Reference Prior Bayesian Analysis for Normal Mean Products

by

Dongchu Sun and Keying Ye
University of Missouri
Virginia Polytechnic Institute and State University

Technical Report # 93–27C

Department of Statistics
Purdue University

June 1993

Reference Prior Bayesian Analysis for Normal Mean Products

Dongchu Sun and Keying Ye

University of Missouri-Columbia and Virginia Polytechnic Institute and State University

ABSTRACT

Two reference priors (Berger and Bernardo, 1992a) for the product of means of n normal distributions with common known variance are developed. One of them induces improper posterior distribution and therefore is not of much interest. The other one is in a generalized form of n=2 case and is selected to compare with the uniform prior (the Jeffreys prior) in posterior inference and some frequentist criterion. The reference prior will be shown better than the uniform prior in the sense of correct frequentist coverages of posterior quantiles by numerical computations. The computations were efficiently done by Gibbs sampling technique for n=3 and n=10. Furthermore, it can be shown that the reference prior selected is one of the asymptotic optimal frequentist coverage prior (Tibshirani, 1989) under a transformation of parameter space such that the parameter of interest and the nuisance parameters are orthogonal (Cox and Reid, 1987).

AMS Subject Classification: 62A15, 62F15.

Keywords: Asymptotic optimal frequentist coverage property; Gibbs sampling; Jeffreys prior; orthogonal parametrization; reference prior.

1. Introduction

Suppose $X_i \sim N(\mu_i, 1)$, for $i=1,\ldots,n$ are independent normal random variables with means $\mu_i > 0$, and common variance 1. The parameter of interest is $\theta = \mu_1 \mu_2 \cdots \mu_n$. When $n=2$, θ can be thought of as the determination of an area based on measurements of length and width, while for $n=3$, θ represents a volume. Note that if X_i has known variance σ_i^2 for each i, the problem can be reduced to the aforementioned by considering $Y_i = X_i / \sigma_i$ and $\mu_i^* = \mu_i / \sigma_i$. Therefore,

$$\theta = \prod_{i=1}^n \mu_i = \prod_{i=1}^n \sigma_i \prod_{i=1}^n \mu_i^* = \left(\prod_{i=1}^n \sigma_i\right)\theta^*,$$

is only a scale transformation between θ and θ^* which has no effect on the discussion of the reference prior approach.

Berger and Bernardo (1989) tackled the problem with reference prior Bayesian analysis in which they successfully applied one reference prior on the problem and showed that the posterior inference using the reference prior is more sensible than that using the uniform prior. However, the reference priors are more difficult to calculate when n is large. When $n=2$, the reference prior obtained by Berger and Bernardo is in the form $\pi(\underline{\mu}) \propto \sqrt{\mu_1^2 + \mu_2^2}$. Several statisticians hence conjectured that the reference prior would be in the form $\pi(\underline{\mu}) \propto \sqrt{\sum_{i=1}^n \mu_i^2}$ when $n>2$ (in personal communication). Yet, under the careful investigation, two reference priors will be derived for $n>2$. The impropriety of the posterior density by one of them will be discussed. The other one is $\pi(\underline{\mu}) \propto \prod_{i=1}^n \mu_i \sqrt{\sum_{i=1}^n \mu_i^{-2}}$ which is also in a generalized form of the prior in $n=2$ case while it is different from the conjectural prior. The derivation of this prior not only depends on selection of compact subsets (Berger and Bernardo, 1992a,b,c) of the parameters, but also depends on the limiting procedure of the boundaries of those subsets.

Posterior inferences made by the reference prior and the uniform prior (Jeffreys prior) will be compared and numerical comparisons made for the frequentist coverage probabilities of posterior credible sets suggests that the reference prior is preferable than the uniform prior.

Furthermore, orthogonal reparametrization in which one of the parameter is θ will be solved by partial differential equations. Tibshirani's asymptotic optimal frequentist coverage prior in which, asymptotically, the frequentist coverage probability of a posterior credible set matches its posterior probability (Tibshirani, 1989), will be derived by this reparametrization. It turns out that the reference prior which has proper posterior is one of those asymptotic optimal frequentist coverage priors. This provides the first multivariate example of Tibshirani's priors in which the orthogonal parameters need to be solved by partial differential equations. Tibshirani (1989) (also see Stein, 1985) considered the "distance of the mean of a multivariate normal distribution from the origin" example in which the natural polar coordinates automatically give one orthogonal reparametrization.

The paper is organized as follows. Section 2 gives two reference priors derived based on one selection of the compact subsets. In section 3, the posterior inferences based on the reference prior which has proper posterior and the uniform prior will be displayed by graphical comparisons. Section 4 will show that the frequentist coverage probabilities of posterior credible sets by the reference prior and the uniform prior. The calculations in both sections 3 and 4 are done by Gibbs sampling technique which is very successful in dealing with the problem even when n=10. In section 5, the asymptotic optimal frequentist coverage priors will be discussed. Section 6 includes discussions and conclusions. In section 7, some necessary proofs for the results or computational results will be given.

2. The derivations of the reference priors

2.1 Reparametrization

Let

$$\begin{cases}\theta=\prod_{j=1}^{n}\mu_j \\ \omega_i=\mu_i/\theta^{1/n}, \quad i=2,3,\ldots,n\end{cases} \text{ so that } \begin{cases}\mu_1=\theta^{1/n}\prod_{j=2}^{n}\omega_j^{-1} \\ \mu_i=\theta^{1/n}\omega_i, \quad i=2,3,\ldots n\end{cases}. \tag{2.1}$$

The purpose of the reparametrization is to separate the factors of θ and ω's in the calculation for the expected Fisher's information matrix. Define

$$\underline{\omega}=(\omega_2,\ldots,\omega_n)',\ p(\underline{\omega})=\prod_{j=2}^{n}\omega_j^{-1},\ \underline{d}_1(\underline{\omega})=(\omega_2^{-1},\ldots,\omega_n^{-1})'. \tag{2.2}$$

The Jacobian matrix of the transformation is given by

$$J(\theta,\underline{\omega})=\begin{pmatrix}\frac{\theta^{1/n-1}}{n}p(\underline{\omega}) & -\theta^{1/n}p(\underline{\omega})\underline{d}_1^t(\underline{\omega}) \\ \frac{\theta^{1/n-1}}{n}\underline{\omega} & \theta^{1/n}I_{n-1}\end{pmatrix},$$

where I_m is the identity matrix of dimension m. Then the expected Fisher's information matrix can be computed as follows:

$$\begin{aligned}I(\theta,\underline{\omega})&=J(\theta,\underline{\omega})'J(\theta,\underline{\omega}) \\ &=\frac{\theta^{2/n-2}}{n^2}\begin{pmatrix}p^2(\underline{\omega})+\underline{\omega}'\underline{\omega} & \theta\, n(\underline{\omega}-p^2(\underline{\omega})\underline{d}_1(\underline{\omega}))' \\ \theta\, n(\underline{\omega}-p^2(\underline{\omega})\underline{d}_1(\underline{\omega})) & \theta^2 n^2(I_{n-1}+p^2(\underline{\omega})\underline{d}_1(\underline{\omega})\underline{d}_1^t(\underline{\omega}))\end{pmatrix}\end{aligned}. \tag{2.3}$$

The inverse matrix of $I(\theta,\underline{\omega})$, which plays an important role in deriving the reference priors, is given by

$$I^{-1}(\theta,\underline{\omega})=\theta^{-2/n}\begin{pmatrix}\theta^2(p^{-2}+\underline{d}_1^t\underline{d}_1) & \theta\,[\underline{d}_1^t-\frac{1}{n}(p^{-2}+\underline{d}_1^t\underline{d}_1)\underline{\omega}'] \\ \theta\,[\underline{d}_1-\frac{1}{n}(p^{-2}+\underline{d}_1^t\underline{d}_1)\underline{\omega}] & I_{n-1}-\frac{1}{n}(\underline{d}_1\underline{\omega}'+\underline{\omega}\underline{d}_1^t)+\frac{1}{n^2}(p^{-2}+\underline{d}_1^t\underline{d}_1)\underline{\omega}\underline{\omega}'\end{pmatrix},$$

where p and $\underline{d}_1$ are the abbreviations of $p(\underline{\omega})$ and $\underline{d}_1(\underline{\omega})$, respectively. In both matrices $I(\theta,\underline{\omega})$ and $I^{-1}(\theta,\underline{\omega})$, θ and $\underline{\omega}$ are separated in each element.

2.2 Selection of the compact subsets and the result of the reference priors

We will display the reference priors of the normal product problem using the algorithm by Berger and Bernardo (1992a). The algorithm is iterative on a sequence of compact subsets of parameter space. When a limiting procedure is processed, the compact subsets will eventually turn out to be the whole parameter space. Therefore, it is possible to have different ways to approach the limits resulting in different reference priors (Berger and Bernardo, 1992c, Ye, 1992).

In many situations, it does not matter what kinds of subsets to choose from and what the limiting processes are. However, for a normal product problem, as shown for n=2 case by Berger and Bernardo (1989), different selections of subsets may yield different results.

Suppose that we are interested in the group $\{\theta,\underline{\omega}\}$ in which only θ is the parameter of interest and $\underline{\omega}$ is treated as a vector of nuisance parameters. For the details of grouping, readers are referred to Berger and Bernardo (1992a). In this case, we have two groups θ and $\underline{\omega}$.

The h functions used (Berger and Bernardo, 1992c) in the calculations are:

$$\begin{aligned} h_2(\theta,\underline{\omega}) &= \theta^{\,2(n-1)/n}[1+p^2\,\underline{d}_1^t\underline{d}_1], \\ h_1(\theta,\underline{\omega}) &= \theta^{\,2/n-2}\textstyle\prod_{i=2}^n \omega_i^{-2}(1+\prod_{i=2}^n \omega_i^{-2}\sum_{i=2}^n \omega_i^{-2})^{-1}. \end{aligned} \tag{2.4}$$

It follows that,

$$\pi_2^l(\underline{\omega}|\theta) = \frac{\sqrt{1+\prod_{i=2}^n \omega_i^{-2}\sum_{i=2}^n \omega_i^{-2}}}{\int_{\Omega_l(\theta)}\sqrt{1+\prod_{i=2}^n \omega_i^{-2}\sum_{i=2}^n \omega_i^{-2}}\,d\underline{\omega}}\, I_{\Omega_l(\theta)}(\underline{\omega}) \quad \text{and}$$

$$\pi_1^l(\theta,\underline{\omega}) = \pi_2^l(\underline{\omega}|\theta)\exp\left\{\frac{1}{2}E^l\left[\log|h_1(\theta,\underline{\omega})||\theta\right]\right\}.$$

where $\{\Omega_l(\theta)\}$ is a sequence of compact sets of $\underline{\omega}$ for any positive integer l such that $\Omega_l(\theta) \to \mathbf{R}^+_{n-1}$ as l tends to ∞. By these forms, the reference priors can be written as

$$\pi(\theta,\underline{\omega}) \propto \sqrt{1+\prod_{i=2}^{n}\omega_i^{-2}\sum_{i=2}^{n}\omega_i^{-2}} \lim_{l\to\infty}\frac{A_l(\theta_0)}{A_l(\theta)}\exp\left\{\frac{1}{2}\lim_{l\to\infty}\left[\frac{B_l^*(\theta)}{A_l(\theta)}-\frac{B_l^*(\theta_0)}{A_l(\theta_0)}\right]\right\}, \tag{2.5}$$

where

$$A_l(\theta) = \int_{\Omega_l(\theta)}\sqrt{1+\prod_{i=2}^{n}\omega_i^{-2}\sum_{i=2}^{n}\omega_i^{-2}}\,d\underline{\omega}, \quad \text{and}$$

(2.6)

$$B_l^*(\theta) = \int_{\Omega_l(\theta)}\sqrt{1+\prod_{i=2}^{n}\omega_i^{-2}\sum_{i=2}^{n}\omega_i^{-2}}\log\left[\theta^{\,2/n-2}\prod_{i=2}^{n}\omega_i^{-2}\left(1+\prod_{i=2}^{n}\omega_i^{-2}\sum_{i=2}^{n}\omega_i^{-2}\right)^{-1}\right]d\underline{\omega},$$

in which θ_0 is an inner point of the interval $(0,\infty)$.

The compact subsets for $\underline{\omega}$ are selected using polar coordinates. First, let $v_i = \omega_i^{-1}$ for i=2,...,n. Then transform to the polar coordinate transformations $(r,\underline{\varphi})$ for $\underline{v}$ in which r is the radius and $\underline{\varphi}$ is a vector of angles. Let each coordinate of $\underline{\varphi}$ range from $1/l_l$ to $\pi/2-1/u_l$ where both l_l and u_l do not depend on θ and tend to ∞. Suppose that r is from $l(\theta)/l_l$ to $u_l u(\theta)$, where $l(\theta)$ and $u(\theta)$ are two functions of θ defined in the proof of *Result 2.1* in section 7. After discussion of the relationship between the boundaries l_l and u_l, we have the following result.

Result 2.1: *Using the transformations described as above, the reference priors could be determined as follows:*

A. For n>2,

(i) as $u_l\log(u_l)/l_l^{n-1}\to 0$, *the reference prior is given by*

$$\pi(\underline{\mu}) \propto \prod_{i=1}^{n}\mu_i\sqrt{\sum_{i=1}^{n}\mu_i^{-2}}\,; \tag{2.7}$$

(ii) as $l_l^{n-1}\log(l_l)/u_l\to 0$, *the reference prior is given by*

$$\pi(\underline{\mu}) \propto \left(\prod_{i=1}^{n} \mu_i\right)^{2/n-1} \sqrt{\sum_{i=1}^{n} \mu_i^{-2}}\ ; \qquad (2.8)$$

(iii) for the other situations, there are no solutions for the reference priors.

B. For n=2, the reference prior is given by

$$\pi(\underline{\mu}) \propto \mu_1 \mu_2 \sqrt{\mu_1^{-2} + \mu_2^{-2}}, \qquad (2.9)$$

as long as both l_i and u_i *tend to* ∞ .

Proof: See section 7. □

The prior in (2.9) is the same as the prior used in Berger and Bernardo (1989) and the prior in (2.7) is in a generalized form of (2.9). In next section, the prior of (2.8) will be proven to result in an improper posterior. Therefore it is not recommended for general use.

3. Posterior distributions

3.1 Conditions for proper posteriors

Consider a prior with the form

$$\pi(\underline{\mu}) \propto \left(\prod_{i=1}^{n} \mu_i\right)^{s} \left(\sum_{i=1}^{n} \mu_i^{-2}\right)^{h}, \qquad (3.1)$$

where h and s are nonnegative numbers. Notice that the priors given by (2.7), (2.8) and the uniform prior are special situations of (3.1). The likelihood function of $\underline{\mu}$ can be expressed by

$$l(\underline{\mu}) \propto \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'(\underline{x}-\underline{\mu})\right]. \qquad (3.2)$$

Therefore, using the prior with form (3.1), the posterior distribution of $\underline{\mu}$ is given by

$$\pi(\underline{\mu}|data) \propto \left(\prod_{i=1}^{n} \mu_i\right)^{s} \left(\sum_{i=1}^{n} \mu_i^{-2}\right)^{h} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})'(\underline{x}-\underline{\mu})\right]. \qquad (3.3)$$

A. When $|\underline{\mu}| \to \infty$, the density in (3.3) is always integrable because of the exponential factors.

B. If, for some k, $\mu_j \neq 0$ for $j \neq k$, then

$$\left(\prod_{i=1}^{n}\mu_i\right)^s\left(\sum_{i=1}^{n}\mu_i^{-2}\right)^h = \left(\mu_k^{s/h-2} + \mu_k^{s/h}\sum_{j\neq k}\mu_i^{-2}\right)^h \prod_{j\neq k}\mu_i^s$$

which is integrable for μ_k at 0 when $s-2h>-1$. If $s-2h>-1$ is true, then

$$\left(\prod_{i=1}^{n}\mu_i\right)^s\left(\sum_{i=1}^{n}\mu_i^{-2}\right)^h = \left(\sum_{i=1}^{n}\mu_i^{s/h-2}\prod_{j\neq i}\mu_j^{s/h}\right)^h,$$

which is always integrable in any finite compact region of $\underline{\mu}$. Therefore, to assure a proper posterior for the prior (3.1), the necessary and sufficient condition should be $s-2h>-1$.

For the prior in (2.7) and (2.9) as well as the uniform prior, $s-2h=0>-1$, so the posteriors are proper. While for the prior given by (2.8), $s-2h=2/n-2\leq-1$ when n≥2. Therefore, the prior (2.8) yields improper posterior.

3.2 Marginal posterior distribution for θ

Consider the reparametrization (2.1), the likelihood of $(\theta,\underline{\omega})$ becomes

$$l(\theta,\underline{\omega}) \propto \exp\left\{-\frac{1}{2}A(\underline{\omega})\left[\theta^{1/n} - B(\underline{x},\underline{\omega})/A(\underline{\omega})\right]^2 + \frac{1}{2}B^2(\underline{x},\underline{\omega})/A(\underline{\omega})\right\},$$

where

$$A(\underline{\omega}) = p^2(\underline{\omega}) + \underline{\omega}'\underline{\omega}, \text{ and } B(\underline{x},\underline{\omega}) = x_1 p(\underline{\omega}) + \underline{x}_2'\underline{\omega}. \tag{3.4}$$

In (3.4), $\underline{\omega}$ and $p(\underline{\omega})$ are defined in (2.2).

Under the transformation (2.1), the prior (3.1) becomes

$$\pi(\theta,\underline{\omega}) \propto \theta^{s-2h/n}p(\underline{\omega})C^h(\underline{\omega}), \tag{3.5}$$

where $C(\underline{\omega}) = p^{-2}(\underline{\omega}) + \underline{d}_1^t(\underline{\omega})\underline{d}_1(\underline{\omega})$ and $\underline{d}_1(\underline{\omega})$ is also defined in (2.2). Hence, the posterior density of $(\theta,\underline{\omega})$ using the prior (3.5) is given by

$$\pi(\theta,\underline{\omega}|\,data) \propto \theta^{s-2h/n} p(\underline{\omega}) C^h(\underline{\omega}) \phi\left(\sqrt{A(\underline{\omega})}\theta^{1/n} - \frac{B(\underline{\omega})}{\sqrt{A(\underline{\omega})}}\right) \exp\left(\frac{B^2(\underline{\omega})}{2A(\underline{\omega})}\right),$$

where $\phi(x)$ is the density function of N(0,1) at x. The following theorem can be easily proven.

Theorem 3.1: *When prior (3.5) is used, the marginal posterior density of* $\theta = \mu_1 \cdots \mu_n$ *is given by*

$$\pi(\theta|\,data) = \frac{\theta^{s-2h/n}\int_0^\infty \cdots \int_0^\infty p(\underline{\omega})C^h(\underline{\omega})\phi\left(\sqrt{A(\underline{\omega})}\theta^{1/n} - \frac{B(\underline{\omega})}{\sqrt{A(\underline{\omega})}}\right)\exp\left(\frac{B^2(\underline{\omega})}{2A(\underline{\omega})}\right)d\underline{\omega}}{\int_0^\infty \theta^{s-2h/n}\int_0^\infty \cdots \int_0^\infty p(\underline{\omega})C^h(\underline{\omega})\phi\left(\sqrt{A(\underline{\omega})}\theta^{1/n} - \frac{B(\underline{\omega})}{\sqrt{A(\underline{\omega})}}\right)\exp\left(\frac{B^2(\underline{\omega})}{2A(\underline{\omega})}\right)d\underline{\omega}}.$$

Furthermore, the marginal posterior cumulative density function of θ *can be expressed by*

$$F(\theta_0) = P(\theta \le \theta_0|\,data) = 1 - \frac{H_{h,s}(\theta_0)}{H_{h,s}(0)},$$

where

$$H_{h,s}(x) = \int_0^\infty \cdots \int_0^\infty p(\underline{\omega})C^h(\underline{\omega})\exp\left(\frac{B^2(\underline{\omega})}{2A(\underline{\omega})}\right)h_{s-2h/n}(x,\underline{\omega})d\underline{\omega},$$

and

$$h_p(x,\underline{\omega}) = \int_x^\infty \theta^p \phi\left(\sqrt{A(\underline{\omega})}\theta^{1/n} - \frac{B(\underline{\omega})}{\sqrt{A(\underline{\omega})}}\right)d\theta. \qquad (3.6)$$

The function $h_p(x,\underline{\omega})$ in (3.6) can be determined further by the probability density function $\phi(x)$ and cumulative distribution function $\Phi(x)$ of N(0,1) distribution. Notice that after transforming $s = \theta^{1/n}$ in (3.6),

$$h_p(x,\underline{\omega}) = nq_{(p+1)n-1}(x^{1/n},\underline{\omega}),$$

where $q_p(x,\underline{\omega})=\int_x^\infty s^p\phi\left(\sqrt{A(\underline{\omega})}s-B(\underline{\omega})/\sqrt{A(\underline{\omega})}\right)ds$ can be evaluated by the following theorem.

Theorem 3.2: *If p is a nonnegative integer, then*

$$q_p(x,\underline{\omega})=\tilde{\psi}_p(x,\underline{\omega})\phi\left(\sqrt{A(\underline{\omega})}x-\frac{B(\underline{\omega})}{\sqrt{A(\underline{\omega})}}\right)+\tilde{h}_p(\underline{\omega})\left[1-\Phi\left(\sqrt{A(\underline{\omega})}x-\frac{B(\underline{\omega})}{\sqrt{A(\underline{\omega})}}\right)\right],$$

where ϕ and Φ *are density and cumulative density of N(0,1) distribution, and* $\tilde{\psi}_p(x,\underline{\omega})$ *and* $\tilde{h}_p(\underline{\omega})$ *are given by*

$$\begin{cases}\tilde{\psi}_p(x)=\left[x^{p-1}+B(\underline{\omega})\tilde{\psi}_{p-1}(x)+(p-1)\tilde{\psi}_{p-2}(x)\right]/A(\underline{\omega}) \\ \tilde{h}_p=\left[B(\underline{\omega})\tilde{h}_{p-1}+(p-1)\tilde{h}_{p-2}\right]/A(\underline{\omega})\end{cases}. \tag{3.7}$$

(3.7) starts at $\tilde{\psi}_0(x)=0$, $\tilde{\psi}_1(x)=1/A(\underline{\omega})$ *and* $\tilde{h}_0=1/\sqrt{A(\underline{\omega})}$, $\tilde{h}_1=B(\underline{\omega})/A^{3/2}(\underline{\omega})$.

Proof: By induction. □

By *Theorem 3.1* and *3.2,* the marginal posterior cumulative distribution function of θ is 1 subtracted by a division of two n-1 dimensional integrals. In general, it is difficult to evaluate multidimensional integrals when the dimension is large. Gibbs sampling method is going to be used to overcome this difficulty in the next subsection and section 4.

3.3 Use of Gibbs sampling to evaluate the marginal posterior distribution for θ

In this section, the reference prior (2.7) and the uniform prior for $\underline{\mu}$ will be considered and the marginal posterior distribution of θ will be evaluated by Gibbs sampling simulation (Gelfand and Smith, 1990). By the likelihood function (3.2) and the prior with form (3.1), the conditional posterior distribution of μ_i given μ_{-i}, where $\mu_{-i}=(\mu_1,\ldots,\mu_{i-1},\mu_{i+1},\ldots,\mu_n)$, is given by

$$\pi(\mu_i|\mu_{-i},data) \propto \mu_i^s\left(\mu_i^{-2}+\Sigma_{j\neq i}\mu_j^{-2}\right)^h \exp\left[-\frac{1}{2}(\mu_i-x_i)^2\right], \quad \text{for } \mu_i>0. \tag{3.8}$$

For the prior (2.7), in which s=1 and h=1/2, (3.8) becomes

$$\pi(\mu_i|\mu_{-i},data) \propto \left(1+\mu_i^2\Sigma_{j\neq i}\mu_j^{-2}\right)^{1/2} \exp\left[-\frac{1}{2}(\mu_i-x_i)^2\right], \quad \text{for } \mu_i>0, \tag{3.9}$$

while for the uniform prior, in which s=0 and h=0, (3.8) is the truncated normal distribution with the density

$$\pi(\mu_i|\mu_{-i},data) \propto \exp\left[-\frac{1}{2}(\mu_i-x_i)^2\right], \quad \text{for } \mu_i>0. \tag{3.10}$$

To simulate the random variates with densities (3.9) and (3.10), rejection methods will be described by the following details.

I. Simulation of X with density $f(x) \propto (1+x^2U)^{1/2}\exp\left[-(x-V)^2/2\right]$ for x>0 and U>0.

Step 1. Consider two regions of u:

(i) $V \le -\sqrt{U}$: If $(du_2-\sqrt{U})\sqrt{V^2-2\log(u_1)} \le 1+V\sqrt{U}$, accept $x=V+\sqrt{V^2-2\log(u_1)}$, where u_1 and u_2 are two U(0,1) random variates and $d=\max\left\{\sqrt{U},-1/V\right\}$.

(ii) $V>-\sqrt{U}$: Let $p=\left(V\sqrt{U}-1+\sqrt{\left(1+V\sqrt{U}\right)^2+4U}\right)/\left(2\sqrt{U}\right)$ and $q=p-V$. Generate x from N(p,1). If $u_1t \le \left(1+x\sqrt{U}\right)\exp(1-qx-1/t)$, where $t=\max\left\{\sqrt{U}/q,1\right\}$ and u_1 is a U(0,1) variate, then accept x.

Step 2. Simulate u_3 from U(0,1). If $u_3 \le (1+x^2U)^{1/2}/\left(1+x\sqrt{U}\right)$ for x being derived in Step 1, then accept x.

II. Simulation of X with density $f(x) \propto \exp\left[-(x-V)^2/2\right]$ for x>0.

Again, consider two separated regions:

(i) $V \ge -1/4$: Generate x from N(V,1). If x>0, then accept x.

(ii) $V < -1/4$, Generate u_1 and u_2 from U(0,1). If $u_2 \le \exp\left[-(\log(u_1) - 1/2)^2\right] / u_1$, then accept $x = \log(u_1)/(2V)$.

Discussion of these rejection rules will be given in section 7.

3.4 Numerical calculations

Using the Gibbs sampling methods described in section 3.3, we are able to deal with the normal mean product problem for large n. In this section, marginal posterior distributions of θ for n=3 and n=10 will be calculated for selected $\underline{x}$ values. Denote π_r as the reference prior given in (2.7) and π_u as the uniform prior.

(Insert Figure 1 and Figure 2 here.)

Figure 1 shows four graphs of the simulated marginal posterior densities of θ for the reference prior and the uniform prior when n=3. When $\underline{X} = (10,10,10)$, both distributions are very close. Hence for large values of x's, both priors provide the similar posterior information about θ. However, when data values are small, the posteriors by uniform prior show the considerable skewness in the information about θ. In section 4, the frequentist coverage probabilities of the posterior credible sets by π_r will be shown closer to its posterior probabilities than those by π_u. Hence, the posterior inference by π_u is inadequate to the parameter θ and the estimation of the standard error could be inappropriate.

Figure 2 displays several posterior densities for n=10. The similar conclusions as for the n=3 can be made in this situation.

Each curve in Figure 1 and 2 is drawn by smoothed Kernel density estimations (see Gelfand and Smith, 1990) for the 30,000 simulated marginal posterior points. For n=10, three out of four graphs are rescaled because the natures of the problem; the vertical values are enlarged by factor 10^9 and the horizontal values are reduced by the factor 10^{-9}.

4. Frequentist coverage probability for the posterior quantile

It has been argued by many authors (cf. Berger and Bernardo, 1989, Efron, 1987, Ghosh and Mukerjee, 1992, Stein, 1985, Welch and Peers, 1963, Ye, 1993 and Ye and Berger, 1991) that a good noninformative prior should have good frequentist properties. One of these properties is that the frequentist coverage probability of a $(1\text{-}\alpha)^{\text{th}}$ posterior quantile should be close to $1\text{-}\alpha$. In this section, this property will be investigated numerically in the light of small sample situation for the prior (2.7) as well as the regular uniform prior for $\underline{\mu}$. Again, Gibbs sampling method described in section 3 will be used.

The computation of Table 1 and 2 is based on the following algorithm for any fixed true $\underline{\mu}_0 = (\mu_{10}, \ldots, \mu_{n0})$ and any predetermined probability value α. Here we choose α as 0.05 and 0.95.

Step 1. Simulate $\underline{x}$ from $N_n(\underline{\mu}_0, I_n)$.

Step 2. Using the Gibbs sampling method given in subsection 3.3 to simulate posterior random vector $\underline{\mu} \mid \underline{x}$. Repeat the simulation m_1 times and check the proportion ρ of which $\prod_{i=1}^n \mu_i \le \theta_0$, where $\theta_0 = \prod_{i=1}^n \mu_{i0}$. This ρ is the estimation of the marginal posterior probability of θ for the interval $(0, \theta_0)$.

The combination of both steps above is called a cycle. Repeat the cycle m_2 times and compute the proportion δ of $\rho \le \alpha$ in these replications. This δ is the estimated frequentist coverage probability of the α^{th} posterior quantile. Table 1 and 2 below show the estimated frequentist coverage probabilities of 0.05(0.95) posterior quantiles for various true values of μ's by the reference prior (2.7) and the uniform prior when n=3 and n=10.

Table 1. Frequentist coverage probabilities for 0.05(0.95) posterior quantiles (n=3)

$\underline{\mu}_0$	(1,1,1)	(1,2,3)	(1,5,10)	(2,2,2)	(3,3,3)	(5,5,5)	(10,10,10)
π_r	.064(1.00)	.033(.995)	.059(.997)	.031(.994)	.037(.972)	.047(.957)	.054(.951)
π_u	.017(.998)	.018(.980)	.042(.995)	.012(.944)	.020(.919)	.028(.925)	.039(.936)

Table 2. Frequentist coverage probabilities for 0.05(0.95) posterior quantiles (n=10)

$\underline{\mu}_0$	(1,2,3,4,5,6,7,8,9,10)	(1,2,3,4,5,6,6,6,6,6)	(5,5,5,5,5,5,5,5,5,5)
π_r	.041(.995)	.040(.995)	.038(.961)
π_u	.012(.948)	.009(.939)	.011(.859)
$\underline{\mu}_0$	(5,5,5,5,5,6,7,8,9,10)	(5,5,5,5,5,10,10,10,10,10)	(10,10,10,10,10,10,10,10,10,10)
π_r	.045(.955)	.043(.954)	.048(.951)
π_u	.015(.882)	.018(.890)	.027(.914)

For the calculations of the entries in both tables, m_1 is 20,000 and m_2 is 10,000. The maximum standard error of Step 2 described above is 0.0035 and the maximum standard error of estimating δ is 0.005.

From Table 1 and 2, clearly the reference prior is better than uniform prior in most of the situations. Therefore, the reference prior with the form (2.7) is more appealing. When each coordinate of $\underline{\mu}_0$ is large, the frequentist coverage results by π_r are almost close to the desired levels. It is not surprised that the posterior quantiles show poor frequentist coverages when the components of $\underline{\mu}_0$ are small because the positive constraints we have on the posterior means.

5. Orthogonal parametrization

Inspired by the work of Stein (1985), Tibshirani (1989) developed a method for calculating noninformative priors which satisfy the asymptotic optimal frequentist coverage property. The prior has such property will have, approximately, 1-α frequentist coverage over its 1-α posterior credible region when sample size gets large. However, the method depends heavily on the orthogonal parametrization, namely the off-diagonal matrix of the expected Fisher information matrix is zero for the parameter of interest and the nuisance parameters (Cox and Reid, 1987). The orthogonal parametrization depends on solutions of partial differential equations (cf. Berger and Roberts, 1992) and is not always solvable.

The frequentist coverages of the 0.05 and 0.95 posterior quantiles by the reference prior (2.7) have been shown for small samples numerically in section 4. It is also interesting to know what kinds of priors will be obtained by Tibshirani's method for the normal mean product problem.

Let $\theta=\prod_{j=1}^{n}\mu_j$ and $\xi_i=\xi_i(\underline{\mu})$ for i=2,...,n. Denote $\xi_i^j=\partial\xi_i(\underline{\mu})/\partial\mu_j$ and $\eta_{(i)}=\prod_{j\neq i}\mu_j$. The Jacobian matrix of this transformation is

$$\frac{\partial(\theta,\underline{\xi})}{\partial(\underline{\mu})}=\begin{pmatrix}\eta_{(1)} & \eta_{(2)} & \cdots & \eta_{(n)}\\ \xi_2^1 & \xi_2^2 & \cdots & \xi_2^n\\ \vdots & \vdots & & \vdots\\ \xi_n^1 & \xi_n^2 & \cdots & \xi_n^n\end{pmatrix}. \tag{5.1}$$

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

$$I^{-1}(\theta,\underline{\xi})=\left(\frac{\partial(\theta,\underline{\xi})}{\partial(\underline{\mu})}\right)\left(\frac{\partial(\theta,\underline{\xi})}{\partial(\underline{\mu})}\right)'=\begin{pmatrix}\sum_{j=1}^{n}\eta_{(j)}^2 & \underline{\phi}'\\ \underline{\phi} & A\end{pmatrix}, \tag{5.2}$$

where $\underline{\phi}=\left(\sum_{j=1}^{n}\eta_{(j)}\xi_2^j,\ldots,\sum_{j=1}^{n}\eta_{(j)}\xi_n^j\right)'$ and A is an (n-1)×(n-1) non-singular matrix. Only $\underline{\phi}=\underline{0}$ is the condition so that θ and $\underline{\xi}$ are orthogonal. Hence, we have n-1 homogeneous linear partial differential equations of first order. Any smooth function with form $\psi(\mu_i^2-\mu_j^2, i<j)$ could be a solution of the equations. For instance, one could take

$$\nu_i=\frac{\mu_1^2-\mu_i^2}{2}, \quad i=2,\ldots,n$$

as the new transformations. Then θ and $\underline{\nu}$ are orthogonal and its Jacobian (5.1) can be rewritten as

$$\frac{\partial(\theta,\underline{\xi})}{\partial(\underline{\mu})}=\begin{pmatrix}\eta_{(1)} & \eta_{(2)} & \cdots & \eta_{(n)}\\ \mu_1 & -\mu_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ \mu_1 & 0 & \cdots & -\mu_n\end{pmatrix},$$

and its determinant is $\sum_{i=1}^{n}\eta_{(i)}^2$. The expected Fisher information matrix, by (5.2), is

$$I(\theta,\underline{\xi})=\begin{pmatrix}\sum_{j=1}^{n}\eta_{(j)}^2 & 0 & \cdots & 0\\ 0 & \mu_1^2+\mu_2^2 & \cdots & \mu_1^2\\ \vdots & \vdots & & \vdots\\ 0 & \mu_1^2 & & \mu_1^2+\mu_n^2\end{pmatrix}^{-1}.$$

Using Tibshirani's method (see also, Berger, 1992), a noninformative prior which has the form

$$\pi(\theta,\underline{\xi})\propto g(\underline{\xi})\left(\textstyle\sum_{i=1}^{n}\eta_{(i)}^2\right)^{-1/2},$$

will achieve the asymptotic optimal frequentist coverage property where $g(\underline{v})>0$ is arbitrary. Transforming back to the original parameter space $\underline{\mu}$, one can derive

$$\pi^*(\underline{\mu})\propto g(\underline{\xi}(\underline{\mu}))\left(\textstyle\sum_{i=1}^{n}\eta_{(i)}^2\right)^{1/2}\propto g(\underline{\xi}(\underline{\mu}))\left(\textstyle\prod_{i=1}^{n}\mu_i\right)\sqrt{\textstyle\sum_{i=1}^{n}\mu_i^{-2}}. \tag{5.3}$$

Note that the prior (2.7) is a special case of (5.3) for g≡1 so it should attain the asymptotic optimal frequentist coverage property.

6. Discussion

It is not easy to compute the reference priors for several normal mean product problem. Also, it can lead to several solutions with different limiting processes when the dimension of the parameter space is higher than 2. Furthermore, some priors obtained can lead to improper posterior distributions. Therefore, careful examination of the prior should be made in the reference prior Bayesian analysis.

However, the derived result is very encouraging because it has the simple form and it also has a good frequentist property comparing to the ad hoc priors such as uniform prior.

Though the posterior inferences need to deal with n-1 dimensional integrations, it is quite successful to use the Gibbs sampling technique to overcome this difficulty. For a comparison, we also calculated the frequentist coverage probabilities of the 0.05(0.95) posterior quantiles for n=3 by Monte Carlo sampling method. The results are very close to the result shown in Table 1. Yet, it consumed much more time than Gibbs sampling method. Therefore, when n is large, the computational difficulty of multidimensional integrations for this particular problem can be successfully resolved by Gibbs sampling technique.

Finally, it has been shown in many examples that the reference prior and Tibshirani's asymptotic optimal frequentist coverage prior are the same except a function of nuisance parameters. However, Berger and Robert (1992) showed that it is not the case for the distance of the multivariate normal mean to the origin with the presence of a non-identity covariance matrix when n=2. Though it is unclear thus far what is the intrinsic relationship between these two approaches, we provided one multivariate example which shows that these two methods give the same priors except a function of nuisance parameters.

7. Necessary proofs

A. Proof of *Result 2.1*

Define

$$B_l(\theta) = \int_{\Omega_l(\theta)} \sqrt{1+\prod_{i=2}^{n}\omega_i^{-2}\sum_{i=2}^{n}\omega_i^{-2}}\,\log\left[\prod_{i=2}^{n}\omega_i^{-2}\left(1+\prod_{i=2}^{n}\omega_i^{-2}\sum_{i=2}^{n}\omega_i^{-2}\right)^{-1}\right] d\underline{\omega}. \quad (7.1)$$

The object here is to evaluate the function $A_l(\theta)$ in (2.6) and $B_l(\theta)$ above. Let $v_i = \omega_i^{-1}$ for i=2,...,n and transform $\underline{v}$ to the polar coordinates $(r,\underline{\varphi})$ where r is the radius and the components of $\underline{\varphi}$ are angles. The expression of $A_l(\theta)$ shows

$$A_l(\theta) = \int_{\Omega_l} f(\underline{\varphi}) g^{-(1+1/n)}(\underline{\varphi}) d\underline{\varphi} \int_{\theta^{1/n}/l_l}^{u_l/\theta^{1/n}} z^{-n}\sqrt{1+z^{2n}}\,dz = A(\Omega_l)T_1(l,\theta), \tag{7.2}$$

where $f(\underline{\varphi}) = |J|/r^{n-2}$ and $g(\underline{\varphi}) = \left[\prod_{i=2}^{n} u_i(r,\underline{\varphi})\right]/r^{2(n-1)}$. Here J is the Jacobian of the transformation. The range for each coordinate of $\underline{\varphi}$ is $[1/l_l, \pi/2 - 1/u_l]$. The compact interval $\left[l(\theta)/l_l, u_l u(\theta)\right]$ is chosen for the radius r such that $l(\theta) = \theta^{1/n}/g(\underline{\varphi})^{1/n}$ and $u(\theta) = 1/g(\underline{\varphi})^{1/n}\theta^{1/n}$. Similarly,

$$\begin{aligned} B_l(\theta) &= A_1(\Omega)T_1(l,\theta) + A\ (\Omega) \int_{\theta^{1/n}/l_l}^{u_l/\theta^{1/n}} z^{-n}\sqrt{1+z^{2n}}\log[z^{2(n-1)}(1+z^{2n})^{-1}]dz \\ &= A_1(\Omega)T_1(l,\theta) + A\ (\Omega)T_2(l,\theta) \end{aligned} \tag{7.3}$$

Both integrals for T_1 and T_2 can be integrated in three regions for z, namely, I: $(\theta^{1/n}/l_l, \varepsilon)$; II: $(\varepsilon, \varepsilon^{-1})$; and III: $(\varepsilon^{-1}, u_l/\theta^{1/n})$. The integrals in II are bounded functions of ε. In the following context, denote $K(\varepsilon)$ as an arbitrary bounded function of ε. Even in a same equation, $K(\varepsilon)$'s are not necessarily to be the same.

(a) For the region I:

$$\int_I z^{-n}\sqrt{1+z^{2n}}\,dz = \int_{\theta^{1/n}/l_l}^{\varepsilon} z^{-n}dz + K(\varepsilon) = \frac{l_l^{n-1}}{(n-1)\theta^{1-1/n}} + K(\varepsilon),$$

and

$$\int_I z^{-n}\sqrt{1+z^{2n}}\log[z^{2(n-1)}(1+z^{2n})^{-1}]dz = 2(n-1)\int_{\theta^{1/n}/l_l}^{\varepsilon} z^{-n}\log(z)dz + K(\varepsilon)$$

$$= \frac{2l_l^{n-1}}{\theta^{1-1/n}}\left[\log(\theta)/n - \log(l_l) + 1/(n-1)\right] + K(\varepsilon)$$

(b) For the region III:

$$\int_{III}\sqrt{1+z^{-2n}}dz = \int_{\varepsilon^{-1}}^{u_l/\theta^{1/n}} dz + K(\varepsilon) = \frac{u_l}{\theta^{1/n}} + K(\varepsilon),$$

and

$$\int_{III} z^{-n}\sqrt{1+z^{2n}}\log[z^{2(n-1)}(1+z^{2n})^{-1}]dz$$

$$= 2(n-1)\int_{\varepsilon^{-1}}^{u_l/\theta^{1/n}}\log(z)dz - 2n\int_{\varepsilon^{-1}}^{u_l/\theta^{1/n}}\log(z)dz + K(\varepsilon).$$

$$= -\frac{2}{\theta^{1/n}}u_l\left[\log(u_l) - \frac{1}{n}\log(\theta) - 1\right] + K(\varepsilon)$$

Therefore,

$$T_1(l,\theta) = \frac{u_l}{\theta^{1/n}} + \frac{l_l^{n-1}}{(n-1)\theta^{1-1/n}} + K(\varepsilon), \text{ and}$$

$$T_2(l,\theta) = \frac{2l_l^{n-1}}{\theta^{1-1/n}}[\log(\theta)/n - \log(l_l) + 1/(n-1)] - \frac{2u_l}{\theta^{1/n}}(\log(u_l) + \log(\theta)/n) + K(\varepsilon)$$

(i) As $u_l\log(u_l)/l_l^{n-1} \to 0$ when $l \to \infty$,

$$\frac{T_1(l,\theta)}{T_1(l,\theta_0)} \to \theta^{1/n-1}, \text{ and } \frac{T_2(l,\theta)}{T_1(l,\theta\ )} - \frac{T_2(l,\theta_0)}{T_1(l,\theta_0)} \to \frac{2(n-1)}{n}\log(\theta). \tag{7.4}$$

(ii) As $l_l^{n-1}\log(l_l)/u_l \to 0$ when $l \to \infty$,

$$\frac{T_1(l,\theta)}{T_1(l,\theta_0)} \to \theta^{-1/n}, \text{ and } \frac{T_2(l,\theta)}{T_1(l,\theta\ )} - \frac{T_2(l,\theta_0)}{T_1(l,\theta_0)} \to -\frac{2}{n}\log(\theta). \tag{7.5}$$

Note that other than the above situations, the limits can not be determined so that there is no exact expression for the reference prior.

Using (7.1), (7.2), (7.3), (7.4), (7.5) and (2.6), the priors for $(\theta,\underline{\omega})$ are

(i) As $u_l\log(u_l)/l_l^{n-1} \to 0$ when $l \to \infty$,

$$\pi(\theta,\underline{\omega}) \propto \theta^{1-1/n}\sqrt{1+\prod_{j=2}^{n}\omega_i^{-2}\sum_{j=2}^{n}\omega_i^{-2}}, \text{ and}$$

(ii) As $l_l^{n-1}\log(l_l)/u_l \to 0$ when $l \to \infty$,

$$\pi(\theta,\underline{\omega}) \propto \theta^{1/n-1}\sqrt{1+\prod_{j=2}^{n}\omega_i^{-2}\sum_{j=2}^{n}\omega_i^{-2}}.$$

The Jacobian of the transformation of (2.1) back to $\underline{\mu}$ is $|J(\theta,\underline{\omega})|^{-1} = p^{-1}(\underline{\omega}) = \theta^{1/n}\mu_1^{-1}$. Notice that $\sqrt{1+\prod_{j=2}^{n}\omega_i^{-2}(\underline{\mu})\sum_{j=2}^{n}\omega_i^{-2}(\underline{\mu})} = \mu_1\sqrt{\sum_{i=1}^{n}\mu_i^{-2}}$. Therefore, the claimed result follows. □

B. Discussion of the rejection rules in section 3.3

I. $f(x) \propto (1+x^2U)^{1/2}\exp(-(x-V)^2/2)$ for x>0 and U>0.

Step 2 of the algorithm is to use a rejection method to generate a variate from $g(x) \propto (1+x\sqrt{U})\exp(-(x-V)^2/2)$ for x>0. The maximum of the ratio $f(x)/g(x)$ when $x \in (0,\infty)$ is ratio of the normalization constants of the two densities. Therefore, when a U(0,1) random number $u_3 \le f(x)/g(x)$ where x is from g(x), then we should accept x.

For part (i) of step 1, let $g_1(x) \propto (x-V)\exp[-(x-V)^2/2]$. Then $g(x)/g_1(x) \propto \sqrt{U} + (1+V\sqrt{U})/(x-V)$ which is monotone for x and the maximum of $g(x)/g_1(x)$ would be proportional to $\max\{\sqrt{U},-1/V\}$. Here, all the proportions mean that we are only ignoring the ratio of the normalization constants between the two density functions $g(x)$ and $g_1(x)$. The algorithm follows the regular rejection method after the maximum of $g(x)/g_1(x)$ has been decided. Part (ii) can be discussed similarly.

II. $f(x) \propto \exp[-(x-V)^2/2]$ for x>0.

When $V \ge -1/4$, we just use the truncated normal simulation. However, when V is far away from 0 to the negative side, the density is exponential like. So, for V<–1/4, the exponential density $g_3(x) = 2V\exp(-2Vx)$ is used and the ordinary rejection method follows.

Acknowledgment: We are grateful to professor James Berger for his helpful comments and suggestions. Sun's work was partially supported by a Research Board Grant from the University of Missouri.

References

Berger, J. (1992). Discussion of the paper by Ghosh, J.K. and Mukerjee, R.. In *Bayesian Statistic 4,* (Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. eds), Oxford University Press: London, 205-206.

Berger, J. and Bernardo, J (1989). Estimating a product of means: Bayesian analysis with reference priors. *J.Am.Statist.Assoc.* **84,** 200-207.

Berger, J. and Bernardo, J. (1992a). On the development of the reference prior method. In *Bayesian Statistic 4,* (Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. eds), Oxford University Press: London, 35-60.

Berger, J. and Bernardo, J. (1992b). Ordered group reference priors with application to multinomial problem. *Biometrika,* **79,** 25-47.

Berger, J. and Bernardo, J. (1992c). Reference priors in a variance components problem. *Proceedings of the Indo-USA Workshop on Bayesian Analysis in Statistics and Econometrics,* (P. Goel, ed.) New York: Springer, 323-340.

Berger, J. and Robert, C. (1992). Estimation of quadratic function (reference priors for non-centrality parameters). Manuscript under preparation.

Bernardo, J. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J.R.Statist.Soc.,* **B 41,** 113-147.

Efron, B. (1986). Why isn't everyone a Bayesian? *Am.Statistician,* **40,** 1-11.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J.Am.Statist.Assoc.,* **85,** 398-409

Ghosh, J.K. and Mukerjee, R. (1992). Noninformative priors. In *Bayesian Statistic 4,* (Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. eds), Oxford University Press: London, 195-210

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics,* Ed. R. Zielinski, 485-514. Warsaw: PWN-Polish Scientific Publishers.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika,* **76,** 604-608.

Welch, B.L. and Peers, H.W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J.R.Statist.Soc.,* **B 25,** 318-329.

Ye, K. (1992). Sensitivity study of the reference priors in a random effect model. *Technical Report, 92-18,* Department of Statistics, Virginia Tech.

Ye, K. (1993). Reference priors when the stopping rule depends on the parameter of interest. *J.Am.Statist.Assoc.,* **88,** 360-363.

Ye, K. and Berger, J. (1991). Noninformative priors for inferences in exponential regression models. *Biometrika,* **78,** 645-656.

Figure 1. Simulated marginal posterior densities for θ when n=3

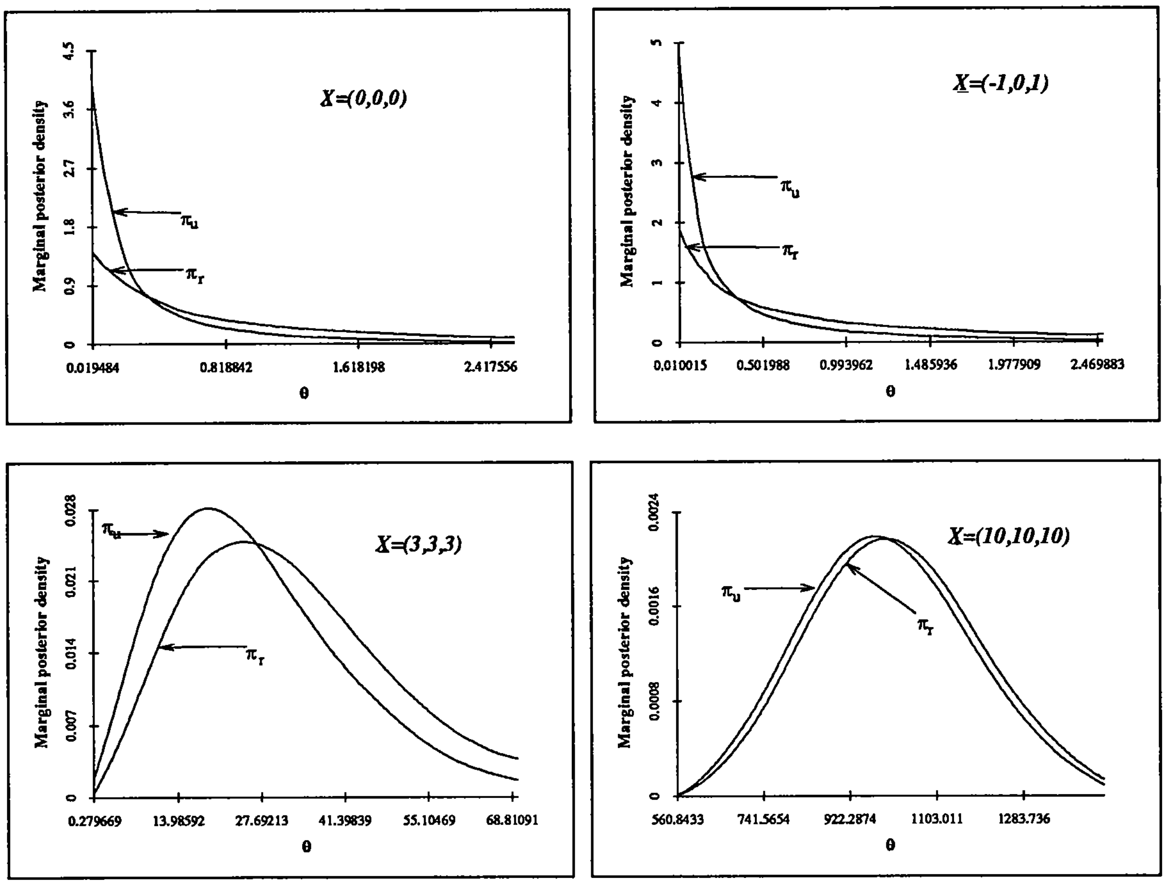

Figure 2. Marginal posterior densities of θ for various x values when n=10