

# Noninformative priors for common scale parameter in the regular Pareto distributions

Sang Gil Kang<sup>1</sup> · Dal Ho Kim<sup>2</sup> · Yongku Kim<sup>3</sup>

<sup>1</sup>Department of Computer and Data Information, Sangji University

<sup>2</sup>Department of Statistics, Kyungpook National University

<sup>3</sup>Department of Statistics, Yeungnam University

Received 6 January 2012, revised 1 February 2012, accepted 13 February 2012

## Abstract

In this paper, we introduce the noninformative priors such as the matching priors and the reference priors for the common scale parameter in the Pareto distributions. It turns out that the posterior distribution under the reference priors is not proper, and Jeffreys' prior is not a matching prior. It is shown that the proposed first order prior matches the target coverage probabilities in a frequentist sense through simulation study.

*Keywords:* Common scale parameter, matching prior, Pareto distribution, reference prior.

## 1. Introduction

The Pareto distribution is a widely used statistical model for an extensive variety of applications such as studies of income, property values, insurance risk, stock prices fluctuations, migration, size of cities and firms, word frequencies, occurrences of natural resources, business failures, service time in queuing systems, error clustering in communications circuits and lifetime data, etc (Arnold and Press, 1983; Fernández, 2008). In a Bayesian point of view, many authors have studied statistical inferences on Pareto distribution (e.g., Arnold and Press, 1983, 1989; Geisser, 1984, 1985; Lwin, 1972; Nigm and Hamdy, 1987; Tiwari, Yang and Zalkikar, 1996; Ko and Kim, 1999; Fernández, 2008; Kim *et al.*, 2009; Kang, 2010).

Consider  $X$  and  $Y$  to be independently distributed random variables with the Pareto distribution  $\mathcal{P}(\alpha, \lambda)$  with the shape parameter  $\alpha$  and the scale parameter  $\lambda$ , and the Pareto distribution  $\mathcal{P}(\beta, \lambda)$  with the shape parameter  $\beta$  and the scale parameter  $\lambda$ , respectively. That is, the parameter  $\lambda$  is the common scale parameter. We focus on developing noninformative

---

<sup>1</sup> Associate professor, Department of Computer and Data Information, Sangji University, Wonju 220-702, Korea.

<sup>2</sup> Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.

<sup>3</sup> Corresponding author: Assistant professor, Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea. E-mail: ykkim@yu.ac.kr

priors for the common scale parameter in Pareto distributions. If sufficient information from past experience, expert opinion or previously collected data is available, subjective priors are ideal. Otherwise noninformative or default priors are to be considered.

The notion of a noninformative prior has attracted much attention in recent years. There are different notions of noninformative prior. Welch and Peers (1963) introduced a probability matching prior which matches the posterior and frequentist probabilities of confidence intervals. Interest in such priors revived with the work of Stein (1985) and Tibshirani (1989). Among others, we may cite the work of Mukerjee and Dey (1993), DiCiccio and Stern (1994), Datta and Ghosh (1995a, 1995b, 1996), Mukerjee and Ghosh (1997). Bernardo (1979) initiated the reference prior approach to maximize the expected Kullback-Leibler divergence of the posterior distribution relative to the prior. Ghosh and Mukerjee (1992), and Berger and Bernardo (1989, 1992) give a general algorithm to derive a reference prior by splitting the parameters into several groups according to their order of inferential importance. This approach is very successful in various practical problems. Quite often reference priors satisfy the matching criterion described earlier.

The Pareto distribution is reverse  $J$ -shaped and positively skewed with a decreasing hazard rate. Although the family was originally applied to analyze socio-economic and natural phenomena with long tail, the family has potential for modeling reliability and lifetime data as well (Arnold and Press, 1983). The Pareto distribution has been used by many authors in a Bayesian viewpoint (e.g., Arnold and Press, 1983, 1989; Geisser, 1984, 1985; Lwin, 1972; Nigm and Hamdy, 1987; Tiwari *et al.*, 1996; Ko and Kim, 1999; Fernández, 2008; Kim *et al.*, 2009; Kang, 2010). For the common scale parameter, Elfessi and Jin (1996) derived a class of improved estimators which uniformly dominates the MLE under a class of convex scale invariant loss functions. However the objective Bayesian inference for the common scale parameter is not fully considered. Therefore there is a necessity for developing the noninformative prior to the objective Bayesian inference of the common scale parameter.

The outline of the remaining sections is as follows. In Section 2, we develop first order and second order probability matching priors for the common scale parameter. The reference priors for the common scale parameter are also derived. However the posterior distribution under the reference is not proper. It turns out that Jeffreys' prior is not a first order matching prior. We provide the propriety of the posterior distribution for the general prior including the matching and Jeffreys' priors. In Section 4, we explore the frequentist coverage probabilities under the proposed prior.

## 2. The noninformative priors

Let  $X$  and  $Y$  be random variables from two independent Pareto distributions with the common scale parameter  $\lambda$ , and shape parameters  $\alpha$  and  $\beta$ , respectively. The Pareto probability density functions of  $X$  and  $Y$  are given by

$$f(x|\alpha, \lambda) = \alpha\lambda(1 + \lambda x)^{-(\alpha+1)}, x > 0, \alpha > 0, \lambda > 0, \quad (2.1)$$

and

$$f(y|\beta, \lambda) = \beta\lambda(1 + \lambda y)^{-(\beta+1)}, y > 0, \beta > 0, \lambda > 0. \quad (2.2)$$

respectively. Let  $x_i, i = 1, \dots, n$  denote observations from the Pareto distribution  $\mathcal{P}(\alpha, \lambda)$ , and  $y_i, i = 1, \dots, m$  denote observations from the Pareto distribution  $\mathcal{P}(\beta, \lambda)$ . Then corre-

sponding likelihood function of  $\lambda, \alpha$  and  $\beta$  is given by

$$L(\lambda, \alpha, \beta | \mathbf{x}, \mathbf{y}) = \alpha^n \beta^m \lambda^{n+m} \prod_{i=1}^n (1 + \lambda x_i)^{-(\alpha+1)} \prod_{i=1}^m (1 + \lambda y_i)^{-(\beta+1)}, \tag{2.3}$$

where  $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_m)$ . The common scale parameter  $\lambda$  is parameter of interest. Now we develop the noninformative priors for  $\lambda$ .

**2.1. The reference priors**

Reference priors introduced by Bernardo (1979), and extended further by Berger and Bernardo (1992) have become very popular over the years for the development of noninformative priors. We derive the reference priors for different groups of ordering of  $(\lambda, \alpha, \beta)$  by the algorithm of Berger and Bernardo (1992).

From the likelihood (2.3), the Fisher information matrix is given by

$$I(\lambda, \alpha, \beta) = \begin{pmatrix} \frac{1}{\lambda^2} \left( \frac{n\alpha}{\alpha+2} + \frac{m\beta}{\beta+2} \right) & \frac{n}{\lambda(\alpha+1)} & \frac{m}{\lambda(\beta+1)} \\ \frac{n}{\lambda(\alpha+1)} & \frac{n}{\alpha^2} & 0 \\ \frac{m}{\lambda(\beta+1)} & 0 & \frac{m}{\beta^2} \end{pmatrix}. \tag{2.4}$$

Firstly, we derived the two group reference prior for the parameter grouping  $\{\lambda, (\alpha, \beta)\}$ . The compact subsets were taken to be Cartesian products of sets of the form

$$\lambda \in [a_1, b_1], \alpha \in [a_2, b_2], \beta \in [a_3, b_3]. \tag{2.5}$$

In the limit  $a_i, i = 1, 2, 3$  will tend to 0, and  $b_i, i = 1, 2, 3$ , will tend to  $\infty$ . For the derivation of the reference prior, from the Fisher information (2.4),

$$h_1 = \frac{1}{\lambda^2} \left( \frac{n\alpha}{(\alpha+1)^2(\alpha+2)} + \frac{m\beta}{(\beta+1)^2(\beta+2)} \right) \text{ and } h_2 = \frac{nm}{\alpha^2\beta^2}.$$

Here, and below, a subscripted  $Q$  denotes a function that is constant and does not depend on any parameters but any  $Q$  may depend on the ranges of the parameters.

*Step 1.* Note that

$$\int_{a_3}^{b_3} \int_{a_2}^{b_2} h_2^{1/2} d\alpha d\beta = \int_{a_3}^{b_3} \int_{a_2}^{b_2} \left( \frac{nm}{\alpha^2\beta^2} \right)^{1/2} d\alpha d\beta = (nm)^{1/2} \log \left( \frac{b_2}{a_2} \right) \log \left( \frac{b_3}{a_3} \right).$$

It follows that

$$\pi_2^l(\alpha, \beta | \lambda) = Q_1^{-1} \alpha^{-1} \beta^{-1},$$

where  $Q_1 = \log(b_2/a_2) \log(b_3/a_3)$ .

*Step 2.* Now

$$\begin{aligned} E^l \{ \log h_1 | \alpha, \beta \} &= \int_{a_3}^{b_3} \int_{a_2}^{b_2} Q_1^{-1} \alpha^{-1} \beta^{-1} \log \left[ \frac{1}{\lambda^2} \left( \frac{n\alpha}{(\alpha+1)^2(\alpha+2)} + \frac{m\beta}{(\beta+1)^2(\beta+2)} \right) \right] d\alpha d\beta \\ &= Q_2 + \log \lambda^{-2}. \end{aligned}$$

It follows that

$$\pi_1^l(\lambda) \propto \exp[E^l\{\log h_1|\alpha, \beta\}/2] = \exp\{Q_2/2\}\lambda^{-1}.$$

Therefore the reference prior is

$$\pi_1(\lambda, \alpha, \beta) = \lim_{l \rightarrow \infty} \frac{\pi_2^l(\alpha, \beta|\lambda)\pi_1^l(\lambda)}{\pi_2^l(\alpha_0, \beta_0|\lambda_0)\pi_1^l(\lambda_0)} \propto \lambda^{-1}\alpha^{-1}\beta^{-1}, \tag{2.6}$$

where  $\lambda_0, \alpha_0$  and  $\beta_0$  are an inner point of the interval  $(0, \infty)$ .

Nextly, we derived the one-at-a-time reference prior for the parameter grouping  $\{\lambda, \alpha, \beta\}$  For the derivation of the reference prior, from the Fisher information (2.4),

$$h_1 = \frac{1}{\lambda^2} \left( \frac{n\alpha(\beta + 1)^2(\beta + 2) + m\beta(\alpha + 1)^2(\alpha + 2)}{(\alpha + 1)^2(\alpha + 2)(\beta + 1)^2(\beta + 2)} \right), h_2 = \frac{n}{\alpha^2} \text{ and } h_3 = \frac{m}{\beta^2}.$$

*Step 1.* Note that

$$\int_{a_3}^{b_3} h_3^{1/2} d\beta = \int_{a_3}^{b_3} \left( \frac{m}{\beta^2} \right)^{1/2} d\beta = m^{1/2} \log \left( \frac{b_3}{a_3} \right).$$

It follows that

$$\pi_3^l(\beta|\lambda, \alpha) = Q_1^{-1}\beta^{-1},$$

where  $Q_1 = \log(b_3/a_3)$ .

*Step 2.* Now

$$E^l\{\log h_2|\lambda, \alpha\} = \int_{a_3}^{b_3} Q_1^{-1}\beta^{-1} \log \left( \frac{n}{\alpha^2} \right) d\beta = \log \left( \frac{n}{\alpha^2} \right).$$

It follows that

$$\int_{a_2}^{b_2} \exp[E^l\{\log h_2|\lambda, \alpha\}/2] d\alpha = n^{1/2} \log \left( \frac{b_2}{a_2} \right).$$

Hence  $\pi_2^l(\alpha, \beta|\lambda) = Q_2^{-1}\alpha^{-1}\beta^{-1}$ , where  $Q_2 = \log(b_2/a_2) \log(b_3/a_3)$ .

*Step 3.* Now

$$\begin{aligned} & E_\theta^l\{\log h_1|\lambda\} \\ &= \int_{a_2}^{b_2} \int_{a_3}^{b_3} K_2^{-1}\alpha^{-1}\beta^{-1} \log \left\{ \frac{1}{\lambda^2} \left( \frac{n\alpha(\beta + 1)^2(\beta + 2) + m\beta(\alpha + 1)^2(\alpha + 2)}{(\alpha + 1)^2(\alpha + 2)(\beta + 1)^2(\beta + 2)} \right) \right\} d\beta d\alpha \\ &= Q_3 + \log \lambda^{-2}. \end{aligned}$$

So

$$\int_{a_1}^{b_1} \exp[E_\lambda^l\{\log h_1|\theta_1\}/2] d\lambda = \exp \left\{ \frac{Q_3}{2} \right\} \log \left( \frac{b_1}{a_1} \right).$$

Thus  $\pi_1^l(\lambda, \alpha, \beta) = Q_3^{-1} \lambda^{-1} \alpha^{-1} \beta^{-1}$ , where  $Q_3 = \log(b_1/a_1) \log(b_2/a_2) \log(b_3/a_3)$ . The reference prior is thus

$$\pi_1(\lambda, \alpha, \beta) \propto \lambda^{-1} \alpha^{-1} \beta^{-1}. \tag{2.7}$$

Note that the one-at-a-time reference prior and the two group reference prior are the same.

**Remark 2.1.** From the Fisher information (2.4), Jeffreys' prior is given by

$$\pi_J(\lambda, \alpha, \beta) \propto \lambda^{-1} \alpha^{-1} \beta^{-1} \left[ \frac{n\alpha}{(1+\alpha)^2(2+\alpha)} + \frac{m\beta}{(1+\beta)^2(2+\beta)} \right]^{\frac{1}{2}}. \tag{2.8}$$

**2.2. The probability matching priors**

For a prior  $\pi$ , let  $\theta_1^{1-\alpha}(\pi; \mathbf{X})$  denote the  $(1 - \alpha)$ th percentile of the posterior distribution of  $\theta_1$ , that is,

$$P^\pi[\theta_1 \leq \theta_1^{1-\alpha}(\pi; \mathbf{X}) | \mathbf{X}] = 1 - \alpha, \tag{2.9}$$

where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_t)^T$  and  $\theta_1$  is the parameter of interest. We want to find priors  $\pi$  for which

$$P[\theta_1 \leq \theta_1^{1-\alpha}(\pi; \mathbf{X}) | \boldsymbol{\theta}] = 1 - \alpha + o(n^{-u}). \tag{2.10}$$

for some  $u > 0$ , as  $n$  goes to infinity. Priors  $\pi$  satisfying (2.10) are called matching priors. If  $u = 1/2$ , then  $\pi$  is referred to as a first order matching prior, while if  $u = 1$ ,  $\pi$  is referred to as a second order matching prior.

In order to find such matching priors  $\pi$ , let

$$\theta_1 = \lambda, \theta_2 = \log(\alpha\lambda) - \alpha^{-1} \text{ and } \theta_3 = \log(\beta\lambda) - \beta^{-1}.$$

The Jacobian matrix of this transformation is

$$\frac{\partial(\theta_1, \theta_2, \theta_3)}{\partial(\lambda, \alpha, \beta)} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{\lambda} & \frac{1}{\alpha^2} + \frac{1}{\alpha} & 0 \\ \frac{1}{\lambda} & 0 & \frac{1}{\beta^2} + \frac{1}{\beta} \end{pmatrix}. \tag{2.11}$$

Therefore the inverse of the expected Fisher information matrix can be written as

$$\begin{aligned} I^{-1}(\theta_1, \theta_2, \theta_3) &= \left( \frac{\partial(\theta_1, \theta_2, \theta_3)}{\partial(\lambda, \alpha, \beta)} \right) I^{-1}(\lambda, \alpha, \beta) \left( \frac{\partial(\theta_1, \theta_2, \theta_3)}{\partial(\lambda, \alpha, \beta)} \right)^t \\ &= \begin{pmatrix} \left[ \frac{n\alpha}{(1+\alpha)^2(2+\alpha)} + \frac{m\beta}{(1+\beta)^2(2+\beta)} \right]^{-1} \lambda^2 & 0 & 0 \\ 0 & \frac{(1+\alpha)^2}{n\alpha^2} & 0 \\ 0 & 0 & \frac{(1+\beta)^2}{m\beta^2} \end{pmatrix}. \end{aligned} \tag{2.12}$$

By (2.12), the Fisher information matrix is

$$I(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} \left[ \frac{n\alpha}{(1+\alpha)^2(2+\alpha)} + \frac{m\beta}{(1+\beta)^2(2+\beta)} \right] \lambda^{-2} & 0 & 0 \\ 0 & \frac{n\alpha^2}{(1+\alpha)^2} & 0 \\ 0 & 0 & \frac{m\beta^2}{(1+\beta)^2} \end{pmatrix}. \tag{2.13}$$

Thus  $\theta_1$  is orthogonal to  $\theta_2$  and  $\theta_3$  in the sense of Cox and Reid (1987). Following Tibshirani (1989), the class of first order probability matching prior is characterized by

$$\pi_m(\theta_1, \theta_2, \theta_3) \propto \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^{\frac{1}{2}} \lambda^{-1} g(\theta_2, \theta_3), \tag{2.14}$$

where  $g(\theta_2, \theta_3) > 0$  is an arbitrary function differentiable in its argument. We may also note that the matching prior in the original parametrization  $(\lambda, \alpha, \beta)$  is given by

$$\begin{aligned} \pi_m(\lambda, \alpha, \beta) &\propto \frac{(1 + \alpha)(1 + \beta)}{\lambda\alpha^2\beta^2} \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^{\frac{1}{2}} \\ &\times g(\log(\alpha\lambda) - \alpha^{-1}, \log(\beta\lambda) - \beta^{-1}). \end{aligned} \tag{2.15}$$

Notice that the matching priors (2.15) include many different matching priors because of the arbitrary selection of the function  $g$ . And for some functions, there does not seem to be any improvement in the coverage probabilities with these posteriors. So we consider a particular first order matching prior where  $g$  is a constant in matching priors (2.15). This prior is given by

$$\pi_m(\lambda, \alpha, \beta) \propto \frac{(1 + \alpha)(1 + \beta)}{\lambda\alpha^2\beta^2} \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^{\frac{1}{2}}. \tag{2.16}$$

### 3. Implementation of the Bayesian procedure

We investigate the propriety of posteriors for a general class of priors which include Jeffreys' prior (2.8) and the matching prior (2.16). We consider the class of priors

$$\pi(\lambda, \alpha, \beta) \propto \lambda^{-1} \alpha^{-a} \beta^{-a} (1 + \alpha)^b (1 + \beta)^b \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^c, \tag{3.1}$$

where  $a > 0, b \geq 0$  and  $c > 0$ . The following general theorem can be proved.

**Theorem 3.1** The posterior distribution of  $(\lambda, \alpha, \beta)$  under the general prior, (3.1), is proper if  $n - a + c > 0, m - a + c > 0$  and  $a + 2c - b > 1$ .

**Proof:** Note that the joint posterior for  $\lambda, \alpha$  and  $\beta$  given  $\mathbf{x}$  and  $\mathbf{y}$  is

$$\begin{aligned} &\pi(\lambda, \alpha, \beta | \mathbf{x}, \mathbf{y}) \\ &\propto \lambda^{n+m-1} \alpha^{n-a} \beta^{m-a} (1 + \alpha)^b (1 + \beta)^b \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^c \\ &\times \prod_{i=1}^n (1 + \lambda x_i)^{-(\alpha+1)} \prod_{i=1}^m (1 + \lambda y_i)^{-(\beta+1)} \\ &\leq \lambda^{n+m-1} \alpha^{n-a} \beta^{m-a} (1 + \alpha)^b (1 + \beta)^b \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^c \\ &\times (1 + \lambda z_1)^{-n(\alpha+1) - m(\beta+1)} \equiv \pi'(\lambda, \alpha, \beta | \mathbf{x}, \mathbf{y}), \end{aligned} \tag{3.2}$$

where  $z_1 = \min\{x_1, \dots, x_n, y_1, \dots, y_m\}$ . Then integrating with respect to  $\lambda$  in (3.2), we can get

$$\begin{aligned} \pi'(\alpha, \beta | \mathbf{x}, \mathbf{y}) &\propto \alpha^{n-a} \beta^{m-a} (1 + \alpha)^b (1 + \beta)^b \left[ \frac{n\alpha}{(1 + \alpha)^2(2 + \alpha)} + \frac{m\beta}{(1 + \beta)^2(2 + \beta)} \right]^c \\ &\quad \times \frac{1}{\prod_{j=0}^{n+m-1} (n\alpha + m\beta + j)}. \end{aligned} \tag{3.3}$$

Now

$$\begin{aligned} \pi'(\alpha, \beta | \mathbf{x}, \mathbf{y}) &\leq c_1 \alpha^{n-a} \beta^{m-a} (1 + \alpha)^b (1 + \beta)^b \frac{\alpha^c \beta^c}{(1 + \alpha)^{2c} (2 + \alpha)^c (1 + \beta)^{2c} (2 + \beta)^c} \\ &\quad \times \frac{1}{\prod_{j=0}^{n+m-1} (n\alpha + m\beta + j)} \\ &\leq c_2 \alpha^{n-a+c-1} \beta^{m-a+c-1} (1 + \alpha)^{-(n-1)+b-3c} (1 + \beta)^{-(m-1)+b-3c} \\ &\equiv \pi''(\alpha, \beta | \mathbf{x}, \mathbf{y}), \end{aligned} \tag{3.4}$$

where  $c_1$  and  $c_2$  are a constant. Thus the posterior (3.4) is proper if  $n - a + c > 0, m - a + c > 0$  and  $a + 2c - b > 1$ . This completes the proof.  $\square$

**Theorem 3.2** The posterior distribution of  $(\lambda, \alpha, \beta)$  under the reference prior, (2.7), is improper.

**Proof:** Note that the joint posterior for  $\lambda, \alpha$  and  $\beta$  given  $\mathbf{x}$  and  $\mathbf{y}$  is

$$\begin{aligned} \pi(\lambda, \alpha, \beta | \mathbf{x}, \mathbf{y}) &\propto \lambda^{n+m-1} \alpha^{n-1} \beta^{m-1} \prod_{i=1}^n (1 + \lambda x_i)^{-(\alpha+1)} \prod_{i=1}^m (1 + \lambda y_i)^{-(\beta+1)} \\ &\geq \lambda^{n+m-1} \alpha^{n-1} \beta^{m-1} (1 + \lambda z)^{-n(\alpha+1) - m(\beta+1)} \equiv \pi'(\lambda, \alpha, \beta | \mathbf{x}, \mathbf{y}), \end{aligned} \tag{3.5}$$

where  $z_2 = \max\{x_1, \dots, x_m, y_1, \dots, y_m\}$ . Then integrating with respect to  $\lambda$  in (3.5), we can get

$$\begin{aligned} \pi'(\alpha, \beta | \mathbf{x}, \mathbf{y}) &\propto \alpha^{n-1} \beta^{m-1} \prod_{j=0}^{n+m-1} (n\alpha + m\beta + j)^{-1} \\ &\geq \alpha^{n-1} \beta^{m-1} (n\alpha + m\beta + n + m - 1)^{-(n+m)} \equiv \pi''(\alpha, \beta | \mathbf{x}, \mathbf{y}). \end{aligned} \tag{3.6}$$

Then integrating with respect to  $\alpha$  in (3.6), we can get

$$\pi''(\beta | \mathbf{x}, \mathbf{y}) \propto \beta^{m-1} (m\beta + n + m - 1)^{-m}. \tag{3.7}$$

Thus

$$\int_0^\infty \pi''(\beta | \mathbf{x}, \mathbf{y}) d\beta = \infty. \tag{3.8}$$

This completes the proof.  $\square$

**Theorem 3.3** Under the prior (3.1), the marginal posterior density of  $\lambda$  is given by

$$\begin{aligned} & \pi(\lambda|\mathbf{x}, \mathbf{y}) \\ & \propto \int_0^\infty \int_0^\infty \lambda^{n+m-1} \alpha^{n-a} \beta^{m-a} (1+\alpha)^b (1+\beta)^b \left[ \frac{n\alpha}{(1+\alpha)^2(2+\alpha)} + \frac{m\beta}{(1+\beta)^2(2+\beta)} \right]^c \\ & \times \prod_{i=1}^n (1+\lambda x_i)^{-(\alpha+1)} \prod_{i=1}^m (1+\lambda y_i)^{-(\beta+1)} d\alpha d\beta. \end{aligned} \quad (3.9)$$

Note that the marginal density of  $\lambda$  requires three dimensional integration. Therefore we have the marginal posterior density of  $\lambda$ , and compute the marginal moment of  $\lambda$  which is a Bayes estimator of  $\lambda$  under the squared error loss function.

#### 4. Numerical studies

We evaluate the frequentist coverage probability by investigating the credible interval of the marginal posterior density of  $\lambda$  under the reference prior given in Section 3 for several configurations  $(\lambda, \alpha, \beta)$  and  $(n, m)$ . That is to say, the frequentist coverage of a  $(1 - \eta)$ th posterior quantile should be close to  $1 - \eta$ . This is done numerically. Table 4.1 gives numerical values of the frequentist coverage probabilities of 0.05 (0.95) posterior quantiles for the proposed prior. The computation of these numerical values is based on the following algorithm for any fixed  $(\lambda, \alpha, \beta)$  and any prespecified probability value  $\eta$ . Here  $\eta$  is 0.05 (0.95). Let  $\lambda^\pi(\eta|\mathbf{X}, \mathbf{Y})$  be the  $\eta$ th posterior quantile of  $\lambda$  given  $\mathbf{X}$  and  $\mathbf{Y}$ . That is,  $F(\lambda^\pi(\eta|\mathbf{X}, \mathbf{Y})|\mathbf{X}, \mathbf{Y}) = \eta$ , where  $F(\cdot|\mathbf{X}, \mathbf{Y})$  is the marginal posterior distribution of  $\lambda$ . Then the frequentist coverage probability of this one sided credible interval of  $\lambda$  is

$$P_{(\lambda, \alpha, \beta)}(\eta; \lambda) = P_{(\lambda, \alpha, \beta)}(0 < \lambda \leq \lambda^\pi(\eta|\mathbf{X}, \mathbf{Y})). \quad (4.1)$$

The estimated  $P_{(\lambda, \alpha, \beta)}(\eta; \lambda)$  when  $\eta = 0.05(0.95)$  is shown in Table 4.1. In particular, for fixed  $n, m$  and  $(\lambda, \alpha, \beta)$ , we take 10,000 independent random samples of  $\mathbf{X} = (X_1, \dots, X_n)$  and  $\mathbf{Y} = (Y_1, \dots, Y_m)$  from the Pareto distributions, respectively.

For the cases presented in Table 4.1, we see that the matching prior  $\pi_m$  matches the target coverage probability much more accurately than Jeffreys' prior  $\pi_J$  for values of  $(\lambda, \alpha, \beta)$  and values of  $(n, m)$ . Note that Jeffreys' prior  $\pi_J$  does not satisfy a first order matching criterion. Thus we recommend to use the matching prior  $\pi_m$  in the sense of asymptotic frequentist coverage property.



**Table 4.1** Frequentist coverage probability of 0.05 (0.95) & posterior quantiles of  $\lambda$

$\alpha, \beta$	$\lambda$	$(n, m)$	$\pi_I$	$\pi_m$
0.5, 0.5	0.5	10,10	0.036 (0.927)	0.044 (0.940)
		10,20	0.040 (0.927)	0.045 (0.938)
		20,20	0.037 (0.934)	0.043 (0.942)
		20,30	0.042 (0.939)	0.046 (0.945)
	1.0	10,10	0.041 (0.938)	0.046 (0.944)
		10,20	0.035 (0.925)	0.043 (0.937)
		20,20	0.044 (0.933)	0.047 (0.943)
		20,30	0.042 (0.936)	0.047 (0.943)
	1.5	10,10	0.038 (0.938)	0.042 (0.946)
		10,20	0.040 (0.940)	0.045 (0.947)
		20,20	0.038 (0.922)	0.048 (0.936)
		20,30	0.036 (0.931)	0.043 (0.942)
3.0	10,10	0.040 (0.936)	0.046 (0.944)	
	10,20	0.044 (0.933)	0.050 (0.941)	
	20,20	0.039 (0.939)	0.044 (0.945)	
	20,30	0.042 (0.941)	0.046 (0.947)	
0.5, 1.0	0.5	10,10	0.033 (0.924)	0.041 (0.937)
		10,20	0.033 (0.922)	0.040 (0.935)
		20,20	0.037 (0.925)	0.043 (0.937)
		20,30	0.039 (0.934)	0.044 (0.940)
	1.0	10,10	0.040 (0.934)	0.045 (0.941)
		10,20	0.040 (0.934)	0.045 (0.941)
		20,20	0.043 (0.935)	0.046 (0.943)
		20,30	0.044 (0.932)	0.048 (0.939)
	1.5	10,10	0.033 (0.921)	0.040 (0.936)
		10,20	0.040 (0.929)	0.045 (0.939)
		20,20	0.040 (0.930)	0.045 (0.938)
		20,30	0.040 (0.934)	0.044 (0.941)
3.0	10,10	0.043 (0.935)	0.046 (0.943)	
	10,20	0.034 (0.922)	0.043 (0.935)	
	20,20	0.036 (0.927)	0.042 (0.938)	
	20,30	0.038 (0.940)	0.043 (0.946)	
0.5, 2.0	0.5	10,10	0.041 (0.934)	0.046 (0.944)
		10,20	0.031 (0.928)	0.037 (0.942)
		20,20	0.032 (0.922)	0.037 (0.932)
		20,30	0.037 (0.932)	0.040 (0.939)
	1.0	10,10	0.036 (0.928)	0.040 (0.934)
		10,20	0.036 (0.936)	0.041 (0.942)
		20,20	0.036 (0.928)	0.040 (0.934)
		20,30	0.036 (0.936)	0.041 (0.942)
	1.5	10,10	0.030 (0.930)	0.036 (0.944)
		10,20	0.033 (0.920)	0.039 (0.932)
		20,20	0.034 (0.938)	0.039 (0.944)
		20,30	0.036 (0.931)	0.041 (0.938)
3.0	10,10	0.037 (0.934)	0.040 (0.939)	
	10,20	0.032 (0.932)	0.037 (0.944)	
	20,20	0.032 (0.920)	0.037 (0.933)	
	20,30	0.037 (0.935)	0.040 (0.942)	
1.0, 1.0	0.5	10,10	0.036 (0.930)	0.041 (0.938)
		10,20	0.036 (0.936)	0.041 (0.938)
		20,20	0.036 (0.930)	0.041 (0.938)
		20,30	0.037 (0.933)	0.041 (0.938)
	1.0	10,10	0.037 (0.933)	0.040 (0.939)
		10,20	0.032 (0.922)	0.037 (0.944)
		20,20	0.032 (0.920)	0.037 (0.933)
		20,30	0.037 (0.935)	0.040 (0.942)
	1.5	10,10	0.036 (0.930)	0.041 (0.938)
		10,20	0.036 (0.930)	0.041 (0.938)
		20,20	0.036 (0.930)	0.041 (0.938)
		20,30	0.037 (0.933)	0.041 (0.938)
3.0	10,10	0.037 (0.933)	0.040 (0.939)	
	10,20	0.030 (0.934)	0.036 (0.946)	
	20,20	0.034 (0.925)	0.040 (0.936)	
	20,30	0.033 (0.932)	0.039 (0.939)	
1.0, 2.0	0.5	10,10	0.038 (0.928)	0.040 (0.935)
		10,20	0.038 (0.928)	0.040 (0.935)
		20,20	0.038 (0.928)	0.040 (0.935)
		20,30	0.038 (0.928)	0.040 (0.935)
	1.0	10,10	0.038 (0.928)	0.040 (0.935)
		10,20	0.038 (0.928)	0.040 (0.935)
		20,20	0.038 (0.928)	0.040 (0.935)
		20,30	0.038 (0.928)	0.040 (0.935)
	1.5	10,10	0.038 (0.928)	0.040 (0.935)
		10,20	0.038 (0.928)	0.040 (0.935)
		20,20	0.038 (0.928)	0.040 (0.935)
		20,30	0.038 (0.928)	0.040 (0.935)
3.0	10,10	0.038 (0.928)	0.040 (0.935)	
	10,20	0.038 (0.928)	0.040 (0.935)	
	20,20	0.038 (0.928)	0.040 (0.935)	
	20,30	0.038 (0.928)	0.040 (0.935)	

## 5. Concluding remarks

In the Pareto models, we have found the first order matching priors and the reference priors for the common scale parameter. It turns out that the posterior distribution of the reference priors is not proper and Jeffreys' prior is not the first order matching. As illustrated in our numerical study, the matching prior met very well the target coverage probabilities. Thus we recommend the use of the matching prior for Bayesian inference of the common scale parameter in two independent Pareto distributions.

## References

- Arnold, B. C. and Press, S. J. (1983). Bayesian inference for Pareto populations. *Journal of Econometrics*, **21**, 287-306.
- Arnold, B. C. and Press, S. J. (1989). Bayesian estimation and prediction for Pareto data. *Journal of the American Statistical Association*, **84**, 1079-1084.
- 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. (1992). On the development of reference priors (with discussion). In *Bayesian Statistics IV*, edited by J. M. Bernardo, et. al., Oxford University Press, Oxford, 35-60.
- Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of Royal Statistical Society B*, **41**, 113-147.
- Cox, D. R. and Reid, N. (1987). Orthogonal parameters and approximate conditional inference (with discussion). *Journal of Royal Statistical Society B*, **49**, 1-39.
- Datta, G. S. and Ghosh, J. K. (1995a). On priors providing frequentist validity for Bayesian inference. *Biometrika*, **82**, 37-45.
- Datta, G. S. and Ghosh, M. (1995b). Some remarks on noninformative priors. *Journal of the American Statistical Association*, **90**, 1357-1363.
- Datta, G. S. and Ghosh, M. (1996). On the invariance of noninformative priors. *The Annals of Statistics*, **24**, 141-159.
- Datta, G. S., Ghosh, M. and Mukerjee, R. (2000). Some new results on probability matching priors. *Calcutta Statistical Association Bulletin*, **50**, 179-192.
- DiCiccio, T. J. and Stern, S. E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihood. *Journal of Royal Statistical Society B*, **56**, 397-408.
- Iffesi, A. and Jin, C. (1996). On robust estimation of the common scale parameter of several Pareto distributions. *Statistics & Probability Letters*, **29**, 345-352.
- Fernández, A. J. (2008). Highest posterior density estimation form multiply censored Pareto data. *Statistical Papers*, **49**, 333-341.
- Geisser, S. (1984). Prediction Pareto and exponential observables. *Canadian Journal of Statistics*, **12**, 143-152.
- Geisser, S. (1985). Interval prediction for Pareto and exponential observables. *Journal of Econometrics*, **29**, 173-185.
- Ghosh, J. K. and Mukerjee, R. (1992). Noninformative priors (with discussion). In *Bayesian Statistics IV*, edited by J. M. Bernardo, et. al., Oxford University Press, Oxford, 195-210.
- Ghosh, J. K. and Mukerjee, R. (1995). Frequentist validity of highest posterior density regions in the presence of nuisance parameters. *Statistics & Decisions*, **13**, 131-139.
- Kang, S. G. (2010). Noninformative priors for the common scale parameter in Pareto distribution. *Journal of the Korean Data & Information Science Society*, **21**, 335-343.
- Kim, D. H., Kang, S. G. and Lee, W. D. (2007). Noninformative priors for the common shape parameter in the gamma distributions. *Journal of the Korean Data & Information Science Society*, **18**, 247-257.
- Kim, D. H., Kang, S. G. and Lee, W. D. (2009). Noninformative priors for Pareto distribution. *Journal of the Korean Data & Information Science Society*, **20**, 1213-1223.
- Ko, J. H. and Kim, Y. H. (1999). Bayesian prediction inference for censored Pareto model. *Journal of the Korean Data & Information Science Society*, **10**, 147-154.
- Lwin, T. (1972). Estimation of the tail of the Paretian law. *Scandinavian Actuarial Journal*, **55**, 170-178.

- Mukerjee, R. and Dey, D. K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter : Higher order asymptotics. *Biometrika*, **80**, 499-505.
- Mukerjee, R. and Ghosh, M. (1997). Second order probability matching priors. *Biometrika*, **84**, 970-975.
- Nigm, A. M. and Hamdy, H. L. (1987). Bayesian prediction bounds for the Pareto lifetime model. *Communications in Statistics: Theory and Methods*, **16**, 1761-1772.
- Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Sequential Methods in Statistics*, Banach Center Publications, **16**, 485-514.
- Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, **76**, 604-608.
- Tiwari, R. C., Yang, Y. and Zalkikar, J. N. (1996). Bayes estimation for the Pareto failure-model using gibbs sampling. *IEEE Transactions on Reliability*, **45**, 471-476.
- Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihood. *Journal of Royal Statistical Society, B*, **25**, 318-329.