

Objective Priors for the Multivariate Normal Model

DONGCHU SUN

Virginia Tech and University of Missouri-Columbia, USA

`sund@vt.edu` and `sund@missouri.edu`

JAMES O. BERGER

Duke University, Durham, NC 27708, USA

`berger@stat.duke.edu`

SUMMARY

Objective Bayesian inference for the multivariate normal distribution is illustrated, using different types of formal objective priors (Jeffreys, invariant, reference and matching), different modes of inference (Bayesian and frequentist), and different criteria involved in selecting optimal objective priors (ease of computation, frequentist performance, marginalization paradoxes, and decision-theoretic evaluation).

In the course of the investigation of the bivariate normal model in Berger and Sun (2006), a variety of surprising results were found, including the availability of objective priors that yield exact frequentist inferences for many functions of the bivariate normal parameters, such as the correlation coefficient. Certain of these results are generalized to the multivariate normal situation.

The prior that most frequently yields exact frequentist inference is the right-Haar prior, which unfortunately is not unique. Two natural proposals are studied for dealing with this non-uniqueness: first, mixing over the right-Haar priors; second, choosing the ‘empirical Bayes’ right-Haar prior, that which maximizes the marginal likelihood of the data. Quite surprisingly, we show that neither of these possibilities yields a good solution. This is disturbing and sobering. It is yet another indication that improper priors do not behave as do proper priors, and that it can be dangerous to apply ‘understandings’ from the world of proper priors to the world of improper priors.

Keywords and Phrases: KULLBACK-LEIBLER DIVERGENCE; JEFFREYS PRIOR; MULTIVARIATE NORMAL DISTRIBUTION; MATCHING PRIORS; REFERENCE PRIORS; INVARIANT PRIORS.

1. INTRODUCTION

Estimating the mean and covariance matrix of a multivariate normal distribution became of central theoretical interest when Stein (1956) and ? showed that standard estimators had significant problems, including inadmissibility from a frequentist perspective. Most problematical were standard estimators of the covariance matrix; see Yang & Berger (1994) and the references therein.

In the Bayesian literature, the most commonly used prior for a multivariate normal distribution is a normal prior for the normal mean and an inverse Wishart prior for the covariance matrix. Such priors are conjugate, leading to easy computation, but lack flexibility and also lead to inferences of the same structure as those shown to be inferior by Stein. More flexible and better performing priors for a covariance matrix were developed by Leonard and Hsu (1992), and Brown (1994) (the generalized inverse Wishart prior). In the more recent Bayesian literature, aggressive shrinkage of eigenvalues, correlations, or other features of the covariance matrix are entertained; see, for example, Daniels & Kass (1999), Daniels & Pourahmadi (2002), Liechty et al. (2004) and the references therein. These priors may well be successful in practice, but they do not seem to be formal objective priors according to any of the common definitions.

Recently, Berger & Sun (2006) considered objective inference for parameters of the bivariate normal distribution and functions of these parameters, with special focus on development of objective confidence or credible sets. In the course of the study, many interesting issues were explored involving objective Bayesian inference, including different types of objective priors (Jeffreys, invariant, reference and matching), different modes of inference (Bayesian and frequentist), and different criteria involved in deciding on optimal objective priors (ease of computation, frequentist performance and marginalization paradoxes).

In this paper, we first generalize some of the bivariate results to the multivariate normal distribution; section 2 presents the generalizations of the various objective priors discussed in Berger & Sun (2006). We particularly focus on reference priors, and show that the right-Haar prior is indeed a one-at-a-time reference prior (Berger & Bernardo (1992a)) for many parameters and functions of parameters.

Section 3 gives some basic properties of the resulting posterior distributions and gives *constructive posterior distributions* for many of the priors. Constructive posteriors are expressions for the posterior distribution which allow very simply simulation from the posterior. Constructive posteriors are also very powerful for proving results about exact frequentist matching. (Exact frequentist matching means that $100(1 - \alpha)\%$ credible sets arising from the resulting posterior are also exact frequentist confidence sets at the specified level.) Results about matching for the right-Haar prior are given in Section 4 for a variety of parameters.

One of the most interesting features of right-Haar priors is that, while they result in exact frequentist matching, they also seem to yield marginalization paradoxes (Dawid et al. (1973)). Thus one is in the philosophical conundrum of having to choose between frequentist matching and avoidance of the marginalization paradox. This is also discussed in Section 4.

Another interesting feature of the right-Haar priors is that they are not unique; they depend on which triangular decomposition of a covariance matrix is employed. In Section 5, two natural proposals are studied to deal with this non-uniqueness. The first is to simply mix over the right-Haar priors. The second is to choose the ‘empirical Bayes’ right-Haar prior, namely that which maximizes the marginal likelihood of the data. Quite surprisingly, it is shown that both of these solutions

gives inferior answers, a disturbing and sobering phenomenon. It is yet another indication that improper priors do not behave as do proper priors, and that it can be dangerous to apply ‘understandings’ from the world of proper priors to the world of improper priors.

2. OBJECTIVE PRIORS FOR THE MULTIVARIATE NORMAL DISTRIBUTION

Consider the p -dimensional multivariate normal population, $\mathbf{x} = (x_1, \dots, x_p)' \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, whose density is given by

$$f(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (1)$$

2.1. Previously Considered Objective Priors

Perhaps the most popular prior for the multivariate normal distribution is the Jeffreys (rule) prior (Jeffreys (1961))

$$\pi_J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-(p+2)/2}. \quad (2)$$

Another commonly used prior is the independence-Jeffreys prior

$$\pi_{IJ}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-(p+1)/2}. \quad (3)$$

It is commonly thought that either the Jeffreys or independence-Jeffreys priors are most natural, and most likely to yield classical inferences. However, Geisser & Cornfield (1963) showed that the prior which is exact frequentist matching for all means and variances (and which also yields Fisher’s fiducial distribution for these parameters) is

$$\pi_{GC}(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-p}. \quad (4)$$

It is simple chance that this prior happens to be the Jeffreys prior for $p = 2$ (and perhaps simple chance that it agrees with π_{IJ} for $p = 1$); these coincidences may have contributed significantly to the popular notion that Jeffreys priors are generally successful.

In spite of the frequentist matching success of π_{GC} for means and variances, the prior seems to be quite bad for correlations, predictions, or other inferences involving a multivariate normal distribution. Thus a variety of other objective priors have been proposed in the literature.

Chang & Eaves (1990) ((7) on page 1605) derived the reference prior for the parameter ordering $(\mu_1, \dots, \mu_p; \sigma_1, \dots, \sigma_p; \boldsymbol{\Upsilon})$,

$$\pi_{CE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\boldsymbol{\mu} d\boldsymbol{\Sigma} = \frac{1}{|\boldsymbol{\Sigma}|^{(p+1)/2} |\mathbf{I}_p + \boldsymbol{\Sigma}^* \boldsymbol{\Sigma}^{-1}|^{1/2}} d\boldsymbol{\mu} d\boldsymbol{\Sigma} \quad (5)$$

$$= 2^p \left[\prod_{i=1}^p \frac{d\mu_i d\sigma_i}{\sigma_i} \right] \left[\frac{1}{|\boldsymbol{\Upsilon}|^{(p+1)/2} |\mathbf{I}_p + \boldsymbol{\Upsilon}^* \boldsymbol{\Upsilon}^{-1}|^{1/2}} \prod_{i < j} d\rho_{ij} \right], \quad (6)$$

where $\boldsymbol{\Upsilon}$ is the correlation matrix and $\mathbf{A}^* \mathbf{B}$ denotes the Hadamard product of the squared matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, whose entries are $c_{ij} = a_{ij} b_{ij}$. In the

bivariate normal case, this prior is the same as Lindley's (1965) prior, derived using certain notions of transformation to constant information, and was derived as a reference prior for the correlation coefficient ρ in Bayarri (1981).

Chang and Eaves (1990) also derived the reference prior for the ordering $(\mu_1, \dots, \mu_p; \lambda_1, \dots, \lambda_p; \mathbf{O})$, where $\lambda_1 > \dots > \lambda_p$ are the ordered eigenvalues of $\mathbf{\Sigma}$, and \mathbf{O} is an orthogonal matrix such that $\mathbf{\Sigma} = \mathbf{O}' \text{diag}(\lambda_1, \dots, \lambda_p) \mathbf{O}$. This reference prior was discussed in detail in Yang & Berger (1994) and has the form,

$$\pi_E(\mu, \mathbf{\Sigma}) d\mu d\mathbf{\Sigma} = \frac{I_{[\lambda_1 > \dots > \lambda_p]}}{|\mathbf{\Sigma}| \prod_{i < j} (\lambda_i - \lambda_j)} d\mu d\mathbf{\Sigma}. \quad (7)$$

Another popular prior is the right Haar prior, which has been extensively studied (see, e.g., Eaton & Sudderth (2002)). It is most convenient to express this prior in terms of a lower-triangular matrix $\mathbf{\Psi}$ with positive diagonal elements and such that

$$\mathbf{\Sigma}^{-1} = \mathbf{\Psi}' \mathbf{\Psi}. \quad (8)$$

(Note that there are many such matrices, so that the right Haar prior is not unique.) The right Haar prior corresponding to this decomposition is given by

$$\pi_H(\mu, \mathbf{\Psi}) d\mu d\mathbf{\Psi} = \prod_{i=1}^p \frac{1}{\psi_{ii}^i} d\mu d\mathbf{\Psi}. \quad (9)$$

We will see in the next subsection that, this prior is one-at-a-time reference prior for various parameterizations.

Because $d\mathbf{\Sigma} = \prod_{i=1}^p \psi_{ii}^{i-2(p+1)} d\mathbf{\Psi}$, the independence Jeffreys prior $\pi_{IJ}(\mu, \mathbf{\Psi})$ corresponding to the left-haar measure is given by

$$\pi_{IJ}(\mu, \mathbf{\Psi}) = \prod_{i=1}^p \frac{1}{\psi_{ii}^{p-i+1}} d\mu d\mathbf{\Psi}. \quad (10)$$

The right-Haar prior and the independence Jeffreys prior are limiting cases of generalized Wishart priors; see Brown (2001) for a review.

We now give the Fisher information matrix and a result about the reference prior for a group ordering related to the right-Haar parameterization; proofs are relegated to the appendix.

Fact 1 (a) The Fisher information matrix for $\{\mu, \psi_{11}, (\psi_{21}, \psi_{22}), \dots, (\psi_{p1}, \dots, \psi_{pp})\}$ is

$$\mathbf{J} = -E \left(\frac{\partial^2 \log f}{\partial \theta \partial \theta'} \right) = \text{diag}(\mathbf{\Sigma}^{-1}, \mathbf{\Lambda}_1, \dots, \mathbf{\Lambda}_p), \quad (11)$$

where, for $i = 1, \dots, p$,

$$\mathbf{\Lambda}_i = \mathbf{\Sigma}_i + \frac{1}{\psi_{ii}^2} \mathbf{e}_i \mathbf{e}_i',$$

with \mathbf{e}_i being the i^{th} unit column vector.

(b) The reference prior of $\mathbf{\Psi}$ for the ordered group $\{\mu_1, \dots, \mu_p, \psi_{11}, (\psi_{21}, \psi_{22}), \dots, (\psi_{p1}, \dots, \psi_{pp})\}$ is given by

$$\pi_{R1}(\mu, \mathbf{\Psi}) \propto \frac{1}{\prod_{i=1}^p \psi_{ii}}. \quad (12)$$

Note that the right-Haar prior, the Jeffreys (rule) prior, the independence Jeffreys prior, the Geisser-Cornfield prior and the reference prior π_{R1} have the form

$$\pi_{\mathbf{a}}(\boldsymbol{\mu}, \boldsymbol{\Psi}) d\boldsymbol{\mu} d\boldsymbol{\Psi} = \prod_{i=1}^p \frac{1}{\psi_{ii}^{a_i}} d\boldsymbol{\mu} d\boldsymbol{\Psi}, \quad (13)$$

where $\mathbf{a} = (a_1, \dots, a_p)$. This class of priors has also received considerable attention in directed acyclic graphical models (cf. Roverato & Consonni (2004)).

Table 1: Summary of objective priors of $(\boldsymbol{\mu}, \boldsymbol{\Psi})$ as special cases of (13)

prior	form	(a_1, \dots, a_p)
π_H	$\prod_{i=1}^p \psi_{ii}^{-i}$	$a_i = i$
π_J	$\prod_{i=1}^p \psi_{ii}^{-(p-i)}$	$a_i = p - i$
π_{IJ}	$\prod_{i=1}^p \psi_{ii}^{-(p-i+1)}$	$a_i = p - i + 1$
π_{GC}	$\prod_{i=1}^p \psi_{ii}^{-(2-i)}$	$a_i = 2 - i$
π_{R1}	$\prod_{i=1}^p \psi_{ii}^{-1}$	$a_i = 1$

2.2. Reference Priors under Alternative Parameterizations

Pourahmadi (1999) considered another decomposition of $\boldsymbol{\Sigma}^{-1}$. Let $\mathbf{T} = (t_{ij})_{p \times p}$ be the $p \times p$ unit lower triangular matrix, where

$$t_{ij} = \begin{cases} 0 & \text{if } i < j, \\ 1 & \text{if } i = j, \\ \frac{\psi_{ij}}{\psi_{ii}} & \text{if } i > j. \end{cases} \quad (14)$$

Pourahmadi (1999) pointed out the statistical interpretations of the below-diagonal entries of \mathbf{T} and the diagonal entries of $\boldsymbol{\Psi}$. In fact, $x_1 \sim N(\mu_1, d_1)$, $x_i \sim N(\mu_i - \sum_{j=1}^{i-1} t_{ij}(x_j - \mu_j), \psi_{ii}^{-2})$, ($j \geq 2$), so the t_{ij} are the negatives of the coefficients of the best linear predictor of x_i based on (x_1, \dots, x_{i-1}) , and ψ_{ii}^2 is the precision of the predictive distribution. Write $\tilde{\boldsymbol{\Psi}} = \text{diag}(\psi_{11}, \dots, \psi_{pp})$. Clearly

$$\boldsymbol{\Psi} = \tilde{\boldsymbol{\Psi}} \mathbf{T}, \quad (15)$$

$$\boldsymbol{\Sigma} = (\mathbf{T}' \tilde{\boldsymbol{\Psi}}^2 \mathbf{T})^{-1}. \quad (16)$$

For $i = 2, \dots, p$, define $\tilde{\boldsymbol{\Psi}}_i = \text{diag}(\psi_{11}, \dots, \psi_{ii})$ and denote the upper and left $i \times i$ submatrix of \mathbf{T} by \mathbf{T}_i . Then

$$\boldsymbol{\Psi}_i = \tilde{\boldsymbol{\Psi}}_i \mathbf{T}_i, \quad (17)$$

$$\boldsymbol{\Sigma}_i = (\mathbf{T}_i' \tilde{\boldsymbol{\Psi}}_i^2 \mathbf{T}_i)^{-1}, \quad i = 2, \dots, p. \quad (18)$$

Fact 2 (a) The Fisher information for $(\boldsymbol{\mu}, \psi_{11}, (t_{21}, \psi_{22}), (t_{31}, t_{32}, \psi_{33}), \dots, (t_{p1}, \dots, t_{p,p-1}, \psi_{ii}))$ is of the form

$$\tilde{\mathbf{J}} = \text{diag}(\mathbf{T}' \tilde{\boldsymbol{\Psi}}^2 \mathbf{T}, \frac{2}{\psi_{11}^2}, \tilde{\mathbf{J}}_2, \dots, \tilde{\mathbf{J}}_p), \quad (19)$$

where for $i = 2, \dots, p$,

$$\tilde{\mathbf{J}}_i = \text{diag}\left(\psi_{ii}^2 \mathbf{T}_{i-1}^{-1} \tilde{\boldsymbol{\Psi}}_{i-1}^{-2} \mathbf{T}_{i-1}'^{-1}, \frac{2}{\psi_{ii}^2}\right). \quad (20)$$

(b) The one-at-a-time reference prior for $\{\mu_1, \dots, \mu_p, \psi_{11}, \psi_{22}, \dots, \psi_{ii}, t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1}\}$, and with any ordering of parameters, is

$$\tilde{\pi}_R(\boldsymbol{\theta}) = \prod_{i=1}^p \frac{1}{\psi_{ii}}. \quad (21)$$

(c) The reference prior in (b) is the same as the right-Haar measure for $\boldsymbol{\Psi}$, given in (9).

Consider the parameterization $\mathbf{D} = \text{diag}(d_1, \dots, d_p)$ and \mathbf{T} , where $d_i = 1/\psi_{ii}^2$. Clearly $\mathbf{D} = \tilde{\boldsymbol{\Psi}}^{-2}$ and $\boldsymbol{\Sigma}^{-1} = \mathbf{T}' \mathbf{D}^{-1} \mathbf{T}$. Also write $\mathbf{D}_i = \boldsymbol{\Psi}_i^{*-2}$.

Fact 3 (a) The Fisher information for $(\boldsymbol{\mu}, d_1, \dots, d_p; t_{21}; t_{31}, t_{32}; \dots, t_{p1}, \dots, t_{p,p-1})$ is of the form

$$\mathbf{J}^\# = \text{diag}(\mathbf{T}' \mathbf{D}^{-1} \mathbf{T}, \frac{1}{d_1^2}, \dots, \frac{1}{d_p^2}, \boldsymbol{\Delta}_2, \dots, \boldsymbol{\Delta}_p), \quad (22)$$

where, for $i = 2, \dots, p$,

$$\boldsymbol{\Delta}_i = \frac{1}{d_i} \mathbf{T}_{i-1}^{-1} \mathbf{D}_{i-1} \mathbf{T}_{i-1}'^{-1}. \quad (23)$$

(b) The one-at-a-time reference prior for $\{\mu_1, \dots, \mu_p, d_1, \dots, d_p, t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1}\}$, and with any ordering, is

$$\tilde{\pi}_R(\boldsymbol{\theta}) \propto \prod_{i=1}^p \frac{1}{d_i}. \quad (24)$$

(c) The reference prior in (b) is the same as the right-Haar measure for $\boldsymbol{\Psi}$, given in (9).

Suppose one is interested in the generalized variance $|\boldsymbol{\Sigma}| = \prod_{i=1}^p d_i$; the one-at-a-time reference prior is also the right-Haar measure π_H . To see this, define

$$\left\{ \begin{array}{lcl} \xi_1 & = & \frac{d_1}{d_2}, \\ \xi_2 & = & \frac{(d_1 d_2)^{1/2}}{d_3}, \\ \dots & & \dots \\ \xi_{p-1} & = & \frac{(\prod_{j=1}^{p-1} d_j)^{1/(p-1)}}{d_p}, \\ \xi_p & = & \prod_{j=1}^p d_j. \end{array} \right. \quad (25)$$

Fact 4 (a) The Fisher information matrix for $(\boldsymbol{\mu}, \xi_1, \dots, \xi_p; t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1})$ is

$$\text{diag}\left(\boldsymbol{\Sigma}^{-1}, \frac{1}{2\xi_1^2}, \frac{2}{3\xi_2^2}, \dots, \frac{p-1}{p\xi_{p-1}^2}, \frac{1}{p\xi_p^2}, \boldsymbol{\Delta}_2, \dots, \boldsymbol{\Delta}_p\right), \quad (26)$$

where $\boldsymbol{\Delta}_i$ is given by (23).

(b) The one-at-a-time reference prior of any ordering for $\{\mu_1, \dots, \mu_p, \xi_1, \dots, \xi_p; t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1}\}$ is

$$\tilde{\pi}_R(\boldsymbol{\theta}) \propto \prod_{i=1}^p \frac{1}{\xi_i}. \quad (27)$$

(c) The reference prior in (b) is π_H , given in (9).

Corollary 0.1 Since $\xi_p = |\boldsymbol{\Sigma}|$, it is immediate that the one-at-a-time reference prior for ξ_p , with nuisance parameters $(\boldsymbol{\mu}, \xi_1, \dots, \xi_{p-1}, t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1})$, is the right-Haar prior π_H .

corollary

Corollary 0.2 One might be interested in $\eta_i \equiv |\boldsymbol{\Sigma}_i| = \prod_{j=1}^i d_i$, the generalized variance of the upper left $i \times i$ submatrix of $\boldsymbol{\Sigma}$. Using the same arguments as in Fact 4, the Fisher information for $(\boldsymbol{\mu}, \xi_1, \dots, \xi_{i-1}, \eta_i, d_{i+1}, \dots, d_p; t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1})$ is

$$\text{diag}\left(\boldsymbol{\Sigma}^{-1}, \frac{1}{2\xi_1^2}, \dots, \frac{i-1}{i\xi_{i-1}^2}, \frac{1}{i\eta_i^2}, \frac{1}{d_{i+1}^2}, \dots, \frac{1}{d_p^2}, \boldsymbol{\Delta}_2, \dots, \boldsymbol{\Delta}_p\right). \quad (28)$$

The one-at-a-time reference prior for $|\boldsymbol{\Sigma}_i|$, with nuisance parameters $\{\mu_1, \dots, \mu_p, \xi_1, \dots, \xi_{i-1}, d_{i+1}, \dots, d_p; t_{21}, t_{31}, t_{32}, \dots, t_{p1}, \dots, t_{p,p-1}\}$ and any parameter order, is the right-Haar prior π_H .

3. POSTERIOR DISTRIBUTIONS

Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The likelihood function of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{n}{2}(\bar{\mathbf{X}}_n - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) - \frac{1}{2} \text{tr}(\mathbf{S} \boldsymbol{\Sigma}^{-1})\right\},$$

where

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, \quad \mathbf{S} = \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)'.$$

Since all the considered priors are constant in $\boldsymbol{\mu}$, the conditional posterior for $\boldsymbol{\mu}$ will be

$$(\boldsymbol{\mu} | \boldsymbol{\Sigma}, \mathbf{X}) \sim N_p(\bar{\mathbf{x}}, \frac{1}{n} \boldsymbol{\Sigma}). \quad (29)$$

Generation from this is standard, so the challenge of simulation from the posterior distribution requires only sampling from the marginal posterior of Σ given \mathbf{S} . Note that the marginal likelihood of Σ based on \mathbf{S} is

$$L_1(\Sigma) = \frac{(2\pi)^{-np/2}}{|\Sigma|^{(n-1)/2}} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} \mathbf{S} \right). \quad (30)$$

Throughout the paper, we assume that \mathbf{S} is positive definite, as this is true with probability one.

3.1. Marginal Posteriors of Σ under π_J , π_{IJ} , π_{CE} and π_E

Marginal Posteriors Under π_J and π_{IJ} : It is immediate that these marginal posteriors for Σ are Inverse Wishart (\mathbf{S}^{-1}, n) and Inverse Wishart $(\mathbf{S}^{-1}, n-1)$, respectively.

Marginal Posterior Under π_{CE} : This marginal posterior distribution is imposing in its complexity. However, rather remarkably there is a simple rejection algorithm that can be used to generate from it:

Step 1. Generate $\Sigma \sim \text{Inverse Wishart}(\mathbf{S}^{-1}, n-1)$.

Step 2. Simulate $u \sim \text{Uniform}(0, 1)$. If $u \leq 2^{p/2} |\mathbf{I}_p + \Sigma^* \Sigma^{-1}|^{-1/2}$, report Σ . Otherwise go back to *Step 1*.

Note that the acceptance probability $2^{p/2} |\mathbf{I}_p + \Sigma^* \Sigma^{-1}|^{-1/2}$ is equal to one if the proposed Σ is diagonal, but is near zero when the proposed Σ is nearly singular. That this algorithm is a valid accept-reject algorithm, based on generation of Σ from the independence Jeffreys posterior, is established in Berger & Sun (2006).

Marginal Posterior Under π_E : It is possible to generate from this posterior using the following Metropolis-Hastings algorithm from Berger et. al. (2005).

Step 1. Generate $\Sigma^* \sim \text{Inverse Wishart}(\mathbf{S}^{-1}, n-1)$.

Step 2. Set $\Sigma' = \begin{cases} \Sigma^* & \text{with probability } \alpha, \\ \Sigma & \text{otherwise,} \end{cases}$ where

$$\alpha = \min \left\{ 1, \frac{\prod_{i < j} (\lambda_i^* - \lambda_j^*)}{\prod_{i < j} (\lambda_i - \lambda_j)} \cdot \frac{|\Sigma|^{(p-1)/2}}{|\Sigma^*|^{(p-1)/2}} \right\}.$$

3.2. Marginal Posterior of Σ under π_a

In the following, we will write $\Sigma = \Sigma_p$, including the dimension in order to derive some useful recursive formulas. Let $\psi_{p,p-1}$ be the $(p-1) \times 1$ vector of the last column of Ψ'_p , excluding ψ_{pp} . Then

$$\Psi_1 = \psi_{11}, \Psi_2 = \begin{pmatrix} \psi_{11} & 0 \\ \psi_{21} & \psi_{22} \end{pmatrix}, \dots, \Psi_p = \begin{pmatrix} \Psi_{p-1} & \mathbf{0} \\ \psi_{p,p-1} & \psi_{pp} \end{pmatrix}.$$

We will also write $\mathbf{S} = \mathbf{S}_p$ and let $\mathbf{s}_{p,p-1}$ represent the $(p-1) \times 1$ vector of the last column of \mathbf{S}_p excluding s_{pp} . Thus

$$\mathbf{S}_1 = s_{11}, \mathbf{S}_2 = \begin{pmatrix} s_{11} & s_{21} \\ s_{21} & s_{22} \end{pmatrix}, \dots, \mathbf{S}_p = \begin{pmatrix} \mathbf{S}_{p-1} & \mathbf{s}_{p,p-1} \\ \mathbf{s}_{p,p-1} & s_{pp} \end{pmatrix}. \quad (31)$$

Fact 5 Under the prior $\pi_{\mathbf{a}}$ in (13), the marginal posterior of Ψ is given as follows.
(a) For given $(\psi_{11}, \dots, \psi_{pp})$, the conditional posteriors of the off-diagonal vector $\psi_{i-1,i}$ are independent normal,

$$(\psi_{i,i-1} \mid \psi_{ii}, \mathbf{S}) \sim N(-\psi_{ii} \mathbf{S}_{i-1}^{-1} \mathbf{s}_{i,i-1}, \mathbf{S}_{i-1}^{-1}). \quad (32)$$

(b) The marginal posteriors of ψ_{ii}^2 ($1 \leq i \leq p$) are independent gamma $((n - a_i)/2, w_i/2)$, where

$$w_i = \begin{cases} s_{11}, & \text{if } i = 1, \\ \frac{|\mathbf{S}_i|}{|\mathbf{S}_{i-1}|} = s_{ii} - \mathbf{s}_{i,i-1}' \mathbf{S}_{i-1}^{-1} \mathbf{s}_{i-1,i}, & \text{if } i = 2, \dots, p. \end{cases} \quad (33)$$

(c) The marginal likelihood of \mathbf{Y} (or the normalizing constant) is

$$\begin{aligned} M &\equiv \int L(\boldsymbol{\mu}, \Psi) \pi_{\mathbf{a}}(\boldsymbol{\mu}, \Psi) d\boldsymbol{\mu} d\Psi \\ &= \frac{\prod_{i=1}^p \Gamma(\frac{1}{2}(n - a_i)) 2^{(n-a_i)/2}}{2^p (2\pi)^{(n-p)p/2} n^{p/2}} \frac{\prod_{i=1}^{p-1} |\mathbf{S}_i|^{(a_i - a_{i+1} - 1)/2}}{|\mathbf{S}|^{(n-a_p)/2}}. \end{aligned} \quad (34)$$

3.2.1. Constructive Posteriors

In the remainder of the paper we use, without further comment, the notation that * appended to a random variable denotes randomness arising from the constructive posterior (i.e., from the random variables used in simulation from the posterior), while a random variable without a * refers to randomness arising from the (frequentist) distribution of a statistic. Also, let Z_{ij} denote standard normal random variables. Whenever several of these occur in an expression, they are all independent (except that random variables of the same type and with the same index refer to the same random variable). Finally, we reserve quantile notation for posterior quantiles, with respect to the * distributions.

Fact 6 Consider the prior $\pi_{\mathbf{a}}$ in (13). Let $\chi_{n-a_i}^{2*}$ denote independent draws from chi-squared distributions with the indicated degree of freedoms, and $\mathbf{z}_{i,i-1}^*$ denote independent draws from $N_{i-1}(\mathbf{0}, \mathbf{I}_{i-1})$. The constructive posterior of $(\psi_{11}, \dots, \psi_{pp}, \psi_{2,1}, \dots, \psi_{p,p-1})$ given \mathbf{X} can be expressed as

$$\psi_{ii}^* = \sqrt{\frac{\chi_{n-a_i}^{2*}}{w_i}}, \quad i = 1, \dots, p, \quad (35)$$

$$\begin{aligned} \psi_{i,i-1}^* &= \mathbf{S}_{i-1}^{-1/2} \mathbf{z}_{i,i-1}^* - \psi_{ii}^* \mathbf{S}_{i-1}^{-1} \mathbf{s}_{i,i-1} \\ &= \mathbf{S}_{i-1}^{-1/2} \mathbf{z}_{i,i-1}^* - \sqrt{\frac{\chi_{n-a_i}^{2*}}{w_i}} \mathbf{S}_{i-1}^{-1} \mathbf{s}_{i,i-1}, \quad i = 2, \dots, p. \end{aligned} \quad (36)$$

Letting $\Psi^* = (\psi_{ij}^*)$, the constructive posterior of Σ is simply $\Sigma^* = \Psi^{*-1}(\Psi^{*-1})'$.

Alternatively, let \mathbf{V} be the Cholesky decomposition of \mathbf{S} , i.e., \mathbf{V} is the lower-triangular matrix with positive diagonal elements such that $\mathbf{S} = \mathbf{V}\mathbf{V}'$. It is easy to see that

$$w_i = t_{ii}^2, \quad i = 1, \dots, p. \quad (37)$$

We will also write $\mathbf{V} = \mathbf{V}_p$ and let $\mathbf{v}_{p,p-1}$ represent the $(p-1) \times 1$ vector of the last column of \mathbf{V}_p excluding v_{pp} . We have

$$\mathbf{V}_1 = v_{11}, \quad \mathbf{V}_2 = \begin{pmatrix} v_{11} & 0 \\ v_{21} & v_{22} \end{pmatrix}, \quad \dots, \quad \mathbf{V}_p = \begin{pmatrix} \mathbf{V}_{p-1} & 0 \\ \mathbf{v}_{p,p-1}' & v_{pp} \end{pmatrix}. \quad (38)$$

Corollary 0.3 *Under the prior π_a in (13),*

$$(\psi_{i,i-1} \mid \psi_{ii}, \mathbf{V}) \sim N(-\psi_{ii} \mathbf{V}_{i-1}^{-1'} \mathbf{v}_{i,i-1}, (\mathbf{V}_{i-1} \mathbf{V}_{i-1}')^{-1}), \quad (39)$$

$$(\psi_{ii}^2 \mid \mathbf{V}) \sim \text{inverse gamma}((n - a_i)/2, v_{ii}^2/2). \quad (40)$$

Proof. This follows from Fact 5 and the equality $\mathbf{S}_{i-1}^{-1} \mathbf{s}_{i,i-1} = \mathbf{V}_{i-1}^{-1'} \mathbf{v}_{i,i-1}$. \square

The following fact is immediate.

Fact 7 *Under the assumptions of Fact 6, the constructive posterior of $(\psi_{11}, \dots, \psi_{pp}, \psi_{21}, \dots, \psi_{p,p-1})$ given \mathbf{X} can be expressed as*

$$\psi_{ii}^* = \frac{\sqrt{\chi_{n-a_i}^{2*}}}{v_{ii}}, \quad i = 1, \dots, p, \quad (41)$$

$$\psi_{i,i-1}^* = \mathbf{V}_{i-1}^{-1'} \mathbf{z}_{i,i-1}^* - \frac{\sqrt{\chi_{n-a_i}^{2*}}}{v_{ii}} \mathbf{V}_{i-1}^{-1'} \mathbf{v}_{i,i-1}, \quad i = 2, \dots, p. \quad (42)$$

3.2.2. Posterior Means of Σ and Σ^{-1}

Fact 8 (a) *If $n - a_i > 0$, $i = 1, \dots, p$, then*

$$E(\Sigma^{-1} \mid \mathbf{X}) = E(\Psi' \Psi \mid \mathbf{X}) = \mathbf{V}^{-1'} \text{diag}(g_1, \dots, g_p) \mathbf{V}^{-1}, \quad (43)$$

where $g_i = n - a_i + p - i$, $i = 1, \dots, p$.

(b) *If $n - a_i > 2$, $i = 1, \dots, p$, then*

$$E(\Sigma \mid \mathbf{X}) = E((\Psi' \Psi)^{-1} \mid \mathbf{X}) = \mathbf{V} \text{diag}(h_1, \dots, h_p) \mathbf{V}', \quad (44)$$

where $h_1 = u_1$, $h_j = u_j \prod_{i=1}^{j-1} (1 + u_i)$, $j = 2, \dots, p$, with $u_i = 1/(n - a_i - 2)$, $i = 1, \dots, p$.

Proof. Letting $\mathbf{Y} = \Psi \mathbf{V}$, then $\mathbf{Y} = (y_{ij})_{p \times p}$ is still lower-triangular and

$$[\mathbf{Y} \mid \mathbf{X}] \propto \prod_{i=1}^p (y_{ii}^2)^{(n-a_i-1)/2} \exp\left\{-\frac{1}{2} \text{tr}(\mathbf{Y} \mathbf{Y}')\right\}. \quad (45)$$

From above, we know that all y_{ij} , $1 \leq i \leq j \leq p$, are independent and

$$\begin{aligned} y_{ij} &\sim N(0, 1), \quad 1 \leq j < i \leq p; \\ y_{ii} &\sim (y_{ii}^2)^{(n-a_i-1)/2} \exp\left(-\frac{1}{2} y_{ii}^2\right), \quad 1 \leq i \leq p. \end{aligned}$$

If $n - a_i > 0$, $i = 1, \dots, p$, then $(y_{ii}^2 | \mathbf{X}) \sim \text{gamma}((n - a_i)/2, 1/2)$ and $E(y_{ii}^2 | \mathbf{X})$ exists. Thus it is straightforward to get (43). For (44), we just need to show $E\{(\mathbf{Y}\mathbf{Y}')^{-1} | \mathbf{X}\} = \text{diag}(h_1, \dots, h_p)$. Under the condition $n - a_i > 1$, $E(y_{ii}^{-2} | \mathbf{X})$ exists and is equal to u_i , $i = 1, \dots, p$. Thus we obtain the result using the same procedure as in Eaton and Olkin (1987). \square

4. FREQUENTIST COVERAGE AND MARGINALIZATION PARADOXES

4.1. Frequentist Coverage Probabilities and Exact Matching

In this subsection we compare the frequentist properties of posterior credible intervals for various quantities under the prior $\pi_{\mathbf{a}}$, given in (13). As is customary in such comparisons, we study one-sided intervals $(\theta_L, q_{1-\alpha}(\mathbf{x}))$ of a parameter θ , where θ_L is the lower bound on the parameter θ (e.g., 0 or $-\infty$) and $q_{1-\alpha}(\mathbf{x})$ is the posterior quantile of θ , defined by

$$P(\theta < q_{1-\alpha}(\mathbf{x}) | \mathbf{x}) = 1 - \alpha.$$

Of interest is the frequentist coverage of the corresponding confidence interval, i.e.,

$$P(\theta < q_{1-\alpha}(\mathbf{X}) | \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

The closer this coverage is to the nominal $1 - \alpha$, the better the procedure (and corresponding objective prior) is judged to be.

Berger & Sun (2006) showed that, when $p = 2$, the right-Haar prior is exact matching prior for many functions of parameters of the bivariate normal distribution. Here we generalize the results to the multivariate normal distribution.

To prove frequentist matching, note first that $(\mathbf{S} | \boldsymbol{\Sigma}) \sim \text{Wishart}(n - 1, \boldsymbol{\Sigma})$. It is easy to see that the joint density for \mathbf{V} (the Chelosky decomposition of \mathbf{S}), given $\boldsymbol{\Psi}$, is

$$f(\mathbf{V} | \boldsymbol{\Psi}) \propto \prod_{i=1}^p v_{ii}^{n-i-1} \text{etr}\left(-\frac{1}{2}\boldsymbol{\Psi}\mathbf{V}\mathbf{V}'\boldsymbol{\Psi}'\right). \quad (46)$$

The following technical lemmas are also needed. The first lemma follows from the expansion

$$\text{tr}(\boldsymbol{\Psi}\mathbf{V}\mathbf{V}'\boldsymbol{\Psi}') = \sum_{i=1}^p \psi^2 v_{ii}^2 + \sum_{i=1}^p \sum_{j=1}^{i-1} \left(\sum_{k=1}^i \psi_{ik} v_{kj} \right)^2. \quad (47)$$

The proofs for both lemmas are straightforward and are omitted.

Lemma 1 For $n \geq p$ and given $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}'$, the following random variables are independent and have the indicated distributions:

$$Z_{ij} = \psi_{ii} \left(v_{ij} + \sum_{k=1}^{i-1} t_{ik} v_{kj} \right) \sim N(0, 1), \quad (48)$$

$$\psi_{ii} v_{ii} = \chi_{n-i}^2. \quad (49)$$

Lemma 2 Let $Y_{1-\alpha}$ denote the $1 - \alpha$ quantile of any random variable Y .

(a) If $g(\cdot)$ is a monotonically increasing function, $[g(Y)]_{1-\alpha} = g(Y_{1-\alpha})$ for any $\alpha \in (0, 1)$.

(b) For any $a > 0, b \in \mathbb{R}$, $(aY + b)_{1-\alpha} = aY_{1-\alpha} + b$.

(c) If W is a positive random variable, $(WY)_{1-\alpha} \geq 0$ if and only if $Y_{1-\alpha} \geq 0$.

Theorem 1 (a) For any $\alpha \in (0, 1)$ and fixed $i = 1, \dots, p$, the posterior $1 - \alpha$ quantile of ψ_{ii} has the expression

$$(\psi_{ii}^*)_{1-\alpha} = \frac{\sqrt{(\chi_{n-a_i}^{2*})_{1-\alpha}}}{v_{ii}}. \quad (50)$$

(b) For any $\alpha \in (0, 1)$ and any $(\boldsymbol{\mu}, \boldsymbol{\Psi})$, the frequentist coverage probability of the credible interval $(0, (\psi_{ii}^*)_{1-\alpha})$ is

$$P(\psi_{ii} < (\psi_{ii}^*)_{1-\alpha} \mid \boldsymbol{\mu}, \boldsymbol{\Psi}) = P(\chi_{n-i}^2 < (\chi_{n-a_i}^{2*})_{1-\alpha}), \quad (51)$$

which does not depend on $(\boldsymbol{\mu}, \boldsymbol{\Psi})$ and equals $1 - \alpha$ if and only if $a_i = i$.

Corollary 1.1 For any $\alpha \in (0, 1)$, the posterior quantile of $d_i = \text{var}(x_i \mid x_1, \dots, x_{i-1})$ is $(d_i^*)_{1-\alpha} = \frac{v_{ii}^2}{(\chi_{n-a_i}^{2*})_\alpha}$. For any $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the frequentist coverage probability of the credible interval $(0, (d_i^*)_{1-\alpha}) = (\chi_{n-i}^2 < (\chi_{n-a_i}^{2*})_{1-\alpha})$, is a constant $P(\chi_{n-i}^2 > (\chi_{n-a_i}^{2*})_\alpha)$, and equals $1 - \alpha$ if and only if $a_i = i$.

Observing that $|\boldsymbol{\Sigma}_i| = \prod_{j=1}^i d_j$ yields the following result.

Theorem 2 (a) For any α , the posterior $1 - \alpha$ quantile of $|\boldsymbol{\Sigma}_i|$ has the expression

$$(|\boldsymbol{\Sigma}_i|)_{1-\alpha} = \frac{\prod_{j=1}^i v_{jj}^2}{\left(\prod_{j=1}^i \chi_{n-a_j}^{2*} \right)_\alpha}. \quad (52)$$

(b) For any $\alpha \in (0, 1)$ and any $(\boldsymbol{\mu}, \boldsymbol{\Psi})$, the frequentist coverage probability of the credible interval $(0, (|\boldsymbol{\Sigma}_i|)_{1-\alpha})$ is

$$P(|\boldsymbol{\Sigma}_i| < (|\boldsymbol{\Sigma}_i|)_{1-\alpha} \mid \boldsymbol{\mu}, \boldsymbol{\Psi}) = P\left(\prod_{j=1}^i \chi_{n-j}^2 > \left(\prod_{j=1}^i \chi_{n-a_j}^{2*} \right)_\alpha \right), \quad (53)$$

which is a constant and equals $1 - \alpha$ if and only if (a_1, \dots, a_i) is a permutation of $(1, \dots, i)$.

For the bivariate normal case, Berger & Sun (2006) showed that the right-Haar measure is the exact matching prior for ψ_{21} and t_{12} . We also expect that, for the multivariate normal distribution, the right-Haar prior is exact matching for all ψ_{ij} and t_{ij} .

5. FREQUENTIST COVERAGE AND MARGINALIZATION PARADOXES

5.1. Frequentist Coverage Probabilities and Exact Matching

In this subsection we compare the frequentist properties of posterior credible intervals for various quantities under the prior $\pi_{\mathbf{a}}$, given in (13). As is customary in such comparisons, we study one-sided intervals $(\theta_L, q_{1-\alpha}(\mathbf{x}))$ of a parameter θ , where θ_L is the lower bound on the parameter θ (e.g., 0 or $-\infty$) and $q_{1-\alpha}(\mathbf{x})$ is the posterior quantile of θ , defined by

$$P(\theta < q_{1-\alpha}(\mathbf{x}) \mid \mathbf{x}) = 1 - \alpha.$$

Of interest is the frequentist coverage of the corresponding confidence interval, i.e.,

$$P(\theta < q_{1-\alpha}(\mathbf{X}) \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

The closer this coverage is to the nominal $1 - \alpha$, the better the procedure (and corresponding objective prior) is judged to be.

Berger & Sun (2006) showed that, when $p = 2$, the right-Haar prior is exact matching prior for many functions of parameters of the bivariate normal distribution. Here we generalize the results to the multivariate normal distribution.

To prove frequentist matching, note first that $(\mathbf{S} \mid \boldsymbol{\Sigma}) \sim \text{Wishart}(n - 1, \boldsymbol{\Sigma})$. It is easy to see that the joint density for \mathbf{V} (the Chelosky decomposition of \mathbf{S}), given $\boldsymbol{\Psi}$, is

$$f(\mathbf{V} \mid \boldsymbol{\Psi}) \propto \prod_{i=1}^p v_{ii}^{n-i-1} \exp\left(-\frac{1}{2} \boldsymbol{\Psi} \mathbf{V} \mathbf{V}' \boldsymbol{\Psi}'\right). \quad (54)$$

The following technical lemmas are also needed. The first lemma follows from the expansion

$$\text{tr}(\boldsymbol{\Psi} \mathbf{V} \mathbf{V}' \boldsymbol{\Psi}') = \sum_{i=1}^p \psi_{ii}^2 v_{ii}^2 + \sum_{i=1}^p \sum_{j=1}^{i-1} \left(\sum_{k=1}^i \psi_{ik} v_{kj} \right)^2. \quad (55)$$

The proofs for both lemmas are straightforward and are omitted.

Lemma 3 *For $n \geq p$ and given $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi} \boldsymbol{\Psi}'$, the following random variables are independent and have the indicated distributions:*

$$Z_{ij} = \psi_{ii} \left(v_{ij} + \sum_{k=1}^{i-1} t_{ik} v_{kj} \right) \sim N(0, 1), \quad (56)$$

$$\psi_{ii} v_{ii} = \chi_{n-i}^2. \quad (57)$$

Lemma 4 *Let $Y_{1-\alpha}$ denote the $1 - \alpha$ quantile of any random variable Y .*

(a) *If $g(\cdot)$ is a monotonically increasing function, $[g(Y)]_{1-\alpha} = g(Y_{1-\alpha})$ for any $\alpha \in (0, 1)$.*

(b) *For any $a > 0, b \in \mathbb{R}$, $(aY + b)_{1-\alpha} = aY_{1-\alpha} + b$.*

(c) *If W is a positive random variable, $(WY)_{1-\alpha} \geq 0$ if and only if $Y_{1-\alpha} \geq 0$.*

Theorem 3 (a) For any $\alpha \in (0, 1)$ and fixed $i = 1, \dots, p$, the posterior $1 - \alpha$ quantile of ψ_{ii} has the expression

$$(\psi_{ii}^*)_{1-\alpha} = \frac{\sqrt{(\chi_{n-a_i}^{2*})_{1-\alpha}}}{v_{ii}}. \quad (58)$$

(b) For any $\alpha \in (0, 1)$ and any $(\boldsymbol{\mu}, \boldsymbol{\Psi})$, the frequentist coverage probability of the credible interval $(0, (\psi_{ii}^*)_{1-\alpha})$ is

$$P(\psi_{ii} < (\psi_{ii}^*)_{1-\alpha} \mid \boldsymbol{\mu}, \boldsymbol{\Psi}) = P(\chi_{n-i}^2 < (\chi_{n-a_i}^{2*})_{1-\alpha}), \quad (59)$$

which does not depend on $(\boldsymbol{\mu}, \boldsymbol{\Psi})$ and equals $1 - \alpha$ if and only if $a_i = i$.

Corollary 3.1 For any $\alpha \in (0, 1)$, the posterior quantile of $d_i = \text{var}(x_i \mid x_1, \dots, x_{i-1})$ is $(d_i^*)_{1-\alpha} = \frac{v_{ii}^2}{(\chi_{n-a_i}^{2*})_\alpha}$. For any $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the frequentist coverage probability of the credible interval $(0, (d_i^*)_{1-\alpha}) = (\chi_{n-i}^2 < (\chi_{n-a_i}^{2*})_{1-\alpha})$, is a constant $P(\chi_{n-i}^2 > (\chi_{n-a_i}^{2*})_\alpha)$, and equals $1 - \alpha$ if and only if $a_i = i$.

Observing that $|\boldsymbol{\Sigma}_i| = \prod_{j=1}^i d_j$ yields the following result.

Theorem 4 (a) For any α , the posterior $1 - \alpha$ quantile of $|\boldsymbol{\Sigma}_i|$ has the expression

$$(|\boldsymbol{\Sigma}_i|)_{1-\alpha} = \frac{\prod_{j=1}^i v_{jj}^2}{\left(\prod_{j=1}^i \chi_{n-a_j}^{2*} \right)_\alpha}. \quad (60)$$

(b) For any $\alpha \in (0, 1)$ and any $(\boldsymbol{\mu}, \boldsymbol{\Psi})$, the frequentist coverage probability of the credible interval $(0, (|\boldsymbol{\Sigma}_i|)_{1-\alpha})$ is

$$P(|\boldsymbol{\Sigma}_i| < (|\boldsymbol{\Sigma}_i|)_{1-\alpha} \mid \boldsymbol{\mu}, \boldsymbol{\Psi}) = P\left(\prod_{j=1}^i \chi_{n-j}^2 > \left(\prod_{j=1}^i \chi_{n-a_j}^{2*} \right)_\alpha \right), \quad (61)$$

which is a constant and equals $1 - \alpha$ if and only if (a_1, \dots, a_i) is a permutation of $(1, \dots, i)$.

For the bivariate normal case, Berger & Sun (2006) showed that the right-Haar measure is the exact matching prior for ψ_{21} and t_{12} . We also expect that, for the multivariate normal distribution, the right-Haar prior is exact matching for all ψ_{ij} and t_{ij} .

5.2. Marginalization Paradoxes

While the Bayesian credible intervals for many parameters under the right-Haar measure are exact matching priors, it can be seen that the prior can suffer marginalization paradoxes. The basis for such paradoxes (Dawid et al. (1973)) is that any proper prior has the property: if the marginal posterior distribution for a parameter

θ depends only on a statistic T – whose distribution in turn depends only on θ – then the posterior of θ can be derived from the distribution of T together with the marginal prior for θ . While this is a basic property of any proper Bayesian prior, it can be violated for improper priors, with the result then called a marginalization paradox.

In Berger & Sun (2006), it was shown that, when using the right-Haar prior, the posterior distribution of the correlation coefficient ρ for a bivariate normal distribution depends only on the sample correlation coefficient r . Brillinger (1962) showed that there does not exist a prior $\pi(\rho)$ such that the this posterior density equals $f(r|\rho)\pi(\rho)$, where $f(r|\rho)$ is the density of r given ρ . This thus provides an example of a marginalization paradox.

Here is another marginalization paradox in the bivariate normal case. We know from Berger & Sun (2006) that the right-Haar prior π_H is exact matching prior for ψ_{21} . Note that the constructive posterior of ψ_{21} is

$$\frac{Z^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi_{n-2}^{2*}}}{\sqrt{s_{11}}} \frac{r}{\sqrt{1-r^2}}, \quad (62)$$

which clearly depends only on (s_{11}, r) .

It turns out that the joint density of (s_{11}, r) depends only on (σ_{11}, ρ) . Note that the posterior of (σ_{11}, ρ) based on the product of $f(s_{11}, r | \sigma_{11}, \rho)$ and the marginal prior for (σ_{11}, ρ) based on π_H is different from the marginal posterior of (σ_{11}, ρ) based on π_H .

Consequently, the posterior distribution of ψ_{21} from the right haar provides another example of the marginalization paradox.

It is somewhat controversial as to whether violation of the marginalization paradox is a serious problem. For instance, in the bivariate normal problem, there is probably no proper prior distribution that yields a marginal posterior distribution of ρ which depends only on r , so the relevance of an unattainable property of proper priors could be questioned.

In any case, this situation provides an interesting philosophical conundrum of a type that we have not previously seen: a complete objective Bayesian and frequentist unification can be obtained for inference about the usual parameters of the bivariate normal distribution, but only if violation of the marginalization paradox is accepted.

The prior π_{CE} does avoid the marginalization paradox for ρ_{12} , but is not exact frequentist matching. We, alas, know of no way to adjudicate between the competing goals of exact frequentist matching and avoidance of the marginalization paradox, and so will simply present both as possible objective Bayesian approaches.

6. ON THE NON-UNIQUENESS OF RIGHT-HAAR PRIORS

While the right-Haar priors seem to have some very nice properties, the fact that they depend on the particular lower triangular matrix decomposition of Σ^{-1} that is used is troubling. In the bivariate case, for instance, both

$$\pi_1(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{\sigma_2^2(1-\rho^2)} \quad \text{and} \quad \pi_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{\sigma_1^2(1-\rho^2)}$$

are right-Haar priors (expressed with respect to $d\mu_1 d\mu_2 d\sigma_1 d\sigma_2 d\rho$).

There are several natural proposals for dealing with this non-uniqueness. One is to mix over the right-Haar priors. Another is to choose the ‘empirical Bayes’

right-Haar prior, that which maximizes the marginal likelihood of the data. These proposals are developed in the next two subsections. The last subsection shows, quite surprisingly, that neither of these solutions works! For simplicity, we restrict attention to the bivariate normal case.

6.1. Symmetrized Right-Haar Priors

Consider the symmetrized right-Haar prior

$$\begin{aligned}\tilde{\pi}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) &= \pi_1(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) + \pi_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \\ &= \frac{1}{\sigma_1^2(1-\rho^2)} + \frac{1}{\sigma_2^2(1-\rho^2)}.\end{aligned}\quad (63)$$

This can be thought of as a 50-50 mixture of the two right-Haar priors.

Fact 9 *The joint posterior of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ under the prior $\tilde{\pi}$ is given by*

$$\begin{aligned}\tilde{\pi}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho \mid \mathbf{X}) &= C\pi_1(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho \mid \mathbf{X}) \\ &+ (1-C)\pi_2(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho \mid \mathbf{X}),\end{aligned}\quad (64)$$

where

$$C = \frac{s_{11}^{-1}}{s_{11}^{-1} + s_{22}^{-1}}. \quad (65)$$

and $\pi_1(\cdot \mid \mathbf{X})$ and $\pi_2(\cdot \mid \mathbf{X})$ are the posteriors under the priors π_1 and π_2 , respectively.

Proof. Let $p = 2$ and $(a_1, a_2) = (1, 2)$ in (34). We get

$$\begin{aligned}C_j &= \int L(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \pi_j(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) d\mu_1 d\mu_2 d\sigma_1 d\sigma_2 d\rho \\ &= \frac{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-2}{2})2^{(n-2)/2}s_{jj}^{-1}}{\pi^{(n-3)/2}|\mathbf{S}|^{(n-2)/2}},\end{aligned}\quad (66)$$

for $j = 1, 2$. The result is immediate. \square

For later use, note that, under the prior $\tilde{\pi}$, the posterior mean of Σ has the form

$$\hat{\Sigma}_S = E(\Sigma \mid \mathbf{X}) \equiv E(\Sigma \mid \mathbf{X}) = C\hat{\Sigma}_1 + (1-C)\hat{\Sigma}_2, \quad (67)$$

where $\hat{\Sigma}_i$ is the posterior mean under $\pi_i(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, given by

$$\hat{\Sigma}_i = \frac{1}{n-3}(\mathbf{S} + \mathbf{G}_i), \quad (68)$$

where

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{2}{n-4}\left(s_{22} - \frac{s_{12}^2}{s_{11}}\right) \end{pmatrix}, \quad \mathbf{G}_2 = \begin{pmatrix} \frac{2}{n-4}\left(s_{11} - \frac{s_{12}^2}{s_{22}}\right) & 0 \\ 0 & 0 \end{pmatrix}. \quad (69)$$

Here Σ_1 is a special case of (44) when $p = 2$ and $(a_1, a_2) = (1, 2)$.

6.2. The Empirical Bayes Right-Haar Prior

The right-Haar priors above were essentially just obtained by coordinate permutation. More generally, one can obtain other right-Haar priors by orthonormal transformations of the data. In particular, define the orthonormal matrix

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix},$$

where the γ_i are orthonormal row vectors. Consider the transformation of the data $\mathbf{\Gamma x}$, so that the resulting sample covariance matrix is

$$\mathbf{S}^* = \mathbf{\Gamma S \Gamma}' = \begin{pmatrix} s_{11}^* & s_{12}^* \\ s_{12}^* & s_{22}^* \end{pmatrix} = \begin{pmatrix} \gamma_1 \mathbf{S} \gamma_1' & \gamma_1 \mathbf{S} \gamma_2' \\ \gamma_2 \mathbf{S} \gamma_1' & \gamma_2 \mathbf{S} \gamma_2' \end{pmatrix}. \quad (70)$$

The right-Haar prior can be defined in this transformed problem, so that each $\mathbf{\Gamma}$ defines a different right-Haar prior.

A commonly employed technique when facing a class of priors, as here, is to choose the ‘empirical Bayes’ prior, that which maximizes the marginal likelihood of the data. This is given in the following lemma.

Lemma 5 *The empirical Bayes right-Haar prior is given by that $\mathbf{\Gamma}$ for which*

$$\begin{aligned} s_{11}^* &= \frac{1}{2}(s_{11} + s_{22}) - \frac{1}{2}\sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}, \\ s_{12}^* &= 0, \\ s_{22}^* &= \frac{1}{2}(s_{11} + s_{22}) + \frac{1}{2}\sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}. \end{aligned}$$

(Note that the two eigenvalues of \mathbf{S} are s_{11}^* and s_{22}^* . Thus this is the orthonormal transformation such that the sample variance of the first coordinate is the smallest eigenvalue.)

Proof. Noting that $|\mathbf{S}^*| = |\mathbf{S}|$, it follows from (66) that the marginal likelihood of $\mathbf{\Gamma}$ is proportional to s_{11}^{*-1} . Hence we simply want to find an orthonormal $\mathbf{\Gamma}$ to minimize $\gamma_1 \mathbf{S} \gamma_1'$. It is standard matrix theory that the minimum is the smallest eigenvalue of \mathbf{S} , with γ_1 being the associated eigenvector. Since $\mathbf{\Gamma}$ is orthonormal, the remainder of the lemma also follows directly. \square

Lemma 6 *Under the empirical Bayes right-Haar prior, the posterior mean of $\mathbf{\Sigma}$ is $\hat{\mathbf{\Sigma}}_E = \mathbf{E}(\mathbf{\Sigma} | \mathbf{X})$ and given by*

$$\begin{aligned} \hat{\mathbf{\Sigma}}_E &= \frac{1}{n-3} \left(\mathbf{S} + \frac{s_{22}^*}{n-4} \left(\mathbf{I} + \frac{1}{s_{22}^* - s_{11}^*} \begin{pmatrix} s_{11} - s_{22} & 2s_{12} \\ 2s_{12} & s_{22} - s_{11} \end{pmatrix} \right) \right) \\ &= \frac{1}{n-3} \left(\mathbf{S} + \frac{s_{22}^*}{n-4} \left(\mathbf{I} + \frac{1}{s_{22}^* - s_{11}^*} \mathbf{S} - \frac{1}{\frac{1}{s_{11}^*} - \frac{1}{s_{22}^*}} \mathbf{S}^{-1} \right) \right). \end{aligned}$$

Proof. Under the empirical Bayes right-Haar prior, the posterior mean of $\Sigma^* = \Gamma \Sigma \Gamma'$ is

$$E(\Sigma^* | \mathbf{X}) = \frac{1}{n-3}(\mathbf{S}^* + \mathbf{G}^*),$$

where

$$\mathbf{G}^* = \begin{pmatrix} 0 & 0 \\ 0 & g^* \end{pmatrix}, \quad g^* = \frac{2}{n-4} \left(s_{22}^* - \frac{s_{12}^{*2}}{s_{11}^*} \right) = \frac{2s_{22}^*}{n-4}.$$

So the corresponding estimate of Σ is

$$E(\Sigma | \mathbf{X}) = \Gamma' E(\Sigma^* | \mathbf{X}) \Gamma = \frac{1}{n-3}(\mathbf{S} + \Gamma' \mathbf{G}^* \Gamma).$$

Computation yields that the eigenvector γ_2 is such that

$$\begin{aligned} \gamma_{21}^2 &= \frac{1}{2} + \frac{1}{2} \frac{s_{11} - s_{22}}{\sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}}, \\ \gamma_{22}^2 &= \frac{1}{2} - \frac{1}{2} \frac{s_{11} - s_{22}}{\sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}}, \\ \gamma_{21}\gamma_{22} &= \frac{s_{12}}{\sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}}. \end{aligned}$$

Thus

$$\begin{aligned} \Gamma' \mathbf{G}^* \Gamma &= g^* \gamma_2' \gamma_2 \\ &= \frac{s_{22}^*}{n-4} \left(\mathbf{I} + \frac{1}{\sqrt{(s_{11} - s_{22})^2 + 4s_{12}^2}} \begin{pmatrix} s_{11} - s_{22} & 2s_{12} \\ 2s_{12} & s_{22} - s_{11} \end{pmatrix} \right) \\ &= \frac{s_{22}^*}{n-4} \left(\mathbf{I} + \frac{1}{s_{22}^* - s_{11}^*} \begin{pmatrix} s_{11} - s_{22} & 2s_{12} \\ 2s_{12} & s_{22} - s_{11} \end{pmatrix} \right). \end{aligned}$$

The last expression in the lemma follows from algebra. \square

6.3. Decision-Theoretic Evaluation

To study the effectiveness of the symmetrized right-Haar prior and the empirical Bayes right-Haar prior, we turn to a decision theoretic evaluation, utilizing a natural invariant loss function.

For a multivariate normal distribution $N_p(\mu, \Sigma)$ with unknown (μ, Σ) , a natural loss to consider is the entropy loss, defined by

$$\begin{aligned} L(\hat{\mu}, \hat{\Sigma}; \mu, \Sigma) &= 2 \int \log \left\{ \frac{f(\mathbf{X} | \mu, \Sigma)}{f(\mathbf{X} | \hat{\mu}, \hat{\Sigma})} \right\} f(\mathbf{X} | \mu, \Sigma) d\mathbf{X} \\ &= (\hat{\mu} - \mu)' \hat{\Sigma}^{-1} (\hat{\mu} - \mu) + \text{tr}(\hat{\Sigma}^{-1} \Sigma) - \log |\hat{\Sigma}^{-1} \Sigma| - p. \quad (71) \end{aligned}$$

Clearly, the entropy loss has two parts, one is related to the means $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}$ (with $\hat{\boldsymbol{\Sigma}}$ as the weight matrix), and the other is related to $\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}$. The last three terms of this expression are related to ‘Stein’s loss,’ and is the most commonly used losses for estimation of a covariance matrix (cf. James & Stein (1961), Haff (1977)).

Lemma 7 *Under the loss (71) and for any of the priors considered in this paper, the generalized Bayesian estimator of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is*

$$\hat{\boldsymbol{\mu}}_B = E(\boldsymbol{\mu} | \mathbf{X}) = (\bar{x}_1, \bar{x}_2)', \quad (72)$$

$$\hat{\boldsymbol{\Sigma}}_B = E(\boldsymbol{\Sigma} | \mathbf{X}) + E\{(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu}) | \mathbf{X}\} = \frac{n+1}{n}E(\boldsymbol{\Sigma} | \mathbf{X}). \quad (73)$$

Proof. For the priors we consider in the paper,

$$[\boldsymbol{\mu} | \boldsymbol{\Sigma}, \mathbf{X}] \sim N_2 \left((\bar{x}_1, \bar{x}_2)', \frac{1}{n}\boldsymbol{\Sigma} \right), \quad (74)$$

so that (72) is immediate. Furthermore, it follows that

$$E((\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu})'\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_B - \boldsymbol{\mu}) | \mathbf{X}) = \frac{1}{n}tr(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) \quad (75)$$

so that the remaining goal is to choose $\hat{\boldsymbol{\Sigma}}$ so as to minimize

$$\begin{aligned} & E\left((1 + \frac{1}{n})tr(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) - \log |\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}| - p | \mathbf{X}\right) \\ &= E\left(tr(\hat{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\Sigma}}) - \log |\hat{\boldsymbol{\Sigma}}^{-1}\tilde{\boldsymbol{\Sigma}}| - p | \mathbf{X}\right) + \log(1 + \frac{1}{n}), \end{aligned} \quad (76)$$

where $\tilde{\boldsymbol{\Sigma}} = (1 + \frac{1}{n})\boldsymbol{\Sigma}$. It is standard (see, e.g., Eaton (1989)) that the first term on the right hand side of the last expression is minimized at

$$\hat{\boldsymbol{\Sigma}} = E(\tilde{\boldsymbol{\Sigma}} | \mathbf{X}) = (1 + \frac{1}{n})E(\boldsymbol{\Sigma} | \mathbf{X}), \quad (77)$$

from which the result is immediate. \square

We now turn to frequentist decision-theoretic evaluation of the various posterior estimates that arise from the reference priors considered in the paper. Thus we now change perspective and consider $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to be given, and consider the frequentist risk of the posterior estimates $\hat{\boldsymbol{\mu}}_B(\mathbf{X})$ and $\hat{\boldsymbol{\Sigma}}_B(\mathbf{X})$, now considered as functions of \mathbf{X} . Thus we evaluate the frequentist risk

$$R(\hat{\boldsymbol{\mu}}_B, \hat{\boldsymbol{\Sigma}}_B; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = EL(\hat{\boldsymbol{\mu}}_B(\mathbf{X}), \hat{\boldsymbol{\Sigma}}_B(\mathbf{X}); \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (78)$$

where the expectation is over \mathbf{X} given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The following lemma states that we can reduce the frequentist risk comparison to a comparison of the frequentist risks of the various posterior means for $\boldsymbol{\Sigma}$ under Stein’s loss. It’s proof is virtually identical to that of Lemma 7, and is omitted.

Lemma 8 *For frequentist comparison of the various Bayes estimators considered in the paper, it suffices to compare the frequentist risks of the $\widehat{\Sigma}(\mathbf{X}) = \mathbb{E}(\Sigma \mid \mathbf{X})$, with respect to*

$$R(\widehat{\Sigma}(\mathbf{X}); \boldsymbol{\mu}, \Sigma) = \mathbb{E} \left(\text{tr}(\widehat{\Sigma}^{-1}(\mathbf{X})\Sigma) - \log |\widehat{\Sigma}^{-1}(\mathbf{X})\Sigma| - p \right), \quad (79)$$

where the expectation is with respect to \mathbf{X} .

Lemma 9 *Under the right haar prior π_H , the risk function (79) is a constant, given by*

$$R(\widehat{\Sigma}(\mathbf{X}); \boldsymbol{\mu}, \Sigma) = \sum_{j=1}^p \log(h_j) + \sum_{j=1}^p \mathbb{E} \log(\chi_{n-j}^2). \quad (80)$$

where h_j is given by (44).

The proof can be found from Eaton (1989). Table 2 gives these risks for the estimates arising from the two right-Haar priors, $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$, the estimate $\widehat{\Sigma}_S$ arising from the symmetrized right-Haar prior, the estimate $\widehat{\Sigma}_E$ arising from the empirical Bayes right-Haar prior, and an estimate in the spirit of Dey & Srinivasan (1985) that will be discussed shortly.

Table 2: *Frequentist risks of various estimates of Σ when $n = 10$ and for various choices of Σ . These were computed by simulation, using 10,000 generated values of \mathbf{S}*

true Σ	$R(\widehat{\Sigma}_1; \Sigma)$	$R(\widehat{\Sigma}_2; \Sigma)$	$R(\widehat{\Sigma}_S; \Sigma)$	$R(\widehat{\Sigma}_E; \Sigma)$	$R(\widehat{\Sigma}_D; \Sigma)$
$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	0.4287	0.4288	0.4452	0.6052	0.3833
$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$	0.4278	0.4270	0.4424	0.5822	0.3859
$\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$	0.4285	0.4287	0.4391	0.5404	0.3989
$\begin{pmatrix} 1 & 0 \\ 0 & 50 \end{pmatrix}$	0.4254	0.4250	0.4272	0.5100	0.4194
$\begin{pmatrix} 1 & .1 \\ .1 & 1 \end{pmatrix}$	0.4255	0.4266	0.4424	0.5984	0.3810
$\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$	0.4274	0.4275	0.4403	0.5607	0.3906
$\begin{pmatrix} 1 & .9 \\ .9 & 1 \end{pmatrix}$	0.4260	0.4255	0.4295	0.5159	0.4134

It follows from Lemma 9 that if $p = 2$, the risk for the two right-Haar priors is

$$\log(n-2) - 2\log(n-3) - \log(n-4) + E\log(\chi_{n-1}^2) + E\log(\chi_{n-2}^2).$$

When $n = 10$, it is approximately 0.4271448. The simulated risks are given in the Table 2 instead because the comparison between priors is best done by comparing simulated values (rather than a simulated value with an analytic value).

The first surprise here is that the risk of $\hat{\Sigma}_S$ is actually worse than the risk of the right-Haar prior estimates. This is in contradiction to the usual belief that, if considering alternate priors, utilization of a mixture of the two priors will give superior performance.

This would also seem to be in contradiction to the known fact for a convex loss function (such as Stein's loss) that, if two estimators $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ have equal risk functions, then an average of the two estimators will have lower risk. But this refers to a constant average of the two estimators, not a data-weighted average as in $\hat{\Sigma}_S$. What is particularly striking is that the data-weighted average arises from the posterior marginal likelihoods corresponding to the two different priors, so the posterior seems to be 'getting it wrong,' weighting the 'bad' prior more than the 'good' prior.

This is indicated in even more dramatic fashion by $\hat{\Sigma}_E$, the empirical Bayes version, which is based on that right-Haar prior which is 'most likely' for given data. In that the risk of $\hat{\Sigma}_E$ is much worse than even the risk of $\hat{\Sigma}_S$, it seems that empirical Bayes has selected the worst of all of the right-Haar priors!

The phenomenon arising here is disturbing and sobering. It is yet another indication that improper priors do not behave as do proper priors, and that it can be dangerous to apply 'understandings' from the world of proper priors to the world of improper priors. (Of course, the same practical problems could arise from use of vague proper priors, so use of such is not a solution to the problem.)

From a formal objective Bayesian position (e.g., the viewpoint from the reference prior perspective), there is no issue here. The various reference priors we considered are (by definition) the correct objective priors for the particular contexts (choice of parameterization and parameter of interest) in which they were derived. It is use of these priors – or modifications of them based on 'standard tricks' – out of context that is being demonstrated to be of concern.

APPENDIX A: PROOFS

Proof of Fact 1. The likelihood function of (μ, Ψ) is

$$f(\mathbf{x} \mid \mu, \Psi) \propto |\Psi' \Psi|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)' \Psi' \Psi (\mathbf{x} - \mu)\right),$$

and the log-likelihood is then

$$\log f = \text{const} + \sum_{i=1}^p \log(\psi_{ii}) - \frac{1}{2} \sum_{i=1}^p \left(\sum_{j=1}^i \psi_{ij} (x_j - \mu_j) \right)^2.$$

For any fixed $i = 1, \dots, p$, let Σ_i be the variance and covariance matrix of $(x_1, \dots, x_i)'$. Also, let \mathbf{e}_i be the $i \times 1$ vector whose i th element is 1 and 0 otherwise. The Fisher information matrix of θ is then (11). ■

Note that $\text{Var}(\mathbf{x}) = \mathbf{\Sigma} = \mathbf{\Psi}^{-1}\mathbf{\Psi}'^{-1}$. Let $\mathbf{\Psi}_i$ be the $i \times i$ left and top sub-matrix of $\mathbf{\Psi}$. It is easy to verify that $\mathbf{\Sigma}_i = \mathbf{\Psi}_i^{-1}\mathbf{\Psi}_i'^{-1}$. Using the fact that $|\mathbf{B} + \mathbf{a}\mathbf{a}'| = |\mathbf{B}|(1 + \mathbf{a}'\mathbf{B}^{-1}\mathbf{a})$ where \mathbf{B} is invertible and \mathbf{a} is a vector, we can show that

$$|\mathbf{\Lambda}_i| = 2 \prod_{j=1}^i \frac{1}{\psi_{jj}^2}. \quad (81)$$

From (11) and (81), the reference prior of $\mathbf{\Psi}$ for the ordered group $\{\boldsymbol{\mu}, \psi_{11}, (\psi_{21}, \psi_{22}), \dots, (\psi_{p1}, \dots, \psi_{pp})\}$, is easy to obtain as (12) according to the algorithm in Berger and Bernardo (1992b).

Proof of Fact 2. For $i = 2, \dots, p$, denote $\mathbf{t}_{i,i-1} = \boldsymbol{\psi}_{i,i-1}/\psi_{ii}$. Clearly, the Jacobian from $(\boldsymbol{\psi}_{i,i-1}, \psi_{ii})$ to $(\mathbf{t}_{i,i-1}, \psi_{ii})$ is

$$\mathbf{J}_i = \frac{\partial(\boldsymbol{\psi}_{i,i-1}, \psi_{ii})}{\partial(\mathbf{t}_{i,i-1}, \psi_{ii})} = \begin{pmatrix} \psi_{ii}\mathbf{I}_{i-1} & \mathbf{t}_{i,i-1} \\ \mathbf{0}' & 1 \end{pmatrix}. \quad (82)$$

The Fisher information for $\tilde{\boldsymbol{\theta}}$ has the form (19), where

$$\tilde{\mathbf{\Lambda}}_i = \mathbf{J}_i' \mathbf{\Lambda}_i \mathbf{J}_i = \mathbf{J}_i' \left(\mathbf{\Psi}_i^{-1} \mathbf{\Psi}_i'^{-1} + \frac{1}{\psi_{ii}^2} \mathbf{e}_i \mathbf{e}_i' \right) \mathbf{J}_i. \quad (83)$$

Note that

$$\mathbf{\Psi}_i = \begin{pmatrix} \mathbf{\Psi}_{i-1} & \mathbf{0} \\ \psi_{ii}\mathbf{t}_{i,i-1}' & \psi_{ii} \end{pmatrix} \quad \text{and} \quad \mathbf{\Psi}_i^{-1} = \begin{pmatrix} \mathbf{\Psi}_{i-1}^{-1} & \mathbf{0} \\ -\mathbf{t}_{i,i-1}' \mathbf{\Psi}_{i-1}^{-1} & \frac{1}{\psi_{ii}} \end{pmatrix}.$$

We have that

$$\mathbf{J}_i' \mathbf{\Psi}_i^{-1} = \begin{pmatrix} \psi_{ii}\mathbf{I}_{i-1} & \mathbf{t}_{i,i-1} \\ \mathbf{0}' & 1 \end{pmatrix} \begin{pmatrix} \mathbf{\Psi}_{i-1}^{-1} & \mathbf{0} \\ -\mathbf{t}_{i,i-1}' \mathbf{\Psi}_{i-1}^{-1} & \frac{1}{\psi_{ii}} \end{pmatrix} = \begin{pmatrix} \psi_{ii}\mathbf{\Psi}_{i-1}^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{1}{\psi_{ii}} \end{pmatrix}. \quad (84)$$

Substituting (84) into (83) and using the fact that $\mathbf{J}_i' \mathbf{e}_i = \mathbf{e}_i$,

$$\tilde{\mathbf{J}}_i = \begin{pmatrix} \psi_{ii}^2 \mathbf{\Psi}_{i-1}^{-1} \mathbf{\Psi}_{i-1}'^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2}{\psi_{ii}^2} \end{pmatrix} = \begin{pmatrix} \psi_{ii}^2 \mathbf{T}_{i-1}^{-1} \tilde{\mathbf{\Psi}}_{i-1}^{-2} \mathbf{T}_{i-1}'^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2}{\psi_{ii}^2} \end{pmatrix}. \quad (85)$$

Part (a) holds. It is easy to see that the upper and left $i \times i$ submatrix of $\mathbf{\Lambda}_i^*$ does not depend on $\mathbf{t}_{i,i-1}$. Part (b) can be proved using the algorithm of Berger and Bernardo (1992b). Furthermore, part (c) holds because of $|\mathbf{J}_i| = \psi_{ii}^{i-1}$.

$$\tilde{\pi}_R(\tilde{\boldsymbol{\theta}}) \prod_{i=2}^p \frac{1}{|\mathbf{J}_i|} = \prod_{i=1}^p \frac{1}{\psi_{ii}^i} = \pi_H(\mathbf{\Psi}).$$

Proof of Fact 4. Note that (25) is equivalent to

$$\left\{ \begin{array}{lcl} d_1 & = & \xi_1^{\frac{1}{2}} \xi_2^{\frac{1}{3}} \dots\dots \xi_{p-2}^{\frac{1}{p-1}} (\xi_{p-1} \xi_p)^{\frac{1}{p}}, \\ d_2 & = & \xi_1^{-\frac{1}{2}} \xi_2^{\frac{1}{3}} \dots\dots \xi_{p-2}^{\frac{1}{p-1}} (\xi_{p-1} \xi_p)^{\frac{1}{p}}, \\ d_3 & = & \xi_2^{-\frac{2}{3}} \dots\dots \xi_{p-2}^{\frac{1}{p-1}} (\xi_{p-1} \xi_p)^{\frac{1}{p}}, \\ \dots & & \dots\dots\dots \\ d_{p-1} & = & \xi_{p-2}^{-\frac{p-2}{p-1}} (\xi_{p-1} \xi_p)^{\frac{1}{p}}, \\ d_p & = & \xi_{p-1}^{-\frac{p-2}{p}} \xi_p^{\frac{1}{p}}. \end{array} \right.$$

Then, the Hessian is

$$\begin{aligned}
H &= \frac{\partial(d_1, \dots, d_p)}{\partial(\xi_1, \dots, \xi_p)} \\
&= \begin{pmatrix} \frac{d_1}{2\xi_1} & \frac{d_1}{3\xi_2} & \frac{d_1}{4\xi_3} & \cdots & \frac{d_1}{(p-1)\xi_{p-2}} & \frac{d_1}{p\xi_{p-1}} & \frac{d_1}{p\xi_p} \\ -\frac{d_2}{2\xi_1} & \frac{d_2}{3\xi_2} & \frac{d_2}{4\xi_3} & \cdots & \frac{d_2}{(p-1)\xi_{p-2}} & \frac{d_2}{p\xi_{p-1}} & \frac{d_2}{p\xi_p} \\ 0 & \frac{d_3}{3\xi_2} & \frac{d_3}{4\xi_3} & \cdots & \frac{d_3}{(p-1)\xi_{p-2}} & \frac{d_3}{p\xi_{p-1}} & \frac{d_3}{p\xi_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{(p-2)d_{p-1}}{(p-1)\xi_{p-2}} & \frac{d_{p-1}}{p\xi_{p-1}} & \frac{d_{p-1}}{p\xi_p} \\ 0 & 0 & 0 & \cdots & 0 & -\frac{(p-1)d_{p-1}}{p\xi_{p-1}} & \frac{d_{p-1}}{p\xi_p} \end{pmatrix} \\
&= DQ\Xi,
\end{aligned}$$

where $\Xi = \text{diag}(\xi_1, \dots, \xi_p)$ and

$$Q = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{p-1} & \frac{1}{p} & \frac{1}{p} \\ -\frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{p-1}{p-1} & \frac{1}{p} & \frac{1}{p} \\ 0 & -\frac{2}{3} & \frac{1}{4} & \cdots & \frac{1}{p-1} & \frac{1}{p} & \frac{1}{p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{p-2}{p-1} & \frac{1}{p} & \frac{1}{p} \\ 0 & 0 & 0 & \cdots & 0 & -\frac{p-1}{p} & \frac{1}{p} \end{pmatrix}.$$

Note that the Fisher information matrix for (d_1, \dots, d_p) is $(\mathbf{D}^2)^{-1}$. The Fisher information matrix for $(\xi \cdots, \xi_p)$ is then

$$H'D^{-2}H = \Xi'Q'DD^{-1}D^{-1}DQ\Xi = \Xi'Q'Q\Xi.$$

It is easy to verify that

$$Q'Q = \text{diag}\left(\frac{1}{2}, \frac{2}{3}, \dots, \frac{p-1}{p}, \frac{1}{p}\right).$$

We have that

$$H'D^{-2}H = diag\left(\frac{1}{2\xi_1^2}, \frac{2}{3\xi_2^2}, \dots, \frac{p-1}{p\xi_{p-1}^2}, \frac{1}{p\xi_p^2}\right).$$

This proves part (a). Parts (b) and (c) are immediate.

Proof of Fact 5. We have

$$\begin{aligned} M &\equiv \int L(\boldsymbol{\mu}, \boldsymbol{\Psi}) \pi_{\mathbf{a}}(\boldsymbol{\mu}, \boldsymbol{\Psi}) d\boldsymbol{\mu} d\boldsymbol{\Psi} \\ &= \int \frac{\prod_{i=1}^p (\psi_{ii}^2)^{(n-a_i-1)/2}}{(2\pi)^{(n-1)p/2} n^{p/2}} \exp\left(-\frac{1}{2} \text{tr}(\boldsymbol{\Psi} \mathbf{S} \boldsymbol{\Psi}')\right) d\boldsymbol{\Psi}. \end{aligned} \quad (86)$$

Note that $w_i = s_{ii} - \mathbf{s}_{i,i-1}' \mathbf{S}_{i-1}^{-1} \mathbf{s}_{i,i-1} > 0$ for $i = 2, \dots, p$. Also let $\mathbf{g}_i = -\psi_{ii} \mathbf{S}_{i-1}^{-1} \mathbf{s}_{i,i-1}$. We then have a recursive formula,

$$\begin{aligned} \text{tr}(\boldsymbol{\Psi}_p \mathbf{S}_p \boldsymbol{\Psi}_p') &= \text{tr}(\boldsymbol{\Psi}_{p-1} \mathbf{S}_{p-1} \boldsymbol{\Psi}_{p-1}') + (\psi_{p,p-1} - \mathbf{g}_p)' \mathbf{S}_{p-1} (\psi_{p,p-1} - \mathbf{g}_p) \\ &= \sum_{i=1}^p \psi_{ii}^2 w_i + \sum_{i=2}^p (\psi_{i,i-1} - \mathbf{g}_i)' \mathbf{S}_{i-1} (\psi_{i,i-1} - \mathbf{g}_i). \end{aligned}$$

Then

$$M = \int \frac{\prod_{i=1}^p (\psi_{ii}^2)^{(n-a_i-1)/2}}{(2\pi)^{(n-1)p/2 - (p-1)p/2} n^{p/2} \prod_{j=1}^{p-1} |\mathbf{S}_j|^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^p \psi_{ii}^2 w_i\right) \prod_{i=1}^p d\psi_{ii}. \quad (87)$$

Let $\delta_i = \psi_{ii}^2$. The right hand side of (87) is equal to

$$\begin{aligned} &\frac{\prod_{j=1}^{p-1} |\mathbf{S}_j|^{-1/2}}{2^p (2\pi)^{(n-p)p/2} n^{p/2}} \prod_{i=1}^p \int_0^\infty \delta_i^{(n-a_i)/2-1} \exp\left(-\frac{w_i}{2} \delta_i\right) d\delta_i \\ &= \frac{\prod_{i=1}^p \Gamma(\frac{1}{2}(n-a_i)) 2^{(n-a_i)/2}}{2^p (2\pi)^{(n-p)p/2} n^{p/2}} \prod_{j=1}^{p-1} |\mathbf{S}_j|^{-1/2} \frac{1}{s_{11}^{(n-a_1)/2}} \prod_{i=2}^p \left(\frac{|\mathbf{S}_{i-1}|}{|\mathbf{S}_i|}\right)^{(n-a_i)/2}. \end{aligned}$$

The fact holds.

REFERENCES

- Bayarri, M. J. (1981), ‘Inferencia bayesiana sobre el coeficiente de correlaci n de una poblaci n normal bivalente’, *Trabajos de Estadistica e Investigacion Operativa* **32**, 18–31.
- Berger, J. O. & Bernardo, J. M. (1992a), On the development of reference priors (with discussion), in ‘Bayesian Statistics 4’, Oxford Univ. Press: London, pp. 35–60.
- Berger, J. O. & Bernardo, J. M. (1992b), Reference priors in a variance components problem, in ‘Bayesian analysis in statistics and econometrics’, pp. 177–194.
- Berger, J. O. & Sun, D. (2006), ‘Objective priors for a bivariate normal model’.

- Brillinger, D. R. (1962), ‘Examples bearing on the definition of fiducial probability with a bibliography’, *Annals of Mathematical Statistics* **33**, 1349–1355.
- Brown, P. J. (2001), The generalized inverted wishart distribution, in ‘Encyclopedia of Environmetrics’.
- Chang, T. & Eaves, D. (1990), ‘Reference priors for the orbit in a group model’, *Annals of Statistics* **18**, 1595–1614.
- Daniels, M. & Kass, R. (1999), ‘Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models’, *Journal of the American Statistical Association* **94**, 1254–1263.
- Daniels, M. & Pourahmadi, M. (2002), ‘Bayesian analysis of covariance matrices and dynamic models for longitudinal data’, *Biometrika* **89**, 553–566.
- Dawid, A. P., Stone, M. & Zidek, J. V. (1973), ‘Marginalization paradoxes in Bayesian and structural inference (with discussion)’, *Journal of the Royal Statistical Society, Series B* **35**, 189–233.
- Dey, D. & Srinivasan, C. (1985), ‘Estimation of a covariance matrix under stein’s loss’, *Annals of Statistics* **13**, 1581–1591.
- Eaton, M. L. (1989), *Group invariance applications in statistics*, Institute of Mathematical Statistics.
- Eaton, M. L. & Sudderth, W. (2002), ‘Group invariant inference and right haar measure’, *Journal of Statistical Planning and Inference* **103**, 87–99.
- Geisser, S. & Cornfield, J. (1963), ‘Posterior distributions for multivariate normal parameters’, *J. Roy. Statist. Soc. B* **25**, 368–376.
- Haff, L. (1977), ‘Minimax estimators for a multivariate precision matrix’, *J. Multivariate Analysis* **7**, 374–385.
- James, W. & Stein, C. (1961), Estimation with quadratic loss, in ‘Proc Fourth Berkely Symp. Math. Statist. Probability, 1’, University of California Press, pp. 361–380.
- Jeffreys, H. (1961), *Theory of Probability*, Oxford University Press, London.
- Liechty, J., Liechty, M. & Müller, P. (2004), ‘Bayesian correlation estimation’, *Biometrika* **91**, 1–14.
- Lindley, D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press: Cambridge.

- Pourahmadi, M. (1999), ‘Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation’, *Biometrika* **86**, 677–690.
- Roverato, A. & Consonni, G. (2004), ‘Compatible prior distributions for dag models’, *Journal of the Royal Statistical Society, Series B, Methodological* **66**, 47–61.
- Stein, C. (1956), Some problems in multivariate analysis. part i., Technical Report 6, Department of Statistics, Stanford University.
- Yang, R. & Berger, J. O. (1994), ‘Estimation of a covariance matrix using the reference prior’, *Annals of Statistics* **22**, 1195–1211.