

ASYMPTOTIC DISTRIBUTION OF THE MAXIMUM CUMULATIVE SUM OF INDEPENDENT RANDOM VARIABLES

KAI LAI CHUNG

The limiting distribution of the maximum cumulative sum¹ of a sequence of independent random variables has been discussed recently by Erdős-Kac² and Wald.³ Erdős and Kac treated the case where each random variable has zero mean, while Wald considered more general cases.

We shall show that the problem can be treated by a uniform method starting with a classical combinatorial formula due to De Moivre.⁴ A careful application of Stirling's formula does the trick in all cases, but it is interesting that quite different transformations are needed in different cases. It is also to be noted that this is the usual method of arriving at the normal approximation to the binomial distribution given in elementary textbooks (frequently without rigor).

By this method we obtain easily a remainder term to the approximation. In the Bernoullian case this is the best possible order of magnitude. In this general case we have to use a recent theorem of H. Bergström⁵ concerning the remainder term in the k -dimensional central limit theorem. Although it was A. C. Berry and Esseen (independently of each other) who first obtained this result in the one-dimensional case and their methods can be extended to higher dimensions, Bergström's result is more precise in the determination of the dependence on k .

For the sake of simplicity we assume that each variable has unit variance. It is easy to remove this restriction but it is cumbersome to do so.

Presented to the Society, April 17, 1948; received by the editors January 28, 1948.

¹ We use the term "maximum cumulative sum" for $\max S_k$; we have used the term "maximum partial sum" for $\max |S_k|$ (*On the maximum partial sums of sequences of independent random variables*, Trans. Amer. Math. Soc. vol. 64 (1948) pp. 205-253).

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 A. Wald, *Limit distribution of the maximum and minimum of successive cumulative sums of random variables*, Bull. Amer. Math. Soc. vol. 53 (1947) pp. 142-153.

⁴ See, for example, Uspensky, *Introduction to mathematical probability*, New York, McGraw-Hill, 1937, p. 153.

⁵ H. Bergström, *On the central limit theorem in the space R_k , $k > 1$* , Skandinavisk Aktuarietidskrift vol. (1945) pp. 106-127. The result there was stated for random variables having the same distribution, but no change is needed to carry over into the present case. I owe to Dr. Bergström the present form of his result.

We state our results as follows. Let X_1, \dots, X_n be a sequence of independent random variables with

$$E(X_k) = \mu_k, \quad E(X_k^2) = 1 + \mu_k^2, \\ E(|X_k - \mu_k|^3) = \gamma_k = O(1).$$

Let

$$S_n = \sum_{k=1}^n X_k, \quad \bar{S}_n = \max_{1 \leq k \leq n} S_k.$$

Then the asymptotic distributions of \bar{S}_n as n tends to infinity are given below:

Case (i): $\mu_k = 0, k = 1, \dots, n.$

$$\Pr(\bar{S}_n < an^{1/2}) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^a e^{-x^2/2} dx + O(n^{-1/26} \log n).$$

Case (ii): $\mu_k = dn^{-1/2} (d \neq 0), k = 1, \dots, n.$

$$\Pr(\bar{S}_n < an^{1/2}) = \int_{a^2/2}^\infty \frac{ade^{ad}}{2\pi^{1/2}x^{3/2}} e^{-a^2d^2/4x-x} dx + O(n^{-1/26} \log n).$$

Case (iii): $\mu_k = \mu > 0, k = 1, \dots, n.$

$$\Pr(\bar{S}_n < n\mu + an^{1/2}) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^a e^{-x^2/2} dx + O(n^{-1/26} \log n).$$

In the Bernoullian case the remainder term in each case is to be replaced by $O(n^{-1/2})$.

1. **The Bernoullian case.** Let each $X_\nu, \nu = 1, \dots, n,$ be distributed as follows:

$$(1) \quad X_\nu = \begin{cases} +g & \text{with probability } p = 2^{-1}(1 + \mu/g), \\ -g & \text{with probability } q = 2^{-1}(1 - \mu/g), \end{cases}$$

where

$$(2) \quad g^2 = 1 + \mu^2.$$

Then

$$E(X_\nu) = \mu, \quad E(X_\nu - \mu)^2 = 1.$$

For definitiveness we assume that n and b are both even. From a classical formula⁶ we have

⁶ See Uspensky, op. cit. p. 151, formula (12), p and q being interchanged.

$$(3) \quad \Pr (\bar{S}_n \geq bg) = p^b \sum_{r=0}^{(n-b)/2} \frac{b(b+2r-1)!}{r!(b+r)!} (pq)^r.$$

Since

$$\lim_{n \rightarrow \infty} \Pr (\bar{S}_n \geq bg) = 1$$

we have

$$(4) \quad \Pr (\bar{S}_n < bg) = p^b \sum_{r > (n-b)/2} \frac{b(b+2r-1)!}{r!(b+r)!} (pq)^r.$$

Using Stirling's formula we have

$$(5) \quad \begin{aligned} & \log \left[\frac{b(b+2r-1)!}{r!(b+r)!} (pq)^r p^b \right] \\ &= \log b + \left(b + 2r - \frac{1}{2} \right) \log (b + 2r - 1) \\ & \quad - \left(r + \frac{1}{2} \right) \log r - \left(b + r + \frac{1}{2} \right) \log (b + r) \\ & \quad + 1 - \frac{1}{2} \log 2\pi + r \log pq + b \log p + O\left(\frac{1}{b+r}\right). \end{aligned}$$

Assuming $b = o(r)$, $\mu = o(1)$, we can write (5) as

$$(6) \quad \begin{aligned} & \log b + \left(b + 2r - \frac{1}{2} \right) \left[\log 2r + \frac{b-1}{2r} - \frac{(b-1)^2}{8r^2} + O\left(\frac{b^3}{r^3}\right) \right] \\ & \quad - \left(r + \frac{1}{2} \right) \log r \\ & \quad - \left(b + r + \frac{1}{2} \right) \left[\log r + \frac{b}{r} - \frac{b^2}{2r^2} + O\left(\frac{b^3}{r^3}\right) \right] \\ & \quad + 1 - \frac{1}{2} \log 2\pi - r(\log 4 + \mu^2 + O(\mu^4)) \\ & \quad - b \log 2 + b \left[\frac{\mu}{g} + O(\mu^2) \right] + O\left(\frac{1}{b+r}\right) \\ & = \log \frac{b}{2\pi^{1/2}r^{3/2}} - \frac{b^2}{4r} + \frac{b\mu}{g} - r\mu^2 + O\left(\frac{b^3}{r^2} + \frac{b}{r} + r\mu^4 + b\mu^2\right). \end{aligned}$$

Now let

$$\mu = dn^{-1/2}, \quad b = an^{1/2},$$

where a and d are constants, $a > 0$. Then if $2^{-1}n \leq r \leq n^{3/2}$, (6) gives

$$\begin{aligned} & \frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b \\ (7) \quad &= \frac{an^{1/2}}{2\pi^{1/2}r^{3/2}g} \exp\left(-\frac{a^2n}{4g^2r} - \frac{d^2r}{n} + \frac{ad}{g^2}\right) \cdot (1 + O(n^{-1/2})) \\ &= \frac{an^{1/2}}{2\pi^{1/2}r^{3/2}} \exp\left(-\frac{a^2n}{4r} - \frac{d^2r}{n} + ad\right) \cdot (1 + O(n^{-1/2})). \end{aligned}$$

If $r > n^{3/2}$, (6) gives

$$(8) \quad \frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b = O(\exp(-Ar^{1/3})),$$

where A , as later, is a positive constant which need not be the same each time it appears. Therefore we obtain from (7) and (8)

$$\begin{aligned} & \sum_{r > (n-b)/2} \frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b \\ (9) \quad &= \sum_{n/2 \leq r \leq n^{3/2}} \frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b + O\left(\frac{bn^{1/2}}{n^{3/2}} e^A\right) \\ & \quad + \sum_{r > n^{3/2}} \exp(-Ar^{1/3}) \\ &= \sum_{r \geq n/2} \frac{an^{1/2}}{2\pi^{1/2}r^{3/2}} \exp\left(-\frac{a^2n}{4r} - \frac{d^2r}{n} + ad\right) + O(n^{-1/2}) \\ &= \int_{n/2}^{\infty} \frac{an^{1/2}}{2\pi^{1/2}y^{3/2}} \exp\left(-\frac{a^2n}{4y} - \frac{d^2y}{n} + ad\right) dy + O(n^{-1/2}). \end{aligned}$$

If $d \neq 0$, letting

$$x = d^2y/n$$

we can write (9) as

$$\begin{aligned} (10) \quad & \int_{d^2/2}^{\infty} \frac{ade^{ad}}{2\pi^{1/2}x^{3/2}} \exp\left(-\frac{a^2d^2}{4x} - x\right) dx + O(n^{-1/2}) \\ &= I(a) + O(n^{-1/2}). \end{aligned}$$

If $d = 0$, letting

$$a^2n/2r = y^2$$

we can write (9) as

$$(11) \quad \left(\frac{2}{\pi}\right)^{1/2} \int_0^a e^{-y^2/2} dy + O(n^{-1/2}).$$

2. The method of Erdős-Kac. Let $X_\nu, \nu=1, \dots, n$, be independent, and

$$E(X_\nu) = \mu = dn^{-1/2}, \quad E(X_\nu - \mu)^2 = 1, \quad E(|X_\nu - \mu|^3) = \gamma_\nu.$$

We also assume that

$$\gamma_\nu = O(1).$$

Let

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

We shall take

$$k = k(n) \sim n^{1/13},$$

$$\epsilon = \epsilon(n) = \left(\frac{\log n}{k}\right)^{1/2} \sim \frac{(\log n)^{1/2}}{n^{1/26}}.$$

According to the method of Erdős-Kac, we have

$$(12) \quad \Pr(\bar{S}_n < an^{1/2}) \geq \Pr\left(\max_{1 \leq j < k} S_{n_j} < (a - \epsilon)n^{1/2}\right) + \max_{1 \leq i < k} \max_{n_i < r \leq n_{i+1}} \Pr(|S_{n_{i+1}} - S_r| \geq \epsilon n^{1/2}).$$

Let $n_i < r \leq n_{i+1}$. Let $B = B_n$ be a function of n to be determined later.

If $n_{i+1} - r \leq B$, then

$$(13) \quad \Pr(|S_{n_{i+1}} - S_r| \geq \epsilon n^{1/2}) \leq \frac{n_{i+1} - r + (n_{i+1} - r)^2 \mu^2}{\epsilon^2 n} \leq \frac{B + B^2 \mu^2}{\epsilon^2 n}.$$

If $n_{i+1} - r > B$, then

$$(14) \quad \Pr(|S_{n_{i+1}} - S_r| \geq \epsilon n^{1/2}) = \frac{1}{(2\pi)^{1/2}} \left\{ \int_{v_1}^\infty + \int_{-\infty}^{-v_2} e^{-u^2/2} du \right\} + O((n_{i+1} - r)^{-1/2})$$

where

$$v_1 = \frac{\epsilon n^{1/2} - (n_{i+1} - r)\mu}{(n_{i+1} - r)^{1/2}}, \quad v_2 = \frac{\epsilon n^{1/2} + (n_{i+1} - r)\mu}{(n_{i+1} - r)^{1/2}}.$$

If $n > n_0$, a fixed constant, we have

$$\begin{aligned} n_{i+1} - r &\leq n_{i+1} - n_i < 2n/k, \\ k\epsilon &> 4d; \end{aligned}$$

hence

$$\begin{aligned} \epsilon n^{1/2} - (n_{i+1} - r)\mu &= \epsilon n^{1/2} \left[1 - \frac{(n_{i+1} - r)d}{\epsilon n} \right] \\ &\geq \epsilon n^{1/2} \left[1 - \frac{2d}{k\epsilon} \right] \geq \frac{\epsilon}{2} n^{1/2}, \\ v_2 > v_1 &> \frac{\epsilon n^{1/2}}{2(2nk^{-1})^{1/2}} \geq \frac{k^{1/2}\epsilon}{2^{3/2}} \geq \left(\frac{\log n}{8} \right)^{1/2}, \\ \left(\int_{v_1}^{\infty} + \int_{-\infty}^{-v_2} \right) e^{-u^2/2} du &= O\left(\exp\left(-\frac{\log n}{16} \right) \right) = O(n^{-1/16}). \end{aligned}$$

Now we choose B so that

$$\frac{B}{\epsilon^2 n} = \frac{1}{B^{1/2}}, \quad B = \epsilon^{4/3} n^{2/3}.$$

Then

$$\frac{B + B^2 \mu^2}{\epsilon^2 n} = O(n^{-1/4}), \quad B^{-1/2} = O(n^{-1/4}).$$

Thus combining (13) and (14) we obtain

$$(15) \quad \Pr(|S_{n_{i+1}} - S_r| \geq \epsilon n^{1/2}) = O(n^{-1/16}).$$

Hence from (12)

$$(16) \quad \Pr(\bar{S}_n < an^{1/2}) \geq \Pr\left(\max_{1 \leq j \leq k} S_{n_j} < (a - \epsilon)n^{1/2} \right) + O(n^{-1/16}).$$

3. The estimate of Bergström. Write

$$V_i^{(p)} = \sum_{j \equiv p \pmod{n_1}, j \leq n_i} X_j$$

and consider the k -dimensional vector

$$V^{(p)} = (V_1^{(p)}, \dots, V_k^{(p)});$$

we have

$$(17) \quad \sum_{p=1}^{n_1} V^{(p)} = (S_{n_1}, \dots, S_{n_k}).$$

Now

$$E(|V_i^{(p)} - E(V_i^{(p)})|^3) \leq A k^2 \sum_{j \equiv p \pmod{n_1}, j \leq n_i} E(|X_j - \mu|^3),$$

$$E(|V_i^{(p)} - E(V_i^{(p)})|^2) = \sum_{j \equiv p \pmod{n_1}, j \leq n_i} 1.$$

Let $F(x_1, \dots, x_k)$ be the distribution of (17) and let $\phi(x_1, \dots, x_k)$ be that of the k -dimensional normal distribution having the same first and second order moments as F . According to the result of Bergström,⁴ we have

$$|F - \phi| < \frac{C(k)}{n_1^{1/2}} \max_p \sum_{i=1}^k \frac{E(|V_i^{(p)} - E(V_i^{(p)})|^3)}{E(|V_i^{(p)} - E(V_i^{(p)})|^2)}$$

$$< \frac{C(k)}{n_1^{1/2}} A k^{5/2} \max_{1 \leq j \leq n} \gamma_j < A \frac{C(k) k^{5/2}}{n_1^{1/2}}.$$

The $C(k)$ in this result can be taken to be $A k^{7/2} \log k$, hence

$$(18) \quad |F - \phi| \leq A \frac{k^6 \log k}{n_1^{1/2}} \leq A \frac{\log n}{n_1^{1/26}}.$$

Remembering that

$$F(x, \dots, x) = \Pr \left(\max_{1 \leq j \leq k} S_{n_j} \leq x \right),$$

we obtain from (16) and (18) the following inequality

$$(19) \quad \Pr(\bar{S}_n < an^{1/2}) \geq \phi((a - \epsilon)n^{1/2}, \dots, (a - \epsilon)n^{1/2})$$

$$+ An^{-1/26} \log n.$$

On the other hand, we have

$$(20) \quad \Pr(\bar{S}_n < an^{1/2}) \leq \Pr \left(\max_{1 \leq j \leq k} S_{n_j} < an^{1/2} \right)$$

$$= \phi(an^{1/2}, \dots, an^{1/2}) + An^{-1/26} \log n,$$

where the last equation follows from (18).

4. **The final result.** As before let $k \sim n^{1/18}$. Let $\phi_k(x)$ denote $\phi(x_1, \dots, x_k)$ with $x_1 = \dots = x_k = x$. Then we have obtained in (19) and (20)

$$(21) \quad \begin{aligned} \phi_k((a - \epsilon)n^{1/2}) - \frac{A \log n}{n^{1/26}} &\leq \Pr(\bar{S}_n < an^{1/2}) \\ &\leq \phi_k(an^{1/2}) + \frac{A \log n}{n^{1/26}}. \end{aligned}$$

Applying these inequalities to the Bernoullian case in §1, and considering the case $d \neq 0$ for the sake of definitiveness, we obtain

$$\begin{aligned} \phi_k((a - \epsilon)n^{1/2}) - \frac{A \log n}{n^{1/26}} &\leq I(a) + O\left(\frac{1}{n^{1/2}}\right) \\ &\leq \phi_k(an^{1/2}) + \frac{A \log n}{n^{1/26}}. \end{aligned}$$

Consequently we have

$$(22) \quad I(a - \epsilon) - \frac{A \log n}{n^{1/26}} \leq \Pr(\bar{S}_n < an^{1/2}) \leq I(a + \epsilon) + \frac{A \log n}{n^{1/26}}.$$

Since it is easy to see that

$$(23) \quad I(a) - I(a \pm \epsilon) = O(\epsilon) = O\left(\frac{(\log n)^{1/2}}{n^{1/26}}\right),$$

we obtain from (22) and (23)

$$\Pr(\bar{S}_n < an^{1/2}) = I(a) + O\left(\frac{\log n}{n^{1/26}}\right).$$

5. **The case $\mu_k = \mu \neq 0$.** We have left out this case in order not to interrupt the argument. We shall now briefly indicate the transformations needed here. Instead of making the variance one in this case we shall follow the more usual way and assume that

$$X_\nu = \begin{cases} +1 & \text{with probability } p, \\ -1 & \text{with probability } q \end{cases}$$

where

$$p - q = \mu.$$

Then

$$E(X_r) = \mu, \quad E(X_r - \mu)^2 = 4pq.$$

In formula (3) we write

$$(24) \quad \begin{aligned} b &= n\mu + a(4pqn)^{1/2}, \\ r &= nq + s = nq + t(pqn)^{1/2} \end{aligned}$$

where q is a constant and t varies between $-(p^{-1}qn)^{1/2}$ and $-a$.

We find, upon using Stirling's formula, that if $t = O(n^{1/4})$

$$\begin{aligned} \frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b &= \left(\frac{p\mu^2}{2\piqn}\right)^{1/2} \exp\left(-\frac{(\mu t - 2qa)^2}{2}\right) \\ &\cdot \left(1 + O\left(\frac{t}{n^{1/2}}\right)\right) \end{aligned}$$

and if $n^{1/4} = O(t)$,

$$\frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b = O(\exp(-An^{1/2})).$$

Thus we can write, t running through the values defined by (24) as r runs through integers,

$$\begin{aligned} &\sum_{r=0}^{(n-b)/2} \frac{b(b + 2r - 1)!}{r!(b + r)!} (pq)^r p^b \\ &= \sum_{r \leq n^{3/4}} + \sum_{r > n^{3/4}} = \sum_{-n^{1/4} < t < -a} \left(\frac{p\mu^2}{2\piqn}\right)^{1/2} \exp\left(-\frac{(\mu t - 2qa)^2}{2}\right) \\ &\quad \cdot \left(1 + O\left(\frac{t}{n^{1/2}}\right)\right) + O(n \exp(-An^{1/2})) \\ &= \frac{1}{(2\pi)^{1/2}} \int_a^\infty e^{-x^2/2} dx + O\left(\frac{1}{n^{1/2}}\right). \end{aligned}$$