Periodica Mathematica Hungarica
https://doi.org/10.1007/s10998-024-00600-6

# A limit theorem for runs containing two types of contaminations

István Fazekas[1] · Borbála Fazekas[2] · Michael Ochieng Suja[3]

Accepted: 6 January 2024


**Abstract**
In this paper, sequences of trials having three outcomes are studied. The outcomes are labelled as success, failure of type I and failure of type II. A run is called at most $1+1$ contaminated, if it contains at most one failure of type I and at most one failure of type II. The accompanying distribution for the length of the longest at most $1+1$ contaminated run is obtained. The proof is based on a powerful lemma of Csáki, Földes and Komlós. Besides a mathematical proof, simulation results supporting our theorem are also presented.

**Keywords** Coin tossing · Longest run · Accompanying distribution · Rate of convergence

**Mathematics Subject Classification** 60C05 · 60F05

## 1 Introduction

The problem of the length of the longest head run for $n$ Bernoulli random variables was first raised by T. Varga. The first answer for the case of a fair coin was given in the classical paper by Erdős and Rényi [1]. A surprisingly more precise answer, the almost sure limit result for the length of the longest runs containing at most $T$ tails was given by Erdős and Révész [2]. Their result is the following. Consider the usual coin tossing experiment with a fair coin. Let $Z_T(N)$ denote the longest head run containing at most $T$ tails in the first $N$ trials. Let log denote the logarithm to base 2 and let [.] denote the integer part. Let $h(N) = \log N + T \log\log N - \log\log\log N - \log T! + \log\log e$. Let $\varepsilon$ be an arbitrary positive number. Then for almost all $\omega \in \Omega$ there exists a finite $N_0 = N_0(\omega)$ such that

✉ István Fazekas
fazekas.istvan@inf.unideb.hu

Borbála Fazekas
borbala.fazekas@science.unideb.hu

Michael Ochieng Suja
michael.suja@science.unideb.hu

1 Faculty of Informatics, University of Debrecen, Kassai Street 26, Debrecen 4028, Hungary

2 Institute of Mathematics, University of Debrecen, Egyetem Square 1, Debrecen 4032, Hungary

3 Doctoral School of Mathematical and Computational Sciences, University of Debrecen, Egyetem Square 1, Debrecen 4032, Hungary

Published online: 06 July 2024

$Z_T(N) \geq [h(N)-2-\varepsilon]$ if $N \geq N_0$, moreover, there exists an infinite sequence $N_i = N_i(\omega)$ of integers such that $Z_T(N_i) < [h(N_i) - 1 + \varepsilon]$.

These results later inspired renewed interest in this research area and several subsequent papers came up. Asymptotic results for the distribution of the number of $T$-contaminated head runs, the first hitting time of a $T$-contaminated head run having a fixed length, and the length of the longest $T$-contaminated head run were presented by Földes [3]. For the asymptotic distribution of $Z_T(N)$, Földes [3] presented the following result. For any integer $k$ and $T \geq 0$

$$P(Z_T(N) - [\log N + T \log\log N] < k) = \exp\left(-\frac{2^{-(k+1-\{\log N + T \log\log N\})}}{T!}\right) + o(1)$$

as $N \to \infty$, where $\{.\}$ denotes the fractional part.

By applying extreme value theory, Gordon et al. [4] obtained the asymptotic behaviour of the expectation and the variance of the length of the longest $T$-contaminated head run. Also, in the same article, accompanying distributions were obtained for the length of the longest $T$-contaminated head run.

Fazekas and Suja [5] showed that the accompanying distribution initially obtained by Gordon et al. [4] could as well be arrived at using the approach given by Földes [3]. After some probabilistic calculations and algebraic manipulations of the approximation of the length of the longest $T$-contaminated run and using the main lemma in Csáki et al. [6], a convergence rate was obtained for an accompanying distribution of the longest $T$-contaminated head run by Fazekas et al. [7] for $T = 1$ and $T = 2$.

Different authors have given in depth considerations to experiments involving sequences of runs emerging from trinary trials, where the Markov chain approach is used in their analysis. Such sequences include system theoretic applications, where components might exist in the following states: 'perfect functioning', 'partial functioning' and 'complete failure' (see [8, 9]).

In this paper, we study sequences of trials having three outcomes: success, failure of type I and failure of type II. We shall say that a run is *at most* $1 + 1$ *contaminated* if it includes at most 1 failure of type I and at most 1 failure of type II. Section 2 contains the main results. We give accompanying distributions for the appropriately centralized length of the longest at most $1 + 1$ contaminated run, see Theorem 2.1. In Sect. 2, we also present simulation results supporting our theorem. The proofs are presented in Sect. 3. The proofs are based on a powerful lemma by Csáki et al. [6]. For the reader's convenience, we quote it in Sect. 3. We mention that the manuscript Fazekas et al. [10] is an extended version of the present paper. In that manuscript one can find minor details of the proofs, additional simulation examples and some further results.

## 2 The longest at most 1 + 1 contaminated run

Let $X_1, X_2, \ldots, X_N$ be a sequence of independent random variables with three possible outcomes; $0, +1$ and $-1$ labelled as success, failure of type I and failure of type II, respectively, with the distribution

$$P(X_i = 0) = p, \quad P(X_i = +1) = q_1 \text{ and } P(X_i = -1) = q_2,$$

where $p + q_1 + q_2 = 1$ and $p > 0, q_1 > 0, q_2 > 0$.

A sequence of length $m$ is called a pure run if it contains only 0 values. It is called a one-type contaminated run if it contains precisely one nonzero element: either a $+1$ or a $-1$. On the other hand, it is called a two-type contaminated run if it contains precisely one $+1$, and one $-1$ while the rest of the elements are 0's.

A run is called at most one+one contaminated (shortly at most $1+1$ contaminated) if it is either pure, one-type contaminated, or two-type contaminated.

So, for an arbitrary fixed $m$, let $A_n = A_{n,m}$ denote the occurrence of the event at the $n^{th}$ step, that is, there is an at most $1+1$ contaminated run in the sequence $X_n, X_{n+1}, \ldots, X_{n+m-1}$ and $\bar{A}_n$ is its non-occurrence. Clearly,

$$P(A_1) = p^m + m(1-p)p^{m-1} + m(m-1)p^{m-2}q_1q_2.$$

Let $\mu(N)$ be the length of the longest at most $1+1$ contaminated run in $X_1, X_2, \ldots, X_N$. Then, $\{\mu(N) < m\}$ if and only if no run of length $m$ in $X_1, X_2, \ldots, X_N$ is two-type contaminated, one-type contaminated or pure.

In what follows, we shall use the notation

$$\alpha = \frac{C_0 + \frac{1}{m}C_1 + \frac{1}{m(m-1)}C_2}{1 + \frac{p(1-p)}{(m-1)q_1q_2} + \frac{p^2}{m(m-1)q_1q_2}}, \tag{2.1}$$

where

$$C_0 = (q_1+q_2),\ C_1 = \frac{p(q_1^2+q_2^2)}{q_1q_2} - 1,$$
$$C_2 = \frac{(q_1^2+q_2^2)p^2}{q_1q_2(p-1)} + \frac{p}{p-1} + \frac{2(2p+1)q_1q_2}{(p-1)^3}.$$

We also need the notation

$$K = \frac{2C_0C_2 - C_1^2 - C_0^2}{2CC_0^2},$$

where $C = \ln\frac{1}{p}$ and ln is the logarithm to base $e$. Let

$$\begin{aligned}
m(N) = &\log N + 2\log\log N + \frac{4\log\log N}{C\log N}\\
&+ \frac{C_1 - C_0}{CC_0}\frac{1}{\log N} - \frac{4}{C}\frac{(\log\log N)^2}{(\log N)^2}\\
&+ \left(\frac{8}{C^2} - \frac{2(C_1 - C_0)}{CC_0}\right)\frac{\log\log N}{(\log N)^2}\\
&+ \left(\frac{2(C_1 - C_0)}{C^2C_0} + K\right)\frac{1}{(\log N)^2} + \frac{16}{3C}\frac{(\log\log N)^3}{(\log N)^3}\\
&+ \left(-\frac{16}{C^2} + \frac{4(C_1 - C_0)}{CC_0}\right)\frac{(\log\log N)^2}{(\log N)^3}\\
&- \left(4K + \frac{8(C_1 - C_0)}{C^2C_0}\right)\frac{\log\log N}{(\log N)^3}\\
&+ \frac{16}{C^3}\frac{\log\log N}{(\log N)^3} - \frac{8}{C^2}\frac{(\log\log N)^2}{(\log N)^3} - \frac{4(C_1 - C_0)}{C^2C_0}\frac{\log\log N}{(\log N)^3},
\end{aligned}$$

where log denotes the logarithm to base $1/p$. Let $[m(N)]$ denote the integer part of $m(N)$ and let $\{m(N)\} = m(N) - [m(N)]$ denote its fractional part. Introduce the function

$$\begin{aligned} H(x) =& -x + \frac{2}{C \log N} x - \frac{4}{C} \frac{\log \log N}{(\log N)^2} x - \frac{C_1 - C_0}{C C_0} \frac{1}{(\log N)^2} x \\ &+ \left( \frac{4(C_1 - C_0)}{C C_0} - \frac{8}{C^2} \right) \frac{\log \log N}{(\log N)^3} x + \frac{8}{C} \frac{(\log \log N)^2}{(\log N)^3} x \\ &- \frac{1}{C} \frac{1}{(\log N)^2} x^2 + \frac{4}{C} \frac{\log \log N}{(\log N)^3} x^2. \end{aligned} \tag{2.2}$$

**Theorem 2.1** *Let $p > 0$, $q_1 > 0$, $q_2 > 0$ be fixed with $p + q_1 + q_2 = 1$. Let $\mu(N)$ be the length of the longest at most $1+1$ contaminated run in $X_1, X_2, \ldots, X_N$. Then, for any integer $k$,*

$$\begin{aligned} &P(\mu(N) - [m(N)] < k) \\ &= \exp\left( -p^{-(\log(C_0 p^{-2} q_1 q_2) + H(k - \{m(N)\}))} \right) \left( 1 + \mathrm{O}\left( \frac{1}{(\log N)^3} \right) \right), \end{aligned} \tag{2.3}$$

*where $f(N) = \mathrm{O}(g(N))$ if $f(N)/g(N)$ is bounded as $N \to \infty$.*

Analysing the beginning of the proof of Theorem 2.1, we can see that the lemma of Csáki et al. [6] offers a good approximation if $p$ is small, but it offers a worse approximation if $p$ is close to 1. However, our simulation studies show that the approximation for the longest run is very good for small values of $p$, but it is still appropriate if $p$ is close to 1.

**Example 2.2** We performed several computer simulation studies for certain fixed values of $p$, $q_1$, and $q_2$. Below $N$ denotes the length of the sequences generated by us and $s$ denotes the number of repetitions of the $N$-length sequences.

Figures 1, 2, 3 and 4 present the results of the simulations. Each figure shows the empirical distribution of the longest at most $1+1$ contaminated run and its approximation suggested by Theorem 2.1. An asterisk (i.e., $*$) denotes the result of the simulation, i.e., the empirical distribution function of the longest at most $1+1$ contaminated run and circle ($\circ$) denotes the approximation offered by Theorem 2.1.

## 3 Proofs

The following lemma of Csáki, Földes and Komlós plays a fundamental role in our proofs.

**Lemma 3.1** (Main lemma, stationary case, finite form of Csáki et al. [6]) *Let $X_1, X_2, \ldots$ be any sequence of independent random variables, and let $\mathcal{F}_{n,m}$ be the $\sigma$-algebra generated by the random variables $X_n, X_{n+1}, \ldots, X_{n+m-1}$. Let $m$ be fixed and let $A_n = A_{n,m} \in \mathcal{F}_{n,m}$. Assume that the sequence of events $A_n = A_{n,m}$ is stationary, that is, $P(A_{i_1+d} A_{i_2+d} \cdots A_{i_k+d})$ is independent of $d$.*

*Assume that there is a fixed number $\alpha$, $0 < \alpha \le 1$, such that the following three conditions hold for some fixed $k$ with $2 \le k \le m$ and fixed $\varepsilon$ with $0 < \varepsilon < \min\{p/10, 1/42\}$:*

(SI)

$$|P(\bar{A}_2 \cdots \bar{A}_k | A_1) - \alpha| < \varepsilon,$$

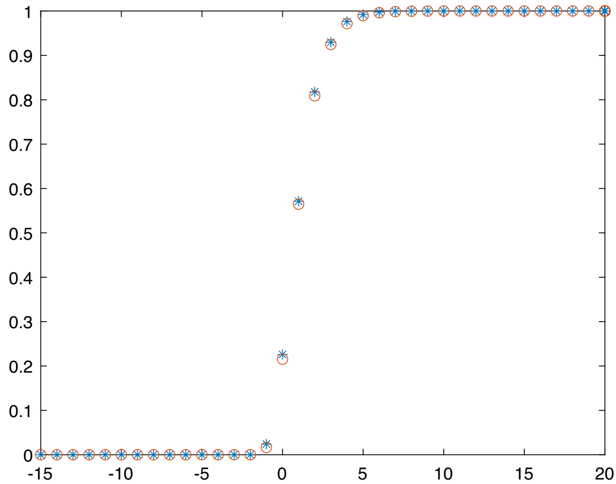


**Fig. 1** Longest at most 1 + 1 contaminated run and its approximation, $p = 1/3$, $q_1 = 1/3$, $q_2 = 1/3$, $N = 3 \times 10^6$, $s = 3000$

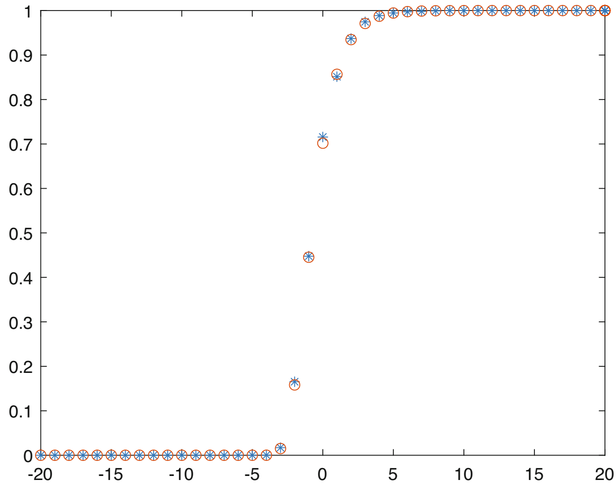


**Fig. 2** Longest at most 1 + 1 contaminated run and its approximation, $p = 0.4$, $q_1 = 0.3$, $q_2 = 0.3$, $N = 3 \times 10^6$, $s = 3000$

(SII)

$$\sum_{k+1 \le i \le 2m} P(A_i | A_1) < \varepsilon,$$

(SIII)

$$P(A_1) < \varepsilon / m.$$

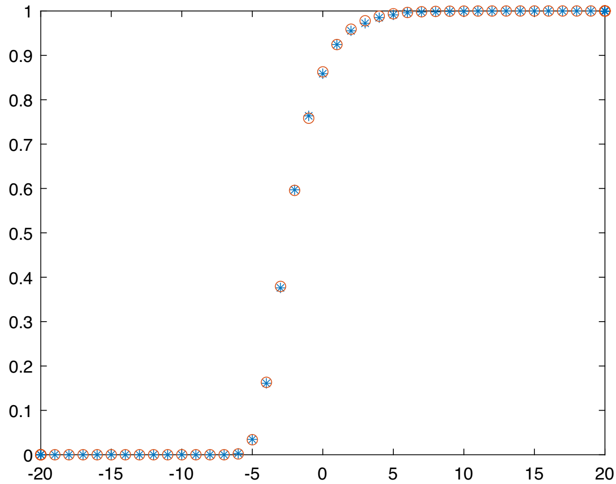


**Fig. 3** Longest at most $1 + 1$ contaminated run and its approximation, $p = 0.5$, $q_1 = 0.4$, $q_2 = 0.1$, $N = 4 \times 10^6$, $s = 3000$

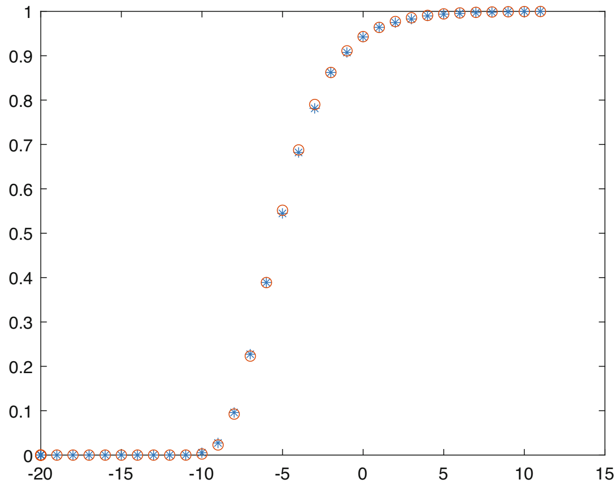


**Fig. 4** Longest at most $1 + 1$ contaminated run and its approximation, $p = 0.6$, $q_1 = 0.2$, $q_2 = 0.2$, $N = 4 \times 10^6$, $s = 3000$

*Then, for all* $N > 1$,

$$\left| \frac{P(\bar{A}_2 \cdots \bar{A}_N | A_1)}{P(\bar{A}_2 \cdots \bar{A}_N)} - \alpha \right| < 7\varepsilon$$

*and*

$$e^{-(\alpha+10\varepsilon)NP(A_1)-2mP(A_1)} < P(\bar{A}_1 \cdots \bar{A}_N) < e^{-(\alpha-10\varepsilon)NP(A_1)+2mP(A_1)}. \tag{3.1}$$

Before proceeding with the proof, we shall consider the fulfilment of some conditions given in the main Lemma for the case of $k = m$ (for fixed $m$) and $0 < p < 1$ and for some $\varepsilon > 0$.

**Remark 3.2** First, we consider condition (SIII) and show that it is true for any large enough $m$. We have

$$P(A_1) = m(m-1)p^{m-2}q_1q_2 \left\{1 + \frac{p(1-p)}{(m-1)q_1q_2} + \frac{p^2}{m(m-1)q_1q_2}\right\} \le \frac{\varepsilon}{m}. \quad (3.2)$$

This inequality is true for any positive $\varepsilon$ if $m$ is large enough.
If $m \approx \log N$, then $p^m \approx p^{\log N} = \frac{1}{N}$ and then $\varepsilon \approx \frac{(\log N)^3}{N}$.

**Remark 3.3** Now, consider condition (SII). If $i > m$, then $A_i$ and $A_1$ are independent, therefore

$$\sum_{i=m+1}^{2m} P(A_i|A_1) = mP(A_1) < \varepsilon, \quad (3.3)$$

which gives precisely the previous assumption in Remark 3.2.

**Lemma 3.4** *Condition (SI) is satisfied for $k = m$ in the following form:*

$$|P(\bar{A}_2\bar{A}_3\cdots\bar{A}_m|A_1) - \alpha| < \varepsilon, \quad (3.4)$$

*where $\alpha$ is given by (2.1) and $\varepsilon = \mathrm{O}(p^m)$.*

***Proof*** To begin, we shall be required to divide the event $A_1$ into the following pairwise disjoint parts:

$$A_1 = A_1^{(0)} \bigcup \left(\bigcup_{i=1}^{m} A_1^{(+)}(i)\right) \bigcup \left(\bigcup_{i=1}^{m} A_1^{(-)}(i)\right) \bigcup \left(\bigcup_{i,j=1, i\neq j}^{m} A_1^{(2)}(i,j)\right),$$

where $A_1^{(0)}$ is the event that $X_1, X_2, \ldots, X_m$ is a pure run,
$A_1^{(+)}(i)$ denotes that $X_i = +1$, while the rest are zeros,
$A_1^{(-)}(i)$ denotes that $X_i = -1$, while the rest are zeros,
finally, $A_1^{(2)}(i,j)$ denotes that $X_i = +1$, $X_j = -1$, while the rest are zeros.
Then

$$\begin{aligned} P(A_1\bar{A}_2\cdots\bar{A}_m) =& P(A_1^{(0)}\bar{A}_2\cdots\bar{A}_m) + \sum_{i=1}^{m} P(A_1^{(+)}(i)\bar{A}_2\cdots\bar{A}_m) \\ &+ \sum_{i=1}^{m} P(A_1^{(-)}(i)\bar{A}_2\cdots\bar{A}_m) + \sum_{i<j}^{m} P(A_1^{(2)}(i,j)\bar{A}_2\cdots\bar{A}_m) \\ &+ \sum_{i>j}^{m} P(A_1^{(2)}(i,j)\bar{A}_2\cdots\bar{A}_m) \\ :=& Y^{(0)} + \sum_{i=1}^{m} Y_i^{(+)} + \sum_{i=1}^{m} Y_i^{(-)} + \sum_{i<j}^{m} Y_{i,j} + \sum_{i>j}^{m} Y_{i,j}. \end{aligned}$$

Here, we can obtain the formula for $\sum_{i=1}^{m} Y_i^{(-)}$ by interchanging the role of $q_1$ and $q_2$ in the corresponding formula $\sum_{i=1}^{m} Y_i^{(+)}$. Similarly, we can obtain $\sum_{i>j}^{m} Y_{i,j}$ by interchanging the role of $q_1$ and $q_2$ in the corresponding formula $\sum_{i<j}^{m} Y_{i,j}$.

1. $Y^{(0)} = 0$, since the event is impossible.
2. Now, calculate $Y_i^{(+)} = P(A_1^{(+)}(i)\bar{A}_2 \cdots \bar{A}_m)$.
We want to evaluate probabilities corresponding to different values of $i$ as follows.

   (a) If $i = 1$, then the event is impossible. So $Y_1^{(+)} = 0$.
   (b) Let $1 < i < m$. Then the $m + 1$ position should be $+1$. Furthermore, in positions $m + 2, \ldots, m + i$, it is not possible that all elements are zeros and also not possible that there is a $-1$ and the rest of the elements are zeros. So, for this case,

$$Y_i^{(+)} = q_1 p^{m-1} q_1 \left(1 - p^{i-1} - (i-1) q_2 p^{i-2}\right), \quad \text{if} \quad 1 < i < m. \tag{3.5}$$

   (c) If $i = m$, then $X_{m+1}$ should be $+1$ and the remaining elements are arbitrary. So

$$Y_m^{(+)} = q_1 p^{m-1} q_1. \tag{3.6}$$

3. Now, let us turn to $Y_{i,j}$; first we consider the case when $i < j$.

   (a) If $i = 1$ and $j = m$, then $X_{m+1}$ should be $-1$ and the remaining elements are arbitrary. So

$$Y_{1,m} = q_1 q_2 p^{m-2} q_2. \tag{3.7}$$

   (b) Now, let $i = 1$ and $j < m$. Then $X_{m+1}$ should be $-1$. Moreover, it is not possible that all elements in positions $m + 2, \ldots, m + j$ are zeros; nor is it possible that one is +1 and the rest are zeros. So

$$Y_{1,j} = q_1 q_2 p^{m-2} q_2 \left(1 - p^{j-1} - p^{j-2}(j-1) q_1\right) \quad \text{if} \quad 1 < j < m. \tag{3.8}$$

   (c) Now, let $i > 1$ and $j = m$. Then $X_{m+1}$ can either be +1 or $-1$.
If $X_{m+1}$ is $-1$, then the remaining elements are arbitrary. However, if $X_{m+1}$ is +1, then in positions $m + 2, \ldots, m + i$ there should be at least one nonzero element. So

$$Y_{i,m} = q_1 q_2 p^{m-2} \left(q_1 \left(1 - p^{i-1}\right) + q_2\right) \quad \text{if} \quad i > 1, \quad j = m. \tag{3.9}$$

   (d) Consider now the case $i > 1$ and $j < m$. We divide this event into two parts.
First, let $X_{m+1} = +1$. Then it is not possible that the elements in positions $m + 2, \ldots, m + i$ are all zeros. It also impossible that there is one $-1$ among $m + 2, \ldots, m + i$ while all $m + i + 1, \ldots, m + j$ are zeros. So this part of $Y_{i,j}$ is

$$q_1 q_2 p^{m-2} q_1 \left(1 - p^{i-1} - (i-1) p^{i-2} q_2 p^{j-i}\right) \quad \text{if} \quad i > 1, \quad j < m. \tag{3.10}$$

Finally, let $X_{m+1} = -1$. Then it is not possible that all elements in positions $m + 2, \ldots, m + j$ are zeros, nor is it possible that among $m + 2, \ldots, m + j$ there is one +1 and the rest are zeros. So this second part of $Y_{i,j}$ is

$$q_1 q_2 p^{m-2} q_2 \left(1 - p^{j-1} - (j-1) q_1 p^{j-2}\right) \quad \text{if} \quad i > 1, \quad j < m. \tag{3.11}$$

Summing (3.5) and (3.6), we get $Y_i^{(+)}$ and consequently by interchanging the roles of $q_1$ and $q_2$ we obtain $Y_j^{(-)}$ as follows:

$$\begin{aligned}
&\sum_{i=1}^{m} Y_i^{(+)} + \sum_{i=1}^{m} Y_i^{(-)} \\
&\quad= \sum_{i=2}^{m-1} q_1^2 p^{m-1}\left(1 - p^{i-1} - (i-1)q_2 p^{i-2}\right) + q_1^2 p^{m-1} \\
&\qquad+ \sum_{i=2}^{m-1} q_2^2 p^{m-1}\left(1 - p^{i-1} - (i-1)q_1 p^{i-2}\right) + q_2^2 p^{m-1} \\
&\quad= (m-1)p^{m-1}\left(q_1^2+q_2^2\right) - \left(q_1^2+q_2^2\right)p^{m-1}\sum_{i=2}^{m-1} p^{i-1} \\
&\qquad- p^{m-1}\left(q_1^2 q_2 + q_2^2 q_1\right)\sum_{i=2}^{m-1}(i-1)p^{i-2} \\
&\quad= p^{m-1}\left\{(q_1^2+q_2^2)\left[(m-1) + \frac{p}{p-1}\right] + q_1 q_2 \frac{1}{p-1}\right. \\
&\qquad\left. - (q_1^2+q_2^2)\frac{p^{m-1}}{p-1} + q_1 q_2\left[(m-2)p^{m-2} - \frac{p^{m-2}}{p-1}\right]\right\}.
\end{aligned} \tag{3.12}$$

Here we applied

$$\sum_{i=a}^{b} i p^{i-1} = \frac{bp^b - ap^{a-1}}{p-1} + \frac{p^a - p^b}{(p-1)^2},$$

which can be obtained by differentiating the known formula for the sum of a geometric sequence. Similarly, summing (3.7), (3.8), (3.9), (3.10) and (3.11) together with their corresponding versions we get by interchanging the roles of $q_1$ and $q_2$, we obtain

$$\begin{aligned}
&\sum_{i<j} Y_{i,j} + \sum_{i>j} Y_{i,j} = q_1 q_2 p^{m-2}(q_1+q_2) \\
&\quad+ q_1 q_2 p^{m-2}\sum_{j=2}^{m-1}\left(q_2\left(1 - p^{j-1} - p^{j-2}(j-1)q_1\right) + q_1\left(1 - p^{j-1} - p^{j-2}(j-1)q_2\right)\right) \\
&\quad+ q_1 q_2 p^{m-2}\left[\sum_{i=2}^{m-1}\left(q_1\left(1-p^{i-1}\right)+q_2\right) + \sum_{i=2}^{m-1}\left(q_2\left(1-p^{i-1}\right)+q_1\right)\right] \\
&\quad+ \sum_{i=2}^{m-2}\sum_{j=i+1}^{m-1}\left[q_1 q_2 p^{m-2} q_1\left(1 - p^{i-1} - (i-1)p^{i-2}q_2 p^{j-i}\right)\right. \\
&\quad+ q_1 q_2 p^{m-2} q_2\left(1 - p^{j-1} - (j-1)q_1 p^{j-2}\right) \\
&\quad+ q_1 q_2 p^{m-2} q_2\left(1 - p^{i-1} - (i-1)p^{i-2}q_1 p^{j-i}\right) \\
&\quad\left.+ q_1 q_2 p^{m-2} q_1\left(1 - p^{j-1} - (j-1)q_2 p^{j-2}\right)\right] \\
&= q_1 q_2 p^{m-2}\left\{(q_1+q_2)\left(m - 1 - \frac{p^{m-1}-p}{p-1}\right) - 2q_1 q_2\sum_{j=2}^{m-1}(j-1)p^{j-2}\right.
\end{aligned}$$

$$
\begin{aligned}
&+2(q_1+q_2)(m-2)-(q_1+q_2)\frac{p^{m-1}-p}{p-1}+\sum_{i=2}^{m-2}\sum_{j=i+1}^{m-1}\Big[(q_1+q_2)(1-p^{i-1})\\
&-2q_1q_2(i-1)p^{j-2}+(q_1+q_2)(1-p^{j-1})-2q_1q_2(j-1)p^{j-2}\Big]\Big\}\\
&=q_1q_2p^{m-2}\Big\{(q_1+q_2)(3m-5)-2(q_1+q_2)\frac{p^{m-1}-p}{p-1}\\
&-2q_1q_2\left(\frac{(m-2)p^{m-2}-1}{p-1}+\frac{p-p^{m-2}}{(p-1)^2}\right)\\
&+\sum_{i=2}^{m-2}\Big[(q_1+q_2)(1-p^{i-1})(m-i-1)-2q_1q_2(i-1)\frac{p^{m-2}-p^{i-1}}{p-1}\\
&+(q_1+q_2)(m-i-1)-(q_1+q_2)\frac{p^{m-1}-p^i}{p-1}\\
&-2q_1q_2\left(\frac{(m-2)p^{m-2}-ip^{i-1}}{p-1}+\frac{p^i-p^{m-2}}{(p-1)^2}\right)\Big]\Big\}.
\end{aligned}
$$

So

$$
\begin{aligned}
&\sum_{i<j}Y_{i,j}+\sum_{i>j}Y_{i,j}\\
&=q_1q_2p^{m-2}\Big\{\ (q_1+q_2)(3m-5)+2(p^{m-1}-p)\\
&-2q_1q_2\left(\frac{(m-2)p^{m-2}-1}{p-1}+\frac{p-p^{m-2}}{(p-1)^2}\right)+(q_1+q_2)(m-3)\frac{m-2}{2}\\
&-(q_1+q_2)m\sum_{i=2}^{m-2}p^{i-1}+(q_1+q_2)\sum_{i=2}^{m-2}p^{i-1}(i+1)\\
&-2q_1q_2\frac{p^{m-2}}{p-1}(m-3)\frac{m-2}{2}+2\frac{q_1q_2}{p-1}\sum_{i=2}^{m-2}(i-1)p^{i-1}+(q_1+q_2)(m-3)\frac{m-2}{2}\\
&-(q_1+q_2)\frac{p^{m-1}}{p-1}(m-3)+\frac{q_1+q_2}{p-1}\sum_{i=2}^{m-2}p^i-\frac{2q_1q_2}{p-1}(m-2)p^{m-2}(m-3)\\
&+\frac{2q_1q_2}{p-1}\sum_{i=2}^{m-2}ip^{i-1}-\frac{2q_1q_2}{(p-1)^2}\sum_{i=2}^{m-2}p^i+\frac{2q_1q_2}{(p-1)^2}p^{m-2}(m-3)\ \Big\}\\
&=q_1q_2p^{m-2}\Big\{(q_1+q_2)(3m-5+(m-3)(m-2))+2\left(p^{m-1}-p\right)\\
&-2q_1q_2\frac{(m-2)p^{m-2}}{p-1}+2q_1q_2\frac{1}{p-1}-\frac{2q_1q_2p}{(p-1)^2}+\frac{2q_1q_2p^{m-2}}{(p-1)^2}-(q_1+q_2)m\frac{p^{m-2}-p}{p-1}\\
&+(q_1+q_2)\left(\frac{1}{p}.\frac{(m-1)p^{m-1}-3p^2}{p-1}+\frac{1}{p}\frac{p^3-p^{m-1}}{(p-1)^2}\right)-2q_1q_2\frac{p^{m-2}}{p-1}\frac{(m-3)(m-2)}{2}\\
&+\frac{2q_1q_2}{p-1}\left(\frac{(m-3)p^{m-3}-1}{p-1}+\frac{p-p^{m-3}}{(p-1)^2}\right)p+p^{m-1}(m-3)-\frac{p^{m-1}-p^2}{p-1}\\
&-\frac{2q_1q_2p^{m-2}}{p-1}(m-2)(m-3)+\frac{2q_1q_2}{p-1}\left(\frac{(m-2)p^{m-2}-2p}{p-1}+\frac{p^2-p^{m-2}}{(p-1)^2}\right)\\
&-\frac{2q_1q_2}{(p-1)^2}\frac{p^{m-1}-p^2}{p-1}+\frac{2q_1q_2}{(p-1)^2}p^{m-2}(m-3)\Big\}.
\end{aligned}
$$

Finally,

$$\sum_{i<j} Y_{i,j} + \sum_{i>j} Y_{i,j} =$$
$$= q_1 q_2 p^{m-2} \Bigg\{ m(m-1)(q_1+q_2) - (m-1) + \frac{2(2p+1)q_1 q_2}{(p-1)^3}$$
$$+ \frac{q_1 q_2 p^{m-2}}{(p-1)^3} \Big( -3(p-1)^2 m^2 + m(p-1)(13p-7)$$
$$+(-14p^2+12p-4) \Big) + p^{m-1}(m-1) \Bigg\}. \qquad (3.13)$$

Therefore, combining (3.12) and (3.13) and by some simplification, we obtain

$$P(A_1 \bar{A}_2 \cdots \bar{A}_m)$$
$$= p^{m-1} \left\{ (q_1^2+q_2^2) \left[ (m-1) + \frac{p}{p-1} \right] + q_1 q_2 \frac{1}{p-1} + \mathrm{O}(mp^m) \right\}$$
$$+ q_1 q_2 p^{m-2} \left\{ m(m-1)(q_1+q_2) - (m-1) + \frac{2(2p+1)q_1 q_2}{(p-1)^3} + \mathrm{O}(m^2 p^m) \right\}$$
$$= m(m-1)p^{m-2} q_1 q_2 \left\{ C_0 + \frac{1}{m} C_1 + \frac{1}{m(m-1)} C_2 + \mathrm{O}(p^m) \right\}.$$

So,

$$P(\bar{A}_2 \cdots \bar{A}_m \mid A_1) = \frac{C_0 + \frac{1}{m} C_1 + \frac{1}{m(m-1)} C_2 + \mathrm{O}(p^m)}{1 + \frac{p(1-p)}{(m-1)q_1 q_2} + \frac{p^2}{m(m-1)q_1 q_2}}.$$

We therefore satisfy from Lemma 3.4

$$|P(\bar{A}_2 \cdots \bar{A}_m | A_1) - \alpha| < \varepsilon,$$

where $\varepsilon = \mathrm{O}(p^m)$ and $\alpha$ is given by (2.1). □

***Proof of Theorem 2.1*** Let $N_1 = N - m + 1$, where $m$ will be specified so that $m \sim \log N$. Then

$$P(\mu(N) < m) = P(\bar{A}_1 \cdots \bar{A}_{N_1}) \sim e^{-(\alpha \pm 10\varepsilon) N_1 P(A_1) \pm 2m P(A_1)}$$
$$= e^{-\alpha N_1 P(A_1)} e^{\pm 10 \varepsilon N_1 P(A_1)} e^{\pm 2m P(A_1)}.$$

As $m P(A_1) \sim m^3 p^m \sim \frac{(\log N)^3}{N}$, so $e^{\pm 2 m P(A_1)} = 1 + \mathrm{O}\left(\frac{(\log N)^3}{N}\right)$. Similarly, as $\varepsilon = \mathrm{O}(p^m)$ and $m \approx \log N$,

$$e^{\pm 10 \varepsilon N_1 P(A_1)} \sim e^{\pm 10 P(A_1)} \sim e^{\pm c (\log N)^2 / N} = 1 + \mathrm{O}\left(\frac{(\log N)^2}{N}\right).$$

Therefore, we can calculate

$$e^{-\alpha N_1 P(A_1)} = e^{-\alpha N P(A_1)} \underbrace{e^{+\alpha (m-1) P(A_1)}}_{1+\mathrm{O}\left(\frac{(\log N)^3}{N}\right)}.$$

So we have to calculate

$$e^{-\alpha N P(A_1)} = e^{-l},$$

where

$$\begin{aligned} l &= \alpha N P(A_1) \\ &= \frac{C_0 + \frac{1}{m} C_1 + \frac{1}{m(m-1)} C_2}{1 + \frac{p(1-p)}{(m-1)q_1q_2} + \frac{p^2}{m(m-1)q_1q_2}} N m(m-1) p^{m-2} (q_1q_2) \\ &\quad \left(1 + \frac{p(1-p)}{(m-1)q_1q_2} + \frac{p^2}{m(m-1)q_1q_2}\right) \\ &= N p^{m-2} q_1 q_2 \left(m^2 C_0 + m(C_1 - C_0) + C_2 - C_1\right). \end{aligned}$$

Our aim is to find $m(N)$ so that the asymptotic behaviour of $P(\mu(N) - [m(N)] < k)$ can be obtained. Then

$$P(\mu(N) - [m(N)] < k) = P(\mu(N) < m(N) + k - \{m(N)\}).$$

Let $m = m(N) + k - \{m(N)\}$. So

$$P(\mu(N) - [m(N)] < k) = P(\mu(N) < m) = e^{-l}\left(1 + \mathrm{O}\left(\frac{(\log N)^3}{N}\right)\right).$$

We want to find $m(N)$ so that the remainder term in the exponent $l$ be small. We shall do it step by step using several Taylor expansions. We try to find $m(N)$ as $\log N + A$, where $A$ will be specified later. So $m = \log N + A + k - \{m(N)\}$.

Then, using the Taylor expansion $\log(x_0 + y) = \log x_0 + \frac{y}{Cx_0} - \frac{1}{2C}\frac{y^2}{x_0^2} + \frac{1}{3C}\frac{y^3}{\tilde{x}^3}$, where $\tilde{x}$ is between $x_0$ and $x_0 + y$ and $x_0 > 0$, $x_0 + y > 0$, we obtain

$$\begin{aligned} L = &\ \log l = \log\left(p^{-2} q_1 q_2\right) - m + \log N + \log(m^2 C_0) + \frac{m(C_1 - C_0) + (C_2 - C_1)}{C m^2 C_0} \\ &- \frac{1}{2C} \frac{(m(C_1 - C_0) + (C_2 - C_1))^2}{(C m^2 C_0)^2} + \mathrm{O}\left(\frac{1}{m^3}\right) \\ = &\ \log\left(p^{-2} q_1 q_2\right) - m + \log N + 2\log m + \log C_0 + \frac{C_1 - C_0}{C C_0 m} + \frac{1}{m^2} K + \mathrm{O}\left(\frac{1}{m^3}\right). \end{aligned}$$

Now since $m = \log N + A + k - \{m(N)\}$, we have

$$\begin{aligned} L = &\ \log\left(C_0 p^{-2} q_1 q_2\right) - \log N - A - (k - \{m(N)\}) + \log N \\ &+ 2\log\left(\log N + A + k - \{m(N)\}\right) + \frac{C_1 - C_0}{C C_0 m} + \frac{1}{m^2} K + \mathrm{O}\left(\frac{1}{m^3}\right). \end{aligned}$$

Again applying the above Taylor expansion, we infer

$$\begin{aligned} L = &\ \log(C_0 p^{-2} q_1 q_2) - A - (k - \{m(N)\}) + 2\log\log N + \frac{2(A + k - \{m(N)\})}{C \log N} \\ &- \frac{1}{C} \frac{(A + k - \{m(N)\})^2}{(\log N)^2} + \frac{2}{3C} \frac{(A + k - \{m(N)\})^3}{(\log N)^3} + \frac{C_1 - C_0}{C C_0 m} + \frac{K}{m^2} + \mathrm{O}\left(\frac{1}{(\log N)^3}\right). \end{aligned}$$

Now, we let $A = 2\log\log N + B$. Then

$$L = \log(C_0 p^{-2} q_1 q_2) - B - (k - \{m(N)\}) + \frac{4\log\log N}{C\log N} + \frac{2B}{C\log N} + \frac{2(k-\{m(N)\})}{C\log N}$$
$$-\frac{1}{C}\frac{A^2}{(\log N)^2} - \frac{2}{C}\frac{A(k-\{m(N)\})}{(\log N)^2} - \frac{1}{C}\frac{(k-\{m(N)\})^2}{(\log N)^2} + \frac{2}{3C}\frac{A^3}{(\log N)^3}$$
$$+\frac{2A^2(k-\{m(N)\})}{C(\log N)^3} + \frac{2A(k-\{m(N)\})^2}{C(\log N)^3} + \frac{C_1 - C_0}{CC_0 m} + \frac{K}{m^2} + \mathrm{O}\left(\frac{1}{(\log N)^3}\right).$$

Let $B = \frac{4\log\log N}{C\log N} + D$. Then

$$L = \log(C_0 p^{-2} q_1 q_2) - D - (k - \{m(N)\}) + \frac{8\log\log N}{C^2(\log N)^2} + \frac{2D}{C\log N} + \frac{2(k-\{m(N)\})}{C\log N}$$
$$-\frac{1}{C}\frac{(2\log\log N + B)^2}{(\log N)^2} - \frac{2}{C}\frac{(2\log\log N + B)(k-\{m(N)\})}{(\log N)^2} - \frac{1}{C}\frac{(k-\{m(N)\})^2}{(\log N)^2}$$
$$+\frac{2}{3C}\frac{(2\log\log N + B)^3}{(\log N)^3} + \frac{2(2\log\log N + B)^2(k-\{m(N)\})}{C(\log N)^3}$$
$$+\frac{2(2\log\log N + B)(k-\{m(N)\})^2}{C(\log N)^3} + \frac{C_1 - C_0}{CC_0 m} + \frac{K}{m^2} + \mathrm{O}\left(\frac{1}{(\log N)^3}\right).$$

We shall use the Taylor expansion $\frac{1}{x_0+x} = \frac{1}{x_0} - \frac{x}{x_0^2} + \frac{x^2}{x_0^3} - \frac{x^3}{\tilde{x}^4}$, where $\tilde{x}$ is between $x_0$ and $x_0 + x$ and where $x_0 > 0$, $x_0 + x > 0$. Since $m = \log N + A + k - \{m(N)\}$,

$$\frac{1}{m} = \frac{1}{\log N} - \frac{A + k - \{m(N)\}}{(\log N)^2} + \frac{(A + k - \{m(N)\})^2}{(\log N)^3} + \mathrm{O}\left(\frac{1}{(\log N)^3}\right)$$

and

$$\frac{1}{m^2} = \frac{1}{(\log N)^2} - \frac{2\log N(A + k - \{m(N)\})}{(\log N)^4} + \mathrm{O}\left(\frac{1}{(\log N)^3}\right).$$

Now let

$$D = \frac{C_1 - C_0}{CC_0}\frac{1}{\log N} + E.$$

Then

$$
\begin{aligned}
L = & \log(C_0 p^{-2} q_1 q_2) - \frac{C_1 - C_0}{C C_0} \frac{1}{\log N} - E - (k - \{m(N)\}) + \frac{8 \log\log N}{C^2 (\log N)^2} \\
& + \frac{2(C_1 - C_0)}{C^2 C_0} \frac{1}{(\log N)^2} + \frac{2E}{C \log N} + \frac{2(k - \{m(N)\}}{C \log N} - \frac{4}{C} \frac{(\log\log N)^2}{(\log N)^2} \\
& - \frac{4B}{C} \frac{\log\log N}{(\log N)^2} - \frac{4}{C} \frac{(\log\log N)(k - \{m(N)\})}{(\log N)^2} - \frac{2B}{C} \frac{(k - \{m(N)\})}{(\log N)^2} \\
& - \frac{1}{C} \frac{(k - \{m(N)\})^2}{(\log N)^2} + \frac{16}{3C} \frac{(\log\log N)^3}{(\log N)^3} + \frac{8B}{C} \frac{(\log\log N)^2}{(\log N)^3} \\
& + \frac{8}{C} \frac{(\log\log N)^2 (k - \{m(N)\})}{(\log N)^3} + \frac{4}{C} \frac{(\log\log N)(k - \{m(N)\})^2}{(\log N)^3} \\
& + \frac{C_1 - C_0}{C C_0} \left( \frac{1}{\log N} - \frac{A + k - \{m(N)\}}{(\log N)^2} + \frac{(2 \log\log N + k - \{m(N)\})^2}{(\log N)^3} \right) \\
& + K \left( \frac{1}{(\log N)^2} - \frac{2(A + k - \{m(N)\})}{(\log N)^3} \right) + \mathrm{O}\left( \frac{1}{(\log N)^3} \right).
\end{aligned}
$$

So we have obtained

$$
\begin{aligned}
L = & \log(C_0 p^{-2} q_1 q_2) - E - (k - \{m(N)\}) + \frac{8 \log\log N}{C^2 (\log N)^2} + \frac{2(C_1 - C_0)}{C^2 C_0} \frac{1}{(\log N)^2} \\
& + \frac{2E}{C \log N} + \frac{2(k - \{m(N)\})}{C \log N} - \frac{4}{C} \frac{(\log\log N)^2}{(\log N)^2} - \frac{1}{C^2} \frac{(4 \log\log N)^2}{(\log N)^3} - \frac{C_1 - C_0}{C^2 C_0} \frac{4 \log\log N}{(\log N)^3} \\
& - \frac{4}{C} \frac{(\log\log N)(k - \{m(N)\})}{(\log N)^2} - \frac{8}{C^2} \frac{(\log\log N)(k - \{m(N)\})}{(\log N)^3} - \frac{1}{C} \frac{(k - \{m(N)\})^2}{(\log N)^2} \\
& + \frac{16}{3C} \frac{(\log\log N)^3}{(\log N)^3} + \frac{8}{C} \frac{(\log\log N)^2 (k - \{m(N)\})}{(\log N)^3} + \frac{4}{C} \frac{\log\log N (k - \{m(N)\})^2}{(\log N)^3} \\
& + \frac{C_1 - C_0}{C C_0} \left( - \frac{2 \log\log N}{(\log N)^2} - \frac{4 \log\log N}{C (\log N)^3} - \frac{k - \{m(N)\}}{(\log N)^2} + \frac{4 (\log\log N)^2}{(\log N)^3} \right. \\
& \left. + \frac{4 \log\log N (k - \{m(N)\})}{(\log N)^3} \right) + \frac{K}{(\log N)^2} - \frac{4K \log\log N}{(\log N)^3} + \mathrm{O}\left( \frac{1}{(\log N)^3} \right).
\end{aligned}
$$

Now let

$$
\begin{aligned}
E = & -\frac{4}{C} \frac{(\log\log N)^2}{(\log N)^2} + \left( \frac{8}{C^2} - \frac{2(C_1 - C_0)}{C C_0} \right) \frac{\log\log N}{(\log N)^2} \\
& + \left( \frac{2(C_1 - C_0)}{C^2 C_0} + K \right) \frac{1}{(\log N)^2} \\
& + \frac{16}{3C} \frac{(\log\log N)^3}{(\log N)^3} + \left( -\frac{16}{C^2} + \frac{4(C_1 - C_0)}{C C_0} \right) \frac{(\log\log N)^2}{(\log N)^3} \\
& - \left( 4K + \frac{8(C_1 - C_0)}{C^2 C_0} \right) \frac{\log\log N}{(\log N)^3} + F.
\end{aligned}
$$

To collect those terms which contain $k - \{m(N)\}$, we use the function $H(x)$ introduced in (2.2). Inserting these expressions, we obtain

$$\begin{aligned} L = &\log(C_0 p^{-2} q_1 q_2) + H(k - \{m(N)\}) \\ &- F + \frac{16}{C^3}\frac{\log\log N}{(\log N)^3} - \frac{8}{C^2}\frac{(\log\log N)^2}{(\log N)^3} \\ &- \frac{4(C_1 - C_0)}{C^2 C_0}\frac{\log\log N}{(\log N)^3} + \mathrm{O}\left(\frac{1}{(\log N)^3}\right). \end{aligned}$$

So choosing

$$F = \frac{16}{C^3}\frac{\log\log N}{(\log N)^3} - \frac{8}{C^2}\frac{(\log\log N)^2}{(\log N)^3} - \frac{4(C_1 - C_0)}{C^2 C_0}\frac{\log\log N}{(\log N)^3},$$

we have

$$L = \log(C_0 p^{-2} q_1 q_2) + H(k - \{m(N)\}) + \mathrm{O}\left(\frac{1}{(\log N)^3}\right).$$

So we obtain

$$\begin{aligned} &P(\mu(N) - [m(N)] < k) \\ &= e^{-l}\left(1 + \mathrm{O}\left(\frac{(\log N)^3}{N}\right)\right) \\ &= e^{-(1/p)^L}\left(1 + \mathrm{O}\left(\frac{(\log N)^3}{N}\right)\right) \\ &= \exp\left(-p^{-(\log(C_0 p^{-2} q_1 q_2) + H(k - \{m(N)\}))}\right)\left(1 + \mathrm{O}\left(\frac{1}{(\log N)^3}\right)\right). \end{aligned}$$

□

## 4 Discussion

In this paper we studied experiments, where two types of failures may occur. Repeating the experiment several times, we considered those runs which contain at most one failure of type I and at most one failure of type II. We were able to find a good approximation for the length distribution of the longest such kind of runs.

**Acknowledgements** The authors would like to thank the referee for a careful reading of the paper and for valuable suggestions.

**Funding** Open access funding provided by University of Debrecen.

## References


1. P. Erdős, A. Rényi, On a new law of large numbers. Anal. Math. **23**, 103–111 (1970)
2. P. Erdős, P. Révész, On the length of the longest head-run, in *Topics in Information Theory. Colloq. Math. Soc. János Bolyai*, vol. 16, (North-Holland, Amsterdam, 1975), pp.219–228
3. A. Földes, The limit distribution of the length of the longest head-run. Period. Math. Hung. **10**(4), 301–310 (1979)
4. L. Gordon, M.F. Schilling, M.S. Waterman, An extreme value theory for long head runs. Probab. Theory Relat. Fields **72**(2), 279–287 (1986)
5. I. Fazekas, M. Suja, Limit theorems for contaminated runs of heads. Ann. Univ. Sci. Budapest Sect. Comp. **52**, 131–146 (2021)
6. E. Csáki, A. Földes, J. Komlós, Limit theorems for Erdős–Rényi type problems. Studia Sci. Math. Hungar. **22**, 321–332 (1987)
7. I. Fazekas, B. Fazekas, M.O. Suja, Convergence rate for the longest T-contaminated runs of heads (paper with detailed proofs). (2023) arXiv:2302.06657
8. S. Eryilmaz, M. Gong, M. Xie, Generalized sooner waiting time problems in a sequence of trinary trials. Stat. Prob. Lett. **115**, 70–78 (2016)
9. M. Koutras, V. Alexandrou, Sooner waiting time problems in a sequence of trinary trials. J. Appl. Probab. **34**(3), 593–609 (1997)
10. I. Fazekas, B. Fazekas, M.O. Suja, Limit theorems for runs containing two types of contaminations (Paper with detailed proofs) (2023). arXiv:2309.11602


**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.