ESTIMATION OF QUADRATIC FUNCTIONS: NONINFORMATIVE PRIORS FOR NON-CENTRALITY PARAMETERS

by

James O. Berger
ISDS, Duke University

Anne Philippe
Université de Rouen

and

Christian P. Robert
Université de Rouen and
CREST-INSEE

Technical Report #96-10C

Department of Statistics
Purdue University
West Lafayette, IN USA

March 1996

◇

Estimation of Quadratic Functions: Noninformative Priors for Non-Centrality Parameters*

JAMES O. BERGER[1], ANNE PHILIPPE[2] AND CHRISTIAN P. ROBERT[2,3]

[1] *ISDS, Duke University*
[2] *URA CNRS 1378, Université de Rouen*
[3] *Laboratoire de Statistique, CREST–INSEE, Paris*

Abstract

The estimation of quadratic functions of a multivariate normal mean is an inferential problem which, while being simple to state and often encountered in practice, leads to surprising complications both from frequentist and Bayesian points of view. The drawbacks of Bayesian inference using the constant noninformative prior are now well established and we consider in this paper the advantages and the shortcomings of alternative noninformative priors. We take into account frequentist coverage probability of confidence sets arising from these priors. Lastly, we derive some optimality properties of the associated Bayes estimators in the special case of independent components under quadratic loss.

Keywords Quantile approximation; chi-squared distribution; orthogonalization; nuisance parameter; Gibbs sampling; importance sampling.

AMS Subject Classification (1991) 62C10, 62C15, 62F10, 62F15, 62F25.

1. INTRODUCTION

Given a normal vector $x \sim \mathcal{N}(\theta, \Sigma)$, with a known covariance matrix Σ, the estimation of $\eta = ||\theta||^2$ is a situation for which the determination of a noninformative prior is troublesome, as pointed out by Stein (1959). For instance, the constant prior on θ (which is the Jeffreys prior here) leads to the definitely suboptimal estimator of η,

$$\delta_0(x) = ||x||^2 + tr(\Sigma),$$

which is uniformly dominated in terms of frequentist risk under squared error loss by $||x||^2 - tr(\Sigma)$. Several authors have addressed this estimation problem (under weighted squared error loss) from a classical point of view, including Perlman and Rassmussen (1975), Neff and Strawderman (1976), Saxena and Alam (1982), Gelfand (1983), Chow (1987) and Kubokawa, Robert and Saleh (1993).

*Research supported, in part, by the U.S. National Science Foundation under grant DMS–9303556. The third author is currently on leave from the Université de Rouen. The authors are grateful to the referees for their helpful and constructive remarks which helped to improve the paper.

Following the reference prior approach proposed in Bernardo (1979), which is designed to deal with nuisance parameters in noninformative settings, we construct in Section 2 two explicit (and competing) noninformative priors when η is the parameter of interest. The particular case $\Sigma = I$ was alluded to in Bernardo (1979) (see also Fernandez, 1982) as an example where reference priors could provide reasonable answers in situations where marginalization paradoxes occur. We compare, in Section 3, the behavior of these two reference priors with each other and with the more naive prior, $\pi(\eta) = 1/\sqrt{\eta}$, which is the reference prior when $\Sigma = I$. The appendix establishes the validity of these priors in terms of posterior propriety. In the particular case when $\Sigma = I$, we study in Section 5 the Bayes estimators associated with $\pi(\eta) = \eta^c$ and show that the reference prior estimator, which corresponds to $c = -1/2$, has certain optimality properties.

Sections 3 and 4 consider the frequentist *confidence behavior* of credible sets arising from these reference priors. Starting with Welch and Peers (1963) and Peers (1965), this has become a common way to study the properties of noninformative priors. For $\Sigma =$ I, Stein (1985) and Tibshirani (1989) (see also Ghosh and Mukerjee (1992, 1993) and Datta and Ghosh (1995)) showed that the reference prior, $\pi(\eta) = 1/\sqrt{\eta}$, is a *frequentist probability matching* prior, in the sense of yielding one-sided credible sets for η (of posterior probability $1-\alpha$) which have frequentist coverage of $1-\alpha$ up to $O(n^{-1})$. For $\Sigma \neq I$, we show that the reference priors from Section 2 are *not* probability matching. However, obtaining any probability matching priors for which the resulting posterior is proper is a formidable task when $p = 2$, and a nearly impossible task for $p \geq 3$. And, even when one can be found, its coverage properties for small n are suspect, as we show in a numerical example. Hence we feel that the practical advantages remain with the reference priors.

The estimation of quadratic functions is of importance in many areas. In astronomy, $||\theta||^2$ arises in the measurement of a celestial object or as an indicator of the accuracy of a measure. Kariya, Giri and Perron (1988) give examples where the norm of the mean appears in the variance of the distribution. Finally, let us mention the following predictive application: if $x^t\theta$ is to be predicted, with $\mathbb{E}[x] = \mu$ and $\mathbb{E}[(x-\mu)(x-\mu)^t] = \Sigma$, we have $\mathbb{E}[x^t\theta] = \mu^t\theta$ and $\text{var}(x^t\theta) = \theta^t\Sigma\theta$, which involves a quadratic function of θ. Berger, Smith and Andrews (1995) address this inferential problem in a car fuel economy study.

2. The direct and reverse reference priors

2.1. Preliminaries. Let $x \sim \mathcal{N}_p(\theta, \Sigma)$. Of interest is inference concerning the parameter $\eta = ||\theta||^2$. First, we can assume without loss of generality that Σ is a diagonal matrix, $diag(\lambda_1, \ldots, \lambda_p)$, even when we estimate $\theta^t Q\theta$ for any p.d. matrix Q. A reparameterization of θ in polar coordinates $(\eta, \varphi_1, \ldots, \varphi_{p-1})$, i.e.

$$\theta_1 = \sqrt{\eta}\cos\varphi_1,$$

$$
\begin{aligned}
\theta_2 &= \sqrt{\eta}\cos\varphi_2\sin\varphi_1,\\
&\;\dots\\
\theta_{p-1} &= \sqrt{\eta}\cos\varphi_{p-1}\sin\varphi_{p-2}\dots\sin\varphi_1,\\
\theta_p &= \sqrt{\eta}\sin\varphi_{p-1}\dots\sin\varphi_1,
\end{aligned} \tag{2.1}
$$

will be used to define the nuisance parameter $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_{p-1})$. For this set of parameters, the information matrix is given by

$$I(\eta, \varphi) = H\Sigma^{-1}H^t = \begin{bmatrix} i_{11} & I_{12}^t \\ I_{12} & I_{22} \end{bmatrix},$$

where i_{11} is a scalar and H is the Jacobian matrix

$$
\begin{aligned}
H &= \frac{D(\theta_1, \dots, \theta_p)}{D(\eta, \varphi_1, \dots, \varphi_{p-1})}\\
&= \begin{bmatrix}
\cos\varphi_1/2\sqrt{\eta} & \cos\varphi_2\sin\varphi_1/2\sqrt{\eta} & \dots & \sin\varphi_{p-1}\dots\sin\varphi_1/2\sqrt{\eta}\\
-\sqrt{\eta}\sin\varphi_1 & \sqrt{\eta}\cos\varphi_1\cos\varphi_2 & & \sqrt{\eta}\sin\varphi_{p-1}\dots\cos\varphi_1\\
0 & -\sqrt{\eta}\sin\varphi_2\sin\varphi_1 & & \dots\\
 & & \dots & \\
0 & 0 & & \sqrt{\eta}\cos\varphi_{p-1}\dots\sin\varphi_1
\end{bmatrix}.
\end{aligned}
$$

Note also that H can be written

$$H = \begin{bmatrix} A^t/\sqrt{\eta} \\ \sqrt{\eta}B \end{bmatrix},$$

where $A \in \mathbb{R}^p$ and B, a $(p-1) \times p$ matrix, are both functions of φ only. By convention, we denote by $|H|$ the absolute value of the determinant of H.

2.2. The direct reference prior. The construction of the reference prior proceeds as follows, using the algorithm in Berger and Bernardo (1992 a,b). First, the conditional distribution of φ given η is $\pi^d(\varphi|\eta) \propto |I_{22}|^{1/2}$. Since Σ is diagonal, we get

$$
\begin{aligned}
I &= \begin{bmatrix} C^t/\sqrt{\eta} \\ \sqrt{\eta}D \end{bmatrix} [\, A/\sqrt{\eta} \quad \sqrt{\eta}B^t \,]\\
&= \begin{bmatrix} C^tA/\eta & (BC)^t \\ DA & \eta DB^t \end{bmatrix},
\end{aligned} \tag{2.2}
$$

with obvious notations; thus, $I_{22} = \eta DB^t$ and $\pi^d(\varphi|\eta)$ does not depend on η (and, indeed, is a proper distribution). The marginal distribution of η is given by

$$\pi^d(\eta) \propto \exp\left\{ \mathbb{E}\left[\frac{1}{2}\log\left(\frac{|I|}{|I_{22}|}\right) \middle| \eta \right] \right\},$$

where $\mathbb{E}$ stands for expectation w.r.t. $\pi^d(\varphi|n)$. Clearly

$$|I| = C^t A|DB^t|\eta^{p-2} + \sum_{i=1}^{p}(DA)_i\omega_i\eta^{p-2},$$
$$|I_{22}| = \eta^{p-1}|DB^t|\ ,$$

as shown by (2.2). Thus

$$\frac{|I|}{|I_{22}|} \propto \frac{1}{\eta}.$$

Because $\pi^d(\varphi|\eta)$ is proper and does not depend on η, the following proposition is immediate.

Proposition 2.1 *The direct reference prior on* (η, φ) *is*

$$\pi^d(\eta, \varphi) = \frac{|I_{22}|_{\eta=1}^{1/2}}{\eta^{1/2}} = \frac{|I_{22}|^{1/2}}{\eta^{p/2}}.$$

The reference posterior is

$$\pi^d(\eta, \varphi|x) \propto \frac{|I_{22}|^{1/2}}{\eta^{p/2}} \exp\{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2\}, \tag{2.3}$$

where $\theta = (\theta_1, \ldots, \theta_p)$ *is given by (2.1).*

The prior $\pi^n(\eta, \varphi) = \eta^{-1/2}|H|_{\eta=1}$, i.e. $\pi^n(\theta) = ||\theta||^{-(p-1)}$, which corresponds to the case $\Sigma = I$, will be called the *naive prior.*

2.3. The reverse reference prior. Another reference prior, π^r, can be constructed for this model; π^r is called the *reverse* reference prior because it considers the parameter of interest, η, and the nuisance parameter, φ, in the reverse order during the derivation. We thus condition first on the nuisance parameters in order to derive the distribution of the parameter of interest. In this case, we have

$$I(\varphi, \eta) = \begin{bmatrix} I_{22} & I_{12} \\ I_{12}^t & i_{11} \end{bmatrix}$$

and

$$i_{11} = \frac{1}{4\eta}(\lambda_1^{-1}\cos^2\varphi_1 + \lambda_2^{-1}\sin^2\varphi_1\cos^2\varphi_2 + \ldots + \lambda_p^{-1}\sin^2\varphi_1\ldots\sin^2\varphi_{p-1}).$$

Therefore, the distribution of η, conditional on φ and on any compact of $\mathbb{R}_+^*$, is

$$\pi^r(\eta|\varphi) \propto \sqrt{i_{11}} \propto 1/\sqrt{\eta}.$$

The marginal reference distribution of φ is then

$$\pi^r(\varphi) \propto \exp\left\{\mathbb{E}\left[(1/2)\log(|I|/i_{11})|\varphi\right]\right\},$$

where $\mathbb{E}$ stands for expectation w.r.t. $\pi^r(\eta|\varphi)$ on the compact. Noting that

$$\frac{|I|}{i_{11}} = \eta^{p-1}\left.\left(\frac{|I|}{i_{11}}\right)\right|_{\eta=1},$$

this marginal distribution is given by

$$\pi^r(\varphi) \propto \exp\left\{(1/2)\log(|I|/i_{11})|_{\eta=1}\right\} = \left.\sqrt{|I|/i_{11}}\right|_{\eta=1}.$$

The following proposition follows directly.

Proposition 2.2 *The reverse reference prior on (η, φ) is*

$$\pi^r(\eta, \varphi) = \frac{|I|^{1/2}_{\eta=1}}{\sqrt{\eta}\,\sqrt{i_{11}}\big|_{\eta=1}}$$

and the associated posterior is

$$\pi^r(\eta, \varphi|x) \propto \frac{|I|^{1/2}}{\sqrt{i_{11}}\eta^{p/2}} \exp\{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2\}. \tag{2.4}$$

Obviously, the two priors (direct and reverse) coincide when $I_{12} = 0$. In particular, this is the case when $\Sigma = I_p$. We show in the appendix that the three priors lead to proper posterior distributions.

3. Computational Issues

The reference priors, π^d and π^r, as well as the naive prior, π^n, allow for a rather straightforward numerical computation of the posterior distributions through Gibbs and importance sampling. In fact, as distributions of θ (in the natural parameterization), these three posteriors involve the expression

$$\frac{e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2}}{||\theta||^{p-1}}, \tag{3.1}$$

since

$$\pi^d(\theta|x) \propto \frac{|I_{22}|^{1/2}_{||\theta||=1}}{||\theta||} \frac{e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2}}{|H|}$$

$$= \frac{|I_{22}|^{1/2}_{||\theta||=1}}{|H|_{||\theta||=1}} \frac{e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2}}{||\theta||^{p-1}},$$

$$\pi^r(\theta|x) \propto \frac{|I|^{1/2}}{|H|\sqrt{i_{11}}} \frac{e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2}}{||\theta||^p}$$

$$\propto \frac{1}{\sqrt{i_{11}}\Big|_{||\theta||=1}} \frac{e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2}}{||\theta||^{p-1}},$$

$$\pi^n(\theta|x) \propto \frac{e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2}}{||\theta||^{p-1}},$$

and $|I| = |H|^2/|\Sigma|$. This decomposition suggests the following simulation lemma.

Lemma 3.1 *Posterior expectations $\mathbb{E}^\pi[h(\theta)|x]$ with respect to one of the three reference priors can be approximated by providing a simulated sample from (3.1) and using, as importance sampling weights, the quantities*

$$\omega^d(\theta) = \frac{|I_{22}|^{1/2}_{||\theta||=1}}{|H|_{||\theta||=1}}, \qquad \omega^r(\theta) = \frac{||\theta||}{\sqrt{i_{11}}} = \frac{1}{\sqrt{i_{11}}\Big|_{||\theta||=1}} \quad \textit{and} \quad \omega^n(\theta) = 1.$$

Proof. Usual importance-sampling arguments can indeed be invoked in this case since the weights are bounded. For the reverse reference prior, we have from (2.5)

$$\left(\sqrt{i_{11}}\Big|_{||\theta|=1}\right)^{-1} \le 2 \sup_i \sqrt{\lambda_i}.$$

The case of the direct reference prior is slightly more intricate. Following the representation $I = H\Sigma^{-1}H^t$, note that we can write I_{22} under the form

$$I_{22} = H_1\Sigma^{-1}H_1^t,$$

where H_1 is a $(p-1)\times p$ matrix such that $H^t = (h_0^t, H_1^t)$. Now, for $\eta = 1$, $|H| = |\sin^{p-2}\varphi_1 \ldots \sin\varphi_{p-2}| = |\Delta|$, where

$$\Delta = \operatorname{diag}(1, |\sin\varphi_1|, |\sin\varphi_1\sin\varphi_2|, \ldots, |\sin\varphi_1\sin\varphi_2\ldots\sin\varphi_{p-2}|).$$

Hence, for $\eta = 1$,

$$\Delta^{-1}H_1 = \begin{bmatrix} -\sin\varphi_1 & \cos\varphi_1\cos\varphi_2 & & \cos\varphi_1\ldots\cos\varphi_{p-1} & \cos\varphi_1\ldots\sin\varphi_{p-1} \\ 0 & \mp\sin\varphi_2 & & \pm\cos\varphi_2\ldots\cos\varphi_{p-1} & \pm\cos\varphi_2\ldots\sin\varphi_{p-1} \\ & & \cdots & & \\ 0 & 0 & & \mp\sin\varphi_{p-1} & \pm\cos\varphi_{p-1} \end{bmatrix}.$$

Therefore, $|I_{22}|^{1/2}/|H| = |(\Delta^{-1}H_1)\Sigma^{-1}(H_1^t\Delta^{-1})|^{1/2}$ is indeed bounded. ■■

For later reference, note that, when $\lambda_2 = \ldots = \lambda_p = 1$, the weights are

$$\omega^d = \sqrt{\cos^2\varphi_1 + \lambda_1^{-1}\sin^2\varphi_1}, \qquad \omega^r = \left\{\lambda_1^{-1}\cos^2\varphi_1 + \sin^2\varphi_1\right\}^{-1/2},$$

since

$$I_{22} = \begin{bmatrix} \cos^2\varphi_1 + \lambda_1^{-1}\sin^2\varphi_1 & 0 & & 0 \\ 0 & \sin^2\varphi_1 & & 0 \\ & & \cdots & \\ 0 & & & \sin^2\varphi_1\ldots\sin^2\varphi_{p-2} \end{bmatrix}.$$

One incentive for using importance sampling is that the three posterior distributions can be evaluated from the same sample, generated from (3.1). Note that the Metropolis algorithm could also be used in this setup with (3.1) as a proposal distribution, since the weights are bounded and this property guarantees geometric convergence (see Mengersen and Tweedie, 1996).

A sample $\theta_1, \ldots, \theta_m$ from (3.1) can be obtained by Gibbs sampling using the following 'hidden mixture' representation:

$$\pi(\theta|x) = \int_0^{+\infty} \pi(\theta, z|x)dz,$$

with

$$\begin{aligned} \pi(\theta, z|x) &\propto e^{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2} e^{-||\theta||^2 z} z^{(p-1)/2-1} \\ &\propto \exp\left\{-\frac{1}{2}(\theta - (\Sigma^{-1}+2zI)^{-1}\Sigma^{-1}x)^t(\Sigma^{-1}+2zI)(\theta - (\Sigma^{-1}+2zI)^{-1}\Sigma^{-1}x)\right\} z^{\frac{p-1}{2}-1}. \end{aligned}$$

The full conditional distributions are then easily available; indeed

$$\pi(z|\theta, x) \quad \text{is} \quad \mathcal{G}a((p-1)/2, ||\theta||^2), \tag{3.2}$$

$$\pi(\theta|z, x) \quad \text{is} \quad \mathcal{N}\left((\Sigma^{-1}+2zI)^{-1}\Sigma^{-1}x, (\Sigma^{-1}+2zI)^{-1}\right), \tag{3.3}$$

and simulation from both distributions is straightforward, especially when Σ is diagonal. The Gibbs sampler then produces a sample $(\theta_1, z_1), \ldots, (\theta_t, z_t)$ by successive simulations from (3.2) and (3.3), and the chain (θ_t) converges to (3.1) in distribution (see, e.g., Robert, 1994).

Lemma 3.1 provides the correction weights ω^d and ω^r for both reference priors and allows us to use the same sample $\theta_1, \theta_2, \ldots$ from the naive prior. One-sided α–credible regions for η are then easily constructed for the three priors and their frequentist coverage can be evaluated by regular Monte Carlo simulation. For instance, Figure 3.1 illustrates

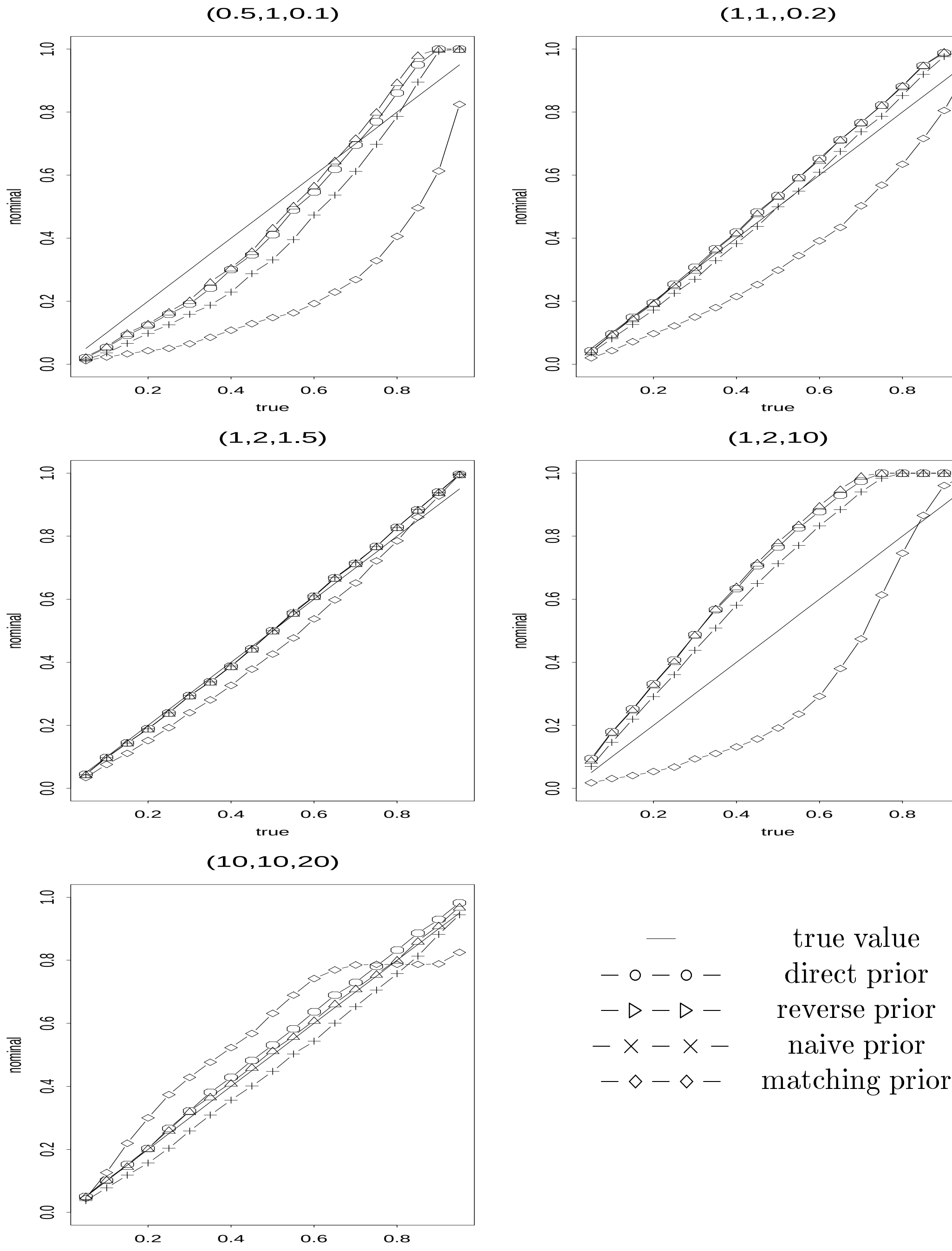

Figure 3.1 – Nominal versus true coverage probabilities of the one-sided α credible set for η when $p = 2$ and $\lambda_2 = 1$, for different values of $(\theta_1, \theta_2, \lambda_1)$. (The frequentist probability matching prior is discussed at the end of Section 4.) The Gibbs sample is of size 5000 and the Monte Carlo evaluation of the coverage is based on 2000 replications.

the behavior of the different priors when $p = 2$ and $\lambda_2 = 1$, for different values of θ_1, θ_2, and λ_1. Small values of the θ_i and λ_1 show a clear domination of both reference priors over the naive prior for most values of α. It is only when α gets close to 1 that the reference credible regions are less satisfactory, the nominal coverage probability being too pessimistic. When the θ_i and λ_1 are large enough, the coverage properties of both reference priors are uniformly very good and slightly dominate those of the naive prior. In addition, comparison between the direct and reverse reference priors turns slightly in favor of the reverse prior. There are, however, cases where the frequentist coverage properties of the three priors are very poor: this occurs for small values of the θ_i when λ_1 is large. In these cases, the three priors behave similarly and the nominal coverage probability is too optimistic.

4. Asymptotically optimal coverage

Recent developments in the noninformative prior literature have considered the frequentist coverage properties of various classes of improper priors in order to determine priors such that one-sided posterior α credible sets for η have also approximately α frequentist coverage; such priors are often called *frequentist probability matching* priors. References for this approach are Welch and Peers (1963), Peers (1965), Stein (1985), Tibshirani (1989), Ghosh and Mukerjee (1992, 1993), and Datta and Ghosh (1995) (see also Berger and Bernardo, 1992b). In particular, from Peers (1965) (see also Stein, 1985, and Tibshirani, 1989) it immediately follows that, if $\theta = (\eta, \omega)$ corresponds to an orthogonal parameterization of θ (in the sense that the Fisher information matrix is block diagonal), with η a scalar, then priors in the class

$$\pi(\eta, \omega) \propto g(\omega) |I_{\eta\eta}(\eta, \omega)|^{1/2} \tag{4.1}$$

lead (asymptotically in the sample size) to the proper frequentist coverage behavior for one-sided credible sets for η. In (4.1), $I_{\eta\eta}$ denotes the sub-matrix of $I(\eta, \omega)$, the Fisher information matrix, which corresponds to the parameter of interest, η. Note that, under the orthogonality assumption, $I(\eta, \omega)$ can be written

$$I(\eta, \omega) = \begin{bmatrix} I_{\eta\eta}(\eta, \omega) & 0 \\ 0 & I_{\omega\omega}(\eta, \omega) \end{bmatrix}. \tag{4.2}$$

It is easy to see from (4.2) why the reference prior approach does not necessarily lead to a distribution satisfying (4.1). Note that the Jeffreys prior associated with (4.2) is

$$\pi^J(\eta, \omega) \propto |I_{\eta\eta}(\eta, \omega)|^{1/2} \, |I_{\omega\omega}(\eta, \omega)|^{1/2},$$

which satisfies (4.1) only if $I_{\omega\omega}$ does not depend on η. A very interesting fact, however, is that a reference prior considering the order η–ω, i.e. a *reverse* reference prior, leads to

$$\pi^r(\eta, \omega) \propto \pi^r(\omega) \, |I_{\eta\eta}(\eta, \omega)|^{1/2}, \tag{4.3}$$

which satisfies (4.1).

Unfortunately, the use of priors satisfying (4.1) calls for an orthogonal parameterization of θ in (η, ω). This is not always possible (see Cox and Reid, 1987) and, moreover, when ω is multidimensional, the solution of $I_{\eta\omega} = 0$, i.e. of several partial differential equations, is typically infeasible. In the following, we only consider the case $p = 2$ and show that the reference priors do not qualify as frequentist matching priors, although their coverage properties are usually acceptable, as shown by the earlier simulation.

Let (η, ω) be in one-to-one correspondence with (θ_1, θ_2). If $\omega_{(i)}$ denotes $\partial\omega/\partial\theta_i$ ($i = 1, 2$), the Jacobian matrix is

$$\frac{D(\eta,\omega)}{D(\theta_1,\theta_2)} = \begin{bmatrix} 2\theta_1 & 2\theta_2 \\ \omega_{(1)} & \omega_{(2)} \end{bmatrix}$$

and the Fisher information is given by

$$\begin{aligned} I(\eta,\omega) &= \begin{bmatrix} \omega_{(2)} & -\omega_{(1)} \\ -2\theta_2 & 2\theta_1 \end{bmatrix} \frac{\Sigma^{-1}}{4(\omega_{(2)}\theta_1 - \omega_{(1)}\theta_2)^2} \begin{bmatrix} \omega_{(2)} & -2\theta_2 \\ -\omega_{(1)} & 2\theta_1 \end{bmatrix} \\ &= \frac{1}{4(\omega_{(2)}\theta_1 - \omega_{(1)}\theta_2)^2} \begin{bmatrix} \omega_{(2)}^2\lambda_1^{-1} + \omega_{(1)}^2\lambda_2^{-1} & -2(\omega_{(2)}\theta_2\lambda_1^{-1} + \omega_{(1)}\theta_1\lambda_2^{-1}) \\ -2(\omega_{(2)}\theta_2\lambda_1^{-1} + \omega_{(1)}\theta_1\lambda_2^{-1}) & 4(\theta_2^2\lambda_1^{-1} + \theta_1^2\lambda_2^{-1}) \end{bmatrix}. \end{aligned}$$

Moreover, the orthogonality requirement leads to the partial differential equation

$$\lambda_2\omega_{(2)}\theta_2 + \lambda_1\omega_{(1)}\theta_1 = 0, \tag{4.4}$$

which is only satisfied by functions of the form

$$\omega = \psi\left(\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\right) \tag{4.5}$$

in each quadrant of $\mathbb{R}^2$.

Note that (4.5) implies

$$\omega_{(1)} = \psi'\left(\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\right)\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\frac{\lambda_2}{\theta_1} \qquad \text{and} \qquad \omega_{(2)} = \psi'\left(\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\right)\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\frac{-\lambda_1}{\theta_2}.$$

Therefore, the prior distributions in (4.1) are of the form

$$\begin{aligned} \pi^o(\eta,\omega) &\propto g(\omega)\,|I_{\eta\eta}(\eta,\omega)|^{1/2} \\ &\propto \frac{\sqrt{\lambda_2\omega_{(2)}^2 + \lambda_1\omega_{(1)}^2}}{2|\omega_{(2)}\theta_1 - \omega_{(1)}\theta_2|}g(\omega) \\ &\propto g(\omega)\frac{\sqrt{\lambda_2\lambda_1^2/\theta_1^2 + \lambda_1\lambda_2^2/\theta_2^2}}{|\lambda_2\theta_2/\theta_1 + \lambda_1\theta_1/\theta_2|} \\ &\propto \frac{g(\omega)}{\sqrt{\lambda_1\theta_1^2 + \lambda_2\theta_2^2}}. \end{aligned}$$

In terms of (θ_1, θ_2), this gives the following family of priors:

$$
\begin{aligned}
\pi^o(\theta_1, \theta_2) &\propto g(\omega(\theta_1, \theta_2)) \frac{1}{\sqrt{\lambda_1 \theta_1^2 + \lambda_2 \theta_2^2}} |\omega_{(2)}\theta_1 - \omega_{(1)}\theta_2| \\
&\propto g(\omega(\theta_1, \theta_2)) \frac{1}{\sqrt{\lambda_1 \theta_1^2 + \lambda_2 \theta_2^2}} \left| \psi' \left(\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}} \right) \right| \frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}} \left| \lambda_1 \frac{\theta_1}{\theta_2} + \lambda_2 \frac{\theta_2}{\theta_1} \right| \\
&\propto h(\omega(\theta_1, \theta_2)) \frac{\sqrt{\lambda_1 \theta_1^2 + \lambda_2 \theta_2^2}}{|\theta_1 \theta_2|}. \qquad (4.6)
\end{aligned}
$$

The reference priors developed in §2.2 and §2.3 do not satisfy (4.6) and, therefore, are not "acceptable" under the frequentist probability matching criterion.

Lemma 4.1 *The direct and reference priors cannot be written in the form (4.6), unless* $\Sigma \propto I$.

Proof. (a) The direct reference prior is given by

$$
\begin{aligned}
\pi^d(\theta_1, \theta_2) &= |I_{22}|^{1/2} \Big/ ||\theta||^2 \\
&= \sqrt{\lambda_1^{-1}\theta_2^2 + \lambda_2^{-1}\theta_1^2} \Big/ ||\theta||^2.
\end{aligned}
$$

Were π^d of the form (4.6), there would exist a function h such that

$$
\sqrt{\lambda_1 \theta_1^2 + \lambda_2 \theta_2^2}\; h\left(|\theta_1|^{\lambda_2}/|\theta_2|^{\lambda_1}\right) \Big/ |\theta_1 \theta_2| = \sqrt{\lambda_1^{-1}\theta_2^2 + \lambda_2^{-1}\theta_1^2}/||\theta||^2,
$$

i.e.

$$
h\left(\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\right) = \frac{|\theta_1 \theta_2|}{||\theta||^2} = \left(\frac{|\theta_1|}{|\theta_2|} + \frac{|\theta_2|}{|\theta_1|}\right)^{-1} = t\left(\frac{|\theta_1|}{|\theta_2|}\right).
$$

This is impossible, unless $\lambda_1 = \lambda_2$.

(b) In the same way, if (4.6) was satisfied by π^r, there would exist a function h such that

$$
\sqrt{\lambda_1 \theta_1^2 + \lambda_2 \theta_2^2}\; h\left(|\theta_1|^{\lambda_2}/|\theta_2|^{\lambda_1}\right) \Big/ |\theta_1 \theta_2| = \frac{1}{\sqrt{\lambda_1^{-1}\theta_1^2 + \lambda_2^{-1}\theta_2^2}},
$$

i.e.

$$
h\left(\frac{|\theta_1|^{\lambda_2}}{|\theta_2|^{\lambda_1}}\right) = \frac{|\theta_1 \theta_2|}{\sqrt{\lambda_1 \theta_2^2 + \lambda_2 \theta_1^2}\sqrt{\lambda_1 \theta_1^2 + \lambda_2 \theta_2^2}} = t\left(\frac{|\theta_1|}{|\theta_2|}\right). \qquad (4.7)
$$

Again, (4.7) cannot be satisfied unless $\lambda_1 = \lambda_2$. ■■

When $\lambda_1 \neq \lambda_2$, it is difficult to find a prior of the form (4.6) for which the posterior is proper. A lengthy search uncovered the following solution. Choose ψ to be the identity function in (4.5), so that

$$w = |\theta_1|^{\lambda_2}/|\theta_2|^{\lambda_1}.$$

Then, in the second line of (4.6), choose

$$g(w) = w^{(\frac{1}{\lambda_2}-1)}/[(1+w^{\frac{1}{\lambda_2}})(1+(\log\ w^{\frac{1}{\lambda_2}})^2)],$$

where we assume, *w.l.o.g.*, that $\lambda_1 \leq \lambda_2$. Then the last line of (4.6) can be written (defining $\gamma = \lambda_1/\lambda_2$)

$$\pi^0(\theta_1,\theta_2) \propto \frac{[1+\gamma\theta_1^2/\theta_2^2]^{1/2}}{(|\theta_1|+|\theta_2|^\gamma)[1+(\log\ |\theta_1| - \gamma\log\ |\theta_2|)^2]}\ . \tag{4.8}$$

As proved in Appendix A.3, this frequentist probability matching prior leads to a proper posterior distribution. Figure 3.1 shows, however, that the performance of this prior is inferior to that of the reference priors for the cases considered. Given this poor small sample ($n = 1$) performance and the difficulty of finding a usable probability matching prior, even for $p = 2$, the clear edge would seem to lie with the reference prior methodology here.

5. The particular case of the identity matrix

When Σ is the identity matrix, I_p, it is possible to provide a closed form expression for the Bayes estimator, $\mathbb{E}^\pi[\eta|x]$. In this case, the three reference priors discussed above (naive, direct and reverse) are identical and equal to $\pi^r(\eta,\varphi) = \eta^{-1/2}|H|_{\eta=1}$.

Proposition 5.1 *The Bayes estimator associated with the prior* $\pi_c(\eta,\varphi) = |H|_{\eta=1}\eta^{-c}$ $(c < 1)$ *under squared error loss is*

$$\delta_c^\pi(x) = 2(1-c)\frac{{}_1F_1(2-c;p/2;||x||^2/2)}{{}_1F_1(1-c;p/2;||x||^2/2)}, \tag{5.1}$$

where ${}_1F_1$ *is the confluent hypergeometric function.*

Proof. Let us denote $z = ||x||^2$. Then, $z \sim \chi_p^2(\eta)$ and

$$\delta_c^\pi(z) = \frac{\int \eta(z/\eta)^{(p-2)/4}e^{-(z+\eta)/2}I_{\frac{p-2}{2}}(\sqrt{z\eta})\eta^{-c}d\eta}{\int (z/\eta)^{(p-2)/4}e^{-(z+\eta)/2}I_{\frac{p-2}{2}}(\sqrt{z\eta})\eta^{-c}d\eta}\ ,$$

where I_ν denotes the modified Bessel function (see Abramowitz and Stegun, 1964, or Gradshteyn and Ryzhik, 1980). We have

$$\int_0^{+\infty} (z/\eta)^{(p-2)/4}e^{-(z+\eta)/2}I_{\frac{p-2}{2}}(\sqrt{z\eta})\eta^{-c}d\eta$$

$$\propto \int_0^{+\infty} \eta^{-q} e^{-\eta/2} I_{\frac{p-2}{2}}(\sqrt{z\eta})d\eta$$

where $q = (p-2+4c)/4$

$$\propto \int_0^{+\infty} \frac{u^{-2q}}{z^{-q}} e^{-u^2/2z} I_{\frac{p-2}{2}}(u) u du$$

$$= z^q \int_0^{+\infty} u^{1-2q} e^{-u^2/2z} I_{\frac{p-2}{2}}(u) du$$

$$= z^q \frac{\Gamma\left(\frac{p-2}{4}+1-q\right)}{\Gamma(p/2)2^{p/2}\left(1/\sqrt{2z}\right)^{2-2q+(p-2)/2}} \, {}_1F_1\left(\frac{p-2+4-4q}{2}; p/2; z/2\right)$$

(Gradshteyn and Ryzhik, 1980)

$$= z^{(p+2)/4} \frac{\Gamma(1-c)}{\Gamma(p/2)2^{p/2}(1/2)^{1-c}} \, {}_1F_1\left(1-c; p/2; z/2\right) \; .$$

Therefore,

$$\delta_c^\pi(z) = \frac{z^{(p+2)/4}\Gamma(2-c)2^{1-(c-1)}}{z^{(p+2)/4}\Gamma(1-c)2^{1-c}} \frac{{}_1F_1(2-c; p/2; z/2)}{{}_1F_1(1-c; p/2; z/2)}$$
$$= 2(1-c)\frac{{}_1F_1(2-c; p/2; z/2)}{{}_1F_1(1-c; p/2; z/2)}. \qquad \blacksquare\blacksquare$$

We are thus able to directly compute the Bayes estimator in this case. (Note that the confluent hypergeometric function is now implemented in most packages, including *Mathematica* and *Maple*. For instance, Figure 5.1, which gives δ_c^π for several values of c, is obtained by *Maple*) The expression (5.1) also leads to the following asymptotic approximation to δ_c^π:

Corollary 5.2 *As* $z = ||x||^2$ *goes to infinity,*

$$\delta_c^\pi(z) = z - (p + 4c - 4) + o(1).$$

Proof. It follows from Abramowitz and Stegun (1964) that

$${}_1F_1(d; p/2; z/2) \simeq \frac{\Gamma(p/2)}{\Gamma(d)} e^{z/2}(z/2)^{d-(p/2)} \left(1 + \frac{(1-d)(p/2-d)}{(z/2)}\right).$$

Thus,

$$\frac{{}_1F_1(2-c; p/2; z/2)}{{}_1F_1(1-c; p/2; z/2)} \simeq \frac{\Gamma(1-c)}{\Gamma(2-c)}(z/2)^{2-c-(1-c)} \frac{1 + (2(c-1)(p/2+c-2)/z)}{1 + (2c(p/2+c-1)/z)}$$

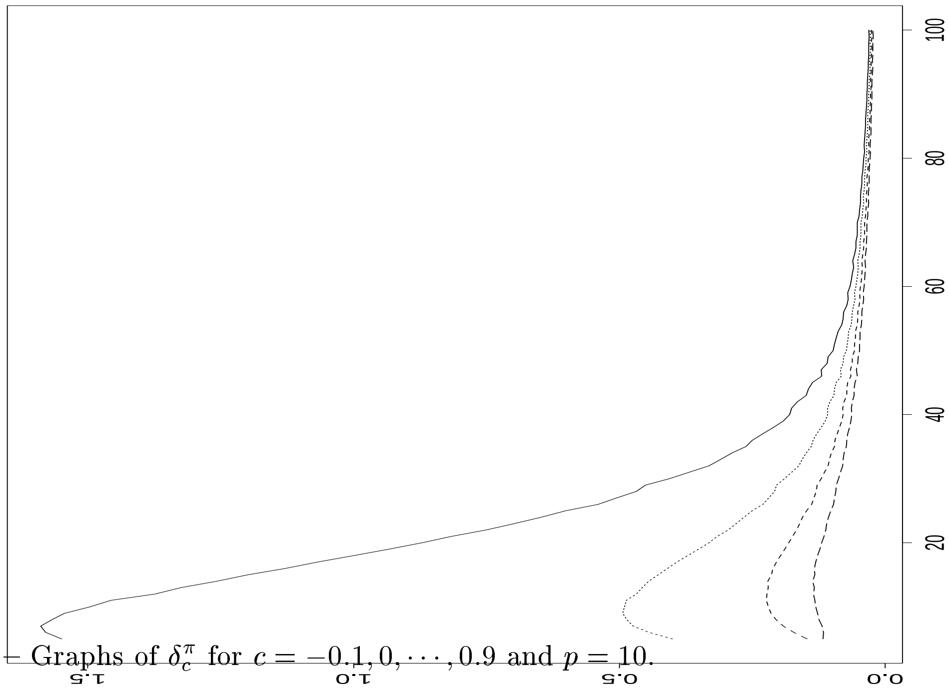

Figure 5.1 – Graphs of δ_c^π for $c = -0.1, 0, \cdots, 0.9$ and $p = 10$.

and

$$\begin{aligned}
\delta_c^\pi(z) &\simeq z\left(1 + 2\frac{(c-1)(c-2+(p/2))}{z}\right)\left(1 - 2\frac{c(c-1+(p/2))}{z}\right) \\
&\simeq z\left(1 + 2\frac{(c-1)(c-2+(p/2)) - c(c-1+(p/2))}{z}\right) \\
&= z - 2(2c - 2 + (p/2)).
\end{aligned}$$

■■

It is of interest to consider the frequentist risk of δ_c^π under squared error loss and for large values of η. Indeed, using Corollary 5.2 it can be shown that, for large η,

$$\mathbb{E}_\eta(\delta_c^\pi(x) - \eta)^2 \simeq 2p + 4\eta + 16(c-1)^2.$$

Hence it would appear that $c = 1$ is an attractive choice. However, the posterior resulting from this choice is not proper, so that δ_1^π is not even defined. Furthermore, it is natural here to consider a weighted loss such as

$$(\delta(x) - \eta)^2/(2p + 4\eta),$$

and for this loss it can be shown that $\pi_0(\eta, \varphi) = 1$ (i.e., $c = 0$) is optimal for large η. Not only is the reference prior ($c = 1/2$) intermediate between these two extremes, but the resulting Bayes rule can be shown to be optimal for large η under the "intermediate" loss

$$(\delta(x) - \eta)^2/\sqrt{2p + 4\eta}.$$

Appendix

A.1. *The naive prior results in a proper posterior.*

According to (3.1), the posterior for the naive prior is proportional to

$$g(\theta) = ||\theta||^{1-p} \exp\left\{-(x-\theta)^t\Sigma^{-1}(x-\theta)/2\right\}.$$

Consider the polar coordinates of θ as in §2.1; then $\theta = \sqrt{\eta}\,\alpha = r\,\alpha$ with $\alpha = (\alpha_1(\varphi), \cdots, \alpha_p(\varphi))$ $= (\alpha_1, \cdots, \alpha_p)$. The associated Jacobian is

$$\Delta_p = r^{p-1}\sin^{p-2}(\varphi_1)\cdots\sin(\varphi_{p-2}).$$

Therefore,

$$\begin{aligned}\int_{\mathbb{R}^p} g(\theta)\,d\theta &= \int_\Phi\int_{\mathbb{R}^+}\Delta_p \exp\left\{-(x-r\alpha)'\Sigma^{-1}(x-r\alpha)\right\} r^{1-p}\,dr\,d\varphi \\ &= \int_\Phi \sin^{p-2}(\varphi_1)\cdots\sin(\varphi_{p-2})\int_{\mathbb{R}^+}\exp\left\{-(x-r\alpha)'\Sigma^{-1}(x-r\alpha)\right\}\,dr\,d\varphi\end{aligned}$$

Denoting $<\alpha, x> = \alpha'\Sigma^{-1}x$ and $N(\alpha) = <\alpha,\alpha>$, we get

$$\begin{aligned}&\int_{\mathbb{R}^+}\exp\left\{-(x-r\alpha)'\Sigma^{-1}(x-r\alpha)/2\right\}\,dr \\ &\quad= \int_{\mathbb{R}^+}\exp\left\{-(N(x) - 2r<\alpha,x> + r^2N(\alpha))/2\right\}\,dr \\ &\quad= \exp\{-N(x)/2\}\int_{\mathbb{R}^+}\exp\left\{-\frac{N(\alpha)}{2}\left(-2r\frac{<\alpha,x>}{N(\alpha)} + r^2\right)\right\}\,dr \\ &\quad= \exp\{-N(x)/2\}\exp\left\{\frac{<\alpha,x>^2}{2N(\alpha)}\right\}\int_{\mathbb{R}^+}\exp\left\{-N(\alpha)\left(r - \frac{<\alpha,x>}{N(\alpha)}\right)^2\right\}\,dr \\ &\quad\le \exp\{-N(x)/2\}\exp\{<\alpha,x>^2/2N(\alpha)\}\sqrt{\frac{2\pi}{N(\alpha)}},\end{aligned}$$

which implies

$$\int_{\mathbb{R}^p} g(\theta)\,d\theta \le I\mathrm{e}^{-N(x)}\int_\Phi \sin^{p-2}(\varphi_1)\cdots\sin(\varphi_{p-2})\exp\left\{\frac{<\alpha,x>^2}{N(\alpha)}\right\}(N(\alpha))^{-1/2}\,d\varphi \qquad (A.1)$$

where I is a constant. Since α is on the unit sphere,

$$0 < \inf_{i\le p}\lambda_i^{-1} \le N(\alpha) \le \sup_{i\le p}\lambda_i^{-1} < \infty\,.$$

Therefore $(A.1)$ is bounded and the posterior is proper.

A.2. *The reverse and direct reference priors result in proper posteriors.*

Since, according to Lemma 3.1,

$$\pi^d(\theta|x) \propto \omega^d(\theta)g(\theta) \qquad \text{and} \qquad \pi^r(\theta|x) \propto \omega^r(\theta)g(\theta),$$

with $\omega^r(\theta)$ and $\omega^d(\theta)$ bounded, the posterior distributions associated with both direct and reverse reference priors are proper.

A.3. *The frequentist probability matching prior results in a proper posterior.*

We want to prove that

$$\int \int \exp\{\frac{-1}{2}[\lambda_1^{-1}(x_1-\theta_1)^2+\lambda_2^{-1}(x_2-\theta_2)^2]\}\, \pi^0(\theta_1,\theta_2)d\theta_1 d\theta_2 < \infty.$$

For simplicity, we only integrate over $\theta_1 > 0$ and $\theta_2 > 0$. Note that, since $\lambda_1 \leq \lambda_2$,

$$[\lambda_1\theta_2^{-2}+\lambda_2\theta_1^{-2}]^{\frac{1}{2}} \leq \lambda_2[\theta_1^{-1}+\theta_2^{-1}]. \tag{A.2}$$

It is immediate that $\pi^0(\theta_1,\theta_2) \leq \lambda_2$ for $\theta_2 > 1$, so the integral is clearly finite over this region. For $\theta_2 < 1$,

$$(\theta_1+\theta_2^{\gamma})^{-1} \leq (\theta_1+\theta_2)^{-1},$$

which together with $(A.2)$ implies that

$$\pi^0(\theta_1,\theta_2) \leq \frac{\lambda_2}{\theta_2[1+(\log\,\theta_1-\gamma\log\theta_2)^2]}\ .$$

Thus

$$\int_0^{\infty}\int_0^{1} \exp\left\{\frac{-1}{2\lambda_1}(x_1-\theta_1)^2-\frac{1}{2\lambda_2}(x_2-\theta_2)^2\right\}\pi^0(\theta_1,\theta_2)d\theta_1 d\theta_2$$

$$\leq \lambda_2 \int_0^{\infty}\int_0^{1}\ e^{-(x_1-\theta_1)^2/2\lambda_1}\frac{1}{\theta_2[1+(\log\theta_1-\gamma\log\theta_2)^2]}d\theta_2 d\theta_1.$$

Defining $\xi = \log\theta_2$ and changing variables yields the equivalent integral

$$\lambda_2\int_0^{\infty}\int_{-\infty}^{0}\ e^{-(x_1-\theta_1)^2/2\lambda_1}\ \frac{1}{[1+(\log\theta_1-\gamma\xi)^2]}d\xi d\theta_1,$$

which is clearly finite.

References

Abramowitz, M. and Stegun, I. (1964) *Handbook of Mathematical Functions.* Dover.

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

Berger, J. and Bernardo, J. (1992a) Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics*, P.K. Goel and N.S. Iyengar (Eds.). Springer-Verlag, New York.

Berger, J. and Bernardo, J. (1992b) On the development of the reference prior method. In *Bayesian Statistics 4*, J. Berger, J. Bernardo, P. Dawid and A.F. Smith (Eds.), 35-49. Oxford University Press, London.

Berger, J., Smith, M. and Andrews, R. (1995) A system for estimating fuel economy due to technology improvement. In *Simulation Approaches to Statistics and Econometrics*, G. Geweke and A. Monfort (Eds.). North-Holland, Amsterdam.

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

Chow, M. (1987) A complete class theorem for estimating a non-centrality parameter. *Ann. Statist.* **15**, 869-876.

Cox, D. and Reid, N. (1987) Orthogonal parameters and approximate conditional inference (with discussion). *J. Royal Statist. Soc. B* **49**, 1-39.

Fernandez, J. (1982) Una solución Bayesiana a la paradoja de Stein. *Trab. Estadist.* **33**, 31-46.

Datta, G. S. and Ghosh, J. K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika* **82**, 37–45.

Gelfand, A. (1983) Estimation in noncentral distributions. *Comm. Statist.* A **12**(4), 463-475.

Ghosh, J. K. and Mukerjee, R. (1992) Non-informative prior. In *Bayesian Statistics 4*, J. Berger, J. Bernardo, P. Dawid and A. F. Smith (Eds.), 195-203. Oxford University Press.

Ghosh, J. K. and Mukerjee, R. (1993) Frequentist validity of highest posterior density regions in the multiparameter case. *Ann. Inst. Statis. Math.* **45**, 293-302.

Gradshteyn, I. and Ryzhik, I. (1980) *Table of Integrals, Series and Products.* Academic Press, New-York.

Kariya, T., Giri, N. and Perron, F. (1988) Invariant estimation of mean vector μ of $\mathcal{N}(\mu, \Sigma)$ with $\mu'\Sigma^{-1}\mu = 1$ or $\Sigma^{-1/2}\mu = C$ or $\Sigma = \delta^2\mu'\mu I$. *J. Multivariate Anal.* **27**, 270-283.

Kubokawa, T., Robert, C. and Saleh, A. (1993) Estimating noncentrality parameters. *Canad. J. Statist.*, **21**, 54-58.

Mengersen, K. L. and Tweedie, R. L. (1996) Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101-121.

Neff, N. and Strawderman, W. (1976) Further remarks on estimating the parameter of a noncentral chi squared distribution. *Comm. Stat. A* **5**, 65-76.

Peers, H. W. (1965). On confidence points and Bayesian probability points in the case of several parameters. *J. Royal Statist. Soc. B* **27**, 9–16.

Perlman, M. and Rassmussen, V. (1975) Some remarks on estimating a noncentrality parameter. *Comm. Stat. A* **4**, 455-468.

Robert, C. P. (1994) *The Bayesian Choice.* Springer-Verlag, New York.

Saxena, K. and Alam, K. (1982) Estimation of the non-centrality parameter of a chi-squared distribution. *Ann. Statist.* **10**, 1012-1016.

Stein, C. (1959) An examination of wide discrepancy between fiducial and confidence intervals. *Ann. Math. Statist.***30**, 877-880.

Stein, C. (1985) On the coverage of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach Center Publication, **16**. PWN-Polish publishers, Warsaw.

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

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