Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 53 (5) (2024), 1408 – 1418
DOI : 10.15672/hujms.1221343

Research Article

On the recent-k-record of discrete random variables

Anshui Li

School of Mathematics,Physics and Information, Shaoxing University, 312000, Shaoxing, Zhejiang, P. R. China

Abstract

Let $X_1,\ X_2,\cdots$ be a sequence of independent and identically distributed random variables which are supposed to be observed in sequence. The nth value in the sequence is a k-record value if exactly k of the first n values (including X_n) are at least as large as it. Let $\mathbf{R}_k$ denote the ordered set of k-record values. The famous Ignatov's Theorem states that the random sets $\mathbf{R}_k(k=1,2,\cdots)$ are independent with common distribution. We introduce one new record named recent-k-record in this paper: X_n is a j-recent-k-record if there are exactly j values at least as large as X_n in $X_{n-k},\ X_{n-k+1},\cdots,\ X_{n-1}$. It turns out that recent-k-record brings many interesting problems and some novel properties such as prediction rule and Poisson approximation are proved in this paper. One application named "No Good Record" via the Lovász Local Lemma is also provided. We conclude this paper with some possible extensions for future work.

Mathematics Subject Classification (2020). 60C05, 60F05, 62A01

Keywords. Ignatov's theorem, recent-k-record, Poisson approximation, the Lovász Local Lemma

1. Introduction

Let $X_1,\ X_2,\cdots$ be a sequence of independent and identically distributed (i.i.d) random variable following the common probability mass function $\mathbb{P}(X=j)=p_j,\ j\in\mathbb{Z}^+$. For a set A, the number of its elements is denoted by $|A|$. Suppose that these random variables are observed one by one and X_n is called a $k-$*record value* if

$$|\{i\in\{1,2,\cdots,n\}:X_i\geq X_n\}|=k.$$

In other words, X_n is one value with exactly k values (including itself) as large as it in the sequence $X_1,\ X_2,\cdots,X_n$. For fixed k, a random ordered set $\mathbf{R}_k$ which includes all the k-record values in the sequence can be defined. In fact, the set

$$\mathbf{R}_1=\{R_1,\ R_2,\ R_3,\cdots\}$$

Email addresses: anshuili@usx.edu.cn (A. Li)
Received: 20.12.2022; Accepted: 31.07.2024

can be regarded as the observation values that are the largest yet seen when they appear, and one can think about the set $\mathbf{R}_2$ of observation values that are the *second* largest on their appearance, and so on. For instance, if the data sequence is

$$X_1 = 2,\ X_2 = 5,\ X_3 = 1,\ X_4 = 6,\ X_5 = 9,\ X_6 = 8,\ X_7 = 3,\ X_8 = 4,\ X_9 = 1,\ X_{10} = 7.$$

Then

$$\mathbf{R}_1 = \{X_1,\ X_2,\ X_4,\ X_5\}, \mathbf{R}_2 = \{X_6\}, \mathbf{R}_3 = \{X_3, X_{10}\}, \mathbf{R}_5 = \{X_7, X_8\}, \mathbf{R}_9 = \{X_9\}.$$

The famous result which is called Ignatov's theorem states that not only do the sequences of k-record values share the same probability distribution for all k, but also these sequences are independent of each other. One can easily identify the $\mathbf{R}_k$ for given sequence observed via one technique used in the proof of the famous Ignatov's Theorem by defining a series of subsequence of the data sequence $X_1,\ X_2, \cdots$, for example, see [13]. Later there are many variants and developments related to this topic, see [3, 6, 9, 10, 12, 14, 17].

In this paper, we will introduce one novel random variable called recent-k-record (RkR) for some fixed integer $k \geq 1$: instead of considering the whole past story, we only consider the values of $X_{n-k},\ X_{n-k+1}, \cdots,\ X_{n-1}$, i.e., the k values before X_n(do not include itself). And let us define X_n be a j-RkR if there are exactly j values at least as large as X_n in $X_{n-k},\ X_{n-k+1}, \cdots,\ X_{n-1}$. In other words, X_n is a j-RkR if

$$|\{p : X_{n-p} \geq X_n, 1 \leq p \leq k\}| = j.$$

We will denote that $i \in R_j^k$ if i is a j-RkR. In other words, there exists a subsequence with length $k+1$ such that $X_{n_0} = i$ and

$$|\{p : X_{n_0-p} \geq i, 1 \leq p \leq k\}| = j.$$

for some $n_0 \geq k+1$. We can consider the usual j-record as one "dynamic version" of j-RkR, i.e., $k = n$ for X_n in that case.

Actually, RkR can be found applications in many areas: for example, to assess one athlete's recent condition and achievements, one proper way is to check the results in his recent records and not necessary to get the whole story(it may be nothing with his records ten years or even five years before). For the k-records application in statistics for athletes, see [15].

The remainder of this paper is organized as follows. In Section 2, we calculate the conditional probability for RkR. The Poisson approximation for RkR and one interesting application via the Lovász Local Lemma are presented in Section 3 and Section 4 respectively. We conclude this paper with some possible extensions for future work in Section 5.

2. Prediction probability for RkR

Theorem 2.1. *Let $X_1,\ X_2, \cdots$ be a sequence of i.i.d random variable following the common probability mass function $\mathbb{P}(X = j) = p_j,\ j \in \mathbb{Z}^+$. Moreover, let $k \geq 1$ and $0 \leq j \leq k, n \geq k+1$, $S_i = \mathbb{P}(X \geq i) = \sum_{s \geq i} p_s$ and $C_i = \mathbb{P}(X \leq i) = \sum_{j=1}^{i} p_j$. We have the following observations*

(i) $\mathbb{P}(i \in R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1})) = \binom{k}{j} S_i^j (C_{i-1})^{k-j} p_i$;

(ii) $\mathbb{P}(i \in R_{j+1}^k \text{ in } (X_2, X_3, \cdots, X_{k+2}) \mid i \in R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1})) = (C_{i-1}) p_i$;

(iii) $\mathbb{P}(i \in R_j^k \text{ in } (X_2, X_3, \cdots, X_{k+2}) \mid i \in R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1})) = S_i p_i$;

(iv) $\mathbb{P}(i \notin R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1})) = \left(1 - \binom{k}{j} S_i^j (C_{i-1})^{k-j}\right) p_i + (1 - p_i)$;

(v) *An upper bound for the probability of the event $A = \{i \in R_j^k \text{ in } (X_1, X_2, \cdots, X_n)\}$ is*

$$\mathbb{P}(A) \leq (n-k) \binom{k}{j} S_i^j (C_{i-1})^{k-j} p_i.$$

Proof. (i) From the definition of j-RkR, the event

$$\{i \in R_j^k \ in \ (X_1, X_2, \cdots, X_{k+1})\}$$

means: $X_{k+1} = i$ and there are j elements in $(X_1, X_2, \cdots, X_k)$ which are not smaller than i.

(ii) From the definition of j-RkR, the event

$$\{i \in R_{j+1}^k \ in \ (X_2, X_3, \cdots, X_{k+2}) \mid i \in R_j^k \ in \ (X_1, X_2, \cdots, X_{k+1})\}$$

means $\{X_1 < i, X_{k+2} = i\}$.

(iii) From the definition of j-RkR, the event

$$\{i \in R_j^k \ in \ (X_2, X_3, \cdots, X_{k+2}) \mid i \in R_j^k \ in \ (X_1, X_2, \cdots, X_{k+1})\}$$

means $\{X_1 \geq i, X_{k+2} = i\}$.

(iv) From the definition of j-RkR, the event

$$\{i \notin R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1})\}$$

$$= \{i \notin R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1}), X_{k+1} = i\} \cup \{i \notin R_j^k \text{ in } (X_1, X_2, \cdots, X_{k+1}), X_{k+1} \neq i\}.$$

(v) It is easy to see that

$$A = \cup_{m=k+1}^n A_m,$$

in which $A_m = \{i \in R_j^k \text{ in } (X_{m-k}, X_{m-k+1}, \cdots, X_{m-1}, X_m)\}$ and then the result can be obtained by

$$\mathbb{P}(A) = \mathbb{P}(\cup_{m=k+1}^n A_m) \leq (n-k)\binom{k}{j} S_i^j (C_{i-1})^{k-j} p_i.$$

□

Next, we present Theorem 2.2, which gives the probability of the nth observation X_n will be some j-RkR, as well as some conditional probability related.

Theorem 2.2. *With same conditions as in Theorem 2.1, we have*

$$\mathbb{P}(X_n \in R_j^k) = \binom{k}{j} \sum_{l=1}^{+\infty} S_l^j (C_{l-1})^{k-j} p_i.$$

As a result, we have

$$q_i = \mathbb{P}(X_n = i \mid X_n \in R_j^k) = \frac{\binom{k}{j} S_i^j (C_{i-1})^{k-j} p_i}{\sum_{l=1}^{+\infty} \binom{k}{j} S_l^j (C_{l-1})^{k-j} p_l}, \quad i = 1, 2, \cdots.$$

Proof. The result is easy to get by conditioning on X_n,

$$\begin{aligned} \mathbb{P}(X_n \in R_j^k) &= \sum_{l=1}^{+\infty} \mathbb{P}(X_n \in R_j^k \mid X_n = l)\mathbb{P}(X_n = l) \\ &= \sum_{l=1}^{+\infty} \binom{k}{j} S_l^j (C_{l-1})^{k-j} p_l \qquad (2.1) \\ &= \binom{k}{j} \sum_{l=1}^{+\infty} S_l^j (C_{l-1})^{k-j} p_l. \end{aligned}$$

And the second result is obtained by Bayes' rule. □

From Theorem 2.2, we can assert that the k random variables

$$R_1^k, R_2^k, \cdots, R_k^k$$

do not have the same distribution and of course they are not independent either. In other words, our result here is completely different with the famous Ignatov's Theorem.

In the following result, we predict X_{n+1} based on the states of X_n for $n \geq k+1$.

Theorem 2.3 (Prediction rule)**.** *With same conditions as in Theorem 2.1,*

$$\mathbb{P}(X_{n+1} \in R_j^k \mid X_n \in R_j^k) = \sum_{i=1}^{+\infty} q_i \left(S_i p_i + p_m \left(\sum_{m>i} \left(\frac{S_m}{S_i}\right)^j \frac{k-j}{k} + \sum_{m<i} \left(\frac{C_{m-1}}{C_{i-1}}\right)^{k-j} \frac{j}{k} \right)\right).$$

Proof.

$$\begin{aligned} &\mathbb{P}(X_{n+1} \in R_j^k \mid X_n \in R_j^k) \\ &= \mathbb{P}(X_{n+1} \in R_j^k, X_n = X_{n+1} \mid X_n \in R_j^k) \\ &+ \mathbb{P}(X_{n+1} \in R_j^k, X_n \neq X_{n+1} \mid X_n \in R_j^k) \end{aligned} \tag{2.2}$$

Then we use the following formula of conditional probability

$$\mathbb{P}(A \mid B) = \sum_{C_i} \mathbb{P}(AC_i \mid B) = \sum_{C_i} \mathbb{P}(A \mid C_i B)\mathbb{P}(C_i \mid B)$$

in which $\{C_i\}_{i\geq 1}$ is a partition of the corresponding sample space Ω.

Then the equation (2.2) can be written by letting $\Omega = \sum_i (X_n = i)$

$$\begin{aligned} &= \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n = X_{n+1}, X_n = i \mid X_n \in R_j^k) \\ &+ \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n \neq X_{n+1}, X_n = i \mid X_n \in R_j^k) \\ &= \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n = X_{n+1} \mid X_n = i, X_n \in R_j^k)\mathbb{P}(X_n = i \mid X_n \in R_j^k) \\ &+ \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n \neq X_{n+1} \mid X_n = i, X_n \in R_j^k)\mathbb{P}(X_n = i \mid X_n \in R_j^k) \\ &= \sum_i \mathbb{P}(X_{n+1} = i \in R_j^k \mid X_n = i \in R_j^k) q_i + \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n \neq X_{n+1} \mid X_n = i \in R_j^k) q_i \\ &= \sum_i \mathbb{P}(X_{n+1} = i \in R_j^k \mid X_n = i \in R_j^k) q_i + \sum_i \sum_{m \neq i} \mathbb{P}(X_{n+1} = m \in R_j^k \mid X_n = i \in R_j^k) q_i \end{aligned} \tag{2.3}$$

Actually, the first part of the formula 2.3 is easy and we have

$$\sum_i \mathbb{P}(X_{n+1} = i \in R_j^k \mid X_n = i \in R_j^k)\mathbb{P}(X_n = i \mid X_n \in R_j^k) = \sum_i S_i p_i q_i. \tag{2.4}$$

Then we will analyze the second part in several steps as follows:

$$\begin{aligned} \sum_{m \neq i} \mathbb{P}(X_{n+1} = m \in R_j^k \mid X_n = i \in R_j^k) = &\sum_{m>i} \mathbb{P}(X_{n+1} = m \in R_j^k \mid X_n = i \in R_j^k) \\ &+ \sum_{m<i} \mathbb{P}(X_{n+1} = m \in R_j^k \mid X_n = i \in R_j^k) \end{aligned} \tag{2.5}$$

Then we discuss the two different cases accordingly:

(i) For the case $m > i$: we have

$$\begin{aligned}
&\sum_{m>i} \mathbb{P}(X_{n+1} = m \in R_j^k \mid X_n = i \in R_j^k) \\
&= \sum_{m>i} \mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} < i \mid X_n = i \in R_j^k) \\
&+ \sum_{m>i} \mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} \geq i \mid X_n = i \in R_j^k)
\end{aligned} \tag{2.6}$$

(a) Conditioning on the event $\{X_{n-k} < i\}$: the event $\{X_n = i \in R_j^k\}$ indicates that there are exactly j elements which are at least as large as $X_n = i$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$; the event $\{X_{n+1} = m \in R_j^k\}$ indicates that there are exactly j elements which are at least as large as $X_{n+1} = m$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$. To sum up: the event $\{X_{n+1} = m \in R_j^k, X_{n-k} < i, X_n = i \in R_j^k\}$ means: there are j elements which are at least as large as $X_{n+1} = m$ and $k-1-j$ elements which are strictly less than i in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$, and $X_{n-k} < i, X_n = i, X_{n+1} = m$. i.e.,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} < i \mid X_n = i \in R_j^k) \\
&= \frac{\mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} < i, X_n = i \in R_j^k)}{\mathbb{P}(X_n = i \in R_j^k)} \\
&= \frac{\binom{k-1}{j} S_m^j (C_{i-1})^{k-1-j} C_{i-1} p_i p_m}{\binom{k}{j} S_i^j (C_{i-1})^{k-j} p_i} \\
&= \left(\frac{S_m}{S_i}\right)^j \left(\frac{k-j}{k}\right) p_m
\end{aligned} \tag{2.7}$$

(b) Conditioning on the event $\{X_{n-k} \geq i\}$: the event $\{X_n = i \in R_j^k\}$ indicates that there are exactly $j-1$ elements which are at least as large as $X_n = i$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$; the event $\{X_{n+1} = m \in R_j^k\}$ and $X_n = i < m$ indicates that there are exactly j elements which are at least as large as $X_{n+1} = m > i$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$. This is not possible, so the probability is zero.

(ii) For the case $m < i$: we have

$$\begin{aligned}
&\sum_{m<i} \mathbb{P}(X_{n+1} = m \in R_j^k \mid X_n = i \in R_j^k) \\
&= \sum_{m<i} \mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} < i \mid X_n = i \in R_j^k) \\
&+ \sum_{m<i} \mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} \geq i \mid X_n = i \in R_j^k)
\end{aligned} \tag{2.8}$$

(a) Conditioning on the event $\{X_{n-k} < i\}$: the event $\{X_n = i \in R_j^k\}$ indicates that there are exactly j elements which are at least as large as $X_n = i$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$, which means there will be $j+1$ elements which are as large as m in $X_{n-k+1}, X_{n-k+2}, \cdots, X_n$, contradicting the event $\{X_{n+1} = m \in R_j^k\}$.

(b) Conditioning on the event $\{X_{n-k} \geq i\}$: the event $\{X_n = i \in R_j^k\}$ indicates that there are exactly $j-1$ elements which are at least as large as $X_n = i$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$; the event $\{X_{n+1} = m \in R_j^k\}$ indicates that there are exactly $j-1$ elements which are at least as large as $X_{n+1} = m$ in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$. To sum up: the event $\{X_{n+1} = m \in R_j^k, X_{n-k} < i, X_n = i \in R_j^k\}$ means: there are $j-1$ elements which are at least as large as $X_n = i > m$ and $k-1-j$ elements which are strictly less than m in $X_{n-k+1}, X_{n-k+2}, \cdots, X_{n-1}$, and $X_{n-k} \geq i, X_n = i, X_{n+1} = m$. i.e.,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} \geq i \mid X_n = i \in R_j^k) \\
&= \frac{\mathbb{P}(X_{n+1} = m \in R_j^k, X_{n-k} \geq i, X_n = i \in R_j^k)}{\mathbb{P}(X_n = i \in R_j^k)} \\
&= \frac{\binom{k-1}{j-1} S_i^{j-1} (C_{m-1})^{k-j} S_i p_i p_m}{\binom{k}{j} S_i^j (C_{i-1})^{k-j} p_i} \\
&= \left(\frac{C_{m-1}}{C_{i-1}}\right)^{k-j} \left(\frac{j}{k}\right) p_m
\end{aligned} \tag{2.9}$$

Finally, we put all the pieces together, we can have

$$\begin{aligned}
\mathbb{P}(X_{n+1} \in R_j^k \mid X_n \in R_j^k) &= \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n = X_{n+1}, X_n = i \mid X_n \in R_j^k) \\
&\quad + \sum_i \mathbb{P}(X_{n+1} \in R_j^k, X_n \neq X_{n+1}, X_n = i \mid X_n \in R_j^k) \\
&= \sum_i q_i \left(S_i p_i + p_m \left(\sum_{m>i} \left(\frac{S_m}{S_i}\right)^j \frac{k-j}{k} + \sum_{m<i} \left(\frac{C_{m-1}}{C_{i-1}}\right)^{k-j} \frac{j}{k} \right)\right).
\end{aligned} \tag{2.10}$$

□

Remark 2.4. Actually, the more general case of the conditional probability

$$\mathbb{P}(X_{n+1} \in R_{j_1}^k \mid X_n \in R_{j_2}^k)$$

for $j_1 \neq j_2$ is a little complicated. To see this fact, we have

$$\begin{aligned}
\mathbb{P}(X_{n+1} \in R_{j_1}^k \mid X_n \in R_{j_2}^k) &= \frac{\mathbb{P}(X_{n+1} \in R_{j_1}^k, X_n \in R_{j_2}^k)}{\mathbb{P}(X_n \in R_{j_2}^k)} \\
&= \frac{\sum_{i=1}^\infty \mathbb{P}(X_{n+1} \in R_{j_1}^k, X_n \in R_{j_2}^k, X_n = i)}{\sum_{i=1}^\infty \mathbb{P}(X_n \in R_{j_2}^k, X_n = i)} \\
&= \frac{\sum_{i=1}^\infty \mathbb{P}(X_{n+1} \in R_{j_1}^k, X_n \in R_{j_2}^k \mid X_n = i)\mathbb{P}(X_n = i)}{\sum_{i=1}^\infty \mathbb{P}(X_n \in R_{j_2}^k \mid X_n = i)\mathbb{P}(X_n = i)} \\
&= \frac{\sum_{i=1}^\infty \mathbb{P}(X_{n+1} \in R_{j_1}^k, X_n \in R_{j_2}^k \mid X_n = i) p_i}{\sum_{i=1}^\infty \binom{k}{j_2} S_i^{j_2} C_{i-1}^{k-j_2} p_i} \\
&= \frac{\sum_{i=1}^\infty \mathbb{P}(X_{n+1} \in R_{j_1}^k, X_n = i \in R_{j_2}^k) p_i}{\sum_{i=1}^\infty \binom{k}{j_2} S_i^{j_2} C_{i-1}^{k-j_2} p_i} \\
&= \frac{\sum_{i=1}^\infty \sum_{j=1}^\infty \mathbb{P}(X_{n+1} = j \in R_{j_1}^k, X_n = i \in R_{j_2}^k) p_i p_j}{\sum_{i=1}^\infty \binom{k}{j_2} S_i^{j_2} C_{i-1}^{k-j_2} p_i}.
\end{aligned} \tag{2.11}$$

The probability $\mathbb{P}(X_{n+1} = j \in R_{j_1}^k, X_n = i \in R_{j_2}^k)$ is very difficult to obtain and have no neat expression.

3. Poisson approximation for R_j^k

The asymptotic properties of sum of random variables are very important in probability and statistics. It is well known that convergence to a Poisson distribution can occur if the individual means of Bernoulli random variable are all small even if they are not independent, more detailed information can be found in [8]. In this section, we will give the Poisson approximation for R_j^k using the Stein-Chen method, see [7].

We will give the definition of dependency graph first and then give the Poisson approximation Lemma based the dependency graph.

3.1. Dependency graph in general and Poisson approximation Lemma

Let (I, E) be a graph with finite or countable vertex set I and edge set E. For $i, j \in I$, we denote $i \sim j$ if $(i, j) \in E$. For $i \in I$, let $\mathcal{N}_i = \{i\} \cup \{j \in I : i \sim j\}$. The graph $(I, \sim)$ is called a dependency graph for a collection of random variables $(\xi_i, i \in I)$ if for any two disjoint subsets I_1, I_2 of I such that there are no edges connecting I_1 to I_2, the collection of random variables $\{\xi_i, i \in I_1\}$ is independent of $\{\xi_i, i \in I_2\}$. The notion of dependency graphs gives a very useful to express some rare-independence, which is a technique to generalize the independence.

The Lemma below gives the total variance of two distributions by Stein-Chen technique with the help of the dependency graphs.

Lemma 3.1 ([5]). *Suppose $\{\xi_i, i \in I\}$ is a finite collection of Bernoulli random variables with dependency graph $(I, \sim)$. Set $p_i := \mathbb{P}(\xi_i = 1) = \mathbb{E}(\xi_i)$, and set $p_{ij} := \mathbb{E}(\xi_i \xi_j)$. Let $\lambda := \sum_{i \in I} p_i$, and suppose λ is finite, let $W := \sum_{i \in I} \xi_i$. Then*

$$d_{TV}(W, Po(\lambda)) \leq \min(3, \lambda^{-1}) \left(\sum_{i \in I} \sum_{j \in \mathcal{N}(i) \setminus \{i\}} p_{ij} + \sum_{i \in I} \sum_{j \in \mathcal{N}(i)} p_i p_j \right).$$

In which $d_{TV}(\xi, \eta) = \sup_{A \subset \mathbb{Z}} \mid \mathbb{P}(\xi \in A) - \mathbb{P}(\eta \in A) \mid$ for two integer-valued random variables ξ, η and $Po(\lambda)$ is the Poisson distribution with parameter λ.

3.2. Poisson approximation for RkR

For fixed $i_0 \in \mathbb{Z}^+$, we define a series of random variables as follows:

$$\xi_i = \begin{cases} 1 & \text{if } i_0 \in R_j^k \ \text{ in } \ (X_i, X_{i+1}, \cdots, X_{i+k}), \\ 0 & \text{else.} \end{cases}$$

Which means the random variables ξ_i are indexed by the $(k+1)$-set $\{X_i, X_{i+1}, \cdots, X_{i+k}\}$. For instance, set $i_0 = 3, k = 3, j = 1$ and the data sequence is

$$X_1 = 2,\ X_2 = 5,\ X_3 = 1,\ X_4 = 1,\ X_5 = 2,\ X_6 = 8,\ X_7 = 3,\ X_8 = 2,\ X_9 = 1,\ X_{10} = 3.$$

Then we have

$$\xi_4 = \xi_7 = 1; \quad \xi_j = 0, \text{ for } j \notin \{4, 7\}.$$

There are many interesting properties on these random variables ξ_i.

Theorem 3.2. *For fixed $i_0 \in \mathbb{Z}^+$, we have the following results:*

(i) $\mathbb{E}(\xi_i) = \mathbb{P}(\xi_i = 1) = \binom{k}{j} S_{i_0}^j (1 - S_{i_0})^{k-j} p_{i_0}$;

(ii) $\mathbb{E}(\xi_i \xi_{i+1}) = \binom{k-1}{j-1} S_{i_0}^j (1 - S_{i_0})^{k-j} p_{i_0}^2$;

(iii) *For $|i_1 - i_2| = m > k$,*

$$\mathbb{P}(\xi_{i_1} = 1, \xi_{i_2} = 1) = \mathbb{P}(\xi_{i_1} = 1)\mathbb{P}(\xi_{i_2} = 1), \qquad \mathbb{E}(\xi_{i_1} \xi_{i_2}) = \mathbb{E}(\xi_{i_1})\mathbb{E}(\xi_{i_2});$$

(iv) *For $|i_1 - i_2| = m \in \{1, 2, \cdots, k\}$,*

$$\phi_m = \mathbb{E}(\xi_{i_1}\xi_{i_2})$$
$$= \sum_{t=\max\{0,j-m-1\}}^{\min\{k-m,j-1\}} \binom{m}{j-t} S_{i_0}^{j-t}(1-S_{i_0})^{m-j+t} \binom{m-1}{j-t-1} S_{i_0}^{j-t-1}(1-S_{i_0})^{m-j+t}$$
$$\times \binom{k-m}{t} S_{i_0}^{t}(1-S_{i_0})^{k-m-t} p_{i_0}^2 \tag{3.1}$$
$$= \sum_{t=\max\{0,j-m-1\}}^{\min\{k-m,j-1\}} \binom{m}{j-t}\binom{m-1}{j-t-1}\binom{k-m}{t} S_{i_0}^{2j-t-1}(1-S_{i_0})^{m-2j+t+k} p_{i_0}^2.$$

Proof. (i) It is easy to see that $\{\xi_i = 1\}$ means that: $X_{i+k} = i_0$ and there at j random variables in $X_i,\ X_{i+1}, \cdots,\ X_{i+k-1}$ which are not smaller than i_0. That is,

$$\mathbb{E}(\xi_i) = \mathbb{P}(\xi_i = 1) = \binom{k}{j} S_{i_0}^{j}(1-S_{i_0})^{k-j} p_{i_0}.$$

(ii) It is easy to see that $\{\xi_i = 1, \xi_{i+1} = 1\}$ means that: $X_{i+k} = X_{i+k+1} = i_0$, and there are $j-1$ random variables in $X_i + 1,\ X_{i+1}, \cdots,\ X_{i+k-1}$ which are not smaller than i_0 while X_i should also not smaller than i_0. In other words,

$$\mathbb{P}(\xi_i = 1, \xi_{i+1} = 1) = S_{i_0}\binom{k-1}{j-1} S_{i_0}^{j-1}(1-S_{i_0})^{k-j} p_{i_0}^2.$$

(iii) From definition of ξ_i, we can see that the value of ξ_i is determined by the status of $X_i,\ X_{i+1}, \cdots,\ X_{i+k}$. To be specific, the value of random variable ξ_{i_1} depends on $X_{i_1}, X_{i_1+1}, \cdots, X_{i_1+k}$ and ξ_{i_2} depends on $X_{i_2}, X_{i_2+1}, \cdots, X_{i_2+k}$. $X_{i_1}, X_{i_1+1}, \cdots, X_{i_1+k}, X_{i_2}, X_{i_2+1}, \cdots, X_{i_2+k}$ are *i.i.d.* when $|i_1 - i_2| > k$.

(iv) It is easy to see that $\{\xi_{i_1} = 1\}$ means $X_{i_1+k} = i_0$ and there are j values which are not not smaller than i_0 in $X_{i_1}, X_{i_1+1}, \cdots, X_{i_1+k-1}$; the event $\{\xi_{2_1} = 1\}$ means $X_{i_2+k} = i_0$ and there are j values which are not not smaller than i_0 in $X_{i_2}, X_{i_2+1}, \cdots, X_{i_2+k_1}$. The events $\{\xi_{i_1} = 1\}$ and $\{\xi_{i_2} = 1\}$ are not independent when $|i_1 - i_2| = m \le k$ since they both depend on the status of $X_{i_2} = X_{i_1+m}, \cdots, X_{i_1+k}$. We classify the event $\{\xi_{i_1} = 1, \xi_{i_2} = 1\}$ into different cases by the number of values $t$which are not smaller than i_0 in $X_{i_1+m}, \cdots, X_{i_1+k-1}$. To be more precise, if there are t variables which are not smaller than i_0, $\{\xi_{i_1} = 1, \xi_{i_2} = 1\}$ implies there are $j-t$ values which are not smaller than i_0 in $X_{i_1}, \cdots, X_{i_1+m-1}$ and there are $j-t-1$ values which are not smaller than i_0 in $X_{i_1+k+1}, \cdots, X_{i_2+k-1}$ as well as $X_{i_1+k} = X_{i_2+k} = i_0$. The result is obtained by summing all the different cases.

□

We then define the dependency graph for RkR as follows:

Let $\mathcal{I}_n$ be the set of all $(k+1)$-sets $\{X_i, X_{i+1}\cdots, X_{i+k}\}$ of $\{X_1, X_2, \cdots, X_{n+k}\}$. It is easy to see that the size of $\mathcal{I}_n$ is n. For each element $\mathbf{i} \in \mathcal{I}_n$, let $\mathcal{N}_i$ be the set of $\mathbf{j} \in \mathcal{I}_n$ such that $\mathbf{i}$ and $\mathbf{j}$ have at least one element in common. And let $\mathbf{i} \sim \mathbf{j}$ if $\mathbf{j} \in \mathcal{N}_i$ but $\mathbf{i} \ne \mathbf{j}$. In other words, $\mathcal{N}_i = \{\mathbf{i}\} \cup \{\mathbf{j} \in \mathcal{I}_n : \mathbf{i} \sim \mathbf{j}\}$.Then ξ_i is independent of ξ_j except when $\mathbf{j} \in \mathcal{N}_i$, and as a result the graph $(\mathcal{I}_n, \sim)$ is a dependency graph for $\xi_i,\ i = 1, 2, \cdots, n$.

As a consequence of Lemma 3.1 and Theorem 3.2, we can get the following result easily:

Theorem 3.3. *Let $i_0 \in \mathbb{Z}^+$ be fixed, then the number of $i_0 \in R_j^k$ in the sequence $(X_1, X_2, \cdots, X_{n+k})$ is $\xi = \sum_{i=1}^{n} \xi_i$, which has an asymptotic Poisson distribution with* *parameter* $\lambda := n\binom{k}{j} S_{i_0}^j (1 - S_{i_0})^{k-j} p_{i_0}$. *To be more precise, we have*

$$d_{TV}(\xi, Po(\lambda)) \le n \min\{3, \lambda^{-1}\} \left(\sum_{s=1}^{k} \phi_s + (k+1)p^2 \right).$$

In which $\phi_s = \mathbb{E}(\xi_{i_1}\xi_{i_2})$ *when* $|i_1 - i_2| = s \le k$ *and* $p = \binom{k}{j} S_{i_0}^j (1 - S_{i_0})^{k-j} p_{i_0}$.

Proof. It is easy to get that $\mathbb{P}(\xi_i = 1) = \binom{k}{j} S_{i_0}^j (1 - S_{i_0})^{k-j} p_{i_0} = p$, leading to

$$\lambda = \mathbb{E}(\xi) = \sum_{i=1}^{n} \mathbb{E}(\xi_i) = np.$$

We then get

$$\sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{N}(i) \setminus \{i\}} p_{ij} = \sum_{i \in \mathcal{I}_n} \sum_{s=1}^{k} \phi_s = n \sum_{s=1}^{k} \phi_s$$

and

$$\sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{N}(i)} p_i p_j = n \sum_{j \in \mathcal{N}(i)} \mathbb{E}(\xi_i = 1)\mathbb{E}(\xi_j = 1) = n(k+1)p^2.$$

Then by Lemma 3.1, we complete the proof. □

4. "No Good Record" via the Lovász Local Lemma

Let E_i be the event that $i_0 \in R_j^k$ in $(X_i, X_{i+1}, \cdots, X_{i+k}), (i = 1, 2, \cdots, n)$. In this section, we will show that there are positive probability that i_0 will not be one RkR in the sequence $\{X_1, X_2, \cdots, X_{n+k}\}$, i.e., the events $\cap_{i=1}^n \bar{E}_i$ can happen with positive probability once p_{i_0} is chosen properly. Our result bases mainly on one version of the famous Lovász Local Lemma which can be checked in [11].

Lemma 4.1 (Lovász Local Lemma)**.** *Let* $E_1, \cdots, E_n$ *be a set of events, and assume that the following hold:*

(i) *for all* i, $\mathbb{P}(E_i) \le p$*;*
(ii) *the degree of the dependency graph given by* $E_1, \cdots, E_n$ *is bounded by* d*;*
(iii) $4dp \le 1$.

Then

$$\mathbb{P}(\cap_{i=1}^n \bar{E}_i) > 0.$$

Our result goes as follows:

Theorem 4.2 ("No Good Result" Theorem)**.** *Let* $E_i = \{i_0 \in R_j^k$ *in* $(X_i, X_{i+1}, \cdots, X_{i+k})\}$, *where* $i = 1, 2, \cdots, n$. *There exists some* $p_{i_0} > 0$, *such that*

$$\mathbb{P}(\cap_{i=1}^n \bar{E}_i) > 0.$$

In other words, there exists some one with "no good record" in the whole story with positive probability.

Proof. Let $p_i = \mathbb{E}(E_i)$, then we have

$$\begin{aligned}
4kp_i &= 4k\binom{k}{j} S_{i_0}^j (1-S_{i_0})^{k-j} p_{i_0} \\
&\le 4k\left(\frac{ke}{j}\right)^j S_{i_0}^j (1-S_{i_0})^{k-j} p_{i_o} \\
&= 4k\left(\frac{ke}{j}\right)^j \left(\frac{jS_{i_0} + (k-j)(1-S_{i_0})}{k}\right)^k p_{i_0} \\
&= 4k\left(\frac{ke}{j}\right)^j \left(\frac{k-j+S_{i_0}(2j-k)}{k}\right)^k p_{i_0} \\
&\le 4k\left(\frac{ke}{j}\right)^j \max\left\{\left(\frac{k-j}{k}\right)^k, \left(\frac{j}{k}\right)^k\right\} p_{i_0}. \\
&:= C(k,j)p_{i_0}.
\end{aligned} \tag{4.1}$$

Since $C(k,j)$ is some constant depending on k, j, one can choose p_{i_0} accordingly to make sure

$$C(k,j)p_{i_0} < 1.$$

□

5. Conclusion

Records and related problems are very interesting topics in applied probability as well as general mathematics, see [1,4]. One novel record named recent-k-record was introduced in this paper and some interesting properties of the recent-k-record were explored. It will be glad to see more variants of records in the classical case and their corresponding statistical properties like ones mentioned in [2,16] in the future.

Acknowledgements

We would like to thank the anonymous referees and the editor for many helpful comments that have greatly improved the paper.

Author contributions. All the co-authors have contributed equally in all aspects of the preparation of this submission.

Conflict of interest statement. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. This work is supported by National Natural Science Foundation of China (Grant No.11901145).

Data availability. No data was used for the research described in the article.

References

[1] M. Ahsanullah and V.B. Nevzorov, *Records via probability theory*, Atlantis, 2015.

[2] G.R. AL-Dayian, A.A. EL-Helbawy, H. Hozaien and E. Gawdat, *Estimation of the topp-leone alpha power weibull distribution based on lower record values*, Comput. J. Math. Stat. Sci. **3** (1), 145-160, 2024.

[3] D. Aldous and P. Diaconis, *Longest increasing subsequences: from patience sorting to the baik-deift-johansson theorem*, Bull. Am. Math. Soc. **36** (4), 413-432, 1999.

[4] B.C. Arnold, N. Balakrishnan and H.N. Nagaraja, *Records*, John Wiley & Sons, 2011.

[5] R. Arratia, L. Goldstein and L. Gordon, *Two moments suffice for poisson approximations: The chen-stein method*, Ann. Probab. **17** (1), 9-25, 1989.

[6] J. Bunge and C.M. Goldie, *Record sequences and their applications*, Handbook of Statistics **19**, 277-308, 2001.

[7] L.H. Chen, L. Goldstein and Q. Shao, *Normal approximation by Steins method*, Springer Science & Business Media, 2010.

[8] A. DasGupta, *Asymptotic theory of statistics and probability*, Springer, New York, 2008.

[9] R. Engelen, P. Tommassen and W. Vervaat, *Ignatov's theorem: A new and short proof*, J. Appl. Probab. **26** (A), 229-236, 1988.

[10] A.V. Gnedin, *Best choice from the planar poisson process*, Stoch. Process. Their Appl. **111** (2), 317-354, 2002.

[11] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*, Cambridge University Press, 2005.

[12] E.A. Pekoz, *Ignatov's theorem and correlated record values*, Stat. Res. Lett. **43** (2), 107-111, 1999.

[13] S. M. Ross, *Introduction to Probability Models: Eleventh Edition*, Elsevier, 2014.

[14] G. Simons, L. Yang and Y. Yao, *Doob, ignatov and optional skipping*, Ann. Probab. **30** (4), 1933-1958, 2002.

[15] R.L. Smith, *Statistics for exceptional athletics records*, J. R. Stat. Soc. **46** (1), 123-128, 1997.

[16] Z. Vidović, J. Nikolić and Z. Perić, *Properties of k-record posteriors for the weibull model*, Stat. Theory Relat. Fields, 1-11, 2024.

[17] Y. C. Yao, *On independence of k-record processes: Ignatovs̈ theorem revisited*, Ann. Probab. **7** (3), 815-821, 1997.