ESTIMATING THE MEAN OF A MULTIVARIATE NORMAL POPULATION WITH GENERAL QUADRATIC LOSS FUNCTION[1]

By P. K. Bhattacharya[2]

Stanford University

1. Introduction and summary. Let X be a p-variate random vector, $p \geqq 3$, having a normal distribution with mean vector θ and covariance matrix $\sigma^2 I$. For estimating θ with loss function

$$L_1(\theta, \hat{\theta}) = (\theta - \hat{\theta})'(\theta - \hat{\theta}) = \|\theta - \hat{\theta}\|^2,$$

and known σ^2, Stein [2] has shown the inadmissibility of the usual estimator X by considering an alternative estimator with uniformly smaller risk than that of X, the improvement being substantial for θ close to the origin. The problem of estimating θ with the same loss function when σ^2 is unknown has been treated similarly by James and Stein [1], when an observation is available on another random variable which is distributed as $\sigma^2\chi_n^2$ independently of X. James and Stein have also demonstrated the inadmissibility of the usual estimator for θ under the loss function

$$L_2(\theta, \hat{\theta}) = (\theta - \hat{\theta})'D(\theta - \hat{\theta}) \tag{1}$$

for the case of known σ^2, where D is a diagonal matrix with unequal diagonal elements. This result has been proved even without the normality assumption, but no explicit formulas for alternative estimators have been given which improve on the usual estimator in some parts of the parameter space under normality.

In this note an estimator for θ is obtained for the case when σ^2 is unknown and the loss function is L_2. An upper bound for the risk function of this estimator is given, which always remains below the risk function of the usual estimator and is substantially smaller for θ close to the origin. This estimator coincides with the estimator given by James and Stein when the diagonal elements of D become equal. An application of this estimator gives an improvement on the least squares estimator of the parameter vector in a usual linear observational model with normal errors.

2. An estimator for the multivariate normal mean when the loss function is L_2. The different steps in constructing the estimator given in this section can be outlined as follows:

(i) The original problem with loss function L_2 is decomposed into a number of problems of different dimensionalities with loss function L_1.

Received 27 June 1963; revised 28 July 1966.

[1] Research supported by the National Science Foundation, Grant 214.

[2] Now at the University of Arizona.

(ii) The estimator of James and Stein is used for each of these problems followed by a suitable randomization. This yields a randomized estimator for the original problem whose risk function is computed easily from the risk functions of the corresponding estimators of James and Stein for the various problems with loss function L_1.

(iii) By the convexity of the loss function, a non-randomized estimator is then constructed which is everywhere better than this randomized estimator.

Let X and V have the joint density function

$$p(x, v) = (2\pi\sigma^2)^{-(p/2)} \exp[-(2\sigma^2)^{-1}(x - \theta)'(x - \theta)] \cdot[(2\sigma^2)^{n/2}\Gamma(n/2)]^{-1}e^{-(v/2\sigma^2)}v^{n/2-1},$$

$x \varepsilon E_p$ (the p-dimensional Euclidean space) and $0 < v < \infty$. We want to estimate θ from (X, V) with loss function L_2. Let λ_i be the ith diagonal element of the matrix D in (1) and suppose $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p > 0$. The usual estimator

$$\varphi_1(x, v) = x$$

has risk function

$$r(\theta, \sigma^2; \varphi_1) = \sigma^2 \textstyle\sum_{i=1}^{p} \lambda_i$$

for all θ.

Define $\alpha_i = \lambda_i - \lambda_{i+1} \geqq 0$, $i = 1, \cdots, p - 1$, and $\alpha_p = \lambda_p > 0$. Then

$$\lambda_i = \textstyle\sum_{j=i}^{p} \alpha_j, \qquad i = 1, \cdots, p, \tag{2}$$
$$\textstyle\sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} i\alpha_i .$$

Define $X'_{(i)} = (X_1, \cdots, X_i)$, $\theta'_{(i)} = (\theta_1, \cdots, \theta_i)$, and

$$f_i(x, v) = 1, \qquad i = 1,$$
$$= 1 - [(i - 2)/(n + 2)]\cdot v/\|x_{(i)}\|^2, \qquad i = 2, \cdots, p.$$

Using the result of James and Stein [1] for the estimation of $\theta_{(i)}$ from $(X_{(i)}, V)$ subject to the loss function

$$L^{(i)}(\theta_{(i)}, \hat{\theta}_{(i)}) = \alpha_i \|\theta_{(i)} - \hat{\theta}_{(i)}\|^2,$$

we see that when θ and σ^2 are the true parameters, the risk of the estimator $f_i(X, V)\cdot X_{(i)}$ is

$$\alpha_i E_{\theta,\sigma^2} \|f_i(X, V)\cdot X_{(i)} - \theta_{(i)}\|^2$$
$$= i\alpha_i\sigma^2, \qquad i = 1, \tag{3}$$
$$= \alpha_i\sigma^2[i - [n(i - 2)^2/(n + 2)]\cdot E(i - 2 + 2K_i)^{-1}], \quad i = 2, \cdots, p,$$

where K_i is a Poisson variate with mean $\|\theta_{(i)}\|^2/2\sigma^2$.

Now for every given (x, v), let $G_1(x, v), \cdots, G_p(x, v)$ be random variables whose marginal distributions are given by

$$P[G_i(x, v) = f_j(x, v)] = \alpha_j / \textstyle\sum_{k=i}^{p} \alpha_k, \qquad j = i, \cdots, p. \tag{4}$$

Any joint distribution of $G_1(x, v), \cdots, G_p(x, v)$ with marginal distributions satisfying (4) will serve our purpose.

Consider the randomized estimator

$$\varphi_2(x, v) = \begin{pmatrix} G_1(x, v) \cdot x_1 \\ \cdots \\ G_p(x, v) \cdot x_p \end{pmatrix}.$$

The risk function of this estimator is given by

$$\begin{aligned} r(\theta, \sigma^2; \varphi_2) &= E_{\theta,\sigma^2} E[\textstyle\sum_{i=1}^{p} \lambda_i \{G_i(X, V) \cdot X_i - \theta_i\}^2 \mid X, V] \\ &= \textstyle\sum_{i=1}^{p} E_{\theta,\sigma^2} (\sum_{j=i}^{p} \alpha_j) E[\{G_i(X, V) \cdot X_i - \theta_i\}^2 \mid X, V] \\ &= \textstyle\sum_{i=1}^{p} E_{\theta,\sigma^2} (\sum_{j=i}^{p} \alpha_j) \sum_{j=i}^{p} (\alpha_j / \sum_{k=i}^{p} \alpha_k) \{f_j(X, V) \cdot X_i - \theta_i\}^2 \\ &= \textstyle\sum_{i=1}^{p} E_{\theta,\sigma^2} \sum_{j=i}^{p} \alpha_j \{f_j(X, V) \cdot X_i - \theta_i\}^2 \\ &= \textstyle\sum_{i=1}^{p} \alpha_i E_{\theta,\sigma^2} \|f_i(X, V) \cdot X_{(i)} - \theta_{(i)}\|^2 \\ &= \sigma^2 \textstyle\sum_{i=1}^{p} i\alpha_i - [n\sigma^2/(n + 2)] \sum_{i=3}^{p} (i - 2)^2 \alpha_i \cdot E(i - 2 + 2K_i)^{-1}, \end{aligned}$$

using (2), (3) and (4). Since $\alpha_i \geqq 0$, $i = 1, \cdots, p$, and $\alpha_p > 0$,

$$[n\sigma^2/(n + 2)] \textstyle\sum_{i=3}^{p} (i - 2)^2 \alpha_i \cdot E(i - 2 + 2K_i)^{-1} > 0$$

and

$$r(\theta, \sigma^2, \varphi_2) < \sigma^2 \textstyle\sum_{i=1}^{p} \lambda_i = r(\theta, \sigma^2, \varphi_1).$$

Consider now the non-randomized estimator

$$\varphi_3(x, v) = \begin{pmatrix} h_1(x, v) \cdot x_1 \\ \cdots \\ h_p(x, v) \cdot x_p \end{pmatrix},$$

where $h_i(x, v) = \sum_{j=i}^{p} \alpha_j f_j(x, v) / \sum_{j=i}^{p} \alpha_j$. Then it follows from the convexity of the loss function that for all θ and σ^2

$$r(\theta, \sigma^2; \varphi_3) \leqq r(\theta, \sigma^2; \varphi_2).$$

Hence φ_3 is uniformly better than the usual estimator φ_1. The improvement achieved at $\theta = 0$ is

$$\begin{aligned} r(0, \sigma^2; \varphi_1) - r(0, \sigma^2; \varphi_3) &\geqq r(0, \sigma^2; \varphi_1) - r(0, \sigma^2; \varphi_2) \\ &= [n\sigma^2/(n + 2)] \textstyle\sum_{i=3}^{p} (i - 2)\alpha_i \\ &= [n\sigma^2/(n + 2)] \textstyle\sum_{i=3}^{p} \lambda_i \end{aligned}$$

by (2).

3. An improvement on the least squares estimator for the parameter vector

in a linear observational model with normal errors. Suppose Z is a known $N \times p$ matrix where $N > p \geqq 3$ and rank $(Z) = p$; Y is an N-dimensional random vector having a normal distribution with mean $Z\eta$ and covariance matrix $\sigma^2 I$ where $\eta' = (\eta_1, \cdots, \eta_p)$ and $\sigma^2 > 0$ are unknown, i.e., Y has the density function

$$p(y) = (2\pi\sigma^2)^{-N/2} \exp [-(2\sigma^2)^{-1}(y - Z\eta)'(y - Z\eta)]. \tag{5}$$

We want to estimate η subject to the loss function

$$L_3(\eta, \hat{\eta}) = (\eta - \hat{\eta})'\Delta(\eta - \hat{\eta})$$

where Δ is a known $p \times p$ symmetric positive definite matrix.

Let B be an $N \times (N - p)$ matrix such that $Z'B = 0$ and $B'B = I$. If we make the transformation $U = (Z'Z)^{-1}Z'Y$ and $W = B'Y$, and let $V = W'W$, $n = N - p$, then U is distributed as p-variate normal distribution with mean vector η and covariance matrix $\sigma^2(Z'Z)^{-1}$, and V is distributed independently of U as $\sigma^2\chi_n^2$.

We can now find a $p \times p$ non-singular matrix C such that

$$Z'Z = C'C \quad \text{and} \quad \Delta = C'DC, \tag{6}$$

where D is a diagonal matrix whose ith diagonal element λ_i, $i = 1, \cdots, p$, are such that $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p > 0$.

Now make the transformation $CU = X$ and $C\eta = \theta$. Then X has a p-variate normal distribution with mean vector θ and covariance matrix $\sigma^2 I$, and V is distributed independently of X as $\sigma^2\chi_n^2$, as in Section 2. Consider now the problem of estimating θ subject to loss function L_2 in which the matrix D is the same as in (6), and define for every estimate $\hat{\theta}$ for θ, an estimate $\hat{\eta} = C^{-1}\hat{\theta}$ for η. Then

$$L_2(\theta, \hat{\theta}) = L_3(\eta, \hat{\eta}).$$

Thus the problem of estimating η from (U, V) subject to loss function L_3 is the same as that of estimating θ from (X, V) subject to loss function L_2. In this correspondence, the usual estimator $\varphi_1(X, V) = X$ for θ is equivalent to the estimator

$$\psi_1(Y) = C^{-1}\varphi_1(C(Z'Z)^{-1}Z'Y, V) = (Z'Z)^{-1}Z'Y$$

for η which is the least squares estimator. If we now define

$$\psi_3(Y) = C^{-1}\varphi_3(C(Z'Z)^{-1}Z'Y, V),$$

then it follows from Section 2 and from the correspondence mentioned above, that ψ_3 is better than the least squares estimator ψ_1.

We have been assuming that Δ is a positive definite matrix in order to ensure that $\lambda_p > 0$. However if Δ is positive semi-definite with $\lambda_r > 0$ for some $r \geqq 3$ and $\lambda_{r+1} = \cdots = \lambda_p = 0$, then $\alpha_r > 0$ and $\alpha_{r+1} = \cdots = \alpha_p = 0$. In that case we define $h_1(x, v), \cdots, h_r(x, v)$ in the same way as in Section 2 and fix

$h_{r+1}(x, v), \cdots, h_p(x, v)$ arbitrarily. The resulting estimator ψ_3 will behave in the same way as mentioned above.

For simplicity let us assume $\Delta = I$. Then the quantities $\lambda_1, \cdots, \lambda_p$ entering into the estimator ψ_3 are the latent roots of the inverse of the matrix of normal equations $Z'Z$. Suppose we believe in the loss function under consideration and agree to use the estimator ψ_3. Suppose further that we have an approximate knowledge of $\eta_1/\sigma, \cdots, \eta_p/\sigma$. Let us now look at the design aspect of the problem. Our optimality criterion now becomes the minimization of

$$\sum_{i=1}^{p} \lambda_i(Z) - [(N-p)/(N-p+2)] \sum_{i=3}^{p} (i-2)^2 \alpha_i(Z) \cdot E(i-2+2K_i)^{-1}$$

over all available $N \times p$ observational matrices Z where $\lambda_1(Z), \cdots, \lambda_p(Z)$ are the latent roots of $(Z'Z)^{-1}$ and $\alpha_i(Z)$ are related to $\lambda_i(Z)$ by (2). If we use a design which is optimal in this sense and if we estimate η by ψ_3 we shall do better than using a design which minimizes tr $(Z'Z)^{-1}$ subject to the condition of estimability and estimating η by the least squares estimator.

When Z is a random matrix such that $Z'Z$ is positive definite with probability 1 and the conditional distribution of Y given Z has the density function given in (5), we consider the conditional risk given Z and then the argument for fixed Z applies to the conditional risk functions for almost all Z. Thus by treating the data as if Z were non-stochastic and computing the estimator ψ_3, we do better than the least squares estimator in terms of the conditional risk functions given Z for almost all Z and therefore we do better also in terms of the unconditional risk function.

4. Generalizations. (i) Let Z be a random vector whose probability distribution belongs to the family $\{P_\omega, \omega \varepsilon \Omega\}$. Let $\theta_i(\omega)$, $i = 1, \cdots, p$, be real-valued functions on Ω and let $\theta_{(i)}(\omega)'$ denote the row-vector $(\theta_1(\omega), \cdots, \theta_i(\omega))$, $i = 1, \cdots, p$. Consider the problems of estimating $\theta_{(i)}(\omega)$ subject to the loss function

$$L_i(\theta_{(i)}, \hat{\theta}_{(i)}) = \|\theta_{(i)} - \hat{\theta}_{(i)}\|^2,$$

$i = 1, \cdots, p$. Suppose $\theta_{(i)}^*(Z)' = (\theta_1^*(Z), \cdots, \theta_i^*(Z))$ and $\hat{\theta}_{(i)}(Z)' = (\theta_{i1}(Z), \cdots, \theta_{ii}(Z))$, $i = 1, \cdots, p$, are such that for each $i = 1, \cdots, p$, $\hat{\theta}_{(i)}(Z)$ is at least as good an extimate of $\theta_{(i)}$ under the loss function L_i as $\theta_{(i)}^*(Z)$ is and that for some i, $\hat{\theta}_{(i)}(Z)$ is strictly better than $\theta_{(i)}^*(Z)$ for some $\omega \varepsilon \Omega$, i.e.

$$E_\omega \|\hat{\theta}_{(i)}(Z) - \theta_{(i)}(\omega)\|^2 \leqq E_\omega \|\theta_{(i)}^*(Z) - \theta_{(i)}(\omega)\|^2,$$

$i = 1, \cdots, p$, and for all $\omega \varepsilon \Omega$, and the inequality is strict for some i and for some $\omega \varepsilon \Omega$. Consider now the problem of estimating $\theta(\omega)$ subject to the loss function

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})'D(\theta - \hat{\theta}),$$

where $D = \operatorname{diag}\{\lambda_1, \cdots, \lambda_p\}$, $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_p > 0$. Following the arguments in Section 2, it can be shown that $\phi(Z)$ defined as

$$\phi(Z)' = (\phi_1(Z), \cdots \phi_p(Z))$$

where $\phi_i(Z) = \sum_{j=i}^{p} \alpha_j \hat{\theta}_{ij}(Z) / \sum_{j=i}^{p} \alpha_j$, $\alpha_1, \cdots, \alpha_p$ being related to $\lambda_1, \cdots, \lambda_p$ by (2) is a strictly better estimate of $\theta(\omega)$ under the loss function L than $\theta^*(Z) = \theta^*_{(p)}(Z)$.

(ii) In Section 2, the random vector X is p-variate normal with mean vector θ and covariance matrix $\sigma^2 I$, and the loss function is

$$L_2(\theta, \hat{\theta}) = (\theta - \hat{\theta})' D(\theta - \hat{\theta})$$

where D is a diagonal matrix. However, no additional difficulty is presented if in the above formulation of the problem the covariance matrix of X is $\sigma^2 A$ instead of $\sigma^2 I$ and the diagonal matrix D involving in the loss function is replaced by B where A and B are arbitrary but known positive definite matrices. How this slightly more general problem can be reduced to the simpler problem of Section 2, is illustrated in Section 3.

5. Acknowledgment. The author is indebted to Ingram Olkin for some helpful comments and to the referee at whose suggestion Section 4 is written.

REFERENCES

[1] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berk. Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.

[2] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berk. Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.