A BAYES APPROACH FOR COMBINING CORRELATED ESTIMATES

SEYMOUR GEISSER
National Institute of Arthritis and Metabolic Diseases

A Bayes solution is supplied for an estimation problem involving a sample from a multivariate normal population having an arbitrary unknown covariance matrix, but a vector mean whose components are all equal. Assuming that a particular unnormed prior density is a convenient expression for displaying prior ignorance, it is then demonstrated that a posterior interval for this common mean can be based on Student's t distribution. If prior information can be conveniently represented by a natural conjugate prior density, the posterior interval will also depend on Student's t. An extension is made to the case of estimating the constant difference between two parallel profiles.

1. INTRODUCTION

HALPERIN [7] discussed the frequently occurring problem of combining averages from normal populations where each average was an estimate of a common mean μ. He derived confidence intervals when the estimates were either correlated or independent. A Bayesian solution for the independent case has already been given by Jeffreys [8, p. 199], therefore, we restrict ourselves to a Bayesian approach for the correlated case.

2. CORRELATED ESTIMATES

The underlying model discussed by Halperin, assumed that the k-vector observations $\mathbf{x}_1, \cdots, \mathbf{x}_n$ have joint likelihood

$$L(\mu, \mathbf{\Sigma}) \propto \left| \mathbf{\Sigma}^{-1} \right|^{n/2} \exp\left(-\tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1} \sum_{j=1}^{n} (\mathbf{x}_j - \mu \mathbf{e})(\mathbf{x}_j - \mu \mathbf{e})' \right) \quad (2.1)$$

where μ is a scalar, $\mathbf{e}' = (1, \cdots, 1)$, $\mathbf{\Sigma}$ is the $k \times k$ covariance matrix and $n > k$.

We assume in this section that in the absence of any prior objective sample evidence that we may assign a diffuse joint prior density to $\mathbf{\Sigma}^{-1}$ and μ, which may be conveniently represented by

$$g(\mathbf{\Sigma}^{-1}, \mu) d\mathbf{\Sigma}^{-1} d\mu \propto \left| \mathbf{\Sigma} \right|^{(k+1)/2} d\mathbf{\Sigma}^{-1} d\mu \quad (2.2)$$

where $d\mathbf{\Sigma}^{-1} = \prod_{i \geq j} d\sigma^{ij}$ and σ^{ij} are the elements of $\mathbf{\Sigma}^{-1}$. Therefore the posterior distribution of $\mathbf{\Sigma}^{-1}$ and μ is

$$P(\mathbf{\Sigma}^{-1}, \mu) \propto L(\mu, \mathbf{\Sigma}) g(\mathbf{\Sigma}^{-1}, \mu) \quad (2.3)$$

or

$$P(\mathbf{\Sigma}^{-1}, \mu) \propto \left| \mathbf{\Sigma}^{-1} \right|^{(n-k-1)/2} \exp\left(-\tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1} \sum_{j=1}^{n} (\mathbf{x}_j - \mu \mathbf{e})(\mathbf{x}_j - \mu \mathbf{e})' \right). \quad (2.4)$$

The marginal posterior distribution of μ is found by integrating out $\mathbf{\Sigma}^{-1}$ in (2.4) and yields

$$P(\mu) \propto \left| \sum_{j=1}^{n} (\mathbf{x}_j - \mu e)(\mathbf{x}_j - \mu e)' \right|^{-n/2} \tag{2.5}$$

Now let the sample mean vector

$$\bar{\mathbf{x}} = n^{-1} \sum_{j=1}^{n} \mathbf{x}_j \quad \text{and} \quad S = \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$$

so that

$$\sum_{j=1}^{n} (\mathbf{x}_j - \mu e)(\mathbf{x}_j - \mu e)' = S + n(\bar{\mathbf{x}} - \mu e)(\bar{\mathbf{x}} - \mu e)'. \tag{2·6}$$

Further recall that

$$| I + n(\bar{\mathbf{x}} - \mu e)(\bar{\mathbf{x}} - \mu e)' S^{-1} | = 1 + n(\bar{\mathbf{x}} - \mu e)' S^{-1} (\bar{\mathbf{x}} - \mu e). \tag{2.7}$$

If we define $u = e'S^{-1}\bar{\mathbf{x}}$, $z = e'S^{-1}e$, $w = \bar{\mathbf{x}}'S^{-1}\bar{\mathbf{x}}$ we may then obtain from (2.5)

$$P(\mu) \propto \left[1 + \frac{nz\left(\mu - \frac{u}{z}\right)^2}{1 + nw - nu^2 z^{-1}} \right]^{-n/2}. \tag{2.8}$$

This yields the result that

$$t = \left(\mu - \frac{u}{z}\right) \left[\frac{(n-1)nz}{1 + nw - nu^2 z^{-1}} \right]^{1/2} \tag{2.9}$$

where t is distributed as the "Student" t distribution with $n-1$ degrees of freedom.

Due to the symmetry of the t distribution one would choose a $1-2\alpha$ Bayesian probability interval on μ to be

$$P\left\{ \frac{u}{z} - st_\alpha \leq \mu \leq \frac{u}{z} + st_\alpha \right\} = 1 - 2\alpha \tag{2.10}$$

where

$$s^2 = \frac{1 + nw - nu^2 z^{-1}}{(n-1)nz} \tag{2.11}$$

and t_α is the αth percentage point of the t distribution. Note also that u/z is the posterior mode or mean of μ and as shown by Halperin [7], the maximum likelihood estimate of μ.

3. CHOICE OF THE PRIOR DISTRIBUTION

As regards μ a location parameter we may take the view of Jeffreys and assume that it is uniformly distributed throughout its domain of definition and independent of the elements of $\mathbf{\Sigma}^{-1}$. We can then take for the prior distribution of $\mathbf{\Sigma}^{-1}$ the Wishart family

$$g_v(\mathbf{\Sigma}^{-1})d\mathbf{\Sigma}^{-1} \propto |\mathbf{\Sigma}|^{v/2} \exp\left(-\tfrac{1}{2}\operatorname{tr}(k-v+1)\mathbf{\Lambda}\mathbf{\Sigma}^{-1}\right)d\mathbf{\Sigma}^{-1} \tag{3.1}$$

where $\mathbf{\Lambda}$ is positive definite, so that $\mathbf{\Sigma}^{-1}$ is $W[(k-v+1)^{-1}\mathbf{\Lambda}^{-1}, k-v+1]$; i.e. Wishart with parameter $[(k-v+1)^{-1}\mathbf{\Lambda}^{-1}, k-v+1]$. Hence $\mathbf{\Sigma}^{-1}$ has characteristic function [Anderson (1, p. 160)]

$$E \exp(i \operatorname{tr} \mathbf{\Sigma}^{-1}\mathbf{\theta}) = |I - 2i(k-v+1)^{-1}\mathbf{\Lambda}^{-1}\mathbf{\theta}|^{-(k-v+1)/2}. \tag{3.2}$$

While the density exists only for $v<2$, the characteristic function exists for $v<k+1$. Since $\mathbf{\Sigma}^{-1}$ is Wishart distributed with $k-v+1$ degrees of freedom, it seems reasonable to argue, in the sense of Jeffreys, that $k-v+1=0$ degrees of freedom may represent a particular display of ignorance, diffuse enough to be substantially modified by a small number of observations. Hence we let $v=k+1$ and take for our unnormed prior (2.2). This is much like having a single vectorial observation on each of m multivariate normal populations having unknown and different vector means, but the same covariance matrix $\mathbf{\Sigma}$. In this case, from the sampling theory point of view, there is no information available for a joint inference on the elements of $\mathbf{\Sigma}$ no matter how large m. On the other hand if n_i observations were available on the i-th population ($i=1, \cdots, m$) and though $n_i<k$, so that no single sample provides us with a joint inference, it is quite clear that as soon as

$$\sum_{i=1}^{m} (n_i - 1) \geq k$$

our pooled estimate is informative. We may note in passing that a single sample of size n_i where $1<n_i<k$ is jointly uninformative but would provide information for a marginal inference—say on a particular variance. However, the case $n_i=1$ is uninformative jointly or marginally for the elements of $\mathbf{\Sigma}$, and moreover the pooling of m such observations as in the case discussed before still does not yield any further information jointly or marginally. We have attempted here to give some intuitive "justification" for the particular diffuse density that was employed by a heuristic argument based on sampling theory considerations. Other "justifications" appear in some of the papers cited in section 5.

More generally, if prior information subjective or otherwise may conveniently be incorporable into the following distribution family

$$\begin{aligned} g_v(\mu, \mathbf{\Sigma}^{-1} \mid \delta, \theta, \mathbf{\Lambda}) \propto \exp\left(-\frac{\delta}{2}(\mu-\theta)^2 \operatorname{tr} \mathbf{\Sigma}^{-1}ee'\right) |\mathbf{\Sigma}|^{v/2} \\ \cdot \exp\left(-\tfrac{1}{2}\operatorname{tr}(k-v+1)\mathbf{\Lambda}\mathbf{\Sigma}^{-1}\right) \end{aligned} \tag{3.3}$$

we obtain as the marginal posterior of μ

$$\begin{aligned} &P_v(\mu \mid \delta, \theta, \mathbf{\Lambda}) \\ &\quad \propto |\delta(\mu-\theta)^2 ee' + (k-v+1)\mathbf{\Lambda} + \textstyle\sum_j (\mathbf{x}_j - \mu e)(\mathbf{x}_j - \mu e)'|^{-(n+k-v+1)/2} \end{aligned}$$

Note that for $\delta>0$, $v<2$ and $\mathbf{\Lambda}$ positive definite, (3.3) is a natural conjugate density in the sense defined in Raiffa and Schlaifer [9]. Further (3.3) reduces to (2.5) for $\delta=0$ and $v=k+1$. Now if we let S be defined as in the previous section and we further define

$$U = (k - v + 1)\mathbf{\Lambda} + S + n(n + \delta)^{-1}\delta(\bar{\mathbf{x}} - \theta e)(\bar{\mathbf{x}} - \theta e)' \tag{3.5}$$

$$\bar{y} = \frac{(n\bar{\mathbf{x}} + \delta\theta e)}{(n + \delta)} \tag{3.6}$$

then (3.4) may be written as

$$P_v(\mu \mid U, \bar{y}) \propto [1 + (n + \delta)(\bar{y} - \mu e)' U^{-1}(\bar{y} - \mu e)]^{-(n+k-v+1)/2} \tag{3.7}$$

If we then let $u = e'U^{-1}\bar{y}$, $z = e'U^{-1}e$, $w = y'U^{-1}y$ then

$$t = \left(\mu - \frac{u}{z}\right)\left[\frac{(n + k - v)(n + \delta)z}{1 + (n + \delta)w - (n + \delta)u^2 z^{-1}}\right]^{1/2} \tag{3.8}$$

where t is distributed at the "Student" t with $n+k-v$ degrees of freedom. Obviously this reduces to (2.9) for $v = k+1$ and $\delta = 0$.

4. AN EXTENSION TO THE PARALLEL PROFILE CASE

Suppose we have a set of k-vectorial observations $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ on $N(\mathbf{n}, \mathbf{\Sigma})$ and a second set $y_1, \cdots, y_m$ on $N(\mathbf{n} + \mu e, \mathbf{\Sigma})$, where $\mathbf{n}$ is a mean vector and μ a scalar. We seek an estimate for μ, the assumed constant difference between two parallel profiles, vide Greenhouse and Geisser [6]. The joint likelihood

$$L(\mu, \mathbf{n}, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}^{-1}|^{(n+m)/2}$$
$$\cdot \exp\left(-\tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1}[S + n(\bar{\mathbf{x}} - \mathbf{n})(\bar{\mathbf{x}} - \mathbf{n})' + m(\bar{y} - \mathbf{n} - \mu e)(\bar{y} - \mathbf{n} - \mu e)']\right) \tag{4.1}$$

where

$$S = \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + \sum_{j=1}^{m} (y_j - \bar{y})(y_j - \bar{y})'$$

and

$$\bar{\mathbf{x}} = n^{-1} \sum_{j=1}^{n} \mathbf{x}_j, \qquad \bar{y} = m^{-1} \sum_{j=1}^{m} y_j.$$

We then multiply $L(\mu, \mathbf{n}, \Sigma)$ by the prior density

$$g(\mu, \mathbf{n}, \mathbf{\Sigma}^{-1}) d\mu d\mathbf{n} d\mathbf{\Sigma}^{-1} \propto |\mathbf{\Sigma}|^{(k+1)/2} d\mu d\mathbf{n} d\mathbf{\Sigma}^{-1} \tag{4.2}$$

which yields

$$P(\mu, \mathbf{n}, \mathbf{\Sigma}^{-1}) \propto |\mathbf{\Sigma}^{-1}|^{(n+m-k-1)/2}$$
$$\cdot \exp\left(-\tfrac{1}{2} \operatorname{tr} \mathbf{\Sigma}^{-1}[S + n(\bar{\mathbf{x}} - \mathbf{n})(\bar{\mathbf{x}} - \mathbf{n})' + m(\bar{y} - \mathbf{n} - \mu e)(\bar{y} - \mathbf{n} - \mu e)']\right. \tag{4.3}$$

We then integrate out $\mathbf{\Sigma}^{-1}$ and obtain

$$P(\mathbf{n}, \mu) \propto |S + n(\bar{\mathbf{x}} - \mathbf{n})(\bar{\mathbf{x}} - \mathbf{n})' + m(\bar{y} - \mathbf{n} - \mu e)(\bar{y} - \mathbf{n} - \mu e)'|^{-(m+n)/2} \tag{4.4}$$

which may be rewritten as

$$P(\mathbf{n}, \mu)$$
$$\propto \left| S + \frac{nm}{n + m} (\mu e - d)(\mu e - d)' + (n + m)(\bar{f} - \mathbf{n})(\bar{f} - \mathbf{n})' \right|^{-(m+n)/2} \tag{4.5}$$

where

$$\bar{f} = \frac{n\bar{x} + m\bar{y} - m\mu e}{n+m}$$

and $\bar{d} = \bar{y} - \bar{x}$. Hence the posterior density of μ is given as

$$P(\mu) \propto \left| S + \frac{nm}{n+m}(\mu e - \bar{d})(\mu e - \bar{d})' \right|^{-(n+m-1)/2} \tag{4.6}$$

or that

$$t = \left(\mu - \frac{u}{z}\right)\left[\frac{(n+m-2)\dfrac{nmz}{n+m}}{1 + \dfrac{nm}{n+m}(w - u^2z^{-1})}\right]^{1/2} \tag{4.7}$$

where $u = e'S^{-1}(\bar{y} - \bar{x})$, $z = e'S^{-1}e$, $w = (\bar{y} - \bar{x})'S^{-1}(\bar{y} - \bar{x})$, and t is the "Student" t with $n+m-2$ degrees of freedom.

A natural conjugate density

$$\begin{aligned} &g_v(\mu, \mathbf{n}, \Sigma^{-1} \mid \delta, \theta, \Lambda, \beta, \boldsymbol{\tau}) \\ &\quad \propto \mid \Sigma \mid^{v/2} \exp\left(-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}[\delta(\mu - \theta)^2 ee' + (k - v + 1)\Lambda + \beta(\mathbf{n} - \boldsymbol{\tau})(\mathbf{n} - \boldsymbol{\tau})']\right) \end{aligned} \tag{4.8}$$

may also be employed here for the prior density. Utilization of (4.8) leads to a marginal posterior density for μ

$$P(\mu) \propto \mid U + c(\mu e - \bar{h})(\mu e - \bar{h})' \mid^{-(n+m-v)/2} \tag{4.9}$$

where

$$\begin{aligned} U = {} & S + (k - v + 1)\Lambda + \beta n(\beta + n)^{-1}(\bar{x} - \boldsymbol{\tau})(\bar{x} - \boldsymbol{\tau})' \\ & + \frac{\delta(\beta + n)^{-1}m}{\delta(\beta + n + m) + (\beta + n)m}(n(\bar{y} - \bar{x} - \theta e) + \beta(\bar{y} - \boldsymbol{\tau} - \theta e)) \\ & \cdot (n(\bar{y} - \bar{x} - \theta e) + \beta(\bar{y} - \boldsymbol{\tau} - \theta e))', \end{aligned} \tag{4.10}$$

$$c = \delta + \frac{(\beta + n)m}{\beta + n + m}, \tag{4.11}$$

$$\bar{h} = \frac{m[n(\bar{y} - \bar{x}) + \beta(\bar{y} - \boldsymbol{\tau})] + (\beta + n + m)\delta\theta e}{\delta(\beta + n + m) + (\beta + n)m}. \tag{4.12}$$

Further

$$t = \left(\mu - \frac{u}{z}\right)\left[\frac{(n + m - v + k - 1)cz}{1 + c(w - u^2z^{-1})}\right]^{1/2} \tag{4.13}$$

has the t distribution with $n+m-v+k-1$ degrees of freedom where $u = e'U^{-1}\bar{h}$, $z = e'U^{-1}e$, $w = \bar{h}'U^{-1}\bar{h}$. It is clear that (4.13) reduces to (4.7) when $\delta = \beta = 0$ and $v = k+1$.

5. REMARKS

The same type of approach, with and without minor variations, has also been utilized for sundry other multivariate normal problems in the following series of papers: Geisser and Cornfield [5], Geisser [3], Tiao and Zellner [10], Geisser [4], Ando and Kaufman [2], Zellner and Chatty [11].

REFERENCES

[1] Anderson, T. W. (1958), *Introduction to Multivariate Statistical Analysis,* New York: John Wiley and Sons, Inc.

[2] Ando, A., and Kaufman, G. (1965), "Bayesian analysis of the independent multinormal process"—neither mean nor precision known, *Journal of the American Statistical Association* (in press) Vol. 60, No. 309, pp. 347–58 March 1965

[3] Geisser, S. (1964), "Posterior odds for multivariate normal classifications," *Journal of the Royal Statistical Society,* B, Vol. 26, No. 1, pp. 69–76.

[4] Geisser, S. (1965), "Bayesian estimation in multivariate analysis," *Annals of Mathematical Statistics,* Vol. 60, pp. 150–59, March 1965)

[5] Geisser, S., and Cornfield, J. (1963), "Posterior distributions for multivariate normal parameters," *Journal of the Royal Statistical Society,* B, Vol. 25, No. 2, pp. 368–376.

[6] Greenhouse, S. and Geisser, S. (1959), "On methods in the analysis of profile data," *Psychometrika,* Vol. 24, No. 2, pp. 95–112.

[7] Halperin, M. (1961), "Almost linearly-optimum combination of unbiased estimates," *Journal of the American Statistical Association,* Vol. 56, No. 293, pp. 36–43.

[8] Jeffreys, H. (1961), *Theory of Probability,* Oxford: Clarendon Press, third edition

[9] Raiffa, H. and Schlaifer, R. (1961), *Applied Statistical Decision Theory,* Boston: Harvard University Graduate School of Business Administration.

[10] Tiao, G. and Zellner, A. (1964), "On the Bayesian estimation of multivariate regression," *Journal of the Royal Statistical Soeity,* B, Vol. 26, No. 2, pp. 277–285.

[11] Zellner, A. and Chatty, V. (1965), "Prediction problems in regression models from the Bayesian point of view," *Journal of the American Statistical Association* (in press).