

# Consistency and strong inconsistency of group-invariant predictive inferences

MORRIS L. EATON\* and WILLIAM D. SUDDERTH\*\*

*School of Statistics, University of Minnesota, 206 Church Street SE, Minneapolis MN 55455, USA. E-mail: \*eaton@stat.umn.edu; \*\*bill@stat.umn.edu*

Consider a statistical model which is invariant under a group of transformations that acts transitively on the parameter space. In this situation, the problem of constructing invariant predictive distributions is considered. It is shown, under certain assumptions, that Fisherian pivoting and the use of right Haar measure as an improper prior distribution both yield the same invariant predictive distribution. Furthermore, it is shown that any other invariant predictive distribution is strongly inconsistent in the sense of Stone.

*Keywords:* Fisherian pivoting; improper prior distributions; invariant predictive distribution; proper action; right Haar measure, strong inconsistency

## 1. Introduction

Let  $P(dx, dz|\theta)$ ,  $\theta \in \Theta$ , be the joint distribution for the data variables  $X, Z$  given the unknown parameter  $\theta$ . The problem we consider in this paper is the prediction of  $Z$  based on observed data  $X = x$ . A *predictive inference* for  $Z$ , say  $Q(dz|x)$ , is a distribution for  $Z$ , given the data  $X = x$ . In other words,  $Q(dz|x)$  is the inferrer's guess at the distribution of  $Z$ , given  $X = x$ . Standard methods of constructing predictive inferences include: (i) the Bayes or formal Bayes method where  $Q(dz|x)$  is a posterior distribution of  $Z$  given  $X = x$ ; (ii) methods based on pivoting arguments; (iii) methods based on substituting estimators of  $\theta$  into the conditional distribution of  $Z$  given  $X = x$  and  $\theta$ . A discussion of the Bayes/formal Bayes method can be found in Geisser (1993), while the recent paper of Barndorff-Nielsen and Cox (1996) contains elements of both pivoting arguments and estimator substitution methods.

A predictive inference  $Q(dz|x)$  is called *strongly inconsistent* if, for some bounded measurable function  $f(x, z)$ ,

$$\inf_x \int f(x, z)Q(dz|x) > \sup_{\theta} \int f(x, z)P(dx, dz|\theta). \quad (1)$$

As will be explained below, a  $Q$  that is strongly inconsistent leads to uniformly inadmissible predictions. Such a  $Q$  is also incoherent in the sense of Lane and Sudderth (1984). As discussed in Section 2, it is our opinion that strongly inconsistent inferences should be excluded from serious consideration when one is trying to solve an inference problem. So we are interested in finding conditions for a predictive inference to be *consistent*, by which we

mean that it is not strongly inconsistent. (Interesting examples of strong inconsistency in an inferential setting were given by Stone (1976), who introduced the terminology.)

In this paper, we assume that the model  $P(dx, dz|\theta)$  is invariant under the action of a group  $G$ , and we study predictive inferences  $Q$  that are also invariant under  $G$ . Throughout,  $G$  is assumed to act transitively on the parameter space  $\Theta$ . (Definitions are in Section 3.) In particular, we concentrate on the two most popular techniques of invariant prediction, namely, Fisherian pivoting and the use of a group-invariant prior distribution, typically an improper prior.

Here is a sketch of the paper and the major results. The next section is a discussion of strong inconsistency. Section 3 presents some necessary definitions and a few examples. In Section 4 we describe a general form of the Fisherian pivotal argument and use it to define a particular invariant inference  $Q_0(dz|x)$ . Section 5 treats general invariant inferences, and in Section 6 we show that any such invariant inference is strongly inconsistent if it is ‘essentially different’ from the pivotal inference  $Q_0$  (Theorem 6.1). In Section 7 we introduce right Haar measure as a (possibly improper) prior and give conditions under which the formal Bayes inference  $Q^*(dz|x)$  is well defined. Under the hypothesis that the group  $G$  is amenable,  $Q^*$  is shown to be consistent (Theorem 8.1). Finally, we show in Section 9 that  $Q_0$  and  $Q^*$  are the same if  $G$  acts ‘properly’ on  $\mathcal{X} \times \mathcal{Z}$  (Theorem 9.1). The final two sections are devoted to examples and discussion.

In essence, our results show that for many standard invariant prediction problems, when the group acts transitively on the parameter space, there is only one invariant predictive distribution, say  $Q_0$ , which can be consistent (assuming the model and the group are given). When possible,  $Q_0$  is obtained via a Fisherian pivoting argument, but in other situations  $Q_0$  is obtained by a formal application of Bayes’s theorem, using a right Haar measure as an improper prior distribution. When the group is amenable,  $Q_0$  is consistent, but may be strongly inconsistent in standard examples when the group is not amenable.

Our formulation of the prediction problem is similar in spirit to that in Aitchison and Dunsmore (1975) and Geisser (1993). This point of view is also adopted in the more theoretical works of Eaton (1982; 1992), Lane and Sudderth (1984), and Eaton and Sudderth (1993; 1995). A discussion of many aspects of group invariance in statistical problems can be found in Eaton (1989) and Wijsman (1990).

The assumption made in Section 4, which we call FP (Fisherian pivoting), implies a variety of things including assumptions 1 and 2 in the paper of Kiefer (1975) where a version of the invariant minimax theorem is established. The FP assumption also appears in Fraser (1968) under the name ‘unitary’.

In a decision-theoretic setting, the use of the right Haar measure to generate best invariant decision rules is a well-known technique – for example, see Stein (1965), Berger (1985, p. 413) or Eaton (1989, Chapter 6). It is this result, together with the invariant minimax theorem for amenable groups, which suggested some of our results in Sections 7, 8 and 9. A survey of the role of amenability in certain statistical problems can be found in Bondar and Milnes (1981).

The use of the Bourbaki notion of proper action in invariant statistical problems first appeared in Andersson (1982) (see also Andersson *et al.* 1983). An application in statistical decision theory appears in Eaton (1989, Chapter 6). The monograph of Wijsman (1990) contains a detailed discussion of proper action and gives some statistical applications.

## 2. Remarks on strong inconsistency

Suppose that the predictive inference  $Q$  is strongly inconsistent so that (1) holds for some bounded, measurable  $f$ . We can rewrite (1) in the form

$$\inf_x \int f(x, z)Q(dz|x) = \varepsilon + \sup_{\theta} E_{\theta}f(X, Z) \tag{2}$$

where  $\varepsilon > 0$ .

Then, for every proper prior distribution  $\pi$  defined on the Borel sets of  $\Theta$ ,

$$\iint f(x, z) Q(dz|x) M_{\pi}(dx) \geq \varepsilon + E f(X, Z),$$

where  $E$  denotes expectation under  $\pi$  and the model, and  $M_{\pi}$  is the marginal distribution of  $X$ . Thus strong inconsistency implies that there is no world (prior opinion on  $\theta$ ) for which the expectation of  $f$  can be consistent with its expectation computed under  $Q(\cdot|x)$  and the marginal for  $x$ . It follows also that  $Q$  is incoherent in the sense of Lane and Sudderth (1984). This means, roughly, that if a bookie uses  $Q$  to post odds on  $z$  given  $x$ , then there will be a finite betting scheme for a gambler that results in a uniformly positive expected pay-off.

Suppose now that we use the strongly inconsistent  $Q(\cdot|x)$  to predict  $f(x, z)$ . Under quadratic loss, our predictor is

$$\hat{f}(x) = \int f(x, z)Q(dz|x).$$

However, when (2) holds, it is easy to show that the predictor  $f^*(x) = \hat{f}(x) - \varepsilon$  satisfies

$$E_{\theta}(f^*(X) - f(X, Z))^2 \leq -\varepsilon^2 + E_{\theta}(\hat{f}(X) - f(X, Z))^2$$

for all  $\theta$ . Thus the mean squared error of  $f^*$  is at least  $\varepsilon^2$  smaller than that of  $\hat{f}$ , for all  $\theta$ .

## 3. Some definitions and examples

The range spaces  $\mathcal{X}$ ,  $\mathcal{Z}$  for the variables  $X$ ,  $Z$  are assumed to be Polish and locally compact as are the parameter space  $\Theta$  and the group  $G$ . ‘Measurable’ always means ‘Borel measurable’ and we denote the Borel  $\sigma$ -field of any space, for example  $\mathcal{X}$ , by  $\mathcal{B}(\mathcal{X})$ . We assume that the group  $G$  acts continuously from the left on  $\mathcal{X}$ ,  $\mathcal{Z}$  and  $\Theta$  and that the model  $P(dx, dz|\theta)$  is invariant in the sense that

$$\iint f(g^{-1}x, g^{-1}z)P(dx, dz|g\theta) = \iint f(x, z)P(dx, dz|\theta) \tag{3}$$

for all  $\theta \in \Theta$ ,  $g \in G$  and bounded, measurable  $f$ .

There is a multitude of examples available. Here are three.

**Example 3.1.** Let  $X_1, X_2, \dots, X_n, Z$  be independent  $N_1(\theta, 1)$ , and take  $X = (X_1, X_2, \dots, X_n)$ . The model is invariant under the translation group of reals with action on  $\mathcal{X} = \mathbb{R}^n$ , for

example,  $g(x_1, \dots, x_n) = (x_1 + g, \dots, x_n + g)$ . Of course, the  $N_1(\theta, 1)$  distribution can be replaced by any translation family  $f(x - \theta)$  where  $f$  is known.

**Example 3.2.** Let  $X_1, X_2, \dots, X_n, Z$  be independent  $N_1(\mu, \sigma^2)$ , and again take  $X = (X_1, X_2, \dots, X_n)$ . This time the natural group is the affine group  $Al_1$  of transformations on  $\mathbb{R}^n$  of the form  $gx = (a, b)x = (ax_1 + b, \dots, ax_n + b)$ . The action of  $Al_1$  on  $\mathcal{Z} = \mathbb{R}^1$  is obvious and it acts on the parameter space  $\mathbb{R} \times \mathbb{R}^+$  by  $(a, b)(\mu, \sigma^2) = (a\mu + b, a^2\sigma^2)$ .

**Example 3.3.** Let  $X_1, \dots, X_n, Z$  be independent column vectors which are  $N_p(0, \Sigma)$ , where the dimension  $p$  can be larger than 1. The unknown  $p \times p$  covariance matrix  $\Sigma$  is assumed to be positive definite. It is assumed that  $n \geq p$ . The sample space  $\mathcal{X}$  for this example is taken to be the set of all  $n \times p$  real matrices whose first  $p$  rows are linearly independent, and the data,  $X_1, \dots, X_n$  are written as

$$x = \begin{pmatrix} X'_1 \\ X'_2 \\ \vdots \\ X'_n \end{pmatrix} \in \mathcal{X}.$$

Note that  $\mathcal{X}$  is *not* the usual sample space consisting of all  $n \times p$  matrices of rank  $p$ . The reason for this choice occurs in the next section (see Example 3.3 (continued)).

The space  $\mathcal{Z}$  is just  $\mathbb{R}^p$ . There are at least two groups of interest for this example. Let  $G_2$  be the group of all  $p \times p$  non-singular matrices and let  $G_1(\subseteq G_2)$  be the group of all  $p \times p$  lower-triangular matrices which have positive diagonal elements. The action of  $G_2$  is

$$\begin{cases} x \rightarrow xg', & x \in \mathcal{X} \\ z \rightarrow gz, & z \in \mathcal{Z} \\ \Sigma \rightarrow g\Sigma g'. \end{cases}$$

It is easy to check that the model is invariant under  $G_2$  and hence under  $G_1$  since  $G_1 \subseteq G_2$ . Both  $G_1$  and  $G_2$  are transitive on the parameter space.

By definition, a *predictive inference*, or just *inference* for brevity, is a Markov kernel  $Q(dz|x)$  on  $\mathcal{B}(\mathcal{Z}) \times \mathcal{X}$ . Such a  $Q$  is also called a *predictive distribution*. It is *invariant* under the group  $G$  if

$$\int f(g^{-1}z)Q(dz|gx) = \int f(z)Q(dz|x)$$

for all  $x \in \mathcal{X}$ ,  $g \in G$  and bounded, measurable  $f$ .

**Remark 3.1.** The predictive framework of this paper includes the usual problem of inference about  $\theta$  as a special case. Just take  $\mathcal{Z} = \Theta$  and let  $P(X \in A, Z = \theta|\theta) = P(X \in A|\theta)$  for all  $A \in \mathcal{B}(\mathcal{X})$ ,  $\theta \in \Theta$ .

### 4. The Fisherian pivotal prediction

In this section, we formulate an abstract version of Fisherian pivotal inference. We will always assume that  $G$  is transitive on  $\Theta$ ; that is, given  $\theta_1, \theta_2 \in \Theta$ , there exists  $g \in G$  with  $g\theta_1 = \theta_2$ . In addition, the following assumption plays a crucial role.

**Assumption FP.** *There is a measurable mapping  $\tau : \mathcal{X} \rightarrow G$  such that  $\tau(gx) = g\tau(x)$  for all  $x \in \mathcal{X}, g \in G$ .*

It follows from Assumption FP that  $G$  is exactly transitive on each orbit in  $\mathcal{X}$ ; that is, if  $x = gx$ , then the group element  $g$  is the identity. Aside from measurability issues, the converse is also true – namely, if  $G$  is exactly transitive on each orbit, then FP holds. Exactly transitive group actions are also called free group actions (see Bourbaki 1966, p. 254).

In many examples there is an obvious choice for  $\tau$ .

**Example 3.1 (continued).** For  $x = (x_1, \dots, x_n) \in \mathcal{X}$ , let  $\tau(x) = \bar{x}$  correspond to translation by  $\bar{x}$ .

**Example 3.2 (continued).** Let  $\tau(x) = (s, \bar{x})$  be the element of  $Al_1$  corresponding to multiplication by the sample sum of squares  $s$  and translation by  $\bar{x}$ . Here,

$$s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

which is assumed to be positive.

**Example 3.3 (continued).** To construct the map  $\tau$  for  $G_2$  in this example, consider  $x \in \mathcal{X}$  and partition  $x$  as

$$x = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

where  $\alpha$  is  $p \times p$  of rank  $p$ , and  $\beta$  is  $(n - p) \times p$ . For  $g \in G_2$ , recall that

$$g(x) = xg'$$

is the action of  $G_2$  (and hence  $G_1$ ) on  $x \in \mathcal{X}$ . Here the left-hand side denotes the group action and the right-hand side denotes matrix multiplication. Now, define  $\tau_2$  by

$$\tau_2(x) = \alpha' \in G_2.$$

Then  $\tau_2(g(x)) = \tau_2(xg') = (\alpha g')' = g\alpha' = g \circ \tau_2(x)$  where ‘ $\circ$ ’ denotes group multiplication. Thus Assumption FP holds for  $G_2$ .

The construction of a  $\tau_1$  for  $G_1$  requires a bit more care. For  $x \in \mathcal{X}$ , the matrix  $S = x'x$  has rank  $p$  since  $x$  has rank  $p$ . Let  $t(S)$  denote the unique element of  $G_1$  which satisfies  $S = t(S)(t(S))'$  and recall that  $t(hSh') = ht(S)$  for  $h \in G_1$  (see Eaton 1983, Chapter 5 for a proof). Now, define  $\tau_1$  by

$$\tau_1(x) = t(x'x).$$

Then, for  $h \in G_1$ ,

$$\tau_1(h(x)) = \tau_1(xh') = t(hx'xh') = ht(x'x) = h \circ \tau_1(x).$$

Thus Assumption FP holds for  $G_1$ .

Return now to the general framework. A function  $f$  defined on  $\mathcal{X}$  is *invariant* if  $f(x) = f(gx)$  for all  $x$  and  $g$ . Let  $\psi$  be a measurable mapping from  $\mathcal{X}$  onto a measurable space  $\mathcal{U}$  that is a *maximal invariant under the action of  $G$  on  $\mathcal{X}$* ; that is,  $\psi$  is invariant and  $\psi(x_1) = \psi(x_2)$  implies  $gx_1 = x_2$  for some  $g$ . (The space  $\mathcal{U}$  is just some version of the quotient space  $\mathcal{X}/G$  with  $\psi$  mapping each  $x$  to its orbit.) Then every invariant function  $f$  can be written in the form  $f(x) = \phi(\psi(x))$  for some function  $\phi$ .

Define now the mapping  $\lambda : \mathcal{X} \times \mathcal{Z} \rightarrow \mathcal{U} \times \mathcal{Z}$  by

$$\lambda(x, z) = (u, w) = (\psi(x), \tau^{-1}(x)z).$$

Note that  $\lambda$  is onto since  $\psi$  is onto by assumption.

**Lemma 4.1.** *The mapping  $\lambda$  is a maximal invariant under the action of  $G$  on  $\mathcal{X} \times \mathcal{Z}$ . (The action of  $G$  on  $\mathcal{X} \times \mathcal{Z}$  is given by  $g(x, z) = (gx, gz)$ .)*

**Proof.** The invariance of  $\lambda$  is obvious. To show  $\lambda$  is maximal, suppose that

$$(\psi(x_1), \tau^{-1}(x_1)z_1) = (\psi(x_2), \tau^{-1}(x_2)z_2).$$

Then  $\psi(x_1) = \psi(x_2)$  and, by the maximality of  $\psi$ ,  $gx_1 = x_2$  for some  $g$ . Hence, by Assumption FP,

$$\tau^{-1}(x_1)z_1 = \tau^{-1}(x_2)z_2 = \tau^{-1}(x_1)g^{-1}z_2$$

and therefore  $gz_1 = z_2$ . □

Because  $(U, W) = (\psi(X), \tau^{-1}(X)Z)$  is an invariant statistic and  $G$  is transitive on  $\Theta$ , the joint distribution of  $(U, W)$  induced by the model  $P(dx, dz|\theta)$  does not depend on  $\theta \in \Theta$ . Let  $H(du, dw)$  be this induced distribution and let

$$H(du, dw) = R_0(dw|u)S_0(du) \tag{4}$$

be a disintegration of  $H$  into its marginal  $S_0$  for  $U$  and conditional distribution  $R_0$  for  $W$  given  $U$ . (The disintegration is possible because  $W$  takes its values in the Polish space  $\mathcal{Z}$ .)

We now define the *Fisherian pivotal inference*  $Q_0$  by

$$Q_0(B|x) = R_0(\tau^{-1}(x)B|\psi(x)) \tag{5}$$

for all  $x \in \mathcal{X}$ ,  $B \in \mathcal{B}(\mathcal{Z})$ . Thus the predictive distribution of  $Z$  given  $X = x$  under  $Q_0$  corresponds to the conditional distribution of  $\tau(x)W$  given  $U = \psi(x)$  under the model.

In simple examples like 3.1 and 3.2, we recover the usual pivotal predictions.

**Example 3.1 (continued).** Take  $\tau(x) = \bar{x}$  and let  $\psi(x) = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$  be the

maximal invariant for the action of the translation group on  $\mathcal{X} = \mathbb{R}^n$ . Under the model,  $W = Z - \bar{X}$  and  $U = \psi(X)$  are independent. Since  $W \sim N(0, 1 + \frac{1}{n})$ , the pivotal prediction given  $x$  is the distribution of  $W + \bar{x} \sim N(\bar{x}, 1 + \frac{1}{n})$ .

**Example 3.2 (continued).** Take  $\tau(x) = (s, \bar{x})$  and  $\psi(x) = s^{-1}(x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x})$ . An application of Basu's theorem shows that again  $W = (Z - \bar{X})/S$  and  $U = \psi(X)$  are independent. Let  $t(m)$  denote a variable with the Student  $t$  distribution with  $m$  degrees of freedom and set  $c = (1 + 1/n)^{1/2}/(n - 1)^{1/2}$ . Since the distribution of  $W$  is that of  $ct(n - 1)$ , the pivotal prediction, given  $x$ , is the distribution of  $sW + \bar{x}$  which is that of  $sct(n - 1) + \bar{x}$ . For this example, the sample space  $\mathcal{X}$  is the set of  $x \in \mathbb{R}^n$  such that  $s^2 = \sum_1^n (x_i - \bar{x})^2 > 0$ .

**Example 3.3 (continued).** First, we describe the prediction distribution of  $Z$  given  $x$  when the group under consideration is  $G_1$ . For this case  $\tau_1(x) = t(S)$  where  $S = x'x$  is positive definite and  $t(S)$  is the unique element of  $G_1$  which satisfies  $S = t(S)t(S)'$ . Thus,

$$\psi(x) = x[(t(S))^{-1}]'$$

and

$$W = (t(S))^{-1}Z.$$

An easy application of Basu's theorem shows that under the model,  $W$  and  $U$  are independent. Thus, the predictive density of  $Z$  given  $x$  is the density of  $t(S)W$  with  $S$  fixed. The density of  $W$  is given in Eaton and Sudderth (1993, Theorem 3.1, p. 485). Hence we have the Fisherian pivotal predictive distribution when the group  $G_1$  is used.

Rather than give the FP predictive distribution here when  $G_2$  is the group, we will do so in Section 10 after an alternative method of calculation is established. However, it should be pointed out that the  $G_1$  and  $G_2$  invariant predictive distributions are different.

Assumption FP is quite strong and can fail to hold for at least two reasons – namely, a reduction by sufficiency has occurred and changed the sample space, or the 'wrong' group was chosen to describe the invariance of the problem. The following example illustrates both of these reasons for the failure of FP to hold.

**Example 4.1.** As in Example 3.1, consider  $X_1, X_2, \dots, X_n, Z$  that are independent  $N_1(\theta, 1)$ . With  $e$  denoting the vector of ones in  $\mathbb{R}^n$ , the sample space of the  $X$ s is  $\mathcal{X} = \mathbb{R}^n \cap (\text{span}\{e\})^c$ . That is,  $\mathcal{X}$  is the space  $\mathbb{R}^n$  with the one-dimensional subspace  $(\text{span}\{e\})$  removed. Of course,  $\Theta = \mathcal{Z} = \mathbb{R}^1$  in this example. There are three groups of interest in this example. The first is  $H_1$  which is the translation group on  $\mathbb{R}^1$  and acts on  $x \in \mathcal{X}$  in the natural way:  $x \rightarrow x + be$ , where  $b \in \mathbb{R}^1$ . The action of  $H_1$  on  $\Theta = \mathcal{Z}$  is just  $\theta \rightarrow \theta + b$ .

The second group is  $H_2$  which consists of pairs  $(a, b)$  with  $a = \pm 1, b \in \mathbb{R}^1$  and group composition given by

$$(a_1, b_1)(a_2, b_2) = (a_1a_2, a_1b_2 + b_1).$$

Then  $H_2$  acts on  $\mathcal{X}$  via  $x \rightarrow ax + be$  and on  $\Theta = \mathcal{Z}$  via  $\theta \rightarrow a\theta + b$ .

The third group is  $H_3$  which consists of all pairs  $(\Gamma, b)$  where  $b \in \mathbb{R}^1$  and  $\Gamma$  is an  $n \times n$  orthogonal matrix satisfying  $\Gamma e = e$ . The group operation is

$$(\Gamma_1, b_1)(\Gamma_2, b_2) = (\Gamma_1\Gamma_2, b_1 + b_2),$$

the action on  $\mathcal{X}$  is  $x \rightarrow \Gamma x + be$  while the action on  $\Theta = \mathcal{Z}$  is  $\theta \rightarrow \theta + b$ . It is easily checked that the prediction problem is invariant under  $H_i, i = 1, 2, 3$ .

Choosing  $\tau_1(x) = \bar{x}$ , it is clear that Assumption FP holds when the group is  $H_1$ . When the group is  $H_2$ , the situation is a bit more complicated. First, assume  $n \geq 2$ . For  $x \in \mathcal{X}$ , let  $\psi(x)$  be the first coordinate of  $x$  which is not equal to  $x_1$ , where  $x = (x_1, x_2, \dots, x_n)$ . With  $\tau_2(x) = (\text{sgn}(x_1 - \psi(x)), \bar{x})$ , it is not hard to check that FP holds. However, if  $n = 1$ , a direct calculation shows that FP cannot hold. Thus if we had reduced to the sufficient statistic  $\bar{X}$  and used the group  $H_2$ , then FP would not have held. This circumstance makes Assumption FP rather unappealing, in spite of its rather natural use in constructing predictive distributions.

Finally, consider the group  $H_3$  and assume that  $n \geq 3$ . For  $x \in \mathcal{X}$ , the equation

$$\Gamma x + be = x$$

implies that  $b = 0$  (just multiply by  $e'$  to obtain  $nb = 0$ ) so that  $\Gamma x = x$ . Because  $n \geq 3$ , there are always solutions other than  $\Gamma = I_n$  to the equation  $\Gamma x = x$ , where  $x \in \mathcal{X}$ ,  $\Gamma e = e$  and  $\Gamma$  is orthogonal. Thus the action on  $\mathcal{X}$  is not free and so FP cannot hold.

Even though FP may not hold, it is often possible to use the formal Bayes method (with right Haar measure as an improper prior) to obtain an invariant predictive inference. When the group  $G$  is amenable, this method will produce a consistent (not strongly inconsistent) predictive inference. A detailed discussion of this occurs in Sections 7 and 8. For the normal example above considered invariant under  $H_i$ , the formal Bayes method produces a consistent inference which is the same for each  $i, i = 1, 2, 3$ .

## 5. Describing invariant inferences

Some general properties of invariant inferences will be established in this section and will be used in the next section to compare such inferences to the Fisherian pivotal inference. We continue to assume FP and follow the notation of the previous section.

**Lemma 5.1.** *Given a Markov kernel  $R(dz|u)$  on  $\mathcal{Z}$  given  $U$ , the inference  $Q$  defined by*

$$Q(B|x) = R(\tau^{-1}(x)B|\psi(x)), \quad x \in \mathcal{X}, \quad B \in \mathcal{B}(\mathcal{Z}), \tag{6}$$

*is invariant. Conversely, given an invariant inference  $Q$ , there exists a unique Markov kernel  $R(dz|u)$  satisfying the same formula.*

**Proof.** It is straightforward to verify the first assertion.

For the converse, let  $Q(dz|x)$  be an invariant inference. Define another inference  $H(dz|x)$  by



$$H(B|x) = Q(\tau(x)B|x), \quad x \in \mathcal{X}, \quad B \in \mathcal{B}(\mathcal{Z}).$$

The invariance of  $Q$  and Assumption FP imply that the function  $x \rightarrow H(B|x)$  is invariant under the action of  $G$  on  $\mathcal{X}$  and must therefore be a function of the maximal invariant  $\psi(x)$ . Say

$$H(B|x) = R(B|\psi(x)).$$

Clearly, equality (6) holds and it is again straightforward to verify that  $R$  is a Markov kernel. (To see that  $R(B|u)$  is measurable in  $u$  for fixed  $B$ , identify  $\mathcal{U}$  with  $\mathcal{X}/G$  and assume that  $\mathcal{U}$  has the quotient  $\sigma$ -field  $\mathcal{S}$  defined by  $A \in \mathcal{S}$  if and only if  $\psi^{-1}(A) \in \mathcal{B}(\mathcal{X})$ .) The uniqueness of  $R$  is immediate because  $\psi$  maps  $\mathcal{X}$  onto  $\mathcal{U}$  and, for each  $x, B \rightarrow \tau^{-1}(x)B$  maps  $\mathcal{B}(\mathcal{Z})$  onto itself.  $\square$

For a given invariant inference  $Q$ , let  $R_Q$  denote the unique  $R$  that satisfies (6). If  $h: \mathcal{Z} \rightarrow \mathbb{R}^1$  is bounded and measurable, then

$$\int h(z) Q(dz|x) = \int h(\tau(x)z) R_Q(dz|\psi(x)). \tag{7}$$

This equality is just (6) when  $h$  is the indicator of a set  $B \in \mathcal{B}(\mathcal{Z})$ . The extension to functions is obvious.

Suppose now that  $f(x, z)$  is an invariant function on  $\mathcal{X} \times \mathcal{Z}$ . By Lemma 4.1,  $\lambda(x, z) = (\psi(x), \tau^{-1}(x)z)$  is maximal invariant. So we can write

$$f(x, z) = f^*(\psi(x), \tau^{-1}(x)z) \tag{8}$$

for some function  $f^*$  on  $\mathcal{U} \times \mathcal{Z}$ . Conversely, given any  $f^*$  on  $\mathcal{U} \times \mathcal{Z}$ , formula (8) defines an invariant function  $f$  on  $\mathcal{X} \times \mathcal{Z}$ .

**Lemma 5.2.** *Let  $Q$  be an invariant inference and let  $f : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}^1$  be invariant, bounded and measurable. Then, for  $f^*$  as in (8),*

$$\inf_x \int f(x, z) Q(dz|x) = \inf_u \int f^*(u, z) R_Q(dz|u). \tag{9}$$

**Proof.** By (7) and (8),

$$\int f(x, z) Q(dz|x) = \int f^*(\psi(x), z) R_Q(dz|\psi(x)), \tag{10}$$

and the mapping  $\psi$  is onto.  $\square$

For the pivotal inference  $Q_0$ , we see from (5) and (6) that  $R_0 = R_{Q_0}$ . Also recall from Section 4 that  $R_0(dw|u)S_0(du)$  is the distribution of  $(U, W) = (\psi(X), \tau^{-1}(X)Z)$  under every value of the parameter  $\theta$ . The next lemma is then immediate.

**Lemma 5.3.** *Let  $f^* : \mathcal{U} \times \mathcal{Z} \rightarrow \mathbb{R}^1$  be bounded and measurable. Define  $f : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}^1$  by (8). Then*

$$\sup_{\theta} E_{\theta} f(X, Z) = \iint f^{*}(u, z) R_0(dz|u) S_0(du). \tag{11}$$

### 6. Non-pivotal invariant inferences are strongly inconsistent

Continue to assume FP holds. Every invariant predictive inference  $Q$  for  $Z$  given  $X$  can be regarded as a predictive inference  $R_Q$  about  $\tau(X)Z$  given  $\psi(X)$  by (6). Now  $S_0$  is the distribution of the maximal invariant  $\psi(X)$  under every parameter value  $\theta$ . Thus it is natural to regard two such inferences  $Q_1$  and  $Q_2$  as being equivalent if  $R_{Q_1}(\cdot|u) = R_{Q_2}(\cdot|u)$  for  $S_0$ -almost every  $u$ .

**Definition.** Two invariant predictive inferences  $Q_1$  and  $Q_2$  are essentially different if  $\{u : R_{Q_1}(\cdot|u) \neq R_{Q_2}(\cdot|u)\}$  has positive  $S_0$ -measure.

**Theorem 6.1.** If Assumption FP holds and  $Q$  is an invariant predictive inference that is essentially different from the Fisherian pivotal inference  $Q_0$ , then  $Q$  is strongly inconsistent.

**Proof.** Because  $\mathcal{Z}$  is Polish,  $\mathcal{B}(\mathcal{Z})$  is countably generated. So we can apply Theorem A.1 (the separation theorem) in the Appendix to  $R_Q(\cdot|u)$  and  $R_0(\cdot|u)$  to obtain a bounded, measurable function  $f^{*}(u, z)$  such that

$$\inf_u \int f^{*}(u, z) R_Q(dz|u) > \iint f^{*}(u, z) R_0(dz|u) S_0(du). \tag{12}$$

Define  $f(x, z)$  by (8). Then, by Lemma 5.2, Lemma 5.3 and (12), we have

$$\inf_x \int f(x, z) Q(dz|x) > \sup_{\theta} E_{\theta} f(X, Z).$$

□

An obvious question, left unanswered by Theorem 6.1, is whether the pivotal inference  $Q_0$  is itself consistent. Sufficient conditions for its consistency will be given in Section 8 after the next section introduces another method of invariant prediction. However, there are standard examples in multivariate analysis where  $Q_0$  is strongly inconsistent (see Section 10).

### 7. The formal Bayes inference from right Haar measure

Invariant predictive inferences can often be obtained by the usual Bayes algorithm from invariant prior distributions. In this section we introduce as a prior a right Haar measure  $\nu$  on the locally compact group  $G$ . The non-trivial measure  $\nu$  is defined on  $\mathcal{B}(G)$  and is uniquely determined up to a positive constant by its property that

$$\int f(gh)\nu(dg) = \int f(g)\nu(dg) \tag{13}$$

for all  $h \in G$  and measurable  $f : G \rightarrow \mathbb{R}^+$ . We continue to make all the assumptions of Section 3 and also assume that  $G$  acts transitively on  $\Theta$ . Consequently, we can fix an element  $\theta_0 \in \Theta$  and write the model as  $P(dx, dz|g\theta_0)$ , where  $g$  ranges over  $G$ .

To obtain the Bayes predictive inference from  $\nu$ , we no longer need the structural assumption FP of Section 4, but we will use two assumptions that were not needed to obtain the pivotal inference  $Q_0$ .

Let  $M$  be the marginal measure on  $\mathcal{X}$  determined by  $\nu$  and the model; that is,

$$M(B) = \int_G \int_{\mathcal{Z}} P(B, dz|g\theta_0)\nu(dg). \tag{14}$$

**Assumption H1.** *The measure  $M$  is  $\sigma$ -finite.*

It follows from this assumption that we can write

$$P(dx, dz|g\theta_0)\nu(dg) = Q_1(dg, dz|x)M(dx), \tag{15}$$

where  $Q_1$  is a Markov kernel on  $G \times \mathcal{Z}$  given  $\mathcal{X}$  and is uniquely determined up to a set of  $M$ -measure zero. (This fact is not difficult and is presumably well known. The only reference that we know for it is Johnson (1991).)

**Assumption H2.** *The Markov kernel  $Q_1$  is  $G$ -invariant; that is,*

$$Q_1(hA, hB|hx) = Q_1(A, B|x), \quad h \in G, \quad x \in \mathcal{X}, \quad A \in \mathcal{B}(G), \quad B \in \mathcal{B}(\mathcal{Z}). \tag{16}$$

In all the applications that we know, Assumption H2 is satisfied. However, we do not know a theorem which implies that  $Q_1$  can be so chosen. Rather than discuss this issue here, we have chosen to make it an assumption. (It is not difficult to show that  $Q_1$  must be  $G$ -invariant for  $M$ -almost all  $x$ .)

A consequence of H2 is that the predictive inference  $Q^*(dz|x)$  defined by

$$Q^*(B|x) = Q_1(G \times B|x), \quad x \in \mathcal{X}, \quad B \in \mathcal{B}(\mathcal{Z}), \tag{17}$$

is invariant. The inference  $Q^*$  is the *formal Bayes predictive inference* from  $\nu$  (or the *right Haar inference* for short).

In the special case when  $X$  and  $Z$  are independent under the model for each parameter value  $\theta$ ,  $Q^*$  can be calculated by first finding the formal posterior  $\Pi(dg|x)$  which is characterized by the formula

$$P(dx|g\theta_0)\nu(dg) = \Pi(dg|x)M(dx)$$

and then integrating as

$$Q^*(B|x) = \int P[Z \in B|g\theta_0]\Pi(dg|x). \tag{18}$$

Here is another look at our examples.

**Example 3.1 (continued).** We can identify the group  $G = \mathbb{R}^1$  with the parameter space  $\Theta$ . Take  $\theta_0 = 0$  and  $\nu(d\theta) = d\theta$ . Under  $\Pi(d\theta|x)$ ,  $\theta \sim N(\bar{x}, 1/n)$  and, given  $\theta$ ,  $Z \sim N(\theta, 1)$  under the model. So, by (7.6), under  $Q^*(dz|x)$ ,  $Z \sim N(\bar{x}, 1 + \frac{1}{n})$ . Thus  $Q^* = Q_0$ .

**Example 3.2 (continued).** Again we can identify the group  $G = A1_1$  with the parameter space. Take  $\theta_0 = (1, 0)$  and  $\nu(d(\sigma, \mu)) = \sigma^{-1}d\mu d\sigma$ . A calculation will show that  $Q^*$  again agrees with  $Q_0$ .

**Example 3.3 (continued).** Let  $Q_i^*$  be the right Haar inferences corresponding to the groups  $G_i$ ,  $i = 1, 2$ . The  $Q_i^*$  are calculated explicitly in Eaton and Sudderth (1993). The results in Section 9 of this paper show that, for both  $G_1$  and  $G_2$ , the FP inference and the right Haar inferences are the same. However,  $Q_1^*$  and  $Q_2^*$  are different.

Now, let  $\tilde{G}$  be the group of all  $p \times p$  non-singular lower-triangular matrices. In this case, the right Haar inference exists and is easily shown to be  $Q_1^*$ . However, for  $\tilde{G}$ , we do not know whether or not FP holds, but we suspect not.

## 8. Amenable groups and consistent predictive inferences

There are a number of equivalent definitions of amenability for groups (cf. Bondar and Milnes 1981). Here is the one most convenient for our purposes.

**Assumption AM.** *There is a sequence of densities  $\{p_n(g)\}$  with respect to right Haar measure  $\nu$  such that*

$$\lim_{n \rightarrow \infty} \int |p_n(gh) - p_n(g)| \nu(dg) = 0$$

for all  $h \in G$ .

Since  $G$  is Polish and locally compact, it is  $\sigma$ -compact. That is why we are able to use a sequence rather than a net in Assumption AM.

**Theorem 8.1.** *Assume H1 and H2 so that the right Haar inference  $Q^*$  is well defined. If Assumption AM also holds, then  $Q^*$  is consistent.*

**Proof.** Let  $f : \mathcal{X} \times \mathcal{Z} \rightarrow \mathbb{R}^1$  be bounded and measurable. We must show

$$\inf_x \int f(x, z) Q^*(dz|x) \leq \sup_{\theta} E_{\theta} f(X, Z). \tag{19}$$

Fix  $\theta_0$  as in the previous section and rewrite

$$\sup_{\theta} E_{\theta} f(X, Z) = \sup_g E_{g\theta_0} f(X, Z).$$

Let  $p(g)$  be an arbitrary probability density with respect to  $\nu$  on  $G$ . Set

$$M_p(dx) = \int_G \int_{\mathcal{Z}} P(dx, dz|g\theta_0) p(g) \nu(dg). \tag{20}$$

By (15),

$$M_p(dx) = \int_G \int_{\mathcal{Z}} p(g) Q_1(dg, dz|x) M(dx). \tag{21}$$

Define

$$\hat{Q}(dg|x) = \int_{\mathcal{Z}} Q_1(dg, dz|x), \tag{22}$$

so that

$$m_p(x) = \int_G \int_{\mathcal{Z}} p(g) Q_1(dg, dz|x) = \int_G p(g) \hat{Q}(dg|x) \tag{23}$$

is the Radon–Nikodym derivative of  $M_p$  with respect to  $M$ .

Now

$$\inf_x \int f(x, z) Q^*(dz|x) \leq \iint f(x, z) Q^*(dz|x) M_p(dx)$$

and

$$\int E_{g\theta_0} f(X, Z) p(g) \nu(dg) \leq \sup_g E_{g\theta_0} f(X, Z).$$

So (19) will follow if we can show that

$$\inf_p \Delta_p = 0, \tag{24}$$

where

$$\Delta_p = \left| \int E_{g\theta_0} f(X, Z) p(g) \nu(dg) - \iint f(x, z) Q^*(dz|x) M_p(dx) \right|.$$

By (15),

$$\int E_{g\theta_0} f(X, Z) p(g) \nu(dg) = \iiint f(x, z) p(g) Q_1(dg, dz|x) M(dx)$$

and, by (17) and (23),

$$\iint f(x, z) Q^*(dz|x) M_p(dx) = \iiint f(x, z) m_p(x) Q_1(dg, dz|x) M(dx).$$

Hence, with  $K = \sup_{x,z} |f(x, z)|$ ,

$$\begin{aligned} \Delta_p &\leq K \iiint |m_p(x) - p(g)| Q_1(dg, dz|x) M(dx) \\ &= K \iiint |m_p(x) - p(g)| P(dx, dz|g\theta_0) \nu(dg) \\ &= K \iiint |m_p(gx) - p(g)| P(dx, dz|\theta_0) \nu(dg). \end{aligned}$$

The last two lines use (15) and (3), respectively. Because  $Q_1$  is invariant by Assumption H2,

$$\begin{aligned} m_p(gx) &= \iint p(h) Q_1(dh, dz|gx) \\ &= \iint p(gh) Q_1(dh, dz|x) = \int p(gh) \hat{Q}(dh|x). \end{aligned}$$

Hence,

$$\Delta_p \leq K \iiint |p(gh) - p(g)| \nu(dg) \hat{Q}(dh|x) P(dx, dz|\theta_0). \tag{25}$$

For any  $p$ ,

$$\int |p(gh) - p(g)| \nu(dg) \leq 2.$$

So it follows from Assumption AM, (25) and the dominated convergence theorem that  $\lim \Delta_{p_n} = 0$ . This completes the proof of (24) and the theorem.  $\square$

In Examples 3.1 and 3.2, the group  $G$  is amenable. So the inference  $Q^* = Q_0$  is consistent. In Example 3.3, the group  $G_1$  is amenable and, hence,  $Q_1^*$  is consistent. The group  $G_2 = Glp$  is not amenable and, indeed, it was shown in Eaton and Sudderth (1993) that the right Haar inference  $Q_2^*$  is strongly inconsistent.

## 9. The right Haar inference is the same as Fisherian pivoting

Suppose that Assumptions FP, H1, and H2 hold so that both  $Q_0$  and  $Q^*$  are well defined. If we impose the additional assumption that  $G$  acts ‘properly’ on  $\mathcal{X}$ , then  $Q_0$  and  $Q^*$  are the same by Theorem 9.1 below. Before stating it, we need to define ‘proper action’ and briefly review some of its implications.

Suppose  $G$  acts continuously on the Polish space  $\mathcal{Y}$ . Then  $G$  is said to *act properly* on  $\mathcal{Y}$  if the mapping  $(g, y) \rightarrow (gy, y)$  from  $G \times \mathcal{Y}$  to  $\mathcal{Y} \times \mathcal{Y}$  is a *proper* mapping in the sense that the inverse images of compact sets are compact. An important implication of proper action is that the quotient space  $\mathcal{Y}/G$  is a nice topological space (Hausdorff and second countable) (see Andersson 1982). Let  $\mathcal{H}(\mathcal{Y})$  be the set of all real-valued,

continuous functions with compact support defined on  $\mathcal{Y}$  and let  $\mathcal{H}(\mathcal{Y}/G)$  be the collection of such functions on  $\mathcal{Y}/G$ . For  $h \in \mathcal{H}(\mathcal{Y})$ , define

$$(Th)(\pi y) = \int h(gy)v(dg)$$

where  $\pi$  is the orbit projection from  $\mathcal{Y}$  onto  $\mathcal{Y}/G$  and  $v$  is a right Haar measure on  $G$ . Another important fact is that  $T$  maps  $\mathcal{H}(\mathcal{Y})$  onto  $\mathcal{H}(\mathcal{Y}/G)$  (for discussion, see Andersson 1982).

**Assumption P.** *The group  $G$  acts properly on  $\mathcal{X}$ .*

Because  $G$  acts continuously on  $\mathcal{Z}$  by assumption, we conclude from Bourbaki (1966, Proposition 5, p. 232) that  $G$  then acts properly on  $\mathcal{X} \times \mathcal{Z}$  when Assumption P holds.

Assumption P is satisfied by many of the standard statistical models that we know.

We will represent the quotient space  $(\mathcal{X} \times \mathcal{Z})/G$  as before by  $\mathcal{U} \times \mathcal{Z}$  using the maximal invariant  $\lambda(x, z) = (\psi(x), \tau^{-1}(x)z)$  as in Lemma 4.1.

**Theorem 9.1.** *Under Assumptions FP, H1, H2, and P, the Fisherian pivotal inference  $Q_0(\cdot|x)$  and the right Haar inference  $Q^*(\cdot|x)$  are equal for  $M$ -almost all  $x$  where  $M$  is the marginal determined by the model and right Haar measure  $v$  as in (14).*

**Proof.** For each  $f \in \mathcal{H}(\mathcal{X} \times \mathcal{Z})$ , let

$$I = \iint f(x, z)Q^*(dz|x)M(dx), \quad J = \iint f(x, z)Q_0(dz|x)M(dx).$$

It suffices to show  $I = J$ . In fact, we will show that

$$I = J = \iint (Tf)(u, w)R_0(dw|u)S_0(du),$$

where  $Tf$  is the invariant function

$$(Tf)(\lambda(x, z)) = \int f(gx, gz)v(dg),$$

with  $\lambda$  as in Lemma 4.1, and  $R_0, S_0$  as in (4).

We first calculate  $I$ . By (15) and (17),

$$\begin{aligned}
 I &= \iiint f(x, z)P(dx, dz|g\theta_0)\nu(dg) \\
 &= \iiint f(gx, gz)\nu(dg)P(dx, dz|\theta_0) \\
 &= \iint (Tf)(\psi(x), \tau^{-1}(x)z)P(dx, dz|\theta_0) \\
 &= \iint (Tf)(u, w)R_0(dw|u)S_0(du).
 \end{aligned}$$

For  $J$ , use (15) to write

$$\begin{aligned}
 J &= \iiint \iiint f(x, z)Q_0(dz|x)P(dx, d\tilde{z}|g\theta_0)\nu(dg) \\
 &= \iiint \iiint f(gx, z)Q_0(dz|gx)P(dx, d\tilde{z}|\theta_0)\nu(dg) \\
 &= \iiint \iiint f(gx, gz)\nu(dg)Q_0(dz|x)P(dx, d\tilde{z}|\theta_0) \\
 &= \iint (Tf)(\psi(x), \tau^{-1}(x)z)Q_0(dz|x)P(dx|\theta_0) \\
 &= \iint (Tf)(u, w)R_0(dw|u)S_0(du).
 \end{aligned}$$

□

The fact that  $Q_0$  and  $Q^*$  are the same can be useful in calculations as well as theory, for one of the two can be easier to calculate in particular examples.

**Remark 9.1.** Suppose  $Q_1(\cdot|x)$  and  $Q_2(\cdot|x)$  are both  $G$ -invariant predictive distributions. Since FP holds by assumption,  $Q_i(\cdot|x)$  can be represented by  $R_i(\cdot|u)$ ,  $i = 1, 2$ , as in Section 5. Let

$$A = \{x|Q_1(\cdot|x) \neq Q_2(\cdot|x)\}$$

and

$$E = \{u|R_1(\cdot|u) \neq R_2(\cdot|u)\}.$$

It can be shown that  $A$  has  $M$ -measure zero if and only if  $E$  has  $S_0$ -measure zero. Thus, equality a.e.- $M$  in Theorem 9.1 implies  $Q_1$  and  $Q_2$  are not essentially different as defined in Section 6.



### 10. Examples continued

For Examples 3.1 and 3.2, all of our assumptions (FP, H1, H2, amenability and proper action) hold. Thus using right Haar measure as an improper prior yields an invariant predictive distribution which is consistent. Further, any other invariant predictive distribution is strongly inconsistent. In short, there is only one sensible invariant predictive inference. In addition, the normality assumptions play essentially no role in these conclusions. For example, if  $X_1, \dots, X_n$  are independently and identically distributed from a distribution with a density  $f(x - \theta)$ ,  $x \in \mathbb{R}^p$ ,  $\theta \in \mathbb{R}^p$  and  $Z$  is also from  $f(x - \theta)$ , then using the improper prior ‘ $d\theta$ ’ on  $\mathbb{R}^p$  yields a consistent invariant predictive distribution.

The situation for Example 3.3 is somewhat more complicated. First consider the group  $G_2$  of all  $p \times p$  non-singular matrices. As remarked earlier, the prediction problem of Example 3.3 is invariant under  $G_2$ . Further, all of the assumptions listed above, *except* amenability, are satisfied. Thus, one can use the right Haar measure on  $G_2$  to obtain the right Haar inference  $Q_2^*$ . Of course,  $Q_2^*$  agrees with what one obtains from Fisherian pivoting, using the group  $G_2$ . An explicit formula for  $Q_2^*$  is well known and can be found, for example, in Eaton and Sudderth (1993). It follows from Theorem 6.1 that any  $G_2$ -invariant predictive inference which differs from  $Q_2^*$  is strongly inconsistent.

Next, consider the group  $G_1$ . Now, all of our assumptions including amenability hold. Using right Haar measure as an improper prior leads to a  $G_1$ -invariant inference, say  $Q_1^*$ . Of course  $Q_1^*$  agrees with the inference obtained by Fisherian pivoting, using the group  $G_1$ , and  $Q_1^*$  is consistent. An explicit formula for  $Q_1^*$  is given in Eaton and Sudderth (1993).

The above discussion implies the following:

*For  $p \geq 2$  in Example 3.3, all  $G_2$  invariant predictive distributions are strongly inconsistent.* (26)

To see this, let  $Q$  be some  $G_2$ -invariant inference. If  $Q$  is different from  $Q_2^*$ , it is strongly inconsistent by Theorem 6.1. However, Theorem 6.1 also implies that any  $G_1$ -invariant inference which differs from  $Q_1^*$  is strongly inconsistent. A direct comparison of  $Q_2^*$  and  $Q_1^*$  shows that  $Q_2^*$  differs from  $Q_1^*$ . Hence, (26) follows. In short, if attention is restricted to  $G_1$ -invariant inferences, then  $Q_1^*$  seems to be the only viable inference – all others are strongly inconsistent by Theorem 6.1. In particular, all  $G_2$ -invariant inferences share this fate.

But there is one troubling aspect of the above situation – namely, the inference  $Q_1$  is coordinate-dependent. To see this, consider the case of  $p = 2$  and let  $\Gamma$  be the  $2 \times 2$  matrix

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

Thus,  $\Gamma$  interchanges coordinates of vectors in  $\mathbb{R}^2$ . Given the original data,  $X_1, \dots, X_n$  and the predictand  $Z$ , consider  $\tilde{X}_i = \Gamma X_i$ ,  $i = 1, \dots, n$ , and  $\tilde{Z} = \Gamma Z$ . Then, using  $G_1$  and  $\tilde{X}_1, \dots, \tilde{X}_n$ , construct  $\tilde{Q}_1^*$  as the ‘good’ (consistent) inference for  $\tilde{Z}$ . The marginal predictive distribution for the first coordinate of  $Z$  is *not* the same as the marginal predictive distribution for the second coordinate of  $\tilde{Z}$  (which equals the first coordinate of  $Z$ ). This is an

example of what is meant by saying ‘ $Q_1^*$  is coordinate-dependent’. The implications of this are currently under study.

## 11. Discussion

The focus of this paper has been on two methods of constructing invariant predictive distributions when the underlying model is invariant – namely, Fisherian pivoting and the formal Bayes method using the right Haar measure as a prior. In most examples that we know, both methods yield the same answer. Further, when the group under consideration is amenable, the inference is consistent. One consequence of this is that within the class of relatively invariant prior distributions (these tend to yield invariant inferences), the right Haar measure stands out as the one which produces the only possible candidate for consistency. This follows from Theorem 6.1 combined with Theorem 9.1. The conclusions here bear some implications for the more general discussion concerning improper priors in the recent survey paper of Kass and Wasserman (1996). In particular, our Example 3.3 is an instance where invariance arguments and the Jeffreys prescription produce strong inconsistency in a standard statistical model.

A widely applicable alternative method of constructing invariant inferences is first to estimate the parameter  $\theta$  using data  $x$  (typically by maximum likelihood) and then use the estimate  $\hat{\theta}$  to ‘estimate’ the distribution of  $Z$ . A discussion of this method within the framework of asymptotics appears in Barndorff-Nielsen and Cox (1996). The discussion here will focus on this method. To simplify things a bit, it is assumed that given  $\theta$ , the data  $X$  and the predictand  $Z$  are independent with distributions  $P_1(\cdot|\theta)$  and  $P_2(\cdot|\theta)$ , respectively. First, observe that if  $\pi$  is a proper prior distribution for  $\theta$ , then the predictive distribution for  $Z$  given  $X = x$  is

$$Q_\pi(dz|x) = \int_{\Theta} P_2(dz|\theta) Q_\pi(d\theta|x), \quad (27)$$

where  $Q_\pi(d\theta|x)$  is the posterior distribution of  $\theta$  given the data  $x$ . Now, let

$$E = \{P_2(\cdot|\theta) | \theta \in \Theta\}$$

and let  $C_2$  be the convex set of probability measures generated by  $E$ . Observe that  $Q_\pi(dz|x)$ , being an average of points in  $E$ , tend to be ‘inside’  $C_2$ . However, if  $\hat{\theta}$  is any estimator of  $\theta$ , the predictive distribution obtained by substituting  $\hat{\theta}$  for  $\theta$  in  $P_2(dz|\theta)$ , that is  $Q_2(dz|x) = P_2(dz|\hat{\theta}(x))$ , is always in the set  $E$ . Since  $E$  will usually be contained in the boundary of  $C_2$ , we conclude that  $P_2(dz|\hat{\theta}(x))$  cannot ‘look like’ a Bayes rule  $Q_\pi(dz|x)$  for any  $\pi$ . This suggests that substituting estimators into  $P_2(dz|\theta)$  to obtain predictive distributions is ill advised. This argument is certainly very soft, but does suggest that predictive distributions of the form  $P_2(dz|\hat{\theta}(x))$  may not perform well.

Now, maintain the independence assumption made in the previous paragraph. Assume, further, that  $G$  is a group which acts on  $X$ ,  $\mathcal{Z}$  and  $\Theta$ , that  $G$  is transitive on  $\Theta$ , and that both  $P_1(dx|\theta)$  and  $P_2(dz|\theta)$  are invariant families of distributions. Recall that an estimator  $\hat{\theta}(x)$  is invariant if  $\hat{\theta}(gx) = g\hat{\theta}(x)$ . Under rather mild assumptions, the maximum likelihood

estimator is invariant in the situation under consideration here (for discussion, see Eaton 1983, Chapter 7). Assuming that  $\hat{\theta}(x)$  is invariant, note that the predictive distribution

$$Q_3(dz|x) = P_2(dz|\hat{\theta}(x))$$

is an invariant predictive inference as defined in Section 3. In all of the examples we have checked (including Examples 3.1, 3.2 and 3.3),  $Q_3(dz|x)$  is not the same as the predictive inference obtained from Fisherian pivoting (assuming FP holds). Thus, by Theorem 6.1,  $Q_3(dz|x)$  will be strongly inconsistent.

In this paper, the prediction problem has been formulated as one of ‘producing a distribution for  $Z$  after seeing the data’, rather than a problem of producing a point predictor for  $Z$ . The only evaluative criterion considered here has been strong inconsistency (and its negation, consistency). In invariant situations, our results show that there is essentially one technique for finding a consistent invariant predictive distribution (when the group is amenable) – namely, one should use the right Haar measure as an improper prior, and calculate the formal posterior distribution of  $Z$  given the data.

The more stringent evaluation of predictive distributions using decision-theoretic notions (minimaxity, admissibility, etc.) has received very little attention in the literature. A few results can be found in Eaton (1982; 1992), but a body of work providing hard evidence – that is, i.e. theorems – that specific predictive distributions will perform well in particular situations is, in the main, lacking. The results in Barndorff-Nielsen and Cox (1996) do provide an asymptotic justification for likelihood-based inference, but fixed sample size results are few. Providing such presents a real challenge.

## Appendix: A separation theorem

Let  $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$  and  $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$  be measurable spaces. Assume singletons  $\{u\}$  are in  $\mathcal{B}(\mathcal{U})$ . Fix a Markov kernel  $R_0(dz|u)$  and a probability measure  $S_0(du)$ . Consider a second Markov kernel  $R_1(dz|u)$ .

**Theorem A.1.** *If  $\mathcal{B}(\mathcal{Z})$  is countably generated and*

$$S_0\{u : R_0(\cdot|u) \neq R_1(\cdot|u)\} > 0,$$

*then there exists a bounded,  $\mathcal{B}(\mathcal{U}) \times \mathcal{B}(\mathcal{Z})$ -measurable function  $g : \mathcal{U} \times \mathcal{Z} \rightarrow \mathbb{R}^1$  such that*

$$\inf_u \int g(u, z) R_1(dz|u) > \iint g(u, z) R_0(dz|u) S_0(du). \tag{28}$$

The proof is an application of a separation theorem (Dunford and Schwartz 1957, Theorem V.2.10). For the application it is convenient to rewrite (28) as

$$\inf_m \iint g(u, z) R_1(dz|u) m(du) > \iint g(u, z) R_0(dz|u) S_0(du), \tag{29}$$

where  $m$  ranges over the collection  $\mathcal{P}(\mathcal{U})$  of probability measures defined on  $\mathcal{B}(\mathcal{U})$ .

Let  $L$  be the linear space of all finite signed measures on  $\mathcal{B}(\mathcal{U}) \times \mathcal{B}(\mathcal{Z})$  and give  $L$  the weak topology induced by the linear functionals

$$\mu \rightarrow \int g \, d\mu,$$

where  $g : \mathcal{U} \times \mathcal{Z} \rightarrow \mathbb{R}^1$  is bounded and  $\mathcal{B}(\mathcal{U}) \times \mathcal{B}(\mathcal{Z})$ -measurable. The collection  $\Gamma$  of all such linear functionals is a total space of functions on  $L$  (Dunford and Schwartz, 1957, p. 418) so that, by Lemma V.3.3 of Dunford and Schwartz,  $L$  is a locally convex linear topological space in its  $\Gamma$  topology.

Let  $E$  be the singleton set  $\{R_0(dz|u)S_0(du)\}$  and let  $C = \{R_1(dz|u)m(du) : m \in \mathcal{P}(\mathcal{U})\}$ . Obviously  $E$  is a compact, convex subset of  $L$  and  $C$  is a convex subset of  $L$ . The closure  $\bar{C}$  of  $C$  is closed and convex. So, in order to apply Theorem V.2.10 from Dunford and Schwartz, we need only show that  $R_0(dz|u)S_0(du) \notin \bar{C}$ . Suppose, to the contrary, that  $R_0(dz|u)S_0(du) \in \bar{C}$ . Then there exists a net  $\mu_\alpha(d(u, z)) = R_1(dz|u)m_\alpha(du)$  of elements of  $C$  such that

$$\int g \, d\mu_\alpha \rightarrow \iint g(u, z)R_0(dz|u)S_0(du) \tag{30}$$

for all bounded, measurable  $g : \mathcal{U} \times \mathcal{Z} \rightarrow \mathbb{R}^1$ . In particular,

$$\int h(u)m_\alpha(du) \rightarrow \int h(u)S_0(du) \tag{31}$$

for all bounded, measurable  $h : \mathcal{U} \rightarrow \mathbb{R}^1$ .

Here is a lemma that will help us reach a contradiction to (30).

**Lemma A.1.** *There exist  $A \in \mathcal{B}(\mathcal{Z})$  and  $\varepsilon > 0$  such that*

$$S_0\{u|R_1(A|u) \geq R_0(A|u) + \varepsilon\} > 0.$$

**Proof.** Let  $\mathcal{S} = \{A_1, A_2, \dots\}$  be a countable generating set for the  $\sigma$ -field  $\mathcal{B}(\mathcal{Z})$  and assume, without loss of generality, that  $\mathcal{S}$  is closed under finite intersections. Then two probability measures  $p$  and  $q$  on  $\mathcal{B}(\mathcal{Z})$  are unequal if and only if for some positive integers  $i$  and  $j$ ,

$$|p(A_i) - q(A_i)| \geq 1/j.$$

Thus, by the hypothesis of the theorem, there exist  $i$  and  $j$  such that

$$S_0\{u||R_1(A_i|u) - R_0(A_i|u)| \geq 1/j\} > 0.$$

To complete the proof of the lemma, take  $\varepsilon = 1/j$  and  $A$  to be either  $A_i$  or its complement. □

Let  $A$  and  $\varepsilon$  be as in Lemma A.1 and set

$$B = \{u|R_1(A|u) \geq R_0(A|u) + \varepsilon\}.$$

Define  $g$  to be the indicator function of  $B \times A$ . Then

$$\begin{aligned} \int g \, d\mu_\alpha &= \int_B \int_A R_1(dz|u) m_\alpha(du) \\ &= \int_B R_1(A|u) m_\alpha(du) \\ &\geq \int_B R_0(A|u) m_\alpha(du) + \varepsilon m_\alpha(B) \\ &\rightarrow \int_B R_0(A|u) S_0(du) + \varepsilon S_0(B) \\ &= \iint g(u, z) R_0(dz|u) S_0(du) + \varepsilon S_0(B). \end{aligned}$$

(The convergence step above is by (31).) We have reached a contradiction to (30).

The theorem now follows immediately from Theorems V.2.10 and V.3.9 of Dunford and Schwartz.

## Acknowledgement

The first author's research was supported in part by National Science Foundation grants DMS 93-00973 and 96-26601. The second author's research was supported in part by National Science Foundation grant DMS 94-23009. The authors would like to thank the referees and the Associate Editor for a variety of constructive and insightful suggestions. These prompted the inclusion of Example 4.1.

## References

- Aitchison, J. and Dunsmore, I.R. (1975) *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.
- Andersson, S. (1982) Distributions of maximal invariants using quotient measures. *Ann. Statist.*, **10**, 955–961.
- Andersson, S., Brøns, H. and Jensen, S.T. (1983) Distribution of eigenvalues in multivariate statistical analysis. *Ann. Statist.*, **11**, 392–415.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1996) Prediction and asymptotics. *Bernoulli*, **2**, 319–340.
- Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. New York: Springer-Verlag.
- Bondar, J.V. and Milnes, P. (1981) Amenability: A survey for statistical applications of Hunt–Stein and related conditions on groups. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **57**, 103–128.
- Bourbaki, N. (1966) *Elements of Mathematics. General Topology, Part 1*. Reading, MA: Addison Wesley.
- Dunford, N. and Schwartz, J. (1957) *Linear Operators, Part 1*. New York: Interscience.

- Eaton, M.L. (1982) A method for evaluating improper prior distributions. In S.S. Gupta and J.O. Berger (eds), *Statistical Decision Theory and Related Topics III*, 1, pp. 329–352. New York: Academic Press.
- Eaton, M.L. (1983) *Multivariate Statistics*. New York: Wiley.
- Eaton, M.L. (1989) *Group Invariance Applications in Statistics*. Regional Conference Series in Probability and Statistics 1. Hayward, CA: Institute of Mathematical Statistics.
- Eaton, M.L. (1992) A statistical diptych: Admissible inferences – recurrence of symmetric Markov chains. *Ann. Statist.*, **20**, 1147–1179.
- Eaton, M.L. and Sudderth, W. (1993) Prediction in a multivariate normal setting: Coherence and incoherence. *Sankhyā Ser. A*, Special Volume **55**, Part 3, 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.*, **2**, 251–270.
- Fraser, D.A.S. (1968) *The Structure of Inference*. New York: Wiley.
- Geisser, S. (1993) *Predictive Inference: An Introduction*. New York: Chapman & Hall.
- Johnson, B.W. (1991) On the admissibility of improper Bayes inferences in fair Bayes decision problems. Ph.D. thesis, University of Minnesota.
- Kass, R. and Wasserman, L. (1996) The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.*, **91**, 1343–1370.
- Kiefer, J. (1975) Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.*, **28**, 573–601.
- Lane, D. and Sudderth, W. (1984) Coherent predictive inference. *Sankhyā, Ser. A*, **46**, 166–185.
- Stein, C. (1965) Approximation of improper prior measures by prior probability measures. In L.M. LeCam and J. Neyman (eds), *Bernoulli (1713) – Bayes (1763) – Laplace (1813)*, pp. 217–240. New York: Springer-Verlag.
- Stone, M. (1976) Strong inconsistency from uniform priors. *J. Amer. Statist. Assoc.*, **71**, 114–119.
- Wijsman, R.A. (1990) *Invariant Measures on Groups and Their Use in Statistics*. IMS Lecture Notes – Monograph Series 14. Hayward, CA: Institute of Mathematical Statistics.

Received January 1997 and revised January 1998