$X_{N-i+1} - X_i$ by $w_{(i)}$ and $w_{(1)} = w$. The unbiased estimate of the type $s' = k' \left(\sum w_{(i)}\right)$, where the summation is over the subset of all $w_{(i)}$ which gives minimum variance, is indicated in Table II. The column headed "Eff." refers to the comparison with the unbiased sample standard deviation. The final column gives the ratio of the variance of the best linear systematic statistic as given in [2] to the variance of s'. By examining this ratio we can see that the loss in efficiency due to the use of "zero or one" weights for each range rather than the optimum weights given in [2], is not great.

REFERENCES

[1] A. K. Gupta, "Estimation of the mean and standard deviation of a normal population from a censored sample," *Biometrika*, Vol. 39 (1952), pp. 260–273.
[2] A. E. Sarhan and B. G. Greenberg, "Estimation of location and scale parameters by order statistics from singly and doubly censored samples," Part 1, *Ann. Math. Stat.*, Vol. 27 (1956), pp. 427–451.
[3] Dan Teichroew, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.

THE INDIVIDUAL ERGODIC THEOREM OF INFORMATION THEORY[1]

By Leo Breiman

University of California, Berkeley

1. Introduction. Information theory is largely concerned with stationary stochastic processes $\cdots x_{-1}, x_0, x_1, \cdots$ taking values in a finite "alphabet," $a_1, \cdots, a_s$. In addition, it is usually assumed that the processes are ergodic, that is to say, the shift operator T, defined on the sequence space Ω of the process by shifting each coordinate of a sequence once to the right, is metrically transitive with respect to the probability measure p on Ω.

A question of importance in information theory regarding these processes is the nature and existence, in some sense, of the expression

$$\text{(a)} \qquad \lim_n \left(-\frac{1}{n} \log_2 p(x_0, \cdots, x_{n-1}) \right).$$

In 1948 Shannon [1] showed that for stationary, ergodic Markov chains (a) exists as a limit in probability and is equal to a constant. This limiting constant was termed by Shannon the "entropy" of the process. In 1953 McMillan [2] lifted the restriction to Markov chains and proved that if the process is merely stationary and ergodic, then (a) exists as a limit in L_1 mean and is constant. The purpose of this note is to prove that under the same conditions the limit (a) exists almost surely (a.s.).

Received October 15, 1956.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

2. The modified Birkhoff theorem. The heart of the matter is the following modification of the individual ergodic theorem.

THEOREM 1. *Let T be a metrically transitive $1 - 1$ measure preserving transformation of the probability space $(\Omega, \mathcal{B}, p)$ onto itself. Let $g_0(\omega), g_1(\omega), \cdots$ be a sequence of measurable functions on Ω converging* a.s. *to the function $g(\omega)$ such that $E(\sup_k | g_k |) < \infty$. Then*

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} g_k(T^k\omega) = Eg \text{ a.s.}$$

Proof. We write

$$\frac{1}{n} \sum_{k=0}^{n-1} g_k(T^k\omega) = \frac{1}{n} \sum_{k=0}^{n-1} g(T^k\omega) + \frac{1}{n} \sum_{k=0}^{n-1} [g_k(T^k\omega) - g(T^k\omega)].$$

The conditions of the theorem imply that $E|g| < \infty$ and by Birkhoff's ergodic theorem (see, for example, [3], pp. 464–469), the first term on the right above converges a.s. to Eg. It remains to show that the second term converges a.s. to zero. Let $G_N(\omega) = \sup_{k \geqq N} |g_k(\omega) - g(\omega)|$, then for every fixed N

$$\overline{\lim} \left| \frac{1}{n} \sum_{k=0}^{n-1} [g_k(T^k\omega) - g(T^k\omega)] \right| \leqq \overline{\lim} \frac{1}{n} \sum_{k=0}^{n-1} | g_k(T^k\omega) - g(T^k\omega) |$$

$$\leqq \overline{\lim} \frac{1}{n} \sum_{k=0}^{n-1} G_N(T^k\omega) = EG_N \quad \text{a.s.}$$

The sequence $\{G_N\}$ converges monotonically to zero and

$$EG_0 \leqq E(\sup_k |g_k| + |g|) < \infty,$$

so by the monotone convergence theorem $EG_N \to 0$, which proves the theorem.

THEOREM 2. *Let $\cdots, x_{-1}, x_0, x_1, \cdots$ be a stationary ergodic process ranging over a finite number of values $a_1, \cdots, a_s$. Then there is a constant H such that*

$$\lim_n \left(-\frac{1}{n} \log_2 p(x_0, \cdots, x_{n-1}) \right) = H \quad \text{a.s.}$$

Proof. Let

$$g_0(\omega) = -\log_2 p(x_0),$$

$$g_k(\omega) = -\log_2 \frac{p(x_{-k}, x_{-k+1}, \cdots, x_0)}{p(x_{-k}, x_{-k+1}, \cdots, x_1)}, \qquad k \geqq 1.$$

Then, letting T be the shift operator,

$$-\frac{1}{n} \log_2 p(x_0, \cdots, x_{n-1}) = \frac{1}{n} \sum_{k=0}^{n-1} g_k(T^k\omega).$$

Since T is $1 - 1$, measure preserving and metrically transitive, we apply Theorem 1 and our work will be done as soon as we show that the sequence $\{g_k\}$ converges a.s. and that $E(\sup_k g_k) < \infty$.

To do this we use the inequality established by McMillan [2],

$$\text{(i)} \qquad \int_{\{m \leqq g_k < m+1\}} g_k \leqq s(m+1)2^{-m}.$$

We confine our attention to the cylinder set $Z_i \subset \Omega$, $Z_i = \{\omega;\, x_0 = a_i\}$. On Z_i we have

$$g_k(\omega) = -\log_2 p(x_0 = a_i \mid x_{-1}, \cdots, x_{-k}).$$

Since $p(x_0 = a_i \mid x_{-1}, \cdots, x_{-k})$ is a martingale, it follows from the convexity of $-\log$ and inequality (i) that the sequence $\{g_k\}$ is a semi-martingale (see [3], p. 295). Therefore, g_k converges a.s. on Z_i and hence on Ω.

Furthermore, by a semi-martingale inequality, [3] p. 317, we have, on Z_i,

$$\int_{Z_i} \left(\sup_{0 \leqq k \leqq n} g_k\right) \leqq \frac{e}{e-1} + \frac{e}{e-1} \int_{Z_i} (g_n \log^+ g_n).$$

By using inequality (i) again, we bound the last term on the above right;

$$\int_{Z_i} (g_n \log^+ g_n) = \sum_{m=0}^{\infty} \int_{Z_i\{m \leqq g_n < m+1\}} (g_n \log^+ g_n)$$
$$\leqq \sum_{m=0}^{\infty} s(m+1) \log (m+1) 2^{-m}.$$

Therefore $\int_{Z_i} (\sup_k g_k) < \infty$, by addition $E(\sup_k g_k) < \infty$, and the theorem is proved.

It is a pleasure to acknowledge our debt to Professor David Blackwell who suggested to us the problem treated herein.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, Vol. 27 (1948), pp. 379–423, pp. 623–656.

[2] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 196–219.

3] J. L. Doob, *Stochastic Processes*, John Wiley & Sons, Inc., New York, 1953.

A COUNTEREXAMPLE TO A THEOREM OF KOLMOGOROV[1,2]

By Leo Breiman

University of California, Berkeley

1. Introduction. In 1928 Kolmogorov [1] presented the now well-known degenerate convergence theorem (weak law of large numbers) as follows (see, for

Received August 10, 1956; revised January 18, 1957.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

[2] After this note was submitted the author was informed that C. Derman had constructed a similar counterexample. While the note was in proof, a similar counterexample appeared in a paper by Hartley Rogers, Jr., *Proc. Am. Math. Soc.*, Vol. 8 (1957), pp. 518–520.