

ON THE NUMBER OF POSITIVE SUMS OF INDEPENDENT RANDOM VARIABLES

P. ERDÖS AND M. KAC¹

1. Introduction. In a recent paper² the authors have introduced a method for proving certain limit theorems of the theory of probability. The purpose of the present note is to apply this general method in order to establish the following:

THEOREM. Let X_1, X_2, \dots be independent random variables each having mean 0 and variance 1 and such that the central limit theorem is applicable. Let $s_k = X_1 + X_2 + \dots + X_k$ and let N_n denote the number of s_k 's, $1 \leq k \leq n$, which are positive. Then,

$$\lim \text{Prob.} \left\{ \frac{N_n}{n} < \alpha \right\} = \frac{2}{\pi} \arcsin \alpha^{1/2}, \quad 0 \leq \alpha \leq 1.$$

This theorem has been proved in the binomial case

$$\text{Prob.} \{X_j = 1\} = \text{Prob.} \{X_j = -1\} = 1/2$$

by P. Lévy³ who also indicated that it is true for more general random variables.

However, the authors felt that their proof is of independent interest, especially since it follows an already established pattern.²

2. The invariance principle. We first prove the following:

If the theorem can be established for *one particular* sequence of independent random variables Y_1, Y_2, \dots satisfying the conditions of the theorem then the conclusion of the theorem holds for *all* sequences of independent random variables which satisfy the conditions of the theorem. In other words, if the limiting distribution exists it is independent of the distributions of the individual X 's.

Let

$$\psi(s) = \begin{cases} 1 & \text{if } s > 0, \\ 0 & \text{if } s \leq 0, \end{cases}$$

Received by the editors March 6, 1947.

¹ This paper was written while both authors were John Simon Guggenheim Memorial Fellows.

² P. Erdős and M. Kac, *On certain limit theorems of the theory of probability*, Bull. Amer. Math. Soc. vol. 52 (1946) pp. 292-302.

³ P. Lévy, *Sur certains processus stochastiques homogènes*, Compositio Math. vol. 7 (1939) pp. 283-339, in particular, Corollaire 2 on pp. 303-304.

and note that $N_n = \sum_{r=1}^n \psi(s_r)$. Let now

$$n_j = \left[j \frac{n}{k} \right], \quad j = 0, 1, \dots, k,$$

and put

$$D_n = \frac{1}{n} \sum_{i=1}^k \sum_{r=n_{i-1}+1}^{n_i} (\psi(s_{n_i}) - \psi(s_r)).$$

We have

$$E(|D_n|) \leq \frac{1}{n} \sum_{i=1}^k \sum_{r=n_{i-1}+1}^{n_i} E(|\psi(s_{n_i}) - \psi(s_r)|)$$

and we wish now to estimate $E(|\psi(s_{n_i}) - \psi(s_r)|)$ for $n_{i-1}+1 \leq r \leq n_i$. Notice that

$$\begin{aligned} E(|\psi(s_{n_i}) - \psi(s_r)|) &= \text{Prob. } \{s_{n_i} > 0, s_r \leq 0\} \\ &\quad + \text{Prob. } \{s_{n_i} \leq 0, s_r > 0\} \end{aligned}$$

and that for $\epsilon > 0$

$$\begin{aligned} \text{Prob. } \{s_{n_i} > 0, s_r \leq 0\} &= \text{Prob. } \{s_{n_i} \geq \epsilon n_i^{1/2}, s_r \leq 0\} \\ &\quad + \text{Prob. } \{0 < s_{n_i} < \epsilon n_i^{1/2}, s_r \leq 0\} \\ &\leq \text{Prob. } \{s_{n_i} - s_r \geq \epsilon n_i^{1/2}\} \\ &\quad + \text{Prob. } \{0 < s_{n_i} < \epsilon n_i^{1/2}\}. \end{aligned}$$

By Chebysheff's inequality

$$\text{Prob. } \{s_{n_i} - s_r \geq \epsilon n_i^{1/2}\} \leq \frac{n_i - r}{\epsilon^2 n_i}$$

and hence

$$\text{Prob. } \{s_{n_i} > 0, s_r \leq 0\} \leq \frac{n_i - r}{\epsilon^2 n_i} + \text{Prob. } \{0 < s_{n_i} < \epsilon n_i^{1/2}\}.$$

In exactly the same way we obtain

$$\text{Prob. } \{s_{n_i} \leq 0, s_r > 0\} \leq \frac{n_i - r}{\epsilon^2 n_i} + \text{Prob. } \{-\epsilon n_i^{1/2} < s_{n_i} \leq 0\}$$

and thus (for $n_{i-1}+1 \leq r \leq n_i$)

$$E(|\psi(s_{n_i}) - \psi(s_r)|) \leq 2 \frac{n_i - r}{\epsilon^2 n_i} + \text{Prob.} \{ -\epsilon n_i^{1/2} < s_{n_i} < \epsilon n_i^{1/2} \}.$$

Finally,

$$\begin{aligned} E(|D_n|) &\leq \frac{2}{n\epsilon^2} \sum_{i=1}^k \frac{1}{n_i} \sum_{r=n_{i-1}+1}^{n_i} (n_i - r) \\ &\quad + \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \text{Prob.} \{ -\epsilon n_i^{1/2} < s_{n_i} < \epsilon n_i^{1/2} \} \\ &= \frac{1}{n\epsilon^2} \sum_{i=1}^k \frac{(n_i - n_{i-1})(n_i - n_{i-1} - 1)}{n_i} \\ &\quad + \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \text{Prob.} \{ -\epsilon n_i^{1/2} < s_{n_i} < \epsilon n_i^{1/2} \} \\ &= R(n, \epsilon, k). \end{aligned}$$

We note that, letting $n \rightarrow \infty$ while keeping k and ϵ fixed, we get

$$\lim_{n \rightarrow \infty} R(n, \epsilon, k) = \frac{1}{k\epsilon^2} \sum_{i=1}^k \frac{1}{i} + (2\pi)^{-1/2} \int_{-\epsilon}^{\epsilon} e^{-u^2/2} du < \frac{1 + \log k}{k\epsilon^2} + \epsilon$$

(using the central limit theorem). Let $\delta > 0$; we have

$$\text{Prob.} \{ |D_n| \geq \delta \} \leq R(n, \epsilon, k)/\delta$$

or, recalling the definition of D_n ,

$$\text{Prob.} \left\{ \left| \frac{1}{n} \sum_{r=1}^n \psi(s_r) - \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \psi(s_{n_i}) \right| \geq \delta \right\} \leq \frac{R(n, \epsilon, k)}{\delta}.$$

We have furthermore

$$\begin{aligned} \text{Prob.} \left\{ \frac{1}{n} \sum_{r=1}^n \psi(s_r) < \alpha \right\} &= \text{Prob.} \left\{ \frac{1}{n} \sum_{r=1}^n \psi(s_r) < \alpha, |D_n| < \delta \right\} \\ &\quad + \text{Prob.} \left\{ \frac{1}{n} \sum_{r=1}^n \psi(s_r) < \alpha, |D_n| \geq \delta \right\} \\ &\leq \text{Prob.} \left\{ \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \psi(s_{n_i}) < \alpha + \delta \right\} \\ &\quad + \frac{R(n, k, \epsilon)}{\delta}. \end{aligned}$$

In the same way we obtain

$$\begin{aligned} \text{Prob. } \left\{ \frac{1}{n} \sum_{r=1}^n \psi(s_r) < \alpha \right\} \\ \cong \text{Prob. } \left\{ \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \psi(s_{n_i}) < \alpha - \delta \right\} - \frac{R(n, \epsilon, k)}{\delta}. \end{aligned}$$

Combining the above inequalities we have

$$\begin{aligned} \text{Prob. } \left\{ \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \psi(s_{n_i}) < \alpha - \delta \right\} - \frac{R(n, \epsilon, k)}{\delta} \\ (1) \quad \cong \text{Prob. } \left\{ \frac{N_n}{n} < \alpha \right\} \\ \cong \text{Prob. } \left\{ \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \psi(s_{n_i}) < \alpha + \delta \right\} + \frac{R(n, \epsilon, k)}{\delta}. \end{aligned}$$

Let now G_1, G_2, \dots be independent, normally distributed random variables each having mean 0 and variance 1 and let $R_j = G_1 + \dots + G_j$. It then follows, almost immediately, from the multidimensional central limit theorem that

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Prob. } \left\{ \frac{1}{n} \sum_{i=1}^k (n_i - n_{i-1}) \psi(s_{n_i}) < \beta \right\} \\ = \text{Prob. } \left\{ \frac{1}{k} \sum_{i=1}^k \psi(R_i) < \beta \right\} = p_k(\beta). \end{aligned}$$

If in (1) we let $n \rightarrow \infty$ while keeping k, ϵ and δ fixed we obtain

$$\begin{aligned} \rho_k(\alpha - \delta) - \frac{1 + \log k}{k\epsilon^2\delta} - \frac{\epsilon}{\delta} \cong \liminf_{n \rightarrow \infty} \text{Prob. } \left\{ \frac{N_n}{n} < \alpha \right\} \\ (2) \quad \cong \limsup_{n \rightarrow \infty} \text{Prob. } \left\{ \frac{N_n}{n} < \alpha \right\} \\ \cong \rho_k(\alpha + \delta) + \frac{1 + \log k}{k\epsilon^2\delta} + \frac{\epsilon}{\delta}. \end{aligned}$$

Notice that the above inequality holds for all random variables satisfying the conditions of our theorem. Thus if for some particular sequence Y_1, Y_2, \dots we can prove that

$$\lim_{n \rightarrow \infty} \text{Prob. } \left\{ \frac{N_n}{n} < \alpha \right\} = \frac{2}{\pi} \arcsin \alpha^{1/2} = \rho(\alpha), \quad 0 \leq \alpha \leq 1,$$

then, making use of (2), with α replaced by $\alpha - \delta$ and by $\alpha + \delta$ we get at once

$$\rho(\alpha - \delta) - \frac{1 + \log k}{k\epsilon^2\delta} - \frac{\epsilon}{\delta} \leq \rho_k(\alpha) \leq \rho(\alpha + \delta) + \frac{1 + \log k}{k\epsilon^2\delta} + \frac{\epsilon}{\delta}.$$

Making use of (2) again we are led to

$$\begin{aligned} \rho(\alpha - 2\delta) - 2 \frac{1 + \log k}{k\epsilon^2\delta} - \frac{2\epsilon}{\delta} &\leq \liminf_{n \rightarrow \infty} \text{Prob.} \left\{ \frac{N_n}{n} < \alpha \right\} \\ &\leq \limsup_{n \rightarrow \infty} \text{Prob.} \left\{ \frac{N_n}{n} < \alpha \right\} \\ &\leq \rho(\alpha + 2\delta) + 2 \frac{1 + \log k}{k\epsilon^2\delta} + \frac{2\epsilon}{\delta}. \end{aligned}$$

Since this inequality holds for all k , ϵ and δ we can put, for instance,

$$\delta = k^{-1/10}, \quad \epsilon = k^{-1/5}$$

and obtain, by letting $k \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \text{Prob.} \left\{ \frac{N_n}{n} < \alpha \right\} = \rho(\alpha).$$

Use has been made of the fact that $\rho(\alpha)$ is continuous for all α . The invariance principle stated at the beginning of this section has thus been proved.

3. Explicit calculations for a particular sequence of random variables. Having established the invariance principle we could appeal to P. Lévy's result concerning the binomial case and thus complete the proof of the general theorem. We prefer, however, to work with another sequence of independent random variables because we feel that the calculations are of some independent interest and because they again emphasize the usefulness of integral equations in proving limit theorems.⁴

We consider independent random variables Y_1, Y_2, \dots such that

$$\text{Prob.} \{Y_j < u\} = \frac{1}{2} \int_{-\infty}^u e^{-|v|} dv, \quad j = 1, 2, \dots,$$

and we set $s_j = Y_1 + \dots + Y_j$.

It should be noted that although $E(Y_j^2) \neq 1$ we are justified in

⁴ See M. Kac, *On the average of a certain Wiener functional and a related limit theorem in calculus of probability*, Trans. Amer. Math. Soc. vol. 59 (1946) pp. 401-414.

using the Y 's inasmuch as N_n is independent of the variance of the Y 's as long as the variance is different from 0. In fact, N_n remains unaltered if one multiplies all the Y 's by the same positive constant.

We are going to calculate

$$\phi_n(u) = E(\exp(uN_n)) = E\left(\exp\left(u\sum_{j=1}^n \psi(s_j)\right)\right).$$

Setting

$$p(y) = 2^{-1} \exp(-|y|)$$

we have

$$\begin{aligned} \phi_n(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(u\sum_{j=1}^n \psi\left(\sum_{r=1}^j y_r\right)\right) \\ \cdot p(y_1) \cdots p(y_n) dy_1 \cdots dy_n \end{aligned}$$

or, introducing the variables

$$\begin{aligned} s_1 &= y_1, \\ s_2 &= y_1 + y_2, \\ &\dots \dots \dots \dots \dots, \\ s_n &= y_1 + y_2 + \cdots + y_n, \end{aligned}$$

$$\begin{aligned} \phi_n(u) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(u\sum_{j=1}^n \psi(s_j)\right) \\ \cdot p(s_1)p(s_2 - s_1) \cdots p(s_n - s_{n-1}) ds_1 ds_2 \cdots ds_n. \end{aligned}$$

We also introduce the auxiliary functions

$$\begin{aligned} F_n(u, s_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(u\sum_{j=1}^n \psi(s_j)\right) \\ \cdot p(s_1)p(s_2 - s_1) \cdots p(s_n - s_{n-1}) ds_1 \cdots ds_{n-1}, \quad n > 1, \end{aligned}$$

and

$$F_1(u, s_1) = \exp(u\psi(s_1))p(s_1).$$

We have, for $n > 1$,

$$F_n(u, s_n) = \exp(u\psi(s_n)) \int_{-\infty}^{\infty} p(s_n - s_{n-1}) F_{n-1}(u, s_{n-1}) ds_{n-1}$$

or, in more convenient notation,

$$(3) \quad F_n(u, s) = \exp(u\psi(s)) \int_{-\infty}^{\infty} p(s-t)F_{n-1}(u, t)dt.$$

We also have

$$(4) \quad \phi_n(u) = \int_{-\infty}^{\infty} F_n(u, s)ds$$

and

$$|F_n(u, s)| \leq \exp(\operatorname{Re} u) \max_{-\infty < t < \infty} |F_{n-1}(u, t)|.$$

Since

$$|F_1(u, s)| \leq \exp(\operatorname{Re} u)$$

we get

$$|F_n(u, s)| \leq \exp(n \operatorname{Re} u).$$

Thus the series

$$G(u, s; z) = \sum_{n=1}^{\infty} F_n(u, s)z^{n-1}$$

converges for sufficiently small $|z|$.

Using (3), we obtain, almost immediately,

$$\frac{G(u, s; z) - \exp(u\psi(s))p(s)}{z} = \exp(u\psi(s)) \int_{-\infty}^{\infty} p(s-t)G(u, t; z)dt$$

or

$$(5) \quad \begin{aligned} \exp(-u\psi(s))G(u, s; z) - 2^{-1} \exp(-|s|) \\ = \frac{z}{2} \int_{-\infty}^{\infty} \exp(-|s-t|)G(u, t; z)dt. \end{aligned}$$

For $s > 0$ we have

$$(6) \quad \begin{aligned} e^{-u}G(u, s; z) - 2^{-1}e^{-s} &= \frac{z}{2} e^{-s} \int_{-\infty}^s e^t G(u, t; z) dt \\ &+ \frac{z}{2} e^s \int_s^{\infty} e^{-t} G(u, t; z) dt. \end{aligned}$$

It is easily seen that, for sufficiently small $|z|$, $G(u, t; z)$ is an absolutely integrable function of t .

Differentiating (6) twice with respect to s we are readily led to

$$\frac{d^2G}{ds^2} + (ze^u - 1)G = 0.$$

In the same way we get for $s < 0$

$$\frac{d^2G}{ds^2} + (z - 1)G = 0.$$

Thus we get

$$G = A \exp((1 - ze^u)^{1/2}s) + B \exp(-(1 - ze^u)^{1/2}s), \quad s > 0,$$

and

$$G = C \exp((1 - z)^{1/2}s) + D \exp(-(1 - z)^{1/2}s), \quad s < 0,$$

where A, B, C, D are functions of u and z . Since G is an integrable function of s we must have $A \equiv D \equiv 0$.

Thus

$$(7) \quad G = B \exp(-(1 - ze^u)^{1/2}s), \quad s > 0,$$

$$(8) \quad G = C \exp((1 - z)^{1/2}s), \quad s < 0.$$

From the absolute integrability of G and the integral equation (5) it follows that

$$\exp(-u\psi(s))G(u, s; z)$$

is continuous for all s and, in particular, for $s=0$. This observation yields

$$(9) \quad Be^{-u} = C.$$

Rewriting (6) in the form

$$e^{-u}G(s) - 2^{-1}e^{-s} = \frac{z}{2}e^{-s} \int_{-\infty}^0 e^{tG(t)} dt + \frac{z}{2}e^{-s} \int_0^s e^{tG(t)} dt \\ + \frac{z}{2}e^s \int_s^{\infty} e^{-tG(t)} dt$$

and substituting expressions (7) and (8) we get after a few elementary transformations

$$\frac{z}{1 + (1 - z)^{1/2}} C - \frac{z}{1 - (1 - ze^u)^{1/2}} B = -1.$$

This together with (9) yields

$$B = \frac{(1-z)^{1/2} - (1-ze^u)^{1/2}}{z(e^u-1)} e^u, \quad C = \frac{(1-z)^{1/2} - (1-ze^u)^{1/2}}{z(e^u-1)}$$

and thus G is completely determined.⁵

In particular, we get

$$\int_{-\infty}^{\infty} G(u, s; z) ds = \frac{(1-z)^{1/2} - (1-ze^u)^{1/2}}{z(e^u-1)} \left[\frac{e^u}{(1-ze^u)^{1/2}} + \frac{1}{(1-z)^{1/2}} \right].$$

On the other hand (see (4))

$$\int_{-\infty}^{\infty} G(u, s; z) ds = \sum_{n=1}^{\infty} \phi_n(u) z^{n-1}$$

and hence $\phi_n(u)$ is the coefficient of z^{n-1} in the power series expansion of

$$\frac{(1-z)^{1/2} - (1-ze^u)^{1/2}}{z(e^u-1)} \left[\frac{e^u}{(1-ze^u)^{1/2}} + \frac{1}{(1-z)^{1/2}} \right].$$

A simple computation gives finally

$$E(\exp(uN_n)) = \phi_n(u) = \frac{1}{e^u - 1} \sum_{k+l=n} C_{1/2,k} C_{-1/2,l} (-1)^{k+l} (1 - e^{ku})(1 + e^{(l+1)u}).$$

Setting

$$u = \frac{i\xi}{n}$$

we obtain the characteristic function of N_n/n .

For large k and l we have

$$C_{1/2,k} \sim (-1)^{k-1} \frac{1}{2k-1} \frac{1}{(\pi k)^{1/2}}, \quad C_{-1/2,l} \sim (-1)^l \frac{1}{(\pi l)^{1/2}},$$

and it is easily seen that, in the limit $n \rightarrow \infty$, we can use these asymptotic expressions in the sum.

⁵ Having found an explicit expression for G it is not too difficult to find explicit expressions for $F_n(u, s)$ and verify the recursive relation (3). Since this recursive relation together with the formula for $F_1(u, s)$ determine G uniquely one might substitute the verification process for the derivation of the formula for G .

Thus

$$\begin{aligned}
 & \lim_{n \rightarrow \infty} E \left(\exp \left(\frac{i\xi}{n} N_n \right) \right) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{\pi(1 - e^{i\xi/n})} \sum_{k=1}^n \frac{1}{2k-1} \frac{1}{(k(n-k))^{1/2}} \\
 & \qquad \qquad \qquad \cdot (1 - e^{i\xi k/n})(1 + e^{i\xi(1-k/n+1/n)}) \\
 &= \frac{1}{2\pi i\xi} \int_0^1 \frac{(1 - e^{i\xi x})(1 + e^{i\xi(1-x)})}{x(x(1-x))^{1/2}} dx = \frac{1}{\pi} \int_0^1 e^{i\xi x} \frac{dx}{(x(1-x))^{1/2}} \\
 &= \int_0^1 e^{i\xi x} d \left(\frac{2}{\pi} \arcsin x^{1/2} \right).
 \end{aligned}$$

We complete the proof by appealing to the continuity theorem for Fourier-Stieltjes transforms.

SYRACUSE UNIVERSITY AND
CORNELL UNIVERSITY