COMMUN. STATIST.-THEORY METH., 19(5), 1803-1810 (1990)

UNBIASED ESTIMATION OF THE COMMON MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

K. Krishnamoorthy
Department of Statistics
Temple University
Philadelphia, PA 19122

Vijay K. Rohatgi
Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio 43403

Key Words and Phrases: Unbiased estimator; Common mean; Multivariate normal; Multiple regression.

ABSTRACT

The problem of unbiased estimation of the common mean of a multivariate normal population is considered. An unbiased estimator is proposed which has a smaller variance than the usual estimator over a large part of the parameter space.

1. INTRODUCTION

Let $(u_1, ..., u_{p+1})'$ be a $(p+1)$-variate normal random vector with mean $(\mu, ..., \mu)'$ and covariance matrix Θ. The problem of estimation of μ was first considered by Halperin (1961) who derived the maximum likelihood estimators of μ and Θ. This problem is equivalent to the estimation of the intercept in multiple regression with random regressors. Let $A = (a_{ij})$ be a $(p+1) \times (p+1)$ matrix with $a_{i1} = 1$ for

$i = 1, ..., p+1$, $a_{ii} = -1$ for $i = 2, ..., p+1$ and $a_{ij} = 0$ elsewhere. Consider the transformation $(y, x_1, ..., x_p)' = (y, X')' = A(u_1, ..., u_{p+1})'$. Then $(y, X')'$ has a $(p+1)$-variate normal distribution with mean $(\mu, 0, ..., 0)'$ and covariance matrix $\Sigma = A\Theta A'$. In this formulation the problem arises in discrete event simulation where one uses control variables to reduce variance by exploiting correlations between output response y and associated auxiliary variables $x_1, ..., x_p$ observed during the course of each simulation run. Typically the mean μ_i of x_i is known (and may be taken to be zero) and the object is to estimate $\mu = E(y)$. This problem has been considered, amongst others, by Baranchik (1973) and Gleser (1987).

For ease in comparison we will use the regression formulation in this paper. Let

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \sigma'_{Xy} \\ \sigma_{Xy} & \Sigma_{XX} \end{pmatrix}$$

be unknown. Suppose that we have n independent observations on (y, X'). Let $(\bar{y}, \bar{X}') = n^{-1}\sum_{i=1}^{n}(y_i, X'_i)$ be the sample mean vector and

$$W = \begin{pmatrix} w_{yy} & w'_{Xy} \\ w_{Xy} & W_{XX} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n}(y_i - \bar{y})^2 & \sum_{i=1}^{n}(y_i - \bar{y})(X_i - \bar{X})' \\ \sum_{i=1}^{n}(y_i - \bar{y})(X_i - \bar{X}) & \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})' \end{pmatrix} \tag{1.1}$$

be the sample cross-product matrix. Then the usual estimators of μ are $\bar{y}$ and the maximum likelihood estimator given in (1.2) below.

Suppose first that Σ is known and let $\beta = \Sigma_{XX}^{-1}\sigma_{Xy}$. Consider the estimator $\bar{y}(\alpha) = \bar{y} - \sum_{i=1}^{p}\alpha_i\bar{x}_i$ where $\alpha_1, ..., \alpha_p$ are real numbers. Then $var(\bar{y}(\alpha))$ is minimized for $\alpha = \beta$ and $\bar{y}(\beta)$ is the best linear unbiased estimator of μ. When β is unknown replacing it by $b = W_{XX}^{-1}w_{Xy}$ leads to the maximum likelihood estimator

$$\bar{y}(b) = \bar{y} - b'\bar{X}. \tag{1.2}$$

The estimator $\bar{y}(b)$ is unbiased for μ and has variance

$$var(\bar{y}(b)) = n^{-1}(1 + \frac{p}{n-p-2})\sigma_{yy.X} \tag{1.3}$$

for $n > p+2$ where $\sigma_{yy.X} = \sigma_{yy}(1 - \rho_{y.X}^2)$ and $\rho_{y.X}^2 = (\sigma'_{Xy}\Sigma_{XX}^{-1}\sigma_{Xy})/\sigma_{yy}$. It is easy to check that

$$var(\bar{y}(b)) < var(\bar{y}) \text{ if and only if } n \geq p+3 \text{ and } \rho_{y.X}^2 > p/(n-2). \tag{1.4}$$

A typical approach has been to improve upon b as an estimator of β. The discussion has been limited to the classes of estimators that have smaller variance than $\bar{y}(b)$. While this appraoch shows that $\bar{y}(b)$ is inadmissible, it has not led to any practical alternative to $\bar{y}$. In this paper we focus attention on $\bar{y}$ and suggest an unbiased estimator $\hat{\mu}_c$ which, with proper choice of c, dominates $\bar{y}$ when a positive lower bound for the squared population multiple correlation is known.

2. PROPOSED ESTIMATOR AND ITS VARIANCE

The estimator b of β in (1.2) uses the cross product matrix $\sum(X_i - \bar{X})(X_i - \bar{X})'$ even though the mean vector of X_i is known to be zero. It is reasonable to expect in this case that

$$b_o = c(\sum_{i=1}^{n} X_i X_i')^{-1} w_{Xy} = c(W_{XX} + n\bar{X}\bar{X}')^{-1} w_{Xy}, \tag{2.1}$$

where c is a constant, will lead to a better estimator of μ. Let

$$\hat{\mu}_c = \bar{y} - b_o'\bar{X}. \tag{2.2}$$

Then we have

Theorem 2.1. The estimator $\hat{\mu}_c$ is unbiased for μ.

Proof. Using the identity

$$(W_{XX} + n\bar{X}\bar{X}')^{-1} = W_{XX}^{-1} - \frac{nW_{XX}^{-1}\bar{X}\bar{X}'W_{XX}^{-1}}{1 + n\bar{X}'W_{XX}^{-1}\bar{X}}$$

we can write $\hat{\mu}_c$ as

$$\hat{\mu}_c = \bar{y} - c(\frac{1}{1+T^2})b'\bar{X} \tag{2.3}$$

where $T^2 = n\bar{X}'W_{XX}^{-1}\bar{X}$. Since $\bar{X} \overset{d}{\sim} - \bar{X}$and $T^2 = n(-\bar{X})'W_{XX}^{-1}(-\bar{X})$ we have

$$E\left(\frac{1}{1+T^2}\right) b'\bar{X} = -E\left(\frac{1}{1+T^2}\right) b'\bar{X}$$

and the assertion follows.

In order to compute $var(\hat{\mu}_c)$ we need the following lemmas.

Lemma 2.1. (i) The conditional distribution of b given $(X_1, ..., X_n)$ is $N_p(\beta, \sigma_{yy.X} W_{XX}^{-1})$.
(ii) $Z'Z = n\bar{X}'\Sigma_{XX}^{-1}\bar{X} \sim \chi_p^2$ independently of $\bar{X}$ (or Z).
(iii)$V = (\bar{X}'\Sigma_{XX}^{-1}\bar{X})/(\bar{X}'W_{XX}^{-1}\bar{X}) \sim \chi_{n-p}^2$ independently of $\bar{X}$(or Z).
(iv) $T^2 = Z'Z/V \sim \chi_p^2/\chi_{n-p}^2$ where the chi-squared random variables in the numerator and the denominator are independent.

For the proof of Lemma 2.1 see Muirhead (1982, Chapters 1 and 3).

Lemma 2.2. Let $X \sim \chi_p^2$ and Y be independent of X. Suppose that f is a real valued function and $k \geq 1$ is an integer. Write $(a)_k = a(a+1)...(a+k-1)$. Then

$$E\{X^k f(X+Y)\} = 2^k (p/2)_k E\{f(Z+Y)\},$$

provided the indicated expectations exist. Here Z is a χ_{p+2k}^2 random variable.

Lemma 2.2 is easy to prove.

Lemma 2.3. Let $Z \sim N_p(0, I)$ where I is the $p \times p$ identity matrix, and let $A = (a_{ij})$ be a $p \times p$ square matrix. Then for any real valued function f

$$E\{f(Z'Z)(Z'AZ)\} = tr(A)E\{f(\chi_{p+2}^2)\}$$

provided the indicated expectations exist.

Proof. We note that

$$E\{f(Z'Z)(Z'AZ)\} = E\{f(\sum_{i=1}^p z_i^2)(\sum_{i=1}^p a_{ii} z_i^2 + \sum_{i \neq j} a_{ij} z_i z_j)\}$$

Since z_i's are *i.i.d.* $N(0,1)$ random variables, the *rhs* is equal to

$$E\{f(\sum_{i=1}^p z_i^2)(\sum_{i=1}^p a_{ii} z_i^2)\} = a_{11} E(z_1^2 f(\sum_{i=1}^p z_i^2)) + ... + a_{pp} E(z_p^2 f(\sum_{i-1}^p z_i^2)).$$

Also, as z_i^2's are independent χ_1^2 random variables, it follows from Lemma 2.2 that, for each j, $E(z_j^2 f(\sum_{i=1}^p z_i^2)) = E(f(z_{j*}{}^2 + \sum_{i \neq j} z_i^2)) = Ef(\chi_{p+2}^2)$, where $z_{j*}{}^2 \sim \chi_3^2$ independently of z_i^2's. Thus, the result follows from the above equation.

Lemma 2.4. For $r > 0$,

$$E\left[\frac{\beta'\bar{X}\bar{X}'\beta}{(1+T^2)^r}\right] = \frac{\sigma_{yy}\rho_{y.X}^2 \Gamma(\frac{n-p}{2} + r)\Gamma(\frac{n+2}{2})}{n\Gamma(\frac{n-p}{2})\Gamma(\frac{n+2}{2} + r)}.$$

Proof. Let $A = \Sigma_{XX}^{-1/2}\sigma_{Xy}\sigma'_{Xy}\Sigma_{XX}^{-1/2}$. Then using Lemmas 2.1 and 2.3 we have

$$E\left[\frac{\beta'\bar{X}\bar{X}'\beta}{(1+T^2)^r}\right] = n^{-1}E\left[\frac{(Z'AZ)V^r}{(Z'Z+V)^r}\right] = n^{-1}tr(A)E\left[\frac{\chi^2_{n-p}}{\chi^2_{p+2}+\chi^2_{n-p}}\right]^r$$

which proves the lemma.

Theorem 2.2. For $n > p$ the variance of $\hat{\mu}_c$ is given by

$$Var(\hat{\mu}_c) = n^{-1}\sigma_{yy} - 2c\frac{n-p}{n(n+2)}\sigma_{yy}\rho^2_{y.X} + \frac{c^2(n-p)}{n^2}\left[\frac{p}{n+2} + \frac{n-2p}{n+4}\rho^2_{y.X}\right]\sigma_{yy}. \quad (2.4)$$

Proof. Rewrite $\hat{\mu}_c$ as

$$\hat{\mu}_c = (\bar{y} - \beta'\bar{X}) - \left[\frac{cb'}{1+T^2} - \beta'\right]\bar{X}.$$

Using the fact that $\bar{y} - \beta'\bar{X}$ is independent of $(\bar{X}, b')$ we get

$$var(\hat{\mu}_c) = E(\bar{y}-\beta'\bar{X}-\mu)^2 + c^2E\left(\frac{b'\bar{X}\bar{X}'b}{(1+T^2)^2}\right) - 2cE\left(\frac{\beta'\bar{X}\bar{X}'b}{1+T^2}\right) + \beta'E(\bar{X}\bar{X}')\beta. \quad (2.5)$$

Now

$$\begin{aligned} E(\bar{y} - \beta'\bar{X} - \mu)^2 &= n^{-1}\sigma_{yy}(1 - \rho^2_{y.X}), \\ \beta'E(\bar{X}\bar{X}')\beta &= n^{-1}\sigma_{yy}\rho^2_{y.X} \end{aligned}$$

and

$$\begin{aligned} E\left[\frac{\beta'\bar{X}\bar{X}'b}{1+T^2}\right] &= E\left[E\left(\frac{\beta'\bar{X}\bar{X}'b}{1+T^2}|X_1,...,X_n\right)\right] \\ &= E\left[\frac{\beta'\bar{X}\bar{X}'\beta}{1+T^2}\right] = \frac{(n-p)}{n(n+2)}\sigma_{yy}\rho^2_{y.X}. \end{aligned}$$

The last two reductions follow respectively from Lemma 2.1(i) and Lemma 2.4. Similarly, using Lemmas 2.1 and 2.4,

$$\begin{aligned} E\left[\frac{b'\bar{X}\bar{X}'b}{(1+T^2)^2}\right] &= E\left(E\left[\frac{b'\bar{X}\bar{X}'b}{(1+T^2)^2}|X_1,...,X_n\right]\right) \\ &= n^{-1}\sigma_{yy.X}E\left(\frac{T^2}{(1+T^2)^2}\right) + E\left(\frac{\beta'\bar{X}\bar{X}'\beta}{(1+T^2)^2}\right] \\ &= \frac{p(n-p)\sigma_{yy.X}}{n^2(n+2)} + \frac{(n-p)(n-p+2)}{n(n+2)(n+4)}\sigma_{yy.X}\rho^2_{y.X}. \end{aligned}$$

Substituting in (2.5) we get (2.4) after some simplification.

Corollary. For $c > 0$

$$var(\hat{\mu}_c) < var(\bar{y}) \text{ if and only if } \rho^2_{y.X} > n^{-1}p\left(\tfrac{p}{n} + \tfrac{2}{c} - \tfrac{n-p+2}{n+4}\right)^{-1} \quad (2.6)$$

For $\alpha = .1$ and $c_o = 2n(n+4)/[9p(n+4) + n(n-p+2)]$

Table 1: Efficiencies of $\hat{\mu}_1$ and $\bar{y}(b)$ relative to $\bar{y}$.

$E_1 = nvar(\hat{\mu}_1)/\sigma_{yy}$, $E_2 = nvar(\bar{y}(b))/\sigma_{yy}$

$\rho_{y.x}$	E_1	E_2	E_1	E_2	E_1	E_2
	(n=10,	p=2)	(n=10,	p=4)	(n=10,	p=6)
0.1	1.1234	1.3200	1.1909	1.9800	1.1928	3.9600
0.2	1.0937	1.2800	1.1634	1.9200	1.1710	3.8400
0.3	1.0442	1.2133	1.1177	1.8200	1.1349	3.6400
0.4	0.9749	1.1200	1.0537	1.6800	1.0842	3.3600
0.5	0.8857	1.0000	0.9714	1.5000	1.0190	3.0000
0.6	0.7768	0.8533	0.8709	1.2800	0.9394	2.5600
0.7	0.6480	0.6800	0.7520	1.0200	0.8453	2.0400
0.8	0.4994	0.4800	0.6149	0.7200	0.7368	1.4400
0.9	0.3310	0.2533	0.4594	0.3800	0.6137	0.7600
	(n=20,	p=3)	(n=20,	p=4)	(n=20,	p=10)
0.1	1.1054	1.1880	1.1349	1.2729	1.2182	2.2275
0.2	1.0739	1.1520	1.1033	1.2343	1.1909	2.1600
0.3	1.0214	1.0920	1.0505	1.1700	1.1455	2.0475
0.4	0.9480	1.0080	0.9767	1.0800	1.0818	1.8900
0.5	0.8535	0.9000	0.8818	0.9643	1.0000	1.6875
0.6	0.7380	0.7680	0.7658	0.8229	0.9000	1.4400
0.7	0.6016	0.6120	0.6287	0.6557	0.7818	1.1475
0.8	0.4442	0.4320	0.4705	0.4629	0.6455	0.8100
0.9	0.2657	0.2280	0.2913	0.2443	0.4909	0.4275
	(n=30,	p=4)	(n=30,	p=8)	(n=30,	p=15)
0.1	1.0977	1.1550	1.1726	1.3860	1.2250	2.1323
02.	1.0658	1.1200	1.1404	1.3440	1.1969	2.0677
0.3	1.0126	1.0617	1.0868	1.2740	1.1500	1.9600
0.4	0.9381	0.9800	1.0116	1.1760	1.0844	1.8092
0.5	0.8423	0.8750	0.9151	1.0500	1.0000	1.6154
0.6	0.7252	0.7467	0.7970	0.8960	0.8969	1.3785
0.7	0.5869	0.5950	0.6575	0.7140	0.7750	1.0985
0.8	0.4272	0.4200	0.4966	0.5040	0.6344	0.7754
0.9	0.2463	0.2217	0.3142	0.2660	0.4750	0.4092

$$var(\hat{\mu}_{c_o}) < var(\bar{y}) \text{ for } \rho^2_{y.X} > .1.$$

Comparing $\hat{\mu}_{c_o}$ with $\bar{y}(b)$ we find that when $n = 20$, $p = 4$, $\hat{\mu}_{c_o}$ has smaller variance than $\bar{y}(b)$ for $\rho^2_{y.X} < .502$. Thus $\hat{\mu}_{c_o}$ dominates both $\bar{y}(b)$ and $\bar{y}$ for $.1 < \rho^2_{y.X} < .5$.

3. CONCLUDING REMARKS

The unbiased estimator $\hat{\mu}_1$ has a smaller variance than $\bar{y}$ over a wide range of parameter values. Thus for $n = 20$, $p = 4$, $var(\hat{\mu}_1) < var(\bar{y})$ for $|\rho_{y.X}| > .37$ and for $n = 20$, $p = 10$,$var(\hat{\mu}_1) < var(\bar{y})$ for $|\rho_{y.X}| > .50$. In fact if $p/n = \lambda$, and n is large then for $\lambda \leq .50$ the parameter set over which $var(\hat{\mu}_1) < var(\bar{y})$ includes the set $|\rho_{y.X}| > .50$. In the worst case when $n = p + 3$, the smallest possible value of n, the parameter set over which $var(\hat{\mu}_1) < var(\bar{y})$ includes the set $|\rho_{y.X}| > .58$.

In Table 1 we have computed the efficiencies of $\hat{\mu}_1$ and $\bar{y}$ for some selected values of n and p, and $\rho_{y.X} = .1(.1).9$. It is clear that $\bar{y}(b)$ dominates $\hat{\mu}_1$ in risk only for values of $\rho_{y.X}$ close to 1. On the other hand $\hat{\mu}_1$ dominates both $\bar{y}$ and $\bar{y}(b)$ for moderate values of $\rho_{y.X}$. Even for smaller values of $\rho_{y.X}$ the loss in efficiency when $\hat{\mu}_1$ is used instead of $\bar{y}$ is not substantial. Thus when $n = 20$, $p = 4$, the loss is 13.49% for $\rho_{y.X} = 1$, and 10.33% for $\rho_{y.X} = .2$ but the gain in efficiency is 37.13% for $\rho_{y.X} = .70$.

If a positive lower bound for $\rho^2_{y.X}$ is known then we can choose c and use $\hat{\mu}_c$ to dominate $\bar{y}$ in risk. Suppose $|\rho^2_{y.X}| > \alpha$, where $0 < \alpha < 1$ is known. Then we can choose c such that

$$\frac{p}{n}\left[\frac{p}{n} + \frac{2}{c} - \frac{n-p+2}{n+4}\right]^{-1} \leq \alpha$$

giving

$$0 < c \leq \frac{2\alpha n(n+4)}{p(1-\alpha)(n+4) + \alpha n(n-p+2)}.$$

In particular with $c_0 = 2\alpha n(n+4)/[p(1-\alpha)(n+4) + \alpha n(n-p+2)]$

$$var(\hat{\mu}_{c_0}) < var(\bar{y}) \text{ for } \rho^2_{y.X} > \alpha.$$

Similarly $\hat{\mu}_{c_o}$ dominates both $y(b)$ and y for $.1 < \rho^2_{y.X} < .64$ when $n = 20$, $p = 10$.

ACKNOWLEDGEMENT

The authors wish to thank the referee and the associate editor Professor R.P. Gupta for their comments and suggestions.

REFERENCES

BARANCHIK, A.J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. *Ann.Statist.* 1, 312-321.

HALPERIN, M. (1961). Almost linearly-optimum linear combination of unbiased estimates. *Amer.Stat.Assoc.* 56, 36-43.

GLESER, L.J. (1987). Improved estimation of mean response in simulation when control variates are used. Tech. Report #87 – 34. Department of Statistics, Purdue University.

MUIRHEAD, R.J.(1982). Aspects of Multivariate Statistical Theory. New York: John Wiley.

Received January 1989; Revised April 1990.

Recommended by R. P. Gupta, Dalhousie University, Halifax, CANADA.

Refereed by L. Edward, University of North Carolina, Chapel Hill, NC. and Anonymously.