NONINFORMATIVE PRIORS FOR MAXIMAL INVARIANT PARAMETER IN GROUP MODELS

by

Gauri Sankar Datta and Jayanta Kumar Ghosh
University of Georgia
Purdue University and
Indian Statistical Institute

Technical Report #94-24C

Department of Statistics
Purdue University

September 1994

◇

NONINFORMATIVE PRIORS FOR MAXIMAL INVARIANT PARAMETER IN GROUP MODELS

GAURI SANKAR DATTA*
Department of Statistics
University of Georgia
Athens, Georgia 30602-1952, USA

JAYANTA KUMAR GHOSH†
Department of Statistics
Purdue University
West Lafayette, Indiana
and Indian Statistical Institute
Calcutta, India

ABSTRACT

For an Euclidean group G acting freely and transitively on the parameter space, we derive, among several noninformative priors, the reference priors of Berger-Bernardo and Chang-Eaves for our parameter of interest θ_1, a maximal invariant parametric function. Identifying the nuisance parameter vector with the group element, we derive a simple structure of the information matrix which is used to obtain different noninformative priors. We compare these priors using the marginalization paradox and the probabilty matching criteria. The Chang-Eaves and the Berger-Bernardo reference priors appear to be the most attractive choice. Several illustrative examples are considered.

AMS 1990 Subject Classification: 62F15

Keywords and phrases: Right invariant Haar density, left invariant Haar density, reference prior, information matrix, marginalization paradox, probability matching prior.

* Research partially supported by NSF Grant DMS 89-230-71 of Professor Jim Berger of Purdue University.

† Research partially supported by NSF Grant Number DMS 93-077-27.

Typeset by $\mathcal{A}_\mathcal{M}\mathcal{S}$-TEX

1. INTRODUCTION

There has been some recent interest in noninformative priors which are often used in Bayesian analysis, at least as a first step, vide Berger and Bernardo (1992a). We study various noninformative priors when there is a group G that leaves the model invariant and the parameter of interest is a scalar maximal invariant. Our focus is on probability matching properties of posterior credibility regions of the parameter of interest upto $O(n^{-1})$ as in Peers(1965), Stein (1985), Tibshirani (1989) and Ghosh and Mukerjee (1992a) just to name a few. We also examine issues related to the marginalization paradox, treated in Dawid et al. (1973), the order in which the Berger-Bernardo (1989, 1992a,b) algorithm is to be implemented (i.e., whether one should have reference or reverse reference priors, see Berger (1992)), and the choice between nuisance parameters taken in a single group or in multiple groups (vide Berger and Bernardo (1992a,b)).

Our main interest is in studying the properties of the Berger-Bernardo reference prior in view of the relative ease with which it can be calculated and the many examples such as product of two normal means (Berger and Bernardo (1989)) where the use of reference prior can be seen to lead to satisfactory inference. We also study the reference prior of Chang and Eaves (1990) obtained by applying the Berger-Bernardo algorithm (1989) starting with a given conditional prior distribution which is given by the right invariant Haar measure on the group G. Our study indicates that in the present context the Chang-Eaves reference prior seems to be the most attractive both from the point of view of eliminating marginalization paradox (Dawid et al., 1973, p 198, Chang and Eaves, 1990, p 1597) and probability matching upto $O(n^{-1})$, and the Berger-Bernardo reference prior is very often, though not always, identical with the Chang-Eaves prior. Of course, unlike the Chang-Eaves prior, it can be applied even when there is no group leaving the model invariant.

The paper is organized as follows. In Section 2 we make assumptions similar to those in Dawid et al. (1973) that the group G acts transitively and freely on the parameter space and obtain a simple structure of the information matrix of the parameter by identifying the group element as the vector of nuisance parameters. Using this structure of the information matrix we derive in a fairly direct way the Berger-Bernardo, Chang-Eaves, Ghosh-Mukerjee and the reverse reference priors. The original Chang-Eaves computations are much more delicate but do not involve the strong assumptions of Section 2. Among other results we also show that the Berger-Bernardo reference prior, when maximal invariant parameter in the group model is the parameter of interest and all the nuisance parameters are taken in a single group, takes the left invariant Haar density on the group G as the conditional distribution of the nuisance parameters given the parameter of interest. In fact one of the advantages of taking nuisance parameters in multiple groups instead of in a single group in the Berger-Bernardo algorithm is to get a unimodular group (of transformations) at each stage so that the algorithm leads to the Chang-Eaves reference prior. For parameter group ordering one may, for example, refer to Berger and Bernardo (1992a,b).

In Section 3 we compare different noninformative priors by using the marginalization paradox and probability matching criteria. In Section 3.1 we show by examples that the marginalization paradox may occur if the Berger-Bernardo reference prior is used with all the nuisance parameters forming one group. However multiple grouping of nuisance parameters in the same examples will lead to a Berger-Bernardo refernce prior free from the paradox. In Section 3.2 we present the probability matching equation upto $O(n^{-1})$ for our parameter of interest. We also derive a set of equations for the right invariant Haar density which implies that the Chang-Eaves reference prior, unlike the Berger-Bernardo reference prior, always satisfies the probability matching equation. It is also clear that in a certain sense the converse is true.

In Section 4 we indicate why under the present assumptions the Berger-Bernardo reference prior is to be preferred to the reverse reference prior. Several other interesting examples, including the Hotelling's T^2 and the two others due to Stein, are treated in Section 5.

2. REFERENCE PRIORS IN GROUP MODELS

2.1 Notations, Assumptions and Information Matrix

Suppose X_i's $i = 1, \ldots, n$ are independently and identically distributed with density $f_{\boldsymbol{\theta}}$ (with respect to the Lebesgue measure) $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_{(2)}) \in \Theta$, and $\boldsymbol{\theta}_{(2)} = (\theta_2, \ldots, \theta_p) \in \Theta_{(2)}(\subset R^{p-1})$, $\theta_1 \in \Theta_1(\subset R)$ and $\Theta = \Theta_1 \times \Theta_{(2)}$. We assume that there is a group of transformations H on range of X_1 which induces a group of transformations $G = \{g\}$ on Θ. We assume that G is a Lie group, and the given decomposition of Θ is such that $\Theta_{(2)} = G$ and G acts on Θ freely and transitively by left multiplication in G. Then it follows from Dawid et al. (1973) that θ_1 is maximal invariant and $\boldsymbol{\theta}_{(2)}$ is equivariant under G, and the group element g may be taken as identical to $\boldsymbol{\theta}_{(2)}$. We also assume that θ_1 is our parameter of interest and $\boldsymbol{\theta}_{(2)} = \boldsymbol{g}$ is the vector of nuisance parameters. We also make the same assumptions as in Ghosh and Mukerjee (1992b, p 868).

We will now derive a simple structure of the information matrix in $(\theta_1, \boldsymbol{g})$ parametrization as a consequence of our assumptions. Let $I_{\boldsymbol{\theta}}$ be the per unit observation information matrix of $\boldsymbol{\theta}$, a $p \times p$ positive definite matrix for all $\boldsymbol{\theta}$. For $\boldsymbol{g} \in G$ and $\boldsymbol{\psi} = (\psi_1, \boldsymbol{\psi}_{(2)}) \in \Theta$, consider the transformation $\boldsymbol{\theta}$ to $\boldsymbol{\psi}$ defined by the group action $\boldsymbol{\theta} \to \boldsymbol{g}\boldsymbol{\theta}$ where $\psi_1 = \theta_1, \boldsymbol{\psi}_{(2)} = \boldsymbol{g}\boldsymbol{\theta}_{(2)}$. Let $J_{\boldsymbol{g}}(\boldsymbol{\theta} \to \boldsymbol{\psi})$, given by $((\frac{\partial \psi_j}{\partial \theta_i}))_{i,j=1,\dots,p}$, be the matrix of Jacobian of transformation $\boldsymbol{\theta} \to \boldsymbol{\psi}$ and $I_{\boldsymbol{\psi}}$ be the information matrix of $\boldsymbol{\psi}$. Denoting the $(p-1) \times (p-1)$ matrix of Jacobian of transformation $\boldsymbol{\theta}_{(2)} \to \boldsymbol{g}\boldsymbol{\theta}_{(2)}$ by $J_{\boldsymbol{g}}(\boldsymbol{\theta}_{(2)} \to \boldsymbol{g}\boldsymbol{\theta}_{(2)})$, we have

$$J_{\boldsymbol{g}}(\boldsymbol{\theta} \to \boldsymbol{\psi}) = \textit{Block diagonal}\Big(1, J_{\boldsymbol{g}}(\boldsymbol{\theta}_{(2)} \to \boldsymbol{g}\boldsymbol{\theta}_{(2)})\Big). \tag{2.1}$$

We make conforming partition of the information matrix $I_{\boldsymbol{\theta}}$ of $(\theta_1, \boldsymbol{\theta}_{(2)})$ as

$$I_{\boldsymbol{\theta}} = \begin{bmatrix} I_{11}(\boldsymbol{\theta}) & \boldsymbol{I}_{12}(\boldsymbol{\theta}) \\ \boldsymbol{I}_{21}(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) \end{bmatrix}.$$

Transforming the information matrix of $\boldsymbol{\theta}$ to that of $\boldsymbol{\psi}$ as in Berger and Bernardo (1989), for example, we get after some simplifications

$$I_{\boldsymbol{\psi}}(\psi_1, \boldsymbol{\psi}_{(2)}) = \begin{bmatrix} I_{11}(\psi_1, \boldsymbol{g}^{-1}\boldsymbol{\psi}_{(2)}) & \boldsymbol{I}_{12}(\psi_1, \boldsymbol{g}^{-1}\boldsymbol{\psi}_{(2)})(J_{\boldsymbol{g}}^T(\boldsymbol{\psi}_{(2)}))^{-1} \\ J_{\boldsymbol{g}}^{-1}(\boldsymbol{\psi}_{(2)})\boldsymbol{I}_{21}(\psi_1, \boldsymbol{g}^{-1}\boldsymbol{\psi}_{(2)}) & J_{\boldsymbol{g}}^{-1}(\boldsymbol{\psi}_{(2)})I_{22}(\psi_1, \boldsymbol{g}^{-1}\boldsymbol{\psi}_{(2)})(J_{\boldsymbol{g}}^T(\boldsymbol{\psi}_{(2)}))^{-1} \end{bmatrix}$$

where, for brevity, $J_{\boldsymbol{g}}(\boldsymbol{\psi}_{(2)}) = J_{\boldsymbol{g}}(\boldsymbol{\theta}_{(2)} \to \boldsymbol{g}\boldsymbol{\theta}_{(2)})|_{\boldsymbol{\theta}_{(2)} = \boldsymbol{g}^{-1}\boldsymbol{\psi}_{(2)}}$ and $\boldsymbol{g}^{-1}$ is the group inverse of $\boldsymbol{g}$. If we take $\boldsymbol{\theta}_{(2)} = \boldsymbol{e}$, the group identity, we have $\boldsymbol{\psi}_{(2)} = \boldsymbol{g}$. Now if we replace ψ_1 by θ_1, from the information matrix given above we get the information matrix of $(\theta_1, \boldsymbol{g})$. From now on we represent the parameter vector by $(\theta_1, \boldsymbol{g})$ whose information matrix is given by

$$I(\theta_1, \boldsymbol{g}) = \begin{bmatrix} I_{11}(\theta_1, \boldsymbol{e}) & \boldsymbol{I}_{12}(\theta_1, \boldsymbol{e})(J_{\boldsymbol{g}}^T(\boldsymbol{g}))^{-1} \\ J_{\boldsymbol{g}}^{-1}(\boldsymbol{g})\boldsymbol{I}_{21}(\theta_1, \boldsymbol{e}) & J_{\boldsymbol{g}}^{-1}(\boldsymbol{g})I_{22}(\theta_1, \boldsymbol{e})(J_{\boldsymbol{g}}^T(\boldsymbol{g}))^{-1} \end{bmatrix} \tag{2.2}$$

where $J_{\boldsymbol{g}}(\boldsymbol{g}) = J_{\boldsymbol{g}}(\boldsymbol{\theta}_{(2)} \to \boldsymbol{g}\boldsymbol{\theta}_{(2)})|_{\boldsymbol{\theta}_{(2)} = \boldsymbol{e}}$.

Remark 1. We will see later in our examples that it is always advantageous to compute the information matrix $I(\theta_1,\ \boldsymbol{g})$ by using equation (2.2) since it is much easier to compute $I(\theta_1,\ \boldsymbol{e})$, and the matrix $J_{\boldsymbol{g}}^{-1}(\boldsymbol{g})$ is easy to find.

2.2 Different Noninformative Priors

We will now derive from the information matrix given above the Berger-Bernardo, Chang-Eaves and Ghosh-Mukerjee reference priors, treating θ_1 as the parameter of interest and $\boldsymbol{g}$ as nuisance parameter vector. Due to the special structure of the information matrix in (2.2), in each of Berger-Bernardo, Chang-Eaves and Ghosh-Mukerjee reference priors, the marginal prior density function of θ_1 does not depend on the given conditional prior density function $\pi_2(\boldsymbol{g}|\theta_1)$. We thus have the following simple proposition which follows immediately from Ghosh and Mukerjee (1992a, p 197).

PROPOSITION 1.

For any given conditional prior density function $\pi_2(\boldsymbol{g}|\theta_1)$, the marginal prior density function $\pi_1(\theta_1)$ of θ_1 which asymptotically maximizes the expected Kullback-Leibler divergence between the marginal posterior and the prior density functions of θ_1 is given by

$$\pi_1(\theta_1) \propto I_{11.2}^{1/2}(\theta_1) \tag{2.3}$$

where $I_{11.2}(\theta_1) = I_{11}(\theta_1, e) - \boldsymbol{I}_{12}(\theta_1, \boldsymbol{e}) I_{22}^{-1}(\theta_1, \boldsymbol{e}) \boldsymbol{I}_{21}(\theta_1, \boldsymbol{e})$.

Before we find the Berger-Bernardo and the Chang-Eaves reference prior we will define the left and the right invariant Haar density on G. When the group G is Euclidean, as assumed by us, it follows from Berger (1985, pp 408-409) that the left and the right invariant Haar densities $h^l(\boldsymbol{g})$ and $h^r(\boldsymbol{g})$ on G are respectively given by

$$h^l(\boldsymbol{g}) = \frac{1}{abs|J_{\boldsymbol{g}}(\boldsymbol{g})|}, \; h^r(\boldsymbol{g}) = \frac{1}{abs|J_{\boldsymbol{g}}^r(e)|} \tag{2.4}$$

where $J_{\boldsymbol{g}}^r(\boldsymbol{x})$ is the $(p-1) \times (p-1)$ matrix of Jacobian of transformation $\boldsymbol{x} \to \boldsymbol{xg}$.

In Berger-Bernardo reference prior, the conditional prior density function $\pi_2(\boldsymbol{g}|\theta_1)$ is given by

$$\pi_2(\boldsymbol{g}|\theta_1) \propto |I_{22}(\theta_1, \boldsymbol{g})|^{1/2} \propto abs|J_{\boldsymbol{g}}(\boldsymbol{g})|^{-1}. \tag{2.5}$$

From $(2.3) - (2.5)$, the Berger-Bernardo reference prior $\pi_{BB}(\theta_1, \boldsymbol{g})$ is given by

$$\pi_{BB}(\theta_1, \boldsymbol{g}) \propto I_{11.2}^{1/2}(\theta_1) \times h^l(\boldsymbol{g}). \tag{2.6}$$

Even if G is not Euclidean but only locally so, one can show that the Berger-Bernardo conditional distribution of $\boldsymbol{g}$ is invariant under the one-to-one transformation of the nuisance parameter and in particular under $\boldsymbol{g} \to \boldsymbol{hg}$ for $\boldsymbol{h} \in G$. Hence it must be the left invariant Haar density.

The Chang-Eaves reference prior $\pi_{CE}(\theta_1, \boldsymbol{g})$ uses the right invariant Haar density $h^r(\boldsymbol{g})$ on G as the conditional prior density function of $\boldsymbol{g}$ given θ_1 and is given by

$$\pi_{CE}(\theta_1, \boldsymbol{g}) \propto I_{11.2}^{1/2}(\theta_1) \times h^r(\boldsymbol{g}). \tag{2.7}$$

Ghosh and Mukerjee (1992a) obtained a reference prior by maximizing the Kullback-Leibler divergence mentioned earlier less a penalty term for the deviation of the conditional prior density function from a uniform prior. This reference prior, to be denoted by $\pi_{GM}(\theta_1, \boldsymbol{g})$, is given by

$$\pi_{GM}(\theta_1, \boldsymbol{g}) \propto I_{11.2}^{1/2}(\theta_1) \tag{2.8}$$

Remark 2. From the discussion of Dawid et al. (1973, p 197) it follows that statisticians will often want their prior probability density function to factorize into $\pi_1(\theta_1)$ and $\pi_2(\boldsymbol{g})$, where $\pi_2(\boldsymbol{g})$ is at least relatively invariant. From (2.6) – (2.8), we see that in each case the reference prior factorizes as mentioned above. Also note that while the Chang-Eaves reference prior uses the right Haar density as the conditional probability density function of $\boldsymbol{g}$, the Berger-Bernardo prior uses the left Haar density and the Ghosh-Mukerjee reference prior uses the uniform density which is often relatively invariant, vide Dawid et al. (1973, p 196). In our subsequent discussion we will assume this structure for our prior probability density function. Incidentally, we should mention that Jeffreys' prior in group model, given by, $\pi_J(\theta_1, \boldsymbol{g}) \propto |I(\theta_1, \boldsymbol{e})|^{1/2} h^l(\boldsymbol{g})$, also has this structure.

3. COMPARISON OF DIFFERENT NONINFORMATIVE PRIORS

We will compare our different noninformative priors by using marginalization paradox and the probability matching criteria.

3.1 Marginalization Paradox

It is well known from Dawid et al. (1973) and Chang and Eaves (1990) that for any prior $\pi(\theta_1, \boldsymbol{g})$ of the form $\pi(\theta_1, \boldsymbol{g}) \propto \pi_1(\theta_1) \times h^r(\boldsymbol{g})$, and in particular the Chang-Eaves prior, does not suffer from marginalization paradox. However, the following examples show that this may not necessarily be true for the other reference priors when the nuisance parameters are taken together in a single group.

Example 1. $\boldsymbol{X}_i = (X_{1i},\ X_{2i})^T$, $i = 1, \ldots, n$ are i.i.d. $N_2(\boldsymbol{\mu},\ \sigma^2 I_2)$ where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and I_2 is a 2×2 identity matrix. Let $\theta_1 = (\mu_1 - \mu_2)/\sigma$ be the parameter of interest, and $\theta_2 = \mu_2$ and $\theta_3 = \sigma$. For the group of transformations $H = \{\boldsymbol{g} | \boldsymbol{g} = (g_2,\ g_3), -\infty < g_2 < \infty,\ g_3 > 0\}$ in the range of $\boldsymbol{X}_1$ defined by $\boldsymbol{g}X_1 = g_3 \boldsymbol{X}_1 + g_2 \mathbf{1}$, the induced group of transformations on the parameter space is $G = \{\boldsymbol{g}\}$ with the transformation defined by $\boldsymbol{g\theta} = (\theta_1,\ g_3\ \theta_2\ +\ g_2,\ g_3\ \theta_3)$. The maximal invariant parameter is θ_1. For $\boldsymbol{g} = (g_2, g_3), \boldsymbol{h} = (h_2, h_3) \in G$, the group operation is defined by $\boldsymbol{gh} = (g_3 h_2 + g_2, g_3 h_3)$. Since our assumptions about G are satisfied, we can identify $\theta_2 = g_2$ and $\theta_3 = g_3$. It can be checked that

$$I(\theta_1, \boldsymbol{e}) = \begin{bmatrix} 1 & 1 & \theta_1 \\ 1 & 2 & \theta_1 \\ \theta_1 & \theta_1 & 4 + \theta_1^2 \end{bmatrix} \tag{3.1}$$

and the per unit observation information matrix is $I(\theta_1, g_2, g_3) = Diag(1, g_3^{-1}, g_3^{-1}) I(\theta_1, \boldsymbol{e})$ $Diag(1, g_3^{-1}, g_3^{-1})$, and $I_{11.2}(\theta_1) = 4/(8 + \theta_1^2)$. Thus the Berger-Bernardo and the Ghosh-Mukerjee priors for the parameter grouping $\{\theta_1, (g_2, g_3)\}$ are given by

$$\pi_{BB}(\theta_1, \boldsymbol{g}) \propto (8 + \theta_1^2)^{-1/2} \times g_3^{-2},\ \pi_{GM}(\theta_1, \boldsymbol{g}) \propto (8 + \theta_1^2)^{-1/2} \tag{3.2}$$

none of which takes $h^r(\boldsymbol{g}) \propto g_3^{-1}$ as the conditional probability density function of $\boldsymbol{g}$.

Define $\bar{X}_j = n^{-1} \sum_{i=1}^n X_{ji},\ j = 1, 2,\ S^2 = (2n-2)^{-1} \sum_{j=1}^2 \sum_{i=1}^{n} (X_{ji} - \bar{X}_j)^2$ and $t = (\bar{X}_1 - \bar{X}_2)/S$. Then it can be shown that for priors of the form $\pi_1(\theta_1) \times g_3^{-r}$, the marginal posterior density function of θ_1 depends on the data only through t and is proportional to

$$\pi_1(\theta_1) \times \int_0^\infty u^{2n+r-3} exp[-\{(n-1)u^2 + \frac{n}{4}(tu - \theta_1)^2\}] du. \tag{3.3}$$

It can also be shown that the sampling density of t is proportional to

$$\int_0^\infty u^{2n-2} exp[-\{(n-1)u^2 + \frac{n}{4}(tu - \theta_1)^2\}] du \tag{3.4}$$

From the last two equations it is obvious that only priors of the form $\pi_1(\theta_1) \times g_3^{-1}$ will be free from the paradox. In particular, the Berger-Bernardo and the Ghosh-Mukerjee prior will generate the paradox.

Remark 3. For the one at a time parameter grouping $\{\theta_1, g_2, g_3\}$ *or* $\{\theta_1, g_3, g_2\}$, Berger-Bernardo refence prior will be identical to the Chang-Eaves prior. Since the groups are commutative, the left and the right Haar densities will be identical and at each stage the conditional probability density function will come out as the Haar density. Finally when combined together they will give the right Haar density as the conditional probability density function for the nuisance parameter vector.

Example 2. $\boldsymbol{X}_i = (X_{1i}, X_{2i})^T$, $i = 1, \dots, n$ are i.i.d. $N_2(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. Let $\theta_1 = \rho$ be the parameter of interest, and $\theta_2 = \mu_1, \theta_3 = \mu_2, \theta_4 = \sigma_1$ and $\theta_5 = \sigma_2$. Consider the affine linear group of transformations on R^2 denoted by $H = \{g | g = (\boldsymbol{a}, A)\}$ in the range of $\boldsymbol{X}_1$, where g consists of the pair $(\boldsymbol{a}, A)$ described by $\boldsymbol{a} = (g_2, g_3)^T$, a 2-vector and $A = Diag(g_4, g_5)$, a 2×2 matrix where $-\infty < g_2 < \infty, -\infty < g_3 < \infty, 0 < g_4 < \infty, 0 < g_5 < \infty$, and the group action is $g\boldsymbol{X}_1 = A\boldsymbol{X}_1 + \boldsymbol{a}$. Induced transformation on the parameter space is $G = \{g\}$ defined by $g\theta = (\theta_1,\ g_4\theta_2 + g_2,\ g_5\theta_3 + g_3,\ g_4\theta_4,\ g_5\theta_5)$ and the θ_1 parameter is maximal invariant. For $g, h \in G$, the group operation is defined by $gh = (g_4h_2 + g_2, g_5h_3 + g_3, g_4h_4, g_5h_5)$. Since our assumptions about G are satisfied we can identify $\theta_2 = g_2, \theta_3 = g_3, \theta_4 = g_4$ and $\theta_5 = g_5$. It can be directly found that

$$I(\theta_1, e) = (1 - \theta_1^2)^{-1} \begin{bmatrix} \frac{1+\theta_1^2}{1-\theta_1^2} & 0 & 0 & -\theta_1 & -\theta_1 \\ 0 & 1 & -\theta_1 & 0 & 0 \\ 0 & -\theta_1 & 1 & 0 & 0 \\ -\theta_1 & 0 & 0 & 2-\theta_1^2 & -\theta_1^2 \\ -\theta_1 & 0 & 0 & -\theta_1^2 & 2-\theta_1^2 \end{bmatrix}. \tag{3.5}$$

Then by (2.2) and (3.5) the per unit observation information matrix is $I(\theta_1, g_2, g_3, g_4, g_5)$ $= Diag(1, g_4^{-1}, g_5^{-1}, g_4^{-1}, g_5^{-1})\ I(\theta_1, e) Diag(1, g_4^{-1}, g_5^{-1}, g_4^{-1}, g_5^{-1})$. Also note that $I_{11.2}(\theta_1)$ $= (1 - \theta_1^2)^{-2}$, $h^r(g) = (g_4g_5)^{-1}$ and $h^l(g) = (g_4g_5)^{-2}$. So the Chang-Eaves prior is given by $(1 - \theta_1^2)^{-1}h^r(g)$, the Berger-Bernardo reference prior for the parameter grouping $\{\theta_1, (g_2, g_3, g_4, g_5)\}$ is given by $(1 - \theta_1^2)^{-1}h^l(g)$ and the Ghosh-Mukerjee reference prior is given by $(1 - \theta_1^2)^{-1}$.

Let r denote the sample correlation coefficient. It can be checked that for priors of the form $\pi_1(\theta_1)(g_4g_5)^{-s}$, the marginal posterior density function of θ_1 depends only on r and is proportional to

$$\pi_1(\theta_1)(1 - \theta_1^2)^{(n+2s-3)/2} \int_0^\infty \frac{dz}{z}(z - 2\theta_1 r + z^{-1})^{-(n+s-2)}. \tag{3.6}$$

It can be found from Muirhead (1982, p 153) after simplifications that the marginal sampling probability density function of r is proportional to

$$(1 - \theta_1^2)^{(n-1)/2}(1 - r^2)^{(n-4)/2} \int_0^\infty \frac{dz}{z}(z - 2\theta_1 r + z^{-1})^{-(n-1)}. \tag{3.7}$$

From the last two equations it is obvious that for $s = 1$ in the given class of priors there will be no marginalization paradox. In particular, the Berger-Bernardo reference prior which, in this case, is same as the Jeffreys' prior, and the Ghosh-Mukerjee reference prior will all generate the paradox.

Remark 4. The Chang-Eaves reference prior is identical to the prior proposed by Lindley (1965) and Bayarri(1981). A different prior for this problem, proposed by Geisser (1965) and Dawid et al. (1973), expressed in our notations, is proportional to $(1-\theta_1^2)^{-3/2}g_4^{-1}g_5^{-1}$ and avoids the paradox. However, we prove in Section 3.2 that among all these priors only the Chang-Eaves prior satisfies the probability matching equation for θ_1.

Remark 5. It can be checked that for the parameter grouping $\{\theta_1, (g_2, g_3), (g_4, g_5)\}$, or any other grouping with θ_1 as the first ordered group and any permutation or further splitting of any one or both the groups (g_2, g_3) and (g_4, g_5), the Berger-Bernardo reference prior will come out same as the Chang-Eaves reference prior.

3.2 Probability Matching of Noninformative Priors

In this section we study the probability matching properties of the noninformative priors considered in the previous section. A prior density $\pi(\theta_1, \boldsymbol{g})$ which matches the posterior and frequentist probabilities of the set $\{\frac{\sqrt{n}(\theta_1 - \hat{\theta}_1)}{\sqrt{b_1}} \leq z\}$ for all z upto $O_p(n^{-1})$ and for all $\boldsymbol{\theta} = (\theta_1, \boldsymbol{g})$ in compact sets is called a probability matching prior for θ_1. Here $\hat{\boldsymbol{\theta}}$ is the MLE or the posterior mode of $\boldsymbol{\theta}$ corresponding to the prior π and b_1 is the asymptotic posterior variance of $\sqrt{n}(\theta_1 - \hat{\theta}_1)$, upto $O_p(n^{-1})$. Such a prior may be sought in an attempt to reconcile a frequentist and Bayesian approach as in Peers(1965), or to find or in some sense to validate a noninformative prior as in Ghosh and Mukerjee (1992a) and Tibshirani (1989), or to construct frequentist confidence sets as in Stein (1985). Berger and Bernardo (1989) and Ye and Berger (1991) use simulations instead of probability matching equations to compare different noninformative priors. Under the regularity assumptions of Section 2, it follows that π will be probability matching for θ_1 if and only if

$$\frac{\partial}{\partial \theta_1}\left(\sqrt{I^{11}}(\theta_1, \boldsymbol{g})\pi(\theta_1, \boldsymbol{g})\right) + \sum_{i=2}^{p} \frac{\partial}{\partial g_i}\left(\frac{I^{i1}(\theta_1, \boldsymbol{g})}{\sqrt{I^{11}}(\theta_1, \boldsymbol{g})}\pi(\theta_1, \boldsymbol{g})\right) = 0$$

where $I^{ij}(\theta_1, \boldsymbol{g})$ is the $(i, j)th$ element of $I^{-1}(\theta_1, \boldsymbol{g})$. This equation is due to Peers (1965).

Now define $\boldsymbol{a}(\theta_1)$ and $s_i(\theta_1, \boldsymbol{g})$ for i=2,...,p by

$$\boldsymbol{a}(\theta_1) = (a_2(\theta_1), \ldots, a_p(\theta_1))^T = I_{22}^{-1}(\theta_1, \boldsymbol{e})\boldsymbol{I}_{21}(\theta_1, \boldsymbol{e}), s_i(\theta_1, \boldsymbol{g}) = \sum_{j=2}^{p} a_j(\theta_1) J_{\boldsymbol{g}ji}(\boldsymbol{g}). \quad (3.8)$$

By (3.8) and the fact that $I^{11}(\theta_1, \boldsymbol{g}) = I_{11.2}^{-1}(\theta_1)$, the last differential equation reduces to

$$\frac{\partial}{\partial \theta_1}\left(I_{11.2}^{-1/2}(\theta_1)\pi(\theta_1, \boldsymbol{g})\right) = \sum_{i=2}^{p} \frac{\partial}{\partial g_i}\left(I_{11.2}^{-1/2}(\theta_1)\ s_i(\theta_1, \boldsymbol{g})\pi(\theta_1, \boldsymbol{g})\right).$$

For priors of the form $\pi_1(\theta_1)\pi_2(\boldsymbol{g})$, this equation simplifies to

$$\frac{\partial}{\partial \theta_1}\left(I_{11.2}^{-1/2}(\theta_1)\pi_1(\theta_1)\pi_2(\boldsymbol{g})\right) = \sum_{j=2}^{p} I_{11.2}^{-1/2}(\theta_1)\pi_1(\theta_1)a_j(\theta_1)\sum_{i=2}^{p}\frac{\partial}{\partial g_i}\left(J_{\boldsymbol{g}ji}(\boldsymbol{g})\pi_2(\boldsymbol{g})\right). \quad (3.9)$$

In particular, since for Berger-Bernardo or Chang-Eaves or Ghosh-Mukerjee priors $\pi_1(\theta_1)$ is proportional to $I_{11.2}^{1/2}(\theta_1)$, the last equation reduces to

$$\sum_{j=2}^{p} a_j(\theta_1) \sum_{i=2}^{p} \frac{\partial}{\partial g_i}\left(J_{\boldsymbol{g}ji}(\boldsymbol{g})\ \pi_2(\boldsymbol{g})\right) = 0. \quad (3.10)$$

A sufficient condition for (3.10) to hold is that the following $(p-1)$ equations hold.

$$\sum_{i=2}^{p} \frac{\partial}{\partial g_i}\left(J_{\boldsymbol{g}ji}(\boldsymbol{g})\ \pi_2(\boldsymbol{g})\right) = 0 \quad \text{for} \quad j = 2, \ldots, p. \quad (3.11)$$

We now show that (3.11) holds when $\pi_2(\boldsymbol{g}) = h^r(\boldsymbol{g})$. For $\boldsymbol{h} \in G$ and for any smooth function $\phi(\boldsymbol{g})$ vanishing outside a compact set, define a function on G by

$$u(\boldsymbol{h}) = \int_G \phi(\boldsymbol{gh})h^r(\boldsymbol{g})d\boldsymbol{g}.$$

Note that $u(\boldsymbol{h})$ is a constant and consequently its derivative with respect to any component of $\boldsymbol{h}$ vanishes everywhere and in particular at $\boldsymbol{h} = \boldsymbol{e}$. Since this is true for all such ϕ mentioned above, one can show by integration by parts that $\pi_2 = h^r$ is a solution of (3.11). We can summarize this result in the following theorem.

THEOREM 1. *The Chang-Eaves reference prior $\pi_{CE}(\theta_1, \boldsymbol{g})$ given in (2.7) is a probability matching prior for θ_1.*

Remark 6. The only solution of (3.11) is $h^r(\boldsymbol{g})$ by reversing the steps above. By Theorem 1, for any family of distributions remaining invariant under a group H with the induced group G on the parameter space, the Chang-Eaves reference prior will always be probability matching for θ_1. On the other hand, if a prior of the form $I_{11.2}^{1/2}(\theta_1)\ \pi_2(\boldsymbol{g})$ has to be probability matching for θ_1 for all families of distributions remaining invariant under a group H with the induced group G on the parameter space, then (3.10) will imply (3.11) and so it must be the Chang-Eaves reference prior. In particular, the Berger-Bernardo reference prior for parameter grouping $\{\theta_1, \boldsymbol{g}\}$ and the Ghosh-Mukerjee prior may not always be probability matching for θ_1. We have the following examples.

Example 1(continued). Here $p = 3$ and from Example 1 we get $\boldsymbol{a}(\theta_1) = (8 + \theta_1^2)^{-1}(4,\ \theta_1)^T$ and $J_{\boldsymbol{g}}(\boldsymbol{g}) = g_3 I_2$. The probability matching equation for priors of the form $I_{11.2}^{1/2}(\theta_1)\pi_2(\boldsymbol{g})$ is

$$4\frac{\partial}{\partial g_2}\Big(g_3\pi_2(g)\Big) + \theta_1\frac{\partial}{\partial g_3}\Big(g_3\pi_2(g)\Big) = 0 \quad \text{for all} \quad \theta_1 \tag{3.12}$$

and it is solved only by $h^r(\boldsymbol{g})$. In particular, the Berger-Bernardo and the Ghosh-Mukerjee reference priors are not probability matching.

Example 2(continued). Here $p = 5$ and from Example 2 we get $\boldsymbol{a}(\theta_1) = -\frac{\theta_1}{1-\theta_1^2}(0\ 0\ 1\ 1)^T$ and $J_{\boldsymbol{g}}(\boldsymbol{g}) = Diag(g_4, g_5, g_4, g_5)$. The probability matching equation for priors of the form $I_{11.2}^{1/2}(\theta_1)\pi_2(\boldsymbol{g})$ is

$$\frac{\partial}{\partial g_4}\Big(g_4\pi_2(\boldsymbol{g})\Big) + \frac{\partial}{\partial g_5}\Big(g_5\pi_2(\boldsymbol{g})\Big) = 0 \tag{3.13}$$

which is solved by $\pi_2(\boldsymbol{g}) = h^r(\boldsymbol{g})d(g_2, g_3)$ where $d(g_2, g_3)$ is arbitrary. In this example neither the Berger-Bernardo nor the Ghosh-Mukerjee reference prior is probability matching. Also since the Chang-Eaves reference prior is probability matching for θ_1, it follows easily that the prior $(1-\theta_1^2)^{-3/2}g_4^{-1}g_5^{-1}$ in Remark 4 is not probability matching.

4. REVERSE REFERENCE PRIORS

The reverse reference prior, to be denoted by $\pi_{RR}(\theta_1, \boldsymbol{g})$, for the parameter grouping $\{\theta_1, \boldsymbol{g}\}$ is obtained by reversing the ordered group parameters to $\{\boldsymbol{g}, \theta_1\}$ and applying the Berger-Bernardo algorithm for the new grouping (vide discussion of Ghosh and Mukerjee (1992a) by Berger (1992)). It can be checked that for a general information matrix structure

$$\pi_{RR}(\theta_1, \boldsymbol{g}) \propto I_{11}^{1/2}(\theta_1, \boldsymbol{g}) \times \pi_2(\boldsymbol{g}). \tag{4.1}$$

This is probability matching by equation (4) of Tibshirani (1989), provided θ_1 and $\boldsymbol{g}$ are orthogonal in the sense of Cox and Reid (1987). Also examples of orthogonal parameters are known for which the Berger-Bernardo reference prior is not probabilty matching. These two facts together appear to make the reverse reference prior a reasonable competitor to the Berger-Bernardo reference prior. We discuss this issue of choice below.

In the present setup for group model it can be easily checked that

$$\pi_{RR}(\theta_1, \boldsymbol{g}) \propto I_{11}^{1/2}(\theta_1, \boldsymbol{e}) \times h^l(\boldsymbol{g}), \tag{4.2}$$

which, like Berger-Bernardo reference prior, uses the left invariant Haar density for $\boldsymbol{g}$. Note that while the refence priors of Berger-Bernardo, Chang-Eaves and Ghosh-Mukerjee all lead to the same marginal probability density function for θ_1, namely, $\pi_1(\theta_1) \propto I_{11.2}^{1/2}(\theta_1)$, making this an attractive and probably the right choice, the reverse reference prior, fails to pick it up and, instead, leads to a different marginal probability density function for θ_1, namely, $\pi_1(\theta_1) \propto I_{11}^{1/2}(\theta_1, \boldsymbol{e})$.

Also, the reverse reference prior, in general, will neither be probability matching (without orthogonality) nor eliminate the marginalization paradox, as evident from Examples 1 and 2. In Example 1, the reverse reference prior for parameter grouping $\{\theta_1, (g_2, g_3)\}$ is same as Jeffreys' prior and is given by g_3^{-2}. In Example 2, the reverse reference prior for the parameter grouping $\{\theta_1, (g_2, \dots, g_5)\}$ is given by $(1+\theta_1^2)^{1/2}(1-\theta_1^2)^{-1}(g_4 g_5)^{-2}$. The reverse reference prior in both the examples generate marginalization paradox and are not probability matching.

It can further be checked that in these examples with multiple (in particular for one at a time) parameter grouping of the nuisance parameters, the reverse reference prior, like the Berger-Bernardo reference prior, will have $\pi_2(\boldsymbol{g}) \propto h^r(\boldsymbol{g})$ and hence be free from marginalization paradox. However, it can be seen from (3.11) and (3.15) that, unlike the Berger-Bernardo reference prior, this in general can not be probability matching since it has the marginal probability density function of θ_1, $\pi_1(\theta_1) \propto I_{11}^{1/2}(\theta_1, \boldsymbol{e})$. However, if $I_{11}(\theta_1, \boldsymbol{e}) \propto I_{11.2}(\theta_1)$ then these two reference priors will be identical (in none of the examples this is true). This condition holds if the parameters θ_1 and $\boldsymbol{g}$ are orthogonal.

On the basis of these results and similar results in many other examples, the Berger-Bernardo reference prior seems preferable to the reverse reference prior at least when orthogonality does not hold. Also we should note the odd fact that the Jeffreys' prior is proportional to the geometric mean of the Berger-Bernardo reference prior and the reverse refrence prior.

5. MORE EXAMPLES

Example 3. As in Example 2, $\boldsymbol{X}_i = (X_{1i}, X_{2i})^T$, $i = 1, \ldots, n$ are i.i.d. $N_2(\boldsymbol{\mu}, \Sigma)$. Let $\theta_1 = (\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu})^{1/2}$ be the parameter of interest. Reparametrize $(\boldsymbol{\mu}, \Sigma)$ to (θ_1, B) where $\mathrm{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ is non-singular with $\boldsymbol{\mu} = \theta_1 \boldsymbol{b}_1$, $\Sigma = BB^T$ and $\boldsymbol{b}_1$ is the first column of B. For $|B| > 0$, the transformation $(\boldsymbol{\mu}, \Sigma)$ to $\theta = (\theta_1, B)$ is one-to-one. Under the group of transformations on R^2 denoted by $H = \{ g \mid g = A, |A| > 0\}$, a subgroup of non-singular transformations, in the range of $\boldsymbol{X}_1$ where $\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the induced group of transformations on the parameter space of θ is $G = \{g\}$ defined by $g\theta = (\theta_1, AB)$. The group operation is matrix multiplication and maximal invariant parameter is θ_1. Note that this group is not *amenable*. Write $\boldsymbol{\theta}_{(2)} = (\theta_2, \theta_3, \theta_4, \theta_5)$ where $\theta_2 = b_{11}$, $\theta_3 = b_{12}$, $\theta_4 = b_{21}$, $\theta_5 = b_{22}$. Also write $g_2 = a_{11}$, $g_3 = a_{12}$, $g_4 = a_{21}$, $g_5 = a_{22}$. Since our assumptions about G are satisfied we can identify g_i as θ_i, $i = 2, \ldots, 5$. It can be found by direct computation that $I(\theta_1, \boldsymbol{e}) = \bigoplus_{i=1}^{3} (D_i)$ where $D_1 = \begin{bmatrix} 1 & \theta_1 \\ \theta_1 & \theta_1^2 + 2 \end{bmatrix}$, $D_2 = \begin{bmatrix} 1 & 1 \\ 1 & \theta_1^2 + 1 \end{bmatrix}$, $D_3 = (2)$ and $\bigoplus_{i=1}^{3} D_i = \textit{Block diagonal}(D_1, D_2, D_3)$. Hence by equation (2.2), the per unit observation information matrix is $I(\theta_1, g) = \textit{Block diagonal}\left(1, J_{\boldsymbol{g}}^{-1}(\boldsymbol{g})\right) I(\theta_1, \boldsymbol{e})\ \textit{Block diagonal}\left(1, (J_{\boldsymbol{g}}^T(\boldsymbol{g}))^{-1}\right)$ where $J_{\boldsymbol{g}}(\boldsymbol{g}) = A^T \otimes I_2$; $\otimes$ is the Kronecker product. Here $I_{11.2}(\theta_1) = 2/(\theta_1^2 + 2)$. In this case, the left and the right Haar densities on G are identical and are given by $|A|^{-2}$. Consequently, the Berger-Bernardo and the Chang-Eaves reference priors are identical and are given by $\pi_{BB}(\theta_1, \boldsymbol{g}) = \pi_{CE}(\theta_1, \boldsymbol{g}) \propto (\theta_1^2 + 2)^{-1/2} |A|^{-2}$ whereas the reverse reference prior is given by $\pi_{RR}(\theta_1, \boldsymbol{g}) \propto |A|^{-2}$, the Ghosh-Mukerjee prior by $\pi_{GM}(\theta_1, \boldsymbol{g}) \propto (\theta_1^2 + 2)^{-1/2}$ and Jeffreys' prior by $\pi_J(\theta_1, \boldsymbol{g}) \propto \theta_1 |A|^{-2}$. All of the above priors except the Ghosh-Mukerjee prior use the right invariant Haar density, and so will be free from marginalization paradox. It can be checked that Ghosh-Mukerjee prior will generate the marginalization paradox. The Berger-Bernardo and Chang-Eaves priors are probability matching by Theorem 1. However, it is easy to see that the reverse reference, Jeffreys' and Ghosh-Mukerjee priors are not probability matching. This and the last two examples amply demonstrate the superiority of the Berger-Bernardo and Chang-Eaves reference priors over the other noninformative priors.

It can be checked that for $n \geq 2$ all these priors will lead to proper posterior. In the original parametrization, these priors are given by

$$\pi_{BB}(\boldsymbol{\mu}, \Sigma) = \pi_{CE}(\boldsymbol{\mu}, \Sigma) \propto \frac{|\Sigma|^{-2} (\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu})^{-1/2} \, d\boldsymbol{\mu} \, d\Sigma}{(\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} + 2)^{1/2}}, \; \pi_{RR}(\boldsymbol{\mu}, \Sigma) \propto \frac{|\Sigma|^{-2} \, d\boldsymbol{\mu} \, d\Sigma}{(\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu})^{1/2}},$$

$$\pi_J(\boldsymbol{\mu}, \Sigma) \propto \frac{d\boldsymbol{\mu} \, d\Sigma}{|\Sigma|^2}, \; \pi_{GM}(\boldsymbol{\mu}, \Sigma) \propto \frac{|\Sigma|^{-1} (\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu})^{-1/2} \, d\boldsymbol{\mu} \, d\Sigma}{(\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} + 2)^{1/2}}.$$

Example 4. This example is due to Stein and it has been discussed in Berger (1985, p 420). We assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent with $\boldsymbol{X} \sim N_m(\boldsymbol{0}, \Sigma)$ and $\boldsymbol{Y} \sim N_m(\boldsymbol{0}, \Delta\Sigma)$ where $\Delta > 0$ and Σ positive definite are unknown and $m \geq 2$. Estimation of Δ has been considered in Berger (1985) and it is shown there that the Hunt-Stein Theorem does not apply for the full linear group of transformations. We will consider the problem under the group of transformations $H = \{g | \; g = A, \; A \; is \; m \times m \; lower \; triangular \; matrix \; with \; positive \; diagonals\}$ where the group action is defined by $g(\boldsymbol{X}, \boldsymbol{Y}) = (A\boldsymbol{X}, A\boldsymbol{Y})$. For a lower triangular matrix $L \in H$, we reparametrize (Δ, Σ) to $\theta = (\theta_1, L)$ where $\Delta = \theta_1$ and $\Sigma = LL^T$. Induced group of transformations on the parameter space θ is $G = \{g\}$, defined by $g\theta = (\theta_1, AL)$ and the maximal invariant parametric function is θ_1. The group action on G is matrix multiplication. Note that this group is *amenable*. Write $p = 1 + m(m+1)/2$, $\boldsymbol{\theta}_{(2)} = (l_{11}, \ldots, l_{m1}, l_{22}, \ldots, l_{m2}, \ldots, l_{mm})^T$. It can be shown that $J_{\boldsymbol{g}}(\boldsymbol{g}) = \bigoplus_{i=1}^{m} U_i$ where $U_i^T = lower \; (m-i+1) \times (m-i+1) \; submatrix \; of \; A$. Writing $\boldsymbol{g} = (g_2, \ldots, g_p)$, since our assumptions about G are satisfied, we can identify g_i as θ_i, $i = 2, \ldots, p$. We find the information matrix $I(\theta_1, \boldsymbol{g})$ finding $I(\theta_1, \boldsymbol{e})$ first and using (2.2). It can be checked that $I_{11}(\theta_1, \boldsymbol{e}) = m/(2\theta_1^2)$, $\boldsymbol{I}_{12}(\theta_1, \boldsymbol{e}) = \theta_1^{-1} \boldsymbol{v}^T$ where $\boldsymbol{v}^T = (\boldsymbol{\rho}_{1,m}^T, \ldots, \boldsymbol{\rho}_{1,1}^T)$ and $\boldsymbol{\rho}_{1,i}$ is a i-vector with the first element one and rest zeroes. It is clear that $I_{22}(\theta_1, \boldsymbol{e}) = 2 \; I^X$ where I^X is the information matrix of L based on X evaluated at the group identity $L = I_m$. By direct calculations, we obtain $I^X = \bigoplus_{i=1}^{m} H_i$ where $H_i = Diag(2, 1, \ldots, 1)$ is $(m-i+1) \times (m-i+1)$. It can also be checked that $\boldsymbol{a}(\theta_1) = (4\theta_1)^{-1}\boldsymbol{v}$, $I_{11.2}(\theta_1) = m/(4\theta_1^2)$. From Berger (1985, p 429), the left and the right Haar densities are given by

$$h^l(\boldsymbol{g}) = \prod_{i=1}^{m} a_{ii}^{-i}, \; h^r(\boldsymbol{g}) = \prod_{i=1}^{m} a_{ii}^{-(m-i+1)}. \tag{5.1}$$

Then the Chang-Eaves reference prior is given by $\pi_{CE}(\theta_1, \boldsymbol{g}) \propto \theta_1^{-1} \prod_{i=1}^{m} a_{ii}^{-(m-i+1)}$. Since $I_{11}(\theta_1, \boldsymbol{e}) \propto I_{11.2}(\theta_1)$ and $|I(\theta_1, \boldsymbol{e})| \propto \theta_1^{-2}$, the Berger-Bernardo and the reverse reference prior are identical to the Jeffreys' prior and are proportional to $\theta_1^{-1} \prod_{i=1}^m a_{ii}^{-i}$. The Ghosh-Mukerjee reference prior is proportional to θ_1^{-1}. All these priors are of the form $I_{11.2}^{1/2}(\theta_1) \times \pi_2(\boldsymbol{g})$ and from Theorem 1, we know that the Chang-Eaves reference prior is probability matching for θ_1. For the other priors, the probability matching equation is given by (3.9) and in this example after simplification reduces to

$$\sum_{j=1}^{m} \sum_{i=j}^{m} \frac{\partial}{\partial a_{ij}} \Big(\pi_2(\boldsymbol{g})\ a_{ij}\Big) \ = \ 0. \tag{5.2}$$

Since for $\pi_2(\boldsymbol{g}) = h^l(\boldsymbol{g})$, $\sum_{i=j}^m \frac{\partial}{\partial a_{ij}}\big(\pi_2(\boldsymbol{g})\ a_{ij}\big) = (m-2j+1)\ h^l(\boldsymbol{g})$, for $j = 1, \ldots, m$, the last equation is satisfied. However, for $\pi_2(\boldsymbol{g}) = 1$, $\sum_{i=j}^m \frac{\partial}{\partial a_{ij}}\big(\pi_2(\boldsymbol{g})\ a_{ij}\big) = m-j+1$, for $j = 1, \ldots, m$, and the last equation fails. Consequently, except the Ghosh-Mukerjee prior all others are probability matching.

Define for $i = 1, \ldots, m$, $\Sigma_i = ((\sigma_{\alpha,\beta}))_{\alpha,\beta} = 1, \ldots, i$. Then in original parametrization, it can be shown that

$$\pi_{CE}(\Delta, \Sigma) \propto \Delta^{-1} \prod_{i=1}^{m} |\Sigma_i|^{-1},\ \pi_{BB}(\Delta, \Sigma) \propto \Delta^{-1} |\Sigma|^{-(m+1)/2},$$

$$\pi_{GM}(\Delta, \Sigma) \propto \Delta^{-1} \prod_{i=1}^{m} |\Sigma_i|^{-1/2}.$$

Remark 7. We can alternatively compute the information matrix I^X by using the properties of maximum likelihood estimate (MLE). For independently and identically distributed $\boldsymbol{X}_i \sim N_m(\boldsymbol{0}, \Sigma)$, $i = 1, \ldots, n$, MLE of Σ is $\hat{\Sigma}_{ML} = n^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^T = W/n$ (say). Use Bartlett's decomposition (e.g., Muirhead, 1982, p100) to write $W = TT^T$ where T is an $m \times m$ lower triangular matrix with nonnegative diagonals. Then $\hat{L} = T/\sqrt{n}$ is MLE of L where we recall that $LL^T = \Sigma$. Under $\Sigma = I_m$ (equivalently, $L = I_m$), it is known from Theorem 3.2.15 of Muirhead (1982) that $T_{ij}'s$ $1 \le j \le i \le m$ are all independent with $T_{ij} \sim N(0,1)$ for $j < i$ and $T_{ii}^2 \sim \chi^2(n-i+1)$. Noting that asymptotic variance of $\sqrt{\chi^2(n)}$ is 1/2, we get the asymptotic variance of $\sqrt{n}\ vec(\hat{L})$ under $L = I_m$ as $\overset{m}{\underset{i=1}{\oplus}} H_i^{-1}$ where H_i is defined earlier. By asymptotic theory of MLE, however, we know that asymptotic variance of $\sqrt{n}\ vec(\hat{L})$ under $L = I_p$ is given by $(I^X)^{-1}$. From these two variance expressions, we get $I^X = \overset{m}{\underset{i=1}{\oplus}} H_i$.

Example 5. This example is considered in Stein (1985). We assume that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, I_p)$, $p \ge 2$ and $\theta_1 = \sqrt{\boldsymbol{\mu}^T \boldsymbol{\mu}}$ is the parameter of interest. Under the orthogonal group of transformations G, θ_1 is maximal invariant. In this example G is neither Euclidean nor does it act freely, but it is compact. The Berger-Bernardo conditional distribution of g is invariant under one to one smooth transformations and hence of $g \to hg$ for $h \in G$, i.e., under the orthogonal group. Let the orbit $O_r = \{\boldsymbol{\mu} : \theta_1 = r\}$. We define a conditional prior distribution on an orbit given $\theta_1 = r$ by lifting the Haar measure from G to O_r:

$$\pi_2\{\boldsymbol{\mu} \in O_r \cap B | \theta_1 = r\} = \nu\{g : g\boldsymbol{\mu}(r) \in O_r \cap B\} \tag{5.3}$$

where $\boldsymbol{\mu}(r)$ is a fixed element of O_r, ν is the Haar measure on G and B is any Borel set. This conditional distribution induces the uniform distribution for $(\frac{\mu_1}{\theta_1}, \dots, \frac{\mu_p}{\theta_1})$ on the unit sphere.

Now consider the polar transformation $(\mu_1, \dots, \mu_p) \to (\theta_1, g_2, \dots, g_p)$ given by $\mu_1 = \theta_1 cos g_2, \dots, \mu_p = \theta_1 sin g_2 \cdots sin g_p$. Using Theorem 1.5.6 of Muirhead (1982, p 37-38) it follows that $(\frac{\mu_1}{\theta_1}, \dots, \frac{\mu_p}{\theta_1})$ have the uniform distribution on the unit sphere if and only if $\boldsymbol{g} = (g_2, \dots, g_p)$ has the joint density $\pi_2(\boldsymbol{g}) \propto s_2^{p-2} \cdots s_{p-1}$ where $s_i = sin g_i, i = 2, \dots, p-1$. For this new parametrization, the information matrix is given by $I(\theta_1, \boldsymbol{g}) = \theta_1^2 Diag(\theta_1^{-2}, 1, s_2^2, \dots, s_2^2 \dots s_{p-1}^2)$ and $I_{11.2}(\theta_1) = 1$. Since G is compact, ν is both left and right invariant; consequently the Chang-Eaves and the Berger-Bernardo reference priors are identical, and are given by

$$\pi_{BB}(\theta_1, \boldsymbol{g}) = \pi_{CE}(\theta_1, \boldsymbol{g}) \propto s_1^{p-2} \dots s_{p-2}. \tag{5.4}$$

It can be checked that the reverse refernce prior will be identical to the one given above. All these priors are probability matching and free from the marginalization paradox. In the original parametrization, they transform into a prior proportional to $(\boldsymbol{\mu}^T \boldsymbol{\mu})^{-(p-1)/2}$, which was also obtained by Stein (1985) and Tibshirani (1989). It can also be checked that the Ghosh-Mukerjee prior is identical to the Berger-Bernardo reference prior for one at a time parameter grouping $\{\theta_1, g_2, \dots, g_p\}$ which, while uniform in the new parametrization, is proportional to $\prod_{i=1}^{p-1} (\sum_{j=i}^{p} \mu_j^2)^{-1/2}$ in the original parametrization. It is well known that Jeffreys' prior is uniform in the original parametrization.

ACKNOWLEDGEMENT

This work was started while the first author was visiting the Department of Statistics at Purdue University in the Fall of 1993. He would like to acknowledge to Professor Jim Berger for many valuable discussions, and for the generous support through his NSF grant DMS-8923071. He is also thankful to Professor S.S. Gupta and other members of the department for their warm hospitality.

REFERENCES

[1] Bayarri, M.J. (1981). Inferencia Bayesiana sobre el coefficiente de correlación de una población normal bivariante. *Trabajos Estadíst. Investigación Oper.* **32**, 18-31.

[2] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Second Edition, Springer-Verlag, New York.

[3] Berger, J. (1992). Discussion of Ghosh and Mukerjee. In *Bayesian Statistics* **4**, J.M. Bernardo, J.O. Berger, D.V. Lindley, and A.F.M. Smith (eds.). London: Oxford University Press, 205-6.

[4] Berger, J. and Bernardo, J.M. (1989). Estimating a product of means: Bayesian analysis with reference priors. *J. Am. Statist. Assoc.* **84**, 200-7.

[5] Berger, J. and Bernardo, J.M. (1992a). On the development of the reference prior method. In *Bayesian Statistics* **4**, J.M. Bernardo, J.O. Berger, D.V. Lindley, and A.F.M. Smith (eds.). London: Oxford University Press, 35-60.

[6] Berger, J. and Bernardo, J.M. (1992b). Ordered group reference priors with applications to the multinomial problem. *Biometrika* **79**, 25-37.

[7] Chang, T. and Eaves, D. (1990). Reference priors for the orbit in a group model. *Ann. Statist.* **18**, 1595-614.

[8] Cox, D.R. and Reid, N.(1987). Orthogonal parameters and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser B* **49**, 1-39.

[9] Dawid, A.P., Stone, M. and Zidek, J.V. (1973). Marginalization paradoxes in Bayesian and structural inferences (with discussion). *J. Roy. Statist. Soc. Ser B* **35**, 189-233.

[10] Geisser, S. (1965). Bayesian estimation in multivariate analysis. *Ann. Math. Statist.* **36**, 150-9.

[11] Ghosh, J.K. and Mukerjee, R. (1992a). Non-informative priors (with discussion).

In *Bayesian Statistics* **4**, J.M. Bernardo, J.O. Berger, D.V. Lindley, and A.F.M. Smith (eds.). London: Oxford University Press, 195-210.

[12] Ghosh, J.K. and Mukerjee, R. (1992b). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *J. Roy. Statist. Soc. Ser B* **54**, 867-75.

[13] Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian viewpoint.* Cambridge University Press.

[14] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[15] 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.

[16] Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Sequential Methods in Statistics. Banach Center Publications* **16**, Warsaw PWN, Polish Scientific Publishers, 485-514.

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

[18] Ye, K. and Berger, J. (1991). Noninformative priors for inferences in exponential regression models. *Biometrika* **78**, 645-56.