Ann Inst Stat Math (2016) 68:1055–1072
DOI 10.1007/s10463-015-0525-x

On monotonicity of expected values of some run-related distributions

Sigeo Aki[1] · **Katuomi Hirano**[2]

Received: 9 May 2014 / Revised: 11 May 2015 / Published online: 8 July 2015

Abstract We prove that the expectation of the binomial distribution of order k with success probability p is monotonically increasing with respect to p for all n and k. The result is extended to the problems on exchangeable random sequences and expectations of distributions of mixtures of binomial distributions of order k are studied. If the mixing measure is stochastically increasing with respect to its parameter, the expectation of the mixture of binomial distributions of order k becomes nondecreasing. As examples of mixing measures two submodels of beta distributions are examined and the resulting expectation of the mixture distribution is monotonically strictly increasing. Further, we prove some properties on the expectation of the ℓ-overlapping 1-runs in a sequence of independent and identically distributed n trials.

Keywords Binomial distribution of order k · Method of moments · Exchangeability · Negative hypergeometric distribution of order k · Beta distribution

1 Introduction

The study on exact distributions of runs and scans has been much developed recently. Above all, distributions of number of runs in random sequences of finite length are

This research was partially supported by the Kansai University Grant-in-Aid for progress of research in graduate course.

✉ Sigeo Aki
aki@kansai-u.ac.jp

1 Department of Mathematics, Kansai University, 3-3-35 Yamate-cho, Suita, Osaka 564-8680, Japan

2 Department of Mathematics, Josai University, 1-1 Keyakidai, Sakado, Saitama 350-0295, Japan

studied by many authors (see, for example, Balakrishnan and Koutras 2002; Eryilmaz and Demir 2007; Eryilmaz and Yalçin 2011; Balakrishnan et al. 2014; Yalcin and Eryilmaz 2014). Let $X_1, X_2, \ldots, X_n$ be $\{0, 1\}$-valued independent identically distributed random variables with $P(X_i = 1) = p = 1 - q$. The distribution of the number of non-overlapping 1-runs of length k in $X_1, X_2, \ldots, X_n$ is called the binomial distribution of order k and denoted by $B_k(n, p)$.

The present authors proved that the expectation $m_n(p)$ of $B_k(n, p)$ is expressed as

$$m_n(p) = \sum_{j=1}^{[\frac{n}{k}]} \{(n - jk + 1)p^{jk} - (n - jk)p^{jk+1}\}, \tag{1}$$

in Aki and Hirano (1988). In order to show that the moment estimator of p is uniquely determined, they tried to prove that the expectation is monotonically increasing with respect to p in Aki and Hirano (1989). In spite of the simple form, it was not easy to show the monotonicity with respect to p for all n and k. However, as the expectation is represented by a polynomial function with respect to p for given n and k, they checked the monotonicity for $k = 2, 3, \ldots, 10$ and $n = k, k + 1, \ldots, 100$ using a computer algebra system with an algorithm based on Sturm's theorem. Though it is not so difficult to check the monotonicity with respect to p for given n with $n > 100$, it is still unknown whether the expectation is monotonically increasing with respect to p for all n and k with $k \leq n$.

In this paper, we show that the mean of the binomial distribution of order k is monotonically increasing with respect to p for all n and k. The idea of the proof is a kind of the special construction of random variables. As the monotonicity of the expectation with respect to p is a distributional property, it may be shown from the formula (1) directly. In this paper, however, we construct specially on a common probability space a family of random variables $\{Y^{(p)}(\omega) | 0 < p < 1\}$, each of which follows the $B_k(n, p)$. When $0 < p_1 < p_2 < 1$, the inequality $Y^{(p_1)}(\omega) \leq Y^{(p_2)}(\omega)$ is shown for every ω. Since taking expectation (integral) preserves the order, we obtain the monotonicity of expectation. The result will be extended to exchangeable random sequences. Further, we shall prove some properties on the mean of the ℓ-overlapping number of 1-runs. In the last section, we shall provide some properties of the binomial distributions of order k. We shall show that some random variables observed in various random sequences, which are not necessarily independent sequences, follow a binomial distribution of a specified order. Similar properties were discussed in Aki and Hirano (2000), where the binomial distributions of order k was slightly generalized. However, the results of the present paper will be proved without generalizing the binomial distributions of order k.

2 Expectations of the number of runs

First, we prove that the mean of the binomial distribution of order k is monotonically increasing with respect to p for all n and k.

Theorem 1 *Let n and k be integers with $1 \leq k \leq n$. Then, the mean of the binomial distribution of order k is monotonically strictly increasing with respect to p.*

Proof The mean of the binomial distribution of order k is a nonzero polynomial of p having degree at most n. Let $U_1, \ldots, U_n$ be independent random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ having uniform distribution on the unit interval [0, 1]. For $0 < p < 1$, we set $X_i^{(p)}(\omega) = I(U_i(\omega) < p)$, where $I(A)$ is the indicator of the statement A. Let $Y^{(p)}(\omega)$ be the number of non-overlapping 1-runs in the independent trials $\{X_i^{(p)}(\omega)\}_{i=1}^n$. Then, $Y^{(p)}(\omega)$ follows the binomial distribution of order k. Suppose that $0 < p_1 < p_2 < 1$. Then, by definition, for every ω, "$X_i^{(p_1)}(\omega) = 1$" implies "$X_i^{(p_2)}(\omega) = 1$". Therefore, we obtain that $Y^{(p_1)}(\omega) \leq Y^{(p_2)}(\omega)$ for every ω. Thus, the monotonicity of the integral (or taking expectation in other words) implies $m_n(p_1) \leq m_n(p_2)$. Noting that $m_n(p)$ is a polynomial function of p, it is differentiable and the derivative $m_n'(p)$ is also a polynomial function of p. Therefore, the set of zeros of $m_n'(p)$ is finite from the fundamental theorem of algebra. Consequently, $m_n(p)$ is monotonically strictly increasing with respect to p, i.e., "$0 < p_1 < p_2 < 1$" implies "$m_n(p_1) < m_n(p_2)$". □

It is not difficult to extend the above result to exchangeable sequences. Let π be a probability measure on the unit interval [0, 1] and let $\{X_i\}$ be the infinite exchangeable sequence with the de Finetti measure (or the mixing measure) π. We denote by $MB_k(n, \pi)$ the distribution of the non-overlapping number of 1-runs in $X_1, X_2, \ldots, X_n$. $MB_k(n, \pi)$ is also regarded as a mixture of binomial distributions of order k (see Johnson et al. 2005). On the mean of the distribution, we obtain the next proposition.

Proposition 1 *The mean of the distribution $MB_k(n, \pi)$ can be written as*

$$\sum_{j=1}^{[\frac{n}{k}]} \{(n - jk + 1)M_{jk} - (n - jk)M_{jk+1}\},$$

where M_m is the m-th moment of the probability measure π.

Proof From de Finetti's theorem (see, for example, Durrett 2010) and the formula (1), the mean of the distribution $MB_k(n, \pi)$ is represented as

$$\sum_{j=1}^{[\frac{n}{k}]} \{(n - jk + 1) \int_0^1 p^{jk} \pi(dp) - (n - jk) \int_0^1 p^{jk+1} \pi(dp)\}.$$

This completes the proof. □

Next, when the above mixing measure has a parameter which defines a stochastic ordering (see Lehmann and Romano 2005), we study the monotonicity of the mean of the binomial distribution of order k with respect to the new parameter.

Theorem 2 *Let $\{\pi_\theta\}$ be a family of distributions with parameter θ on the unit interval $[0, 1]$. $F_\theta(x)$ denotes the cumulative distribution function of π_θ. Suppose that the function $F_\theta(x)$ is continuous and strictly increasing with respect to x. Further, we assume that the family $\{F_\theta\}$ is stochastically increasing, i.e.,*

$$\theta_1 < \theta_2 \text{ implies } F_{\theta_1}(x) \geq F_{\theta_2}(x) \text{ for every } x \in [0, 1].$$

Then, the mean of the distribution $MB_k(n, \pi_\theta)$ is nondecreasing with respect to θ.

Proof From the assumption of the theorem, for every θ, $F_\theta(x)$ has its inverse function $F_\theta^{-1}(y)$ which is continuous and strictly increasing. Suppose that $\theta_1 < \theta_2$. Then, for every $y \in [0, 1]$, x_1 and x_2 satisfying $y = F_{\theta_1}(x_1) = F_{\theta_2}(x_2)$ are uniquely determined. Since $F_{\theta_1}(x_1) \geq F_{\theta_2}(x_1)$ holds, we have $x_1 \leq x_2$, i.e., $F_{\theta_1}^{-1}(y) \leq F_{\theta_2}^{-1}(y)$. Suppose that independent random variables $V, U_1, \ldots, U_n$ defined on the probability space Ω follow the uniform distribution on the unit interval $[0, 1]$. For every $\omega \in \Omega$, we set $P_\theta(\omega) = F_\theta^{-1}(V(\omega))$ and define

$$X_i^{(\theta)}(\omega) = \begin{cases} 1 & \text{if } U_i(\omega) \leq P_\theta(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y_n^\theta(\omega)$ be the number of non-overlapping 1-runs in $\{X_i^{(\theta)}(\omega)\}_{i=1}^n$. Then $Y_n^\theta(\omega)$ follows the distribution $MB_k(n, \pi_\theta)$.

Suppose that $\theta_1 < \theta_2$. For every $\omega \in \Omega$, we see that

$$\begin{aligned} X_i^{(\theta_1)}(\omega) = 1 &\iff U_i(\omega) \leq F_{\theta_1}^{-1}(V(\omega)) \\ &\Longrightarrow U_i(\omega) \leq F_{\theta_2}^{-1}(V(\omega)) \\ &\iff X_i^{(\theta_2)}(\omega) = 1. \end{aligned}$$

Therefore, for every ω, we have $Y_n^{(\theta_1)}(\omega) \leq Y_n^{(\theta_2)}(\omega)$. Consequently, from the monotonicity of integral, we have $E[Y_n^{(\theta_1)}] \leq E[Y_n^{(\theta_2)}]$. □

Here, we show two examples as applications of Theorem 2. These examples are given by one-parameter sub-families of beta distributions $Beta(\alpha, \beta)$ as the mixing measures. These are distributions of numbers of non-overlapping runs based on the Pólya–Eggenberger sampling schemes and are called the beta-binomial distribution of order k or the negative hypergeometric distribution of order k (see, Panaretos and Xekalaki 1986; Johnson et al. 2005).

Example 1 Let c be a constant with $c \geq 1$. For $\theta \geq 1$, we define

$$F_\theta(x) = \int_0^x \frac{1}{B(\theta, c)} t^{\theta-1}(1-t)^{c-1} dt \quad (0 \leq x \leq 1),$$

where $B(\alpha, \beta)$ is the beta function. Then, the family of distribution functions $\{F_\theta(x); \theta \geq 1\}$ satisfies the assumption of Theorem 2. Let θ_1 and θ_2 be parameters satisfying $1 \leq \theta_1 < \theta_2$. Let Y_1, Y_2 and Y_3 be independent random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ following the gamma distributions $Gamma(\theta_1)$, $Gamma(\theta_2 - \theta_1)$ and $Gamma(c)$, respectively, where the gamma distribution $Gamma(\alpha)$ is a distribution on the interval $(0, \infty)$ having the probability density function

$$\frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}, \quad (x > 0),$$

where $\Gamma(\alpha)$ is the gamma function. Noting that $\frac{Y_1(\omega)}{Y_1(\omega)+Y_3(\omega)}$ and $\frac{Y_1(\omega)+Y_2(\omega)}{Y_1(\omega)+Y_2(\omega)+Y_3(\omega)}$ follow the $Beta(\theta_1, c)$ and $Beta(\theta_2, c)$, respectively (see, for example, Wilks 1962). Then, for $x \in (0, 1)$, it holds that

$$F_{\theta_1}(x) = E\left[I\left(\frac{Y_1(\omega)}{Y_1(\omega) + Y_3(\omega)} \leq x \right) \right],$$

and

$$F_{\theta_2}(x) = E\left[I\left(\frac{Y_1(\omega) + Y_2(\omega)}{Y_1(\omega) + Y_2(\omega) + Y_3(\omega)} \leq x \right) \right].$$

Since for almost all $\omega \in \Omega$, $Y_1(\omega)$, $Y_2(\omega)$ and $Y_3(\omega)$ are positive,

$$\frac{Y_1(\omega)}{Y_1(\omega) + Y_3(\omega)} \leq \frac{Y_1(\omega) + Y_2(\omega)}{Y_1(\omega) + Y_2(\omega) + Y_3(\omega)}$$

holds and, therefore, we have

$$I\left(\frac{Y_1(\omega)}{Y_1(\omega) + Y_3(\omega)} \leq x \right) \geq I\left(\frac{Y_1(\omega) + Y_2(\omega)}{Y_1(\omega) + Y_2(\omega) + Y_3(\omega)} \leq x \right).$$

Thus, the monotonicity of integral implies $F_{\theta_1}(x) \geq F_{\theta_2}(x)$ and we have seen that the family of the distribution functions $\{F_\theta(x); \theta \geq 1\}$ satisfies the assumption of Theorem 2.

Therefore, from Theorem 2, we see that the mean $M(\theta)$ of the distribution $MB_k(n, \pi_\theta)$ is nondecreasing with respect to θ. Further, we have

$$M_{jk} = \int_0^1 p^{jk} \frac{1}{B(\theta, c)} p^{\theta-1}(1-p)^{c-1} dp = \frac{B(\theta + jk, c)}{B(\theta, c)}.$$

Noting that

$$\frac{B(\theta + jk, c)}{B(\theta, c)} = \frac{\Gamma(\theta + jk)}{\Gamma(\theta)} \frac{\Gamma(\theta + c)}{\Gamma(\theta + c + jk)} = \frac{(\theta)_{jk\uparrow}}{(\theta + c)_{jk\uparrow}},$$

we see that every moment M_{jk} is a rational function of θ. Here, $(x)_{n\uparrow}$ denotes the nth factorial power of x with increment 1. Since the expectation $M(\theta)$ of $MB_k(n, \pi_\theta)$

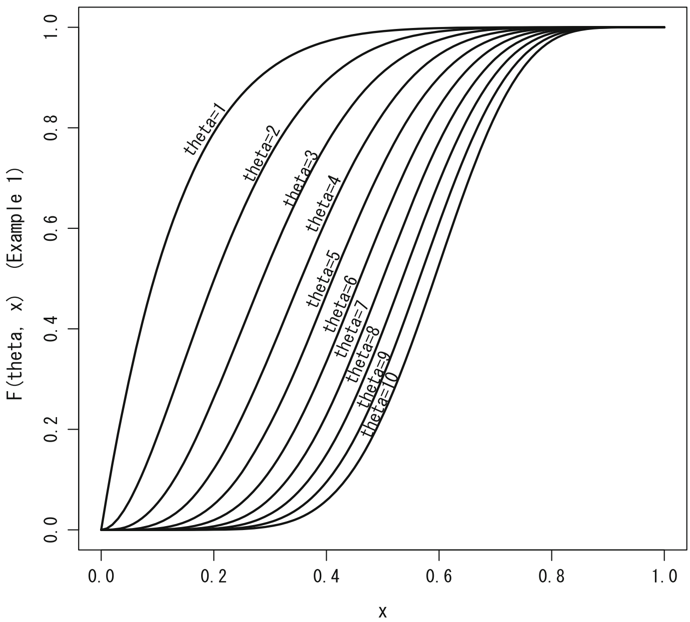

Fig. 1 Distribution functions $F_\theta(x)$ with $c = 7$ and $\theta = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$

is a linear combination of the moments of π_θ from Proposition 1, $M(\theta)$ is also a rational function of θ. Then, the expectation $M(\theta)$ is differentiable with respect to θ and $\frac{d}{d\theta}M(\theta) \geq 0$ holds by Theorem 2. Since $\frac{d}{d\theta}M(\theta)$ is also a rational function of θ, the set of zeros of $\frac{d}{d\theta}M(\theta)$ is finite by the fundamental theory of algebra. Therefore, the expectation $M(\theta)$ of $MB_k(n, \pi_\theta)$ is strictly increasing with respect to θ.

Figure 1 illustrates the distribution functions $F_\theta(x)$ with $c = 7$ and $\theta = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ which are stochastically increasing. Figure 2 plots the mean of the $MB_k(n, \pi_\theta)$, $M(\theta)$, where $k = 3$, $n = 14$, $c = 7$ and the distribution function of π_θ is given by the $F_\theta(x)$ of Example 1. We observe that the mean $M(\theta)$ is strictly increasing with respect to θ.

The next model which is a modification of Example 1 has a similar property.

Example 2 Let c be a constant with $c \geq 2$. For every θ with $1 \leq \theta \leq c-1$, we define

$$F_\theta(x) = \int_0^x \frac{1}{B(\theta, c-\theta)} t^{\theta-1}(1-t)^{c-\theta-1} dt \quad (0 \leq x \leq 1).$$

Then, the family of distribution functions $\{F_\theta(x); 1 \leq \theta \leq c-1\}$ satisfies the assumption of Theorem 2. We can show it in the same manner of Example 1. The difference from Example 1 is that the random variable Y_3 is assumed to follow the gamma distribution $Gamma(c - \theta_2)$ and only to show

$$\frac{Y_1(\omega)}{Y_1(\omega) + Y_2(\omega) + Y_3(\omega)} \leq \frac{Y_1(\omega) + Y_2(\omega)}{Y_1(\omega) + Y_2(\omega) + Y_3(\omega)}$$

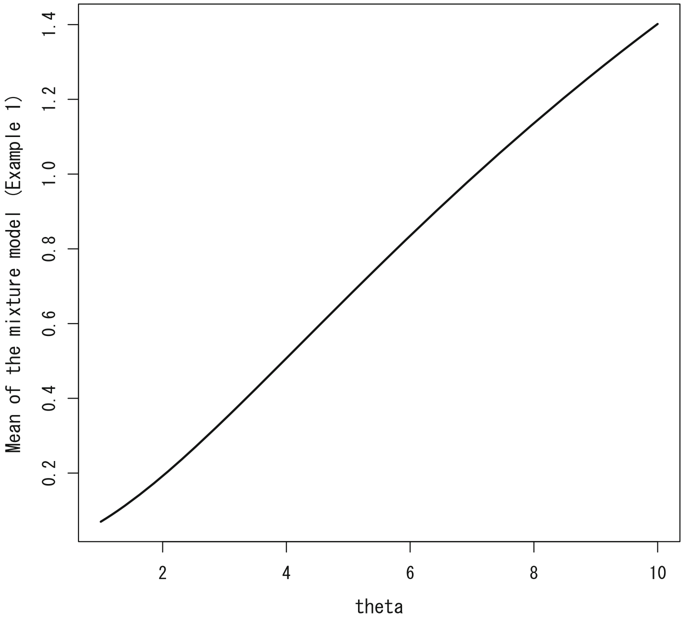

Fig. 2 Mean of the $MB_k(n, \pi_\theta)$, $M(\theta)$, is strictly increasing with respect to θ, where $k = 3$, $n = 14$, $c = 7$ and the distribution function of π_θ is given by the $F_\theta(x)$ of Example 1

for every $\omega \in \Omega$.

Theorem 2 implies that the expectation $M(\theta)$ of $MB_k(n, \pi_\theta)$ is nondecreasing with respect to θ. Further, by observing that

$$M_{jk} = \int_0^1 p^{jk} \frac{1}{B(\theta, c-\theta)} p^{\theta-1}(1-p)^{c-\theta-1} dp = \frac{B(\theta + jk, c-\theta)}{B(\theta, c-\theta)}$$

and

$$\frac{B(\theta + jk, c-\theta)}{B(\theta, c-\theta)} = \frac{\Gamma(\theta + jk)}{\Gamma(\theta)} \frac{\Gamma(c)}{\Gamma(+c + jk)} = \frac{(\theta)_{jk\uparrow}}{(c)_{jk\uparrow}},$$

we can see that the expectation $M(\theta)$ of $MB_k(n, \pi_\theta)$ is strictly increasing with respect to θ in the same manner of Example 1.

Figure 3 illustrates the distribution functions $F_\theta(x)$ with $c = 7$ and $\theta = 1, 2, 3, 4, 5, 6$ which are stochastically increasing. Figure 4 presents the mean of the $MB_k(n, \pi_\theta)$, $M(\theta)$, where $k = 3$, $n = 14$, $c = 7$ and the distribution function of π_θ is given by the $F_\theta(x)$ of Example 2. We observe that the mean $M(\theta)$ is strictly increasing with respect to θ.

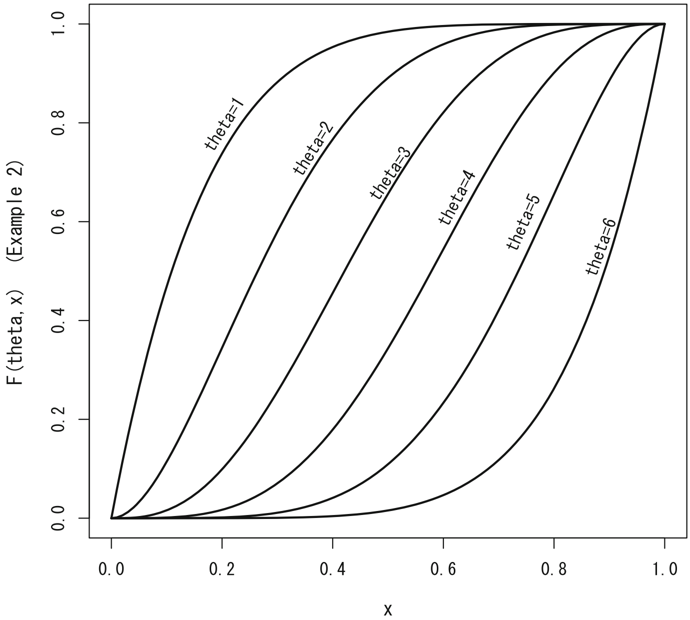

Fig. 3 Distribution functions $F_\theta(x)$ of Example 2 with $c = 7$ and $\theta = 1, 2, 3, 4, 5, 6$

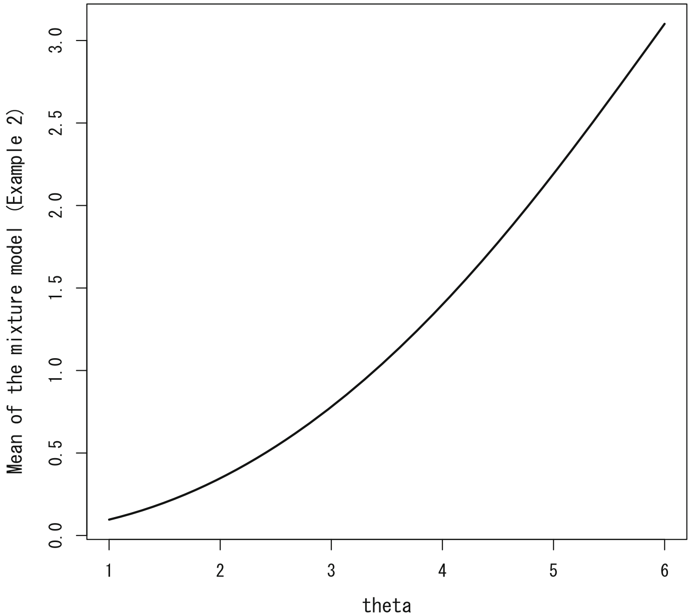

Fig. 4 Mean of the $MB_k(n, \pi_\theta)$, $M(\theta)$, is strictly increasing with respect to θ, where $k = 3$, $n = 14$, $c = 7$ and the distribution function of π_θ is given by the $F_\theta(x)$ of Example 2

3 Estimating the parameter of mixtures of binomial distributions of order *k*

In Examples 1 and 2 in the previous section, we proved that the expectations of the mixing models are strictly increasing with respect to the parameter θ. Therefore, we can estimate the parameter θ of the examples by the method of moments at least theoretically. Since the expectations are expressed as rational functions of θ, we show the feasibility of estimating the parameter by simulation experiments. Note that these examples are subclasses of the beta-binomial or negative hypergeometric distributions of order k, because the de Finetti (or mixing) measure of the model are one-parameter subclasses of the beta distributions.

Initially, we need to construct simulated data which follow the beta-binomial distribution of order k. A simple method is to repeat generating a random number of $B_k(n, p)$ with p from the beta distribution. However, it takes much time to generate random variables of $B_k(n, p)$ for varying values of p at all times.

Here, we use the inverse cumulative distribution function method for discrete distributions (see, for example, Gentle 2003) after calculating the probability of $MB_k(n, \pi_\theta)$.

The probability generating function of the negative hypergeometric distribution of order k, which is equal to the mixture of binomial distributions of order k with the mixing measure $Beta(a, b)$, is expressed as

$$\sum_{m=0}^{k-1} \sum_{x_1+2x_2+\cdots+kx_k=n-m} \binom{x_1+\cdots+x_k}{x_1,\ldots,x_k} \frac{(b)_{(\sum x_i)\uparrow}(a)_{(n-\sum x_i)\uparrow}}{(a+b)_{n\uparrow}} \times F\left(-x_k, -a+n-\sum x_i; b+1-\sum x_i; t\right),$$

where $F(\alpha, \beta; \gamma; z)$ is the hypergeometric function (see Aki and Hirano 1988). However, it is not so easy to evaluate it from the above form.

For numerical evaluation of the function, the following recurrence relation is convenient:

$$\phi(a,b,k,n) = 1 \text{ if } n < k$$
$$\phi(a,b,k,n) = \frac{(a)_{k\uparrow}}{(a+b)_{k\uparrow}} t + \left(1 - \frac{(a)_{k\uparrow}}{(a+b)_{k\uparrow}}\right) \text{ if } n = k$$
$$\phi(a,b,k,n) = \sum_{m=1}^{k} \frac{(a)_{(m-1)\uparrow} b}{(a+b)_{m\uparrow}} \phi(a+m-1, b+1, k, n-m) + \frac{(a)_{k\uparrow}}{(a+b)_{k\uparrow}} t\phi(a+k, b, k, n-k) \text{ if } n > k,$$

where $\phi(a, b, k, n)$ is the probability generating function of the negative hypergeometric distribution of order k.

Alternatively, it is also convenient for numerical evaluation of the probability generating function to integrate the probability generating function of the binomial distribution of order k after multiplying the density of the beta distribution $Beta(a, b)$ with respect to p (see Panaretos and Xekalaki 1986).

We illustrate the feasibility of estimation of the parameter of $MB_k(n, \pi_\theta)$ by the following numerical examples:

Example 3 (Continuation of Example 1) We consider the mixture model of Example 1 of Sect. 2 with $k = 3$, $n = 14$ and $c = 7$. By setting $\theta = 8.0$, we have constructed a simulated sample which follow the mixture model $MB_k(n, \pi_\theta)$. By using the method stated above, we see that the probability of $MB_3(14, \pi_8)$ as follows. Suppose that X follows the distribution $MB_3(14, \pi_8)$. Then, we have that

$$\begin{aligned}
P(X = 0) &= 0.31935657558218\\
P(X = 1) &= 0.35097411046799\\
P(X = 2) &= 0.2222973033607\\
P(X = 3) &= 0.089014522666427\\
P(X = 4) &= 0.018357487922705.
\end{aligned}$$

Using this discrete distribution, we constructed the simulated sample by the method of inverse cumulative distribution function with sample size = 50. Based on the sample, we calculated the moment estimate of θ (=8). Repeating the process 100,000 times, the sample mean and sample variance of the 100,000 estimates are obtained as 8.035496 and 1.066979, respectively. Figure 5 shows the estimated density function of the moment estimator with sample size 50 based on the 100,000 estimates.

Example 4 (Continuation of Example 2) We consider the mixture model of Example 2 of Sect. 2 with $k = 3$, $n = 14$ and $c = 7$. By setting $\theta = 4.5$, we have constructed a simulated sample which follows the mixture model $MB_k(n, \pi_\theta)$. Suppose that X follows the distribution $MB_3(14, \pi_{4.5})$. Then, we have that

$$\begin{aligned}
P(X = 0) &= 0.19290482997894\\
P(X = 1) &= 0.25314967600363\\
P(X = 2) &= 0.248562717012\\
P(X = 3) &= 0.19545306318573\\
P(X = 4) &= 0.10992971381971.
\end{aligned}$$

Similarly as Example 3, we constructed the simulated sample by the method of inverse cumulative distribution function with sample size = 50. Based on the sample, we calculated the moment estimate of θ (=4.5). Repeating the process 100,000 times, the sample mean and sample variance of the 100,000 estimates are obtained as 4.495461 and 0.05145771, respectively. Figue 6 shows the estimated density function of the moment estimator with sample size 50 based on the 100,000 estimates.

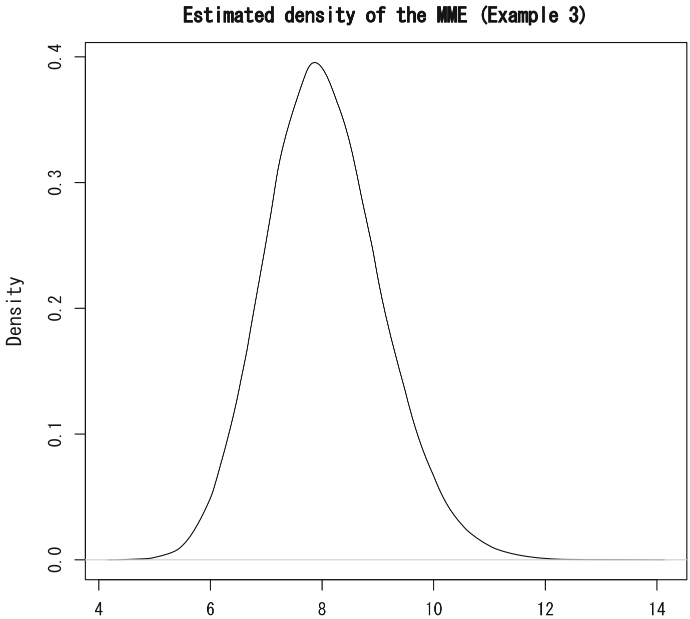

Fig. 5 Estimated density of the moment estimator with sample size 50 of Example 3 based on the 100,000 estimates

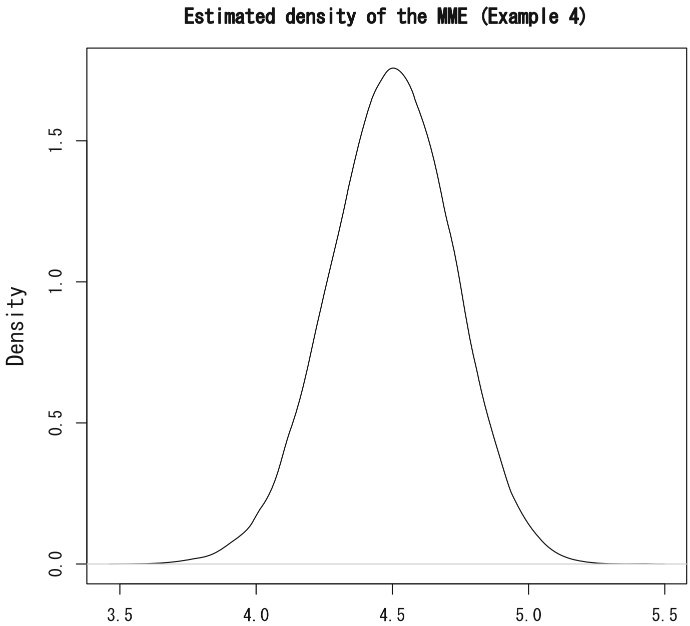

Fig. 6 Estimated density of the moment estimator with sample size 50 of Example 4 based on the 100,000 estimates

4 Number of 1-runs under the ℓ-overlapping enumeration scheme

There are several enumeration schemes of counting number of 1-runs in random {0, 1}-sequences. Let ℓ be an integer satisfying $0 \le \ell < k$. The ℓ-overlapping number of 1-runs of length k is the number of 1-runs each of which may have overlapping part of length at most ℓ with the previous 1-run of length k that has been enumerated (see Definition 1.1 of Aki and Hirano 2000). Then, 0-overlapping and $(k-1)$-overlapping numbers of 1-runs of length k mean the usual non-overlapping and overlapping numbers of 1-runs of length k, respectively. The $(k-1)$-overlapping number of 1-runs of length k was introduced and studied by Ling (1988). To make the meaning of ℓ-overlapping counting clear, we assume that $n = 15$ and outcomes of the sequence of length n are

111101111111010.

Let us enumerate number of 1-runs of length 3. Then, the non-overlapping (0-overlapping) number of 1-runs of length 3 in the sequence (111)10(111)(111)1010 is 3. The overlapping (2-overlapping) number of 1-runs of length 3 in the sequence (1[11)1]0(1[1{1)(1][1}1)1]010 is 7. The 1-overlapping number of 1-runs of length 3 in the sequence (111)10(11[1)1(1]11)010 is 4.

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed {0, 1}-valued random variables with $P(X_i = 1) = p = 1 - q$ with $0 < p < 1$. What we have discussed in the previous section can be studied under the general counting schemes. Let $X_{(\ell)}$ be the number of ℓ-overlapping 1-runs in $X_1, X_2, \ldots, X_n$. The next result on the expectation of $X_{(\ell)}$ was proved in Proposition 2.1 of Makri and Philippou (2005). In Makri et al. (2013) an even simpler formula for the expectation of $X_{(\ell)}$ is given.

Theorem 3 *Let $n_\ell = \left[\frac{n-k}{k-\ell}\right]$. Then, the expectation of $X_{(\ell)}$ can be written as*

$$E[X_{(\ell)}] = \sum_{j=0}^{n_\ell} \{(n-(k+j(k-\ell))+1)p^{k+j(k-\ell)} - (n-(k+j(k-\ell)))p^{k+j(k-\ell)+1}\}.$$

In particular, setting $\ell = 0$ in Theorem 3, $X_{(0)}$ follows the binomial distribution of order k, $B_k(n, p)$. Hence, we have the next corollary immediately.

Corollary 1 *The mean of the binomial distribution of order k, $B_k(n, p)$ can be written as*

$$E[X_{(0)}] = \sum_{j=1}^{[\frac{n}{k}]} \{(n - jk + 1)p^{jk} - (n - jk)p^{jk+1}\}.$$

As the expectation of $X_{(\ell)}$ is expressed as a polynomial of p from Theorem 3, the next theorem can be proved similarly as Theorem 1.

Theorem 4 *Let n, k and ℓ be integers with $1 \le k \le n$ and $0 \le \ell < k$. Then, the mean of the number of ℓ-overlapping 1-runs in $X_1, X_2, \ldots, X_n$ is monotonically strictly increasing with respect to p.*

Let $B_k^{(\ell)}(n, p)$ denote the distribution of $X_{(\ell)}$, i.e., the distribution of the number of ℓ-overlapping 1-runs in the i.i.d. sequence $X_1, X_2, \ldots, X_n$. As in the previous section, let π_θ be a probability measure on the unit interval [0, 1] with parameter θ. Let $\xi_1, \xi_2, \ldots$ be an infinite exchangeable sequence with the mixing measure π_θ. To be precise, let Π be a [0, 1]-valued random variable with distribution π_θ and let $\xi_1, \xi_2, \ldots$ be conditionally independent identically distributed {0, 1}-valued random variables with $P(\xi_i = 1|\Pi = p) = p$ given that $\Pi = p$. Let $X_{(\ell)}^{(\theta)}$ be the number of ℓ-overlapping 1-runs in $\xi_1, \ldots, \xi_n$ and let $MB_k^{(\ell)}(n, \pi_\theta)$ denote the distribution of $X_{(\ell)}^{(\theta)}$.

Then, we obtain the next theorem similarly as Theorem 2.

Theorem 5 *Let $\{\pi_\theta\}$ be a family of distributions with parameter θ on the unit interval [0, 1]. $F_\theta(x)$ denotes the cumulative distribution function of π_θ. Suppose that the function $F_\theta(x)$ is continuous and strictly increasing with respect to x. Further, we assume that the family $\{F_\theta\}$ is stochastically increasing, i.e.,*

$$\theta_1 < \theta_2 \textit{ implies } F_{\theta_1}(x) \ge F_{\theta_2}(x) \textit{ for every } x \in [0, 1].$$

Then, the mean of the distribution $MB_k^{(\ell)}(n, \pi_\theta)$ is nondecreasing with respect to θ.

5 Some examples related to the binomial distributions of order *k*

First, let $X_1, X_2, \ldots$ be {0, 1}-valued independent identically distributed random variables with $P(X_i = 1) = p = 1 - q$ $(0 < p < 1)$.

The next proposition was proved by Aki and Hirano (1988).

Let X be the number of non-overlapping 1-runs in $X_1, X_2, \ldots, X_n$. The distribution of X is the binomial distribution of order k. We denote by $\phi(k, n; t)$ the probability generating function of X.

Proposition 2 (Aki and Hirano 1988) *$\phi(k, n; t)$ satisfies the following recurrence relations:*

$$\begin{cases} \phi(k, n; t) = \sum_{m=1}^{k} p^{m-1} q \phi(k, n-m; t) + p^k t \phi(k, n-k; t) & \textit{if } n > k, \\ \phi(k, n; t) = (1 - p^k) + p^k t & \textit{if } n = k, \\ \phi(k, n; t) = 1 & \textit{if } n < k. \end{cases} \quad (2)$$

Then, we obtain the next proposition.

Proposition 3 *Let $k > 1$. Let N be the number of 1-overlapping 1-runs of length k until the n-th occurrence of "1" in $X_1, X_2, \ldots$. Then, N follows the binomial distribution of order $(k-1)$, $B_{k-1}(n-1, p)$.*

Proof Let $\phi_n(t)$ be the probability generating function of N. If $n < k$, $\phi_n(t) = 1$, since a 1-run of length k does not occur. If $n = k$, a 1-run of length k occurs if and only if a 1-run of length k occurs just after a 0-run of arbitrary length in $X_1, X_2, \ldots$. Since the probability of the event is

$$\left(\sum_{\ell=0}^{\infty} q^{\ell}\right) \times p^k = \frac{1}{1-q} \times p^k = p^{k-1},$$

it follows that $\phi_k(t) = p^{k-1}t + (1 - p^{k-1})$. Let us consider the case of $n > k$. If we observe the first trial, "1" or "0" necessarily occurs and hence we see that

$$\phi_n(t) = p\phi_{n-1}^{(1)}(t) + q\phi_n(t),$$

where $\phi_m^{(\ell)}(t)$ is the conditional probability generating function of number of 1-overlapping 1-runs of length k until the m-th "1" starting from the condition that the current length of 1-run which we are observing is just ℓ. Here, if $\ell = 0$, we prefer to write $\phi_m(t)$ rather than $\phi_m^{(0)}(t)$. Thus, we see that $\phi_n(t) = \phi_{n-1}^{(1)}(t)$ and we have the following system of equations:

$$\begin{cases} \phi_{n-1}^{(1)}(t) & = p\phi_{n-2}^{(2)}(t) + q\phi_{n-2}^{(1)}(t) \\ \phi_{n-2}^{(2)}(t) & = p\phi_{n-3}^{(3)}(t) + q\phi_{n-3}^{(1)}(t) \\ \cdots & \cdots \\ \phi_{n-(k-1)}^{(k-1)}(t) & = pt\phi_{(n-1)-(k-1)}^{(1)}(t) + q\phi_{(n-1)-(k-1)}^{(1)}(t). \end{cases}$$

From the equations of conditional probability generating functions, we easily obtain the following recurrence relation:

$$\phi_{n-1}^{(1)}(t) = q\phi_{n-2}^{(1)}(t) + pq\phi_{n-3}^{(1)}(t) + \cdots + p^{k-2}q\phi_{(n-1)-(k-1)}^{(1)}(t) + p^{k-1}t\phi_{(n-1)-(k-1)}^{(1)}(t).$$

Comparing this with (2) of Proposition 2, we see that the distribution of $\phi_{n-1}^{(1)}(t)$ is the binomial distribution of order $(k-1)$, $B_{k-1}(n-1, p)$. □

Though we have given the above proof using the traditional method of conditional probability generating functions, we give an alternative proof of Proposition 3 by constructing a new sequence of Bernoulli trials. The idea leads us more intuitive understanding.

Proof (Alternative proof of Proposition 3) In $X_1, X_2, \ldots$, we transform just after the first "1" to the following $\{S, F\}$-sequence such as

$$00\cdots01SFSF\cdots,$$

where S is "1" and F is a finite sequence of 0-run of positive length followed by "1". In the sequence $X_1, X_2, \ldots$, the probabilities of occurrence of S and F are, respectively,

$$P(S) = p$$
$$P(F) = \sum_{\ell=1}^{\infty} q^{\ell} p = p\frac{q}{1-q} = q.$$

When $\underbrace{S \cdots S}_{k-1}$ occurs for the first time, just before this pattern is the first "1" or F. Hence, at the corresponding trial of $X_1, X_2, \ldots$, 1-run of length k occurs. When the next $\underbrace{S \cdots S}_{k-1}$ occurs under the non-overlapping counting, the outcome just before the pattern is F or $\underbrace{S \cdots S}_{k-1}$. Therefore, at the corresponding trial of $X_1, X_2, \ldots$, 1-run of length k occurs under the 1-overlapping counting. Noting that just one "1" is included in S and F and the first "1" is not included in S. Consequently, the distribution of N is equal to that of the number of S-run of length $(k-1)$ in the independent $\{S, F\}$-sequence of length $(n-1)$, that is, the binomial distribution of order $(k-1)$, $B_{k-1}(n-1, p)$. □

Using the idea of the above alternative proof, Proposition 3 can be generalized further.

Proposition 4 *Let ℓ be an integer satisfying $1 \leq \ell < k$. Let N be the number of ℓ-overlapping 1-runs of length k until the n-th overlapping occurrence of 1-run of length ℓ in $X_1, X_2, \ldots$. Then, N follows the binomial distribution of order $(k-\ell)$, $B_{k-\ell}(n-1, p)$.*

Proof In $X_1, X_2, \ldots$, we transform just after the first 1-run of length ℓ to the following $\{S, F\}$-sequence:

$$X_1, X_2, \ldots, X_{\tau_\ell}, Y_1, Y_2, \ldots.$$

Here, τ_ℓ is the waiting time for the first 1-run of length ℓ, and Y_i is the following $\{S, F\}$-valued independent random variable, where $S =$ "1" and $F =$ "$00101 \cdots 0\underbrace{1 \cdots 1}_{\ell}$", i.e., F is a finite subsequence which starts from "0" until the next 1-run of length ℓ. As we assume that $0 < p < 1$, the next 1-run of length ℓ necessarily occurs, and hence we have $P(F) = q$ from $P(S) = p$. For rigorous proof of the statement, we need the second Borel–Cantelli lemma (see Remark 1 for the details). When $\underbrace{S \cdots S}_{k-\ell}$ occurs for the first time, the value just before it is 1-run of length ℓ or F. Therefore, at the corresponding trial of the sequence of $\{X_i\}$, 1-run of length k necessarily occurs. When the next $\underbrace{S \cdots S}_{k-\ell}$ occurs under the non-overlapping counting, just before it F or $\underbrace{S \cdots S}_{k-\ell}$ must have occurred. Hence, at the corresponding trial of the sequence of $\{X_i\}$, using 1-run of length ℓ which has occurred just before, a 1-run of length k occurs under ℓ-overlapping counting. Though each S and F correspond to overlapping 1-run of length ℓ, respectively, only the first 1-run of length ℓ does not correspond to any of S or F. Therefore, the distribution of N is equal to the number of non-overlapping 1-runs of length $(k-\ell)$ in independent $\{S, F\}$-sequence of length $(n-1)$. Consequently, N follows the binomial distribution of order $(k-\ell)$, $B_{k-\ell}(n-1, p)$. □

Remark 1 As $0 < p < 1$ is assumed, $\{0, 1\}$-pattern of finite length occurs infinitely often. For example, let the pattern be a 1-run of length m. For $k = 1, 2, \ldots$, we define the sequence of events $A_k = \{X_{m(k-1)+1} = 1, X_{m(k-1)+2} = 1, \ldots, X_{m(k-1)+m} = 1\}$ $(k \geq 1)$. Then, they are mutually independent and $\sum_{k=1}^{\infty} P(A_k) = \infty$ holds, since $P(A_k) = p^m$. Therefore, we see that $P(A_k \text{ i.o. }) = 1$ from the second Borel-Cantelli lemma.

Random variates which follow a binomial distribution of some order are observed not only on independent sequences but also on dependent sequences. We can obtain the corresponding results to Proposition 4 on time homogeneous $\{0, 1\}$-valued mth order Markov chain. For a rigorous statement, we need the second conditional Borel-Cantelli lemma.

Lemma 1 (Conditional second Borel-Cantelli lemma) *Let $\mathcal{F}_n$ $Cn \geq 0$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\{A_n\}$, $n \geq 1$ a sequence of events with $A_n \in \mathcal{F}_n$. Then*

$$\{A_n \ \ i.o.\ \} = \left\{ \sum_{n=1}^{\infty} P(A_n | \mathcal{F}_{n-1}) = \infty \right\}.$$

For a proof of the lemma, see, for example, Durrett (2010) or Chandra (2012).

Remark 2 Let $X_{-m+1}, \ldots, X_0, X_1, X_2, \ldots$ be time homogeneous $\{0, 1\}$-valued mth order Markov chain and for every positive integer i we set

$$p_{i_1 i_2 \ldots i_m} = P(X_i = 1 | X_{i-m} = i_1, X_{i-m+1} = i_2, \ldots, X_{i-1} = i_m),$$
$$q_{i_1 i_2 \ldots i_m} = P(X_i = 0 | X_{i-m} = i_1, X_{i-m+1} = i_2, \ldots, X_{i-1} = i_m),$$

where $q_{i_1 i_2 \ldots i_m} = 1 - p_{i_1 i_2 \ldots i_m}$. Here, we assume that $0 < p_{i_1 i_2 \ldots i_m} < 1$ for every $(i_1, i_2, \ldots, i_m) \in \underbrace{\{0, 1\} \times \cdots \times \{0, 1\}}_{m}$. Hence, if we set $\alpha = \min_{i_1, i_2, \ldots, i_m} p_{i_1 i_2 \ldots i_m}$, $\alpha > 0$ holds.

Let $m \leq \nu$ and we set a sequence of events $A_n = \{X_n = 1, X_{n+1} = 1, \ldots, X_{n+\nu} = 1\}$ for $n = 1, 2, \ldots$. Let $\mathcal{F}_n$ be the σ-field generated by $\{X_{-m+1}, \ldots, X_n\}$ and let $B_k = A_{\nu k+1}$ and $\mathcal{G}_k = \mathcal{F}_{\nu(k+1)}$ for $k = 1, 2, \ldots$. Then, since, for each k, $P(B_k | \mathcal{G}_{k-1}) \geq \alpha^{\nu}$ holds, we see that $\sum_{k=1}^{\infty} P(B_k | \mathcal{G}_{k-1}) = \infty$. Therefore, from the conditional second Borel-Cantelli lemma $P(B_k \text{ i.o. }) = 1$ holds and hence we obtain $P(A_n \text{ i.o. }) = 1$.

Though the study like this was attempted intuitively in Aki et al. (1996), we have stated rigorously here that a pattern of finite length occurs infinitely often almost surely.

Proposition 5 *Let $X_{-m+1}, \ldots, X_0, X_1, X_2, \ldots$ be a time-homogeneous $\{0, 1\}$-valued m-th order Markov chain satisfying for every positive integer i,*

$$p_{i_1 i_2 \ldots i_m} = P(X_i = 1 | X_{i-m} = i_1, X_{i-m+1} = i_2, \ldots, X_{i-1} = i_m),$$
$$q_{i_1 i_2 \ldots i_m} = P(X_i = 0 | X_{i-m} = i_1, X_{i-m+1} = i_2, \ldots, X_{i-1} = i_m),$$

where $q_{i_1 i_2 \ldots i_m} = 1 - p_{i_1 i_2 \ldots i_m}$. Here, assume that $0 < p_{i_1 i_2 \ldots i_m} < 1$ for every $(i_1, i_2, \ldots, i_m) \in \underbrace{\{0,1\} \times \cdots \times \{0,1\}}_{m}$. Let $1 \le m \le \ell < k$ and let N be the ℓ-overlapping number of 1-runs of length k until the n-th overlapping occurrence of 1-run of length ℓ in $X_1, X_2, \ldots$. Then, N follows the binomial distribution of order $(k-\ell)$, $B_{k-\ell}(n-1, p_{11\cdots 1})$.

Proof In $X_1, X_2, \ldots$, we transform just after the first 1-run of length ℓ to the following independent $\{S, F\}$-sequence

$$X_1, X_2, \ldots, X_{\tau_\ell}, Y_1, Y_2, \ldots,$$

where τ_ℓ is the waiting time for the first 1-run of length ℓ, and Y_i is the following $\{S, F\}$-valued independent random variable. Here, $S =$ "1" and $F =$ "$00101 \cdots 0 \underbrace{1 \cdots 1}_{\ell}$", i.e., F is a finite subsequence which starts from "0" until the next 1-run of length ℓ. From the assumption of the proposition and Remark 2, we observe the next 1-run of length ℓ necessarily. Noting that $m \le \ell$ and $P(S) = p_{11\ldots 1}$, we see that $P(F) = q_{11\ldots 1}$. When we observe the first $\underbrace{S \cdots S}_{k-\ell}$, just before it the first 1-run of length ℓ or F must have occurred. Therefore, at the corresponding trial of the sequence $\{X_i\}$, 1-run of length k occurs under the ℓ-overlapping counting. When the next $\underbrace{S \cdots S}_{k-\ell}$ occurs under the non-overlapping counting, just before it F or $\underbrace{S \cdots S}_{k-\ell}$ must has occurred. Therefore, at the corresponding trial of the sequence $\{X_i\}$, using 1-run of length ℓ just before the trial, a 1-run of length k occurs under the ℓ-overlapping counting. Note that every S or F corresponds to a 1-run of length ℓ under overlapping counting. But, only the first 1-run of length ℓ does not correspond to any of S or F. Therefore, N follows the binomial distribution of order $(k-\ell)$ in independent $\{S, F\}$-sequence of length $(n-1)$, $B_{k-\ell}(n-1, p_{11\ldots 1})$. □

We have the corresponding result on infinite exchangeable sequences.

Proposition 6 *Let $X_1, X_2, \ldots$ be an exchangeable sequence of $\{0, 1\}$-valued random variables with mixing measure π. Let $0 \le \ell < k$ and let N be the ℓ-overlapping number of occurrences of 1-run of length k until the n-th overlapping occurrence of 1-run of length ℓ. Then, N follows the mixture of the binomial distribution of order $(k-\ell)$, $MB_{k-\ell}(n-1, \pi)$.*

Proof From de Finetti's theorem, there exists a random variable Π with $0 \leq \Pi \leq 1$ such that $X_1, X_2, \ldots$ are conditionally independently identically distributed with $P(X_i = 1|\Pi = p) = p$ given $\Pi = p$ (see Billingsley 1995). From Proposition 4, the conditional distribution of N given $\Pi = p$ is $B_{k-\ell}(n-1, p)$. Therefore, N follows the mixture of the binomial distribution of order $(k-\ell)$, $MB_{k-\ell}(n-1, \pi)$. □

References

Aki, S., Hirano, K. (1988). Some characteristics of the binomial distribution of order k and related distributions. In K. Matusita (Ed.), *Statistical theory and data analysis II* (pp. 211–222). North-Holland: Amsterdam.

Aki, S., Hirano, K. (1989). Estimation of parameters in the discrete distributions of order k. *Annals of the Institute of Statistical Mathematics*, *41*, 47–61.

Aki, S., Hirano, K. (2000). Numbers of success-runs of specified length until certain stopping time rules and generalized binomial distributions of order k. *Annals of the Institute of Statistical Mathematics*, *52*, 767–777.

Aki, S., Balakrishnan, N., Mohanty, S. G. (1996). Sooner and later waiting time problems for success and failure runs in higher order Markov dependent trials. *Annals of the Institute of Statistical Mathematics*, *48*, 773–787.

Balakrishnan, N., Koutras, M. V. (2002). *Runs and scans with applications*. New York: Wiley.

Balakrishnan, N., Koutras, M. V., Milienos, F. S. (2014). Some binary start-up demonstration tests and associated inferential methods. *Annals of the Institute of Statistical Mathematics*, *66*, 759–787.

Billingsley, P. (1995). *Probability and measure* (3rd ed.). New York: Wiley.

Chandra, T. K. (2012). *The Borel–Cantelli lemma*. New York: Springer.

Durrett, R. (2010). *Probability: theory and examples* (4th ed.). New York: Cambridge University Press.

Eryilmaz, S., Demir, S. (2007). Success runs in a sequence of exchangeable binary trials. *Journal of Statistical Planning and Inference*, *137*, 2954–2963.

Eryilmaz, S., Yalçin, F. (2011). Distribution of run statistics in partially exchangeable processes. *Metrika*, *73*, 293–304.

Gentle, J. E. (2003). *Random number generation and Monte Carlo methods* (2nd ed.). New York: Springer.

Johnson, N. L., Kemp, A. W., Kotz, S. (2005). *Univariate discrete distributions* (3rd ed.). New York: Wiley.

Lehmann, E. L., Romano, J. P. (2005). *Testing statistical hypotheses* (3rd ed.). New York: Springer.

Ling, K. D. (1988). On binomial distribution of order k. *Statistics and Probability Letters*, *6*, 247–250.

Makri, F. S., Philippou, A. N. (2005). On binomial and circular binomial distributions of order k for ℓ-overlapping success runs of length k. *Statistical Papers*, *46*, 411–432.

Makri, F. S., Psillakis, Z. M., Arapis, A. N. (2013). Counting runs of ones with overlapping parts in binary strings ordered linearly and circularly. *International Journal of Statistics and Probability*, *2*(3), 50–60.

Panaretos, J., Xekalaki, E. (1986). On some distributions arising certain generalized sampling schemes. *Communications in Statistics Theory and Methods*, *15*, 873–891.

Wilks, S. S. (1962). *Mathematical statistics*. New York: Wiley.

Yalcin, F., Eryilmaz, S. (2014). q-geometric and q-binomial distributions of order k. *Journal of Computational and Applied Mathematics*, *271*, 31–38.