The Annals of Statistics
2008, Vol. 36, No. 2, 963–982
DOI: 10.1214/07-AOS501

OBJECTIVE PRIORS FOR THE BIVARIATE NORMAL MODEL

BY JAMES O. BERGER[1] AND DONGCHU SUN[2]

Duke University and University of Missouri-Columbia

Study of the bivariate normal distribution raises the full range of issues involving objective Bayesian inference, including the different types of objective priors (e.g., Jeffreys, invariant, reference, matching), the different modes of inference (e.g., Bayesian, frequentist, fiducial) and the criteria involved in deciding on optimal objective priors (e.g., ease of computation, frequentist performance, marginalization paradoxes). Summary recommendations as to optimal objective priors are made for a variety of inferences involving the bivariate normal distribution.

In the course of the investigation, 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, including the correlation coefficient.

1. Introduction and prior distributions.

1.1. *Notation and problem statement.* The bivariate normal distribution of $(x_1, x_2)'$ has mean parameters $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where ρ is the correlation between x_1 and x_2. The density is

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \exp\left\{-\frac{\sigma_2^2(x_1-\mu_1)^2 + \sigma_1^2(x_2-\mu_2)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2)}{2\sigma_1^2\sigma_2^2(1-\rho^2)}\right\}.$$

The data consists of an independent random sample $\mathbf{X} = (\mathbf{x}_k = (x_{1k}, x_{2k}), k = 1, \ldots, n)$ of size $n \geq 3$, for which the sufficient statistics are

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \sum_{k=1}^{n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})' = \begin{pmatrix} s_{11} & r\sqrt{s_{11}s_{22}} \\ r\sqrt{s_{11}s_{22}} & s_{22} \end{pmatrix}, \tag{1}$$

[1]Supported in part by NSF Grant DMS-01-03265.
[2]Supported in part by NSF Grant SES-03-51523 and NIH Grant R01-MH071418.

AMS 2000 subject classifications. Primary 62F10, 62F15, 62F25; secondary 62A01, 62E15, 62H10, 62H20.

Key words and phrases. Reference priors, matching priors, Jeffreys priors, right-Haar prior, fiducial inference, frequentist coverage, marginalization paradox, rejection sampling, constructive posterior distributions.

where, for $i, j = 1, 2$,

$$\overline{x}_i = n^{-1}\sum_{j=1}^{n} x_{ij}, \qquad s_{ij} = \sum_{k=1}^{n}(x_{ik}-\overline{x}_i)(x_{jk}-\overline{x}_j) \quad \text{and} \quad r = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}.$$

We will denote prior densities as $\pi(\mu_1, \mu_2, \sigma_1\ \sigma_2,\ \rho)$, and the corresponding posterior densities as $\pi(\mu_1, \mu_2, \sigma_1\ \sigma_2,\ \rho \mid \mathbf{X})$ (all with respect to $d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2\, d\rho$).

We consider 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. Section 1.2 introduces many of the key issues to be covered, through a summary of some of the most interesting results involving priors yielding exact frequentist procedures; this section also raises interesting historical and philosophical issues. For easy access, Section 1.3 presents our summary recommendations as to which priors to utilize.

Often, the posteriors for the recommended priors are essentially available in computational closed form, allowing direct Monte Carlo simulation. Section 2 provides simple accept-reject schemes for computing with the recommended priors in other cases. Sections 3 and 4 develop the needed theory, concerning what are called reference priors and matching priors, respectively, and also present various simulations that were conducted to enable summary recommendations to be made.

Notation: In addition to $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, the following parameters will be considered:

$$\eta_1 = \frac{1}{\sigma_1}, \qquad \eta_2 = \frac{1}{\sigma_2\sqrt{1-\rho^2}}, \qquad \eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}, \tag{2}$$

$$\begin{aligned} &\theta_1 = \frac{\rho\sigma_2}{\sigma_1}, \qquad \theta_2 = \sigma_2^2(1-\rho^2), \qquad \theta_3 \equiv |\mathbf{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2),\\ &\theta_4 = \frac{\sigma_2\sqrt{1-\rho^2}}{\sigma_1}, \end{aligned} \tag{3}$$

$$\begin{aligned} &\theta_5 = \frac{\mu_1}{\sigma_1}, \qquad \theta_6 = \sigma_1^2\sigma_2^2, \qquad \theta_7 = \frac{\sigma_2}{\sigma_1}, \qquad \theta_8 = \frac{\mu_2}{\sigma_2},\\ &\theta_9 \equiv \sigma_{12} = \rho\sigma_1\sigma_2, \end{aligned} \tag{4}$$

$$\theta_{10} = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2, \tag{5}$$

$$\theta_{11} = \mathbf{d}'\mathbf{\Sigma}\mathbf{d} \qquad [\mathbf{d}' = (d_1, d_2) \text{ not proportional to } (0, 1)], \tag{6}$$

$$\lambda_1 = ch_{\max}(\mathbf{\Sigma}), \qquad \lambda_2 = ch_{\min}(\mathbf{\Sigma}). \tag{7}$$

Some of these parameters have straightforward statistical interpretations. Since $(x_2 \mid x_1, \boldsymbol{\mu}, \mathbf{\Sigma}) \sim N(\mu_2 + \theta_1(x_1 - \mu_1), \theta_2)$, it is clear that θ_1 is a regression coefficient, θ_2 is a conditional variance, and η_2^2 is the corresponding precision. For the marginal distribution of x_1, η_1^2 is the precision and θ_5 is the reciprocal of the coefficient of variation. θ_3 is usually called the generalized variance. (η_1, η_2, η_3) gives a type of Cholesky decomposition of the precision matrix $\mathbf{\Sigma}^{-1}$ [see (13) in Section 2.1]. θ_{10} is the variance of $x_1 - x_2$, and θ_{11} is the variance of $d_1x_1 + d_2x_2$. Finally, λ_1 and λ_2 are the largest and smallest eigenvalues of $\mathbf{\Sigma}$.

Technical issue. We will assume that $|\rho| < 1$ and $|r| < 1$ in virtually all expressions and results that follow. This is because, if either equals 1 in absolute value, then $\rho = \{\text{sign of } r\}$ with probability 1 (either frequentist or Bayesian posterior, as relevant). Indeed, the situation then essentially collapses to the univariate version of the problem, which is standard.

1.2. *Matching, constructive posteriors and fiducial distributions.* The bivariate normal distribution has been extensively studied from frequentist, fiducial and objective Bayesian perspectives. Table 1 summarizes a number of interesting results.

- For a variety of parameters, it presents objective priors (discussed below) for which the resulting Bayesian posterior credible sets of level $1-\alpha$ are also exact frequentist confidence sets at the same level; in this case, the priors are said to be *exact frequentist matching*. This is a very desirable situation: see [23] and [2] for general discussion and the many earlier references.
- For $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ, the constructive posterior distributions are also the fiducial distributions for the parameters, as found in Fisher [14, 15] and [21].
- Posterior distributions are presented as *constructive random distributions*, that is, by a description of how to simulate from them. Thus to simulate from the posterior distribution of σ_1, given the data (actually, only s_{11} is needed), one draws independent χ^2_{n-1} random variables and simply computes the corresponding $\sqrt{s_{11}/\chi^2_{n-1}}$; this yields an independent sample from the fiducial/posterior distribution of σ_1.

Table 1 also lists the objective prior distributions that yield the indicated objective posterior. The notation π_{ab} in the table stands for the important class of prior densities (a subclass of the *generalized Wishart distributions* of [8])

$$\pi_{ab}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{\sigma_1^{3-a}\sigma_2^{2-b}(1-\rho^2)^{2-b/2}}. \tag{8}$$

Special cases of this class are the *Jeffreys-rule* prior $\pi_J = \pi_{10}$, the *right-Haar* prior $\pi_H = \pi_{12}$, the *independence Jeffreys* prior $\pi_{IJ} = \pi_{21} = \sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-3/2}$ and π_{RO} which has $a = b = 1$. The independence Jeffreys prior follows from using a constant prior for the means, and then the Jeffreys prior for the covariance matrix with means given.

We highlight the results about ρ in Table 1 because they are interesting from practical, historical and philosphical perspectives. First, it does not seem to be

TABLE 1

Parameters with exact matching priors of the form π_{ab}, and associated constructive posteriors: Here Z^* is a standard normal random variable, and χ^{2*}_{n-1} and χ^{2*}_{n-2} are chi-squared random variables with the indicated degrees of freedom, all random variables being independent. For $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ, the indicated posteriors are also fiducial distributions

Parameter	**Prior**	**Posterior**
μ_1	$\pi_{1b}, \forall b$ (including π_J and π_H)	$\bar{x}_1 + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{s_{11}}{n}}$
μ_2	$\pi_J = \pi_{10}$	$\bar{x}_2 + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{s_{22}}{n}}$
$\mathbf{d}'\binom{\mu_1}{\mu_2}$, $\mathbf{d} \in \mathbb{R}^2$	$\pi_J = \pi_{10}$ and π_{H^*} (see Table 4)	$\mathbf{d}'(\bar{x}_1, \bar{x}_2)' + \frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}\sqrt{\frac{\mathbf{d}'\mathbf{S}\mathbf{d}}{n}}$
σ_1	$\pi_{1b}, \forall b$ (including π_J and π_H)	$\sqrt{\frac{s_{11}}{\chi^{2*}_{n-1}}}$
ρ	$\pi_H = \pi_{12}$	$\psi(\frac{-Z^*}{\sqrt{\chi^{2*}_{n-1}}} + \frac{\sqrt{\chi^{2*}_{n-2}}}{\sqrt{\chi^{2*}_{n-1}}}\frac{r}{\sqrt{1-r^2}})$ $\psi(y) = y/\sqrt{1+y^2}$
$\eta_3 = -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}}$	$\pi_{a2}, \forall a$ (including π_H)	$\frac{Z^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi^{2*}_{n-2}}}{\sqrt{s_{11}}}\frac{r}{\sqrt{1-r^2}}$
$\theta_1 = \frac{\rho\sigma_2}{\sigma_1}$	$\pi_{a2}, \forall a$ (including π_H)	$\frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}} - \frac{Z^*}{\sqrt{\chi^{2*}_{n-2}}}\frac{\sqrt{1-r^2}\sqrt{s_{22}}}{\sqrt{s_{11}}}$
$\theta_2 = \sigma_2^2(1-\rho^2)$	$\pi_{a2}, \forall a$ (including π_H)	$\frac{s_{22}(1-r^2)}{\chi^{2*}_{n-2}}$
$\theta_3 = \lvert\mathbf{\Sigma}\rvert$	$\pi_H = \pi_{12}$ and $\pi_{IJ} = \pi_{21}$	$\frac{\lvert\mathbf{S}\rvert}{\chi^{2*}_{n-1}\chi^{2*}_{n-2}}$
$\theta_4 = \frac{\sigma_2\sqrt{1-\rho^2}}{\sigma_1}$	$\pi_H = \pi_{12}$	$\frac{\sqrt{\chi^{2*}_{n-1}}}{\sqrt{\chi^{2*}_{n-2}}}\frac{\sqrt{s_{22}(1-r^2)}}{\sqrt{s_{11}}}$
$\theta_5 = \frac{\mu_1}{\sigma_1}$	$\pi_{1b}, \forall b$ (including π_J and π_H)	$\frac{Z^*}{\sqrt{n}} + \frac{\bar{x}_1\sqrt{\chi^{2*}_{n-1}}}{\sqrt{s_{11}}}$
$\mathbf{d}'\mathbf{\Sigma}\mathbf{d}$	$\pi_J = \pi_{10}$ and π_{H^*} (see Table 4)	$\sqrt{\frac{\mathbf{d}'\mathbf{S}\mathbf{d}}{\chi^{2*}_{n-1}}}$

known that the indicated prior for ρ is exact frequentist matching (proved here in Theorem 2). Indeed, standard statistical software utilizes various approximations to arrive at frequentist confidence sets for ρ, missing the fact that a simple exact confidence set exists, even for $n = 3$. It was, of course, known that exact frequentist confidence procedures could be constructed (cf. Exercise 54, Chapter 6 of [18]), but explicit expressions do not seem to be available.

The historically interesting aspect of this posterior for ρ is that it is also the fiducial distribution of ρ. Geisser and Cornfield [16] studied the question of whether the fiducial distribution of ρ could be reproduced as an objective Bayesian posterior, and they concluded that this was most likely not possible. The strongest evidence for this arose from Brillinger [7], which used results from [19] and a difficult analytic argument to show that there does not exist a prior $\pi(\rho)$ such that the fiducial density of ρ equals $f(r \mid \rho)\pi(\rho)$, where $f(r \mid \rho)$ is the density of r given ρ. Since the fiducial distribution of ρ only depends on r, it was certainly reasonable to speculate that if it were not possible to derive this distribution from the density of r and a prior, then it would not be possible to do so in general. The above result, of course, shows that this speculation was incorrect.

The philosophically interesting aspect of this situation is that Brillinger's result does show that the fiducial/posterior distribution for ρ provides another example of the marginalization paradox ([13]). This leads to an interesting philosophical conundrum of a type that we have not previously seen: a complete fiducial/objective Bayesian/frequentist unification can be obtained for inference about ρ, but only if violation of the marginalization paradox is accepted. We will shortly introduce a prior distribution that avoids the marginalization paradox for ρ, but which is not exactly frequentist matching. We 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. (Note that the same conundrum also arises for $\theta_5 = \mu_1/\sigma_1$; the exact frequentist matching prior results in a marginalization paradox, as shown in [24].) Some interesting examples of improper priors resulting in marginalization paradox can be found from Ghosh and Yang [17] and Datta and Ghosh [10, 11].

1.3. *Recommended priors.* It is actually rare to have exact matching priors for parameters of interest. Also, one is often interested in very complex functions of parameters (e.g., predictive distributions) and/or joint distributions of parameters. For such problems it is important to have a general objective prior that seems to perform reasonably well for all quantities of interest. Furthermore, it is unappealing to many Bayesians to change the prior according to which parameter is declared to be of interest, and an objective prior that performs well overall is often sought.

The five priors we recommend for various purposes are π_J, π_H,

$$\pi_{R\rho} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}, \qquad \pi_{R\sigma} \propto \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)} \tag{9}$$

and

$$\pi_{R\lambda} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{(\sigma_1/\sigma_2-\sigma_2/\sigma_1)^2+4\rho^2}}. \tag{10}$$

The first prior in (9) was developed in [20] and was studied extensively in [1], where it was shown to be a one-at-a-time reference prior (see Section 3). The second prior in (9) is new and is derived in Section 3. $\pi_{R\lambda}$ was developed as a one-at-a-time reference prior in [25].

With these definitions, we can make our summary recommendations. Table 2 gives the four objective priors that are recommended for use, and indicates for

TABLE 2

Recommendations of objective priors for various parameters in the bivariate normal model: $\square$ *indicates that the posterior will not be exact frequentist matching*. (*For μ_2 and parameters with σ_1 replaced by σ_2, use the right-Haar prior with the variances interchanged.*)

Prior	**Parameter**
$\pi_{R\rho}$	$\boxed{\rho}$, $\boxed{\frac{\sigma_1}{\sigma_2}}$, $\boxed{\textbf{general use}}$
π_H	$\mu_1, \sigma_1, \rho, \eta_3, \frac{\rho\sigma_2}{\sigma_1}, \sigma_2^2(1-\rho^2), \lvert\mathbf{\Sigma}\rvert, \frac{\sigma_2}{\sigma_1}\sqrt{1-\rho^2}, \frac{\mu_1}{\sigma_1}$
$\tilde{\pi}_H$ (see Table 4)	$\mathbf{d}'(\mu_1, \mu_2)'$, $\mathbf{d}'\mathbf{\Sigma}\mathbf{d}$
$\pi_{R\lambda}$	$\boxed{ch_{\max}(\mathbf{\Sigma})}$
$\pi_{R\sigma}$	$\boxed{\sigma_{12} = \rho\sigma_1\sigma_2}$

which parameters (or functions thereof) they are recommended. These recommendations are based on three criteria: (i) the degree of frequentist matching, discussed in Section 4; (ii) being a one-at-a-time reference prior, discussed in Section 3; and (iii) ease of computation. The rationale for each of the entries in the table, based on these criteria, is given in Section 4.5.

Another commonly used prior is the "scale prior," $\pi_S \propto (\sigma_1\sigma_2)^{-1}$. The motivation that is often given for this prior is that it is "standard" to use σ_i^{-1} as the prior for a standard deviation σ_i, while $-1 < \rho < 1$ is on a bounded set and so one can use a constant prior in ρ. We do not recommend this prior, but do consider its performance in Section 4.5.

2. Computation. In this paper, a constant prior is always used for (μ_1, μ_2), so that

$$\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} \Big| \mathbf{\Sigma}, \mathbf{X}\right) \sim N_2\left(\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\end{pmatrix}, n^{-1}\mathbf{\Sigma}\right). \tag{11}$$

Generation from this conditional posterior distribution is standard, so the challenge of simulation from the posterior distribution requires only sampling from $(\sigma_1, \sigma_2, \rho \mid \mathbf{X})$.

The marginal likelihood of $(\sigma_1, \sigma_2, \rho)$ satisfies

$$L_1(\sigma_1, \sigma_2, \rho) \propto \frac{1}{|\mathbf{\Sigma}|^{(n-1)/2}} \exp\left(-\frac{1}{2}\operatorname{trace}(\mathbf{S}\mathbf{\Sigma}^{-1})\right). \tag{12}$$

It is immediate that, under the priors π_J and π_{IJ}, the marginal posteriors of $\mathbf{\Sigma}$ are Inverse Wishart $(\mathbf{S}^{-1}, n)$ and Inverse Wishart $(\mathbf{S}^{-1}, n-1)$, respectively.

Berger, Strawderman and Tang [4] gave a Metropolis–Hastings algorithm to generate from $(\sigma_1, \sigma_2, \rho \mid \mathbf{X})$ based on the prior $\pi_{R\lambda}$. The following sections deal with the other priors we consider.

TABLE 3

Ratio π/π_{IJ}, upper bound M, rejection step and acceptance probability for $\rho = 0.80, 0.95, 0.99$, when $\pi = \pi_{R\rho}, \pi_{R\sigma}, \tilde{\pi}_{R\sigma}, \pi_S$ and π_{MS}

		Bound		**Acceptance probability**		
Prior	**Ratio** $\frac{\pi}{\pi_{IJ}}$	M	***Rejection Step***	$\rho = 0.80$	$\rho = 0.95$	$\rho = 0.99$
$\pi_{R\rho}$	$\sqrt{1-\rho^2}$	1	$u \le \sqrt{1-\rho^2}$	0.6000	0.3122	0.1410
$\pi_{R\sigma}$	$\sqrt{1-\rho^4}$	1	$u \le \sqrt{1-\rho^4}$	0.7684	0.4307	0.1985
$\tilde{\pi}_{R\sigma}$	$\sqrt{\frac{1-\rho^2}{2-\rho^2}}$	$\frac{1}{\sqrt{2}}$	$u \le \sqrt{\frac{2(1-\rho^2)}{2-\rho^2}}$	0.7276	0.4215	0.1975
π_S	$(1-\rho^2)^{3/2}$	1	$u \le (1-\rho^2)^{3/2}$	0.2160	0.0304	0.0028

2.1. *Marginal posteriors of* $(\sigma_1, \sigma_2, \rho)$ *under* $\pi_{R\rho}$, $\pi_{R\sigma}$, $\tilde{\pi}_{R\sigma}$, *and* π_S. For these priors, an independent sample from $\pi(\sigma_1, \sigma_2, \rho \mid \mathbf{X})$ can be obtained by the following acceptance-rejection algorithm:

Simulation step. Generate $(\sigma_1, \sigma_2, \rho)$ from the independence Jeffreys posterior $\pi_{IJ}(\sigma_1, \sigma_2, \rho \mid \mathbf{X})$ [the Inverse Wishart $(\mathbf{S}^{-1}, n-1)$ distribution] and, independently, sample $u \sim \text{Uniform}(0, 1)$.

Rejection step. Suppose $M \equiv \sup_{(\sigma_1,\sigma_2,\rho)} \frac{\pi(\sigma_1,\sigma_2,\rho)}{\pi_{IJ}(\sigma_1,\sigma_2,\rho)} < \infty$. If $u \le \pi(\sigma_1, \sigma_2, \rho)/[M\pi_{IJ}(\sigma_1, \sigma_2, \rho)]$, accept $(\sigma_1, \sigma_2, \rho)$; else, return to *Simulation step*.

For each of the priors listed in Table 3, the key ratio, π/π_{IJ}, is listed in the table, along with the upper bound M, the *Rejection step* and the resulting acceptance probability for $\rho = 0.80, 0.95, 0.99$. The rejection algorithm is quite efficient for sampling these posteriors. Indeed, for $\rho \approx 0$, the algorithms accept with probability near one and, even for large $|\rho|$, the acceptance probabilities are very reasonable for the priors $\pi_{R\rho}$, $\pi_{R\sigma}$, and $\tilde{\pi}_{R\sigma}$. For large $|\rho|$, the algorithm is less efficient for the posteriors under the prior π_S, but even these acceptance rates may well be fine in practice, given the simplicity of the algorithm.

2.2. *Computation under* π_{ab}. The most interesting prior of this form (besides the Jeffreys and independence Jeffreys priors) is the right-Haar prior π_H, although other priors such as π_{11} arise as reference priors, and hence are potentially of interest. While Table 1 gave an explicit form for the most important marginal posteriors arising from priors of this form, it is of considerable interest that essentially closed form generation from the full posterior of any prior of this form is possible (see, e.g., [8]). This is briefly reviewed in this section, since the expressions for the resulting constructive posteriors are needed for later results on frequentist coverage.

It is most convenient to work with the parameters (η_1, η_2, η_3) given in (2). This parameterization gives a type of Cholesky decomposition of the precision matrix $\mathbf{\Sigma}^{-1}$,

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \eta_1 & \eta_3 \\ 0 & \eta_2 \end{pmatrix} \begin{pmatrix} \eta_1 & 0 \\ \eta_3 & \eta_2 \end{pmatrix}, \tag{13}$$

which accounts for the simplicity of ensuing computations. Note that (2) is equivalent to

$$\sigma_1 = \frac{1}{\eta_1}, \qquad \sigma_2 = \frac{\sqrt{\eta_1^2 + \eta_3^2}}{\eta_1 \eta_2}, \qquad \rho = -\frac{\eta_3}{\sqrt{\eta_1^2 + \eta_3^2}}. \tag{14}$$

The prior π_{ab} of (8) for $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ transforms to the extended conjugate class of priors for $(\mu_1, \mu_2, \eta_1, \eta_2, \eta_3)$, given by $\pi_{ab}(\mu_1, \mu_2, \eta_1, \eta_2, \eta_3) = \eta_1^{-a} \eta_2^{-b}$.

LEMMA 1. *Consider the prior π_{ab}.*

(a) *The marginal posterior of η_3 given $(\eta_1, \eta_2; \mathbf{X})$ is $N(-\eta_2 r \sqrt{s_{22}/s_{11}}, 1/s_{11})$.*
(b) *The marginal posterior distributions of η_1 and η_2 are independent and*

$$(\eta_1^2 \mid \mathbf{X}) \sim \text{Gamma}\big(\tfrac{1}{2}(n-a), \tfrac{1}{2} s_{11}\big);$$
$$(\eta_2^2 \mid \mathbf{X}) \sim \text{Gamma}\big(\tfrac{1}{2}(n-b), \tfrac{1}{2} s_{22}(1-r^2)\big).$$

See [5] for a proof of this result. We next present the constructive posteriors of (η_1, η_2, η_3), and from these derive the constructive posteriors of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ and other parameters. All results follow directly from Lemma 1 and (14).

In presenting the constructive posteriors, we will use a star to represent a random draw from the implied distribution; thus μ_1^* will represent a random draw from its posterior distribution, Z_1^*, Z_2^*, Z_3^* will be independent draws from the standard normal distribution, and χ_{n-a}^{2*} and χ_{n-b}^{2*} will be independent draws from chi-squared distributions with the indicated degrees of freedom. Recall that these constructive posteriors are not only useful for simulation, but will be the key to proving exact frequentist matching results.

FACT 1. (a) *The constructive posterior of (η_1, η_2, η_3) given $\mathbf{X}$ can be expressed as*

$$\begin{aligned} \eta_1^* &= \sqrt{\frac{\chi_{n-a}^{2*}}{s_{11}}}, \qquad \eta_2^* = \sqrt{\frac{\chi_{n-b}^{2*}}{s_{22}(1-r^2)}}, \\ \eta_3^* &= \frac{Z_3^*}{\sqrt{s_{11}}} - \frac{\sqrt{\chi_{n-b}^{2*}}}{\sqrt{s_{11}}} \frac{r}{\sqrt{1-r^2}}. \end{aligned} \tag{15}$$

(b) *The constructive posterior of* $(\sigma_1, \sigma_2, \rho)$ *given* $\mathbf{X}$ *can be expressed as*

$$\sigma_1^* = \sqrt{\frac{s_{11}}{\chi_{n-a}^{2*}}}, \tag{16}$$

$$\sigma_2^* = \sqrt{s_{22}(1-r^2)} \sqrt{\frac{1}{\chi_{n-b}^{2*}} + \frac{1}{\chi_{n-a}^{2*}} \left(\frac{Z_3^*}{\sqrt{\chi_{n-b}^{2*}}} - \frac{r}{\sqrt{1-r^2}} \right)^2}, \tag{17}$$

$$\rho^* = \psi(Y^*), \qquad Y^* = -\frac{Z_3^*}{\sqrt{\chi_{n-a}^{2*}}} + \frac{\sqrt{\chi_{n-b}^{2*}}}{\sqrt{\chi_{n-a}^{2*}}} \frac{r}{\sqrt{1-r^2}}, \tag{18}$$

where $\psi(x) = x/\sqrt{1+x^2}$.

(c) *The constructive posterior for* μ_1 *and* μ_2 *can be written*

$$\mu_1^* = \overline{x}_1 + \frac{Z_1^*}{\sqrt{\chi_{n-a}^{2*}}} \sqrt{\frac{s_{11}}{n}}, \tag{19}$$

$$\mu_2^* = \overline{x}_2 + \frac{Z_1^*}{\sqrt{\chi_{n-a}^{2*}}} \frac{r\sqrt{s_{22}}}{\sqrt{n}} + \left(\frac{Z_2^*}{\sqrt{\chi_{n-b}^{2*}}} - \frac{Z_3^*}{\sqrt{\chi_{n-b}^{2*}}} \frac{Z_1^*}{\sqrt{\chi_{n-a}^{2*}}} \right) \sqrt{\frac{s_{22}(1-r^2)}{n}}. \tag{20}$$

3. Reference priors. This paper began with an effort to derive and catalogue the possible reference priors for the bivariate normal distribution. The reference prior theory (cf. Bernardo [6] and Berger and Bernardo [3]) has arguably been the most successful technique for deriving objective priors. Reference priors depend on (i) specification of a parameter of interest; (ii) specification of nuisance parameters; (iii) specification of a grouping of parameters; and (iv) ordering of the groupings. These are all conveyed by the shorthand notation used in Table 4. Thus, $\{(\mu_1, \mu_2), (\sigma_1, \sigma_2, \rho)\}$ indicates that (μ_1, μ_2) is the parameter of interest, with the others being nuisance parameters, and there are two groupings with the indicated ordering. (The resulting reference prior is the *independence Jeffreys* prior, π_{IJ}.) As another example, $\{\lambda_1, \lambda_2, \vartheta, \mu_1, \mu_2\}$ introduces the eigenvalues $\lambda_1 > \lambda_2$ of $\mathbf{\Sigma}$ as being primarily of interest, with ϑ (the angle defining the orthogonal matrix that diagonalizes $\mathbf{\Sigma}$), μ_1 and μ_2 being the nuisance parameters.

Based on experience with numerous examples, the reference priors that are typically judged to be best are one-at-a-time reference priors, in which each parameter is listed separately as its own group. Hence we will focus on these priors. It turns out to be the case that, for the one-at-a-time reference priors, the ordering of μ_1 and μ_2 among the variables is irrelevant. Hence if μ_1 and μ_2 are omitted from a listing in Table 4, the resulting reference prior is to be viewed as any one-at-a-time reference prior with the indicated ordering of other variables, with the μ_i being inserted anywhere in the ordering.

TABLE 4
Reference priors for the bivariate normal model (*where* $\tilde{\mu}_1 = \mathbf{d}'(\mu_1, \mu_2)'$, $(\tilde{\sigma}_1)^2 = \theta_7$, $\tilde{\rho} = \mathbf{d}'\mathbf{\Sigma}(0, 1)'/(\sigma_1\sqrt{\theta_7})$, $\tilde{\theta}_2 = \sigma_2^2[1 - (\tilde{\rho})^2]$ *and* $\tilde{\theta}_1 = \tilde{\rho}\sigma_2/\tilde{\sigma}_1$); $\{\{\ \}\}$ *indicates that any ordering of the parameters yields the same reference prior*

Prior $\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$	**For parameter ordering**	**Has form (8) with**
$\pi_J \propto \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}$	$\{(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\}$	$(a, b) = (1, 0)$
$\pi_{IJ} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}$	$\{(\mu_1, \mu_2), (\sigma_1, \sigma_2, \rho)\}$	$(a, b) = (2, 1)$
$\pi_{R\rho} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}$	$\{\rho, \sigma_1, \sigma_2\}$, $\{\theta_7, \theta_6, \rho\}$	
$\pi_{R\sigma} \propto \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)}$	$\{\sigma_1, \sigma_2, \rho\}$	
$\tilde{\pi}_{R\sigma} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{2-\rho^2}}$	$\{\sigma_1, \rho, \sigma_2\}$	
	$\{\sigma_1, \eta_3, \theta_2\}$	
$\pi_{RO} \propto \frac{1}{\sigma_1^2\sigma_2(1-\rho^2)^{3/2}}$	$\{\sigma_1, \theta_2, \eta_3\}$	$(a, b) = (1, 1)$
$\pi_{R\lambda} \propto \frac{[((\sigma_1/\sigma_2)-(\sigma_2/\sigma_1))^2+4\rho^2]^{-1/2}}{\sigma_1\sigma_2(1-\rho^2)}$	$\{\lambda_1, \lambda_2, \vartheta\}$	
$\pi_H \propto \frac{1}{\sigma_1^2(1-\rho^2)}$	$\{\{\sigma_1, \theta_1, \theta_2\}\}$, $\{\{\theta_1, \theta_3, \theta_4\}\}$	$(a, b) = (1, 2)$
	$\{\{\eta_1, \eta_2, \theta_1\}\}$, $\{\{\eta_1, \theta_1, \theta_2\}\}$	
$\tilde{\pi}_H \propto \frac{d\tilde{\mu}_1 d\mu_2 d\tilde{\sigma}_1 d\sigma_2 d\tilde{\rho}}{(\tilde{\sigma}_1)^2[1-(\tilde{\rho})^2]}$	$\{\{\mathbf{d}'(\mu_1, \mu_2)', \mu_2, \theta_{11}, \tilde{\theta}_2, \tilde{\theta}_1\}\}$	

We are interested in finding one-at-a-time reference priors for the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \eta_3, \theta_1, \ldots, \theta_9$ and λ_1. This is done in [5], with the results summarized in Table 4, for all these parameters (i.e., the parameter appears as the first entry in the parameter ordering) except η_3, σ_{12}, and μ_i/σ_i; finding one-at-a-time reference priors for these parameters is technically challenging. (We do not explicitly list the reference priors for σ_2 in the table, since they can be found by simply switching with σ_1 in the various expressions.)

4. Comparisons of priors via frequentist matching.

4.1. *Frequentist coverage probabilities and exact matching.* Suppose a posterior distribution is used to create one-sided credible intervals $(\theta_L, \theta_{1-\alpha}(\mathbf{X}))$, where θ_L is the lower limit in the relevant parameter space and $\theta_{1-\alpha}(\mathbf{X})$ is the posterior quantile of the parameter θ of interest, defined by $P(\theta < \theta_{1-\alpha}(\mathbf{X}) \mid \mathbf{X}) = \mathbf{1} - \alpha$. (Here θ is the random variable.) Of interest is the frequentist coverage of the corresponding confidence interval, that is, $C(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = P(\theta < \theta_{1-\alpha}(\mathbf{X}) \mid \boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\sigma_1}, \sigma_2, \rho)$. (Here $\mathbf{X}$ is the random variable.) The closer $C(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is to the nominal $1 - \alpha$, the better the procedure (and corresponding objective prior) is judged to be.

The main results about exact matching are given in Theorems 1 through 8. The proofs of Theorems 1, 2 and 8 are given in Section 5; the rest can be found in [5].

The following technical lemmas will be repeatedly utilized. The first lemma is from (3d.2.8) in [22]. Lemma 3 is easy.

LEMMA 2. *For $n \geq 3$ and given σ_1, σ_2, ρ, the following three random variables are independent and have the indicated distributions*:

$$T_2 = \left[\frac{s_{11}}{\sigma_2^2(1-\rho^2)}\right]^{1/2}\left[\frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}} - \frac{\rho\sigma_2}{\sigma_1}\right] \equiv Z_3 \qquad (\textit{standard normal}), \tag{21}$$

$$T_3 = \frac{s_{22}(1-r^2)}{\sigma_2^2(1-\rho^2)} \equiv \chi^2_{n-2}, \tag{22}$$

$$T_5 = \frac{s_{11}}{\sigma_1^2} \equiv \chi^2_{n-1}. \tag{23}$$

LEMMA 3. *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) *If W is a positive random variable*, $(WY)_{1-\alpha} \geq 0$ *if and only if* $Y_{1-\alpha} \geq 0$.

We will reserve quantile notation for posterior quantiles, with respect to the $*$ distributions. Thus the quantile $[(\sigma_1 Z_3^* - rZ_3)/\chi^2_{n-1} + \rho\sqrt{s_{11}}\chi^{2*}_{n-b}]_{1-\alpha}$ would be computed based on the joint distribution of (Z_3^*, χ^{2*}_{n-b}), while holding $(\sigma_1, \rho, r, s_{11}, Z_3, \chi^2_{n-1})$ fixed.

4.2. *Credible intervals for a class of functions of* $(\sigma_1, \sigma_2, \rho)$. We consider the one-sided credible intervals of σ_1, σ_2 and ρ and some functions of the form

$$\theta = \sigma_1^{d_1}\sigma_2^{d_2} g(\rho), \tag{24}$$

for $d_1, d_2 \in \mathbb{R}$ and some function $g(\cdot)$. We also consider a class of scale-invariant priors

$$\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) \propto \frac{h(\rho)}{\sigma_1^{c_1}\sigma_2^{c_2}}, \tag{25}$$

for some $c_1, c_2 \in \mathbb{R}$ and a positive function h.

THEOREM 1. *Denote the $1-\alpha$ posterior quantile of θ by $\theta_{1-\alpha}(\mathbf{X})$ under the prior* (25). *For any fixed $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, the frequentist coverage of the credible interval $(\theta_L, \theta_{1-\alpha}(\mathbf{X}))$ depends only on ρ. Here θ_L is the lower boundary of the parameter space for θ.*

Note that parameters $\rho, \eta_1, \eta_2, \eta_3, \theta_1, \dots, \theta_4$ are all functions of the form (24). From Theorem 1, under any of the priors $\pi_J, \pi_{IJ}, \pi_{R\sigma}, \pi_{R\rho}, \pi_{RO}, \pi_H, \pi_S$, the frequentist coverage probabilities of credible intervals for any of these parameters will depend only on ρ. We will show that the frequentist coverage probabilities could be exact under the prior π_{ab}. Since $\eta_1(\eta_2)$ is a monotone function of $\sigma_1(\theta_2)$, we consider only ρ and the last 5 parameters.

4.3. *Coverage probabilities under π_{ab}.*

THEOREM 2. (a) *For ψ defined in* (18), *the posterior $1-\alpha$ quantile of ρ is $\rho^*_{1-\alpha} = \psi(Y^*_{1-\alpha})$.* (b) *For any $\alpha \in (0,1)$, $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ and $\rho \in (-1,1)$,*

$$\begin{aligned} &P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \\ &\quad = P\biggl(\frac{\sqrt{1-\rho^2}Z_3 + \rho\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} > \biggl(\frac{\sqrt{1-\rho^2}Z_3^* + \rho\sqrt{\chi^{2*}_{n-a}}}{\sqrt{\chi^{2*}_{n-b}}}\biggr)_\alpha \Bigm| \rho\biggr). \end{aligned} \tag{26}$$

(c) (26) *equals $1-\alpha$ if and only if the right Haar prior is used, that is,* $(a,b) = (1,2)$.

THEOREM 3. (a) *For any $\alpha \in (0,1)$, $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ and $\rho \in (-1,1)$,*

$$\begin{aligned} &P\bigl(\eta_3 < (\eta_3^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho\bigr) \\ &\quad = P\biggl(\frac{Z_3 + \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} < \biggl(\frac{Z_3^* + \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^{2*}_{n-b}}}\biggr)_{1-\alpha} \Bigm| \rho\biggr). \end{aligned} \tag{27}$$

(b) (27) *equals $1-\alpha$ for any $-1 < \rho < 1$ if and only if $b = 2$.*

THEOREM 4. (a) *The constructive posterior of $\theta_1 = \rho\sigma_2/\sigma_1$ has the expression*

$$\theta_1^* = \frac{r\sqrt{s_{22}}}{\sqrt{s_{11}}} - \frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\frac{\sqrt{1-r^2}\sqrt{s_{22}}}{\sqrt{s_{11}}}.$$

(b) *For any $\alpha \in (0,1)$, $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ and $\rho \in (-1,1)$,*

$$P\bigl(\theta_1 < (\theta_1^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho\bigr) = P\biggl(t_{n-2} < \sqrt{\frac{n-2}{n-b}}(t^*_{n-b})_{1-\alpha}\biggr), \tag{28}$$

which does not depend on ρ. Furthermore, (28) *equals $1-\alpha$ if and only if $b = 2$.*

THEOREM 5. (a) *The constructive posterior of $\theta_2 = \sigma_2^2(1-\rho^2)$ is $\theta_2^* = s_{22}(1-r^2)/\chi^{2*}_{n-b}$.*

(b) *For any $\alpha \in (0,1)$, $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ and $\rho \in (-1,1)$,*

$$P\bigl(\theta_2 < (\theta_2^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho\bigr) = P\bigl(\chi^2_{n-2} > (\chi^{2*}_{n-b})_\alpha\bigr), \tag{29}$$

which does not depend on ρ. Furthermore, (29) *equals $1-\alpha$ if and only if $b = 2$.*

THEOREM 6. (a) *The constructive posterior of* $\theta_3 = |\boldsymbol{\Sigma}|$ *is* $\theta_3^* = |\mathbf{S}|/(\chi_{n-a}^{2*}\chi_{n-b}^{2*})$.

(b) *For any* $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ *and* $\rho \in (-1, 1)$,

$$P\big(\theta_3 < (\theta_3^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho\big) = P\big(\chi_{n-1}^2\chi_{n-2}^2 > (\chi_{n-a}^{2*}\chi_{n-b}^{2*})_\alpha\big), \tag{30}$$

which does not depend on ρ. *Furthermore*, (30) *equals* $1-\alpha$ *iff* (a, b) *is* $(1, 2)$ *or* $(2, 1)$.

THEOREM 7. (a) *The constructive posterior of* θ_4 *is*

$$\theta_4^* = \sqrt{\frac{\chi_{n-a}^{2*}}{\chi_{n-b}^{2*}}}\sqrt{\frac{s_{22}(1-r^2)}{s_{11}}}.$$

(b) *For any* $\boldsymbol{\xi} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$ *and* $\rho \in (-1, 1)$,

$$P\big(\theta_4 < (\theta_4^*)_{1-\alpha} \mid \boldsymbol{\xi}, \rho\big) = P\big(\chi_{n-1}^2/\chi_{n-2}^2 < (\chi_{n-a}^{2*}/\chi_{n-b}^{2*})_{1-\alpha}\big), \tag{31}$$

which does not depend on ρ. *Furthermore*, (31) *equals* $1-\alpha$ *iff* $(a, b) = (1, 2)$.

An interesting function of $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ not of the form (24) is $\theta_5 = \mu_1/\sigma_1$.

THEOREM 8. (a) *The constructive posterior of* $\theta_5 = \mu_1/\sigma_1$ *is*

$$\theta_5^* = \frac{Z_1^*}{\sqrt{n}} + \frac{\overline{x}_1}{\sqrt{s_{11}}}\sqrt{\chi_{n-a}^{2*}}.$$

(b) *For any* $\alpha \in (0, 1)$, *the frequentist coverage of the credible interval* $(-\infty, (\theta_5^*)_{1-\alpha})$ *is*

$$\begin{aligned} &P\big(\theta_5 < (\theta_5^*)_{1-\alpha} \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho\big) \\ &\quad = P\left(\frac{Z_1 - \theta_5\sqrt{n}}{\sqrt{\chi_{n-1}^2}} < \left(\frac{Z_1^* - \theta_5\sqrt{n}}{\sqrt{\chi_{n-a}^{2*}}}\right)_{1-\alpha} \Big| \theta_5\right), \end{aligned} \tag{32}$$

which depends on θ_5 *only and equals* $1-\alpha$ *if and only if* $a = 1$.

4.4. *First order asymptotic matching.* Datta and Mukerjee [9] and Datta and Ghosh [12] discuss how to determine first-order matching priors for functions of parameters; these are priors such that the frequentist coverage of a one-sided credible interval is equal to the Bayesian coverage up to a term of order n^{-1}. For each of the nine objective priors π_J, π_{IJ}, $\pi_{R\rho}$, $\tilde{\pi}_{R\sigma}$, π_{RO}, $\pi_{R\lambda}$, π_H, π_S and $\pi_{R\sigma}$, [5] determines if it is a first-order matching prior for each of the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho, \eta_3, \theta_1, \ldots, \theta_{10}$. The results are listed in Table 5. For example, π_J is a first order matching prior for $\mu_1, \mu_2, \sigma_1, \sigma_2$, θ_1, θ_5, θ_7, θ_8, and θ_{10}, but not for η_3, θ_2, θ_3 and θ_9.

TABLE 5

The first-order asymptotic matching of objective priors for $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$, $\mu_1 - \mu_2$, η_3, θ_j, $j = 1, \dots, 10$. Here a boldface letter indicates exact matching

	Asymptotic matching	
Prior $\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$	**Yes**	**No**
$\pi_J \propto \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)^2}$	$\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, \boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}$	ρ
	$\mu_1 - \mu_2, \theta_1, \boldsymbol{\theta_5}, \theta_7, \boldsymbol{\theta_8}, \theta_{10}$	$\eta_3, \theta_2, \theta_3, \theta_9$
$\pi_{IJ} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}$	μ_1, μ_2	σ_1, σ_2, ρ
	$\mu_1 - \mu_2, \theta_1, \boldsymbol{\theta_3}, \theta_7$	$\eta_3, \theta_2, \theta_5, \theta_8, \theta_9, \theta_{10}$
$\pi_{R\rho} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)}$	μ_1, μ_2, ρ	σ_1, σ_2
	$\mu_1 - \mu_2, \theta_3, \theta_7$	$\eta_3, \theta_1, \theta_2, \theta_5, \theta_8, \theta_9, \theta_{10}$
$\tilde{\pi}_{R\sigma} \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)\sqrt{2-\rho^2}}$	μ_1, μ_2	σ_1, σ_2, ρ
	$\mu_1 - \mu_2, \eta_3, \theta_3, \theta_7$	$\theta_1, \theta_2, \theta_5, \theta_8, \theta_9, \theta_{10}$
$\pi_{RO} \propto \frac{1}{\sigma_1^2\sigma_2(1-\rho^2)^{3/2}}$	$\boldsymbol{\mu_1}, \mu_2, \boldsymbol{\sigma_1}$	σ_2, ρ
	$\mu_1 - \mu_2, \theta_1, \boldsymbol{\theta_5}$	$\eta_3, \theta_2, \theta_3, \theta_7, \theta_8, \theta_9, \theta_{10}$
$\pi_{R\lambda} \propto \frac{[\sigma_1\sigma_2(1-\rho^2)]^{-1}}{\sqrt{((\sigma_1/\sigma_2)-(\sigma_2/\sigma_1))^2+4\rho^2}}$	μ_1, μ_2	σ_1, σ_2, ρ
	$\mu_1 - \mu_2, \theta_3$	$\eta_3, \theta_1, \theta_2, \theta_5,$ $\theta_7, \theta_8, \theta_9, \theta_{10}$
$\pi_H \propto \frac{1}{\sigma_1^2(1-\rho^2)}$	$\boldsymbol{\mu_1}, \mu_2, \boldsymbol{\sigma_1}, \boldsymbol{\rho}$	σ_2
	$\mu_1 - \mu_2, \boldsymbol{\eta_3}, \boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}, \boldsymbol{\theta_4}, \boldsymbol{\theta_5}$	$\theta_7, \theta_8, \theta_9, \theta_{10}$
$\pi_S \propto \frac{1}{\sigma_1\sigma_2}$	μ_1, μ_2	σ_1, σ_2, ρ
	$\mu_1 - \mu_2, \theta_3, \theta_7$	$\eta_3, \theta_1, \theta_2, \theta_5, \theta_8, \theta_9, \theta_{10}$
$\pi_{R\sigma} \propto \frac{\sqrt{1+\rho^2}}{\sigma_1\sigma_2(1-\rho^2)}$	μ_1, μ_2	σ_1, σ_2, ρ
	$\mu_1 - \mu_2, \theta_3, \theta_7, \theta_9$	$\theta_1, \theta_2, \eta_3, \theta_5, \theta_8, \theta_{10}$

4.5. *Numerically computed coverage and recommendations.* First-order matching is only an asymptotic property, and finite sample performance is also crucial. We thus also implemented a modest numerical study, comparing the numerical values of frequentist coverages of the one-sided credible sets $P(\theta > q_{0.05})$ and $P(\theta < q_{0.95})$, for the parameters, θ, listed in Table 6 and for the eight objective priors $\pi_J, \pi_{IJ}, \pi_{R\rho}, \pi_{R\sigma}, \pi_{RO}, \pi_{R\lambda}, \pi_H$ and π_S. As usual, $q_\alpha = q_\alpha(\mathbf{X})$ is the posterior α-quantile of θ, and the coverage probability is computed based on the sampling distribution of $q_\alpha(\mathbf{X})$ for the fixed parameter $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ and ρ. Many of the coverage probabilities depend only on ρ, which was thus chosen to be the x-axis in the graphs. We considered the case $n = 3$ (the minimal possible sample size and hence the most challenging in terms of obtaining good coverage) and the two scenarios Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, and Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$.

TABLE 6
Performance of objective priors for each of the parameters

Parameter	**Prior**		
	Bad	**Medium**	**Good**
μ_1		rest	π_{RO}, π_H, π_J
$\mu_1 - \mu_2$		rest	π_J, π_{RO}
σ_1	π_{IJ}	rest	$\pi_H, \pi_{R\lambda}, \pi_{MS}$
σ_2	$\pi_H, \pi_{RO}, \pi_{IJ}$	rest	π_J
ρ	$\pi_J, \pi_{IJ}, \pi_S, \pi_{RO}$		$\pi_{R\rho}, \pi_{R\sigma}, \pi_{R\lambda}, \pi_H, \pi_{MS}$
λ_1	rest	$\pi_J, \pi_{R\lambda}, \pi_{RO}$	
$\theta_3 = \lvert\mathbf{\Sigma}\rvert$	π_{RO}, π_J	rest	π_{IJ}, π_H
$\theta_7 = \frac{\sigma_2}{\sigma_1}$	$\pi_H, \pi_J, \pi_{RO}, \pi_{R\lambda}$	rest	
$\theta_9 = \sigma_{12}$	π_J, π_{IJ} (due to size)	rest	$\pi_H, \pi_{R\rho}, \pi_{R\sigma}$

Here we present the numerical results concerning coverage for only two of the parameters: ρ in Figure 1 and $\theta_7 = \sigma_2/\sigma_1$ in Figure 2. Table 6 summarizes the results from the entire numerical study, the details of which can be found in [5]. The recommendations made in Table 2 for the boxed parameters are justified from these numerical results as follows.

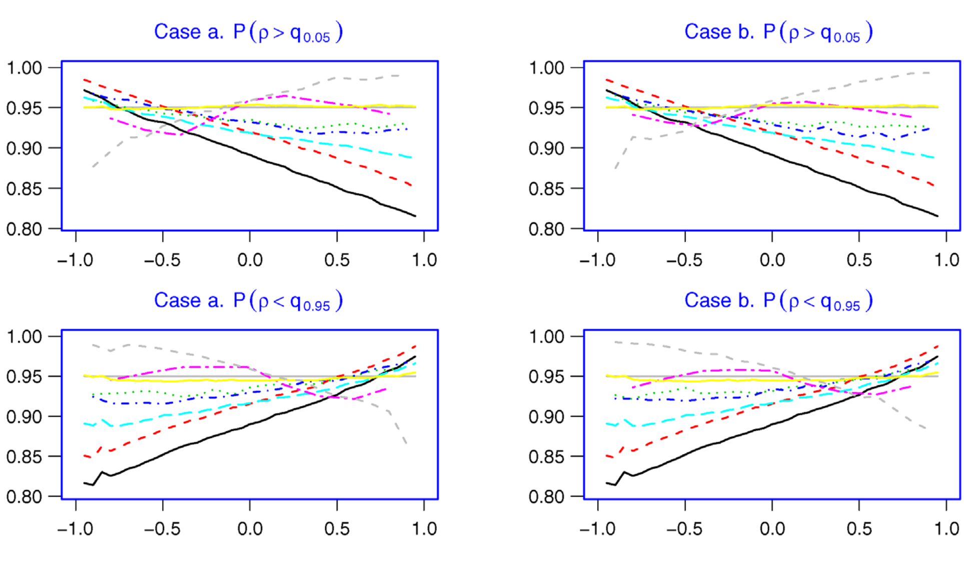

FIG. 1. *Frequentist coverages for ρ, where* Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$, *and* Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. *The x-axis is for $\rho \in (-1, 1)$.*

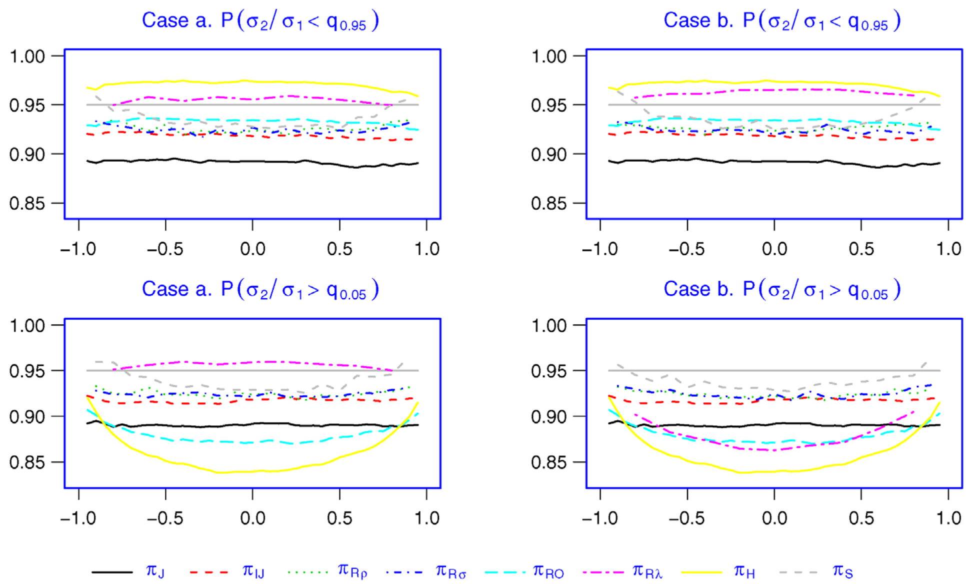

FIG. 2. *Frequentist coverages for* $\theta_7 = \sigma_2/\sigma_1$, *where* Case a: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 1, 1)$ *and* Case b: $(\mu_1, \mu_2, \sigma_1, \sigma_2) = (0, 0, 2, 1)$. *The x-axis is for* $\rho \in (-1, 1)$.

The inferences involving the nonboxed parameters in Table 2 are given in closed form in Table 1 (and so are computationally simple), and are exact frequentist matching. Furthermore, with the exception of μ_1/σ_1 and η_3, the nonboxed parameters have the indicated priors as one-at-a-time reference priors, so all three criteria point to the indicated recommendation.

For ρ, we recommend using $\pi_{R\rho}$, since this prior is a one-at-a-time-reference for ρ, first-order matching (as shown in Table 5), and has excellent numerical coverage as shown in Figure 1. Note that some might prefer to use the right-Haar prior because of its exact matching for ρ (even though it exhibits a marginalization paradox). For σ_2/σ_1, the one-at-a-time reference prior was also $\pi_{R\rho}$. As this was first-order frequentist matching and among the best in terms of numerical coverage (see Figure 2), we also recommend it for this parameter.

For λ_1, the situation is unclear. The one-at-a-time reference prior is $\pi_{R\lambda}$ and is hence our recommendation, but first-order matching results for this parameter are not known, and the numerical coverages of all priors were rather bad. For σ_{12}, the only first-order matching prior among our candidates is $\pi_{R\sigma}$. It also had the best numerical coverages, and so is a clear recommendation. Note, however, that we were not able to determine if it is a one-at-a-time reference prior for σ_{12}, so the recommendation should be considered tentative.

The most interesting question is what to recommend for general use, as an all-purpose prior. Looking at Table 2, it might seem that π_H or even π_J would be good choices, since they are optimal for so many parameters. However, both these priors can also give quite bad coverages, as indicated in Figure 2 for π_H and in Figures 1 and 2 for π_J. Indeed, from Table 6, the only priors that did not have significantly poor performance for at least one parameter (other than λ_1, for which no prior gave good coverages) were $\pi_{R\rho}$ and $\pi_{R\sigma}$. The numerical coverages for $\pi_{R\rho}$ and $\pi_{R\sigma}$ are virtually identical for all the parameters, so there is no principled way to choose between them. $\pi_{R\rho}$ is a commonly used prior and somewhat simpler, so it becomes our recommended choice for a general prior.

5. Proofs. Due to space limitations, we give only the proofs of Theorems 1, 2 and 8, because their proofs are quite different. The proofs of the other theorems in Section 4 are relatively easy consequences of Fact 1 and Lemmas 1–3. For details of these other proofs, see [5].

5.1. *Proof of Theorem* 1. With the constant prior for (μ_1, μ_2), the marginal likelihood of $(\sigma_1, \sigma_2, \rho)$ depends on $\mathbf{S}$ and is proportional to

$$|\boldsymbol{\Sigma}|^{-(n-1)/2} \exp\{-\tfrac{1}{2} \operatorname{trace}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\}.$$

Define

$$\mathcal{D} = \{(\sigma_1^*, \sigma_2^*, \rho^*) : \sigma_1^{*d_1}\sigma_2^{*d_2} g(\rho^*) < \sigma_1^{d_1}\sigma_2^{d_2} g(\rho)\},$$

$$G(\mathbf{X}, \sigma_1, \sigma_2, \rho) = \int_{\mathcal{D}} \pi(\sigma_1^*, \sigma_2^*, \rho^* \mid \mathbf{S})\, d\sigma_1^*\, d\sigma_2^*\, d\rho^*.$$

Clearly, the frequentist coverage probability is

$$P\{\theta < \theta_{1-\alpha}(\mathbf{X}) \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho\} = P\{G(\mathbf{S}, \sigma_1, \sigma_2, \rho) < 1 - \alpha \mid \sigma_1, \sigma_2, \rho\}.$$

Under the prior (25),

$$G(\mathbf{X}, \sigma_1, \sigma_2, \rho) = \frac{\int\int\int_{\mathcal{D}} \frac{h(\rho^*) \exp(-0.5 \operatorname{trace}(\mathbf{S}\boldsymbol{\Sigma}^{*-1}))}{\sigma_1^{*(n-1+c_1)} \sigma_2^{*(n-1+c_2)} (1-\rho^{*2})^{(n-1)/2}}\, d\sigma_1^*\, d\sigma_2^*\, d\rho^*}{\int\int\int \frac{h(\rho^*) \exp(-0.5 \operatorname{trace}(\mathbf{S}\boldsymbol{\Sigma}^{*-1}))}{\sigma_1^{*(n-1+c_1)} \sigma_2^{*(n-1+c_2)} (1-\rho^{*2})^{(n-1)/2}}\, d\sigma_1^*\, d\sigma_2^*\, d\rho^*},$$

where $\boldsymbol{\Sigma}^*$ is the 2×2 symmetric matrix, whose diagonal elements are σ_1^{*2} and σ_2^{*2}, and off-diagonal element is $\sigma_1^*\sigma_2^*\rho^*$. Denote $\boldsymbol{\Xi} = \operatorname{diag}(1/\sigma_1, 1/\sigma_2)$ and make transformations

$$\mathbf{T} = \boldsymbol{\Xi}\mathbf{S}\boldsymbol{\Xi} = \begin{pmatrix} \dfrac{S_{11}}{\sigma_1^2} & \dfrac{S_{12}}{\sigma_1\sigma_2} \\ \dfrac{S_{12}}{\sigma_1\sigma_2} & \dfrac{S_{22}}{\sigma_2^2} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega} = \boldsymbol{\Xi}\boldsymbol{\Sigma}^*\boldsymbol{\Xi} = \begin{pmatrix} \omega_1^2 & \omega_1\omega_2\rho^* \\ \omega_1\omega_2\rho^* & \omega_2^2 \end{pmatrix}.$$

Clearly $\operatorname{trace}(\mathbf{S}\boldsymbol{\Sigma}^{*-1}) = \operatorname{trace}(\mathbf{T}\boldsymbol{\Omega}^{-1})$, and then

$$G(\mathbf{X}, \sigma_1, \sigma_2, \rho) = \frac{\int\int\int_{\widetilde{\mathcal{D}}} \frac{h(\rho^*) \exp(-0.5 \operatorname{trace}(\mathbf{T}\boldsymbol{\Omega}^{-1}))}{\omega_1^{n-1+c_1} \omega_2^{n-1+c_2} (1-\rho^{*2})^{(n-1)/2}}\, d\omega_1\, d\omega_2\, d\rho^*}{\int\int\int \frac{h(\rho^*) \exp(-0.5 \operatorname{trace}(\mathbf{T}\boldsymbol{\Omega}^{-1}))}{\omega_1^{n-1+c_1} \omega_2^{n-1+c_2} (1-\rho^{*2})^{(n-1)/2}}\, d\omega_1\, d\omega_2\, d\rho^*},$$

where $\widetilde{\mathcal{D}} = \{(\omega_1, \omega_2, \rho^*) : \omega_1^{d_1}\omega_2^{d_2} g(\rho^*) < g(\rho)\}$. Since the sampling distribution of $\mathbf{T}$ depends only on ρ, so does the sampling distribution of $G(\mathbf{X}, \sigma_1, \sigma_2, \rho)$. Also $\widetilde{\mathcal{D}}$ depends on ρ only. The result thus holds.

5.2. *Proof of Theorem 2.* It follows from (18) and Lemma 3 (a) that

$$P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) = P\left\{\left[\psi\left(\frac{-Z_3^*}{\sqrt{\chi^{2*}_{n-a}}} + \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{\chi^{2*}_{n-a}}}\frac{r}{\sqrt{1-r^2}}\right)\right]_{1-\alpha} > \rho \,\middle|\, \rho\right\},$$

Note that ψ, defined in (18), is invertible, and $\psi^{-1}(\rho) = \rho/\sqrt{1-\rho^2}$, for $|\rho| < 1$. It follows from Lemma 3 (a) and (b) that

$$\begin{aligned}
P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) &= P\left(\left(\frac{-Z_3^*}{\sqrt{\chi^{2*}_{n-a}}} + \frac{\sqrt{\chi^{2*}_{n-b}}}{\sqrt{\chi^{2*}_{n-a}}}\frac{r}{\sqrt{1-r^2}} - \frac{\rho}{\sqrt{1-\rho^2}}\right)_{1-\alpha} > 0 \,\middle|\, \rho\right) \\
&= P\left(\left(\frac{-Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} - \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^{2*}_{n-a}}}{\sqrt{\chi^{2*}_{n-b}}}\right)_{1-\alpha} + \frac{r}{\sqrt{1-r^2}} > 0 \,\middle|\, \rho\right).
\end{aligned}$$

It follows from (21)–(23) that

$$\begin{aligned}
\frac{r}{\sqrt{1-r^2}} &= \frac{s_{12}/\sqrt{s_{11}}}{\sqrt{s_{22}(1-r^2)}} \\
&= \frac{\sigma_2\sqrt{1-\rho^2}Z_3 + (\rho\sigma_2/\sigma_1)\sqrt{s_{11}}}{\sigma_2\sqrt{1-\rho^2}\sqrt{\chi^2_{n-2}}} \\
&= \frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&P(\rho < \rho^*_{1-\alpha} \mid \boldsymbol{\xi}, \rho) \\
&\quad = P\left(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^2_{n-1}}}{\sqrt{\chi^2_{n-2}}} < \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}} + \frac{\rho}{\sqrt{1-\rho^2}}\frac{\sqrt{\chi^{2*}_{n-a}}}{\sqrt{\chi^{2*}_{n-b}}}\right)_{1-\alpha} \,\middle|\, \rho\right).
\end{aligned}$$

This completes the proof of part (a). For part (b), if (26) equals to $1-\alpha$ for any $-1 < \rho < 1$, choose $\rho = 0$ and get

$$P\left(\frac{Z_3}{\sqrt{\chi^2_{n-2}}} < \left(\frac{Z_3^*}{\sqrt{\chi^{2*}_{n-b}}}\right)_{1-\alpha}\right) = 1-\alpha,$$

which implies that $b = 2$. Substituting $b = 2$ into (26) shows that $a = 1$.

5.3. *Proof Theorem* 8. Part (a) is obvious. For part (b), since $\overline{x}_1 = \mu_1 + Z_1\sigma_1/\sqrt{n}$ and Z_1 and χ^2_{n-1} are independent, we have

$$(\theta_5 < (\theta_5^*)_{1-\alpha}) = \left(\left[\frac{Z_1^*}{\sqrt{n}} + \theta_5\left(\sqrt{\frac{\chi^{2*}_{n-a}}{\chi^2_{n-1}}} - 1\right) + \frac{Z_1}{\sqrt{n}}\sqrt{\frac{\chi^{2*}_{n-a}}{\chi^2_{n-1}}}\right]_{1-\alpha} > 0\right).$$

It follows from Lemma 3 (a) and (b) that

$$\begin{aligned}(\theta_5 < (\theta_5^*)_{1-\alpha}) &= \left(\left[\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}} + \theta_5\left(\frac{\sqrt{n}}{\sqrt{\chi^2_{n-1}}} - \frac{\sqrt{n}}{\sqrt{\chi^{2*}_{n-a}}}\right) + \frac{Z_1}{\sqrt{\chi^2_{n-1}}}\right]_{1-\alpha} > 0\right) \\ &= \left(-\frac{Z_1}{\sqrt{\chi^2_{n-1}}} - \theta_5\frac{\sqrt{n}}{\sqrt{\chi^2_{n-1}}} < \left(\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}} - \theta_5\frac{\sqrt{n}}{\sqrt{\chi^{2*}_{n-a}}}\right)_{1-\alpha}\right).\end{aligned}$$

Because Z_1 and $-Z_1$ have the same distribution and Z_1 and χ^2_{n-1} are independent, (32) holds. If (32) equals $1-\alpha$ for any θ_5, choose $\theta_5 = 0$,

$$P\left(\frac{Z_1}{\sqrt{\chi^2_{n-1}}} < \left(\frac{Z_1^*}{\sqrt{\chi^{2*}_{n-a}}}\right)_{1-\alpha}\right) = 1-\alpha,$$

which implies that $a = 1$. The result holds.

Acknowledgments. The authors are grateful to Fei Liu for performing the numerical frequentist coverage computations, to Xiaoyan Lin for computing the matching priors in Table 5, and to Susie Bayarri for helpful discussions. The authors gratefully acknowledge the very constructive comments of the editor, an associate editor and two referees.

REFERENCES

[1] BAYARRI, M. J. (1981). Inferencia bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trabajos de Estadistica e Investigacion Operativa* **32** 18–31. MR0697200

[2] BAYARRI, M. J. and BERGER, J. (2004). The interplay between Bayesian and frequentist analysis. *Statist. Sci.* **19** 58–80. MR2082147

[3] BERGER, J. O. and BERNARDO, J. M. (1992). On the development of reference priors (with discussion). In *Bayesian Statistics 4* 35–60. Oxford Univ. Press. MR1380269

[4] BERGER, J. O., STRAWDERMAN, W. and TANG, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann. Statist.* **33** 606–646. MR2163154

[5] BERGER, J. O. and SUN, D. (2006). Objective priors for a bivariate normal model with multivariate generalizations. Technical Report 07-06, ISDS, Duke Univ.

[6] BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 113–147. MR0547240

[7] BRILLINGER, D. R. (1962). Examples bearing on the definition of fiducial probability with a bibliography. *Ann. Math. Statist.* **33** 1349–1355. MR0142183

[8] BROWN, P., LE, N. and ZIDEK, J. (1994). Inference for a covariance matrix. In *Aspects of Uncertainty: A Tribute to D. V. Lindley* (P. R. Freeman and A. F. M. Smith, eds.) 77–92. Wiley, Chichester. MR1309689

[9] DATTA, G. and MUKERJEE, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. Springer, New York. MR2053794

[10] DATTA, G. S. and GHOSH, J. K. (1995a). On priors providing frequentist validity for Bayesian inference. *Biometrika* **82** 37–45. MR1332838

[11] DATTA, G. S. and GHOSH, J. K. (1995b). Noninformative priors for maximal invariant parameter in group models. *Test* **4** 95–114. MR1365042

[12] DATTA, G. S. and GHOSH, M. (1995c). Some remarks on noninformative priors. *J. Amer. Statist. Assoc.* **90** 1357–1363. MR1379478

[13] DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **35** 189–233. MR0365805

[14] FISHER, R. A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.* **26** 528–535.

[15] FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh.

[16] GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. Ser. B* **25** 368–376. MR0171354

[17] GHOSH, M. and YANG, M.-C. (1996). Noninformative priors for the two sample normal problem. *Test* **5** 145–157. MR1410459

[18] LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed. Wiley, New York. MR0852406

[19] LINDLEY, D. V. (1961). The use of prior probability distributions in statistical inference and decisions. *Proc. 4th Berkeley Sympos. Math. Statist. Probab.* **1** (J. Neyman and E. L. Scott, eds.) 453–468. Univ. California Press, Berkeley. MR0156437

[20] LINDLEY, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Cambridge Univ. Press.

[21] PRATT, J. W. (1963). Shorter confidence intervals for the mean of a normal distribution with known variance. *Ann. Math. Statist.* **34** 574–586. MR0148150

[22] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York. MR0346957

[23] SEVERINI, T. A., MUKERJEE, R. and GHOSH, M. (2002). On an exact probability matching property of right-invariant priors. *Biometrika* **89** 952–957. MR1946524

[24] STONE, M. and DAWID, A. P. (1972). Un-Bayesian implications of improper Bayes inference in routine statistical problems. *Biometrika* **59** 369–375. MR0431449

[25] YANG, R. and BERGER, J. (2004). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22** 1195–1211. MR1311972

ISDS
DUKE UNIVERSITY
BOX 90251
DURHAM, NORTH CAROLINA NC 27708-0251
USA
E-MAIL: berger@stat.duke.edu
URL: www.stat.duke.edu/~berger

DEPARTMENT OF STATISTICS
UNIVERSITY OF MISSOURI-COLUMBIA
146 MIDDLEBUSH HALL
COLUMBIA, MISSOURI 65211-6100
USA
E-MAIL: sund@missouri.edu
URL: www.stat.missouri.edu/~dsun