The Annals of Statistics
1981, Vol. 9, No. 3, 686–688

THE QUADRATIC LOSS OF ISOTONIC REGRESSION UNDER NORMALITY

BY CHU-IN CHARLES LEE

Memorial University of Newfoundland

The maximum likelihood estimator $\hat{\mu}$ of a nondecreasing regression function has been studied in detail in the literature. However, little is known about its quadratic loss pointwise. This paper shows that the mean square error of $\hat{\mu}_i$ is less than that of the usual estimator $\bar{X}_i$ for each i when $\bar{X}_1, \cdots, \bar{X}_k$ are independent normal variates.

1. Introduction. Let $X_1, \cdots, X_k$ be independent normal variates with unknown means μ_i satisfying $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ and with variances $\text{Var}(X_i) = \sigma^2/w_i$ where w_i are given positive weights. The isotonic regression $(\hat{\mu}_1, \cdots, \hat{\mu}_k)$ of unknown parameters $(\mu_1, \cdots, \mu_k)$ is defined as the optimal solution to the least squares problem

$$\min_{(Y_1, \cdots, Y_k)} \textstyle\sum_{i=1}^k (X_i - Y_i)^2 w_i$$

subject to the condition $Y_1 \leq Y_2 \leq \cdots \leq Y_k$. This optimal solution $(\hat{\mu}_1, \cdots, \hat{\mu}_k)$ which is also the maximum likelihood estimator of $(\mu_1, \cdots, \mu_k)$ can be easily manipulated by the Pool-Adjacent-Violators algorithm proposed by Ayer et al. (1955). Most of the applications of statistical inference under order restrictions which appeared in the literature prior to 1972 can be found in Barlow et al. (1972).

Brunk (1965) showed that

$$\textstyle\sum_{i=1}^k (X_i - \mu_i)^2 w_i \geq \sum_{i=1}^k (X_i - \hat{\mu}_i)^2 w_i + \sum_{i=1}^k (\hat{\mu}_i - \mu_i)^2 w_i.$$

Thus the total mean square error of the maximum likelihood estimator, $\sum_{i=1}^k E(\hat{\mu}_i - \mu_i)^2 w_i$, is strictly less than that of the usual estimator, $\sum_{i=1}^k E(X_i - \mu_i)^2 w_i$. The aim of this paper is to show that the inequality

$$E(X_i - \mu_i)^2 > E(\hat{\mu}_i - \mu_i)^2 \tag{1.1}$$

holds pointwise. Consequently the inequality also holds if $\mu_i > \mu_{i+1} > \mu_i - c_i$ for some i. For instance, if $k = 2$ and $w_1 = w_2 = 1$, then c_1 can be as large as $1.118\ \sigma$.

2. The inequality. We shall verify the inequality (1.1) by mathematical induction. But first let us consider the isotonic regression in the absence of X_k and that in the presence of X_k. We shall let $(\hat{\nu}_1, \cdots, \hat{\nu}_{k-1})$ and $(\hat{\mu}_1, \cdots, \hat{\mu}_k)$ denote the isotonic regressions based upon $(X_1, \cdots, X_{k-1})$ and $(X_1, \cdots, X_k)$, respectively. By the min-max formula, i.e., Equation (1.11), page 19 of Barlow et al. (1972), we have that for each $i < k$,

$$\hat{\mu}_i = \min_{i \leq t} \max_{s \leq i} Av(s, t) \leq \min_{i \leq t < k} \max_{s \leq i} Av(s, t) = \hat{\nu}_i \tag{2.1}$$

where

$$Av(s, t) = \textstyle\sum_{j=s}^t X_j w_j / \sum_{j=s}^t w_j.$$

It follows that if $\hat{\nu}_i > \hat{\mu}_i$, then

Received November, 1978; revised April, 1979.
AMS 1970 *subject classifications*. Primary 62F10; secondary 62A10.
Key words and phrases. Isotonic regression, maximum likelihood estimator, mean square error.

(2.2) $$\hat{\mu}_i = \max_{s \le i} Av(s, k) \ge Av(1, k).$$

By the independence and the normality of the variates $X_1, \cdots, X_k$, $Av(m, n) - Av(s, t)$ and $Av(1, k)$ are stochastically independent. The event $[\hat{\mu}_i = Av(s, t)]$ can be represented as the intersection of $\cap_{j=1}^{s-1}[Av(j, t) \le Av(s, t)]$ and $\cap_{j=s}^{k}[Av(s, j) \ge Av(s, t)]$ by the min-max formula and by the max-min formula. Therefore the indicator $1_{[\hat{\mu}_i = Av(s,t)]}$ and $Av(1, k)$ are stochastically independent; so are $1_{[\hat{\nu}_i = Av(m,n)]}$ and $Av(1, k)$. It follows that for $i < k$

(2.3)
$$\begin{aligned} E[(\hat{\nu}_i - \hat{\mu}_i)Av(1, k)] &= \sum E\{[Av(m, n) - Av(s, t)]Av(1, k)1_{[\hat{\nu}_i = Av(m,n)]}1_{[\hat{\mu}_i = Av(s, t)]}\} \\ &= \sum E\{[Av(m, n) - Av(s, t)]1_{[\hat{\nu}_i = Av(m,n)]}1_{[\hat{\mu}_i = Av(s, t)]}\}E[Av(1, k)] \\ &= E(\hat{\nu}_i - \hat{\mu}_i)E[Av(1, k)]. \end{aligned}$$

THEOREM. *Let $X_1, \cdots, X_k$ be independent normal variates with unknown means μ_i satisfying $\mu_1 \le \mu_2 \le \cdots \le \mu_k$ and with variances* $\mathrm{Var}(X_i) = \sigma^2/w_i$ *where w_i are given positive weights and $k \ge 2$. Then for each i we have*

$$E(X_i - \mu_i)^2 > E(\hat{\mu}_i - \mu_i)^2,$$

where $(\hat{\mu}_i, \cdots, \hat{\mu}_k)$ is the maximum likelihood estimator of $(\mu_1, \cdots, \mu_k)$.

PROOF. Assume that the result holds for $k - 1$ so that

(A) $E(X_i - \mu_i)^2 \ge E(\hat{\nu}_i - \mu_i)^2$ where $\hat{\nu}_i$ is the isotonic estimator of μ_i based upon $(X_1, \cdots, X_{k-1})$, $i = 1, 2, \cdots, k - 1$, and

(B) $E(X_i - \mu_i)^2 \ge E(\hat{\nu}_i - \mu_i)^2$ where $\hat{\nu}_i$ is the isotonic estimator of μ_i based upon $(X_2, \cdots, X_k)$, $i = 2, 3, \cdots, k$.

CASE 1. Let $\mu_i \le E[Av(1, k)]$, $i < k$. From Condition (A), we have that

$$\begin{aligned} E(X_i - \mu_i)^2 &\ge E(\hat{\nu}_i - \mu_i)^2 \\ &= E(\hat{\mu}_i - \mu_i)^2 + E(\hat{\nu}_i - \hat{\mu}_i)^2 + 2E(\hat{\nu}_i - \hat{\mu}_i)(\hat{\mu}_i - \mu_i). \end{aligned}$$

From (2.1) and (2.2),

$$E(\hat{\nu}_i - \hat{\mu}_i)[\hat{\mu}_i - Av(1, k)] = E(\hat{\nu}_i - \hat{\mu}_i)[\hat{\mu}_i - Av(1, k)]1_{[\hat{\nu}_i > \hat{\mu}_i]} \ge 0.$$

From (2.3) and (2.1),

$$E(\hat{\nu}_i - \hat{\mu}_i)[Av(1, k) - \mu_i] = E(\hat{\nu}_i - \hat{\mu}_i)E[Av(1, k) - \mu_i] \ge 0.$$

Therefore, $E(X_i - \mu_i)^2 \ge E(\hat{\nu}_i - \mu_i)^2 > E(\hat{\mu}_i - \mu_i)^2$. When $k = 2$, we have identities in Conditions (A) and (B).

CASE 2. Let $\mu_i \ge E[Av(1, k)]$, $i > 1$. Use Condition (B) and results analogous to (2.1), (2.2) and (2.3) but with the inequalities reversed.

One consequence of the proof of the theorem is that one can show

$$\begin{aligned} &E(X_i - \mu_i)^2 - E(\hat{\mu}_i - \mu_i)^2 \\ &\qquad > \min\{E[X_i - Av(i - 1, i)]^2 1_{[X_{i-1} \ge X_i]}, E[X_i - Av(i, i + 1)]^2 1_{[X_i \ge X_{i+1}]}\} \end{aligned}$$

for $1 < i < k$. See Wright (1978) for numerical examples of pointwise mean square errors $E(\hat{\mu}_i - \mu_i)^2$.

Acknowledgment. The author wishes to thank the referee and an associate editor for their helpful comments. Thanks are also due to Professor Norman F. Rehner for his valuable suggestions.

REFERENCES

AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

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

BRUNK, H. D. (1965). Conditional expectation given a σ-lattice and applications. *Ann. Math. Statist.* **36** 1339–1350.

WRIGHT, F. T. (1978). Estimating strictly increasing regression functions. *J. Amer. Statist. Assoc.* **73** 636–639.

DEPARTMENT OF MATHEMATICS AND STATISTICS,
MEMORIAL UNIVERSITY OF NEWFOUNDLAND,
ST. JOHN'S, NEWFOUNDLAND,
CANADA. A1B 3X7