ON A NEW LAW OF LARGE NUMBERS

By

PAUL ERDÖS AND ALFRED RÉNYI

in Budapest, Hungary

§1. Introduction

We shall prove first (in §2) the new law of large numbers for the simplest special case, that is for independent repetitions of a fair game. For this special case the theorem can be stated as follows: if the game is played N times, the maximal average gain of a player over $[C \log_2 N]$ consecutive games* ($C \geqq 1$), tends with probability one to the limit α, where α is the only solution in the interval $0 < \alpha \leqq 1$ of the equation

$$\frac{1}{C} = 1 - \left(\frac{1+\alpha}{2}\right)\log_2\left(\frac{2}{1+\alpha}\right) - \left(\frac{1-\alpha}{2}\right)\log_2\left(\frac{2}{1-\alpha}\right).$$

In §3 we generalize this result to an arbitrary sequence η_n ($n = 1, 2, \cdots$) of independent, identically distributed random variables with expectation 0, the common distribution of which satisfies the condition, that its moment-generating function $\phi(t) = E(e^{\eta_n t})$ exists in an open interval around the origin. We prove that for every α in a certain interval $0 < \alpha < \alpha_0$ one has

$$\text{(1.1)} \quad P\left(\lim_{N \to +\infty} \max_{0 \leqq n \leqq N - [C \log N]} \frac{\eta_{n+1} + \eta_{n+2} + \cdots + \eta_{n+[C \log N]}}{[C \log N]} = \alpha\right) = 1,$$

where $C = C(\alpha)$ is defined by the equation

$$\text{(1.2)} \qquad e^{-(1/C)} = \min_t \phi(t)\, e^{-\alpha t} .$$

* Here and in what follows $[x]$ denotes the integral part of x.

In § 4 we discuss the special case of Gaussian random variables, in which case our result is essential.y equivalent to a previous result of *Paul Lévy* about the Brownian movement process.

In § 5 we give as an application of the result of § 3, a new proof of the theorem of *P. Bártfai* on the "stochastic geyser problem", using the fact that the functional dependence between C and α in (1.1) determines the distribution of the variables uniquely (Theorem 3). The result of § 2 can also be applied in probabilistic number theory; as a matter of fact it was such an application which led the first named author to raise the problem which is solved in the present paper.

§2. The maximal average gain of a player over a short period.

Let $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ be a sequence of independent random variables, each taking on the values ± 1 with probability $1/2$. We may interpret ξ_n as the gain of one of the players in the n^{th} repetition of a fair game. Let us put $S_0 = 0$,

$$S_n = \xi_1 + \xi_2 + \cdots + \xi_n \qquad (n = 1, 2, \cdots) \tag{2.1}$$

and

$$\vartheta(N, K) = \max_{0 \leqq n \leqq N-K} \frac{S_{n+K} - S_n}{K}. \tag{2.2}$$

Let us introduce the notation

$$h(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x} \quad \text{for } 0 < x < 1; \tag{2.3}$$

i.e. $h(x)$ is the entropy of the probability distribution $(x, 1-x)$. We shall prove the following

Theorem 1. *For every fixed $c \geqq 1$ we have**

$$P(\lim_{N \to +\infty} \vartheta(N, [c \log_2 N]) = \alpha) = 1, \tag{2.4}$$

* Here and what follows $P(\ldots.)$ denotes the probability of the event in the brackets.

where $\alpha = \alpha(c)$ *is the only solution with* $0 < \alpha \leqq 1$ *of the equation*

$$\frac{1}{c} = 1 - h\left(\frac{1+\alpha}{2}\right). \tag{2.5}$$

Remark. It is easy to see that $\alpha(c)$ is a decreasing function of c, further $\alpha(1) = 1$ and $\lim_{c \to +\infty} \alpha(c) = 0$.

Proof of Theorem 1. We shall use the following estimates, which follow immediately from Stirling's formula: If $\frac{1}{2} \leqq \gamma < 1$

$$A_1 \cdot n^{-1/2} \cdot 2^{n(h(\gamma)-1)} \leqq 2^{-n} \sum_{n\gamma \leqq K \leqq n} \binom{n}{K} \leqq B_1 \cdot n^{-1/2} \cdot 2^{n(h(\gamma)-1)} \tag{2.6}$$

where A_1 and B_1 are positive constants, depending only on γ. Let $c \geqq 1$ be fixed, and let α be the unique solution of the equation (2.5) with $0 < \alpha \leqq 1$. Let ε be an arbitrary small positive number and put $\alpha' = \alpha + \varepsilon$. It follows from (2.6) that

$$P(\vartheta(N, [c \log_2 N]) \geqq \alpha') \leqq B_1 N^{-\delta_1} \tag{2.7}$$

where δ_1 is a positive number, depending only on α and ε. Thus the series

$$\sum_{j=1}^{+\infty} P(\vartheta(2^{(j+1)/c} - 1, j) \geqq \alpha') \tag{2.8}$$

is convergent, and therefore by the Borel-Cantelli lemma one has

$$\vartheta(2^{(j+1)/c} - 1, j) < \alpha' \tag{2.9}$$

with probability 1 for all but a finite number of values of j. As however

$$\vartheta(N, [c \log_2 N]) \leqq \vartheta(2^{(j+1)/c} - 1, j) \text{ for } 2^{j/c} \leqq N \leqq 2^{(j+1)/c} - 1, \tag{2.10}$$

it follows that with probability one, for all but a finite number of values of N one has

$$\vartheta(N, [c \log_2 N]) < \alpha' . \tag{2.11}$$

As $\varepsilon > 0$ is arbitrary, we obtain

$$P(\limsup_{N\to+\infty} \vartheta(N, [c \log_2 N]) \leqq \alpha) = 1 . \tag{2.12}$$

Now let again ε be an arbitrary small positive number, $0 < \varepsilon < \alpha$ and put $\alpha'' = \alpha - \varepsilon$. As

$$P(\vartheta(N, K) \leqq \alpha'') \leqq P\left(\frac{S_{(r+1)K} - S_{rK}}{K} \leqq \alpha'', 0 \leqq r \leqq \frac{N}{K} - 1\right) \tag{2.13}$$

and because of the independence of the random variables $S_{(r+1)K} - S_{rk}$ $(r = 0, 1, \cdots)$ it follows that

$$P(\vartheta(N, [c \log_2 N]) \leqq \alpha'') \leqq \left(1 - \frac{A_1 N^{\delta_2}}{N}\right)^{N/([c\log_2 N]) - 1} \leqq e^{-(A_2 N^{\delta_2})/\log N} \tag{2.14}$$

where A_2 and δ_2 are positive constants. Thus the series

$$\sum_{N=1}^{\infty} P(\vartheta(N, [c \log_2 N]) \leqq \alpha'') \tag{2.15}$$

is convergent and using again the Borel-Cantelli lemma we get

$$P \liminf_{N\to+\infty} \vartheta(N, [c \log_2 N]) \geqq \alpha) = 1. \tag{2.16}$$

As (2.12) and (2.16) imply (2.4), Theorem 1 is proved.

It should be remarked, that the same argument as that used to prove (2.12) can be used to show that if $K(N)$ is an integer-valued function of N such that $\frac{K(N)}{\log N} \to +\infty$ we have

$$P\left(\lim_{N\to+\infty} \vartheta(N, K(N)) = 0\right) = 1). \tag{2.17}$$

This result can be interpreted as follows: if $K(N)$ grows faster than $\log N$, then the ordinary law of large numbers applies. On the other hand if $K(N) \leqq c \log_2 N$ with $0 < c < 1$ then with probability 1 for all except for a finite number of values of N there exists at least one $n \leqq N - K(N)$ such that $\xi_{n+1} = \xi_{n+2} = \cdots = \xi_{n+K(N)} = 1$, which of course implies $\vartheta(N, K(N)) = 1$. Thus the case of real interest is just when $K(N) \sim c \log_2 N$ with $c \geqq 1$, and Theorem 1 gives an answer to the question what happens in this case.

§3. The general case.

We shall prove now the following

Theorem 2. *Let $\eta_1, \eta_2, \cdots, \eta_n, \cdots$, be a sequence of independent, identically distributed nondegenerate random variables. We suppose that the moment generating function*

$$(3.1) \qquad \phi(t) = E(e^{t\eta_n})$$

of the common distribution of the η_n exists for $t \in I$ where I is an open interval** containing $t = 0$. Let us suppose that*

$$(3.2) \qquad E(\eta_n) = 0.$$

Let α be any positive number such that the function $\phi(t)e^{-\alpha t}$ takes on its minimum in some point in the open interval I and let us put

$$(3.3) \qquad \min_{t \in I} \phi(t)e^{-\alpha t} = \phi(\tau)e^{-\alpha\tau} = e^{-(1/C)}.$$

Then $C > 0$ and putting $S_0 = 0$,

$$(3.4) \qquad S_n = \eta_1 + \eta_2 + \cdots + \eta_n \qquad \text{for } n \geqq 1$$

and

$$(3.5) \qquad \vartheta(N, K) = \max_{0 \leqq n \leqq N-K} \frac{S_{n+K} - S_n}{K} \qquad (1 \leqq K \leqq N),$$

*we have****

$$(3.6) \qquad P\left(\lim_{N \to +\infty} \vartheta(N, [C \log N]) = \alpha\right) = 1.$$

* $E(\ldots.)$ denotes the expectation of the random variable in the brackets.
** We suppose that I is the *largest* open interval in which $\phi(t)$ exists.
*** In this and the following §§ $\log N$ denotes the natural logarithm of N.

Proof of Theorem 2. Let us notice first that $\psi(t) = \phi(t)e^{-\alpha t}$ is a strictly convex function: thus τ in (3.3) is determined uniquely. As clearly $\psi(0) = 1$ and in view of (3.2) $\psi'(0) = -\alpha < 0$ it follows that $\tau > 0$ and $\psi(\tau) < 1$ and thus $C > 0$. Let us mention that the condition that $\psi(t)$ takes on its minimum in the interval I is satisfied if for instance $P(\eta_n > \alpha) > 0$ because in this case $\psi(t)$ tends to $+\infty$ if t tends to the upper endpoint of I (which may be the point $+\infty$). We have evidently

$$\frac{\phi'(\tau)}{\phi(\tau)} = \alpha.$$

The proof of Theorem 2 follows exactly that of Theorem 1, only instead of (2.6) we have to use the following result, which under some restrictions is due to *H. Cramér* (see [1]), and in the form needed for our purpose is due to *R. R. Bahadur* and *R. Ranga Rao* (see [2], Theorem 1):

$$P(S_n > \alpha n) = \frac{e^{-(n/C)}}{\sqrt{2\pi n}} b_n \cdot (1 + o(1)) \tag{3.7}$$

where b_n is a sequence of positive numbers such that $0 < b \leqq b_n \leqq B$; if the η_n are not lattice variables, b_n does not depend on n.

Remark. In the special case when $P(\eta_n = \pm 1) = 1/2$, we have $\phi(t) = \frac{1}{2}(e^t + e^{-t})$ therefore if $0 < \alpha < 1$ $\tau = \frac{1}{2}\log\frac{1+\alpha}{1-\alpha}$ and $\frac{1}{C} = \frac{1+\alpha}{2}\log(1+\alpha) + \frac{1-\alpha}{2}\log(1-\alpha)$. Passing to logarithms with base 2 it is easily seen that $e^{-(1/C)} = 2^{h((1+\alpha)/2)-1} = 2^{-(1/c)}$ i.e. $c = C\log 2$. Thus the statement of Theorem 1 for $c > 1$ is contained as a special case in Theorem 2.

§4. The Gaussian case.

Let us consider the special case in which the random variables have a normal distribution with mean 0 and variance 1. (In this case of course S_n is also normally distributed and we do not even need the result (3.7).) As regards the connection between C and α this can be explicitly determined in this special case: we have evidently for every $\alpha > 0$ $C = \frac{2}{\alpha^2}$, and thus we get from (3.6)

$$(4.1) \qquad P(\lim_{N \to +\infty} \vartheta(N, [C \log N] = \sqrt{\frac{2}{C}} = 1 \text{ for every } C > 0.$$

From (4.1) one can deduce the following remarkable theorem, due to *P. Lévy* (see [3]): Let $x(t)$ be a Brownian movement process, then

$$(4.2) \lim_{h \to 0} P\left(\left| x(t+h) - x(t) \right| < \lambda \sqrt{2h \log \frac{1}{h}} \text{ for } 0 \leqq t \leqq 1 - h \right) = \begin{cases} 1 \text{ if } \lambda > 1 \\ 0 \text{ if } \lambda < 1. \end{cases}$$

Notice that if the variance of the random variables η_n is equal to 1 then we have in general for $\alpha \to 0$ $C \sim 2/\alpha^2$; as a matter of fact we have for $t \to 0$ $\phi'(t) \sim t$ and thus for $\alpha \to 0$ we get $\tau \sim \alpha$ and therefore $C \sim 2/\alpha^2$. Thus for very small values of α the relation between α and C in Theorem 2 becomes in the limit independent from the distribution of the variables η_n; however for a fixed not too small value of α the functional relation between α and C depends essentially on the distribution of the random variables η_n. Clearly the reason why the relation between α and C in Theorem 2 depends on the distribution of the variables η_n, is that Theorem 2 is a theorem about big deviations, while the reason for the disappearance of this dependence in the limit if $\alpha \to 0$ is that if α is decreasing we approach the domain of validity of the central limit theorem.

§5. **An application.**

Let η_n $(n = 1, 2, \cdots)$ be a sequence of independent and identically distributed random variables and let $F(x)$ denote their common distribution function. Let us put

$$(5.1) \qquad \xi_n = S_n + r_n$$

where S_n is defined by (3.4) and r_n $(n = 1, 2, \cdots)$ is an arbitrary sequence of bounded random variables such that

$$(5.2) \qquad |r_n| \leqq R_n \quad \text{where} \quad R_n = o(\log n)$$

(Nothing is supposed concerning the dependence between the variables S_n and r_n). *P. Bártfai* has proved (see [4]) that if the moment generating function

$$\phi(t) = \int_{-\infty}^{+\infty} e^{tx}\, dF(x) \tag{5.3}$$

of the variables η_n exists in a neighbourhood of $t = 0$, then given the values ξ_n $(n = 1, 2, \cdots)$ the distribution function $F(x)$ is thereby uniquely determined with probability one. A new proof of this result of *Bártfai* can be obtained from Theorem 2 as follows: We may suppose without restricting the generality that $E(\eta_n) = 0$; in this case all conditions of Theorem 2 are satisfied and thus it follows that for $0 < \alpha < a$ where a is a sufficiently small positive number we have (in view of (5.2)) with probability one

$$\lim_{N \to +\infty} \left(\max_{0 \leqq n \leqq N - [c \log N]} \frac{\zeta_{n+[c \log N]} - \zeta_n}{[c \log N]} \right) = \alpha. \tag{5.4}$$

Thus knowing the sequence ζ_n we can determine the functional dependence between α and c.

To prove *Bártfai*'s theorem we shall need the following

Theorem 3. *The functional dependence between α and $c = c(\alpha)$ in Theorem 2 determines the distribution of the random variables η_n uniquely.*

Proof. If the function $c = c(\alpha)$ is given for $0 < \alpha < a$, we can determine the function

$$\lambda(\alpha) = e^{-(1/c(\alpha))} \tag{5.5}$$

and thus also the function

$$\frac{\lambda'(\alpha)}{\lambda(\alpha)} = -\tau . \tag{5.6}$$

As clearly $\tau = \tau(\alpha)$ is an increasing function of α, its inverse function $\alpha = \alpha(\tau)$ can also be determined. This means however that we can determine the function

$$\phi(\tau) = \alpha(\alpha(\tau))e^{\tau\alpha(\tau)} \tag{5.7}$$

in some interval $0 \leqq \tau \leqq \tau_0$. As it is well known that the moment-generating function $\phi(t)$ determines the distribution function $F(x)$ uniquely, (even if $\phi(t)$ is given only in some interval, it being an analytic function if it exists), the statement of Theorem 3 follows.

It follows from Theorem 3 that in the stochastic geyser problem if we know a single realization of the sequence ζ_n $(n = 1, 2, \cdots)$ we can determine the distribution function $F(x)$ with probability one; this proves *Bártfai's* theorem.

References

1. H. Cramér, *Sur un nouveau théorème-limite de la théorie des probabilités*. Actualités Scientifiques et Industrielles, No 736, Hermann et Cie, Paris, 1938.

2. R. R. Bahadur and R. Ranga Rao, On deviations of the sample mean, *Annals of Mathematical Statistics*, **31** (1960), 1015–1027.

3. P. Lévy, *Théorie de l'addition des variables aléatoires indépendantes*, Paris, Gauthier-Villars.

4. P. Bártfai, Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. *Studia Scientiarum Mathematicarum Hungarica*, **1** (1966), 161–168.

Mathematical Institute
Hungarian Academy of Sciences
Budapest, Hungary