MULTINOMIAL RATIO [PAUL ERDÖS SOLVES A PROBLEM]

Kai Lai Chung

1. Historical origin

The asymptotic distribution of sums $S_n = \sum_{k=1}^{n} X_k$ of integer-valued, independent random variables with a common distribution was the grand theme of classical probability, associated with names like Jakob Bernoulli, DeMoivre, and Laplace. If the mean is zero and the variance is finite, then for each integer c, when $P(S_n = c) > 0$, it is asymptotically equal to ($\sim$) a constant times $n^{-1/2}$. When the variance is infinite, it becomes $o(n^{-1/2})$. The standard proof employs the Fourier transform, alias characteristic function (ch. f.); and in the latter case, a finer tuning of the analytic technique is required which is not often given in the literature. No sharp asymptotic result is known without very special conditions on the distribution; the mere finiteness of an absolute moment of order between 1 and 2 is not sufficient. On the other hand, under the sole condition that the mean is zero, Wolfgang Fuchs and I proved a result including, as a particular integer-valued case, the divergence of the series $\sum_{n=1}^{\infty} P(S_n = 0)$. These circumstances led me to consider the asymptotic behavior of the ratio

$$\frac{P(S_n = c)}{P(S_n = c')}$$

and to venture the optimistic guess that it should go to 1 as $n \to \infty$. To my pleasant surprise, this was quickly confirmed when the underlying common distribution is symmetric, or equivalently when its ch. f. is real-valued and symmetric. Details of this are given in the final section of this article. But in the general case under "zero mean", the complex-valued ch. f. appeared intractable, at least for me. This was particularly frustrating because in the proof of the divergence of the series mentioned above such ch. f.'s were dealt with successfully.

When I told Erdős about the problem around 1949, he was instantly interested, thought it *nice*, and was tickled by the failure of the Fourier method which he disdained as needless. Here the much abused adjective "nice" is used in the proper classic sense, meaning "requiring precision or care, fine, subtle".

Fifty years being a long time, I can no longer recall the speed of Erdős's performance, but I can still hear the tone of his announcing it, more or less:

> "Look: if S_n can take the value c and the value c', then we can change c into c' by swapping a few of the possible values of the X_ks. Now the law of large numbers says that each of these values is taken a very large number of times (proportional to n, in fact); hence the few (fixed) swaps can't make much difference."

From left to right: Mother of Paul Erdős, Paul, Charles Loewner, the author, and author's son Daniel.

An early version of this argument is given in [3]. On rereading it, I was astonished and embarrassed by the miserable misprints that I had failed to correct in that poorly printed, newly launched Memoirs series, though I found no real mathematical mistake. I am grateful to have a *second chance* in this memorial to Erdős.

Epilogue. William Feller was known to hold strong opinions. I recall his tirade about Theorem 3: "True, it may be hard to prove it if we have only the silly mean zero; but if we assume any $1 + \varepsilon$ moment, it is trivial!" When I eventually made bold to ask him about the proof, after trying it in vain, he burst into a huge smile. On another occasion, he called Kolmogorov's three-series theorem a trivial deduction of the L^2 convergence in Hilbert space, after he heard the latter in Harry Pollard's class. Harry was astounded too.

2. An integer representation theorem

Euclid's algorithm determines the greatest common divisor (GCD) of two positive integers, and hence by iteration that of a finite number of them. The GCD of an infinite set of positive integers is the GCD of a certain finite subset. One can also extend the definition to any set of integers by simply replacing the negative ones with their absolute values. It is a consequence of Euclid's algorithm that if d is the GCD of $\{a_j\}$ then there exist integers x_j such that

$$d = \sum x_j a_j \tag{1}$$

where as below the sum is over all j. In fact, one can *define* the GCD of the set $\{a_j\}$ as the smallest positive member of the class of all sums of the form given in (1) with arbitrary integers $\{x_j\}$ — that class has the name "modul."

A well-known theorem asserts that there exists a positive integer M such that for all $n \geq M$, we have

$$nd = \sum x_j a_j \qquad \text{with all } x_j \geq 0. \tag{2}$$

The proof is often omitted but here it is. Starting with the representation in (1), we have

$$cd = \sum |x_j| a_j \tag{3}$$

for an integer $c > 0$ because d divides all a_j. For any integer $n \geq c^2$, we have $n = qc + r$ where $q \geq c$ and $0 \leq r < c$. Therefore, we have, using both (1) and (3),

$$nd = (qc + r)d = \sum (q|x_j| + rx_j)a_j. \tag{4}$$

Since $q \geq c$, all the coefficients of a_j in (4) are positive, so the assertion is proved with $M = c^2$.

The next result may be viewed as a *deepening* of the above, which Erdős considered as "obvious." At that epoch, circa 1949, I seemed also to think so. But I did write down a detailed proof and mentioned it in a footnote (see [1]), and kept the notes somewhere for years. A recent search failed to uncover it. Not without some effort was I able to supply the version given below.

Theorem 1. [1] *Let $\{a_j\}$ be a finite or infinite set of integers. The two conditions:*

(i) *there are both positive and negative integers in the set $\{a_j\}$;*

(ii) *the GCD of the set of all differences $\{a_i - a_j\}$ is 1;*
are necessary and sufficient for the following contingency:

(iii) *for any integer c, there is $M(c) > 0$ such that for any integer $n \geq M(c)$, there exist integers $x_j \geq 0$, such that*

$$\sum x_j = n \qquad and \qquad \sum x_j a_j = c. \tag{5}$$

---

[1]Since the article was written, I have asked several number-theorists and algebraists (including Halberstam via Philipp) as well as probabilists about Theorem 1, in the hope of finding a reference. Whereas the GCD in (ii) is well-known as the "span" of a "lattice distribution," no trace of the result was found. An elegant proof of Theorem 1 in the Bourbaki manner by Giorgio Letta (pisa) and Luca Pratelli will appear in *Rivista dei XL*.

Proof. The necessity of (i) for (iii) is obvious. Next, take $c = 1$ in (iii), and let d be the GCD in (ii). If (iii) holds for $c = 1$, then for some positive integer n, we have $x_j \geq 0$ satisfying

$$\sum x_j = nd \qquad \text{and} \qquad \sum x_j a_j = 1. \tag{6}$$

It follows that

$$nda_1 - 1 = \sum x_j(a_1 - a_j). \tag{7}$$

Each summand on the right side is divisible by d, hence d must divide 1 on the left side. Thus $d = 1$ as asserted.

Now assume (i) and (ii). We shall first prove (iii) for $c = 0$. The set of differences

$$a_j - a_i \quad \text{for all pairs } (j, i) \text{ such that } a_j > 0 > a_i \tag{8}$$

has the same GCD as the full set of differences in (ii). This is easy to see since any difference in (ii) is the difference of two of the restricted differences in (8). Hence, by the old theorem proved above, there exists $N > 0$ such that for any integer $n \geq N$, we have

$$n = \sum{}^* m_{ji}(a_j - a_i) \qquad \text{with} \qquad m_{ji} \geq 0 \tag{9}$$

where $\sum^*$ denotes the summation over the pairs (j, i) restricted as in (8). Now for each such pair (j, i), we have

$$0 = a_j a_i + |a_i| a_j. \tag{10}$$

It follows that

$$0 = \sum{}^* m_{ji}(a_j a_i + |a_i| a_j). \tag{11}$$

In view of (9), valid for $n = N$, we may take N to be the $M(0)$ required in (iii), thus proving (iii) in the case $c = 0$.

Once this is done, in order to prove (iii) for any integer c, it is sufficient to prove the existence of $x_j \geq 0$ such that

$$c = \sum x_j a_j. \tag{12}$$

We will prove this for $c = -1$. The GCD of the set $\{|a_j|\}$ being a divisor of the GCD in (ii) must be 1. Let us denote by d^- and d^+, respectively, the GCD of all $-a_j$ for $a_j < 0$ and the GCD of all a_j for $a_j > 0$. Then the GCD of d^- and d^+ is 1.

By the old theorem, there exists $M > 0$ such that for any integer $n \geq M$, we have simultaneously

$$-nd^- = \sum{}^- m_j a_j \qquad \text{and} \qquad nd^+ = \sum{}^+ m_j a_j \tag{13}$$

where all m_j are positive integers, and the sums $\sum^-$ and $\sum^+$ extend, respectively, over the negative and positive a_j's. Since 1 is the GCD of d^- and d^+, we have by Euclid

$$-1 = ad^- + bd^+ \tag{14}$$

where a and b are integers. Hence, for any integer s, we have also

$$-1 = (sd^+ + a)d^- + (-sd^- + b)d^+. \tag{15}$$

Choose an s such that

$$sd^+ + a \leq -M \qquad \text{and} \qquad -sd^- + b \geq M$$

where M is defined above (13). Then we have by (13)

$$(sd^+ + a)d^- = \sum{}^{-} m_j a_j, \qquad (-sd^- + b)d^+ = \sum{}^{+} m_j a_j \tag{16}$$

where as before all $m_j \geq 0$. Thus by (15), we have achieved a representation of -1 as a linear form of the $\{a_j\}$ with integer coefficients $m_j \geq 0$. Let $m = \sum m_j$. Recall the meaning of $M(c)$ in (iii). Since $-1 + 0 = -1$, we may take $M(-1)$ to be $m + M(0)$.

This proves (iii) for $c = -1$. It is clear that we may take, for any negative integer $c, M(c)$ to be $|c|M(-1)$.

The case of (iii) for $c > 0$ is exactly similar to that for $c < 0$ since the argument detailed above is neutral as regards the sign of the integer. I have deliberately chosen the presumably less pretty sign in the exposition. $\qquad\square$

3. The probability setting

Let X be an integer-valued random variable. We say that the integer z is a "possible value" of X iff $P(X = z) > 0$. Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables with the common distribution of X, as follows:

$$P(X = z) = p_z, \qquad \text{for all integers } z. \tag{17}$$

Let $\{a_j\}$ denote the set of all possible values of X, so that $p_z > 0$ if and only if $z = a_j$ for some j. Put $p_{a_j} = q_j$.

Let $S_n = \sum_{k=1}^{n} X_k$. Then an integer c is a possible value of S_n if and only if there exist integers $m_j \geq 0$ satisfying

$$\sum m_j = n \qquad \text{and} \qquad \sum m_j a_j = c. \tag{18}$$

In fact, if we denote by $A_n(z)$ the number among the summands $X_1, X_2, \ldots, X_n$ taking the value z, we have almost surely $A_n(z) = 0$ unless z is a possible value of X, and therefore,

$$S_n = \sum A_n(a_j)a_j. \tag{19}$$

Now by independence and combinatorics, we have

$$P\big(A_n(a_j) = m_j \text{ for all } j\big) = n! \prod_j \frac{q_j^{m_j}}{m_j!}. \tag{20}$$

Therefore, we have

$$P(S_n = c) = \sum n! \prod_j \frac{q_j^{m_j}}{m_j!} \tag{21}$$

where the product is over all j, and the sum is over all $\{m_j\}$ with $m_j \geq 0$ satisfying the conditions in (18). Observe that an $m_j = 0$ contributes 1 to the product in (20) and plays only a notational role. The probability distribution exhibited in (21) is called the multinomial; when n is fixed, it is denoted by $M(n; \{q_j\})$.

Thus, the probability $P(S_n = c)$ is positive if and only if there is a solution in nonnegative integers $\{m_j\}$ of the two equations in (18) where the set $\{a_j\}$ is given. Theorem 1 may be restated in probabilistic terms as follows.

Theorem 1*. *In order that each integer c be a possible value of S_n for all sufficiently large n, it is necessary and sufficient that* (i) *there exist both positive and negative possible values of X, and* (ii) *the GCD of the set of all differences of the possible values of X be equal to 1.*

It is curious that, notwithstanding the appearance of the condition (ii) in the classical literature, e.g., [4, pp.232–233], and its apparent intent, I have not been able to locate a statement tantamount to Theorem 1*. But, of course, the old archives are vast and my scanning is limited. As for its algebraic form, Theorem 1, I wonder if it is known or utile in another context?

If (i) is true while the GCD in (ii) is the integer $d > 1$, then each integer c is the possible value of S_{nd} for all sufficiently large n. A celebrated case of this is the symmetric Bernoullian random walk in which $p_1 = p_{-1} = 1/2$ and $d = 2$.

4. Laplace's asymptotic and a lower estimate

Laplace's Theorem. *Let X and S_n be as in the above. Suppose that $E(X) = 0$ and $E(X^2) = v < \infty$. Then under conditions* (i) *and* (ii) *in §1, we have for each integer c,*

$$P(S_n = c) \sim (2\pi v n)^{-1/2} \qquad as \ n \to \infty. \tag{22}$$

This stemmed from a method of Laplace's that I remember as "steepest descent"; see [5, p.78]. A probabilistic presentation is given in [4, p.233]. I can no longer check these documents, but will refer the reader to a more accessible text [7]. Suffice it to say that the proof employs the Fourier transform, alias the characteristic function in probability theory. The result (together with some related results) is also given as an exercise in [2, p.177]. I recall that Erdős told me that he could prove it without any such gimmick which he sneered at in an offhand manner. I do not know if there is any record of his way of doing these things and should like to be informed.

Theorem 2. *Suppose* (iii) *in Theorem 1 holds, and that*

$$E(X^+) = E(X^-) \le +\infty \tag{23}$$

where $X^+ = \max\{X, 0\}$ and $X^- = \max\{-X, 0\}$. Then, for any ε in $(0,1)$, we have for any integer c:

$$\lim_{n \to \infty} P(S_n = c)(1 - \varepsilon)^{-n} = +\infty. \tag{24}$$

Proof. We may suppose that the set of possible values of X is unbounded, for otherwise, Theorem 2 follows from Laplace's. Hence, by considering X or $-X$, we may suppose that the set is unbounded from above.

We use the notation in (17). An arbitrary ε in (0,1) is given. There exists $N = N(\varepsilon)$ such that

$$\sum_{|n|>N} p_n < \varepsilon. \tag{25}$$

Under the hypothesis (23), there exists an $M > N$ such that

$$-\sum_{n=-M}^{0} n p_n > \sum_{n=1}^{N} n p_n. \tag{26}$$

Then there exists an $L > N$ such that

$$\sum_{n=1}^{L} np_n \geq - \sum_{n=-M}^{0} np_n > \sum_{n=1}^{L-1} np_n. \tag{27}$$

Hence, there exists p'_L such that

$$0 < p'_L \leq p_L \tag{28}$$

and

$$\sum_{n=-M}^{L-1} np_n + Lp'_L = 0. \tag{29}$$

Put

$$b = \sum_{n=-M}^{L-1} p_n + p'_L. \tag{30}$$

By the choice of N, M, and L, we have

$$b > 1 - \varepsilon. \tag{31}$$

Define

$$r_n = \begin{cases} b^{-1}p_n & \text{for } -M \leq n \leq L-1, \\ b^{-1}p'_L & \text{for } n = L, \\ 0 & \text{for } n < -M \text{ or } n > L. \end{cases} \tag{32}$$

Then we have, by (29) and (30),

$$\sum r_n = 1 \qquad \text{and} \qquad \sum nr_n = 0. \tag{33}$$

We define a random variable X' with the distribution

$$P(X' = n) = r_n \tag{34}$$

for all integers n. Then $|X'|$ is bounded by $\max\{M, L\}$ and

$$E(X') = 0. \tag{35}$$

Since X' is bounded, we have

$$E((X')^2) = v < +\infty. \tag{36}$$

Let $\{X'_n, n \geq 1\}$ be a sequence of independent random variables with the common distribution of X' indicated in (34), and let $S'_n = \sum_{k=1}^{n} X'_k$. In order that the contingency in Theorem 1* holds for S'_n as well as for S_n, the conditions (i) and (ii) should hold for the set of possible values of X' (as well as for that of X, which we have assumed). We need to go back to the choice of $N(\varepsilon)$ in (25) to make it large enough — we let the reader figure out this little amendment.

The distribution of S'_n is $M(n; \{r_j\})$. Thus $P(S'_n = c)$ is given by an expression similar to the right-hand member of (21) with the $q_j = p_{a_j}$ replaced by r_{a_j}, and the sum is subject to (18) exactly as for $P(S_n = c)$. Hence, by the definition of $\{r_n\}$ in (32), each term is not greater than b^{-n} times the corresponding term exhibited in (21); it may be less when one of the r_ns happens to be $b^{-1}p'_L$ and $p'_L < p_L$. Therefore, the sum of all these terms for $P(S'_n = c)$ is not greater than b^{-n} times the corresponding sum for $P(S_n = c)$. But the possible values of X' are a subset of the possible value of

X, by our construction. Hence, in general, there are solutions of (18) which are not available when the q_js are replaced by r_{a_j}s because the latter may be zero. In other words, in general, there are terms in the multinomial (21) that aren't there in the corresponding expression for $P(S_n' = c)$. This tedious and perhaps needless discussion is meant as a reminder of what is often left unsaid. The conclusion is

$$P(S_n = c) \geq b^n P(S_n' = c) > (1 - \varepsilon)^n P(S_n' = c). \tag{37}$$

By virtue of (35) and (36), Laplace's theorem is applicable to S_n', so that (22) holds when S_n is replaced by S_n'. Since $n^{-1/2}$ is of infinitely greater order of magnitude in comparison with $(1 - \varepsilon)^n$, (37) implies the assertion (24) after a tiny modification of ε. $\square$

Observe that the condition (23), optimally stated here, includes the case where $E(X) = 0$. In the latter case, Theorem 2 can be proved indirectly by using the recurrence theorem which yields

$$\sum_n P(S_n = c) = +\infty; \tag{38}$$

see [2, §8.3]. The proof given above was attributed to Shizuo Kakutani. He was a member of the small group gathered at Urbana, Illinois in the summer of 1950, about which time Erdős and I must have discussed Theorem 3 below. Simple as Theorem 2 may seem, no other proof, by Fourier transform or otherwise, has been found. Some time ago while contemplating this memorial, I wrote Wolfgang Fuchs to see if he could supply a Fourier argument. Several months later, I received a letter informing me that Fuchs had died.

The main result, Theorem 3 below, was originally considered under the hypothesis $E(X) = 0$. It was Erdős who saw that the condition $E(X^+) = E(X^-) = +\infty$ "works" as well. Theorem 2 was also stated under these two diverse hypotheses. Only a short while ago when I decided on a slow-paced redrafting of Theorem 2, did I realize their true import in the form given in (23). Observe that this condition is needed in Theorem 3 only via Theorem 2.

5. The main argument

From here on, we assume the hypotheses in both Theorem 1* and Theorem 2. Furthermore, let $\{a_j, 1 \leq j \leq k\}$ be a finite set of the possible values of X that satisfy the conditions (i) and (ii).

As early as 1713, Jakob Bernoulli published the following "incredibly" sharp inequality for his Law of Great Numbers.

Bernoulli's Theorem. *For any $\varepsilon > 0$, we have for any integer z:*

$$P\{|A_n(z) - np_z| \geq n\varepsilon\} \leq 2e^{-n\varepsilon^2/2} \tag{39}$$

(uniformly in all z).

For a proof, see [6, p.102ff]. I do not know if there is a modern exposition of the result; the "exponential decay" of the estimate is essential for our application. I was able to derive such an estimate by using the standard approximation of binomial coefficients (due apparently not to Stirling but to DeMoivre according to some historian).

As a consequence of the theorem above and Theorem 2, we can state the

Lemma. *For each integer c, we have asymptotically as $n \to \infty$:*

$$P(S_n = c) \sim P(S_n = c, \text{ and } |A_n(a_j) - nq_j| < n\varepsilon \text{ for } 1 \le j \le k). \quad (40)$$

To see this, note that the right-hand member of (39) is of the order of magnitude $(1 - f(\varepsilon))^n$ where $0 < f(\varepsilon) < 1$, but $f(\varepsilon)$ may vanish more rapidly than ε, for instance, $f(\varepsilon) = \varepsilon^3$. Now the other ε in (24) is arbitrary and, therefore, may be taken smaller than the present $f(\varepsilon)$. Then the lower estimate for $P(S_n = c)$, given by Theorem 2, dominates the upper estimate in (39), and (40) follows. Thus, in the asymptotic evaluation of $P(S_n = c)$, we may proceed as if the supplementary conditions in the second probability in (40) are all in force.

Take ε so that

$$\varepsilon < \min_{1 \le j \le k} q_j. \quad (41)$$

Theorem 3. *Assume the conditions in Theorem 2. Let c and c' be any two distinct integers. Then, as $n \to \infty$,*

$$P(S_n = c) \sim P(S_n = c'). \quad (42)$$

Proof. By the condition (ii) on $\{a_j, 1 \le j \le k\}$ and Euclid's theorem, there exist integers x_{ij} such that

$$c' - c = \sum_{i=1}^{k} \sum_{j=1}^{k} x_{ij}(a_i - a_j). \quad (43)$$

Note that the double entry of $a_i - a_j$ and $a_j - a_i$ causes no problem but affords symmetry, so that we obtain from (43) the desired representation

$$c' - c = \sum_{j=1}^{k} e_j a_j \qquad \text{with} \qquad \sum_{j=1}^{k} e_j = 0. \quad (44)$$

We leave it to the reader to derive the integers e_j from the x_{ij} in (43), and thus to verify the second equation in (44). The possibility of the representation of any $c' - c$, as shown in (44), is crucial in the proof of Theorem 3.

On second thought, the possibility of an expression like that in (44) for an arbitrary integer $z = c' - c$ is indeed an "obvious" consequence of Theorem 1* (or its equivalent Theorem 1). Apply (18) to both c and c', and then subtract!

Now consider the representation given in (18) where in the two sums the index j ranges over all positive integers, and let's not forget $m_j \ge 0$ for all j. To each such representation, we shall make correspond a similar representation in which c is replaced by c', as follows:

$$\sum m'_j = n \qquad \text{and} \qquad \sum m'_j a_j = c'. \quad (45)$$

This is done by means of (44); specifically, we put

$$m'_j = \begin{cases} m_j & \text{for } j > k, \\ m_j + e_j & \text{for } 1 \le j \le k. \end{cases} \quad (46)$$

In order that all these integers m'_j be nonnegative, it suffices that the m_j's be large because the e_j's are fixed integers.

Now the restrictions placed on $A_n(a_j)$ in the second probability in (40) makes it greater than $(q_j - \varepsilon)n$ which is positive for all large n owing to (41). Hence, for all

j, when n is large enough, all m'_j defined in (46) are "realizable" as $A_n(a_j)$ to yield $S_n = c'$ according to (45). The probability of such an event is given as in (20) when m_j is replaced by m'_j, namely,

$$n! \prod_j \frac{q_j^{m'_j}}{m'_j!}. \tag{47}$$

When we sum these probabilities over all $\{m'_j\}$ satisfying (45), we obtain a number not greater than $P(S_n = c')$ because, in general, there are solutions $\{m'_j\}$ of the equations in (45) that do not correspond to the solutions $\{m_j\}$ of the equations in (18) in the manner prescribed by (46). However, we shall see that this does not matter.

We now must compare the two numbers given in (20) and (47). Take the ratio

$$\prod_{j=1}^k \frac{m'_j!}{m_j!} q_j^{m_j - m'_j}. \tag{48}$$

Here, in the product, the index ranges from 1 to k because for $j > k$ we have $m'_j - m_j = 0$. If $e_j \geq 0$, then the number in (48) does not exceed

$$(m_j + e_j) \cdots (m_j + 1) q_j^{-e_j} \leq \left(\frac{m_j + e_j}{q_j} \right)^{e_j}. \tag{49}$$

If $e_j < 0$, the said number does not exceed

$$\frac{q_j^{-e_j}}{m_j(m_j - 1) \cdots (m_j + e_j + 1)} \leq \left(\frac{q_j}{m_j + e_j} \right)^{|e_j|}. \tag{50}$$

Under the restrictions

$$(q_i + \varepsilon)n > m_j > (q_i - \varepsilon)n, \tag{51}$$

we have, since e_j is a constant and whether $e_j \geq 0$ or $e_j < 0$,

$$(q_j + \varepsilon)n > m_j + e_j > (q_i - \varepsilon)n \tag{52}$$

with indeed the same ε, provided that n is large enough, which we do assume. Then, the right-hand members of (49) and (50) are both not greater than

$$\left(\frac{q_j}{q_j - \varepsilon} \right)^{|e_j|} n^{e_j} \tag{53}$$

where we have made a minor and unnecessary simplification by using $(q_j + \varepsilon)(q_j - \varepsilon) < q_j^2$. The crucial point is the exponent of n there. Therefore, when we take the product of (53) from $j = 1$ to $j = k$, we obtain the inequality

$$P\big(A_n(a_j) = m_j \text{ for } j \geq 1\big) \leq \prod_{j=1}^k \left(\frac{q_j}{q_j - \varepsilon} \right)^{|e_j|} P\big(A_n(a_j) = m'_j \text{ for } j \geq 1\big) \tag{54}$$

because by (44)

$$\prod_{j=1}^k n^{e_j} = n^0 = 1. \tag{55}$$

Summing (54) over $\{m_j\}$ satisfying (18) and (51), with $\{m'_j\}$ determined by (46), we obtain, for all large values of n, the inequality

$$P\big(S_n = c; |A_n(a_j) - nq_j| < n\varepsilon \text{ for } 1 \le j \le k\big)$$

$$\le \prod_{j=1}^{k} \left(\frac{q_j}{q_j - \varepsilon}\right)^{|e_j|} P(S_n = c'). \tag{56}$$

Together with the previous relation (40), we conclude, since ε is arbitrary, that

$$\limsup_{n \to \infty} \frac{P(S_n = c)}{P(S_n = c')} \le 1. \tag{57}$$

Since the two integers c and c' can be exchanged, we have in fact proved the assertion (42). $\qquad\square$

Theorem 3 plus an argument of the kind used in its proof establishes the following companion.

Theorem 4. *For each integer c, we have*

$$\lim_{n \to \infty} \frac{P(S_{n+1} = c)}{P(S_n = c)} = 1.$$

It is odd that Theorem 4 is not deducible directly from Theorem 3 by using

$$P(S_{n+1} = c) = \sum_j q_j P(S_n = c - a_j),$$

for lack of dominated convergence!

6. Fourier analysis in the symmetric case

Denote the ch. f. of X by

$$f(t) \equiv E\big(e^{itX}\big) = \sum_{z=-\infty}^{\infty} p_z e^{itz}, \qquad -\infty < t < +\infty. \tag{58}$$

When X is symmetric, $p_z = p_{-z}$ for all z, and so

$$f(t) = p_0 + 2 \sum_{z=1}^{\infty} p_z \cos tz. \tag{59}$$

Thus f is real-valued, symmetric, continuous, and periodic with period at most 2π; the ch. f. of S_n is

$$E\big(e^{itS_n}\big) = \prod_{k=1}^{n} E\big(e^{itX_k}\big) = f(t)^n. \tag{60}$$

We have by the Fourier inversion formula, for each integer c,

$$P(S_n = c) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)^n e^{-itc} dt$$

$$= \frac{1}{\pi} \int_{0}^{\pi} f(t)^n \cos tc \, dt. \tag{61}$$

Put

$$t_0 = \min\{t > 0 : |f(t)| = 1\}, \tag{62}$$

$$d_0 = GCD\{a_j - a_i\}; \tag{63}$$

see §1 and §2 for notation. Then we have

$$t_0 = \frac{2\pi}{d_0}. \tag{64}$$

This follows easily from (58); see [4, p.175 and p.176, Exercise 21]. Under the condition (ii) in Theorem 1*, $d_0 = 1, t_0 = 2\pi$. Since $f(0) = 1$, it follows by continuity that for each $\delta \in (0, \pi)$, we have

$$\max_{\delta \leq t \leq \pi} |f(t)| = \theta(\delta) < 1. \tag{65}$$

Now choose the δ above so that $f(t) > 0$ for $t \in [0, \delta]$, then we have by (65)

$$\left| \int_\delta^\pi f(t)^n \cos tc\, dt \right| \leq \pi \theta(\delta)^n. \tag{66}$$

For a symmetric X, the condition (23) holds and so by Theorem 2,

$$\theta(\delta)^n = o(P(S_n = c)). \tag{67}$$

Let us observe that the proof of Theorem 2 in the symmetric case is much simpler than given there in the general case; we need only truncate X at $\pm N$ for a sufficiently large N.

It follows from (61), (66), and (67) that

$$P(S_n = c) \sim \frac{1}{\pi} \int_0^\delta f(t)^n \cos tc\, dt. \tag{68}$$

Applying (61) and (66) , we obtain

$$|P(S_n = 0) - P(S_n = c)| \leq \frac{1}{\pi} \int_0^\delta f(t)^n [1 - \cos tc] dt + 2\theta(\delta)^n. \tag{69}$$

For any $\varepsilon > 0$, we may choose $\delta = \delta(\varepsilon)$, so that

$$0 \leq 1 - \cos tc \leq \varepsilon \qquad \text{for} \qquad 0 \leq t \leq \delta. \tag{70}$$

With this choice, we have by (69), and (67), (68) with $c = 0$:

$$|P(S_n = 0) - P(S_n = c)| \leq \varepsilon \frac{1}{\pi} \int_0^\delta f(t)^n dt + o(P(S_n = 0))$$
$$= (\varepsilon + o(1))P(S_n = 0). \tag{71}$$

Thus (42) follows.

For a "purely analytic" solution of the problem (Theorem 3), an announcement was published in Math. Scand **2** (1954), p.345. The problem remains open as of July 18, 1997.

Editor's note.

In 1992, Professor K. L. Chung and I were invited to a summer school for graduate students in Dalian, China, which was organised by the State Education Commission. Professor Chung and I had not met previously, and soon after the start of the summer school, he learned that I was an asymptotic analyst.

He told me that he had a mathematical problem which was essentially of an asymptotic nature and had been open since 1949. Although he had mentioned this problem to several well-known analysts at that time, it was Paul Erdős who helped him solve the problem by using a probability argument. Together they published the solution to this problem in the Memoirs of the American Mathematical Society in 1951. Since Professor Chung felt that the problem was really an analysis problem, he posted it in the Problem Section of Mathematica Scandinavica in 1954, asking for an analytic solution.

The interesting history of the problem led me to pursue this problem. Unfortunately, I also was unable to come up with a solution using analysis. Several of my colleagues and postdoctoral fellows read the paper by Erdős and Chung to see whether it would be possible to translate the probability argument into an analysis argument. However, they found the paper difficult to read because of typographical errors and some incomplete arguments.

About eight months ago, Professor Chung wrote to me and indicated that he wanted to do a new exposition of the results, emphasizing the multinomial expansion in the use of probability. I encouraged Professor Chung to write up his new findings, for publication in Methods and Applications of Analysis. On July 18 this year, Professor Chung handed me the manuscript which two analysts and I now have read carefully. Although we found the argument in this paper essentially similar to that given in his original paper with Erdős, many more details have been provided. Since the problem is very interesting and challenging from an asymptotic point of view, and still awaits an analytic solution, the paper has been accepted for publication with the hope that it will motivate more analysts to look at the problem.

— R. S. C. Wong, January, 1998

References

1. K. L. Chung, *Fluctuations of sums of independent random variables*, Ann. Math. **51** (1950), 697–706.
2. _____, *A Course in Probability Theory*, Second edition, Academic Press, 1974.
3. K. L. Chung and P. Erdős, *Probability limit theorems assuming only the first moment I*, Memoirs A.M.S. **6** (1950), 1–19.
4. B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, (English translation by K. L. Chung), Second revised edition, Addison-Wesley, 1968.

5. G. Pólya and G. Szegő, *Aufgaben und Lehrsätze aus der Analysis.* Volume I, Dover Publications, 1945.
6. J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, 1937.
7. R. Wong, *Asymptotic Approximations of Integrals*, Academic Press, 1989.

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, STANFORD, CA, U.S.A.