J. Appl. Math. & Informatics Vol. **42**(2024), No. 3, pp. 709 - 723
https://doi.org/10.14317/jami.2024.709

MODIFIED GEOMETRIC DISTRIBUTION OF ORDER k AND ITS APPLICATIONS

JUNGTAEK OH, KYEONG EUN LEE*

Abstract. We study the distributions of waiting times in variations of the geometric distribution of order k. Variation imposes length on the runs of successes and failures. We study two types of waiting time random variables. First, we consider the waiting time for a run of k consecutive successes the first time no sequence of consecutive k failures occurs prior, denoted by $T^{(k)}$. Next, we consider the waiting time for a run of k consecutive failures the first time no sequence of k consecutive successes occurred prior, denoted by $J^{(k)}$. In addition, we study the distribution of the weighted average. The exact formulae of the probability mass function, mean, and variance of distributions are also obtained.

AMS Mathematics Subject Classification: 05A99, 60C05, 62E15.
Key words and phrases : Geometric distribution of order k, modified geometric distribution of order k, runs and patterns statistics, weighted average.

1. Introduction

The distribution theory of runs and patterns has been considerably developed in the last decades owing to its theoretical relevance and applications in various research areas, such as hypothesis testing, system reliability, quality control, physics, psychology, radar astronomy, molecular biology, computer science, insurance, and finance. In the past few decades, meaningful progress has been demonstrated on runs and related statistics in [3] as well as in [6]. More recently, contributions on the topic have been reported, such as [2], [5], [7], [8], and [1]. Waiting time distributions related to the runs of Bernoulli trials have recently received immense interest in applied probability. One of the well-known and extensively researched waiting time distributions is the geometric distribution of order k, which is defined as the distribution of the number of trials until the first

Received January 23, 2024. Revised May 3, 2024. Accepted May 10, 2024. *Corresponding author.

occurrence of k consecutive successes in Bernoulli trials with success probability p. This definition is as per [10].
In weighted averages, some data points have a stronger effect than others, i.e., the relative importance or frequencies of some factors or the varying degrees of the numbers in a data set are incorporated. To calculate the weighted average, each number in the data set is multiplied by weights that predetermine the relative importance of each data point. A weighted average can be more accurate than a simple average, in which all numbers in a data set are treated equally and assigned equal weights.
The remainder of this paper is organized as follows. In Section 2, we introduce basic definitions and necessary notations that will be useful throughout this article. In Section 3, we study the waiting time for a run of k consecutive successes the first time no sequence of k failures occurred prior. We derive the exact probability mass function (PMF), mean, and variance of $T^{(k)}$ using combinatorial analysis. In Section 4, we study the waiting time for a run of k consecutive failures the first time no sequence of k successes occurs prior. Using combinatorial analysis, we derive the exact PMF, mean, and variance of $J^{(k)}$. In Section 5, we study the weighted average of $T^{(k)}$ and $J^{(k)}$.

2. Preliminaries

We first recall some definitions and notations used throughout this paper. We suppose that $0 < q < 1$. First, we introduce the following notations:

- $L_n^{(1)}$: the length of the longest run of successes in $X_1, X_2, \ldots, X_n$;
- $L_n^{(0)}$: the length of the longest run of failures in $X_1, X_2, \ldots, X_n$;
- S_n : the total number of successes in $X_1, X_2, \ldots, X_n$;
- F_n : the total number of failures in $X_1, X_2, \ldots, X_n$.

We consider a modified geometric distribution of order k. We study the two types of waiting time random variables, which represent the waiting time for a run of k consecutive successes the first time no sequence of k failures occurs prior and the waiting time for a run of k consecutive failures the first time no sequence of k successes occurs prior, denoted by $T^{(k)}$ and $J^{(k)}$, respectively. We obtain the recursive scheme for the random variables $T^{(k)}$ and $J^{(k)}$. More specifically, by taking advantage of the recursive scheme, we obtain nonrecursive formulae for the computation of the mean and variance of the random variables $T^{(k)}$ and $J^{(k)}$. Let us consider the random variable X to be a p parameter Bernoulli random variable, which is independent of $T^{(k)}$. To obtain a run of k consecutive successes, we must perform a run of $k-1$ consecutive successes. This means that we have already waited for $T^{(k-1)}$. Then, we consider two cases:

(1) If we obtain a new success event with probability p, we have k consecutive successes. In this case, $T^{(k)}$ is given by $T^{(k-1)}+1$, with probability p. That is given by the parcel $X(T^{(k-1)}+1)$. Note that $X = 1$ with probability p.

(2) If we obtain a failure event with probability $q = 1 - p$, we must wait for a new complete sequence of k consecutive successes. That is represented by the parcel $(1-X)(T^{(k-1)}+1+T^{(k)})$. Note that $X = 0$ with probability $q = 1 - p$.

We obtain the following recursive scheme satisfied by $T^{(k)}$ from the above mentioned cases:

$$T^{(k)} \stackrel{d}{=} X(T^{(k-1)} + 1) + (1 - X)(T^{(k-1)} + 1 + T^{(k)}). \tag{1}$$

Similarly, we obtain the following recursive scheme satisfied by $J^{(k)}$:

$$J^{(k)} \stackrel{d}{=} (1 - X)(J^{(k-1)} + 1) + X(J^{(k-1)} + 1 + J^{(k)}). \tag{2}$$

Further, the (1) and (2) can be used for the recursive evaluation of the expectation of $T^{(k)}$ and $J^{(k)}$ respectively. Furthermore, using a weighted average, we consider a new random variable

$$W_a(k) = aT^{(k)} + (1 - a)J^{(k)}, \ 0 < a < 1.$$

Before proceeding with the main result, note that the total number of integer solutions is $x_1 + x_2 + \cdots + x_a = c$ such that $0 < x_i < b$ for $i = 1, 2, \ldots, a$. Alternatively, the number of ways of distributing c identical balls into a different cells with no containing more or equal than b balls is described using [4]

$$S(a,\ b,\ c) = \sum_{j=0}^{min(a,\left[\frac{c-a}{b-1}\right])} (-1)^j \binom{a}{j} \binom{c - j(b-1) - 1}{a - 1}.$$

3. Distribution of $T^{(k)}$

Let us begin our study of the modified geometric distribution of order k by presenting some results related to the closed formulae for the PMF, mean, and variance of the random variables $T^{(k)}$. More specifically, our derivations for finding the mean and variance of the random variables $T^{(k)}$ are mainly based on the recursive scheme for the random variables $T^{(k)}$. The exact formulae for PMF, mean, and variance of the distributions are obtained as follows:

3.1. PMF of $T^{(k)}$.

This section derives the PMF of $T^{(k)}$. The following theorem presents the PMF of $T^{(k)}$.

Theorem 3.1. *The PMF* $f_T^{(1)}(n) = P(T^{(k)} = n)$ *for* $n \geq k$ *is given by* $f_T^{(1)}(n) = p^k$ *and*

$$f_T^{(1)}(n) = \sum_{i=1}^{n-k} p^{n-i}(1-p)^i \sum_{s=1}^{i} S(s,\ k,\ i)\Big[S(s-1, k, n-k-i) + S(s, k, n-k-i)\Big],\ n > k.$$

Proof. We use arguments similar to those used in [9]. We begin with the study of $f_T^{(1)}(n)$. Clearly, $f_T^{(1)}(k) = p^k$. We now assume $n > k$ and write $f_T^{(1)}(n)$ as follows:

$$f_T^{(1)}(n) = P\left(L_{n-k}^{(1)} < k \ \wedge \ L_{n-k}^{(0)} < k \ \wedge \ X_{n-k} = 0 \ \wedge \ X_{n-k+1} = \cdots = X_n = 1\right).$$

We partition the event $T^{(k)} = n$ into disjoint events given by $F_{n-k} = i$, for $i = 1, \ldots, n-k$. Adding the probabilities, we obtain

$$\begin{aligned} f_T^{(1)}(n) = \sum_{i=1}^{n-k} P\Big(&L_{n-k}^{(1)} < k \ \wedge \ L_{n-k}^{(0)} < k \ \wedge \ F_{n-k} = i \ \wedge \ X_{n-k} = 0 \ \wedge \\ &X_{n-k+1} = \cdots = X_n = 1\Big). \end{aligned}$$

If the number of 0's in the first $n-k$ trials is equal to i, that is, $F_{n-k} = i$, then in each of the $(n-k+1)$ to n-th trials, the probability of success is

$$p_{n-k+1} = \cdots = p_n = p.$$

If we use $E_{n,i}^{(0)}$ to represent the event $\left\{L_n^{(1)} < k \ \wedge \ L_n^{(0)} < k \ \wedge \ X_n = 0 \ \wedge \ F_n = i\right\}$, we obtain

$$\begin{aligned} f_T^{(1)}(n) &= \sum_{i=1}^{n-k} P\Big(L_{n-k}^{(1)} < k \ \wedge \ L_{n-k}^{(0)} < k \ \wedge \ F_{n-k} = i \ \wedge \ X_{n-k} = 0\Big) \\ &\qquad \times P\Big(X_{n-k+1} = \cdots = X_n = 1\Big) \\ &= \sum_{i=1}^{n-k} P\Big(E_{n-k,i}^{(0)}\Big) p^k. \end{aligned}$$

We focus on the event $E_{n-k,i}^{(0)}$. For $i = 1, \ldots, n-k$, a typical element of the event $\{L_{n-k}^{(1)} < k \wedge L_{n-k}^{(0)} < k \wedge F_{n-k} = i\}$ is an ordered sequence consisiting of $n-k-i$ successes and i failures with the longest success and failure runs having lengths less than k. We can derive the number of these sequences as follows: first, we distribute the i failures. Let s $(1 \le s \le i)$ be the number of failure runs in the typical element of the event $E_{n-k,i}^{(0)}$. Next, we distribute the $n-k-i$ successes. We divide it into two cases: starting with a failure run or starting with a success run. Thus, we distinguish between the two types of sequences in the event $\left\{L_{n-k}^{(1)} < k \wedge \ L_{n-k}^{(0)} < k \wedge \ F_{n-k} = i \wedge X_{n-k} = 0\right\}$, named $(s-1,\ s)$-type and $(s,\ s)$-type, respectively, which are defined as follows:

$$(s-1,\ s)\text{-type} \ : \quad \overbrace{0\ldots0}^{y_1} \mid \overbrace{1\ldots1}^{x_1} \mid \overbrace{0\ldots0}^{y_2} \mid \overbrace{1\ldots1}^{x_2} \mid \ldots \mid \overbrace{0\ldots0}^{y_{s-1}} \mid \overbrace{1\ldots1}^{x_{s-1}} \mid \overbrace{0\ldots0}^{y_s},$$

with i 0's and $n-k-i$ 1's, where x_j $(j = 1, \ldots, s-1)$ represents the length of a run of 1's and y_j $(j = 1, \ldots, s)$ represents the length of a run of 0's. Further, all integers $x_1, \dots, x_{s-1}$ and $y_1, \dots, y_s$ satisfy the conditions

$$0 < x_j < k \text{ for } j = 1, ..., s-1, \text{ and } x_1 + \cdots + x_{s-1} = n - k - i,$$

$$0 < y_j < k \text{ for } j = 1, ..., s, \text{ and } y_1 + \cdots + y_s = i.$$

$$(s,\ s)\text{-type} \ : \quad \overbrace{1 \dots 1}^{x_1} \mid \overbrace{0 \dots 0}^{y_1} \mid \overbrace{1 \dots 1}^{x_2} \mid \overbrace{0 \dots 0}^{y_2} \mid \overbrace{1 \dots 1}^{x_3} \mid \dots \mid \overbrace{0 \dots 0}^{y_{s-1}} \mid \overbrace{1 \dots 1}^{x_s} \mid \overbrace{0 \dots 0}^{y_s},$$

with i 0's and $n-k-i$ 1's, where x_j $(j = 1, \dots, s)$ represents the length of a run of 1's and y_j $(j = 1, \dots, s)$ represents the length of a run of 0's. Here, all of $x_1, \dots, x_s$ and $y_1, \dots, y_s$ are integers, and satisfy

$$0 < x_j < k \text{ for } j = 1, ..., s, \text{ and } x_1 + \cdots + x_s = n - k - i,$$

$$0 < y_j < k \text{ for } j = 1, ..., s, \text{ and } y_1 + \cdots + y_s = i.$$

Then, the probability of the event $E^{(0)}_{n-k,i}$ is given by

$$P\Big(E^{(0)}_{n-k,i}\Big) = p^{n-k-i} q^i \sum_{s=1}^{i} S(s,\ k,\ i) \Big[S(s-1,\ k,\ n-k-i) + S(s,\ k,\ n-k-i) \Big].$$

Therefore, we can compute the probability of the event $W^{(1)}_S = n$ as follows:

$$f^{(1)}_T(n) = \sum_{i=1}^{n-k} p^{n-i} q^i \sum_{s=1}^{i} S(s,\ k,\ i) \Big[S(s-1,\ k,\ n-k-i) + S(s,\ k,\ n-k-i) \Big].$$

Thus, the proof is completed. □

3.2. Closed formulae for the mean of $T^{(k)}$. In this section, we derive the closed form of the expectation of $T^{(k)}$. We first derive the closed form of the expectation of $T^{(2)}$ prior.

Lemma 3.2. *The closed form of the expectation $E[T^{(2)}]$ is given by*

$$E\big[T^{(2)}\big] = \frac{2p^2}{(1-pq)^2} + \frac{qp^2}{1-qp} + \frac{2qp^2}{(1-qp)^2}. \tag{3}$$

Proof. First, we consider the number of cases where a sequence of two consecutive successes occurs for the first time on the nth trial and no sequence of two consecutive failures occurs before the nth trial for $n = 2, 3, \dots$. Examples of sequences are "SS" for $n = 2$, "FSS" for $n = 3$, "SFSS" for $n = 4$, "FSFSS" for $n = 5$, "SFSFSS" for $n = 6$, "FSFSFSS" for $n = 7$, etc. Using this argument, we obtain the expectation of $T^{(2)}$ as follows.

$$E[T^{(2)}] = 2p^2 + 3qp^2 + 4pqp^2 + 5qpqp^2 + 6pqpqp^2 + 7qpqpqp^2 + \cdots \tag{4}$$

We divide (4) into odd and even-numbered terms. Applying typical mathematical algebraic arguments, we obtain the expectation of $T^{(2)}$. Thus, the proof is completed. □

Next, using Lemma 3.2, we derive the closed form of the expectation of $T^{(k)}$.

Theorem 3.3. *The closed form of the expectation of $T^{(k)}$ is given by*

$$E[T^{(k)}] = \frac{1}{p^k}\left(\frac{p^k - p^2}{1-p} + \frac{p^4[2+q(3-pq)]}{(1-qp)^2}\right). \tag{5}$$

Proof. Taking the expectation of (1), using the linearity of expectations, and through rearrangement, we obtain: $E[T^{(k)}] = E[T^{(k-1)}]+1+E[T^{(k)}]-pE[T^{(k)}]$. Applying typical mathematical algebraic arguments, we obtain the recurrence relation of the expectation of $T^{(k)}$ as follows:

$$E[T^{(k)}] = \frac{1}{p}E[T^{(k-1)}] + \frac{1}{p}. \tag{6}$$

The closed form of the expectation of $T^{(k)}$ can be deduced by iterating (6) and applying typical mathematical algebraic arguments as follows:

$$E[T^{(k)}] = \frac{1}{p^{k-2}}E[T^{(2)}] + \frac{1}{p^{k-2}} + \frac{1}{p^{k-3}} + \cdots + \frac{1}{p^2} + \frac{1}{p}$$

Using Lemma 3.2, we obtain the closed form of the expectation of $T^{(k)}$; thus, the proof is completed. □

The following gives the generating function of the means $E[T^{(k)}]$.

Corollary 3.4. *The generating function of the means $E[T^{(k)}]$ is given by*

$$\sum_{k=0}^{\infty} E\big[T^{(k)}\big]z^k = \frac{A(p,q)z - B(p,q)}{(q+p^2)^2(z-p)(z-1)}, \tag{7}$$

where

$$A(p,q) = 1 - pq(1-p-2p^3+4p^4-p^5+p^6) \text{ and}$$
$$B(p,q) = pq^2(1+p+2p^2+4p^3+p^5).$$

We multiply both sides of the generating function by the denominator of the right-hand side (RHS) and perform classical analysis on the resulting power series to yield the following expression for the mean $v_k = E[T^{(k)}]$ that satisfies the recurrence relation:

$$v_k = \frac{1+p}{p}v_{k-1} - \frac{1}{p}v_{k-2},\ k \geq 2. \tag{8}$$

Equation (7) may also be used to develop nonrecursive expressions for v_k. In particular, using the geometric series for

$$\left(1-\frac{z}{p}\right)^{-1} = \sum_{i=0}^{\infty}\left(\frac{z}{p}\right)^i \text{ and } (1-z)^{-1} = \sum_{j=0}^{\infty} z^j, \tag{9}$$

we obtain

$$\sum_{k=0}^{\infty} E\left[T^{(k)}\right]z^k = \left[\frac{A(p,q)z - B(p,q)}{(q+p^2)^2(z-p)(z-1)}\right]\sum_{k=0}^{\infty} c_k z^k, \tag{10}$$

where

$$c_k = \sum_{j=0}^{k} \frac{p^j}{p^k} = \frac{1-p^{k+1}}{p^k(1-p)}. \tag{11}$$

Thus,

$$v_k = \frac{A(p,q)}{p(q+p^2)^2} c_{k-1} - \frac{B(p,q)}{p(q+p^2)^2} c_k \text{ for all } k \geq 2.$$

3.3. Closed formulae for the variance of $T^{(k)}$. To obtain the closed form of the expectation of $Var[T^{(k)}]$, we derive $Var[T^{(2)}]$. We know that $Var[T^{(2)}] = E[(T^{(2)})^2] - (E[T^{(2)}])^2$, from the definition of variance. First, we obtain $E[(T^{(2)})^2]$ and derive the variance of $T^{(2)}$.

Lemma 3.5. *The closed form of the expectation of* $\left(T^{(2)}\right)^2$ *is given by*

$$E\left[\left(T^{(2)}\right)^2\right] = \frac{4p^2}{(1-pq)^3} - \frac{4p^2}{(1-pq)^2} + \frac{p^2q}{1-pq} + \frac{8p^2q}{(1-qp)^3} \tag{12}$$

Proof. First, we consider the number of cases where a sequence of two consecutive successes occurs for the first time on the nth trial and no sequence of two consecutive failures occurs before the nth trial for $n = 2, 3, \ldots$. Examples of sequences are "SS" for $n = 2$, "FSS" for $n = 3$, "SFSS" for $n = 4$, "FSFSS" for $n = 5$, "SFSFSS" for $n = 6$, "FSFSFSS" for $n = 7$, etc. Using this argument, we obtain the expectation of $E[(T^{(2)})^2]$ as follows:

$$E\left[(T^{(2)})^2\right] = 2^2p^2 + 3^2qp^2 + 4^2pqp^2 + 5^2qpqp^2 + 6^2pqpqp^2 + \cdots \tag{13}$$

We divide (13) into odd and even-numbered terms and then apply typical mathematical algebraic arguments to obtain the closed form of the expectation of $\left(T^{(2)}\right)^2$. Thus, the proof is completed. □

Next, we derive the closed form of the variance of $T^{(2)}$ using Lemma 3.5.

Lemma 3.6. *The closed form of the variance of* $T^{(2)}$ *is given by*

$$Var\left[T^{(2)}\right] = \frac{p^2q(p^7 - 2p^6 + 6p^5 - 11p^4 + 16p^3 - 11p^2 - 7p + 13}{(1-pq)^4}. \tag{14}$$

Proof. Using Lemmas 3.2 and 3.5 and applying typical mathematical algebraic arguments, we obtain

$$\begin{aligned} Var\left[T^{(2)}\right] =& \frac{8p^2 + 8p^2q}{(1-pq)^3} - \frac{4p^2}{(1-pq)^2} + \frac{p^2q}{1-pq} - \frac{4p^4}{(1-pq)^4} - \frac{p^4q^2}{(1-qp)^2} \\ & - \frac{4p^4q^2}{(1-qp)^4} - \frac{4p^4q}{(1-qp)^3} - \frac{4p^4q^2}{(1-qp)^3} - \frac{8p^4q}{(1-qp)^4}. \end{aligned}$$

Thus, the proof is completed. □

We now obtain the closed form of the variance of $T^{(k)}$.

Theorem 3.7. *The closed form of the variance of $T^{(k)}$ is given by*

$$V\left[T^{(k)}\right] = \frac{A(p,q)}{p^{2k}q^2(1-pq)^4}, \tag{15}$$

where

$$\begin{aligned} A(p,q) =& -p^{2k}(1-pq)^4 + p^5\{-1+p(2+2p-7p^2+5p^3-3p^4+p^5)\}^2 \\ &+ p^{k+2}q\{5-2k+(8k-20)p+(43-10k)p^2-(6k+27)p^3 \\ &+ (40k-98)p^4+(274-72k)p^5+(76k-326)p^6 \\ &+ (256-56k)p^7+(28k-140)p^8+(55-10k)p^9 \\ &+ (2k-15)p^{10}+2p^{11}\}. \end{aligned}$$

Proof. We begin with the study of $E[(T^{(k)})^2]$. First, by squaring both sides of the Eq.(1), taking expectation, expanding the right side, and applying the linearity of expectations, we obtain

$$\begin{aligned} E\left[(T^{(k)})^2\right] =& E\left[\{X(T^{(k-1)}+1)\}^2\right] + E\left[\{(1-X)(T^{(k-1)}+1+T^{(k)})\}^2\right] \\ &+ 2E\left[X(1-X)(T^{(k-1)}+1)(T^{(k-1)}+1+T^{(k)})\right]. \end{aligned}$$

Since either $X=0$ or $(1-X)=0$, we have

$$E\left[(T^{(k)})^2\right] = E\left[\{X(T^{(k-1)}+1)\}^2\right] + E\left[\{(1-X)(T^{(k-1)}+1+T^{(k)})\}^2\right].$$

As X and $T^{(k)}$ are independent, we obtain

$$\begin{aligned} E\left[(T^{(k)})^2\right] =& E\left[X^2\right]E\left[(T^{(k-1)}+1)^2\right] + E\left[(1-X)^2\right]E\left[(T^{(k-1)}+1+T^{(k)})^2\right] \\ =& pE\left[(T^{(k-1)}+1)^2\right] + (1-p)E\left[(T^{(k-1)}+1+T^{(k)})^2\right] \end{aligned}$$

Expanding the squared terms, applying the linearity of expectations, and algebraically simplifying the expression, we obtain

$$\begin{aligned} E\left[(T^{(k)})^2\right] =& E\left[(T^{(k-1)})^2+2T^{(k-1)}+1\right] \\ &+ (1-p)E\left[(T^{(k)})^2+2T^{(k)}+2T^{(k-1)}\cdot T^{(k)}\right] \end{aligned} \tag{16}$$

Next, we focus on $\left(E[T^{(k)}]\right)^2$, taking the expectation of (1), and applying typical mathematical algebraic arguments, we have

$$\begin{aligned} \left(E[T^{(k)}]\right)^2 =& \left(E[T^{(k-1)}]\right)^2 + 1 + (1-p)^2\left(E[T^{(k)}]\right)^2 + 2E[T^{(k-1)}] \\ &+ 2(1-p)E[T^{(k)}] + 2(1-p)E[T^{(k-1)}]\cdot E[T^{(k)}]. \end{aligned} \tag{17}$$

Using (16) and (17) and algebraically simplifying the expression, we obtain

$$\begin{aligned}Var\left[T^{(k)}\right] =&E\left[(T^{(k-1)})^2\right] - \left(E[T^{(k-1)}]\right)^2 + (1-p)E\left[(T^{(k)})^2\right]\\ &- (1-p)^2\left(E[T^{(k)}]\right)^2.\end{aligned}$$

Using the definition of variance, we can rewrite the above equations as follows:

$$Var[T^{(k)}] = Var[T^{(k-1)}] + (1-p)Var[T^{(k)}] + p(1-p)\left(E[T^{(k)}]\right)^2. \tag{18}$$

The following recursive scheme is deduced using (18):

$$Var\left[T^{(k)}\right] =\frac{1}{p}Var\left[T^{(k-1)}\right] + (1-p)\left(E[T^{(k)}]\right)^2. \tag{19}$$

We iterate (19), and obtain the closed form of the variance of $T^{(k)}$ as follows:

$$Var\left[T^{(k)}\right] = \frac{1}{p^{k-2}}Var\left[T^{(2)}\right] + (1-p)\sum_{i=0}^{k-3}\frac{1}{p^i}\left(E[T^{(k-i)}]\right)^2.$$

Using Theorem 3.7 and Lemma 3.6, we obtain the closed form of the variance of $T^{(k)}$ as follows:

$$\begin{aligned}Var\left[T^{(k)}\right] =&\frac{1}{p^{k-2}}\left[\frac{4p^2+8p^2q}{(1-pq)^3} - \frac{4p^2}{(1-pq)^2} + \frac{p^2q}{1-pq} - \frac{4p^4}{(1-pq)^4}\right.\\ &\left. - \frac{p^4q^2}{(1-qp)^2} - \frac{4p^4q^2}{(1-qp)^4} - \frac{4p^4q}{(1-qp)^3} - \frac{4p^4q^2}{(1-qp)^3} - \frac{8p^4q}{(1-qp)^4}\right]\\ &+ \frac{(1-p)}{p^{2k}}\sum_{i=0}^{k-3}p^i\left[\frac{p^{k-i}-p^2}{1-p} + \frac{p^4\{2+q(3-pq)\}}{(1-pq)^2}\right]^2.\end{aligned}$$

Thus, the proof is completed. □

4. Distribution of $J^{(k)}$

Let us begin our study of the modified geometric distribution of order k by presenting some results related to the closed formulae of the PMF, mean, and variance of the random variables $J^{(k)}$. More specifically, our derivations for finding the mean and variance of the random variables $J^{(k)}$ are mainly based on the recursive scheme of the random variables $J^{(k)}$. The exact formulae of the PMF, mean, and variance of distributions are obtained as follows.

4.1. PMF of $J^{(k)}$. In this section, we shall study PMF of $J^{(k)}$. The following theorem presents the PMF of $J^{(k)}$.

Theorem 4.1. *The PMF* $f_J^{(0)}(n) = P(J^{(k)} = n)$ *for* $n \geq k$ *is given by* $f_J^{(0)}(n) = q^k$ *and*

$$f_J^{(0)}(n) = \sum_{i=1}^{n-k} p^{n-k-i}\ q^{i+k} \sum_{s=1}^{n-k-i} S(s,k,n-k-i)\Big[S(s-1,k,i) + S(s,k,i)\Big].$$

Proof. We use arguments similar to those used in [9]. We begin with the study of $f_J^{(0)}(n)$. Clearly, $f_J^{(0)}(k) = q^k$. We now assume $n > k$ and write $f_J^{(0)}(n)$ as follows:

$$f_J^{(0)}(n) = P\left(L_{n-k}^{(1)} < k\ \wedge\ L_{n-k}^{(0)} < k\ \wedge\ X_{n-k} = 1\ \wedge\ X_{n-k+1} = \cdots = X_n = 0\right).$$

We partition the event $J^{(k)} = n$ into disjoint events given by $F_{n-k} = i$, for $i = 1, \ldots, n-k$. Adding the probabilities, we have

$$f_J^{(0)}(n) = \sum_{i=1}^{n-k} P\Big(L_{n-k}^{(1)} < k\ \wedge\ L_{n-k}^{(0)} < k\ \wedge\ F_{n-k} = i\ \wedge\ X_{n-k} = 1\ \wedge$$
$$X_{n-k+1} = \cdots = X_n = 0\Big).$$

If we use $E_{n,i}^{(1)}$ to represent the event $\Big\{L_n^{(1)} < k\ \wedge\ L_n^{(0)} < k\ \wedge\ X_n = 1\ \wedge\ F_n = i\Big\}$, we obtain

$$\begin{aligned} f_J^{(0)}(n) &= \sum_{i=1}^{n-k} P\Big(L_{n-k}^{(1)} < k\ \wedge\ L_{n-k}^{(0)} < k\ \wedge\ F_{n-k} = i\ \wedge\ X_{n-k} = 1\Big) \\ &\qquad \times P\Big(X_{n-k+1} = \cdots = X_n = 0\Big) \\ &= \sum_{i=1}^{n-k} P\Big(E_{n-k,i}^{(1)}\Big) q^{i+k}. \end{aligned}$$

We focus on the event $E_{n-k,i}^{(1)}$. For $i = 1, \ldots, n-k$, a typical element of the event $\{L_{n-k}^{(1)} < k \wedge\ L_{n-k}^{(0)} < k \wedge\ F_{n-k} = i\}$ is an ordered sequence that consists of $n-k-i$ successes and i failures in a way that the longest success and failure runs have lengths shorter than k. We can derive the number of these sequences as follows.: first, we distribute the i failures. Let s $(1 \leq s \leq i)$ be the number of failure runs in the typical element of the event $E_{n-k,i}^{(1)}$. Next, we distribute the $n-k-i$ successes. We divide it into two cases: starting with a success run and starting with a failure run. Thus, we distinguish between the two types of sequences in the event $\Big\{L_{n-k}^{(1)} < k \wedge\ L_{n-k}^{(0)} < k \wedge\ F_{n-k} = i\Big\}$, named $(s-1,\ s)$-type and $(s,\ s)$-type, respectively, which are defined as follows:

$$(s,\ s-1)\text{-type}\ :\quad \overbrace{1\ldots1}^{x_1} \mid \overbrace{0\ldots0}^{y_1} \mid \overbrace{1\ldots1}^{x_2} \mid \overbrace{0\ldots0}^{y_2} \mid \ldots \mid \overbrace{0\ldots0}^{y_{s-1}} \mid \overbrace{1\ldots1}^{x_s},$$

with i 0's and $n-k-i$ 1's, where x_j $(j=1,\dots,s)$ represents the length of a run of 1's and y_j $(j=1,\dots,s-1)$ represents the length of a run of 0's. Further, all integers $x_1,\dots,x_s$ and $y_1,\dots,y_{s-1}$ satisfy the conditions

$$0<x_j<k \text{ for } j=1,...,s, \text{ and } x_1+\cdots+x_s=n-k-i,$$

$$0<y_j<k \text{ for } j=1,...,s-1, \text{ and } y_1+\cdots+y_{s-1}=i.$$

$$(s,\ s)\text{-type}\ :\ \overbrace{0\dots0}^{y_1}\,|\,\overbrace{1\dots1}^{x_1}\,|\,\overbrace{0\dots0}^{y_2}\,|\,\overbrace{1\dots1}^{x_2}\,|\,\overbrace{0\dots0}^{y_3}\,|\,\dots\,|\,\overbrace{0\dots0}^{y_s}\,|\,\overbrace{1\dots1}^{x_s},$$

with i 0's and $n-k-i$ 1's, where x_j $(j=1,\dots,s)$ represents the length of a run of 1's and y_j $(j=1,\dots,s)$ represents the length of a run of 0's. Here, all of $x_1,\dots,x_s$ and $y_1,\dots,y_s$ are integers and satisfy

$$0<x_j<k \text{ for } j=1,...,s, \text{ and } x_1+\cdots+x_s=n-k-i,$$

$$0<y_j<k \text{ for } j=1,...,s, \text{ and } y_1+\cdots+y_s=i.$$

Then, the probability of the event $E^{(1)}_{n-k,i}$ is given by

$$P\Big(E^{(1)}_{n-k,i}\Big)=p^{n-k-i}\ q^i\sum_{s=1}^{n-k-i}S(s,\ k,\ n-k-i)\Big[S(s-1,\ k,\ i)+S(s,\ k,\ i)\Big].$$

Therefore, we can compute the probability of the event $J^{(k)}=n$ as follows:

$$\begin{aligned} f_J^{(0)}(n)&=\sum_{i=1}^{n-k}P\Big(E^{(1)}_{n-k,i}\Big)q^k\\ &=\sum_{i=1}^{n-k}p^{n-k-i}\ q^{i+k}\sum_{s=1}^{n-k-i}S(s,\ k,\ n-k-i)\Big[S(s-1,\ k,\ i)+S(s,\ k,\ i)\Big]. \end{aligned}$$

Thus, the proof is completed. □

4.2. Closed formulae for the mean of $J^{(k)}$. In this section, we derive the closed form of the expectation of $J^{(k)}$. We first derive the closed form of the expectation of $J^{(2)}$.

Lemma 4.2. *The closed form of the expectation of $J^{(2)}$ is given by*

$$E\left[J^{(2)}\right]=\frac{2q^2}{(1-qp)^2}+\frac{pq^2}{1-pq}+\frac{2pq^2}{(1-pq)^2}. \tag{20}$$

Proof. We use arguments similar to those used in Lemma 3.2 for the proof. □

Next, using Lemma 4.2, we derive the closed form of the expectation of $T^{(k)}$.

Theorem 4.3. *The closed form of the expectation of $E[J^{(k)}]$ is given by*

$$E\left[J^{(k)}\right]=\frac{1}{q^k}\left[\frac{q^k-q^2}{1-q}+\frac{q^4\{2+p(3-pq)\}}{(1-pq)^2}\right], \tag{21}$$

Proof. We use arguments similar to those used in Theorem 3.3. □

The following gives the generating function of the mean $E[J^{(k)}]$.

Corollary 4.4. *The generating function of the mean* $E[J^{(k)}]$ *is given by*

$$\sum_{k=0}^{\infty} E\left[J^{(k)}\right] z^k = \frac{C(p,q)z + D(p,q)}{(q+p^2)^2(z-q)(z-1)}, \tag{22}$$

where

$$C(p,q) = 1 - pq(1-p-2p^3+4p^4-p^5+p^6) \text{ and}$$
$$D(p,q) = p^2q(p^5-5p^4+14p^3-24p^2+22p-9).$$

We multiply both sides of the generating function by the denominator of the RHS and perform classical analysis on the resulting power series to yield the following expression for the mean $u_k = E[J^{(k)}]$ that satisfies the recurrence relation:

$$u_k = \frac{1+q}{q} u_{k-1} - \frac{1}{q} u_{k-2}, \; k \geq 2. \tag{23}$$

Equation (23) may also be used to develop nonrecursive expressions for u_k. In particular, using the geometric series for

$$\left(1 - \frac{z}{q}\right)^{-1} = \sum_{i=0}^{\infty} \left(\frac{z}{q}\right)^i \text{ and } (1-z)^{-1} = \sum_{j=0}^{\infty} (z)^j, \tag{24}$$

we obtain

$$\sum_{k=0}^{\infty} E\left[J^{(k)}\right] z^k = \Big[C(p,q)z + D(p,q)\Big] \sum_{k=0}^{\infty} b_k z^k, \tag{25}$$

where

$$C(p,q) = \frac{\{1 - pq(1-p-2p^3+4p^4-p^5+p^6)\}}{q(q+p^2)^2},$$
$$D(p,q) = \frac{p^2q(p^5-5p^4+14p^3-24p^2+22p-9)}{q(q+p^2)^2} \text{ and}$$
$$b_k = \sum_{j=0}^{k} \frac{q^j}{p^k} = \frac{1-q^{k+1}}{q^k(1-q)}.$$

Thus

$$u_k = \frac{\{1 - pq(1-p-2p^3+4p^4-p^5+p^6)\}}{q(q+p^2)^2} b_{k-1} + \frac{p^2q(p^5-5p^4+14p^3-24p^2+22p-9)}{q(q+p^2)^2} b_k,$$

for all $k \geq 2$.

4.3. Closed formulae of the variance of $J^{(k)}$. To obtain the closed form of the expectation of $Var[J^{(k)}]$, we need to derive of $Var[J^{(2)}]$. We already known that $Var[J^{(2)}] = E[(J^{(2)})^2] - (E[J^{(2)}])^2$ from the definition of variance. First, we obtain $E[(J^{(2)})^2]$ and derive the variance of $J^{(2)}$.

Lemma 4.5. *The closed form of the expectation of $E[(J^{(2)})^2]$ is given by*

$$E\left[\left(J^{(2)}\right)^2\right] = \frac{8q^2}{(1-qp)^3} - \frac{4q^2}{(1-qp)^2} + \frac{pq^2}{1-pq} + \frac{8pq^2}{(1-pq)^3} \tag{26}$$

Proof. We use arguments similar to those used in Theorem 3.5. □

Next, we derive the closed form of the variance of $T^{(2)}$ using Lemma 4.5.

Lemma 4.6. *The closed form of the variance of $J^{(2)}$ is given by*

$$Var\left[J^{(2)}\right] = \frac{pq^2\left\{5 + p^2(22 - 27p + 24p^2 - 15p^3 + 5p^4 - p^5)\right\}}{(1-pq)^4}. \tag{27}$$

Proof. We use arguments similar to those used in Lemma 3.6. □

We now obtain the closed form of the variance of $J^{(k)}$.

Theorem 4.7. *The closed form of the variance of $J^{(k)}$ is given by*

$$V\left[J^{(k)}\right] = \frac{B(p,q)}{p^2q^{2k}(1-pq)^4}, \tag{28}$$

where

$$\begin{aligned} B(p,q) = &-q^{2k}(1-pq)^4 + q^5(1-4p+4p^2+3p^3-5p^4+3p^5-p^6)^2 \\ &+ pq^{2+k}\{9-2k-2(23-6k)p+5(19-6k)p^2-(71-38k)p^3 \\ &-2(9+10k)p^4+4(18-5k)p^5-12(5-4k)p^6+24(1-2k)p^7 \\ &+2(5+14k)p^8-5(3+2k)p^9+(7+2k)p^{10}-2p^{11}\} \end{aligned}$$

Proof. We use arguments similar to those used in Theorem 3.7. □

5. Weighted average of T_k and J_k

This section studies the weighted average of T_k and J_k. Let $W_a(k) = aT^{(k)} + (1-a)J^{(k)}$, $0 < a < 1$. Next, we obtain the expectation and variance of $W_a(k)$. First, we study the expectation of $W_a(k)$.

Theorem 5.1. *For $0 < a < 1$, the expectation of $W_a(k)$ is given by*

$$\begin{aligned} E\left[W_a(k)\right] = &\frac{a}{p^{k-2}}\left[\frac{2p^2}{(1-pq)^2} + \frac{p^2q}{1-pq} + \frac{2p^2q}{(1-pq)^2} + \frac{1-p^{k-2}}{1-p}\right] \\ &+ \frac{(1-a)}{q^{k-2}}\left[\frac{2q^2}{(1-pq)^2} + \frac{pq^2}{1-pq} + \frac{2pq^2}{(1-pq)^2} + \frac{1-q^{k-2}}{1-q}\right]. \end{aligned}$$

Proof. Applying the linearity of expectations, we obtain $E[aT^{(k)} + (1-a)J^{(k)}] = aE[T^{(k)}] + (1-a)E[J^{(k)}]$, which can be proved using Theorems 3.3 and 4.3. □

Next, we obtain the variance of $W_a(k)$.

Theorem 5.2. *For $0 < a < 1$, the variance of $W_a(k)$ is given by*

$$Var\,[W_a(k)] = \frac{a^2 A(p,q) + (1-a)^2 B(p,q)}{p^2 q^{2k}(1-pq)^4}, \tag{29}$$

where

$$\begin{aligned} A(p,q) = &- p^{2k}(1-pq)^4 + p^5\{-1 + p(2 + 2p - 7p^2 + 5p^3 - 3p^4 + p^5)\}^2 \\ &+ p^{k+2}q\{5 - 2k + (8k-20)p + (43-10k)p^2 - (6k+27)p^3 \\ &+ (40k-98)p^4 + (274-72k)p^5 + (76k-326)p^6 \\ &+ (256-56k)p^7 + (28k-140)p^8 + (55-10k)p^9 + (2k-15)p^{10} + 2p^{11}\}. \end{aligned}$$

and

$$\begin{aligned} B(p,q) = &- q^{2k}(1-pq)^4 + q^5(1 - 4p + 4p^2 + 3p^3 - 5p^4 + 3p^5 - p^6)^2 \\ &+ pq^{2+k}\{9 - 2k - 2(23-6k)p + 5(19-6k)p^2 - (71-38k)p^3 \\ &- 2(9+10k)p^4 + 4(18-5k)p^5 - 12(5-4k)p^6 + 24(1-2k)p^7 \\ &+ 2(5+14k)p^8 - 5(3+2k)p^9 + (7+2k)p^{10} - 2p^{11}\} \end{aligned}$$

Proof. Based on the definition of variance, we have $Var[W_a(k)] = E[\{(aT^{(k)} + (1-a)J^{(k)}) - (a\mu_k + (1-a)\nu_k)\}^2]$. Because E is linear, e.g., $E[aX + b] = aE[X] + b$. Hence, $Var[W_a(k)] = a^2E[(T^{(k)} - \mu_k)^2] + (1-a)^2E[(J^{(k)} - \nu_k)^2] + a(1-a)E[(T^{(k)} - \mu_k)(J^{(k)} - \nu_k)]$. As the random variables $T^{(k)}$ and $J^{(k)}$ are independent, $Var[W_a(k)] = a^2Var[T^{(k)}] + (1-a)^2Var[J^{(k)}]$, which can be proved using Theorems 3.7 and 4.7. □

Conflicts of interest : The authors declare no conflict of interest.

Acknowledgments: The author would like to thank Emeritus Prof. Dr. Chongjin Park whose comments led to significant improvements in this manuscript.

References

1. S. Aki, *Waiting time for consecutive repetitions of a pattern and related distributions*, Ann. Inst. Statist. Math. **71** (2019), 307-325.
2. A.N. Arapis, F.S. Makri and Z.M. Psillakis, *Distributions of statistics describing concentration of runs in non homogeneous Markov-dependent trials*, Comm. Statist. Theory Methods **47** (2018), 2238-2250.
3. N. Balakrishnan, M.V. Koutras, *Runs and scans with applications*, J. Wiley and Sons, New York, 2003.
4. Charalambos A. Charalambides, *Enumerative combinatorics*, Chapman and Hall/CRC, New York, 2002.
5. S. Eryilmaz, *On success runs in a sequence of dependent trials with a change point*, Statist. Probab. Lett. **132** (2018), 91-98.

6. J.C. Fu and W.Y.W. Wendy, *Distribution theory of runs and patterns and its applications: a finite Markov chain imbedding approach*, World Scientific, 2003.
7. Y. Kong, *Joint distribution of rises, falls, and number of runs in random sequences*, Comm. Statist. Theory Methods **48** (2019), 493-499.
8. F.S. Makri, Z.M. Psillakis and A.N. Arapis, *On the concentration of runs of ones of length exceeding a threshold in a Markov chain*, J. Appl. Stat. **46** (2019), 85-100.
9. J. Oh, *Run and Frequency quotas for q-binary trials*, arXiv preprint arXiv:2206.13053. https://doi.org/10.48550/arXiv.2206.13053
10. A.N. Philippou, C. Georghiou and G.N. Philippou, *A generalized geometric distribution and some of its properties*, Statist. Probab. Lett. **1** (1983), 171-175.

Jungtaek Oh received Ph.D. at Kyungpook National University. His research interests are Mathematical Statistics, Distribution theory of Runs, Scans and Patterns with applications, Graphical models, Causal inference, Statistical learning, Bayesian statistics and Uncertainty quantification.

Department of Biomedical Science, School of Medicine, Kyungpook National University.
Clinical Omics Center, Kyungpook National University, Daegue, 41566, Republic of Korea.
The Institute of Industrial Technology, Changwon National University, Republic of Korea.
e-mail: jungtaekoh0191@gmail.com

Kyeong Eun Lee received Ph.D. from Texas A& M University. She has been an associate professor at Kyungpook National University since 2005. Her research interests are Bayesian variable selection, survival data analysis, and propensity score analysis.

Department of Statistics, Kyungpook National University, Daegue, 41566, Republic of Korea.
e-mail: artlee@knu.ac.kr