ON THE LENGTH OF THE LONGEST HEAD-RUN

P. ERDŐS – P. RÉVÉSZ

1§. INTRODUCTION

Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables with $\mathsf{P}(X_1 = 0) = \mathsf{P}(X_1 = 1) = \frac{1}{2}$ and let $S_0 = 0$, $S_n = X_1 + X_2 + \dots + X_n$ $(n = 1, 2, \dots)$ and

$$I(N, K) = \max_{0 \leq n \leq N-K} (S_{n+K} - S_n) \qquad (N \geq K).$$

Define the r.v.'s Z_N $(N = 1, 2, \dots)$ as follows: let Z_N be the largest integer for which

$$I(N, Z_N) = Z_N.$$

This Z_N is the length of the longest head-run. Studying the properties Z_N resp. $I(N, K)$ Erdős and Rényi proved the following:

Theorem A. ([1]) *Let* $0 < C_1 < 1 < C_2 < \infty$ *then for almost all* $\omega \in \Omega$ *(Ω is the basic space) there exists a finite* $N_0 = N_0(\omega, C_1, C_2)$

*such that**

$$[C_1 \log N] \leqslant Z_N \leqslant [C_2 \log N]$$

if $N \geqslant N_0$.

The aim of this paper is to get sharper bounds of Z_N. In connection with this problem our first result is

Theorem 1. *Let* ϵ *be any positive number. Then for almost all* $\omega \in \Omega$ *there exists a finite* $N_0 = N_0(\omega, \epsilon)$ *such that*

$$Z_N \geqslant [\log N - \log\log\log N + \log\log e - 2 - \epsilon] = \alpha_1(N) = \alpha_1$$

if $N \geqslant N_0$.

This result is quite near to the best possible one in the following sense:

Theorem 2. *Let* ϵ *be any positive number. Then for almost all* $\omega \in \Omega$ *there exists an infinite sequence* $N_i = N_i(\omega, \epsilon)$ $(i = 1, 2, \ldots)$ *of integers such that*

$$Z_{N_i} < [\log N_i - \log\log\log N_i + \log\log e - 1 + \epsilon] = \alpha_2(N) = \alpha_2.$$

Theorems 1 and 2 together say that the length of the longest head-run is larger than α_1 but in general not larger than α_2. Clearly enough for some N the length of the longest head-run can be much larger than α_2. In our next theorems the largest possible values of Z_N are investigated.

Theorem 3. *Let* $\{\gamma_n\}$ *be a sequence of positive numbers for which*

$$\sum_{n=1}^{\infty} 2^{-\gamma_n} = \infty.$$

Then for almost all $\omega \in \Omega$ *there exists an infinite sequence* $N_i = N_i(\omega, \{\gamma_n\})$ $(i = 1, 2, \ldots)$ *of integers such that*

$$Z_{N_i} \geqslant \gamma_{N_i}.$$

This result is the best possible in the following sense:

*Here and in what follows log means logarithm with base 2; $[x]$ is the integral part of x.

Theorem 4. *Let* $\{\delta_n\}$ *be a sequence of positive numbers for which*

$$\sum_{n=1}^{\infty} 2^{-\delta_n} < \infty.$$

Then for almost all $\omega \in \Omega$ *there exists a positive integer* $N_0 = N_0(\omega, \{\delta_n\})$ *such that*

$$Z_N < \delta_N$$

if $N \geqslant N_0$.

Theorems 1-4 are characterizing the length of the longest run containing no tail at all. One can ask about the length of the longest run containing at most T tails. In order to formulate our results precisely introduce the following notation: Let $Z_N(T)$ be the largest integer for which

$$I(N, Z_N(T)) \geqslant Z_N(T) - T.$$

This $Z_N(T)$ is the length of the longest run containing at most T tails.

Our Theorems 1-4 can be easily generalized for this case as follows:

Theorem 1*. *Let* ϵ *be any positive number. Then for almost all* $\omega \in \Omega$ *there exists a finite* $N_0 = N_0(\omega, T, \epsilon)$ *such that*

$$Z_N(T) \geqslant [\log N + T \log\log N - \log\log\log N - \log T! +$$
$$+ \log\log e - 2 - \epsilon] = \alpha_1(N, T)$$

if $N \geqslant N_0$.

Theorem 2*. *Let* ϵ *be any positive number. Then for almost all* $\omega \in \Omega$ *there exists an infinite sequence* $N_i = N_i(\omega, T, \epsilon)$ *of integers such that*

$$Z_{N_i}(T) < \alpha_2(N_i, T) =$$
$$= [\log N_i + T \log\log N_i - \log\log\log N_i - \log T! + \log\log e -$$
$$- 1 + \epsilon].$$

Theorem 3*. *Let* $\{\gamma_n\}$ *be a sequence of positive integers for which*

$$\sum_{n=1}^{\infty} \gamma_n^T 2^{-\gamma_n} = \infty.$$

Then for almost all $\omega \in \Omega$ *there exists an infinite sequence* $N_i = N_i(\omega, T, \{\gamma_n\})$ *of integers such that*

$$Z_{N_i}(T) \geqslant \gamma_{N_i}.$$

Theorem 4*. *Let* $\{\delta_n\}$ *be a sequence of positive integers for which*

$$\sum_{n=1}^{\infty} \delta_n^T 2^{-\delta_n} < \infty.$$

Then for almost all $\omega \in \Omega$ *there exists a positive integer* $N_0 = N_0(\omega, T, \{\delta_n\})$ *such that*

$$Z_N(T) < \delta_N$$

if $N \geqslant N_0$.

The last two Theorems clearly can be reformulated as follows:

Theorem 3.** *Let* $\{\gamma_n\}$ *be a sequence of positive integers for which*

$$\sum_{n=1}^{\infty} \gamma_n^T 2^{-\gamma_n} = \infty.$$

Then for almost all $\omega \in \Omega$ *there exists a sequence* $N_i = N_i(\omega, \{\gamma_n\})$ *of integers such that*

$$S_{N_i} - S_{N_i - \gamma_{N_i}} \geqslant \gamma_{N_i} - T.$$

Theorem 4.** *Let* $\{\delta_n\}$ *be a sequence of positive integers for which*

$$\sum_{n=1}^{\infty} \delta_n^T 2^{-\delta_n} < \infty.$$

Then for almost all $\omega \in \Omega$ *there exists a positive integer* $N_0 = N_0(\omega, T, \{\delta_n\})$ *such that*

$$S_N - S_{N-\delta_N} < \delta_N - T$$

if $N \geqslant N_0$.

2§. A THEOREM ON THE DISTRIBUTION OF $I(N, K)$

The proofs of Theorems 1-4 are based on the following

Theorem 5. *We have*

$$\left(1 - 2^{-K-1} \frac{K^{T+1}}{T!} (1 + o_K(1))\right)^{[\frac{N-2K}{K}]+1} \leqslant$$

$$\leqslant \mathsf{P}(I(N, K) < K - T) \leqslant$$

$$\leqslant \left(1 - 2^{-K-1} \frac{K^{T+1}}{T!} (1 + o_K(1))\right)^{[\frac{1}{2}[\frac{N-2K}{K}]]+1}$$

if $N \geqslant 2K$.

Before the proof of this Theorem we prove our

Lemma 1. *We have*

$$\mathsf{P}(I(2N, N) \geqslant N - T) = \begin{cases} 2^{-N-1}(N+2) & \text{if } T = 0, \\ 2^{-N-1}(N^2 + 4 - 2^{-N+1}) & \text{if } T = 1, \\ 2^{-N-1} \frac{N^{T+1}}{T!}(1 + o(1)) & \text{if } T > 1. \end{cases}$$

Proof. Let

$$A = A(T) = \{I(2N, N) \geqslant N - T\},$$

$$A_k = A_k(T) = \{S_{k+N} - S_k \geqslant N - T\} \qquad (k = 0, 1, 2, \ldots, N),$$

and

$$S_{-j} = -\infty \qquad (j = 1, 2, \ldots).$$

Then we clearly have

$$A = A_0 + \bar{A}_0 A_1 + \bar{A}_0 \bar{A}_1 A_2 + \ldots + \bar{A}_0 \bar{A}_1 \ldots \bar{A}_{N-1} A_N$$

where

$$\mathsf{P}(A_0) = \sum_{j=0}^{T} \binom{N}{j} 2^{-N},$$

and

$$p_k = \mathsf{P}(\bar{A}_0 \bar{A}_1 \ldots \bar{A}_{k-1} A_k) =$$

$$= \sum_{k+1 \leqslant l_1 < l_2 < \ldots < l_T \leqslant k+N} \mathsf{P}(\bar{A}_0 \bar{A}_1 \ldots \bar{A}_{k-1} A_k,$$

$$X_k = X_{l_1} = X_{l_2} = \ldots = X_{l_T} = 0) =$$

$$= \sum_{k+1 \leqslant l_1 < l_2 < \ldots < l_T \leqslant k+N} \mathsf{P}(A_k, X_k = X_{l_1} = X_{l_2} = \ldots$$

$$\ldots = X_{l_T} = 0,\ S_{k-1} - S_{l_T - N - 1} < k - l_T + N,$$

$$S_{k-1} - S_{l_{T-1} - N - 1} < k - l_{T-1} + N - 1, \ldots$$

$$\ldots, S_{k-1} - S_{l_1 - N - 1} < k - l_1 + N - (T-1)) =$$

$$= 2^{-N-1} \sum_{k+1 \leqslant l_1 < l_2 < \ldots < l_T \leqslant k+N} \mathsf{P}(S_{k-1} - S_{l_T - N - 1} <$$

$$< k - l_T + N,\ S_{k-1} - S_{l_{T-1} - N - 1} < k - l_{T-1} + N - 1, \ldots$$

$$\ldots, S_{k-1} - S_{l_1 - N - 1} < k - l_1 + N - (T-1)).$$

Especially if

(i) $T = 0$ then $p_k = 2^{-N-1}$

(ii) $T = 1$ then $p_k = 2^{-N-1}(N - 2 + 2^{-k+1})$

(iii) $T > 1$ then $p_k = 2^{-N-1}\binom{N}{T}(1 + o(1))$

what clearly implies our Lemma.

Proof of Theorem 5. Let

$$B_k = \{S_{k+K} - S_k \geqslant k - T\} \qquad (k = 0, 1, 2, \ldots, N-K),$$

$$C_l = \sum_{k=lK}^{(l+1)K}{}' B_k \qquad \left(l = 0, 1, 2, \ldots, \left[\frac{N-2K}{K}\right]\right),$$

$$D_0 = C_0 + C_2 + \ldots + C_{2[\frac{1}{2}[\frac{N-2K}{K}]]},$$

$$D_1 = C_1 + C_3 + \ldots + C_{2[\frac{1}{2}([\frac{N-2K}{K}]-1)]+1}.$$

Then by Lemma 1

$$\mathsf{P}(C_l) = 2^{-K-1}\,\frac{K^{T+1}}{T!}\,(1 + o_K(1))$$

and since the events $C_0, C_2, \ldots$ are independent we have

$$\mathsf{P}(\bar{D}_0) = \mathsf{P}(\bar{C}_0)\,\mathsf{P}(\bar{C}_2)\ldots\mathsf{P}\left(\bar{C}_{2[\frac{1}{2}[\frac{N-2K}{K}]]}\right) =$$

$$= \left(1 - 2^{-K-1}\,\frac{K^{T+1}}{T!}\,(1 + o_K(1))\right)^{[\frac{1}{2}[\frac{N-2K}{K}]]+1}$$

and similarly

$$\mathsf{P}(\bar{D}_1) = \left(1 - 2^{-K-1}\,\frac{K^{T+1}}{T!}\,(1 + o_K(1))\right)^{[\frac{1}{2}([\frac{N-2K}{K}]-1)]+1}.$$

Clearly

$$D_0 \subset \{I(N, K) \geqslant K - T\} = D_0 + D_1$$

and

$$\mathsf{P}\{I(N, K) < K - T\} = \mathsf{P}(\overline{D_0 + D_1}) = \mathsf{P}(\bar{D}_0\bar{D}_1) \geqslant \mathsf{P}(\bar{D}_0)\,\mathsf{P}(\bar{D}_1).$$

what proves Theorem 5. The right side of the last inequality follows from the simple inequality

$$\mathsf{P}(D_1 \mid B_k) \geqslant \mathsf{P}(D_1) \qquad (k = 0, 1, 2, \ldots, N - K).$$

§3. THE PROOFS OF THEOREMS $1^* - 4^*$

The following two Lemmas are trivial consequences of Theorem 5.

Lemma 2. *Let* $N_j = N_j(T)$ *be the smallest integer for which* $\alpha_1(N_j, T) = j$. *Then*

$$\sum_{j=1}^{\infty} \mathsf{P}\{Z_{N_j}(T) < \alpha_1(N_j, T)\} =$$

$$= \sum_{j=1}^{\infty} \mathsf{P}\{I(N_j, \alpha_1(N_j, T)) < \alpha_1(N_j, T) - T\} < \infty.$$

Lemma 3. *Let* δ *be a positive number and let* $N_j = N_j(T, \delta)$ *be the smallest integer for which* $\alpha_2(N_j, T) = [j^{1+\delta}]$. *Then*

$$\sum_{j=1}^{\infty} \mathsf{P}\{I(N_j, \alpha_2(N_j, T)) < \alpha_2(N_j, T) - T\} = \infty$$

if δ *is small enough.*

Now Theorem 1* follows immediately from Lemma 2.

In order to prove Theorem 2* the following version of the Borel – Cantelli lemma will be applied:

Lemma A. ([2]) *If* $A_1, A_2, \ldots$ *are arbitrary events, fulfilling the conditions*

$$\sum_{n=1}^{\infty} \mathsf{P}(A_n) = \infty$$

and

$$\liminf_{n \to \infty} \frac{\sum_{k=1}^{n} \sum_{l=1}^{n} \mathsf{P}(A_k A_l)}{\left(\sum_{k=1}^{n} \mathsf{P}(A_k)\right)^2} = 1. \tag{1}$$

Then there occur with probability 1 *infinitely many of the events* A_n.

Hence Theorem 2* will follow from

Lemma 4. *If the event* A_j *is defined as*

$$A_j = \{I(N_j, \alpha_2(N_j, T)) < \alpha_2(N_j, T) - T\}$$

then (1) *holds true.*

Proof of Lemma 4. *Let*

$$B_{ij} = \{ \max_{0 \leq k \leq N_i - \alpha_2(N_j, T)} (S_{k+\alpha_2(N_j,T)} - S_k) < \alpha_2(N_j, T) - T\} \qquad (i < j),$$

$$C_{ij} = \{ \max_{N_i \leq k \leq N_j - \alpha_2(N_j, T)} (S_{k+\alpha_2(N_j,T)} - S_k) < \alpha_2(N_j, T) - T\} \qquad (i < j).$$

Then

$$\mathsf{P}(A_i A_j) = \mathsf{P}(A_i)\mathsf{P}(C_{ij})(1 + o(1))$$

and

$$\mathsf{P}(A_j) = \mathsf{P}(B_{ij})\mathsf{P}(C_{ij})(1 + o(1))$$

hence

$$\mathsf{P}(A_i A_j) = \frac{\mathsf{P}(A_i)\mathsf{P}(A_j)}{\mathsf{P}(B_{ij})}(1 + o(1)).$$

By Theorem 5 we also have: $\mathsf{P}(B_{ij}) = 1 + o(1)$ what proves Lemma 4 and Theorem 2 at the same time.

Since

$$\mathsf{P}(S_n - S_{n-a} \geq a - T) = \sum_{j=0}^{T} \binom{a}{j} \frac{1}{2^a} \approx \frac{a^T}{T!} \frac{1}{2^a}.$$

Theorem 4** follows from the Borel – Cantelli Lemma and Theorem 3** is a simple consequence of Lemma A. (To check the conditions of Lemma A is quite easy.)

REFERENCES

[1] P. Erdős – A. Rényi, On a new law of large numbers. *Journ. Analyse Math.*, 22 (1970), 103-111.

[2] A. Rényi, *Probability Theory,* Akadémiai Kiadó, Budapest, 1970, p. 391.

P. Erdős – P. Révész

Math. Inst. of Hungarian Academy of Sci. Budapest, Reáltanoda u. 13-15, Hungary.