OBJECTIVE PRIORS UNDER THE ALPHA-DIVERGENCE MEASURES

DAYU SUN

M.Phil

The Hong Kong Polytechnic University

2016

The Hong Kong Polytechnic University

Department of Applied Mathematics

Objective Priors under the Alpha-Divergence Measures

Dayu Sun

A thesis submitted in partial fulfilment of the requirements

for the degree of Master of Philosophy

April 2016

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_______________________________(Signed)

__________Dayu Sun__________(Name of student)

Dedicate to my parents.

Abstract

Without prior knowledge, objective priors are derived from an assumed model for Bayesian inference. A popular approach is to elicit an objective prior which asymptotically maximizes the expected divergence measure between a prior and its corresponding posterior. The class of α-divergence measures has been used for unified prior selection recently as this class includes many widely used divergence measures. However, there is a lack of rigorous derivation of objective priors in the presence of nuisance parameters under the α-divergence measure. Neglecting nuisance parameters may lead to unsatisfactory objective Bayesian inference. In this thesis, a rigorous method is developed to elicit objective priors in the presence of nuisance parameters using the α-divergence measure. This new class of objective priors is called alpha priors. Alpha priors with independence assumption between model parameters and those with model parameters divided into multiple ordered groups are also derived. Several examples are presented to illustrate the alpha prior method. A simulation study based on a two-parameter regression model is performed to compare alpha priors with different values of α. A real example from an extracorporeal membrane oxygenation experiment demonstrates an application of the alpha prior.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my chief supervisor Dr. Xingqiu Zhao for she gave me a valuable opportunity to be a research student. I am also grateful for her continuous support of my M.Phil. study and related research, patience, motivation, and immense knowledge.

Furthermore, I would like to express my deep thanks to my co-supervisor Dr. Zhisheng Ye for his guidance helped me in all the time of research and writing of scientific papers as well as this thesis. I benefited much from his rigorous and diligent attitude to scientific research. I could not have imagined having a better supervisor and mentor for my M.Phil. study.

Besides my supervisors, I would like to thank my schoolmates, for their discussion, insightful comments and encouragement on my research.

Last but not least, I would like to thank my parents for supporting me spiritually and financially throughout my study and life in general.

Contents

List of Tables

List of Figures

List of Notations

Symbols

$\mathbb{R}$	The set of real numbers.
$D^{\alpha}(p\|q)$	The α-divergence measure between two densities $p(\cdot)$ and $q(\cdot)$.
$\boldsymbol{x}_n$	A random sample with size n.
θ	Unknown model parameters.
Θ	A compact parameter space.
$\mathbf{I}(\theta)$	The Fisher information matrix.
$\lvert\cdot\rvert$	The determinant of a matrix.

Abbreviations

PDF	Probability density function.
CDF	Cumulative density function.
ECMO	Extracorporeal membrane oxygenation.

Chapter 1

Introduction

1.1 Background

Bayesian analysis increasingly draws statisticians' attention recently due to its wide application. Implementation of Bayesian methods requires a prior distribution for model parameters, which is often summarized from prior knowledge. Then, combining the prior with the information obtained from observed data, one can derive a posterior distribution for the model parameters and use it to draw inference.

However, in practice, it is usually hard to elicit an appropriate prior subjectively due to lacking prior information. In the absence of prior knowledge, it is possible to adopt a prior which adds little information to sample information. The prior is called "objective" which are also known as "default" or "noninformative" prior. Objective priors are often constructed by some formal rules, depending on the model only. Some recent comprehensive reviews of objective priors include Ghosh (2011) and Kass and Wasserman (1996). Two main methods to construct objective priors are described as follows.

1.1.1 Jeffreys prior

A common objective prior is the Jeffreys prior (see Jeffreys, 1961). It is proportional to the square root of the determinant of the Fisher information matrix. The Jeffreys prior enjoys many optimality properties in the one-parameter case. It is invariant, that is, the Jeffreys prior remains same under one-to-one transformation of model parameters. Nevertheless, it suffers some problems when dealing with nuisance parameters. For example, the Jeffreys prior may lead to an inconsistent posterior estimation of parameters of interest. This happens in many location-scale modeled problems, such as the Neyman-Scott problem (Neyman and Scott, 1958).

1.1.2 Reference prior

To overcome difficulties resulting from the Jeffreys prior, Bernardo (1979) proposed the reference prior approach. He selected objective priors that asymptotically maximize the expected Kullback-Leibler divergence between a prior and its corresponding posterior. The divergence between the prior and the posterior can be regarded as the measure of the missing information about the unknown parameters. In this manner, data from a random sample may have the largest amount of influence on the behavior of the posterior distribution. In another word, the larger the divergence measure, the lower the information that the prior contains. This approach has been further developed by Berger and Bernardo (1989, 1992a,b). In the presence of nuisance parameters, Berger and Bernardo (1992a,b) developed a stepwise procedure to derive the reference prior. Thus, the general form of the reference prior depends on the inferential importance of model parameters. The reference prior has been shown to be successful to dealing many problems in the presence of nuisance parameters.

1.2 Literature review

Apart from the Kullback-Leibler divergence, other divergence measures have been used to derive objective priors recently. Clarke and Sun (1997) obtained objective priors under the chi-squared distance for a one-dimensional exponential family, leading to a prior proportional to the reciprocal of the Jeffreys prior. They argued that this prior weights more difficultly-discriminable parameters than the Jeffreys prior does. Shemyakin (2014) proposed the Hellinger prior which is proportional to the square root of the Hellinger information constructed by the Hellinger distance. The Hellinger priors coincide with Bernardo-Berger's reference priors under certain regular conditions and in many irregular cases.

Apart from the individual divergence measures mentioned above, there exists a more general class of divergence measures, known as α-divergence measures (see Amari, 1985; Cichocki and Amari, 2010). The class of α-divergence measures includes different types of divergence measures, such as the Kullback-Leibler divergence measure and the squared Hellinger divergence measure. Hence, the α-divergence measure allows us to derive a unified general objective priors applicable to different divergence measures. We can also compare objective Bayesian inference under different divergence measures in the class of α-divergence measures to find a better prior for a specific problem. Ghosh et al. (2011) derived objective priors in the one-parameter case using the α-divergence measure. Liu et al. (2014) then generalized their results to the multi-parameter case. In the absence of nuisance parameters, Liu et al. (2014) found that, except the chi-squared divergence measure, the α-divergence measures lead to the Jeffreys prior in the regular continuous case. Due to the limitation of the Jeffreys prior, how to handle nuisance parameters under the α-divergence measure needs investigation. Liu et al. (2014) considered an approximation to the α-divergence measure between the marginal prior and the marginal posterior of a parameter of interest. They then derived a marginal objective prior of the parameter of interest by maximizing the approximation. However, they considered a two-parameter case only and the derivation of the approximation is not rigorous.

1.3 Goals

In this thesis, we fill the gap in the previous research and rigorously develop objective priors in the presence of nuisance parameters using the class of α-divergence measures. Here, we will confine ourselves to the regular continuous case. In contrast to the case where nuisance parameters are absent, the general form of these objective priors depends on the value of α and hence the choice of divergence measures. Therefore, we have a wider choice of objective priors and call these objective priors alpha priors.

A prerequisite for the prior elicitation is the asymptotic behavior of the expected α-divergence between a marginal prior and its corresponding marginal posterior. The other one is the asymptotic behavior of the expected α-divergence between a conditional prior and its corresponding conditional posterior. We will establish the two prerequisites first and then derive the alpha prior. Sometimes, the independence between model parameters is assumed. We also consider alpha priors with the independence assumption.

Berger and Bernardo (1992b,a) pointed out that it is insufficient to separate model parameters into parameters of interest and nuisance parameters for some problems. They instead suggested to split model parameters into multiple ordered groups for these problems. Therefore, we also propose an algorithm to elicit alpha priors when model parameters are divided into multiple ordered groups.

1.4 Outline

The main purpose of this thesis is to derive the alpha prior, a new class of objective priors beyond Bernardo-Berger's reference prior. The rest of the thesis is organized as follows.

Chapter 2 provides the asymptotic behavior of the expected α-divergence between a marginal prior and its corresponding marginal posterior. The limiting behavior of the expected divergence in the conditional prior case is also established.

Chapter 3 proposes a complete approach to derive alpha priors based on the results in Chapter 2. We also extend the approach to the case where parameters are independent and the multiple ordered group case.

In Chapter 4, we illustrate the alpha prior approach by several examples, including a clinical trial of extracorporeal membrane oxygenation (ECMO). We also compare alpha priors with different values of α by a simulation study and an illustrative example.

Chapter 5 contains some concluding remarks and suggestions for future work.

Chapter 2

Asymptotics of the expected α-divergence measures

2.1 Notations and definitions

The general α-divergence measure between two densities $p(\cdot)$ and $q(\cdot)$ is defined as

$$D^{\alpha}(p\|q) = \frac{1 - \int p^{\alpha}(x)\, q^{1-\alpha}(x)\, \mathrm{d}x}{\alpha(1-\alpha)}, \tag{2.1}$$

where $\alpha \in \mathbb{R}\backslash\{0,1\}$. Several well-known divergence measures are special cases of (2.1). For example, the squared Hellinger distance is $D^{1/2}(p\|q)$. The Kullback-Leibler divergence measure is defined as $D^{0}(p\|q)$ because it equals $\lim_{\alpha\to 0} D^{\alpha}(p\|q)$.

Let $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ be an observed random sample from a distribution P_θ with a probability density function (PDF) $p(x|\theta)$, where $\theta = (\theta_1, \theta_2)$ and the model parameters θ_1 and θ_2 are vectors of dimensions d_1 and d_2, respectively. Let $p(\boldsymbol{x}_n|\theta)$ denote the joint density function of $\boldsymbol{x}_n$. Let $\pi(\theta)$ and $\pi_i(\theta_i)$ $(i = 1, 2)$ denote the joint prior density of (θ_1, θ_2) and the marginal prior density of θ_i, respectively. Besides, $\pi_i(\theta_i|\theta_{3-i})$ denotes the conditional prior density of θ_i given θ_{3-i}, where $i = 1, 2$. The posterior density of θ and the marginal density of $\boldsymbol{x}_n$ are denoted by $p(\theta|\boldsymbol{x}_n)$ and $m(\boldsymbol{x}_n)$, respectively. Also, we use $p(\theta_i|\boldsymbol{x}_n)$ to denote the marginal posterior density of θ_i, where $i = 1, 2$. Define the marginal model for θ_1 as $p(\boldsymbol{x}_n|\theta_1) =$

$\int p(\boldsymbol{x}_n|\theta_1,\theta_2)\,\pi(\theta_2|\theta_1)\mathrm{d}\theta_2$ and the conditional posterior density of θ_2 given θ_1 as $p(\theta_2|\theta_1,\boldsymbol{x}_n) = p(\boldsymbol{x}_n|\theta)\pi(\theta_2|\theta_1)/p(\boldsymbol{x}_n|\theta_1)$.

The per-observation Fisher information matrix for θ is

$$\mathbf{I}(\theta) = \begin{pmatrix} \mathbf{I}_{11}(\theta) & \mathbf{I}_{12}(\theta) \\ \mathbf{I}_{21}(\theta) & \mathbf{I}_{22}(\theta) \end{pmatrix},$$

where $\mathbf{I}_{ij}(\theta) = -E\left[\partial^2 \log(p(x|\theta))/\partial\theta_i\partial\theta_j\right]$ and $i = 1, 2$. Specifically, $\mathbf{I}_{ii}(\theta)$ is the per-observation Fisher information matrix for θ_i, given that θ_{3-i} is held fixed. Let $|\cdot|$ be the determinant of a matrix.

Let θ_1 and θ_2 be a parameter of interest and a nuisance parameter, respectively. From (2.1), the α-divergence measure between $\pi_1(\theta_1)$ and $p(\theta_1|\boldsymbol{x}_n)$ is defined as

$$D^{\alpha}(\pi_1(\theta_1)\,\|p(\theta_1|\boldsymbol{x}_n)) = \frac{1 - \int \pi_1^{\alpha}(\theta_1)\,p^{1-\alpha}(\theta_1|\boldsymbol{x}_n)\,\mathrm{d}\theta_1}{\alpha(1-\alpha)},$$

which is a function of $\boldsymbol{x}_n$. The expected α-divergence between the marginal prior density and the marginal posterior density of θ_1 is

$$J^{\alpha}(\pi_1(\theta_1)) = \frac{1 - \int\int \pi_1^{\alpha}(\theta_1)\,p^{1-\alpha}(\theta_1|\boldsymbol{x}_n)\,\mathrm{d}\theta_1 m(\boldsymbol{x}_n)\,\mathrm{d}\boldsymbol{x}_n}{\alpha(1-\alpha)},$$

where $\alpha \in (-\infty, 0) \cup (0, 1)$. We then consider the expected α-divergence between the conditional prior $\pi_2(\theta_2|\theta_1)$ and the corresponding conditional posterior $p(\theta_2|\theta_1,\boldsymbol{x}_n)$. Similarly, it is defined as

$$J^{\alpha}(\pi_2(\theta_2|\theta_1)) = \frac{1 - \int\int\int \pi_2^{\alpha}(\theta_2|\theta_1)\,p^{1-\alpha}(\theta_2|\theta_1,\boldsymbol{x}_n)\,\mathrm{d}\theta_2 p(\boldsymbol{x}_n,\theta_1)\,\mathrm{d}\theta_1\mathrm{d}\boldsymbol{x}_n}{\alpha(1-\alpha)},$$

where $p(\boldsymbol{x}_n,\theta_1) = \int p(\boldsymbol{x}_n|\theta_1,\theta_2)\,\pi(\theta_1,\theta_2)\,\mathrm{d}\theta_2 = p(\boldsymbol{x}_n|\theta_1)\,\pi_1(\theta_1)$.

We follow the stepwise approach introduced by Bernardo (1979) and Berger and Bernardo (1992b) in this Chapter. The main idea of Bernardo-Berger's strategy is to first find $\pi_2(\theta_2|\theta_1)$, the normalized conditional alpha prior of θ_2 given θ_1. The next step is to derive the marginal alpha prior for $\pi_1(\theta_1)$, based on $\pi_2(\theta_2|\theta_1)$. To derive $\pi_1(\theta_1)$ and $\pi_2(\theta_2|\theta_1)$, we establish the respective asymptotic behaviors of $J^{\alpha}(\pi_1(\theta_1))$ and $J^{\alpha}(\pi_2(\theta_2|\theta_1))$ in Section 2.2. Then, $\pi_1(\theta_1)$ and $\pi_2(\theta_2|\theta_1)$ are asymptotic maximizers of $J^{\alpha}(\pi_1(\theta_1))$ and $J^{\alpha}(\pi_2(\theta_2|\theta_1))$, respectively. We will present the algorithm for deriving the alpha prior in Section 3.1. This algorithm is extended to the ordered group case in Section 3.3.

2.2 Asymptotics of the expected α-divergence measures

In this section, we provide some asymptotic results of $J^{\alpha}(\pi_1(\theta_1))$ and $J^{\alpha}(\pi_2(\theta_2|\theta_1))$ under some regularity conditions, respectively. The limiting behavior of $J^{\alpha}(\pi_1(\theta_1))$ is shown in the following theorem.

Theorem 2.1. *Let the prior density $\pi(\theta)$ be positive and three times continuously differentiable on a compact parameter space Θ. The corresponding compact parameter spaces of θ_1 and θ_2 are Θ_1 and Θ_2, respectively. Both $p(\boldsymbol{x}_n|\,\theta)$ and $p(\boldsymbol{x}_n|\,\theta_1)$ satisfy the conditions (AI)-(AV) of Ghosh et al. (1982) (see the appendix for details). Then,*

$$\lim_{n\to\infty}\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_1}{2}}\left\{1-\alpha(1-\alpha)J^{\alpha}(\pi_1(\theta_1))\right\} = (1-\alpha)^{-\frac{d_1}{2}}\int_{\Theta_1}\pi^{\alpha+1}(\theta_1)\eta(\theta_1)\,\mathrm{d}\theta_1, \tag{2.2}$$

where

$$\eta(\theta_1) = \int_{\Theta_2}\left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{22}(\theta)|}\right)^{-\frac{\alpha}{2}}\pi_2(\theta_2|\theta_1)\,\mathrm{d}\theta_2,$$

and $\alpha \in \{-\infty, 1\}\backslash\{0\}$.

Proof. First,

$$
\begin{aligned}
1-\alpha\left(1-\alpha\right)J^{\alpha}\left(\pi_1\left(\theta_1\right)\right) &= \int\int_{\Theta_1} \pi^{\alpha}\left(\theta_1\right)p^{1-\alpha}\left(\theta_1|\boldsymbol{x}_n\right)\mathrm{d}\theta_1 m\left(\boldsymbol{x}_n\right)\mathrm{d}\boldsymbol{x}_n \\
&= \int\int_{\Theta_1} p\left(\theta_1|\boldsymbol{x}_n\right)m\left(\boldsymbol{x}_n\right)\left(\frac{\pi\left(\theta_1\right)}{p\left(\theta_1|\boldsymbol{x}_n\right)}\right)^{\alpha}\mathrm{d}\theta_1\mathrm{d}\boldsymbol{x}_n \\
&= \int\int_{\Theta_1} p\left(\boldsymbol{x}_n|\,\theta_1\right)\pi_1(\theta_1)\left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta_1\right)}\right)^{\alpha}\mathrm{d}\theta_1\mathrm{d}\boldsymbol{x}_n \qquad (2.3) \\
&= \int_{\Theta}\pi\left(\theta\right)\int p\left(\boldsymbol{x}_n|\theta\right)\left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta\right)}\Big/\frac{p\left(\boldsymbol{x}_n|\theta_1\right)}{p\left(\boldsymbol{x}_n|\theta\right)}\right)^{\alpha}\mathrm{d}\boldsymbol{x}_n\mathrm{d}\theta.
\end{aligned}
$$

The third equality (2.3) is due to the definition of conditional probabilities and the forth equality is because of the definition of $p\left(\boldsymbol{x}_n|\,\theta_1\right)$. Note that (2.3) can also be expressed as

$$
E\left[\left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta_1\right)}\right)^{\alpha}\right] = E\left[E\left[\left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta\right)}\Big/\frac{p\left(\boldsymbol{x}_n|\theta_1\right)}{p\left(\boldsymbol{x}_n|\theta\right)}\right)^{\alpha}\Bigg|\,\theta\right]\right].
$$

By Theorem 2.1 in Clarke and Barron (1990),

$$
\log\frac{p\left(\boldsymbol{x}_n|\theta\right)}{m\left(\boldsymbol{x}_n\right)} - \frac{d_1+d_2}{2}\log\frac{n}{2\pi} + \frac{1}{2}S_n^T\mathbf{I}^{-1}\left(\theta\right)S_n \to \frac{1}{2}\log\left|\mathbf{I}\left(\theta\right)\right| + \log\frac{1}{\pi\left(\theta\right)},
$$

in $L^1\left(P_\theta\right)$ and P_θ probability as $n\to\infty$ for each θ in the interior of the support of $\pi\left(\theta\right)$. Here, $S_n = \left(1/\sqrt{n}\right)\nabla\log\left(p\left(\boldsymbol{x}_n|\theta\right)\right)$ is the standardized score function. It can be seen that the assumptions (AI)-(AV) of Ghosh et al. (1982) for $p\left(\boldsymbol{x}_n|\,\theta\right)$ are sufficient for the conditions 1-3 in Clarke and Barron (1990). By the continuity of the exponential function, the continuous mapping theorem ensures that

$$
\left(\frac{2\pi}{n}\right)^{-\frac{d_1+d_2}{2}}\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta\right)}\exp\left(-\frac{1}{2}S_n^T\mathbf{I}^{-1}\left(\theta\right)S_n\right) \to \left|\mathbf{I}\left(\theta\right)\right|^{-\frac{1}{2}}\pi\left(\theta\right), \qquad (2.4)
$$

in probability as $n\to\infty$. The marginal model of θ_1 $p\left(\boldsymbol{x}_n|\theta_1\right)$ can be regarded as a marginal density over the parameter space of $\theta_2|\theta_1$. Under this conditional parameter space,

$$\left(\frac{2\pi}{n}\right)^{-\frac{d_2}{2}} \frac{p\left(\boldsymbol{x}_n|\theta_1\right)}{p\left(\boldsymbol{x}_n|\theta\right)} \exp\left(-\frac{1}{2}S_n^T A S_n\right) \to \left|\mathbf{I}_{22}\left(\boldsymbol{\theta}\right)\right|^{-\frac{1}{2}} \pi_2\left(\theta_2|\theta_1\right), \tag{2.5}$$

in probability as $n \to \infty$ where $A = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & I_{22}^{-1}\left(\boldsymbol{\theta}\right) \end{pmatrix}$. Incorporating the two limits (2.4) and (2.5), we have

$$\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_1}{2}} \left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta_1\right)}\right)^{\alpha} \exp\left(-\frac{\alpha}{2}S_n^T\left(\mathbf{I}^{-1}\left(\theta\right)-A\right)S_n\right) \to \left(\frac{\left|\mathbf{I}\left(\theta\right)\right|}{\left|\mathbf{I}_{22}\left(\theta\right)\right|}\right)^{-\frac{\alpha}{2}} \pi_1^{\alpha}\left(\theta_1\right),$$

in probability as $n \to \infty$.

It is well-known that $S_n \xrightarrow{d} \boldsymbol{Z}$ under P_θ where $\boldsymbol{Z}$ is distributed as $\mathcal{N}\left(0, \mathbf{I}\left(\theta\right)\right)$. Therefore, $\exp\left(\frac{\alpha}{2}S_n^T\left(\mathbf{I}^{-1}\left(\theta\right)-A\right)S_n\right)$ converges to $\exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T\left(\mathbf{I}^{-1}\left(\theta\right)-A\right)\boldsymbol{Z}\right)$ in distribution by the continuity of quadratic forms. Multiplying (2.4) by $\exp\left(\frac{\alpha}{2}S_n^T\left(\mathbf{I}^{-1}\left(\theta\right)-A\right)S_n\right)$, we have that

$$\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_1}{2}} \left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta_1\right)}\right)^{\alpha} \to \left(\frac{\left|\mathbf{I}\left(\theta\right)\right|}{\left|\mathbf{I}_{22}\left(\theta\right)\right|}\right)^{-\frac{\alpha}{2}} \pi^{\alpha}\left(\theta_1\right) \exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T\left(\mathbf{I}^{-1}\left(\theta\right)-A\right)\boldsymbol{Z}\right), \tag{2.6}$$

in distribution as $n \to \infty$ by Slutsky's theorem.

Second, we assume the marginal model $p\left(\boldsymbol{x}_n|\,\theta_1\right)$ satisfies the conditions (AI)-(AV). Thus, Theorem 3.1 in Liu et al. (2014) implies

$$E\left[\left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta_1\right)}\right)^{\alpha}\right] = \left(\frac{2\pi}{n}\right)^{d_1\alpha/2} C_n\left(\theta_1, \alpha\right), \tag{2.7}$$

when $\alpha < 1$. In (2.7), $C_n\left(\theta_1, \alpha\right)$ is bounded for any n. Therefore, the left hand side of (2.6) is uniformly integrable over the joint probability measure of θ and $\boldsymbol{x}_n$ since

$$\sup_n E\left[\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_1}{2}} \left(\frac{m\left(\boldsymbol{x}_n\right)}{p\left(\boldsymbol{x}_n|\theta_1\right)}\right)^{\alpha}\right]^{1+\delta} = \sup_n \left[C_n\left(\theta, \alpha\right)\right]^{1+\delta} < \infty,$$

for some $\delta > 0$.

Finally, by (2.6) and the uniform integrability of its left hand side,

$$\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_1}{2}} E\left[\left(\frac{m(\boldsymbol{x}_n)}{p(\boldsymbol{x}_n|\theta_1)}\right)^{\alpha}\right] \to$$

$$E_{\theta}\left[\left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{22}(\theta)|}\right)^{-\frac{\alpha}{2}} \pi_1^{\alpha}(\theta_1) E\left[\exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T\left(\mathbf{I}^{-1}(\theta) - A\right)\boldsymbol{Z}\right)|\theta\right]\right].$$

It is seen that $\boldsymbol{Z}^T(\mathbf{I}^{-1}(\theta) - A)\boldsymbol{Z}$ follows chi-squared distribution $\chi^2_{d_1}$. According to the moment generating function of $\chi^2_{d_1}$, $E\left[\exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T(\mathbf{I}^{-1}(\theta) - A)\boldsymbol{Z}\right)|\theta\right]$ is $(1-\alpha)^{-\frac{d_1}{2}}$. Note that this moment generating function only exists when $\alpha < 1$ because the expectation does not converge when $\alpha/2 \geqslant 1/2$. Therefore, the uniform integrability cannot be proved and Theorem 2.1 does not hold. This is the reason that Theorem 3.1 of Liu et al. (2014) and Theorem 2.1 in this thesis restrict α to be less than 1.

Finally, we have

$$\begin{aligned}
&\lim_{n\to\infty}\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_1}{2}}\{1-\alpha(1-\alpha)J^{\alpha}(\pi(\theta_1))\} \\
&= (1-\alpha)^{-\frac{d_1}{2}}\int_{\Theta_1}\pi_1^{\alpha+1}(\theta_1)\int_{\Theta_2}\pi(\theta_2|\theta_1)\left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{22}(\theta)|}\right)^{-\frac{\alpha}{2}}\mathrm{d}\theta_2\mathrm{d}\theta_1 \\
&= (1-\alpha)^{-\frac{d_1}{2}}\int_{\Theta_1}\pi_1^{\alpha+1}(\theta_1)\eta(\theta_1)\,\mathrm{d}\theta_1,
\end{aligned}$$

which completes the proof. □

The asymptotic behavior of $J^{\alpha}(\pi_2(\theta_2|\theta_1))$ is presented in the next theorem.

Theorem 2.2. *Let the prior density $\pi(\theta)$ be positive and three times continuously differentiable on a compact parameter space Θ. The corresponding compact parameter* *spaces of θ_1 and θ_2 are Θ_1 and Θ_2, respectively. Conditions (AI)-(AV) of Ghosh et al. (1982) are valid for $p(\boldsymbol{x}_n|\,\theta)$. Then*

$$\lim_{n\to\infty}\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_2}{2}}\{1-\alpha(1-\alpha)J^\alpha(\pi_2(\theta_2|\theta_1))\} =$$

$$(1-\alpha)^{-\frac{d_1}{2}}\int_{\Theta_1}\pi(\theta_1)\,\pi_2^{\alpha+1}(\theta_2|\theta_1)\,|\mathbf{I}_{22}(\theta)|^{-\frac{\alpha}{2}}\,\mathrm{d}\theta_2\mathrm{d}\theta_1. \quad (2.8)$$

Proof. The proof is similar to Theorem 2.1. First,

$$\begin{aligned}1-\alpha(1-\alpha)J^\alpha(\pi_2(\theta_2|\theta_1)) &= \int\int\int_\Theta \pi_2^\alpha(\theta_2|\theta_1)\,p^{1-\alpha}(\theta_2|\theta_1\boldsymbol{x}_n)\,\mathrm{d}\theta_2 p(\boldsymbol{x}_n,\theta_1)\,\mathrm{d}\theta_1\mathrm{d}\boldsymbol{x}_n \\ &= \int\int_\Theta p(\boldsymbol{x}_n,\theta)\left(\frac{\pi_2(\theta_2|\theta_1)}{p(\theta_2|\theta_1\boldsymbol{x}_n)}\right)^\alpha \mathrm{d}\theta\mathrm{d}\boldsymbol{x}_n \\ &= \int\int_\Theta p(\boldsymbol{x}_n,\theta)\left(\frac{p(\boldsymbol{x}_n|\theta_1)}{p(\boldsymbol{x}_n|\theta)}\right)^\alpha \mathrm{d}\theta\mathrm{d}\boldsymbol{x}_n \\ &= E\left[E\left[\left(\frac{p(\boldsymbol{x}_n|\theta_1)}{p(\boldsymbol{x}_n|\theta)}\right)^\alpha\middle|\,\theta\right]\right].\end{aligned}$$

The third equality follows from the definition of $p(\theta_2|\theta_1\boldsymbol{x}_n)$. Then, from (2.5) and the method of deriving (2.6) in Theorem 2.1, we have that

$$\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_2}{2}}\left(\frac{p(\boldsymbol{x}_n|\theta_1)}{p(\boldsymbol{x}_n|\theta)}\right)^\alpha \to |\mathbf{I}_{22}(\theta)|^{-\frac{\alpha}{2}}\,\pi_2^\alpha(\theta_2|\theta_1)\exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T A\boldsymbol{Z}\right),$$

in distribution as $n\to\infty$. Under the assumptions of this theorem, similarly to Theorem 2.1,

$$\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_2}{2}} E\left[\left(\frac{p(\boldsymbol{x}_n|\theta_1)}{p(\boldsymbol{x}_n|\theta)}\right)^\alpha\right] \to E_\theta\left[|\mathbf{I}_{22}(\theta)|^{-\frac{\alpha}{2}}\,\pi_2^\alpha(\theta_2|\theta_1)\,E\left[\exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T A\boldsymbol{Z}\right)|\theta\right]\right].$$

Under P_θ, $\boldsymbol{Z}^T A\boldsymbol{Z}$ follows $\chi^2_{d_2}$. Hence, $E\left[\exp\left(\frac{\alpha}{2}\boldsymbol{Z}^T A\boldsymbol{Z}\right)|\theta\right] = (1-\alpha)^{-\frac{d_2}{2}}$ by the moment generating function of the chi-squared distribution. Therefore,

$$\lim_{n\to\infty}\left(\frac{2\pi}{n}\right)^{-\frac{\alpha d_2}{2}}\left\{1-\alpha\left(1-\alpha\right)J^{\alpha}\left(\pi_2\left(\theta_2|\theta_1\right)\right)\right\}=$$

$$\left(1-\alpha\right)^{-\frac{d_2}{2}}\int_{\Theta_1}\pi_1\left(\theta_1\right)\int_{\Theta_2}\pi_2^{\alpha+1}\left(\theta_2|\theta_1\right)\left|\mathbf{I}_{22}\left(\theta\right)\right|^{-\frac{\alpha}{2}}\mathrm{d}\theta_2\mathrm{d}\theta_1.$$

□

We restrict α to less than 1 because $\left(1-\alpha\right)^{-\frac{d_1}{2}}$ may be complex when $\alpha>1$. Therefore, (2.2) and (2.8) do not hold when $\alpha>1$. In the proof of Theorem 2.1, we will show that the left hand side of (2.2) may not converge when $\alpha>1$.

Theorem 2.1 immediately implies that $J^{\alpha}\left(\pi\left(\theta_1\right)\right)$ can be well approximated by

$$\frac{1-\left(\frac{2\pi}{n}\right)^{\frac{\alpha d_1}{2}}\left(1-\alpha\right)^{-\frac{d_1}{2}}\int_{\Theta_1}\pi^{\alpha+1}\left(\theta_1\right)\eta\left(\theta_1\right)\mathrm{d}\theta_1}{\alpha\left(1-\alpha\right)},\tag{2.9}$$

when n is large. When $\alpha\to 0$, $J^{\alpha}\left(\pi_1\left(\theta_1\right)\right)$ converges to the expected Kullback-Leibler divergence which has been extensively studied. Ghosh and Mukerjee (1992) first gave the limiting behavior of $J^{0}\left(\pi_1\left(\theta_1\right)\right)$ as

$$\frac{d_1}{2}\log\left(\frac{n}{2\pi e}\right)+\int_{\Theta_1}\pi_1\left(\theta_1\right)\log\left\{\frac{\eta^{\star}\left(\theta_1\right)}{\pi_1\left(\theta_1\right)}\right\}\mathrm{d}\theta_1,\tag{2.10}$$

when $n\to\infty$, where

$$\eta^{\star}\left(\theta_1\right)=\exp\left\{\frac{1}{2}\int_{\Theta_2}\pi_2\left(\theta_2|\theta_1\right)\log\left(\frac{\left|\mathbf{I}\left(\theta\right)\right|}{\left|\mathbf{I}_{22}\left(\theta\right)\right|}\right)\theta_2\right\}.\tag{2.11}$$

It is easily seen that (2.9) converges to (2.10) when $\alpha\to 0$.

Likewise, the large sample approximation to $J^{\alpha}\left(\pi_2\left(\theta_2|\theta_1\right)\right)$ is

$$\frac{1-\left(\frac{2\pi}{n}\right)^{\frac{\alpha d_2}{2}}\left(1-\alpha\right)^{-\frac{d_2}{2}}\int\int_{\Theta}\pi\left(\theta_1\right)\pi_2^{\alpha+1}\left(\theta_2|\theta_1\right)\left|\mathbf{I}_{22}\left(\theta\right)\right|^{-\frac{\alpha}{2}}\mathrm{d}\theta_2\mathrm{d}\theta_1}{\alpha\left(1-\alpha\right)}.\tag{2.12}$$

Chapter 3

Derivation of alpha priors

In this chapter, we develop alpha priors in Section 3.1. We also consider alpha priors when parameters are assumed to be independent in Section 3.2. We will show that the independence assumption is used for a real example in Section 4.8. The alpha priors are generalized to multiple ordered group case in Section 3.3.

3.1 Alpha priors in the presence of nuisance parameters

In this section, we derive $\pi_1(\theta_1)$ and $\pi_2(\theta_2|\theta_1)$ and then propose an algorithm for the alpha prior based on these two priors.

3.1.1 Derivation of $\pi_1(\theta_1)$ and $\pi_2(\theta_2|\theta_1)$

We seek the marginal alpha prior of θ_1 by maximizing (2.9) with respect to $\pi_1(\theta_1)$. It is equivalent to minimizing

$$\int_{\Theta_1} \frac{\pi_1^{\alpha+1}(\theta_1)}{\alpha(1-\alpha)} \eta(\theta_1)\, d\theta_1, \tag{3.1}$$

over choices of $\pi_1(\theta_1)$. The following theorem gives the marginal alpha prior $\pi_1(\theta_1)$.

Theorem 3.1. *On a compact set of θ, Θ, the marginal alpha prior of θ_1 is of the form*

$$\pi_1(\theta_1) \propto \eta^{-\frac{1}{\alpha}}(\theta_1), \tag{3.2}$$

when $\alpha \in (-1,0) \cup (0,1)$.

Proof. Let

$$\int_{\Theta_1} \frac{\pi_1^{\alpha+1}(\theta_1)}{\alpha(1-\alpha)} \eta(\theta_1)\, \mathrm{d}\theta_1 = \int_{\Theta_1} f(\theta_1, \pi_1(\theta_1))\, \mathrm{d}\theta_1,$$

where

$$f(y,u) = \frac{u^{\alpha+1}}{\alpha(1-\alpha)} \eta(y).$$

With the constraint

$$\int_{\Theta_1} \pi_1(\theta_1)\, \mathrm{d}\theta_1 = 1,$$

our target is to minimize

$$\int_{\Theta_1} f(\theta_1, \pi_1(\theta_1))\, \mathrm{d}\theta_1 + \lambda \left(\int_{\Theta_1} \pi_1(\theta_1)\, \mathrm{d}\theta_1 - 1 \right), \tag{3.3}$$

where λ is a Lagrange multiplier. The corresponding Euler-Lagrange equation of (3.3) is

$$\frac{(\alpha+1)\,\pi_1^{\alpha}(\theta_1)}{\alpha(1-\alpha)} \eta(\theta_1) + \lambda = 0. \tag{3.4}$$

The unique solution of (3.4) is

$$\pi_1(\theta_1) = \frac{\eta^{-\frac{1}{\alpha}}(\theta_1)}{\int_{\Theta_1} \eta^{-\frac{1}{\alpha}}(\theta_1)\, \mathrm{d}\theta_1}, \tag{3.5}$$

and is equivalent to (3.2).

When $\alpha \in (-1,1)\backslash\{0\}$, $u \to f(y,u)$ is strictly convex for every fixed $y \in \Theta$. Accordingly, (3.5) is the unique minimizer for (3.1). When $\alpha < -1$, $u \to f(y,u)$ is strictly concave and (3.5) uniquely maximizes (3.1). The minimizers of (3.1) do not exist under this case. If we assume the existence of the minimizers, the minimizers of (3.1) coincide with (3.5) since any stationary point necessarily satisfies (3.4). This contradicts that (3.5) is the unique maximizer. When $\alpha = -1$, $\pi_1^{1+\alpha}(\theta_1)$ is constant. Hence, we cannot find the alpha prior in this case by our method and leave it as an open question for future research. In conclusion, $\pi_1(\theta_1)$ has the form (3.2) when $\alpha \in (-1, 0) \cup (0, 1)$. □

Sun and Berger (1998) gave a seemingly different result under the Kullback-Leibler divergence. Their marginal reference prior is proportional to (2.11). It is seen that $\eta^{-\frac{1}{\alpha}}(\theta_1)$ converges to (2.11) if α goes to zero. This consequence is consistent with the view that the Kullback-Leibler divergence is $D^0(p\|q)$. Liu et al. (2014) presented, without a rigorous argument, the marginal objective prior of θ_1 in a two-parameter case as

$$\pi_1(\theta_1) \propto \psi^{-\frac{1}{\alpha}}(\theta_1),$$

where $\psi(\theta_1) = \int (I^{11}(\theta))^{\alpha/2} \pi(\theta_2|\theta_1) \mathrm{d}\theta_2$ and $I^{11}(\theta)$ is the $(1,1)$-th element of $\mathbf{I}^{-1}(\theta)$. This objective prior is a special case of Theorem 3.1. It is seen that $\eta(\theta_1)$ coincides with $\psi(\theta_1)$ when $|\mathbf{I}_{22}(\theta)| / |\mathbf{I}(\theta)| = I^{11}(\theta)$ in the two-parameter case. However, this equality is not valid for the multi-parameter case.

The next theorem presents the conditional alpha prior $\pi_2(\theta_2|\theta_1)$.

Theorem 3.2. *On a compact set Θ of θ, the conditional alpha prior of θ_2 given θ_1 is of the form*

$$\pi_2(\theta_2|\theta_1) \propto |\mathbf{I}_{22}(\theta)|^{\frac{1}{2}},$$

when $\alpha \in (-1, 0) \cup (0, 1)$.

Proof. To maximize (2.12) with respect to $\pi_2(\theta_2|\theta_1)$, it is equivalent to minimize

$$\frac{1}{\alpha(1-\alpha)} \int \pi_2^{\alpha+1}(\theta_2|\theta_1) |\mathbf{I}_{22}(\theta)|^{-\frac{\alpha}{2}} \mathrm{d}\theta_2,$$

because $\pi(\theta_1)$ is fixed. Following the conclusion of Theorem 3.1, the maximizer of (2.12) is proportional to $\left[|\mathbf{I}_{22}(\theta)|^{-\frac{\alpha}{2}}\right]^{-\frac{1}{\alpha}}$ when $\alpha \in (-1, 0) \cup (0, 1)$. Hence,

$$\pi_2(\theta_2|\theta_1) \propto |\mathbf{I}_{22}(\theta)|^{\frac{1}{2}}.$$

□

Remark 3.1. *The conditional objective prior $\pi_2(\theta_2|\theta_1)$ is also known as the conditional Jeffrey prior which does not involve α.*

3.1.2 Algorithm for the alpha priors

After obtaining $\pi_1(\theta_1)$ and $\pi_2(\theta_2|\theta_1)$, we propose an algorithm to drive alpha priors. One common assumption in Theorems 3.1 and 3.2 is that the parameter space is compact. If the original parameter space is not compact (e.g., the set of real numbers $\mathbb{R}$) and $\pi(\theta_2|\theta_1)$ is improper, we may operate on a sequence of compact supports of the model parameters suggested by Berger and Bernardo (1989, 1992a). We propose the following algorithm for finding alpha priors.

Step 1. Choose a nested sequence $\Theta^1 \subset \Theta^2 \subset \cdots$ of compact subsets of the parameter space Θ of θ such that $\bigcup_{k=1}^{\infty} \Theta^k = \Theta$.

Step 2. Let $\pi_2(\theta_2|\theta_1)$ be proportional to $|\mathbf{I}_{22}(\theta)|^{\frac{1}{2}}$. Normalize $\pi_2(\theta_2|\theta_1)$ on each Θ_k, obtaining

$$\pi_2^k(\theta_2|\theta_1) = \frac{|\mathbf{I}_{22}(\theta)|^{\frac{1}{2}}}{\int_{\Theta^k} |\mathbf{I}_{22}(\theta)|^{\frac{1}{2}}\, \mathrm{d}\theta_2}.$$

Step 3. Define the marginal alpha prior $\pi_1^k(\theta_1)$ by

$$\pi_1^k(\theta_1) = \left\{ \int_{\Theta^k} \pi_2^k(\theta_2|\theta_1) \left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{22}(\theta)|} \right)^{-\frac{\alpha}{2}} \mathrm{d}\theta_2 \right\}^{-\frac{1}{\alpha}}. \tag{3.6}$$

Step 4. Let $\pi^k(\theta) = \pi_1^k(\theta_1)\,\pi_2^k(\theta_2|\theta_1)$. We then determine the alpha prior $\pi(\theta)$ by

$$\pi(\theta) = \lim_{k\to\infty} \frac{\pi^k(\theta)}{\pi^k(\theta^\star)},$$

where $\theta^\star$ is any fixed point in Θ.

If $\pi(\theta_2|\theta_1)$ is proper, we can directly obtain $\pi(\theta_2|\theta_1)$ from Theorem 3.2 and substitute it into (3.2) to obtain the marginal alpha prior $\pi(\theta_1)$. Note that $\pi(\theta_2|\theta_1)$ must be in its normalized form. As a result,

$$\pi(\theta) \propto |\mathbf{I}_{22}(\theta)|^{\frac{1}{2}} \left\{ \int_\Theta \pi(\theta_2|\theta_1) \left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{22}(\theta)|} \right)^{-\frac{\alpha}{2}} \mathrm{d}\theta_2 \right\}^{-\frac{1}{\alpha}}.$$

The general form of alpha priors depends on the value of α. However, an alpha prior possibly reduces to a form independent of α for a particular problem. The following theorem provides a sufficient condition that alpha priors are independent of α.

Theorem 3.3. *If (a) θ_1 and θ_2 are variation independent (i.e. the parameter space of θ_2 does not depend on that of θ_1) and (b) both $(|\mathbf{I}(\theta)|\,/\,|\mathbf{I}_{22}(\theta)|)^{\frac{1}{2}}$ and $|\mathbf{I}_{22}(\theta)|^{\frac{1}{2}}$ can be factorized as*

$$(|\mathbf{I}(\theta)|\,/\,|\mathbf{I}_{22}(\theta)|)^{\frac{1}{2}} = g_1(\theta_1)\,g_2(\theta_2), \qquad |\mathbf{I}_{22}(\theta)|^{\frac{1}{2}} = h_1(\theta_1)\,h_2(\theta_2),$$

then the alpha prior $\pi(\theta) \propto g_1(\theta_1)\,h_2(\theta_2)$.

Proof. Under the assumptions of this theorem, the conditional alpha prior

$$\pi_2^k(\theta_2|\theta_1) = \frac{h_1(\theta_1)\,h_2(\theta_2)}{\int_{\Theta^k} h_1(\theta_1)\,h_2(\theta_2)\,\mathrm{d}\theta_2} = \frac{h_2(\theta_2)}{\int_{\Theta^k} h_2(\theta_2)\,\mathrm{d}\theta_2},$$

which does not depend on θ_1 as θ_1 and θ_2 are variation independent. Equation (3.6) then becomes

$$\pi_1^k(\theta_1) = \left\{ \int_{\Theta^k} \frac{h_2(\theta_2)}{\int_{\Theta^k} h_2(\theta_2)\,\mathrm{d}\theta_2} \left(g_1(\theta_1)\, g_2(\theta_2)\right)^{-\alpha} \mathrm{d}\theta_2 \right\}^{-\frac{1}{\alpha}}$$

$$= g_1(\theta_1) \left\{ \int_{\Theta^k} \frac{h_2(\theta_2)}{\int_{\Theta^k} h_2(\theta_2)\,\mathrm{d}\theta_2} \left(g_2(\theta_2)\right)^{-\alpha} \mathrm{d}\theta_2 \right\}^{-\frac{1}{\alpha}}. \tag{3.7}$$

Since θ_1 and θ_2 are variation independent, Equation (3.7) reduces to $\pi_1^k(\theta_1) = c g_1(\theta_1)$ where c is some constant. Finally,

$$\frac{\pi^k(\theta)}{\pi^k(\theta^\star)} = \frac{h_2(\theta_2)\, g_1(\theta_1)}{h_2(\theta_2^\star)\, g_1(\theta_1^\star)},$$

which does not involve Θ^k. Therefore,

$$\pi(\theta) = \lim_{k\to\infty} \frac{\pi^k(\theta)}{\pi^k(\theta^\star)} \propto g_1(\theta_1)\, h_2(\theta_2).$$

□

3.2 Independent alpha priors

Sometimes, independence between model parameters is assumed for some problems (see Section 4.8). Sun and Berger (1998) proposed an iterative algorithm to derive Berger-Bernardo's reference prior under the independence assumption between model parameters. The algorithm for deriving independent alpha priors is proposed as follows.

Step 0. Choose any initial nonzero marginal prior density for θ_2, $\pi_2^{(0)}(\theta_2)$.

Step 1. Define an interim prior density for θ_1 by

$$\pi_1^{(1)}(\theta_1) \propto \left\{ \int \left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{22}(\theta)|} \right)^{-\frac{\alpha}{2}} \pi_2^{(0)}(\theta_2)\,\mathrm{d}\theta_2 \right\}^{-1/\alpha}.$$

Step 2. Define an interim prior density for θ_2 by

$$\pi_2^{(1)}(\theta_2) \propto \left\{ \int \left(\frac{|\mathbf{I}(\theta)|}{|\mathbf{I}_{11}(\theta)|} \right)^{-\frac{\alpha}{2}} \pi_1^{(1)}(\theta_1)\, \mathrm{d}\theta_1 \right\}^{-1/\alpha}.$$

Next, replace $\pi_2^{(0)}$ in Step 0 by $\pi_2^{(1)}$ and repeat Step 1 and Step 2 to obtain $\pi_1^{(2)}$ and $\pi_2^{(2)}$. Finally, we have two sequences $\left\{\pi_1^{(i)}\right\}_{i\geqslant 1}$ and $\left\{\pi_2^{(i)}\right\}_{i\geqslant 1}$. Thus, the independent marginal alpha priors are the limits $\lim_{i\to\infty} \pi_j^{(i)}(\theta_j)(j=1,2)$. It is possibly necessary to operate on compact sets for the parametric space if the interim priors chosen are improper. Sun and Berger (1998) pointed that the sequences of $\pi_j^{(i)}(\theta_j)\,(j=1,2)$ may not converge sometimes. However, we can found some sufficient condition that ensures the convergence of the sequences and yields independent alpha priors without the direct iterations.

Proposition 3.1. *If* $|\mathbf{I}(\theta)| / |\mathbf{I}_{22}(\theta)|$ *does not depend on* θ_2*, then the two independent marginal alpha priors are*

$$\pi_1(\theta_1) \propto \left(|\mathbf{I}(\theta)| / |\mathbf{I}_{22}(\theta)|\right)^{\frac{1}{2}},$$

$$\pi_2(\theta_2) \propto \left\{ \int \pi_1(\theta_1) \left(|\mathbf{I}(\theta)| / |\mathbf{I}_{11}(\theta)|\right)^{-\frac{\alpha}{2}} \mathrm{d}\theta_1 \right\}^{-\frac{1}{\alpha}}.$$

Proof. If $|\mathbf{I}(\theta)| / |\mathbf{I}_{22}(\theta)|$ is independent of θ_2, $\pi_1(\theta_1)$ follows from the Step 1 of the algorithm for independent alpha priors. Then, $\pi_2(\theta_2)$ is obtained from one iteration of the algorithm. □

3.3 Ordered group alpha priors

In this section, we extend the algorithm in Section 3.1 to the ordered group case. The algorithm for ordered group alpha priors is inspired by Berger and Bernardo

(1992a). Instead of dividing parameters in two groups as parameters of interests and nuisance parameters, parameters are separated into m groups of sizes $d_1, \ldots, d_m$. Let $\theta = \left(\theta_{(1)}, \ldots, \theta_{(m)}\right)$ be the m groups of parameters and the size of each group $\theta_{(i)} = \left(\theta_{i_1}, \ldots \theta_{i_{d_i}}\right)$ is d_i. We also define

$$\theta_{[i]} = \left(\theta_{(1)}, \ldots, \theta_{(i)}\right), \qquad \theta_{[-i]} = \left(\theta_{(i+1)}, \ldots, \theta_{(m)}\right),$$

for $i = 1, 2, \ldots, m$. Let $S(\theta)$ be the inverse of the Fisher information matrix and write it as

$$S(\theta) = I^{-1}(\theta) = \begin{pmatrix} S_{11} & S_{21}^T & \ldots & S_{m1}^T \\ S_{21} & S_{22} & \ldots & S_{m2}^T \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \ldots & S_{mm} \end{pmatrix}$$

such that S_{ij} is the corresponding decomposition for $\theta_{(i)}$ and $\theta_{(j)}$. Define

$$S_i = \begin{pmatrix} S_{11} & S_{21}^T & \ldots & S_{i1}^T \\ S_{21} & S_{22} & \ldots & S_{i2}^T \\ \vdots & \vdots & \ddots & \vdots \\ S_{i1} & S_{i2} & \ldots & S_{ii} \end{pmatrix}, \qquad H_i = S_i^{-1} = \begin{pmatrix} H_{11} & H_{21}^T & \ldots & H_{i1}^T \\ H_{21} & H_{22} & \ldots & H_{i2}^T \\ \vdots & \vdots & \ddots & \vdots \\ H_{i1} & H_{i2} & \ldots & H_{ii} \end{pmatrix},$$

where H_{ij} is a $d_i \times d_j$ matrix and $h_i = H_{ii}$, for $i = 1, 2, \ldots, m$. We follow the same notation of Θ^k in Section 3.1 and define

$$\Theta^k\left(\theta_{[i]}\right) = \left\{\theta_{(i+1)} : \left(\theta_{[i]}, \theta_{(i+1)}, \theta^\star_{[-(i+1)]}\right) \in \Theta^k\right\} \text{ for some fixed } \theta^\star_{[-(i+1)]}.$$

An indicator function is denoted by

$$\mathbf{1}_\Omega(x) = \begin{cases} 1 & x \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

The algorithm for ordered group alpha priors is as follows:

Start. We define

$$\pi_m^k\left(\theta_{[-(m-1)]} | \theta_{[m-1]}\right) = \frac{|h_m(\theta)|^{\frac{1}{2}} \mathbf{1}_{\Theta^k\left(\theta_{[m-1]}\right)}\left(\theta_{(m)}\right)}{\int |h_m(\theta)|^{\frac{1}{2}} \mathbf{1}_{\Theta^k\left(\theta_{[m-1]}\right)}\left(\theta_{(m)}\right) \mathrm{d}\theta_{(m)}}.$$

Iteration. For $i = m-1, m-2, \ldots, 1$, define

$$\pi_i^k\left(\theta_{[-(i-1)]}|\theta_{[i-1]}\right) = \frac{\pi_{i+1}^k\left(\theta_{[-i]}|\theta_{[i]}\right)\left\{E_i^k\left[\left|h_i\left(\theta\right)\right|^{-\frac{\alpha}{2}}|\theta_{[i]}\right]\right\}^{-\frac{1}{\alpha}}\mathbf{1}_{\Theta^k\left(\theta_{[i-1]}\right)}\left(\theta_{(i)}\right)}{\int\left\{E_i^k\left[\left|h_i\left(\theta\right)\right|^{-\frac{\alpha}{2}}|\theta_{[i]}\right]\right\}^{-\frac{1}{\alpha}}\mathbf{1}_{\Theta^k\left(\theta_{[i-1]}\right)}\left(\theta_{(i)}\right)\mathrm{d}\theta_{(i)}}. \tag{3.8}$$

Here,

$$E_i^k\left[\left|h_i\left(\theta\right)\right|^{-\frac{\alpha}{2}}|\theta_{[i]}\right] = \int\left|h_i\left(\theta\right)\right|^{-\frac{\alpha}{2}}\pi_{i+1}^k\left(\theta_{[-i]}|\theta_{[i]}\right)\mathrm{d}\theta_{[-i]}, \tag{3.9}$$

where the integral is over the range $\left\{\theta_{[-i]} : \left(\theta_{[i]}, \theta_{[-i]}\right) \in \Theta^k\right\}$. For $i = 1$, one interprets $\theta_{[-0]}$ as θ and $\theta_{[0]}$ as vacuous, and writes $\pi^k\left(\theta\right) = \pi^k\left(\theta_{[-0]}|\theta_{[0]}\right)$.

Finish. The ordered group alpha prior is

$$\pi\left(\theta\right) = \lim_{k\to\infty}\frac{\pi^k\left(\theta\right)}{\pi^k\left(\theta^\star\right)},$$

assuming the limit exists, where $\theta^\star$ is some fixed point in Θ.

Chapter 4

Examples

In this chapter, we consider some examples of alpha priors. The first example concerns the location-scale family and the log-location-scale family. The second example is from the two-parameter exponential family. The third example considers a multiparameter case, the bivariate normal distribution. We find that the alpha priors for these four families reduce to Bernardo-Berger's reference prior. The fourth example is the Neyman-Scott problem. We show that the alpha prior for the Neyman-Scott problem leads to a consistent estimate for the parameter of interest. The fifth example is the beta distribution whose alpha prior is different from Bernardo-Berger's reference prior and depends on the value of α. The sixth example is the Dirichlet distribution which is a multivariate generalization of the beta distribution. We use the Dirichlet distribution to illustrate ordered group alpha priors. The seventh example considers a two-parameter logistic regression. A simulation study is carried out based on the two-parameter logistic regression model to compare alpha priors under different values of α. The last example comes from a clinical trial about ECMO which is a real application of the two-parameter logistic regression. With the real data, we illustrate the influence of the value of α on the posterior distribution of the parameter of interest.

4.1 The location-scale family and the log-location-scale family

The location-scale family and the log-location-scale family are widely used in statistical inference. These two families include many important distributions, for instance, the normal, lognormal and Weibull distributions. The PDF of a location-scale distribution is defined as

$$p(x|\mu,\sigma) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right), \quad x,\sigma > 0$$

where $\phi(\cdot)$ is the derivative of the cumulative density function of a standard distribution $\Phi(\cdot)$. Here, μ is the location parameter and σ is the scale parameter. Similarly, the PDF of a log-location-scale distribution is

$$p(x|\mu,\sigma) = \frac{1}{x\sigma}\phi\left(\frac{\log(x)-\mu}{\sigma}\right).$$

The Fisher information matrices of the two distributions are of the same form

$$\mathbf{I}(\mu,\sigma) = \frac{1}{\sigma^2}\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix},$$

where C_{ij} $(i,j=1,2)$ are constant independent of model parameters. It is seen that $|\mathbf{I}(\theta)| / |\mathbf{I}_{22}(\theta)| = \sigma^{-2}(C_{11} - C_{12}C_{21}/C_{22})$.

In many applications, the location parameter is of interest. For example, in medical or industrial experiments, the location parameter represents the expected life of an experiment unit. Thus, it attracts more attention than the scale parameter does. The next proposition presents the alpha prior of the location-scale family and the log-location-scale family if the location parameter is of interest.

Proposition 4.1. *If μ is the parameter of interest, the alpha prior of the location-scale family and the log-location-scale family is $\pi(\mu,\sigma) \propto \sigma^{-1}$.*

Proof. The alpha prior immediately follows Theorem 3.3 as the Fisher Information matrix does not involve μ. □

4.2 Two-parameter exponential family

Bar-Lev and Reiser (1982) proposed the following two-parameter exponential family density

$$p\left(x|\theta_1,\theta_2\right)=a\left(x\right)\exp\left\{\theta_1U_1\left(x\right)-\theta_1G_2'\left(\theta_2\right)U_2\left(x\right)-\psi\left(\theta_1,\theta_2\right)\right\},$$

where $\theta_1<0$, $\theta_2=E_{(\theta_1,\theta_2)}\left[U_2\left(X\right)\right]$, $U_2\left(\cdot\right)$ is a one-to-one function on the support of the density, the $G_i\left(\cdot\right)$s, are infinitely differentiable functions with $G_i''\left(\cdot\right)>0$, and $\psi\left(\theta_1,\theta_2\right)=-\theta_1\left\{\theta_2G_2'\left(\theta_2\right)-G_2\left(\theta_2\right)\right\}+G_1\left(\theta_2\right)$. This class of distributions includes many popular distributions such as the normal, inverse normal, gamma, and inverse gamma distribution. Table 4.1 connects the general two-parameter exponential family to some special distributions.

	$G_1\left(\theta_1\right)$	$G_2\left(\theta_2\right)$	$U_1\left(x\right)$	$U_2\left(x\right)$	θ_1	θ_2
Normal (μ,σ)	$-\frac{1}{2}\log\left(-2\theta_1\right)$	θ_2^2	x^2	x	$-1/\left(2\sigma^2\right)$	μ
Inverse Gaussian	$-\frac{1}{2}\log\left(-2\theta_1\right)$	$1/\theta_2$	$1/x$	x	$-\alpha/2$	$\sqrt{\alpha/\mu}$
Gamma	$h\left(\theta_1\right)$	$-\log\theta_2$	$-\log x$	x	$-\alpha$	μ
Inverse Gamma	$h\left(\theta_1\right)$	$-\log\theta_2$	$\log x$	$1/x$	$-\alpha$	μ

Table 4.1: Special cases for the two-parameter exponential family

According to Berger et al. (2015), the Fisher information matrix of $\left(\theta_1,\theta_2\right)$ is

$$\mathbf{I}\left(\theta_1,\theta_2\right)=\begin{pmatrix}G_1''\left(\theta_1\right) & 0\\ 0 & -\theta_1G_2''\left(\theta_2\right)\end{pmatrix}.$$

The following proposition presents the alpha priors for the two-parameter exponential family.

Proposition 4.2. *If θ_1 or θ_2 is the parameter of interest, the alpha prior of the two-parameter exponential family is in the form*

$$\pi(\theta_1, \theta_2) \propto \sqrt{G_1''(\theta_1) G_2''(\theta_2)}.$$

Proof. The alpha prior follows Theorem 3.3 because $\sqrt{-\theta_1 G_2''(\theta_2)}$ can be factorized as $\sqrt{-\theta_1}\sqrt{G_2''(\theta_2)}$. □

4.3 Bivariate normal distribution

Consider that (X_1, X_2) follows a bivariate normal distribution with mean $\theta_1 = (\mu_1, \mu_2)$ and covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where ρ is the correlation between X_1 and X_2. Let $\theta_2 = (\sigma_1, \sigma_2, \rho)$. The Fisher information matrix is

$$\mathbf{I}(\theta_1, \theta_2) = \begin{pmatrix} \mathbf{I}_{11}(\theta_2) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{22}(\theta_2) \end{pmatrix},$$

where

$$\mathbf{I}_{11}(\theta_2) = \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} / \left(1 - \rho^2\right),$$

and

$$\mathbf{I}_{22}(\theta_2) = \begin{pmatrix} \frac{2-\rho^2}{\sigma_1^2} & -\frac{\rho^2}{\sigma_1\sigma_2} & -\frac{\rho}{\sigma_1} \\ -\frac{\rho^2}{\sigma_1\sigma_2} & \frac{2-\rho^2}{\sigma_2^2} & -\frac{\rho}{\sigma_2} \\ -\frac{\rho}{\sigma_1} & -\frac{\rho}{\sigma_2} & \frac{1+\rho^2}{1-\rho^2} \end{pmatrix} / \left(1 - \rho^2\right).$$

Proposition 4.3. *If θ_1 is the parameter of interest, then the alpha prior has the form*

$$\pi(\theta_1, \theta_2) \propto 1 / \left(\sigma_1\sigma_2\left(1 - \rho^2\right)^{\frac{3}{2}}\right).$$

Proof. Clearly, $|\mathbf{I}(\theta)|/|\mathbf{I}_{22}(\theta_2)| = 1/(\sigma_1^2\sigma_2^2(1-\rho^2))$ and $|\mathbf{I}_{22}(\theta_2)| = 4/\left(\sigma_1^2\sigma_2^2(1-\rho^2)^3\right)$ are independent of θ_1. Thus, the result follows Theorem 3.3. □

Remark 4.1. *The prior in Proposition 4.3 is the independent Jeffreys prior as stated in Berger and Sun (2008).*

4.4 Neyman-Scott problem

Consider the Neyman-Scott problem (Neyman and Scott, 1958) that x_{ij} $(i = 1, \ldots, n; j = 1, 2)$ are independent observations where x_{ij} follows a normal distribution with mean μ_i and variance σ^2. The Fisher information matrix for $(\mu_1, \ldots, \mu_n, \sigma^2)$ is given by

$$\mathbf{I}\left(\mu_1, \ldots, \mu_n, \sigma^2\right) = \text{diag}\left(2\sigma^{-2}, \ldots, 2\sigma^{-2}, n\sigma^{-4}\right).$$

In Neyman-Scott problem, σ^2 or σ is the parameter of interest. We summarize the conclusions on the alpha prior distribution in the next proposition.

Proposition 4.4. *If σ^2 is the parameter of interest, the alpha prior has the form*

$$\pi\left(\mu_1 \ldots, \mu_n, \sigma^2\right) \propto \sigma^{-2},$$

or equivalently,

$$\pi\left(\mu_1 \ldots, \mu_n, \sigma\right) \propto \sigma^{-1}.$$

Proof. Note that

$$\mathbf{I}_{22}\left(\mu_1, \ldots, \mu_n\right) = \text{diag}\left(2\sigma^{-2}, \ldots, 2\sigma^{-2}\right),$$

and $|\mathbf{I}(\theta)|/|\mathbf{I}_{22}(\theta)| = n\sigma^{-4}$, independent of $(\mu_1, \ldots, \mu_n)$. According to Theorem 3.3, we have the first conclusion. The second result follows simple reparametrization. □

The Jeffreys prior $\pi^J(\mu_1 \ldots, \mu_n, \sigma) \propto \sigma^{-(n+1)}$ leads to a posterior mean of σ^2 $E(\sigma^2|x) = s^2/(2n-2)$, where s^2 is the sample variance. This is an inconsistent estimation for σ^2. The alpha prior is the same as the reference prior presented in Bernardo (1979); Berger and Bernardo (1992a) which leads to a consistent posterior mean for σ^2.

4.5 Beta distribution

Consider the PDF of the beta distribution

$$p\left(x|\gamma,\beta\right)=\frac{\Gamma\left(\gamma+\beta\right)}{\Gamma\left(\gamma\right)\Gamma\left(\beta\right)}x^{\gamma-1}\left(1-x\right)^{\beta-1},\quad 0<x<1,$$

where $\gamma>0$ and $\beta>0$. The Fisher information matrix for (γ,β) is

$$I\left(\gamma,\beta\right)=\begin{pmatrix}\psi'\left(\gamma\right)-\psi'\left(\gamma+\beta\right) & -\psi'\left(\gamma+\beta\right)\\ -\psi'\left(\gamma+\beta\right) & \psi'\left(\beta\right)-\psi'\left(\gamma+\beta\right)\end{pmatrix}$$

where $\psi'(\beta)=(\mathrm{d}^2/\mathrm{d}\beta^2)\log\{\Gamma(\beta)\}$ is the trigamma function. If γ is the parameter of interest, we note that the Fisher information matrix does not satisfy the conditions in Theorem 3.3. This implies that the alpha prior of the beta distribution involves α. The following proposition gives the closed-form expression of the alpha prior.

Proposition 4.5. *A sequence of sets* Θ^k *for* (γ,β) *is chosen as* $(l_k,u_k)\times(l_k,u_k)$ *where* $l_k=1/u_k$ *and* $u_k\to\infty$ *as* $k\to\infty$*. Then, the alpha prior of the beta distribution when* β *is the nuisance parameter is*

$$\pi\left(\gamma,\beta\right)\propto\left\{\psi'\left(\gamma\right)\gamma-1\right\}^{\frac{1}{2}}\left\{\frac{\gamma^{\frac{1+\alpha}{2}}}{\sqrt{\gamma}+1}\right\}^{-\frac{1}{\alpha}}\frac{\sqrt{\psi'\left(\beta\right)-\psi'\left(\gamma+\beta\right)}}{\sqrt{\gamma}+1},\tag{4.1}$$

for $\alpha\in(-1,1)\backslash\{0\}$.

Proof. Sun and Berger (1998) showed that

$$\pi^k(\beta|\gamma)=\frac{\left\{\psi'\left(\beta\right)-\psi'\left(\gamma+\beta\right)\right\}^{1/2}}{\sqrt{\gamma}\log\left(u_k\right)-\log\left(l_k\right)+O(1)},\tag{4.2}$$

for any sufficiently small $\epsilon>0$ and k large enough such that $l_k<\epsilon<1/\epsilon<u_k$. Here, $f(x)=O(g(x))$ means there exist two positive constants M and δ such that

$|f(x)| \leqslant M|g(x)|$ for $|x| \leqslant \delta$ and particularly $O(1)$ represents some bounded function of l_k and u_k. The marginal alpha prior of γ, $\pi^k(\gamma)$, has the form

$$\pi^k(\gamma) = \left\{ \int_{l_k}^{u_k} \left\{ \frac{|I(\gamma,\beta)|}{|I_{22}(\gamma,\beta)|} \right\}^{-\frac{\alpha}{2}} \frac{\{\psi'(\beta) - \psi'(\gamma+\beta)\}^{1/2}}{\sqrt{\gamma}\log(u_k) - \log(l_k) + O(1)} \mathrm{d}\beta \right\}^{-\frac{1}{\alpha}}, \tag{4.3}$$

where

$$\frac{|I(\gamma,\beta)|}{|I_{22}(\gamma,\beta)|} = \frac{\psi'(\gamma)\psi'(\beta) - \psi'(\gamma)\psi'(\gamma+\beta) - \psi'(\gamma+\beta)\psi'(\beta)}{\psi'(\beta) - \psi'(\gamma+\beta)}.$$

In the Appendix, we show that

$$\pi^k(\gamma) = \left\{ \frac{\{\psi'(\gamma)\gamma - 1\}^{-\frac{\alpha}{2}} \gamma^{\frac{1+\alpha}{2}} \log(u_k) + O(\varepsilon)}{\sqrt{\gamma}\log(u_k) - \log(l_k) + O(1)} \right\}^{-\frac{1}{\alpha}}, \tag{4.4}$$

for any $\alpha \in (-1,1)\backslash\{0\}$. Since $l_k = 1/u_k$,

$$\lim_{k\to\infty} \frac{\log(l_k)}{\log(u_k)} = -1.$$

Hence, the limit $\lim_{k\to\infty} \pi^k(\gamma,\beta)/\pi^k(\gamma_0,\beta_0)$ is (4.1). □

It is also of interest to investigate how the choice of the sequence of sets Θ^k influences the form of objective priors. We choose $l_k = 1/u_k$ as it is natural for positive parameters and has been used by many authors such as Sun and Berger (1998) and Bernardo (2005). If we choose some unusual compact sets, the final results will be different. For example, if $l_k = e^{-u_k}$ and $u_k \to \infty$ as $k \to \infty$, then

$$\pi(\gamma,\beta) \propto \{\psi'(\gamma)\gamma - 1\}^{\frac{1}{2}} \gamma^{-\frac{1+\alpha}{2\alpha}} \sqrt{\psi'(\beta) - \psi'(\gamma+\beta)},$$

for $\alpha \in (-1,1)\backslash\{0\}$.

4.6 Dirichlet distribution

Consider the PDF of the Dirichlet distribution

$$p\left(x_1,\ldots,x_p|\beta_1,\ldots,\beta_p\right)=\frac{\Gamma\left(\sum_{i=1}^{p}\beta_i\right)}{\prod_{i=1}^{p}\Gamma(\beta_i)}\prod_{i=1}^{p}x_i^{\beta_i-1}.$$

The Fisher information matrix is

$$\mathbf{I}\left(\beta_1,\ldots,\beta_p\right)=\begin{pmatrix}\psi'\left(\beta_1\right)-\psi'\left(\beta_0\right) & -\psi'\left(\beta_0\right) & \ldots & -\psi'\left(\beta_0\right)\\ -\psi'\left(\beta_0\right) & \psi'\left(\beta_2\right)-\psi'\left(\beta_0\right) & & \vdots\\ \vdots & & \ddots & -\psi'\left(\beta_0\right)\\ -\psi'\left(\beta_0\right) & \ldots & -\psi'\left(\beta_0\right) & \psi'\left(\beta_p\right)-\psi'\left(\beta_0\right)\end{pmatrix}.$$

where $\beta_0=\sum_{i=1}^{p}\beta_i$. Define

$$B\left(\Omega\right)=\psi'(\beta_0)\prod_{j\in\Omega}\psi'(\beta_j)\left(-\prod_{j\in\Omega}\frac{1}{\psi'(\beta_j)}+\frac{1}{\psi'(\beta_0)}\right),$$

where Ω is a subset of $\{1,2,\ldots,p\}$. The determinant of the Fisher information matrix is $\left|\mathbf{I}\left(\beta_1,\ldots,\beta_p\right)\right|=B\left(\{1,2,\ldots,p\}\right)$. We consider a p-group ordering $(\beta_1,\ldots,\beta_p)$. After some tedious calculation, $S=\{S_{ij}\}$ where

$$S_{ij}=\begin{cases}\frac{B(\{1,2,\ldots,p\}\backslash\{i\})}{B(\{1,2,\ldots,p\})} & i=j,\\ \frac{\prod_{k=0}^{p}\psi'(\beta_j)}{\psi'(\beta_i)\psi'(\beta_j)B(\{1,2,\ldots,p\})} & i\neq j,\end{cases}$$

for $i,j=1,2,\ldots,p$. Hence,

$$h_i=\begin{cases}\frac{B(\{i,i+1,\ldots,p\})}{B(\{i+1,\ldots,p\})}=\psi'\left(p\right)-\frac{\prod_{j=i+1}^{p}\psi'(\beta_j)}{B(\{i+1,\ldots,p\})} & 1\leqslant i\leqslant p-1,\\ B\left(\{p\}\right) & i=p.\end{cases}$$

The derivation of the p-group alpha prior is intractable for large p. For illustrative purposes, we consider the case of $p=3$. A sequence of sets Θ^k for $(\beta_1,\beta_2,\beta_3)$ is chosen as $(l_{1k},u_{1k})\times(l_{2k},u_{2k})\times(l_{3k},u_{3k})$ and $u_{jk}\to\infty$, $l_{jk}\to 0$ for $j=1,2,3$. Since the detailed derivation of the alpha prior of the Dirichlet distribution is complicated, we only present the expressions of $\pi_3^k(\beta_3|\beta_1,\beta_2)$, $\pi_2^k(\beta_2,\beta_3|\beta_1)$ and $\pi_1^k(\beta_1,\beta_2,\beta_3)$ for sufficiently large k and sufficiently small $\epsilon > 0$ such that $l_{2k}, l_{3k} < \epsilon < 1/\epsilon < u_{2k}, u_{3k}$. Based on these asymptotic expressions, we sketch out the derivation of the alpha prior. The complete derivation is in the Appendix.

The conditional alpha prior $\pi_3^k(\beta_3|\beta_1,\beta_2)$ is

$$\pi_3^k(\beta_3|\beta_1,\beta_2) = \frac{\{\psi'(\beta_3) - \psi'(\beta_1+\beta_2+\beta_3)\}^{\frac{1}{2}}}{\sqrt{\beta_1+\beta_2}\log(u_{3k}) - \log(l_{3k}) + C_3(\varepsilon,\beta_1,\beta_2)},$$

where

$$C_3(\varepsilon,\beta_1,\beta_2) = \int_{\varepsilon}^{1/\varepsilon} \{h_3(\beta)\}^{\frac{1}{2}}\,\mathrm{d}\beta_3 + O(\varepsilon),$$

can be regarded as a boundary function of l_{2k}, u_{2k}, l_{3k} and u_{3k}.

Next, we obtain $\pi_2^k(\beta_2,\beta_3|\beta_1)$ as

$$\begin{aligned}\pi_2^k(\beta_2,\beta_3|\beta_1) = \pi_3^k(\beta_3|\beta_1,\beta_2)\Big\{&\{\psi'(\beta_2)(\beta_1+\beta_2)-1\}^{-\frac{\alpha}{2}}\{\beta_1+\beta_2\}^{\frac{1+\alpha}{2}}\log(u_{3k}) \\ &- \{\psi'(\beta_2)-\psi'(\beta_1+\beta_2)\}^{-\frac{\alpha}{2}}\log(l_{3k}) + C_{21}(\varepsilon,\beta_1,\beta_2)\Big\}^{-\frac{1}{\alpha}} \\ &\times\Big\{\sqrt{\beta_1+\beta_2}\log(u_{3k}) - \log(l_{3k}) + C_3(\varepsilon,\beta_1,\beta_2)\Big\}^{\frac{1}{\alpha}} \\ &\times\left\{\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\log(u_{2k}) - \log(l_{2k}) + C_{22}(\varepsilon,\beta_1)\right\}^{-1}\end{aligned}$$

where $C_{21}(\varepsilon,\beta_1,\beta_2)$, and $C_{22}(\varepsilon,\beta_1)$ are defined in the Appendix.

Finally, we have

$$\pi_1^k(\beta_1,\beta_2,\beta_3) \propto \pi_2^k(\beta_2,\beta_3|\beta_1)\left\{\left\{\psi'(\beta_1)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\log(u_{2k})+C_{12}^k(\varepsilon,\beta_1)\right.$$

$$\left.-\left\{\psi'(\beta_1)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\sqrt{\beta_1}\log(u_{3k})\left\{\sqrt{\beta_1}\log(u_{3k})-\log(l_{3k})\right\}^{-1}\log(l_{2k})\right\}^{-\frac{1}{\alpha}}$$

$$\left\{\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\log(u_{2k})-\log(l_{2k})+C_{22}(\varepsilon,\beta_1)\right\}^{\frac{1}{\alpha}}$$

where $C_{12}^k(\varepsilon,\beta_1)$ is defined in the Appendix. If we choose that $l_{3k}=1/u_{3k}$ and $l_{2k}=1/u_{2k}$, the p-group alpha prior has the form

$$\pi(\beta_1,\beta_2,\beta_3)$$

$$=\lim_{k\to\infty}\frac{\pi_1^k(\beta_1,\beta_2,\beta_3)}{\pi_1^k(1,1,1)}$$

$$\propto\left\{\left\{\sqrt{\beta_1+\beta_2}+1\right\}\left\{1+\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\right\}\right\}^{\frac{1-\alpha}{\alpha}}\left\{\psi'(\beta_3)-\psi'(\beta_1+\beta_2+\beta_3)\right\}^{\frac{1}{2}}$$

$$\times\left\{\left\{\psi'(\beta_2)(\beta_1+\beta_2)-1\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1+\alpha}{2}}+\left\{\psi'(\beta_2)-\psi'(\beta_1+\beta_2)\right\}^{-\frac{\alpha}{2}}\right\}^{-\frac{1}{\alpha}}$$

$$\times\left\{\left\{\psi'(\beta_1)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}+\left\{\psi'(\beta_1)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\right\}^{-\frac{1}{\alpha}}.$$

4.7 Two-parameter logistic regression

The binary response models are widely used in bioassay experiments and reliability tests. In this section, we consider a two-parameter logistic regression model for the binary response model.

Suppose $\boldsymbol{x}=(x_1,\dots x_{n_1})$ are i.i.d. Bernoulli observations whose probability of success is $p_1=e^{\beta_1}/(1+e^{\beta_1})$ and $\boldsymbol{y}=(y_1,\dots y_{n_2})$ is another group of i.i.d. Bernoulli observations whose probability of success is $p_2=e^{\beta_1+\beta_2}/(1+e^{\beta_1+\beta_2})$.

The likelihood function of $\beta = (\beta_1, \beta_2)$ is

$$L(\beta|\boldsymbol{x}, \boldsymbol{y}) = \frac{e^{\beta_1 \sum_{i=1}^{n_1} x_i}}{(1 + e^{\beta_1})^{n_1}} \frac{e^{(\beta_1+\beta_2) \sum_{i=1}^{n_2} y_i}}{(1 + e^{\beta_1+\beta_2})^{n_2}},$$

and the Fisher information matrix is

$$\mathbf{I}(\beta_1, \beta_2) = \begin{pmatrix} \frac{n_1 e^{\beta_1}}{(1+e^{\beta_1})^2} + \frac{n_2 e^{\beta_1+\beta_2}}{(1+e^{\beta_1+\beta_2})^2} & \frac{n_2 e^{\beta_1+\beta_2}}{(1+e^{\beta_1+\beta_2})^2} \\ \frac{n_2 e^{\beta_1+\beta_2}}{(1+e^{\beta_1+\beta_2})^2} & \frac{n_2 e^{\beta_1+\beta_2}}{(1+e^{\beta_1+\beta_2})^2} \end{pmatrix}.$$

4.7.1 Alpha prior elicitation

In many real problems, β_2 is of interest as it represents the effect of different experiment settings. We provide the alpha prior of the two-parameter logistic regression in the following proposition.

Proposition 4.6. *When β_2 is the parameter of interest, the conditional alpha prior $\pi_1(\beta_1|\beta_2)$ has the from*

$$\pi_1(\beta_1|\beta_2) \propto \sqrt{\frac{n_1 e^{\beta_1}}{(1 + e^{\beta_1})^2} + \frac{n_2 e^{\beta_1+\beta_2}}{(1 + e^{\beta_1+\beta_2})^2}}, \tag{4.5}$$

and the marginal alpha prior $\pi_2(\beta_2)$ has the form

$$\pi_2(\beta_2) \propto \left\{ \int_{-\infty}^{\infty} \pi_1(\beta_1|\beta_2) \left\{ \frac{n_1 n_2 e^{\beta_1+\beta_2}}{n_1 (1 + e^{\beta_1+\beta_2})^2 + n_2 e^{\beta_2} (1 + e^{\beta_1})^2} \right\}^{-\frac{\alpha}{2}} \mathrm{d}\beta_1 \right\}^{-\frac{1}{\alpha}}. \tag{4.6}$$

Proof. Since β_2 is of interest, we first derive

$$|\mathbf{I}_{11}(\beta)| = \frac{n_1 e^{\beta_1}}{(1 + e^{\beta_1})^2} + \frac{n_2 e^{\beta_1+\beta_2}}{(1 + e^{\beta_1+\beta_2})^2}, \tag{4.7}$$

and

$$|\mathbf{I}(\beta)| / |\mathbf{I}_{11}(\beta)| = \frac{n_1 n_2 e^{\beta_1+\beta_2}}{n_1 (1 + e^{\beta_1+\beta_2})^2 + n_2 e^{\beta_2} (1 + e^{\beta_1})^2}.$$

According to Theorem 3.2, the conditional alpha prior $\pi_1(\beta_1|\beta_2)$ is proportional to the square root of (4.7). Note that

$$\begin{aligned}\int_{-\infty}^{\infty}|\mathbf{I}_{11}(\beta)|\mathrm{d}\beta_1 &\leqslant \int_{-\infty}^{\infty}\sqrt{2\max\left(\frac{n_1e^{\beta_1}}{(1+e^{\beta_1})^2},\frac{n_2e^{\beta_1+\beta_2}}{(1+e^{\beta_1+\beta_2})^2}\right)}\mathrm{d}\beta_1\\ &\leqslant \sqrt{2}\max\left(\int_{-\infty}^{\infty}\sqrt{\frac{n_1e^{\beta_1}}{(1+e^{\beta_1})^2}}\mathrm{d}\beta_1,\int_{-\infty}^{\infty}\sqrt{\frac{n_2e^{\beta_1+\beta_2}}{(1+e^{\beta_1+\beta_2})^2}}\mathrm{d}\beta_1\right)\\ &= \sqrt{2}\max\left(\sqrt{n_1}\pi,\sqrt{n_2}\pi\right)<\infty,\end{aligned}$$

where $\max(a,b)$ means the maximum value of a and b. Therefore, $\pi_1(\beta_1|\beta_2)$ is proper and it is unnecessary to operate on the compact subsets of the parameter space. Hence, (4.5) follows and the marginal alpha prior of β_1 is given by (4.6) due to (3.2). □

4.7.2 Simulation study

We compare the alpha priors under different values of α using a frequentist criterion. The frequentist coverage properties of priors have been used to compare objective priors by many authors (Stein, 1985; Berger et al., 2001). A common idea is to compute the frequentist coverage probabilities of equal-tailed Bayesian credible intervals for parameters of interest. The coverage probability yielded by the most favorable prior would be closest to the nominal level of the credible intervals.

For the two-parameter logistic regression, we compare the frequentist coverage probabilities of 95% equal-tailed Bayesian credible intervals of β_2 under different choices of α, namely -0.9, -0.5, -0.1, 0, 0.1, 0.5, 0.9. Here, $\alpha = 0$ denotes the Bernardo-Berger's reference prior used as a benchmark for the comparison. We consider two different combinations of n_1 and n_2, namely $(5,5)$ and $(20,20)$, as well as five choices of (β_1,β_2), i.e., $(0,0)$, $(0,1)$, $(0,-1)$, $(1,-1)$, $(1,0)$, $(1,1)$. For each of the ten conditions, we generated 10000 replications of $\boldsymbol{x}$ and $\boldsymbol{y}$. For each alpha prior (including the reference prior), we computed the equal-tailed 95% credible interval of β_2 for each pair of $\boldsymbol{x}$ and $\boldsymbol{y}$. The credible intervals are calculated based on a sample from the posterior of β_2 generated by a Markov Chain Monte Carlo (MCMC) algorithm (the Metropolis-Hastings algorithm) with 100000 iterations. Table 4.2 presents the frequentist coverage probabilities for these intervals estimated from the 10000 replications and Table 4.3 demonstrates the corresponding expected lengths of these intervals. All the numerical integration involved in the simulation is computed by the Gauss-Kronrod quadrature algorithm.

Table 4.2: The coverage probabilities of the equal-tailed 95% credible intervals of β_2 under different combinations of sample sizes and parameters (β_1, β_2) when α is -0.9, -0.5, -0.1, 0, 0.1, 0.5, 0.9.

Sample sizes	Values of α						
and (β_1, β_2)	-0.9	-0.5	-0.1	0	0.1	0.5	0.9
n_1=5, n_2=5							
(0,-1)	0.9507	0.9529	0.9547	0.9520	0.9499	0.9590	0.9564
(0,0)	0.9541	0.9566	0.9380	0.9585	0.9574	0.9590	0.9616
(0,1)	0.9459	0.9464	0.9529	0.9532	0.9537	0.9581	0.9574
(1,-1)	0.9517	0.9415	0.9550	0.9416	0.9531	0.9564	0.9558
(1,0)	0.9470	0.9464	0.9445	0.9451	0.9445	0.9507	0.9574
(1,1)	0.9332	0.9300	0.9682	0.9436	0.9357	0.9448	0.9438
n_1=20, n_2=20							
(0,-1)	0.9471	0.9538	0.9559	0.9477	0.9460	0.9491	0.9503
(0,0)	0.9455	0.9447	0.9559	0.9454	0.9400	0.9419	0.9489
(0,1)	0.9493	0.9477	0.9558	0.9475	0.9501	0.9508	0.9461
(1,-1)	0.9520	0.9501	0.9567	0.9515	0.9476	0.9417	0.9502
(1,0)	0.9456	0.9456	0.9444	0.9451	0.9436	0.9500	0.9497
(1,1)	0.9404	0.9462	0.9345	0.9481	0.9390	0.9531	0.9495

From Table 4.2, we can see the coverage probabilities for all α are closer to the nominal level 95% even for small sample sizes. This implies that the alpha prior may be robust to the choice of α. However, from Table 4.3, compared with Bernardo-Berger's reference prior, the expected lengths of the credible intervals are smaller for a majority of the alpha priors with positive values of α. The larger the alpha

Table 4.3: The expected lengths of the equal-tailed 95% credible intervals of β_2 under different combinations of sample sizes and parameters (β_1, β_2) when α is -0.9, -0.5, -0.1, 0, 0.1, 0.5, 0.9.

Sample sizes and (β_1, β_2)	Values of α -0.9	-0.5	-0.1	0	0.1	0.5	0.9
n_1=5, n_2=5							
(0,-1)	7.1068	7.0791	8.7970	6.5081	6.4584	6.3367	6.2244
(0,0)	5.9939	6.0921	7.0869	5.7241	5.6718	5.6185	5.5455
(0,1)	6.8691	7.0115	8.9565	6.4697	6.4204	6.3106	6.2175
(1,-1)	7.0542	7.0249	8.9155	6.5282	6.4104	6.3376	6.2232
(1,0)	8.1622	8.0447	8.7448	7.3718	7.3613	7.2398	7.0075
(1,1)	10.1641	10.1426	10.4270	9.0203	8.8349	8.5333	8.3238
n_1=20, n_2=20							
(0,-1)	2.7691	2.7671	2.7643	2.7484	2.7371	2.7315	2.7124
(0,0)	2.5158	2.5160	2.5073	2.5031	2.5064	2.4925	2.4818
(0,1)	2.7536	2.7587	2.7441	2.7399	2.7332	2.7197	2.7050
(1,-1)	2.7635	2.7552	2.7545	2.7346	2.7298	2.7212	2.7057
(1,0)	2.9986	3.0068	2.9975	2.9756	2.9648	2.9376	2.9087
(1,1)	4.1811	4.2497	4.8729	4.0451	4.0289	3.9569	3.9399

is, the shorter the expected length is. This encourages adopting the alpha priors with positive α rather than Bernardo-Berger's reference prior. The improvement on Bernardo-Berger's reference prior is smaller compared with the large sample case. This is possibly because that the credible intervals with a smaller sample are more biased.

It is also seen that even though the alpha prior with $\alpha = 0.9$ generally enjoys better expected lengths than Bernardo-Berger's reference prior, it is not optimal and the coverage probabilities are not as satisfied as those of Bernardo-Berger's reference prior. The worse performance is possibly because the credible intervals are more biased towards zero and shorter when $\alpha = 0.9$. Even though the "universally" optimal value of α does not exist, we can still conclude that positive α is favored for the two-parameter regression model.

This simulation study is based on the two-parameter logistic regression model only. The limited simulation cannot guarantee that the alpha priors with positive α have a general frequentist advantage. Nonpositive α may be preferred for other problems. However, this study indeed demonstrates that alpha priors provide better frequentist performance than Bernardo-Berger's reference prior for some value of alpha.

In the next section, we use a clinical trial to illustrate the application of alpha priors for the two-parameter logistic regression.

4.8 Real example: a clinical trial of ECMO

ECMO is a treatment that uses a pump to circulate blood through an artificial lung back into the bloodstream of a very ill baby. This system provides heart-lung bypass support outside of the baby's body. It may help support a child who is awaiting a heart or lung transplant. Ware (1989) first introduced Bayesian analysis for a clinical trial about ECMO. Nine patients were treated with ECMO while ten patients in a control group were given conventional therapy. The experiment result is shown in Table 4.4.

Table 4.4: The numbers of survival and death from the ECMO group and the control group.

	ECMO	Control
Lived	9	6
Died	0	4

Let p_1 be the probability of survival under the conventional therapy and p_2 be the probability of survival under ECMO. Then $\beta_1 = \log\{p_1/(1-p_1)\}$ and $\beta_2 = \log\{p_2/(1-p_2)\} - \beta_1$. It is seen that the model of the ECMO problem is actually a two-parameter logistic regression discussed in Section 4.7. Experimenters were interested in comparing the performances of the two treatments and they used the posterior probability that β_2 is positive $\Pr(\beta_2 > 0|\text{data})$ to make a conclusion. If

$\Pr(\beta_2 > 0|\text{data})$ is large enough, say, more than 0.95, the experimenters can conclude that the ECMO therapy is effective. Thus, β_2 is of interest and β_1 is a nuisance parameter.

This example has been analyzed by many researchers. Kass and Greenhouse (1989) considered 84 different subjective priors, all with assumed independence between β_1 and β_2. Other authors, including Lavine et al. (1991) and Berger and Moreno (1994), argued that an independence assumption between β_1 and β_2 is necessary. Only under this assumption, $\Pr(\beta_2 > 0|\text{data})$ is large enough for a reasonable class of priors.

Following them, we derive the independent alpha priors for the ECMO experiment in the following proposition using the algorithm that we proposed in Section 3.2.

Proposition 4.7. *Under the independence assumption for the marginal alpha priors of β_1 and β_2, the marginal alpha priors are of the form*

$$\pi_1(\beta_1) = \frac{e^{\beta_1/2}}{\pi\left(1 + e^{\beta_1}\right)}, \tag{4.8}$$

and

$$\pi_2(\beta_2) \propto \left\{\frac{1}{\pi}\int_0^1 \{t(1-t)\}^{-\frac{1+\alpha}{2}} \left[n_1 + n_2\left\{(1-t)\,e^{-\beta_2/2} + te^{\beta_2/2}\right\}^2\right]^{\frac{\alpha}{2}} \mathrm{d}t\right\}^{-\frac{1}{\alpha}}. \tag{4.9}$$

Proof. Note that $|\mathbf{I}(\beta)| / |\mathbf{I}_{22}(\beta)| = e^{\beta_1} / \left(1 + e^{\beta_1}\right)^2$ is independent of β_2. By Proposition 3.1, $\pi_1(\beta_1) \propto e^{\beta_1/2} / \left(1 + e^{\beta_1}\right)$. Shown in Sun and Berger (1998), the normalizing constant of $\pi_1(\beta_1)$ is π and (4.8) follows.

For (4.9),

$$\pi_2(\beta_2) \propto \left\{ \int_{-\infty}^{\infty} \pi_1(\beta_1) \left(|\mathbf{I}(\beta)| / |\mathbf{I}_{11}(\beta)| \right)^{-\frac{\alpha}{2}} \mathrm{d}\beta_1 \right\}^{-\frac{1}{\alpha}}$$

$$= \left\{ \int_{-\infty}^{\infty} \frac{e^{\beta_1/2}}{\pi(1+e^{\beta_1})} \left\{ \frac{n_1 n_2 e^{\beta_1+\beta_2}}{n_1(1+e^{\beta_1+\beta_2})^2 + n_2 e^{\beta_2}(1+e^{\beta_1})^2} \right\}^{-\frac{\alpha}{2}} \mathrm{d}\beta_1 \right\}^{-\frac{1}{\alpha}}$$

$$= \left\{ \int_{-\infty}^{\infty} \frac{e^{\beta_1/2}}{\pi(1+e^{\beta_1})} \left\{ n_1 e^{-(\beta_1+\beta_2)} \left(1+e^{\beta_1+\beta_2}\right)^2 + n_2 e^{-\beta_1} \left(1+e^{\beta_1}\right)^2 \right\}^{\frac{\alpha}{2}} \mathrm{d}\beta_1 \right\}^{-\frac{1}{\alpha}}.$$

Let $t = e^{\beta_1} / \left(1+e^{\beta_1}\right)$, and then we have

$$\pi_2(\beta_1) \propto \left\{ \int_0^1 \frac{\{t(1-t)\}^{-\frac{1}{2}}}{\pi} \left[\frac{n_1(1-t)}{te^{\beta_1}} \left(1+\frac{te^{\beta_1}}{1-t}\right) + \frac{n_2(1-t)}{te^{\beta_1}} \left(1+\frac{t}{1-t}\right) \right]^{\frac{\alpha}{2}} \mathrm{d}t \right\}^{-\frac{1}{\alpha}}$$

$$= \left\{ \frac{1}{\pi} \int_0^1 \{t(1-t)\}^{-\frac{1}{2}} \left[\frac{n_1 + n_2 \left\{(1-t) e^{-\beta_1/2} + te^{\beta_1/2}\right\}^2}{t(1-t)} \right]^{\frac{\alpha}{2}} \mathrm{d}t \right\}^{-\frac{1}{\alpha}}$$

$$= \left\{ \frac{1}{\pi} \int_0^1 \{t(1-t)\}^{-\frac{1+\alpha}{2}} \left[n_1 + n_2 \left\{(1-t) e^{-\beta_1/2} + te^{\beta_1/2}\right\}^2 \right]^{\frac{\alpha}{2}} \mathrm{d}t \right\}^{-\frac{1}{\alpha}}.$$

Hence (4.9) is proved. □

We evaluate the performances of the independent alpha priors and the dependent alpha priors we developed in Section 4.7. We compare $\Pr(\beta_2 > 0|\text{data})$, the posterior standard deviations (s.d.), the posterior medians and the posterior means of β_2 under each choice of α. We also include Bernardo-Berger's reference prior in comparison as a benchmark and report all the results in Table 4.5. For convenience, the reference prior is denoted by $\alpha = 0$. Figure 4.1 illustrates the marginal prior densities of β_2 we proposed in Section 4.7 and their corresponding posterior densities of β_2 under different choices of α. Figure 4.2 presents the marginal prior densities of β_2 under the independence assumption and their corresponding posterior densities of β_2 under different choices of α. All the numerical integration for calculating the

Table 4.5: The posterior probability that $\beta_2 > 0$, the standard deviations, the medians and the means of the posterior distributions of β_2 for the independent alpha priors and the dependent alpha priors, respectively, when α is -0.9, -0.5, -0.1, 0, 0.1, 0.5 and 0.9.

Prior Types	α	$\Pr(\beta_2 > 0 \mid \text{data})$	s.d.	Median	Mean
Dependent	-0.9	0.9777	1.580	2.432	2.621
	-0.5	0.9782	1.550	2.406	2.589
	-0.1	0.9773	1.526	2.362	2.537
	0	0.9768	1.537	2.361	2.543
	0.1	0.9762	1.533	2.343	2.520
	0.5	0.9749	1.500	2.278	2.452
	0.9	0.9738	1.479	2.216	2.388
Independent	-0.9	0.9910	2.373	3.455	3.882
	-0.5	0.9926	2.442	3.609	4.050
	-0.1	0.9932	2.488	3.736	4.184
	0	0.9936	2.544	3.796	4.258
	0.1	0.9936	2.699	3.943	4.449
	0.5	0.9941	2.750	4.025	4.549
	0.9	0.9931	2.522	3.764	4.226

alpha priors is computed by the Gauss-Kronrod quadrature algorithm. All the prior and posterior samples are generated from 1,000,000 MCMC iterations. We choose this very large number of iterations in order to draw smooth curves of the prior and posterior densities.

From Figure 4.1 and Figure 4.2, the dependent alpha priors and independent alpha priors behave similarly. All marginal alpha priors of β_2 look like normal or the Cauchy distributions. The alpha prior density of β_2 with a smaller value of α is flatter. The marginal alpha prior density of β_2 becomes sharper while the marginal posterior density of β_2 is more skewed to 0 with a higher value of α. This suggests the alpha prior increasingly dominates the likelihood function when the value of α increases. The difference between the two types of alpha priors is that the dependent marginal alpha priors are sharper and their modes are closer to 0. It implies that the dependent marginal alpha priors tend to have more influence on the marginal posterior densities of β_2. By the definition of objective or noninformative priors, the independent alpha priors are preferred for the ECMO data.

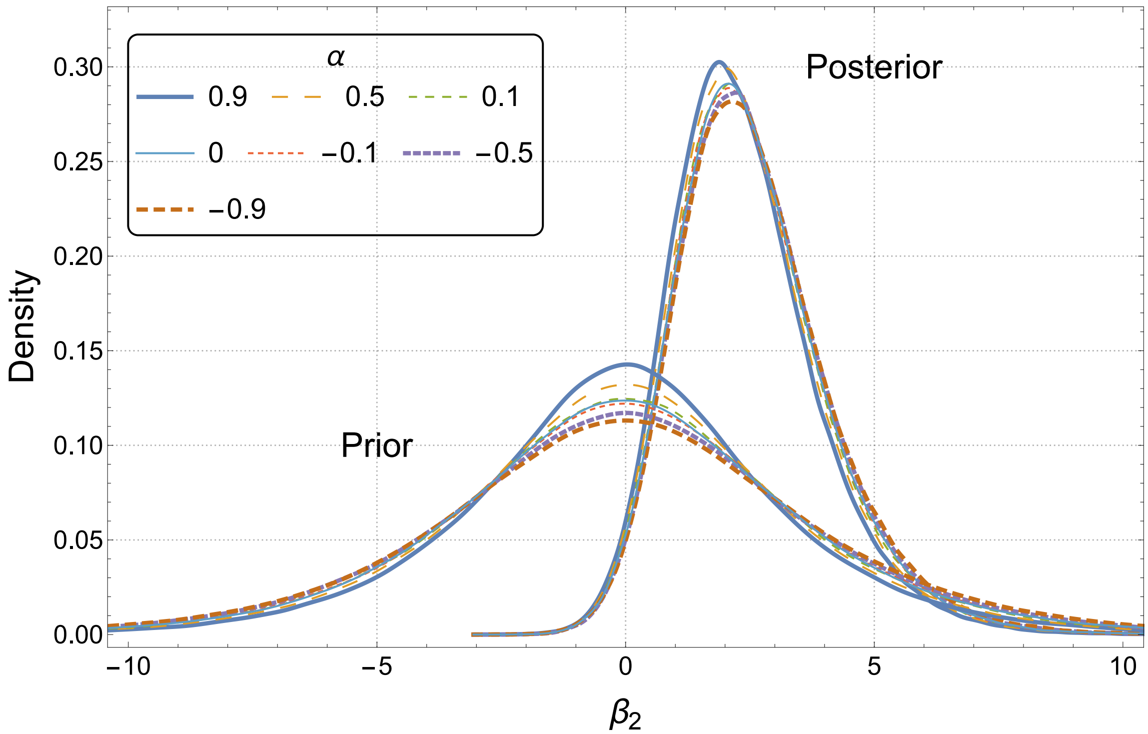

Figure 4.1: The marginal alpha prior and the marginal posterior densities of β_2 without the independence assumption when α is -0.9, -0.5, -0.1, 0, 0.1, 0.5 and 0.9.

Sun and Berger (1998) demonstrated that $\Pr(\beta > 0|\text{data})$ was approximately 0.99 under the independent reference priors. This value is larger than that obtained by Kass and Greenhouse (1989) which is approximately 0.95. From Table 4.5, the posterior probabilities that β_2 is positive are still approximately 0.99 for the independent alpha priors while those become between 0.97 and 0.98 for the dependent alpha priors. The independent alpha priors are preferred due to their larger $\Pr(\beta_2 > 0|\text{data})$. Generally, $\Pr(\beta_2 > 0|\text{data})$ decreases for the dependent alpha priors with the value of α increasing while it increases for the independent alpha priors. The standard deviations of the posterior density of β_2 become smaller for the dependent alpha priors with a larger value of α while they grow for the independent objective priors.

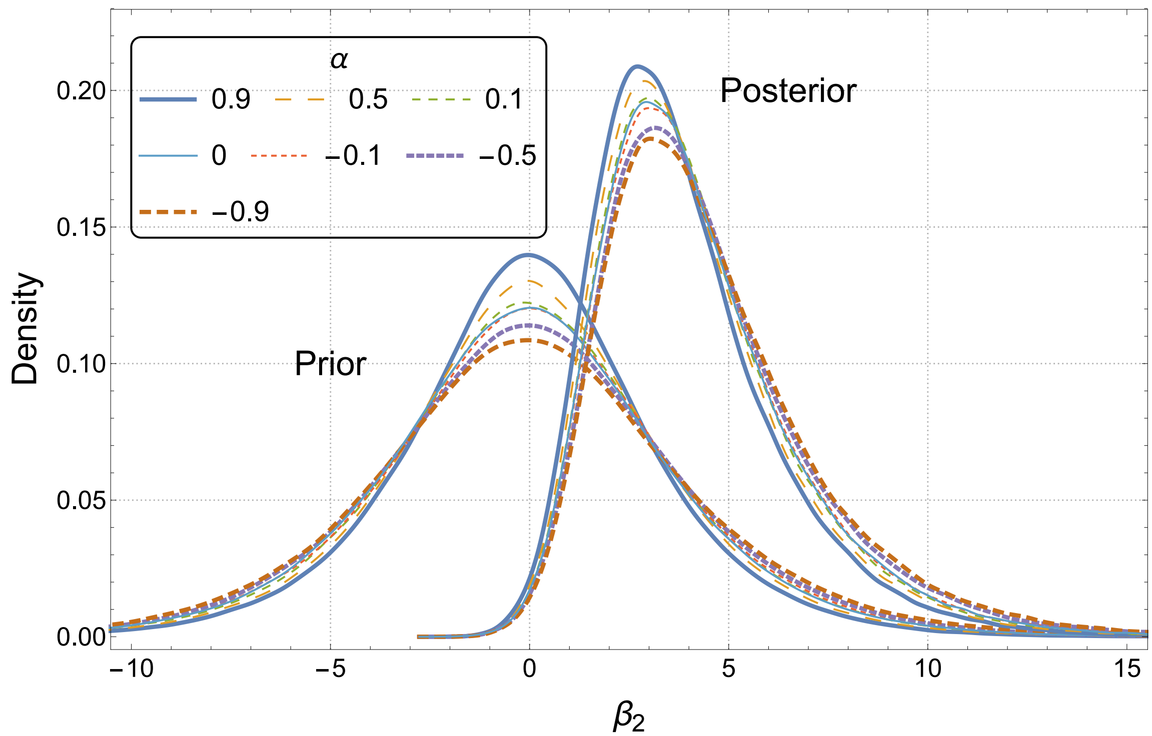

Figure 4.2: The marginal alpha prior and the marginal posterior densities of β_2 under the independence assumption when α is -0.9, -0.5, -0.1, 0, 0.1, 0.5 and 0.9.

Both the means and the medians behave similarly to the standard deviation. They are closer to 0 with a larger value of α under the dependent alpha priors while they tend upwards under the independent alpha priors. It is obvious that effect of the value of α on the posterior distribution of β_2 under the dependent alpha priors is contrary to that under the independent alpha priors. Since all the priors are peaked at 0, it suggests that dependent alpha priors contribute more to the posterior of β_2. This implication is consistent with the aforementioned conclusion from Figure 4.1 and Figure 4.2. However, there are exceptions to the general behavior of the posterior distribution of β_2, namely, $\alpha = 0.9$. From Table 4.5, all properties under the independent alpha prior with $\alpha = 0.9$ are similar to those under Bernardo-Berger's reference prior. This may be explained by the conclusion in Section 4.7 that the alpha priors with a large value of α may yield highly biased Bayesian estimation towards zero. Besides, $\Pr(\beta > 0|\text{data})$ under the dependent alpha prior with $\alpha = -0.9$ is smaller than that with $\alpha = -0.5$. The irregularity may discourage using alpha priors with α close to -1 or 1. Therefore, we recommend α whose absolute values is small, say, less than or equal to 0.5 for this experiment.

Chapter 5

Conclusions and future work

This chapter draws conclusions on the thesis, and points out some possible future research directions related to the work done in this thesis.

5.1 Conclusions

In this thesis, we have developed the asymptotic approximations to $J^{\alpha}\left(\pi_1\left(\theta_1\right)\right)$ and $J^{\alpha}\left(\pi_2\left(\theta_2|\theta_1\right)\right)$ in the regular continuous case. Based on these approximations, we have proposed objective priors under the class of α-divergence measures in the presence of nuisance parameters, which we called alpha priors. The general form of the alpha prior involves α, enlarging the class of objective priors beyond the Jeffreys prior and Bernardo-Berger's reference prior. We have also developed a general approach to elicit alpha priors in the ordered group case.

The alpha prior might provide new insights into how the choice of divergence measures affects inference. One may deliberately choose the values of α to achieve better inference for a specific problem. Alpha priors may also allow us to investigate the robustness of objective Bayesian inference with respect to the choice of divergence measures. In some cases, the alpha priors are independent of α. However, the choice of α indeed affects the inference in some exponential family distributions such as the beta distribution and the logistic regression model. An illustrative example showed that the independent alpha prior with α not far from 0 is suitable for analyzing the ECMO experiment.

5.2 Future work

Related topics for the future research work are listed below.

1. In this thesis, we have considered the regular continuous case only. It would be useful to extend our method to the discrete case or other non-regular cases in future research.

2. We consider all possible values of α except $\alpha = -1$, i.e. the chi-squared divergence. In the presence of nuisance parameters, it is still necessary to study objective priors under the chi-squared divergence. The second order expansion of $p(\theta_1|\boldsymbol{x}_n)$ may be used to generate terms involving $\pi_1(\theta_1)$ in the asymptotic approximation, as Liu et al. (2014) suggested. In this way, $J^{-1}(\pi_1(\cdot))$ can be expressed as a function of $\pi_1(\theta_1)$ and we can find the corresponding alpha prior.

3. The simulation study is not enough to provide a comprehensive guidance on choosing a proper value of α for different problems. It would be helpful to further investigate how to find optimal α based on frequentist criteria or Bayes risk.

4. It seems that the figures of the simulation results for different α are in a small range. It is likely that alpha priors are robust with respect to the choice of α. Further research can also be conducted on the robustness of alpha priors.

Appendix A

Derivation of (4.4)

In this appendix, we describe how to derive Equation (4.4).

From (4.3), $\pi^k(\gamma)$ can be expressed as

$$\pi^k(\gamma) = \left\{ \frac{\int_{l_k}^{u_k} W^k(\gamma, \beta)\mathrm{d}\beta}{\sqrt{\gamma}\log(u_k) - \log(l_k) + O(1)} \right\}^{-\frac{1}{\alpha}},$$

where

$$W^k(\gamma, \beta) = \{\psi'(\gamma)\{\psi'(\beta) - \psi'(\gamma+\beta)\} - \psi'(\gamma+\beta)\psi'(\beta)\}^{-\frac{\alpha}{2}} \times \{\psi'(\beta) - \psi'(\gamma+\beta)\}^{\frac{1+\alpha}{2}}.$$

To obtain (4.4), we need the following lemma.

Lemma A.1. *Define $\psi'(x)$ is the trigamma function, then for positive x and y*

$$\lim_{y\to 0} \psi'(x+y)\psi'(y)y^2 = \psi'(x), \tag{A.1}$$

$$\lim_{y\to \infty} \psi'(x+y)\psi'(y)y^2 = 1, \tag{A.2}$$

$$\lim_{y\to 0} \{\psi'(y) - \psi'(x+y)\}y^2 = 1, \tag{A.3}$$

and

$$\lim_{y\to \infty} \{\psi'(y) - \psi'(x+y)\}y^2 = x. \tag{A.4}$$

Proof. The trigamma function can be expanded as $\psi'(x) = \sum_{j=0}^{\infty}(x+j)^{-2}$ (Bownan and Shenton, 1982). Let $h(x) = \sum_{j=1}^{\infty}(x+j)^{-2}$, for $x \geqslant 0$. Since $h'(x) = -2\sum_{j=1}^{\infty}(x+j)^{-3} < 0$ for $x \geqslant 0$, $h(x)$ is a decreasing function of x. Then,

$$h(x) \geqslant \int_{1}^{\infty}(x+y)^{-2}\mathrm{d}y = (x+1)^{-1}.$$

On the other hand,

$$\frac{1}{(x+j)^2} \leqslant \frac{1}{(x+j)^2 - \frac{1}{4}} = \frac{1}{(x+j-\frac{1}{2})} - \frac{1}{(x+j+\frac{1}{2})} = \int_{j-\frac{1}{2}}^{j+\frac{1}{2}} \frac{1}{(x+y)^2}\mathrm{d}y,$$

for $j \geqslant 1$. Thus,

$$h(x) \leqslant \sum_{j=1}^{\infty}\int_{j-\frac{1}{2}}^{j+\frac{1}{2}} \frac{1}{(x+y)^2}\mathrm{d}y = \int_{\frac{1}{2}}^{\infty}(x+y)^{-2}\mathrm{d}y = (x+\frac{1}{2})^{-1}.$$

Consequently,

$$x^{-2} + (x+1)^{-1} \leqslant \psi'(x) \leqslant x^{-2} + (x+1/2)^{-1}. \tag{A.5}$$

We then have

$$\psi'(x+y)\left(1+\frac{y^2}{y+1}\right) \leqslant \psi'(x+y)\,\psi'(y)\,y^2 \leqslant \psi'(x+y)\left(1+\frac{y^2}{y+\frac{1}{2}}\right).$$

Let y go to 0, and hence

$$\psi'(x) \leqslant \lim_{y\to 0}\psi'(x+y)\,\psi'(y)\,y^2 \leqslant \psi'(x),$$

and hence (A.1) follows.

To prove (A.2), we first show that, according to (A.5),

$$\left(\frac{y}{(x+y)^2} + \frac{y}{x+y+1}\right)\left(\frac{1}{y} + \frac{y}{y+1}\right) \leqslant \psi'(x+y)\,\psi'(y)\,y^2$$
$$\leqslant \left(\frac{y}{(x+y)^2} + \frac{y}{x+y+\frac{1}{2}}\right)\left(\frac{1}{y} + \frac{y}{y+\frac{1}{2}}\right).$$

Let y go to infinity, we conclude that $\lim_{y\to\infty} \psi'(x+y)\,\psi'(y)\,y^2 = 1$. Equation (A.1) and Equation (A.2) are proved.

Sun and Berger (1998) showed that

$$x\left\{\frac{x+2y}{y^2\,(x+y)^2}+\frac{1}{(y+1)\,(x+y+1)}\right\}\leqslant \psi'(y)-\psi'(x+y)$$
$$\leqslant x\left\{\frac{x+2y}{y^2\,(x+y)^2}+\frac{1}{\left(y+\frac{1}{2}\right)\left(x+y+\frac{1}{2}\right)}\right\}.$$

Therefore,

$$x\left\{\frac{x+2y}{(x+y)^2}+\frac{y^2}{(y+1)\,(x+y+1)}\right\}\leqslant \{\psi'(y)-\psi'(x+y)\}\,y^2$$
$$\leqslant x\left\{\frac{x+2y}{(x+y)^2}+\frac{y^2}{\left(y+\frac{1}{2}\right)\left(x+y+\frac{1}{2}\right)}\right\}. \quad \text{(A.6)}$$

The both sides of (A.6) go to 1 if $y \to 0$ and go to x if $y \to \infty$. Thus, we prove (A.3) and (A.4).

□

For any sufficiently small $\epsilon > 0$ and k large enough such that $l_k < \epsilon < 1/\epsilon < u_k$, we have

$$\int_{l_k}^{u_k} W^k(\gamma,\beta)\mathrm{d}\beta = \int_{l_k}^{\varepsilon} W^k(\gamma,\beta)\mathrm{d}\beta + O(1) + \int_{1/\varepsilon}^{u_k} W^k(\gamma,\beta)\mathrm{d}\beta. \quad \text{(A.7)}$$

Here, $O(1)$ represents a boundary function of l_k and u_k. For the first term on the right hand side of (A.7)

$$\int_{l_k}^{\varepsilon} W^k(\gamma,\beta)\mathrm{d}\beta = \int_{l_k}^{\varepsilon} \frac{1}{\beta} \left\{\psi'(\gamma)\left\{\psi'(\beta)-\psi'(\gamma+\beta)\right\}\beta^2 - \psi'(\gamma+\beta)\,\psi'(\beta)\,\beta^2\right\}^{-\frac{\alpha}{2}}$$

$$\times \left\{\left\{\psi'(\beta)-\psi'(\gamma+\beta)\right\}\beta^2\right\}^{\frac{1+\alpha}{2}} \mathrm{d}\beta$$

$$= \int_{l_k}^{\varepsilon} \frac{1}{\beta} \left\{\left\{\left\{\psi'(\gamma)-\psi'(\gamma+\beta)\right\}\beta^2\psi'(\gamma) - \psi'(\beta)\,\psi'(\gamma+\beta)\,\beta^2\right\}^{-\frac{\alpha}{2}} + O(\beta)\right\} \mathrm{d}\beta$$

$$= \int_{l_k}^{\varepsilon} \frac{1}{\beta} \left\{0 + O(\beta)\right\} \mathrm{d}\beta$$

$$= O(\varepsilon - l_k)$$

$$= O(\varepsilon).$$

The second and the third equation follows (A.1) and (A.3). For the third term on the right hand side of (A.7)

$$\int_{1/\varepsilon}^{u_k} W^k(\gamma,\beta)\mathrm{d}\beta = \int_{1/\varepsilon}^{u_k} \frac{1}{\beta} \left\{\psi'(\gamma)\left\{\psi'(\beta)-\psi'(\gamma+\beta)\right\}\beta^2 - \psi'(\gamma+\beta)\,\psi'(\beta)\,\beta^2\right\}^{-\frac{\alpha}{2}}$$

$$\times \left\{\left\{\psi'(\beta)-\psi'(\gamma+\beta)\right\}\beta^2\right\}^{\frac{1+\alpha}{2}} \mathrm{d}\beta$$

$$= \int_{1/\varepsilon}^{u_k} \frac{1}{\beta} \left\{\left\{\psi'(\gamma)\,\gamma - 1\right\}^{-\frac{\alpha}{2}} \gamma^{\frac{1+\alpha}{2}} + O\left(\frac{1}{\beta}\right)\right\} \mathrm{d}\beta$$

$$= \left\{\psi'(\gamma)\,\gamma - 1\right\}^{-\frac{\alpha}{2}} \gamma^{\frac{1+\alpha}{2}} \left\{\log(u_k) + \log(\varepsilon)\right\} + O\left(\varepsilon - u_k^{-1}\right)$$

$$= \left\{\psi'(\gamma)\,\gamma - 1\right\}^{-\frac{\alpha}{2}} \gamma^{\frac{1+\alpha}{2}} \log(u_k) + O(\varepsilon).$$

The second equation follows (A.2) and (A.4).

Finally,

$$\int_{l_k}^{u_k} W^k(\gamma,\beta)\mathrm{d}\beta = \left\{\psi'(\gamma)\,\gamma - 1\right\}^{-\frac{\alpha}{2}} \gamma^{\frac{1+\alpha}{2}} \log(u_k) + O(\varepsilon),$$

and we obtain (4.4).

Appendix B

The alpha prior of the Dirichlet distribution

In this appendix, we provide the details of deriving the alpha prior of the Dirichlet distribution.

Before we calculate the alpha prior of the Dirichlet distribution, we need the following lemma.

Lemma B.1. *Let $\psi'(x)$ denote the trigamma function, then for positive x and y*

$$\lim_{y \to 0} \{\psi'(y)(x+y) - 1\} y^2 = x, \tag{B.1}$$

and

$$\lim_{y \to \infty} \{\psi'(y)(x+y) - 1\} y = \frac{1}{2} + x. \tag{B.2}$$

Proof. From (A.5), we have

$$y^2 \left(\frac{x+y}{y^2} + \frac{x-1}{y+1} \right) \leqslant \{\psi'(y)(x+y) - 1\} y^2 \leqslant y^2 \left(\frac{x+y}{y^2} + \frac{x-\frac{1}{2}}{y+\frac{1}{2}} \right),$$

$$x + y + \frac{y^2(x-1)}{y+1} \leqslant \{\psi'(y)(x+y) - 1\} y^2 \leqslant x + y + \frac{y^2\left(x-\frac{1}{2}\right)}{y+\frac{1}{2}}.$$

Let y go to 0, we obtain

$$\lim_{y\to 0} x + y + \frac{y^2\,(x-1)}{y+1} \leqslant \lim_{y\to 0} \{\psi'(y)\,(x+y) - 1\}\, y^2 \leqslant \lim_{y\to 0} x + y + \frac{y^2\left(x-\frac{1}{2}\right)}{y+\frac{1}{2}},$$

$$x \leqslant \lim_{y\to 0} \{\psi'(y)\,(x+y) - 1\}\, y^2 \leqslant x.$$

Hence, (B.1) follows.

To prove the second limit (B.2), we first determine the upper bound of $\{\psi'(y)\,(x+y)-1\}\,y$. By (A.5), we have

$$\{\psi'(y)\,(x+y)-1\}\,y \leqslant \frac{x+y}{y} + \frac{y\left(x-\frac{1}{2}\right)}{y+\frac{1}{2}},$$

and then,

$$\lim_{y\to\infty} \{\psi'(y)\,(x+y)-1\}\,y \leqslant x + \frac{1}{2}.$$

Batir (2008) showed that $\psi'(x) - \frac{1}{x} - \frac{1}{2x^2} > 0$. Thus, we obtain the lower bound of $\{\psi'(y)\,(x+y)-1\}\,y$ as

$$\{\psi'(y)\,(x+y)-1\}\,y \geqslant x + \frac{x+y}{2y},$$

and $\lim_{y\to\infty} \{\psi'(y)\,(x+y)-1\}\,y \geqslant x + \frac{1}{2}$. Hence, $\lim_{y\to\infty} \{\psi'(y)\,(x+y)-1\}\,y = x + \frac{1}{2}$. □

We derive $\pi_3^k\,(\beta_3|\beta_1,\beta_2)$ first. Follow the algorithm in Section 3.3,

$$\pi_3^k\,(\beta_3|\beta_1,\beta_2) = \frac{\{h_3\,(\beta)\}^{\frac{1}{2}}}{W_3^k\,(\beta_1,\beta_2)},$$

where

$$W_3^k\,(\beta_1,\beta_2) = \int_{l_{3k}}^{u_{3k}} \{h_3\,(\beta)\}^{\frac{1}{2}}\,\mathrm{d}\beta_3.$$

Similarly to (4.2), for any ε small enough and k large enough such that $l_{3k} < \epsilon < 1/\epsilon < u_{3k}$, we have

$$W_3^k(\beta_1, \beta_2) = \sqrt{\beta_1 + \beta_2} \log(u_{3k}) - \log(l_{3k}) + C_3(\varepsilon, \beta_1, \beta_2),$$

and hence

$$\pi_3^k(\beta_3|\beta_1, \beta_2) = \frac{\{\psi'(\beta_3) - \psi'(\beta_1 + \beta_2 + \beta_3)\}^{\frac{1}{2}}}{\sqrt{\beta_1 + \beta_2} \log(u_{3k}) - \log(l_{3k}) + C_3(\varepsilon, \beta_1, \beta_2)},$$

where

$$C_3(\varepsilon, \beta_1, \beta_2) = \int_{\varepsilon}^{1/\varepsilon} \{h_3(\beta)\}^{\frac{1}{2}} \mathrm{d}\beta_3 + O(\varepsilon).$$

Note that $C_3(\varepsilon, \beta_1, \beta_2)$ is a boundary function of l_{3k} and u_{3k} for fixed ε, β_1 and β_2.

Secondly, we derive $\pi_2^k(\beta_2, \beta_3|\beta_1)$. From (3.8),

$$\pi_2^k(\beta_2, \beta_3|\beta_1) = \frac{\pi_3^k(\beta_3|\beta_1, \beta_2)\left\{E_2^k\left[|h_2(\beta)|^{-\frac{\alpha}{2}}|\beta_{[2]}\right]\right\}^{-\frac{1}{\alpha}}}{W_2^k(\beta_1)},$$

where

$$W_2^k(\beta_1) = \int_{l_{2k}}^{u_{2k}} \left\{E_2^k\left[|h_2(\beta)|^{-\frac{\alpha}{2}}|\beta_{[2]}\right]\right\}^{-\frac{1}{\alpha}} \mathrm{d}\beta_2.$$

We calculate $E_2^k\left[|h_2(\beta)|^{-\frac{\alpha}{2}}|\beta_{[2]}\right]$ and $W_2^k(\beta_1)$ one by one for any ε small enough and k large enough such that $l_{3k} < \varepsilon < 1/\varepsilon < u_{3k}$.

From the algorithm in Section 3.3, we have

$$\begin{aligned} E_2^k\left[|h_2(\beta)|^{-\frac{\alpha}{2}}|\beta_{[2]}\right] &= \int_{l_{3k}}^{u_{3k}} \{h_2(\beta)\}^{-\frac{\alpha}{2}} \frac{\{h_3(\beta)\}^{\frac{1}{2}}}{W_3^k(\beta_1, \beta_2)} \mathrm{d}\beta_3 \\ &= \frac{1}{W_3^k(\beta_1, \beta_2)} \int_{l_{3k}}^{u_{3k}} g_{21}(\beta_1, \beta_2, \beta_3) \mathrm{d}\beta_3, \end{aligned}$$

where

$$g_{21}\left(\beta_1,\beta_2,\beta_3\right)=\left\{\psi'\left(\beta_2\right)-\frac{\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)}{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)}\right\}^{-\frac{\alpha}{2}}\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}^{\frac{1}{2}}.$$

From Lemma A.1, we obtain following limits

$$\lim_{\beta_3\to 0}\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2 = 1, \tag{B.3}$$

$$\lim_{\beta_3\to \infty}\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2 = \beta_1+\beta_2, \tag{B.4}$$

$$\lim_{\beta_3\to 0}\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2 = \psi'\left(\beta_1+\beta_2\right), \tag{B.5}$$

$$\lim_{\beta_3\to \infty}\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2 = 1. \tag{B.6}$$

For any ε small enough and k large enough such that $l_{2k}<\epsilon<1/\epsilon<u_{2k}$, we have

$$\int_{l_{3k}}^{u_{3k}} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3=\int_{1/\varepsilon}^{u_{3k}} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3+\int_{\varepsilon}^{1/\varepsilon} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3 + \int_{l_{3k}}^{\varepsilon} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3. \tag{B.7}$$

For the first term on the right hand side of (B.7), by (B.4) and (B.6), we have

$$\begin{aligned}\int_{1/\varepsilon}^{u_{3k}} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3 &= \int_{1/\varepsilon}^{u_{3k}}\frac{1}{\beta}\left\{\psi'\left(\beta_2\right)-\frac{\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2}{\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2}\right\}^{-\frac{\alpha}{2}}\\ &\quad\times\left\{\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2\right\}^{\frac{1}{2}}\mathrm{d}\beta_3\\ &=\int_{1/\varepsilon}^{u_{3k}}\frac{1}{\beta_3}\left\{\left\{\psi'\left(\beta_2\right)-\frac{1}{\beta_1+\beta_2}\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1}{2}}+O\left(\frac{1}{\beta_3}\right)\right\}\mathrm{d}\beta_3\\ &=\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1+\alpha}{2}}\log\left(u_{3k}\right)+O\left(\varepsilon\right).\end{aligned}$$

For the third term on the right hand side of (B.7), by (B.3) and (B.5), we have

$$\int_{l_{3k}}^{\varepsilon} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3 = \int_{l_{3k}}^{\varepsilon} \frac{1}{\beta}\left\{\psi'\left(\beta_2\right) - \frac{\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2}{\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2}\right\}^{-\frac{\alpha}{2}}$$

$$\times\left\{\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2\right\}^{\frac{1}{2}}\mathrm{d}\beta_3$$

$$= \int_{l_{3k}}^{\varepsilon} \frac{1}{\beta_3}\left\{\left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}^{-\frac{\alpha}{2}} 1 + O\left(\beta_3\right)\right\}\mathrm{d}\beta_3$$

$$= -\left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right) + O\left(\varepsilon\right).$$

Therefore,

$$E_2^k\left[\left|h_2\left(\beta\right)\right|^{-\frac{\alpha}{2}}\left|\beta_{[2]}\right.\right] = \frac{1}{W_3^k\left(\beta_1,\beta_2\right)}\left\{\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1+\alpha}{2}}\log\left(u_{3k}\right)\right.$$

$$\left. - \left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right) + C_{21}\left(\varepsilon,\beta_1,\beta_2\right)\right\},$$

where

$$C_{21}\left(\varepsilon,\beta_1,\beta_2\right) = \int_{\varepsilon}^{1/\varepsilon} g_{21}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3 + O\left(\varepsilon\right).$$

Similarly,

$$W_2^k\left(\beta_1\right) = \int_{l_{2k}}^{u_{2k}} g_{22}^k\left(\beta_1,\beta_2\right)\mathrm{d}\beta_2,$$

where

$$g_{22}^k\left(\beta_1,\beta_2\right) = \left\{\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1+\alpha}{2}}\log\left(u_{3k}\right)\right.$$

$$\left. - \left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right) + C_{21}\left(\varepsilon,\beta_1,\beta_2\right)\right\}^{-\frac{1}{\alpha}}$$

$$\left\{\sqrt{\beta_1+\beta_2}\log\left(u_{3k}\right) - \log\left(l_{3k}\right) + C_3\left(\varepsilon,\beta_1,\beta_2\right)\right\}^{\frac{1}{\alpha}}.$$

For u_{2k}, u_{3k} large enough and l_{2k}, l_{3k} small enough, $C_{21}\left(\varepsilon,\beta_1,\beta_2\right)$ and $C_3\left(\varepsilon,\beta_1,\beta_2\right)$ in $g_{22}\left(\beta_1,\beta_2\right)$ can be neglected as they are bounded functions of u_{2k}, u_{3k}, l_{2k} and l_{3k}. By Lemma A.1 and Lemma B.1, we have

$$\lim_{\beta_2 \to 0} g_{22}^k(\beta_1, \beta_2)\beta_2 = \left\{ \left\{ \{\psi'(\beta_2)(\beta_1+\beta_2) - 1\} \beta_2^2 \right\}^{-\frac{\alpha}{2}} \{\beta_1+\beta_2\}^{\frac{1+\alpha}{2}} \log(u_{3k}) \right.$$

$$\left. - \left\{ \{\psi'(\beta_2) - \psi'(\beta_1+\beta_2)\} \beta_2^2 \right\}^{-\frac{\alpha}{2}} \log(l_{3k}) \right\}^{-\frac{1}{\alpha}}$$

$$\left\{ \sqrt{\beta_1+\beta_2} \log(u_{3k}) - \log(l_{3k}) \right\}^{\frac{1}{\alpha}}$$

$$= \left\{ \frac{\sqrt{\beta_1} \log(u_{3k}) - \log(l_{3k})}{\sqrt{\beta_1} \log(u_{3k}) - \log(l_{3k})} \right\}^{-\frac{1}{\alpha}}$$

$$= 1,$$

and

$$\lim_{\beta_2 \to \infty} g_{22}^k(\beta_1, \beta_2)\beta_2 = \lim_{\beta_2 \to \infty} \left\{ \frac{1}{\sqrt{\beta_1+\beta_2}\log(u_{3k}) - \log(l_{3k})} \right.$$

$$\times \left\{ \{\psi'(\beta_2)(\beta_1+\beta_2) - 1\} \beta_2 \right\}^{-\frac{\alpha}{2}} \left\{ \frac{\beta_1}{\beta_2} + 1 \right\}^{\frac{1+\alpha}{2}} \beta_2^{\frac{1}{2}} \log(u_{3k})$$

$$\left. - \left\{ \{\psi'(\beta_2) - \psi'(\beta_1+\beta_2)\} \beta_2^2 \right\}^{-\frac{\alpha}{2}} \log(l_{3k}) \right\}^{-\frac{1}{\alpha}}$$

$$= \lim_{\beta_2 \to \infty} \left\{ \frac{\beta_2^{\frac{1}{2}}}{\sqrt{\beta_1+\beta_2}\log(u_{3k}) - \log(l_{3k})} \right.$$

$$\times \left\{ \{\psi'(\beta_2)(\beta_1+\beta_2) - 1\} \beta_2 \right\}^{-\frac{\alpha}{2}} \left\{ \frac{\beta_1}{\beta_2} + 1 \right\}^{\frac{1+\alpha}{2}} \log(u_{3k})$$

$$\left. - \left\{ \{\psi'(\beta_2) - \psi'(\beta_1+\beta_2)\} \beta_2^2 \right\}^{-\frac{\alpha}{2}} \beta_2^{-\frac{1}{2}} \log(l_{3k}) \right\}^{-\frac{1}{\alpha}}$$

$$= \left(\frac{1}{2} + \beta_1 \right)^{\frac{1}{2}}.$$

Then,

$$W_2^k(\beta_1) = \int_{l_{2k}}^{\varepsilon} \frac{1}{\beta_2} \{1 + O(\beta_2)\} \, \mathrm{d}\beta_2 + \int_{1/\varepsilon}^{u_{2k}} \frac{1}{\beta_2} \left\{ \left(\frac{1}{2} + \beta_1 \right)^{\frac{1}{2}} + O(\beta_2) \right\} \mathrm{d}\beta_2$$

$$+ \int_{\varepsilon}^{1/\varepsilon} g_{22}^k(\beta_1, \beta_2) \, \mathrm{d}\beta_2$$

$$= \left(\frac{1}{2} + \beta_1 \right)^{\frac{1}{2}} \log(u_{2k}) - \log(l_{2k}) + C_{22}(\varepsilon, \beta_1),$$

where

$$C_{22}^k(\beta_1) = \int_{\varepsilon}^{1/\varepsilon} g_{22}^k(\beta_1, \beta_2) \, \mathrm{d}\beta_2 + O(\varepsilon).$$

Therefore,

$$\pi_2^k(\beta_2, \beta_3 | \beta_1) = \pi_3^k(\beta_3 | \beta_1, \beta_2) \left\{ \{\psi'(\beta_2)(\beta_1 + \beta_2) - 1\}^{-\frac{\alpha}{2}} \{\beta_1 + \beta_2\}^{\frac{1+\alpha}{2}} \log(u_{3k}) \right.$$

$$\left. - \{\psi'(\beta_2) - \psi'(\beta_1 + \beta_2)\}^{-\frac{\alpha}{2}} \log(l_{3k}) + C_{21}(\varepsilon, \beta_1, \beta_2) \right\}^{-\frac{1}{\alpha}}$$

$$\times \left\{ \sqrt{\beta_1 + \beta_2} \log(u_{3k}) - \log(l_{3k}) + C_3(\varepsilon, \beta_1, \beta_2) \right\}^{\frac{1}{\alpha}}$$

$$\times \left\{ \left(\frac{1}{2} + \beta_1 \right)^{\frac{1}{2}} \log(u_{2k}) - \log(l_{2k}) + C_{22}(\varepsilon, \beta_1) \right\}^{-1}.$$

Finally, we have

$$\pi_1^k(\beta_1, \beta_2, \beta_3) = \frac{\pi_2^k(\beta_2, \beta_3 | \beta_1) \left\{ E_1^k \left[|h_1(\beta)|^{-\frac{\alpha}{2}} | \beta_{[1]} \right] \right\}^{-\frac{1}{\alpha}}}{\int_{l_{1k}}^{u_{1k}} \left\{ E_1^k \left[|h_1(\beta)|^{-\frac{\alpha}{2}} | \beta_{[1]} \right] \right\}^{-\frac{1}{\alpha}} \mathrm{d}\beta_1}.$$

Note that we need to investigate $E_1^k \left[|h_1(\beta)|^{-\frac{\alpha}{2}} | \beta_{[1]} \right]$ only as $\int_{l_{1k}}^{u_{1k}} \left\{ E_1^k \left[|h_1(\beta)|^{-\frac{\alpha}{2}} | \beta_{[1]} \right] \right\}^{-\frac{1}{\alpha}} \mathrm{d}\beta_1$

is a constant. From the algorithm in Section 3.3,

$$E_1^k\left[\left|h_1(\beta)\right|^{-\frac{\alpha}{2}}|\beta_{[1]}\right] = \int_{l_{3k}}^{u_{3k}}\int_{l_{2k}}^{l_{2k}} \{h_1(\beta)\}^{-\frac{\alpha}{2}} \pi_2^k(\beta_2,\beta_3|\beta_1)\,\mathrm{d}\beta_2\mathrm{d}\beta_3$$

$$= \frac{1}{W_2^k(\beta_1)}\int_{l_{2k}}^{u_{2k}}\int_{l_{3k}}^{u_{3k}} g_{11}(\beta_1,\beta_2,\beta_3)\,\mathrm{d}\beta_3\frac{g_{22}^k(\beta_1,\beta_2)}{W_3^k(\beta_1,\beta_2)}\mathrm{d}\beta_2,$$

where

$$g_{11}(\beta_1,\beta_2,\beta_3) =$$

$$\left\{\psi'(\beta_1) - \frac{\psi'(\beta_2)\psi'(\beta_3)\psi'(\beta_1+\beta_2+\beta_3)}{\psi'(\beta_2)\psi'(\beta_3) - \psi'(\beta_2)\psi'(\beta_1+\beta_2+\beta_3) - \psi'(\beta_3)\psi'(\beta_1+\beta_2+\beta_3)}\right\}^{-\frac{\alpha}{2}}$$

$$\times\{\psi'(\beta_3) - \psi'(\beta_1+\beta_2+\beta_3)\}^{\frac{1}{2}}.$$

For any ε small enough and k large enough such that $l_{2k} < \epsilon < 1/\epsilon < u_{2k}$, we expand $\int_{l_{3k}}^{u_{3k}} g_1(\beta_1,\beta_2,\beta_3)\,\mathrm{d}\beta_3$ as

$$\int_{l_{3k}}^{u_{3k}} g_{11}(\beta_1,\beta_2,\beta_3)\,\mathrm{d}\beta_3 = \int_{1/\varepsilon}^{u_{3k}} g_{11}(\beta_1,\beta_2,\beta_3)\,\mathrm{d}\beta_3 + \int_{\varepsilon}^{1/\varepsilon} g_{11}(\beta_1,\beta_2,\beta_3)\,\mathrm{d}\beta_3$$

$$+ \int_{l_{3k}}^{\varepsilon} g_{11}(\beta_1,\beta_2,\beta_3)\,\mathrm{d}\beta_3.$$

Then, by Lemma A.1 and Lemma B.1, we have

$$\int_{1/\varepsilon}^{u_{3k}} g_{11}\left(\beta_1, \beta_2, \beta_3\right) \mathrm{d}\beta_3$$

$$= \int_{1/\varepsilon}^{u_{3k}} \left\{\psi'\left(\beta_1\right) - \frac{\psi'\left(\beta_2\right)\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)}{\psi'\left(\beta_2\right)\psi'\left(\beta_3\right) - \psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2+\beta_3\right) - \psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)}\right\}^{-\frac{\alpha}{2}}$$

$$\times \left\{\psi'\left(\beta_3\right) - \psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}^{\frac{1}{2}} \mathrm{d}\beta_3$$

$$= \int_{1/\varepsilon}^{u_{3k}} \frac{1}{\beta_3} \left\{\psi'\left(\beta_1\right) - \frac{\psi'\left(\beta_2\right)\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2}{\psi'\left(\beta_2\right)\left\{\psi'\left(\beta_3\right) - \psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2 - \psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2}\right\}^{-\frac{\alpha}{2}}$$

$$\times \left\{\left\{\psi'\left(\beta_3\right) - \psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2\right\}^{\frac{1}{2}} \mathrm{d}\beta_3$$

$$= \int_{1/\varepsilon}^{u_{3k}} \frac{1}{\beta_3} \left\{\left\{\psi'\left(\beta_1\right) - \frac{\psi'\left(\beta_2\right)}{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right) - 1}\right\}^{-\frac{\alpha}{2}} \left(\beta_1+\beta_2\right)^{\frac{1}{2}} + O\left(\frac{1}{\beta_3}\right)\right\} \mathrm{d}\beta_3$$

$$= \left\{\psi'\left(\beta_1\right) - \frac{\psi'\left(\beta_2\right)}{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right) - 1}\right\}^{-\frac{\alpha}{2}} \left(\beta_1+\beta_2\right)^{\frac{1}{2}} \log\left(u_{3k}\right) + O\left(\varepsilon\right),$$

and

$$
\begin{aligned}
&\int_{l_{3k}}^{\varepsilon} g_{11}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3 \\
&=\int_{l_{3k}}^{\varepsilon}\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)}{\psi'\left(\beta_2\right)\psi'\left(\beta_3\right)-\psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)-\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)}\right\}^{-\frac{\alpha}{2}} \\
&\quad\times\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}^{\frac{1}{2}}\mathrm{d}\beta_3 \\
&=\int_{l_{3k}}^{\varepsilon}\frac{1}{\beta_3}\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2}{\psi'\left(\beta_2\right)\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2-\psi'\left(\beta_3\right)\psi'\left(\beta_1+\beta_2+\beta_3\right)\beta_3^2}\right\}^{-\frac{\alpha}{2}} \\
&\quad\times\left\{\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}\beta_3^2\right\}^{\frac{1}{2}}\mathrm{d}\beta_3 \\
&=\int_{l_{3k}}^{\varepsilon}\frac{1}{\beta_3}\left\{\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2\right)}{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)}\right\}^{-\frac{\alpha}{2}}+O\left(\beta_3\right)\right\}\mathrm{d}\beta_3 \\
&=-\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2\right)}{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)}\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right)+O\left(\varepsilon\right).
\end{aligned}
$$

Hence, we obtain

$$
\begin{aligned}
\int_{l_{3k}}^{u_{3k}} g_{11}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3 &= \left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)}{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1}\right\}^{-\frac{\alpha}{2}}\left(\beta_1+\beta_2\right)^{\frac{1}{2}}\log\left(u_{3k}\right) \\
&\quad-\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2\right)}{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)}\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right)+C_{11}\left(\varepsilon,\beta_1,\beta_2\right),
\end{aligned}
$$

where $C_{11}\left(\varepsilon,\beta_1,\beta_2\right)=\int_{\varepsilon}^{1/\varepsilon} g_{11}\left(\beta_1,\beta_2,\beta_3\right)\mathrm{d}\beta_3+O\left(\varepsilon\right)$.

Next, we have

$$
E_1^k\left[\left|h_1\left(\beta\right)\right|^{-\frac{\alpha}{2}}\left|\beta_{[1]}\right.\right]=\frac{\int_{l_{2k}}^{u_{2k}} g_{12}\left(\beta_1,\beta_2\right)\mathrm{d}\beta_2}{W_2^k\left(\beta_1\right)},
$$

where

$$g_{12}^{k}\left(\beta_1,\beta_2\right)=\left\{\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)}{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1}\right\}^{-\frac{\alpha}{2}}\left(\beta_1+\beta_2\right)^{\frac{1}{2}}\log\left(u_{3k}\right)\right.$$

$$\left.-\left\{\psi'\left(\beta_1\right)-\frac{\psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2\right)}{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)}\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right)+C_{11}\left(\varepsilon,\beta_1,\beta_2\right)\right\}$$

$$\times\left\{\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1+\alpha}{2}}\log\left(u_{3k}\right)\right.$$

$$\left.-\left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right)+C_{21}\left(\varepsilon,\beta_1,\beta_2\right)\right\}^{-\frac{1}{\alpha}}$$

$$\times\left\{\sqrt{\beta_1+\beta_2}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)+C_3\left(\varepsilon,\beta_1,\beta_2\right)\right\}^{\frac{1-\alpha}{\alpha}}.$$

For any ε small enough and k large enough such that $l_{2k}<\epsilon<1/\epsilon<u_{2k}$, we can neglect $C_{11}\left(\varepsilon,\beta_1,\beta_2\right)$, $C_{21}\left(\varepsilon,\beta_1,\beta_2\right)$ and $C_3\left(\varepsilon,\beta_1,\beta_2\right)$. According to Lemma A.1 and Lemma B.1, we obtain

$$\int_{l_{2k}}^{\varepsilon}g_{12}^{k}\left(\beta_1,\beta_2\right)\mathrm{d}\beta_2=\int_{l_{2k}}^{\varepsilon}\left\{\left\{\psi'\left(\beta_1\right)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\beta_1^{\frac{1}{2}}\log\left(u_{3k}\right)-\left\{\psi'\left(\beta_1\right)-\psi'\left(\beta_1\right)\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right)\right\}$$

$$\times\frac{1}{\beta_2}\left\{\beta_1^{-\frac{\alpha}{2}}\beta_1^{\frac{1+\alpha}{2}}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)\right\}^{-\frac{1}{\alpha}}$$

$$\times\left\{\sqrt{\beta_1}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)\right\}^{\frac{1-\alpha}{\alpha}}\mathrm{d}\beta_2+O\left(\varepsilon\right)$$

$$=-\left\{\left\{\psi'\left(\beta_1\right)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\beta_1^{\frac{1}{2}}\log\left(u_{3k}\right)\right\}$$

$$\times\left\{\sqrt{\beta_1}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)\right\}^{-1}\log\left(l_{2k}\right)+O(\varepsilon),$$

and

$$\int_{1/\varepsilon}^{u_{2k}} g_{12}^{k}\left(\beta_1,\beta_2\right)\mathrm{d}\beta_2$$

$$= \int_{1/\varepsilon}^{u_{2k}} \left\{\left\{\psi'\left(\beta_1\right) - \frac{\psi'\left(\beta_2\right)\beta_2}{\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}\beta_2}\right\}^{-\frac{\alpha}{2}}\left(\frac{\beta_1}{\beta_2}+1\right)^{\frac{1}{2}}\beta_2^{\frac{1}{2}}\log\left(u_{3k}\right)\right.$$

$$-\left\{\psi'\left(\beta_1\right) - \frac{\psi'\left(\beta_2\right)\psi'\left(\beta_1+\beta_2\right)\beta_2^2}{\left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}\beta_2^2}\right\}\log\left(l_{3k}\right)\Bigg\}$$

$$\times\left\{\left\{\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}\beta_2\right\}^{-\frac{\alpha}{2}}\beta_2^{\frac{\alpha}{2}}\left\{\frac{\beta_1}{\beta_2}+1\right\}^{\frac{1+\alpha}{2}}\beta_2^{\frac{1+\alpha}{2}}\log\left(u_{3k}\right)\right.$$

$$\left.-\beta_2^{\alpha}\left\{\left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}\beta_2^2\right\}^{-\frac{\alpha}{2}}\log\left(l_{3k}\right)\right\}^{-\frac{1}{\alpha}}$$

$$\times\left\{\sqrt{\frac{\beta_1}{\beta_2}+1}\beta_2^{\frac{1}{2}}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)\right\}^{\frac{1-\alpha}{\alpha}}\mathrm{d}\beta_2$$

$$= \int_{1/\varepsilon}^{u_{2k}} \frac{1}{\beta_2}\left\{\left\{\psi'\left(\beta_1\right)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\log\left(u_{3k}\right)\right\}\left\{\left\{\frac{1}{2}+\beta_1\right\}^{-\frac{\alpha}{2}}\log\left(u_{3k}\right)\right\}^{-\frac{1}{\alpha}}$$

$$\times\left\{\log\left(u_{3k}\right)\right\}^{\frac{1-\alpha}{\alpha}}\mathrm{d}\beta_2 + O(\varepsilon)$$

$$= \int_{1/\varepsilon}^{u_{2k}} \frac{1}{\beta_2}\left\{\psi'\left(\beta_1\right)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\mathrm{d}\beta_2 + O(\varepsilon)$$

$$= \left\{\psi'\left(\beta_1\right)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\log\left(u_{2k}\right) + O(\varepsilon).$$

Eventually, we have

$$\int_{l_{2k}}^{u_{2k}} g_{12}^{k}\left(\beta_1,\beta_2\right)\mathrm{d}\beta_2 = \left\{\psi'\left(\beta_1\right)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\log\left(u_{2k}\right) + C_{12}^{k}\left(\varepsilon,\beta_1\right)$$

$$-\left\{\psi'\left(\beta_1\right)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\sqrt{\beta_1}\log\left(u_{3k}\right)\left\{\sqrt{\beta_1}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)\right\}^{-1}\log\left(l_{2k}\right),$$

where

$$C_{12}^{k}\left(\varepsilon,\beta_1\right)=\int_{\varepsilon}^{1/\varepsilon} g_{12}^{k}\left(\beta_1,\beta_2\right)\mathrm{d}\beta_2+O\left(\varepsilon\right).$$

Therefore,

$$E_1^k\left[\left|h_1\left(\beta\right)\right|^{-\frac{\alpha}{2}}\left|\beta_{[1]}\right.\right]=\frac{1}{W_2^k\left(\beta_1\right)}\left\{\left\{\psi'\left(\beta_1\right)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\log\left(u_{2k}\right)+C_{12}^{k}\left(\varepsilon,\beta_1\right)\right.$$
$$\left.-\left\{\psi'\left(\beta_1\right)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\sqrt{\beta_1}\log\left(u_{3k}\right)\left\{\sqrt{\beta_1}\log\left(u_{3k}\right)-\log\left(l_{3k}\right)\right\}^{-1}\log\left(l_{2k}\right)\right\}.$$

By the last step of the algorithm in Section 3.3, we obtain the alpha prior by

$$\pi\left(\beta_1,\beta_2,\beta_3\right)=\lim_{k\to\infty}\frac{\pi_1^k\left(\beta_1,\beta_2,\beta_3\right)}{\pi_1^k\left(1,1,1\right)}.$$

If we choose that $l_{3k}=1/u_{3k}$ and $l_{2k}=1/u_{2k}$, the alpha prior $\pi\left(\beta_1,\beta_2,\beta_3\right)$ has the form

$$\pi\left(\beta_1,\beta_2,\beta_3\right)=\left\{\left\{\sqrt{\beta_1+\beta_2}+1\right\}\left\{1+\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}\right\}\right\}^{\frac{1-\alpha}{\alpha}}\left\{\psi'\left(\beta_3\right)-\psi'\left(\beta_1+\beta_2+\beta_3\right)\right\}^{\frac{1}{2}}$$
$$\times\left\{\left\{\psi'\left(\beta_2\right)\left(\beta_1+\beta_2\right)-1\right\}^{-\frac{\alpha}{2}}\left\{\beta_1+\beta_2\right\}^{\frac{1+\alpha}{2}}+\left\{\psi'\left(\beta_2\right)-\psi'\left(\beta_1+\beta_2\right)\right\}^{-\frac{\alpha}{2}}\right\}^{-\frac{1}{\alpha}}$$
$$\times\left\{\left\{\psi'\left(\beta_1\right)-\frac{1}{\frac{1}{2}+\beta_1}\right\}^{-\frac{\alpha}{2}}\left(\frac{1}{2}+\beta_1\right)^{\frac{1}{2}}+\left\{\psi'\left(\beta_1\right)-\frac{1}{\beta_1}\right\}^{-\frac{\alpha}{2}}\right\}^{-\frac{1}{\alpha}}.$$

Appendix C

The conditions (AI)-(AV) of Ghosh et al. (1982)

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having common distribution function $F(x, \theta)$, parametrized by $\theta \in \Theta$, Θ being an open interval of $\mathbb{R}$. Let $f(x, \theta)$ be the density of $F(x, \theta)$ w.r.t. a σ-finite measure μ. Let the parameter θ have a prior distribution which has density $p(\cdot)$ w.r.t. Lebesgue measure. Let $[a_0, b_0] \subset \Theta$ be such that $p(\theta) > 0$ on (a_0, b_0) and $p(\theta) = 0$ on $(a_0, b_0)^c$. Let $c < a_0$ and $d > b_0$ be such that $[c, d] \subset \Theta$. We need to make the following assumptions.

AI: $f(x, \theta)$ is measurable in x for each $\theta \in [c, d]$.

AII: $\int |f(x, \theta) - f(x, \theta')| \, \mathrm{d}\mu(x) > 0 \ \forall \theta, \theta' \in [c, d], \theta \neq \theta'$

AIII: For each x, $f(x, \theta)$ admits partial derivatives w.r.t. θ of order up to two which are continuous in $[c, d]$.

AIV:

1. For each x, $f(x, \theta)$ admits partial derivatives of order upto K+3$(K \geqslant 0)$ which are continuous in $[c, d]$.
2. For every $\theta \in [a_0, b_0]$ there exists a neighbourhood (nhb) U_θ such that

$$\sup_{\theta' \in U_\theta} \mathbb{E}_{\theta'} \left| \frac{\mathrm{d}^{\mathrm{i}}}{\mathrm{d}\theta^i} \log f(X, \theta') \right|^{r_1} < \infty, \quad 1 \leqslant i \leqslant K + 2.$$

3. For every $\theta \in [a_0, b_0]$,∃a nhbd U_θ and measurable functions $H_\theta(x)$ and $A_\theta(x)$ such that

   (a)

$$\left|\frac{\mathrm{d}^{\mathrm{K}+3}}{\mathrm{d}\theta^{K+3}} \log f(X, \theta')\right| \leqslant H_\theta(x),$$

   $\forall \theta' \in U_\theta$ and $\forall x$,

   (b)

$$\left|\frac{\mathrm{d}^{\mathrm{K}+3}}{\mathrm{d}\theta^{K+3}} \log f(X, \theta') - \frac{\mathrm{d}^{\mathrm{K}+3}}{\mathrm{d}_\theta^{K+3}} \log f(X, \theta'')\right| \leqslant |\theta' - \theta''| A_\theta(x), \quad \forall \theta', \theta'' \in U_\theta \text{ and } \forall x.$$

   (c)

$$\sup_{\theta' \in U_\theta} E_{\theta'}\left(H_{\theta'}(x)^{r_1}\right) < \infty.$$

   (d)

$$\sup_{\theta' \in U_\theta} E_{\theta'}\left(A_{\theta'}(x)^{r_1}\right) < \infty.$$

AV:

1. For every $\theta \in (c, d)$ ∃ a nhdb U_θ such that

$$\sup_{\theta' \in U_\theta} \mathbb{E}_{\theta'}\left(|\log f(X, \theta')|^{r_2+1}\right) < \infty.$$

2.

$$\mathbb{E}_\theta\left(\frac{\mathrm{d}}{\mathrm{d}\theta} \log f(X, \theta)\right) = 0, \forall \theta \in (c, d).$$

3. For every $\theta \in [c, d]$, $\theta' \in (c, d)$, ∃ nhbds V_θ and W_θ such that for all nhbds $V \subset V_\theta$ of θ,

$$\sup_{\alpha \in W_{\theta'}} \mathbb{E}_\alpha\left(\left|\sup_{\alpha \in V} \log f(X, \theta')\right|^{r_2+1}\right) < \infty.$$

4. For every $\theta \in (c, d)$

(a)

$$I(\theta) = E_\theta\left(-\frac{d^2}{d\theta^2}\log f(X,\theta)\right) > 0.$$

(b)

$$I_1(\theta) = E_\theta\left(-\frac{d}{d\theta}\log f(X,\theta)\right)^2 > 0.$$

5. $I(\theta)$and $I_1(\theta)$ are continuous on (c,d).

6. For every $\theta \in (c,d)$, $\exists$ a nhbd U_θ and a measurable function $m(x,\theta)$ such that

   (a)

$$\left|\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X,\theta') - \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X,\theta'')\right| \leqslant |\theta' - \theta''|\, m(x,\theta), \quad \forall\theta',\theta'' \in U_\theta \text{ and } \forall x.$$

   (b)

$$\sup_{\theta'\in U_\theta} \mathbb{E}_{\theta'}\left(m(x,\theta)^{r_2+1}\right) < \infty$$

   (c)

$$\sup_{\theta'\in U_\theta} \mathbb{E}_{\theta'}\left(\left|\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log f(X,\theta')\right|^{r_2+1}\right) < \infty$$

Note than if AIV holds, then on $[a_0, b_0]$ $I(\theta)$ equals $I_1(\theta)$ and is continuous.

Bibliography

Amari, S. (1985), *Differential-Geometrical Methods in Statistics*, Springer Verlag: Berlin, Germany.

Bar-Lev, S. K. and Reiser, B. (1982), "An exponential subfamily which admits UMPU tests based on a single test statistic," *The Annals of Statistics*, 10, 979–989.

Batir, N. (2008), "On some properties of the gamma function," *Expositiones Mathematicae*, 26, 187–196.

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

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

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

Berger, J. O. and Moreno, E. (1994), "Bayesian robustness in bidimensional models: Prior independence," *Journal of Statistical Planning and Inference*, 40, 161–176.

Berger, J. O. and Sun, D. (2008), "Objective priors for the bivariate normal model," *The Annals of Statistics*, 36, 963–982.

Berger, J. O., De Oliveira, V., and Sansó, B. (2001), "Objective Bayesian analysis of spatially correlated data," *Journal of the American Statistical Association*, 96, 1361–1374.

Berger, J. O., Bernardo, J. M., and Sun, D. (2015), "Overall objective priors," *Bayesian Analysis*, 10, 189–221.

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

Bernardo, J. M. (2005), "Reference Analysis," in *Bayesian Thinking Modeling and Computation*, eds. D. Dey and C. Rao, vol. 25 of *Handbook of Statistics*, pp. 17 – 90, Elsevier.

Bownan, K. O. and Shenton, L. R. (1982), "Properties of estimators for the gamma distribution," *Communications in Statistics - Simulation and Computation*, 11, 377–519.

Chen, M.-H. and Shao, Q.-M. (1999), "Monte Carlo estimation of Bayesian credible and HPD intervals," *Journal of Computational and Graphical Statistics*, 8, 69–92.

Cichocki, A. and Amari, S. (2010), "Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities," *Entropy*, 12, 1532–1568.

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

Clarke, B. and Sun, D. (1997), "Reference priors under the chi-squared distance," *Sankhyā: The Indian Journal of Statistics, Series A*, 59, 215–231.

Ghosh, J. and Mukerjee, R. (1992), "Noninformative priors (with discussion)," in *Bayesian Statistics*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, vol. 4, pp. 195–210, Oxford University Press.

Ghosh, J., Sinha, B., and Joshi, S. (1982), "Expansion for posterior probability and integrated Bayes risk," in *Statistical Decision Theory and Related Topics*, eds. S. Gupta and J. Berger, vol. 1 of *III*, pp. 403–456, Academic Press.

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

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.

Jeffreys, H. (1961), *Probability Theory*, Oxford University Press, New York.

Kass, R. E. and Greenhouse, J. B. (1989), "[Investigating therapies of potentially great benefit: ECMO]: Comment: A Bayesian perspective," *Statistical Science*, 4, 310–317.

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

Lavine, M., Wasserman, L., and Wolpert, R. L. (1991), "Bayesian inference with specified prior marginals," *Journal of the American Statistical Association*, 86, 964–971.

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

Neyman, J. and Scott, E. L. (1958), "Statistical approach to problems of cosmology," *Journal of the Royal Statistical Society. Series B (Methodological)*, 20, 1–43.

Shemyakin, A. (2014), "Hellinger distance and non-informative priors," *Bayesian Analysis*, 9, 923–938.

Stein, C. M. (1985), "On the coverage probability of confidence sets based on a prior distribution," in *Sequential Methods in Statistics*, ed. R. Zielinski, vol. 16, pp. 485–514, PWN Polish Scientific Publishers.

Sun, D. and Berger, J. O. (1998), "Reference priors with partial information," *Biometrika*, 85, 55–71.

Ware, J. H. (1989), "Investigating therapies of potentially great benefit: ECMO," *Statistical Science*, 4, 298–306.