A LAW OF LARGE NUMBERS FOR THE MAXIMUM IN A STATIONARY GAUSSIAN SEQUENCE[1]

By Simeon M. Berman

Columbia University

1. Introduction. Let X_1, X_2, $\cdots$ be sequence of random variables which are unbounded above, and let

$$Z_n = \max (X_1, \cdots, X_n).$$

The law of large numbers (LLN) is said to hold for the sequence $\{Z_n\}$ if there exists a sequence of constants $\{A_n\}$ such that

$$(1) \qquad Z_n - A_n \to 0 \qquad \text{in probability.}$$

The necessary and sufficient conditions for the LLN for Z_n in the case where $\{X_n\}$ is a sequence of mutually independent random variables with a common d.f. $F(x)$ were found by B. V. Gnedenko [2]. In particular, he mentioned that the standard normal distribution satisfies the conditions and that (1) holds with

$$(2) \qquad A_n = (2 \log n)^{\frac{1}{2}}.$$

The main result of this paper is that if $\{X_n : n \geqq 1\}$ is a stationary Gaussian process with

$$EX_i = 0, \qquad EX_i^2 = 1, \qquad EX_1X_i = r_i,$$

then Z_n satisfies (1) with A_n given by (2), under the condition that $nr_n \to 0$.

Lemma 1 furnishes a condition for a stationary process under which the maximum behaves (in probability) almost as if the underlying random variables were mutually independent. Lemma 2 generalizes a result of G. S. Watson [3] on the tail of a bivariate normal d.f. The results of Lemma 2 are used to show that the given stationary Gaussian process satisfies the conditions of Lemma 1.

2. Gnedenko's conditions. It has been shown by Gnedenko [2] that (1) holds if and only if for every $\epsilon > 0$,

$$(3) \qquad \begin{aligned} \lim_{n\to\infty} n(1 - F(A_n + \epsilon)) &= 0 \\ \lim_{n\to\infty} n(1 - F(A_n - \epsilon)) &= \infty. \end{aligned}$$

This can be seen from the fact that (1) holds if and only if for every $\epsilon > 0$,

$$(4) \qquad \begin{aligned} 1 &= \lim_{n\to\infty} P\{A_n - \epsilon < Z_n \leqq A_n + \epsilon\} \\ &= \lim_{n\to\infty} F^n(A_n + \epsilon) - \lim_{n\to\infty} F^n(A_n - \epsilon), \end{aligned}$$

Received April 19, 1961; revised August 21, 1961.

[1] Work done under a grant from the National Science Foundation. This is a section of the author's Ph.D. dissertation written at Columbia University.

and the equivalence of (3) and (4) follows from the logarithmic expansion of the terms in the last part of (4).

3. Preliminaries.

LEMMA 1. *Let* $\{X_n : n \geqq 1\}$ *be a stationary sequence, with the marginal* d.f. $F(x)$, *which satisfies* (3) *for some sequence* $\{A_n\}$ *and for every* $\epsilon > 0$; *let* $Z_n = \max(X_1, \cdots, X_n)$. *If for every* $\epsilon > 0$,

$$(5) \qquad \lim_{n\to\infty} \frac{2}{n^2} \sum_{j=2}^{n} (n - j + 1) \frac{P\{X_1 > A_n - \epsilon, X_j > A_n - \epsilon\}}{P^2\{X_1 > A_n - \epsilon\}} = 1,$$

then (1) *holds.*

PROOF. From the relations

$$P\{Z_n > A_n + \epsilon\} = P\left(\bigcup_{i=1}^{n} \{X_i > A_n + \epsilon\}\right)$$
$$\leqq \sum_{i=1}^{n} P\{X_i > A_n + \epsilon\} = n(1 - F(A_n + \epsilon))$$

and from (3), it follows that

$$P\{Z_n > A_n + \epsilon\} \to 0.$$

The proof of the lemma will be completed by showing that

$$(6) \qquad P\{Z_n \leqq A_n - \epsilon\} \to 0.$$

Let $I(H)$ denote the indicator function of the event H. It follows from (3), (5), and the stationarity of the sequence $\{X_n\}$ that

$$\frac{E\left(\sum_{i=1}^{n} I[X_i > A_n - \epsilon]\right)^2}{n^2(1 - F(A_n - \epsilon))^2} = \frac{1}{n(1 - F(A_n - \epsilon))}$$
$$+ \frac{\sum_{i\neq j} P\{X_i > A_n - \epsilon, X_j > A_n - \epsilon\}}{n^2(1 - F(A_n - \epsilon))^2} = \frac{1}{n(1 - F(A_n - \epsilon))}$$
$$+ \frac{2}{n^2} \sum_{j=2}^{n} (n - j + 1) \frac{P\{X_1 > A_n - \epsilon, X_j > A_n - \epsilon\}}{P^2\{X_1 > A_n - \epsilon\}} \to 1;$$

hence,

$$\text{l.i.m.} \frac{\sum_{i=1}^{n} I[X_i > A_n - \epsilon]}{n(1 - F(A_n - \epsilon))} = 1,$$

and from (3),

$$\sum_{i=1}^{n} I[X_i > A_n - \epsilon] \to \infty \qquad \text{in probability.}$$

By the application of the elementary inequality

$$1 - x \leqq e^{-x}, \qquad x \geqq 0,$$

and the bounded convergence theorem, one may now conclude that

$$P\{Z_n \leqq A_n - \epsilon\} = E \prod_{i=1}^{n} (1 - I[X_i > A_n - \epsilon])$$

$$\leqq E\left\{\exp\left[-\sum_{i=1}^{n} I[X_i > A_n - \epsilon]\right]\right\} \to 0;$$

therefore, (6) is verified.

LEMMA 2. *If X and Y have a bivariate normal distribution with expectations* 0, *unit variances, and correlation coefficient r, then*

$$\lim_{c\to\infty} \frac{P\{X > c, Y > c\}}{[2\pi(1 - r)^{1/2}c^2]^{-1} \exp\left\{-\frac{c^2}{1 + r}\right\} (1 + r)^{3/2}} = 1$$

uniformly for all r such that $|r| \leqq \delta$, *for any* δ, $0 < \delta < 1$.

PROOF.

$$P\{X > c, Y > c\}$$

$$= \frac{1}{2\pi(1 - r^2)^{1/2}} \int_c^\infty \int_c^\infty \exp\left\{-\frac{1}{2(1 - r^2)} (x^2 - 2rxy + y^2)\right\} dxdy.$$

After the change of variables

$$x = [w(1 + r)/c] + c; \qquad y = [z(1 + r)/c] + c,$$

the integral becomes

$$\frac{(1 + r)^{3/2} \exp\{-c^2/(1 + r)\}}{2\pi(1 - r)^{1/2}c^2}$$

$$\cdot \int_0^\infty \int_0^\infty \exp\left\{-\frac{1 + r}{2(1 - r)c^2} (w^2 - 2rwz + z^2)\right\} e^{-w-z}\, dwdz.$$

The first exponent in the integrand is never positive; hence, as $c \to \infty$, it follows from the bounded convergence theorem that

$$\left|\int_0^\infty \int_0^\infty \exp\left\{-\frac{1 + r}{2(1 - r)c^2} (w^2 - 2rwz + z^2)\right\} e^{-w-z}\, dwdz - 1\right|$$

$$\leqq \int_0^\infty \int_0^\infty \left(1 - \exp\left[-\frac{1 + \delta}{2(1 - \delta)c^2} (w^2 + 2\delta wz + z^2)\right]\right) e^{-w-z}\, dwdz \to 0,$$

where the convergence is independent of r.

4. The main result.

THEOREM. *Let* $\{X_n\}$ *be a stationary Gaussian process such that*

$$EX_i = 0, \qquad EX_i^2 = 1, \qquad i = 1, 2, \cdots,$$
$$EX_1X_i = r_i, \qquad i = 2, 3, \cdots.$$

If

$$\lim_{n\to\infty} nr_n = 0, \tag{7}$$

then $Z_n - (2 \log n)^{\frac{1}{2}} \to 0$ *in probability.*

REMARK. The theorem requires that the covariance sequence tend to zero faster than n^{-1}. This holds, e.g., for the Markov process where $r_n = r^{n-1}$ for some r such that $0 < r < 1$.

PROOF. Condition (7) and the stationarity of the sequence imply that $|r_n| < 1$ for all n; therefore, condition (7) also implies the existence of a δ, $0 < \delta < 1$, such that $|r_n| \leqq \delta$ for all n. To prove the theorem, it will be shown that (5) holds for A_n given by (2).

From Lemma 2 and the well-known asymptotic expression for the tail of the univariate normal d.f.

$$P\{X > c\} \sim (2\pi)^{-\frac{1}{2}}c^{-1} \exp(-\tfrac{1}{2}c^2),$$

it follows that the expression corresponding to the left side of (5) is asymptotic to

$$\left(\sum_{j=2}^{[\log n]} + \sum_{j=[\log n]+1}^{n}\right)\frac{2}{n^2}(n-j+1) \exp\left[(2\log n - 2\epsilon(2\log n)^{\frac{1}{2}} + \epsilon^2)\frac{r_j}{1+r_j}\right]\frac{(1+r_j)^{\frac{1}{2}}}{(1-r_j)^{\frac{1}{2}}}, \tag{8}$$

since the convergence in Lemma 2 is uniform in r.

The first sum in (8) tends to zero; since

$$r/(1+r) \qquad \text{and} \qquad (1+r)^{\frac{1}{2}}(1-r)^{-\frac{1}{2}}$$

are increasing functions of r, the first sum is bounded above by

$$\frac{2}{n^2}\exp\left\{\epsilon^2\frac{\delta}{1+\delta}\right\}\frac{(1+\delta)^{\frac{1}{2}}}{(1-\delta)^{\frac{1}{2}}}\cdot n^{2\delta/(1+\delta)}\sum_{j=2}^{[\log n]}(n-j+1),$$

which tends to zero.

The second sum in (8) converges to 1. Since $r_n \to 0$, the factors

$$(1+r_j)^{\frac{1}{2}}(1-r_j)^{-\frac{1}{2}}$$

and $r_j/(1+r_j)$ are uniformly close to 1 and 0, respectively, for sufficiently large j; furthermore, for $j > [\log n]$, from (7),

$$|r_j/(1+r_j)|\, 2\log n \sim 2|r_j| \log n \leqq 2|r_j|j \to 0.$$

The entire second sum in (8) is therefore asymptotic to

$$\frac{2}{n^2} \sum_{j=[\log n]+1}^{n} (n - j + 1) \to 1.$$

5. Concluding remarks. The referee has pointed out that the assumption of stationarity in the theorem is not critically used; condition (7) may be replaced by

$$\lim_{n\to\infty} nEX_iX_{i+n} = 0 \qquad \text{for all } i,$$

and the proofs go through without difficulty after a suitable modification of Lemma 1. The referee's suggestions were also helpful in the elimination of unnecessary calculations in an earlier version of the paper.

The author thanks Professor Gisiro Maruyama for valuable suggestions and Professors E. J. Gumbel and Ronald Pyke for helpful discussions.

REFERENCES

[1] Geffroy, Jean, "Contribution a la théorie de valeurs extrêmes," *Publ. Inst. Stat. Univ. Paris*, Vol. 7 (1958), pp. 37–123; Vol. 8 (1959), pp. 3–65.

[2] Gnedenko, B. V., "Sur la distribution limite du terme maximum d'une série aléatoire," *Ann. Math.*, Vol. 44 (1943), pp. 423–453.

[3] Watson, G. S., "Extreme values in samples from m-dependent stationary stochastic processes," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 798–800.