The Annals of Statistics
1990, Vol. 18, No. 4, 1595–1614

REFERENCE PRIORS FOR THE ORBIT IN A GROUP MODEL

BY TED CHANG[1] AND DAVID EAVES[2]

University of Virginia and Simon Fraser University

For a group model in which the group **G** acts freely on the parameter space $\mathbf{\Omega}$, this paper considers a prior which is a product of right Haar measure on **G** and a limiting form of Jeffreys' prior for the maximal invariant. When the parameter of interest is the orbit of **G** in $\mathbf{\Omega}$, it is shown that such a prior is the reference prior defined by Bernardo. A method of calculating this reference prior is given which avoids the necessity of working in a parameterization of $\mathbf{\Omega}$ which expresses $\mathbf{\Omega}$ as a product of **G** and a cross section. Examples of the multivariate normal distribution, with the parameter of interest being the correlation matrix or the eigenvalues of the covariance matrix, are discussed.

Introduction. In this paper we consider the choice of a noninformative prior in group models in which the parameter of interest indexes the orbits in the parameter space. An example of this situation is the bivariate normal distribution model with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix **V**. This model is naturally a group model using the affine linear group which consists of the pairs $(\mathbf{a}, \mathbf{A})$ where **A** is a 2×2 matrix and **a** is a 2-vector.

By letting the group **G** vary over various subgroups of the affine linear group, one obtains several useful examples. If **G** consists of the pairs $(\mathbf{a}, \mathbf{A})$ for which **A** is diagonal, the orbits are indexed by the correlation coefficient ρ. If instead, **G** consists of the pairs $(\mathbf{a}, \mathbf{A})$ for which **A** is orthogonal (that is, the columns of **A** form an orthonormal basis of Euclidean 2-space R^2), the orbits are indexed by the eigenvalues of **V**, that is, by the variances of the population principal components. Finally, if we restrict the parameter space to pairs $(\boldsymbol{\mu}, \mathbf{V})$ with **V** of the form $\sigma^2 I$ and let **G** consist of the pairs $(\mathbf{a}, \mathbf{A})$ with $\mathbf{a} = 0$ and **A** being a multiple of an orthogonal matrix, the orbits are indexed by the noncentrality parameter $\delta = \boldsymbol{\mu}^t \boldsymbol{\mu} / \sigma^2$.

Let $\mathbf{\Omega}$ denote the parameter space and $\mathbf{\Omega}/\mathbf{G}$ the orbit space: that is, the collection of orbits of **G** in $\mathbf{\Omega}$. We shall use the notation ω for the elements of $\mathbf{\Omega}$, g for the elements of **G**, θ for the elements of $\mathbf{\Omega}/\mathbf{G}$, $g \cdot \omega$ for the action of g on ω, **X** for the sample space and **x** for the generic element of **X**. Multiplication in **G** of g_1 and g_2 will be denoted $g_1 g_2$. If **G** acts freely on $\mathbf{\Omega}$ (that is, $g \cdot \omega = \omega$ for some ω implies $g = 1$), one can usually find (at least locally in $\mathbf{\Omega}$) a decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$ with **G** acting on $\mathbf{\Omega}$ by left multiplication

Received May 1986; revised October 1989.
[1]Research conducted while visiting the Kansas Geological Survey of the University of Kansas. Supported in part by NSF Grant DMS-89-01494.
[2]Supported in part by a grant from the National Science and Engineering Research Council (Canada).
AMS 1980 *subject classifications*. Primary 62A05, 62F15; secondary 62H20, 62H25.
Key words and phrases. Noninformative priors, Bayesian inference in group models, maximal invariant.

in the first factor. The prior we propose will be defined by:

1. The conditional distribution of $p(g|\theta)$ of g given θ will be a right Haar measure, independent of θ. That is, although Haar measure on $\mathbf{G}$ is defined only up to multiplication by a constant, the same right Haar measure will be used for each θ.
2. The marginal prior $p(\theta)$ of θ will be the limit (as $n \to \infty$) of Jeffreys' prior for the sampling distribution of the maximal invariant $\mathbf{y}_n$ of the $\mathbf{G}$ action on a random sample of size n, that is, the orbits of the action of $\mathbf{G}$ on the Cartesian product $\mathbf{X}^n$. More precisely, if $I_n(\theta)$ is the information matrix of $\mathbf{y}_n$, $p(\theta) = \lim_{n \to \infty} \sqrt{\det(I_n(\theta))/n}$.

We show, in the Appendix, that the prior defined by conditions 1 and 2 is minimally informative about θ in the sense of Bernardo (1979). Berger and Bernardo (1989) contain a clarification of the Bernardo approach when, as in the present case, nuisance parameters are present. Berger and Bernardo also use this approach to find a reference prior for a pair of independent normal distributions when the parameter of interest is the product of the means.

For example, in the bivariate normal case if θ is the correlation coefficient ρ [and $\mathbf{G}$ consists of pairs $(\mathbf{a}, \mathbf{A})$ for which $\mathbf{A}$ is diagonal], the prior defined by these constraints is the Lindley (1965) prior $(1 - \rho^2)^{-1}\sigma_1^{-1}\sigma_2^{-1}\, d\mu_1\, d\mu_2\, d\sigma_1\, d\sigma_2\, d\rho$. Lindley's prior has also been derived by Bayarri (1981) using the Bernardo approach.

If one starts with the parameter θ, conditions 1 and 2 appear to depend upon several arbitrary choices: the choice of the group $\mathbf{G}$ and the choice of the decomposition $\boldsymbol{\Omega} \approx \mathbf{G} \times \boldsymbol{\Omega}/\mathbf{G}$.

With respect to the choice of the group $\mathbf{G}$, the authors believe that many statistical parameters θ are interesting precisely because they represent the invariants (also known as orbits) of a specific important group $\mathbf{G}$. In other words, is not an uncommon situation that θ has a natural group $\mathbf{G}$ attached to it. For example, the correlation coefficient is interesting because it is a measurement of linear relationship *which does not depend upon the location and scale of the components of* $\mathbf{x}$. Thus the group $\mathbf{G} = \{(\mathbf{a}, \mathbf{A})|\mathbf{A} \text{ is diagonal}\}$, which represents the location and scale transformations in the components of $\mathbf{x}$, is the natural defining group of the correlation coefficient. Principal components analysis arises out of an attempt (after first centering the variables by their means) to rotate the axes in the data space so as to capture certain important information: the direction of the maximum variation in $\mathbf{x}$. Thus the group $\mathbf{G} = \{(\mathbf{a}, \mathbf{A})|\mathbf{A} \text{ is orthogonal}\}$ is the natural defining group for the variances of the population principal components. Finally, the noncentrality parameter δ arises in the study uncorrelated normal variables (namely the components of $\mathbf{x}$) with equal variances σ^2. The theory of the normal linear model is vitally tied into the metric and linear properties of Euclidean space (scaled by the constant σ). Since the orthogonal transformations are exactly the distance preserving linear transformations of Euclidean space, the scaled orthogonal transformations are the natural defining group of the noncentrality parameter δ.

With reference to the apparent dependence of condition 1 on the decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$, let (g_1, θ) and (g_2, θ) be the representations of ω under two decompositions of $\mathbf{\Omega}$. It can be shown (see Proposition 4 in the Appendix) that $g_2 = g_1 s(\theta)$ for some map s: $\mathbf{\Omega}/\mathbf{G} \to \mathbf{G}$. Thus if $p(g|\theta)$ is right Haar measure in some decomposition of $\mathbf{\Omega}$ as $\mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$, it will be the same right Haar measure in all such decompositions. Priors satisfying condition 1 have the additional advantage [see Dawid, Stone and Zidek (1973)] that marginalization paradoxes for θ will not occur. That is: starting with a prior in ω and using the sampling density of $f(\mathbf{x}|\omega)$ has the same marginal posterior for θ as marginalizing the prior and using the sampling density $f(\mathbf{y}|\theta)$.

As the prior defined by conditions 1 and 2 does not depend upon the decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$, it would appear aesthetically advantageous to have a way of computing it which does not depend upon using such a decomposition. Such a method is described in greater detail in Section 1 together with the example of the noncentrality parameter in a noncentral spherically symmetric bivariate distribution. Indeed, a differentiable decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$ often does not exist globally on $\mathbf{\Omega}$, but rather must be pieced together from local decompositions. Because condition 1 has a meaning independent of the particular choice of decomposition, this lack of a global decomposition does not affect the validity of condition 1 as a defining characteristic for a prior. Nevertheless, the common lack of such a global decomposition indicates that a method of computing the prior defined by conditions 1 and 2, which is decomposition independent, would often be more convenient. Section 1 contains an example, the well known Hopf fibration from topology recast into a statistical context, which vividly illustrates the problem of the lack of a generic decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$.

The reference prior we propose is a modification of Jeffreys' prior. One of the motivations for the Jeffreys' prior is its independence of the choice of coordinates for $\mathbf{\Omega}$ in which it is expressed [see Jeffreys (1939, 1946)]. In a similar spirit, our approach also emphasizes a coordinate-independent reference prior. Kass (1989) gives an extremely lucid introduction to the geometric ideas underlying both the Jeffreys' prior and the present paper.

Section 2 contains certain specific calculations for use in multivariate normal models. These calculations can be extended to other multivariate location and scale models if the information matrix can be calculated. The calculations in Section 2 are applied in Section 3 to find noninformative priors for the correlation matrix and for the eigenvalues of the covariance matrix in the multivariate normal model. Appendix A describes in substantial detail the method topologists have devised to handle the generic lack of a global decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$. The remaining proofs are contained in Appendix B.

We note that a theorem of Stein [see Chang and Villegas (1986)] states that if a prior satisfying condition 1 is used and if equivariant credibility regions are used, then credibility conditional on θ coincides with coverage probability conditional on the maximal invariant $\mathbf{y}$. Although the prior we define will satisfy the hypotheses of the Stein theorem, the Stein theorem is actually a statement about inference on the $\mathbf{G}$ factor in a decomposition of $\mathbf{\Omega}$.

1. Calculation of the reference prior for an orbit. We assume that the group $\mathbf{G}$ acts freely on $\mathbf{\Omega}$. We will also assume that $\mathbf{G}$ acts freely almost everywhere on $\mathbf{X}$.

The condition that $\mathbf{G}$ acts freely almost everywhere on $\mathbf{X}$ can usually be arranged by taking a big enough sample. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are in $\mathbf{X}$, then $g \cdot (\mathbf{x}_1, \mathbf{x}_2) = (g \cdot \mathbf{x}_1, g \cdot \mathbf{x}_2)$. It follows that the isotropy subgroup of $(\mathbf{x}_1, \mathbf{x}_2)$ [that is, the elements of $\mathbf{G}$ that fix $(\mathbf{x}_1, \mathbf{x}_2)$] is the intersection of the isotropy subgroups of $\mathbf{x}_1$ and of $\mathbf{x}_2$. Thus, unless some element (besides 1) of $\mathbf{G}$ fixes every element of $\mathbf{X}$, a general position argument can usually establish that $\mathbf{G}$ acts freely almost everywhere on $\mathbf{X}$. For example, if $\mathbf{G} = \mathrm{GL}(p)$, the nonsingular $p \times p$ matrices acts on the space $\mathbf{X} = R^p$ in the usual fashion, then as long as we take a sample of size n at least p, the action will be free except on the set of measure zero consisting of those n-tuples of p-vectors which do not contain a spanning set of R^p.

If some element (besides 1) of $\mathbf{G}$ fixes every element of $\mathbf{X}$, the collection $\mathbf{K}$ of such elements (known in the literature as the ineffective kernel) will form a normal subgroup of $\mathbf{G}$. If $g \in \mathbf{K}$, then $g \cdot \omega$ and ω will yield the same distribution for every $\omega \in \mathbf{\Omega}$. It follows that one should replace $\mathbf{G}$ by $\mathbf{G}/\mathbf{K}$ and $\mathbf{\Omega}$ by $\mathbf{\Omega}/\mathbf{K}$ (the orbits of $\mathbf{\Omega}$ under $\mathbf{K}$).

We denote by $\mathscr{L}(\mathbf{G})$ the vector space of tangent vectors to $\mathbf{G}$ at the identity 1 of $\mathbf{G}$. Let $A \in \mathscr{L}(\mathbf{G})$ and let $\alpha(t)$ be a curve in $\mathbf{G}$ with $\alpha(0) = 1$ and $\alpha'(0) = A$. Define a vector field A^* on $\mathbf{\Omega}$ by

$$A^*(\omega) = \frac{d}{dt}\bigg|_{t=0} \alpha(t) \cdot \omega, \quad \text{for } \omega \in \mathbf{\Omega}.$$

We further define for $g \in \mathbf{G}$, the linear transformation $\mathrm{ad}(g)$: $\mathscr{L}(\mathbf{G}) \to \mathscr{L}(\mathbf{G})$ by

$$\mathrm{ad}(g) \cdot A = \frac{d}{dt}\bigg|_{t=0} g\alpha(t)g^{-1}.$$

Let $A_1, \dots, A_k$ be a basis of $\mathscr{L}(\mathbf{G})$, $k = \dim \mathbf{G}$ and define

$$J(\omega) = \det A_i^*(\omega)^t \cdot I(\omega) \cdot A_j^*(\omega),$$

where $I(\omega)$ is the Fisher information matrix at ω. Here $I(\omega)$ and $A_i^*(\omega)$ are matrices calculated using any parameterization of $\mathbf{\Omega}$. It is not necessary to calculate them with a parameterization of $\mathbf{\Omega}$ which decomposes $\mathbf{\Omega}$ in the form $\mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$.

PROPOSITION 1. (a) *If $\alpha_i(t)$ are curves in $\mathbf{G}$ with $\alpha_i(0) = 1$ and $\alpha_i'(0) = A_i$, then*

$$A_i^*(\omega)^t \cdot I(\omega) \cdot A_j^*(\omega) = E_\omega\left[\frac{\partial \log f(x; \alpha_i(t) \cdot \omega)}{\partial t}\bigg|_{t=0} \frac{\partial \log f(x; \alpha_j(s) \cdot \omega)}{\partial s}\bigg|_{s=0}\right].$$

(b) *$J(\omega)$ does not depend upon the choice of parameterization used to calculate $I(\omega)$.*

(c) *If* $\mathbf{G}$ *acts freely on* $\mathbf{\Omega}$, *then* $J(\omega) > 0$.

(d) *Let $\tilde{J}(\omega)$ be calculated using the basis $\tilde{A}_1, \dots, \tilde{A}_k$ and let C be the matrix defined by $C \cdot A_i = \tilde{A}_i$. Then $\tilde{J}(\omega) = (\det C)^2 J(\omega)$.*

(e) $J(g \cdot \omega) = \det[\mathrm{ad}(g^{-1})]^2 J(\omega)$.

(f) *If* $\mathbf{G}$ *is unimodular, then* $J(g \cdot \omega) = J(\omega)$.

THEOREM 2. *Suppose* $\mathbf{G}$ *acts freely on* $\mathbf{\Omega}$ *and freely almost everywhere on* $\mathbf{X}$. *Let $\tilde{p}(\omega)$ be overall Jeffreys' prior. Let $p(\omega)$ be the prior defined by*:

(i) *The conditional distribution of $p(g|\theta)$ of g given θ is a right Haar measure, independent of θ.*

(ii) *The marginal prior $p(\theta)$ of θ is the limit of Jeffreys' prior for the sampling distribution of the maximal invariant.*

Then $p(\omega)$ is the Bernardo minimally θ-informative prior. Furthermore, it can be calculated by

$$p(\omega) = \tilde{p}(\omega) J(\omega)^{-1/2}.$$

EXAMPLE (Noncentrality parameter in a spherically symmetric noncentral bivariate distribution). Consider a bivariate distribution of the form

$$f(\mathbf{x}|\mu, \sigma) \propto \sigma^{-2} \cdot \exp\left[g\left((\mathbf{x} - \mu)^t (\mathbf{x} - \mu)/\sigma^2\right)\right].$$

The parameter of interest is the noncentrality parameter $\delta = \boldsymbol{\mu}^t \boldsymbol{\mu}/\sigma^2$. If

$$\mathbf{G} = \left\{ \begin{bmatrix} \rho \cos\theta & -\rho \sin\theta \\ \rho \sin\theta & \rho\cos\theta \end{bmatrix} = A(\rho, \theta) \right\}$$

acting on $\mathbf{X}$ by matrix multiplication and on $\mathbf{\Omega}$ by $A(\rho, \theta) \cdot (\mu, \sigma) = (A(\rho, \theta) \cdot \mu, \rho\sigma)$, we have a group model and δ parameterizes the orbits of $\mathbf{G}$ in $\mathbf{\Omega}$.

In the coordinate system (μ_1, μ_2, σ), the information matrix is of the form $\sigma^{-2} \operatorname{diag}[k_1 \quad k_1 \quad k_2]$, where

$$\begin{aligned} k_1 &= -2E[g'(z) + zg''(z)], \\ k_2 &= -2E\left[1 + 3zg'(z) + 2z^2 g''(z)\right], \\ z &= (\mathbf{x} - \boldsymbol{\mu})^t (\mathbf{x} - \boldsymbol{\mu})/\sigma^2. \end{aligned}$$

Note that k_1 and k_2 do not depend upon (μ_1, μ_2, σ).

Let $\alpha_1(t) = A(1, t)$ and $\alpha_2(t) = A(e^t, 0)$. The tangent vectors at the identity of $\mathbf{G}$ to α_1 and α_2 are $A_1 = \partial/\partial\theta$ and $A_2 = \partial/\partial\rho$. Therefore

$$A_1^*(\mu_1, \mu_2, \sigma) = \frac{d}{dt}\bigg|_{t=0} A(1, t) \cdot (\mu_1, \mu_2, \sigma) = [-\mu_2, \mu_1, 0]^t,$$

$$A_2^*(\mu_1, \mu_2, \sigma) = \frac{d}{dt}\bigg|_{t=0} A(e^t, 0) \cdot (\mu_1, \mu_2, \sigma) = [\mu_1, \mu_2, \sigma]^t$$

and hence

$$J(\mu_1, \mu_2, \sigma) = \det\begin{bmatrix} k_1\delta & 0 \\ 0 & k_2 + k_1\delta \end{bmatrix} = k_1\delta(k_2 + k_1\delta).$$

Overall, Jeffreys' prior is $\tilde{p}(\mu_1, \mu_2, \sigma) = \sigma^{-3} k_1 \sqrt{k_2}$ and hence the minimally δ-informative prior is

$$p(\mu_1, \mu_2, \sigma) = \sigma^{-3}\sqrt{k_1 k_2}\left(k_2\delta + k_1\delta^2\right)^{-1/2}.$$

If $g(z) = -z/2$, so that the underlying distribution is normal, $k_1 = 1$ and $k_2 = 4$. After a **G**-equivariant sufficiency reduction, the orbits of the action of **G** on $\mathbf{X}^n$ can be parameterized by $t = \bar{\mathbf{x}}^t\bar{\mathbf{x}}/s_p^2$ where $s_p^2 = \Sigma(\mathbf{x}_i - \bar{\mathbf{x}})^t(\mathbf{x}_i - \bar{\mathbf{x}})/(2n-2)$. $\frac{1}{2}nt$ has a noncentral F distribution with noncentrality parameter $n\delta$ and $(2, 2n-2)$ degrees of freedom. It follows that $\sqrt{n}(t - \delta)$ has a limiting normal distribution with mean 0 and variance $4\delta + \delta^2$. Thus Jeffreys' prior for δ has a limiting form of $d\delta/\sqrt{4\delta + \delta^2}$.

$\mathbf{\Omega}$ can be parameterized globally as $\mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$ using $\rho = \sqrt{\mu_1^2 + \mu_2^2}$, $\varphi = \arctan(\mu_2/\mu_1)$ and δ. Then making the change of variables,

$$p(\mu_1, \mu_2, \sigma)\, d\mu_1\, d\mu_2\, d\sigma = \frac{2\, d\mu_1\, d\mu_2\, d\sigma}{\sigma^3\sqrt{4\delta + \delta^2}} = \frac{d\delta}{\sqrt{4\delta + \delta^2}}\,\frac{d\rho\, d\varphi}{\rho}.$$

We recognize $d\rho\, d\varphi/\rho$ as Haar measure on **G** thus verifying directly properties (i) and (ii) in Theorem 2.

One might consider for this problem the prior $d\mu_1\, d\mu_2\, d\sigma/\sigma$. Since $d\mu_1\, d\mu_2\, d\sigma/\sigma = \rho\, d\rho\, d\varphi \cdot d\delta/(2\delta)$, the conditional distribution along an orbit is not Haar and hence a marginalization paradox with respect to δ can occur.

Example. (The Hopf Fibration). The Hopf fibration is a low dimension example which is commonly used to illustrate the techniques topologists have evolved to handle the usual lack of a global decomposition $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$. In this example we discuss a close cousin of the Hopf fibration in a statistical context which could plausibly arise in tectonic studies.

Let S^2 denote the unit sphere in Euclidean three-dimensional space and assume $\mathbf{v}_1, \dots, \mathbf{v}_n$ are independent points on S^2 with each $\mathbf{v}_i$ having a Fisher distribution with modal vector $\mathbf{A}\mathbf{u}_i$ and a known concentration parameter κ. Here $\mathbf{u}_1, \dots, \mathbf{u}_n$ are known fixed points on S^2 and $\mathbf{A}$ is a rotation of S^2 written as a 3×3 matrix. It is well known that $\mathbf{A}$ is a rotation if and only if $\mathbf{A}\mathbf{A}^t = I$ and the determinant of $\mathbf{A}$ is 1; the collection of such matrices is usually denoted by SO(3). In our notation $\mathbf{\Omega} = \mathrm{SO}(3)$. Suppose the parameter of interest is $\mathbf{A}^t\mathbf{e}_3$, where $\mathbf{e}_3 = [0 \quad 0 \quad 1]^t$.

Chang (1986) discusses inference on $\mathbf{A}$. The problem arises in connection with the determination of the errors in the reconstruction of the past position of tectonic plates. In that context, $\mathbf{A}$ represents the rotation which rotates the present position of a tectonic plate into its past position and we are now interested in finding a minimally informative prior for the present location $\mathbf{A}^t\mathbf{e}_3$ of a point whose past location is $\mathbf{e}_3$.

This statistical model is naturally a group model with group SO(3). SO(3) acts on $\mathbf{\Omega}$ and on each $\mathbf{v}_i$ by left matrix multiplication. It follows that Jeffreys' prior is Haar measure on $\mathbf{\Omega}$. Thinking of SO(3) as an embedded surface in the nine-dimensional Euclidean space of 3×3 matrices, $\mathscr{L}(\mathrm{SO}(3)) = \{\mathbf{E} \mid \mathbf{E} + \mathbf{E}^t = \mathbf{0}\}$. Given $\mathbf{E}$, $\alpha(t) = \exp t\mathbf{E} \equiv \sum_r t^r \mathbf{E}^r / r!$ is a curve in SO(3) with $\alpha(0) = I$ and $\alpha'(0) = \mathbf{E}$. If $\mathbf{A} \in \mathbf{\Omega}$, $\mathbf{A} \exp t\mathbf{E}$ is a curve in $\mathbf{\Omega}$ and it is shown in Chang (1986) that

$$\frac{d}{dt}\bigg|_{t=0} (\mathbf{A} \exp t\mathbf{E})^t \cdot I(\mathbf{A}) \cdot \frac{d}{dt}\bigg|_{t=0} (\mathbf{A} \exp t\mathbf{E}) = -n(\kappa \coth \kappa - 1)\mathrm{Tr}(\mathbf{E}\mathbf{\Sigma}\mathbf{E}),$$

where $\mathbf{\Sigma} = \sum_i u_i u_i^t / n$. In our notation,

$$\mathbf{E}^*(\mathbf{A}) = \frac{d}{dt}\bigg|_{t=0} (\exp t\mathbf{E})\mathbf{A} = \frac{d}{dt}\bigg|_{t=0} \mathbf{A} \exp(t\mathbf{A}^t\mathbf{E}\mathbf{A}).$$

Therefore

$$\mathbf{E}^*(\mathbf{A})^t \cdot I(\mathbf{A}) \cdot \mathbf{E}^*(\mathbf{A}) = -n(\kappa \coth \kappa - 1)\mathrm{Tr}(\mathbf{E}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^t\mathbf{E}).$$

It is easily shown that $\mathbf{A}^t\mathbf{e}_3 = \mathbf{B}^t\mathbf{e}_3$ if and only if $\mathbf{B} = \mathbf{C}\mathbf{A}$ for some

$$\mathbf{C} \in \mathbf{G} = \left\{ \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}.$$

$\mathbf{G}$ is the collection of rotations which fix $\mathbf{e}_3$. Restrict the action of SO(3) on $\mathbf{\Omega}$ to an action of $\mathbf{G}$ on $\mathbf{\Omega}$. Then the orbit space $\mathbf{\Omega}/\mathbf{G}$ consists of the right cosets $\mathbf{G}\mathbf{A}$ of $\mathbf{G}$ in SO(3) and $\mathbf{\Omega}/\mathbf{G} \approx S^2$ with the correspondence taking the right coset $\mathbf{G}\mathbf{A}$ into $\mathbf{A}^t\mathbf{e}_3$.

$$\mathbf{E} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is a basis for $\mathscr{L}(\mathbf{G})$ and hence

$$\begin{aligned} J(\mathbf{A}) &= -n(\kappa \coth \kappa - 1)\mathrm{Tr}(\mathbf{E}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^t\mathbf{E}) \\ &= n(\kappa \coth \kappa - 1)(\mathbf{A}^t\mathbf{e}_3)^t (I - \mathbf{\Sigma})(\mathbf{A}^t\mathbf{e}_3). \end{aligned}$$

Since the overall Jeffreys' prior on $\mathbf{\Omega} = \mathrm{SO}(3)$ is Haar measure, the minimally informative prior about $\mathbf{A}^t\mathbf{e}_3$ has density $J(\mathbf{A})^{-1/2}$ with respect to Haar measure. Since the eigenvector corresponding to the largest eigenvalue of $\mathbf{\Sigma}$ is at the center of the points u_i, we see that this minimally informative prior has greatest density at rotations $\mathbf{A}$ which move the center of the points $\mathbf{u}_i$ to $\mathbf{e}_3$.

In topology, a close cousin of the projection π: SO(3) $\to$ SO(3)/$\mathbf{G}$ is known as the Hopf fibration. Topologically, SO(3) is a real projective 3-space and $\mathbf{G}$ is a circle S^1. Since $S^1 \times S^2$ is topologically quite different from SO(3), no topological decomposition SO(3) $\approx S^1 \times S^2$ is possible. The Hopf fibration is described in Steenrod [(1951), pages 105–108] in its usual form $S^3 \to S^3/S^1 \approx S^2$; the relationship between SO(3) and S^3 is given in Goldstein [(1950), pages 109–111].

Let $U_+ = S^2 - \{-\mathbf{e}_3\}$ and $U_- = S^2 - \{\mathbf{e}_3\}$ be the complements of $-\mathbf{e}_3$ and $\mathbf{e}_3$, respectively. Let $\varphi_+: \mathbf{G} \times U_+ \to \pi^{-1}(U_+)$ and $\varphi_-: \mathbf{G} \times U_- \to \pi^{-1}(U_-)$ be defined by

$$\varphi_+(g, \mathbf{u}) = g \cdot \left[\frac{(\mathbf{e}_3 + \mathbf{u})(\mathbf{e}_3 + \mathbf{u})^t}{\mathbf{e}_3^t \mathbf{u} + 1} - \mathbf{I} \right]$$

and

$$\varphi_-(g, \mathbf{u}) = g \cdot \left[\frac{(\mathbf{e}_3 - \mathbf{u})(\mathbf{e}_3 - \mathbf{u})^t}{\mathbf{e}_3^t \mathbf{u} - 1} + \mathbf{I} \right],$$

respectively. $\varphi_\pm$ are local trivializations of π: SO(3) $\to$ SO(3)$/\mathbf{G}$. Notice that $\varphi_\pm$ satisfy:

(a) $\varphi_\pm$ are differentiable, one-to-one, onto maps of $\mathbf{G} \times U_\pm \to \pi^{-1}(U_\pm)$ with differentiable inverses (that is, $\varphi_\pm$ are diffeomorphisms).

(b) $\pi \varphi_\pm(g, \mathbf{u}) = \mathbf{u}$, that is, $\varphi_\pm$ takes each $\mathbf{G} \times \mathbf{u}$ one-to-one onto the fiber $\pi^{-1}(\mathbf{u})$.

(c) $\tilde{g} \cdot \varphi_\pm(g, \mathbf{u}) = \varphi_\pm(\tilde{g}g, \mathbf{u})$, that is, $\varphi_\pm^{-1}$ takes the action of $\mathbf{G}$ on $\mathbf{\Omega}$ into group multiplication in $\mathbf{G}$.

(d) $\varphi_-^{-1}\varphi_+(g, \mathbf{u}) = (gs(\mathbf{u}), \mathbf{u})$, where $s: U_+ \cap U_- \to \mathbf{G}$ is

$$s(\mathbf{u}) = \left[\frac{(\mathbf{e}_3 + \mathbf{u})(\mathbf{e}_3 + \mathbf{u})^t}{\mathbf{e}_3^t \mathbf{u} + 1} - \mathbf{I} \right] \left[\frac{(\mathbf{e}_3 - \mathbf{u})(\mathbf{e}_3 - \mathbf{u})^t}{\mathbf{e}_3^t \mathbf{u} - 1} + \mathbf{I} \right].$$

π: SO(3) $\to$ SO(3)$/\mathbf{G}$ is typical of a general phenomenon. Usually differentiable trivializations $\mathbf{\Omega} \approx \mathbf{G} \times \mathbf{\Omega}/\mathbf{G}$ can only be found locally and even local trivializations [that is, maps φ which satisfy (a), (b) and (c) above] may only be definable, as in this example, in some arbitrary manner. Appendix A contains a proof that (under mild conditions) a locally trivial structure to the orbit projection $\pi: \mathbf{\Omega} \to \mathbf{\Omega}/\mathbf{G}$ always exists and that any two local trivializations will be related by a condition analogous to (d) above. In topology, $\pi: \mathbf{\Omega} \to \mathbf{\Omega}/\mathbf{G}$ is said to have the structure of a fiber bundle. In view of Proposition 1(b), Theorem 2 gives a method for calculating $p(\omega)$ which does not depend upon the choice of parameterization of $\mathbf{\Omega}$. Thus we can be sure that the prior $p(\omega)$ is intrinsic to the fiber bundle $\pi: \mathbf{\Omega} \to \mathbf{\Omega}/\mathbf{G}$ and does not depend upon any choices of local trivialization.

2. Some calculations for the multivariate normal distribution. We parameterize the p-dimensional multivariate normal distribution by $(\boldsymbol{\mu}, \mathbf{V})$, where $\boldsymbol{\mu}$ is the mean and $\mathbf{V}$ is the covariance matrix. Let AL(p) be the group $\{(\mathbf{a}, \mathbf{A})\}$, where $\mathbf{a} \in R^p$ and $\mathbf{A} \in$ GL(p). AL(p) acts on the sample space $\mathbf{X} = R^p$ by $(\mathbf{a}, \mathbf{A}) \cdot \mathbf{x} = \mathbf{a} + \mathbf{A}\mathbf{x}$. If $n \geq p + 1$, the resulting action of AL(p) on $\mathbf{X}^n$ is free except on the set of measure zero consisting of all n-tuples

$(\mathbf{x}_1, \dots, \mathbf{x}_n)$ such that $\mathbf{x}_2 - \mathbf{x}_1, \dots, \mathbf{x}_n - \mathbf{x}_1$ span a subset of R^p of dimension less than p. The action of $\mathrm{AL}(p)$ on $\boldsymbol{\Omega}$ is $(\mathbf{a}, \mathbf{A}) \cdot (\boldsymbol{\mu}, \mathbf{V}) = (\mathbf{a} + \mathbf{A}\boldsymbol{\mu}, \mathbf{AVA}^t)$. Overall, Jeffreys' prior on $\boldsymbol{\Omega}$ is

$$\tilde{p}(\boldsymbol{\mu}, \mathbf{V}) = (\det \mathbf{V})^{-(p+2)/2}. \tag{1}$$

The tangent space $\mathscr{L}(\mathrm{AL}(p))$ at the identity $(\mathbf{0}, \mathbf{I})$ of $\mathrm{AL}(p)$ is a product $R^p \times M(p)$, where $M(p)$ is the collection of $p \times p$ matrices. If $\mathbf{a}, \mathbf{b} \in R^p$, let $\alpha(t, s) = (t\mathbf{a} + s\mathbf{b}, \mathbf{I}) \in \mathrm{AL}(p)$. Then $d/dt|_{t=0}\, \alpha(t, 0) = (\mathbf{a}, \mathbf{0})$ and $d/dt|_{s=0}\, \alpha(0, s) = (\mathbf{b}, \mathbf{0})$. It follows that

$$\begin{aligned}
&((\mathbf{a}, \mathbf{0})^*(\boldsymbol{\mu}, \mathbf{V}))^t \cdot I(\boldsymbol{\mu}, \mathbf{V}) \cdot ((\mathbf{b}, \mathbf{0})^*(\boldsymbol{\mu}, \mathbf{V})) \\
&\quad = E_{\mathbf{x}}\left[\frac{\partial \log f(\mathbf{x}; \boldsymbol{\mu} + t\mathbf{a}, \mathbf{V})}{\partial t}\bigg|_{t=0} \frac{\partial \log f(\mathbf{x}; \boldsymbol{\mu} + s\mathbf{b}, \mathbf{V})}{\partial s}\bigg|_{s=0} \right] \\
&\quad = -E_{\mathbf{x}}\left[\frac{\partial^2 \log f(\mathbf{x}; \boldsymbol{\mu} + t\mathbf{a} + s\mathbf{b}, \mathbf{V})}{\partial t\, \partial s}\bigg|_{s,t=0} \right] \qquad (2) \\
&\quad = -E_{\mathbf{x}}\left[\frac{\partial^2}{\partial t\, \partial s}\bigg|_{s,t=0} - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - t\mathbf{a} - s\mathbf{b})^t \mathbf{V}^{-1} (\mathbf{x} - \boldsymbol{\mu} - t\mathbf{a} - s\mathbf{b}) \right] \\
&\quad = E_{\mathbf{x}}\left[\frac{1}{2}\mathbf{a}^t \mathbf{V}^{-1}\mathbf{b} + \frac{1}{2}\mathbf{b}^t \mathbf{V}^{-1}\mathbf{a} \right] = \mathbf{a}^t \mathbf{V}^{-1}\mathbf{b}.
\end{aligned}$$

If $\mathbf{A} \in M(p)$, then $\alpha(t) = (0, \exp t\mathbf{A}) \equiv (0, \Sigma_r t^r \mathbf{A}^r / r!)$ is a curve in $\mathrm{AL}(p)$ with $\alpha(0) = (\mathbf{0}, \mathbf{I})$ and $\alpha'(0) = (\mathbf{0}, \mathbf{A})$. Thus

$$\begin{aligned}
&((\mathbf{a}, \mathbf{0})^*(\boldsymbol{\mu}, \mathbf{V}))^t \cdot I(\boldsymbol{\mu}, \mathbf{V}) \cdot ((\mathbf{0}, \mathbf{A})^*(\boldsymbol{\mu}, \mathbf{V})) \\
&\quad = -E_{\mathbf{x}}\left[\frac{\partial^2}{\partial t\, \partial s}\bigg|_{s,t=0} \log f\big(\mathbf{x} | (\exp t\mathbf{A})\boldsymbol{\mu} + s\mathbf{a}, (\exp t\mathbf{A})\mathbf{V}(\exp t\mathbf{A})^t\big) \right] \\
&\quad = -E_{\mathbf{x}}\bigg[\frac{\partial^2}{\partial t\, \partial s}\bigg|_{s,t=0} - \log\det\exp t\mathbf{A} - \frac{1}{2}(\mathbf{x} - (\exp t\mathbf{A})\boldsymbol{\mu} - s\mathbf{a})^t \qquad (3) \\
&\qquad\qquad \times (\exp - t\mathbf{A}^t)\mathbf{V}^{-1}(\exp - t\mathbf{A})(\mathbf{x} - (\exp t\mathbf{A})\boldsymbol{\mu} - s\mathbf{a}) \bigg] \\
&\quad = \frac{1}{2}\boldsymbol{\mu}^t \mathbf{A}^t \mathbf{V}^{-1}\mathbf{a} + \frac{1}{2}\mathbf{a}^t \mathbf{V}^{-1}\mathbf{A}\boldsymbol{\mu} = \mathbf{a}^t \mathbf{V}^{-1}\mathbf{A}\boldsymbol{\mu}.
\end{aligned}$$

In (3) we have used the identities $\exp(-t\mathbf{A}) = (\exp t\mathbf{A})^{-1}$ and $\det(\exp t\mathbf{A}) = \exp(t \operatorname{Tr} \mathbf{A})$. Similarly, letting $\alpha(t, s) = (\mathbf{0}, \exp(t\mathbf{A} + s\mathbf{B})) = (\mathbf{0}, \mathbf{I} + t\mathbf{A} +$

$s\mathbf{B} + \frac{1}{2}ts(\mathbf{AB} + \mathbf{BA}) + \cdots)$, we get

$$((\mathbf{0},\mathbf{A})^*(\boldsymbol{\mu},\mathbf{V}))^t \cdot I(\boldsymbol{\mu},\mathbf{V}) \cdot ((\mathbf{0},\mathbf{B})^*(\boldsymbol{\mu},\mathbf{V}))$$

$$\begin{aligned}
&= -E_{\mathbf{x}}\left[\frac{\partial^2}{\partial t\,\partial s}\bigg|_{s,t=0} - \operatorname{Tr}(t\mathbf{A} + s\mathbf{B}) - \frac{1}{2}\operatorname{Tr}(\mathbf{x} - (\exp t\mathbf{A} + s\mathbf{B})\boldsymbol{\mu})\right.\\
&\qquad \times(\mathbf{x} - (\exp t\mathbf{A} + s\mathbf{B})\boldsymbol{\mu})^t(\exp(-t\mathbf{A} - s\mathbf{B})^t)\mathbf{V}^{-1}\\
&\qquad \left.\times(\exp(-t\mathbf{A} - s\mathbf{B}))\right] \qquad (4)\\
&= \frac{1}{2}\bigg\{\boldsymbol{\mu}^t\mathbf{A}^t\mathbf{V}^{-1}\mathbf{B}\boldsymbol{\mu} + \boldsymbol{\mu}^t\mathbf{B}^t\mathbf{V}^{-1}\mathbf{A}\boldsymbol{\mu}\\
&\qquad + \operatorname{Tr}\bigg[\mathbf{VA}^t\mathbf{V}^{-1}\mathbf{B} + \mathbf{VB}^t\mathbf{V}^{-1}\mathbf{A}\\
&\qquad + \frac{1}{2}\mathbf{V}(\mathbf{A}^t\mathbf{B}^t + \mathbf{B}^t\mathbf{A}^t)\mathbf{V}^{-1} + \frac{1}{2}\mathbf{VV}^{-1}(\mathbf{AB} + \mathbf{BA})\bigg]\bigg\}\\
&= \operatorname{Tr}[\mathbf{AB} + \mathbf{VA}^t\mathbf{V}^{-1}\mathbf{B}] + \boldsymbol{\mu}^t\mathbf{A}^t\mathbf{V}^{-1}\mathbf{B}\boldsymbol{\mu}.
\end{aligned}$$

Multiplication in AL(p) is $(\mathbf{b},\mathbf{B}) \cdot (\mathbf{a},\mathbf{A}) = (\mathbf{b} + \mathbf{Ba}, \mathbf{BA})$. Letting $\alpha(t)$ be the curve in AL(p) defined by $\alpha(t) = (t\mathbf{a}, \exp t\mathbf{A})$, where $(\mathbf{a},\mathbf{A}) \in R^p \times M(p)$, then,

$$\begin{aligned}
\operatorname{ad}(\mathbf{b},\mathbf{B}) \cdot (\mathbf{a},\mathbf{A}) &= \frac{d}{dt}\bigg|_{t=0} (\mathbf{b},\mathbf{B})(t\mathbf{a}, \exp t\mathbf{A})(\mathbf{b},\mathbf{B})^{-1}\\
&= \frac{d}{dt}\bigg|_{t=0} (\mathbf{b} + t\mathbf{Ba} - \mathbf{B}(\exp t\mathbf{A})\mathbf{B}^{-1}\mathbf{b}, \mathbf{B}(\exp t\mathbf{A})\mathbf{B}^{-1}) \qquad (5)\\
&= (\mathbf{Ba} - \mathbf{BAB}^{-1}\mathbf{b}, \mathbf{BAB}^{-1}).
\end{aligned}$$

REMARK. These calculations can be extended, in a manner similar to that used in the previous section, to models of the form

$$f(\mathbf{x}|\boldsymbol{\mu},\mathbf{V}) = c(\det\mathbf{V})^{-1/2}\exp g\big((\mathbf{x} - \boldsymbol{\mu})^t\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\big). \qquad (6)$$

Let $\mathbf{W} = \mathbf{V}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ and write $\mathbf{W} = [w_1, \ldots, w_p]^t$. The density of $\mathbf{W}$ is $f(\cdot\,|\mathbf{0},\mathbf{I})$. Define constants k_1, k_2, k_3 and k_4 by

$$\begin{aligned}
k_1 &= -4E\big(g''(\mathbf{W}^t\mathbf{W})w_1^2\big) - 2E\big(g'(\mathbf{W}^t\mathbf{W})\big),\\
k_2 &= -2E\big(g'(\mathbf{W}^t\mathbf{W})w_1^2\big),\\
k_3 &= -4E\big(g''(\mathbf{W}^t\mathbf{W})w_1^2w_2^2\big),\\
k_4 &= -4E\big(g''(\mathbf{W}^t\mathbf{W})w_1^4\big) - 3k_3.
\end{aligned}$$

The constants k_1, k_2, k_3 and k_4 are components of the information matrix I.

Then

$$((\mathbf{a},\mathbf{0})^*(\boldsymbol{\mu},\mathbf{V}))^t \cdot I(\boldsymbol{\mu},\mathbf{V}) \cdot ((\mathbf{b},\mathbf{0})^*(\boldsymbol{\mu},\mathbf{V})) = k_1 \mathbf{a}^t \mathbf{V}^{-1}\mathbf{b},$$

$$((\mathbf{a},\mathbf{0})^*(\boldsymbol{\mu},\mathbf{V}))^t \cdot I(\boldsymbol{\mu},\mathbf{V}) \cdot ((\mathbf{0},\mathbf{A})^*(\boldsymbol{\mu},\mathbf{V})) = k_1 \mathbf{a}^t \mathbf{V}^{-1}\mathbf{A}\boldsymbol{\mu},$$

$$\begin{aligned}((\mathbf{0},\mathbf{A})^*(\boldsymbol{\mu},\mathbf{V}))^t &\cdot I(\boldsymbol{\mu},\mathbf{V}) \cdot ((\mathbf{0},\mathbf{B})^*(\boldsymbol{\mu},\mathbf{V}))\\ &= (k_2 + k_3)\mathrm{Tr}[\mathbf{AB} + \mathbf{VA}^t\mathbf{V}^{-1}\mathbf{B}], + k_1\boldsymbol{\mu}^t\mathbf{A}^t\mathbf{V}^{-1}\mathbf{B}\boldsymbol{\mu} + k_3 \,\mathrm{Tr}\,\mathbf{B}\,\mathrm{Tr}\,\mathbf{A}\\ &\quad + k_4 \,\mathrm{Tr}\big((\mathbf{V}^{-1/2}\mathbf{B}\mathbf{V}^{1/2})^*(\mathbf{V}^{-1/2}\mathbf{A}\mathbf{V}^{1/2})\big).\end{aligned}$$

Here $\mathbf{A}^*\mathbf{B}$ denotes the Hadamard product of the square matrices $\mathbf{A}$ and $\mathbf{B}$: namely, the matrix $\mathbf{C}$, whose entries are $c_{ij} = a_{ij}b_{ij}$.

Since $\mathrm{AL}(p)$ acts transitively on $\boldsymbol{\Omega}$, (1) is still Jeffreys' prior for group models of the form (6).

3. Some noninformative priors for the multivariate normal distribution. Let $\mathbf{G} \subseteq \mathrm{AL}(p)$ be the subgroup of matrices $\{(\mathbf{a},\mathbf{A}) | \mathbf{A}$ is diagonal with positive entries$\}$. Then $\mathbf{G}$ acts freely on $\boldsymbol{\Omega} = \{(\boldsymbol{\mu},\mathbf{V})\}$ and the orbit space is parameterized by the correlation matrix $\boldsymbol{\rho}$.

Given $(\boldsymbol{\mu},\mathbf{V}) \in \boldsymbol{\Omega}$ consider $(-\boldsymbol{\mu},\mathbf{I}) \in \mathbf{G}$. Now $\mathscr{L}(\mathbf{G}) = \{(\mathbf{a},\mathbf{A}) | \mathbf{A}$ is diagonal$\}$ and by (5), $\mathrm{ad}(-\boldsymbol{\mu},\mathbf{I}) \cdot (\mathbf{a},\mathbf{A}) = (\mathbf{a} - \mathbf{A}\boldsymbol{\mu},\mathbf{A})$. It follows that $\det(\mathrm{ad}(-\boldsymbol{\mu},\mathbf{I})) = 1$. Since $(-\boldsymbol{\mu},\mathbf{I}) \cdot (\boldsymbol{\mu},\mathbf{V}) = (\mathbf{0},\mathbf{V})$, we have by Proposition 1(e)

$$J(\boldsymbol{\mu},\mathbf{V}) = J(\mathbf{0},\mathbf{V}).$$

Using equations (2) and (3)

$$((\mathbf{a},\mathbf{0})^*(\mathbf{0},\mathbf{V}))^t \cdot I(\mathbf{0},\mathbf{V}) \cdot ((\mathbf{b},\mathbf{0})^*(\mathbf{0},\mathbf{V})) = \mathbf{a}^t\mathbf{V}^{-1}\mathbf{b},$$

$$((\mathbf{a},\mathbf{0})^*(\mathbf{0},\mathbf{V}))^t \cdot I(\mathbf{0},\mathbf{V}) \cdot ((\mathbf{0},\mathbf{A})^*(\mathbf{0},\mathbf{V})) = 0.$$

Let $\mathbf{E}_j$ be the diagonal matrix with all 0's except for a 1 in the (j,j)th spot. Then using (4),

$$\begin{aligned}((\mathbf{0},\mathbf{E}_i)^*(\mathbf{0},\mathbf{V}))^t \cdot I(\mathbf{0},\mathbf{V}) \cdot \big((\mathbf{0},\mathbf{E}_j)^*(\mathbf{0},\mathbf{V})\big) &= \mathrm{Tr}\big[\mathbf{E}_i\mathbf{E}_j + \mathbf{V}\mathbf{E}_i\mathbf{V}^{-1}\mathbf{E}_j\big]\\ &= \delta_{ij} + v_{ji}v^{ij},\end{aligned}$$

where $\mathbf{V} = [v_{ij}]$ and $\mathbf{V}^{-1} = [v^{ij}]$.

Thus $J(\mathbf{0},\mathbf{V}) = \det(\mathbf{I} + \mathbf{V}^*\mathbf{V}^{-1})/\det\mathbf{V}$ and using (1),

$$\begin{aligned}p(\boldsymbol{\mu},\mathbf{V})\,d\boldsymbol{\mu}\,d\mathbf{V} &= (\det\mathbf{V})^{-(p+1)/2}\big(\det(\mathbf{I} + \mathbf{V}^*\mathbf{V}^{-1})\big)^{-1/2}\,d\boldsymbol{\mu}\,d\mathbf{V}\\ &= 2^p\Big[\prod_i\big(\sigma_i^{-1}\,d\mu_i\,d\sigma_i\big)\Big]\\ &\quad\times\Big[(\det\boldsymbol{\rho})^{-(p+1)/2}\big(\det(\mathbf{I} + \boldsymbol{\rho}^*\boldsymbol{\rho}^{-1})\big)^{-1/2}\prod_{i<j}d\rho_{ij}\Big].\end{aligned} \tag{7}$$

Recognizing $\prod_i(\sigma_i^{-1}\,d\mu_i\,d\sigma_i)$ as right Haar measure on $\mathbf{G}$, we get

$$p(\boldsymbol{\rho}) = (\det\boldsymbol{\rho})^{-(p+1)/2}\big(\det(\mathbf{I} + \boldsymbol{\rho}^*\boldsymbol{\rho}^{-1})\big)^{-1/2}.$$

The prior (7) is inequivalent to priors of the form $(\det \mathbf{V})^r \, d\boldsymbol{\mu}\, d\mathbf{V}$ considered by Geisser (1965) and Dawid, Stone and Zidek (1973). When $p = 2$, marginalization paradoxes with respect to ρ are avoided only when $r = -\frac{3}{2}$ and this is the prior recommended by Geisser. Then

$$(\det \mathbf{V})^{-3/2}\, d\boldsymbol{\mu}\, d\mathbf{V} = 4\, d\mu_1\, d\mu_2 \frac{d\sigma_1}{\sigma_1}\frac{d\sigma_2}{\sigma_2}\frac{d\rho}{(1-\rho^2)^{3/2}},$$

whereas the prior (7) is Lindley's prior

$$2\, d\mu_1\, d\mu_2 \frac{d\sigma_1}{\sigma_1}\frac{d\sigma_2}{\sigma_2}\frac{d\rho}{1-\rho^2}.$$

Now let $\mathbf{G} \subseteq \mathrm{AL}(p)$ be the subgroup of matrices $\{(\mathbf{a}, \mathbf{A}) | \mathbf{A}\mathbf{A}^t = \mathbf{I}\}$. The orbits of $\mathbf{G}$ in $\boldsymbol{\Omega}$ are in one-to-one correspondence with the p-tuples $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$ of eigenvalues of the covariance matrix $\mathbf{V}$. $\mathbf{G}$ does not act freely on $\boldsymbol{\Omega}$. Nevertheless, if $\mathbf{A}\mathbf{A}^t = \mathbf{I}$ and $\mathbf{V}$ is diagonal with distinct entries, then $\mathbf{A}\mathbf{V}\mathbf{A}^t = \mathbf{V}$ implies that $\mathbf{A}$ is diagonal with entries ± 1. Let $\boldsymbol{\Omega}_0 = \{(\boldsymbol{\mu}, \mathbf{V}) | \mathbf{V}$ has distinct eigenvalues$\}$. If $(\boldsymbol{\mu}, \mathbf{V}) \in \boldsymbol{\Omega}_0$, its isotropy group $\mathbf{G}_{(\boldsymbol{\mu}, \mathbf{V})} = \{g \in \mathbf{G} | g \cdot (\boldsymbol{\mu}, \mathbf{V}) = (\boldsymbol{\mu}, \mathbf{V})\}$ is isomorphic to $(Z_2)^p$ and has order 2^p.

Recall that two subgroups $\mathbf{H}_1$ and $\mathbf{H}_2$ of $\mathbf{G}$ are said to be conjugate if there is a $g \in \mathbf{G}$ so that $\mathbf{H}_1 = g\mathbf{H}_2 g^{-1}$. If $\mathbf{G}$ acts on $\boldsymbol{\Omega}$ with all isotropy subgroups conjugate to the finite subgroup $\mathbf{H}$, then all the orbits are of the form $\mathbf{G}/\mathbf{H}$, with the action of $\mathbf{G}$ on an orbit being left multiplication of left cosets. In our example, $\mathbf{H} = (Z_2)^p$. Then under reasonable conditions (see Appendix), one can find local decompositions $\boldsymbol{\Omega} \approx \mathbf{G}/\mathbf{H} \times \boldsymbol{\Omega}/\mathbf{G}$, with $\mathbf{G}$ acting on the first factor only. We have the following extension of Theorem 2:

THEOREM 3. *Suppose $\mathbf{G}$ acts on $\boldsymbol{\Omega}$ with all isotropy subgroups conjugate to the finite subgroup $\mathbf{H}$ and freely almost everywhere on $\mathbf{X}$. Let $\tilde{p}(\omega)$ be overall Jeffreys' prior. Let $p(\omega)$ be the prior defined by*

(i) *In any local decomposition $\boldsymbol{\Omega} \approx \mathbf{G}/\mathbf{H} \times \boldsymbol{\Omega}/\mathbf{G}$, the conditional distribution of $p(g\mathbf{H}|\theta)$ of $g\mathbf{H}$ given θ is the measure induced on $\mathbf{G}/\mathbf{H}$ from a right Haar measure on $\mathbf{G}$ which is independent of θ.*

(ii) *The marginal prior $p(\theta)$ of θ is the limit of Jeffreys' prior for the sampling distribution of the maximal invariant. Then $p(\omega)$ is the Bernardo minimally θ-informative prior. Furthermore it can be calculated by*

$$p(\omega) = \tilde{p}(\omega) J(\omega)^{-1/2}.$$

In our case, $\mathbf{G}$ is unimodular. Hence using Proposition 1(e) and equation (5) to calculate $J(\boldsymbol{\mu}, \mathbf{V})$, we can restrict ourselves to the case $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{V}$ is diagonal. From equations (2) and (3),

$$((\mathbf{a}, \mathbf{0})^*(\mathbf{0}, \mathbf{V}))^t \cdot I(\mathbf{0}, \mathbf{V}) \cdot ((\mathbf{b}, \mathbf{0})^*(\mathbf{0}, \mathbf{V})) = \mathbf{a}^t \mathbf{V}^{-1} \mathbf{b},$$

$$((\mathbf{a}, \mathbf{0})^*(\mathbf{0}, \mathbf{V}))^t \cdot I(\mathbf{0}, \mathbf{V}) \cdot ((\mathbf{0}, \mathbf{A})^*(\mathbf{0}, \mathbf{V})) = 0.$$

$\mathscr{L}(\mathbf{G}) = R^p \times \mathscr{L}(O(p))$, where $\mathscr{L}(O(p)) = \{\mathbf{A} \in \mathbf{M}(p) | \mathbf{A} + \mathbf{A}^t = \mathbf{0}\}$. Let $\mathbf{E}_{ij}$ be the matrix with 1 in the (i, j)th place and -1 in the (j, i)th place. The $\mathbf{E}_{ij}$

for $i < j$ form a basis of $\mathscr{L}(O(p))$. Then using (4),

$$((\mathbf{0}, \mathbf{E}_{ij})^*(\mathbf{0}, \mathbf{V}))^t \cdot I(\mathbf{0}, \mathbf{V}) \cdot ((\mathbf{0}, \mathbf{E}_{kl})^*(\mathbf{0}, \mathbf{V})) = 0, \quad \text{if } i \neq k \text{ or } j \neq l,$$

$$((\mathbf{0}, \mathbf{E}_{ij})^*(\mathbf{0}, \mathbf{V}))^t \cdot I(\mathbf{0}, \mathbf{V}) \cdot ((\mathbf{0}, \mathbf{E}_{ij})^*(\mathbf{0}, \mathbf{V})) = \frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i} - 2,$$

where $\mathbf{V}$ has eigenvalues $\lambda_1, \dots, \lambda_p$.

It follows that

$$J(0, \mathbf{V}) = \left[\prod_i \lambda_i^{-1}\right]\left[\prod_{i<j}\left(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i} - 2\right)\right] = (\det \mathbf{V})^{-p} \prod_{i<j} (\lambda_i - \lambda_j)^2$$

and therefore $p(\mu, \mathbf{V}) = (\det \mathbf{V})^{-1} \Pi_{i<j} |\lambda_i - \lambda_j|^{-1}$.

APPENDIX A

Topological preliminaries. In this section we assume the action of $\mathbf{G}$ on $\mathbf{\Omega}$ has all isotropy groups conjugate to a fixed group $\mathbf{H}$ and that $\mathbf{\Omega}$ is a connected differentiable Cartan $\mathbf{G}$-space [see Palais (1961)].

Following standard notation, we shall write $\mathbf{G}_\omega$ for the isotropy subgroup of $\omega \in \mathbf{\Omega}$ (that is, $\mathbf{G}_\omega = \{g \in \mathbf{G} | g \cdot \omega = \omega\}$). We note that $\mathbf{G}_{g \cdot \omega} = g\mathbf{G}_\omega g^{-1}$. In other words, $g \cdot \omega$ and ω have conjugate isotropy subgroups.

The assumptions ensure that $\mathbf{\Omega}/\mathbf{G}$ is a differentiable manifold and that the map $\pi: \mathbf{\Omega} \to \mathbf{\Omega}/\mathbf{G}$, which takes each $\omega \in \mathbf{\Omega}$ to its orbit, is differentiable. The differential structure on $\mathbf{\Omega}/\mathbf{G}$ is defined in such a way that a real valued function f on $\mathbf{\Omega}/\mathbf{G}$ is differentiable if and only if $f\pi$ is. Indeed, much more can be said about the structure of π. Let $N(\mathbf{H})$ be the normalizer of $\mathbf{H}$ (that is, $N(\mathbf{H}) = \{g \in \mathbf{G} | g\mathbf{H}g^{-1} = \mathbf{H}\}$). $N(\mathbf{H})/\mathbf{H}$ has a natural *right* action on $\mathbf{G}/\mathbf{H}$, the *left* cosets of $\mathbf{H}$. That action is for $g \in \mathbf{G}$ and $n \in N(\mathbf{H})$: $(g\mathbf{H}) \cdot (n\mathbf{H}) = gn\mathbf{H}$. We have:

PROPOSITION 4. *$\pi: \mathbf{\Omega} \to \mathbf{\Omega}/\mathbf{G}$ is a differentiable equivariant fiber bundle with structure group $N(\mathbf{H})/\mathbf{H}$ and fiber $\mathbf{G}/\mathbf{H}$. That is*:

(i) *Every point in $\mathbf{\Omega}/\mathbf{G}$ has an open neighborhood U in $\mathbf{\Omega}/\mathbf{G}$ such that there exists a diffeomorphism φ_U: $\mathbf{G}/\mathbf{H} \times U \to \pi^{-1}(U)$ with the properties*: (a) *If π_2 is the projection of $\mathbf{G}/\mathbf{H} \times U$ onto its second factor, then $\pi_2 = \pi\varphi_U$.* (b) *If $g \in \mathbf{G}$, $\varphi_U(g\tilde{g}\mathbf{H}, \theta) = g \cdot \varphi_U(\tilde{g}\mathbf{H}, \theta)$. That is, under φ_U, the action of $\mathbf{G}$ on $\mathbf{\Omega}$ corresponds to left multiplication of left cosets.*

(ii) *If U and V are open sets in $\mathbf{\Omega}/\mathbf{G}$ with charts φ_U and φ_V, respectively, then there is a differentiable map s: $U \cap V \to N(\mathbf{H})/\mathbf{H}$ such that*

$$\varphi_U(g\mathbf{H}, \theta) = \varphi_V(g\mathbf{H}s(\theta), \theta) \quad \textit{for all } \theta \in U \cap V \textit{ and } g \in \mathbf{G}.$$

(In particular, if $\mathbf{G}$ acts freely on $\mathbf{\Omega}$, s: $U \cap V \to \mathbf{G}$.)

Furthermore, the fiber bundle structure on $\pi: \mathbf{\Omega} \to \mathbf{\Omega}/\mathbf{G}$ *is unique in the sense that any two maximal collections of local trivializations* $\{(U, \varphi_U)\}$ *satisfying* (i) *and* (ii) *are the same.*

PROOF. If $\mathbf{G}$ is compact, Proposition 4 is standard. It can be found, for example, less the statement of uniqueness, as Corollary 2.5 on page 309 of Bredon (1972). The compactness condition is used only to show that slices exist at each point of $\mathbf{\Omega}$. The existence of slices is guaranteed by the assumption of Cartan action [Palais (1961), Proposition 2.2.2].

Let $\omega \in \mathbf{\Omega}$ and pick $g \in \mathbf{G}$ so that $g \cdot \omega$ has isotropy group $\mathbf{H}$. Section 2.2 of Palais shows that the Cartan condition is sufficient to insure that there is a submanifold, called a *slice*, S of $\mathbf{\Omega}$ and a differentiable equivariant f: $\mathbf{G}S = \{g \cdot s | g \in \mathbf{G} \text{ and } s \in S\} \to \mathbf{G}/\mathbf{H}$ such that $\mathbf{G}S$ is open neighborhood of ω in $\mathbf{\Omega}$ and $S = f^{-1}(\mathbf{H})$. Since f is equivariant and all isotropy subgroups are conjugate to $\mathbf{H}$, S intersects each orbit in $\mathbf{G}S$ exactly once and each element of S has isotropy group $\mathbf{H}$. Then (see Palais, Remark 2.2.3) F_S: $\mathbf{G}/\mathbf{H} \times S \to \mathbf{G}S$ defined by $F_S(g\mathbf{H}, \omega) = g \cdot \omega$ is a diffeomorphism.

Let $U = \pi(S) = \pi(\mathbf{G}S)$. Then U is open in the quotient topology of $\mathbf{\Omega}/\mathbf{G}$ and the restriction $\pi|_S$ of π to S is a homeomorphism of S onto U. Define φ_U: $\mathbf{G}/\mathbf{H} \times U \to \pi^{-1}(U) = \mathbf{G}S$ by $\varphi_U(g\mathbf{H}, \theta) = F_S(g\mathbf{H}, (\pi|_S)^{-1}(\theta))$. It is easily shown that φ_U is a homeomorphism and that (i)(a) and (i)(b) hold.

If S_1 is any other slice, then since F_{S_1} is a diffeomorphism and both S_1 and S are fixed by $\mathbf{H}$, there is a differentiable f_{S_1S} from $S \cap \mathbf{G}S_1$ into $N(\mathbf{H})/\mathbf{H}$ such that for each $\omega \in S \cap \mathbf{G}S_1$, $\omega = f_{S_1S}(\omega) \cdot \omega_1$ with $\omega_1 \in S_1 \cap \mathbf{G}\omega$.

We note that $N(\mathbf{H})/\mathbf{H}$ is a group and $(\pi|_{S_1})^{-1}(\pi|_S)(\omega) = f_{S_1S}(\omega)^{-1} \cdot \omega$ on $S \cap \mathbf{G}S_1$. It follows that differentiability on $\mathbf{\Omega}/\mathbf{G}$ can be unambiguously defined by demanding that each $\pi|_S$ be differentiable. It is now easily shown that a function f on $\mathbf{\Omega}/\mathbf{G}$ is differentiable if and only if $f\pi$ is differentiable, that each φ_U is a diffeomorphism and that (ii) holds.

Now suppose φ_U: $\mathbf{G}/\mathbf{H} \times U \to \pi^{-1}(U)$ is any diffeomorphism satisfying (i)(a) and (i)(b). Then $S = \{\varphi_U(\mathbf{H}, u) \text{ for } u \in U\}$ is a slice and the uniqueness of the fiber bundle structure follows. □

REMARK. Proposition 4 does not require the finiteness of $\mathbf{H}$. If $\mathbf{\Omega}$ is not a differentiable Cartan $\mathbf{G}$-space but merely a topological one, Proposition 4 is true without differentiability in its conclusion.

The assumption that all points in $\mathbf{\Omega}$ have isotropy subgroups conjugate to some fixed subgroup $\mathbf{H}$, can also be usually finessed using the principal isotropy group (p.i.g.) or principal orbit type (p.o.t.) theorem. That theorem asserts that if $\mathbf{G}$ is compact, then there is a subgroup $\mathbf{H}$ of $\mathbf{G}$ and an open and dense subset $\mathbf{\Omega}_0$ of $\mathbf{\Omega}$, such that each point of $\mathbf{\Omega}_0$ has isotropy subgroup conjugate to $\mathbf{H}$. Under the assumption of differentiability, the p.i.g. theorem has an elementary proof which can be found in tom Dieck [(1987), page 43] or Bredon [(1972), page 179–180]. The proof only uses the existence of slices and the compactness of all isotropy groups and hence the condition that $\mathbf{G}$ be compact can be replaced by the Cartan condition. The p.i.g. theorem is actually true without differentiability restrictions; see Borel (1960).

APPENDIX B

Proofs. In this paper, we assume:

(i) $\mathbf{G}$ is a Lie group and the action of $\mathbf{G}$ on the sample space $\mathbf{X}$ is assumed to be free and such that a measurable cross section $\mathbf{X} \approx \mathbf{G} \times \mathbf{X}/\mathbf{G}$ exists. Bondar (1976) discusses conditions which ensure the existence of a measurable cross section.

(ii) We assume the necessary regularity conditions to ensure the correctness of asymptotics for both the statistical models $(\mathbf{\Omega}, \mathbf{X})$ and $(\mathbf{\Omega}/\mathbf{G}, \mathbf{X}/\mathbf{G})$. Lehmann [(1983), pages 406–415] gives these conditions in greater detail.

(iii) We assume the action of $\mathbf{G}$ on $\mathbf{\Omega}$ has all isotropy groups conjugate to a fixed group $\mathbf{H}$, that $\mathbf{H}$ is finite and that $\mathbf{\Omega}$ is a connected differentiable Cartan $\mathbf{G}$-space [see Palais (1961)].

It is well known [see Amari (1985), Kass (1989)] that Fisher information can be considered as a Riemannian metric on $\mathbf{\Omega}$: that is, for each $\omega \in \mathbf{\Omega}$, Fisher information defines an inner product on the tangent vectors $T_\omega\mathbf{\Omega}$ to $\mathbf{\Omega}$ at ω. To emphasize this coordinate-free viewpoint, if $X, Y \in T_\omega\mathbf{\Omega}$, we will write $I_\omega\langle X, Y\rangle$ for the Fisher information inner product of X and Y. In the main body of this paper, $I_\omega\langle X, Y\rangle$ was denoted $X^t \cdot I(\omega) \cdot Y$. When there is no danger of confusion, we will usually omit the subscripted ω. Thus if $c_1(t)$ and $c_2(t)$ are curves in $\mathbf{\Omega}$ with $c_1(0) = c_2(0) = \omega$,

$$I\langle X, Y\rangle = E_\omega\left[\frac{\partial \log f(x; c_1(t))}{\partial t}\bigg|_{t=0} \frac{\partial \log f(x; c_2(s))}{\partial s}\bigg|_{s=0}\right]. \tag{8}$$

For $g \in \mathbf{G}$, let $L_g\colon \mathbf{\Omega} \to \mathbf{\Omega}$ be the map $L_g(\omega) = g \cdot \omega$ and $DL_g\colon T_\omega\mathbf{\Omega} \to T_{g\cdot\omega}\mathbf{\Omega}$ be its derivative defined by

$$DL_g(c'(0)) = \frac{d}{dt}\bigg|_{t=0} g \cdot c(t),$$

where we notice that $g \cdot c(t)$ is interpreted as a curve in $\mathbf{\Omega}$. It is well known that I is $\mathbf{G}$-invariant, that is,

$$I_{g\cdot\omega}\langle DL_g X, DL_g Y\rangle = I_\omega\langle X, Y\rangle. \tag{9}$$

Since Jeffreys' overall prior on $\mathbf{\Omega}$ is the volume element on $\mathbf{\Omega}$ induced by I, it is also $\mathbf{G}$-invariant.

For $\omega \in \mathbf{\Omega}$, let $\phi_\omega\colon \mathbf{G}/\mathbf{G}_\omega \to \mathbf{\Omega}$ be defined by

$$\phi_\omega(g\mathbf{G}_\omega) = g \cdot \omega.$$

Since $\mathbf{G}_\omega$ is finite, the Cartan condition implies that ϕ_ω is nonsingular.

Proof of Proposition 1. Part (a) follows trivially from (8) and (b) from (a). The nonsingularity of ϕ_ω implies that (c) is true. The truth of (d) is immediate from the bilinearity of $I\langle\ ,\ \rangle$.

We have for $A \in \mathscr{L}(\mathbf{G})$ and $\alpha(t)$ a curve in $\mathbf{G}$ with $\alpha(0) = 1$ and $\alpha'(0) = A$,

$$A^*(g \cdot \omega) = \frac{d}{dt}\Big|_{t=0} \alpha(t) \cdot (g \cdot \omega) = \frac{d}{dt}\Big|_{t=0} g\big(g^{-1}\alpha(t)g\big) \cdot \omega$$
$$= (DL_g)\big(\mathrm{ad}(g^{-1})A\big)^*(\omega),$$

and (e) follows from (d) and (9). Since $\mathbf{G}$ is unimodular is equivalent to $\det[\mathrm{ad}(g^{-1})] = \pm 1$, (f) is immediate from (e). □

Let $\tilde{T}_\omega\mathbf{\Omega} = \{X \in T_\omega\mathbf{\Omega}$ such that $I\langle X, A^*(\omega)\rangle = 0$ for all $A \in \mathscr{L}(\mathbf{G})\}$. Using (9), DL_g takes $\tilde{T}_\omega\mathbf{\Omega}$ into $\tilde{T}_{g\cdot\omega}\mathbf{\Omega}$. $\tilde{T}_\omega\mathbf{\Omega}$ is a vector subspace of $T_\omega\mathbf{\Omega}$ of dimension equal to $\dim\mathbf{\Omega} - \dim\mathbf{G}$. Since the vectors of the form $A^*(\omega)$ span the vectors tangent to the orbit at ω, $\tilde{T}_\omega\mathbf{\Omega}$ is a complementary subspace to the tangent vectors to the orbit. Define a new inner product $\tilde{I}\langle X, Y\rangle$ on each tangent space to $\mathbf{\Omega}$ by

$$\tilde{I}\langle X, Y\rangle = I\langle X, Y\rangle, \quad \text{if } X \text{ or } Y \text{ is in } \tilde{T}_\omega\mathbf{\Omega},$$

$$\tilde{I}\langle A^*(\omega), B^*(\omega)\rangle = I\langle A^*(\omega), B^*(\omega)\rangle / J(\omega)^{1/k},$$

where $k = \dim\mathbf{G}$. Jeffreys' prior $\tilde{p}(\omega)\,d\omega$ is the volume element of $\mathbf{\Omega}$ using the Riemannian metric I. Let $p(\omega)\,d\omega$ be the volume element of $\mathbf{\Omega}$ using the metric $\tilde{I}$. That is, if $M(\tilde{I}(\omega))$ is the matrix of the Riemannian metric $\tilde{I}$ using *any* parameterization of $\mathbf{\Omega}$,

$$p(\omega)\,d\omega = \det\big(M\big(\tilde{I}(\omega)\big)\big)^{1/2}\,d\omega. \tag{10}$$

LEMMA 5. (a) *Suppose $\mathbf{G}_\omega = \mathbf{H}$. Then the measure induced by $\tilde{I}$ on the orbit through ω is, after pulling back to $\mathbf{G}/\mathbf{H}$ using ϕ_ω, a right Haar measure which does not depend upon ω.*

(b) $p(\omega) = \tilde{p}(\omega)J(\omega)^{-1/2}$.

PROOF OF LEMMA 5. Let $A \in \mathscr{L}(\mathbf{G})$ and $\alpha(t)$ be a curve in $\mathbf{G}$ with $\alpha(0) = 1$ and $\alpha'(0) = A$. Extend A to a right invariant vector field on $\mathbf{G}$ by $A(g) = d/dt|_{t=0}\,\alpha(t)g$. Let $q\colon \mathbf{G} \to \mathbf{G}/\mathbf{H}$ be projection and define a Riemannian metric $I_{G,\omega}$ on $\mathbf{G}$ by

$$I_{G,\omega}\langle A(g), B(g)\rangle = \tilde{I}\langle D(\phi_\omega q)A(g), D(\phi_\omega q)B(g)\rangle.$$

Now $D(\phi_\omega q)A(g) = A^*(g \cdot \omega)$ and hence

$$\det\big[I_{G,\omega}\langle A_i(g), A_j(g)\rangle\big] = \det\big[\tilde{I}\langle A_i^*(g \cdot \omega), A_j^*(g \cdot \omega)\rangle\big] = 1.$$

It follows from the way volume elements are constructed from Riemannian metrics that $I_{G,\omega}$ defines a right Haar measure on $\mathbf{G}$ and that this right Haar measure is independent of ω. In other words, $\tilde{I}$ and ϕ_ω define a measure on $\mathbf{G}/\mathbf{H}$ which pulls back (under q) to a constant right Haar measure on $\mathbf{G}$. This is part (a). Part (b) is immediate. □

Although (10) defines $p(\omega)\,d\omega$ independently of the parameterization of $\mathbf{\Omega}$, to prove Theorems 2 and 3 we can use Proposition 4 and assume that we are working in a fixed local decomposition $\mathbf{G}U \approx \mathbf{G}/\mathbf{H} \times U$ for U open in $\mathbf{\Omega}/\mathbf{G}$.

We will use θ for the generic point of U and will write $\omega = (\bar{g}, \theta)$ to mean that $(g\mathbf{H}, \theta)$ is the representation of ω in this local decomposition.

LEMMA 6. *If the conditional prior on each orbit is a constant right Haar measure, then $p(\omega)$ is the θ-noninformative prior as defined by Bernardo (1979).*

PROOF OF LEMMA 6. Let $\omega = (\bar{g}, \theta)$ be a point in $\mathbf{\Omega}$ and pick local coordinate systems (parameterizations) ψ_1: $R^k \to \mathbf{G}/\mathbf{H}$ and ψ_2: $R^m \to \mathbf{\Omega}/\mathbf{G}$, where $k = \dim \mathbf{G} = \dim \mathbf{G}/\mathbf{H}$ and $m = \dim \mathbf{\Omega} - \dim \mathbf{G}$. Let $\psi_1(v_1) = \bar{g}$ and $\psi_2(v_2) = \theta$. We denote by $D_{(v_1, v_2)}(\psi_1 \times \psi_2)$ the derivative map $R^k \times R^m \to T_\omega \mathbf{\Omega}$ defined by

$$D_{(v_1, v_2)}(\psi_1 \times \psi_2)(\gamma_1'(0), \gamma_2'(0)) = \frac{d}{dt}\bigg|_{t=0} (\psi_1(\gamma_1(t)), \psi_2(\gamma_2(t))),$$

where $\gamma_1(t)$ and $\gamma_2(t)$ are curves in R^k and R^m, respectively, with $\gamma_1(0) = v_1$ and $\gamma_2(0) = v_2$.

Let $M(I(v_1, v_2))$ be the information matrix (using these coordinate systems) at (v_1, v_2). When $\mathbf{A}$ is a $(k+m) \times (k+m)$ matrix, let $\mathbf{A}_{22}$ be the lower $m \times m$ submatrix of $\mathbf{A}$.

Let ρ: $T_\omega \mathbf{\Omega} \to \tilde{T}_\omega \mathbf{\Omega}$ be orthogonal projection under I. The matrix of inner product $\langle\ ,\ \rangle_{(v_1, v_2)}$ on $R^m \times R^m$ given by the composition

$$R^m \times R^m \xrightarrow{\approx} (\{0\} \times R^m) \times (\{0\} \times R^m) \subseteq (R^k \times R^m) \times (R^k \times R^m)$$
$$\xrightarrow{D_{(v_1, v_2, v_1, v_2)}(\psi_1 \times \psi_2 \times \psi_1 \times \psi_2)} T_\omega \mathbf{\Omega} \times T_\omega \mathbf{\Omega} \xrightarrow{\rho \times \rho} \tilde{T}_\omega \mathbf{\Omega} \times \tilde{T}_\omega \mathbf{\Omega} \xrightarrow{I} R,$$

is

$$M(I(v_1, v_2))_{22} - M(I(v_1, v_2))_{21} \cdot \left[M(I(v_1, v_2))_{11}\right]^{-1} \cdot M(I(v_1, v_2))_{12}$$
$$= \left[M(I(v_1, v_2))^{-1}\right]_{22}^{-1}.$$

We claim that $\langle\ ,\ \rangle_{(v_1, v_2)}$ and hence $[M(I(v_1, v_2))^{-1}]_{22}^{-1}$ do not depend upon v_1. Indeed, given v_1', let g be so that $\psi_1(v_1') = g \cdot \psi_1(v_1)$ and let $\alpha_1(t)$ and $\alpha_2(t)$ be curves in R^m with $\alpha_i(0) = v_2$. Then using the invariance of I [equation (9)] and that in each trivialization given by Proposition 4, the action of $\mathbf{G}$ on $\mathbf{\Omega}$ is left multiplication in the first factor

$$\begin{aligned}
\rho\big(D_{(v_1', v_2)}(\psi_1 \times \psi_2)(0, \alpha_i'(0))\big) &= \rho\left(\frac{d}{dt}\bigg|_{t=0} [\psi_1(v_1'), \psi_2(\alpha_i(t))]\right) \\
&= \rho\left(\frac{d}{dt}\bigg|_{t=0} [g \cdot \psi_1(v_1), \psi_2(\alpha_i(t))]\right) \\
&= \rho\left(DL_g \frac{d}{dt}\bigg|_{t=0} [\psi_1(v_1), \psi_2(\alpha_i(t))]\right) \\
&= DL_g\left\{\rho\left(\frac{d}{dt}\bigg|_{t=0} [\psi_1(v_1), \psi_2(\alpha_i(t))]\right)\right\} \\
&= DL_g\big\{\rho\big(D_{(v_1, v_2)}(\psi_1 \times \psi_2)(0, \alpha_i'(0))\big)\big\}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\langle \alpha_1'(0), \alpha_2'(0) \rangle_{(v_1', v_2)} \\
&\quad = I_{g\cdot\omega} \left\langle \rho\big(D_{(v_1', v_2)}(\psi_1 \times \psi_2)(0, \alpha_1'(0))\big), \rho\big(D_{(v_1', v_2)}(\psi_1 \times \psi_2)(0, \alpha_2'(0))\big) \right\rangle \\
&\quad = I_{g\cdot\omega} \left\langle DL_g \rho\big(D_{(v_1, v_2)}(\psi_1 \times \psi_2)(0, \alpha_1'(0))\big), \right. \\
&\qquad\qquad \left. DL_g \rho\big(D_{(v_1, v_2)}(\psi_1 \times \psi_2)(0, \alpha_2'(0))\big) \right\rangle \\
&\quad = I_\omega \left\langle \rho\big(D_{(v_1, v_2)}(\psi_1 \times \psi_2)(0, \alpha_1'(0))\big), \rho\big(D_{(v_1, v_2)}(\psi_1 \times \psi_2)(0, \alpha_2'(0))\big) \right\rangle \\
&\quad = \langle \alpha_1'(0), \alpha_2'(0) \rangle_{(v_1, v_2)}.
\end{aligned}$$

Furthermore we note that ρ is also an orthogonal projection under $\tilde{I}$ and I and $\tilde{I}$ coincide on $\tilde{T}_\omega \mathbf{\Omega}$. Thus we can write

$$\text{(11)} \qquad \left[M(I(v_1, v_2))^{-1}\right]_{22}^{-1} = \left[M(\tilde{I}(v_1, v_2))^{-1}\right]_{22}^{-1}.$$

Since the matrices (11) do not depend upon v_1, we shall usually write them with v_1 deleted.

Let $\bar{p}(v_1)$ be right Haar measure on $\mathbf{G}/\mathbf{H}$ pulled back to R^k using ψ_1. Let $\bar{p}(v_2)$ be the Bernardo (1979) θ-noninformative prior if $\bar{p}(v_1|v_2)$ is set to $\bar{p}(v_1)$. We will show that $\bar{p}(v_2)\bar{p}(v_1|v_2)\, dv_1\, dv_2 = p(v_1, v_2)\, dv_1\, dv_2$, where p is defined using (10).

Following the development of Bernardo, let $p^*(v_2|\mathbf{x}_n)$ be the asymptotic posterior distribution of v_2. We have that $p^*(v_2|\mathbf{x}_n)$ is normal with mean $\hat{v}_2$ and precision matrix $n \cdot [M(I(\cdot, \hat{v}_2))^{-1}]_{22}^{-1}$. Its entropy $H\{p^*(v_2|\mathbf{x}_n)\}$ is (see Section 3.3 of Bernardo)

$$H\{p^*(v_2|\mathbf{x}_n)\} = (m/2)\log(2\pi e/n) - \tfrac{1}{2}\log\det\left(\left[M(I(\cdot, \hat{v}_2))^{-1}\right]_{22}^{-1}\right) + o(1).$$

Given a prior $\bar{\bar{p}}(v_1|v_2)$, the $\pi_n(v_2)$ of Bernardo's equation (17) is proportional to

$$\begin{aligned}
&\exp\left(\tfrac{1}{2}\iint \log\det\left(\left[M(I(\cdot, \hat{v}_2))^{-1}\right]_{22}^{-1}\right) f(\mathbf{x}_n|v_1, v_2)\bar{\bar{p}}(v_1|v_2)\, dv_1\, d\mathbf{x}_n\right) + o(1) \\
&\qquad \propto \exp\left(\tfrac{1}{2}\int \bar{\bar{p}}(v_1|v_2)\log\det\left(\left[M(I(\cdot, v_2))^{-1}\right]_{22}^{-1}\right) dv_1\right) + o(1).
\end{aligned}$$

Thus if $\bar{\bar{p}}(v_1|v_2)$ does not depend upon v_2, the Bernardo prior $\bar{p}(v_2)$ is

$$\bar{p}(v_2)\, dv_2 \propto \det\left(\left[M(I(\cdot, v_2))^{-1}\right]_{22}^{-1}\right)^{1/2} dv_2.$$

By Lemma 5, if $\bar{p}(v_1|v_2)\, dv_1$ is a right Haar measure independent of v_2, then

$$\bar{p}(v_1|v_2)\, dv_1 \propto \det\left(M(\tilde{I}(v_1, \cdot))_{11}\right)^{1/2} dv_1.$$

For a square matrix, $\det A = \det A_{11} \cdot \det[(A^{-1})_{22}^{-1}]$. Using (11),

$$\bar{p}(v_1|v_2)\bar{p}(v_2)\, dv_1\, dv_2 \propto \det\left(M(\tilde{I}(v_1, v_2))\right) dv_1\, dv_2 = p(v_1, v_2)\, dv_1\, dv_2.$$

In other words, the volume element determined by the Riemannian metric $\tilde{I}$ is Bernardo's minimally informative prior about θ. □

Although we will continue to work in some local trivialization of $\mathbf{\Omega}$, we will henceforth eliminate explicit reference to the parameterizations ψ_1 and ψ_2. In particular, we will use the notation $[M(I(\cdot,\theta))^{-1}]_{22}$ for the matrix which was denoted $[M(I(\cdot,v_2))^{-1}]_{22}$ in the previous proof. We will continue to denote by $\mathbf{y}_n$ the maximal invariant of the $\mathbf{G}$ action on a random sample of size n, that is, the orbits of the action of $\mathbf{G}$ on the Cartesian product $\mathbf{X}^n$. Let $M(I_n^G(\theta))$ denote the information matrix calculated from the sampling density $f_n(\mathbf{y}_n|\theta)$ of $\mathbf{y}_n$. $f_n(\mathbf{y}_n|\theta)$ is induced by the quotient map $\mathbf{X}^n \to \mathbf{X}^n/\mathbf{G}$. Let $d\bar{g}_R$ denote right Haar measure on $\mathbf{G}/\mathbf{H}$.

LEMMA 7. *Suppose that for all n sufficiently large, $\mathbf{G}$ acts freely on $\mathbf{X}^n$ except for a set of measure zero. As $n \to \infty$, $M(I_n^G(\theta))/n$ has a limit $M(I^G(\theta))$ and in a local trivialization, $p(\omega)\,d\omega = \det(M(I^G(\theta)))^{1/2}\,d\theta\,d\bar{g}_R$.*

REMARK. In the notation of the proof of Lemma 6, $M(I^G(\theta)) = [M(I(\cdot,\theta))^{-1}]_{22}^{-1}$.

PROOF. Let θ_0 be fixed and let $\mathbf{x}_i$ have the density $f(\mathbf{x}|1,\theta_0)$. Let $p(\theta)\,d\theta$ be an arbitrary prior in θ and use the prior $p(\theta)\,d\theta\,d\bar{g}_R$ on $\mathbf{\Omega}$. Let $(\hat{g},\hat{\theta})$ be the MLE calculated from $\Pi_i f(\mathbf{x}_i|1,\theta_0)$. Asymptotically, the marginal posterior of θ is normal with mean $\hat{\theta}$ and precision matrix $n\cdot[M(I(\cdot,\hat{\theta}))^{-1}]_{22}^{-1}$. Since right Haar measure has been used on $\mathbf{G}/\mathbf{H}$, this posterior can be calculated using the f_n. [The condition (i) has been used here.] By calculating the log of this posterior using an expansion of $\log f_n$ about $\hat{\theta}$, we have for each θ,

$$\frac{1}{n}\left[\frac{\partial \log f_n}{\partial\theta}(\mathbf{y}_n|\hat{\theta})\right]^t(\theta-\hat{\theta}) + \frac{1}{2n}(\theta-\hat{\theta})^t\left[\frac{\partial^2 \log f_n}{\partial\theta_i\,\partial\theta_j}(\mathbf{y}_n|\hat{\theta})\right](\theta-\hat{\theta})$$

$$= -\frac{1}{2}(\theta-\hat{\theta})^t\cdot\left[M(I(\cdot,\hat{\theta}))^{-1}\right]_{22}^{-1}\cdot(\theta-\hat{\theta}) + o_p(1).$$

Therefore

$$\frac{1}{n}\left[\frac{\partial \log f_n}{\partial\theta}(\mathbf{y}_n|\hat{\theta})\right]$$

is $o_p(1)$ and

$$\frac{1}{2n}\left[\frac{\partial^2 \log f_n}{\partial\theta_i\,\partial\theta_j}(\mathbf{y}_n|\hat{\theta})\right] = -\frac{1}{2}\left[M(I(\cdot,\hat{\theta}))^{-1}\right]_{22}^{-1} + o_p(1).$$

Since $\hat{\theta} = \theta_0 + o_p(1)$,

$$\frac{1}{n}\left[\frac{\partial^2 \log f_n}{\partial\theta_i\,\partial\theta_j}(\mathbf{y}_n|\theta_0)\right] = -\left[M(I(\cdot,\theta_0))^{-1}\right]_{22}^{-1} + o_p(1)$$

and hence taking expected values

$$\frac{1}{n}M\left(I_n^G(\theta_0)\right) = \left[M(I(\cdot,\theta_0))^{-1}\right]_{22}^{-1} + o(1). \qquad \square$$

Lemmas 5, 6 and 7 complete the proof of Theorem 3 and a fortiori Theorem 2.

Acknowledgment. The authors wish to thank the referee whose many helpful comments greatly improved the exposition.

REFERENCES

AMARI, S.-C. (1985). *Differential-Geometrical Methods in Statistics*. Springer, Berlin.

BAYARRI, M. J. (1981). Inferencia Bayesiana sobre el coeficiente de correlación de una población normal bivariante. *Trabajos Estadíst. Investigación Oper.* **32** 18–31.

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

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **41** 113–128.

BONDAR, J. (1976). Borel cross-sections and maximal invariants. *Ann. Statist.* **4** 866–877.

BOREL, A. (1960). *Seminar on Transformation Groups. Ann. of Math. Stud.* No. 46. Princeton Univ. Press.

BREDON, G. (1972). *Introduction to Compact Transformation Groups*. Academic, New York.

CHANG, T. (1986). Spherical regression. *Ann. Statist.* **14** 907–924.

CHANG, T. and VILLEGAS, C. (1986). On a theorem of Stein relating Bayesian and classical inferences in group models. *Canad. J. Statist.* **14** 289–296.

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. Ser. B* **35** 189–213.

GEISSER, S. (1965). Bayesian estimation in multivariate analysis. *Ann. Math. Statist.* **36** 150–159.

GOLDSTEIN, H. (1950). *Classical Mechanics*. Addison-Wesley, Cambridge, Mass.

JEFFREYS, H. (1939). *Theory of Probability*. Clarendon, Oxford.

JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London Ser. A* **186** 453–461.

KASS, R. E. (1989). The geometry of asymptotic inference (with discussion). *Statist. Sci.* **4** 188–234.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

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

PALAIS, R. (1961). On the existence of slices for actions of noncompact Lie groups. *Ann. Math. (2)* **73** 295–323.

STEENROD, N. (1951). *The Topology of Fibre Bundles*. Princeton Univ. Press.

TOM DIECK, T. (1987). *Transformation Groups*. de Gruyter, Berlin.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF VIRGINIA
MATHEMATICS-ASTRONOMY BUILDING
CHARLOTTESVILLE, VIRGINIA 22903-3199

DEPARTMENT OF MATHEMATICS
AND STATISTICS
SIMON FRASER UNIVERSITY
BURNABY, BRITISH COLUMBIA
CANADA V5A 1S6