BAYESIAN ESTIMATION IN MULTIVARIATE ANALYSIS

By Seymour Geisser[1]

National Institute of Arthritis and Metabolic Diseases

1. Introduction. In this paper the Bayesian approach to multivariate analysis taken by Geisser and Cornfield [6] and Geisser [5] is extended and given a more comprehensive treatment. Most of the classical multivariate estimation problems are here considered from the Bayesian standpoint.

In Section 2, Bayesian estimation procedures are obtained for: a vector mean, linear combinations of the elements of a vector mean; simple and partial variances; simple, partial, and multiple correlation coefficients. The posterior distributions of the canonical correlations and of the principal components are also discussed. Section 3 is devoted essentially to linear combinations of independent vector means when a common covariance matrix is assumed and also when the covariance structure is different for each population. For the general multivariate linear hypothesis we demonstrate, in Section 4, that the joint Bayesian posterior region for the elements of the regression matrix is equivalent to the usual confidence region for these parameters. Further the joint predictive density of a set of future observations generated by the linear hypothesis is also obtained thus enabling one to specify the probability that a set of future observations will be contained in a particular region based only on previous data.

Essentially no new distributions are necessary for this Bayesian approach, since all the results are couched in terms of familiar densities. While some of the posterior regions are equivalent to well established confidence regions, others are not. They may differ either as to the degrees of freedom involved or because of the fact that certain "confidence distributions" are non-Bayesian inversions e.g. the correlation coefficient, Brillinger [2].

The prior densities or weight functions used here are basically those of [5] and [6] and purport to reflect to a large degree prior ignorance or relative diffuseness. These unnormed densities or weight functions presumably may be "justified" by various rules, e.g. invariance, conjugate families, stable estimation, etc., or heuristic arguments. Although their utilization here does not necessarily preclude other contenders which may also be conceived of as displaying a measure of ignorance, it is our view that no others at present seem to be either more appropriate or as convenient. The fact that their application yields in many instances the same regions as those of classical confidence theory is certainly no detriment to their use, but in fact provides a Bayesian interpretation for these well established procedures.

2. One Population. Let $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be independent observations on a

Received 12 August 1964.

[1] Sponsored in part by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No.: DA-11-022-ORD-2059.

p-variate $N(\mathbf{\mu}, \mathbf{\Sigma})$ population. Hence,

$$L(\mathbf{\mu}, \mathbf{\Sigma}^{-1}) \propto |\mathbf{\Sigma}^{-1}|^{N/2} \exp \{-\tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1}[(N-1)\mathbf{S} + N(\bar{\mathbf{x}} - \mathbf{\mu})(\bar{\mathbf{x}} - \mathbf{\mu})']\}, \tag{2.1}$$

where

$$\bar{\mathbf{x}} = N^{-1}\textstyle\sum_{\alpha=1}^{N} \mathbf{x}_\alpha, \qquad (N-1)\mathbf{S} = \textstyle\sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \bar{\mathbf{x}})(\mathbf{x}_\alpha - \bar{\mathbf{x}})'$$

and

$$\textstyle\sum_{\alpha=1}^{N} (\mathbf{x}_\alpha - \mathbf{\mu})(\mathbf{x}_\alpha - \mathbf{\mu})' = (N-1)\mathbf{S} + N(\bar{\mathbf{x}} - \mathbf{\mu})(\bar{\mathbf{x}} - \mathbf{\mu})'. \tag{2.2}$$

If we assume as in Geisser and Cornfield [6] (setting $v = p + 1$) that the prior density which displays ignorance is

$$g(\mathbf{\mu}, \mathbf{\Sigma}^{-1})\, d\mathbf{\mu}\, d\mathbf{\Sigma}^{-1} \propto |\mathbf{\Sigma}|^{(p+1)/2}\, d\mathbf{\mu}\, d\mathbf{\Sigma}^{-1}, \tag{2.3}$$

then

$$P(\mathbf{\mu}, \mathbf{\Sigma}^{-1} \mid \mathbf{S}, \bar{\mathbf{x}}) \propto |\mathbf{\Sigma}^{-1}|^{(N-p-1)/2} \exp \{-\tfrac{1}{2} \operatorname{tr} [(N-1)\mathbf{S} + N(\bar{\mathbf{x}} - \mathbf{\mu})(\bar{\mathbf{x}} - \mathbf{\mu})']\} \tag{2.4}$$

with marginal densities

$$P(\mathbf{\Sigma}^{-1} \mid \mathbf{S}) \propto |\mathbf{\Sigma}^{-1}|^{(N-p-2)/2} \exp [-\tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1}(N-1)\mathbf{S}] \tag{2.5}$$

and

$$P(\mathbf{\mu} \mid \bar{\mathbf{x}}, \mathbf{S}) \propto |(N-1)\mathbf{S} + N(\bar{\mathbf{x}} - \mathbf{\mu})(\bar{\mathbf{x}} - \mathbf{\mu})'|^{-N/2}. \tag{2.6}$$

A posterior region can be constructed for $\mathbf{\mu}$ through the relation

$$T^2 = N(\bar{\mathbf{x}} - \mathbf{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \mathbf{\mu}), \tag{2.7}$$

where the posterior distribution of T^2 is $[p(N-1)/(N-p)]F(p, N-p)$ as shown in [6]. Hence

$$\Pr [T^2(\mathbf{\mu}) \leqq [p(N-1)/(N-p)]F_\alpha(p, N-p)] = 1 - \alpha \tag{2.8}$$

yields a posterior ellipsoid for $\mathbf{\mu}$. Note that this is equivalent to the confidence region for $\mathbf{\mu}$.

Intervals on linear combinations of the elements of $\mathbf{\mu}$ can also be obtained. Let $\mathbf{a}$ be a non-null real vector of constants, then it was shown [6] that *a posteriori*

$$\mathbf{a}'(\mathbf{\mu} - \bar{\mathbf{x}})N^{\frac{1}{2}}/(\mathbf{a}'\mathbf{S}\mathbf{a})^{\frac{1}{2}} \tag{2.9}$$

is distributed like $(N-1)^{\frac{1}{2}}(N-p)^{-\frac{1}{2}}t(N-p)$, where $t(N-p)$ is Student's t with $N-p$ degrees of freedom. Thus an interval on $\mathbf{a}'\mathbf{\mu}$ is obtained through

$$\begin{aligned} \Pr \{\mathbf{a}'\bar{\mathbf{x}} - (\mathbf{a}'\mathbf{S}\mathbf{a})^{\frac{1}{2}}[(N-1)/(N-p)N]^{\frac{1}{2}}t_{\alpha/2} &\leqq \mathbf{a}'\mathbf{\mu} \\ &\leqq \mathbf{a}'\bar{\mathbf{x}} + (\mathbf{a}'\mathbf{S}\mathbf{a})^{-1}[(N-1)/(N-p)N]^{\frac{1}{2}}t_{\alpha/2}\} = 1 - \alpha. \end{aligned} \tag{2.10}$$

This is in contrast to the confidence region based on the sampling density of (2.9) which is $t(N-1)$.

Another problem of interest is the ratio of the means from a bivariate normal population. Let $p = 2$ in (2.6). This yields for the joint posterior density of μ_1 and μ_2,

$$P(\mu_1, \mu_2) \propto [1 + k_{11}(\mu_1 - \bar{x}_1)^2 + k_{12}(\mu_1 - \bar{x}_1)(\mu_2 - \bar{x}_2) + k_{22}(\mu_2 - \bar{x}_2)^2]^{-N/2}, \tag{2.11}$$

where k_{ij} depends on sample variances and covariance. The problem then is to find suitable limits for $\eta = \mu_1/\mu_2$. The calculation of the posterior density of η is rather messy and involves several parameters i.e. fixed values of the statistics, so that finding reasonable approximations would be very helpful. Of course, as N increases the density of (2.11) tends to bivariate normality and the work of Creasy [4] would apply in the asymptotic case. If a suitable approximation to the density of η or some function of η can be obtained, this would provide a Bayesian solution for one of the possible models (bivariate normality with $\rho \neq 0$ and unknown) involved in the Fieller-Creasy problem.

Before turning our attention to the derivation of posterior densities for some of the interesting functions of the elements of $\mathbf{\Sigma}$, we shall briefly review some of the properties concerning marginal densities of certain functions of variables which are jointly Wishart distributed. Let $\mathbf{C} = \{c_{ij}\}$ be a $p \times p$ symmetric positive definite random matrix whose elements have the Wishart density with ν degrees of freedom based on a given positive definite matrix $\mathbf{\Omega} = \{\omega_{ij}\}$. $\mathbf{C}$ then is $W(\mathbf{\Omega}, \nu)$, i.e.,

$$f(\mathbf{C} \mid \mathbf{\Omega}, \nu) \propto |\mathbf{C}|^{(\nu-p-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \mathbf{\Omega}^{-1}\mathbf{C}\}. \tag{2.11a}$$

Further let $\mathbf{C}$ and $\mathbf{\Omega}$ be partitioned thus:

$$\mathbf{C} = \begin{matrix} & q & p-q \\ q & \mathbf{C}_{11} & \mathbf{C}_{12} \\ p-q & \mathbf{C}_{21} & \mathbf{C}_{22} \end{matrix}, \qquad \mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{pmatrix}, \tag{2.12}$$

$$\mathbf{C}^{11} = (\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21})^{-1} = \mathbf{C}_{11\cdot 2}^{-1}; \tag{2.13}$$

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{pmatrix}, \qquad \mathbf{\Omega}^{-1} = \begin{pmatrix} \mathbf{\Omega}^{11} & \mathbf{\Omega}^{12} \\ \mathbf{\Omega}^{21} & \mathbf{\Omega}^{22} \end{pmatrix}, \tag{2.14}$$

$$\mathbf{\Omega}^{11} = (\mathbf{\Omega}_{11} - \mathbf{\Omega}_{12}\mathbf{\Omega}_{22}^{-1}\mathbf{\Omega}_{21})^{-1} = \mathbf{\Omega}_{11\cdot 2}^{-1}. \tag{2.15}$$

From Wishart distribution theory we may now state the following well-known results:

I. The marginal density of $\mathbf{C}_{11}$ is $W(\mathbf{\Omega}_{11}, \nu)$.

II. The marginal density of $\mathbf{C}_{11\cdot 2}$ is $W(\mathbf{\Omega}_{11\cdot 2}, \nu - (p - q))$.

III. The marginal density of $r_{12} = c_{12}/(c_{11}c_{22})^{\frac{1}{2}}$ is

$$h(r_{12} \mid \rho_{12}, \nu) \propto (1 - r_{12})^{(\nu-3)/2}(1 - \rho_{12}^2)^{\nu/2} I_\nu(\rho_{12}r_{12}) \tag{2.16}$$

where

$$\rho_{12} = \omega_{12}/(\omega_{11}\omega_{22})^{\frac{1}{2}} \quad \text{and} \quad I_\nu(\rho r) = \int_0^\infty [(\cosh y - \rho r)^\nu]^{-1}\,dy.$$

IV. The joint density of the roots $c_1, \cdots, c_p$ of $\mathbf{C}$ depends only on the roots $\omega_1^{-1}, \cdots, \omega_p^{-1}$ of $\mathbf{\Omega}^{-1}$; (Roy [10], p. 188), and let it be represented by

$$h(c_1, \cdots, c_p \mid \omega_1^{-1}, \cdots, \omega_p^{-1}). \tag{2.17}$$

Although this is a rather complicated expression it has been evaluated by James [8] in terms of zonal polynomials.

V. The joint density of the q roots $b_1, \cdots, b_q$ of the $q \times q$ matrix $\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$ where $q \leqq p - q$, depends only on the roots $\delta_1, \cdots, \delta_q$ of $\mathbf{\Omega}_{11}^{-1}\mathbf{\Omega}_{12}\mathbf{\Omega}_{22}^{-1}\mathbf{\Omega}_{21}$, (Roy [10] p. 189) and let it be represented by

$$h(b_1, \cdots, b_q \mid \delta_1, \cdots, \delta_q). \tag{2.18}$$

This complicated density has been evaluated by Constantine [3] in terms of zonal polynomials. In particular when $q = 1$, let $c_1 = r^2$ and $\delta_1 = \rho^2$ (the square of the sample and population multiple correlation coefficient respectively), then

$$h(r^2 \mid \rho^2) \propto (1 - r^2)^{(\nu-p-1)/2}(1 - \rho^2)^{\nu/2} \cdot \sum_{j=0}^{\infty} \frac{(\rho^2)^j (r^2)^{(p-1)/2+j-1}\Gamma^2(\nu/2 + j)}{j!\Gamma[(p - 1)/2 + j]}. \tag{2.19}$$

Now in (2.5) let $(N - 1)\mathbf{S} = \mathbf{A}$ so that the posterior marginal density of $\mathbf{\Sigma}^{-1}$ is $W(\mathbf{A}, N - 1)$. This immediately implies that $\mathbf{\Sigma}_{11\cdot 2}^{-1}$ is $W(\mathbf{A}_{11\cdot 2}^{-1}, N - 1)$ and $\mathbf{\Sigma}_{11}^{-1}$ is $W(\mathbf{A}_{11}^{-1}, N - 1 - (p - q))$ by identifying $\mathbf{\Sigma}^{-1}$ with $\mathbf{C}$ and $\mathbf{A}$ with $\mathbf{\Omega}$ and making the indicated partitions of $\mathbf{\Sigma}$ and $\mathbf{A}$. Further for $q \geqq 2$, let

$$\mathbf{\Sigma}_{11\cdot 2} = \begin{array}{c} \\ 2 \\ q-2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 2 & \quad q-2 \end{array} \\ \begin{pmatrix} \mathbf{\Sigma}_{11\cdot 2\cdot 11} & \mathbf{\Sigma}_{11\cdot 2\cdot 12} \\ \mathbf{\Sigma}_{11\cdot 2\cdot 21} & \mathbf{\Sigma}_{11\cdot 2\cdot 22} \end{pmatrix} \end{array} \tag{2.20}$$

$$\mathbf{A}_{11\cdot 2} = \begin{array}{c} \\ 2 \\ q-2 \end{array}\!\!\begin{array}{c} \begin{array}{cc} 2 & \quad q-2 \end{array} \\ \begin{pmatrix} \mathbf{A}_{11\cdot 2\cdot 11} & \mathbf{A}_{11\cdot 2\cdot 12} \\ \mathbf{A}_{11\cdot 2\cdot 21} & \mathbf{A}_{11\cdot 2\cdot 22} \end{pmatrix} \end{array}. \tag{2.21}$$

Hence we get that $\mathbf{\Sigma}_{11\cdot 2\cdot 11}^{-1}$ is $W(\mathbf{A}_{11\cdot 2\cdot 11}^{-1}, N + 1 - q)$. From the preceding results it is clear that a_{11}/σ_{11} is χ^2_{N-p} by using the result for $\mathbf{\Sigma}_{11}^{-1}$ and letting $q = 1$ so that $\mathbf{\Sigma}_{11}^{-1} = \sigma_{11}^{-1}$. From $\mathbf{\Sigma}_{11\cdot 2}^{-1} = \sigma_{11\cdot 2}^{-1}$, i.e., $q = 1$, we get that $a_{11\cdot 2}/\sigma_{11\cdot 2}$ is χ^2_{N-1}. Note that in the two preceding cases the posterior densities are in the reverse order of the sampling densities. The marginal density of ρ_{12} is found from $\mathbf{\Sigma}_{11}^{-1}$ where $q = 2$ and the relation

$$(\sigma^{12})^2/\sigma^{11}\sigma^{22} = \sigma_{12}^2/\sigma_{11}\sigma_{22} = \rho_{12}^2. \tag{2.22}$$

This yields, by application of III, that

$$(2.23)\qquad h(\rho_{12} \mid r_{12}, N - p + 1)$$

where $h(\cdot \mid \cdot)$ is the function defined in (2.16). Observe here that the density of ρ_{12} depends on p, the original number of variables, in contrast to the sampling density of r_{12} which is independent of p although of the same form.

Now $\mathbf{\Sigma}_{11\cdot 2}$ is the partial covariance matrix of the first q variates holding fixed the last $p - q$ variates, i.e., $q + 1, q + 2, \cdots, p$. Hence from the marginal density of $\mathbf{\Sigma}_{11\cdot 2\cdot 11}^{-1}$ we can obtain, by applying III, the posterior density of the partial correlation coefficient $\rho_{12\cdot q+1,\cdots,p}$,

$$(2.24)\qquad h(\rho_{12\cdot q+1,\cdots,p} \mid r_{12\cdot q+1,\cdots,p}, N + 1 - q)$$

where $h(\cdot \mid \cdot)$ is given by (2.16). This density depends on the number of variates which are not held constant in contradistinction to the sampling density of $r_{12\cdot q+1,\cdots,p}$, which is from (2.16)

$$(2.25)\qquad h(r_{12\cdot q+1,\cdots,p} \mid \rho_{12\cdot q+1,\cdots,p}, N - 1 - (p - q))$$

and consequently depends only on the number of variates held constant. Actually (2.23) is a special case of (2.24) with $q = p$. Approximate posterior limits on $\rho_{12\cdot q+1,\cdots,p}$ can be obtained through Fisher's transformation. Let

$$(2.26)\qquad z = \tfrac{1}{2}\log[(1 + r)/(1 - r)], \qquad \zeta = \tfrac{1}{2}\log[(1 + \rho)/(1 - \rho)];$$

then *a posteriori* $(N - q - 1)^{\frac{1}{2}}(\zeta - z)$ is asymptotically $N(0, 1)$ and will be suitable for large sample sizes. For smaller sample sizes several good approximations are given by Hotelling [7]. His approximations of course deal with $h(r \mid \rho)$ and depend on ρ. His results can be readily utilized since for the Bayesian r and ρ are interchanged and hence the computations will depend on the known statistic r.

The principal components are the roots $\sigma_1, \cdots, \sigma_p$ of $\mathbf{\Sigma}$. From (2.5) and IV it is clear that the density of $\sigma_1^{-1}, \cdots, \sigma_p^{-1}$ is

$$(2.27)\qquad h(\sigma_1^{-1}, \cdots, \sigma_p^{-1} \mid a_1, \cdots, a_p)$$

where this is of the form discussed in (2.17) and $a_1, \cdots, a_p$ are the roots of $\mathbf{A}$.

The squares of the canonical correlations are the latent roots $\lambda_1, \lambda_2, \cdots, \lambda_q$ of $\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$. Recall that

$$(2.28)\qquad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}; \qquad \mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12} \\ \mathbf{\Sigma}^{21} & \mathbf{\Sigma}^{22} \end{pmatrix}.$$

Hence

$$(2.29)\qquad \mathbf{\Sigma}_{11}^{-1} = \mathbf{\Sigma}^{11} - \mathbf{\Sigma}^{12}(\mathbf{\Sigma}^{22})^{-1}\mathbf{\Sigma}^{21}.$$

Multiplication of both sides of (2.29) on the left by

$$(2.30)\qquad (\mathbf{\Sigma}^{11})^{-1} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$$

leads to

$$(\Sigma^{11})^{-1}\Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21} = \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}, \tag{2.31}$$

and the fact that the latent roots of the right hand side of (2.31) are equal to the latent roots of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Hence applying V, the joint posterior density of the latent roots is

$$h(\lambda_1, \cdots, \lambda_q \mid d_1, \cdots, d_q) \tag{2.32}$$

of the form of (2.18) where $d_1, \cdots, d_q$ are the latent roots of $\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$. For the particular case $q = 1$, the posterior density of the square of the multiple correlation coefficient, ρ^2, is obtained,

$$h(\rho^2 \mid r^2) \propto (1-\rho^2)^{(N-p-2)/2}(1-r^2)^{(N-1)/2} \cdot \sum_{j=0}^{\infty} \frac{(r^2)^j(\rho^2)^{(p-1)/2+j-1}\,\Gamma^2[(N-1)/2+j]}{j!\,\Gamma[(p-1)/2+j]}. \tag{2.33}$$

This of course is the exact form of the sampling density of r^2 given ρ^2 with r^2 and ρ^2 interchanged. Since ρ^2 is asymptotically normal *a posteriori*, where $r^2 \neq 0$, with mean

$$r^2 + [(p-1)/(N-1)](1-r^2) - 2(N-p)r^2(1-r^2)/(N^2-1) + O(N^{-2}) \tag{2.34}$$

and variance

$$4r^2(1-r^2)(N-p)^2/(N^2-1)(N+3) + O(N^{-2}), \tag{2.35}$$

we may use these facts to obtain approximate limits for ρ (see Kendall and Stuart [9] p. 341).

3. Several Populations. Let $\mathbf{x}_1^{(j)}, \cdots, \mathbf{x}_{N_j}^{(j)}$ be a set of independent observations on $N(\boldsymbol{\mu}^{(j)}, \Sigma)$ for $j = 1, \cdots, k$. In other words we have k samples with differing vector means but a common covariance matrix. The joint likelihood is proportional to

$$|\Sigma^{-1}|^{N/2} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}[\mathbf{A} + \textstyle\sum_{j=1}^{k} N_j(\bar{\mathbf{x}}^{(j)} - \boldsymbol{\mu}^{(j)})(\bar{\mathbf{x}}^{(j)} - \boldsymbol{\mu}^{(j)})']\}, \tag{3.1}$$

where

$N = \sum_{j=1}^{k} N_j$, $\bar{\mathbf{x}}^{(j)} = N_j^{-1}\sum_{\alpha=1}^{N_j}\mathbf{x}_\alpha^{(j)}$, $\mathbf{A}_j = \sum_{\alpha=1}^{N_j}(\mathbf{x}_\alpha^{(j)} - \bar{\mathbf{x}}^{(j)})(\mathbf{x}_\alpha^{(j)} - \bar{\mathbf{x}}^{(j)})'$,

and $\mathbf{A} = \sum_{j=1}^{k}\mathbf{A}_j$. We assume that the prior density is

$$g(\boldsymbol{\mu}^{(1)}, \cdots, \boldsymbol{\mu}^{(k)}, \Sigma^{-1}) \propto |\Sigma|^{(p+1)/2}, \tag{3.2}$$

which yields for the posterior density

$$|\Sigma^{-1}|^{(N-p-1)/2} \exp\{-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}[\mathbf{A} + \textstyle\sum_{\alpha=1}^{k} N_j(\bar{\mathbf{x}}^{(j)} - \boldsymbol{\mu}^{(j)})(\bar{\mathbf{x}}^{(j)} - \boldsymbol{\mu}^{(j)})']\}. \tag{3.3}$$

Further by integrating out Σ^{-1} we obtain the marginal posterior density of

$\mathbf{\mu}^{(1)}, \cdots, \mathbf{\mu}^{(k)}$ proportional to

$$|\mathbf{A} + \textstyle\sum_{j=1}^{k} N_j(\bar{\mathbf{x}}^{(j)} - \mathbf{\mu}^{(j)})(\bar{\mathbf{x}}^{(j)} - \mathbf{\mu}^{(j)})'|^{-N/2}. \tag{3.4}$$

A joint region can be obtained on $\mathbf{\mu}^{(1)}, \cdots, \mathbf{\mu}^{(k)}$ by means of the relation

$$\Pr[U \leqq U_{\alpha,p,k,N-k}] = 1 - \alpha, \tag{3.5}$$

where

$$U = |\mathbf{A}|/|\mathbf{A} + \textstyle\sum_{j=1}^{k} N_j(\bar{\mathbf{x}}^{(j)} - \mathbf{\mu}^{(j)})(\bar{\mathbf{x}}^{(j)} - \mathbf{\mu}^{(j)})'|, \tag{3.6}$$

since it is easy to show through the calculation of the moments of U that U has the $U_{p,k,N-k}$ distribution as given in Anderson [1] p. 194.

It is often of interest to calculate a region on the non-null linear combination $\mathbf{n} = \sum_{j=1}^{k} c_j \mathbf{\mu}^{(j)}$. The posterior marginal density is easily found from (3.3) to be

$$P(\mathbf{n}) \propto |\mathbf{A} + b^{-1}(\mathbf{n} - \textstyle\sum_{j=1}^{k} c_j\bar{\mathbf{x}}^{(j)})(\mathbf{n} - \sum_{j=1}^{k} c_j\bar{\mathbf{x}}^{(j)})'|^{(N-k+1)/2}, \tag{3.7}$$

where $b = \sum_{j=1}^{k} c_j^2 N_j^{-1}$. Hence, the quadratic form

$$b^{-1}(\mathbf{n} - \textstyle\sum_{j=1}^{k} c_j\bar{\mathbf{x}}^{(j)})'\mathbf{A}^{-1}(\mathbf{n} - \sum_{j=1}^{k} c_j\bar{\mathbf{x}}^{(j)}) \tag{3.8}$$

has posterior density $[p/(N - k + 1 - p)]F(p, N - k + 1 - p)$, which may be used for obtaining an ellipsoidal region on $\mathbf{n}$. This of course is identical to the confidence region on $\mathbf{n}$.

The problem becomes more complicated if we assume that the covariance matrix is different for each population, i.e., $N(\mathbf{\mu}^{(j)}, \mathbf{\Sigma}_j)$ $j = 1, \cdots, k$. If we assume a prior density

$$g(\mathbf{\mu}^{(1)}, \cdots, \mathbf{\mu}^{(k)}, \mathbf{\Sigma}_1^{-1}, \cdots, \mathbf{\Sigma}_k^{-1}) \propto \textstyle\prod_{j=1}^{k} |\mathbf{\Sigma}_j|^{(p+1)/2}, \tag{3.9}$$

we may calculate the posterior density of $\mathbf{\mu}^{(1)}, \cdots, \mathbf{\mu}^{(k)}$ to be proportional to

$$\textstyle\prod_{j=1}^{k} |\mathbf{A}_j + N_j(\bar{\mathbf{x}}^{(j)} - \mathbf{\mu}^{(j)})(\bar{\mathbf{x}}^{(j)} - \mathbf{\mu}^{(j)})'|^{-N_j/2}. \tag{3.10}$$

Now to calculate the density of $\mathbf{n}$ is rather involved (this is essentially the multivariate extension of the Fisher-Behrens distribution) since $\mathbf{n}$ is distributed as a linear combination of independent generalized Student densities.

A crude approximation that may not at all be too bad for moderately large sample sizes can be based on the multivariate normal approximation to the generalized Student density using the first two moments of $\mathbf{n}$. Since

$$E\mathbf{\mu}^{(j)} = \bar{\mathbf{x}}^{(j)} \quad \text{and} \quad V(\mathbf{\mu}^{(j)}) = \mathbf{A}_j/N_j(N_j - p - 2), \tag{3.11}$$

$\mathbf{n}$ is approximately multivariate normal with

$$E\mathbf{n} = \textstyle\sum_{j=1}^{k} c_j\bar{\mathbf{x}}^{(j)} = \bar{\mathbf{y}} \quad \text{and} \quad V(\mathbf{n}) = \sum_{j=1}^{k} (c_j^2 N_j^{-1}/N_j - p - 2)\mathbf{A}_j = \mathbf{B}; \tag{3.12}$$

and the quadratic form

$$(\mathbf{n} - \bar{\mathbf{y}})'\mathbf{B}^{-1}(\mathbf{n} - \bar{\mathbf{y}}) \tag{3.13}$$

is asymptotically χ_p^2. One then may obtain an approximate region on $\mathbf{n}$ using (3.13).

4. The General Linear Hypothesis. Suppose $\mathbf{x}_1, \cdots, \mathbf{x}_N$ are a set of N independent p-component observations where $\mathbf{x}_\alpha$ is $N(\boldsymbol{\beta}\mathbf{z}_\alpha, \boldsymbol{\Sigma})$, $\mathbf{z}_\alpha$ is a known vector of q components, $\boldsymbol{\beta}$ is a $p \times q$ regression matrix whose elements are unknown. Let $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_N)$ be of rank q, $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_N)$ of rank p where $N \geqq p + q$. Since

$$(4.1) \qquad (\mathbf{X} - \boldsymbol{\beta}\mathbf{Z})(\mathbf{X} - \boldsymbol{\beta}\mathbf{Z})' = \mathbf{A} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{Z}\mathbf{Z}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})',$$

where

$$(4.2) \qquad \hat{\boldsymbol{\beta}} = \mathbf{X}\mathbf{Z}'(\mathbf{Z}\mathbf{Z}')^{-1}; \qquad \mathbf{A} = (\mathbf{X} - \hat{\boldsymbol{\beta}}\mathbf{Z})(\mathbf{X} - \hat{\boldsymbol{\beta}}\mathbf{Z})',$$

then

$$(4.3) \qquad L(\boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{N/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}[\mathbf{A} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{Z}\mathbf{Z}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})']\}.$$

Let the prior density of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{-1}$ be

$$(4.4) \qquad g(\boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1})\, d\boldsymbol{\beta}\, d\boldsymbol{\Sigma}^{-1} \propto |\boldsymbol{\Sigma}|^{(p+1)/2}\, d\boldsymbol{\beta}\, d\boldsymbol{\Sigma}^{-1}.$$

Hence the posterior density of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}^{-1}$ is

$$(4.5) \qquad \begin{aligned} &P(\boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1} \mid \mathbf{X}, \mathbf{Z}) \\ &\qquad \propto |\boldsymbol{\Sigma}^{-1}|^{(N-p-1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}[\mathbf{A} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{Z}\mathbf{Z}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})']\}. \end{aligned}$$

Since

$$(4.6) \qquad \int \exp[-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{Z}\mathbf{Z}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})']\, d\boldsymbol{\beta} \propto |\boldsymbol{\Sigma}|^{q/2}|\mathbf{Z}\mathbf{Z}'|^{-p/2},$$

we obtain

$$(4.7) \qquad P(\boldsymbol{\Sigma}^{-1} \mid \mathbf{A}) \propto |\boldsymbol{\Sigma}^{-1}|^{(N-q-p-1)/2} \exp(-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}\mathbf{A})$$

and

$$(4.8) \qquad P(\boldsymbol{\beta} \mid \mathbf{X}, \mathbf{Z}) \propto |\mathbf{A} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{Z}\mathbf{Z}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'|^{-N/2}.$$

Let

$$(4.9) \qquad U = |\mathbf{A}|/|\mathbf{A} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{Z}\mathbf{Z}'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'|.$$

Calculation of the moments of U from the density of $\boldsymbol{\beta}$ reveals that U is distributed like $U_{p,q,N-q}$, i.e. as a product of beta variables and defined by Anderson [1], p. 194. In other words the posterior distribution of U (where $\boldsymbol{\beta}$ is the set of random variables and the other quantities are fixed) is the same as the sampling distribution of U (where $\boldsymbol{\beta}$ is fixed and $\mathbf{A}$ and $\hat{\boldsymbol{\beta}}$ are the sets of random variables). Hence a posterior region on the elements of $\boldsymbol{\beta}$ is given through the relation

$$(4.10) \qquad \Pr[U(\boldsymbol{\beta}) \leqq U(_{\alpha,p,q,N-q})] = 1 - \alpha$$

where $U_{\alpha,p,q,N-q}$ is the αth percentage point. Therefore the Bayesian region on $\boldsymbol{\beta}$ is equivalent to the confidence region.

Suppose now we wish to predict where a set of future observations $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M$ will lie, where $\mathbf{y}_\alpha$ is $N(\boldsymbol{\beta}\mathbf{w}_\alpha, \boldsymbol{\Sigma})$ and $\mathbf{w}_\alpha$ is a known vector of q components,

$$(4.11) \qquad \mathbf{Y} = (\mathbf{y}_1, \cdots, \mathbf{y}_M); \qquad \mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_M).$$

The predictive density of **Y** is

$$P(\mathbf{Y} \mid \mathbf{W}, \mathbf{Z}, \mathbf{X}) = \iint P(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\beta} \mid \mathbf{X}, \mathbf{Z}) f(\mathbf{Y} \mid \mathbf{W}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1})\, d\boldsymbol{\beta}\, d\boldsymbol{\Sigma}^{-1}, \tag{4.12}$$

where

$$f(\mathbf{Y} \mid \mathbf{W}, \boldsymbol{\beta}, \boldsymbol{\Sigma}^{-1}) \propto |\boldsymbol{\Sigma}^{-1}|^{M/2} \exp[-\tfrac{1}{2} \operatorname{tr} \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\beta}\mathbf{W})(\mathbf{Y} - \boldsymbol{\beta}\mathbf{W})']. \tag{4.13}$$

Define

$$\mathbf{V} = (\mathbf{X}, \mathbf{Y}), \qquad \mathbf{U} = (\mathbf{Z}, \mathbf{W}), \qquad \hat{\boldsymbol{\beta}}_0 = \mathbf{V}\mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}; \tag{4.14}$$

then we may easily obtain

$$P(\mathbf{Y} \mid \mathbf{W}, \mathbf{Z}, \mathbf{X}) \propto |(\mathbf{V} - \hat{\boldsymbol{\beta}}_0\mathbf{U})(\mathbf{V} - \hat{\boldsymbol{\beta}}_0\mathbf{U})'|^{-(N+M-q)/2}. \tag{4.15}$$

Using the following relations,

$$\begin{aligned} \mathbf{V} - \hat{\boldsymbol{\beta}}_0\mathbf{U} &= (\mathbf{X}, \mathbf{Y})[\mathbf{I} - \mathbf{U}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{U}] \\ (\mathbf{U}\mathbf{U}')^{-1} &= (\mathbf{Z}\mathbf{Z}')^{-1}[\mathbf{I} - \mathbf{W}\mathbf{W}'(\mathbf{U}\mathbf{U}')^{-1}] \\ \mathbf{U}\mathbf{U}' &= [\mathbf{I} + \mathbf{W}\mathbf{W}'(\mathbf{Z}\mathbf{Z}')^{-1}]\mathbf{Z}\mathbf{Z}', \end{aligned} \tag{4.16}$$

we may without any difficulty establish

$$\begin{aligned} &P(\mathbf{Y} \mid \mathbf{X}, \mathbf{Z}, \mathbf{W}) \\ &\quad \propto |\mathbf{A} + (\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{W})(\mathbf{I} - \mathbf{W}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{W})(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{W})'|^{-(N+M-q)/2}. \end{aligned} \tag{4.17}$$

The predictive density of **Y** is then of the same type as the posterior density of $\boldsymbol{\beta}$ so that

$$U = |\mathbf{A}|/|\mathbf{A} + (\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{W})(\mathbf{I} - \mathbf{W}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{W})(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{W})'| \tag{4.18}$$

is distributed as $U_{p,M,N-q}$. If we are in the univariate case $p = 1$, then $\mathbf{A} = a_{11}$ is a scalar and

$$a_{11}^{-1}(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{W})(\mathbf{I} - \mathbf{W}'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{W})(\mathbf{Y} - \hat{\boldsymbol{\beta}}\mathbf{W})' \tag{4.19}$$

is a quadratic form distributed like $[M/(N - q)]F(M, N - q)$. In the case $M = 1$ we may find a predictive ellipsoid for the single future vector observation $\mathbf{y}_1$ since the quadratic form

$$(1 - \mathbf{w}_1'(\mathbf{U}\mathbf{U}')^{-1}\mathbf{w}_1)(\mathbf{y}_1 - \hat{\boldsymbol{\beta}}\mathbf{w}_1)'\mathbf{A}^{-1}(\mathbf{y}_1 - \hat{\boldsymbol{\beta}}\mathbf{w}_1) \tag{4.20}$$

is distributed like $[p/(N - q - p + 1)]F(p, N - q - p + 1)$. Therefore by means of either (4.18), (4.19) or (4.20) depending on the particular situation, predictive regions can be obtained for future observations.

We now return to the posterior density of $\boldsymbol{\beta}$ and derive the marginal density of $\boldsymbol{\beta}_1$, a $q_1 \times p$ matrix. Define

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2), \qquad \hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2), \qquad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \tag{4.21}$$

where $\mathbf{Z}_1$ is $q_1 \times N$. Utilization of a fundamental decomposition in the general linear hypothesis

$$(4.22)\quad (\beta - \hat{\beta})\mathbf{ZZ}'(\beta - \hat{\beta})' = (\beta_1 - \hat{\beta}_1)\mathbf{F}(\beta_1 - \hat{\beta}_1)' + (\beta_2 - \Delta)\mathbf{Z}_2\mathbf{Z}_2'(\beta_2 - \Delta)',$$

where $\mathbf{F} = \mathbf{Z}_1[\mathbf{I} - \mathbf{Z}_2'(\mathbf{Z}_2\mathbf{Z}_2')^{-1}\mathbf{Z}_2]\mathbf{Z}_1'$ and Δ does not involve β_1 or β_2, permits us to rewrite (4.8) as

$$(4.23)\quad P(\beta_1, \beta_2) \propto |\mathbf{A} + (\beta_1 - \hat{\beta}_1)\mathbf{F}(\beta_1 - \hat{\beta}_1)' + (\beta_2 - \Delta)\mathbf{Z}_2\mathbf{Z}_2'(\beta_2 - \Delta)'|^{-N/2}.$$

We may then easily integrate out β_2 since (4.23) is of the same basic form as (4.8). This yields

$$(4.24)\qquad P(\beta_1) \propto |\mathbf{A} + (\beta_1 - \hat{\beta}_1)\mathbf{F}(\beta_1 - \hat{\beta}_1)'|^{-[N-(q-q_1)]/2}.$$

Hence by the same argument as before

$$(4.25)\qquad U = |\mathbf{A}|/|\mathbf{A} + (\beta_1 - \hat{\beta}_1)\mathbf{F}(\beta_1 - \hat{\beta}_1)'|$$

is distributed *a posteriori* as $U_{p,q_1,N-q}$ and the posterior region on β_1 is equivalent to the confidence region.

5. Acknowledgment. I am indebted to the Mathematics Research Center, United States Army, the University of Wisconsin for providing the facilities for the preparation of the final draft of this paper. After obtaining these results, I learned that Tiao and Zellner [11] have also independently derived the result given by [4.8].

REFERENCES

[1] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] BRILLINGER, D. R. (1963). Necessary and sufficient conditions for a statistical problem to be invariant under a Lie group. *Ann. Math. Statist.* **34** 492–500.

[3] CONSTANTINE, A. G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

[4] CREASY, M. A. (1954). Limits for the ratio of means. *J. Roy. Statist. Soc. Ser. B* **16** 186–194.

[5] GEISSER, S. (1964). Posterior odds for multivariate normal classifications. *J. Roy. Statist. Soc. Ser. B* **26** 69–76.

[6] GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. Ser. B* **25** 368–376.

[7] HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *J. Roy. Statist. Soc. Ser. B* **15** 193–225.

[8] JAMES, A. T. (1960). The distribution of the latent roots of the covariance matrix. *Ann. Math. Statist.* **31** 151–158.

[9] KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics.* **2.** Hafner, New York.

[10] ROY, S. N. (1957). *Some Aspects of Multivariate Analysis.* Wiley, New York.

[11] TIAO, G. C. and ZELLNER, A. (1964). The Bayesian estimation of multivariate regression. To be published in J. Roy. Statist. Soc.