Statistics & Probability Letters 19 (1994) 33–38
North-Holland

3 January 1994

Unbiased equivariant estimation of a common normal mean vector with one observation from each population

K. Krishnamoorthy and Nabendu Pal

Department of Statistics, University of Southwestern Louisiana, Lafayette, LA, USA

Received November 1992
Revised February 1993

Abstract: Let X_1 be a random observation from a p-variate normal population with mean vector θ and covariance matrix proportional to identity matrix, $N_p(\theta, \sigma_1^2 I_p)$. In addition to X_1, there is another observation X_2 from $N_p(\theta, \sigma_2^2 I_p)$. In this note, an unbiased estimator which combines both X_1 and X_2 is developed and its risk behavior is studied. Then, assuming that σ_1^2 is known, a motivation for the best shrinkage estimator in a class of estimators that shrink X_1 toward X_2 is given. It is shown that such shrinkage estimators are unbiased and location equivariant. Also, for a shrinkage estimator from this class the risk improvements over X_1 and the one that shrinks toward the origin are studied.

Keywords: Location equivariant; loss function; unbiased estimator.

1. Introduction

Suppose that there are two independent observations X_1 and X_2 such that the first one is from a p-variate normal population with mean vector θ and covariance matrix $\sigma_1^2 I_p$, $N_p(\theta, \sigma_1^2 I_p)$ and the second one is from $N_p(\theta, \sigma_2^2 I_p)$. The parameters $\theta \in \mathbb{R}^p$, $\sigma_1^2 > 0$, and $\sigma_2^2 > 0$ are unknown. In this paper, the problem of estimation of the common mean vector based on X_1 and X_2 is considered. The merit of an estimator is evaluated by the sum of the squared error losses

$$L(\theta, \hat{\theta}) = \sum_{i=1}^{p} \left(\theta_i - \hat{\theta}_i\right)^2 = \|\theta - \hat{\theta}\|^2, \tag{1.1}$$

where $\hat{\theta} = (\hat{\theta}_1, \ldots, \theta_p)$ is an estimator of θ.

While both σ_1^2 and σ_2^2 are unknown, one may estimate θ based on some prior knowledge on the variabilities of X_1 and X_2 by

$$\delta_\eta = \eta X_1 + (1 - \eta) X_2, \tag{1.2}$$

where η is a known number in the interval (0, 1). George (1992) proved that the estimator δ_η is dominated by a shrinkage version of it when $p \geqslant 3$. This result is further improved by Krishnamoorthy

Correspondence to: Nabendu Pal, Department of Statistics, University of Southwestern Louisiana, P.O. Box 41006, Lafayette, LA 70504-1006, USA.

0167-7152/94/$07.00 © 1994 – Elsevier Science B.V. All rights reserved
SSDI 0167-7152(93)E0079-9

(1992) and extended to the k-sample situation by Sarkar (1992). However, there is no estimator based on X_1 and X_2 that dominates either of them under the loss (1.1).

Ideally, the problem of interest here is to develop a combined estimator that dominates both X_1 and X_2. If σ_1^2 and σ_2^2 are known, then

$$\delta_{(\sigma_1^2,\sigma_2^2)} = \sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}X_2 + \sigma_2^2(\sigma_1^2+\sigma_2^2)^{-1}X_1 \tag{1.3}$$

is the best linear unbiased estimator (BLUE) and dominates both X_1 and X_2. Therefore, when σ_1^2 and σ_2^2 are unknown, it is intuitive that a combined estimator of the form $wX_2+(1-w)X_1$, where w is an estimator of $\sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}$, would be preferable to both X_1 and X_2. However, such an estimator proposed in the following section neither dominates X_1 and X_2 uniformly under the loss (1.1) for any p. Interestingly, when σ_1^2 is known, a combined estimator of the form $wX_2+(1-w)X_1$ turns out to be the shrinkage estimator that shrinks X_1 toward X_2. This combined as well as shrinkage estimator not only improves X_1 but also is unbiased and location equivariant. As a result, its risk is independent of θ and it improves X_1 significantly. To understand the impact of X_2 on improving X_1, this combined estimator is compared with the one that shrinks X_1 toward the origin. The study shows that the former dominates the latter over a large parameter space.

The example given by George (1992) can be used as a motivation for this problem too. Suppose that the value of each of p parcels of real estate was assesed by an assesser who is known for several years and as a result the variability of his assessment X_1 is known. Also to get a second opinion, another assessor is called and the variability of his assessment X_2 is unknown. In this setup, one may consider X_1 as a reliable estimate of the true value as its varibility is know. Now the goal is to improve X_1 using the estimate X_2. Again, as pointed out by George, it may be necessary to use transformed units to obtain constant variance for each assessor.

Finally, it is to be mentioned that several authors have considered related versions of the above problem when there are more than one observations from each of the populations. For details see Chiou and Cohen (1985), Kubokawa (1989) and Krishnamoorthy (1991).

2. Main result

Suppose that both σ_1^2 and σ_2^2 are known. Then, $(\sigma_2^2X_1+\sigma_1^2X_2)/(\sigma_1^2+\sigma_2^2)$ is the best linear unbiased estimator of θ. Therefore, if the variances are unknown, replacing $\sigma_1^2(\sigma_1^2+\sigma_2^2)$ and $\sigma_2^2/(\sigma_1^2+\sigma_2^2)$ respectively by suitable estimators one can develop an estimator of θ. The following lemma is needed to develop such an estimator and also to study its statistical properties.

Lemma 2.1. *Let* $X_1 \sim \mathrm{N}_p(\theta, \sigma_1^2 I_p)$ *independently of* $X_2 \sim \mathrm{N}_p(\theta, \sigma_2^2 I_p)$ *and* $Z = X_1 - X_2$. *Then,*

(i) $$\begin{pmatrix} X_1 \\ Z \end{pmatrix} \sim \mathrm{N}_{2p}\left[\begin{pmatrix}\theta \\ 0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 I_p & \sigma_1^2 I_p \\ \sigma_1^2 I_p & (\sigma_1^2+\sigma_2^2)I_p\end{pmatrix}\right],$$

(ii) $$X_1 | Z \sim \mathrm{N}_p\left(\theta + \sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}Z,\ \sigma_1^2\sigma_2^2(\sigma_1^2+\sigma_2^2)^{-1}I_p\right),$$

(iii) $$E\left[X_1'Z\|Z\|^{-2}\right] = \sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1},$$

(iv) $$E\left[-X_2'Z\|Z\|^{-2}\right] = \sigma_2^2(\sigma_1^2+\sigma_2^2)^{-1},$$

(v) $$E(\theta'Z/\|Z\|)^2 = \theta'\theta/p,$$

(vi) $$E(X_1'Z/\|Z\|)^2 = \sigma_1^2(p\sigma_1^2+\sigma_2^2)(\sigma_1^2+\sigma_2^2)^{-1} + \theta'\theta/p.$$

Proof. (i) and (ii) are well known.

(iii) The conditional expectation of $(X_1'Z\,\|Z\|^{-2})$ given Z yields

$$E\left[\theta' Z\|Z\|^{-2}\right] + \sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1} = \sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}$$

since $Z\,\|Z\|^2$ is distributed as $-Z\,\|Z\|^2$ implies that $E(\theta' Z\,\|Z\|^{-2}) = 0$.

(iv) follows from (iii).

(v) $E(\theta' Z/\|Z\|)^2 = E(\theta' ZZ'\theta/\|Z\|^2)$ and as $Z \sim N_p(0, (\sigma_1^2+\sigma_2^2)I_p)$ it can be easily verified that $E(ZZ'/\|Z\|^2) = I_p/p$.

(vi) Noting that $(X_1'Z/\|Z\|)^2 = \operatorname{tr}(X_1X_1'ZZ'/\|Z\|^2)$ and taking conditional expectation given Z yields

$$E\ \operatorname{tr}\left\{\left[\theta\theta' + \frac{\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)}I_p + \frac{\sigma_1^4}{(\sigma_1^2+\sigma_2^2)}ZZ'\right]ZZ'/\|Z\|^2\right\} = \frac{\sigma_1^2(p\sigma_1^2+\sigma_2^2)}{\sigma_1^2+\sigma_2^2} + \frac{\theta'\theta}{p}. \qquad \square$$

From the above lemma (iii) and (iv) it is clear that one can estimate $\sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}$ and $\sigma_2^2(\sigma_1^2+\sigma_2^2)^{-1}$ respectively by $X_1'(X_1-X_2)/\|X_1-X_2\|^2$ and $X_2'(X_2-X_1)/\|X_1-X_2\|^2$ and hence one can propose

$$\hat{\theta}_s = \frac{X_1'(X_1-X_2)}{\|X_1-X_2\|^2}X_2 + \frac{X_2'(X_2-X_1)}{\|X_1-X_2\|^2}X_1 = X_1 - \frac{X_1'Z}{\|Z\|^2}Z \tag{2.1}$$

as an estimator of θ. Also, its expectation

$$E\hat{\theta}_s = \theta - E\left\{E\left[\frac{X_1'Z}{\|Z\|^2}Z|ZZ\right]\right\} = \theta - E\left[\frac{\theta' Z}{\|Z\|^2}Z\right] = (p-1)\theta/p.$$

Adjusting for bias yields that $p\hat{\theta}_s/(p-1)$ is an unbiased estimator of θ. In order to compare various estimators in the class $\{c\hat{\theta}_s$: c is a real $\neq 0\}$, the risk of $c\hat{\theta}_s$ is obtained in the following theorem.

Theorem 2.1. *The risk of $c\hat{\theta}_s$, where c is a real number, is given by*

$$R(\theta, c\hat{\theta}_s) = c^2(p-1)\sigma_1^2\sigma_2^2/(\sigma_1^2+\sigma_2^2) + \left[(c-1)^2(p-1)+1\right]\theta'\theta/p. \tag{2.2}$$

Proof. It can be easily shown that

$$R(\theta, c\hat{\theta}_s) = E(\|cX_1-\theta\|^2) - c^2E(X_1'Z/\|Z\|)^2 + 2cE\left[(\theta' ZZ'X_1)/\|Z\|^2\right]. \tag{2.3}$$

After substituting $E[E(\theta' ZZ'X_1\,\|Z\|^{-2}Z)] = E(\theta' ZZ'\theta\,\|Z\|^{-2})$ and using Lemma 2.1, (v) and (vi), (2.3) can be expressed as

$$R(\theta, c\hat{\theta}_s) = E(\|cX_1-\theta\|^2) - c^2\left[\frac{\sigma_1^2(p\sigma_1^2+\sigma_2^2)}{\sigma_1^2+\sigma_2^2} + \frac{\theta'\theta}{p}\right] + 2c\theta'\theta/p. \tag{2.4}$$

Substituting $E(\|cX_1-\theta\|^2) = c^2\sigma_1^2p + (c-1)^2\theta'\theta$ in (2.4) and after simplification one can get (2.2). $\square$

It is clear from (2.2) that for any given c the risk of $c\hat{\theta}_s$ is an increasing function of $\|\theta\|$. Hence, as the risk of X_1 is $p\sigma_1^2$, none of the estimator of the form $c\hat{\theta}_s$ dominates X_1. Also, because of symmetry, $c\hat{\theta}_s$ does not beat X_2. The estimator $\hat{\theta}_s$ has lower risk than that of $p\hat{\theta}_s/(p-1)$ although the latter is an unbiased estimator of θ. At present, it is not clear how to improve over both X_1 and X_2 simultaneously when σ_1^2 and σ_2^2 are unknown. However, if one of the variances is known then one can develop a shrinkage estimator which is location equivariant, unbiased and also substantially better than the X_i whose variance is known.

When σ_1^2 is known, note that $\sigma_1^2(p-2)\| X_1 - X_2 \|^{-2}$ is an unbiased estimator of $\sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}$. So, for the same reason given for (2.1) one can propose

$$\begin{aligned}\hat{\theta}_0 &= \sigma_1^2(p-2)\|X_1-X_2\|^{-2}X_2 + \left(1-\sigma_1^2(p-2)\|X_1-X_2\|^{-2}\right)X_1\\ &= X_1 - \sigma_1^2(p-2)\|X_1-X_2\|^{-2}(X_1-X_2)\\ &= X_2 + \left[1-\sigma_1^2(p-2)\|X_1-X_2\|^{-2}\right](X_1-X_2) \qquad (2.5)\end{aligned}$$

as an estimator of θ. Note that this estimator (2.5) is nothing but the usual James–Stein (1961) estimator that shrinks X_1 toward X_2 instead the origin. Although by shrinking X_1 toward any arbitrary fixed vector one can improve X_1, in this situation it is more reasonable to shrink X_1 toward X_2 as it is also an unbiased estimator of θ. Also, it follows from (2.5) that this shrinkage estimator $\hat{\theta}_0$ can be interpreted as a weighted average of X_1 and X_2. Further, as X_1 and X_2 are independent usual conditional argument yields that $\hat{\theta}_0$ dominated X_1.

Considering (2.5), whether σ_1^2 is known or not, one can propose a class of estimators of the form

$$\hat{\theta}_r = X_1 - \left[r\left(\|Z\|^2\right)/\|Z\|^2\right]Z \qquad (2.6)$$

where $r(\cdot)$ is a real valued and differentiable function. Clearly, the estimator (2.6) is location equivariant. Also, using the fact that Z is distributed as $-Z$ it can be easily verified that $\hat{\theta}_r$ is unbiased.

Next, in order to compute the risk of the estimators of the form (2.6), the following chi-square identity which is a special case of Haff's (1979) Wishart identity is needed.

Lemma 2.2 (Chi-square identity). *Let $V \sim \sigma^2\chi_m^2$. For a suitable function $h(V)$,*

$$E\left[\sigma^{-2}h(V)\right] = 2E\left[\partial h(V)/\partial V\right] + (m-2)E\left[V^{-1}h(V)\right], \qquad (2.7)$$

provided all the expectations exist. □

Theorem 2.2. *The risk of the estimator* (2.6) *can be expressed as*

$$R\left(\theta, \hat{\theta}_r\right) = R(\theta, X_1) + E\left\{\|Z\|^{-2}r\left(\|Z\|^2\right)\left[r\left(\|Z\|^2\right) - 2(p-2)\sigma_1^2\right]\right\} - 4\sigma_1^2E\left[r'\left(\|Z\|^2\right)\right].$$

Proof. The risk function of $\hat{\theta}_r$ is

$$R\left(\theta, \hat{\theta}_r\right) = R(\theta, X_1) + E\left[r^2\left(\|Z\|^2\right)\|Z\|^{-2}\right] - 2E\left[r\left(\|Z\|^2\right)(X_1-\theta)'Z\|Z\|^{-2}\right]$$

and so the risk difference $\mathrm{RD} = R(\theta, \hat{\theta}_r) - R(\theta, X_1)$ is

$$\mathrm{RD} = E\left[r^2\left(\|Z\|^2\right)\|Z\|^{-2}\right] - 2E_ZE\left[r\left(\|Z\|^2\right)(X_1-\theta)'Z\|Z\|^{-2}|Z\right]. \qquad (2.8)$$

Using the conditonal expectation $E[(X_1-\theta)\mid Z] = \sigma_1^2(\sigma_1^2+\sigma_2^2)^{-1}Z$, (2.8) can be simplified as

$$\mathrm{RD} = E\left[r^2\left(\|Z\|^2\right)\|Z\|^{-2}\right] - 2\sigma_1^2\left(\sigma_1^2+\sigma_2^2\right)^{-1}E\left[r\left(\|Z\|^2\right)\right]. \qquad (2.9)$$

Noting that $\|Z\|^2 \sim (\sigma_1^2+\sigma_2^2)\chi_p^2$ and using (2.7) the second term of (2.9) can be expressed as

$$E\left[\left(\sigma_1^2+\sigma_2^2\right)^{-1}r\left(\|Z\|^2\right)\right] = 2E\left[r'\left(\|Z\|^2\right)\right] + (p-2)E\left[\|Z\|^{-2}r\left(\|Z\|^2\right)\right]. \qquad (2.10)$$

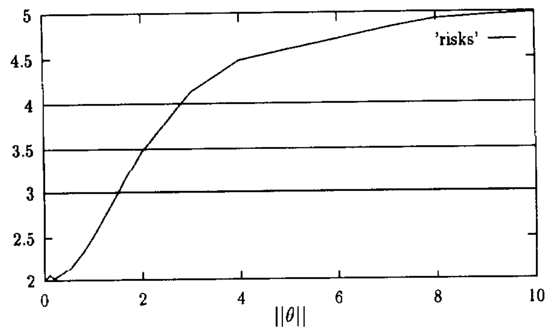

Fig. 2.1.

Substituting (2.10) in (2.9), the RD can be written as

$$\mathrm{RD} = E\left\{\|Z\|^{-2} r\left(\|Z\|^2\right)\left[r\left(\|Z\|^2\right) - 2(p-2)\sigma_1^2\right]\right\} - 4\sigma_1^2 E\left[r'\left(\|Z\|^2\right)\right] \tag{2.11}$$

and this completes the proof. □

Remark 2.1. It is clear from (2.11) that $\mathrm{RD} \leqslant 0$ for all θ in $\mathbb{R}^p$ and σ_2^2 in $\mathbb{R}^+$ provided $r(\cdot)$ is nondecreasing and $0 \leqslant r(\cdot) \leqslant 2(p-2)\sigma_{1\mathrm{o}}^2$ where $\sigma_{1\mathrm{o}}^2$ is a known lower bound of σ_1^2. An obvious choice of $r(\cdot)$ is the constant function. Another choice is $r(\|Z\|^2) = c_\mathrm{o}/(1 + c_\mathrm{o}\|Z\|^{-2})$ with the corresponding estimator being $X_1 - c_\mathrm{o}Z/\|Z\|^2 + c_\mathrm{o})$ where $c_\mathrm{o} = (p-2)\sigma_{1\mathrm{o}}^2$.

Remark 2.2. If $r(\|Z\|^2)$ is a constant function and σ_1^2 is known, then it follows from (2.11) that the best choice of $r = (p-2)\sigma_1^2$. Interestingly, this leads to the estimator which is the weighted average $\hat{\theta}_\mathrm{o}$ given in (2.5).

Remarks 2.3. For comparison purpose, without loss of generality, σ_1^2 is assumed to be equal to one. Then, the risks $R(\theta, X_1) = p$ and $R(\theta, \hat{\theta}_{\mathrm{JS}}) = p - (p-2)^2 E\|X_1\|^2$ where $\hat{\theta}_{\mathrm{JS}} = (1 - (p-2)\|X_1\|^{-2})X_1$ is the James–Stein shrinkage estimator. Also, it can be easily seen from (2.8) that $R(\theta, \hat{\theta}_\mathrm{o}) = p - (p-2)(1+\sigma_2^2)^{-1}$. Further,

$$\text{risk of } \hat{\theta}_{\mathrm{JS}} \to 2 \quad \text{as } \|\theta\|^2 \to 0; \quad \text{and } \to p \quad \text{as } \|\theta\|^2 \to \infty$$

and

$$\text{risk of } \hat{\theta}_\mathrm{o} \to 2 \quad \text{as } \sigma_2^2 \to 0; \quad \text{and } \to p \quad \text{as } \sigma_2^2 \to \infty.$$

Thus, there is no clear cut winner between $\hat{\theta}_\mathrm{o}$ and $\hat{\theta}_{\mathrm{JS}}$. However, it is evident from the risk expressions and the following graph that $\hat{\theta}_\mathrm{o}$ is preferable to $\hat{\theta}_{\mathrm{JS}}$ when σ_2^2 is small and/or $\|\theta\|^2$ is moderately large. In Figure 2.1, the four horizontal lines represent the $R(\theta, \hat{\theta}_\mathrm{o}) = 3.0$, 3.5, 4.0 and 5 respectively at $\sigma_2^2 = 0.5$, 1.0, 2.0 and ∞, and the curve represents the risk of $\hat{\theta}_{\mathrm{JS}}$ when $p = 5$.

Acknowledgement

The authors gratefully acknowledge some useful comments made by the referee.

References

Chiou, W. and A. Cohen, (1985), On estimating a common multivariate normal mean vector, *Ann. Inst. Statist. Math.* **37**, 499–506.

George, E.I. (1991), Shrinkage domination in a multivariate common mean problem, *Ann. Statist.* **19**, 952–960.

Haff, L.R. (1979), An identity for the Wishart distribution with applications, *J. Multivariate Anal.* **9**, 531–544.

James, W. and C. Stein (1961), Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1**, 361–379.

Krishnamoorthy, K. (1991), Estimation of a common multivariate normal mean vector, *Ann. Inst. Statist. Math.* **43**, 761–771.

Krishnamoorthy, K. (1992), On a shrinkage estimator of a normal common mean vector, *J. Multivariate Anal.* **40**, 109–114.

Kubokawa, T. (1989), Estimating common parameters of growth curve models under a quadratic loss, *Comm. Statist.—Theory Methods* **18**, 3149–3155.

Sarkar, S.K. (1992), Shrinkage estimation of some usual estimators of the common mean of several multivariate normal populations, Tech. Rept., Temple Univ. (Philadelphia, PA).