Ann. Inst. Statist. Math.
Vol. 44, No. 3, 537–550 (1992)

MINIMAX ESTIMATORS FOR LOCATION VECTORS IN ELLIPTICAL DISTRIBUTIONS WITH UNKNOWN SCALE PARAMETER AND ITS APPLICATION TO VARIANCE REDUCTION IN SIMULATION*

M. TAN[1] AND L. J. GLESER[2]

[1] *Department of Biostatistics and Epidemiology/P88, The Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195-5196, U.S.A.*
[2] *Department of Mathematics and Statistics, University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA 15260-0001, U.S.A.*

(Received December 17, 1990; revised October 14, 1991)

Abstract. In this paper, we give an ever wider and new class of minimax estimators for the location vector of an elliptical distribution (a scale mixture of normal densities) with an unknown scale parameter. Then its application to variance reduction for Monte Carlo simulation when control variates are used is considered. The results obtained thus extend (i) Berger's result concerning minimax estimation of location vectors for scale mixtures of normal densities with known scale parameter and (ii) Strawderman's result on the estimation of the normal mean with common unknown variance.

Key words and phrases: Minimax estimators, shrinkage estimators, elliptical distributions, scale mixture of normal, Monte Carlo simulation, variance reduction.

1. Introduction

In this paper, we give an ever wider class of minimax estimators for the location vector of an elliptical distribution involving an unknown scale parameter. And then we consider its application to variance reduction for Monte Carlo simulation. The actual model we shall consider is of the form

$$f(z \mid \theta) = \int_0^\infty \frac{|\Sigma|^{-1/2}}{(2\pi\tau^2 v)^{p/2}} e^{-(z-\theta)'\Sigma^{-1}(z-\theta)/2\tau^2 v} dF(v), \tag{1.1}$$

where $\theta_{p\times 1}$ and $\tau^2 > 0$ are unknown parameters, Σ is a known positive matrix and $F(v)$ is a known c.d.f. on $(0, \infty)$. This distribution is often called a scale

* Research partially supported by National Science Foundation, Grant #DMS 8901922.

mixture of normals. It includes important cases such as the usual normal density, the ϵ-contaminated normal density and the p-variate elliptical t-distribution (see Muirhead (1982)) and consequently the multivariate t-error terms in regression (Zellner (1976)). Model (1.1) is a more general model than that considered in Theorem 1 of Berger (1975) since it involves an unknown parameter τ^2, which complicates the problem of finding improved estimators for θ considerably.

Various classes of estimators which improve upon the invariant estimator for the known scalar case have been found. See Brandwein and Strawderman (1990) for an overview and the references therein. For the case with an unknown scale parameter, more recently, Bravo and MacGibbon (1988) obtain a class of improved estimators for θ under model (1.1) which also extends Theorem 1 of Berger (1975). Bravo and MacGibbon (1990) further develop improved estimators for more general elliptical distributions. Thus Theorem 2 in Berger (1975) is generalized. Cellier *et al.* (1989) develop improved estimators for a class of elliptical distributions, but their estimator of the scale parameter is obtained from the residual vector and is not independent of z unless z is normal. Also Gleser and Tan (1989) generalize Theorem 2 in Berger (1975) assuming the scalar is estimated by an estimator which is independent of z and which may or may not be χ^2 distributed.

Considered in this paper is model (1.1). In Section 2, under the same assumption that there is an independent estimate of τ^2, a new and wider class of improved estimators for θ is given, extending the results in both Berger (1975) and Strawderman (1973). The former considers minimax estimation of location vectors for densities of the form $f((x-\theta)'\Sigma^{-1}(x-\theta))$ with $\Sigma_{p\times p}$ known and $p \geq 3$ and the latter concerns estimation of the normal mean with common unknown variance. More specifically, generalizations of Theorem 1 of both Berger (1975) and Strawderman (1973) are obtained via estimating the unknown scale parameter τ^2 by a multiple of a χ^2 random variable and using a shrinkage function with two arguments (used in Strawderman (1973) for the normal case). A special case is that the shrinkage function depends only on the first argument. In this special case our class of estimators is however different from that of Bravo and MacGibbon (1988) and it gives adaptive estimators in the control variates problem (see Section 4).

In Section 3, as an application of the result in Section 2, the problem of variance reduction in Monte Carlo simulation when control variates are used is considered. Consequently improved estimators of the mean response in simulation are obtained when control variates are used. This serves on the other hand as an example where an independent estimator of τ^2 is available, which may not always be the case for the elliptical distributions (1.1). In fact, it is well known in the spherically symmetric case that a χ^2 distributed estimate of τ^2 can be easily found but the independence of the estimates and z is in general no longer guaranteed unless z is $N(\theta, \tau^2 I)$ (see Kelker ((1970), p. 428)).

The final section gives a brief discussion on various improved estimators and their implications and some other topics of both practical and theoretical interest, which partly motivate the consideration of a two-argument shrinkage function.

2. Minimax estimators

Suppose $f(z \mid \theta)$ is of the form (1.1) and we want to estimate θ. Since τ^2 is unknown, we suppose it is estimated by w where $w = c_1 w_1$ and $\tau^{-2} w_1$ is distributed as a chi-square r.v. Consider the estimation of θ under the loss

$$L(\delta; \theta, \tau^2) = \frac{(\delta - \theta)' Q (\delta - \theta)}{\tau^2}, \tag{2.1}$$

where $Q_{p \times p}$ is a known positive definite matrix. When $p \geq 3$, we can take advantage of the Stein effect and obtain a class of estimators which improve upon z in risk. Consequently those estimators are all minimax estimators of θ, since z is minimax in the present problem.

In the known $-\tau^2$ case, Berger's Σ corresponds to our $\tau^2\Sigma$, and Q to our $\tau^{-2}Q$. Thus, the class of estimators shown by Berger to dominate z in the known-τ^2 case is of the form

$$\left(I_p - h\left(\frac{z'\Sigma^{-1}Q^{-1}\Sigma^{-1}z}{\tau^2} \right) Q^{-1}\Sigma^{-1} \right) z, \tag{2.2}$$

where the function h is often called shrinkage function. Since τ^2 is unknown, we replace τ^2 by w, and consider estimators of the form

$$\begin{aligned} \delta_h(z, w) &= \left(I_p - h\left(\frac{z'\Sigma^{-1}Q^{-1}\Sigma^{-1}z}{w}, w \right) Q^{-1}\Sigma^{-1} \right) z \\ &= \left(I_p - \frac{w r\left(\frac{z'\Sigma^{-1}Q^{-1}\Sigma^{-1}z}{w}, w \right)}{z'\Sigma^{-1}Q^{-1}\Sigma^{-1}z} Q^{-1}\Sigma^{-1} \right) z, \end{aligned} \tag{2.3}$$

where w has the same distribution as $c_1 w_1$ with $w_1 \sim \tau^2 \chi^2_m$ and $c_1 = 1/(m+2)$ and the function $h(t, w)$ is a bivariate function from $[0, \infty) \times [0, \infty)$ to $[0, \infty)$ and satisfies the following three requirements:

$$\begin{aligned} &\text{(i) } h(u, w) \text{ is nonincreasing in } u \geq 0 \text{ for fixed } w, \\ &\text{(ii) } r(u, w) = uh(u, w) \text{ is nondecreasing in } u \geq 0 \text{ for fixed } w, \\ &\text{(iii) and } r(u, w) = uh(u, w) \text{ is nonincreasing in } w \text{ for any } u \geq 0. \end{aligned} \tag{2.4}$$

THEOREM 2.1. *Suppose that $Ez'z$ and $E(z'z)^{-1}$ are finite when $\theta = 0$, w and $\delta_h(z, w)$ are defined in (2.3), w and z are independent, $w = c_1 w_1$, $w_1 \sim \tau^2 \chi^2_m$, $r(t, w)$ satisfies condition (2.4) and $r_1 = \sup_{t \geq 0} r(t, w)$ is a constant. Then $\delta_h(z, w)$ dominates z in risk (and is hence minimax estimator of θ) provided that*

$$0 \leq r_1 = \sup_{t \geq 0} r(t, w) \leq \frac{2}{E_{\theta=0, \tau^2=1}(z'\Sigma^{-1}z)^{-1}}. \tag{2.5}$$

Note. In fact $\sup_{t\geq 0} r(t,w)$ is in many cases a constant free of w. A useful special case is when $r(t,w) = r(t)$, i.e., $r(t,w)$ is only a function of t. Other cases may include the Bayes estimator of θ under a spherical symmetric prior on θ given τ^2 and a conjugate prior on τ^2 (Strawderman (1973), DasGupta and Rubin (1988)).

PROOF OF THEOREM 2.1. As in Berger (1975), let $A_1 \geq A_2 \geq \cdots \geq A_p \geq 0$ be the eigenvalues of ΣQ^{-1}. Then there exists a nonsingular matrix B such that $B'QB = I_p$, $B'\Sigma^{-1}B = A^{-1}$, where $A = \operatorname{diag}(A_1, \ldots, A_p)$. Transforming $z \to B^{-1}z$, $\theta \to B^{-1}\theta$, yields a "canonical" estimation problem in which the distribution of z has parameter $\Sigma = A$, the loss function in (2.1) has centering matrix $Q = I_p$, and the estimators $\delta_h(z,w)$ have the form

$$\delta_h(z,w) = \left(I_p - h\left(\frac{z'A^{-2}z}{w}, w\right)A^{-1}\right)z. \tag{2.6}$$

Let $\Delta(\theta)$ be the difference in risks between z and $\delta_h(z,w)$, then

$$\Delta(\theta) = E_\theta[(z-\theta)'(z-\theta) - (\delta_h(z,w)-\theta)'(\delta_h(z,w)-\theta)].$$

The goal is then to show that $\Delta(\theta) \geq 0$ for all θ.

Since w and z are independent, the expectations

$$E^{(wz)}\cdot = E^w E^{z|w}\cdot = E^w E^z.$$

Let $h_w(t) = h(t/w, w)$. It is then clear that $h_w(t)$ (as a function of t) satisfies the first two requirements in (2.4), that is, $h_w(t)$ is nonincreasing in $t \in [0,\infty)$, and that $r_w(t) = th_w(t)$ is nondecreasing in $t \in [0,\infty)$. Since $r_w(t) = wr(t/w,w)$, if $r_1 = \sup_{t\geq 0} r(t,w)$, then $\sup_{t\geq 0} r_w(t) = r_1 w$. Now the estimator is

$$\delta_h(z,w) = (I_p - h_w(z'A^{-2}z, w)A^{-1})z.$$

Notice that (1.1) can be written as a two-stage model (in canonical form):

$$z \mid v \sim N(\theta, \tau^2 vA) \quad \text{and} \quad v \sim F(v).$$

Thus conditioning on w (in fact, since w and z are independent, w can be treated as fixed), and conditioning on v and treating $v\tau^2$ as σ^2, we have $z \mid v \sim N(\theta, \sigma^2 A)$. Then using the integration by parts identity (Lemma A.1 in the Appendix) and following the steps on pp. 1320–1322 in Berger (1975) yield that

$$\begin{aligned}\Delta(\theta) &\geq E^w\left(\int_0^\infty \left(2(p-2) - \frac{r_1 w}{v\tau^2}\right)dF(v)\right)T_\theta(w)\\ &= 2(p-2)E^w T_\theta(w) - \frac{1}{\tau^2}E^w wT_\theta(w)\int \frac{1}{v}dF(v),\end{aligned}$$

where

(2.7) $T_\theta(w) =$

$$\int_0^\infty \int_{R^p} \frac{(\tau^2 v)^{-(p-2)/2} h\left(\frac{z'A^{-2}z}{w}, w\right) \exp\left\{-\frac{1}{2\tau^2 v}(z-\theta)'A^{-1}(z-\theta)\right\}}{(2\pi)|A|^{1/2}} dz dF(v).$$

Our goal is then to find an upper bound b_0 such that

$$EwT_\theta(w) \leq b_0 ET_\theta(w). \tag{2.8}$$

Let $v_1 = \int_0^\infty (1/\nu) dF(\nu) = E(1/v)$, then

$$\Delta(\theta) \geq 2(p-2)ET_\theta(w) - \frac{r_1 v_1}{\tau^2} EwT_\theta(w). \tag{2.9}$$

Integration by parts using Lemma A.2 (in the Appendix) for w_1 gives (note $w = c_1 w_1$)

$$\begin{aligned} EwT_\theta(w) &= \tau^2 c_1 E^{w_1} w_1 T_\theta(w) \\ &= \tau^2 c_1(\{mE^{w_1}T_\theta(w) + 2c_1 E^{w_1} w_1 T'_\theta(w)\}) \\ &= \tau^2(c_1 m ET_\theta(w) + 2Ec_1 w_1 T'_\theta(w)) \\ &= \tau^2 c_1(mET_\theta(w) + 2EwT'_\theta(w)). \end{aligned} \tag{2.10}$$

Note in fact

$$T_\theta(w) = E^v E^{z|v} v \frac{wr\left(\frac{z'A^{-2}z}{w}, w\right)}{z'A^{-2}z} = E^v E^{z|v} vh(s/w, w),$$

where $s = z'A^{-2}z$, v and c.d.f. $F(v)$, and $z \mid v \sim N(\theta, v\tau^2 A)$. Since $z \mid v$ has normal density, the derivative can be taken inside the integral. So

$$T'_\theta(w) = E^v E^{z|v} v \frac{\partial}{\partial w}\left(\frac{wr\left(\frac{z'A^{-2}z}{w}, w\right)}{z'A^{-2}z}\right).$$

Then the usual chain rule of the derivative of a compound function gives

$$\begin{aligned} \frac{\partial}{\partial w}\left[\frac{wr\left(\frac{z'A^{-2}z}{w}, w\right)}{z'A^{-2}z}\right] &= \frac{\partial}{\partial w}\left[\frac{wr(s/w, w)}{s}\right] \\ &= \frac{1}{s} r(s/w, w) + \frac{w}{s}\left(r'_{(1)}\left(-\frac{s}{w^2}\right) + r'_{(2)}\right) \\ &= \frac{1}{s} r(s/w, w) - \frac{1}{w} r'_{(1)}(s/w, w) + \frac{w}{s} r'_{(2)}(s/w, w), \end{aligned}$$

where $r'_{(i)}$ $(i = 1, 2)$ is the derivative with respect to the i-th argument of the function r, then

$$wT'_\theta(w) = v\left[h(s/w, w) - r'_{(1)}(s/w, w) + \frac{w^2}{s} r'_{(2)}(s/w, w)\right].$$

Then the monotonicity properties (conditions (ii) and (iii) in (2.4)) of the function $r(s/w, w)$ imply that

$$r'_{(1)}(s/w, w) \geq 0 \quad \text{and} \quad r'_{(2)}(s/w, w) \leq 0.$$

Therefore,

$$EwT'_\theta(w) \leq ET_\theta(w).$$

So

$$EwT_\theta(w) \leq \tau^2 c_1(m+2)ET_\theta(w).$$

Meanwhile note the fact when $\theta = 0$, $v^{-1}z'A^{-1}z \sim \tau^2\chi^2_p$, which is independent of v. Then

$$E_{\theta=0}(z'A^{-1}z)^{-1} = E(\tau^2 v)^{-1}\frac{1}{(\tau^2 v)^{-1}z'A^{-1}z} = \frac{\tau^2 Ev^{-1}}{p-2}$$

so

$$v_1 = Ev^{-1} = (p-2)E_{\theta=0}(\tau^{-2}z'A^{-1}z)^{-1} = (p-2)E_{\theta=0,\tau^2=1}(z'A^{-1}z)^{-1}.$$

Combining this with (2.9) yields

$$\begin{aligned}\Delta(\theta) &\geq 2(p-2)ET_\theta(w) - r_1 v_1(m+2)c_1 ET_\theta(w)\\ &\geq (2(p-2) - r_1 v_1)ET_\theta(w)\\ &\geq 0,\end{aligned}$$

provided that (2.5) is satisfied. □

To see the magnitude of the potential gain in risk, a Monte Carlo simulation is used to evaluate the integrals involved. A random sample (of size 300) of a p-variate elliptical t-distribution is hence generated, where $p = 5$, $\Sigma_{xx} = I_5$, $k = 11$, the mixture distribution $v^{-1} \sim \chi^2_k$, let $m = k-1$ and $w_1 \sim \tau^2\chi^2_{k-1}$, $w = w_1/(k+1)$, then $E_{\theta=0,\tau^2=1}(z'z)^{-1} = k/(p-2)$. Let $\delta_0 = z$ be the usual maximum likelihood estimator and the James-Stein estimator be defined as

$$\delta_1(z, w) = \left(1 - \frac{p-2}{k(k+1)}\frac{w}{z'z}\right)z.$$

It is well known this estimator itself is inadmissible as an estimator of θ. It is dominated by the positive-part James-Stein estimator defined by

$$\delta_2(z, w) = \left(1 - \frac{p-2}{k(k+1)}\frac{w}{z'z}\right)^+ z,$$

where $a^+ = a$ if $a \geq 0$ and $a^+ = 0$ if $a < 0$ for any number a. This last estimator is also known to be inadmissible, but serves as an adequate approximation to an admissible estimator. Then the risk of the standard estimator $\delta_0 = z$ is $E(z-\theta)'(z-\theta) = E^v E^{z|v}(z-\theta)'(z-\theta) = p/(k-2)$. And the risks of the improved estimators for various $|\theta|$ are numerically calculated, normalized so that the risk of z is 1, and plotted in Fig. 1. In fact the graph of the risks $R(\theta, \delta_2)$ versus $|\theta|$ is given in Fig. 1. As what happened in the normal mean estimation problem, the numerical results indicate that when $|\theta|$ is small, the potential improvement in risk is very large, while as $|\theta|$ becomes very large the improvement in risk vanishes. In other words, if all the θ_is are at or near 0, then the shrinkage estimator will greatly improve the usual estimator.

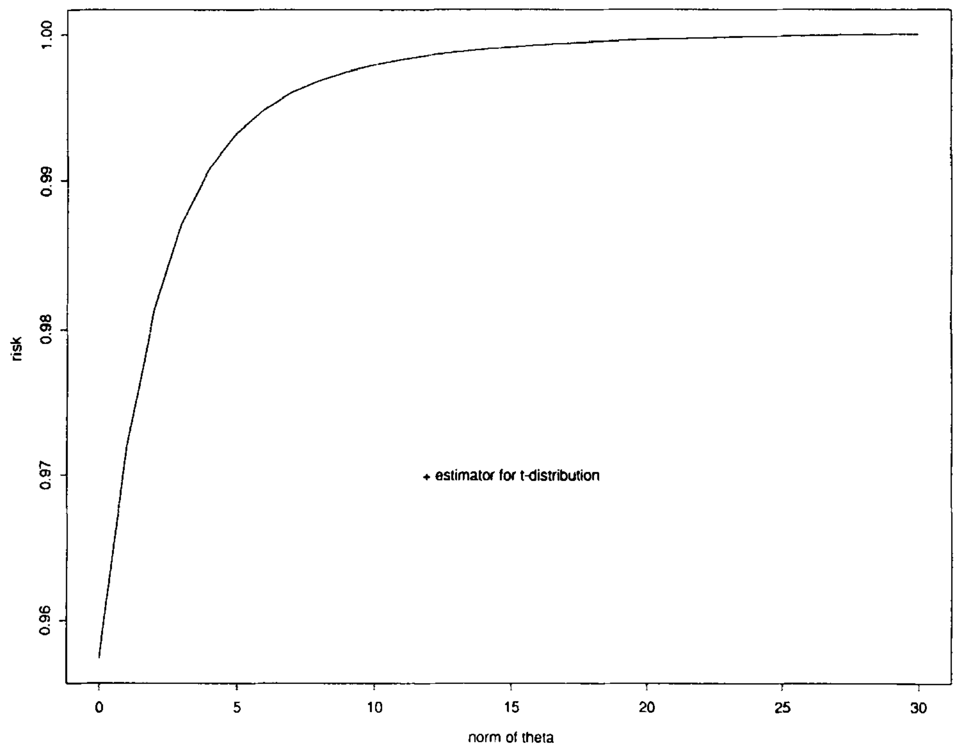

Fig. 1. Normalized risk versus the norm of theta.

3. Application to variance reduction in Monte Carlo simulation when control variates are used

Variance reduction has been an important aspect in Monte Carlo simulation. Various techniques have been proposed (see, e.g., Nelson (1987) and Wilson (1984)). Among them control variates method is one of the most promising variance reduction methods and has been widely used in simulation studies as a means for improving efficiency in the estimation of the characteristics of the output variables (Lavenberg and Welch (1981), Fishman (1989)). This technique collects sample data not only on the response (say y) but also on certain ancillary phenomena whose true means are known. This extra sample information is then used to construct an unbiased estimator of the response mean which has smaller variance than the estimator $\bar{y}$. For example, consider a r.v. y (a response variable for one population) with unknown mean μ_y which is the quantity to be estimated. Let $x = (x_1, \dots, x_p)'$ be a random vector with known mean vector $\mu_x = (\mu_1, \dots, \mu_p)'$ which is thought to be correlated with y. The elements $x_1, \dots, x_p$ of x are called control variates.

Since on each independent run of the simulation y and x result from a common probabilistic structure (e.g., a multiserver queue), $(y, x')'$ can often be assumed to have joint normal distribution (Lavenberg and Welch (1981)) with mean $(\mu_y, \mu_x')'$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \sigma_{yx} \\ \sigma_{xy} & \Sigma_{xx} \end{pmatrix}.$$

Then n repetitions of the simulation experiment yield statistically independent observations

$$(y_i, x_{1i}, x_{2i}, \dots, x_{pi})', \qquad i = 1, 2, \dots, n.$$

The usual regression (with random regressors) theory gives the LS (ML also) estimator of μ

$$\hat{\mu}(b) = \bar{y} - b'(\bar{x} - \mu_x)$$

where

$$W = \begin{pmatrix} w_{yy} & w_{yx} \\ w_{xy} & W_{xx} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n}(y_i - \bar{y})^2 & \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})' \\ \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) & \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})' \end{pmatrix},$$

and $b = W_{xx}^{-1} w_{xy}$. This estimator is unbiased and

$$\text{(3.1)} \qquad \operatorname{Var}(\hat{\mu}(b)) \le \operatorname{Var}(\bar{y}) \quad \text{if} \quad n \ge p + 2 \quad \text{and} \quad \rho_{y \cdot x}^2 > \frac{p}{n-2},$$

where $\rho_{y \cdot x} = (\sigma_{yx} \Sigma_{xx}^{-1} \sigma_{yx}' / \sigma_{yy})^{1/2}$ is the multiple correlation coefficient of determination.

Also it is worth noting that the problem treated here is in fact the regression problem with random regressors (or with concomitant variables), a problem which occurs very often in practice. Point estimators of intercepts and slopes for this problem have been considered by Stein (1960), Baranchik (1973), Takada (1979) and Zidek (1978) under special conditions.

In addition to the usual variance reduction, the use of the shrinkage methodology further increases the efficiency of Monte Carlo simulation studies in terms of reducing the variances of the estimators for the response mean. Let $b_h = b_h(W)$ be any vector valued function of W, then $\hat{\mu}(b_h) = \bar{y} - b_h'(\bar{x} - \mu_x)$ is an unbiased estimator of μ_y as long as the expectation of $b(W)$ exists. Then consider any estimator of μ_y of the form

$$\hat{\mu}(b_h) = \bar{y} - b_h'(W)(\bar{x} - \mu_x). \tag{3.2}$$

Define $\sigma_{yy\cdot x} = \sigma_{yy} - \sigma_{yx}\Sigma_{xx}^{-1}\sigma_{xy}$ and notice that $\sigma_{yy\cdot x} = \sigma_{yy}(1 - \rho_{y\cdot x}^2)$, then it is easy to show that

$$\mathrm{Var}[\hat{\mu}(b_h)] = \frac{1}{n}\sigma_{yy}(1 - \rho_{y\cdot x}^2)\left[1 + \frac{E(b_h - \beta)'\Sigma_{xx}(b_h - \beta)}{\sigma_{yy\cdot x}}\right], \tag{3.3}$$

where $\beta = \Sigma_{xx}^{-1}\sigma_{xy}$. Define the loss function

$$L(\delta; \beta, \sigma_{yy\cdot x}, \Sigma_{xx}) = \frac{(\delta - \beta)'\Sigma_{xx}(\delta - \beta)}{\sigma_{yy\cdot x}}, \tag{3.4}$$

and note that any improved estimator $b_h(W)$ of the slope β under this loss function gives us estimators for the mean response μ with smaller variances than the simple control variates estimator $\hat{\mu}(b) = \bar{y} - b'(\bar{x} - \mu_x)$. Indeed, Stein (1960), Baranchik (1973), Zidek (1978), Takada (1979) and Brown (1990) derive classes of improved estimators for β, and thus for μ_y. Since in simulation studies using control variates it is often the case that the covariance among the control variates is also known (Bauer and Wilson (1989)), in the following, we shall assume that the covariance matrix of the control variates is known. Using the results in Section 2, class of improved estimators (say, $\hat{b}$) for β can be readily obtained such that each of them dominates $\hat{b}$ in risk when $p \geq 3$, $n \geq p + 2$. Each member of this class thus yields an unbiased estimator $\hat{\mu}(\hat{b})$ of μ which has smaller variance than $\hat{\mu}(b)$.

Note that define $w_{yy\cdot x} = w_{yy} - w_{yx}W_{xx}^{-1}w_{yx} = w_{yy} - b'W_{xx}b$. Then $w_{yy\cdot x}$ is independent of (b, W_{xx}), and

$$b \mid W_{xx} \sim N(\beta, \sigma_{yy\cdot x}W_{xx}^{-1}),$$
$$W_{xx} \sim \mathcal{W}_p(n-1, \Sigma_{xx}), \qquad w_{yy\cdot x} \sim \sigma_{yy\cdot x}^2\chi_{n-p-1}^2,$$

where $\mathcal{W}_p(k, \Sigma)$ is the Wishart distribution with dimension p, degree of freedom k and covariance matrix Σ. Consequently, when Σ_{xx} is known, the distribution of b is simply a p-variate elliptical t-distribution with $n - p$ degrees of freedom, location parameter β and scale matrix $\sigma_{yy\cdot x}\Sigma_{xx}^{-1}/(n-p)$ (see Muirhead (1982)). It has density

$$f(b) = \frac{\Gamma\left(\frac{n}{2}\right)|\Sigma_{xx}|^{1/2}}{\Gamma\left(\frac{n-p}{2}\right)(\pi\sigma_{yy\cdot x})^{p/2}}\left[1 + \frac{(b-\beta)'\Sigma_{xx}(b-\beta)}{\sigma_{yy\cdot x}}\right]^{-n/2}. \tag{3.5}$$

The class of estimators of β considered is defined by

$$b_h = \left[1 - h\left(\frac{b'\Sigma_{xx}b}{w_{yy\cdot x}}, c_1 w_{yy\cdot x}\right)\right]b, \tag{3.6}$$

where $h(u,w) = r(u,w)/u$ and the functions $h(u,w)$ and $r(u,w)$ satisfy (2.4).

THEOREM 3.1. *Assume that $p \geq 3$, $n \geq p+2$. Then if*

$$r(t,w) \leq \frac{2(p-2)}{(n-p)(n-p+1)},$$

b_h has risk everywhere (over β, $\sigma_{yy\cdot x}$) at least as small as that of b under the loss function (3.4). As a result, $\hat{\mu}(b_h) = \bar{y} - b_h'(W)(\bar{x} - \mu_x)$, the unbiased estimator of μ_x, has variance no more than the variance of $\bar{y} - b'(\bar{x} - \mu_x)$.

Note that the distribution of b is in fact a normal mixture with inverse-chi-square. That is,

$$b \mid v \sim N(\beta, \sigma_{yy\cdot x} v \Sigma_{xx}^{-1}) \quad \text{and} \quad v^{-1} \sim \chi^2_{n-p}.$$

PROOF. Direct application of Theorem 2.1. In fact, make the following correspondences between the notation of Theorem 2.1 and that of 3.1:

$$b \sim z, \quad \Sigma_{xx}^{-1} \sim \Sigma, \quad \beta \sim \theta, \quad \sigma_{yy\cdot x} \sim \tau^2, \quad \frac{w_{yy\cdot x}}{n-p+1} \sim w.$$

Further, the loss function used in Theorem 3.1 has $Q = \Sigma_{xx}$. Then it is clear that Theorem 3.1 is a special case of Theorem 2.1. □

Interestingly, the estimators of the form $\hat{\mu}(b)$ are adaptive estimators of μ_y when $h(\cdot,\cdot)$ is simply a function of the first argument. In fact, as a result of (3.1), $\rho^2_{y\cdot x}$ needs to be fairly large if $\hat{\mu}(b)$ is to be superior to $\bar{y}$ as an estimate of μ_y. In particular, we would want to use $\bar{y}$ in preference to $\hat{\mu}(b)$ as an estimator of μ_y when $\rho^2_{y\cdot x} = 0$ (equivalently, $\beta = 0$). An appropriate (likelihood ratio) test statistic for testing $H_0 : \rho^2_{y\cdot x} = 0$ when Σ_{xx} is known is

$$T = \frac{b'\Sigma_{xx} b}{w_{yy\cdot x}}.$$

We reject H_0 for sufficiently large values of T. Thus, we might naively try to use the value of T to choose between $\bar{y}$ and $\hat{\mu}(b)$ as an estimator of μ_y. Let

$$b_h = \left(1 - h\left(\frac{b'\Sigma_{xx} b}{w_{yy\cdot x}}\right)\right) b = (1 - h(T))b$$

be any of the estimators (3.2). Then

$$\begin{aligned} \hat{\mu}(b_h) &= \bar{y} - b_h'(\bar{x} - \mu_x) \\ &= h(T)\bar{y} + (1 - h(T))\hat{\mu}(b). \end{aligned} \tag{3.7}$$

Since $h(T)$ is nonincreasing in T, large values of T (indicating that $H_0 : \rho^2_{y\cdot x} = 0$ may be false) cause greater weight to be placed on $\hat{\mu}(b)$ in (3.7). Small values of

T cause greater weight to be placed on $\bar{y}$. Indeed, T is stochastically increasing in $\rho^2_{yy\cdot x}$. Consequently, each member of the class of estimators $\hat{\mu}(b_h)$ defined by (3.7) is adaptive to the information provided by the data concerning the magnitude of $\rho^2_{yy\cdot x}$. This property of these estimators in part explains why some of these estimators have lower variance than $\hat{\mu}(b)$.

To obtain some idea of the magnitude of the variance reduction, define *the relative improvement in variance* (*riv* in short) by

$$\frac{\mathrm{Var}[\hat{\mu}(b)] - \mathrm{Var}[\hat{\mu}(b_h)]}{\mathrm{Var}[\hat{\mu}(b)]} = \frac{n-p-2}{n-2}\left[\frac{p}{n-p-2} - \frac{(b_h-\beta)'\Sigma_{xx}(b_h-\beta)}{\sigma_{yy\cdot x}}\right],$$

and in particular consider the somewhat crude estimator

$$b^* = \left(1 - \frac{p-2}{(n-p)(n-p+1)T}\right) b,$$

which nevertheless dominates b in risk. Then Fig. 1 can be viewed as an illustration of the possible *riv* of the improved estimator for μ_y. Furthermore, when $\beta = 0$, it can be shown directly that the risk of b^* is

$$\frac{2(n-2)}{(n-p)(n-p-2)} + \frac{2(p-2)}{(n-p)(n-p+1)}$$

and the variance of $\hat{\mu}(b^*)$ is

$$n^{-1}\sigma_{yy\cdot x}\left(1 + \frac{2(n-2)}{(n-p)(n-p-2)} + \frac{2(p-2)}{(n-p)(n-p+1)}\right).$$

Similarly, the variance of $\hat{\mu}(b)$ is

$$n^{-1}\sigma_{yy\cdot x}\left(1 + \frac{p}{n-p-2}\right).$$

When the number of replications n is even modestly large, and $\beta = 0$, the improvement in variance of $\hat{\mu}(b^*)$ over $\hat{\mu}(b)$ is small, mainly because of the n^{-1} term appearing in both variances. However, the *riv* is impressive. In fact, when $\beta = 0$,

$$riv = \frac{(n-p-2)(p-2)}{(n-2)(n-p)}\frac{n-p-1}{n-p+1}.$$

For example, when $n = 20$, $p = 10$, $riv = 0.29$, indicating a 29% reduction in variance. When $n = 40$, $p = 25$, there is a 46% reduction in variance for $\hat{\mu}(b^*)$ over $\hat{\mu}(b)$.

4. Comments

1. In the special case that r depends only on its first argument, the estimator equivalent to that considered in Bravo and MacGibbon (1988) is as follows

$$\delta_h(z,w) = \left(I_p - \frac{wr(z'\Sigma^{-1}Q^{-1}\Sigma^{-1}z)}{z'\Sigma^{-1}Q^{-1}\Sigma^{-1}z}Q^{-1}\Sigma^{-1}\right)z.$$

This is different from (2.3) (even when $r(s,t)$ is independent of t), since we have used w to replace τ^2 in the estimator obtained by assuming τ^2 is known.

2. It is worth noting that the variance of the estimator for μ_y can always be reduced as long as estimator for the slope which dominates the least square slope under (3.4) is used in (3.2). Therefore, as the referee remarked, other classes of improved estimators for the slope β obtained by other authors (e.g., Bravo and MacGibbon (1988)) can be applied to (3.3). However, other classes of improved estimators may not be adaptive estimators which seem to be intuitively appealing in the control variates problem.

3. In simulation studies using control variates, investigators typically report a confidence interval for μ_y, rather than merely giving a point estimator. It seems that improved confidence set estimator for μ_y can also be obtained by using ideas from Hwang and Casella (1982). In fact numerical evidence (Tan (1990)) has suggested that this is the case.

4. One advantage to consider the estimator of the form (2.3) is that it offers the potential to have a class of improved estimators large enough to allow the selection of the appropriate one from the class. For instance, the restricted risk Bayes estimators (DasGupta and Rubin (1988)) may be obtained, which is a problem of more practical importance but a technically extremely difficult one. Some approximation or numerical verification may have to be used in order to show the optimality of the improved estimators. The Monte Carlo sampling (or Baysian sampling) offers hope for solution. These issues are currently under investigation.

Acknowledgements

We are extremely grateful to the three anonymous referees for their detailed comments and suggestions which significantly improve the quality of the paper. The first author thanks Prof. Anirban DasGupta at Purdue University for helpful suggestions and encouragement. He also wishes to thank the Department of Biostatistics and Epidemiology at The Cleveland Clinic Foundation for its support.

Appendix

LEMMA A.1. (Integration by Parts Identity) *Let* $x \sim N_p(\theta,\Sigma)$, $\gamma = (\gamma_1(x), \ldots, \gamma_p(x))'$, *and* $\gamma(x)$ *satisfy the regularity conditions for integration by parts* (*see p.* 362 *in Berger* (1985)). *Then we have*

$$E(x-\theta)'\gamma(x) = E\,\mathrm{tr}(\Sigma J_{\gamma(x)}(x)),$$

where $J_{\gamma(x)}(x) = ((\partial/\partial_{x_i})\gamma_j(x))$, *provided that the integral at the right side exists.*

LEMMA A.2. *Let* $s \sim \tau^2\chi_l^2$. *If* $g(s)$ *is sufficiently regular in* s *for integration by parts, and if* $Esg(s)$ *exists, then*

$$Esg(s) = l\tau^2 Eg(s) + 2\tau^2 Esg'(s).$$

REFERENCES

Baranchik, A. J. (1973). A family of minimax estimators of the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **41**, 642–645.

Bauer, K. W. and Wilson, J. (1989). Control-variate selection criteria, SMS 89-29, Department of Statistics and School of Industrial Engineering, Purdue University, Indiana.

Berger, J. (1975). Minimax estimation of location vectors for wide class of densities, *Ann. Statist.*, **10**, 81–92.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer, New York.

Brandwein, A. C. and Strawderman, W. E. (1990). Stein estimation: the spherically symmetric case, *Statist. Sci.*, **5**, 358–368.

Bravo, G. and MacGibbon, B. (1988). Improved shrinkage estimators for the mean vector of a scale mixture of normals with unknown variance, *Canad. J. Statist.*, **16**, 237–245.

Bravo, G. and MacGibbon, B. (1990). Improved estimators of a location vector with unknown scale parameter, *Comm. Statist. Theory Methods*, **19**, 3657–3670.

Brown, L. D. (1990). An ancillary paradox which appears in multiple regression, *Ann. Statist.*, **18**, 471–493.

Cellier, D., Fourdrinier, D. and Robert, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions, *J. Multivariate Anal.*, **29**, 39–52.

DasGupta, A. and Rubin, H. (1988). Bayesian estimation subject to minimaxity of the mean of a multivariate normal distribution in the case of a common unknown variance: a case for Bayesian robustness, *Statistical Decisions & Related Topics* (eds. J. Berger and S. Gupta), Springer, New York.

Fishman, G. S. (1989). Monte Carlo, control variates, and stochastic ordering, *SIAM J. Sci. Statist. Comput.*, **10**, 187–204.

Gleser, L. J. and Tan, M. (1989). Minimax estimation of location vectors in elliptical distributions with unknown scale parameter, Mimeo. Series, # 89-37, Department of Statistics, Purdue University, Indiana.

Hwang, J. T. and Casella, G. (1982). Minimax confidence sets for the mean vector of a multivariate normal distribution, *Ann. Statist.*, **10**, 868–881.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization, *Sankhyā Ser. A*, **32**, 419–430.

Lavenberg, S. S. and Welch, P. D. (1981). A perspective on the use of control variates to increase the efficiency of Monte Carlo simulation, *Management Sci.*, **27**, 332–335.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Nelson, B. L. (1987). A perspective on variance reduction in dynamic simulation experiments, *Comm. Statist. Simulation Comput.*, **16**, 385–426.

Stein, C. (1960). Multiple regression, *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling* (eds. I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann), Stanford University Press, California.

Strawderman, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances, *Ann. Statist.*, **1**, 1189–1194.

Takada, Y. (1979). A family of minimax estimators in some multiple regression problems, *Ann. Statist.*, **7**, 1144–1147.

Tan, M. (1990). Shrinkage, GMANOVA, control variates and their applications, Ph.D. Thesis, Department of Statistics, Purdue University, Indiana.

Wilson, J. R. (1984). Variance reduction techniques for digital simulation, *Mathematics and Management Science*, **1**, 227–312.

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate student t-error terms, *J. Amer. Statist. Assoc.*, **71**, 400–405.

Zidek, J. (1978). Deriving unbiased risk estimators of multinormal mean and regression coefficient estimators using zonal polynomials, *Ann. Statist.*, **6**, 769–782.