PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 13, Fasc. 2 (1992), pp. 215–228

# ORDERING AND COMPARING ESTIMATORS BY MEANS OF GINI SEPARATION INDEX

BY

VIERA GÁFRIKOVÁ (WARSZAWA)
AND MAGDALENA NIEWIADOMSKA-BUGAJ (MORGANTOWN, WV)

*Abstract.* The suggested ordering of estimators uses the separation measure between the population with true and estimated values of parameter. It allows us to choose one estimator as better in the cases where variances are equal or where information contained in variances is not satisfactory.

**1. Introduction.** The purpose of this paper is to suggest a new method of comparing estimators. The general motivating idea may briefly be described as follows. Let $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ be a family of distributions of the observed random variable. If $\theta$ is the true value of the parameter and our estimate is $\theta'$ (i.e. $\theta'$ is the observed value of the estimator, say $\hat{\theta}$), then the relevant question should be "How close are the distributions $P_\theta$ and $P_{\theta'}$?" rather than "How close are $\theta$ and $\theta'$?" Now, closeness between distributions can be expressed in a number of ways. We propose to regard distributions as "close" if they are difficult to discriminate. Thus we can take as the measure of closeness any index of discrimination or separability. In this paper we investigate the possibilities of taking the index $ar(F, G)$ based on the Gini Index of discrimination. If $\hat{\theta}$ is an estimator of $\theta$, then the performance of $\hat{\theta}$ can be judged by the properties of the random variable $AR_\theta(\hat{\theta})$, defined for a sample point $x$ as $ar(P_{\hat{\theta}(x)}, P_\theta)$.

While from the point of view of the general statistical theory this approach consists simply of taking $ar(P_{\hat{\theta}(x)}, P_\theta)$ as a loss function $L(\hat{\theta}(x), \theta)$, the nature of index $ar$ allows for more than the usual approach based on the concept of risk $\mathrm{E}\{AR_\theta(\hat{\theta})\}$ and the corresponding notion of admissibility. One namely can order estimators by requiring that the random variables $AR_\theta(\hat{\theta})$ be stochastically ordered for each $\theta$. Such an ordering is stronger than that based on domination of risk functions.

What is perhaps more important, in some cases the distribution of $AR_\theta(\hat{\theta})$ does not depend on $\theta$. This opens up an interesting possibility of comparing estimators of different parameters in unrelated families on one universal scale of performance of estimators.

**2. The Lorenz order.** Let $\mathscr{L}$ be the class of all non-negative random variables with positive finite expectations. For any random variable $X$ in $\mathscr{L}$ the Lorenz curve $L_X$ is defined by

$$(1) \qquad L_X(u) = \mu^{-1} \int_0^u F_X^{-1}(y)dy, \qquad u \in [0, 1],$$

where $\mu$ is the mean value of $X$ and the inverse distribution function $F_X^{-1}$ is given by

$$F_X^{-1}(y) = \sup\{x\colon F_X(x) \leqslant y\}.$$

The Lorenz partial order $\leqslant_L$ on $\mathscr{L}$ is defined as follows:

$$X \leqslant_L Y \Leftrightarrow L_X(u) \geqslant L_Y(u), \qquad u \in [0, 1].$$

Let $X_1, X_2, \ldots$ be i.i.d. random variables from $\mathscr{L}$ and let $\bar{X}_n$, $n = 1, 2, \ldots$, and $X_{i:n}$, $i = 1, \ldots, n$, denote the corresponding sample mean and the $i$-th order statistic. Arnold and Villaseñor [2] showed that for any $n$ we have

(i) $\bar{X}_n \leqslant_L \bar{X}_{n-1}$;

(ii) if the common density of the $X_i$'s is symmetric on the interval $[0, \Theta]$, then $X_{n+2:2n+3} \leqslant_L X_{n+1:2n+1}$;

(iii) in certain cases the sample median is Lorenz ordered with respect to the sample mean.

Since sample means, and also medians, are estimators of the population mean, Arnold and Villaseñor attempt to use the corresponding Lorenz curves to evaluate and compare variability of these estimators. The ordering proposed in the present paper, although different from $\leqslant_L$, refers in a way to $\leqslant_L$ and to the summary measure of inequality called the *Gini Index* [1, p. 35]. We shall analyze properties of this ordering for estimators of location parameters. In particular, we obtain some results concerning ordering of sample means and medians which are analogous to the results given by Arnold and Villaseñor in the case of $\leqslant_L$ for sample means and medians.

**3. A separation measure analogous to the Gini Index.** The degree of separation between two distributions can be defined and then estimated in various ways. For instance, Kowalczyk [5] used a decision theoretic approach to this problem, treating it as a discrimination between two classes. This idea will be shortly presented in the sequel. Let $I$ be the classified variable, $Z$ be the observed random variable, and let $\mathscr{Z}$ be the support of $Z$. Furthermore, let $f_i$ and $F_i$ denote the density (with respect to some measure $\nu$) and cdf of random variable $Z_i$ defined as $Z|I = i$ $(i = 1, 2)$. We shall consider randomized decision rules $\delta = (\delta_1, \delta_2)$, where $\delta_i(z)$ is the probability that the classified object is allocated to class $i$ when $z$ is observed, so that $\delta_1+\delta_2 \equiv 1$. Let $a_{ij}(\delta) = \int_{\mathscr{X}} f_i(z)\delta_j(z)d\nu(z)$ be the probability that an object from class $i$ is classified as belonging to class $j$. A natural ordering of decision rules is given by

(2) $$\delta \leqslant \delta' \Leftrightarrow a_{ij}(\delta) \geqslant a_{ij}(\delta') \text{ for } i \neq j \ (i, j = 1, 2).$$

A rule $\delta$ will be called a *threshold* rule with respect to a function $h: \mathscr{X} \to \boldsymbol{R}$ if for some $s \in [0, 1]$ and $\eta$ it takes the form

$$\delta_1(z) = \begin{cases} 1 & \text{if } h(z) < \eta, \\ s & \text{if } h(z) = \eta, \\ 0 & \text{if } h(z) > \eta. \end{cases}$$

The most important case is where $h$ is the likelihood ratio

$$h(z) = \begin{cases} f_2(z)/f_1(z) & \text{if } f_1(z) > 0, \\ +\infty & \text{if } f_1(z) = 0. \end{cases}$$

It has been shown (see [3]) that the class $\varDelta$ of threshold rules with respect to $h$ is the minimal class of rules admissible with respect to the ordering (2). Now, for any set $U$ of decision rules $\delta = (\delta_1, \delta_2)$ we can consider the subset of unit squares consisting of all possible pairs of error rates $(a_{21}(\delta), a_{12}(\delta))$, $\delta \in U$. If $U$ is the set of all threshold rules with respect to some function $h$, the error rates will form a curve in unit square, joining points (0, 1) and (1, 0). In particular, if $U = \varDelta$, then this curve, denoted by $C_{F_1,F_2}$, is called a *divergence curve* (see [4]). Specifically,

(3) $$C_{F_1,F_2} = \{F_2^h(z)+s(F_2^h(z+)-F_2^h(z)),\ 1-F_1^h(z)-s(F_1^h(z+)-F_1^h(z))\},$$

where $z \in \mathscr{X}$, $s \in (0, 1)$, and $F_i^h$ is the cdf of $h(Z)|I = i$ $(i = 1, 2)$. Pairs of distributions can be ordered in the following way:

(4) $$(F_1, F_2) \leqslant_C (F_1^*, F_2^*) \Leftrightarrow C_{F_1^*,F_2^*} \leqslant C_{F_1,F_2}.$$

The latter inequality means that for any pair $(x^*, y^*) \in C_{F_1^*,F_2^*}$ there exists $y$ such that $y \geqslant y^*$ and $(x^*, y) \in C_{F_1,F_2}$.

If $h(Z)|I = i$ for $i = 1, 2$ are continuous random variables with the same support, then $C_{F_1,F_2}$ is equal to the Lorenz curve for $h(Z)|I = 1$ and the ordering (4) is equal to the Lorenz order for $h(Z)|I = 1$ and $h^*(Z^*)|I = 1$. It is also natural to consider an analogue of the Gini Index, denoted by $ar$ and defined as twice the area between the divergence curve and the segment joining the points (0, 1) and (1, 0). If $F_1^h$, $F_2^h$ are continuous, then

(5) $$ar(F_1, F_2) = 1-2\int_0^1 F_2^h((F_1^h)^{-1}(t))dt,$$

where $(F_1^h)^{-1}(t) = \sup\{x\colon F_1^h(x) \leqslant t\}$. Without danger of confusion, we shall use an alternative notation $C_{X,Y}$ for the divergence curve of distributions of

$X$ and $Y$; similarly, we shall write $ar(X, Y)$ instead of $ar(F_1, F_2)$ if it is clear that $F_1$ is the cdf of $X$ and $F_2$ is the cdf of $Y$. It can be shown that

(i) $0 \leqslant ar(F_1, F_2) \leqslant 1$ with $ar(F_1, F_2) = 0$ iff $F_1 = F_2$ and $ar(F_1, F_2) = 1$ if $\inf\{x: F_1(x) = 1\} \leqslant \sup\{x: F_2(x) = 0\}$;

(ii) if $(F_1, F_2) \leqslant_C (F_1^*, F_2^*)$, then $ar(F_1, F_2) \leqslant ar(F_1^*, F_2^*)$;

(iii) for any increasing function $\varphi$ we have $C_{Z_1, Z_2} = C_{\varphi(Z_1), \varphi(Z_2)}$, hence also $ar(Z_1, Z_2) = ar(\varphi(Z_1), (Z_2))$.

**4. Evaluating the estimators by the separability index** ***ar*.** Let $\mathscr{P} = \{P_\theta;\ \theta \in \Theta\}$ be a family of univariate distributions and let $F_\theta$ be the cdf of $P_\theta$. Let $\hat{\theta}(X^{(n)})$ be an estimator of $\theta$, $X^{(n)}$ being the random sample of size $n$. For any $\theta \in \Theta$ and any observed sample $x^{(n)}$, $ar$ can be used to measure how much the distribution function $F_{\hat{\theta}(x^{(n)})}$ differs from $F_\theta$. Thus, we deal with a random variable defined as

$$AR_\theta(\hat{\theta}) = ar(F_{\hat{\theta}(X^{(n)})}, F_\theta).$$

The distribution of $AR_\theta(\hat{\theta})$ serves to evaluate the quality of $\hat{\theta}$ in $\{P_\theta;\ \theta \in \Theta\}$. Generally, the smaller are the values of $ar(F_{\hat{\theta}(X^{(n)})}, F_\theta)$, the closer are the distributions $P_\theta$ and $P_{\hat{\theta}(X^{(n)})}$, hence the better is the estimator $\hat{\theta}(X^{(n)})$ of $\theta$.

Accordingly, $ar(F_{\hat{\theta}(X^{(n)})}, F_\theta)$ may play the role of the loss function, and the analogue of the risk function is

$$\mathrm{E}_\theta\{AR_\theta(\hat{\theta})\} = \int AR_\theta(u)\, dG_{\hat{\theta},\theta}(u),$$

where $G_{\hat{\theta},\theta}(u) = P_\theta\{\hat{\theta} \leqslant u\}$. The expectation exists, since $0 \leqslant AR_\theta(\hat{\theta}) \leqslant 1$.

The estimators of $\theta$ within the family $\mathscr{P}$ can now be partially ordered by the relation $\leqslant_{ar}$ defined as follows:

Definition 4.1. We say that $\hat{\theta}_1$ is *AR-dominated* by $\hat{\theta}_2$ (to be denoted as $\hat{\theta}_2 \leqslant_{ar} \hat{\theta}_1$) iff for all $\theta \in \Theta$ we have

$$AR_\theta(\hat{\theta}_2) \leqslant_{st} AR_\theta(\hat{\theta}_1),$$

where $X \leqslant_{st} Y$ means that $Y$ is stochastically larger than $X$. We say that $\hat{\theta}_1$ and $\hat{\theta}_2$ are *AR-equivalent* (to be denoted as $\hat{\theta}_1 \sim_{ar} \hat{\theta}_2$) if $\hat{\theta}_1 \leqslant_{ar} \hat{\theta}_2$ and $\hat{\theta}_2 \leqslant_{ar} \hat{\theta}_1$. If $\hat{\theta}_1 \leqslant_{ar} \hat{\theta}_2$ but not $\hat{\theta}_2 \leqslant_{ar} \hat{\theta}_1$, we write $\hat{\theta}_1 <_{ar} \hat{\theta}_2$.

We may now introduce the concept of $AR$-admissibility:

Definition 4.2. The estimator $\hat{\theta}$ is *AR-admissible* if there is no estimator $\hat{\theta}^*$ such that $\hat{\theta}^* <_{ar} \hat{\theta}$.

A weaker partial order in the class of estimators may be based on the risk function.

Definition 4.3. We say that $\hat{\theta}_1$ is *risk-dominated* by $\hat{\theta}_2$ iff

$$\mathrm{E}_\theta\{AR_\theta(\hat{\theta}_2)\} \leqslant \mathrm{E}_\theta\{AR_\theta(\hat{\theta}_1)\}$$

for all $\theta \in \Theta$. In this case we write $\hat{\theta}_2 \leqslant_R \hat{\theta}_1$.

Clearly, $\leqslant_{ar}$ is stronger than $\leqslant_R$, that is: if $\hat{\theta}_2 \leqslant_{ar} \hat{\theta}_1$, then $\hat{\theta}_2 \leqslant_R \hat{\theta}_1$. The relation $<_R$ and the risk-admissibility are defined in the usual way.

An important special case, illustrated by examples in this paper, occurs when the distribution of $AR_\theta(\hat{\theta})$ does not depend on $\theta$, so that we may write $H_{\hat{\theta}}(u) = P_\theta\{AR_\theta(\hat{\theta}) \leqslant u\}$. In this case, $\hat{\theta}_2 \leqslant_{ar} \hat{\theta}_1$ if $H_{\hat{\theta}_2}(u) \geqslant H_{\hat{\theta}_1}(u)$ for every $u$, and $\hat{\theta}_2 \leqslant_R \hat{\theta}_1$ if $\mathrm{E}_\theta[AR_\theta(\hat{\theta}_2)] \leqslant \mathrm{E}_\theta[AR_\theta(\hat{\theta}_1)]$, where in the last inequality both sides are independent of $\theta$.

Suppose now that we have two families of distributions, $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ and $\mathscr{Q} = \{Q_\xi : \xi \in \Xi\}$, and let $\hat{\theta}$ and $\hat{\xi}$ be estimators of $\theta$ and $\xi$ in the families $\mathscr{P}$ and $\mathscr{Q}$, respectively. Assume that the distribution of $AR_\theta(\hat{\theta})$ does not depend on $\theta$, and the distribution of $AR_\xi(\hat{\xi})$ does not depend on $\xi$. This situation makes it possible to compare (according to the relation $\leqslant_R$, and possibly also according to the relation $\leqslant_{ar}$) the estimators $\hat{\theta}$ and $\hat{\xi}$, even if $\mathscr{P}$ and $\mathscr{Q}$ concern two different sample spaces. Thus we have an intriguing possibility of comparing the quality of estimators of unrelated parameters on an absolute scale.

**5. Evaluating and comparing the estimators of location and scale.** In this section we consider families of continuous univariate random variables indexed by a one-dimensional parameter $\theta$ such that for any $x$ and for some density function $f$

$$f_\theta(x) = f(x-\theta) \tag{6}$$

and

(7) $f(x-\theta')/f(x-\theta)$ is a non-decreasing function of $x$ for $\theta < \theta'$.

If (6) holds, then $\theta$ is called a *location parameter*; if both (6) and (7) hold, then we say that the family has a *monotone likelihood ratio* with respect to the location parameter $\theta$. Condition (7) is satisfied iff $-\log f(x)$ is a convex function in some open interval $(a, b)$ such that $-\infty \leqslant a < b \leqslant +\infty$ and $\int_a^b f(x)dx = 1$. We shall prove the following theorem:

THEOREM 5.1. *Let $\mathscr{P} = \{P_\theta, \theta \in \Theta\}$ be a family with a monotone likelihood ratio with respect to location parameter $\theta$. Let $\hat{\theta}$ be any estimator of $\theta$. Then the following conditions hold*:

(i) *$AR_\theta(\hat{\theta})$ is an increasing function of $|\hat{\theta}-\theta|$.*

(ii) *If $\hat{\theta}$ is a weakly (strongly) consistent estimator of $\theta$, then $AR_\theta(\hat{\theta})$ converges weakly (strongly) to* 0.

(iii) *If for any $c$ we have*

$$\hat{\theta}(X_1+c, \ldots, X_n+c) = c+\hat{\theta}(X_1, \ldots, X_n), \tag{8}$$

*then the distribution of $AR_\theta(\hat{\theta})$ does not depend on $\theta$.*

(iv) *$\hat{\theta}_2 \leqslant_{ar} \hat{\theta}_1$ iff for any $\theta$ we have $|\hat{\theta}_2-\theta| \leqslant_{st} |\hat{\theta}_1-\theta|$.*

Proof. Let $F$ denote the distribution function corresponding to density $f$. Conditions (6) and (7) imply that for any $\theta_1, \theta_2 \in \Theta$

$$ar(F_{\theta_1}, F_{\theta_2}) = 1-2\int_0^1 F\big(F^{-1}(t)-|\theta_1-\theta_2|\big)dt. \tag{9}$$

Indeed, observe first that since $h$ is monotone by (7), we may use (5) with $h$ omitted. If $\theta_1 < \theta_2$, then we take $F_2(x) = F(x-\theta_2)$ and $F_1(x) = F(x-\theta_1)$; hence $F_1^{-1}(t) = F^{-1}(t)+\theta_1$. Therefore

$$ar(F_1, F_2) = 1-2\int_0^1 F\big(F^{-1}(t)+\theta_1-\theta_2\big)dt.$$

A similar formula is obtained when $\theta_2 < \theta_1$, which proves (9). It follows that for any estimator $\hat{\theta}$

$$AR_\theta(\hat{\theta}) = 1-2\int_0^1 F\big(F^{-1}(t)-|\hat{\theta}-\theta|\big)dt \tag{10}$$

and properties (i) and (ii) follow easily. Property (iii) follows from the fact that the density of $\hat{\theta}$ is of the form $\tau_\theta(u) = \tau(u-\theta)$ for some density function $\tau$, while (iv) follows from the fact that the same increasing function serves to define $AR_\theta(\hat{\theta}_1)$ and $AR_\theta(\hat{\theta}_2)$. ∎

If the carrier of the distribution is bounded by $\theta$, we have the following theorem:

THEOREM 5.2. *Let $\mathcal{P}$ be a family of univariate continuous distributions indexed by a real parameter $\theta \in (a, b)$, where $-\infty \leqslant a < b \leqslant \infty$. Let $r$ be a strictly positive function defined on $[a, b]$ such that $\int_a^b r(x)dx < \infty$.*

(i) *If*

$$f_\theta(x) = \begin{cases} k(\theta)r(x) & \text{for } \theta \leqslant x \leqslant b, \\ 0 & \text{otherwise}, \end{cases}$$

*where $k(\theta) = 1/\int_\theta^b r(u)du$, then*

$$AR_\theta(\hat{\theta}) = 1-\frac{\min[k(\theta), k(\hat{\theta})]}{\max[k(\theta), k(\hat{\theta})]}.$$

(ii) *If*

$$f_\theta(x) = \begin{cases} k^*(\theta)r(x) & \text{for } a \leqslant x \leqslant \theta, \\ 0 & \text{otherwise}, \end{cases}$$

*where $k^*(\theta) = 1/\int_a^\theta r(u)du$, then*

$$AR_\theta(\hat{\theta}) = \frac{\min[k^*(\theta), k^*(\hat{\theta})]}{\max[k^*(\theta), k^*(\hat{\theta})]}.$$

Proof. (i) The proof follows from the fact that for any $\theta_1 < \theta_2$ the function $h(x)$ takes the form

$$h(x) = \begin{cases} 0 & \text{for } \theta_1 \leqslant x < \theta_2, \\ k(\theta_2)/k(\theta_1) & \text{for } x \geqslant \theta_2. \end{cases}$$

Consequently, $F_1^h$ is concentrated at two points, 0 and $k(\theta_2)/k(\theta_1)$, and its cdf is

$$F_1^h(x) = \begin{cases} 0 & \text{for } x \leqslant 0, \\ k(\theta_1)\int_{\theta_1}^{\theta_2} r(u)du & \text{for } 0 < x \leqslant k(\theta_2)/k(\theta_1), \\ 1 & \text{for } x > k(\theta_2)/k(\theta_1). \end{cases}$$

On the other hand, $F_2^h$ is concentrated at one point, $k(\theta_2)/k(\theta_1)$.

The divergence curve (3) consists of two segments: one vertical segment joining points (0, 1) and $(0, 1-k(\theta_1)\int_{\theta_1}^{\theta_2} r(u)du)$, and a linear segment joining the latter point with (1, 0). Observe that

$$1-k(\theta_1)\int_{\theta_1}^{\theta_2} r(u)du = k(\theta_1)/k(\theta_2).$$

The index *ar* is twice the area between the divergence curve and the line connecting (0, 1) and (1, 0), hence equals $1-k(\theta_2)/k(\theta_1)$ (see Fig. 1).

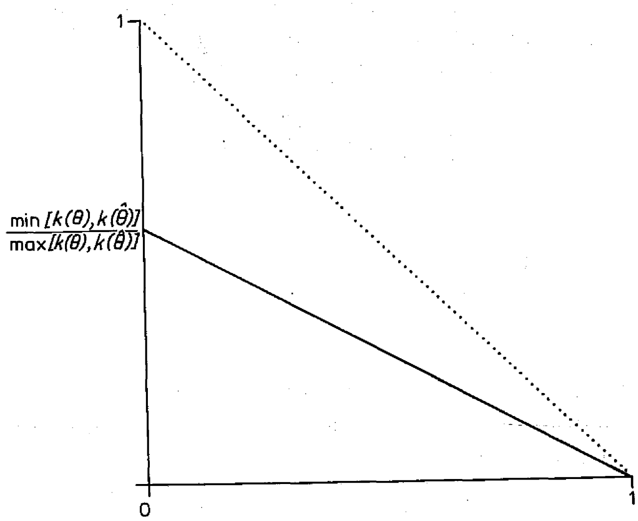



Fig. 1. The divergence curve for $(F_{\theta_1}, F_{\theta_2})$

In general, therefore,

$$ar(F_{\theta_1}, F_{\theta_2}) = 1-\frac{\min[k(\theta_1), k(\theta_2)]}{\max[k(\theta_1), k(\theta_2)]},$$

and taking $\theta = \theta_1$, $\hat{\theta} = \theta_2$ we obtain the proof of (i).

The proof of (ii) is analogous. ∎

Let $\boldsymbol{a}$, $\boldsymbol{b} \in \boldsymbol{R}^n$ and let $a_{[1]}, \ldots, a_{[n]}$ and $b_{[1]}, \ldots, b_{[n]}$ be the coordinates of $\boldsymbol{a}$ and $\boldsymbol{b}$ arranged in a non-increasing order. We say that $\boldsymbol{b}$ majorizes $\boldsymbol{a}$, to be written as $\boldsymbol{a} \leqslant_m \boldsymbol{b}$, if

$$\sum_{i=1}^{k} a_{[i]} \leqslant \sum_{i=1}^{k} b_{[i]} \text{ for } k = 1, \ldots, n-1 \quad \text{and} \quad \sum_{i=1}^{n} a_{[i]} = \sum_{i=1}^{n} b_{[i]}.$$

THEOREM 5.3. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with density function $f_\theta$. If the family $\{f_\theta;\ \theta \in \Theta\}$ has a monotone likelihood ratio with respect to the location parameter $\theta$, and for any $x$ we have*

$$f_\theta(x) = f(x-\theta) = f(-x+\theta),$$

*then for $n = 2, 3, \ldots$*

(i) *for any $\boldsymbol{a}$, $\boldsymbol{b} \in \boldsymbol{R}^n$ such that $a_i$, $b_i \geqslant 0$, $i = 1, \ldots, n$, and $\boldsymbol{a} \leqslant_m \boldsymbol{b}$*

$$\sum_{i=1}^{n} a_i X_i \leqslant_{ar} \sum_{i=1}^{n} b_i X_i;$$

*in particular, $\bar{X}_n \leqslant_{ar} \bar{X}_{n-1}$;*

(ii) $X_{n+2:2n+3} \leqslant_{ar} X_{n+1:2n+1}$.

Proof. Part (i) is a corollary to Theorem 5.1 (iv) and to the following theorem (see [6]): If $X_1, \ldots, X_n$ are i.i.d. random variables with density function $\psi$ symmetric about 0 and such that $-\log\psi(x)$ is convex, then for any $\boldsymbol{\alpha}$, $\boldsymbol{\beta} \in \boldsymbol{R}^n$ such that $\boldsymbol{\alpha} \leqslant_m \boldsymbol{\beta}$ and $\alpha_i$, $\beta_i \geqslant 0$, $i = 1, \ldots, n$, we have

$$\Big|\sum_{i=1}^{n} \alpha_i X_i\Big| \leqslant_{st} \Big|\sum_{i=1}^{n} \beta_i X_i\Big|;$$

in particular,

$$\left(\frac{1}{n}, \ldots, \frac{1}{n}\right) \leqslant_m \left(\frac{1}{n-1}, \ldots, \frac{1}{n-1}, 0\right),$$

and so $\bar{X}_n \leqslant_{ar} \bar{X}_{n-1}$.

To prove (ii), we first check by straightforward calculations that the graphs of the density functions of $X_{n+1:2n+1}$ and $X_{n+2:2n+3}$ are symmetric about $\theta$ and have two intersection points:

$$x_1 = F_\theta^{-1}\left(\tfrac{1}{2} - \sqrt{(N-4)/N}\right), \qquad x_2 = F_\theta^{-1}\left(\tfrac{1}{2} + \sqrt{(N-4)/N}\right),$$

where $N = (4n+6)/(n+1)$. Then

$$|X_{n+2:2n+3} - \theta| \leqslant_{st} |X_{n+1:2n+1} - \theta|,$$

and (ii) follows from Theorem 5.1 (iv). ∎

We shall now give some examples when one can obtain useful comparisons of unbiased estimators (with the same variance) by using the index *AR*, even in the case where the estimators are not stochastically ordered.

These pairs of estimators are such that one of them is inadmissible (as based on a statistic which is not sufficient). It turns out, in the second example, that the inadmissible estimator is better according to the new criterion (for some sample sizes). This shows that the order according to the suggested criterion differs from the order according to the usual risk function.

EXAMPLE 5.1. Let us consider the family of distributions with the density

$$f_\theta(x) = \begin{cases} \exp[-(x-\theta)] & \text{for } x \geqslant \theta, \\ 0 & \text{for } x < \theta. \end{cases} \tag{11}$$

Then for any estimator $\hat{\theta}$ of $\theta$ we have, using Theorem 5.2 (i):

$$AR_\theta(\hat{\theta}) = 1-\exp(-|\hat{\theta}-\theta|).$$

Let $\hat{\theta}_1(X_1, \ldots, X_n) = X_{1:n}-1/n$ and $\hat{\theta}_2(X_1, \ldots, X_m) = \bar{X}_m-1$. Then

$$X_{1:n} \sim G(1, 1/n, \theta) \quad \text{and} \quad \bar{X}_m \sim G(m, 1/m, \theta),$$

where the density function of $G(a, b, c)$ is given by

$$f(x) = \frac{(x-c)^{a-1}\exp[-(x-c)/b]}{b^a\Gamma(a)}, \qquad a > 0, b > 0, x \geqslant c.$$

It follows that $\mathrm{E}(\hat{\theta}_1) = \mathrm{E}(\hat{\theta}_2) = \theta$; $\mathrm{Var}(\hat{\theta}_1) = 1/n^2$, $\mathrm{Var}(\hat{\theta}_2) = 1/m$. Thus, $\mathrm{Var}(\hat{\theta}_1)$ for sample size $n$ is the same as $\mathrm{Var}(\hat{\theta}_2)$ for sample size $n^2$.

The densities of $AR_\theta(\hat{\theta}_1)$ and $AR_\theta(\hat{\theta}_2)$ are:

$$g_1(x) = \begin{cases} (n/e)\{(1-x)^{n-1}+(1-x)^{-n-1}\} & \text{if } 0 < x \leqslant 1-e^{-1/n}, \\ (n/e)(1-x)^{n-1} & \text{if } 1-e^{-1/n} < x < 1 \end{cases}$$

and

$$g_2(x) = \begin{cases} \dfrac{[\ln e(1-x)]^{m-1}(1-x)^{-m-1}+[\ln e/(1-x)]^{m-1}(1-x)^{m-1}}{\Gamma(m)(e/m)^m} \\ \qquad\qquad\qquad\qquad \text{if } 0 < x < 1-1/e, \\ \dfrac{[\ln e/(1-x)]^{m-1}(1-x)^{m-1}}{\Gamma(m)(e/m)^m} \quad \text{if } 1-1/e \leqslant x < 1. \end{cases}$$

As may be seen from Figs. 2*a*, *b* which show the densities and the corresponding cdf's for sample sizes giving the same variances (5 and 25, respectively), the estimators $\hat{\theta}_1$ (solid line) and $\hat{\theta}_2$ (dotted line) are not stochastically ordered. However, one could argue that $\hat{\theta}_1$ is better than $\hat{\theta}_2$ because, below a certain threshold, smaller values of $AR$, hence less differentiation between distributions corresponding to a true and estimated parameter are more likely under $\hat{\theta}_1$ than under $\hat{\theta}_2$.

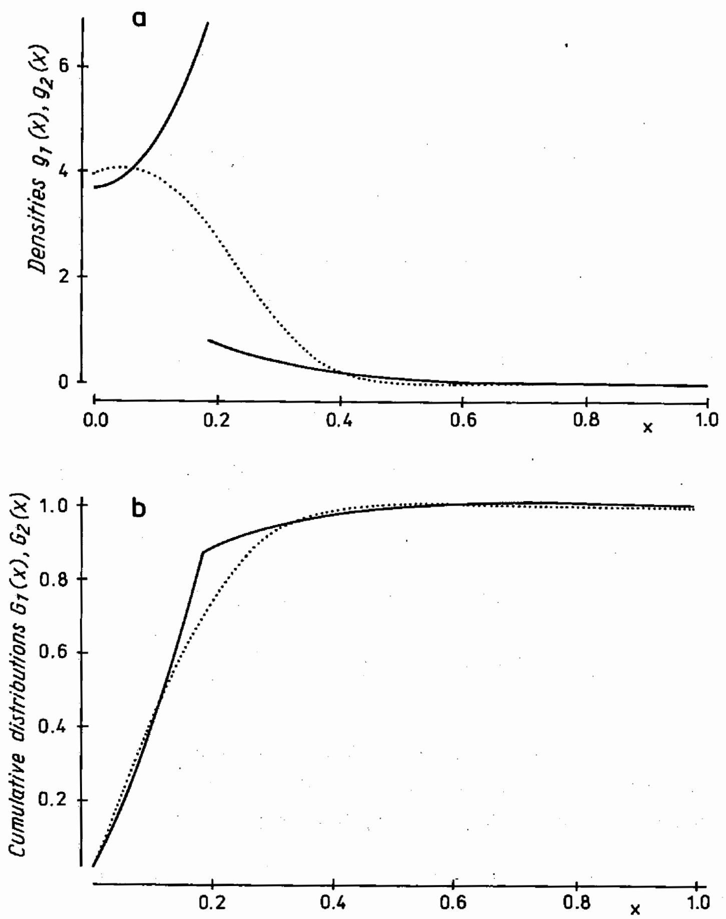



Fig. 2

$a$ — densities of $AR_\theta(\hat{\theta}_1)$ (solid line) and $AR_\theta(\hat{\theta}_2)$ (dotted line) for sample sizes 5 and 25, respectively
$b$ — cumulative distributions of $AR_\theta(\hat{\theta}_1)$ (solid line) and $AR_\theta(\hat{\theta}_2)$ (dotted line) for sample sizes 5 and 25, respectively

Table 1 gives some numerical comparisons of $\hat{\theta}_1$ and $\hat{\theta}_2$ with respect to the relation $\leqslant_R$ defined through $AR$-risks $E_\theta\{AR_\theta(\hat{\theta})\}$. Sample sizes

Table 1. Comparison of means and standard deviations of $AR_\theta(\hat{\theta}_i)$, $i = 1, 2$

| $AR_\theta(\hat{\theta}_1)$ | | | $AR_\theta(\hat{\theta}_2)$ | | |
|---|---|---|---|---|---|
| sample size $n$ | mean | standard deviation | sample size $n^2$ | mean | standard deviation |
| 2 | 0.2775 | 0.1724 | 4 | 0.2948 | 0.1809 |
| 3 | 0.2011 | 0.1383 | 9 | 0.2168 | 0.1424 |
| 4 | 0.1577 | 0.1152 | 16 | 0.1710 | 0.1162 |
| 5 | 0.1299 | 0.0985 | 25 | 0.1410 | 0.0978 |
| 6 | 0.1103 | 0.0861 | 36 | 0.1199 | 0.0843 |
| 7 | 0.0959 | 0.0764 | 49 | 0.1043 | 0.0740 |
| 8 | 0.0849 | 0.0687 | 64 | 0.0923 | 0.0660 |

are adjusted so that variances of estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ are equal in each row.

As may be seen, the minimum of five observations (adjusted to remove bias) is a better estimate (in the sense of average $AR$) than the mean of twenty five observations (adjusted to remove bias).

EXAMPLE 5.2. Let

$$f_\theta(x) = \begin{cases} 1/\theta & \text{for } 0 \leqslant x \leqslant \theta, \\ 0 & \text{otherwise.} \end{cases}$$

Consider two unbiased estimators of $\theta$, namely

$$\hat{\theta}_3 = [(n+1)/n] X_{n:n} \quad \text{and} \quad \hat{\theta}_4 = 2X_{m+1:2m+1},$$

so that $\hat{\theta}_4$ is the double median for sample size $2m+1$. Elementary calculations give here

$$\mathrm{Var}(\hat{\theta}_3) = \theta^2/[n(n+2)] \quad \text{and} \quad \mathrm{Var}(\hat{\theta}_4) = \theta^2/(2m+3),$$

so that variances are equal if

$$m = [n(n+2)-3]/2. \tag{12}$$

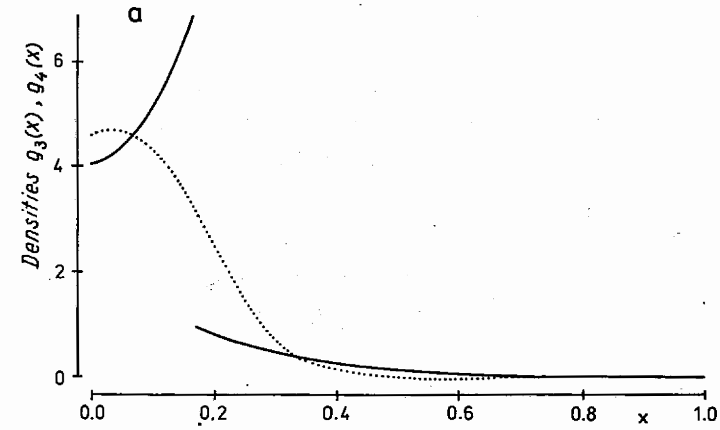
Now, we have $AR_\theta(\hat{\theta}) = 1 - \min(\theta, \hat{\theta})/\max(\theta, \hat{\theta})$, which leads to the following densities of $AR_\theta(\hat{\theta}_3)$ and $AR_\theta(\hat{\theta}_4)$:

$$g_3(x) = \begin{cases} \left(\dfrac{n}{n+1}\right)^n \{n(1-x)^{-(n-1)} + n(1-x)^{n-1}\} & \text{if } 0 < x < \dfrac{1}{n+1}, \\ \left(\dfrac{n}{n+1}\right)^n n(1-x)^{n-1} & \text{if } \dfrac{1}{n+1} \leqslant x \leqslant 1 \end{cases}$$

and

$$g_4(x) = \begin{cases} \dfrac{(2m+1)!}{(m!)^2 2^{2m+1}} \left\{ \dfrac{(1-2x)^m}{(1-x)^{2m+2}} + (1-x^2)^m \right\} & \text{if } 0 < x < \dfrac{1}{2}, \\ \dfrac{(2m+1)!}{(m!)^2 2^{2m+1}} (1-x^2)^m & \text{if } \dfrac{1}{2} \leqslant x < 1. \end{cases}$$

Figs. 3*a*, *b* show the graphs of the above densities and the corresponding cdf's. As may be seen, the estimators $\hat{\theta}_3$ (solid line) and $\hat{\theta}_4$ (dotted line) for $n = 5$ and $m = 33$ adjusted so as to equalize the variances are not stochastically ordered. This time, however, it is hard to decide which estimator has smaller mean of $AR$. Numerical comparison of $\hat{\theta}_3$ and $\hat{\theta}_4$ for various sample sizes is given in the following table.



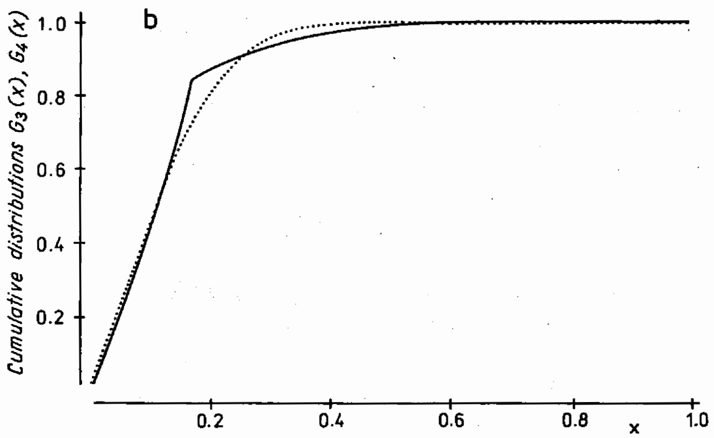




Fig. 3

$a$ — densities of $AR_\theta(\hat{\theta}_3)$ (solid line) and $AR_\theta(\hat{\theta}_4)$ (dotted line) for sample sizes 5 and 33, respectively
$b$ — cumulative distributions of $AR_\theta(\hat{\theta}_3)$ (solid line) and $AR_\theta(\hat{\theta}_4)$ (dotted line) for sample sizes 5 and 33, respectively

Table 2. Comparison of means and standard deviations of $AR_\theta(\hat{\theta}_3)$ and $AR_\theta(\hat{\theta}_4)$

| $AR_\theta(\hat{\theta}_3)$ | | | $AR_\theta(\hat{\theta}_4)$ | | |
|---|---|---|---|---|---|
| sample size $n$ | mean | standard deviation | sample size $n(n+2)-2$ | mean | standard deviation |
| 3 | 0.1917 | 0.1454 | 13 | 0.1853 | 0.1296 |
| 5 | 0.1259 | 0.1020 | 33 | 0.1245 | 0.0890 |
| 7 | 0.0939 | 0.0784 | 61 | 0.0943 | 0.0682 |
| 9 | 0.0749 | 0.0636 | 97 | 0.0760 | 0.0554 |

This time the order of relation $\leqslant_R$ becomes reversed with increase of sample size. The maximum of 5 observations (adjusted for unbiasedness) is inferior to double median of sample of 33. However, adjusted maximum of 7 observations is already better than double sample median from sample of size 61.

To conclude, let us remark that the preceding considerations concerning location parameters are easily transferable to the case of scale parameters. We say that $\beta$ is a *scale parameter* if for any $x$

$$g_\beta(x) = (1/\beta)g(x/\beta), \quad \beta > 0, \tag{13}$$

for some density function $g$; the family $(g_\beta : \beta \in \mathscr{B})$ has a monotone likelihood ratio with respect to $\beta$ if

$$g(x/\beta')/g(x/\beta) \text{ is a non-decreasing function of } x \text{ for } \beta < \beta'. \tag{14}$$

Then for any $\hat{\beta}$ we have

$$AR_\beta(\hat{\beta}) = \left|1 - 2\int_0^1 G\big([\min(\beta, \hat{\beta})/\max(\beta, \hat{\beta})]\, G^{-1}(t)\big)dt\right|,$$

where $G$ is the distribution function corresponding to $g$. If $X$ is continuous and non-negative with $g_\beta$ satisfying (13) and (14), then the family of distributions of $Y = \ln X$ satisfies (6) and (7) for the location parameter $\alpha = \ln\beta$. Let $\hat{\alpha}$ be any estimator of $\alpha \in A = \{\alpha;\, \alpha = \ln\beta;\, \beta \in B\}$ such that $\hat{\alpha}$ is valued in $A$. Let $\hat{\beta}$ be the estimator of $\beta$, defined by

$$\hat{\beta}(X_1, \ldots, X_n) = \exp[\hat{\alpha}(Y_1, \ldots, Y_n)].$$

Then the distributions of $AR_\beta(\hat{\beta}(X_1, \ldots, X_n))$ and of $AR_\alpha(\hat{\alpha}(Y_1, \ldots, Y_n))$ are equal, which means that $\hat{\beta}$ and $\hat{\alpha}$ are in this sense equivalent estimators of $\beta$ and $\alpha$, respectively.

**Acknowledgement.** The authors wish to thank Professor Elżbieta Pleszczyńska for her helpful comments.

## REFERENCES

[1] B. C. Arnold, *Majorization and the Lorenz Order: A Brief Introduction*, Lecture Notes in Statist. 43 (1987).

[2] — and J. A. Villaseñor, *Lorenz ordering of means and medians*, Statist. Probab. Lett. 4 (1986), pp. 47–49.

[3] T. Bromek and M. Niewiadomska-Bugaj, *Threshold rules in two-class discrimination problems*, Probab. Math. Statist. 8 (1987), pp. 11–16.

[4] D. M. Ciffarelli and E. Regazzini, *On a general definition of concentration function*, Sankhyā 49 (1987), pp. 307–319.

[5] T. Kowalczyk and J. Mielniczuk, *Neyman–Pearson curves. Properties and estimation*, ICS PAS Report No 683, Warszawa 1990.

[6] F. Proshan, *Peakedness of distributions of convex combinations*, Ann. Math Statist. 36 (1965), pp. 1703–1706.

Institute of Mathematics
Technical University of Warsaw
pl. Politechniki 1
00-661 Warszawa, Poland

Dept. Statistics and
Computer Science
West Virginia University
Knapp Hall
Morgantown, WV 26505, USA

*Received on 12.12.1991;*
*revised version on 15.7.1992*