ON THE STRONG LAW OF LARGE NUMBERS[1]

JOHN SLIVKA[2] AND NORMAN C. SEVERO

1. Introduction and summary. Let $\langle X_i\rangle$ be a sequence of independent and identically distributed random variables and let $S_n = \sum_1^n X_i$. The Strong Law of Large Numbers asserts that if X_i has an expectation μ, then $n^{-1}S_n \to \mu$ with probability one: hereafter it is assumed that μ exists. More precisely, this law states that for any fixed positive value of λ, with probability one the inequality $|S_n - n\mu| \geqq \lambda n$ will be fulfilled for only finitely many n-values. It is interesting to note, as was done by Wold [8, 22], that the "finitely many n-values" mentioned here is a random variable on the sample space of infinitely long realizations of the sequence $\langle S_n\rangle$. Let $\langle Y_k(\lambda)\rangle$ be the sequence of indicator variables given by $Y_k(\lambda) = 1$ if $|S_k - k\mu| \geqq \lambda k$, and 0 otherwise, and let $N_m(\lambda) = \sum_1^m Y_k(\lambda)$. Then $N_\infty(\lambda) \equiv \sum_1^\infty Y_k(\lambda)$ is precisely the "finitely many" random variable of the Strong Law of Large Numbers. Indeed, this law may be formulated in terms of this counting variable as in the following.

STRONG LAW OF LARGE NUMBERS. *If $\langle X_i\rangle$ is a sequence of independent and identically distributed random variables having a finite expectation, then for any fixed positive value of* λ, $P\{N_\infty(\lambda) < \infty\} = 1$. Thus, this fundamental law of probability is equivalent to the assertion that $N_\infty(\lambda)$ is an honest random variable or that $N_\infty(\lambda)$ has a proper distribution, and knowledge about $N_\infty(\lambda)$ will provide further insight into the nature of chance fluctuations of sums of random variables. In what follows it is shown that for fixed $t \geqq 1$, the existence of the $(t+1)$st moment of X_i is a sufficient condition for the existence of the tth moment of $N_\infty(\lambda)$ and that this result is in a particular sense "best." Further, the expected value of $N_\infty(\lambda)$ is shown to lie between $\sigma^2\lambda^{-2} - 1$ and $\sigma^2\lambda^{-2}$ when X_i has a normal distribution with variance σ^2. Bounds on the tail probability $P\{N_\infty(\lambda) \geqq j\}$ are derived under mild conditions on the X_i. With respect to the existence of moments of $N_\infty(\lambda)$, the result here is to be contrasted with that of Slivka [7] in which it is shown that even if the X_i possess moments of

Received by the editors May 19, 1969.

[1] This paper is based on a portion of the first author's doctoral dissertation at the State University of New York at Buffalo under the supervision of the second author, May 1968.

[2] Now at State University College at Buffalo.

all orders the corresponding variable for the celebrated Law of the Iterated Logarithm possesses no moments of positive order.

2. **Moments.** It is easily verified that the density and distribution functions of $N_m(\lambda)$ tend to those of $N_\infty(\lambda)$ respectively as $m \to \infty$. Indeed, when X_i has finite variance it can be shown that this convergence is uniform with respect to the range variable. The following theorem provides a useful criterion for the existence of the moments of $N_\infty(\lambda)$.

THEOREM 1. *For any fixed $t \geqq 1$ and $\lambda > 0$, the existence of $E(X_i^{t+1})$ is a sufficient condition for the existence of $E(N_\infty^t(\lambda))$.*

PROOF. Defining $N_0(\lambda) = 0$, using the notion of a telescoping sum, and conditioning on the value of $Y_k(\lambda)$ one may verify that for every positive integer m

$$\begin{aligned} E(N_m^t(\lambda)) &= \sum_{k=1}^{m} [E(N_k^t(\lambda)) - E(N_{k-1}^t(\lambda))] \\ &= \sum_{k=1}^{m} E[(N_{k-1}(\lambda) + Y_k(\lambda))^t - N_{k-1}^t(\lambda)] \\ &= \sum_{k=1}^{m} E[(N_{k-1}(\lambda) + 1)^t - N_{k-1}^t(\lambda) \mid Y_k(\lambda) = 1] P\{Y_k(\lambda) = 1\} \\ &\leqq \sum_{k=1}^{m} [k^t - (k-1)^t] P\{Y_k(\lambda) = 1\} \end{aligned}$$

since $\max_{0 \leq x \leq k-1} [(x+1)^t - x^t] = k^t - (k-1)^t$ when $t \geqq 1$. By the Convergence Theorem in Loève [6, 183], it follows that

$$\begin{aligned} E(N_\infty^t(\lambda)) &\leqq \liminf_{m \to \infty} E(N_m^t(\lambda)) \\ &\leqq \liminf_{m \to \infty} \sum_{k=1}^{m} [k^t - (k-1)^t] P\{Y_k(\lambda) = 1\} \\ &= \sum_{k=1}^{\infty} [k^t - (k-1)^t] P\{Y_k(\lambda) = 1\}. \end{aligned}$$

Since $k^t - (k-1)^t \sim t k^{t-1}$, a sufficient condition for the existence of $E(N_\infty^t(\lambda))$ is the convergence of $\sum_1^\infty k^{t-1} P\{Y_k(\lambda) = 1\}$. M. Katz [5] has shown that a necessary and sufficient condition for the convergence of the latter sum is the existence of $E(X_i^{t+1})$.

COROLLARY. *For any fixed* $\lambda>0$, *the existence of moments of all orders of* X_i *is a sufficient condition for the existence of moments of all orders of* $N_\infty(\lambda)$.

For general $t\geqq 1$, Theorem 1 cannot be improved upon as the following example for $t=1$ shows. Suppose X_i has a continuous distribution function F such that $F(0)=0$ and $1-F(x)\sim cx^{-2}$, where c is a positive constant. Then

$$E(X_i^\delta) = \int_0^\infty x^\delta dF(x) = \delta\int_0^\infty x^{\delta-1}[1-F(x)]dx, \qquad \delta>0,$$

shows that all moments of positive order $\delta<2$ of X_i exist. However, if $\beta\equiv\mu+\lambda$, so that $\beta>0$, then

$$E(N_\infty(\lambda)) \geqq \lim_{m\to\infty} E(N_m(\lambda)) = \sum_{k=1}^\infty P\{Y_k(\lambda)=1\} \geqq \sum_{k=1}^\infty P\{S_k\geqq\beta k\}$$

$$\geqq \sum_{k=1}^\infty P\{X_i\geqq\beta k \quad \text{for some } i=1,2,\cdots,k\}.$$

Employing one of Bonferroni's Inequalities stated by Feller [**3**, 100], one finds that the general term of the latter sum is not less than

$$\sum_{i=1}^k P\{X_i\geqq\beta k\} - \sum_{i=1}^{k-1}\sum_{j=i+1}^k P\{X_i\geqq\beta k \quad\text{and}\quad X_j\geqq\beta k\}$$

$$= k[1-F(\beta k)] - [\tfrac{1}{2}k(k-1)][1-F(\beta k)]^2 \sim c\beta^{-2}k^{-1}.$$

Therefore $E(N_\infty(\lambda))$ fails to exist even though all moments of order less than 2 of X_i exist.

3. **The special case of a normal distribution.** Since the central limit theorem provides conditions under which sums of independent random variables are asymptotically normally distributed, the special case in which X_i has a normal distribution with mean μ and variance σ^2 is especially attractive, for then $(S_n-n\mu)/(\sigma n^{1/2})$ has precisely the standard normal distribution. For practical purposes the theorem below provides rather precise magnitudes of $E(N_\infty(\lambda))$ in this special case.

THEOREM 2. *If* X_i *has a normal distribution with mean* μ *and variance* σ^2, *then for any* $\rho>0$, $\rho^{-2}-1\leqq E(N_\infty(\lambda))\leqq\rho^{-2}$ *where* $\rho=\lambda/\sigma$, *that is,* λ *measured in units of the standard deviation.*

PROOF. If Φ denotes the distribution function of the standard normal distribution, then

$$E(N_m(\lambda)) = \sum_{k=1}^{m} P\{Y_k(\lambda) = 1\} = 2\sum_{k=1}^{m} \Phi(-\rho k^{1/2}),$$

where ρ is defined as in the theorem. By the Euler-Maclaurin sum formula [2, 124],

$$\begin{aligned} E(N_m(\lambda)) + 1 &= 2\sum_{k=0}^{m} \Phi(-\rho k^{1/2}) \\ &= 2\int_0^m \Phi(-\rho x^{1/2})dx + \Phi(0) + \Phi(-\rho m^{1/2}) \\ &\quad + \rho\int_0^m P_1(x)x^{-1/2}\phi(-\rho x^{1/2})dx, \end{aligned}$$

where ϕ is the density function of the standard normal distribution and $P_1(x) = [x] - x + \frac{1}{2}$, $[x]$ denoting the greatest integer not exceeding x. Letting $m \to \infty$ and noting that

$$2\int_0^\infty \Phi(-\rho x^{1/2})dx = 4\rho^{-2}\int_0^\infty y[1 - \Phi(y)]dy = \rho^{-2},$$

one finds that

$$E(N_\infty(\lambda)) - \rho^{-2} + \tfrac{1}{2} = \rho\int_0^\infty P_1(x)x^{-1/2}\phi(-\rho x^{1/2})dx.$$

Since $|P_1(x)| \leqq \frac{1}{2}$,

$$\begin{aligned} |E(N_\infty(\lambda)) - \rho^{-2} + \tfrac{1}{2}| &\leqq \tfrac{1}{2}\rho\int_0^\infty x^{-1/2}\phi(\rho x^{1/2})dx \\ &= \tfrac{1}{2}\pi^{-1/2}\int_0^\infty y^{-1/2}e^{-y}dy = \tfrac{1}{2}, \end{aligned}$$

from which the desired result follows.

An application of the preceding theorem when $\rho = .01$ shows that $E(N_\infty(.01\sigma))$ lies between 9999 and 10000. Utilization of a more extended form of the Euler-Maclaurin sum formula shows that $\lim_{\rho\to 0} |E(N_\infty(\rho\sigma)) - (\rho^{-1/2} - \frac{1}{2})| = 0$.

4. **Tail probabilities.** The large magnitude of $E(N_\infty(\lambda))$ for small positive λ when X_i has a normal distribution suggests that the counting variable can often assume large values with high probability. The

following theorem provides upper bounds on the tail probabilities of $N_\infty(\lambda)$ under mild conditions on the X_i. Note that the bound in the first inequality is of order j^{-1}, while that in the second, under stricter assumptions, decreases geometrically in j. For large j the latter bound is more restrictive although this may well not be the case for moderate j.

THEOREM 3. *For any positive integer j, if X_i has finite variance σ^2, then $P\{N_\infty(\lambda)\geqq j\}\leqq 2\sigma^2 j^{-1}\lambda^{-2}$, while if the moment generating function $E(\exp(\tau X_i))$ of X_i exists, then $P\{N_\infty(\lambda)\geqq j\}\leqq B_j(\lambda)$, where*

$$B_j(\lambda) = [h(\mu+\lambda)]^j[1-h(\mu+\lambda)]^{-1} + [h(\mu-\lambda)]^j[1-h(\mu-\lambda)]^{-1}$$

and $h(a)=\inf e^{-a\tau}E(\exp(\tau X_i))$, the infimum being taken with respect to real values of τ.

PROOF. First note that $P\{N_\infty(\lambda)\geqq j\}=P\{\sum_1^\infty Y_k(\lambda)\geqq j\}\leqq P\{\sum_j^\infty Y_k(\lambda)>0\}$. If X_i has finite variance σ^2, then the Hájek-Rényi inequality [4] yields

$$\begin{aligned} P\left\{\sum_j^\infty Y_k(\lambda)>0\right\} &= P\left\{\sup_{k\geqq j}\left|k^{-1}S_k-\mu\right|\geqq\lambda\right\} \\ &\leqq \sigma^2\lambda^{-2}\left[j^{-1}+\sum_{j+1}^\infty k^{-2}\right]\leqq 2\sigma^2 j^{-1}\lambda^{-2}. \end{aligned}$$

Suppose now the moment generating function of X_i exists, i.e., converges for all τ in a neighborhood of the origin. Then, since

$$P\{Y_k(\lambda)=1\} = P\{S_k\geqq k(\mu+\lambda)\} + P\{S_k\leqq k(\mu-\lambda)\},$$

Chernoff [1] has found that for every positive integer k

$$P\{Y_k(\lambda)=1\}\leqq [h(\mu+\lambda)]^k+[h(\mu-\lambda)]^k, \tag{1}$$

since $\lambda>0$, where $h(a)$ is defined in the theorem and both $h(\mu+\lambda)$ and $h(\mu-\lambda)$ are less than 1. Note that if $\sup(X_i)$ is finite, set

$$\begin{aligned} h(\mu+\lambda) &= 0 && \text{if } \lambda>\sup(X_i)-\mu \\ &= P\{X_i=\sup(X_i)\} && \text{if } \lambda=\sup(X_i)-\mu, \end{aligned}$$

with similar remarks applying to $h(\mu-\lambda)$ if $\inf(X_i)$ is finite. Thus, from (1), one finds

$$\begin{aligned} P\left\{\sum_j^\infty Y_k(\lambda)>0\right\} &= P\{Y_k(\lambda)=1 \quad \text{for some } k\geqq j\} \\ &\leqq \sum_j^\infty P\{Y_k(\lambda)=1\}\leqq B_j(\lambda), \end{aligned}$$

by performing the indicated summation.

REFERENCES

1. H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist. **23** (1952), 493–507. MR **15**, 241.

2. H. Cramér, *Mathematical methods of statistics*. Princeton Math. Series, vol. **9**, Princeton Univ. Press, Princeton, N. J., 1946. MR **8**, 39.

3. W. Feller, *An introduction to probability theory and its applications*. Vol. I, 2nd ed., Wiley, New York, and Chapman and Hall, London, 1957. MR **19**, 466.

4. J. Hájek and A. Rényi, *Generalization of an inequality of Kolmogorov*, Acta Math. Acad. Sci. Hungar. **6** (1955), 281–283. MR **17**, 864.

5. M. Katz, *The probability in the tail of a distribution*, Ann. Math. Statist. **34** (1963), 312–318. MR **26** #1914.

6. M. Loève, *Probability theory*, 3rd ed., Van Nostrand, Princeton, N. J., 1963. MR **34** #3596.

7. J. Slivka, *On the law of the iterated logarithm*, Proc. Nat. Acad. Sci. U.S.A. **63** (1969), 289–291.

8. H. Wold (Editor), "A graphic introduction to stochastic processes," in *Bibliography on time series and stochastic processes*, M.I.T. Press, Cambridge, Mass., 1966. MR **34** #8467.

NATIONAL CANCER INSTITUTE AND
STATE UNIVERSITY OF NEW YORK AT BUFFALO