ON THE STRONG LAW OF LARGE NUMBERS

JOHN SLIVKA 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 \( S_n = \sum_{i=1}^{n} X_i \). The Strong Law of Large Numbers asserts that if \( X_i \) has an expectation \( \mu \), then \( n^{-1} S_n \rightarrow \mu \) with probability one: hereafter it is assumed that \( \mu \) exists. More precisely, this law states that for any fixed positive value of \( \lambda \), with probability one the inequality \( |S_n - n\mu| \geq \lambda n \) 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 \( |S_k - k\mu| \geq \lambda k \), and 0 otherwise, and let \( N_\alpha(\lambda) = \sum_{k=1}^{\infty} Y_k(\lambda) \). Then \( N_\alpha(\lambda) \equiv \sum_{k=1}^{\infty} Y_k(\lambda) \) 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 \( \lambda \), \( P\{N_\alpha(\lambda) < \infty \} = 1 \). Thus, this fundamental law of probability is equivalent to the assertion that \( N_\alpha(\lambda) \) is an honest random variable or that \( N_\alpha(\lambda) \) has a proper distribution, and knowledge about \( N_\alpha(\lambda) \) will provide further insight into the nature of chance fluctuations of sums of random variables. In what follows it is shown that for fixed \( t \geq 1 \), the existence of the \((t+1)\)st moment of \( X_i \) is a sufficient condition for the existence of the \( t \)th moment of \( N_\alpha(\lambda) \) and that this result is in a particular sense “best.” Further, the expected value of \( N_\alpha(\lambda) \) is shown to lie between \( \sigma^{t+1} \lambda^{-2} - 1 \) and \( \sigma^t \lambda^{-2} \) when \( X_i \) has a normal distribution with variance \( \sigma^2 \). Bounds on the tail probability \( P\{N_\alpha(\lambda) \geq j \} \) are derived under mild conditions on the \( X_i \). With respect to the existence of moments of \( N_\alpha(\lambda) \), 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 \( t \geq 1 \) and \( \lambda > 0 \), the existence of \( E(X_i^{t+1}) \)
is a sufficient condition for the existence of \( E(N_\infty(\lambda)) \).

**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 \( m \)

\[
E(N_m(\lambda)) = \sum_{k=1}^{m} \left[ E(N_k(\lambda)) - E(N_{k-1}(\lambda)) \right]
= \sum_{k=1}^{m} E\left[ (N_{k-1}(\lambda) + Y_k(\lambda))^t - N_{k-1}^t(\lambda) \right]
\leq \sum_{k=1}^{m} E\left[ (N_{k-1}(\lambda) + 1)^t - N_{k-1}^t(\lambda) \cdot Y_k(\lambda) = 1 \right] P\{ Y_k(\lambda) = 1 \}
\leq \sum_{k=1}^{m} [k^t - (k-1)^t] P\{ Y_k(\lambda) = 1 \}
\]

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

\[
E(N_\infty(\lambda)) \leq \lim_{m \to \infty} E(N_m(\lambda)) \leq \lim_{m \to \infty} \sum_{k=1}^{m} [k^t - (k-1)^t] P\{ Y_k(\lambda) = 1 \}
= \sum_{k=1}^{\infty} [k^t - (k-1)^t] P\{ Y_k(\lambda) = 1 \}.
\]

Since \( k^t - (k-1)^t \sim t k^{t-1} \), a sufficient condition for the existence
of \( E(N_\infty(\lambda)) \) is the convergence of \( \sum_{t=0}^{\infty} k^{t-1} P\{ Y_k(\lambda) = 1 \} \). 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 $N^n(\lambda)$.

For general $i \geq 1$, Theorem 1 cannot be improved upon as the following example for $i = 1$ 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

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

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

$$E(N^n(\lambda)) \geq \lim_{m \to \infty} E(N_m(\lambda)) = \sum_{k=1}^\infty P\{Y_k(\lambda) = 1\} \geq \sum_{k=1}^\infty P\{S_k \geq \beta k\} \geq \sum_{k=1}^\infty P\{X_i \geq \beta k\} \text{ for some } i = 1, 2, \ldots, k\}.$$

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

$$\sum_{i=1}^k P\{X_i \geq \beta k\} - \sum_{i=1}^{k-1} \sum_{j=i+1}^k P\{X_i \geq \beta k \text{ and } X_j \geq \beta k\} = k[1 - F(\beta k)] - \left[\frac{1}{2}k(k - 1)\right][1 - F(\beta k)]^2 \sim c\beta^{-2}k^{-1}.$$

Therefore $E(N^n(\lambda))$ 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 $\mu$ and variance $\sigma^2$ is especially attractive, for then $(S_n - n\mu)/(\sigma\sqrt{n})$ has precisely the standard normal distribution. For practical purposes the theorem below provides rather precise magnitudes of $E(N^n(\lambda))$ in this special case.

Theorem 2. If $X_i$ has a normal distribution with mean $\mu$ and variance $\sigma^2$, then for any $\rho > 0$, $\rho^{-2} - 1 \leq E(N^n(\lambda)) \leq \rho^{-2}$ where $\rho = \lambda/\sigma$, that is, $\lambda$ measured in units of the standard deviation.
Proof. If \( \Phi \) 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 \( \rho \) is defined as in the theorem. By the Euler-Maclaurin sum formula [2, 124],

\[
E(N_m(\lambda)) + 1 = 2 \sum_{k=0}^{m} \Phi(-\rho k^{1/2})
= 2 \int_{0}^{m} \Phi(-\rho x^{1/2}) dx + \Phi(0) + \Phi(-\rho m^{1/2})
+ \rho \int_{0}^{m} P_1(x) x^{-1/2} \phi(-\rho x^{1/2}) dx,
\]

where \( \phi \) 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

\[
E(N_\infty(\lambda)) = \rho^{-2} + \frac{1}{2} = \rho \int_{0}^{\infty} P_1(x) x^{-1/2} \phi(-\rho x^{1/2}) dx.
\]

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

\[
\left| E(N_\infty(\lambda)) - \rho^{-2} + \frac{1}{2} \right| \leq \frac{1}{2} \rho \int_{0}^{\infty} x^{-1/2} \phi(\rho x^{1/2}) dx
= \frac{1}{2} \pi^{-1/2} \int_{0}^{\infty} y^{-1/2} e^{-y} dy = \frac{1}{2},
\]

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 \( \lim_{\rho \to 0} \left| E(N_\infty(\rho\sigma)) - (\rho^{-1/2} - \frac{1}{2}) \right| = 0. \)

4. Tail probabilities. The large magnitude of \( E(N_\infty(\lambda)) \) for small positive \( \lambda \) when \( X_i \) has a normal distribution suggests that the counting variable can often assume large values with high probability. The
Theorem 3. For any positive integer \(j\), if \(X_i\) has finite variance \(\sigma^2\), then \(P\{N_\infty(\lambda) \geq j\} \leq 2\sigma^2 j^{-1} \lambda^{-2}\), while if the moment generating function \(E(\exp(\tau X_i))\) of \(X_i\) exists, then \(P\{N_\infty(\lambda) \geq j\} \leq B_j(\lambda)\), where

\[
B_j(\lambda) = [h(\mu + \lambda)]^2[1 - h(\mu + \lambda)]^{-1} + [h(\mu - \lambda)]^2[1 - h(\mu - \lambda)]^{-1}
\]

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

Proof. First note that \(P\{N_\infty(\lambda) \geq j\} = P\{\sum_{k=1}^{\infty} Y_k \geq j\} \leq P\{\sum_{k=1}^{j_\infty} Y_k > 0\}\). If \(X_i\) has finite variance \(\sigma^2\), then the Hájek-Rényi inequality \([4]\) yields

\[
P\left\{\sum_{j} Y_k(\lambda) > 0\right\} = P\left\{\sup_{k \geq j} | k^{-1} S_k - \mu | \geq \lambda\right\}
\leq \sigma^2 \lambda^{-2} \left[ j^{-1} + \sum_{j+1}^{\infty} k^{-2}\right] \leq 2\sigma^2 j^{-1} \lambda^{-2}.
\]

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

\[
P\{Y_k(\lambda) = 1\} = P\{S_k \geq k(\mu + \lambda)\} + P\{S_k \leq k(\mu - \lambda)\},
\]

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

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

since \(\lambda > 0\), where \(h(\alpha)\) 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

\[
h(\mu + \lambda) = 0 \quad \text{if } \lambda > \sup (X_i) - \mu
\]

\[
h(\mu - \lambda) = P\{X_i = \sup(X_i)\} \quad \text{if } \lambda = \sup(X_i) - \mu,
\]

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

\[
P\left\{\sum_{j} Y_k(\lambda) > 0\right\} = P\{Y_k(\lambda) = 1\} \text{ for some } k \geq j\]
\leq \sum_{j} P\{Y_k(\lambda) = 1\} \leq B_j(\lambda),
\]
by performing the indicated summation.

References


National Cancer Institute and
State University of New York at Buffalo