and

$$a_0^{2N} \mid A_N \mid = a_0^4[(a_0^2 - a_2^2)^2 - (a_0 - a_2)^2 a_1^2]^{-1}.$$

It may be mentioned here that Ulf Gernander and Murray Rosenblatt ([2], pp. 238–239) have considered asymptotic properties of A_N^{-1} as N tends to infinity. They, however, do not attempt to determine the k^2 elements standing in the first k rows and the first k columns of A_N^{-1}, although they suggest a method of orthogonalization of the vector X_N.

REFERENCES

[1] J. WISE, "The autocorrelation function and the spectral density function," *Biometrika*, Vol. 42 (1955), pp. 151–159.
[2] ULF GERNANDER AND MURRAY ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley and Sons, New York, 1957.

A PROBLEM OF BERKSON, AND MINIMUM VARIANCE ORDERLY ESTIMATORS[1]

BY JOHN W. TUKEY

Princeton University

1. Summary. The distinction between efficiency in the asymptotic sense originally introduced by Fisher ([2], 1925, p. 703), and the finite sample sense sometimes used by others has been recently stressed by various writers (e.g., Berkson [1]). The technique of proof used below was originally developed to provide a simple example where the maximum likelihood estimate of location, though asymptotically efficient, was not of minimum variance for any finite sample size whatever. The (symmetrical) double exponental distribution with known scale, where the sample median is the maximum likelihood estimator of location, could easily be shown to be such an example. (While this result is useful in deflating unwarranted views about minimum variance properties of maximum likelihood estimates, Fisher's ([2], p. 716) results about intrinsic accuracy in the same situation are of more basic interest.)

On examination, however, the technique used to provide this rather isolated and special result was found capable of showing, for a class of distributions with suitable monotony properties (in particular all distributions for which $f'(y)/f(y)$ is monotone decreasing, and all normal, exponential, gamma and beta distributions), that the covariances of the order statistics in a sample of any chosen size are monotone in either index separately.

Received March 22, 1957.
[1] Prepared in connection with research sponsored by the Office of Naval Research. Based, in part, on Memorandum Report 11 of the Statistical Research Group, Princeton University, which was issued 25 October 1948.

2. Complete regression. We shall say the distribution of z given y *shows complete negative regression on* y if the cumulative distribution $F(z \mid y)$ satisfies

$$F(z \mid y'') \leqq F(z \mid y') \qquad \text{for } y'' \leqq y',$$

provided the equality does not always hold. We define complete positive regression analogously. We notice that:

(A_1) *If the distribution of* z *given* y *shows complete negative regression* y, *and* z_1 *is an order statistic from a random sample of* z*'s, then the distribution of* z_1 *given* y *shows complete negative regression on* y.

(A_2) *If the distribution of* w *given* z *shows complete negative regression on* z, *and the distribution of* z *given* y *shows complete positive regression on* x, *and the distribution of* w *given* z *is unaffected by giving* y, *then the distribution of* w *given* y *shows complete negative regression on* y.

(A_3) *If the distribution of* z *given* y *shows complete negative regression on* y *then* $\operatorname{cov}\{z, y\} < 0$.

The first result follows from the beta-function formula for an order statistic cumulative,

$$G_{k \mid n}(w) = k \binom{n}{k} \int_0^{F(w)} t^{k-1}(1 - t)^{n-k}\, dt$$

which follows from the interpretation of $G_{k \mid n}(w)$ as the probability of k or more out of n falling in an interval of probability $F(w)$, and which shows that the cumulative of $z_{k \mid n}$ is a monotone function of the cumulative of z. The second follows easily by introducing the monotone *representing function* [3] $z_y(u)$ corresponding to $F(z \mid y)$ such that if u is uniformly distributed on [0, 1], then $z_y(u)$ is distributed according to $F(z \mid y)$. The hypothesis of complete positive regression is equivalent to $z_{y'}(u) \geqq z_{y''}(u)$ for $y' \geqq y''$, and we have

$$H(w \mid y') = \int G(w \mid z_{y'}(u))\, du \leqq \int G(w z_{y''}(u))\, du = H(w \mid y'') \qquad \text{for } y' \leqq y''$$

which we desired to show. The third result follows from the fact that

$$\operatorname{ave}\{z \mid y'\} \geqq \operatorname{ave}\{z \mid y''\}$$

which follows directly from the inequality for the representing functions.

3. Subexponential distributions. We shall say that a cumulative distribution is *subexponential to the right* if

$$\frac{F(z + h) - F(h)}{1 - F(h)} = 1 - \frac{1 - F(z + h)}{1 - F(h)}$$

is monotonically decreasing for fixed $z > 0$ as h increases. We notice that this is equivalent to stating that, *referred to the point of truncation*, the distribution of z after truncation on the left shows complete negative regression on the point of truncation. We define subexponential on the left, or in both directions, analogously. We are now prepared to demonstrate:

(B_1) *If* $F(q) = F(y - \theta)$ *is subexponential on the right, and* $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ *are an ordered sample of* y's *(and* $q_1 \leqq q_2 \leqq \cdots \leqq q_n$ *are an ordered sample of* q's*), then for any* $j < k$ *we have*

$$\text{cov}\,\{y_k - y_j\,,\, y_j\} = \text{cov}\,\{q_k - q_j\,,\, q_j\} < 0.$$

(The analogs for "on the left" or "in both directions" clearly follow by symmetry.) The proof rests on Wald's principle ([4], p. 536) according to which the distribution of q_k given q_j is that of the $(k - j - 1)$st order statistic from a sample of $n - j$ from the result of truncating the original distribution at q_j. The distribution of $q_k - q_j$ for q_j fixed is that of a similar order statistic from the truncated distribution referred to its point of truncation as origin—and as remarked above this latter distribution shows complete negative regression on q_j. By (A_1) the same is true of any distribution of an order statistic, and hence for the distribution of $q_k - q_j$. The negativity of the covariance follows from (A_3).

This result (and its analogs) can easily be extended to

(B_2) *Under the hypotheses of* (B_1), *if* $h \leqq j < k$, *then*

$$\text{cov}\,\{y_k - y_j\,,\, y_h\} = \text{cov}\,\{q_k - q_j\,,\, q_h\} < 0.$$

For, since the distribution of q_k given q_j is not affected by giving q_h, and the distribution of q_j given q_h shows complete positive regression on q_h, we may apply (A_2) to complete the proof.

As a corollary we have the curiously simple results:

(B_3) *If the distribution of* q *is subexponential in both directions, then the covariance of any two order statistics is less than the variance of either.*

(B_4) *If the distribution of* q *is subexponential in both directions, the covariance between order statistics* q_j, q_k *is monotone in* j *and* k *separately, decreasing as* j *and* k *separate from one another.*

The interest of these results is enhanced when we observe normal, exponential, gamma and beta distributions, pristine or truncated, are all subexponential in both directions.

4. Monotone location-scores. By definition, a distribution $F(q)$ is subexponential to the right if

$$G(y \mid h) = \frac{F(y + h) - F(h)}{1 - F(h)} = 1 - \frac{1 - F(y + h)}{1 - F(h)}$$

is monotone decreasing as h increases for every y. This is equivalent to

$$\log\,\{1 - F(y + h)\} - \log\,\{1 - F(h)\}$$

being monotone increasing, or, granting differentiability, to

$$\frac{f(h)}{1 - F(h)} - \frac{f(y + h)}{1 - F(y + h)} > 0,$$

where $y > 0$, and hence to

$$-\frac{f(u)}{1 - F(u)} = \frac{d}{du}\log\,(1 - F(u)) = \frac{\int_u^\infty f'(u)\,du}{\int_u^\infty f(u)\,du}$$

being monotone decreasing. This will follow from the monotone decreasing character of $(\log f(u))' = f'(u)/f(u)$ since $f(u) \geqq 0$. It is thus sufficient, but not necessary, for subexponentiality on the right that $f'(u)/f(u)$, which is the location score associated with the specification consisting of all translations of $F(u)$, be monotone decreasing.

The class of distributions with monotone scores for location is immediately seen to be closed under formal multiplication of densities, so that if

(i) $F(u)$, $G(u)$ have monotone scores for location,

(ii) $F(u) = \int f(u)\,du$, $G(u) = \int g(u)\,du$,

(iii) $H(u) = (\text{constant})\,f(u)g(u)\,du$,

then $H(u)$ also has a monotone score for location. The class is also closed under truncation at one or both ends. It is immediately seen to include all distributions whose shapes are single exponential, double exponential (balanced or not), normal, (incomplete) gamma, (incomplete) beta, and their formal products and truncations. (It does *not* include distributions of Cauchy shape.)

Since the large-sample optimum weight to be assigned to an order statistic is the negative of the derivative of the score for location at the typical point of distribution, it seems both peculiarly appropriate and highly reasonable that the minimum variance orderly estimator of location will actually have all its coefficients positive for any distribution with monotone score for location.

5. Orderly estimates. We now turn to a location specification $F(y \mid \theta) = F(y - \theta)$ and to orderly estimates of θ, by which we mean linear combinations of order statistics of total weight 1, i.e.,

$$\tilde{y} = \sum w_i y_i + c, \qquad \sum w_i = 1.$$

(Notice that the variance and bias of $\tilde{y}$ as an estimator of θ are exactly the variance and average value of $\tilde{q}$, where $\tilde{q} = \sum w_i q_i$ are order statistics in a sample from $F(q)$.) We begin with a general result, applicable to *any* convex (i.e., closed under $\alpha t' + (1 - \alpha)t''$ for $0 < \alpha < 1$) class of estimates of u which contains all order statistics (and thus surely contains all orderly estimates).

(C) *If t is the minimum variance estimate in any convex class containing the order statistics, and y_k is any order statistic (or any linear combination of order statistics of total weight one)*

$$\text{cov}\,\{y_k - t, t\} = 0.$$

This follows easily by considering the variance of $t + \lambda(y_k - t)$, where, in particular, if cov $\{y_k - t, t\} < 0$, a value of $\lambda > 0$ will provide lesser variance than for $\lambda = 0$.

From the result it is easy to show that:

(D) *If $F(q)$ is subexponential to the right, then no single order statistic, except possibly the righthandmost, is of minimum variance among orderly estimates of location.*

(Again, the analogs with "to the left ... the lefthandmost" or "in both directions ... statistic," follow by symmetry.) For if y_j were of minimum variance, and y_n the righthandmost, then by

$$(\text{B}_1) \text{ cov } (y_n - y_j, y_j) = \text{cov } (q_n - q_j, q_j) < 0,$$

and by (C) y_j is not of minimum variance. It is reasonable to anticipate that, actually, all coefficients must be positive (particularly for distributions with monotone scores), but the elementary methods used here do not seem to show this easily.

REFERENCES

[1] J. BERKSON, "Question 21," *Amer. Statistician*, Vol. 3 (Oct., 1949), p. 12.

[2] R. A. FISHER, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, Vol. 22 (1923–25), pp. 700–725.

[3] C. HASTINGS, F. MOSTELLER, J. W. TUKEY AND C. P. WINSOR, "Low moments for small samples: a comparative study of order statistics," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 413–426.

[4] J. W. TUKEY, "Non-parametric estimation, II. Statistical equivalent blocks and tolerance regions ... the continuous case," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 529–539.

AN ELEMENTARY THEOREM CONCERNING STATIONARY ERGODIC PROCESSES

BY LEO BREIMAN

University of California, Berkeley

1. Introduction. The purpose of this note is to state and prove a theorem concerning strictly stationary, ergodic processes and to give some of its applications. Although the theorem itself is a simple consequence of the ergodic theorem, its applications include a proof of the consistency of the maximum likelihood estimates for stationary distributions and an extension of the zero-one law for symmetric sets given by Hewitt and Savage [1].

THEOREM 1. *Let $\cdots x_{-1}, x_0, x_1, \cdots$ be a strictly stationary process such that every set invariant under shifts has measure zero or one. Let $\{\phi_n\}$ be a sequence of real-valued functions, ϕ_n being a measurable function of $n + 1$ variables. Then if the sequence $\phi_n(x_0, \cdots, x_n)$ and the sequence $\phi_n(x_{-n}, \cdots, x_0)$ both converge in probability, their limits are almost surely constant and equal.*

Received June 25, 1957; revised October 25, 1957.