LOWER BOUNDS FOR RISK FUNCTIONS IN ESTIMATION*

BY M. M. RAO

CARNEGIE INSTITUTE OF TECHNOLOGY

Communicated by S. Bochner, October 20, 1959

1. Introduction.—Let $T$ be an estimator of $\theta$, of a probability density function $p(x, \theta)$ where $x = (x_1, \ldots, x_n)$ is a vector of observations on a (vector or sequence) random variable (r.v.) $X$, and $\theta$ is an unknown parameter (a constant). If $R(T, \theta) = E_\theta W(T, \theta)$ is the risk function, where $E_\theta$ stands for the mathematical expectation when $\theta$ is the true parameter, and $W(T, \theta)$ is the loss in choosing $T$ when $\theta$ is the true value, then it is of considerable interest in theoretical statistics to obtain the best lower bounds for $R(T, \theta)$, without actually evaluating it. The actual computation of $R(T, \theta)$ is much more difficult since it is related to the solution of the distribution problem of $T$. If $W(T, \theta)$ is quadratic or $W(T, \theta) = |T - \theta|^p$, $p \geq 1$, several lower bounds have been obtained in the past, most of them, using the unbiased estimators, $T$ of $\theta$. In this paper, all these results are unified and extended without assuming unbiasedness. Non-sequential problems only are considered. As usual, the r.v.'s are denoted by capitals, and the values assumed by them by the corresponding small letters.

2. Assumptions.—Let $\theta = (\theta_1, \ldots, \theta_k)$ be a point in $A$, a bounded non-empty subset of a real Euclidean $k$-space. For any $x$ in $\mathcal{S}$, a real Euclidean space and $\theta$ in $A$, there exists a density $p(x, \theta)$ with respect to a fixed $\sigma$-finite measure $\mu$ defined on the $\sigma$-algebra of subsets of $\mathcal{S}$. It will be assumed that the carrier of $p(x, \theta)$, say $S(\subset \mathcal{S})$, remains invariant for all $\theta$ in $A$. Further the following regularity conditions are imposed on $p(x, \theta)$.

Condition I: All partial derivatives of $p(x, \theta)$ with respect to $\theta_i$ ($i = 1, \ldots, k$) exist for all $x$ in $\mathcal{S}$ and $\theta$ in $A$. (For later analysis, arrange these derivatives in some order. For definiteness, group the first order derivatives first, then the second order etc., and within groups use a lexicographic ordering. Then eliminate all those terms that are linearly dependent on the preceding ones, and divide the resulting sequence by $p(x, \theta)$. Denote the $i^{th}$ term by $D_i(x, \theta)$ or $D_i$ for short.

Condition II: For each $i$, $|p(x, \theta)D_i(x, \theta)| < M_i(x)$ and $\int S M_i(x) d\mu < \infty$ all $\theta$ in $\bar{A}$ (closure of $A$).

Condition III: The r.v.'s $D_{ij}$, $i = 1, 2, \ldots$, are square integrable $(p(x, \theta)d\mu)$.

Condition IV: If the column vector $T = (T_1, \ldots, T_k)'$ (prime for transpose) is an estimator of $\theta'$, where the $T_i$ have a finite second moment and $E_\theta T_i = \theta_i + b_i(\theta) = \alpha_i(\theta)$ with $b_i(\theta)$ as bias, then all the partial derivatives of $\alpha_i(\theta)$ exist with respect to the $\theta_i$.

It is possible to state several other sets of conditions in place of the set I–IV above.

3. Quadratic Loss.—The loss functions $W(T, \theta)$ to be considered here are of the following two types.

$$W_{\delta}(T, \theta) = [\sum_{i=1}^{k} (T_i - \theta_i)^{\delta}]^{1/\delta}, 1 < \delta \leq 2, \text{and } 2 < \delta \leq 4,$$  \hspace{1cm} (1)

$$W_{\delta}(T, \theta) = \sum_{i,j=1}^{k} (T_i - \theta_i)(T_j - \theta_j)a_{ij},$$  \hspace{1cm} (2)
where $a_i$ are arbitrary real numbers (non-stochastic and known). Note that if $T$ is unbiased, then $R_a(T, \theta) = E_aW_a(T, \theta)$, is the reciprocal of the familiar concentration ellipsoid given by Cramér.\(^3\)

Let $f_m^{(j)} = D_{(m-1)k+j}$, $j = 1, \ldots, k$, and $f_m = (f_m^{(1)}, \ldots, f_m^{(k)})$, $m = 1, 2, \ldots$ (If the sequence of column vectors $\{f_m\}$ is finite, and the last vector happens to be incomplete, then add zeros to complete it.) Let $L^2$ be the space generated by all finite linear combinations of $\{f_m\}$, i.e., $f_m$, $f_n$ in $L^2$ implies $Af_m + Bf_n$ in $L^2$ for all $(k \times k)$ constant matrices $A$ and $B$. If the resulting space is not complete, add the limit points of all such combinations under the norm $\|f_m\|^2 = \text{tr}(f_m f_n) = \sum_{i=1}^k E_{\theta}$.

Using the classical Schmidt procedure the $\{f_m\}$ sequence can be orthogonalized, and denote by $\{\varphi_n\}$, the orthonormal sequence so obtained. Let $A_n = (T, \varphi_n)$. From the above assumptions, $A_n$ can be calculated explicitly. If $\delta = 2$ in (1), one obtains the following

**Theorem 1.** Suppose the density function $p(x, \theta)$ satisfies the conditions I–IV of section 2. Let $T'$ be a (row) vector of estimators and $\theta$ the true parameter value in $A$. If $R(T, \theta) = E_\theta W(T, \theta)$ where $W(T, \theta)$ is given by (1) with $\delta = 2$, or (2), then the corresponding lower bounds are given by

$$R_2(T, \theta) \geq \sum_{n=1}^\infty \text{tr}(A_n A_n'),$$

$$R_\delta(T, \theta) \geq \sum_{n=1}^\infty a' (A_n A_n') a,$$

where $a' = (a_1, \ldots, a_n)$. The lower bounds in (3) and (4) are reached if and only if $(T - \theta')$ lies in $L^2$. The minimum risk estimator is essentially unique.

The above result enables one to obtain the following interesting

**Corollary 1.1** In order that the lower bounds in (3) or (4) (and hence those given by others\(^2-4\)) for $R(T, \theta)$, $\delta = 2$, may be reached, it is necessary that the estimator $T$ of $\theta'$ be unbiased.

**Corollary 1.2.** If a sufficient statistic, for $\theta$, exists, then, when $T - \theta'$ lies in $L^2$, $T$ is a (Borel) function of the sufficient statistic.

The above statements are proved with the help of certain simple facts of the $L^2$ space. In particular, a vector analog of the Bessel-inequality is employed.

It will be noticed that all the classical results\(^2-4\) on these inequalities are contained in (3). If $T$ is unbiased, the results on the concentration ellipsoid\(^2-8\) are subsumed under the bound given in (4).

In the case $1 < \delta < 2$, many difficulties arise to give bounds corresponding to (3) or (4), and restrictive assumptions on $p(x, \theta)$ are needed. If $\delta$ is of a certain form, then the restrictions are somewhat reasonable.

**Theorem 2.** Let $\delta = 2q/(2q - 1)$, $q \geq 1$ integer, and $\mu$ be the Lebesgue measure. Let $D_j\sqrt{p(x, \theta)}$ and $(T_j - \theta_j)\sqrt{p(x, \theta)}$ be in $L^2(S_x, \mu)$ for each $i, j$ (i.e. $\int_S D_j p(x, \theta)^{1/2} d\mu$, and $\int_S (T_j - \theta_j)^p p(x, \theta)^{1/2} d\mu$ exist), and suppose that $M = l \in b p(x, \theta)^{(q-2)/2} < \infty$, for all $\theta$ in $A$. If the conditions I–IV of section 2 hold, then
$R_\theta(T, \theta) = E_\theta[\sum_{i=1}^{\infty} (T_i - \theta_i)^2]^{1/2} \geq M^{-1}[\sum_{n=1}^{\infty} |A_n|^{2/(2q)}, (5)$

where $|A|^{-2} = \text{tr}(AA')$, $A_n$ is defined before, and $A^q = (a_{ij})^q$ if $A = (a_{ij})$.

Remark: If $q = 1$, then $\delta = 2$, and $M = 1$ so that (5) coincides with (3). Analogous result, if $2 < \delta \leq 4$, is given by the following

**Corollary 2.1.** Let the risk function be $\bar{R}_\theta(T, \theta) = E_\theta[\sum_{i=1}^{\infty} (T_i - \theta_i)^2]^{\delta}$, with $1 < \delta \leq 2$, and the density function $p(x, \theta)$ satisfy a Lipschitz condition of order $\alpha$ ($0 < \alpha \leq 1$). Let $\delta = 2q/(2q - 1), q \geq 1$ integer, and $D_i\sqrt{p(x, \theta)}$ and $(T_i - \theta_i)$ $\sqrt{p(x, \theta)}$ be in $L^q$ with $\mu$ as Lebesgue measure. If the conditions I–IV of section 2 hold, then

$$\bar{R}_\theta(T, \theta) \geq (\int_\mathbb{R} p(x, \theta)\delta^{-1}(\mu))^{-1}[\sum_{n=1}^{\infty} |A_n|^{2/(2q)}]^{\delta/q}. (6)$$

The theorem and its corollary are proved upon employing the result of Theorem 1, and an extension of a special form of Hausdorff-Young Theorem for the vector variables.

The requirement that $\mu$ be a Lebesgue measure is not quite essential, and $\mu$ can be, for instance, a counting measure. However, it is necessary that it be translation invariant to obtain the bounds (5) and (6).

If $\delta$ is allowed to vary continuously, further restrictions on $p(x, \theta)$ should be imposed. That bound will not be presented here as the result obtained is not very pleasing.

4. Convex Loss: The most interesting type of loss functions can be subsumed under the following type. Let $y = t - \theta$, and $W_\theta(y)$ stand for a function whose argument depends only on $(t - \theta)$, but which may contain $\theta$ in some other manner. Let $W_\theta(y)$ satisfy the conditions: (a) $W_\theta(y) \geq 0$, for all $\theta$ in $A$, and $y$, (b) $W_\theta(0) = 0$, (c) $W_\theta(y) = W_\theta(-y)$, (d) $W_\theta(y)$ is convex in $y$ for each $\theta$ in $A$, and is measurable in both $y$ and $\theta$, and (e) $E_\theta W_\theta(T) < \infty$ for all $\theta$ in $\bar{A}$, where $T = T - \theta$, and $T$ is an estimator of the (single) parameter $\theta$.

It will be noted that, if $W_\theta(y)$ satisfies the conditions (a)–(e), there exists a function $V_\theta(y)$ satisfying (a)–(e) and such that $V_\theta^k(y) = W_\theta(y)$ for some $k (\geq 1)$. Let $k_0$ be the largest such value of $k$.

**Theorem 3.** Let $T$ be an estimator of $\theta$, in $A$, the true parameter. Suppose $p(x, \theta)$ satisfies conditions I–IV of section 2, for $i = 1$ (i.e. there is only $D_1$), and $D_i\xi$ is integrable $(p(x, \theta)d\mu)$, where $k^{-1} + k'^{-1} = 1$, and $1 \leq k \leq k_0$. If the loss function $W_\theta(y)$ satisfies the conditions (a)–(e) above, then

$$R(T, \theta) = E_\theta[W_\theta(T)] \geq W_\theta\left[\frac{1 + \mu'(\theta)}{E_\theta[|D_1|]}\right][E_\theta[|D_1|]/E_\theta[|D_1|]^{k^{-1}}(|D_1|^{k_0})]^{k}, (7)$$

where if $k_0 = 1$, $E_\theta^{k^{-1}}[D_1]^{k_0}$ is taken as the ess. sup. of $|D_1|$, and $b(\theta)$ is, as usual, the bias with $b'(\theta) = \frac{db(\theta)}{d\theta}$.

This theorem is proved using several inequalities on convex functions, and then showing that the second factor in square brackets on the right of (7) is a monotone
increasing function of $k$. If $W(y) = |y|^2$, (7) reduces to the classical Cramér-Rao inequality.\textsuperscript{3, 4} Other specializations lead to certain other results.

5. An Extension: A further extension of the convex loss is the following which includes some results of Barankin.\textsuperscript{1} Let $\varphi(t)$ and $\psi(t)$ be two non-negative symmetric convex functions such that $\varphi(0) = \psi(0) = 0$, and

$$|ab| \leq \varphi(a) + \psi(b),$$

for any real $a$ and $b$. (8)

It is seen that $\varphi$ and $\psi$ satisfy the conditions (a)-(e) of the preceding section. Define the risk function as

$$R_\varphi(T, \theta) = \|T\|_\varphi = \sup_{\mu} \int_S |\hat{d}_\mu| p(x, \theta) d\mu,$$

with $\int_S \varphi(d_\mu) p(x, \theta) d\mu \leq 1. (9)$

$(\|D_\varphi\|_\psi$ is defined similarly.) If $\varphi(u) = |u|^k/k$ and $\psi(u) = |u|^k/k'$ where $k^{-1} + k'^{-1} = 1$, and $T$ is unbiased, then $R_\varphi(T, \theta) = \|T\|_\varphi = k'^{1/k}T_k = k^{1/k}R_k(T, \theta)$. Barankin\textsuperscript{1} has obtained lower bounds for $R_k^{1/k}(T, \theta)$. The risk function in (9), therefore, includes his.

Let $L^r$ and $L^s$ be the classes of all r.v.'s $T$ and $D_1$ and all of their finite linear combinations such that $\|T\|_\varphi$ and $\|D_1\|_\varphi$ are finite. It is noted that these spaces are complete (or are completed by adding the limit points under their respective norms) linear spaces.

**Theorem 4.** Suppose $M(T)$ is the subspace of $L^r$ that contains the elements $T(= T - \theta)$, where $T$ is an estimator of $\theta$, of $p(x, \theta)$. Then (i) $M(T)$ is non-empty if and only if there exists a positive constant $K$ (independent of $n$) such that for every set of r.v.'s $D_{11}, \ldots, D_{1n}$ in $L^k$, where $\varphi(u)$ and $\psi(u)$ satisfy (9) and $\varphi(2u) \leq C\varphi(u)$, where $C$ is a finite positive constant independent of $u$, and for any set of $n$ real numbers $a_1, \ldots, a_n$, the following inequality obtains:

$$\left| \sum_{j=1}^{n} a_j a^{ij}(\theta) \right| \leq K \left\| \sum_{j=1}^{n} a_{ij} D_{ij} \right\|_{\psi},$$

(10)

where $a(\theta) = E_\theta(T)$, and $a^{ij}(\theta) = \frac{d^2 a(\theta)}{d\theta^j}$, (ii) for every $T$ in $M(T)$, $\|T\|_\varphi \geq C_0$, where $C_0 = g lb K$ satisfying (10), and (iii) if there is a $T^*$ in $M(T)$ such that $\|T^*\|_\varphi = C_0$ then it is essentially unique.

This theorem is a consequence of some results in functional analysis and the Zaanen representation theorem\textsuperscript{4} for linear functionals in the $L^r$ spaces.

The details of the proofs, illustrations, and some extensions will be published separately.

---


\textsuperscript{7} Zygmund, Antoni, Trigonometrical Series (Warsaw: 1935).