ESTIMATORS WITH NONDECREASING RISK: APPLICATION OF A CHI-SQUARED IDENTITY

BU-899-M Revised March, 1989

by George Casella

Biometrics Unit, Cornell University, Ithaca, NY 14853

Abstract

By using an apparently little known fact about concave functions together with a new expectation identity for noncentral chi-squared random variables, a characterization of risk functions of Stein-type estimators is obtained. In particular, concavity of the function appearing in the shrinkage factor is related to the estimator's risk function being nondecreasing.

AMS 1980 *subject classifications.* 62C20, 62F20.

Key words and phrases. Minimax, Stein Estimation, Multivariate Normal.

1. Introduction. Let X be an observation from a p-variate ($p \geq 3$) normal distribution with mean θ and identity covariance matrix. For any estimator $\delta(X)$ of θ, the loss in estimating θ by $\delta(X)$ is

$$L(\theta,\delta) = |\theta - \delta(X)|^2 ,$$

whee $|\cdot|$ denotes Euclidean distance. The risk of $\delta(X)$, $R(\theta,\delta)$, is given by

$$R(\theta,\delta) = E_\theta L(\theta,\delta) = E_\theta|\theta - \delta(X)|^2 ,$$

and $\delta(X)$ is a minimax estimator of θ if and only if $R(\theta,\delta) \leq R(\theta,X) = p$ for all θ. Many Stein-type estimators have the form

$$\delta(X) = \left(1 - \frac{r(|X|^2)}{|X|^2}\right) X ,$$

and theorems about the risk behavior of $\delta(X)$ give conditions on $r(|X|)$. In particular, common conditions are that $r(t)$ be nondecreasing and $r(t)/t$ be nonincreasing. The following lemma relates the property of concavity to these conditions.

LEMMA 1: Let $r:[0,\infty) \rightarrow [0,\infty)$ be concave. Then

i) $r(t)$ is nondecreasing,

ii) $r(t)/t$ is nonincreasing.

PROOF: Since $r(\cdot)$ is concave, it follows that

$$r\Big((1-\lambda)t_1 + \lambda t_2\Big) \geq (1-\lambda)r(t_1) + \lambda r(t_2)$$

for $0 \leq \lambda \leq 1$. Moreover, this inequality is reversed if $\lambda > 1$.

We now prove part i) by contradiction. Suppose $t_1 < t_2$ and $r(t_1) > r(t_2)$. Then the function of λ, $f(\lambda) = (1-\lambda)r(t_1) + \lambda r(t_2)$ is decreasing and, for sufficiently large $\lambda > 1$, it is negative. For sufficiently large λ we then have

$$r\Big((1-\lambda)t_1 + \lambda t_2\Big) \leq (1-\lambda)r(t_1) + \lambda r(t_2) < 0 ,$$

which contradicts the fact that $r(\cdot)$ is nonnegative on $[0,\infty)$ and establishes part i).

To prove part ii) write, for $0 < t_1 < t_2$, $t_1 = \left(1 - \frac{t_1}{t_2}\right) \times 0 + \left(\frac{t_1}{t_2}\right) \times t_2$. By the concavity of $r(\cdot)$ we have

$$r(t_1) = r\left(\left[1 - \frac{t_1}{t_2}\right] \times 0 + \left[\frac{t_1}{t_2}\right] \times t_2\right) \geq \left[1 - \frac{t_1}{t_2}\right] r(0) + \frac{t_1}{t_2}\, r(t_2) \geq \frac{t_1}{t_2}\, r(t_2) ,$$

where the second inequality follows from the fact that $r(0) \geq 0$. Thus $r(t_1)/t_1 \geq r(t_2)/t_2$ for $t_1 < t_2$, establishing part ii). □

Therefore, the two original minimax conditions are tied together through the property of concavity. The converse of Lemma 1 is false, so there can exist minimax estimators satisfying i) and ii) with nonconcave $r(t)$. However, most familiar estimators do, in fact, have concave $r(t)$.

Although the characterization in Lemma 1 may seem new, similar results are well known. For example, Barlow and Proschan (1975) define a function $g(t)$ on $[0,\infty)$ to be *star-shaped* if $g(t)/t$ is increasing. They then give an exercise to show that convex functions on $[0,\infty)$ are star-shaped, which is quite close to the result of Lemma 1.

2. **A Chi-Squared Expectation Identity.** In the following, let χ_p^2 denote a noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter $|\theta|^2/2$. The value of θ is indicated by the subscript on the expectation operator.

LEMMA 2: Let $h:[0,\infty) \rightarrow (-\infty,\infty)$ be differentiable. Then, provided both sides exist,

$$\frac{\partial}{\partial|\theta|^2} E_\theta\left[h(\chi_p^2)\right] = E_\theta\left[\frac{\partial}{\partial\chi_{p+2}^2} h\left(\chi_{p+2}^2\right)\right].$$

PROOF. The lemma is established by equating the results of the well-known integration-by-parts technique with the results of some lesser-known identities for expectations of noncentral chi-squared random variables. We will proceed by evaluating the risk of the estimator $\delta(X) = \left(1 - [h(|X|^2)/|X|^2]\right)X$, where $X \sim N_p(\theta,I)$. The usual integration-by-parts yields

$$E_\theta|\theta - \delta(X)|^2 = p - 4E_\theta h'(|X|^2) + E_\theta\left\{\frac{h(|X|^2)}{|X|^2}\left[h(|X|^2) - 2(p-2)\right]\right\}. \tag{1}$$

We can also write

$$\begin{aligned} E_\theta|\theta - \delta(X)|^2 &= E_\theta\left|(\theta - X) + \left(X-\delta(X)\right)\right|^2 \\ &= p + 2E_\theta\left\{(\theta - X)'Xh(|X|^2)/|X|^2\right\} + E_\theta\left\{h^2(|X|^2)/|X|^2\right\}. \end{aligned} \tag{2}$$

We now employ the following identities, which can be found either in Bock (1975) or Casella (1980). If $h:[0,\infty) \to (-\infty,\infty)$, then provided the expectations exist,

$$E_\theta\left\{Xh(|X|^2)\right\} = \theta E_\theta\left\{h\left(\chi^2_{p+2}\right)\right\}, \tag{3}$$

$$|\theta|^2 E_\theta\left\{h\left(\chi^2_{p+2}\right)/\chi^2_{p+2}\right\} = E_\theta\left\{h\left(\chi^2_{p-2}\right)\right\} - (p-2)E_\theta\left\{h\left(\chi^2_p\right)/\chi^2_p\right\}, \tag{4}$$

$$\frac{\partial}{\partial|\theta|^2} E_\theta\left\{h\left(\chi^2_p\right)\right\} = \tfrac{1}{2}\left\{E_\theta h\left(\chi^2_{p+2}\right) - E_\theta h\left(\chi^2_p\right)\right\}. \tag{5}$$

Now, using (3) and (4) on the first expectation in (2), and rearranging terms, we obtain

$$\begin{aligned} E_\theta|\theta-\delta(X)|^2 &= p-2\left\{E_\theta h\left(\chi^2_p\right) - E_\theta h\left(\chi^2_{p-2}\right)\right\} \\ &\quad + E_\theta\left\{\frac{h(|X|^2)}{|X|^2}[h(|X|^2) - 2(p-2)]\right\}. \end{aligned} \tag{6}$$

We note in passing that, using the fact that the noncentral chi-squared distribution has monotone likelihood ratio in its degrees of freedom, equation (6) provides an immediate proof of the minimaxity of $\delta(X)$ provided $h(\cdot)$ is nondecreasing and $0 \le h \le 2(p-2)$.

Now, equating (1) and (6), cancelling common terms, and using (5) establishes

$$\frac{\partial}{\partial|\theta|^2} E_\theta h\left(\chi^2_{p-2}\right) = E_\theta h'(|X|^2) = E_\theta \frac{\partial}{\partial\chi^2_p} h\left(\chi^2_p\right),$$

proving the lemma. □

3. Nondecreasing Risk Functions. Stein-type estimators of a multivariate normal mean pull the maximum likelihood estimator, X, toward a particular point (which, without loss of generality, can be taken to be zero). This selected point should be interpreted as an experimenter's best (prior) guess at the true mean, and will locate the portion of the parameter space in which the greatest risk improvement will be attained. It should be expected then, that given X is close to the prior guess, there should be good risk improvement in that portion of the parameter space. Work of Berger (1982) addresses the question of how to best take advantage of this fact. The following theorem characterizes a class of minimax estimators whose minimum risk is at $|\theta| = 0$.

THEOREM 1: Let $\delta(X) = (1 - r(|X|^2)/|X|^2)X$, where $r:\{0,\infty) \to [0,\infty)$ is concave. If $0 \leq r(t) \leq 2(p-2)$ then $R(\theta,\delta)$ is a nondecreasing function of $|\theta|$.

PROOF. Assume, for the moment, that r is twice differentiable. Using (1) we have

$$E_\theta|\theta - \delta(X)|^2 = p - 4E_\theta r'(|X|^2) + E_\theta\left\{\frac{r(|X|^2}{|X|^2}\,[r(|X|^2) - 2(p-2)]\right\}. \tag{7}$$

The concavity of r, together with Lemma 1, insure that the function inside the second expectation in (7) is nondecreasing in $|X|^2$, hence the expectation is nondecreasing in $|\theta|^2$. Thus, we only need establish the $E_\theta r'(|X|^2)$ is nonincreasing in $|\theta|^2$. Using Lemma 2, we have

$$\frac{\partial}{\partial|\theta|^2}E_\theta r'(|X|^2) = \frac{\partial}{\partial|\theta|^2}E_\theta r'\left(\chi_p^2\right) = E_\theta r''\left(\chi_{p+2}^2\right) \leq 0$$

by the fact that r is concave. Thus, the theorem is established if r is twice-differentiable. If r is not twice-differentiable, we can take a sequence $\left\{r_n\right\}$ of twice-differentiable concave functions which uniformly approach r. By carrying out the above calculations, and passing to the limit, the theorem is readily established. □

From Lemma 1, it also follows that $\delta(X)$ of Theorem 1 is minimax. This can be seen from expression (7). The properties of $r(|X|^2)$ insure that the first expectation is positive and the second is negative.

Although the concavity condition on r seems rather strong, most familiar estimators satisfy the condition. These include not only the ordinary and postive-part James-Stein estimators, but also the proper Bayes minimax estimators of Strawderman (1971). We also note that Theorem 1 generalizes a result of Efron and Morris (1973), who show that the risk function of the ordinary James-Stein estimator is nondecreasing in $|\theta|$.

Acknowledgement. I would like to thank Roger Berger for pointing out the existence of Lemma 1.

REFERENCES

BARLOW, R.E., and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing.* New York: Holt, Rinehart and Winston, Inc.

BERGER, J. O. (1982). Selecting a minimax estimator of a multivariate mean. *Ann. Statist.* 10, 81-92.

BOCK, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* 3, 209-218.

CASELLA, G. (1980). Minimax ridge regression estimation. *Ann. Statist.* 8, 1036-1056.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *J. Amer. Statist. Assoc.* 68, 117-130.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* 9, 1135-1151.

STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* 42, 385-388.