*The Annals of Statistics*
1989, Vol. 17, No. 2, 937–940

# STOCHASTIC REDUCTION OF LOSS IN ESTIMATING NORMAL MEANS BY ISOTONIC REGRESSION[1]

BY ROBERT E. KELLY

*Pennsylvania State University*

Consider the problem of estimating the ordered means $\mu_1 \le \mu_2 \le \cdots \le \mu_k$ of independent normal random variables, $Y_1, Y_2, \dots, Y_k$. It is shown that the absolute error of each component $\hat{\mu}_i$ of the isotonic regression estimator is stochastically smaller than that of the usual estimator $Y_i$. Thus $\hat{\mu}_i$ is superior to $Y_i$ under any nonconstant loss which is a nondecreasing function of absolute error.

**1. Introduction.** Let $Y_1, \dots, Y_k$ be independent normal random variables with unknown means satisfying $\mu_1 \le \mu_2 \le \cdots \le \mu_k$ and $\mathrm{Var}(Y_i) = \sigma_i^2 = \sigma^2/w_i$, $w_i > 0$, $i = 1, \dots, k$. The order restricted maximum likelihood estimator $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_k)$ of $\mu = (\mu_1, \dots, \mu_k)$ is the value of $\theta$ which minimizes the weighted sum of squares

$$\sum_{i=1}^{k} (Y_i - \theta_i)^2 w_i$$

subject to the simple order restriction $\theta_1 \le \theta_2 \le \cdots \le \theta_k$. This estimator is called the isotonic regression of $Y = (Y_1, \dots, Y_k)$ on the simple order and is given by

$$\hat{\mu}_i = \min_{s \ge i} \max_{r \le i} \mathrm{Av}(Y; r, s), \qquad i = 1, \dots, k, \tag{1.1}$$

where, for $z = (z_1, \dots, z_k) \in \mathbb{R}^k$,

$$\mathrm{Av}(z; r, s) = \sum_{j=r}^{s} w_j z_j \Big/ \sum_{j=r}^{s} w_j \tag{1.2}$$

[cf. Barlow, Bartholomew, Bremner and Brunk (1972) page 19]. The estimator $\hat{\mu}$ has been compared to the unrestricted estimator $Y$, and several results point to the superiority of the isotonic regression estimator when the order restriction holds. It has been shown [cf. Robertson, Wright and Dykstra (1988)] that the total error of the estimator is reduced through isotonic regression on any order restriction in the sense that

$$\sum_{i=1}^{k} |\hat{\mu}_i - \mu_i|^p w_i \le \sum_{i=1}^{k} |Y_i - \mu_i|^p w_i \quad \text{for all } p \ge 1.$$

Lee (1981) considered the mean squared error of the components of the estimator

Received February 1988; revised September 1988.
[1]Research supported in part by NSF Grant DMS-88-01488.
*AMS* 1980 *subject classifications*. Primary 62F10, 62E99; secondary 62C99, 60E15.
*Key words and phrases*. Isotonic regression, loss function, maximum likelihood, order restriction, stochastic ordering.

and showed that for $k \geq 2$,

$$E(\hat{\mu}_i - \mu_i)^2 < E(Y_i - \mu_i)^2, \qquad i = 1, \dots, k,$$

when $\mu$ satisfies the order restriction. It will be shown here that an even stronger result holds componentwise, namely that $|\hat{\mu}_i - \mu_i|$ is strictly stochastically smaller than $|Y_i - \mu_i|$ for $i = 1, \dots, k$, $k \geq 2$, when $\mu$ satisfies the order restriction. From this it follows that, for loss functions $\mathscr{L}$ of the form $\mathscr{L}(\delta; \mu) = \rho(|\delta - \mu_i|)$ with $\rho(t)$ nonconstant and nondecreasing on $t > 0$, $\mathscr{L}(\hat{\mu}_i; \mu)$ is stochastically smaller than $\mathscr{L}(Y_i; \mu)$ for $i = 1, \dots, k$ and as a result, $E\{\mathscr{L}(\hat{\mu}_i; \mu)\} < E\{\mathscr{L}(\hat{\mu}_i; \mu)\}$ for $i = 1, \dots, k$ when the order restriction is satisfied. The stochastic ordering is strict if $\rho$ is strictly increasing. Here we use the following definitions. The random variable $A$ is stochastically smaller than the random variable $B$ if

$$P(A \leq t) \geq P(B \leq t) \quad \text{for all } t, \tag{1.3}$$

with strict inequality for some $t$. If strict inequality holds in (1.3) for all $t$ such that $P(B \leq t) > 0$, then we say $A$ is strictly stochastically smaller than $B$.

## 2. Stochastic ordering of absolute errors.

THEOREM. *For $k \geq 2$, let $Y_1, \dots, Y_k$ be independent normal random variables with $Y_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \dots, k$, where $\sigma_i^2 = \sigma^2/w_i$ and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$. Let $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_k)$ be the isotonic regression of $Y = (Y_1, \dots, Y_k)$ on the simple order with weights $w_i$, $i = 1, \dots, k$. Then $P(|\hat{\mu}_i - \mu_i| \leq t) > P(|Y_i - \mu_i| \leq t)$ for all $t > 0$, $i = 1, \dots, k$.*

PROOF. Assume without loss of generality that

$$\sum_{j=1}^{k} w_j \mu_j = 0. \tag{2.1}$$

Define $X^{(i)} = (X_1^{(i)}, \dots, X_k^{(i)})$, where $X_j^{(i)} = Y_j - Y_i$, for $j = 1, \dots, k$, $i = 1, \dots, k$. Note that $X^{(i)}$ takes values in

$$\mathscr{X} = \{(x_1, \dots, x_k) \in \mathbb{R}^k : x_i = 0\}.$$

In view of (1.1) and (1.2) we have

$$\hat{\mu}_i - Y_i = \min_{s \geq i} \max_{r \leq i} \operatorname{Av}(X^{(i)}; r, s), \tag{2.2}$$

so that $(\hat{\mu}_i - Y_i)$ depends on $Y$ only through $X^{(i)}$. Clearly, $\operatorname{Av}(Y; 1, k)$ is independent of $X^{(i)}$ and has expectation zero by (2.1). Hence, the conditional distribution of $Y_i$ given $X^{(i)}$ is normal with mean

$$m(X^{(i)}) = E(Y_i | X^{(i)}) = -\operatorname{Av}(X^{(i)}; 1, k) = Y_i - \operatorname{Av}(Y; 1, k) \tag{2.3}$$

and variance

$$s^2 = \operatorname{Var}\{\operatorname{Av}(Y; 1, k)\}.$$

We now proceed by induction on $k$. For $k = 1$, clearly $\hat{\mu}_1 = Y_1$. Assume that for dimension $k - 1$,

$$P(|\hat{\mu}_i - \mu_i| \le t) \ge P(|Y_i - \mu_i| \le t) \quad \text{for all } t > 0,\ i = 1, \ldots, k - 1.$$

We shall show that this implies strict stochastic inequality

$$P(|\hat{\mu}_i - \mu_i| \le t) > P(|Y_i - \mu_i| \le t) \quad \text{for all } t > 0,\ i = 1, \ldots, k,$$

for dimension $k$. Two cases are considered.

CASE 1. $\mu_i \leqq 0$ and $i < k$. Let $\hat{\nu} = (\hat{\nu}_1, \ldots, \hat{\nu}_{k-1})$ be the isotonic regression of $(Y_1, \ldots, Y_{k-1})$ on the $(k - 1)$-dimensional simple order with weights $w_1, \ldots, w_{k-1}$. By the induction hypothesis,

$$P(|\hat{\nu}_i - \mu_i| \le t) \ge P(|Y_i - \mu_i| \le t) \quad \text{for all } t > 0,$$

and it suffices to show that for all $t > 0$,

$$P(|\hat{\mu}_i - \mu_i| \le t,\ \hat{\mu}_i \ne \hat{\nu}_i) > P(|\hat{\nu}_i - \mu_i| \le t,\ \hat{\mu}_i \ne \hat{\nu}_i). \tag{2.4}$$

As Lee (1981) pointed out, $\hat{\mu}_i \ne \hat{\nu}_i$ implies that

$$\hat{\mu}_i = \max_{r \le i} \mathrm{Av}(Y; r, k)$$

and hence that

$$\mathrm{Av}(Y; 1, k) \le \hat{\mu}_i < \hat{\nu}_i. \tag{2.5}$$

Define

$$c_1(X^{(i)}) = Y_i - \hat{\mu}_i - m(X^{(i)}) + \mu_i,$$
$$c_2(X^{(i)}) = Y_i - \hat{\nu}_i - m(X^{(i)}) + \mu_i$$

and note that these do indeed depend on $X^{(i)}$ only. In view of (2.3) and (2.5) and the assumption for Case 1, we have

$$c_2(X^{(i)}) < c_1(X^{(i)}) \le 0 \tag{2.6}$$

on the set where $\hat{\mu}_i \ne \hat{\nu}_i$.

Now

$$\begin{aligned} P\big(|\hat{\mu}_i - \mu_i| \le t | X^{(i)}\big) &= P\big(c_1(X^{(i)}) - t \le Y_i - m(X^{(i)}) \le c_1(X^{(i)}) + t\big) \\ &= \Phi\left[\frac{c_1(X^{(i)}) + t}{s}\right] - \Phi\left[\frac{c_1(X^{(i)}) - t}{s}\right], \end{aligned} \tag{2.7}$$

where $\Phi$ is the standard normal distribution function and $m(X^{(i)})$ and $s^2$ are the conditional mean and variance of $Y_i$. Similarly,

$$P\big(|\hat{\nu}_i - \mu_i| \le t | X^{(i)}\big) = \Phi\left[\frac{c_2(X^{(i)}) + t}{s}\right] - \Phi\left[\frac{c_2(X^{(i)}) - t}{s}\right]. \tag{2.8}$$

For $t > 0$, $\Phi((c + t)/s) - \Phi((c - t)/s)$ is strictly increasing in $c$ for $c \le 0$, and hence (2.6)–(2.8) ensure that on the set where $\hat{\mu}_i \neq \hat{\nu}_i$,

$$P\big(|\hat{\mu}_i - \mu_i| \leq t|X^{(i)}\big) > P\big(|\hat{\nu}_i - \mu_i| \leq t|X^{(i)}\big),$$

which implies (2.4).

CASE 2. $\mu_i \geq 0$ and $i > 1$. Replace $Y_1, \ldots, Y_k$ by $-Y_k, \ldots, -Y_1$ and we are back in Case 1. Note that $\mu_i < 0$ and $i = k$ or $\mu_i > 0$ and $i = 1$ are impossible in view of (2.1). □

**Acknowledgment.** I am grateful to the Editor for his many comments and suggestions which have led to a much clearer and more concise presentation of the results.

## REFERENCES

BARLOW, R. F., BARTHOLOMEW, D. D., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

LEE, C. C. (1981). The quadratic loss of isotonic regression under normality. *Ann. Statist.* **9** 686–688.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
219 POND LABORATORY
UNIVERSITY PARK, PENNSYLVANIA 16802