SEPARATING SYSTEMS OF RANDOM SUBSETS

Tamás F. Móri

Eötvös Loránd University, Budapest

28.10.1999

Abstract. Let $A_1, A_2, \dots$ be i.i.d. random subsets of the positive integers generated in such a way that the events $\{i \in A_j\}$, $1 \le i$, $1 \le j$ are independent and of the same probability p. For every $n = 1, 2, \dots$ let $\Omega_n = \{1, 2, \dots, n\}$ and define $A_j^{(n)} = A_j \cap \Omega_n$. Finally, let

$$Y_n = \min\left\{ j : A_1^{(n)}, A_2^{(n)}, \dots, A_j^{(n)} \text{ separate } \Omega_n \right\}.$$

(We say that Ω_n is separated by a family $\mathcal{A}$ of its subsets if for any two elements x, y of Ω_n there exists a subset $A \in \mathcal{A}$ such that either $x \in A,\ y \notin A$ or $y \in A,\ x \notin A$.)

In the paper the following issues are discussed:

- asymptotic distribution of Y_n as $n \to \infty$, with estimation for the accuracy of approximation,
- a.s. limit distribution,
- a.s. asymptotic behaviour, Lévy classes.

1. Introduction

Definiton. Let Ω be an arbitrary nonempty set and $\mathcal{A} \subset 2^{\Omega}$ a family of its subsets. $\mathcal{A}$ is said *to separate* Ω if for any two elements x, y of Ω there exists a subset $A \in \mathcal{A}$ such that either $x \in A,\ y \notin A$ or $y \in A,\ x \notin A$ holds.

Let Ω_n be a fixed set of size n. Select a sequence $A_1^{(n)}, A_2^{(n)}, \dots,$ of i.i.d. random subsets of Ω_n in such a way, that for each subset $A_j^{(n)}$ every element of Ω_n is picked independently and with the same probability p. Stop when they separate. Let Y_n denote the number of subsets selected. We are interested in the asymptotic properties of Y_n as $n \to \infty$. In order that a.s. investigations also make sense we need to define all Y_n in the same probability space.

Let $(X_{ij}, 1 \le i, 1 \le j)$ a two-way infinite array of i.i.d. Bernoulli random variables with $P(X_{ij} = 1) = p$, $P(X_{ij} = 0) = 1 - p = q$. With every column we associate a random subset of positive integers as follows: $A_j = \{i \ge 1 : X_{ij} = 1\}$, $j \ge 1$, that is, $X_{ij} = I(i \in A_j)$. These subsets are independent and identically distributed. Let

Research supported by the Hungarian National Foundation for Scientific Research, Grant No. T-29621

us define $A_j^{(n)}$ as the starting section of A_j: $A_j^{(n)} = \{1, 2, \dots, n\} \cap A_j$. We consider the stopping times

$$Y_n = \min\left\{k : A_1^{(n)}, A_2^{(n)}, \dots, A_k^{(n)} \text{ separate } \{1, \dots, n\}\right\},\ n \ge 1.$$

as well as the inverse quantities

$$T_k = \min\left\{n : A_1^{(n)}, A_2^{(n)}, \dots, A_k^{(n)} \text{ do not separate } \{1, \dots, n\}\right\},\ k \ge 1.$$

If we focus on the first n rows, Y_n will show, how many columns are needed so that these rows become all different. If, instead of rows, we fix k columns, and take rows one after another while they are all different (up to the first k element), then T_k is the number of rows needed for the first repetition, that is, the smallest n for which the k-vectors

$$[X_{11}, \dots, X_{1k}],\ [X_{21}, \dots, X_{2k}],\ \dots,\ [X_{n1}, \dots, X_{nk}]$$

are not all different.

Random variables Y_n and T_k are obviously in strong connection, for $\{Y_n \le k\} \equiv \{T_k > n\}$. There are problems that can be attacked more easily through T_k, while others may appear simpler if the Y_n are dealt with.

2. Asymptotic distribution

The second representation of T_k clearly shows that, as far as limit distribution is concerned, we face a particular case of the generalized birthday problem: i.i.d. random vectors of distribution $P(\boldsymbol{x}) = p^{\sum x_i} q^{k-\sum x_i}$, $\boldsymbol{x} \in \{0,1\}^k$ are taken, one after another, until the first repetition. There exists a huge amount of literature on that problem, here we only mention two papers: the classical work [9], which contains a complete description of possible limit distributions in a more general setup, and a recent preprint [2], which offers a good survey of related results. From the classical theory it follows that T_k, multiplied by the factor

$$\vartheta_k = \left(\sum_{\boldsymbol{x}\in\{0,1\}^k} \left(p^{\sum x_i} q^{k-\sum x_i}\right)^2\right)^{1/2} = \left(p^2 + q^2\right)^{k/2}$$

converges in distribution: $P\left(\vartheta_k T_k > t\right) \to \exp(-t^2/2)$, $t > 0$, as $k \to \infty$. For precise asymptotic analysis we shall also need an estimation for the rate of convergence. As we have already seen, $\{Y_n \le k\}$ means that there are no two identical k-vectors among the first n rows. For $1 \le i < j \le n$ let B_{ij} denote the event that row i is identical to row j (up to the first k element). We need the probability that none of the events B_{ij} occur. Two powerful methods that can be applied with success in similar situations are the graph-sieve of Rényi (see [4]) and the Chen–Stein method of Poisson approximation [1]. They are not equally efficient. The Chen–Stein method, if applicable, usually gives more: a Poisson approximation for the number of occurring events, together with a very sharp estimation for the accuracy measured in total variation of probability distributions. If all events in question are dependent with a complicated dependency structure then the Rényi sieve still works when the Chen–Stein method breaks down, see [5]. But when each event has a relatively small "dependency neighborhood" such that it is independent of all events outside of that, then the proper choice is the Chen–Stein method. This is the case just now: B_{ij} is independent of all events $B_{\ell m}$ that have no indices in common with it.

Let us apply Theorem 1 of [1]. Introduce $H = \{(i,j) : 1 \le i < j \le n\}$, $K_{ij} = \{(\ell,m) \in H : \{i,j\} \cap \{\ell,m\} \neq \emptyset\}$ (neighborhood of dependence), and finally

$$\begin{aligned}
\lambda_0 &= \sum_{(i,j)\in H} P(B_{ij}) = \binom{n}{2}\left(p^2+q^2\right)^k,\\
b_1 &= \sum_{(i,j)\in H}\ \sum_{(\ell,m)\in K_{ij}} P(B_{ij})\,P(B_{\ell m}) = \binom{n}{2}(2n-1)\left(p^2+q^2\right)^{2k},\\
b_2 &= \sum_{(i,j)\in H}\ \sum_{(i,j)\neq(\ell,m)\in K_{ij}} P(B_{ij}\cap B_{\ell m}) = n(n-1)^2\left(p^3+q^3\right)^k.
\end{aligned}$$

Then we immediately obtain the following basic inequality.

$$\left|P(T_k > n) - e^{-\lambda_0}\right| \le \frac{1-e^{-\lambda_0}}{\lambda_0}(b_1+b_2). \tag{2.1}$$

In order to formulate the main result of this section we shall need some more notations. Let

$$\beta = \frac{(p^2+q^2)^{3/2}}{p^3+q^3} > 1, \quad \gamma = \frac{p^2+q^2}{p^3+q^3} \le \frac{1}{p^2+q^2}, \quad \lambda = \frac{1}{2}n^2\left(p^2+q^2\right)^k.$$

Let $F(x) = \exp\left(-\frac{1}{2}\left(p^2+q^2\right)^x\right)$, $x \in \mathbb{R}$, this is the distribution function of an extreme value distribution from the location–scale family of Gumbel distributions. Define $\varrho_i(x) = F(i+x) - F(i+x-1)$, thus $\boldsymbol{\varrho}(x) = (\varrho_i(x) : i \in \mathbb{Z})$ is a parametric family of discretized versions of distribution F. For sake of brevity let us denote the logarithm to the base $(p^2+q^2)^{-1}$ by Log (while log will be reserved for natural logarithm). Let α and N denote the fractional and integer parts of $2\operatorname{Log} n$, resp. Finally, introduce $\pi_i(n) = P(Y_n = N+i)$.

Theorem 2.1.
$$\left|P(Y_n \le k) - e^{-\lambda}\right| \le 4n\gamma^{-k}, \tag{2.2}$$

$$\|\boldsymbol{\pi}(n) - \boldsymbol{\varrho}(-\alpha)\| = O\left(\frac{(\log n)^{\operatorname{Log}\gamma}}{n^{2\operatorname{Log}\gamma-1}}\right) = o\big(n^{-3pq/2}\big), \tag{2.3}$$

where $\|\,.\,\|$ stands for total variation,

$$\sup_x\left|P\left(\left(p^2+q^2\right)^{k/2}T_k > x\right) - \exp\left(-\tfrac{1}{2}x^2\right)\right| = O\Big(\frac{\sqrt{k}}{\beta^k}\Big). \tag{2.4}$$

Proof. From (2.1) it follows that

$$\begin{aligned}
\left|P(Y_n \le k) - e^{-\lambda_0}\right| &\le 2n\left(\left(p^2+q^2\right)^k + \gamma^{-k}\right)\left(1-e^{-\lambda_0}\right)\\
&\le 4n\gamma^{-k}\left(1-e^{-\lambda_0}\right).
\end{aligned}$$

This, together with the inequality

$$\left|e^{-\lambda_0}-e^{-\lambda}\right| \le e^{-\lambda_0}\left(1-\exp\left(-\frac{n}{2}\left(p^2+q^2\right)^k\right)\right) \le e^{-\lambda_0}\frac{n}{2}\left(p^2+q^2\right)^k \le \frac{n}{2}\gamma^{-k}e^{-\lambda_0}$$

gives (2.2).

For the proof of (2.3) let $k=N+i$, then $\gamma^k=\gamma^{2\,\mathrm{Log}\,n+i-\alpha}=n^{2\,\mathrm{Log}\,\gamma}\,\gamma^{i-\alpha}$, and

$$\lambda=\tfrac{1}{2}n^2\left(p^2+q^2\right)^k=\tfrac{1}{2}\left(p^2+q^2\right)^{N+i-2\,\mathrm{Log}\,n}=\tfrac{1}{2}\left(p^2+q^2\right)^{i-\alpha},$$

thus $e^{-\lambda}=F(i-\alpha)$. Hence, with an arbitrarily fixed i_0 we can write

$$\begin{aligned}
\|\boldsymbol{\pi}(n)-\boldsymbol{\varrho}(-\alpha)\| &= \sum_{i\in\mathbb{Z}}|\varrho_i(-\alpha)-\pi_i(n)| = 2\sum_{i\in\mathbb{Z}}\left(\varrho_i(-\alpha)-\pi_i(n)\right)^+\\
&\le 2\sum_{i>i_0}|\varrho_i(-\alpha)-\pi_i(n)| + 2\sum_{i\le i_0}\varrho_i(-\alpha)\\
&\le 4\sum_{i\ge i_0}\Big|P\left(Y_n\le N+i\right)-F(i-\alpha)\Big| + 2F(i_0-\alpha)\\
&\le 16n\sum_{i\ge i_0}\gamma^{-(N+i)}+2F(i_0-\alpha)\\
&= 16n\left(1-\tfrac{1}{\gamma}\right)^{-1}\gamma^{-(N+i_0)}+2F(i_0-\alpha)\\
&= \frac{16\gamma}{\gamma-1}\,n^{1-2\,\mathrm{Log}\,\gamma}\,\gamma^{-(i_0-\alpha)}+2F(i_0-\alpha).
\end{aligned}\tag{2.5}$$

Let $\delta=2\,\mathrm{Log}\,\gamma-1>0$ and i_0 such that

$$n^{-\delta/(p^2+q^2)}<F(i_0-\alpha)\le n^{-\delta}.$$

Such an i_0 does exist, because $F(x+1)=F(x)^{p^2+q^2}$. Since $F(i_0+1-\alpha)>n^{-\delta}$, it follows that $i_0+1-\alpha>\mathrm{Log}(2\delta\log n)$, thus

$$\gamma^{-i_0-\alpha}<\gamma(2\delta\log n)^{\mathrm{Log}\,\gamma}.$$

Plugging this in (2.5) we obtain the first equality of (2.3).

For the second equality of (2.3) we need to estimate $2\,\mathrm{Log}\,\gamma-1$. Since $p^2+q^2=1-2pq$ and $p^3+q^3=1-3pq$, we can write

$$2\,\mathrm{Log}\,\gamma-1=2\,\frac{\log(1-3pq)}{\log(1-2pq)}-3=3\left(\frac{\displaystyle\int_0^{pq}\frac{dt}{1-3t}}{\displaystyle\int_0^{pq}\frac{dt}{1-2t}}-1\right).$$

Here

$$\begin{aligned}
\frac{1}{pq}\int_0^{pq}\frac{dt}{1-3t} > \frac{1}{pq}\int_0^{pq}\frac{1+t}{1-2t}\,dt > \frac{1}{pq}\int_0^{pq}(1+t)dt\;\frac{1}{pq}\int_0^{pq}\frac{dt}{1-2t}\\
=\frac{1}{pq}\left(1+\frac{pq}{2}\right)\int_0^{pq}\frac{dt}{1-2t},
\end{aligned}$$

consequently, $2\,\mathrm{Log}\,\gamma - 1 > \frac{3}{2}\,pq$.

Finally, let x be a fixed positive number, and $n = \left[x\left(p^2+q^2\right)^{-k/2}\right]$. Then

$$P\left(\left(p^2+q^2\right)^{k/2} T_k > x\right) = P\left(T_k > n\right) = P\left(Y_n \le k\right),$$

and from (2.2) we have

$$\left|P\left(Y_n \le k\right) - \exp\left(-\tfrac{1}{2}n^2\left(p^2+q^2\right)^k\right)\right| \le 4n\gamma^{-k} \le 4x\beta^{-k}.$$

On the other hand, $0 \le x^2 - n^2\left(p^2+q^2\right)^k \le 2x\left(p^2+q^2\right)^{k/2}$, which implies

$$\begin{aligned} 0 \le \exp\left(-\tfrac{1}{2}n^2\left(p^2+q^2\right)^k\right) - \exp\left(-\tfrac{1}{2}x^2\right) &\le 1 - \exp\left(-x\left(p^2+q^2\right)^{k/2}\right) \\ &\le x\left(p^2+q^2\right)^{k/2} \le x\,\beta^{-k}. \end{aligned}$$

Hence, for $x \le x_0 = \sqrt{2k\log\beta}$ we have

$$\left|P\left(\left(p^2+q^2\right)^{k/2} T_k > x\right) - \exp\left(-\tfrac{1}{2}x^2\right)\right| \le 5x_0\beta^{-k} = O\Big(\frac{\sqrt{k}}{\beta^k}\Big),$$

while for $x > x_0$

$$\begin{aligned} &\left|P\left(\left(p^2+q^2\right)^{k/2} T_k > x\right) - \exp\left(-\tfrac{1}{2}x^2\right)\right| \le \\ &\qquad \le P\left(\left(p^2+q^2\right)^{k/2} T_k > x_0\right) \vee \exp\left(-\tfrac{1}{2}x_0^2\right) = O\Big(\frac{\sqrt{k}}{\beta^k}\Big). \end{aligned}$$

3. A.s. limit distribution

From (2.3) it is clear that $Y_n - [\mathrm{Log}\,n]$ is stochastically bounded, but does not have a limit distribution as $n \to \infty$, because of the logarithmic periodicity appearing in the asymptotic distribution. This is not just a matter of centering, no other centering sequence could made T_n converge in distribution.

Similar periodicity appears in the asymptotic distribution of random variables inverse to other sequences of waiting times that increase at an exponential rate, see [7]. A typical example is the length of the longest head-run observed during n tosses of a coin. However, in each of those examples the existence of an a.s. limit distribution can be proved.

A sequence of random variables ζ_n is said to have a.s. limit distribution, if for every real x

$$\lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t} \frac{1}{n}\, I(\zeta_n \le x) = G(x) \quad \text{a.s.} \tag{3.1}$$

with some non-degenerate distribution function $G(x)$. Under quite general conditions, (3.1) holds if and only if the sequence of probabilities $P(\zeta_n \le x)$ is logarithmically summable to $G(x)$. This "transfer principle" is supported by the following simple lemma.

Lemma 3.1. [6] *Let $\xi_1, \xi_2, \dots$ be a sequence of uniformly bounded random variables (e.g. $\xi_n = I(\zeta_n \le x) - P(\zeta_n \le x)$), such that $|E(\xi_i \xi_j)| \le h(j/i)$, $1 \le i < j$, where h is a positive decreasing function, and*

$$\int_1^\infty \frac{h(x)}{x \log x}\, dx \le \infty.$$

Then

$$\lim_{t\to\infty} \frac{1}{\log t} \sum_{n=1}^{t} \frac{1}{n}\, \xi_n = 0 \quad a.s.$$

Since logarithmic averaging can eliminate periodicity, a.s. limit distribution may exist even when ordinary limit distribution does not.

In order to apply Lemma 3.1 we first have to estimate $P(Y_n \le k,\, Y_s \le r) = P(T_k > n,\, T_r > s)$, $k \le r$, $n \le s$. Such an estimation will be useful in Section 4, so calculation will be carried out in a little bit more general setup than it is necessary here. The method we are going to apply is the Chen–Stein approximation for the conditional distribution

$$P\left(T_r > s \;\middle|\; X_{ij},\ i \le n,\ j \le k\right).$$

For sake of brevity, let $\mathcal{F} = \sigma\{X_{ij} : i \le n,\, j \le k_1\}$ and let $\mathcal{H}$ denote the set of those pairs (i,j), $1 \le i < j \le n$, that are not separated by $A_1^{(n)}, \dots, A_k^{(n)}$, that is, $[X_{i1}, \dots, X_{ik}] \equiv [X_{j1}, \dots, X_{jk}]$.

$$\mathcal{H} = \big\{(i,j) : 1 \le i < j \le n,\ B_{ij} \text{ occurs}\big\}.$$

Further, let $S_i = X_{i1} + \dots + X_{ik}$, $1 \le i \le n$, they are i.i.d. random variables.

By Theorem 1 of [1], $P\left(T_r > s \mid \mathcal{F}\right)$ is approximately equal to $e^{-\mu}$, where

$$\begin{aligned}\mu &= \sum_{1\le i<j\le s} P\left(B_{ij} \mid \mathcal{F}\right) = \sum_{1\le i<j\le n} + \sum_{n<i<j\le s} + \sum_{1\le i\le n<j\le s} \\ &= \left(p^2+q^2\right)^{r-k} |\mathcal{H}| + \binom{s-n}{2} \left(p^2+q^2\right)^{r} + \left(p^2+q^2\right)^{r-k} \sum_{i=1}^{n} p^{S_i} q^{k-S_i}.\end{aligned}$$

The approximation error is majorized again by

$$\sum_{\{i,j\}\cap\{\ell,m\}\neq\emptyset} P\left(B_{ij} \mid \mathcal{F}\right) P\left(B_{\ell m} \mid \mathcal{F}\right) + \sum_{\substack{\{i,j\}\cap\{\ell,m\}\neq\emptyset\\ (i,j)\neq(\ell,m)}} P\left(B_{ij} \cap B_{\ell m} \mid \mathcal{F}\right)\Big). \tag{3.2}$$

Let us estimate the sums of (3.2) on the event $\{Y_n \le k\} = \{\mathcal{H} = \emptyset\} \in \mathcal{F}$. In the second sum $\big|\{i,j,\ell,m\}\big| = 3$, and $\{i,j,\ell,m\} \cap \{1,\dots,n\} \le 1$. Obviously, on $\mathcal{H}$ $\big|\{1,\dots,n\} \cap \{i,j,\ell,m\}\big| > 1$ cannot happen, because $B_{ij} \cap B_{\ell m}$ means that neither pair from $\{i,j,\ell,m\}$ is separated. Thus the second sum will be divided into two parts.

Case (a): $n < i$, and $n < \ell$. The summands are all equal to $\left(p^3+q^3\right)^r$, and there are $6\binom{s-n}{3}$ of them.

Case (b): either $i \le n$ or $\ell \le n$. The summands are of the form

$$p^{2S_t} q^{2(k-S_t)} \left(p^3+q^3\right)^{r-k},$$

where $t = i \wedge \ell$, and there are $6\binom{s-n}{2}$ of each.

Thus the second sum in (3.2) is estimated by

$$s^3 \left(p^3+q^3\right)^r + 3s^2 \left(p^3+q^3\right)^{r-k} \sum_{t=1}^{n} p^{2S_t} q^{2(k-S_t)}. \tag{3.3}$$

As regards the first sum, we distinguish two (not disjoint) cases according as $t \in \{i,j\} \cap \{\ell, m\}$ falls below or above n (in fact, the two pairs may coincide, then t is not unique).

Case (a): $t \le n$. The contribution of those terms is

$$\left((s-n)\, p^{S_t} q^{k-S_t} \left(p^2+q^2\right)^{r-k}\right)^2.$$

Case (b): $n < t$. The contribution of those terms is

$$\left(\sum_{i=1}^{n} p^{S_i} q^{k-S_i} \left(p^2+q^2\right)^{r-k} + (s-n-1)\left(p^2+q^2\right)^r\right)^2 \le$$
$$\le 2\left(p^2+q^2\right)^{2(r-k)} \left(\sum_{i=1}^{n} p^{S_i} q^{k-S_i}\right)^2 + 2s^2\left(p^2+q^2\right)^{2r}.$$

Thus the first sum in (3.2) is estimated by

$$2s^3\left(p^2+q^2\right)^{2r} + s^2\left(p^2+q^2\right)^{2(r-k)} \sum_{r=1}^{n} p^{2S_r} q^{2(k-S_r)} +$$
$$+ 2s\left(p^2+q^2\right)^{2(r-k)} \left(\sum_{i=1}^{n} p^{S_i} q^{k-S_i}\right)^2. \tag{3.4}$$

By using (3.3), (3.4), and inequality $\left(p^2+q^2\right)^2 \le p^3+q^3$ we obtain the following estimation for the approximation error,

$$3s^3\left(p^3+q^3\right)^r \; + \; 2s\left(p^3+q^3\right)^{r-k} \Sigma_1^2 \; + \; 4s^2\left(p^3+q^3\right)^{r-k} \Sigma_2,$$

where

$$\Sigma_1 = \sum_{i=1}^{n} p^{S_i} q^{k-S_i}, \quad \Sigma_2 = \sum_{i=1}^{n} p^{2S_i} q^{2(k-S_i)}.$$

Let us introduce the event

$$D_{kn} = \left\{ \sum_{i=1}^{n} p^{S_i} q^{k-S_i} \le k^3\left(p^2+q^2\right)^k, \; \sum_{i=1}^{n} p^{2S_i} q^{2(k-S_i)} \le k^3\left(p^3+q^3\right)^k \right\}.$$

The distribution of S_i is binomial, so it is easy to see that

$$E\left(p^{S_i}q^{k-S_i}\right)=\left(p^2+q^2\right)^k,\quad E\left(p^{2S_i}q^{2(k-S_i)}\right)=\left(p^3+q^3\right)^k,$$

hence by the Markov inequality $P\left(\overline{D}_{kn}\right)\le 2k^{-3}$.

On $D_{kn}\cap\{Y_n\le k\}$ we have

$$\binom{s-n}{2}\left(p^2+q^2\right)^r\le\mu\le\left(\binom{s-n}{2}+k^3n\right)\left(p^2+q^2\right)^r, \tag{3.5}$$

and the approximation error can be estimated by

$$s^3\left(p^3+q^3\right)^r\left(3+2k^6+4k^3\right)\le 9s^3\left(p^2+q^2\right)^r k^6\gamma^{-r}.$$

Putting all these together we obtain the following estimation.

Lemma 3.2. *Let $C_{kn}=\{T_k>n\}\cap D_{kn}$. Then*

$$\begin{aligned}\left|P\left(C_{kn}\cap C_{rs}\right)-P\left(C_{kn}\right)P\left(C_{rs}\right)\right|&\le\frac{ns}{(s-n)^2}P\left(C_{kn}\right)+\\&+\left(s+k^3n\right)\left(p^2+q^2\right)^r+4s\gamma^{-r}+9s^3\left(p^2+q^2\right)^r k^6\gamma^{-r}+4r^{-3}.\end{aligned}$$

Proof. Let us start from inequality

$$\begin{aligned}\left|P\left(C_{kn}\cap C_{rs}\right)-P\left(C_{kn}\right)P\left(C_{rs}\right)\right|&\le\left|P\left(C_{kn}\cap C_{rs}\right)-P\left(C_{kn}\cap\{T_r>s\}\right)\right|+\\&+\left|P\left(C_{kn}\cap\{T_r>s\}\right)-P\left(C_{kn}\right)e^{-\mu}\right|+\left|e^{-\mu}-e^{-\lambda}\right|P\left(C_{kn}\right)+\\&+\left|e^{-\lambda}-P\left(C_{rs}\right)\right|P\left(C_{kn}\right),\end{aligned}$$

where $\lambda=\frac12 s^2\left(p^2+q^2\right)^r$, and

$$\left|\mu-\tfrac12(s-n)^2\left(p^2+q^2\right)^r\right|\le\left(s+k^3n\right)\left(p^2+q^2\right)^r$$

by (3.5). Terms in the right-hand side will be estimated separately. Firstly,

$$\left|P\left(C_{kn}\cap\{T_r>s\}\right)-P\left(C_{kn}\cap C_{rs}\right)\right|\le P\left(\overline{D}_{rs}\right)\le 2r^{-3}.$$

Let us integrate $P\left(T_r>s\mid\mathcal{F}\right)$ on the event C_{kn}. There we have

$$\left|P\left(T_r>s\mid\mathcal{F}\right)-e^{-\mu}\right|\le 9s^3\left(p^2+q^2\right)^r k^6\gamma^{-r},$$

hence the same upper bound holds for $\left|P\left(C_{kn}\cap\{T_r>s\}\right)-P\left(C_{kn}\right)e^{-\mu}\right|$. Let η denote $\left(1-\frac{n}{s}\right)^2$, then in the next term

$$\begin{aligned}\left|e^{-\mu}-e^{-\lambda}\right|&\le e^{-\lambda\eta}-e^{-\lambda}+\left|e^{-\mu}-e^{-\lambda\eta}\right|\\&\le e^{-\lambda\eta}\lambda(1-\eta)+\left(s+k^3n\right)\left(p^2+q^2\right)^r\\&\le\frac{1}{e\eta}\cdot\frac{2n}{s}+\left(s+k^3n\right)\left(p^2+q^2\right)^r\\&\le\frac{ns}{(s-n)^2}+\left(s+k^3n\right)\left(p^2+q^2\right)^r.\end{aligned}$$

Finally, from (2.2) it follows that

$$\left|e^{-\lambda}-P\left(C_{rs}\right)\right|\le 4s\gamma^{-r}+2r^{-3}.$$

From all these we get just what we need.

Now we are in a position to prove the main result of this section.

Theorem 3.1. *With probability* 1

$$\lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\, I\left(Y_n - [2\,\mathrm{Log}\, n] = i\right) = \int_0^1 \big(F(i-y) - F(i-1-y)\big)dy, \quad i \in \mathbb{Z},$$

$$\lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\, I\left(Y_n - 2\,\mathrm{Log}\, n \le x\right) = \int_0^1 F(x-y)dy, \quad x \in \mathbb{R}.$$

Proof. We will only prove the first limit relation. The case where the centering sequence is $2\,\mathrm{Log}\, n$ can be treated similarly, and therefore it will be omitted.

Let $k = [2\,\mathrm{Log}\, n] + i$ and $C_n = \{Y_n \le k\} \cap D_{kn}$. We will use Lemma 3.1 with $\xi_n = I(C_n) - P(C_n)$, thus we need to estimate the covariances $E\left(\xi_n \xi_s\right) = P\left(C_n \cap C_s\right) - P\left(C_n\right)P\left(C_s\right)$, $1 \le n < s$. Let $r = [2\,\mathrm{Log}\, s] + i \ge k$, then $s\gamma^{-r} = O\left(\beta^{-r}\right)$, $s^2\left(p^2+q^2\right)^r = O(1)$, and from Lemma 3.2 it is clear that

$$\left|P\left(C_n \cap C_s\right) - P\left(C_n\right)P\left(C_s\right)\right| = O\left(\frac{n}{s} + \frac{1}{\log^3 s}\right)$$

as n and $s-n$ tend to infinity, thus $h(x) = O\left((\log x)^{-3}\right)$ will do. Since $P(C_n) \sim F(i-\alpha)$, Lemma 3.1 implies

$$\lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\, I(C_n) = \lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\, F(i-\alpha).$$

Let the value of N be fixed; it means that n falls between $h_1 = (p^2+q^2)^{-N/2}$ and $h_2 = (p^2+q^2)^{-(N+1)/2}$. The contribution of such terms to the logarithmic sum is

$$\sum_{h_1 \le n < h_2} \frac{1}{n} F(i-\alpha) \sim \int_{h_1}^{h_2} \frac{1}{x} F(i - 2\,\mathrm{Log}\, x)dx.$$

By substitution $y = 2\,\mathrm{Log}\, x - N$ this integral is transformed into

$$-\frac{1}{2}\log\left(p^2+q^2\right)\int_0^1 F(i-y)dy,$$

hence we obtain that

$$\lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\, I(C_n) = \int_0^1 F(i-y)dy.$$

In order to complete the proof of the first relation of Theorem 3.1 it suffices to note that

$$E\left(\sum_{n=1}^{\infty}\frac{1}{n}\, I(D_{kn})\right) \le \sum_{n=1}^{\infty}\frac{2}{nk^3} < \infty,$$

for here

$$\frac{1}{nk^3} = O\left(\frac{1}{n\log^3 n}\right).$$

Consequently, with probability 1,

$$\lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\big(I\left(T_n - [2\,\mathrm{Log}\, n] \le i\right) - I(C_n)\big) \le \lim_{t\to\infty} \frac{1}{\log t}\sum_{n=1}^{t}\frac{1}{n}\, I(D_{kn}) = 0.$$

4. Lévy classes

For the definiton of Lévy classes UUC, ULC, LUC, LLC see Chapter 5 of [8]. The a.s. asymptotic behaviour of the sequence Y_n is better to study through the inverse sequence T_k. First we deal with the upper classes.

Theorem 4.1 (UUC/ULC *of* T_k). *Let* ψ *be a positive increasing function. The probability that* $\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)$ *holds for infinitely many* k *is equal to* 0 *or* 1, *according as the sum*

$$\sum_{k=1}^{\infty} \exp\left(-\tfrac{1}{2}\psi(k)^2\right) \tag{4.1}$$

is finite or infinite.

Proof. Suppose (4.1) is finite. Then $P\left(\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)\right) \sim \exp\left(-\frac{1}{2}\psi(k)^2\right)$, by (2.4), thus

$$\sum_{k=1}^{\infty} P\left(\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)\right) < \infty.$$

The Borel–Cantelli lemma implies that $\psi(k)$ belongs to the upper–upper class of the sequence $\left(p^2+q^2\right)^{k/2} T_k$.

Conversely, assume (4.1) is infinite. We may suppose that $\psi(k) \leq 2\left(\log k\right)^{1/2}$, or else we can replace $\psi(k)$ with $\psi'(k) = \psi(k) \wedge 2\left(\log k\right)^{1/2}$. In this way (4.1) remains infinite, and $\psi(k)$ belongs to the lower–upper class if and only if so does $\psi'(k)$, because $\left(p^2+q^2\right)^{k/2} T_k \geq 2\left(\log k\right)^{1/2}$ cannot occur for sufficiently large k. We may also assume that $\psi(k) \to \infty$, otherwise $\limsup P\left(\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)\right)$ would be positive, which, combined with the 0 or 1 law of Halmos and Savage, would give that $\psi(k) \in \mathrm{LUC}$.

Let $n = n(k) = \lceil \left(p^2+q^2\right)^{-k/2} \psi(k) \rceil - 1$, that is, $\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)$ if and only if $T_k > n$.

This time let $C_k = C_{k,n(k)} = \{T_k > n\} \cap D_{k,n(k)}$, then $P\left(C_k\right) \sim \exp\left(-\frac{1}{2}\psi(k)^2\right)$ again. We will apply the Erdős–Rényi generalization of the Borel–Cantelli lemma (see [3]) to the events C_k. To this end we need an upper estimation for the expression

$$\sigma_M^2 := \sum_{k=1}^{M} \sum_{r=1}^{M} \left(P\left(C_k \cap C_r\right) - P\left(C_k\right) P\left(C_r\right)\right).$$

Let us apply Lemma 3.2 with $r > k$ and $s = n(r) \leq n(k)$. By supposition,

$$n^2 \left(p^2+q^2\right)^k \leq 4 \log k, \quad s^2 \left(p^2+q^2\right)^r \leq 4 \log r, \tag{4.2}$$

hence

$$\begin{aligned} \left|P\left(C_k \cap C_r\right) - P\left(C_k\right) P\left(C_r\right)\right| &\leq \frac{ns}{(s-n)^2} P\left(C_k\right) + \\ &+ 4k^3 \left(\log r\right)^{1/2} \left(p^2+q^2\right)^{r/2} + 8\left(\log r\right)^{1/2} \beta^{-r} + 72k^6 \left(\log r\right)^{3/2} \beta^{-r} + 4r^{-3}. \end{aligned}$$

Here

$$\frac{ns}{(s-n)^2} = \frac{n}{s}\left(1-\frac{n}{s}\right)^{-2}, \qquad \frac{n}{s} = \left(p^2+q^2\right)^{(r-k)/2} + O\left(\left(p^2+q^2\right)^{-r}\right),$$

from which it follows that

$$\begin{aligned}
\sigma_M^2 &\le \sum_{k=1}^{M} P\left(C_k\right) + 2\sum_{1\le k<r\le M} \left|P\left(C_k\cap C_r\right) - P\left(C_k\right)P\left(C_r\right)\right| \\
&\le \sum_{k=1}^{M} P\left(C_k\right) + 2\sum_{1\le k<r\le M} P\left(C_k\right)\frac{ns}{(s-n)^2} + O(1) \\
&= \sum_{k=1}^{M} P\left(C_k\right) + O\left(\sum_{1\le k<r\le M} P\left(C_k\right)\left(p^2+q^2\right)^{(r-k)/2}\right) \\
&= \sum_{k=1}^{M} P\left(C_k\right) + O\left(\sum_{\ell=1}^{M-1}\left(p^2+q^2\right)^{\ell/2}\sum_{k=1}^{M-\ell} P\left(C_k\right)\right) \\
&= O\left(\sum_{k=1}^{M} P\left(C_k\right)\right).
\end{aligned}$$

The Erdős–Rényi lemma implies that, with probability 1, infinitely many of the events C_k occur. Since $\sum P\left(\overline{D}_{k,n(k)}\right) < \infty$, $D_{k,n(k)}$ occurs for every k large enough, thus $\psi(k) \in$ LUC, indeed.

Theorem 4.2 (LUC/LLC *of* T_k). *Let* ψ *be a positive decreasing function, for which* $\left(p^2+q^2\right)^{-k/2}\psi(k)$ *increases. The probability that* $\left(p^2+q^2\right)^{k/2} T_k \le \psi(k)$ *holds for infinitely many* k *is equal to* 0 *or* 1*, according as the sum*

$$\sum_{k=1}^{\infty}\psi(k)^2 \tag{4.3}$$

is finite or infinite.

Proof. The proof goes along the same lines as that of Theorem 4.1. When (4.3) is finite, then $P\left(\left(p^2+q^2\right)^{k/2} T_k \le \psi(k)\right) \sim \frac{1}{2}\psi(k)^2$, hence the LLC result follows from the ordinary Borel–Cantelli lemma.

When (4.3) is infinite, we can suppose that $P\left(\left(p^2+q^2\right)^{k/2} T_k \le \psi(k)\right) \to 0$, that is, $\psi(k) \to 0$. We can confine ourselves to the case $1/k < \psi(k)$ without loss of generality. Let $n = n(k) = \left[\left(p^2+q^2\right)^{-k/2}\psi(k)\right]$, and $C_k = \{T_k > n\} \cap D_{k,n}$. Again, the Erdős–Rényi lemma will be applied, but this time to the events $\overline{C}_k$. Note that (4.2) is replaced with inequality $n\left(p^2+q^2\right)^{k/2} \ge 1/k$.

For the estimation of

$$\begin{aligned}
\sigma_M^2 &= \sum_{k=1}^{M}\sum_{r=1}^{M}\left(P\left(\overline{C}_k\cap\overline{C}_r\right) - P\left(\overline{C}_k\right)P\left(\overline{C}_r\right)\right) \\
&= \sum_{k=1}^{M}\sum_{r=1}^{M}\left(P\left(C_k\cap C_r\right) - P\left(C_k\right)P\left(C_r\right)\right)
\end{aligned}$$

it is sufficient to deal with $\left|P\left(C_k \cap C_r\right) - P\left(C_k\right) P\left(C_r\right)\right|$ again, but Lemma 3.2 has to be replaced with another, very similar result, namely

$$\begin{aligned}\left|P\left(C_{kn} \cap C_{rs}\right) - P\left(C_{kn}\right) P\left(C_{rs}\right)\right| &\le ns\left(p^2+q^2\right)^r + \\ &+ \left(s + k^3 n\right)\left(p^2+q^2\right)^r + 4s\gamma^{-r} + 9s^3\left(p^2+q^2\right)^r k^6 \gamma^{-r} + 4r^{-3}.\end{aligned}$$

The only difference is in the estimation of $e^{-\lambda\eta} - e^{-\lambda}$. Clearly,

$$\begin{aligned}e^{-\lambda\eta} - e^{-\lambda} &= \left(1 - \left(1 - e^{-\lambda}\right)\right)^{\eta} - e^{-\lambda} \le 1 - \eta\left(1 - e^{-\lambda}\right) - e^{-\lambda} \\ &= (1-\eta)\left(1 - e^{-\lambda}\right) \le \frac{2n}{s}\lambda = ns\left(p^2+q^2\right)^r.\end{aligned}$$

Here

$$ns\left(p^2+q^2\right)^r \le \psi(k)^2\left(p^2+q^2\right)^{(r-k)/2},$$

therefore we can write

$$\begin{aligned}\sigma_M^2 &\le \sum_{k=1}^{M} P\left(\overline{C}_k\right) + 2\sum_{1\le k<r\le M}\left|P\left(C_k \cap C_r\right) - P\left(C_k\right) P\left(C_r\right)\right| \\ &\le \sum_{k=1}^{M} P\left(\overline{C}_k\right) + 2\sum_{1\le k<r\le M}\psi(k)^2\left(p^2+q^2\right)^{(r-k)/2} + O(1) \\ &= O\left(\sum_{k=1}^{M} P\left(\overline{C}_k\right)\right),\end{aligned}$$

completing the proof.

Remark. A sequence x_i of real numbers is called *quasi-increasing* (quasi-decreasing, resp.), if the supremum (infimum) of the set of differences $\{x_i - x_j : 1 \le i < j\}$ is finite. From the proofs it can be seen that the sequence $\psi(k)$ in Theorems 4.1 and 4.2 need not be monotone: it is sufficient to require that $\left(p^2+q^2\right)^{-k/2}\psi(k)$ increases and

- (in Theorem 4.1) $\log\psi(k)$ is quasi-increasing,
- (in Theorem 4.2) $\log\psi(k)$ is quasi-decreasing.

Finally, we adapt our results to the sequence Y_n.

Theorem 4.3 (UUC/ULC *of* Y_n). *Let $k(n)$ be a a non-decreasing sequence of positive integers, for which $k(n) - 2\operatorname{Log} n$ is quasi-increasing. The probability that $Y_n \ge k(n)$ holds for infinitely many n is equal to* 0 *or* 1, *according as the sum*

$$\sum_{n=1}^{\infty} n\left(p^2+q^2\right)^{k(n)} \tag{4.4}$$

is finite or infinite.

Proof. Let $n(k) = \min\{n : k(n) = k\}$, i.e., $k(n) = k$ for $n(k) \le n < n(k+1)$. Define $\psi(k) = \left(p^2+q^2\right)^{k/2}(n(k+2) - 1)$, then $\log\psi(k)$ is quasi-decreasing. Obviously,

$Y_n \geq k(n) \Leftrightarrow T_{k(n)-1} \leq n$, thus $Y_n \geq k(n)$ holds for infinitely many n if and only if $T_k \leq n(k+2)-1$, that is, $\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)$ for infinitely many k. We shall prove that (4.3) and (4.4) are equiconvergent.

Since

$$\frac{1}{4}\Big(n(k+1)^2 - n(k)^2\Big)\left(p^2+q^2\right)^k \leq \sum_{n=n(k)}^{n(k+1)-1} n\left(p^2+q^2\right)^{k(n)}$$
$$\leq \frac{1}{2}\Big(n(k+1)^2 - n(k)^2\Big)\left(p^2+q^2\right)^k,$$

we obtain, on one hand, that

$$\sum_{n=1}^{\infty} n\left(p^2+q^2\right)^{k(n)} \geq \frac{1}{4}\sum_{k=k(1)}^{\infty} n(k)^2\Big(\left(p^2+q^2\right)^{k-1} - \left(p^2+q^2\right)^k\Big) \;-\; n(1)^2$$
$$\geq \frac{pq}{4}\sum_{k=k(1)}^{\infty}\psi(k)^2 \;-\; n(1)^2,$$

thus the finiteness of (4.4) implies that of (4.3).

On the other hand, if $n(k)^2\left(p^2+q^2\right)^k \to 0$, that is, $\psi(k)\to 0$, then

$$\sum_{n=n(k(1)+2)}^{\infty} n\left(p^2+q^2\right)^{k(n)} \leq \frac{1}{2}\sum_{k=k(1)+2}^{\infty} n(k)^2\Big(\left(p^2+q^2\right)^{k-1} - \left(p^2+q^2\right)^k\Big)$$
$$\leq pq\sum_{k=k(1)}^{\infty}\psi(k)^2.$$

If (4.3) is finite, then $\psi(k)\to 0$, therefore (4.4) is also convergent.

Theorem 4.4 (LUC/LLC *of* Y_n). *Set* $f(x) = (1+x)e^{-x}$ *and let* $k(n)$ *be a non-decreasing sequence of positive integers, for which* $k(n) - 2\,\mathrm{Log}\, n$ *is quasi-decreasing. Then the probability that* $Y_n \leq k(n)$ *holds for infinitely many* n *is equal to* 0 *or* 1, *according as the sum*

$$\sum_{n=1}^{\infty}\frac{1}{n} f\left(\tfrac{1}{2}n^2\left(p^2+q^2\right)^{k(n)}\right) \tag{4.5}$$

is finite or infinite.

Proof. As in the proof of Theorem 4.3, let $n(k) = \min\{n : k(n) = k\}$. This time define $\psi(k) = \left(p^2+q^2\right)^{k/2}(n(k)+1)$, then $\log\psi(k)$ is quasi-increasing. Clearly, $Y_n \leq k(n) \Leftrightarrow T_{k(n)} \geq n+1$, thus $Y_n \leq k(n)$ holds for infinitely many n if and only if $T_k \geq n(k)+1$, that is, $\left(p^2+q^2\right)^{k/2} T_k \geq \psi(k)$ for infinitely many k. Under the condition $k(n) - 2\,\mathrm{Log}\, n \to -\infty$ or equivalently, $\psi(k)\to\infty$, we will show that (4.1) and (4.5) are equiconvergent.

On one hand we have

$$2\sum_{n=n(k)}^{n(k+1)-1}\frac{1}{n}f\left(\tfrac12 n^2\left(p^2+q^2\right)^{k(n)}\right)\ge\sum_{n=n(k)}^{n(k+1)-1}n\left(p^2+q^2\right)^k\exp\left(-\tfrac12 n^2\left(p^2+q^2\right)^k\right)$$
$$\ge\int_{n(k)}^{n(k+1)}x\left(p^2+q^2\right)^k\exp\left(-\tfrac12 x^2\left(p^2+q^2\right)^k\right)dx$$
$$=\exp\left(-\tfrac12 n(k)^2\left(p^2+q^2\right)^k\right)-\exp\left(-\tfrac12 n(k+1)^2\left(p^2+q^2\right)^k\right),$$

hence (4.5) is not less than

$$\frac12\sum_{k=1}^{\infty}\Big(\exp\big(-\tfrac12 n(k)^2\left(p^2+q^2\right)^k\big)-\exp\big(-\tfrac12 n(k)^2\left(p^2+q^2\right)^{k-1}\big)\Big).$$

Here

$$\exp\left(-\tfrac12 n(k)^2\left(p^2+q^2\right)^k\right)-\exp\left(-\tfrac12 n(k)^2\left(p^2+q^2\right)^{k-1}\right)\sim$$
$$\sim\exp\left(-\tfrac12 n(k)^2\left(p^2+q^2\right)^k\right)\ge\exp\left(-\tfrac12\psi(k)^2\right).$$

Consequently, if (4.5) is finite, so is (4.1).

On the other hand,

$$\sum_{n=n(k)}^{n(k+1)-1}\frac{1}{n}f\left(\tfrac12 n^2\left(p^2+q^2\right)^{k(n)}\right)$$
$$\le\int_{n(k)-1}^{n(k+1)-1}x\left(p^2+q^2\right)^k\exp\left(-\tfrac12 x^2\left(p^2+q^2\right)^k\right)dx$$
$$\le\exp\left(-\tfrac12(n(k)-1)^2\left(p^2+q^2\right)^k\right).\tag{4.6}$$

If $n(k)\left(p^2+q^2\right)^k\le 1$, (4.6) can be estimated by $\exp\left(2-\frac12\psi(k)^2\right)$.

If $n(k)\left(p^2+q^2\right)^k>1$, (4.6) can be estimated by $\exp\big(-\frac18\left(p^2+q^2\right)^{-k}\big)$, the latter terms produce a convergent series. Thus, if (4.1) is finite, so is (4.5).

Now the proof can be completed by applying Theorem 4.1. If (4.5) is finite, then there exists a subsequence of positive integers along which $n^2\left(p^2+q^2\right)^{k(n)}\to\infty$, that is, $k(n)-2\operatorname{Log}n\to-\infty$. By its quasi-decreasing property, $k(n)-2\operatorname{Log}n$ tends to $-\infty$ along the positive integers. If (4.5) is infinite and $k(n)-2\operatorname{Log}n$ does not tend to $-\infty$, the Hewitt–Savage 0–1 law can be applied in the same way as in the proof of Theorem 4.1.

Corollary 4.1. *With probability* 1,

$$T_n\le 2\operatorname{Log}n+\operatorname{Log}\log n+(1+\varepsilon)\operatorname{Log}\log\log n \qquad \textit{for large } n,$$
$$T_n> 2\operatorname{Log}n+\operatorname{Log}\log n+\operatorname{Log}\log\log n \qquad \textit{infinitely often},$$
$$T_n\le\left[2\operatorname{Log}n-\operatorname{Log}\log\log n-\operatorname{Log}2-2\,\frac{\operatorname{Log}\log\log n}{\log\log n}\right] \qquad \textit{infinitely often},$$
$$T_n\ge\left[2\operatorname{Log}n-\operatorname{Log}\log\log n-\operatorname{Log}2-(2+\varepsilon)\frac{\operatorname{Log}\log\log n}{\log\log n}\right] \qquad \textit{for large } n.$$

Apart from the multiplier 2 of the term $\operatorname{Log} n$, these bounds are very similar to those obtained by Erdős and Révész for the length of the longest success run in a sequence of Bernoulli trials, see [8].

References

1. Arratia, R., Goldstein, L. and Gordon, L., *Two moments suffice for Poisson approximations: The Chen–Stein method*, Ann. Probab. **17** (1989), 9–25.
2. Camarri, M. and Pitman, J., *Limit distributions and random trees derived from the birthday problem with unequal probabilities*, Technical Report No. 253, Department of Statistics, University of California, Berkeley, CA (1998).
3. Erdős, P. and Rényi, A., *On Cantor's series with convergent* $\sum 1/q_n$, Annales Univ. Sci. Budapest., Sectio Math. **2** (1959), 93–109.
4. Galambos, J., *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York, 1978.
5. Móri, T. F., *More on the waiting time till each of some given pattern occurs as a run*, Canad. J. Math. **42** (1990), 915–932.
6. Móri, T. F., *On the strong law of large numbers for logarithmically weighted sums*, Annales Univ. Sci. Budapest., Sectio Math. **36** (1993), 35–46.
7. Móri, T. F., *The a.s. limit distribution of the longest head run*, Canad. J. Math. **45** (1993), 1245–1262.
8. Révész, P., *Random Walk in Random and Non-random Environments*, World Scientific, Singapore, 1990.
9. Zubkov, A. M. and Mikhaĭlov, V. G., *Limit distributions of random variables that are connected with long duplications in a sequence of independent trials*, Teor. Verojatnost. i Primenen. **19** (1974), 173–181. (in Russian)

Department of Probability Theory and Statistics,
Eötvös Loránd University,
Rákóczi út 5, Budapest, Hungary H-1088
E-mail address: moritamas@ludens.elte.hu