JOURNAL OF MULTIVARIATE ANALYSIS 22, 137–143 (1987)

Estimation under $\mathscr{G}$-Invariant Quasi-convex Loss

K. C. MOSLER

Universität der Bundeswehr, Hamburg, West Germany

Communicated by the Editors

The classical point estimation problem is investigated under alternative loss functions which are quasi-convex and symmetric with respect to some subgroup of the orthogonal group in $\mathbb{R}^n$. A characterization of better estimators is proved and applied to scale and translation families of estimators. Finally, it is shown that every minimum variance unbiased normal estimator is best unbiased under arbitrary loss being quasi-convex and symmetric about the origin.

1. INTRODUCTION AND NOTATION

Consider a parametric estimation problem consisting of a triplet $(\mathscr{X}, \mathscr{B}, P_\vartheta)$, $\vartheta \in \theta \subset \mathbb{R}^m$ and a function $g: \theta \to \Omega \subset \mathbb{R}^n$. We are interested in an estimator t of $g(\vartheta)$, $t: \mathscr{X} \to \Omega$ element of some set of estimators D, which for given ϑ minimizes the risk

$$R(L, \vartheta, t) = \int L(t(x) - g(\vartheta))\, dP_\vartheta(x)$$

$$= E_\vartheta L(t(X) - g(\vartheta)),$$

where $L: \mathbb{R}^n \to \mathbb{R}_+$ denotes a loss function and X is the observed random variable. t^* is said to be better than t at ϑ, $t, t^* \in D$, if both risks exist and are finite, and

$$R(L, \vartheta, t^*) \leqslant R(L, \vartheta, t). \tag{1.1}$$

For example, let L be a quadratic error loss, i.e.,

$$L(y) = y^T A y, \qquad y \in \mathbb{R}^n, \tag{1.2}$$

Received October 5, 1982; revised November 23, 1983.

Key words and phrases: quasi-convex loss function, minimum variance unbiased estimator, unimodal density.

AMS 1979 subject classifications: Primary 62H12; secondary 62C05.

0047-259X/87 $3.00

with some positive definite $n \times n$ matrix A. When $t^*(X)$ and $t(X)$ have finite second moments and $E_\vartheta t^*(X) = E_\vartheta t(X)$, it is well known that (1.1) holds for all loss functions of type (1.2) if and only if $\mathrm{Cov}_\vartheta t(X) - \mathrm{Cov}_\vartheta t^*(X)$ is a positive semi-definite matrix.

However, from a decision-theoretic point of view, quadratic error loss (or, equivalently, variance) seems to be a very special device; moreover, second moments of the estimators may not exist. The question whether a given estimator, being best in some class D under quadratic loss, is also best under certain alternative loss functions has been posed early; see, e.g., Blackwell and Girshick [2, p. 111]. Answers have been obtained mainly in the case of a convex loss function: Padmanabhan [5] has shown that a bounded univariate estimator which is uniformly best unbiased (UBUE) under quadratic loss is also UBUE under symmetric convex loss; for related results see Schmetterer [6]. Näther [4] has characterized robustness properties of univariate estimators against quasi-convex loss and of multivariate normal estimators against arbitrary convex loss, while Bamberg and Rauhut [1] provided counterexamples in the non-normal multivariate case.

For any convex loss function, the marginal loss of error is increasing. By this, loss functions like bounded absolute error loss (e.g., $L(y) = 0$ if $\|y\| \leqslant \beta$, $L(y) = 1$ else) are ruled out. In this paper we investigate loss functions of a more general type (where the marginal loss can be arbitrarily chosen), namely loss functions which are quasi-convex and symmetric with respect to some groups of L-measure preserving transformations of $\mathbb{R}^n$.

A loss function $L: \mathbb{R}^n \to \mathbb{R}_+$ is quasi-convex if for all $y, z \in \mathbb{R}^n$ and $\beta \in]0, 1[$,

$$L(\beta y + (1-\beta) z) \leqslant L(y) \vee L(z),$$

or, equivalently, for all $\alpha \in \mathbb{R}_+$ the sets $H(L, \alpha) = \{ y \in \mathbb{R}^n \mid L(y) \leqslant \alpha \}$ are convex. Obviously, every convex loss, hence every quadratic loss, is quasi-convex. Let $\mathcal{O}$ denote the group of orthogonal transformations in $\mathbb{R}^n$, and $\mathcal{G}$ some subgroup of $\mathcal{O}$. A set $K \subset \mathbb{R}^n$ is $\mathcal{G}$-invariant if $y \in K$, $\psi \in \mathcal{G}$ implies $\psi(y) \in K$. L is called $\mathcal{G}$-invariant if

$$L(\psi(y)) = L(y)$$

holds for all $y \in \mathbb{R}^n$ and $\psi \in \mathcal{G}$, or, equivalently, the sets $H(L, \alpha)$ are $\mathcal{G}$-invariant for all $\alpha \in \mathbb{R}_+$.

In the sequel we consider loss functions L with

(i) L quasi-convex

(ii) L $\mathcal{G}$-invariant, and

(iii) $L(o) = 0$.

The class of those losses is denoted by $Q(G)$; the class of all $\mathscr{G}$-invariant convex sets which contain the origin is denoted by $S(\mathscr{G})$. Thus, we can write

$$Q(\mathscr{G}) = \{L: \mathbb{R}^n \to \mathbb{R}_+ \mid H(L, \alpha) \in S(\mathscr{G}), \alpha \in \mathbb{R}_+\}. \tag{1.3}$$

To illustrate the definitions consider the group $\mathscr{G}_1 = \{\text{id}, -\text{id}\}$ consisting of the identity transformation and its negative: any quadratic loss is in $Q(\mathscr{G}_1)$. The extremes are marked by $\mathscr{G}_0 = \{\text{id}\}$ and the full group $\mathcal{O}$; $Q(\mathscr{G}_0)$ is the class of all quasi-convex losses with infimum equal to zero at the origin while $Q(\mathcal{O})$ consists of all L of type

$$L(y) = l(\|y\|) \tag{1.4}$$

with arbitrary non-decreasing $l: \mathbb{R}_+ \to \mathbb{R}_+$, $l(0) = 0$.

In Theorem 1 of Section 2 we characterize (1.1) for all $L \in Q(\mathscr{G})$ by a probability inequality on the sets in $S(\mathscr{G})$; in Theorem 2 the result is specialized to the groups $\mathcal{O}$ and $\mathscr{G}_0$. Three corollaries treat the cases when the distributions of $t^*(X)$ and $t(X)$ differ by a $\mathscr{G}$-transformation, a scale vector, a translation vector, and a convolution, respectively. Finally, in Section 3 it is shown that every minimum variance unbiased normal estimator is best under arbitrary quasi-convex loss L with $L(y) = L(-y)$, $y \in \mathbb{R}^n$.

2. Comparison of Estimators when $L \in Q(\mathscr{G})$

Let $\mathscr{G} \subset \mathcal{O}$ be given, $\vartheta \in \theta$ and $t, t^* \in D$.

THEOREM 1. (1.1) *holds for all* $L \in Q(\mathscr{G})$ *both risks of which are finite if and only if for all* $B \in S(\mathscr{G})$,

$$P_\vartheta(t^*(X) - g(\vartheta) \in B) \geqslant P_\vartheta(t(X) - g(\vartheta) \in B). \tag{2.1}$$

Proof. Let μ denote the probability distribution of $Y = t(X) - g(\vartheta)$ at ϑ, and let ν denote the distribution of $Z = t^*(X) - g(\vartheta)$ at ϑ.

With these notations,

$$R(L, \vartheta, t) - R(L, \vartheta, t^*) = \int L \, d\mu - \int L \, d\nu$$

$$= \int_{-\infty}^{\infty} [P_\vartheta(Z \in H(L, \beta)) - P_\vartheta(Y \in H(L, \beta))] \, d\beta, \tag{2.2}$$

where $H(L, \beta) = \{y \in \mathbb{R}^n \mid L(y) \leqslant \beta\}$.

Assume (2.1) for all $B \in S(\mathscr{G})$; let $L \in Q(\mathscr{G})$ be given, hence $H(L, \beta) \in S(\mathscr{G})$ for all β. We conclude from (2.2) $R(L, \vartheta, t) - R(L, \vartheta, t^*) \geqslant 0$, i.e., (1.1). On the other hand, assume (1.1) for all $L \in Q(\mathscr{G})$, and let $B \in S(\mathscr{G})$ be given; then, the indicator function 1_B is in $-Q(\mathscr{G})$: $0 \leqslant R(1_B, \vartheta, t^*) - R(1_B, \vartheta, t) = P_\vartheta(Z \in B) - P_\vartheta(Y \in B)$. Q.E.D.

From Theorem 1 we follow three corollaries, the first of which states that the risk remains constant under a $\mathscr{G}$-transformation of the estimator and that the risk is isotone in the scale parameter when the estimators are from a scale parameter family; here, $\circ$ denotes componentwise multiplication of vectors in $\mathbb{R}^n$. Let $L \in Q(\mathscr{G})$ be given, and assume that both risks $R(L, \vartheta, t^*)$ and $R(L, \vartheta, t)$ are finite.

Corollary 1. (i) *If $t(X) - g(\vartheta)$ is distributed as $\psi(t^*(X) - g(\vartheta))$ at ϑ for some $\psi \in \mathscr{G}$ then $R(L, \vartheta, t^*) = R(L, \vartheta, t)$.*

(ii) *If $t(X) - g(\vartheta)$ is distributed as $a \circ (t^*(X) - g(\vartheta))$ at ϑ for some $a \in \mathbb{R}^n$, $a_i \geqslant 1$ for all i, then $R(L, \vartheta, t^*) \leqslant R(L, \vartheta, t)$.*

Proof. Part (i) is obvious from $\psi^{-1}(B) = B$, $B \in S(\mathscr{G})$. As every $B \in S(\mathscr{G})$ is convex and contains the origin we have $(1/a_1, ..., 1/a_n)^T \circ B \subset B$, hence (2.1); and part (ii) follows from Theorem 1.

Corollary 2. *Let f_ϑ be a $\mathscr{G}$-invariant probability density in $\mathbb{R}^n$ with $\{y \mid f_\vartheta(y) > \alpha\}$ convex for every $\alpha \in \mathbb{R}$. Assume that for every $s \in \{t, t^*\}$, $s(X)$ has a density $y \mapsto f_\vartheta(y - g(\vartheta) + w_s)$ at ϑ with some translation parameter w_s. Then* (1.1) *holds if $w_{t^*} \in \operatorname{conv}\{\psi(w_t) \mid \psi \in \mathscr{G}\}$, where* conv *denotes the convex hull.*

Proof. For arbitrary $B \in S(\mathscr{G})$ we have to show (2.1), which is equivalent to

$$\int_B f_\vartheta(y + w_{t^*})\, dy \geqslant \int_B f_\vartheta(y + w_t)\, dy. \tag{2.3}$$

Under the assumptions of the corollary, (2.3) holds by a result of Mudholkar [3].

In Corollary 2 the assumptions sufficient for (1.1) may be reworded as follows: $t^*(X) - g(\vartheta)$ and $t(X) - g(\vartheta)$ belong to a translation family with translation parameters w_{t^*} and w_t; they have unimodular densities, and w_{t^*} is in the convex hull of the $\mathscr{G}$-orbit of w_t. Especially, when $\mathscr{G} = \mathscr{G}_1$, the last condition becomes $w_{t^*} = \beta w_t$ for some $\beta \in \mathbb{R}$, $|\beta| \leqslant 1$, since $\operatorname{conv}\{\psi(w_t) \mid \psi \in \mathscr{G}_1\} = \{\beta w_t \mid \beta \in [-1, 1]\}$.

Corollary 3. *Assume $\mathscr{G}_1 \subset \mathscr{G}$, and that the distribution $t^* P_\vartheta$ of $t^*(X)$ at*

ϑ has a density $y \mapsto f_\vartheta(y-g(\vartheta))$, and f_ϑ is as in Corollary 2. Then (1.1) holds if the distribution tP_ϑ of $t(X)$ equals the convolution

$$tP_\vartheta = t^*P_\vartheta * \lambda \tag{2.4}$$

with some probability measure λ on $\mathbb{R}^n$.

Proof. Let ν denote the distribution of $t^*(X)-g(\vartheta)$; and let $B \in S(\mathscr{G})$ be given, $z \in \mathbb{R}^n$. From $\mathscr{G}_1 \subset \mathscr{G}$ follows $o \in \operatorname{conv}\{\psi(-z) \mid \psi \in \mathscr{G}\}$; hence, when setting $w_{t^*}=0$ and $w_t=-z$ in Corollary 2 the premises of that corollary are met. Then, (2.3) yields the inequality

$$\nu(B-z) = \int_B f_\vartheta(y-z)\,dy \leqslant \int_B f_\vartheta(y)\,dy = \nu(B).$$

It follows, by (2.4), that

$$tP_\vartheta(B+g(\vartheta)) = (\nu * \lambda)(B) = \int \nu(B-z)\,\lambda(dz)$$

$$\leqslant \int \nu(B)\,\lambda(dz) = \nu(B) = t^*P_\vartheta(B+g(\vartheta)),$$

hence (2.1).

We conclude the section by specializing Theorem 1 to the groups $\mathcal{O}$ and $\mathscr{G}_0$. Recall that

$$Q(\mathcal{O}) = \{L \mid L(y) = l(\|y\|),\ y \in \mathbb{R}^n,\ l\colon \mathbb{R}_+ \to \mathbb{R}_+ \text{ isotone},\ l(0)=0\},$$

$$Q(\mathscr{G}_0) = \{L \mid L \text{ quasi-convex},\ L(o)=0\}.$$

THEOREM 2. (i) (1.1) *holds for all $L \in Q(\mathcal{O})$ both risks of which are finite if and only if*

$$F_\vartheta^* \geqslant F_\vartheta, \tag{2.5}$$

where F_ϑ^* and F_ϑ denote the distribution functions of $\|t^*(X)-g(\vartheta)\|$ and $\|t(X)-g(\vartheta)\|$ at ϑ, respectively.

(ii) (1.1) *holds for all $L \in Q(\mathscr{G}_0)$ both risks of which are finite if and only if* (2.1) *for all convex B which contain the origin.*

Proof. Note that

$$S(\mathscr{G}_0) = \{B \mid B \text{ convex},\ o \in B\},$$

$$S(\mathcal{O}) = \{B \mid B \text{ open or closed ball around } o\},$$

hence (ii) is obvious. For any closed ball $B \in S(\mathcal{O})$ with radius ρ, (2.1) means

$$F_\vartheta^*(\rho) = P_\vartheta(\|t^*(X) - g(\vartheta)\| \leqslant \rho)$$
$$\geqslant P_\vartheta(\|t(X) - g(\vartheta)\| \leqslant \rho) = F_\vartheta(\rho).$$

Therefore, (2.5) iff (2.1) for all closed $B \in S(\mathcal{O})$. As any open ball is a countable intersection of closed balls, (2.5) iff (2.1) for all $B \in S(\mathcal{O})$.

3. Application to Unbiased Normal Estimators

The results of Section 2 apply to a variety of distributions of known estimators. Here, we restrict ourselves to a first and most important application to normally distributed unbiased estimators.

THEOREM 3. *Let $\vartheta \in \Theta$ be given, and assume that for any $t \in D$ $t(X)$ is unbiased for $g(\vartheta)$ and normally distributed at ϑ. If t^* is minimum variance in D at ϑ then t^* is best in D at ϑ under any loss function which is quasi-convex and symmetric about the origin and yields finite risks.*

Proof. Let $t \in D$, and denote the covariance of tP_ϑ by T, the covariance of t^*P_ϑ by S. Then by assumption $T - S$ is positive semi-definite, thus $tP_\vartheta = t^*P_\vartheta * \lambda$, where λ is a normal distribution with zero mean and covariance $T - S$. Corollary 3 yields the proposition for any loss $L \in Q(\mathscr{G}_1)$, i.e., L quasi-convex and symmetric about $o \in \mathbb{R}^n$, as the centralized normal density is unimodal and $\mathscr{G}_1$-invariant. Q.E.D.

An analogous theorem with convex (not necessarily symmetric) loss functions has been proved by Näther [4]. We give an application of Theorem 3:

Consider the general linear model with normal errors

$$X = C\beta + U, \qquad U \sim N(0, \sigma^2 S_0),$$

where the $n \times p$ design matrix C has rank p, and S_0 is positive definite and known. The usual least-squares estimator is uniformly best linear unbiased not only in terms of minimum variance (by the Gauss–Markov theorem) but also (by Theorem 3) with respect to an arbitrary quasi-convex loss which is symmetric about the origin and yields finite risks.

References

[1] Bamberg, G., and Rauhut, B. (1972). *Lineare Regression bei alternativen Schadensfunktionen.* Methods of Operations Research Vol. 12, pp. 1–10. Athenäum, Hain, and Hanstein, Königstein, West Germany.

[2] Blackwell, D., and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[3] Mudholkar, G. S. (1966). The integral of an invariant unimodal function over an invariant convex set—An inequality and applications. *Proc. Amer. Math. Soc.* **17** 1327–1333.

[4] Näther, W. (1975). Semi-orderings between distribution functions and their application to robustness of parameter estimators. *Math. Operationsforsch. Statist.* **6** 179–188.

[5] Padmanabhan, A. R. (1970). Some results on minimum variance unbiased estimation. *Sankhya Ser. A* **32** 107–114.

[6] Schmetterer, L. (1974). Einige Resultate aus der Theorie der erwartungstreuen Schätzungen. In *Transactions of the 7th Prague Conference*, Vol. B, pp. 489–503.