Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series A, Pt. 3, pp. 363-382

A NEW PREDICTIVE DISTRIBUTION FOR NORMAL MULTIVARIATE LINEAR MODELS*

By MORRIS L. EATON
and
WILLIAM D. SUDDERTH
University of Minnesota, Minneapolis

SUMMARY. Predictive distributions are calculated in the MANOVA problem for a class of relatively invariant improper prior distributions. An invariance argument then shows that many of the predictive distributions recommended in the literature are strongly inconsistent in the sense of M.Stone. In particular, this is true of the formal Bayes prediction based on the Jeffreys prior. We introduce a new predictive distribution that is not strongly inconsistent.

1. **Introduction**

The recent paper by Keyes and Levy (1996) provides an excellent overview of much of the work of the past thirty five years on the problem of prediction in the multivariate analysis of variance context. So that our discussion is easily compared to that in Keyes and Levy, their notation will be followed to the extent practicable. In the MANOVA prediction problem, a data matrix $Y : n \times r$ is available and is assumed to have a normal distribution with density given by

$$p_1(y) = (2\pi)^{\frac{-nr}{2}}|\Sigma|^{\frac{-n}{2}}exp[\tfrac{-1}{2}tr(y - X\beta)\Sigma^{-1}(y - X\beta)'] \qquad \ldots (1.1)$$

Here $X : n \times p$ is a known rank p matrix, $\beta : p \times r$ is a matrix of unknown parameters, and $\Sigma : r \times r$ is an unknown positive definite covariance matrix (the common covariance matrix of the rows of Y). Based on the data Y, the inferential goal is to obtain a predictive distribution for the unknown observable $Z : m \times r$ that is assumed to have a density

---

AMS (1991) *subject classification.* 62H05, 62H99, 62A05, 62A30.
Key words and phrases. Predictive inferences, improper prior distributions, strong inconsistency, invariance, multivariate linear models.
* Research supported in part by National Science Foundation Grants DMS 96-26601 (Eaton) and DMS 97-03285 (Sudderth)

$$p_2(z) = (2\pi)^{\frac{-mr}{2}} |\Sigma|^{\frac{-m}{2}} exp[\tfrac{-1}{2} tr(z - W\beta)\Sigma^{-1}(z - W\beta)'] \qquad \ldots (1.2)$$

where $W : m \times p$ is a known matrix, and β, Σ are the parameters in 1.1. It is assumed that Y and Z are independent, given the parameters β and Σ.

By a *predictive distribution*, $Q(\cdot|y)$, we mean a probability distribution for Z given the data $Y = y$. Thus, $Q(\cdot|y)$ is a Markov kernel. Table 1 on page 193 in Keyes and Levy (1996) lists nine different proposed predictive distributions which are specified by giving their densities. The first seven of these are of direct interest to us since most of these can be derived using an improper prior distribution and a formal Bayes argument. Furthermore, these seven proposals are all "fully invariant" - a term which we will make precise in the next few paragraphs. Some relevant references for these seven proposals are Geisser and Cornfield (1963), Geisser (1965,1971), Zellner and Chetty (1965), Kalbfleish and Sprott (1969), Lauritzen (1974), O'Reilly (1975), Murray (1977), Hinkley (1979), Levy and Perng (1984) and Butler (1986). Some other related references can be found in Keyes and Levy (1996). Using a mean Kullback-Leibler directed divergence as an optimality criterion, Keyes and Levy (1996) advocate using the predictive distribution proposed independently in Geisser (1965) and Zellner and Chetty (1965). This predictive distribution can be obtained using a formal Bayes argument when the improper prior is the so called "Jeffreys prior" for this problem.

In what follows, we will suggest an alternative predictive distribution for Z given $Y = y$, in the case $r > 1$. Our arguments are based on invariance considerations so we now turn to a careful discussion of invariance in the prediction problem described above.

Consider the group G whose elements g consist of all pairs (C, B) where C is an $r \times r$ non-singular matrix and B is an arbitrary $p \times r$ real matrix. The group operation is defined by $(C_1, B_1)(C_2, B_2) = (C_1C_2, B_1 + B_2C_1')$ where C_1' is the transpose of C_1. The group G acts on three different spaces-namely, the observation space of $n \times r$ matrices, the predictand space of $m \times r$ matrices, and the parameter space, which is all pairs (Σ, β) where Σ is an $r \times r$ positive definite matrix and β is a $p \times r$ real matrix. The action of $g = (C, B)$ on these three spaces is

$$\begin{array}{ll} y \rightarrow & yC' + XB \\ z \rightarrow & zC' + WB \\ (\Sigma, \beta) \rightarrow & (C\Sigma C', B + \beta C') \end{array} \qquad \ldots (1.3)$$

where the prime denotes transpose. It is a routine matter to check that the parametric models assumed for Y and Z are invariant under the actions specified in (1.3). It is in this sense that we say "the prediction problem is invariant under G". Recall that a predictive distribution $Q(\cdot|y)$ for Z is invariant under G *(fully* *invariant)* if for all $g \in G$, all $y \in \mathcal{Y}$, and all Borel sets E of the predictand space, we have

$$Q(gE|gy) = Q(E|y) \qquad \ldots (1.4)$$

The set $\mathcal{Y}$ is described precisely following display (1.13). It is easily verified that the first seven predictive distributions specified in Table 1 of Keyes and Levy (1996) are fully invariant.

The discussion in this paper will focus on predictive distributions which are invariant under a particular subgroup of G. To motivate this discussion, first let G_T^+ denote the group of all $r \times r$ lower triangular matrices whose diagonal elements are positive. Then given Σ which is $r \times r$ and positive definite, there exists a a unique $\theta \in G_T^+$ such that $\Sigma = \theta\theta'$. In what follows, we will take the parameter space of the prediction problem to be all pairs (θ, β) where $\theta \in G_T^+$ and β is an $r \times p$ real matrix. Next, let G_0 be the subgroup of G obtained by restricting C in (1.3) to be in G_T^+. Then the action of G_0 on both $\mathcal{Y}$ and the predictand space is given in (1.3) while the action of $g = (t, B) \in G_0$ on the parameter space is now given by

$$(\theta, \beta) \rightarrow (t\theta, B + \beta t') \qquad \ldots (1.5)$$

where $t \in G_T^+$. Since the prediction problem is G invariant, it is also G_0 invariant because G_0 is a subgroup of G.

The class of G_0 invariant predictive distributions of particular concern here are obtained from a formal Bayes calculation using certain improper prior distributions for the parameter pair (θ, β). To describe these improper priors, we now recall some standard results regarding the group G_T^+ - see Eaton (1989, Chapter 1, especially section 1.4) for details. A right invariant measure on G_T^+ is given by

$$\nu_\rho(dt) = \frac{dt}{\prod_{i=1}^r t_{ii}^{r-i+1}} \qquad \ldots (1.6)$$

where

$$t = \begin{pmatrix} t_{11} & 0 & \ldots & 0 \\ t_{21} & t_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ t_{r1} & t_{r2} & \ldots & t_{rr} \end{pmatrix}$$

is an element of G_T^+ and "dt" means Lebesque measure on the non-zero elements of t. The *modulus* of G_T^+ is

$$\Delta(t) = \prod_{i=1}^r t_{ii}^{r-2i+1} \qquad \ldots (1.7)$$

so that

$$\nu_l(dt) = \Delta(t)\nu_\rho(dt) \quad \ldots (1.8)$$

is a left invariant measure on G_T^+. Furthermore, given any real numbers $c_1, \ldots, c_r$, the function

$$\chi_c(t) = \prod_{i=1}^r (t_{ii})^{c_i}, t \in G_T^+ \quad \ldots (1.9)$$

is a *multiplier* on G_T^+. The vector c with coordinates $c_1, \ldots, c_r$ is used to index the multipliers.

Improper prior distributions for(θ, β) of the form

$$\mu_c(d\beta, d\theta) = \chi_c(\theta) d\beta\, \nu_\rho(d\theta) \quad \ldots (1.10)$$

are used below to compute G_0 invariant predictive distributions for Z given $Y = y$. However, the vector c indexing χ_c is restricted to lie in a set $C^* \subseteq R^r$ so that certain integrals are finite. The set C^* is defined by the r inequalities in (4.25). It should be observed that improper priors expressed in terms of (Σ, β) which have the form

$$|\Sigma|^\alpha d\beta d\Sigma \quad \ldots (1.11)$$

can all be written in the form (1.10) for appropriate choice of c (see Section 4 for a discussion). The point here is that all of the relatively invariant improper priors mentioned in Keyes and Levy (1996) have the form (1.11) and thus can be written in the form (1.10). In particular, the standard "Jeffreys prior" $d\beta d\Sigma/|\Sigma|^{\frac{r+1}{2}}$ corresponds to the improper prior $d\beta\, \nu_l(d\theta)$. which involves the left invariant measure on G_T^+. However, we will argue below that the only improper prior of the form (1.10) which deserves serious consideration is the improper prior $d\beta\, \nu_\rho(d\theta)$. Note that $\nu_\rho = \nu_l$ in the univariate case-i.e., r=1. Our arguments will depend on having an explicit expression for the predictive density when the improper prior is μ_c in (1.10). These predictive densities, which depend on $c \in C^*$, can be described as follows. For any $r \times r$ positive definite matrix K, let $\tau(K)$ denote the unique matrix in G_T^+ which satisfies $K = \tau(K)(\tau(K))'$.

Next, for $c \in C^*$, let

$$f_c(z) = \alpha(c)|I_r + z'z|^{\frac{(m+n-p)}{2}} \chi_c(\tau(I_r + z'z))\Delta^{-1}(\tau(I_r + z'z)) \quad \ldots (1.12)$$

be a density on $m \times r$ real matrices z where $\alpha(c)$ is a normalizing constant. The expression for $\alpha(c)$ is given in (4.28).

The arguments in Section 4 give the predictive density for Z given $Y = y$ when using the improper prior μ_c in (1.10) coupled with the formal Bayes calculation. To describe this predictive density, let

$$\begin{aligned} A = &\quad I_m + W(X'X)^{-1}W', \\ S = &\quad y'(I_n - X(X'X)^{-1}X')y, \\ \hat{\beta} = &\quad (X'X)^{-1}X'y. \end{aligned} \qquad \ldots (1.13)$$

Here, y is restricted to the set $\mathcal{Y}$ where S is positive definite (the complement of $\mathcal{Y}$ has Lebesque measure zero in the vector space of all $n \times r$ real matrices). In this notation, we have for $c \in C^*$,

$$Q_c(dz|y) = |S|^{\frac{-m}{2}} |A|^{\frac{-r}{2}} f_c(A^{\frac{-1}{2}}(z - W\hat{\beta})(\tau(S)^{-1})')dz. \qquad \ldots (1.14)$$

This density is used in Section 3 to establish one of the main results in this paper-namely, for $c \neq 0$, the predictive distribution $Q_c(dz|y)$ is strongly inconsistent, a notion we discuss in Section 2. This result shows that all predictive distributions obtained from formal Bayes arguments using improper priors of the form (1.11) are strongly inconsistent. In particular, the predictive distributions listed as numbers 3 through 7 in Keyes and Levy (1996, Table 1 on page 93) are strongly inconsistent.

In brief, this paper is organized as follows. Section 2 contains a discussion of strong inconsistency and gives a Bayesian interpretation. In Section 3, a constructive argument is given which establishes the strong inconsistency. It is also shown that $Q_0(dz|y)$ is not strongly inconsistent. Because $Q_0(dz|y)$ is the only predictive distribution among the $Q_c(dz|y)$, $c \in C^*$ which is not strongly inconsistent, we would recommend its use in practice over all other Q_c's. The detailed calculations resulting in the density of each Q_c are given in Section 4 while some technical issues are treated in an appendix.

2. Strong Inconsistency

The notion of strong inconsistency (SI), introduced by Stone(1976) in an "estimation" context, is a rather undesirable property of an inference. In the prediction context, SI takes the following form. Consider a joint probability model $P(dy, dz|\theta)$ for variables $y \epsilon \mathcal{Y}$ and $z \epsilon \mathcal{Z}$ where $\theta \in \Theta$ is an unknown parameter. The random variable $Y \epsilon \mathcal{Y}$ is to be observed and $Z \epsilon \mathcal{Z}$ is to predicted based on the observed value of Y. By a *predictive inference* we mean a Markov Kernel $Q(dz|y)$ where $Q(\cdot|y)$ is a probability distribution for Z which depends on y (interpreted as an inference about Z after seeing $Y = y$).

DEFINITION 2.1. The predictive inference $Q(dz|y)$ is *strongly inconsistent* (SI) if there exists an $\epsilon > 0$ and a measurable function $k_0(y, z)$ with values in [-1,1] such that

$$\inf_y \int_{\mathcal{Z}} k_0(y, z) Q(dz|y) \geq \epsilon + \sup_\theta E_\theta k_0(Y, Z) \qquad \ldots (2.1)$$

When (2.1) holds, the inference $Q(\cdot|y)$ does not "fit together" with the given model. The left side of (2.1) is the smallest possible expectation of k_0 (under the $Q(\cdot|y)$'s) that one could obtain while the right side of (2.1) is the largest expectation of k_0 (under the model) that one could obtain. Of course, (2.1) implies there is an inconsistency. A Bayesian interpretation of (2.1) is quite interesting. Let π be a prior distribution for θ so that under this prior, the marginal distribution of (Y, Z) is

$$P_\pi(dy, dz) = \int_\Theta P(dy, dz|\theta)\pi(d\theta).$$

The convex set $C_1 = \{P_\pi \,|\pi \text{ is a prior}\}$ is the set of all possible marginal models a Bayesian could contemplate for (Y, Z) based on the given parametric model. Next, let $M_1(\mathcal{Y})$ be all the probability measures on $\mathcal{Y}$ and set

$$C_2 = \{s(dy, dz) = Q(dz|y)m(dy) \,|\, m \in M_1(\mathcal{Y})\}.$$

The convex set C_2 is the collection of all possible joint distributions for (Y, Z) that are consistent with the predictive inference $Q(\cdot|y)$.

Given any bounded signed measure σ on $\mathcal{Y} \times \mathcal{Z}$, recall that the variation norm of σ is defined by

$$||\sigma|| = \sup_{|f|\le 1} |\int f d\sigma| \qquad \ldots (2.2)$$

where

$$\int f d\sigma = \int_{\mathcal{Y}} \int_{\mathcal{Z}} f(y, z)\sigma(dy, dz).$$

In particular, if $s_i \in C_i$, $i = 1, 2$, then $\|s_1 - s_2\|$ is the variation distance between s_1 and s_2.

PROPOSITION 2.1. *If an inference $Q(\cdot|y)$ is SI in the sense of Definition 2.1, then*

$$\inf_{s_1\in C_1} \inf_{s_2\in C_2} ||s_1 - s_2|| \ge \epsilon \qquad \ldots (2.3)$$

where ϵ is given in (2.1).

PROOF. With k_0 and ϵ as in (2.1), consider $s_i \in C_i, i = 1, 2$. Then

$$\begin{aligned}\int\int k_0(y, z)s_2(dy, dz) &\ge \inf_y \int k_0(y, z)Q(dz|y) \\ &\ge \epsilon + \sup_\theta \int\int k_0(y, z)P(dy, dz|\theta) \\ &\ge \epsilon + \int\int k_0(y, z)s_1(dy, dz).\end{aligned}$$

Hence,

$$\int k_0 d(s_2 - s_1) \geq \epsilon$$

so that

$$||s_2 - s_1|| \geq \epsilon.$$

Therefore (2.3) holds and the proof is complete.

Inequality (2.3) makes precise the claim that a strongly inconsistent $Q(\cdot|y)$ does not "fit together" with the model. In particular, there is no element of C_2 (a joint distribution for Y and Z obtained from $Q(\cdot|y)$) which would be plausible for a Bayesian since all possible Bayesian joint distributions are in C_1. Inequality (2.3) implies that the elements of C_2 are uniformly bounded away from the elements in C_1 (in variation distance) by at least ϵ.

One possible criticism of (2.1) as a criterion for exclusion of a predictive inference is that the function k_0 in (2.1) may not be of any particular interest for the problem at hand. For this reason, it is useful to look at (2.1) more closely-especially in the case of invariant inferences. The argument in the next section does this for the MANOVA problem, but it is rather instructive to consider a simple invariant example here so as to isolate the essential features of the argument.

Example 2.1. Suppose that given θ, Y and Z are independent $N(\theta, 1)$ and $\theta \in R^1$. The real valued random variable Z is to be predicted based on the observed value of Y, say $y \in R^1$. This problem is invariant under translations ($Y \rightarrow Y + c, Z \rightarrow Z + c$, and $\theta \rightarrow \theta + c$ where $c \in R^1$). Based on the model $Z - Y$ is $N(0, 2)$ so an obvious predictive inference is $Q_2(\cdot|y) = N(y, 2)$. Now, suppose $Q(\cdot|y)$ is an invariant inference-that is,

$$Q(B|y) = Q(B + c|y + c)$$

for all Borel sets B, all $y \in R^1$ and $c \in R^1$. Setting $c = -y$, we see

$$Q(B|y) = Q(B - y|0) \qquad \ldots (2.4)$$

Using (2.4) it is easy to show that for any bounded function h and for all y,

$$\int h(z - y)Q(dz|y) = \int h(t)Q(dt|0).$$

However, the model assumption implies that

$$E_\theta h(Z - Y) = \int h(t)Q_2(dt|0) \qquad \ldots (2.5)$$

Thus, for the invariant inference Q, inequality (2.1) will hold for any $h, |h| \leq 1$, for which

$$\int h(t)Q(dt|0) > \int h(t)Q_2(dt|0).$$

(just set $k_0(y,z) = h(z-y)$). Such an h will always exist as long as $Q(\cdot|0) \neq Q_2(\cdot|0)$. Therefore in this example there is only one possible invariant inference which can be *consistent* (not SI). The arguments in ES (1996) show that $Q_2(\cdot|y)$ is consistent. Furthermore, there is an $h, |h| \leq 1$, which yields an ϵ equal to the variation distance between $Q(\cdot|0)$ and $Q_2(\cdot|0)$. The particular case when $Q_a(\cdot|y) = N(y,a)$ with $a \neq 2$ is of special interest. Note that $Q_a(\cdot|y)$ is invariant and $Q_a(\cdot|0)$ has a density

$$g_a(t) = \frac{1}{\sqrt{2\pi a}}\, exp\,[-\frac{1}{2a}t^2].$$

Let $B = \{t|g_a(t) > g_2(t)\}$ so that B is a symmetric interval about 0 or the complement of such an interval. With $k_0(y,z) = I_B(z-y) - I_{B^c}(z-y)$, inequality (2.1) holds with ϵ equal to $||Q_2(\cdot|0) - Q_a(\cdot|0)||$. Obviously such k_0's are of interest in this problem since the probability that $''|Z-y| \leq b''$ is a quantity of direct predictive interest. This completes the example.

3. Evaluating the Predictive Inferences

We now turn to an evaluation of the predictive distributions $Q_c(dz|y)$. given by (1.14) for $c \in C^*$. Our first task is to show that for $c \neq 0$, the predictive inference $Q_c(dz|y)$ is SI.

For any observation matrix $y \in \mathcal{Y}$, the residual sum of squares matrix

$$S = y'(I_n - X(X'X)^{-1}X')y \qquad \ldots(3.1)$$

is non-singular, and

$$\hat{\beta} = (X'X)^{-1}X'y \qquad \ldots(3.2)$$

is the maximum likelihood estimator of β based on the model for the observed data. Given S and $\hat{\beta}$, let

$$\gamma(y) = (\tau(S), \hat{\beta}) \qquad \ldots(3.3)$$

be an element of the group G_0. It is easily verified that

$$\gamma(gy) = g \circ \gamma(y)\,,\; for\, g \in G_0, \qquad \ldots(3.4)$$

where $''\circ''$ denotes group multiplication. For this reason, the statistic

$$U = (\gamma(Y))^{-1}Z \, : m \times r \qquad \ldots (3.5)$$

is an ancillary statistic. Here, $(\gamma(Y))^{-1}$ denotes the group inverse of $\gamma(Y)$, and U is the value of $(\gamma(Y))^{-1}$ at Z (see(1.3) for the action of G_0 on $z's$).

PROPOSITION 3.1. *Under the assumed model for Y and Z, the density function of U is given by*

$$|A|^{-\frac{r}{2}} f_0(A^{-\frac{1}{2}}u) \qquad \ldots (3.6)$$

where

$$A = I_m + W(X'X)^{-1}W' \qquad \ldots (3.7)$$

and f_0 is given in (1.12) when $c = 0$.

PROOF. This is a relatively routine multivariate calculation similar to those sketched in Section 4. The details are omitted.

Next consider a bounded function k defined on $m \times r$ real matrices.

PROPOSITION 3.2. *For each $y \in Y$ and $c \in C$,*

$$\begin{aligned} &\int \quad k((\gamma(y))^{-1}z)Q_c(dz|y) = \\ &\int \quad k(u)|A|^{-\frac{r}{2}} f_c(A^{-\frac{1}{2}}u)du \end{aligned} \qquad \ldots (3.8)$$

where A is given in (3.6) and f_c is given in (1.12).

PROOF. The identity (3.8) follows immediately from the expression (1.14) for the predictive distribution $Q_c(dz|y)$.

Propositions 3.1 and 3.2 together with the explicit form for f_c now allow us to specify a particular function k_0 and an $\epsilon > 0$ satisfying (2.1) when $c \neq 0$. To this end, fix $c \neq 0$ and let

$$B = \{u|f_c(A^{-\frac{1}{2}}u) > f_0(A^{-\frac{1}{2}}u)\} \qquad \ldots (3.9)$$

Next, set

$$k_0((\gamma(y))^{-1}z) = I_B((\gamma((y))^{-1}z). \qquad \ldots (3.10)$$

THEOREM 3.3. *With k_0 given by (3.10),*

$$\inf_y \int k_0((\gamma(y))^{-1}z)Q_c(dz|y) - \sup_{\theta,\beta} E_{\theta,\beta}k_0((\gamma(Y))^{-1}Z) = \epsilon_c \qquad \ldots (3.11)$$

where ϵ_c is the variation distance between the probability measures defined by the densities f_c and f_0.

PROOF. A direct application of Propositions (3.1) and (3.2) show that the left hand side of (3.11) is equal to

$$d = \int k_0(u)|A|^{-\frac{r}{2}}(f_c(A^{-\frac{1}{2}}u) - f_0(A^{-\frac{1}{2}}u))du.$$

The choice of k_0 in (3.10) is well known to yield a value of d equal to the variation distance between distributions determined by $|A|^{-\frac{r}{2}}f_c(A^{-\frac{1}{2}}u)$ and $|A|^{-\frac{r}{2}}f_0(A^{-\frac{1}{2}}u)$. This is equal to the claimed value of ϵ_c since variation distance is invariant under one-to-one maps.

Of course Theorem (3.3) establishes the strong inconsistency of Q_c for $c \neq 0$. However the argument yielding this result contains a bit more information. The pivotal quality $U = (\tau(Y))^{-1}Z$ has the density $|A|^{\frac{-m}{2}}f_0(A^{\frac{-1}{2}}u)$ under the model. But equation (3.8) shows that, in some sense, U is assigned the "wrong" distribution under each Q_c, for $c \neq 0$. It is exactly this happenstance that establishes the strong inconsistency of $Q_c, c \neq 0$.

Next, we turn to a discussion of the invariant predictive inference $Q_0(dz|y)$ which arises from the use of the right Haar measure on G_0 as an improper prior.

THEOREM 3.4. *The predictive inference $Q_0(dz|y)$ is consistent (not SI).*

PROOF. We apply some recent results in ES (1996). First, it is well known that the group G_0 is amenable (see Bondar and Milnes (1961, p.114) for a discussion of amenability). The other assumptions of Theorem 8.1 in ES (1996) are easily checked. This theorem implies that Q_0 is consistent.

The results in this section show that if $c \neq 0$, then the predictive distribution $Q_c(\cdot|y)$ is SI while $Q_0(\cdot|y)$ is consistent. Furthermore $Q_0(\cdot|y)$ is a formal Bayes predictive distribution in the sense that it can be obtained as a formal posterior distribution by using the improper prior distribution

$$\mu_0(d\beta, d\theta) = d\beta\nu_\rho(d\theta). \qquad \ldots(3.12)$$

for the parameters β and θ. Since β is a "translation parameter", the statistical interpretation of the $d\beta$ portion of (3.12) is fairly clear–namely, $d\beta$ is the translation invariant improper prior for β. However the statistical interpretation of the $\nu_\rho(d\theta)$ portion of (3.12) is somewhat less clear in part because θ is the lower triangular square root of the covariance matrix. The following remark gives one possible statistical interpretation of $\nu_\rho(d\theta)$ in terms of standard statistical parameters.

REMARK 3.1. Let Δ be an r-dimensional random vector with coordinates $\delta_1, \delta_2, \ldots, \delta_r$, and assume Δ is $N_r(0, \Sigma)$. For each $i = 1, \ldots, r-1$ let $\delta_{(i)}$ be the vector with coordinates $\delta_1, \ldots, \delta_i$. As usual, write $\Sigma = \theta\theta'$ with $\theta \in G_T^+$. Our purpose here is to interpret the improper prior distribution $\nu_\rho(d\theta)$ in terms of regression parameters and conditional variances of the coordinates of Δ. To this end, we first describe the joint distribution of $\delta_1, \delta_2, \ldots, \delta_r$ by specifying the distribution of δ_1, and then specifying the conditional distribution of δ_i given $\delta_{(i-1)}$ for $i = 2, \ldots, r$. Let θ_{ii} be the i^{th} diagonal element of θ and let $\theta_{(ii)}$ be the $i \times i$ upper left hand corner of θ, for $i = 1, \ldots, r$. It is well known that the conditional variance of δ_i given $\delta_{(i-1)}$ is θ_{ii}^2, for $i = 2, \ldots, r$ while θ_{11}^2 is the variance of δ_1. Because of the normality assumption, the conditional distribution of δ_i given $\delta_{(i-1)}$ is $N_1(\alpha_{(i-1)}'\delta_{(i-1)}, \theta_{ii}^2)$ for $i = 2, \ldots, r$ where $\alpha_{(i-1)}$ is an $(i-1)$ dimensional vector of regression coefficients. It is not hard to show that

$$\alpha_{(i-1)} = \theta_{(i-1,i-1)}^{-1}\theta_{(i)} \qquad \ldots (3.13)$$

where $\theta_{(i)}$ is an $(i-1)$ dimensional vector whose coordinates are the first $(i-1)$ elements of the i^{th} row of θ.

In terms of the parameters in $A = \{\alpha_{(i-1)}, \theta_{ii} | i = 2, \ldots, r\} \cup \{\theta_{11}\}$, consider the measure (improper prior)

$$\left(\prod_{i=2}^{r} d\alpha_{i-1}\right) \prod_{i=1}^{r} \frac{d\theta_{ii}}{\theta_{ii}}. \qquad \ldots (3.14)$$

Under this measure, the $\frac{r(r-1)}{2}$ regression parameters have Lebesque measure as their prior distribution while each conditional variance has the standard flat prior $\frac{d\theta_{ii}}{\theta_{ii}}$, for $i = 1, 2, \ldots, r$. Now, it is not hard to show (by induction) that the improper prior distribution $\nu_\rho(d\theta)$, expressed in terms of θ, is transformed into the improper prior (3.14) under the mapping (3.13). Thus the interpretation of ν_ρ is relatively straightforward in terms of the regression coefficients and conditional variances. This ends Remark 3.1.

4. Some Predictive Densities in MANOVA

In this section we give a derivation of the predictive density in a MANOVA model for a wide class of relatively invariant improper prior distributions. Recall that if Y and Z have a joint density $p(y, z|\lambda)$ where λ is a parameter, when the prior distribution is π (proper or improper), the predictive density of Z given Y=y is calculated as

$$q(z|y) = \frac{\int_\Theta p(y, z|\lambda)\pi(d\lambda)}{\int_\Theta \int_z p(y, u|\lambda) du\, \pi(d\lambda)} \qquad \ldots (4.1)$$

as long as both integrals are finite and the denominator is positive. Our task here is to calculate (4.1) in the MANOVA case when π has the form (1.10) described in Section 1.

In the case of the MANOVA problem considered below, Y and Z are conditionally independent given λ. In this case, the denominator (4.1) is just

$$\int_\Theta p(y|\lambda)\pi(d\lambda)$$

where $p(\cdot|\lambda)$ is the marginal density of Y.

The Model. The model assumptions described in Section 1 imply that the pair Y, Z has a joint multivariate normal distribution on $(m+n)\times r$ real matrices. The mean matrix is

$$E\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} X \\ W \end{pmatrix}\beta$$

and the "covariance" is

$$Cov\begin{pmatrix} Y \\ Z \end{pmatrix} = I_{m+n}\otimes(\theta\theta').$$

That is, the rows of the random matrix $\begin{pmatrix} Y \\ Z \end{pmatrix}$ are independent and each row has covariance matrix $\theta\theta'$ with $\theta\epsilon G_T^+$. Thus the joint model can be written

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim N(\begin{pmatrix} X \\ W \end{pmatrix}\beta, I_{m+n}\otimes\theta\theta') \qquad \ldots(4.2)$$

Invariance considerations. The focus in this and the remaining sections will be the invariance of the model (4.2) under the group G_0 (the terminology from Eaton (1989) is used below.) The action of G_0 on sample points specified in (1.3) while the action of G_0 on a parameter pair (θ, β) is given by (1.5). Observe that parameter pairs (θ, β) are in a natural 1-1 correspondence with elements of G_0 and the group action (1.5) is just composition in G_0. For this reason we often write $(\theta, \beta)\in G_0$. When expressed in terms of (θ, β), the (relatively invariant) improper priors considered here have the form

$$\mu_c(d\beta, d\theta) = \chi_c(\theta)d\beta\,\nu_\rho(d\theta), \qquad \ldots(4.3)$$

as discussed in Section 1. It is not difficult to show that all the relatively invariant measures on G_0 have the form (4.3) (up to a constant) (see Eaton (1989, Section 1.4) for a discussion).

From the prediction perspective, most of the improper prior distributions considered in the literature (see Keyes and Levy (1996) for a review) have the form (when written in the (Σ, β) parameterization)

$$|\Sigma|^\alpha d\beta\frac{d\Sigma}{|\Sigma|^{\frac{r+1}{2}}} \qquad \ldots(4.4)$$

for various choices of $\alpha\in R^1$. The choice $\alpha = 0$ gives what is commonly called the Jeffreys prior for the MANOVA prediction problem. However all priors of the form (4.4) can be written in the form (4.3) (up to a constant). To see this, recall that

$$\frac{d\Sigma}{|\Sigma|^{\frac{r+1}{2}}} = 2^r \nu_l(d\theta) \qquad \ldots (4.5)$$

where ν_l is a left Haar measure on G_T^+ given in (1.8). The equality in (4.5) means that

$$\int_{S_r^+} f(\Sigma)\frac{d\Sigma}{|\Sigma|^{\frac{r+1}{2}}} = 2^r \int_{G_T^+} f(\theta\theta')\nu_l(d\theta) \qquad \ldots (4.6)$$

for all non-negative measureable f. Here S_r^+ is the set of all $r \times r$ positive definite matrices. A bit of algebra now shows that

$$|\Sigma|^{\alpha} d\beta \frac{d\Sigma}{|\Sigma|^{\frac{(r+1)}{2}}} = 2^r \chi_c(\theta) d\beta \nu_\rho(d\theta) \qquad \ldots (4.7)$$

with $c_i = 2\alpha + r - 2i + 1, i = 1, \ldots, r$. However, this argument also shows that a prior of the form (4.3) can be written in the form (4.4) iff for some number α, $c_i = 2\alpha + r - 2i + 1$ for $i = 1, \ldots, r$. In particular, the prior

$$d\beta\nu_\rho(d\theta) \qquad \ldots (4.8)$$

which corresponds to $c = 0$ cannot be written in the form (4.4). The prior (4.8) is a right Haar measure on the group G_0.

Two Lemmas. In this section we give two technical lemmas involving the evaluation of integrals over G_T^+ and G_0.

Lemma 4.1. *Given $c \in R^r$ and $s \in R^1$,*

$$\gamma_0(s, c) = \int_{G_T^+} |\theta|^s exp[\frac{-1}{2} tr(\theta\theta')^{-1}]\chi_c(\theta)\nu_\rho(d\theta) \qquad \ldots (4.9)$$

is finite iff

$$s + c_i + i - 1 < 0, \qquad \textit{for } i = 1, \ldots, r. \qquad \ldots (4.10)$$

When (4.9) is finite,

$$\gamma_0(s, c) = (2\pi)^{\frac{r(r-1)}{4}} \prod_{i=1}^{r} [\Gamma(\frac{-s - c_i - i + 1}{2}) 2^{(-s-c_i-i+1)}] \qquad \ldots (4.11)$$

Proof. This is a standard multivariate calculation. The details are omitted.

Now, let k, p and r be positive integers with $k \geq p + r$. Fix a matrix $T : k \times p$ with rank p and consider a random matrix $V : k \times r$ whose distribution is

$$N(T\beta, I_k \otimes (\theta, \theta')) \qquad \ldots (4.12)$$

Here, $\theta \in G_T^+$ and β is a $p \times r$ matrix of parameters. Thus the mean of V is $T\beta$ and the rows of V are independent each with a covariance matrix $\theta\theta'$. The density of V is

$$p(v|\theta, \beta) = |\theta|^{-k} h[tr(v - T\beta)(\theta\theta')^{-1}(v - T\beta)'] \qquad \ldots (4.13)$$

where for $t \in R^1$,

$$h(t) = (\sqrt{2\pi})^{-kr} exp[-\frac{t}{2}] \qquad \ldots (4.14)$$

The next result gives the value of

$$\psi_c(v) = \int\int p(v|\theta, \beta)\chi_c(\theta) d\beta \nu_\rho(d\theta) \qquad \ldots (4.15)$$

where the double integral is over $M_{p,r} \times G_T^+$ and $M_{p,r}$ is the set of all $p \times r$ real matrices. The following notation is useful in describing $\psi_c(v)$. Since T has rank p, the maximum likelihood estimator of β is

$$\tilde{\beta} = (T'T)^{-1}T'v \qquad \ldots (4.16)$$

and the $r \times r$ matrix

$$\tilde{S} = (v - T\tilde{\beta})'(v - T\tilde{\beta}) \qquad \ldots (4.17)$$

has rank r except for a set of v's of Lebesque measure zero (this set of meaure zero is ignored in what follows). Next, let $\tau \in G_T^+$ be the unique matrix which satisfies

$$\tilde{S} = \tau\tau' \qquad \ldots (4.18)$$

Lemma 4.2. *In the above notation, (4.15) is finite iff*

$$p + c_i + i - 1 < k, \qquad \text{for } i = 1, \ldots, k. \qquad \ldots (4.19)$$

When (4.15) is finite, it is given by

$$\psi_c(v) = \gamma_1 |T'T|^{-\frac{r}{2}} |\tau|^{p-k} \chi_c(\tau) \Delta^{-1}(\tau) \qquad \ldots (4.20)$$

where the constant γ_1 *is*

$$\gamma_1 = (\sqrt{2\pi})^{-rp} \gamma_0(-k + p, c) \qquad \ldots (4.21)$$

and $\gamma_0(-k + p, c)$ *is given in (4.11).*

Proof. Some routine algebra shows that

$$tr(v - T\beta)(\theta\theta')^{-1}(v - T\beta)' = tr\tilde{S}(\theta\theta')^{-1} + tr(\beta - \tilde{\beta})'T'T(\beta - \tilde{\beta})(\theta\theta')^{-1}.$$

Setting $\xi = (T'T)^{\frac{1}{2}}(\beta - \tilde{\beta})(\theta')^{-1}$ (as a transformation from β to ξ) and computing the Jacobian yields

$$\psi_c(v) = |T'T|^{-\frac{r}{2}} \int |\theta|^{p-k} h[tr\tilde{S}(\theta\theta')^{-1} + tr\xi'\xi]\chi_c(\theta)\, d\xi \nu_\rho(d\theta)$$

Next, the change of variable $\eta = \tau^{-1}\theta$ (from θ to η) yields

$$\psi_c(v) = |T'T|^{-\frac{r}{2}}|\tau|^{p-k}\chi_c(\tau)\Delta^{-1}(\tau)\gamma_1$$

where the constant γ_1 is

$$\gamma_1 = \int\int |\theta|^{p-k} h[tr(\theta\theta')^{-1} + tr\xi'\xi]\chi_c(\theta) d\xi \nu_\rho(d\theta)$$

Performing the ξ-integration and using Lemma 4.1 yields

$$\gamma_1 = (\sqrt{2\pi})^{-rp}\gamma_0(p-k, c)$$

where γ_0 is given in (4.11). This completes the proof.

The predictive density. Finally we are ready to compute the predictive density (4.1) when the model is (4.2) and the improper prior has the form (4.3). First a bit of notation is needed. Given the model (4.2), there are two relevant "residual sums of squares"–namely

$$S^* = \begin{pmatrix} Y \\ Z \end{pmatrix}(I_{m+n} - P_{x,w})\begin{pmatrix} Y \\ Z \end{pmatrix}$$

and

$$S = Y'(I_n - P_x)Y \qquad \ldots (4.22)$$

where the two orthogonal projections $P_{x,w}$ and P_x are defined in the Appendix. According to Lemma A.2 in the Appendix, the following identity holds (except for a null set):

$$S^* = S + V(z)$$

where

$$V(z) = (z - W\hat{\beta})'(I_m + W(X'X)^{-1}W')^{-1}(z - W\hat{\beta}) \qquad \ldots (4.23)$$

and $\hat{\beta}$ is given by (1.13). Recall that, given any positive definite matrix $K : r \times r$, $\tau(K)$ denotes the unique element in G_T^+ which satisfies

$$K = \tau(K)(\tau(K))' \qquad \ldots (4.24)$$

THEOREM 4.3. *Assume the model (4.2) for the data Y and the predictand Z. Consider an improper prior (4.3) for a vector $c \in R^r$. If*

$$c_i + i - 1 < n \;\; for \;\; i = 1, \ldots, r \qquad \ldots (4.25)$$

then both of the integrals defining the predictive density in (4.1) are finite. When (4.25) holds, the predictive density of Z given $Y = y\,(y \in Y)$ is

$$q_c(z|y) = \alpha(c)|A|^{\frac{-r}{2}}|S|^{\frac{-m}{2}}|U(z)|^{\frac{-(m+n-p)}{2}}\chi_c(\tau(U(z)))\Delta^{-1}(\tau(U(z))) \qquad \ldots (4.26)$$

where

$$U(z) = I_r + (\tau(S))^{-1}V(z)(\tau(S)')^{-1} \qquad \ldots (4.27)$$

and

$$\alpha(c) = \gamma_0(-n-m+p,c)/\gamma_0(-n+p,c) \qquad \ldots (4.28)$$

Here, A and S are given in (1.13), $V(z)$ is given in (4.23) and γ_0 is given in (4.11).

Proof. The first assertion is an easy consequence of Lemma 4.1. Next, apply Lemma 4.2 to the numerator and denominator of (4.1) to see that

$$q_c(z|y) = \frac{\gamma_0(-n-m+p,c)}{\gamma_0(-m+p,c)}\frac{|X'X+W'W|^{-\frac{r}{2}}}{|X'X|^{-\frac{r}{2}}} \times \frac{|\tau(S^*)|^{p-(n+m)}\chi_c\left(\tau(S^*)\right)\Delta^{-1}\left(\tau(S^*)\right)}{|\tau(S)|^{p-n}\chi_c\left(\tau(S)\right)\Delta^{-1}\left(\tau(S)\right)}$$

Now, observe that

$$\frac{|X'X+W'W|^{\frac{-r}{2}}}{|X'X|^{\frac{-r}{2}}} = |I_m + W(X'X)^{-1}W^1|^{\frac{-r}{2}}$$

and, with $Z = z$,

$$\tau(S^*) = \tau(S)\tau(U(z)) \qquad \ldots (4.29)$$

Based on these two identities, the expression for $q_c(z|y)$ follows. This completes the proof.

The predictive density of Z given $Y = y$ obtained from the model and the improper prior (4.3) has the following interpretation. Consider a random matrix $H : m \times r$ which has the density $f_c(h)$ defined in (1.12). An easy change of variable argument shows that Z given $Y = y$ has the same distribution as

$$A^{\frac{1}{2}}H(\tau(S))' + W\hat{\beta} \qquad \ldots (4.30)$$

In other words, Z is an affine transformation (which depends on the design matrices and the data) of H that has $f_c(h)$ as its density.

Two special cases of this predictive density are of particular interest. First, when $\chi_c(\theta) = \Delta(\theta)$, then the density of H is the standard matrix-t density (see

Dickey (1966)). In this case the improper prior distribution is $\Delta(\theta)\, d\beta\, \nu_r(d\theta)$ which is just the usual Jeffreys prior $d\beta d\Sigma/|\Sigma|^{\frac{(r+1)}{2}}$. This is also the improper prior distribution $d\beta\, \nu_l(d\theta)$ which is the *left* invariant measure on G_0. In this case the predictive distribution is strongly inconsistent. The second case of special interest is when $\chi_c(\theta) = 1$ so the improper prior is just the *right* invariant measure $d\beta\nu_\rho(d\theta)$ on G_0. In this case the predictive distribution is consistent. Indeed the arguments in Section 3 show that for any $c \neq 0$ (i.e., $\chi_c \neq 1$), the predictive distribution is strongly inconsistent.

Appendix

The design matrix $\begin{pmatrix} X \\ W \end{pmatrix}$ in the MANOVA model is of full rank since X is assumed to be of full rank. Thus

$$P_{x,w} = \begin{pmatrix} X \\ W \end{pmatrix} \left[\begin{pmatrix} X \\ W \end{pmatrix}' \begin{pmatrix} X \\ W \end{pmatrix} \right]^{-1} \begin{pmatrix} X \\ W \end{pmatrix}' : (n+m) \times (n+m)$$

is the orthogonal projection onto the column space of $\begin{pmatrix} X \\ W \end{pmatrix}$. Similarly,

$$P_x = X(X'X)^{-1}X' : (n \times n)$$

is the orthogonal projection onto the column space of X.

Consider the partitioned matrix

$$H = (-W(X'X)^{-1}X' \;\; I_m) : m \times (m+n)$$

and the matrix

$$C = (I_m + W(X'X)^{-1}W')^{-1}$$

LEMMA A.1. *The following identity holds:*

$$I_{n+m} - P_{x,w} = \begin{pmatrix} I_n - P_x & 0 \\ 0 & 0 \end{pmatrix} + H'CH \qquad \ldots (A.1)$$

PROOF. First observe that

(i) $HH' = C^{-1}$

(ii) $\begin{pmatrix} I_n - P_x & 0 \\ 0 & 0 \end{pmatrix} H' = 0$

(iii) $H \begin{pmatrix} X \\ W \end{pmatrix} = 0$

so $P_{x,w}H' = 0$. Equation (i) shows that $H'CH$ is an orthogonal projection. Equation (ii) shows that the two orthogonal projections $H'CH$ and

$$\begin{pmatrix} I_n - P_x & 0 \\ 0 & 0 \end{pmatrix}$$

are mutually orthogonal. Thus the right hand side of (A.1) is an orthogonal projection of rank $(n-p)+m$. However the right hand side of (A.1) is also a rank $n+m-p$ orthogonal projection. But a vector $\alpha \in R^{n+m}$ is in the null space of $I_{n+m} - P_{x,w}$ iff α has the form

$$\alpha = \begin{pmatrix} X \\ W \end{pmatrix} \delta$$

for some $\delta \in R^p$. For such α's, it is easy to show that

$$H'CH\alpha = 0$$

and

$$\begin{pmatrix} I_n - P_x & 0 \\ 0 & 0 \end{pmatrix} \alpha = 0$$

Thus the null space of the projection $I_{n+m} - P_{x,w}$ is contained in the null space of the projection on the right hand side of (A.1). Since the projections on both sides of (A.1) have the same rank, they are equal.

For the MANOVA model of (4.2), the above Lemma has the following implication. Since

$$\beta^* = \left[\begin{pmatrix} X \\ W \end{pmatrix}' \begin{pmatrix} X \\ W \end{pmatrix} \right]^{-1} \begin{pmatrix} X \\ W \end{pmatrix}' \begin{pmatrix} Y \\ Z \end{pmatrix} \qquad \ldots (A.2)$$

is the maximum likelihood estimator of β based on $\begin{pmatrix} Y \\ Z \end{pmatrix}$ and the model(4.2), the $r \times r$ matrix

$$\begin{aligned} S^* &= \left[\begin{pmatrix} Y \\ Z \end{pmatrix} - \begin{pmatrix} X \\ W \end{pmatrix} \beta^* \right]' \left[\begin{pmatrix} Y \\ Z \end{pmatrix} - \begin{pmatrix} X \\ W \end{pmatrix} \beta^* \right] \\ &= \begin{pmatrix} Y \\ Z \end{pmatrix}' (I_{n+m} - P_{x,w}) \begin{pmatrix} Y \\ Z \end{pmatrix} \end{aligned} \qquad \ldots (A.3)$$

is the unnormalized residual sum of squares. Based only on the data Y, the maximum likelihood estimator for β is

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \ldots (A.4)$$

and the $r \times r$ matrix

$$\begin{aligned} S &= (Y - X\hat{\beta})'(Y - X\hat{\beta}) \\ &= Y'(I_n - P_x)Y \end{aligned} \qquad \ldots (A.5)$$

is the unnormalized residual sum of squares based only on Y.

LEMMA A.2. *In the above notation,*

$$S^* = S + (Z - W\hat{\beta})'(I_m + W(X'X)^{-1}W')^{-1}(Z - W\hat{\beta}) \qquad \ldots (A.6)$$

PROOF. Some routine algebra shows that (A.6) is just a rewrite of the equation

$$\begin{aligned} &\begin{pmatrix} Y \\ Z \end{pmatrix}' (I_{m+n} - P_{x,w}) \begin{pmatrix} Y \\ Z \end{pmatrix} \\ &= \begin{pmatrix} Y \\ Z \end{pmatrix}' \begin{pmatrix} I_n - P_x & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix} + \begin{pmatrix} Y \\ Z \end{pmatrix}' H'CH \begin{pmatrix} Y \\ Z \end{pmatrix} \end{aligned} \qquad \ldots (A.7)$$

However, (A.7) is a direct consequence of (A.1) established in Lemma A.1.

References

BONDAR, J. AND MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups. *Warksheinlichkeitstheorie verw. Gebiete*, **57**, 103–128.

BUTLER, R. (1986). Predictive likelihood inference with applications (with discussion). *Jour. Royal Statist. Soc.,* Ser. B, **48**, 1–38.

DICKEY, J. (1966). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.*, **38**, 511–518.

EATON, M. (1989). Invariance applications in statistics. In *Regional Conference Series in Probability and Statistics I*, IMS, Hayward, CA.

EATON, M. AND SUDDERTH, W. (1996). Consistency and strong inconsistency of group invariant predictive inferences. *Bernoulli*, to appear.

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

GEISSER, S. (1971). The inferential use of predictive distributions. In Godambe, V. and Sprott, D., editors, *Foundations of Statistical Inferences*, pages 459–469. Holt, Rinehart, and Winston, Toronto.

GEISSER, S. AND CORNFIELD, J. (1963). Posterior distributions for multivariate normal classifications. *Jour. Royal Statist.*, **25**, 368–376.

HINKLEY, D. (1979). Predictive likelihood. *Ann. Statit.*, **7**, 718–728.

KALBFLEISCH, S. AND SPROTT, D. (1969). *Applications of Likelihood and Fiducial Probability to Sampling Finite Populations.* eds. N. Johnson and H. Smith Wiley, New York.

KEYES, T. AND LEVY, M. (1996). Goodness of prediction fit for multivariate linear models. *Jour. Amer. Statist. Soc.*, **91**, 191–197.

LAURITZEN, S. (1974). Sufficiency prediction and extreme models. *Scandinavian Journal of Statistics*, **1**, 128–134.

LEVY, M. AND PERNG, S. (1984). A maximum likelihood prediction function for the linear model with consistency results. *Communications in Statistics, Part A–Theory and Methods.*, **13**, 1257–1273.

MURRAY, G. (1977). A note on the estimation of probability density functions. *Biometrika*, **64**, 150–152.

O'Reilly, F. (1975). On a criterion for extrapolation in normal regression. *Ann. Statist.*, **3**, 219–222.

Stone, M. (1976). Strong inconsistency from uniform priors. *Jour. Amer. Statist. Assoc.*, **71**, 114–119.

Zellner, A. and Chetty, V. (1965). Prediction and decision problems in regression models from a bayesian point of view. *Jour. Amer. Statist. Assoc.*, **60**, 608–616.

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