Sankhyā : The Indian Journal of Statistics
Special issue in memory of A. Maitra

Invariance of Posterior Distributions under Reparametrization

By Morris L. Eaton and William D. Sudderth
University of Minnesota, USA

Abstract

In 1946, Sir Harold Jeffreys introduced a prior distribution whose density is the square root of the determinant of Fisher information. The motivation for suggesting this prior distribution is that the method results in a posterior that is invariant under reparametrization. For invariant statistical models when there is a transitive group action on the parameter space, it is shown that all relatively invariant priors have this "Jeffreys Invariance" property. However, this invariance may not prevail when using a subtle modification suggested by an alternative argument.

AMS (2000) *subject classification.* 62F15.
Keywords and phrases. Jeffreys priors, invariance under reparametrization, group invariant models and Jeffreys invariance, coordinate system choices and prior distributions.

1. Introduction and Summary

In 1946, Sir Harold Jeffreys described a method of constructing a prior distribution that is invariant under reparametrization. The calculation of this prior distribution (which may be improper) involves first finding the Fisher information matrix, say $I(\theta)$, and then setting the prior equal to the square root of the determinant of $I(\theta)$. The invariance of this method, which we call "Jeffreys invariance" below, is sometimes considered a basic requirement for the so-called non-informative prior distributions. The reader is referred to Robert (1994, pp. 112-120) for a modern discussion of the issues and some interesting examples. A wide ranging and very informed discussion of topics related to prior distribution selection is Kass and Wasserman (1996).

Statistical models invariant under a group of transformations provide a variety of interesting, and at times, confounding examples. Indeed, consider a random sample from a univariate normal distribution with unknown mean and variance. For this example, the Jeffreys prior distribution produces posterior inferences for the mean that differ from the classical Student-t inferences. However, as Jeffreys

(1946, p. 457) points out, a slight modification of the Jeffreys prior results in a posterior that is in agreement with Student-t inferences. But, in some decision theoretic settings, the use of so-called relatively invariant priors, not the Jeffreys prior, is suggested. For example, see Eaton and Sudderth (2004). The main purpose of this paper is to show that all relatively invariant priors possess Jeffreys invariance, but some care needs to be taken in certain invariant situations.

The paper begins with a review of the Jeffreys prior and its invariance properties. The "Jeffreys invariance" described in Section 2 refers to invariance under reparametrization of the Jeffreys prescription for the prior. This is explained in detail in Section 2.

In Section 3 we consider a statistical model that is invariant under a group G. Here the group acts on both the sample space and the parameter space. The use of "invariance" in this context means something rather different than its use in the phrase "Jeffreys invariance." Throughout this paper, the phrase "invariant situations" and "invariant statistical models" are used in the traditional sense of group invariance as in Chapter 3 of Eaton (1989). Under some regularity conditions, it is shown in Section 3 that relatively invariant priors exist and that all such priors have the Jeffreys invariance property. In addition, it is shown that the Jeffreys prior corresponds to a prior induced by left Haar measure. See Kass and Wasserman (1996) and Eaton and Sudderth (2004) for some relevant discussion.

Finally, in Section 4, we consider a classical invariant multivariate normal model. For this model it is well-known that the Jeffreys prior produces a posterior with some disturbing foundational properties, and it is common to suggest a right Haar alternative. However, our calculations show that the corresponding posterior can depend on the initial choice of a coordinate system for the data variables. This observation suggests that care needs to be taken when using invariance to suggest a prior distribution.

2. The Jeffreys Prior

Here we set notation and review the classical argument that verifies the "Jeffreys-Invariance" property for the Jeffreys prior distribution. Consider a density function $f_1(x|\theta)$ with respect to a σ-finite dominating measure $\mu(dx)$ on a sample space $(\mathcal{X}, \mathcal{B})$. The parameter space Θ is assumed to be a nonempty open subset of k-dimensional space $\mathbb{R}^k$. It is assumed further that the function $\log f_1(x|\theta)$ is, for each x, a well-defined differentiable function of θ. So the **score function**

$$s_1(x|\theta) = \begin{pmatrix} \frac{\partial}{\partial\theta_1} \log f_1(x|\theta) \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial}{\partial\theta_k} \log f_1(x|\theta) \end{pmatrix} \tag{2.1}$$

is also well-defined. The covariance matrix

$$I_1(\theta) = \operatorname{cov}(s_1(\cdot|\theta)) \tag{2.2}$$

is assumed to exist. In this case the **Jeffreys prior** is defined to have the density

$$j_1(\theta) = (\det(I_1(\theta)))^{1/2}. \tag{2.3}$$

This is a prior density with respect to Lebesque measure, $d\theta$, on Θ and it may be the case that it is improper. Obviously, $j_1(\theta)$ depends on the particular θ-coordinate system chosen for the statistical model $\{f_1(\cdot|\theta) \,|\, \theta \in \Theta\}$.

Now we consider a reparametrization $\eta = \phi(\theta)$ where $\phi : \Theta \mapsto H$ is a one-one, onto differentiable mapping from Θ to another open subset H of $\mathbb{R}^k$. The inverse ϕ^{-1} is also assumed to be differentiable. So ϕ and ϕ^{-1} have Jacobians $J_\phi(\theta)$ and $J_{\phi^{-1}}(\eta)$, respectively. In particular,

$$J_\phi(\theta) = \{\frac{\partial \phi_i}{\partial \theta_j} \,|\, i, j = 1, \ldots, k\} \tag{2.4}$$

is a $k \times k$ non-singular matrix. Of course, $J_{\phi^{-1}}(\eta)$ is defined similarly and

$$J_\phi(\phi^{-1}(\eta)) = (J_{\phi^{-1}}(\eta))^{-1}. \tag{2.5}$$

This reparametrization gives a new density $f_2(x|\eta) \equiv f_1(x|\phi^{-1}(\eta))$ with $\eta \in H$, and hence a new score function and a new Jeffreys prior. A routine matrix calculation shows that

$$s_2(x|\eta) = (J_{\phi^{-1}}(\eta))' s_1(x|\phi^{-1}(\eta)) \tag{2.6}$$

where $s_2(x|\eta)$ is the score function for the density $f_2(x|\eta)$. Thus the information matrix in the η-coordinate system is

$$I_2(\eta) = \text{cov}(s_2(x|\eta)) = (J_{\phi^{-1}}(\eta))' I_1(\phi^{-1}(\eta)) J_{\phi^{-1}}(\eta). \tag{2.7}$$

Therefore, the Jeffreys prior for the η-coordinate system is

$$\begin{aligned} j_2(\eta) = (\det(I_2(\eta)))^{1/2} &= |\det(J_{\phi^{-1}}(\eta))||\det(I_1(\phi^{-1}(\eta)))|^{1/2} \\ &= |\det(J_{\phi^{-1}}(\eta))| j_1(\phi^{-1}(\eta)). \end{aligned} \tag{2.8}$$

Since $\theta = \phi^{-1}(\eta)$, we see that

$$d\theta = |\det(J_{\phi^{-1}}(\eta))| d\eta. \tag{2.9}$$

So

$$j_2(\eta) d\eta = j_1(\theta) d\theta \tag{2.10}$$

under the reparametrization $\eta = \phi(\theta)$. It is exactly this relationship that is usually expressed by saying "Jeffreys method of constructiong a prior is invariant under reparametrization." We will refer to this relationship as the **invariance of Jeffreys method** or more briefly as **Jeffreys invariance**.

By a "method" we mean a recipe for assigning a prior, say $\pi_1(d\theta)$, for $\{f_1(x|\theta) \,|\, \theta \in \Theta\}$ and also assigning a prior, say $\pi_2(d\eta)$, for $\{f_2(x|\eta) \,|\, \eta \in H\}$ for every smooth reparametrization $\eta = \phi(\theta)$. Improper priors are allowed. Then the method possesses Jeffreys invariance if

$$\int k(\eta)\,\pi_2(d\eta) = \int k(\phi(\theta))\,\pi_1(d\theta) \tag{2.11}$$

for each such reparametrization $\eta = \phi(\theta)$, and for all non-negative k.

The condition (2.11) implies that the two posteriors obtained from π_1 and π_2, respectively, are equivalent in the sense that the posterior for η is transformed into the posterior for θ under the mapping $\eta = \phi(\theta)$. Thus Jeffreys invariance as expressed by (2.11) implies that the posteriors transform properly. Verification of this is routine.

Note that the posterior distribution obtained from a prior π, whether proper or improper, is the same as that obtained from $c\pi$ for any positive constant c. Thus Jeffreys invariance, in particular (2.11), should be interpreted with the proviso that multiplication by positive constants is allowed.

3. The Relative Invariance Method

3.1 Preliminaries: Throughout this section, we will use the notions and results concerning invariance as described in Nachbin (1965) and Eaton (1982, 1989). In particular, for a locally compact Hausdorff space $\mathcal{Y}$, $K(\mathcal{Y})$ will denote the class of all continuous real-valued functions on $\mathcal{Y}$ that have compact support, and $K^+(\mathcal{Y})$ is the class of nonnegative elements in $K(\mathcal{Y})$. An integral on $K(\mathcal{Y})$ is a non-zero linear functional L that is nonnegative on $K^+(\mathcal{Y})$. In all cases of interest in this paper, $\mathcal{Y}$ will be a nice subset of a Euclidean space and every integral L will correspond to a unique non-zero σ-finite Radon measure m (defined on the Baire sets) via the equation

$$L(k) = \int_{\mathcal{Y}} k(y)\,m(dy), \qquad k \in K(\mathcal{Y}). \tag{3.1}$$

Conversely, every non-zero σ-finite Radon measure defines an integral via (3.1). In our applications, the Baire sets and the Borel sets will be the same (see chapter 10 of Halmos, 1950).

Now consider a topological group G that acts topologically on the space $\mathcal{Y}$. Let ν_l denote a left Haar measure on G so that the integral L_1 on $K(G)$ defined by ν_l satisfies

$$L_1(k) = \int_G k(g)\,\nu_l(dg) \tag{3.2}$$

and

$$L_1(hk) = L_1(k) \tag{3.3}$$

for all $h \in G$ and $k \in K(G)$. As usual (see Nachbin, 1965), $hk \in K(G)$ is defined by

$$(hk)(g) \equiv k(h^{-1}g) \tag{3.4}$$

for $g, h \in G$ and $k \in K(G)$. Recall that a **multiplier** on G, say χ, is a continuous homomorphism from G to the multiplicative group $(0, \infty)$. The special multiplier Δ defined by the equation

$$\int_G k(gh^{-1})\,\nu_l(dg) \;\equiv\; \Delta(h) \int_G k(g)\,\nu_l(dg) \tag{3.5}$$

is called the (right hand) modulus of G (see page 77 of Nachbin, 1965). It is well-known that

$$\nu_r(dg) \;=\; \Delta^{-1}(g)\nu_l(dg) \;=\; \Delta(g^{-1})\nu_l(dg) \tag{3.6}$$

is a right Haar measure on G.

Recall next that an integral L_2 on $K(\mathcal{Y})$ is **relatively invariant with multiplier** χ if

$$L_2(hk) \;=\; \chi(h)L_2(k) \tag{3.7}$$

for all $h \in G$ and $k \in K(\mathcal{Y})$.

The group G acts **transitively** on $\mathcal{Y}$ if, for each pair $y_1, y_2 \in \mathcal{Y}$, there is a $g \in G$ so that $gy_1 = y_2$. When G is transitive in this sense, it is clear that $\mathcal{Y} = \{gy_0 \,|\, g \in G\}$ for each fixed $y_0 \in \mathcal{Y}$. Given $y_0 \in \mathcal{Y}$, the subgroup H of G defined by

$$H \;=\; \{g \,|\, gy_0 = y_0\} \tag{3.8}$$

is called the **isotropy subgroup** of y_0.

The following is a basic result about relatively invariant integrals on $K(\mathcal{Y})$.

THEOREM 3.1. *Suppose G is transitive on $\mathcal{Y}$ and that the isotropy subgroup H in (3.8) is compact. Then, given a multiplier χ on G, the integral*

$$L_2(k) \;=\; \int_G k(gy_0)\chi(g)\,\nu_l(dg) \tag{3.9}$$

is relatively invariant on $K(\mathcal{Y})$ with multiplier χ, and is unique up to a positive constant.

Proof. That L_2 is relatively invariant with multiplier χ is easy to check. Because H is assumed to be compact, Theorem 1 on page 138 of Nachbin (1965) implies that L_2 is unique up to a positive constant. □

EXAMPLE 3.2. *Let $G = Gl_p$, the group of $p \times p$ real non-singular matrices and let $\mathcal{Y}$ be S_p^+, the set of all real $p \times p$ positive definite symmetric matrices. The action of G on $\mathcal{Y}$ is*

$$y \to gyg' \tag{3.10}$$

where g' is the transpose of $g \in G$. With $|g|$ denoting the absolute value of the determinant of g, it is well-known that

$$\nu_l(dg) \;=\; \frac{dg}{|g|^p} \tag{3.11}$$

is a left (and right) Haar measure on G. For $\alpha \in \mathbb{R}^1$, set

$$\chi_\alpha(g) = |g|^\alpha, \qquad g \in G. \tag{3.12}$$

Each χ_α is a multiplier on G and, conversely, every multiplier is a χ_α for some $\alpha \in \mathbb{R}^1$ (see problem 7, page 230 in Eaton, 1983). Let $y_0 = I_p$, the $p \times p$ identity matrix, so that

$$H \;=\; \{g \,|\, g(y_0) = y_0\} \tag{3.13}$$

is just the group of real $p \times p$ orthogonal matrices, and hence compact. Formula (3.9) gives a relatively invariant measure with multiplier $\chi = \chi_\alpha$, but it is more convenient to represent this integral in terms of a Borel measure on $\mathcal{Y}$. Consider $m_\alpha(dy)$ defined by

$$m_\alpha(dy) \;=\; |y|^{\alpha/2} \frac{dy}{|y|^{(p+1)/2}}. \tag{3.14}$$

It is not hard to show that the integral defined by (3.14) is relatively invariant with multiplier χ_α. General theory shows that m_α is unique up to a positive constant. Of course, m_0 is invariant since $\chi_0 \equiv 1$.

EXAMPLE 3.3. *For this example let $G = G_T^+$, the group of all $p \times p$ lower triangular matrices with positive diagonal elements. Also, let $\mathcal{Y}$ be as in the previous example with G_T^+ acting on $\mathcal{Y}$ as Gl_p acts on $\mathcal{Y}$. With $y_0 = I_p$, the isotropy subgroup is just $\{I_p\}$ since G_T^+ is* ***exactly transitive*** *- that is, given $y \in S_p^+$, there is a unique element $g \in G_T^+$ such that $y = gg'$ (for a proof, see page 162 in Eaton, 1983). This uniqueness implies that when $\tau(y)$ is the unique element satisfying $y = \tau(y)(\tau(y))'$, then*

$$\tau(hyh') \;=\; h\tau(y), \qquad for\, h \in G_T^+. \tag{3.15}$$

It is well-known that a left Haar measure on G_T^+ is

$$\nu_l(dg) \;=\; \frac{dg}{\prod_{i=1}^p g_{ii}^i} \tag{3.16}$$

and a right Haar measure on G_T^+ is

$$\nu_r(dg) \;=\; \frac{dg}{\prod_{i=1}^p g_{ii}^{p-i+1}}. \tag{3.17}$$

Also, the (right hand) modulus of G_T^+ is

$$\Delta_r(g) \;=\; \prod_{i=1}^p g_{ii}^{p-2i+1}. \tag{3.18}$$

Further, given any vector $c \in \mathbb{R}^p$, let

$$\chi_c(g) \;=\; \prod_{i=1}^p g_{ii}^{c_i}. \tag{3.19}$$

Then each χ_c is a multiplier on G_T^+, and conversely, every multiplier is a χ_c for some c (see pp. 16-18 in Eaton, !989).

Now, given a vector $c \in \mathbb{R}^p$, consider the measure m_c on $\mathcal{Y}$ given by

$$m_c(dy) \;=\; \chi_c(\tau(y))\frac{dy}{|y|^{(p+1)/2}}, \tag{3.20}$$

where χ_c is from (3.19) and $\tau(y)$ is the unique element in G_T^+ satisfying $y = \tau(y)(\tau(y))'$. Using (3.15), a routine calculation shows that the measure m_c is relatively invariant with multiplier χ_c. However, it should be noted that $\chi_c(\tau(y))$ in (3.20) is a rather complicated function of y. In statistical applications, it is often more convenient to use the space G_T^+ (rather than $\mathcal{Y}$) as a parameter space. When this is done, then the relatively integrals on $K(G_T^+)$ are just

$$\chi_c(g)\nu_l(dg), \qquad c \in \mathbb{R}^p, \tag{3.21}$$

which are relatively simple functions of $g \in K(G_T^+)$. This completes the example.

3.2 Relatively Invariant Priors: In this subsection we study an invariant statistical model where the group in question acts transitively on the parameter space. With some additional regularity conditions, we show that the Jeffreys prior described in Section 2 corresponds to using the left Haar invariant prior (as defined below), and that all relatively invariant priors possess Jeffreys invariance. One justification, based on decision theory, for using relatively invariant priors is given in Eaton and Sudderth (2004). Other arguments for their use can be found in Robert (1994) and Kass and Wasserman (1996).

As in Section 2, we consider a density function $f_1(x|\theta)$ with respect to a σ-finite measure μ on a sample space $(\mathcal{X}, \Theta)$. The parameter space Θ is again assumed to be a nonempty open subset of $\mathbb{R}^k$. Further the regularity conditions necessary for the calculation of the Fisher information matrix (2.2) and the Jeffreys prior (2.3) are assumed to hold.

To introduce invariance, we consider a group G that acts on both $\mathcal{X}$ and Θ, and we assume that the dominating measure μ is relatively invariant with multiplier χ_0. That is,

$$\int k(g^{-1}x)\,\mu(dx) \;=\; \chi_0(g)\int k(x)\,\mu(dx) \tag{3.22}$$

for all integrable k and all $g \in G$. In addition, it is assumed that the density f_1 satisfies the usual invariance condition

$$f_1(x|\theta) \;=\; f_1(gx|g\theta)\chi_0(g). \tag{3.23}$$

Thus the parametric model defined by the density is invariant in the usual sense (see Chapter 3 in Eaton, 1989).

Finally, we assume that the group G is transitive on the parameter space Θ, and that the isotropy subgroup

$$H_0 \;=\; \{g \,|\, g\theta_0 = \theta_0\} \tag{3.24}$$

is compact. Under this compactness assumption, given a multiplier χ on the group G, there is a unique (up to a positive constant) Radon measure π_χ on Θ, that is relatively invariant with multiplier χ. In other words,

$$\int_\Theta k(g^{-1}\theta)\,\pi_\chi(d\theta) \;=\; \chi(g)\int_\Theta k(\theta)\,\pi_\chi(d\theta) \tag{3.25}$$

for all $k \in K(\Theta)$. As discussed above, π_χ is also the unique (up to a positive constant) Radon measure on Θ that satisfies

$$L(k) \;=\; \int_G k(g\theta_0)\chi(g)\,\nu_l(dg) \;=\; \int_\Theta k(\theta)\pi_\chi(d\theta) \tag{3.26}$$

for all $k \in K(\Theta)$.

We now want to argue that the above method of constructing a relatively invariant prior distribution has the property of Jeffreys invariance as expressed in (2.11). To this end, consider a reparametrization $\eta = \phi(\theta)$ where ϕ is a homeomorphism from Θ to H. Recall that each $g \in G$ determines a "natural homeomorphism" from Θ to Θ via left multiplication (see the discussion on page 126 of Nachbin, 1965). Thus

$$G_\phi \;=\; \{\phi g \phi^{-1} \,|\, g \in G\} \tag{3.27}$$

is a homeomorphic image of G under the mapping

$$i : g \mapsto \phi g \phi^{-1} \tag{3.28}$$

and G_ϕ acts in an obvious way on H, namely

$$(\phi g \phi^{-1})(\eta) \;=\; \phi(g(\phi^{-1}(\eta))). \tag{3.29}$$

The isomorphism i sets up a natural one-to-one correspondence between the multipliers χ on G and the multipliers χ^* on G_ϕ as follows;

$$\chi \;=\; \chi^* \circ i. \tag{3.30}$$

It is obvious that

$$\nu_l^* \equiv \nu_l \circ i^{-1} \tag{3.31}$$

also defines a left Haar measure on $K(G_\phi)$. Taking $\eta_0 = \phi(\theta_0)$, the following integral

$$L^*(k^*) \;=\; \int_{G_\phi} k^*(g^*\eta_0)\chi^*(g^*)\,\nu_l^*(dg^*) \tag{3.32}$$

defines what we call a prior distribution on H that is relatively invariant with multiplier χ^*. That is, there is a prior distribution π_{χ^*} on H so that

$$L^*(k^*) \;=\; \int_H k^*(\eta)\,\pi_{\chi^*}(d\eta). \tag{3.33}$$

To show that the relatively invariant priors thus constructed enjoy Jeffreys invariance, we need to verify (2.11) - that is, we need to show that

$$\int_H k^*(\eta)\,\pi_{\chi^*}(d\eta) \;=\; \int_\Theta k^*(\phi(\theta))\,\pi_\chi(d\theta). \tag{3.34}$$

In other words, we need to show that

$$L(k^* \circ \phi) \;=\; L^*(k^*) \tag{3.35}$$

with L^* given by (3.32) and L given by (3.26). However, (3.35) is a direct consequence of (3.30) and (3.31). In summary, we have

THEOREM 3.4. *Given a multiplier χ on G, let π_χ denote the relatively invariant prior defined implicitly by (3.26). This method of prior construction possesses Jeffreys invariance.*

REMARK 3.5. *Given a density function satisfying (3.23) and a relatively invariant prior m_χ on Θ, the usual definition of the posterior density (relative to m_χ) is*

$$q(\theta|x) \;=\; \frac{f(x|\theta)}{r(x)} \tag{3.36}$$

where

$$r(x) \;=\; \int f(x|\theta)\,m_\chi(d\theta) \tag{3.37}$$

is the "marginal" for x. However, there are many examples where $r(x) = \infty$ so (3.36) is not well-defined. Ordinarily, one assumes that $0 < r(x) < \infty$ in order that discussions regarding the posterior make sense. See Eaton and Sudderth (1993, Theorem 3.2) for an example where χ has to be restricted.

3.3 Invariance and the Jeffreys Prior: In this section we again consider an invariant probability density $f_1(x|\theta)$ that satisfies the invariance condition (3.23) under the action of the group G on $\mathcal{X}$ and Θ. As in Section 3.2, G is assumed to act transitively on Θ and the isotropy group (3.24) is assumed to be compact. Thus, for each multiplier χ on G, there will exist a unique (up to a positive constant) relatively invariant prior on Θ with multiplier χ.

Under the additional assumptions that Θ is an open subset of $\mathbb{R}^k$ and $f(x|\theta)$ is differentiable (as in Section 2), we now want to argue that the Jeffreys prior always corresponds to left Haar measure (that is, the relatively invariant prior with multiplier χ identically 1). To this end, we make the further assumption that each g acting on Θ to Θ is a differentiable function whose Jacobian we denote by $J_g(\theta)$. The following result recalls how Fisher information transforms in invariant situations.

LEMMA 3.6. *For each $g \in G$, Fisher information $I(\theta)$ satisfies*

$$I(\theta) \;=\; {J_g}'(\theta) I(g\theta) J_g(\theta). \tag{3.38}$$

Proof. Recall that

$$I(\theta) = \operatorname{cov}(s(x|\theta))$$

where $s(x|\theta)$ is the score function as defined by (2.1). An application of the chain rule gives

$$s(x|\theta) = J_g{}'(\theta)s(gx|g\theta). \tag{3.39}$$

Equality (3.38) now follows from (3.23) and the invariance of the parametric family determined by the density f. □

Now recall that the Jeffreys prior is

$$j(\theta)d\theta = (\det I(\theta))^{1/2}d\theta.$$

Consider the integral

$$L_j(k) = \int k(\theta)j(\theta)\,d\theta$$

on $K^+(\Theta)$. Then, for $g \in G$,

$$\begin{aligned} L_j(gk) = \int k(g^{-1}\theta)j(\theta)\,d\theta &= \int k(\eta)j(g\eta)J_g(\eta)\,d\eta \\ &= \int k(\eta)|I(g\eta)|^{1/2}J_g(\eta)\,d\eta \\ &= \int k(\eta)|I(\eta)|^{1/2}\,d\eta = L_j(k), \end{aligned}$$

where the next to last equality follows from (3.38). Thus the integral L_j is invariant and is unique (up to a positive constant). This yields

THEOREM 3.7. *In invariant situations, the Jeffreys prior defines an invariant integral on $K^+(\Theta)$ and corresponds uniquely (up to a positive constant) to using the left Haar measure to induce a relatively invariant prior with multiplier 1.*

It is well-known that in invariant situations, decision theoretic criteria typically suggest that the right Haar measure is the "correct" relatively invariant prior distribution. This is based on the result that says "when the group is transitive on Θ in fully invariant decision problems, use of the right Haar prior will produce a best invariant decision rule as a formal Bayes rule" - see Eaton and Sudderth (2004) for an extended discussion of this result. One consequence of this is that the Jeffreys prior prescription and that of decision theory tend to be at odds when right and left Haar measure are different. There does not seem to be an obvious way to reconcile this discrepancy. This difficulty has been noted in a variety of particular problems - for example, see Jeffreys (1946) and Kass and Wasserman (1996).

4. A Cautionary Example

In this section we consider an invariant example where the Jeffreys prior leads to rather undesirable inferences and the right Haar prior also has a somewhat cautionary aspect. Some of our discussion is based on Eaton and Sudderth (1993, 1995) and Eaton and Freedman (2004). For an applied problem where the issues below arise, see Eaton et al (2006).

Let $X_1, \ldots, X_n$ be iid multivariate normal in $\mathbb{R}^p$ with mean vector 0 and unknown $p \times p$ positive definite covariance matrix Σ. The normal distribution of the X_i is written $N_p(0, \Sigma)$. Note that the mean vector has been taken to be zero to simplify the discussion. For the case with a non-zero mean, see Eaton and Sudderth (1993, 1995).

We assume $n \geq p$ so that the sufficient statistic statistic

$$S = \sum_{i=1}^{n} X_i {X_i}' \tag{4.1}$$

is non-singular with probability 1 and has a Wishart distribution $W(\Sigma, p, n)$. The density function of S with respect to the measure

$$\mu(dS) = \frac{dS}{|S|^{\frac{p+1}{2}}} \tag{4.2}$$

is

$$f(\Sigma^{-1}S) = c\,|\Sigma^{-1}S|^{\frac{n}{2}} \exp[-\frac{1}{2}\mathrm{tr}\Sigma^{-1}S] \tag{4.3}$$

where c is the Wishart constant (see page 175 of Eaton, 1983).

The Wishart model above is invariant under Gl_p, the group of all $p \times p$ non-singular matrices. The group action is

$$S \to ASA', \qquad \Sigma \to A\Sigma A' \tag{4.4}$$

for $A \in Gl_p$. A right and left Haar measure on Gl_p is

$$\nu_l(dA) = \frac{dA}{|A|^p}. \tag{4.5}$$

As is easily verified, ν_l induces the invariant prior measure

$$\pi_j(d\Sigma) = \frac{\Sigma}{|\Sigma|^{\frac{p+1}{2}}} \tag{4.6}$$

which is the Jeffreys prior for this problem (see Box and Tiao (1973) for an early derivation of (4.6) as the Jeffreys prior). This gives the posterior density with respect to π_j as

$$q_j(\Sigma|S) = c\,|\Sigma^{-1}S|^{\frac{n}{2}} \exp[-\frac{1}{2}\mathrm{tr}\Sigma^{-1}S] \tag{4.7}$$

which we recognize as a $W(S^{-1}, k, n)$ distribution for Σ^{-1}.

That there are problems with the above invariance argument was hinted at quite early by Stein in a 1956 Stanford technical report and was reported in the James and Stein (1961) paper. The fact that the posterior distribution defined by (4.7) has foundational problems was made explicit in Eaton and Sudderth (1993, 1995) and given a simplified proof in Eaton and Freedman (2004). In essence, it is shown in these papers that the posterior of (4.7) is incoherent in de Finetti's sense and is strongly inconsistent in Stone's sense. In addition, decision theoretic deficiencies associated with (4.7) suggest quite forcefully that the Jeffreys prior is inappropriate for the Wishart model (4.3).

An alternative approach is suggested by Stein's early work on the covariance matrix estimation problem (see James and Stein, 1961). Let G_T^+ be the group of all $p \times p$ lower triangular real matrices that have positive diagonal (elements as in Example 3.3). The Wishart model for S is obviously invariant under the action of G_T^+ given by restricting the action in (4.4) to A's in G_T^+. Of course, the Jeffreys prior is still (4.6), but now there are many more relatively invariant priors on Σ. Note that Theorem 3.7 implies that the Jeffreys prior is also given by the left Haar measure on G_T^+ as specified in (3.16).

In invariant situations when the group is transitive on the parameter space, there are a variety of arguments that support using the right Haar prior distribution, within the class of all relatively invariant priors. For an overview of these arguments, see Eaton and Sudderth (2001, 2004). Because we find the case for using the right Haar prior rather compelling, attention is restricted to this prior in what follows.

We now proceed with the derivation of the posterior distribution for Σ using the right Haar prior distribution from G_T^+. Let S_p^+ denote the space of all $p \times p$ positive definite matrices. Note that the measure μ in (4.2) corresponds to the measure $2^p \nu_l(dt)$ on G_T^+ (with the left Haar measure ν_l given by (3.16)) in the sense that for all non-negative functions ψ,

$$\int_{S_p^+} \psi(s)\, \mu(ds) \;=\; \int_{G_T^+} \psi(tt')\, 2^p \nu_l(dt) \tag{4.8}$$

(see Eaton (1983, Proposition 5.18)). Recall that the mapping τ from S_p^+ to G_T^+ defined uniquely by $s = \tau(s)\tau(s)'$ satisfies (3.15). Now a right Haar measure on G_T^+ is

$$\nu_r(dt) \;=\; \Delta(t^{-1})\nu_l(dt) \tag{4.9}$$

with the right hand modulus Δ given by (3.18). By definition, a right Haar prior on the parameter space $\Theta = S_p^+$ is the unique (up to a positive constant) measure m_r that satisfies, for all non-negative ψ,

$$\int_{S_p^+} \psi(g^{-1}\theta)\, m_r(d\theta) \;=\; \frac{1}{\Delta(g)} \int_{S_p^+} \psi(\theta)\, m_r(d\theta) \tag{4.10}$$

since right Haar measure is given by (4.9) (see display (3.25)). In our specific case, $\theta = \Sigma$ is the $p \times p$ covariance matrix and we take $\theta_0 = I_p$. Then the right Haar prior m_r satisfies (see (3.26))

$$\int_{G_T^+} \psi(g\theta_0)\frac{1}{\Delta(g)}\,\nu_l(dg) \;=\; \int_{G_T^+} \psi(gg')\frac{1}{\Delta(g)}\,\nu_l(dg) \;=\; \int_{S_p^+} \psi(\Sigma)\,m_r(d\Sigma). \qquad (4.11)$$

Since $\tau(gg') = g$, we can use (4.2) and (4.8) to write

$$\int_{G_T^+} \psi(gg')\frac{1}{\Delta(g)}\,\nu_l(dg) \;=\; \int_{G_T^+} \psi(gg')\frac{1}{\Delta(\tau(gg'))}\,\nu_l(dg) \;=\; \int_{S_p^+} \psi(\Sigma)\frac{1}{\Delta(\tau(\Sigma))}2^{-p}\frac{d\Sigma}{|\Sigma|^{\frac{p+1}{2}}}.$$

Combining this with (4.11) gives

Proposition 4.1. *A right Haar prior on Θ is*

$$m_r(d\Sigma) \;=\; \frac{1}{\Delta(\tau(\Sigma))}2^{-p}\frac{d\Sigma}{|\Sigma|^{\frac{p+1}{2}}}. \qquad (4.12)$$

Using (4.12), it is now an easy matter to obtain the posterior of Σ given S.

Proposition 4.2. *The posterior distribution of Σ given S when the prior is (4.12) is*

$$Q(d\Sigma|S) \;=\; \frac{\Delta(\tau(S))}{\Delta(\tau(\Sigma))}f(\Sigma^{-1}S)\frac{d\Sigma}{|\Sigma|^{\frac{p+1}{2}}}. \qquad (4.13)$$

Proof. By definition, the posterior is

$$Q(d\Sigma|S) \;=\; \frac{f(\Sigma^{-1}S)\,m_r(d\Sigma)}{v(S)} \qquad (4.14)$$

where

$$v(S) \;=\; \int_{S_p^+} f(\Sigma^{-1}S)\,m_r(d\Sigma).$$

A relatively routine calculation shows that

$$v(S) \;=\; \frac{2^{-p}}{\Delta(\tau(S))}. \qquad (4.15)$$

The substitution of (4.15) and (4.12) into (4.14) immediately yields (4.13). □

Now suppose we transform the original data to $X_i^* = B_0X_i$, $i = 1,\dots,n$ where B_0 is a known full rank $p\times p$ matrix. Then, with $\Sigma^* = B_0\Sigma B_0{}'$, $X_1^*,\dots,X_n^*$ are iid $N_p(0,\Sigma^*)$. So

$$S^* \;=\; \sum_{i=1}^n X_i^*X_i^{*\prime} \;=\; B_0SB_0{}' \qquad (4.16)$$

is $W(\Sigma^*,p,n)$. The invariance argument using G_T^+ suggests we should use the right Haar prior for Σ^*, namely $m_r(d\Sigma^*)$, and compute a posterior for Σ^*. Simply repeating the previous argument gives

Proposition 4.3. *The posterior distribution of* Σ^* *given* S^* *for the prior* $m_r(d\Sigma^*)$ *is*

$$Q^*(d\Sigma^*|S^*) \;=\; \frac{\Delta(\tau(S^*))}{\Delta(\tau(\Sigma^*))} f((\Sigma^*)^{-1}S^*)\frac{d\Sigma^*}{|\Sigma^*|^{\frac{p+1}{2}}}. \tag{4.17}$$

Next we use (4.17) to induce a "posterior distribution" on $\Sigma = B_0^{-1}\Sigma^*(B_0^{-1})'$ given $S = B_0^{-1}S^*(B_0^{-1})'$. Because of the explicit nature of (4.17), this is easy to do.

Proposition 4.4. *Using (4.17) and the prior* $m_r(d\Sigma^*)$*, the induced distribution on* Σ*, given* $S^* = B_0 S {B_0}'$*, is*

$$\tilde{Q}(d\Sigma|S) \;=\; \frac{\Delta(\tau(B_0 S {B_0}'))}{\Delta(\tau(B_0 \Sigma {B_0}'))} f(\Sigma^{-1}S)\frac{d\Sigma}{|\Sigma|^{\frac{p+1}{2}}}. \tag{4.18}$$

Proof. First observe that

$$f((\Sigma^*)^{-1}S^*) \;=\; f(\Sigma^{-1}S)$$

and the measure $d\Sigma/|\Sigma|^{\frac{p+1}{2}}$ is invariant under all transformations $\Sigma \to A\Sigma A'$, for $A \in Gl_p$. Equation (4.18) is now clear. □

The obvious question is now whether the two posterior distributions (4.13) and (4.18) are the same. Because f, Δ, and τ are continuous, they are the same if and only if

$$\frac{\Delta(\tau(S))}{\Delta(\tau(\Sigma))} \;=\; \frac{\Delta(\tau(B_0 S {B_0}'))}{\Delta(\tau(B_0 \Sigma {B_0}'))} \tag{4.19}$$

for all S and Σ in S_p^+. If $B_0 \in G_T^+$, then (4.19) holds since $\tau(B_0 s {B_0}') = B_0\tau(s)$ for all $s \in S_p^+$ and Δ is a multiplier. But there are B_0's of interest which are not in G_T^+ for which (4.19) does not hold.

Example 4.5. *Let* $p = 2$ *and take*

$$B_0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

So B_0 *simply interchanges the two coordinates of each data vector* $X_1, \ldots, X_n$*. For* $S = I_2$*, it is an easy calculation to show that (4.19) does not hold for almost all* Σ*'s. Thus the posteriors (4.13) and (4.19) are different. In other words, the posterior using the right Haar prior depends on the order of the coordinates of the data. Therefore the practitioner needs to take this into account if the right Haar measure is used as "the non-informative prior" for this problem.*

The general message is that there are B_0's of interest which affect the posterior inference when one uses what many feel is the appropriate "non-informative prior." Please see Eaton et al (2006) for an applied context where this issue is relevant.

References

Box, G. E. P and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

Eaton, M. L. (1983, 2007). *Multivariate Statistics: A Vector Space Approach.* Originally published by Wiley, New York. Reprinted by Institute of Mathematical Statistics as vol. 53 in LMNS, Beachwood, Ohio.

Eaton, M. L. (1989). *Group Invariant Applications in Statistics.* Regional Conference Series in Probability and Statistics **1**. Institute of Mathematical Statistics, Beachwood, Ohio.

Eaton, M. L. and Freedman, D. A. (2004). Dutch book against some "objective" priors. *Bernoulli*, **10**, 861-872.

Eaton, M. L., Muirhead, R. J. and Pickering, E. H. (2006). Assessing a vector of clinical observations. *Journal of Statistical Planning and Inference*, **136**, 3383-3414.

Eaton, M. L. and Sudderth, W. D. (1993). Prediction in a multivariate normal setting: coherence and incoherence. *Sankhyā, Series A*, **55**, 481-493.

Eaton, M. L. and Sudderth, W. D. (1995). The formal posterior of a standard flat prior in MANOVA is incoherent. *J. Italian Statist. Soc.*, **4**. 251-270.

Eaton, M. L. and Sudderth, W. D. (2001). Best invariant predictive inferences. In Algebraic Methods in Statistics and Probability, editors, M. Viana and D. Richards. *Contemporary Mathematics* **287**, 49-62, American Mathematical Society, Providence, Rhode Island.

Eaton, M. L. and Sudderth, W. D. (2004). Properties of right Haar predictive inference. *Sankhyā*, **66, Part 3**, 487-512.

Halmos, P. R. (1950). *Measure Theory.* Van Nostrand, New York.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Statist. and Prob.* **1**, 361-380. University of California Press, Berkeley, California.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Royal Soc. of London, Series A*, **186**, 453-461.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assn.*, **91**, 1343-1370.

Nachbin, L. (1965). *The Haar Integral.* Van Nostrand, New York.

Robert, C. P. (1994). *The Bayesian Choice.* Springer-Verlag, New York.

Morris L. Eaton and William D. Sudderth
School of Statistics
University of Minnesota
Minneapolis, Minnesota 55455.
E-mail: eaton@stat.umn.edu, bill@stat.umn.edu