The Annals of Statistics
1996, Vol. 24, No. 1, 141–159

ON THE INVARIANCE OF NONINFORMATIVE PRIORS

By Gauri Sankar Datta[1] and Malay Ghosh[2]

University of Georgia and University of Florida

Jeffreys' prior, one of the widely used noninformative priors, remains invariant under reparameterization, but does not perform satisfactorily in the presence of nuisance parameters. To overcome this deficiency, recently various noninformative priors have been proposed in the literature.

This article explores the invariance (or lack thereof) of some of these noninformative priors including the reference prior of Berger and Bernardo, the reverse reference prior of J. K. Ghosh and the probability-matching prior of Peers and Stein under reparameterization. Berger and Bernardo's m-group ordered reference prior is shown to remain invariant under a special type of reparameterization. The reverse reference prior of J. K. Ghosh is shown not to remain invariant under reparameterization. However, the probability-matching prior is shown to remain invariant under any reparameterization. Also for spherically symmetric distributions, certain noninformative priors are derived using the principle of group invariance.

1. Introduction. Bayesian analysis with noninformative priors is very common when little or no prior information is available. One of the most widely used noninformative priors, introduced by Laplace (1812), is a uniform (possibly improper) distribution over the parameter space. However, a uniform prior lacks invariance under reparameterization since a uniform distribution for one parameterization will not yield, on transformation, another uniform distribution unless the transformation is linear. For example, a uniform prior for the standard deviation σ will not transform into a uniform prior for the variance σ^2. This lack of invariance of the uniform prior often translates into significant variation in the resulting posteriors. To overcome this difficulty, Jeffreys (1961) proposed his prior which, up to a proportionality constant, is given by the square root of the determinant of Fisher's information matrix. Jeffreys' prior remains invariant under any one-to-one reparameterization. Despite its success in one-parameter problems, Jeffreys' prior often runs into serious difficulties in multiparameter problems when only a subset or one or more suitable parametric function(s) of the parameter

Received August 1993; revised May 1994.

[1]Research partially supported by a research grant from the University of Georgia Research Foundation.

[2]Research partially supported by NSF Grant SES-92-01210.

AMS 1991 *subject classifications*. 62F15, 62A05.

Key words and phrases. Jeffreys' prior, reference priors, group ordering, reverse reference prior, probability-matching equation, probability-matching priors, parameter orthogonality, parameter of interest, nuisance parameters, location-scale family, spherically symmetric, group invariance.

vector $\theta(p \times 1)$ are of inferential interest and the remaining are nuisance parameters. For example, in the Neyman–Scott problem, Jeffreys' prior produces an inconsistent Bayes estimator (under squared error loss) of the error variance [Berger and Bernardo (1992a)], in the multinomial problem it lacks marginalization over "nuisance" cell probabilities [Berger and Bernardo (1992b)] and in estimating the sum of squares of a large number of independent normal means with a common variance, it leads to an unsatisfactory posterior, often referred to as Stein's paradox [Stein (1959); Bernardo (1979)].

To overcome these deficiencies of Jeffreys' prior, Berger and Bernardo (1989) expounded the reference prior approach of Bernardo (1979) for deriving noninformative priors in multiparameter situations by dividing θ into parameters of interest and nuisance parameters. This approach not only eliminates the need for an hoc modifications of Jeffreys' prior as suggested by Jeffreys himself in multiparameter problems, it also results in Jeffreys' prior when the whole parameter vector is of interest (both in the one- and multiparameter situations). The idea was further extended and generalized in a series of articles by Berger and Bernardo (1992a–c), who suggested splitting the parameter vector into multiple groups (not necessarily two) according to their order of inferential importance and prescribed a general algorithm for the construction of reference priors.

Welch and Peers (1963), Peers (1965) and Stein (1985) sought to derive noninformative priors when a real-valued parametric function $t(\theta)$ is of interest by requiring the frequentist coverage probability of the posterior credible region of $t(\theta)$ for a sample of size n to match with the nominal level with a remainder of $O(n^{-1})$. As shown by Peers (1965) and Stein (1985), such a prior is obtained by solving a differential equation [see (3.2) in Section 3]. Henceforth, this differential equation will be referred to as the matching equation, and a prior satisfying such an equation will be referred to as a matching prior. Tibshirani (1989) carried this idea further in deriving noninformative priors for $t(\theta)$ by using "orthogonal parameterization" in the sense of Cox and Reid (1987). However, in general, orthogonal parameterization is not needed, as shown by examples in Datta and Ghosh (1995a).

Some recent contributions for the construction of noninformative priors other than those of Berger and Bernardo (1992a, b) are due to Berger (1992), Ghosh and Mukerjee (1992), Clarke and Wasserman (1992), Clarke and Sun (1993) and Mukerjee and Dey (1993). Mukerjee and Dey (1993) proposed a prior by matching the frequentist coverage probability of the posterior credible interval of ψ_1 (the parameter of interest) with the nominal level up to $o(n^{-1})$ when the orthogonal nuisance parameter ψ_2 also contains a single component. This requirement often completely specifies the matching prior by determining the arbitrary component depending on ψ_2 alone [see Tibshirani (1989)]. Also a notion of reverse reference prior, attributed to J. K. Ghosh, came up in Berger's (1992) discussion of Ghosh and Mukerjee (1992). Reverse reference priors are obtained by simply interchanging the roles of the parameters of interest and of the nuisance parameters in the algorithm of Berger and Bernardo (1989, 1992a, b). Such priors satisfy the matching criterion under orthogonal parameterization, but need not necessarily be probability-matching priors otherwise [cf. Datta and Ghosh (1995a)].

It is well known that Jeffreys' prior remains invariant under one-to-one reparameterization. In this article, we will explore the invariance of various other noninformative priors under reparameterization. In particular, in Section 2.2, we establish that the m-group reference prior of Berger and Bernardo (1992a, b) remains invariant under certain one-to-one transformations. This fact is stated without proof in Berger and Bernardo (1992a) when there are only two groups of parameter vectors, but is not established analytically in the general case. Also, in Section 2.3, we have shown by an example that reverse reference priors do not necessarily remain invariant under similar one-to-one transformations.

In Section 3, it is shown that probability-matching priors remain invariant under one-to-one transformations. By "invariance" it is meant that if $\pi_\theta(\theta)$, the pdf of θ, is a probability-matching prior for the parameter of interest $t(\theta)$, and ψ is a one-to-one transformation of θ, then the transformed prior $\pi_\psi(\psi)$ obtained from $\pi_\theta(\theta)$ by change of variables will also be a probability-matching prior for $\tau(\psi) = t(\theta)$ under the ψ parameterization. Next in Section 4, it is shown that the information tradeoff prior of Clarke and Wasserman (1992) remains invariant with respect to the choice of nuisance parameters and any one-to-one transformation of the parameter (possibly vector-valued) of interest. However, we have shown by examples that priors of Ghosh and Mukerjee (1992) and of Clarke and Sun (1993) do not typically remain invariant under reparameterization.

In Section 5, based on the principle of group invariance, we provide a new motivation for generating a class of noninformative priors in some problems pertaining to spherically symmetric distributions. Similar group invariance ideas are used to generate noninformative priors in two other examples involving the general location-scale family and the exponential regression model of Cox and Reid (1987). As we shall see, often in these examples, the resulting priors agree with some of the priors mentioned earlier.

2. Invariance of reference priors.

2.1. *Notations and the algorithm.* We will follow the notations of Berger and Bernardo (1992a, b) and introduce some additional notations needed for development of our results. Suppose $p(x|\theta)$ is the pdf of a random variable X (real- or vector-valued), where $\theta \in \Theta \subset R^p$ is the unknown parameter. We also assume that the Fisher information matrix for θ, denoted by I_θ, is positive definite. We use the subscript θ to indicate that I_θ is the information matrix corresponding to the parameterization θ.

We also assume that the θ_i are separated into m groups of sizes $n_1, \dots, n_m$ and that these groups are given by

$$\theta_{(1)} = (\theta_1, \dots, \theta_{n_1}),$$
$$\theta_{(2)} = (\theta_{N_1+1}, \dots, \theta_{N_2}), \dots, \theta_{(m)} = (\theta_{N_{m-1}+1}, \dots, \theta_p),$$

where $N_j = \Sigma_{i=1}^{j} n_i$ for $j = 1, \dots, m$. Also we define, for $j = 1, \dots, m$,

$$\theta_{[j]} = \left(\theta_{(1)}, \dots, \theta_{(j)}\right), \qquad \theta_{[\sim j]} = \left(\theta_{(j+1)}, \dots, \theta_{(m)}\right)$$

with $\theta_{[m]} = \theta$ and $\theta_{[\sim m]} = \phi$, an empty set. For $n_i \times n_j$ matrix $I_{\theta ij}$, we write the information matrix in partitioned form as

$$I_\theta = \left((I_{\theta ij})\right)_{i=1,\dots,m,\, j=1,\dots,m}.$$

Also we write for $j = 1, \dots, m-1$,

$$\begin{aligned}
I_{\theta[jj]} &= ((I_{\theta ik}))_{i=1,\dots,j,\, k=1,\dots,j},\\
I_{\theta[j|\sim j]} &= ((I_{\theta ik}))_{i=1,\dots,j,\, k=j+1,\dots,m},\\
I_{\theta[\sim j|j]} &= ((I_{\theta ik}))_{i=j+1,\dots,m,\, k=1,\dots,j},\\
I_{\theta[\sim jj]} &= ((I_{\theta ik}))_{i=j+1,\dots,m,\, k=j+1,\dots,m},\\
A_{i|\sim j} &= \left(I_{\theta i,j+1}, \dots, I_{\theta im}\right).
\end{aligned}$$

For $n_i \times n_j$ matrix A_{ij}, write $S = I_\theta^{-1}$ in partitioned form $S = ((A_{ij}))_{i=1,\dots,m,\; j=1,\dots,m}$. Define S_j to be the $N_j \times N_j$ upper left corner of S, with $S_m \equiv S$ and $H_j \equiv S_j^{-1}$. Then the matrices $h_{\theta j}$, defined to be the lower right $n_j \times n_j$ corner of H_j, for $j = 1, \dots, m$, are the central quantities in deriving the reference prior for the group ordering $\{\theta_{(1)}, \dots, \theta_{(m)}\}$ following the algorithm of Berger and Bernardo (1992a, b). In fact, the algorithm involves only $|h_{\theta j}|$, where $|h_{\theta j}| =$ determinant of $h_{\theta j}$. The above definition of $h_{\theta j}$ depends heavily on the inversion of the matrices I_θ and S_j. The following lemma shows that one can express $|h_{\theta j}|$ as the ratio of $|I_{\theta[\sim j-1,j-1]}|$ to $|I_{\theta[\sim jj]}|$.

LEMMA 2.1. *With the notations introduced earlier*,

$$|h_{\theta j}| = \frac{|I_{\theta[\sim j-1,j-1]}|}{|I_{\theta[\sim jj]}|} \quad \textit{for } j = 1, \dots, m, \tag{2.1}$$

where we interpret $|I_{\theta[\sim mm]}| = 1$.

PROOF. Since $H_m = I_\theta$, $h_{\theta m} = I_{\theta mm} = I_{\theta[\sim m-1,m-1]}$. Thus (2.1) follows for $j = m$. For $j = 1, \dots, m-1$, we can write, using the inversion formula for partitioned matrices [Rao (1973), Exercise 2.7, page 33], that

$$H_j = S_j^{-1} = I_{\theta[jj]} - I_{\theta[j|\sim j]} I_{\theta[\sim jj]}^{-1} I_{\theta[\sim j|j]}$$

and, consequently,

$$h_{\theta j} = I_{\theta jj} - A_{j|\sim j} I_{\theta[\sim jj]}^{-1} A_{j|\sim j}^T.$$

Now using the result from Rao [(1973), Exercise 2.5, page 32] and the definition of $A_{j|\sim j}$ and $I_{\theta[\sim jj]}$, we get $|h_{\theta j}| = |I_{\theta[\sim j-1,j-1]}|/|I_{\theta[\sim jj]}|$ and the lemma follows. □

Lemma 2.1 is extremely useful in proving Theorem 2.1, which establishes the invariance of the reference prior for $\{\theta_{(1)}, \dots, \theta_{(m)}\}$ under reparameterization in the regular case.

We now outline the algorithm of Berger and Bernardo (1992a, b), which is used to derive the ordered group reference prior in the regular case. Let $\Theta^l \subset \Theta$, $l = 1, 2, \dots,$ denote the compacts to be chosen for θ. Define for $j = 0, 1, \dots, m-1$,

$$\Theta^l(\theta_{[j]}) = \left\{\theta_{(j+1)} : (\theta_{[j]}, \theta_{(j+1)}, \theta_{[\sim(j+1)]}) \in \Theta^l \quad \text{for some } \theta_{[\sim(j+1)]}\right\}.$$

Let $1(y \in \Omega)$ denote the indicator function that equals 1 if $y \in \Omega$ and 0 otherwise. Following Berger and Bernardo (1992a, b), we define

$$\pi^l_{\theta m}(\theta_{[\sim(m-1)]}|\theta_{[m-1]}) = \pi^l_{\theta m}(\theta_{(m)}|\theta_{[m-1]}) \tag{2.2}$$
$$= \frac{|h_{\theta m}|^{1/2} 1(\theta_{(m)} \in \Theta^l(\theta_{[m-1]}))}{\int_{\theta_{(m)} \in \Theta^l(\theta_{[m-1]})} |h_{\theta m}|^{1/2}\, d\theta_{(m)}}.$$

For $j = m-1, \dots, 1$, define successively

$$\pi^l_{\theta j}(\theta_{[\sim(j-1)]}|\theta_{[j-1]}) \tag{2.3}$$
$$= \frac{\pi^l_{\theta, j+1}(\theta_{[\sim j]}|\theta_{[j]}) \exp\left\{1/2 E^l_{\theta j}\left[(\log|h_{\theta j}|)|\theta_{[j]}\right]\right\} 1(\theta_{(j)} \in \Theta^l(\theta_{[j-1]}))}{\int_{\theta_{(j)} \in \Theta^l(\theta_{[j-1]})} \exp\left\{1/2 E^l_{\theta j}\left[(\log|h_{\theta j}|)|\theta_{[j]}\right]\right\} d\theta_{(j)}},$$

where $E^l_{\theta j}(\cdot)$ denotes the expectation w.r.t. the pdf $\pi_{\theta, j+1}(\theta_{[\sim j]}|\theta_{[j]})$.

Then the reference prior for θ is given by

$$\pi_\theta(\theta) = \lim_{l \to \infty} \frac{\pi^l_{\theta 1}(\theta)}{\pi^l_{\theta 1}(\theta^*)}, \tag{2.4}$$

where $\theta^* \in \Theta$ so that $\pi^l_{\theta 1}(\theta^*) > 0$ for all l.

2.2. *Main result on invariance.* We now establish the invariance of the reference prior for $\{\theta_{(1)}, \dots, \theta_{(m)}\}$ under reparameterization in the regular case. We consider the transformation $\psi = k(\theta)$ of the form

$$\psi_{(1)} = k_1(\theta_{(1)}), \dots, \psi_{(j)} = k_j(\theta_{[j]}), \dots, \psi_{(m)} = k_m(\theta_{[m]}), \tag{2.5}$$

where $\psi_{(j)}$ is the n_j-component. Define $\psi_{[j]} = (\psi_{(1)}, \dots, \psi_{(j)})$ and $k_{[j]}(\theta_{[j]}) = (k_1(\theta_{(1)}), \dots, k_j(\theta_{[j]}))$. We assume that $\psi_{[j]} = k_{[j]}(\theta_{[j]})$ is a one-to-one function of $\theta_{[j]}$ for $j = 1, \dots, m$. This is equivalent to assuming that for fixed $\theta_{[j-1]}$, $\psi_{(j)}$ is a one-to-one function of $\theta_{(j)}$ for $j = 1, \dots, m$. Define for $i = 1, \dots, m$, $j = 1, \dots, m$, the $n_j \times n_i$ matrix $M_{ij} = \partial\psi_{(j)}/\partial\theta_{(i)}$. Then the Jacobian of transformation matrix from θ to ψ is given by

$$M = \frac{\partial\psi}{\partial\theta} = ((M_{ij}))_{i, j = 1, \dots, m}.$$

OBSERVATION 1. Since $M_{ij} = 0$ for $j < i$, M is an upper triangular matrix.

Denoting the absolute value of the determinant of M by $\|M\|$, the Jacobian of transformation from θ to ψ is $\|M\| = \prod_{j=1}^{m} \|M_{jj}\|$. We assume $\|M\| > 0$ so that the Jacobian is nonsingular. This is equivalent to $\|M_{jj}\| > 0$, so that the Jacobian of transformation from $\theta_{(j)}$ to $\psi_{(j)}$ for fixed $\theta_{[j-1]}$ is also nonsingular for $j = 1, \dots, m$.

OBSERVATION 2. We observe that M_{jj} depends only on $\theta_{[j]}$ (or equivalently on $\psi_{[j]}$) and does not depend on $\theta_{[\sim j]}$ (or equivalently on $\psi_{[\sim j]}$, where $\psi_{[\sim j]}$ is defined as before).

Let I_ψ denote the information matrix for a new parameterization ψ. By routine calculations it can be checked that

$$I_\psi = M^{-1} I_\theta (M^{-1})^T. \tag{2.6}$$

Defining $I_{\psi[\sim jj]}$ and $M_{[\sim jj]}$ as in Section 2.1, it can be easily checked from (2.6) and Observation 1 that

$$I_{\psi[\sim jj]} = M_{[\sim jj]}^{-1} I_{\theta[\sim jj]} \left(M_{[\sim jj]}^{-1} \right)^T. \tag{2.7}$$

Defining the $n_j \times n_j$ matrix $h_{\psi j}$ as before for ψ, it can be easily checked from (2.7), by using Lemma 2.1 and the upper triangularity of $M_{[\sim jj]}$ that, for $j = 1, \dots, m$,

$$|h_{\psi j}| = |M_{jj}|^{-2} |h_{\theta j}|. \tag{2.8}$$

Let Θ, the parameter space of θ, be mapped under the transformation $\psi = k(\theta)$ onto the parameter space of ψ, which is given by

$$\Psi = \{\psi : \psi = k(\theta),\ \theta \in \Theta\} = k(\Theta).$$

Let Ψ^l denote the compacts for ψ induced by the transformation, that is,

$$\Psi^l = \{\psi : \psi = k(\theta),\ \theta \in \Theta^l\} = k(\Theta^l).$$

Define

$$\Psi^l(\psi_{[j]}) = \left\{\psi_{(j+1)} : \left(\psi_{[j]}, \psi_{(j+1)}, \psi_{[\sim (j+1)]}\right) \in \Psi^l \text{ for some } \psi_{[\sim (j+1)]}\right\}.$$

Suppose $\pi_{\psi 1}^l(\psi)$ denotes the proper prior of ψ, defined on Ψ^l and derived by following the algorithm of Section 2.1 using I_ψ, the information matrix of ψ. We establish in Theorem 2.1 below the invariance of reference prior under the transformation in (2.5). By "invariance" we mean that $\pi_\psi(\psi)$ can also be obtained from $\pi_\theta(\theta)$ by using (2.5) and the usual Jacobian method. The key step to this result is achieved by the following theorem.

THEOREM 2.1. *Consider the ordered group* $\{\theta_{(1)}, \dots, \theta_{(m)}\}$ *and the transformation* $\psi = k(\theta)$ *given by* (2.5). *Then* $\pi_{\psi_1}^l(\psi)$ *can be obtained from* $\pi_{\theta 1}^l(\theta)$ *by transformation*, *that is*,

$$\pi_{\psi 1}^l(\psi) = \pi_{\theta 1}^l\left(k^{-1}(\psi)\right) \|M\|^{-1}, \tag{2.9}$$

where $k^{-1}(\psi)$ *denotes the inverse transformation of* $\psi = k(\theta)$.

REMARK 2.1. From (2.4) and (2.9), the relation between the reference prior of ψ and the reference prior of θ can be obtained as

$$\begin{aligned}\pi_\psi(\psi) &= \lim_{l\to\infty}\frac{\pi^l_{\psi 1}(\psi)}{\pi^l_{\psi 1}(\psi^*)} = \lim_{l\to\infty}\frac{\pi^l_{\theta 1}(k^{-1}(\psi))\|M\|^{-1}}{\pi^l_{\theta 1}(k^{-1}(\psi^*))\|M_*\|^{-1}}\\ &= \frac{\pi_\theta(k^{-1}(\psi))\|M\|^{-1}}{\|M_*\|^{-1}},\end{aligned} \tag{2.10}$$

where $\psi^* \in \Psi$ is such that $\pi^l_{\theta 1}(k^{-1}(\psi^*)) > 0$ for all l, and $M_* = M$ evaluated at ψ^*.

PROOF OF THEOREM 2.1. We prove the theorem when there are $m = 2$ groups and the general m can be handled similarly by using complex notations, tedious bookkeeping, Lemma 2.1 and (2.8).

For $m = 2$ from (2.2) and (2.3),

$$\begin{aligned}\pi^l_{\theta 1}(\theta) &= \frac{\pi^l_{\theta 2}(\theta_{[\sim 1]}|\theta_{[1]})\exp\{1/2E^l_{\theta 1}[\log|h_{\theta 1}\|\theta_{[1]}]\}1(\theta_{(1)}\in\Theta^l(\theta_{[0]}))}{\int_{\theta_{(1)}\in\Theta^l(\theta_{[0]})}\exp\{1/2E^l_{\theta 1}[\log|h_{\theta 1}\|\theta_{[1]}]\}\,d\theta_{(1)}}\\ &= \frac{|h_{\theta 2}|^{1/2}1(\theta_{(2)}\in\Theta^l(\theta_{[1]}))}{\int_{\theta_{(2)}\in\Theta^l(\theta_{[1]})}|h_{\theta 2}|^{1/2}\,d\theta_{(2)}}\\ &\quad\times\frac{\exp\{1/2E^l_{\theta 1}[\log|h_{\theta 1}\|\theta_{[1]}]\}1(\theta_{(1)}\in\Theta^l(\theta_{[0]}))}{\int_{\theta_{(1)}\in\Theta^l(\theta_{[0]})}\exp\{1/2E^l_{\theta 1}[\log|h_{\theta 1}\|\theta_{[1]}]\}\,d\theta_{(1)}}.\end{aligned} \tag{2.11}$$

Now we note that:

(i) $1(\theta_{(2)}\in\Theta^l(\theta_{[1]}))1(\theta_{(1)}\in\Theta^l(\theta_{[0]})) = 1(\theta\in\Theta^l) = 1(\psi\in\Psi^l) = 1(\psi_{(2)}\in\Psi^l(\psi_{[1]}))1(\psi_{(1)}\in\Psi^l(\psi_{[0]}))$, where $\psi = k(\theta)$.

(ii) $|h_{\theta 2}|^{1/2} = |h_{\psi 2}|^{1/2}\|M_{22}\|$ by (2.8) and for fixed $\theta_{[1]}$, $\theta_{(2)}\to\psi_{(2)}$ is a one-to-one transformation with Jacobian $\|M_{22}\|^{-1}$. Then

$$\int_{\theta_{(2)}\in\Theta^l(\theta_{[1]})}|h_{\theta 2}|^{1/2}\,d\theta_{(2)} = \int_{\psi_{(2)}\in\Psi^l(\psi_{[1]})}|h_{\psi 2}|^{1/2}\,d\psi_{(2)},$$

where $\psi_{[1]} = k_{[1]}(\theta_{[1]}) = k_{(1)}(\theta_{(1)})$.

(iii) Since M_{11} is a function of only $\theta_{[1]}$,

$$\begin{aligned}E^l_{\theta 1}[\log|h_{\theta 1}\|\theta_{[1]}] &= E^l_{\theta 1}[2\log\|M_{11}\| + \log|h_{\psi 1}\|\theta_{[1]}]\\ &= 2\log\|M_{11}\| + E^l_{\theta 1}[\log|h_{\psi 1}\|\theta_{[1]}].\end{aligned} \tag{2.12}$$

Now

$$E^l_{\theta 1}\big[\log|h_{\psi 1}|\|\theta_{[1]}\big] = \frac{\int_{\theta_{(2)}\in\Theta^l(\theta_{[1]})}(\log|h_{\psi 1}|)|h_{\theta 2}|^{1/2}\,d\theta_{(2)}}{\int_{\theta_{(2)}\in\Theta^l(\theta_{[1]})}|h_{\theta 2}|^{1/2}\,d\theta_{(2)}}$$

$$(2.13)\qquad = \frac{\int_{\psi_{(2)}\in\Psi^l(\psi_{[1]})}(\log|h_{\psi 1}|)|h_{\psi 2}|^{1/2}\,d\psi_{(2)}}{\int_{\psi_{(2)}\in\Psi^l(\psi_{[1]})}|h_{\psi 2}|^{1/2}\,d\psi_{(2)}} \quad [\text{as in (ii)}]$$

$$= E^l_{\psi 1}\big[\log|h_{\psi 1}|\|\psi_{[1]}\big].$$

By (2.12) and (2.13),

$$\exp\big\{1/2E^l_{\theta 1}\big[\log|h_{\theta 1}|\|\theta_{[1]}\big]\big\} = \|M_{11}\|\exp\big\{1/2E^l_{\psi 1}\big[\log|h_{\psi 1}|\|\psi_{[1]}\big]\big\}.$$

(iv) Therefore,

$$\int_{\theta_{(1)}\in\Theta^l(\theta_{[0]})}\exp\big\{1/2E^l_{\theta 1}\big[(\log|h_{\theta 1}|)|\theta_{[1]}\big]\big\}\,d\theta_{(1)}$$

$$= \int_{\theta_{(1)}\in\Theta^l(\theta_{[0]})}\|M_{11}\|\exp\big\{1/2E^l_{\psi 1}\big[\log(|h_{\psi 1}|)|\psi_{[1]} = k_{[1]}(\theta_{[1]})\big]\big\}\,d\theta_{(1)}$$

$$= \int_{\psi_{(1)}\in\Psi^l(\psi_{[0]})}\exp\big\{1/2E^l_{\psi 1}\big[(\log|h_{\psi 1}|)|\psi_{[1]}\big]\big\}\,d\psi_{(1)}.$$

The first equality follows by (iii) and since $\psi_{(1)} = \psi_{[1]} = k_{[1]}(\theta_{[1]})$, and the second equality follows since the transformation $\psi_{(1)} = \psi_{[1]} = k_{[1]}(\theta_{[1]})$ is one-to-one with Jacobian $\|M_{11}\|^{-1}$. Then by (2.11) and (i)–(iv),

$$\begin{aligned}\pi^l_{\theta 1}(k^{-1}(\psi)) &= \|M_{11}\|\,\|M_{22}\|\pi^l_{\psi 1}(\psi)\\ &= \|M\|\pi^l_{\psi 1}(\psi)\\ \Leftrightarrow \pi^l_{\psi 1}(\psi) &= \|M\|^{-1}\pi^l_{\theta 1}(k^{-1}(\psi)).\end{aligned}$$

Hence (2.9) is established and the proof is complete. □

REMARK 2.2. For $m = 2$, if we call $\theta_{(1)}$ as the parameter of interest and $\theta_{(2)}$ as the nuisance parameter, then Theorem 2.1 implies that the reference prior remains invariant with regard to the choice of nuisance parameters.

REMARK 2.3. It is true that different reference priors may result under different choices of compact sets [cf. Berger and Bernardo (1989), page 205]. However, the invariance property of reference priors in the setup of Theorem 2.1 continues to hold in the sense that once a reference prior is found with a particular choice of compact sets, a reference prior under reparameterization is obtainable using the usual Jacobian formula.

2.3. *Noninvariance of reverse reference priors*. In this section we consider an example to show that the reverse reference prior does not remain invariant even under one-to-one transformation of the parameter of interest. We consider the estimation of the product of two normal means of Berger and Bernardo (1989). Let X_1 and X_2 be two independent normal random variables with unit variances and means μ_1 (> 0) and μ_2 (> 0), respectively. The parameter of interest is $\theta_1 = \mu_1 \mu_2$. Berger and Bernardo (1989) considered the parameterization $\theta_1 = \mu_1 \mu_2$ and $\theta_2 = \sqrt{\mu_2/\mu_1}$, and they derived the information matrix

$$I_\theta = \begin{bmatrix} \dfrac{\theta_2^2 + \theta_2^{-2}}{4\theta_1} & \dfrac{\theta_2(1 - \theta_2^{-4})}{2} \\ \dfrac{\theta_2(1 - \theta_2^{-4})}{2} & \theta_1(1 + \theta_2^{-4}) \end{bmatrix}. \tag{2.14}$$

The reverse reference prior $\pi_{\mathrm{RR}}(\theta_1, \theta_2)$ is

$$\pi_{\mathrm{RR}}(\theta_1, \theta_2) \propto \theta_1^{-1/2}(1 + \theta_2^4)^{-1/2}. \tag{2.15}$$

The above prior may be obtained by following the usual algorithm of Berger and Bernardo (1989) for the ordered group $\{\theta_2, \theta_1\}$ and using rectangular compacts either for (θ_1, θ_2) or for (μ_1, μ_2). For the orthogonal parameterization $\psi_1 = 2\mu_1 \mu_2$ and $\psi_2 = \mu_2^2 - \mu_1^2$ considered by Tibshirani (1989) and also by Datta and Ghosh (1995a), the information matrix is

$$I_\psi = 1/4(\psi_1^2 + \psi_2^2)^{-1/2} I_2. \tag{2.16}$$

The reverse reference prior for (ψ_1, ψ_2) parameterization, derived by Datta and Ghosh (1995a), is

$$\pi_{\mathrm{RR}}(\psi_1, \psi_2) = (\psi_1^2 + \psi_2^2)^{-1/4}. \tag{2.17}$$

Note that the transformation between ψ and θ is given by $\psi_1 = 2\theta_1$ and $\psi_2 = \theta_1(\theta_2^2 - \theta_2^{-2})$, and the transformed prior of θ, $\pi_{\mathrm{RR}}^T(\theta)$ from $\pi_{\mathrm{RR}}(\psi)$ is given by

$$\pi_{\mathrm{RR}}^T(\theta_1, \theta_2) = \frac{8\theta_1}{\theta_2}\left(\frac{\theta_2^2 + \theta_2^{-2}}{4\theta_1}\right)^{1/2} = 4\theta_1^{1/2}\frac{(1 + \theta_2^4)^{1/2}}{\theta_2^2}. \tag{2.18}$$

It follows easily from (2.15) and (2.18) that the reverse reference prior does not remain invariant. However, the invariance of the usual reference prior follows from Theorem 2.1.

3. Invariance of probability-matching priors. Suppose $X_1, \dots, X_n$ are iid d-component random vectors with density $p(x|\theta)$, where $\theta = (\theta_1, \dots, \theta_p)$. We denote $(X_1, \dots, X_n) = Z$. Inferences are sought concerning a real-valued parametric function $t(\theta)$ which is twice continuously differentiable. Suppose we seek a prior $\pi(\theta)$ so that

$$(3.1)\qquad \begin{aligned} &P_\theta\left[\frac{\sqrt{n}\,\{t(\theta)-t(\hat\theta)\}}{\sqrt{b}}\le u\right]\\ &\quad = P_\pi\left[\frac{\sqrt{n}\,\{t(\theta)-t(\hat\theta)\}}{\sqrt{b}}\le u \,\middle|\, Z\right] + O_p(n^{-1})\end{aligned}$$

for all u, where $\hat\theta$ is the posterior mode or maximum likelihood estimator of θ and b denotes the asymptotic posterior variance of $\sqrt{n}\,(t(\theta)-t(\hat\theta))$ up to $O_p(n^{-1})$. In the above, $P_\theta(\cdot)$ refers to the probability distribution of Z under $P_\pi(\cdot|Z)$ is the posterior probability distribution of θ under π. Priors satisfying (3.1) are referred to as probability-matching priors. It is shown in Datta and Ghosh (1995b) that (3.1) holds if and only if

$$(3.2)\qquad \sum_{\alpha=1}^{p}\frac{\partial}{\partial\theta_\alpha}\left\{\frac{\rho_\alpha^T I_\theta^{-1}(\theta)\,\nabla_t(\theta)}{\sqrt{\nabla_t^T(\theta)\, I_\theta^{-1}(\theta)\,\nabla_t(\theta)}}\,\pi(\theta)\right\} = 0,$$

where

$$\nabla_t(\theta) = \left(\frac{\partial}{\partial\theta_1}t(\theta),\dots,\frac{\partial}{\partial\theta_p}t(\theta)\right)^T$$

and ρ_α is the αth unit column p-vector. Equation (3.2) will be referred to as the probability-matching equation and is similar to Stein's (1985) equation (3.8). Probability-matching priors are extensively discussed in the literature in various contexts by Peers (1965), Stein (1985), Berger and Bernardo (1989), Tibshirani (1989), Ghosh and Mukerjee (1992) and Datta and Ghosh (1995b), just to name a few.

We denote any prior satisfying differential equation (3.2) by $\pi_\theta(\theta)$. Consider a one-to-one transformation $\psi = (\psi_1,\dots,\psi_p)^T = (k_1(\theta),\dots,k_p(\theta))^T$ with nonsingular Jacobian of transformation matrix given by $J(\theta\to\psi) = ((\partial\psi_j/\partial\theta_i))_{i,j=1,\dots,p} = M$ (say). Suppose under this transformation the parametric function $t(\theta)$ is changed to $\tau(\psi)$ and the information matrix for ψ is $I_\psi(\psi)$. We also denote the Jacobian of inverse transformation (i.e., $\psi\to\theta$) matrix by $J(\psi\to\theta) = ((\partial\theta_j/\partial\psi_i))_{i,j=1,\dots,p} = N$ (say).

Let $\pi_\theta(\theta)$, a prior density for θ, satisfy (3.2) when $t(\theta)$ is the parameter of interest. By change of variables, $\pi_\psi(\psi)$, the density for ψ is given by

$$(3.3)\qquad \begin{aligned}\pi_\psi(\psi) &= \pi_\theta(k^{-1}(\psi))\|M\|^{-1}\\ &= \pi_\theta(k^{-1}(\psi))\|N\|.\end{aligned}$$

In the following theorem we show that prior density $\pi_\psi(\psi)$ is a probability-matching prior for ψ when $\tau(\psi)$ is the parameter of interest.

THEOREM 3.1. *A prior density $\pi_\theta(\theta)$ will be probability-matching for $t(\theta)$ if and only if the prior density $\pi_\psi(\psi)$ given in* (3.3) *is probability-matching for $\tau(\psi)$.*

PROOF. We will show that $\pi_\theta(\theta)$ satisfies (3.2) if and only if $\pi_\psi(\psi)$ satisfies

$$\sum_{\beta=1}^{p} \frac{\partial}{\partial \psi_\beta} \left\{ \frac{\rho_\beta^T I_\psi^{-1}(\psi) \nabla_\tau(\psi)}{\sqrt{\nabla_\tau^T(\psi) I_\psi^{-1}(\psi) \nabla_\tau(\psi)}} \pi_\psi(\psi) \right\} = 0. \tag{3.4}$$

It is easy to check that $\nabla_\tau(\psi) = N\,\nabla_t(\theta)$ and $M = N^{-1}$. Then by (2.6),

$$\begin{aligned} I_\psi^{-1}(\psi)\nabla_\tau(\psi) &= M^T I_\theta^{-1}(\theta)\nabla_t(\theta), \\ \nabla_\tau^T(\psi) I_\psi^{-1}(\psi)\nabla_\tau(\psi) &= \nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta). \end{aligned} \tag{3.5}$$

Define $I_\theta^{-1}(\theta)\,\nabla_t(\theta) = (u_1(\theta), \dots, u_p(\theta))^T$. Then $\rho_\beta^T I_\psi^{-1}(\psi)\,\nabla_\tau(\psi) = \sum_{j=1}^p u_j(\theta)(\partial\psi_\beta/\partial\theta_j)$. Consequently, by (3.3),

$$\begin{aligned} &\sum_{\beta=1}^{p} \frac{\partial}{\partial \psi_\beta} \left(\frac{\rho_\beta^T I_\psi^{-1}(\psi) \nabla_\tau(\psi)}{\sqrt{\nabla_\tau^T(\psi) I_\psi^{-1}(\psi) \nabla_\tau(\psi)}} \pi_\psi(\psi) \right) \\ &\quad = \sum_{\beta=1}^{p}\sum_{j=1}^{p} \frac{\partial}{\partial\psi_\beta}\left(\frac{u_j(\theta)\pi_\theta(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \right) \frac{\partial\psi_\beta}{\partial\theta_j}\|N\| \\ &\qquad + \sum_{\beta=1}^{p}\sum_{j=1}^{p} \frac{u_j(\theta)\pi_\theta(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \frac{\partial}{\partial\psi_\beta}\left(\frac{\partial\psi_\beta}{\partial\theta_j}\right)\|N\| \\ &\qquad + \sum_{\beta=1}^{p}\sum_{j=1}^{p} \frac{u_j(\theta)\pi_\theta(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \frac{\partial\psi_\beta}{\partial\theta_j}\frac{\partial}{\partial\psi_\beta}\|N\|. \end{aligned} \tag{3.6}$$

Now, after some simplification, first term on the rhs of (3.6) equals

$$\|N\| \sum_{j=1}^{p} \frac{\partial}{\partial\theta_j}\left(\frac{\rho_j^T I_\theta^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \pi_\theta(\theta)\right) \tag{3.7}$$

and third term on the rhs of (3.6) equals

$$\sum_{j=1}^{p} \frac{u_j(\theta)\pi_\theta(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \frac{\partial}{\partial\theta_j}\|N\|. \tag{3.8}$$

First assume that $|N| > 0$. Then $\|N\| = |N|$. Since $M = N^{-1}$,

$$\begin{aligned} \frac{\partial}{\partial\theta_j}\|N\| &= -|M|^{-2}\frac{\partial}{\partial\theta_j}|M| \\ &= -|N|^2 \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} \frac{\partial}{\partial\theta_j}\left(\frac{\partial\psi_\beta}{\partial\theta_\alpha}\right) c_{\alpha\beta} \end{aligned} \tag{3.9}$$

by using Lemma A.4.5 of Anderson [(1984), page 598], where

$$c_{\alpha\beta} = \text{cofactor of } (\alpha, \beta)\text{th element of } M = |M|\frac{\partial\theta_\alpha}{\partial\psi_\beta}.$$

Using $c_{\alpha\beta}$ in (3.9), we get

$$\frac{\partial}{\partial\theta_j}|N| = -|N| \sum_{\alpha=1}^{p} \sum_{\beta=1}^{p} \frac{\partial^2\psi_\beta}{\partial\theta_j\,\partial\theta_\alpha} \frac{\partial\theta_\alpha}{\partial\psi_\beta}. \tag{3.10}$$

From (3.8)–(3.10), third term on the rhs of (3.6) equals

$$-|N| \sum_{j=1}^{p} \frac{u_j(\theta)\pi_\theta(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \sum_{\alpha=1}^{p}\sum_{\beta=1}^{p} \frac{\partial^2\psi_\beta}{\partial\theta_j\,\partial\theta_\alpha}\frac{\partial\theta_\alpha}{\partial\psi_\beta}. \tag{3.11}$$

Now for $j = 1, \dots, p$, using $\sum_{\beta=1}^{p}(\partial/\partial\psi_\beta)(\partial\psi_\beta/\partial\theta_j) = \sum_{\beta=1}^{p}\sum_{\alpha=1}^{p}(\partial^2\psi_\beta/\partial\theta_\alpha\,\partial\theta_j)(\partial\theta_\alpha/\partial\psi_\beta)$, second term on the rhs of (3.6) equals

$$|N| \sum_{j=1}^{p} \frac{u_j(\theta)\pi_\theta(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}} \sum_{\beta=1}^{p}\sum_{\alpha=1}^{p} \frac{\partial^2\psi_\beta}{\partial\theta_\alpha\,\partial\theta_j}\frac{\partial\theta_\alpha}{\partial\psi_\beta}. \tag{3.12}$$

By (3.6), (3.7), (3.11), (3.12) and using $\partial^2\psi_\beta/\partial\theta_j\,\partial\theta_\alpha = \partial^2\psi_\beta/\partial\theta_\alpha\,\partial\theta_j$ for all α, β and j, the lhs of (3.6) equals

$$|N| \sum_{j=1}^{p} \frac{\partial}{\partial\theta_j}\left(\frac{\rho_j^T I_\theta^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}}\pi_\theta(\theta)\right).$$

Hence, the lhs of (3.4) will be zero if and only if

$$\sum_{j=1}^{p} \frac{\partial}{\partial\theta_j}\left(\frac{\rho_j^T I_\theta^{-1}(\theta)\nabla_t(\theta)}{\sqrt{\nabla_t^T(\theta) I_\theta^{-1}(\theta)\nabla_t(\theta)}}\pi_\theta(\theta)\right) = 0.$$

The same equation results when $|N| < 0$. Hence a prior density $\pi_\theta(\theta)$ is probability-matching for $t(\theta)$ if and only if $\pi_\psi(\psi)$ in (3.3) is probability-matching for $\tau(\psi)$. □

4. Invariance of tradeoff priors. This section addresses the invariance (or lack thereof) of some of the other noninformative priors that are proposed in the literature. First we show that the prior of Ghosh and Mukerjee (1992) does not remain invariant under different choices of the orthogonal nuisance parameter. Suppose $\theta(2 \times 1)$ is orthogonal with information matrix $I(\theta) = \operatorname{diag}(I_{\theta_1\theta_1}, I_{\theta_2\theta_2})$ and θ_1 is the parameter of interest. Then

$$\pi_{\mathrm{GM}}(\theta_1, \theta_2) = I_{\theta_1\theta_1}^{1/2}.$$

For the one-to-one transformation $\psi_1 = h_1(\theta_1)$ and $\psi_2 = h_2(\theta_2)$, the information matrix is $I(\psi) = \operatorname{diag}(\{h_1'(\theta_1)\}^{-2} I_{\theta_1\theta_1}, \{h_2'(\theta_2)\}^{-2} I_{\theta_2\theta_2})$ and, consequently, $\pi_{\mathrm{GM}}(\psi_1, \psi_2) = |h_1'(\theta_1)|^{-1} I_{\theta_1\theta_1}^{1/2}$ and the transformed prior for (θ_1, θ_2) is $\pi_{\mathrm{GM}}^T(\theta_1, \theta_2) = |h_2'(\theta_2)| I_{\theta_1\theta_1}^{1/2} \neq \pi_{\mathrm{GM}}(\theta_1, \theta_2)$ unless $|h_2'(\theta_2)| = 1$. Also, the prior of

Clarke and Sun (1993) is the inverse of the determinant of the Fisher information matrix, and does not remain invariant under reparameterization.

The tradeoff prior of Clarke and Wasserman (1992) possesses the invariance property under transformation of the type used in Theorem 2.1 with $m = 2$. We shall prove this in the special two-parameter case when θ_1 is the parameter of interest and θ_2 is the nuisance parameter. The general case can be proved along the same lines with slightly more complex notations.

First note that the information tradeoff prior of Clarke and Wasserman (1992) is given by

$$\pi_\alpha(\theta_1, \theta_2) \propto \frac{I_{\theta 11.2}^{1/(2\alpha)}(\theta_1, \theta_2)\left|I_\theta(\theta_1, \theta_2)\right|^{1/2}}{\left(\int I_{\theta 11.2}^{1/2\alpha}(\theta_1, \theta_2)\left|I_\theta(\theta_1, \theta_2)\right|^{1/2} d\theta_2\right)^{1/(\alpha+1)}}. \tag{4.1}$$

Consider now the one-to-one transformation $\psi_1 = k_1(\theta_1)$ and $\psi_2 = k_2(\theta_1, \theta_2)$ from (θ_1, θ_2) to (ψ_1, ψ_2), where $k_1(\cdot)$ is also a one-to-one function. Let $\pi_\alpha^T(\psi_1, \psi_2)$ denote the distribution of (ψ_1, ψ_2) derived from $\pi_\alpha(\theta_1, \theta_2)$ using the usual Jacobian technique. Also, let $\pi_\alpha(\psi_1, \psi_2)$ denote the information tradeoff prior using the information matrix I_ψ for (ψ_1, ψ_2). The invariance result is established by proving the following theorem.

THEOREM 4.1. *$\pi_\alpha^T(\psi_1, \psi_2) = \pi_\alpha(\psi_1, \psi_2)$ for all (ψ_1, ψ_2).*

PROOF. Write $\pi_\alpha(\theta_1, \theta_2) = p_{\theta\alpha}(\theta_2|\theta_1)p_{\theta\alpha}(\theta_1)$, where

$$p_{\theta\alpha}(\theta_2|\theta_1) = \frac{I_{\theta 11.2}^{1/(2\alpha)}(\theta_1, \theta_2)\left|I_\theta(\theta_1, \theta_2)\right|^{1/2}}{\int I_{\theta 11.2}^{1/(2\alpha)}(\theta_1, \theta_2)\left|I_\theta(\theta_1, \theta_2)\right|^{1/2} d\theta_2}, \tag{4.2}$$

$$p_{\theta\alpha}(\theta_1) = \frac{\left[\int I_{\theta 11.2}^{1/(2\alpha)}(\theta_1, \theta_2)\left|I_\theta(\theta_1, \theta_2)\right|^{1/2} d\theta_2\right]^{\alpha/(\alpha+1)}}{\int\left[\int I_{\theta 11.2}^{1/(2\alpha)}(\theta_1, \theta_2)\left|I_\theta(\theta_1\theta_2)\right|^{1/2} d\theta_2\right]^{\alpha/(\alpha+1)} d\theta_1}. \tag{4.3}$$

Then

$$\pi_\alpha^T(\psi_1, \psi_2) = p_{\theta\alpha}(u_2(\psi_1, \psi_2)|u_1(\psi_1))p_{\theta\alpha}(u_1(\psi_1))\left|\frac{\partial u_1}{\partial\psi_1}\right|\left|\frac{\partial u_2}{\partial\psi_2}\right|, \tag{4.4}$$

where $(\theta_1 = u_1(\psi_1), \theta_2 = u_2(\psi_1, \psi_2))$ denotes the inverse transformation. Use the facts

$$|I_\theta| = |I_\psi|\left(\frac{\partial u_1}{\partial\psi_1}\right)^{-2}\left(\frac{\partial u_2}{\partial\psi_2}\right)^{-2}, \qquad I_{\theta 11.2} = \left(\frac{\partial u_1}{\partial\psi_1}\right)^{-2} I_{\psi 11.2}. \tag{4.5}$$

Then, after simplification,

$$p_{\theta\alpha}(u_2(\psi_1,\psi_2)|u_1(\psi_1)) = \frac{I_{\psi 11.2}^{1/(2\alpha)}(\psi_1,\psi_2)\left|I_\psi(\psi_1,\psi_2)\right|^{1/2}|\partial u_2/\partial\psi_1|^{-1}}{\int I_{\psi 11.2}^{1/(2\alpha)}(\psi_1,\psi_2)\left|I_\psi(\psi_1,\psi_2)\right|^{1/2}\,d\psi_2}, \tag{4.6}$$

$$p_{\theta\alpha}(u_1(\psi_1)) = \frac{|\partial u_1/\partial\psi_1|^{-1}\left[\int I_{\psi 11.2}^{1/(2\alpha)}(\psi_1,\psi_2)\left|I_\psi(\psi_1,\psi_2)\right|^{1/2}\,d\psi_2\right]^{\alpha/(\alpha+1)}}{\int\left[\int I_{\psi 11.2}^{1/(2\alpha)}\left|I_\psi(\psi_1,\psi_2)\right|^{1/2}\,d\psi_2\right]^{\alpha/(\alpha+1)}d\psi_1}. \tag{4.7}$$

Combining (4.4), (4.6), (4.7) and (4.1),

$$\pi_\alpha^T(\psi_1,\psi_2) = \pi_\alpha(\psi_1,\psi_2). \qquad \square$$

5. Noninformative priors for spherically symmetric distributions. The noninformative priors, for example, the reference priors of Berger and Bernardo, Ghosh and Mukerjee, Clarke and Sun and Clarke and Wasserman considered in the previous sections, are derived based on the maximization of certain divergence functions. In this section, using the principle of group invariance, we derive noninformative priors for certain parametric functions in spherically symmetric distributions. In many cases the priors derived in this way coincide with reference prior or some of the others priors mentioned earlier.

Suppose $X = (X_1,\dots,X_p)^T$ has a spherically symmetric distribution with a pdf

$$f(x;\mu,\sigma) = \frac{1}{\sigma^p} f_0\left(\frac{(x-\mu)^T(x-\mu)}{\sigma^2}\right),$$

where $\mu = (\mu_1,\dots,\mu_p)^T \in R^p$ and $\sigma > 0$. We denote the spherically symmetric distribution of X by $X \sim S_p(\mu,\sigma^2 I_p)$. The information matrix for (μ,σ) is given by

$$I(\mu,\sigma) = \sigma^{-2}\,\mathrm{diag}(c_1 I_p, c_2), \tag{5.1}$$

where

$$c_1 = 4\int_{R^p} u_1^2\left[\frac{f_0'(u^Tu)}{f_0(u^Tu)}\right]^2 f_0(u^Tu)\,du$$

and

$$c_2 = \int_{R^p}\left[p + 2u^Tu\frac{f_0'(u^Tu)}{f_0(u^Tu)}\right]^2 f_0(u^Tu)\,du.$$

We now derive group invariant priors for certain parameters of interest.

EXAMPLE 5.1. We consider deriving noninformative priors when $\psi_1 = (\mu^T\mu)^{1/2}/\sigma$ is the parameter of interest. Note that the estimation problem remains invariant under the orthogonal transformation $(\mu, \sigma) \to (\nu, \tau) = (Q\mu, \sigma)$ for orthogonal $Q(p \times p)$ as well as under the scale transformation $(\nu, \tau) = (c\mu, c\sigma)$ for scalar $c > 0$. Any reasonable prior $\pi(\mu, \sigma)$ for (μ, σ) should also be a reasonable prior for (ν, τ). This requirement leads to the following two conditions on π:

$$\pi(\mu, \sigma) = \pi(Q^T\mu, \sigma); \tag{5.2}$$

$$\pi(\mu, \sigma) = \pi(\mu/c, \sigma/c)c^{-p-1}. \tag{5.3}$$

Simple algebraic manipulations yield

$$\pi(\mu, \sigma) = k(\mu^T\mu/\sigma^2)\sigma^{-p-1} \tag{5.4}$$

for an arbitrary nonnegative function $k(\cdot)$. Using the polar transformation (μ, σ) to (r, θ, τ), where $\theta = (\theta_1, \dots, \theta_{p-1})^T$ and $\sigma = \tau$, we obtain

$$\pi(r, \theta, \tau) = k(r^2/\tau^2)\tau^{-p-1}r^{p-1}s_1^{p-2} \cdots s_{p-2}, \tag{5.5}$$

where $s_i = \sin\theta_i$, $i = 1, \dots, p-2$. Note that $\psi_1 = r/\tau$. Transforming further (r, θ, τ) to (ψ_1, ψ_2, θ) by $\psi_1 = r/\tau$, $\psi_2 = \tau$, $\theta = \theta$, we get from (5.5),

$$\pi(\psi_1, \psi_2, \theta) = k(\psi_1^2)\psi_1^{p-1}\psi_2^{-1}s_1^{p-2} \cdots s_{p-2}. \tag{5.6}$$

We now investigate the condition of $k(\cdot)$ under which the prior given in (5.6) is a probability-matching prior. First observe that the information matrix [under the (ψ_1, ψ_2, θ) parameterization] is given by

$$I(\psi, \theta) = \text{block diagonal}(I_\psi, I_\theta), \tag{5.7}$$

where

$$I_\psi = \psi_2^{-2}\begin{bmatrix} c_1\psi_2^2 & c_1\psi_1\psi_2 \\ c_1\psi_1\psi_2 & c_1\psi_1^2 + c_2 \end{bmatrix},$$

$$I_\theta = c_1\psi_1^2 \operatorname{diag}(1, s_1^2, s_1^2s_2^2, \dots, s_1^2s_2^2 \cdots s_{p-2}^2).$$

Comparing (3.2) and (5.6) we find $k(x) = (c_1x + c_2)^{-1/2}x^{-(p-1)/2}$ and the resulting probability-matching prior is given by $\pi(\psi_1, \psi_2, \theta) = (c_1\psi_1^2 + c_2)^{-1/2}\psi_2^{-1}s_1^{p-2} \cdots s_{p-2}$. It can be checked that this prior is also the reference prior for the grouping $\{\psi_1, \psi_2, \theta\}$ and a matching prior for θ_{p-1} due to Tibshirani (1989). When $p = 2$, θ_1 is one-to-one with μ_2/μ_1, the parameter of interest in the Fieller–Creasy problem and the resulting prior $\pi(\psi_1, \psi_2, \theta_1) = (c_1\psi_1^2 + c_2)^{-1/2}\psi_2^{-1}$ will be a matching prior of θ_1. However, in general, this will not be a matching prior for any of the components $\theta_1, \dots, \theta_{p-2}$.

REMARK 5.1. Suppose σ is known and we want to derive a noninformative prior for $\mu = (\mu_1, \dots, \mu_p)^T$ when the parameter of interest is $\psi = \sum_{i=1}^p \mu_i^2$. This example was considered by Bernardo (1979) and Stein (1985) under normality. For any $p \times p$ orthogonal matrix Q, $Y = QX$ is again spherically symmetric with location parameter $\nu = Q\mu$. Since $\mu^T\mu = \nu^T\nu$, any reasonable prior $\pi(\mu)$ for estimating $\mu^T\mu$ should also be reasonable for estimating $\nu^T\nu$. Since the transformation μ to ν is orthogonal, any reasonable prior $\pi(\mu)$ for μ should be invariant under orthogonal transformation. This is equivalent to $\pi(\mu) = k(\mu^T\mu)$ for some nonnegative function $k(\cdot)$. The information matrix is $I(\mu) = \sigma^{-2}c_1 I_p$. Note that $k(x) = 1$ corresponds to the Jeffreys prior for μ. If we require $\pi(\mu)$ to be a matching prior for ψ, then solving again the differential equation (3.2), we get $k(x) = x^{-(p-1)/2}$ resulting in the prior $\pi(\mu) = (\sum_{i=1}^p \mu_i^2)^{-(p-1)/2}$, which was also derived by Stein (1985). Using a polar transformation, one gets from the above prior of μ the prior of ψ given by $\pi_\psi(\psi) = \psi^{-1/2}$, which is obtained by Bernardo [(1979), page 125] as the reference prior for ψ.

REMARK 5.2. Suppose $p = 1$ in Example 5.1. Then the parameter of interest is $|\mu|/\sigma$. Note that $\psi_1 = |\mu|/\sigma$, $\psi_2 = \sigma$ and the matching prior is $\pi(\psi_1, \psi_2) = (c_1\psi_1^2 + c_2)^{-1/2}\psi_2^{-1}$. Under normality, $c_1 = 1$, $c_2 = 2$ and the resulting prior is identical to the reference prior of Bernardo (1979). See also Tibshirani (1989).

EXAMPLE 5.2. In this example, we assume σ^2 is known and $p = 2$. We want to derive the noninformative prior when the parameter of interest is $\mu_1\mu_2 = \psi_1$ (say). Note that X is equivalent to $Y = \begin{pmatrix} Y_1 \\ Y_1 \end{pmatrix}$, where $Y_1 = (X_1 + X_2)/\sqrt{2}$ and $Y_2 = (X_1 - X_2)/\sqrt{2}$ and Y is spherically symmetric with location parameter $\theta = (\theta_1, \theta_2)$ with $\theta_1 = (\mu_1 + \mu_2)/\sqrt{2}$ and $\theta_2 = (\mu_1 - \mu_2)/\sqrt{2}$. The estimation of $\theta_1\theta_2 = (\mu_1^2 - \mu_2^2)/2 = \psi_2$ (say) is equivalent to the estimation of ψ_1. Also since the information matrices of (μ_1, μ_2) and (θ_1, θ_2) are identical, it is natural to expect that any reasonable prior for (μ_1, μ_2) when $\psi_1 = \mu_1\mu_2$ is the parameter of interest should also be reasonable when $\psi_2 = (\mu_1^2 - \mu_2^2)/2$ is the parameter of interest. Indeed, since any 2×2 orthogonal matrix of rotation is representable as $Q = \begin{pmatrix} a_1 & -a_2 \\ a_2 & a_1 \end{pmatrix}$ with $a_1^2 + a_2^2 = 1$, $X \sim S_2(\mu, \sigma^2 I_2)$ is equivalent to $W = QX \sim S_2(\nu, \sigma^2 I_2)$, where $\nu = (\nu_1, \nu_2)^T = Q\mu$. Also it should be noted that $\nu_1\nu_2$ as well as $(\nu_1^2 - \nu_2^2)/2$ is a linear function of ψ_1 and ψ_2. Due to this observation, any reasonable prior for (μ_1, μ_2) to estimate $\mu_1\mu_2$ should also be reasonable to estimate $\nu_1\nu_2$. This means that since the transformation μ to ν is orthogonal, the prior $\pi(\mu_1, \mu_2)$, say, should remain invariant under orthogonal transformation. This is equivalent to $\pi(\mu_1, \mu_2) = k(\mu_1^2 + \mu_2^2)$ for some nonnegative function $k(\cdot)$. If we require $\pi(\mu_1, \mu_2)$ to be a matching prior for $\psi_1 = \mu_1\mu_2$, then solving the differential equation (3.2), we get $k(x) = \sqrt{x}$, giving Stein's prior $(\mu_1^2 + \mu_2^2)^{1/2}$ for (μ_1, μ_2), which is also the reference prior derived by Berger and Bernardo (1989) [also Datta and Ghosh (1995a)]. Note that the choice $k(x) = 1$ produces the Jeffreys prior for (μ_1, μ_2), which is not a matching prior.

The group invariance structure for deriving a noninformative prior need not be confined to spherically symmetric distributions alone. We provide below two examples where the group invariance structure is exploited to find noninformative priors for parameters of certain nonspherically symmetric distributions.

EXAMPLE 5.3. Suppose X has location-scale pdf $\sigma^{-1}f((x-\mu)/\sigma)$, where $(\mu,\sigma)\in(-\infty,\infty)\times(0,\infty)$ and f is differentiable in its argument. We want to derive a noninformative prior for estimating μ/σ. Note that for any $c>0$, $Y=cX$ has a location-scale pdf with parameters $c\mu$ and $c\sigma$. Let $\theta=(\theta_1, \theta_2)=(\mu,\sigma)$ and $\phi=(\phi_1,\phi_2)=(c\mu,c\sigma)$. Then the estimation of $\psi(\theta)=\theta_1/\theta_2=\mu/\sigma$ is equivalent to the estimation of $\psi(\phi)$. Also since the information matrix of ϕ is a scalar multiple of that of θ, it is expected that any reasonable prior $\pi_\theta(\theta_1,\theta_2)$ for θ for estimating $\psi(\theta)$ should also be reasonable for ϕ for estimating $\psi(\phi)$. That means that if $\pi_\phi(\phi_1,\phi_2)$ denotes such a prior for ϕ, we should have

$$\pi_\theta(u_1,u_2)=\pi_\phi(u_1,u_2) \tag{5.8}$$

for all u_1 and u_2. However, since $\phi=c\theta$, by transformation we get

$$\pi_\phi(u_1,u_2)=\pi_\theta\left(\frac{u_1}{c},\frac{u_2}{c}\right)\frac{1}{c^2} \tag{5.9}$$

for all $c>0$. Combining (5.8) and (5.9) we get $\pi_\theta(u_1,u_2)=u_2^{-2}k(u_1/u_2)$ for some nonnegative function $k(\cdot)$. We choose $k(\cdot)$ by satisfying (3.2) for $\psi(\theta)$.

The information matrix of θ can be found to be

$$I(\theta)=\theta_2^{-2}\begin{bmatrix} c_1 & c_3 \\ c_3 & c_2 \end{bmatrix}, \tag{5.10}$$

where

$$c_1=\int[f'(x)/f(x)]^2 f(x)\,dx,\ c_2=\int[1+xf'(x)/f(x)]^2 f(x)\,dx$$

and

$$c_3=\int x[f'(x)/f(x)]^2 f(x)\,dx.$$

Once again solving (3.2), we obtain $k(x)=(c_2+2c_3x+c_1x^2)^{-1/2}$. The resulting prior is given by

$$\pi_\theta(\theta_1,\theta_2)=\theta_2^{-1}\left(c_2\theta_2^2+2c_3\theta_1\theta_2+c_1\theta_1^2\right)^{-1/2},$$

which is also the reference prior of Berger and Bernardo in this situation. This example generalizes the example of Remark 5.2 and is considered by Mukerjee and Dey (1993) in estimating μ and σ. For a symmetric pdf $f(x)$, $c_3=0$, and if further $f(x)$ is normal, then $c_2=2$ and $c_1=1$.

EXAMPLE 5.4. We consider a version of the exponential regression model of Cox and Reid (1987) discussed by Mukerjee and Dey (1993). Suppose $(X_1, \dots, X_p)$ have the joint pdf

$$f(x_1, \dots, x_p; \theta) = \prod_{i=1}^{p} \left[\theta_2^{-1} \exp(-\theta_1 z_i) \exp\{-x_i \theta_2^{-1} \exp(-\theta_1 z_i)\} \right],$$

$x_1, \dots, x_p > 0$, $-\infty < \theta_1 < \infty$, $\theta_2 > 0$, $p \geq 2$, $z_1, \dots, z_p$ are constants not all equal satisfying $\sum_{i=1}^{p} z_i = 0$. Define $\gamma_2 = \sum_{i=1}^{p} z_i^2$. Here θ_1 is the parameter of interest and θ_2 is a nuisance parameter. From Mukerjee and Dey (1993) the information matrix of θ is given by $I(\theta) = \operatorname{diag}(\gamma_2, p\theta_1^{-2})$. Here Jeffreys' prior, the usual reference prior and the reverse reference prior are all identical and equal to

$$\pi_{\mathrm{J}}(\theta_1, \theta_2) = \pi_{\mathrm{R}}(\theta_1, \theta_2) = \pi_{\mathrm{RR}}(\theta_1, \theta_2) = \theta_2^{-1}. \tag{5.11}$$

Now consider the group of scale transformations $y_i = cx_i$, $i = 1, \dots, p$. Then $(\theta_1, \theta_2) \to (\phi_1, \phi_2) = (\theta_1, c\theta_2)$. The information matrix of $\phi = (\phi_1, \phi_2)$ is $I(\phi) = \operatorname{diag}(\gamma_2, p/\phi_2^2)$, which has structure the same as that of θ. Hence, any reasonable prior for (θ_1, θ_2) to estimate θ_1 should also be a reasonable prior for (ϕ_1, ϕ_2) to estimate ϕ_1. Proceeding as in Example 5.3, any such prior $\pi_\theta(\theta_1, \theta_2)$ should be of the form

$$\pi_\theta(\theta_1, \theta_2) = k(\theta_1)\theta_2^{-1} \tag{5.12}$$

for some arbitrary nonnegative function $k(\cdot)$. Since θ_1 and θ_2 are orthogonal, if we want to choose $k(\cdot)$ such that $\pi_\theta(\theta_1, \theta_2)$ is a matching prior for θ_1, then it follows from Tibshirani (1989) that $k(x) = 1$. The resulting prior is same as that given by (5.11).

REMARK 5.3. Since θ_1 and θ_2 are orthogonal, the prior given by (5.11) is also a matching prior when θ_2 is the parameter of interest.

REMARK 5.4. Example 5.4 reduces to the estimation of the ratio of two exponential means with $p = 2$. In this case $z_1 = -z_2 = z$ (say) and $\mu_1 = \theta_2^{-1} \exp(-\theta_1 z)$ and $\mu_2 = \theta_2^{-1} \exp(\theta_1 z)$. Define $\psi_1 = \mu_2/\mu_1 = \exp(2z\theta_1)$ and $\psi_2 = \mu_1 \mu_2 = \theta_2^{-2}$; ψ_1 is the parameter of interest and ψ_2 is nuisance. The prior for (ψ_1, ψ_2) obtained from $\pi_\theta(\theta_1, \theta_2) = \theta_2^{-1}$ by variable transformation is $\pi_\psi(\psi_1, \psi_2) \propto \psi_1^{-1}\psi_2^{-3/2}\psi_2^{1/2} = (\psi_1\psi_2)^{-1}$. This prior was obtained by Mukerjee and Dey (1993).

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

BERGER, J. (1992). Discussion of "Non-informative priors," by J. K. Ghosh and R. Mukerjee. In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 205–206. Oxford Univ. Press.

BERGER, J. and BERNARDO, J. M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Amer. Statist. Assoc.* **84** 200–207.

BERGER, J. and BERNARDO, J. M. (1992a). On the development of reference priors (with discussion). In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 35–60. Oxford Univ. Press.

BERGER, J. and BERNARDO, J. M. (1992b). Ordered group reference priors with application to the multinomial problem. *Biometrika* **79** 25–37.

BERGER, J. and BERNARDO, J. M. (1992c). Reference priors in a variance components problem. In *Bayesian Analysis in Statistics and Econometrics* (P. K. Goel and N. S. Iyengar, eds.) 177–194. Springer, New York.

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

CLARKE, B. and SUN, D. (1993). Reference priors under the chi-square distance. Unpublished manuscript.

CLARKE, B. and WASSERMAN, L. (1992). Information tradeoff. Technical Report 558, Carnegie Mellon Univ.

COX, D. R. and REID, N. (1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49** 1–39.

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

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

GHOSH, J. K. and MUKERJEE, R. (1992). Non-informative priors (with discussion). In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 195–210. Oxford Univ. Press.

JEFFREYS, H. (1961). *Theory of Probability*. Oxford Univ. Press.

LAPLACE, P. (1812). *Theorie Analytique des Probabilities*. Courcier, Paris.

MUKERJEE, R. and DEY, D. K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics. *Biometrika* **80** 499–505.

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

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. Wiley, New York.

STEIN, C. (1959). An example of wide discrepancy between fiducial and confidence interval. *Ann. Math. Statist.* **30** 877–880.

STEIN, C. (1985). On coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics* 485–514. PWN, Polish Scientific Publishers, Warsaw.

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

WELCH, B. L. and PEERS, H. W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *J. Roy. Statist. Soc. Ser. B* **25** 318–329.

DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602-1952

DEPARTMENT OF STATISTICS
103 GRIFFIN-FLOYD HALL
P.O. BOX 118545
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611-8545