International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 9(Special Issue), 2009, pp 255-271

Probability Matching Priors for Ratio of Variances of a Bivariate Normal Distribution

Malay Ghosh
Department of Statistics
University of Florida
Gainesville, Florida, USA
E-mail: ghoshm@stat.ufl.edu

Bhramar Mukherjee
Department of Biostatistics
School of Public Health
University of Michigan
Ann Arbor, Michigan, USA
E-mail: bhramar@umich.edu

Upasana Santra
Department of Statistics
University of Florida
Gainesville, Florida, USA
E-mail: usantra@stat.ufl.edu

[Received December 30, 2008; Revised June 14, 2009; Accepted June 30, 2009]

Abstract

The paper derives a general class of priors for inference regarding the ratio of standard deviations from a bivariate normal population with nonzero correlation. This class of priors satisfies a matching property in the sense that the coverage probabilities of resultant credible intervals matches asymptotically the corresponding frequentist coverage probabilities up to a high order of approximation. The propriety of resultant posteriors is proved under very mild conditions, and a simulation study suggests that the approximation is valid even for moderate sample sizes.

Keywords and Phrases: Distribution Matching, First Order, HPD Matching, Inversion of Test Statistics, Posterior Propriety, Quantile Matching and Second Order.

AMS Classification: 60E$_{XX}$.

1 Introduction

There are many experimental situations in which an investigator wants to estimate the ratio of variances of two independent normal populations. Study of the ratio of variances dates back to 1920 when Fisher developed the F-statistic for testing the variance ratio. The most well-used example involves testing of the hypothesis that the standard deviations of two normally distributed populations are equal. Although ratio of variances has been vigorously studied in the case of two independent normal samples both in the frequentist and in the Bayesian literature, little study has been done for a possibly correlated bivariate normal population. For testing the equality of variances in a bivariate normal population, Pitman (1939) and Morgan (1939) introduced a variable transformation which reduces the problem to testing a bivariate normal correlation coefficient equal to zero. This same idea can be extended easily to test the null hypothesis whether a variance ratio equals a particular value. Inverting this test statistic, Roy and Potthoff (1958) obtained confidence bounds on the ratio of variances in the correlated bivariate normal distribution. Since the test statistic has a Students's t-distribution under the null hypothesis, the resulting confidence bounds involve percentiles of a Student t-distribution.

Probability matching criterion amounts to the requirement that the coverage probability of a Bayesian credible region is asymptotically equivalent to the coverage probability of the corresponding frequentist confidence region upto a certain order. This has found some appeal to both frequentists and Bayesians. An excellent monograph on this topic is due to Datta and Mukerjee (2004) which provides a thorough and comprehensive discussion of various probability matching criteria. Other review papers are due to Kass and Wasserman (1996), Ghosh and Mukerjee (1998) and Datta and Sweeting (2005).

Again, as one might expect, there are several probability matching criteria. The matching is accomplished through either (a) posterior quantiles, (b) distribution functions, (c) highest posterior density (HPD) regions, or (d) inversion of certain test statistics. However, priors based on (a),(b),(c), or (d) need not always be identical. Specifically, it may so happen that there does not exist any prior satisfying all the four criteria.

The objective of this article is to find a general class of priors which meet all four matching criteria when the ratio of variances in the bivariate normal distribution is the parameter of interest and compare the performance of several competing priors for moderate sample sizes.

The outline of the remaining sections is as follows. In Section 2 of this paper, we have introduced an orthogonal reparameterization of the bivariate normal parameters. Section 3 develops a class of quantile matching priors, whereas Section 4 develops a general class of distribution function matching priors. Matching priors based on HPD regions are given in Section 5 while those based on the inversion of likelihood-ratio test statistics are developed in Section 6. The propriety of the posteriors is established in Section 7. Section 8 undertakes a simulation study. Some final remarks are made in Section 9.

One may wonder about the need for probability matching priors. It is generally agreed upon that with adequate historical data, one should elicit a suitable subjective prior for a given problem. But even in the absence of such information, Bayesian methods can be used very effectively with some objective prior. One criterion of objectivity is to achieve Bayes-frequentist synthesis through the asymptotic equivalence of the coverage probabilities of credible and confidence intervals.

2 The Orthogonal Parameterization

Let $(X_{1i}, X_{2i}), (i = 1, \ldots, n)$ be independent and identically distributed random variables having a bivariate normal distribution with means μ_1 and μ_2, variances ${\sigma_1}^2(> 0)$ and ${\sigma_2}^2(> 0)$, and correlation coefficient ρ $(|\rho| < 1)$. Using the transformation

$$\theta_1 = \sigma_1/\sigma_2,\ \theta_2 = \sigma_1\sigma_2(1-\rho^2)^{1/2} \text{ and } \theta_3 = \rho, \tag{2.1}$$

the bivariate normal pdf can be rewritten as

$$f(X_1, X_2 \mid \mu_1, \mu_2, \theta_1, \theta_2, \theta_3) \propto$$
$$\frac{1}{\theta_2} \exp\left[-\frac{1}{2(1-\theta_3^2)^{1/2}\theta_2}\left\{\frac{(X_1-\mu_1)^2}{\theta_1} + \theta_1(X_2-\mu_2)^2 - 2\theta_3(X_1-\mu_1)(X_2-\mu_2)\right\}\right]. \tag{2.2}$$

Since we are interested only in the ratio of variances, essentially without loss of generality, we can assume that $\mu_1 = \mu_2 = 0$. This involves in practice loss of one degree of freedom which does not affect the asymptotic methodology that we are going to discuss.

With this simplification, the Fisher Information matrix based on (2.2) reduces to

$$\mathbf{I}(\theta_1, \theta_2, \theta_3) = \text{Diag}(\theta_1^{-2}(1-\theta_3^2)^{-1}, \theta_2^{-2}, (1-\theta_3^2)^{-2}). \tag{2.3}$$

This establishes immediately the mutual orthogonality of θ_1,θ_2 and θ_3 in the sense of Huzurbazar (1950) and Cox and Reid (1987). Such orthogonality is often referred to as "Fisher Orthogonality".

The inverse of the information matrix is simply then

$$\mathbf{I}^{-1}(\theta_1, \theta_2, \theta_3) = \text{Diag}(\theta_1^2(1-\theta_3^2), \theta_2^2, (1-\theta_3^2)^2). \tag{2.4}$$

For subsequent sections, we need also a few other results which are collected in the following lemma.

Lemma 2.1 For the bivariate normal density given in (2.2),

$$\mathbf{E}\Big(\frac{\partial^3 \log f}{\partial\theta_1^2 \partial\theta_3}\Big) = -\frac{\theta_3}{\theta_1^2(1-\theta_3^2)^2},\ \mathbf{E}\Big(\frac{\partial^3 \log f}{\partial\theta_1^2 \partial\theta_2}\Big) = \frac{1}{\theta_1^2\theta_2(1-\theta_3^2)}; \tag{2.5}$$

$$\mathbf{E}\Big(\frac{\partial^3 \log f}{\partial\theta_1^3}\Big) = \frac{3}{\theta_1^3(1-\theta_3^2)},\ \mathbf{E}\Big(\frac{\partial \log f}{\partial\theta_1}\Big)^3 = 0; \tag{2.6}$$

$$\mathbf{E}\left(\Big(\frac{\partial \log f}{\partial\theta_1}\Big)\Big(\frac{\partial^2 \log f}{\partial\theta_1^2}\Big)\right) = -\frac{1}{\theta_1^3(1-\theta_3^2)}; \tag{2.7}$$

$$\mathbf{E}\Big(\frac{\partial^3 \log f}{\partial\theta_1\partial\theta_3^2}\Big) = 0, \mathbf{E}\Big(\frac{\partial^3 \log f}{\partial\theta_1\partial\theta_2^2}\Big) = 0; \tag{2.8}$$

We derive the matching priors in the next few sections.

3 Quantile Matching Priors

Here one is interested in the approximate frequentist validity of the posterior quantiles of a one-dimensional interest parameter. The pioneering research in this area was due to Welch and Peers (1963) and Peers (1965), while the recent stimulus came from Stein (1985) and Tibshirani (1989). Specifically, one considers here priors $\pi(.)$ for which the relation

$$P\{\theta_1 \leq {\theta_1}^{1-\alpha}(\pi, X)|\theta\} = 1 - \alpha + o(n^{-\frac{r}{2}}), \tag{3.1}$$

holds for $r = 1$ or 2 and for each $\alpha\,(0 < \alpha < 1)$. In the above, $\theta = (\theta_1, \ldots, \theta_p)^T$ is the unknown parameter, θ_1 is the one-dimensional parameter of interest, ${\theta_1}^{1-\alpha}(X)$ is the asymptotic first order bias-corrected $(1-\alpha)$th posterior quantile of θ_1 based on the prior π and data $X = (X_1, \ldots, X_n)^T$, while $P(.|\theta)$ denotes the conditional probability given θ, the usual frequentist probability. A prior satisfying (3.1) with r=1 is called a *first order* probability matching prior, while one with r=2 is called a *second order* probability matching prior. Clearly, second order probability matching priors constitute a subclass of first order probability matching priors.

Due to orthogonality of θ_1 with (θ_2, θ_3), from Tibshirani(1989), the class of first order matching priors is characterized by

$$\pi(\theta_1, \theta_2, \theta_3) \propto \theta_1^{-1}(1-\theta_3^2)^{-1/2} g(\theta_2, \theta_3). \tag{3.2}$$

A prior of the above form satisfies the second-order quantile matching property if and only if (see (2.5.26) of Datta and Mukerjee (2004, p27)) g satisfies the relation

$$\frac{\partial}{\partial\theta_2}\big\{\theta_1^{-1}(1-\theta_3^2)^{1/2} g\ \theta_1^2\theta_2^2 E\big(\frac{\partial^3 \log f}{\partial\theta_1^2\partial\theta_2}\big)\big\} + \frac{\partial}{\partial\theta_3}\big\{\theta_1^{-1}(1-\theta_3^2)^{1/2} g\ \theta_1^2(1-\theta_3^2)^2 E\big(\frac{\partial^3 \log f}{\partial\theta_1^2\partial\theta_3}\big)\big\}$$
$$+ \frac{1}{6}(1-\theta_3^2)^{-1/2} g \frac{\partial}{\partial\theta_1}\big\{\theta_1^3(1-\theta_3^2)^{3/2} E\big(\frac{\partial \log f}{\partial\theta_1}\big)^3\big\} = 0. \tag{3.3}$$

¿From (2.5)-(2.8), (3.3) simplifies to

$$\theta_1^{-1}(1-\theta_3^2)^{-1/2}\frac{\partial}{\partial\theta_2}\big\{g\ \theta_2\big\} - \theta_1^{-1}\frac{\partial}{\partial\theta_3}\big\{g\ \theta_3(1-\theta_3^2)^{1/2}\big\} = 0. \tag{3.4}$$

Now let g be the class of functions given by $g(\theta_2, \theta_3) = \theta_2^a(\theta_3^2)^b(1-\theta_3^2)^c$. With this choice of g the left hand side of the above equation reduces after some simplification to

$$\theta_1^{-1}\theta_2^a(\theta_3^2)^b(1-\theta_3^2)^{c-\frac{1}{2}}[a - 2b + 2(b+c+1)\theta_3^2] = 0, \tag{3.5}$$

which leads to the solution $b = a/2$ and $c = -(a+2)/2$. Thus every prior $\pi(\theta_1, \theta_2, \theta_3) \propto \theta_1^{-1}\theta_2^a|\theta_3|^a(1-\theta_3^2)^{-\frac{a+3}{2}}$ is a second order probability matching prior for θ_1. Due to the invariance property of such a prior (Mukerjee and Ghosh, 1997), back to the original parameterization, a second order matching prior for $\frac{\sigma_1}{\sigma_2}$ is given by $\pi(\sigma_1, \sigma_2, \rho) \propto \sigma_1^a\sigma_2^a|\rho|^a(1-\rho^2)^{-1}$.

4 Matching Via Distribution Functions

In this section, we target priors π which achieve matching via distribution functions of some standardized variables. More specifically, when θ_1 is the parameter of interest, while $(\theta_2, \ldots, \theta_p)^T$ is the vector of nuisance parameters, writing $\hat{\theta_1}$ as the MLE of θ_1 with $n^{-\frac{1}{2}}I^{11}$ as its asymptotic variance, $\big(I = ((I_{jj'})), I^{-1} = ((I^{jj'}))\big)$, we consider the random variable $y = \sqrt{n}(\theta_1 - \hat{\theta_1})/(I^{11})^{1/2}$. Specifically, if P^π denotes the posterior of y given the data X, what we want to achieve is the asymptotic matching

$$\mathbf{E}[P^\pi(y \le w|X)|\theta] = P(y \le w|\theta) + o(n^{-1}). \tag{4.1}$$

Under orthogonality of θ_1 with $(\theta_2, \ldots, \theta_p)$, it follows from (3.2.5) to (3.2.7) of Datta and Mukerjee (2004) that such a prior π is of the form ${I_{11}}^{1/2}g(\theta_2, \ldots, \theta_p)$, where in addition one needs to satisfy the two differential equations

$$\begin{aligned} A_1 =& \frac{\partial^2}{\partial{\theta_1}^2}\big(I^{11}\pi(\theta)\big) - 2\frac{\partial}{\partial\theta_1}(I^{11}\frac{\partial}{\partial\theta_1}\pi) - \sum_{s=2}^{p}\sum_{v=2}^{p}\frac{\partial}{\partial\theta_s}\Big\{E\big(\frac{\partial^3 \log f}{\partial{\theta_1}^2\partial\theta_s}\big)I^{11}I^{sv}\pi(\theta)\Big\} \\ &- \sum_{s=2}^{p}\sum_{v=2}^{p}\frac{\partial}{\partial\theta_1}\Big\{E\big(\frac{\partial^3 \log f}{\partial\theta_1\partial\theta_s\partial\theta_v}\big)I^{11}I^{sv}\pi(\theta)\Big\} = 0. \end{aligned} \tag{4.2}$$

and

$$A_2 = \sum_{s=2}^{p}\sum_{v=2}^{p}\frac{\partial}{\partial\theta_s}\Big\{E\big(\frac{\partial^3 \log f}{\partial{\theta_1}^2\partial\theta_s}\big)I^{11}I^{sv}\pi(\theta)\Big\} = 0. \tag{4.3}$$

In our context, when $\theta_1 = \frac{\sigma_1}{\sigma_2}$ is the parameter of interest, any prior of the form $\pi \propto \theta_1^{-1}(1-\theta_3^2)^{-1/2}g(\theta_2, \theta_3)$ ensures matching of the posterior and frequentist cumulative distribution functions at the second order if from (4.2) and (4.3)

$$
\begin{aligned}
& g\frac{\partial^2}{\partial\theta_1^2}\{\theta_1^{-1}\theta_1^2(1-\theta_3^2)^{1/2}\} - 2g\,\frac{\partial}{\partial\theta_1}\{\theta_1^2(1-\theta_3^2)^{1/2}\frac{\partial}{\partial\theta_1}\theta_1^{-1}\} \\
& -\frac{\partial}{\partial\theta_2}\{\theta_1^{-1}g\,\theta_1^2(1-\theta_3^2)^{1/2}\theta_2^2E(\frac{\partial^3\log f}{\partial\theta_1^2\partial\theta_2})\} \\
& -\frac{\partial}{\partial\theta_3}\{\theta_1^{-1}g\,\theta_1^2(1-\theta_3^2)^{1/2}(1-\theta_3^2)^2E(\frac{\partial^3\log f}{\partial\theta_1^2\partial\theta_3})\} \qquad (4.4) \\
& -g\,\frac{\partial}{\partial\theta_1}\{E(\frac{\partial^3\log f}{\partial\theta_1\partial{\theta_2}^2})\theta_1^2(1-\theta_3^2)^{1/2}\theta_2^2\theta_1^{-1}\} \\
& -g\,\frac{\partial}{\partial\theta_1}\{E(\frac{\partial^3\log f}{\partial\theta_1\partial{\theta_3}^2})\theta_1{}^2(1-\theta_3^2)^{1/2}(1-\theta_3^2)^2\theta_1^{-1}\} = 0,
\end{aligned}
$$

and

$$
g(1-\theta_3^2)^{-1/2}\,\frac{\partial}{\partial\theta_1}\{E(\frac{\partial^3\log f}{\partial\theta_1^3})\theta_1^{-1}\theta_1^4(1-\theta_3^2)^2\} = 0. \qquad (4.5)
$$

From (2.5) and (2.8) of Lemma 2.1, (4.4) reduces to

$$
-\theta_1^{-1}(1-\theta_3^2)^{-1/2}\frac{\partial}{\partial\theta_2}\{g\,\theta_2\} - \theta_1^{-1}\frac{\partial}{\partial\theta_3}\{g\,\theta_3(1-\theta_3^2)^{1/2}\} = 0 \qquad (4.6)
$$

while the left hand side of (4.5) reduces to $3g(1-\theta_3^2)^{-1/2}\ \frac{\partial}{\partial\theta_1}\{(1-\theta_3^2)\}$ which is clearly 0 for any g. So, we need to find g such that (4.6) is satisfied. In particular, (4.6) is satisfied if we let g once again to be the class of functions $g(\theta_2,\theta_3) = \theta_2^a|\theta_3|^a(1-\theta_3^2)^{-\frac{a+2}{2}}$. In other words, the same class of priors enjoys second order matching for both quantiles as well as distribution functions.

5 Highest Posterior Density Matching Priors

We now turn our attention to HPD matching priors for θ_1. In general, if $\tilde{\theta}$ is the parameter (real or vector-valued) of interest, then a HPD region is of the form $\{\tilde{\theta} : \pi(\tilde{\theta}|X) \geq k\}$, where $\pi(\tilde{\theta}|X)$ is the posterior of $\tilde{\theta}$ under the prior π and data X. We will consider priors which ensure that HPD regions with credibility level $1-\alpha$ also have asymptotically the same frequentist coverage probability, the error of approximation being $o(n^{-1})$. From (4.4.3) of Datta and Mukerjee (2004, p76) any second order matching prior for posterior quantiles of θ_1 is also HPD matching for θ_1 in the special case of models satisfying

$$
\frac{\partial}{\partial\theta_1}(I_{11}^{-3/2}E(\frac{\partial^3\log f}{\partial\theta_1^3})) = 0. \qquad (5.1)
$$

It is easy to check that when $\theta_1 = \frac{\sigma_1}{\sigma_2}$ is the parameter of interest from (2.5) and (2.6), (5.1) holds and hence the second order quantile matching prior $\pi(\theta_1,\theta_2,\theta_3) \propto \theta_1^{-1}\theta_2^a|\theta_3|^a(1-\theta_3^2)^{-\frac{a+3}{2}}$ is also HPD matching.

6 Matching Priors Via Inversion of Test Statistics

One traditional way to derive frequentist confidence intervals is inversion of certain test statistics. The most popular such test is the likelihood ratio test. But tests based on Rao's score statistic or the Wald statistic are also of importance, and are first order equivalent (i.e. upto $o(n^{-1/2})$) to the likelihood ratio tests.

We begin with the general case when θ_1 is the parameter of interest, while $\theta_2, \ldots, \theta_p$ are the nuisance parameters. Let $\theta = (\theta_1, \ldots, \theta_p)$, and let $l(\theta)$ denote the usual log-likelihood. The corresponding profile log-likelihood for θ_1 is given by $l^*(\theta_1) = l(\theta_1, \hat{\theta_2}(\theta_1), \ldots, \hat{\theta_p}(\theta_1))$, where $\hat{\theta_j}(\theta_1)$ is the MLE of θ_j given $\theta_1 (j = 2, \ldots, p)$. Then the likelihood ratio statistic for θ_1 is given by

$$M_{LR}{}^*(\theta_1, X) = 2[l(\hat{\theta}) - l^*(\theta_1)]. \tag{6.1}$$

Then from Yin and Ghosh (1997) (also from (5.2.18) of Datta and Mukerjee), a likelihood ratio matching prior π is obtained by solving

$$\begin{aligned} &\frac{\partial}{\partial\theta_2}\{\pi\ \theta_1^2(1-\theta_3^2)\theta_2^2 E(\frac{\partial^3 \log f}{\partial\theta_1^2\partial\theta_2})\} + \frac{\partial}{\partial\theta_3}\{\pi\ \theta_1^2(1-\theta_3^2)^3 E(\frac{\partial^3 \log f}{\partial\theta_1^2\partial\theta_3})\} \\ &+ \frac{\partial}{\partial\theta_1}\left\{\theta_1^2(1-\psi^2)\left\{\frac{\partial}{\partial\theta_1}\pi - \pi\big(\theta_1^2(1-\psi^2)E\left((\frac{\partial \log f}{\partial\theta_1})(\frac{\partial^2 \log f}{\partial\theta_1^2})\right)\right.\right. \\ &\left.\left. - \theta_2^2 E(\frac{\partial^3 \log f}{\partial\theta_1\partial\theta_2^2}) - (1-\theta_3^2)^2 E(\frac{\partial^3 \log f}{\partial\theta_1\partial\theta_3^2})\big)\right\}\right\} = 0 \end{aligned} \tag{6.2}$$

Then from (2.5),(2.7) and (2.8) of Lemma 2.1, (6.2) reduces to

$$\frac{\partial}{\partial\theta_2}\{\pi\ \theta_2\} + \frac{\partial}{\partial\theta_3}\{\pi\ \theta_3(1-\theta_3^2)\} + \frac{\partial}{\partial\theta_1}\left\{\theta_1^2(1-\theta_3^2)\{\frac{\partial}{\partial\theta_1}\pi + \pi\theta_1^{-1}\}\right\} = 0 \tag{6.3}$$

Consider once again $\pi(\theta_1, \theta_2, \theta_3) \propto \theta_1^{-1}\theta_2^a|\theta_3|^a(1-\theta_3^2)^{-\frac{a+3}{2}}$. Then $\frac{\partial}{\partial\theta_1}\pi + \pi\theta_1^{-1} = 0$, and the left hand side of (6.3) simplifies to

$$\theta_1^{-1}|\theta_3|^a(1-\theta_3^2)^{-\frac{a+3}{2}}\frac{\partial}{\partial\theta_2}\theta_2^{a+1} - \theta_1^{-1}\theta_2^a\frac{\partial}{\partial\theta_3}\{|\theta_3|^{a+1}(1-\theta_3^2)^{-\frac{a+3}{2}+1}\}$$

which when simplified is exactly the same as the left hand side of (3.5), and leads to the same class of matching priors as before. With this we conclude that we have been able to find a class of priors $\pi(\sigma_1, \sigma_2, \rho) \propto \sigma_1^a\sigma_2^a|\rho|^a(1-\rho^2)^{-1}$ which satisfies all the different matching criteria.

7 Propriety of Posteriors

We now find conditions on a which ensure propriety of posteriors for the general class of priors of the form $\pi(\sigma_1, \sigma_2, \rho) \propto (\sigma_1\sigma_2)^a|\rho|^a(1-\rho^2)^{-(a+3)/2}$. Without essential loss of generality, we continue assuming $\mu_1 = \mu_2 = 0$. Then writing $\boldsymbol{S} = \begin{pmatrix} \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} \\ \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}^2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, the joint posterior is given by

$$\pi(\sigma_1, \sigma_2, \rho|\boldsymbol{X}) \propto |\boldsymbol{\Sigma}|^{-n/2}|\rho|^a(1-\rho^2)^{-(a+3)/2}\text{exp}[-(1/2)\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})]. \tag{7.1}$$

Denoting by $0 < \lambda_1 < \lambda_2$ the eigenvalues of $\boldsymbol{S}$, one gets the inequality

$$\text{exp}[-(1/2\lambda_2)\text{tr}(\boldsymbol{\Sigma}^{-1}0] \leq \text{exp}[-(1/2)\text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S})] \leq \text{exp}[-(1/2\lambda_1)\text{tr}(\boldsymbol{\Sigma}^{-1})]. \tag{7.2}$$

In view of (7.1) and (7.2), noting that $\text{tr}(\boldsymbol{\Sigma}^{-1}) = (1-\rho^2)^{-1}(\sigma_1^{-2}+\sigma_2^{-2})$, the propriety of the posterior follows by showing integrability of $[\sigma_1\sigma_2(1-\rho^2)]^{-n}|\rho|^a(1-\rho^2)^{-(a+3)/2}\text{exp}[-(1/2\lambda_1)$ $(1-\rho^2)^{-1}(\sigma_1^{-2}+\sigma_2^{-2})]$ with respect to $\sigma_1(>0)$, $\sigma_2(>0)$ and $\rho \in (-1,1)$. With the transformation $\eta_1 = \sigma_1^{-2}(1-\rho^2)^{-1}$ and $\eta_2 = \sigma_2^{-2}(1-\rho^2)^{-1}$ and $\tau = -\rho$, the Jacobian of this transformation is given by $(1/4)\eta_1^{-3/2}\eta_2^{-3/2}(1-\tau^2)^{-1}$. Hence, the propriety of the posterior will follow from the integrability of

$$\eta_1^{(n-a-3)/2}\eta_2^{(n-a-3)/2}|\tau|^a(1-\tau^2)^{(n-3a-5)/2}\text{exp}[-(1/2\lambda_1)(\eta_1+\eta_2)]$$

with respect to $\eta_1(>0)$, $\eta_2(>0)$ and $\tau \in (-1,1)$. Now assuming $n \geq 4$, from the needed conditions for finiteness of beta and gamma integrals, the integrability follows when $-1 < a < (n-3)/3$. This is the needed condition on a for the propriety of posteriors under the given class of priors.

8 Simulation Study

Using the parameterization $\theta_1 = \sigma_1/\sigma_2$, $\theta_2 = \sigma_1\sigma_2(1-\rho^2)^{1/2}$ and $\theta_3 = \rho$, as before, our parameter of interest is θ_1. A general class of priors was obtained as $\pi \propto \theta_1^{-1}\theta_2^a|\theta_3|^a(1-\theta_3^2)^{(-\frac{a+3}{2})}$. This prior satisfies quantile matching, matching via distribution functions, HPD matching as well as likelihood ratio matching property.

There are three priors that we wish to compare. The first one (Prior 1) is $\pi \propto \theta_1^{-1}$. This was recommended by Staicu (2007) in her PhD dissertation showing that this prior achieves matching up to $O(n^{-3/2})$.

The second one (Prior 2) is $\pi \propto \theta_1^{-1}(1-\theta_3^2)^{-3/2}$. This was suggested by Mukerjee and Reid (2001). This is a special case (a=0) of the class of priors that we obtained satisfying all the matching criteria.

Finally, the third prior (Prior 3) $\pi \propto \theta_1^{-1}\theta_2^{-1}(1-\theta_3^2)^{-1}$ was recommended by Berger and Sun (2007). This is also one-at-a-time reference prior for each one of the parameters θ_1, θ_2 and θ_3 satisfying the first order matching property.

In order to evaluate the three different priors, we undertook a simulation study where data was generated from a bivariate normal distribution with $(\mu_1, \mu_2, \sigma_2, \rho) = (0, 0, 1, 0.5)$ and varying values of σ_1 and varying sample sizes n. The values of θ_1 varied from 0.5 to 2.0.

Since the full conditional distribution of the parameters under any of the three priors do not follow a standard distributional form, we used Gibbs sampling with componentwise Metropolis-Hastings updates at each iteration to generate random numbers from the conditional posterior distributions of each parameter (Robert and Casella, 2001). We ran two chains with different initial values and allowed a burn-in of 10000 each. A random-walk jumping density with normal noise added to the existing value in the chain for the means and log standard deviations was used. The correlation also had a random walk prior by adding a small normal noise to the old values. Each chain was run 40,000 times and convergence was judged by a Gelman-Rubin (Gelman and Rubin, 1992) diagnostic. The trace plot presenting the time history of the last 8000 iterations for all five parameters is presented for a sample simulated dataset with $\theta_1 = 0.7$, under Prior 3 and sample size 20, in Figure 1. Figure 2 presents the plot of Gelman-Rubin diagnostic for the θ_1 chain under the same setting with diagnostic values close to 1 suggesting convergence. Figures 3, 4 and 5 are posterior distributions for θ_1 under three different priors for four different sample sizes $n = 10, 20, 30, 40$. One can immediately make the following observations. Though there are certain numerical differences, the posterior distribution of θ_1 does not seem to vary widely between Priors 1,2 and 3, even for smaller sample sizes, though Prior 2 typically gave smaller posterior standard deviations. As data information increases with sample size, the posterior distributions become very similar under the three priors. Some skewness can be observed in the posterior distributions for smaller sample sizes, which was often noted during our simulation, but the distribution becomes fairly symmetric as n becomes large. The posterior distribution also becomes more concentrated around the true value of θ_1 with increasing n, as expected.

We repeated our Gibbs sampling estimation technique for 500 datasets under each configuration of θ_1 and n. Each time, we computed the posterior mean, the 95% quantile interval (as given by the 2.5th and 97.5th sample percentile of the randomly generated parameter values after the burn-in period) and the 95% HPD interval. Table 1 presents the average of posterior means, the mean squared error, the frequentist coverage of the Bayesian credible intervals (as estimated by the proportion of times the true parameter value falls in the corresponding credible intervals) across the 500 datasets and under three different priors. Some interesting differences can be noted in the behavior for smaller sample sizes. Prior 2 appears to be performing best in terms of both coverage of quantile and HPD intervals and also has excellent point estimation properties in terms of average posterior mean and MSE for smaller sample sizes. For larger sample sizes all three priors become almost indistinguishable in terms of their performances.

9 Discussion

The paper develops a general class of priors for the ratio of variances in a bivariate normal population which matches asymptotically the coverage probabilities of Bayesian credible intervals with the corresponding frequentist confidence intervals. Several matching criteria are used, and a class of priors meeting all these criteria is found. A simulation study shows that these priors perform quite well even for small and moderate samples. It is true that in the era of Markov chain Monte Carlo based numerical estimation one could essentially use any prior which reflects the investigator's a'priori belief regarding the parameters. However, matching priors not only provide insight into the connection of Bayes-frequentist inference, but may be useful in complex frequentist settings, when Bayesian computation techniques could be seen as a way to generate frquentist confidence intervals. The Bayesian procedure with a suitable matching prior could be thought of as an alternative algorithm to yield correct frequentist confidence intervals, when one wants to avoid bootstrap or rely on asymptotic expansions even with a small sample size.

References

[1] Berger, J. and Sun, D. (2007). Objective priors for the bivariate normal model. To appear in the *Annals of Statistics*.

[2] Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with Discussion). *J. R. Statist. Soc. B***53**,79-109.

[3] Datta, G. S. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics.* Lecture notes in Statistics. Springer, New York.

[4] Datta, G. S. and Sweeting, T. J. (2005). *Probability matching priors.* Handbook of Statistics, Vol 25: Bayesian Thinking: Modeling and Computation, Eds.: D. Dey and C.R. Rao, pp. 91–114. Elsevier.

[5] Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**, 457-72.

[6] Huzurbazar, V. S. (1950). Probability distributions and orthogonal parameters. *Proc. Camb. Phil. Soc* **46**, 281-284.

[7] Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Am. Statist. Assoc.* **91**, 1343-70.

[8] Morgan, W.A. (1939). A Test for the Significance of the Difference Between the Two variances in a Sample from a Normal Bivariate Population. *Biometrika* **31**, 13-19.

[9] Mukerjee, R. and Reid, N. (2001). Second-order probability Matching priors for a parametric function with application to Bayesian tolerance limits. *Biometrika* **88**, 587-92.

[10] Peers, H. W. (1965). Confidence properties of Bayesian interval estimates. *J. R. Statist. Soc. B* **30**, 535-44.

[11] Pitman, E.J.G. (1939). A Note on Normal Correlation. *Biometrika* **31**, 9-12.

[12] Roy, S.N. and Potthoff, R.F. (1958). Confidence Bounds on Vector Analogues of the "Ratio of Means" and the "Ratio of Variances" for Two Correlated Normal Variates and Some Associated Tests. *The Annals of Mathematical Statistics* **29**, 829-41.

[13] Robert,C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.

[14] Stein, C, (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, Banach center publications, **16**. Warsaw: PWN-Polish scientific publishers.

[15] Tibshirani, R. J. (1989). Noninformative priors for one parameter of many. *Biometrika* **76**, 604-8.

[16] Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. R. Statist. Soc. B* **35**, 318-29.

[17] Yin, M. and Ghosh, M. (1997). A note on the probability difference between matching priors based on posterior quantiles and on inversion of conditional likelihood ratio statistics. *Calcutta Statist. Assoc. Bull.* **47**, 59-65.

Table 1: Simulation result to compare the three different priors suggested for bivariate normal ratio of standard deviation parameter θ_1. The true parameter settings are $\mu_1 = \mu_2 = 0, \sigma_2 = 1$ and varying values of $\theta_1 = \sigma_1$ as listed. Prior 1: $\pi \propto \theta_1^{-1}$, Prior 2: $\pi \propto (1-\rho^2)^{-3/2}\theta_1^{-1}$, Prior 3: $\pi \propto (1-\psi^2)^{-1}\theta_1^{-1}\theta_2^{-1}$. Results are based on 500 simulated data sets. The target coverage probability is 95%.

θ_1	n	Prior 1				Prior 2				Prior 3			
		$\hat{\theta}_1^*$	MSE	Coverage (Quantile)	Coverage (HPD)	$\hat{\theta}_1$	MSE	Coverage (Quantile)	Coverage (HPD)	$\hat{\theta}_1$	MSE	Coverage (Quantile)	Coverage (HPD)
	10	0.52	0.02	0.93	0.94	0.51	0.02	0.95	0.95	0.52	0.02	0.93	0.93
0.5	20	0.51	0.01	0.94	0.94	0.50	0.01	0.95	0.94	0.51	0.01	0.94	0.94
	30	0.50	0.01	0.94	0.94	0.50	0.01	0.95	0.96	0.5	0.01	0.94	0.94
	40	0.51	0.01	0.96	0.95	0.51	0.01	0.96	0.95	0.51	0.01	0.95	0.95
	10	1.06	0.14	0.93	0.94	1.06	0.12	0.95	0.95	1.07	0.14	0.93	0.93
1.0	20	1.01	0.04	0.96	0.94	1.01	0.03	0.95	0.95	1.01	0.04	0.95	0.94
	30	1.01	0.03	0.96	0.95	1.01	0.02	0.96	0.95	1.01	0.03	0.95	0.95
	40	1.01	0.02	0.95	0.94	1.01	0.02	0.95	0.95	1.01	0.02	0.95	0.94
	10	1.61	0.25	0.92	0.94	1.61	0.23	0.95	0.95	1.61	0.26	0.93	0.94
1.5	20	1.54	0.12	0.94	0.92	1.54	0.11	0.93	0.93	1.54	0.12	0.93	0.93
	30	1.51	0.06	0.94	0.94	1.51	0.05	0.93	0.92	1.51	0.06	0.93	0.93
	40	1.52	0.05	0.94	0.94	1.52	0.05	0.95	0.95	1.52	0.05	0.95	0.95
	10	2.11	0.44	0.94	0.94	2.11	0.42	0.95	0.94	2.11	0.42	0.96	0.94
2.0	20	2.04	0.18	0.95	0.95	2.04	0.18	0.95	0.95	2.04	0.18	0.95	0.95
	30	2.04	0.11	0.95	0.96	2.04	0.10	0.95	0.95	2.04	0.11	0.94	0.95
	40	2.03	0.08	0.95	0.95	2.03	0.07	0.95	0.95	2.03	0.08	0.95	0.94

*: Average value for posterior mean of θ_1, averaged across 500 simulated datasets.

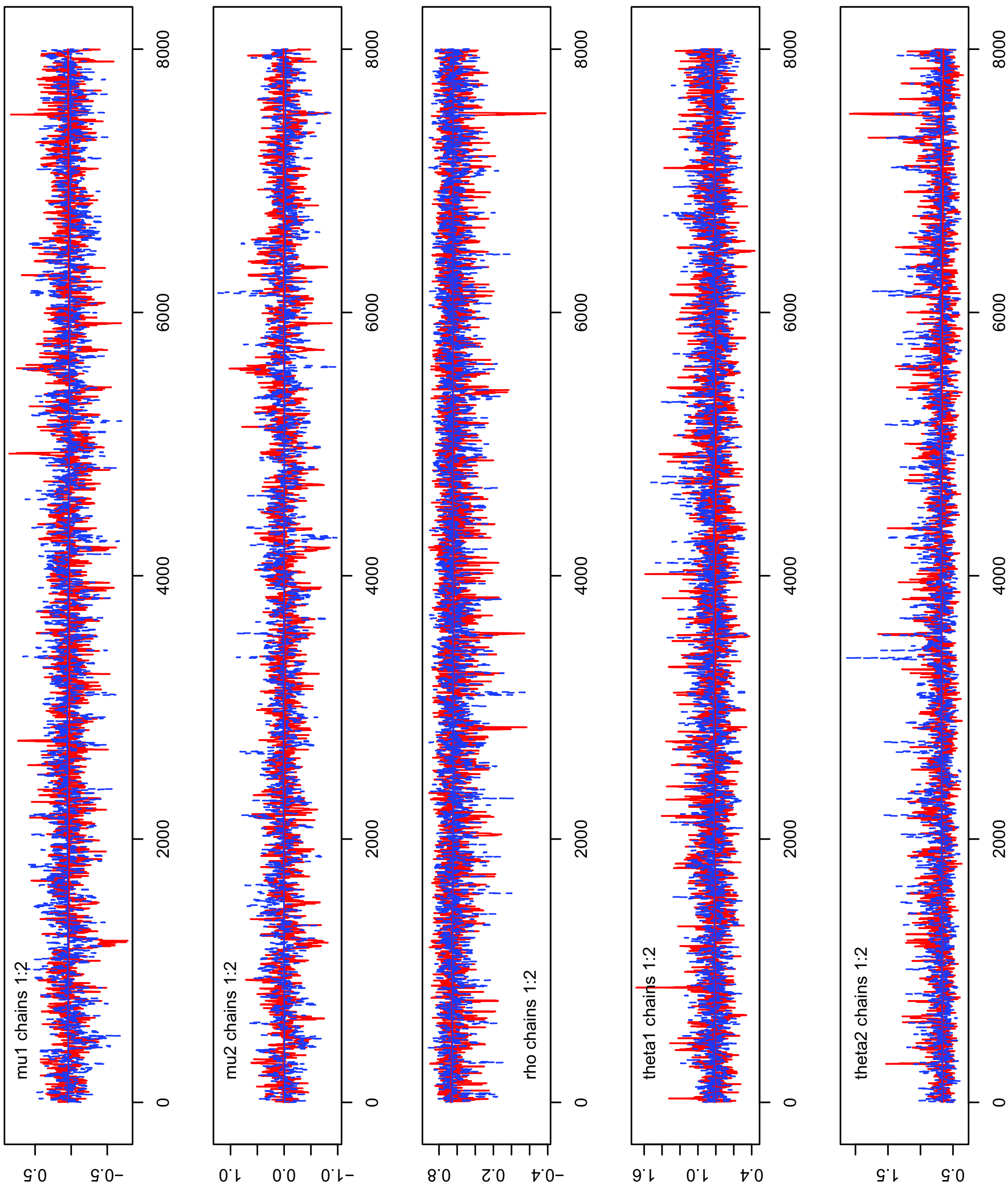

Figure 1: Sample trace plot for all the parameters under Prior 3 for $n = 20$ under the simulation setting of Section 4. True value of $\theta_1 = 0.7$.

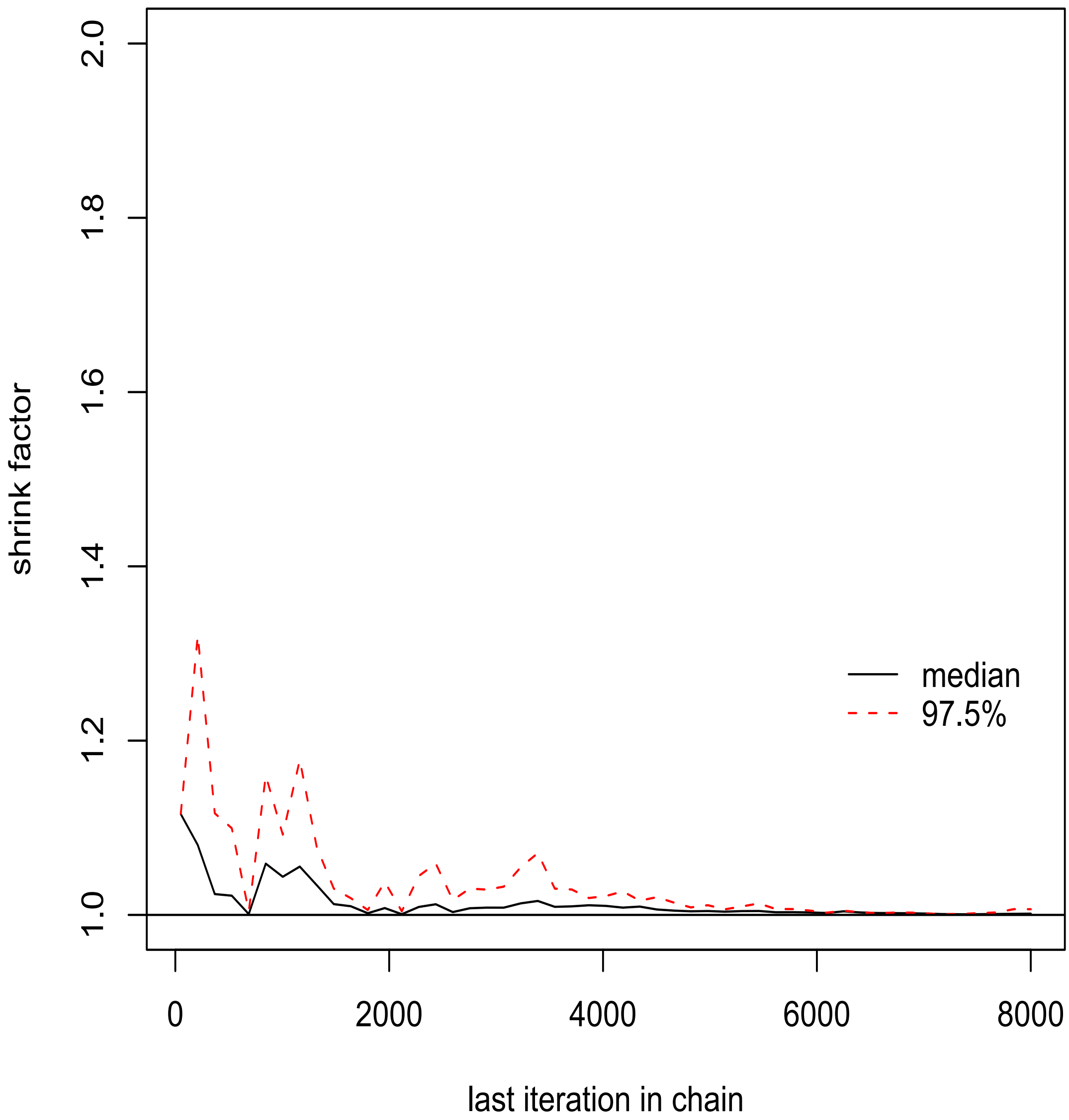

Figure 2: Plot of Gelman-Rubin diagnostic statistics for θ_1 under Prior 3 for $n = 20$ under the simulation seeting of Section 4. True value of $\theta_1 = 0.7$.

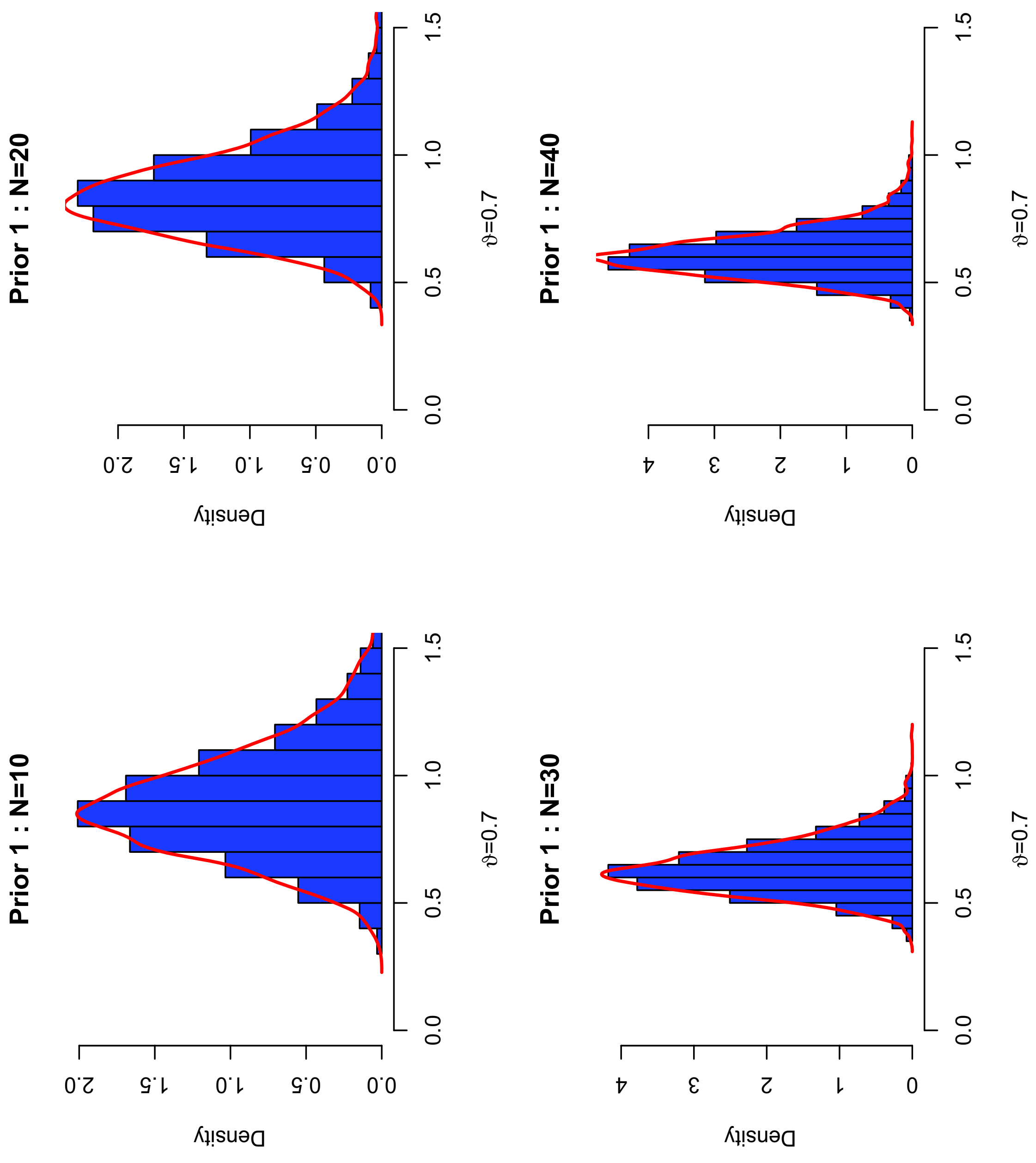

Figure 3: Posterior distribution for θ_1 under Prior 1 for different sample sizes, under the simulation setting of Section 4. True value of $\theta_1 = 0.7$.

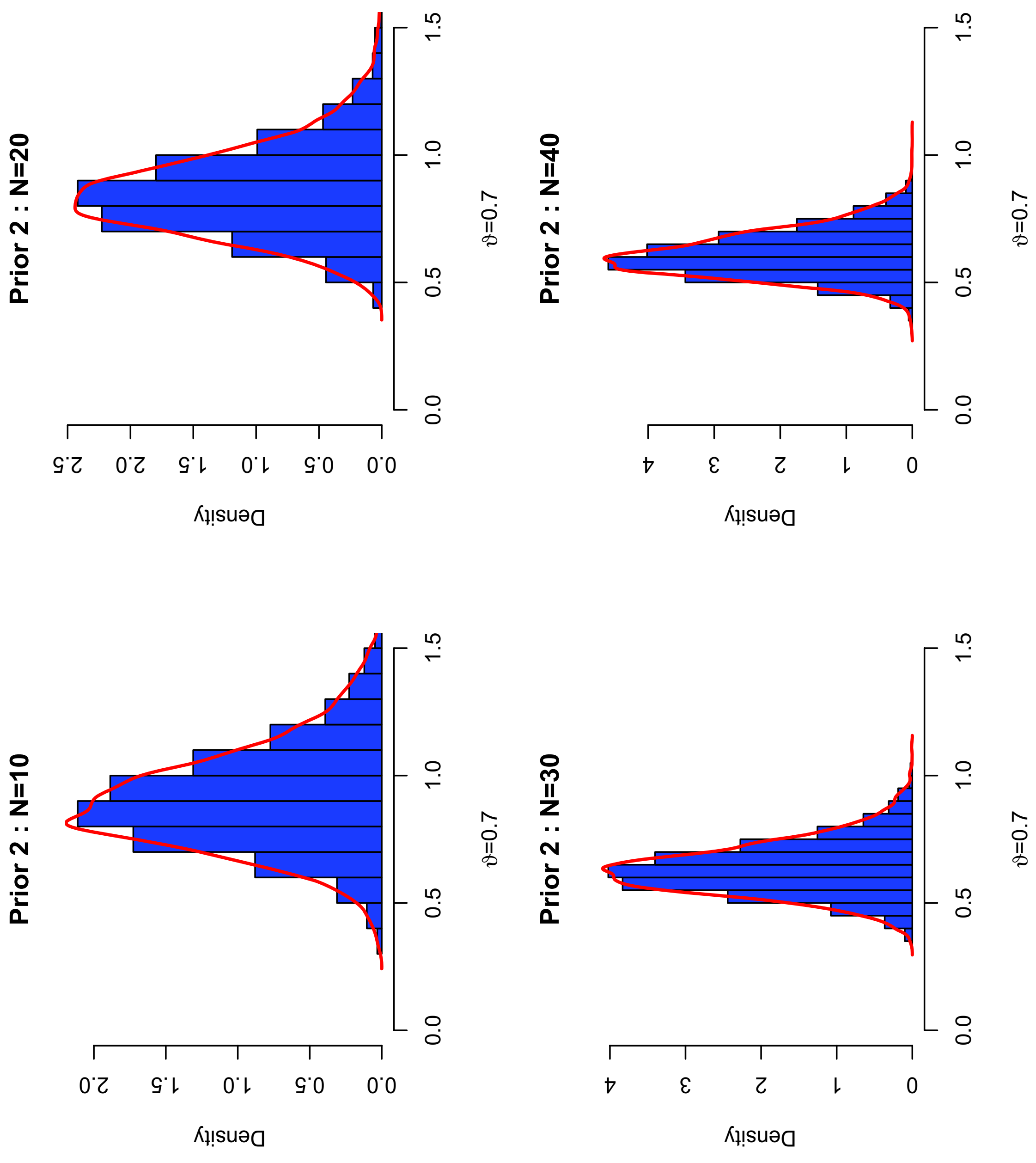

Figure 4: Posterior distribution for θ_1 under Prior 2 for different sample sizes, under the simulation setting of Section 4. True value of $\theta_1 = 0.7$.

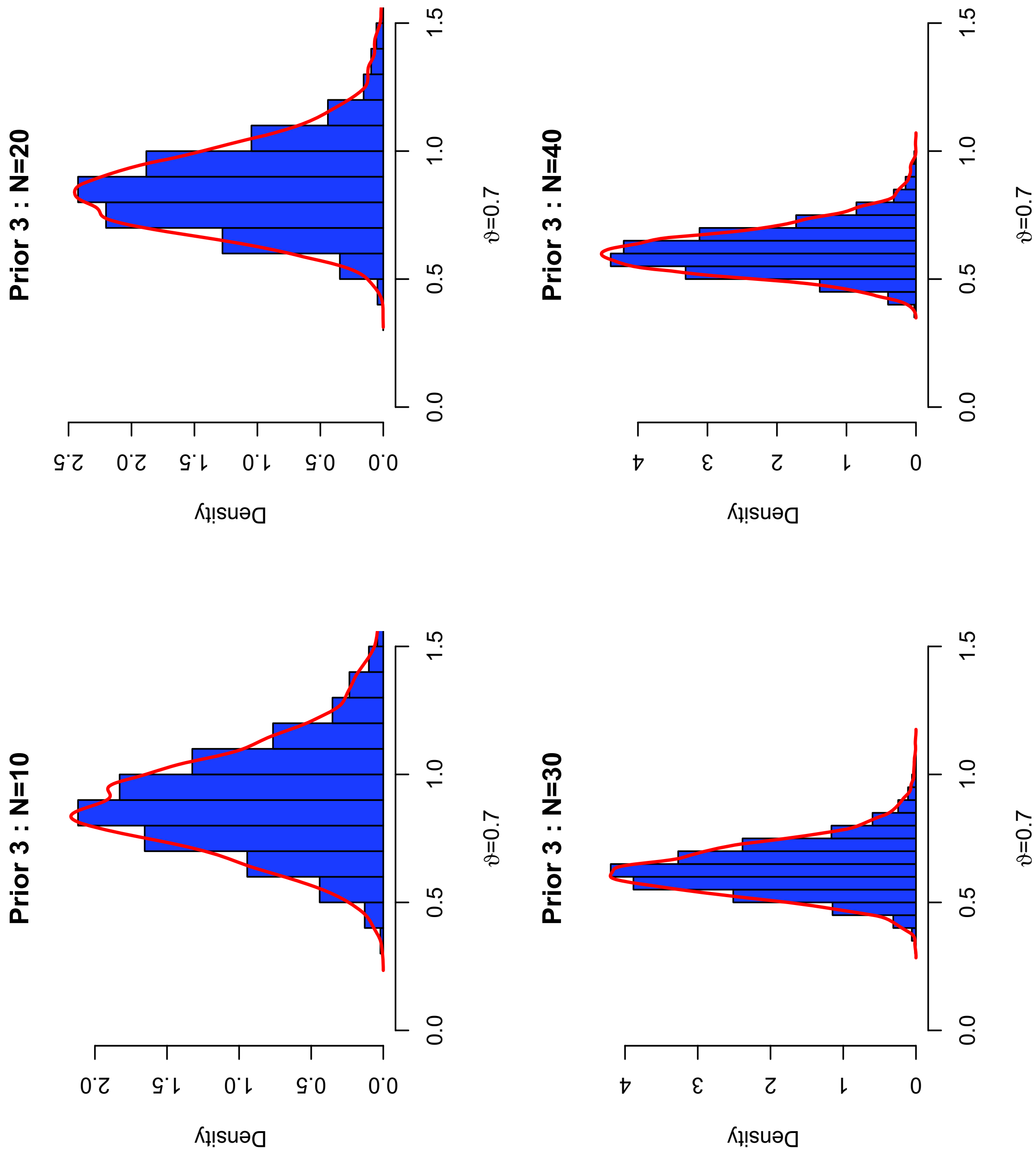

Figure 5: Sample posterior distribution for θ_1 under Prior 3 for different sample sizes, under the simulation setting of Section 4. True value of $\theta_1 = 0.7$.