means that the prior $\pi(\theta) = 1/\theta$ is permissible.

We have already verified all the conditions in the Theorem 3.1. Therefore, the reference prior for the uniform $unif(0, \theta)$ model is $\pi(\theta) = 1/\theta$, which is also consistent with well-known results for reference priors.

Example 3.3. *Uniform($a_1(\theta), a_2(\theta)$) model, where $\Theta = (\theta_0, \infty)$ and $0 < a_1(\theta) < a_2(\theta)$ are both strictly monotonic increasing functions on Θ with derivatives satisfying $0 < a_1'(\theta) < a_2'(\theta)$.*

This is a non-regular model. It has no group invariance structure and is outside of the Ghosal-Samanta non-regular class.

Let $\mathbf{x} = (x_1, \dots, x_k)$ be a random sample from the original uniform distribution on the interval $(a_1(\theta), a_2(\theta))$. Obviously, the sufficient statistic of θ is

$$\mathbf{t}_k \equiv (t_1, t_2) = (x_{(1)}, x_{(k)}) = (\min\{x_1, \dots, x_k\}, \max\{x_1, \dots, x_k\}),$$

with the density

$$p(t_1, t_2 | \theta) = \frac{k(k-1)(t_2 - t_1)^{k-2}}{[a_2(\theta) - a_1(\theta)]^k}, \quad a_1(\theta) < t_1 < t_2 < a_2(\theta).$$

If we choose $\pi^*(\theta) = 1$, the corresponding posterior density of θ is

$$\pi^*(\theta | t_1, t_2) = \frac{1}{[a_2(\theta) - a_1(\theta)]^k m_k(t_1, t_2)}, \quad a_2^{-1}(t_2) < \theta < a_1^{-1}(t_1),$$

where

$$m_k(t_1, t_2) = \int_{a_2^{-1}(t_2)}^{a_1^{-1}(t_1)} \frac{1}{[a_2(s) - a_1(s)]^k} ds.$$

Change variables

$$y_1 = k[a_1^{-1}(t_1) - \theta],$$

$$y_2 = k[\theta - a_2^{-1}(t_2)].$$

It follows from Berger, Bernardo and Sun (2009) that, for fixed $\theta > \theta_0, y_1 > 0, y_2 > 0$, as $k \rightarrow \infty$,

$$p(y_1, y_2 | \theta) \rightarrow p^*(y_1, y_2 | \theta) = \frac{a_1'(\theta)a_2'(\theta)}{[a_2(\theta) - a_1(\theta)]^2} \exp \left[-\frac{a_1'(\theta)y_1 + a_2'(\theta)y_2}{a_2(\theta) - a_1(\theta)} \right].$$

Therefore, as $k \rightarrow \infty$, the variables y_i 's have independent exponential distributions with means $\lambda_i = [a_2(\theta) - a_1(\theta)]/a_i'(\theta), i = 1, 2$.

It also follows from Berger, Bernardo and Sun (2009) that, for fixed $y_1 > 0, y_2 > 0$, as $k \rightarrow \infty$,

$$k[a_2(\theta) - a_1(\theta)]^k m_k(t_1, t_2) \rightarrow \frac{a_2(\theta) - a_1(\theta)}{a_2'(\theta) - a_1'(\theta)} \exp \left[\frac{a_2'(\theta) - a_1'(\theta)}{a_2(\theta) - a_1(\theta)} y_2 \right] \times \left\{ 1 - \exp \left[-\frac{a_2'(\theta) - a_1'(\theta)}{a_2(\theta) - a_1(\theta)} (y_1 + y_2) \right] \right\}.$$

Then, for fixed $\theta > \theta_0$, as $k \rightarrow \infty$,

$$\begin{aligned}
& k^\beta \int_0^\infty \int_0^\infty p(y_1, y_2 | \theta) [\pi^*(\theta | y_1, y_2)]^{-\beta} dy_1 dy_2 \\
&= \int_0^\infty \int_0^\infty p(y_1, y_2 | \theta) \{k[a_2(\theta) - a_1(\theta)]^k m_k(t_1, t_2)\}^\beta dy_1 dy_2 \\
&\rightarrow \int_0^\infty \int_0^\infty p^*(y_1, y_2 | \theta) \left[\frac{a_2(\theta) - a_1(\theta)}{a_2'(\theta) - a_1'(\theta)} \right]^\beta \left(\exp \left[\frac{a_2'(\theta) - a_1'(\theta)}{a_2(\theta) - a_1(\theta)} y_2 \right] \right. \\
&\quad \left. \times \left\{ 1 - \exp \left[-\frac{a_2'(\theta) - a_1'(\theta)}{a_2(\theta) - a_1(\theta)} (y_1 + y_2) \right] \right\} \right)^\beta dy_1 dy_2. \tag{3.74}
\end{aligned}$$

Let $v_i = y_i/\lambda_i, i = 1, 2$, then v_i 's are i.i.d. with the standard exponential distribution.

Define

$$b_i \equiv b_i(\theta) = \frac{a_2'(\theta) - a_1'(\theta)}{a_i'(\theta)}, \quad i = 1, 2,$$

(3.74) becomes

$$\begin{aligned}
k^\beta f_k^{-\beta}(\theta) &\equiv k^\beta \int_0^\infty \int_0^\infty p(y_1, y_2 | \theta) [\pi^*(\theta | y_1, y_2)]^{-\beta} dy_1 dy_2 \\
&\rightarrow \left[\frac{a_2(\theta) - a_1(\theta)}{a_2'(\theta) - a_1'(\theta)} \right]^\beta E \left[e^{\beta b_2 v_2} (1 - e^{-b_1 v_1 - b_2 v_2})^\beta \right] \text{ as } k \rightarrow \infty. \tag{3.75}
\end{aligned}$$

We have

$$\begin{aligned}
& E \left[e^{\beta b_2 v_2} (1 - e^{-b_1 v_1 - b_2 v_2})^\beta \right] \\
&= E \left[e^{\beta b_2 v_2} \left(1 + \sum_{j=1}^{\infty} \binom{\beta}{j} (-1)^j e^{-j b_1 v_1 - j b_2 v_2} \right) \right] \\
&= E(e^{\beta b_2 v_2}) + \sum_{j=1}^{\infty} \binom{\beta}{j} (-1)^j E(e^{-j b_1 v_1}) E(e^{-(j-\beta) b_2 v_2}) \\
&= \frac{1}{1 - \beta b_2} + \sum_{j=1}^{\infty} \binom{\beta}{j} (-1)^j \frac{1}{1 + j b_1} \frac{1}{1 + (j - \beta) b_2}. \tag{3.76}
\end{aligned}$$

On the other hand, using the Euler-Gauss infinite product definition of the gamma function $\Gamma(\cdot)$, we have

$$\begin{aligned}
\binom{\beta}{j} &= \frac{\beta(\beta - 1)(\beta - 2) \cdots (\beta - j + 1)}{j!} \\
&= (-1)^j \frac{1}{j} \frac{(-\beta + j - 1)(-\beta + j - 2) \cdots (-\beta + 1)(-\beta)}{(j - 1)!} \\
&\approx (-1)^j \frac{1}{j} \frac{(j - 1)^{-\beta}}{\Gamma(-\beta)} \\
&\approx \frac{(-1)^j}{j^{\beta+1} \Gamma(-\beta)}, \tag{3.77}
\end{aligned}$$

where the function $\Gamma(\cdot)$ could be extended to negative values by the following equation,

$$\Gamma(x) = \frac{\Gamma(x + 1)}{x}, \quad -1 < x < 0,$$

for example $\Gamma(-1/2) = -2\Gamma(1/2)$.

So,

$$\begin{aligned}
& \sum_{j=1}^{\infty} \binom{\beta}{j} (-1)^j \frac{1}{1+jb_1} \frac{1}{1+(j-\beta)b_2} \\
& \approx \lim_{n \rightarrow \infty} \sum_{j=1}^n \frac{1}{\Gamma(-\beta)} \frac{1}{j^{\beta+1}} \frac{1}{1+jb_1} \frac{1}{1+(j-\beta)b_2} \\
& \approx \frac{1}{\Gamma(-\beta)} \lim_{n \rightarrow \infty} \frac{1}{n^{\beta+2}} \sum_{j=1}^n \frac{1}{n} \frac{1}{(\frac{j}{n})^{\beta+1}} \frac{1}{\frac{1}{n} + (\frac{j}{n})b_1} \frac{1}{\frac{1-\beta b_2}{n} + (\frac{j}{n})b_2} \\
& \approx \frac{1}{\Gamma(-\beta)} \lim_{n \rightarrow \infty} \frac{1}{n^{\beta+2}} \int_{\frac{1}{n}}^1 \frac{1}{x^{\beta+1}} \frac{1}{b_1 x} \frac{1}{b_2 x} dx \\
& \approx \frac{1}{b_1 b_2 \Gamma(-\beta)} \lim_{n \rightarrow \infty} \frac{1}{n^{\beta+2}} \int_{\frac{1}{n}}^1 \frac{1}{x^{\beta+3}} dx \\
& \approx \frac{1}{b_1 b_2 \Gamma(-\beta)} \lim_{n \rightarrow \infty} \frac{1}{n^{\beta+2}} \left[-\frac{1}{\beta+2} (1-n^{\beta+2}) \right] \\
& \approx \frac{1}{b_1 b_2 (\beta+2) \Gamma(-\beta)}. \tag{3.78}
\end{aligned}$$

(3.75), (3.76) and (3.78) imply that the reference prior of θ is

$$\begin{aligned}
\pi(\theta) &= \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)} = \lim_{k \rightarrow \infty} \frac{[k^\beta f_k^{-\beta}(\theta)]^{-1/\beta}}{[k^\beta f_k^{-\beta}(\theta_0)]^{-1/\beta}} \\
&\propto \frac{a'_2(\theta) - a'_1(\theta)}{a_2(\theta) - a_1(\theta)} \left[\frac{1}{1-\beta b_2} + \frac{1}{(\beta+2)\Gamma(-\beta)b_1 b_2} \right]^{-\frac{1}{\beta}}. \tag{3.79}
\end{aligned}$$

Example 3.4. Uniform model with support on (θ, θ^2) , $\theta > 1$, in the case of Hellinger divergence, $\beta = 1/2$.

This is a special case of the previous example where $\theta_0 = 1$, $a_1(\theta) = \theta$, $a_2(\theta) = \theta^2$. Then,

$$\begin{aligned}
a'_1(\theta) &= 1, \quad a'_2(\theta) = 2\theta, \\
b_1 &= \frac{a'_2(\theta) - a'_1(\theta)}{a'_1(\theta)} = 2\theta - 1, \\
b_2 &= \frac{a'_2(\theta) - a'_1(\theta)}{a'_2(\theta)} = \frac{2\theta - 1}{2\theta}.
\end{aligned}$$

Therefore, the reference prior (3.79) becomes

$$\pi(\theta) \propto \frac{(2\theta + 1)^2(2\theta - 1)^5}{\theta^3(\theta - 1)[20\sqrt{\pi}(2\theta - 1)^2 - 2(2\theta + 1)]^2}. \quad (3.80)$$

The figure 3.1 presents the exact reference prior for this problem.

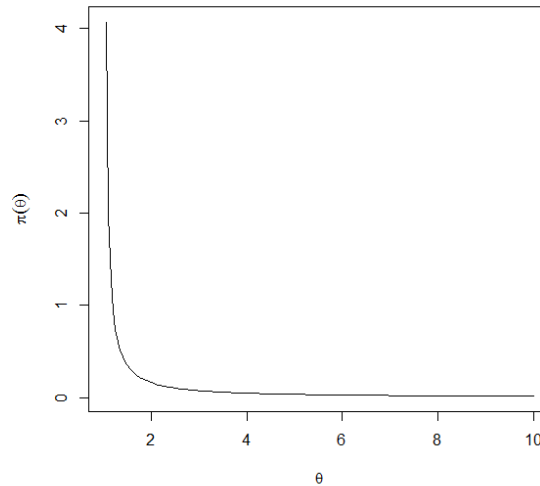


FIGURE 3.1: Exact reference prior for the uniform model on (θ, θ^2)

Now we want to compare this exact reference prior to the numerical reference prior calculated by the algorithm given in the previous section. It is worth to say

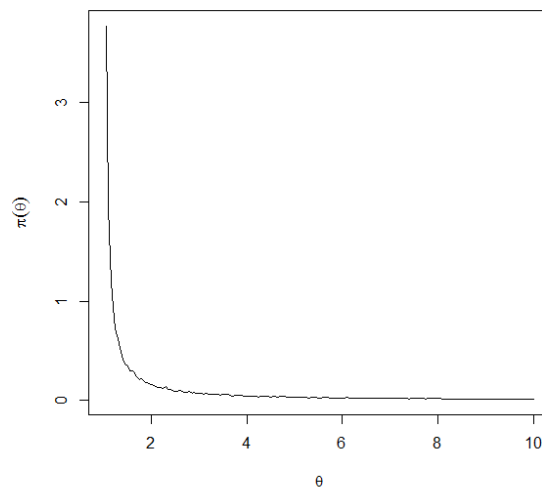
a little bit about how we can generate a random variable from the distribution $p(x | \theta)$ in *Step 2*. We know that if $u \sim \text{uniform}(0, 1)$, then $x = F^{-1}(u) \sim p(x | \theta)$ as desired, where

$$F^{-1}(u) = \theta + (\theta^2 - \theta)u.$$

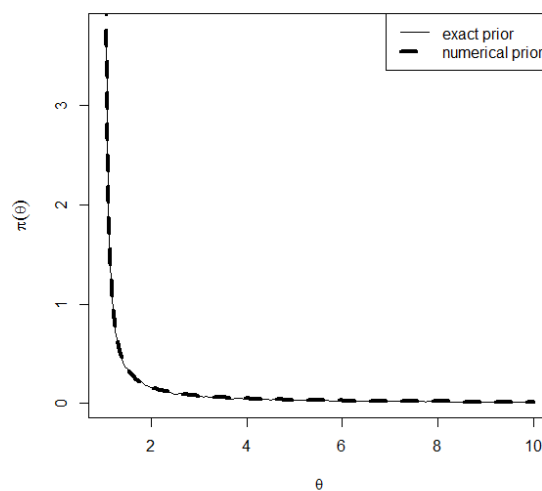
The numerical computation was done for the problem by the following R code and the figure 3.2 presents the numerical reference prior which is rescaled to have $\pi(2) = 0.163$ (since $\pi(2) = 0.163$ for the exact reference prior); $m = 1000$ samples of $k = 500$ observations were used to compute each point $\{\theta, \pi(\theta)\}$.

```
#####
### R code for the figure 3.2###
#####
prior = function(t){
  (2*t+1)^2*(2*t-1)^5/(t^3*(t-1)*(20*sqrt(pi)*(2*t-1)^2-2*(2*t+1))^2)
}
theta = seq(1.05,10,0.05)
true.prior = prior(theta)
true.prior = true.prior/(sum(true.prior*0.05))
plot(theta,true.prior,type="l",xlab=expression(theta),
ylab=expression(pi(theta)))
k = 500
m = 1000
```

```
theta = seq(1.05,10,length=180)
c = rep(NA,m)
r = rep(NA,m)
for (l in 1:length(theta))
{
G = function(y){
theta[l]+(theta[l]^2-theta[l])*y
}
for (j in 1:m){
x = rep(NA,k)
for (i in 1:k){
u = runif(1)
x[i] = G(u)
}
c[j] = integrate(function(t) 1/(t^2-t)^k, sqrt(max(x)),min(x))$value
r[j] = (1/(theta[l]^2-theta[l])^k/c[j])^(-.5)
}
ref[l] = (sum(r)/m)^(-2)
}
ref = ref/ref[theta==2]*0.163
plot(theta,ref,type="l",xlab=expression(theta),ylab=expression(pi(theta)))
lines(theta,true.prior,lty=2,lwd=4)
```

FIGURE 3.2: Numerical reference prior for the uniform model on (θ, θ^2)

When comparing these two plots, exact reference prior and numerical reference prior, we could see that they are obviously almost perfectly the same as shown in the figure 3.3.

FIGURE 3.3: Numerical reference prior vs. exact reference prior for the uniform model on (θ, θ^2)

Example 3.5. Consider the following model,

$$f(x | \theta) = \begin{cases} \frac{g(x)}{c(\theta)}, & \text{if } a_1(\theta) < x < a_2(\theta) \\ 0, & \text{else} \end{cases},$$

where $\Theta = (\theta_0, \infty)$, $c(\theta) = \int_{a_1(\theta)}^{a_2(\theta)} g(x)dx$, $g(x) > 0$ is a non-decreasing function, and $0 < a_1(\theta) < a_2(\theta)$ are both strictly monotonic increasing functions on Θ with derivatives satisfying $0 < a_1'(\theta) < a_2'(\theta)$.

This is the general case of the example 3.3.

Let $\mathbf{x} = (x_1, \dots, x_k)$ be a random sample from the given distribution on the interval $(a_1(\theta), a_2(\theta))$. We also have the sufficient statistic of θ is

$$\mathbf{t}_k \equiv (t_1, t_2) = (x_{(1)}, x_{(k)}) = (\min\{x_1, \dots, x_k\}, \max\{x_1, \dots, x_k\}).$$

Let $F(x | \theta) = \int_{a_1(\theta)}^x f(t | \theta)dt$ is the c.d.f. corresponding the p.d.f. $f(x | \theta)$.

Then,

$$\begin{aligned} p(t_1, t_2 | \theta) &= k(k-1)f(t_1 | \theta)f(t_2 | \theta)[F(t_2 | \theta) - F(t_1 | \theta)]^{k-2} \\ &= k(k-1)g(t_1)g(t_2) \frac{\left[\int_{t_1}^{t_2} g(t)dt \right]^{k-2}}{\left[\int_{a_1(\theta)}^{a_2(\theta)} g(t)dt \right]^k}, \quad a_1(\theta) < t_1 < t_2 < a_2(\theta). \end{aligned} \tag{3.81}$$

Choosing $\pi^*(\theta) = 1$, the corresponding posterior density of θ is

$$\pi^*(\theta | t_1, t_2) = \frac{1}{\left[\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt \right]^k m_k(t_1, t_2)}, \quad a_2^{-1}(t_2) < \theta < a_1^{-1}(t_1),$$

where

$$m_k(t_1, t_2) = \int_{a_2^{-1}(t_2)}^{a_1^{-1}(t_1)} \frac{1}{\left[\int_{a_1(s)}^{a_2(s)} g(t) dt \right]^k} ds. \quad (3.82)$$

Similarly as in the example 3, we change variables as follows

$$\begin{aligned} y_1 &= k[a_1^{-1}(t_1) - \theta], \\ y_2 &= k[\theta - a_2^{-1}(t_2)]. \end{aligned}$$

We then find the asymptotic distribution of (y_1, y_2) as $k \rightarrow \infty$. From the assumption, as k is large enough, for any fixed $y_1 > 0$ and $y_2 > 0$, we have $t_1 = a_1(\theta + y_1/k) < t_2 = a_2(\theta - y_2/k)$. So, from (3.81), the joint density of (y_1, y_2) is

$$\begin{aligned} p(y_1, y_2 | \theta) &= \frac{k-1}{k} \frac{a_1'(\theta + \frac{y_1}{k}) a_2'(\theta - \frac{y_2}{k}) g[a_1(\theta + \frac{y_1}{k})] g[a_2(\theta - \frac{y_2}{k})]}{\left[\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt \right]^2} \\ &\quad \times \left[\frac{\int_{a_1(\theta + \frac{y_1}{k})}^{a_2(\theta - \frac{y_2}{k})} g(t) dt}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right]^{k-2}. \end{aligned} \quad (3.83)$$

By using the facts that, as $k \rightarrow \infty$,

$$\begin{aligned}
a_1(\theta + y_1/k) &\approx a_1(\theta) + \frac{y_1}{k} a_1'(\theta), \\
a_2(\theta - y_2/k) &\approx a_2(\theta) - \frac{y_2}{k} a_2'(\theta), \\
\int_{a_1(\theta)}^{a_1(\theta) + \frac{y_1}{k} a_1'(\theta)} g(t) dt &\approx \frac{y_1}{k} a_1'(\theta) g[a_1(\theta)], \\
\int_{a_2(\theta) - \frac{y_2}{k} a_2'(\theta)}^{a_2(\theta)} g(t) dt &\approx \frac{y_2}{k} a_2'(\theta) g[a_2(\theta)],
\end{aligned} \tag{3.84}$$

we have

$$\begin{aligned}
\left[\frac{\int_{a_1(\theta) + \frac{y_1}{k}}^{a_2(\theta) - \frac{y_2}{k}} g(t) dt}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right]^{k-2} &\approx \left[\frac{\int_{a_1(\theta) + \frac{y_1}{k} a_1'(\theta)}^{a_2(\theta) - \frac{y_2}{k} a_2'(\theta)} g(t) dt}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right]^{k-2} \\
&\approx \left[\frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt - \int_{a_1(\theta)}^{a_1(\theta) + \frac{y_1}{k} a_1'(\theta)} g(t) dt - \int_{a_2(\theta) - \frac{y_2}{k} a_2'(\theta)}^{a_2(\theta)} g(t) dt}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right]^{k-2} \\
&\approx \left\{ 1 - \frac{y_1 a_1'(\theta) g[a_1(\theta)] + y_2 a_2'(\theta) g[a_2(\theta)]}{k \int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right\}^{k-2} \\
&\rightarrow \exp \left\{ - \frac{y_1 a_1'(\theta) g[a_1(\theta)] + y_2 a_2'(\theta) g[a_2(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right\}.
\end{aligned}$$

It follows from (3.83) that, for fixed $\theta > \theta_0$, $y_1 > 0$, $y_2 > 0$, as $k \rightarrow \infty$,

$$\begin{aligned}
p(y_1, y_2 | \theta) &\rightarrow p^*(y_1, y_2 | \theta) \\
&\equiv \frac{a_1'(\theta) a_2'(\theta) g[a_1(\theta)] g[a_2(\theta)]}{\left[\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt \right]^2} \exp \left\{ - \frac{y_1 a_1'(\theta) g[a_1(\theta)] + y_2 a_2'(\theta) g[a_2(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right\}.
\end{aligned}$$

Therefore, as $k \rightarrow \infty$, the variables y_i 's have independent exponential distributions

with means

$$\lambda_i = \frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt}{a'_i(\theta) g[a_i(\theta)]}, \quad i = 1, 2.$$

Now, from (3.82), let $s = \theta + v/k$, we have

$$\begin{aligned} m_k(t_1, t_2) &= \int_{\theta - \frac{y_2}{k}}^{\theta + \frac{y_1}{k}} \frac{1}{\left[\int_{a_1(s)}^{a_2(s)} g(t) dt \right]^k} ds \\ &= \int_{-y_2}^{y_1} \frac{1}{k} \frac{dv}{\left[\int_{a_1(\theta + \frac{v}{k})}^{a_2(\theta + \frac{v}{k})} g(t) dt \right]^k}. \end{aligned}$$

Then, using the similar approximations as in (3.84), as $k \rightarrow \infty$, for fixed $y_1 > 0, y_2 > 0$,

$$\begin{aligned} &k \left[\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt \right]^k m_k(t_1, t_2) \\ &= \int_{-y_2}^{y_1} \left[\frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt}{\int_{a_1(\theta + \frac{v}{k})}^{a_2(\theta + \frac{v}{k})} g(t) dt} \right]^k dv \\ &\approx \int_{-y_2}^{y_1} \left[\frac{\int_{a_1(\theta) + \frac{v}{k} a'_1(\theta)}^{a_2(\theta) + \frac{v}{k} a'_2(\theta)} g(t) dt}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right]^{-k} dv \\ &\approx \int_{-y_2}^{y_1} \left[\frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt - \int_{a_1(\theta)}^{a_1(\theta) + \frac{v}{k} a'_1(\theta)} g(t) dt + \int_{a_2(\theta)}^{a_2(\theta) + \frac{v}{k} a'_2(\theta)} g(t) dt}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right]^{-k} dv \\ &\approx \int_{-y_2}^{y_1} \left\{ 1 - \frac{v a'_1(\theta) g[a_1(\theta)] - v a'_2(\theta) g[a_2(\theta)]}{k \int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \right\}^{-k} dv \\ &\rightarrow \int_{-y_2}^{y_1} \exp \left\{ - \frac{a'_2(\theta) g[a_2(\theta)] - a'_1(\theta) g[a_1(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} v \right\} dv \end{aligned}$$

$$\begin{aligned}
&= \frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt}{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]} \exp \left\{ \frac{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} y_2 \right\} \\
&\times \left\{ 1 - \exp \left[-\frac{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} (y_1 + y_2) \right] \right\}.
\end{aligned}$$

Then, for fixed $\theta > \theta_0$, as $k \rightarrow \infty$,

$$\begin{aligned}
&k^\beta \int_0^\infty \int_0^\infty p(y_1, y_2 | \theta) [\pi^*(\theta | y_1, y_2)]^{-\beta} dy_1 dy_2 \\
&= \int_0^\infty \int_0^\infty p(y_1, y_2 | \theta) \left\{ k \left[\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt \right]^k m_k(t_1, t_2) \right\}^\beta dy_1 dy_2 \\
&\rightarrow \int_0^\infty \int_0^\infty p^*(y_1, y_2 | \theta) \left[\frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt}{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]} \right]^\beta \\
&\times \exp \left\{ \frac{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} \beta y_2 \right\} \\
&\times \left\{ 1 - \exp \left[-\frac{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]}{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt} (y_1 + y_2) \right] \right\}^\beta dy_1 dy_2.
\end{aligned} \tag{3.85}$$

Let $v_i = y_i/\lambda_i, i = 1, 2$, then v_i 's are i.i.d. with the standard exponential distribution.

Define

$$b_i \equiv b_i(\theta) = \frac{a'_2(\theta)g[a_2(\theta)] - a'_1(\theta)g[a_1(\theta)]}{a'_i(\theta)g[a_i(\theta)]}, i = 1, 2,$$

(3.85) becomes

$$\begin{aligned}
k^\beta f_k^{-\beta}(\theta) &\equiv k^\beta \int_0^\infty \int_0^\infty p(y_1, y_2 | \theta) [\pi^*(\theta | y_1, y_2)]^{-\beta} dy_1 dy_2 \\
&\rightarrow \left[\frac{\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt}{a_2'(\theta)g[a_2(\theta)] - a_1'(\theta)g[a_1(\theta)]} \right]^\beta E \left[e^{\beta b_2 v_2} (1 - e^{-b_1 v_1 - b_2 v_2})^\beta \right].
\end{aligned} \tag{3.86}$$

Using the same arguments as in the example 3.3, we also have the following results as in (3.76) and (3.78),

$$E \left[e^{\beta b_2 v_2} (1 - e^{-b_1 v_1 - b_2 v_2})^\beta \right] \approx \frac{1}{1 - \beta b_2} + \frac{1}{b_1 b_2 (\beta + 2) \Gamma(-\beta)}. \tag{3.87}$$

Consequently, (3.86) and (3.87) imply that the reference prior of θ is

$$\begin{aligned}
\pi(\theta) &= \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)} = \lim_{k \rightarrow \infty} \frac{\left[k^\beta f_k^{-\beta}(\theta) \right]^{-1/\beta}}{\left[k^\beta f_k^{-\beta}(\theta_0) \right]^{-1/\beta}} \\
&\propto \frac{a_2'(\theta)g[a_2(\theta)] - a_1'(\theta)g[a_1(\theta)]}{c(\theta)} \left[\frac{1}{1 - \beta b_2} + \frac{1}{(\beta + 2)\Gamma(-\beta)b_1 b_2} \right]^{-\frac{1}{\beta}}.
\end{aligned} \tag{3.88}$$

We now consider an example for which the reference prior is not known and it appears to be extremely difficult to derive the reference prior analytically.

Example 3.6. Consider the following model,

$$p(x | \theta) = \begin{cases} \frac{2x}{\theta}, & \text{if } 0 < x \leq \theta \\ \frac{2(1-x)}{1-\theta}, & \text{if } \theta < x < 1 \end{cases}, 0 < \theta < 1.$$

This model is also known as the non-symmetric standard triangular distribution on $(0, 1)$. It does not possess a useful reduced sufficient statistic and the Jeffrey prior does not exist too. For this example, the analytical derivation of the reference prior does not seem to be feasible and we will use our algorithm to find the numerical reference prior. Again, it is worth to say a little bit about how we can generate a random variable from the distribution $p(x | \theta)$ in *Step 2* of the algorithm. We know that if $u \sim \text{uniform}(0, 1)$, then $x = F^{-1}(u) \sim p(x | \theta)$ as desired, where, for this example,

$$F^{-1}(u) = \begin{cases} \sqrt{\theta u}, & \text{if } 0 < u \leq \theta \\ 1 - \sqrt{(1-\theta)(1-u)}, & \text{if } \theta < u < 1 \end{cases}, 0 < \theta < 1.$$

The numerical computation was done for this example by the following R code and the figure 3.4 presents the numerical reference priors for $\beta = 0, 0.2, 0.5$, and -0.5 which are rescaled to have $\pi(1/2) = 2/\pi$; $m = 1000$ samples of $k = 500$ observations were used to compute each point $\{\theta, \pi(\theta)\}$.

```
#####  
### R code for the figure 3.4###  
#####  
  
# beta = 0  
  
k = 500  
  
m = 1000  
  
r = rep(NA,m)  
  
theta = seq(0.05,0.95,length = 201)  
  
ref = rep(NA,length(theta))  
  
for (l in 1:length(theta)){  
  
G = function(y){ if(y<theta[l]) {sqrt(theta[l]*y)}  
    else {1-sqrt((1-theta[l])*(1-y))}  
  
    }  
  
for (s in 1:m){  
  
x = rep(NA,k)  
  
a = 0  
  
A = 1  
  
B = 1  
  
for (i in 1:k){  
  
u = runif(1)  
  
x[i] = G(u)  
  
if(x[i]<theta[l])  
  
{a = a + 1
```

```
A = A*x[i]}
else {B = B*(1-x[i])}
}
t = seq(0.05,0.95,0.05)
n = length(t)
S = 0
for (j in 1:n){
  a.new = 0
  A.new = 1
  B.new = 1
  for (i in 1:k){
    if(x[i]<t[j])
      {a.new = a.new + 1
       A.new = A.new*x[i]}
    else {B.new = B.new*(1-x[i])}
  }
  S = S + A.new*B.new/(t[j]^a.new*(1-t[j])^(k-a.new))
}
S = S*0.05
r[s] = log(A*B/((theta[l]^a)*((1-theta[l])^(k-a))/S)
}
ref[l] = exp(sum(r)/m)
}
```

```
ref00 = ref

ref00 = ref00/ref00[101]*2/pi

plot(theta,ref00,type="p",pch=21,col="blue",xlab=expression(theta),
ylab=expression(pi(theta)))

the = seq(0.01,0.99,0.01)

lines(the,dbeta(the,0.5,0.5),lty=1,lwd=2)

legend(0.4,1.8,c(expression(beta==0),expression(beta==0.9),
"Beta(0.5,0.5)"),col=c("blue","black","black"),pch=21:22,lty=1)

#beta = 0.5

k = 500

m = 1000

r = rep(NA,m)

beta = 0.5

theta = seq(0.05,0.95,length = 201)

ref = rep(NA,length(theta))

for (l in 1:length(theta)){

G = function(y){ if(y<theta[l]) {sqrt(theta[l]*y)}
  else {1-sqrt((1-theta[l])*(1-y))}

  }

for (s in 1:m){

x = rep(NA,k)

a = 0
```

```
A = 1
B = 1
for (i in 1:k){
  u = runif(1)
  x[i] = G(u)
  if(x[i]<theta[l])
  {a = a + 1
   A = A*x[i]}
  else {B = B*(1-x[i])}
}
t = seq(0.05,0.95,0.05)
n = length(t)
S = 0
for (j in 1:n){
  a.new = 0
  A.new = 1
  B.new = 1
  for (i in 1:k){
    if(x[i]<t[j])
    {a.new = a.new + 1
     A.new = A.new*x[i]}
    else {B.new = B.new*(1-x[i])}
  }
}
```

```
S = S + A.new*B.new/(t[j]^a.new*(1-t[j])^(k-a.new))
}

S = S*0.05

r[s] = (A*B/((theta[l]^a)*((1-theta[l])^(k-a))))/S^(-beta)
}

ref[l] = (sum(r)/m)^(-1/beta)
}

ref0.5 = ref

ref0.5 = ref0.5/ref0.5[101]*2/pi

lines(theta,ref0.5,type="p",pch=22,col="red")

#beta = 0.2

k = 500

m = 1000

r = rep(NA,m)

beta = 0.2

theta = seq(0.05,0.95,length = 201)

ref = rep(NA,length(theta))

for (l in 1:length(theta)){

G = function(y){ if(y<theta[l]) {sqrt(theta[l]*y)}
  else {1-sqrt((1-theta[l])*(1-y))}
}

for (s in 1:m){
```

```
x = rep(NA,k)

a = 0

A = 1

B = 1

for (i in 1:k){

u = runif(1)

x[i] = G(u)

if(x[i]<theta[l])

{a = a + 1

  A = A*x[i]}

else {B = B*(1-x[i])}

}

t = seq(0.05,0.95,0.05)

n = length(t)

S = 0

for (j in 1:n){

a.new = 0

A.new = 1

B.new = 1

for (i in 1:k){

if(x[i]<t[j])

{a.new = a.new + 1

  A.new = A.new*x[i]}
```



```
else {B.new = B.new*(1-x[i])}
}
S = S + A.new*B.new/(t[j]^a.new*(1-t[j])^(k-a.new))
}
S = S*0.05
r[s] = (A*B/((theta[l]^a)*((1-theta[l])^(k-a))))/S)^(-beta)
}
ref[l] = (sum(r)/m)^(-1/beta)
}
ref0.2 = ref
ref0.2 = ref0.2/ref0.2[101]*2/pi
lines(theta,ref0.2,type="p",pch=23,col="green")

#beta = -0.5
k = 500
m = 1000
r = rep(NA,m)
beta = -0.5
theta = seq(0.05,0.95,length = 201)
ref = rep(NA,length(theta))
for (l in 1:length(theta)){
G = function(y){ if(y<theta[l]) {sqrt(theta[l]*y)}
else {1-sqrt((1-theta[l])*(1-y))}}
```

```
    }  
  for (s in 1:m){  
    x = rep(NA,k)  
    a = 0  
    A = 1  
    B = 1  
    for (i in 1:k){  
      u = runif(1)  
      x[i] = G(u)  
      if(x[i]<theta[l])  
      {a = a + 1  
        A = A*x[i]}  
      else {B = B*(1-x[i])}  
    }  
    t = seq(0.05,0.95,0.05)  
    n = length(t)  
    S = 0  
    for (j in 1:n){  
      a.new = 0  
      A.new = 1  
      B.new = 1  
      for (i in 1:k){  
        if(x[i]<t[j])
```

```

{a.new = a.new + 1
  A.new = A.new*x[i]}
else {B.new = B.new*(1-x[i])}
}
S = S + A.new*B.new/(t[j]^a.new*(1-t[j])^(k-a.new))
}
S = S*0.05
r[s] = (A*B/((theta[1]^a)*((1-theta[1])^(k-a))))/S)^(-beta)
}
ref[1] = (sum(r)/m)^(-1/beta)
}
ref.neg.0.5 = ref
ref.neg.0.5 = ref.neg.0.5/ref.neg.0.5[101]*2/pi
lines(theta,ref.neg.0.5,type="p",pch=8,col="purple")
arrows(.3,1.4,.08,1.2)
text(.35,1.41,"Beta(0.5,0.5)")
png("triangular.png", width = 800, height = 600)
par(mfrow=c(2,2))
the = seq(0.01,0.99,0.01)
plot(the,dbeta(the,0.5,0.5),main=expression(beta==0),type="l",
lwd=2,ylim=c(0.5,1.5),xlab=expression(theta),ylab=expression(pi(theta)))
lines(theta,ref00,type="p",pch=21,col="blue")
arrows(.3,1.4,.08,1.2,length = 0.15,angle=15)

```

```
text(.35,1.41,"Beta(0.5,0.5)",cex=0.8)

plot(the,dbeta(the,0.5,0.5),main=expression(beta==0.5),type="l",
lwd=2,ylim=c(0.5,1.5),xlab=expression(theta),ylab=expression(pi(theta)))

lines(theta,ref0.5,type="p",pch=22,col="red")

arrows(.3,1.4,.08,1.2,length = 0.15,angle=15)

text(.35,1.41,"Beta(0.5,0.5)",cex=0.8)

plot(the,dbeta(the,0.5,0.5),main=expression(beta==0.2),type="l",
lwd=2,ylim=c(0.5,1.5),xlab=expression(theta),ylab=expression(pi(theta)))

lines(theta,ref0.2,type="p",pch=23,col="green")

arrows(.3,1.4,.08,1.2,length = 0.15,angle=15)

text(.35,1.41,"Beta(0.5,0.5)",cex=0.8)

plot(the,dbeta(the,0.5,0.5),main=expression(beta== -0.5),type="l",
lwd=2,ylim=c(0.5,1.5),xlab=expression(theta),ylab=expression(pi(theta)))

lines(theta,ref.neg.0.5,type="p",pch=8,col="purple")

arrows(.3,1.4,.08,1.2,length = 0.15,angle=15)

text(.35,1.41,"Beta(0.5,0.5)",cex=0.8)

graphics.off()
```

Interestingly, when compared these numerical reference priors to the $Beta(1/2, 1/2)$ distribution, shown by the continuous line, we could see that they are somehow close.

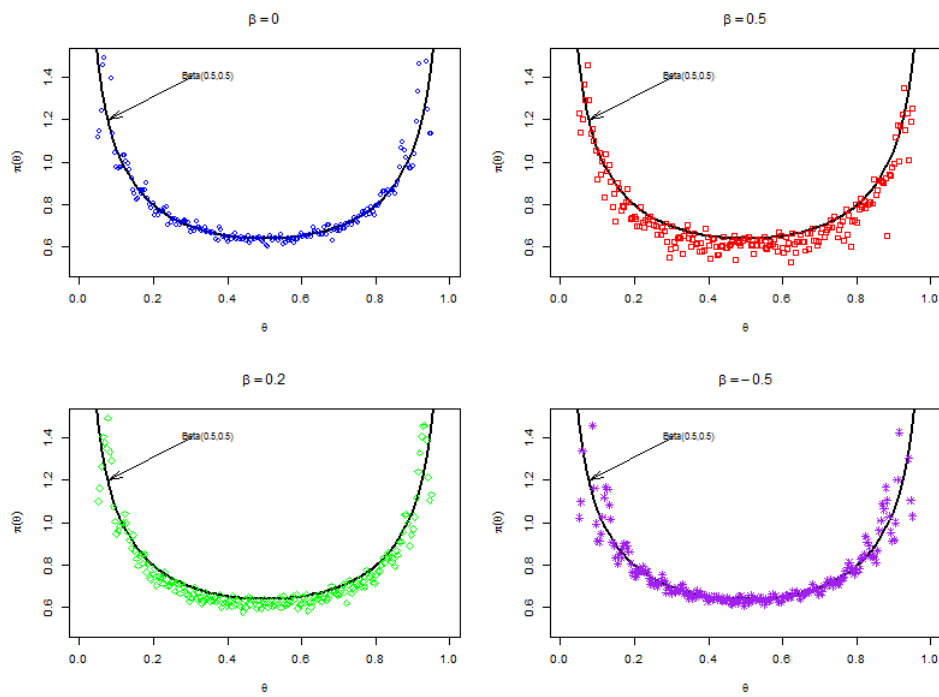


FIGURE 3.4: Numerical reference priors for the Triangular model when $\beta = 0, 0.2, 0.5, -0.5$

Chapter 4

Reference prior under the Chi-squared divergence

In the present Chapter, we consider the reference prior for the remained case when $\beta = -1$, the Chi-squared divergence measure. This could be viewed as the boundary case. For the regular models where there exist the Fisher information numbers, Ghosh, Mergel and Liu (2011) shown that the reference prior is no longer identical to the Jeffreys' prior but it is proportional to the fourth root of the Fisher information number. Our main objective here is to deal with the non-regular models where there do not exist the Fisher information numbers.

Let's look back to the equation (3.44) in the proof of our main Theorem 3.1,

$$\lim_{k \rightarrow \infty} \left[R_{0k}^\beta(\pi_0) - R_{0k}^\beta(p_0) \right] = 0,$$

this means that, as $k \rightarrow \infty$, $R_{0k}^\beta(p_0)$ does not depend on the prior $p(\theta)$ when $\beta = -1$. Therefore, our method there does not point out the reference priors in this case.

Our conjecture is that for this boundary case for the non-regular models we cannot find the general formula for the reference prior. For some models there exist reference priors, for some other models there do not. This conjecture is illustrated through a couple following examples.

4.1 Examples without reference priors when

$$\beta = -1$$

Example 4.1. Consider the following non-regular model,

$$f(x | \theta) = \begin{cases} \frac{g(x)}{G[a_2(\theta)]}, & \text{if } a_1(\theta) < x < a_2(\theta) \\ 0, & \text{else} \end{cases},$$

where $G(x) = \int_{a_1(\theta)}^x g(t)dt$ and $g(x)$ is a positive function .

Notice that $G[a_1(\theta)] = 0$ and $G[a_2(\theta)] = \int_{a_1(\theta)}^{a_2(\theta)} g(t)dt$.

The sufficient statistic of θ also is

$$\mathbf{t}_k \equiv (t_1, t_2) = (x_{(1)}, x_{(k)}) = (\min\{x_1, \dots, x_k\}, \max\{x_1, \dots, x_k\}).$$

From (3.81), we have

$$\begin{aligned}
p(t_1, t_2 | \theta) &= k(k-1)g(t_1)g(t_2) \frac{\left[\int_{t_1}^{t_2} g(t) dt \right]^{k-2}}{\left[\int_{a_1(\theta)}^{a_2(\theta)} g(t) dt \right]^k} \\
&= k(k-1)g(t_1)g(t_2) \frac{[G(t_2) - G(t_1)]^{k-2}}{G^k[a_2(\theta)]}, \quad a_1(\theta) < t_1 < t_2 < a_2(\theta)
\end{aligned} \tag{4.1}$$

When $\beta = -1$, from (2.29), the divergence between the prior $p_0(\theta)$ and the posterior $p_0(\theta | \mathbf{t}_k)$ is

$$\begin{aligned}
R_{0k}^{-1}(p_0) &= -\frac{1}{2} \left\{ 1 - \int_{\Theta_0} p_0(\theta) \int_{\mathcal{T}_k} \left[\frac{p_0(\theta)}{p_0(\theta | \mathbf{t}_k)} \right]^{-1} p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \right\} \\
&= -\frac{1}{2} + \frac{1}{2} \int_{\Theta_0} \int_{\mathcal{T}_k} p_0(\theta | \mathbf{t}_k) p(\mathbf{t}_k | \theta) d\mathbf{t}_k d\theta \\
&= -\frac{1}{2} + \frac{1}{2} \int_{\Theta_0} \int_{\mathcal{T}_k} \frac{p^2(\mathbf{t}_k | \theta) p_0(\theta)}{\int_{\Theta_0} p(\mathbf{t}_k | \theta) p_0(\theta) d\theta} d\mathbf{t}_k d\theta.
\end{aligned} \tag{4.2}$$

We now consider the inner integral, from (4.1),

$$\begin{aligned}
&\int_{\mathcal{T}_k} \frac{p^2(\mathbf{t}_k | \theta) p_0(\theta)}{\int_{\Theta_0} p(\mathbf{t}_k | \theta) p_0(\theta) d\theta} d\mathbf{t}_k \\
&= \int_{\mathcal{T}_k} \frac{\frac{k^2(k-1)^2}{G^{2k}[a_2(\theta)]} g^2(t_1) g^2(t_2) [G(t_2) - G(t_1)]^{2(k-2)} p_0(\theta)}{\int_{\Theta_0} \frac{k(k-1)}{G^k[a_2(\theta)]} g(t_1) g(t_2) [G(t_2) - G(t_1)]^{k-2} p_0(\theta) d\theta} d\mathbf{t}_k \\
&= \frac{\frac{k(k-1)}{G^{2k}[a_2(\theta)]} p_0(\theta)}{\int_{\Theta_0} \frac{p_0(\theta)}{G^k[a_2(\theta)]} d\theta} \int_{\mathcal{T}_k} g(t_1) g(t_2) [G(t_2) - G(t_1)]^{k-2} d\mathbf{t}_k.
\end{aligned} \tag{4.3}$$

Change variables

$$s_1 = G(t_1),$$

$$s_2 = G(t_2),$$

then

$$ds_1 = g(t_1)dt_1,$$

$$ds_2 = g(t_2)dt_2.$$

So,

$$\begin{aligned}
& \int_{\mathcal{T}_k} g(t_1)g(t_2)[G(t_2) - G(t_1)]^{k-2} d\mathbf{t}_k \\
&= \int_{a_1(\theta)}^{a_2(\theta)} \int_{t_1}^{a_2(\theta)} g(t_1)g(t_2)[G(t_2) - G(t_1)]^{k-2} dt_2 dt_1 \\
&= \int_0^{G[a_2(\theta)]} \int_{s_1}^{G[a_2(\theta)]} (s_2 - s_1)^{k-2} ds_2 ds_1 \\
&= \int_0^{G[a_2(\theta)]} \frac{(s_2 - s_1)^{k-1}}{k-1} \Big|_{s_1}^{G[a_2(\theta)]} ds_1 \\
&= \int_0^{G[a_2(\theta)]} \frac{\{G[a_2(\theta)] - s_1\}^{k-1}}{k-1} ds_1 \\
&= -\frac{\{G[a_2(\theta)] - s_1\}^k}{k(k-1)} \Big|_0^{G[a_2(\theta)]} \\
&= \frac{G^k[a_2(\theta)]}{k(k-1)}. \tag{4.4}
\end{aligned}$$

The combination of (4.2), (4.3), and (4.4) gives us

$$\begin{aligned}
R_{0k}^{-1}(p_0) &= -\frac{1}{2} + \frac{1}{2} \int_{\Theta_0} \frac{\frac{k(k-1)}{G^{2k}[a_2(\theta)]} p_0(\theta) G^k[a_2(\theta)]}{\int_{\Theta_0} \frac{p_0(\theta)}{G^k[a_2(\theta)]} d\theta} \frac{1}{k(k-1)} d\theta \\
&= -\frac{1}{2} + \frac{1}{2} \frac{\int_{\Theta_0} \frac{p_0(\theta)}{G^k[a_2(\theta)]} d\theta}{\int_{\Theta_0} \frac{p_0(\theta)}{G^k[a_2(\theta)]} d\theta} \\
&= 0.
\end{aligned} \tag{4.5}$$

This means that, when $\beta = -1$, the divergence $R_{0k}^{-1}(p_0)$ between the prior p_0 and its corresponding posterior is zero for all sample size k and for all prior p . Therefore, we do not have the reference prior for the model given in this example.

4.2 Examples with reference priors when $\beta = -1$

Example 4.2. Consider the following non-regular model,

$$p(x | \theta) = \begin{cases} \frac{\theta}{1 - e^{-\theta^2}} e^{-\theta x}, & \text{if } 0 < x < \theta \\ 0, & \text{else} \end{cases},$$

where θ is unknown parameter in $\Theta = (0, \infty)$.

This example is not separable like the previous one. So we cannot use the analogous arguments as in the previous example to handle this case.

Let $p(\cdot)$ be an any prior of θ on the parameter space Θ . Then an similar expression to (4.2) yields

$$\begin{aligned} R_k^{-1}(p) &= -\frac{1}{2} + \frac{1}{2} \int_{\Theta} \int_{\mathcal{X}} \frac{p^2(\mathbf{x}_k | \theta)p(\theta)}{\int_{\Theta} p(\mathbf{x}_k | \theta)p(\theta)d\theta} d\mathbf{x}_k d\theta \\ &= -\frac{1}{2} + \frac{1}{2} \int_{\Theta} E_{\theta}[p(\theta | \mathbf{x}_k)] d\theta, \end{aligned} \quad (4.6)$$

where

$$E_{\theta}[p(\theta | \mathbf{x}_k)] = \int_{\mathcal{X}} \frac{p^2(\mathbf{x}_k | \theta)p(\theta)}{\int_{\Theta} p(\mathbf{x}_k | \theta)p(\theta)d\theta} d\mathbf{x}_k. \quad (4.7)$$

Therefore, in order to evaluate the divergence $R_k^{-1}(p)$ we first need to evaluate the posterior $p(\theta | \mathbf{x}_k)$.

Now we have

$$\begin{aligned} p(\theta | \mathbf{x}_k) &= \frac{p(\theta)p(\mathbf{x}_k | \theta)}{C_k} \\ &= p(\theta) \frac{\theta^k}{(1 - e^{-\theta^2})^k} e^{-\theta \sum_{i=1}^k x_i} I(x_{(k)} < \theta) / C_k, \end{aligned} \quad (4.8)$$

where

$$C_k = \int_{x_{(k)}}^{\infty} p(\theta) \frac{\theta^k}{(1 - e^{-\theta^2})^k} e^{-\theta \sum_{i=1}^k x_i} d\theta.$$

Change variable to

$$\eta = k(\theta - x_{(k)}).$$

Then $\eta > 0$, $\theta = x_{(k)} + \eta/k$, $d\theta = d\eta/k$, and from (4.8) we have

$$(\eta \mid x_{(k)}) \sim p\left(x_{(k)} + \frac{\eta}{k}\right) \left[\frac{x_{(k)} + \frac{\eta}{k}}{1 - e^{-(x_{(k)} + \frac{\eta}{k})^2}} \right]^k e^{-(x_{(k)} + \frac{\eta}{k}) \sum_{i=1}^k x_i} I(\eta > 0) / D_k, \quad (4.9)$$

where

$$D_k = \int_0^\infty p\left(x_{(k)} + \frac{\eta}{k}\right) \left[\frac{x_{(k)} + \frac{\eta}{k}}{1 - e^{-(x_{(k)} + \frac{\eta}{k})^2}} \right]^k e^{-(x_{(k)} + \frac{\eta}{k}) \sum_{i=1}^k x_i} d\eta. \quad (4.10)$$

Using the following Taylor expansion

$$e^{-x^2} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \dots,$$

we get

$$\frac{1 - e^{-x^2}}{x} = x \left(1 - \frac{x^2}{2!} + \frac{x^4}{3!} - \dots \right).$$

So,

$$\begin{aligned} \left[\frac{x_{(k)} + \frac{\eta}{k}}{1 - e^{-(x_{(k)} + \frac{\eta}{k})^2}} \right]^k &= \left[\frac{1 - e^{-(x_{(k)} + \frac{\eta}{k})^2}}{x_{(k)} + \frac{\eta}{k}} \right]^{-k} \\ &= \left(x_{(k)} + \frac{\eta}{k} \right)^{-k} \left[1 - \frac{1}{2!} \left(x_{(k)} + \frac{\eta}{k} \right)^2 + \frac{1}{3!} \left(x_{(k)} + \frac{\eta}{k} \right)^4 - \dots \right]^{-k} \\ &= x_{(k)}^{-k} \left(1 + \frac{\eta}{x_{(k)} k} \right)^{-k} \left[1 - \frac{1}{2!} \left(x_{(k)} + \frac{\eta}{k} \right)^2 + \frac{1}{3!} \left(x_{(k)} + \frac{\eta}{k} \right)^4 - \dots \right]^{-k}. \end{aligned}$$

As $k \rightarrow \infty$,

$$\begin{aligned} \left(1 + \frac{\eta}{x_{(k)}k}\right)^{-k} &\approx e^{-\frac{\eta}{x_{(k)}}}, \\ \left[1 - \frac{1}{2!} \left(x_{(k)} + \frac{\eta}{k}\right)^2 + \frac{1}{3!} \left(x_{(k)} + \frac{\eta}{k}\right)^4 - \dots\right]^{-k} &\approx \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k}. \end{aligned}$$

Therefore,

$$\left[\frac{x_{(k)} + \frac{\eta}{k}}{1 - e^{-(x_{(k)} + \frac{\eta}{k})^2}}\right]^k \approx x_{(k)}^{-k} e^{-\frac{\eta}{x_{(k)}}} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k}. \quad (4.11)$$

Also,

$$p\left(x_{(k)} + \frac{\eta}{k}\right) \approx p(x_{(k)}) + \frac{\eta}{k}p'(x_{(k)}). \quad (4.12)$$

(4.9), (4.11), and (4.12) yield

$$\begin{aligned} &p\left(x_{(k)} + \frac{\eta}{k}\right) \left[\frac{x_{(k)} + \frac{\eta}{k}}{1 - e^{-(x_{(k)} + \frac{\eta}{k})^2}}\right]^k e^{-(x_{(k)} + \frac{\eta}{k})\sum_{i=1}^k x_i} I(\eta > 0) \\ &\approx p(x_{(k)})x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} e^{-\frac{\eta}{x_{(k)}}} I(\eta > 0) \\ &\quad + p'(x_{(k)})x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} \frac{\eta}{k} e^{-\frac{\eta}{x_{(k)}}} I(\eta > 0). \end{aligned} \quad (4.13)$$

From (4.10) and (4.13),

$$\begin{aligned}
D_k &\approx p(x_{(k)})x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} \int_0^\infty e^{-\frac{\eta}{x_{(k)}}} d\eta \\
&\quad + p'(x_{(k)})x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} \frac{1}{k} \int_0^\infty \eta e^{-\frac{\eta}{x_{(k)}}} d\eta \\
&\approx p(x_{(k)})x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} \cdot x_{(k)} \\
&\quad + p'(x_{(k)})x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} \frac{1}{k} \cdot x_{(k)}^2 \\
&\approx x_{(k)}^{-k} \left(1 - \frac{1}{2!}x_{(k)}^2 + \frac{1}{3!}x_{(k)}^4 - \dots\right)^{-k} e^{-x_{(k)}\sum_{i=1}^k x_i} \left[p(x_{(k)})x_{(k)} + \frac{1}{k}p'(x_{(k)})x_{(k)}^2 \right].
\end{aligned} \tag{4.14}$$

Therefore, the combination of (4.9), (4.13), and (4.14) gives

$$\begin{aligned}
(\eta \mid \mathbf{x}_k) &\approx \frac{p(x_{(k)})e^{-\frac{\eta}{x_{(k)}}} + p'(x_{(k)})\frac{\eta}{k}e^{-\frac{\eta}{x_{(k)}}}}{p(x_{(k)})x_{(k)} + \frac{1}{k}p'(x_{(k)})x_{(k)}^2} I(\eta > 0) \\
&\approx \frac{p(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \left(\frac{1}{x_{(k)}} e^{-\frac{\eta}{x_{(k)}}} \right) + \frac{\frac{1}{k}x_{(k)}p'(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \left(\frac{\eta e^{-\frac{\eta}{x_{(k)}}}}{x_{(k)}^2} \right) \\
&\approx \frac{p(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \cdot \text{Exp}(x_{(k)}) + \frac{\frac{1}{k}x_{(k)}p'(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \cdot \text{gamma}(2, x_{(k)}).
\end{aligned} \tag{4.15}$$

This means that, as $k \rightarrow \infty$, the posterior $(\eta \mid \mathbf{x}_k)$ is approximately a mixed distribution of the two distributions $\text{Exp}(x_{(k)})$ and $\text{gamma}(2, x_{(k)})$.

Transform back to the posterior of θ , $(\theta \mid \mathbf{x}_k)$, we have

$$\begin{aligned}
(\theta \mid \mathbf{x}_k) &\approx \frac{kp(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \frac{1}{x_{(k)}} e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})} \\
&+ \frac{x_{(k)}p'(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \frac{k}{x_{(k)}^2} (\theta - x_{(k)}) e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})}, \quad \theta > x_{(k)} \\
&\approx \frac{kp(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \frac{1}{x_{(k)}} e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})} \\
&+ \frac{x_{(k)}p'(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \left[\frac{k}{x_{(k)}^2} \theta e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})} - \frac{k}{x_{(k)}} e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})} \right], \quad \theta > x_{(k)} \\
&\approx \frac{p(x_{(k)}) - x_{(k)}p'(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \cdot \frac{k}{x_{(k)}} e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})} \\
&+ \frac{\frac{k+1}{k}x_{(k)}p'(x_{(k)})}{p(x_{(k)}) + \frac{1}{k}p'(x_{(k)})x_{(k)}} \cdot \frac{k^2}{(k+1)x_{(k)}^2} \theta e^{-\frac{k}{x_{(k)}}(\theta-x_{(k)})}, \quad \theta > x_{(k)} \tag{4.16}
\end{aligned}$$

We now use the shrinkage argument due to Ghosh (1994) which is described in details in Datta and Mukerjee (2004) to get $E_\theta[p(\theta \mid \mathbf{x}_k)]$. The shrinkage argument consists of three steps as follows.

Step 1. Consider the following prior $\bar{\pi}(\theta)$,

$$\bar{\pi}(\theta) = \begin{cases} h(\theta), & \text{if } \theta \in \Theta_0 = \left[\theta_0 - \frac{1}{m}, \theta_0 + \frac{1}{m} \right], \\ \frac{1}{m}, & \text{if } \theta \in \Theta \setminus \Theta_0 \end{cases},$$

where $h(\theta) > 0$ is any continuous function and θ_0 is a fixed interior point of the parameter space $\Theta = (0, \infty)$.

In this step, we want to obtain the following quantity

$$E^{\bar{\pi}}[p(\theta \mid \mathbf{x}_k)] = \int_{x_{(k)}}^{\infty} p(\theta \mid \mathbf{x}_k) \bar{\pi}(\theta \mid \mathbf{x}_k) d\theta. \tag{4.17}$$

An expression similar to (4.16) gives

$$\begin{aligned} (\bar{\pi} \mid \mathbf{x}_k) &\approx \frac{\bar{\pi}(x_{(k)}) - x_{(k)}\bar{\pi}'(x_{(k)})}{\bar{\pi}(x_{(k)}) + \frac{1}{k}\bar{\pi}'(x_{(k)})x_{(k)}} \cdot \frac{k}{x_{(k)}} e^{-\frac{k}{x_{(k)}}(\theta - x_{(k)})} \\ &+ \frac{\frac{k+1}{k}x_{(k)}\bar{\pi}'(x_{(k)})}{\bar{\pi}(x_{(k)}) + \frac{1}{k}\bar{\pi}'(x_{(k)})x_{(k)}} \cdot \frac{k^2}{(k+1)x_{(k)}^2} \theta e^{-\frac{k}{x_{(k)}}(\theta - x_{(k)})}, \quad \theta > x_{(k)}. \end{aligned}$$

So if let $y_k = x_{(k)}$, then

$$\begin{aligned} E^{\bar{\pi}}[p(\theta \mid \mathbf{x}_k)] &= \int_{y_k}^{\infty} \left[\frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{k}{y_k} e^{-\frac{k}{y_k}(\theta - y_k)} \right. \\ &\quad \left. + \frac{\frac{k+1}{k} y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{k^2}{(k+1)y_k^2} \theta e^{-\frac{k}{y_k}(\theta - y_k)} \right] \\ &\quad \times \left[\frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k}{y_k} e^{-\frac{k}{y_k}(\theta - y_k)} \right. \\ &\quad \left. + \frac{\frac{k+1}{k} y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^2}{(k+1)y_k^2} \theta e^{-\frac{k}{y_k}(\theta - y_k)} \right] d\theta \\ &= J_1 + J_2 + J_3, \end{aligned} \tag{4.18}$$

where

$$\begin{aligned} J_1 &= \frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^2}{y_k^2} \cdot \int_{y_k}^{\infty} e^{-\frac{2k}{y_k}(\theta - y_k)} d\theta \\ &= \frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^2}{y_k^2} \cdot \frac{y_k}{2k} \\ &= \frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k}{2y_k}, \end{aligned} \tag{4.19}$$

$$\begin{aligned}
J_2 &= \frac{\frac{k+1}{k} y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\frac{k+1}{k} y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^4}{(k+1)^2 y_k^4} \cdot \int_{y_k}^{\infty} \theta^2 e^{-\frac{2k}{y_k}(\theta - y_k)} d\theta \\
&= \frac{\frac{k+1}{k} y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\frac{k+1}{k} y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^4}{(k+1)^2 y_k^4} \cdot \frac{2k^2 + 2k + 1}{4k^3} y_k^3 \\
&= \frac{\frac{k+1}{k} y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\frac{k+1}{k} y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k(2k^2 + 2k + 1)}{4(k+1)^2 y_k}, \tag{4.20}
\end{aligned}$$

$$\begin{aligned}
J_3 &= \left[\frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{k}{y_k} \cdot \frac{\frac{k+1}{k} y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^2}{(k+1) y_k^2} \right. \\
&\quad \left. + \frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k}{y_k} \cdot \frac{\frac{k+1}{k} y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{k^2}{(k+1) y_k^2} \right] \int_{y_k}^{\infty} \theta e^{-\frac{2k}{y_k}(\theta - y_k)} d\theta \\
&= \left[\frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{k}{y_k} \cdot \frac{\frac{k+1}{k} y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k^2}{(k+1) y_k^2} \right. \\
&\quad \left. + \frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{k}{y_k} \cdot \frac{\frac{k+1}{k} y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{k^2}{(k+1) y_k^2} \right] \cdot \frac{y_k^2(2k+1)}{4k^2} \\
&= \frac{p(y_k) - y_k p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{\bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{2k+1}{4} \\
&\quad + \frac{\bar{\pi}(y_k) - y_k \bar{\pi}'(y_k)}{\bar{\pi}(y_k) + \frac{1}{k} \bar{\pi}'(y_k) y_k} \cdot \frac{p'(y_k)}{p(y_k) + \frac{1}{k} p'(y_k) y_k} \cdot \frac{2k+1}{4}. \tag{4.21}
\end{aligned}$$

Step 2. Find $E_{\theta} E^{\bar{\pi}}[p(\theta | \mathbf{x}_k)]$.

From *Step 1*, $E^{\bar{\pi}}[p(\theta | \mathbf{x}_k)]$ is a function of y_k , say

$$E^{\bar{\pi}}[p(\theta | \mathbf{x}_k)] = \lambda(y_k). \tag{4.22}$$

It is easy to see the c.d.f. of x_i is

$$F(x) = \frac{1 - e^{-\theta x}}{1 - e^{-\theta^2}}, \quad 0 < x < \theta. \tag{4.23}$$

So the p.d.f. of y_k is

$$\begin{aligned} g_k(y_k) &= k[F(y_k)]^{k-1}f(y_k) \\ &= k\left(\frac{1-e^{-\theta y_k}}{1-e^{-\theta^2}}\right)^{k-1}\frac{\theta}{1-e^{-\theta^2}}e^{-\theta y_k} \\ &= \frac{k\theta}{(1-e^{-\theta^2})^k}(1-e^{-\theta y_k})^{k-1}e^{-\theta y_k}, \quad 0 < y_k < \theta. \end{aligned}$$

We now calculate $E_\theta(y_k)$,

$$E_\theta(y_k) = \int_0^\theta \frac{k\theta}{(1-e^{-\theta^2})^k}(1-e^{-\theta y_k})^{k-1}e^{-\theta y_k}y_k dy_k.$$

Change variable to

$$t = 1 - e^{-\theta y_k},$$

with $y_k = -\log(1-t)/\theta$ and $dt = \theta e^{-\theta y_k} dy_k$.

Then,

$$\begin{aligned} E_\theta(y_k) &= \int_0^{1-e^{-\theta^2}} -\frac{k\theta}{(1-e^{-\theta^2})^k} \frac{t^{k-1}}{\theta^2} \log(1-t) dt \\ &= -\frac{k}{\theta(1-e^{-\theta^2})^k} \int_0^{1-e^{-\theta^2}} t^{k-1} \log(1-t) dt. \end{aligned}$$

Integrate by parts,

$$\begin{aligned} u &= \log(1-t) \Rightarrow du = -\frac{1}{1-t} dt, \\ dv &= t^{k-1} dt \Rightarrow v = \frac{t^k}{k}, \end{aligned}$$

we have

$$\begin{aligned} E_\theta(y_k) &= -\frac{k}{\theta(1-e^{-\theta^2})^k} \cdot \left[\log(1-t) \frac{t^k}{k} \Big|_0^{1-e^{-\theta^2}} + \frac{1}{k} \int_0^{1-e^{-\theta^2}} \frac{t^k}{1-t} dt \right] \\ &= -\frac{k}{\theta(1-e^{-\theta^2})^k} \cdot \left[-\frac{\theta^2}{k} (1-e^{-\theta^2})^k + \frac{1}{k} \int_0^{1-e^{-\theta^2}} \frac{t^k}{1-t} dt \right] \\ &= \theta - \frac{1}{\theta(1-e^{-\theta^2})^k} \cdot \int_0^{1-e^{-\theta^2}} \frac{t^k}{1-t} dt. \end{aligned} \quad (4.24)$$

By using the series $1/(1-x) = \sum_{i=0}^{\infty} x^i$ and $\log(1-x) = -\sum_{j=1}^{\infty} x^j/j$, consider the integral

$$\begin{aligned} \int_0^{1-e^{-\theta^2}} \frac{t^k}{1-t} dt &= \int_0^{1-e^{-\theta^2}} t^k \cdot \sum_{i=0}^{\infty} t^i dt \\ &= \sum_{i=0}^{\infty} \int_0^{1-e^{-\theta^2}} t^{k+i} dt \\ &= \sum_{i=0}^{\infty} \frac{(1-e^{-\theta^2})^{k+i+1}}{k+i+1} \\ &= \sum_{j=1}^{\infty} \frac{(1-e^{-\theta^2})^j}{j} - \sum_{j=1}^k \frac{(1-e^{-\theta^2})^j}{j} \\ &= \theta^2 - \sum_{j=1}^k \frac{(1-e^{-\theta^2})^j}{j}. \end{aligned} \quad (4.25)$$

When $k \rightarrow \infty$,

$$\sum_{j=1}^k \frac{(1 - e^{-\theta^2})^j}{j} = \sum_{j=1}^{\infty} \frac{(1 - e^{-\theta^2})^j}{j} = \theta^2.$$

Therefore, from (4.24) and (4.25),

$$E_{\theta}(y_k) = \theta \text{ as } k \rightarrow \infty. \quad (4.26)$$

Now using the Taylor expansion for the function $\lambda(y_k)$ in (4.22) about θ , we have

$$\lambda(y_k) = \lambda(\theta) + (y_k - \theta)\lambda'(\theta).$$

Then, from (4.26), as $k \rightarrow \infty$,

$$\begin{aligned} E_{\theta}E^{\bar{\pi}}[p(\theta | \mathbf{x}_k)] &= E_{\theta}[\lambda(y_k)] \\ &= \lambda(\theta) + [E_{\theta}(y_k) - \theta]\lambda'(\theta) \\ &= \lambda(\theta), \end{aligned} \quad (4.27)$$

where, from (4.19), (4.20), and (4.21),

$$\begin{aligned}
\lambda(\theta) = & \frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \cdot \frac{\bar{\pi}(\theta) - \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta) + \frac{1}{k} \bar{\pi}'(\theta) \theta} \cdot \frac{k}{2\theta} \\
& + \frac{\frac{k+1}{k} \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \cdot \frac{\frac{k+1}{k} \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta) + \frac{1}{k} \bar{\pi}'(\theta) \theta} \cdot \frac{k(2k^2 + 2k + 1)}{4(k+1)^2 \theta} \\
& + \frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \cdot \frac{\bar{\pi}'(\theta)}{\bar{\pi}(\theta) + \frac{1}{k} \bar{\pi}'(\theta) \theta} \cdot \frac{2k+1}{4} \\
& + \frac{\bar{\pi}(\theta) - \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta) + \frac{1}{k} \bar{\pi}'(\theta) \theta} \cdot \frac{p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \cdot \frac{2k+1}{4}, \quad \theta \in \Theta. \quad (4.28)
\end{aligned}$$

Step 3. Integrate $\lambda(\theta)$ with respect to $\bar{\pi}(\theta)$.

As $k \rightarrow \infty$, from (4.28), we could rewrite $\lambda(\theta)$ as follows,

$$\begin{aligned}
\lambda(\theta) = & \frac{\bar{\pi}(\theta) - \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta) + \frac{1}{k} \bar{\pi}'(\theta) \theta} \left[\frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \frac{k}{2\theta} + \frac{p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \frac{2k+1}{4} \right] \\
& + \frac{\frac{1}{k} \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta) + \frac{1}{k} \bar{\pi}'(\theta) \theta} \left[\frac{2k^2 + 2k + 1}{4} \frac{p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} + \frac{k(2k+1)}{4\theta} \frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \right] \\
\approx & \frac{\bar{\pi}(\theta) - \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta)} \cdot \frac{\frac{k}{2\theta} p(\theta) + \frac{1}{4} p'(\theta)}{p(\theta) + \frac{\theta}{k} p'(\theta)} \\
& + \frac{\frac{1}{k} \theta \bar{\pi}'(\theta)}{\bar{\pi}(\theta)} \left[\frac{2k^2 + 2k + 1}{4} \frac{p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} + \frac{k(2k+1)}{4\theta} \frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta) \theta} \right] \quad (4.29)
\end{aligned}$$

Therefore,

$$\int_0^\infty \lambda(\theta) \bar{\pi}(\theta) d\theta = K_1 + K_2, \quad (4.30)$$

where

$$\begin{aligned}
K_1 &= \int_0^\infty [\bar{\pi}(\theta) - \theta\bar{\pi}'(\theta)] \cdot \frac{\frac{k}{2\theta}p(\theta) + \frac{1}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} d\theta, \\
K_2 &= \int_0^\infty \bar{\pi}'(\theta) \left[\frac{2k^2 + 2k + 1}{4k} \frac{\theta p'(\theta)}{p(\theta) + \frac{1}{k}p'(\theta)\theta} + \frac{2k + 1}{4} \frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k}p'(\theta)\theta} \right] d\theta.
\end{aligned}$$

For K_1 ,

$$K_1 = \int_0^\infty \frac{\frac{k}{2\theta}p(\theta) + \frac{1}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \bar{\pi}(\theta) d\theta - \int_0^\infty \frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \bar{\pi}'(\theta) d\theta.$$

Upon integration by parts, we have

$$\int_0^\infty \frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \bar{\pi}'(\theta) d\theta = \frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \bar{\pi}(\theta) \Big|_0^\infty - \int_0^\infty \frac{d}{d\theta} \left[\frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \right] \bar{\pi}(\theta) d\theta.$$

As m is large enough, $\bar{\pi}(0) = \bar{\pi}(\infty) = 0$, then the first term in the previous equation is zero. So,

$$\int_0^\infty \frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \bar{\pi}'(\theta) d\theta = - \int_0^\infty \frac{d}{d\theta} \left[\frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \right] \bar{\pi}(\theta) d\theta.$$

This leads to

$$K_1 = \int_0^\infty \frac{\frac{k}{2\theta}p(\theta) + \frac{1}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \bar{\pi}(\theta) d\theta + \int_0^\infty \frac{d}{d\theta} \left[\frac{\frac{k}{2}p(\theta) + \frac{\theta}{4}p'(\theta)}{p(\theta) + \frac{\theta}{k}p'(\theta)} \right] \bar{\pi}(\theta) d\theta. \quad (4.31)$$

For K_2 , similarly, upon integration by parts, we also have

$$\begin{aligned}
K_2 &= - \int_0^\infty \frac{d}{d\theta} \left[\frac{2k^2 + 2k + 1}{4k} \frac{\theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta)\theta} + \frac{2k + 1}{4} \frac{p(\theta) - \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta)\theta} \right] \bar{\pi}(\theta) d\theta \\
&= - \int_0^\infty \frac{d}{d\theta} \left[\frac{\frac{2k+1}{4} p(\theta) + \frac{k+1}{4k} \theta p'(\theta)}{p(\theta) + \frac{1}{k} p'(\theta)\theta} \right] \bar{\pi}(\theta) d\theta \\
&\approx - \int_0^\infty \frac{d}{d\theta} \left[\frac{\frac{k}{2} p(\theta) + \frac{\theta}{4} p'(\theta)}{p(\theta) + \frac{\theta}{k} p'(\theta)} \right] \bar{\pi}(\theta) d\theta.
\end{aligned} \tag{4.32}$$

The combination of (4.30), (4.31), and (4.32) yields

$$\int_0^\infty \lambda(\theta) \bar{\pi}(\theta) d\theta = \int_0^\infty \frac{\frac{k}{2\theta} p(\theta) + \frac{1}{4} p'(\theta)}{p(\theta) + \frac{\theta}{k} p'(\theta)} \bar{\pi}(\theta) d\theta. \tag{4.33}$$

As $m \rightarrow \infty$, $\bar{\pi}(\theta)$ converges weakly to the degenerate prior at the true θ_0 .

Therefore, from (4.33), we get

$$E_\theta[p(\theta | \mathbf{x}_k)] = \frac{\frac{k}{2\theta} p(\theta) + \frac{1}{4} p'(\theta)}{p(\theta) + \frac{\theta}{k} p'(\theta)}. \tag{4.34}$$

Now, according to (4.6), in order to maximize the divergence $R_k^{-1}(p)$ we need to maximize the following integral

$$\begin{aligned}
\int_{\Theta} E_\theta[p(\theta | \mathbf{x}_k)] d\theta &= \int_{\Theta} \frac{\frac{k}{2\theta} p(\theta) + \frac{1}{4} p'(\theta)}{p(\theta) + \frac{\theta}{k} p'(\theta)} d\theta \\
&= \int_{\Theta} \left[\frac{k}{2\theta} - \frac{1}{4} \cdot \frac{p'(\theta)}{p(\theta) + \frac{\theta}{k} p'(\theta)} \right] d\theta.
\end{aligned}$$

Writing the integrand as a function of $p(\theta)$ and $p'(\theta)$, $s(p(\theta), p'(\theta))$, the solution is found by solving the Euler-Lagrange equation (Hewitt and Stromberg, 1969),

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

or equivalently,

$$\frac{1}{4} \cdot \frac{p'(\theta)}{[p(\theta) + \frac{\theta}{k}p'(\theta)]^2} + \frac{1}{4} \cdot \frac{d}{d\theta} \left\{ \frac{p(\theta)}{[p(\theta) + \frac{\theta}{k}p'(\theta)]^2} \right\} = 0,$$

or,

$$\frac{p'(\theta)}{[p(\theta) + \frac{\theta}{k}p'(\theta)]^2} + \frac{p'(\theta)[p(\theta) + \frac{\theta}{k}p'(\theta)]^2 - 2p(\theta)[p(\theta) + \frac{\theta}{k}p'(\theta)][p'(\theta) + \frac{p'(\theta)}{k} + \frac{\theta}{k}p''(\theta)]}{[p(\theta) + \frac{\theta}{k}p'(\theta)]^4} = 0.$$

This leads to

$$p'(\theta) \left[p(\theta) + \frac{\theta}{k}p'(\theta) \right] - p(\theta) \left[\frac{k+1}{k}p'(\theta) + \frac{\theta}{k}p''(\theta) \right] = 0,$$

or,

$$p(\theta)p'(\theta) + \theta p(\theta)p''(\theta) - \theta p'^2(\theta) = 0.$$

Divided both sides by $p'^2(\theta)$, we get

$$\frac{p(\theta)}{p'(\theta)} - \theta \cdot \frac{p'^2(\theta) - p(\theta)p''(\theta)}{p'^2(\theta)} = 0,$$

or,

$$\frac{p(\theta)}{p'(\theta)} - \theta \cdot \left[\frac{p(\theta)}{p'(\theta)} \right]' = 0.$$

Let $y(\theta) = p(\theta)/p'(\theta)$. Then,

$$\begin{aligned} y'(\theta) - \frac{1}{\theta}y(\theta) &= 0 \\ \Leftrightarrow \frac{1}{\theta}y'(\theta) - \frac{1}{\theta^2}y(\theta) &= 0 \\ \Leftrightarrow \left[\frac{1}{\theta}y(\theta) \right]' &= 0. \end{aligned}$$

So,

$$\begin{aligned} \frac{1}{\theta}y(\theta) &= \frac{1}{c} \\ \Leftrightarrow \frac{p'(\theta)}{p(\theta)} &= \frac{c}{\theta} \\ \Leftrightarrow [\log p(\theta)]' &= \frac{c}{\theta}, \quad c \in \mathbb{R}. \end{aligned}$$

That is

$$\log p(\theta) = c \log(c'\theta) \Leftrightarrow p(\theta) = c'^c \cdot \theta^c, \quad c', c \in \mathbb{R}.$$

Accordingly, the reference prior in the given example is

$$\pi(\theta) \propto \theta^c, \quad c \in \mathbb{R}.$$

Now, one natural question arising is that how to choose c , which values of c are better? In order to answer this question, we need to calculate the corresponding Bayes risk and choose c such that it minimizes the Bayes risk. First, we find the Bayes estimator, $\delta^\pi(\mathbf{x}_k)$, under the squared error loss as follows. From (4.15), we have

$$\begin{aligned}
 E(\eta \mid \mathbf{x}_k) &= \frac{\pi(x_{(k)})}{\pi(x_{(k)}) + \frac{1}{k}\pi'(x_{(k)})x_{(k)}} \cdot (x_{(k)}) + \frac{\frac{1}{k}x_{(k)}\pi'(x_{(k)})}{\pi(x_{(k)}) + \frac{1}{k}\pi'(x_{(k)})x_{(k)}} \cdot (2x_{(k)}) \\
 &= \frac{x_{(k)}^c}{x_{(k)}^c + \frac{1}{k}cx_{(k)}^{c-1}x_{(k)}} \cdot (x_{(k)}) + \frac{\frac{1}{k}x_{(k)}cx_{(k)}^{c-1}}{x_{(k)}^c + \frac{1}{k}cx_{(k)}^{c-1}x_{(k)}} \cdot (2x_{(k)}) \\
 &= \frac{1 + \frac{2c}{k}}{1 + \frac{c}{k}}x_{(k)}. \tag{4.35}
 \end{aligned}$$

So, the Bayes estimator $\delta^\pi(\mathbf{x}_k)$ is

$$\begin{aligned}
 \delta^\pi(\mathbf{x}_k) &= E(\theta \mid \mathbf{x}_k) \\
 &= x_{(k)} + \frac{1}{k}E(\eta \mid \mathbf{x}_k) \\
 &= x_{(k)} + \frac{1 + \frac{2c}{k}}{1 + \frac{c}{k}}x_{(k)} \\
 &= \frac{k^2 + (c+1)k + 2c}{k^2 + ck}x_{(k)}. \tag{4.36}
 \end{aligned}$$

Then, as $k \rightarrow \infty$, we have

$$\delta^\pi(\mathbf{x}_k) = x_{(k)}, \tag{4.37}$$

which does not depend on c . Therefore, the corresponding Bayes risk also does not depend on c . In other words, there is no better c to answer the above question,

we could choose any arbitrary c in \mathbb{R} to get the reference prior for this example when $\beta = -1$.

Next, another question is that what is the reference prior for this example when $\beta = 0$, the case of Kullback-Leibler divergence? For this situation, we know that the reference prior is given by

$$\begin{aligned}\pi(\theta) &= \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)}, \\ f_k(\theta) &= \exp \left\{ \int p(\mathbf{x}^{(k)} | \theta) \log[\pi^*(\theta | \mathbf{x}^{(k)})] d\mathbf{x}^{(k)} \right\}.\end{aligned}$$

However, analytical derivation of reference prior for this situation is hard to obtain. Instead, we find the reference prior numerically by using the algorithm given in Berger, Bernardo, and Sun (2009) as follows.

Step 1. Choose a moderate value for k ;

Choose an arbitrary positive function $\pi^*(\theta)$, say $\pi^*(\theta) = 1$;

Choose the number of m of samples to be simulated.

Step 2. For any given θ value, repeat, for $j = 1, \dots, m$:

Simulate a random sample $\{x_{1j}, \dots, x_{kj}\}$ of size k from $p(x | \theta)$;

Compute numerically the integral $c_j = \int_{\Theta} \prod_{i=1}^k p(x_{ij} | \theta) \pi^*(\theta) d\theta$;

Evaluate $r_j(\theta) = \log \left[\prod_{i=1}^k p(x_{ij} | \theta) \pi^*(\theta) / c_j \right]$.

Step 3. Compute $\pi(\theta) = \exp \left[m^{-1} \sum_{j=1}^m r_j(\theta) \right]$ and store the pair $\{\theta, \pi(\theta)\}$.

Step 4. Repeat *Step 2* and *Step 3* for all θ values for which the pair $\{\theta, \pi(\theta)\}$ is required.

It is worth to say a little bit about how we can generate a random variable from the distribution $p(x | \theta)$ in *Step 2*. We know that if $u \sim \text{uniform}(0, 1)$, then $x = F^{-1}(u) \sim p(x | \theta)$ as desired, where, from (4.23),

$$F^{-1}(u) = -\frac{1}{\theta} \log[1 - (1 - e^{-\theta^2})u].$$

The numerical computation was done for the problem by the following R code and the figure 4.1 presents the reference prior numerically.

```
#####
### R code for figure 4.1 ###
#####

k = 500

m = 1000

theta = seq(0.05,10,length=100)

c = rep(NA,m)

r = rep(NA,m)

for (l in 1:length(theta))

{

G = function(y){

-1/theta[l]*log(1-(1-exp(-theta[l]^2))*y)
```

```
}  
  
for (j in 1:m){  
  x = rep(NA,k)  
  for (i in 1:k){  
    u = runif(1)  
    x[i] = G(u)  
  }  
  c[j] = integrate(function(t) t^k/(1-exp(-t^2))^k*exp(-t*sum(x)), max(x),Inf)  
  $value  
  r[j] = log(theta[1]^k/(1-exp(-theta[1]^2))^k*exp(-theta[1]*sum(x)))/c[j])  
}  
  
ref[1] = exp(sum(r)/m)  
  
plot(theta,ref,type="l",xlab=expression(theta),ylab=expression(pi(theta)))
```

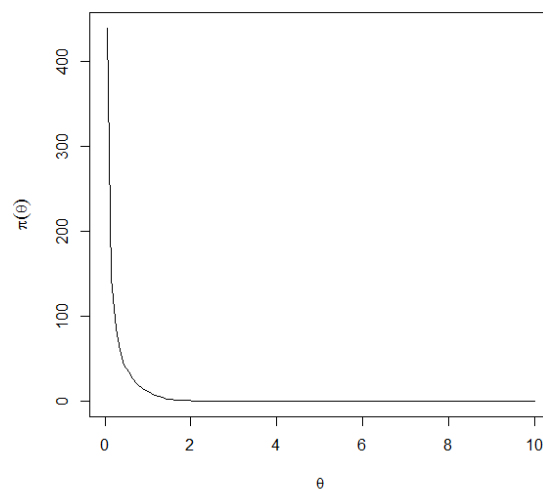


FIGURE 4.1: Numerical reference prior for the example 4.2 when $\beta = 0$, the Kullback-Leibler divergence

Chapter 5

Summary and Comments

In this dissertation, we revisit the problem of using reference analysis to derive objective priors. We generalize the results from Berger, Bernardo and Sun (2009), Ghosh, Mergel and Liu (2011), and make some new contributions based on the use of a very general class of divergence measures, known as β -divergence measures (Amari 1982, Cressie and Read 1984). This class of β -divergence measures includes the Kullback-Leibler divergence, the Hellinger divergence, and the Chi-squared divergence.

We introduce precise the definition of the reference prior with two properties, permissible and maximizing missing information properties.

We show how an explicit expression for the reference prior can be obtained under very weak regularity conditions in the case $0 < |\beta| < 1$ and it can be used to derive new reference prior both analytically and numerically. It is shown that for a model described by density $p(\mathbf{x}^{(k)} | \theta)$, the formula for the reference prior in

this case, $\pi(\theta)$, is

$$\pi(\theta) = \lim_{k \rightarrow \infty} \frac{f_k(\theta)}{f_k(\theta_0)},$$

$$f_k(\theta) = \left\{ \int p(\mathbf{x}^{(k)} | \theta) [\pi^*(\theta | \mathbf{x}^{(k)})]^{-\beta} d\mathbf{x}^{(k)} \right\}^{-1/\beta},$$

where θ_0 is an interior point of the parameter space Θ , $\mathbf{x}^{(k)} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ stands for k replications of \mathbf{x} , and $\pi^*(\theta | \mathbf{x}^{(k)})$ is the posterior distribution corresponding to some fixed arbitrary prior $\pi^*(\theta)$. This formula holds for any type of continuous parameter models, regardless of the asymptotic nature of the posterior. Prior under the Kullback-Leibler divergence is a special case in a limiting sense as $\beta \rightarrow 0$.

In the final part, under the Chi-squared divergence, the boundary case when $\beta = -1$, we find that, for the non-regular models, there is no general formula for the reference priors. For some models there exist reference priors, for some other models there do not.

For each β such that $0 \leq |\beta| < 1$ we have the corresponding divergence measure and from this divergence measure we could construct a new reference prior, enlarging the class of objective priors. These reference priors provide a new insight about the concept of objective priors, low information priors. The study of these priors will help us to see if the inferences based on these priors are robust or not, at least with respect to the reference priors.

Next, one natural arising question: how to choose β ? which β will be the best in this β -divergence class? One way to answer this question is based on the Bayes

risks, we will choose β such that the corresponding Bayes risk is smallest. The routine will be as follows. First, we find the Bayes estimator for θ , $\delta^\pi(\mathbf{x}_k) = E(\theta | \mathbf{x}_k)$, under the squared error loss. Then, we find the risk $R(\theta, \delta^\pi(\mathbf{x}_k)) = E(\delta^\pi(\mathbf{x}_k) - \theta)^2$. Next, calculate the Bayes risk $r(\pi, \delta^\pi(\mathbf{x}_k)) = \int_{\Theta} R(\theta, \delta^\pi(\mathbf{x}_k))\pi(\theta)d\theta$. The Bayes risk $r(\pi, \delta^\pi(\mathbf{x}_k))$ is a function of β and hence we could choose β such that the Bayes risk is smallest. However, for the complex models, it seems to be extremely difficult to calculate the Bayes risk. So we could use the algorithm referred in section 3 to find the numerical reference priors depending on β and see how they are different. For example, consider again the uniform distribution on (θ, θ^2) in the Example 3.4 where now $0 \leq |\beta| < 1$. By using the similar R code as we did in this example for $\beta = 0.5, 0.9, 0.2, -0.5$, we have the figure 5.1, which presents the numerical reference priors for these cases.

```
#####
### R code for figure 5.1 ###
#####

beta = c(0.5,0.9,0.2,-0.5)

theta = seq(1.05,10,length=180)

c = rep(NA,m)

r = rep(NA,m)

ref = rep(NA,length(theta))

for (l in 1:length(theta))

{

G = function(y){
```

```

theta[1]+(theta[1]^2-theta[1])*y
}
for (j in 1:m){
x = rep(NA,k)
for (i in 1:k){
u = runif(1)
x[i] = G(u)
}
c[j] = integrate(function(t) 1/(t^2-t)^k, sqrt(max(x)),min(x))$value
r[j] = (1/(theta[1]^2-theta[1])^k/c[j])^(-beta)
}
ref[1] = (sum(r)/m)^(-1/beta)
}
ref = ref/ref[theta==2]

```

From this figure we can see that the numerical reference priors are clearly almost perfectly the same. Therefore, we could have a guess that there is no better β for this example. We could choose any β such that $0 \leq |\beta| < 1$ to do inferences and these inferences are robust. However, for this example, we have derived the closed form for the reference prior as follows,

$$\pi(\theta) \propto \frac{2\theta - 1}{\theta^2 - \theta} \left[\frac{2\theta}{2\theta - (2\theta - 1)\beta} + \frac{2\theta}{(\beta + 2)\Gamma(-\beta)(2\theta - 1)^2} \right]^{-\frac{1}{\beta}}. \quad (5.1)$$

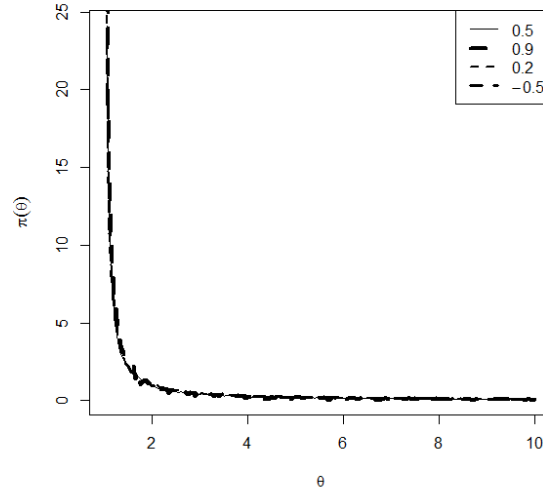


FIGURE 5.1: Numerical reference priors for the uniform model on (θ, θ^2) when $\beta = 0.5, 0.9, 0.2, -0.5$

The corresponding posterior is

$$\pi(\theta | \mathbf{t}_k) \propto \frac{2\theta - 1}{(\theta^2 - \theta)^{k+1}} \left[\frac{2\theta}{2\theta - (2\theta - 1)\beta} + \frac{2\theta}{(\beta + 2)\Gamma(-\beta)(2\theta - 1)^2} \right]^{-\frac{1}{\beta}} \quad (5.2)$$

where $\sqrt{t_2} < \theta < t_1$. Then, the Bayes estimator for θ is

$$\begin{aligned} \delta^\pi(\mathbf{t}_k) &= E(\theta | \mathbf{t}_k) \\ &= \frac{\int_{\sqrt{t_2}}^{t_1} \frac{2\theta - 1}{(\theta - 1)(\theta^2 - \theta)^k} \left[\frac{2\theta}{2\theta - (2\theta - 1)\beta} + \frac{2\theta}{(\beta + 2)\Gamma(-\beta)(2\theta - 1)^2} \right]^{-\frac{1}{\beta}} d\theta}{\int_{\sqrt{t_2}}^{t_1} \frac{2\theta - 1}{(\theta^2 - \theta)^{k+1}} \left[\frac{2\theta}{2\theta - (2\theta - 1)\beta} + \frac{2\theta}{(\beta + 2)\Gamma(-\beta)(2\theta - 1)^2} \right]^{-\frac{1}{\beta}} d\theta}. \end{aligned} \quad (5.3)$$

Obviously, it is extremely difficult to calculate these integrals. So, we will calculate the Bayes risk numerically by the following R code.

```
#####  
### R code for the figure 5.2###  
#####  
beta = seq(-0.95,0.95,0.006)  
k = 500  
theta = seq(1.05,10,length=359)  
bay.risk = rep(NA,length(beta))  
for (j in 1:length(beta))  
{  
  bay.est = rep(NA,length(theta))  
  for (l in 1:length(theta))  
  {  
    G = function(y){  
      theta[l]+(theta[l]^2-theta[l])*y  
    }  
    x = rep(NA,k)  
    for (i in 1:k){  
      u = runif(1)  
      x[i] = G(u)  
    }  
    bay.est[l] = integrate(function(t) (2*t-1)/(t-1)/(t^2-t)^k*(2*t/  
      (2*t-(2*t-1)*beta[j])+ 2*t/((beta[j]+2)*(gamma(1-beta[j])/  
      (-beta[j]))*(2*t-1)^2))^(-1/beta[j]), sqrt(max(x)),min(x))$value/
```

```

integrate(function(t) (2*t-1)/(t^2-t)^(k+1)*(2*t/(2*t-(2*t-1)*beta[j])+
2*t/((beta[j]+2)*(gamma(1-beta[j])/(-beta[j]))*(2*t-1)^2))^(-1/beta[j]),
sqrt(max(x)),min(x))$value
}

prior = function(t){
(2*t-1)/(t^2-t)*(2*t/(2*t-(2*t-1)*beta[j])+ 2*t/((beta[j]+2)*
(gamma(1-beta[j])/(-beta[j]))*(2*t-1)^2))^(-1/beta[j])
}

ref = prior(theta)
ref = ref/(sum(ref*0.025))

risk = function(t){
var(bay.est)+(mean(bay.est)-t)^2
}

Risk = risk(theta)
bay.risk[j] = sum(ref*(Risk))*0.025
}

plot(beta,bay.risk,type="l",xlab=expression(beta),ylab= "bayes risk")
beta[which.min(bay.risk)]

```

The figure 5.2 shows the plot of the Bayes risk versus β and we see that the minimum of the Bayes risk will be obtained when β is around 0.5. This result is different from the previous guess (there is no better β for this example). Actually, the range of the Bayes risk is rather small, (20.6, 21.6). Now, we consider another

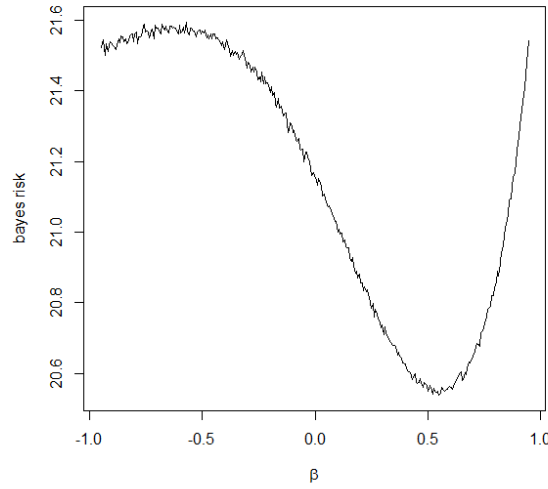


FIGURE 5.2: Numerical Bayes risk for the uniform model on (θ, θ^2)

example where the model as follows,

$$p(x | \theta) = \frac{2x}{\theta^4 - \theta^2}, \quad (5.4)$$

where $\theta < x < \theta^2, \theta > 1$. This is a special case of the example 3.5 with $g(x) = x, a_1(\theta) = \theta$, and $a_2(\theta) = \theta^2$. Then, from (3.88), the closed form for the reference prior in this example is

$$\pi(\theta) \propto \frac{2\theta^2 - 1}{\theta^3 - \theta} \left[\frac{2\theta^2}{2\theta^2 - (2\theta^2 - 1)\beta} + \frac{2\theta^2}{(\beta + 2)\Gamma(-\beta)(2\theta^2 - 1)^2} \right]^{-\frac{1}{\beta}}. \quad (5.5)$$

Similarly as in the previous example, the plot of the Bayes risk versus β is given in the figure 5.3 by using the following R code.

```
#####

### R code for the figure 5.3###

#####

beta = seq(-0.95,0.95,0.006)

k = 500

theta = seq(1.05,3,length=79)

bay.risk = rep(NA,length(beta))

for (j in 1:length(beta))
{

bay.est = rep(NA,length(theta))

for (l in 1:length(theta))
{

G = function(y){

theta[l]*sqrt((theta[l]^2-1)*y+1)

}

x = rep(NA,k)

for (i in 1:k){

u = runif(1)

x[i] = G(u)

}

bay.est[l] = integrate(function(t) (2*t^2-1)/(t^2-1)/(t^4-t^2)^k*(2*t^2/
(2*t^2-(2*t^2-1)*beta[j])+2*t^2/((beta[j]+2)*(gamma(1-beta[j])/(-beta[j]))*
(2*t^2-1)^2))^(-1/beta[j]), sqrt(max(x)),min(x))$value/
```

```

integrate(function(t) (2*t^2-1)/(t^3-t)/(t^4-t^2)^k*(2*t^2/
(2*t^2-(2*t^2-1)*beta[j])+2*t^2/((beta[j]+2)*(gamma(1-beta[j])/(-beta[j]))*
(2*t^2-1)^2))^(-1/beta[j]), sqrt(max(x)),min(x))$value
}

prior = function(t){
(2*t^2-1)/(t^3-t)*(2*t^2/(2*t^2-(2*t^2-1)*beta[j])+ 2*t^2/
((beta[j]+2)*(gamma(1-beta[j])/(-beta[j]))*(2*t^2-1)^2))^(-1/beta[j])
}

ref = prior(theta)
ref = ref/(sum(ref*0.025))

risk = function(t){
var(bay.est)+(mean(bay.est)-t)^2
}

Risk = risk(theta)
bay.risk[j] = sum(ref*(Risk))*0.025
}

plot(beta,bay.risk,type="l",xlab=expression(beta),ylab= "bayes risk")
beta[which.min(bay.risk)]

```

Once again, we have the Bayes risk is minimum when β is around 0.5 and the range for the Bayes risk is rather small (0.87, 0.92).

The main scope of this dissertation is restricted to single parameter models.

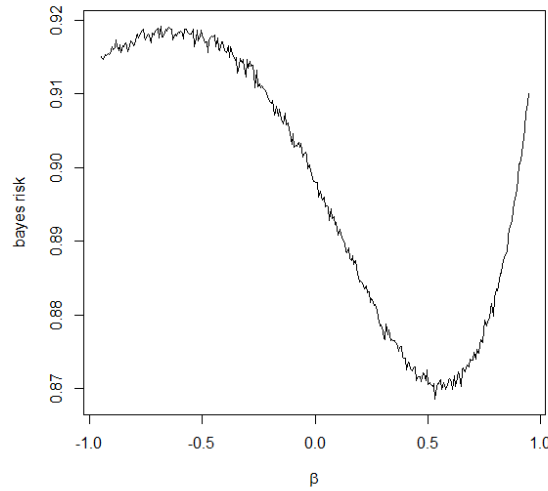


FIGURE 5.3: Numerical Bayes risk

It would be very useful to be able to generalize the results under this general divergence measures class in the presence of nuisance parameters. The permissibility and the MMI property could be generalized to the multi-parameter case. However, the main purpose here is to find the explicit representation for the reference prior as we did in the Theorem 3.1. Unfortunately, as mentioned in Berger, Bernardo, and Sun (2009) for the Kullback-Leibler divergence when $\beta = 0$, it seems that any generalizations require expressions involving limits over approximating compact sequences, the feature we want to avoid when finding reference priors.

Finally, there is an interested note that when $\beta = 0$, the reference prior corresponds to the maximin prior. The maximin prior story was outlined by Clarke (2013) as follows. Under the maximin standpoint we want to find the maximin prior $\pi_{Mn}(\theta)$ achieving

$$\arg \max_{\pi} \left[\min_Q \int \pi(\theta) \Delta^{\beta}(Q, L(\theta)) d\theta \right], \quad (5.6)$$

where $L(\theta)$ is the likelihood function and $\Delta^\beta(Q, L(\theta))$ is the β -distance between $Q(\mathbf{x})$ and $L(\theta)$,

$$\Delta^\beta(Q, L(\theta)) = \frac{1}{\beta(1-\beta)} \left[1 - \int Q^\beta(\mathbf{x}) L^{1-\beta}(\theta) d\mathbf{x} \right]. \quad (5.7)$$

Under the Kullback-Leibler divergence ($\beta = 0$),

$$\min_Q \int \pi(\theta) \Delta^\beta(Q, L(\theta)) d\theta = \int \pi(\theta) \Delta^\beta(m, L(\theta)) d\theta, \quad (5.8)$$

or the marginal density $m(\mathbf{x})$ is the Bayes estimator of the likelihood $L(\theta)$. Then, from the relation $L(\theta)\pi(\theta) = \pi(\theta | \mathbf{x})m(\mathbf{x})$, we have

$$\begin{aligned} & \int \pi(\theta) \Delta^\beta(m, L(\theta)) d\theta \\ &= \int \pi(\theta) \frac{1}{\beta(1-\beta)} \left[1 - \int m^\beta(\mathbf{x}) L^{1-\beta}(\theta) d\mathbf{x} \right] d\theta \\ &= \frac{1}{\beta(1-\beta)} \left\{ 1 - \int \pi(\theta) \left[\int m^\beta(\mathbf{x}) L^{1-\beta}(\theta) d\mathbf{x} \right] d\theta \right\} \\ &= \frac{1}{\beta(1-\beta)} \left\{ 1 - \int \left[\int \pi^\beta(\theta) \pi^{1-\beta}(\theta | \mathbf{x}) d\theta \right] m(\mathbf{x}) d\mathbf{x} \right\} \\ &= R^\beta(\pi). \end{aligned} \quad (5.9)$$

Therefore, from (5.6), (5.8), and (5.9), we could see that if $m(\mathbf{x})$ is the Bayes estimator of the likelihood $L(\theta)$, then the maximin prior corresponds to the reference prior. However, when $\beta \neq 0$ the Bayes estimator of the likelihood $L(\theta)$ is no longer $m(\mathbf{x})$ and the reference prior may be different from the maximin prior. In this situation, we want to see if this difference is small asymptotically or not and that will be what we deal with in the future.

References

- [1] Berger, J. O. (1985) Statistical decision theory and bayesian analysis. *Springer, Berlin*.
- [2] Berger, J. O. and Bernardo, J. M. (1989) Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.*, **84**, 200–207.
- [3] Berger, J. O. and Bernardo, J. M. (1992) On the development of the reference priors. *Bayesian statistics*, **4**, 35–60.
- [4] Berger, J. O. and Bernardo, J. M. (1992) Ordered group reference priors with applications to a multinomial problem. *Biometrika*, **79**, 25–37.
- [5] Berger, J. O. and Bernardo, J. M. (1992) Reference priors in a variance components problem. *Bayesian analysis in statistics and econometrics*, pp. 323–340.
- [6] Berger, J. O., Bernardo, J. M., and Sun, D. (2009) The formal definition of reference priors. *The Annals of Statistics*, **37**, 905–938.
- [7] Bernardo, J. M. (1979) Reference prior distributions for bayesian inference. *J. R. Stat. Soc. Ser. B*, **41**, 113–147.
- [8] Clarke, B. (1999) Asymptotic normality of the posterior in relative entropy. *IEEE Trans. Inform. Theory*, **45**, 165–176.
- [9] Clarke, B. and Barron, A. R. (1994) Jeffreys’ prior is asymptotically least favorable under entropy risk. *J. Statist. Plann. Inference*, **41**, 37–60.
- [10] Clarke, B. and Sun, D. (1997) Reference priors under the chi-squared distance. *The Indian Journal of Statistics*, **59**, 215–231.
- [11] Clarke, B. and Sun, D. (1999) Asymptotics of the expected posterior. *Ann. Inst. Statist. Math.*, **51**, 163–185.

-
- [12] Cressie, N. and Read, T. R. C. (1984) Multinomial goodness-of-fit tests. *J. R. Stat. Soc. Ser. B*, **46**, 440–464.
- [13] Datta, G. S. and Mukerjee, R. (2004) Probability matching priors: higher order asymptotics. *Springer, New York*.
- [14] Ghosal, S. (1997) Reference priors in multiparameter nonregular cases. *Test*, **6**, 159–186.
- [15] Ghosal, S. and Samanta, T. (1997) Expansion of bayes risk for entropy loss and reference prior in nonregular case. *Statist. Decisions*, **15**, 129–140.
- [16] Ghosh, M., Mergel, V., and Liu, R. (2011) A general divergence criterion for prior selection. *Annals of the Institute of Statistical Mathematics*, **63**, 43–58.
- [17] Hewitt, E. and Stromberg, K. (1969) Real and abstract analysis. a modern treatment of the theory of functions of a real variable. *Springer, Berlin*.
- [18] Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.
- [19] Sebah, P. and Gourdon, X. (2002) Introduction to the gamma function. *numbers.computation.free.fr/Constants/constants.html*.
- [20] Sun, D. and Berger, J. O. (1998) Reference priors under partial information. *Biometrika*, **85**, 55–71.

VITA

Tri Minh Le was born in Quang Ngai, Vietnam. He earned his Bachelor degree from the University of Natural Sciences, HCMC, Vietnam in 1999, major in mathematics and informatics. Then in 2000 he went to the University of Natural Sciences, HCMC, Vietnam as a graduate student and received his Master degree in probability and statistics in 2002. Later, he enrolled in the Ph.D. program of the Department of Statistics at the University of Missouri-Columbia, USA in the fall 2009. He is going to get his Ph.D. degree from the University of Missouri-Columbia in the summer of 2014.