Josai Mathematical Monographs
vol. 8 (2015), pp. 25-47

On calculations of exact distributions of waiting times of discrete patterns based on generating functions

Sigeo Aki*

Abstract. The rational generating functions of the probabilities and cumulative probabilities of the geometric distribution of order k are investigated. All the roots of each denominator are rigorously proved to be simple if $p \neq k/(k+1)$. It is also shown that the distribution of the waiting time for (k_1, k_2)-events has the similar property. The results are applied to numerical calculations of probability and cumulative probability of the distributions. Further, explicit expressions for the probability functions of the geometric distributions of order 2,3, and 4 are given by using partial fraction expansion of the generating functions. Moreover, as an example that the denominator of a rational generating function has multiple roots, the negative binomial distribution of order 2 with parameter (r, p) are studied. Explicit expressions of the probability function and the cumulative distribution function are provided. They are also useful for numerical calculations.

1. Introduction

The discrete distribution theory on runs and patterns in random sequences with various dependency, such as Markov, higher-order Markov, exchangeable, or partially exchangeable models (for example, see Balakrishnan and Koutras [**3**], Inoue, Aki, Hirano[**9**], Aki[**1**], Aki and Hirano[**2**]). In the present paper, we focus on calculations of exact probabilities of the distributions based on the probability and cumulative probability generating functions.

For example, the next recurrence formula is useful for the calculation of the geometric distribution of order k, which is the distribution of the waiting time W_k for the first 1-run of length k in independent sequence of $\{0,1\}$-valued random variables $X_1, X_2, \ldots$ with $P(X_i = 1) = p = 1 - q$.

$$G_k(p;x) = \begin{cases} 0 & \text{if } 0 \leq x < k, \\ p^k & \text{if } x = k, \\ p^k q & \text{if } k+1 \leq x \leq 2k, \\ G_k(p;x-1) - p^k q G_k(p;x-k-1) & \text{if } x \geq 2k+1, \end{cases}$$

where $G_k(p;x) = P(W_k = x)$. Usually, we can obtain the probabilities or cumulative probabilities of the distribution by using the above formula very fast. However,

* This research was partially supported by the Kansai University Grant-in-Aid for progress of research in graduate course, 2014.

when the probability p of the occurrence of 1 is small and/or k is large, it takes much time for the calculation of large x, such as the calculation of 90-percentile point of such a distribution. If we are interested in numerical calculations even in such cases, we can select another method, the partial fraction expansion for rational generating functions. Feller[**6**] examined asymptotic evaluation of the probability distribution of the waiting time for the first 1-run of length k in independent binary trials. As the partial fraction expansion of rational generating functions in the complex plane leads exact evaluation of the coefficients as well as their asymptotic evaluation, we can perform sufficiently exact numerical evaluation. By using the method, Shmuelli and Cohen[**12**] calculated probability functions of the geometric distributions of order k and the negative binomial distributions of order k numerically.

First, in the present paper, we shall calculate the probability functions of the geometric distribution of order k after examining Feller's result on the roots of the denominator of the probability generating function. Further, we evaluate the cumulative probabilities of the distribution for some large values by applying the partial fraction expansion. When we use the partial fraction expansion for a rational generating function, we need multiplicity of each root of the denominator. In Section 4 we give other examples of rational generating functions. In the examples, the multiplicity of each root of the denominators can be examined theoretically. The distributions are waiting times of so called (k_1, k_2)-events. We shall prove that all the roots of the denominator of the probability generating function are simple. The partial fraction expansion is very useful in numerical calculations. However, if all the roots of the denominator of a rational generating function are obtained explicitly, we may be able to give an explicit expression of the probability function of the distribution. The denominators of the probability generating functions of the geometric distributions of order 2,3 and 4 have degree less than 5, and hence all the roots can be written explicitly. We give explicit expressions of these distributions in Section 5. In Section 6, we examine the negative binomial distribution of order 2 with parameter (r,p), where r is an arbitrarily given positive integer, as an example that all the roots of the denominator of the probability generating function are given explicitly and the multiplicity of the roots are exactly r.

2. Partial fraction expansion

Let $\Phi(z)$ be the generating function of the sequence of numbers $\{a_n\}$, i.e., $\Phi(z) = \sum_k a_k z^k$. Feller[**6**] proved the following useful proposition.

PROPOSITION 1. *Suppose that the generating function is of the form* $\Phi(z) = \frac{U(z)}{V(z)}$, *where* $U(z)$ *and* $V(z)$ *are polynomials without common roots. Assume that the degree of* $U(z)$ *is lower than that of* $V(z)$, *say* m. *Further, assume that* $V(z) = 0$

has distinct roots $z_1, z_2, \ldots, z_m$. *Then,* a_n *can be written as*

$$a_n = \frac{\rho_1}{z_1^{n+1}} + \frac{\rho_2}{z_2^{n+1}} + \cdots + \frac{\rho_m}{z_m^{n+1}},$$

where $\rho_k = \frac{-U(z_k)}{V'(z_k)}$, *for* $k = 1, 2, \ldots, m$.

After proving the above proposition, Feller discussed approximation of a_n by using the only one root of $V(z)$, which has the minimum absolute value among the roots.

However, the partial fraction expansion provides useful exact calculation of a_n by using all the roots of $V(z)$ even in the case that $V(z)$ has multiple roots. Here, we introduce the general result of the partial fraction expansion of rational generating functions.

PROPOSITION 2. *Suppose that the generating function is of the form* $\Phi(z) = \frac{U(z)}{V(z)}$, *where* $U(z)$ *and* $V(z)$ *are polynomials without common roots. Assume that the degree of* $U(z)$ *is lower than that of* $V(z)$. *Assume that* $V(z)$ *can be written as*

$$V(z) = (z - z_1)^{k_1}(z - z_2)^{k_2} \cdots (z - z_N)^{k_N}.$$

Then, every z_i *is the pole of* $\Phi(z)$ *of order* k_i *and the singular part of* $\Phi(z)$ *at* $z = z_i$ *is*

$$f_i(z) = \frac{a_{-k_i}^{(i)}}{(z - z_i)^{k_i}} + \cdots + \frac{a_{-1}^{(i)}}{z - z_i}.$$

Consequently, the partial fraction expansion of $\Phi(z)$ *is written as*

$$\Phi(z) = \sum_{i=1}^{N} f_i(z).$$

For a proof of the proposition, see, for example, Conway[**4**], Flajolet and Sedgewick[**7**] (Theorem IV.9), or Pemantle and Wilson[**10**]. The coefficient $a_{-j}^{(i)}$ in the above expansion can be obtained by

$$a_{-j}^{(i)} = \frac{1}{(k_i - j)!} \lim_{z \to z_i} \frac{d^{k_i - j}}{dz^{k_i - j}} \frac{U(z)}{V(z)} (z - z_i)^{k_i}.$$

Here, we obtain the coefficient of z^n in the partial fraction expansion given in Proposition 2. By expanding every term of $f_i(z)$, we have

$$\frac{a_{-r}^{(i)}}{(z - z_i)^r} = (-1)^r \frac{a_{-r}^{(i)}}{z_i^r}(1 - \frac{z}{z_i})^{-r}.$$

Here, we can further expand it as

$$\left(1-\frac{z}{z_i}\right)^{-r}=\sum_{j=0}^{\infty}\binom{-r}{j}\left((-1)\frac{z}{z_i}\right)^j.$$

Noting that $(-1)^j\binom{-r}{j}=\binom{r+j-1}{j}$, we obtain

$$\frac{a_{-r}^{(i)}}{(z-z_i)^r}=\sum_{j=0}^{\infty}(-1)^r a_{-r}^{(i)}\binom{r+j-1}{j}\frac{z^j}{z_i^{r+j}}.$$

Therefore, the coefficient of z^n in the expansion of $\Phi(z)$ can be written as

$$\sum_{i=1}^{N}\sum_{r=1}^{k_i}(-1)^r a_{-r}^{(i)}\binom{r+n-1}{n}\frac{1}{z_i^{r+n}}.$$

Since the partial fraction expansion for a rational generating function is an exact formula, we can obtain the sequence exactly by using the expansion if all the roots of the denominator are available. Generally, roots of algebraic equations of degree more than 4 do not have an algebraic expression with a finite number of operations involving just the coefficients of the algebraic equations. Hence, we need repeated methods for getting all the roots until they satisfy sufficient precision.

3. Probability and cumulative probability functions of the geometric distribution of order k

The generating functions of the probability and the cumulative probability of the geometric distribution of order k is written as

$$\phi(t)=\frac{(1-pt)(pt)^k}{1-t+p^kqt^{k+1}}$$

and

$$\psi(t)=\frac{(1-pt)(pt)^k}{(1-t)(1-t+p^kqt^{k+1})}.$$

These functions are written in Balakrishnan and Koutras[**3**]. Let us study the roots of the polynomial $1-t+p^kqt^{k+1}$ which is a factor of the both denominators above. For simplicity, dividing the polynomial by the constant so that its leading coefficient becomes 1, we shall study the roots of the polynomial $f(t)=t^{k+1}-ct+c$, where $c=\frac{1}{p^kq}$.

LEMMA 1. *If $0<p<1$ and $p\neq\frac{k}{k+1}$, then $f(t)=t^{k+1}-ct+c=0$ does not have a multiple root.*

PROOF. Assume that the equation $f(t)=0$ has a multiple root α. Then, $f(\alpha)=f'(\alpha)=0$ holds. Since $f'(t)=(k+1)t^k-c$ and $c>0$, it holds that $\alpha^k=\frac{c}{k+1}$. Thus, there exists a nonnegative integer $\ell\in\{0,1,2,\dots,k-1\}$ such that $\alpha=\sqrt[k]{\frac{c}{k+1}}e^{i\frac{2\pi\ell}{k}}$ holds, where $\sqrt[k]{\frac{c}{k+1}}$ is the positive k-th root of $\frac{c}{k+1}$. $f(\alpha)=0$ implies that

$$\left(\sqrt[k]{\frac{c}{k+1}}\right)^{k+1}e^{i\frac{2\pi\ell(k+1)}{k}}-c\sqrt[k]{\frac{c}{k+1}}e^{i\frac{2\pi\ell}{k}}+c=0.$$

Therefore, by dividing the both sides by c (>0), we obtain

$$\sqrt[k]{\frac{c}{k+1}}e^{i2\pi\frac{\ell}{k}}\left(\frac{1}{k+1}e^{i2\pi\ell}-1\right)+1=0.$$

Here, $e^{i2\pi\ell}=1$ implies

$$\frac{k}{k+1}\sqrt[k]{\frac{c}{k+1}}e^{i2\pi\frac{\ell}{k}}=1.$$

Further, since $\frac{k}{k+1}\sqrt[k]{\frac{c}{k+1}}>0$, $e^{i2\pi\frac{\ell}{k}}$ must be a positive real number. Therefore, $\ell=0$, $e^{i2\pi\frac{\ell}{k}}=1$ and $\frac{k}{k+1}\sqrt[k]{\frac{c}{k+1}}=1$ hold. Consequently, we obtain

$$\frac{1}{c}=p^k(1-p)=\frac{k^k}{(k+1)^{k+1}}.$$

Let us examine whether such a p exists between 0 and 1. We set $g(p)=p^k(1-p)$. Since $g'(p)=kp^{k-1}(1-p)-p^k$, we see that $g(p)$ takes its maximum $g(\frac{k}{k+1})=\frac{k^k}{(k+1)^{k+1}}$ at $p=\frac{k}{k+1}$. Therefore, the value of p is only $p=\frac{k}{k+1}$ which satisfies $\frac{1}{c}=p^k(1-p)=\frac{k^k}{(k+1)^{k+1}}$. Thus, if $p\neq\frac{k}{k+1}$, $f(t)=t^{k+1}-ct+c=0$ does not have a multiple root. This completes the proof. $\square$

REMARK 1. In Feller[**6**] (p.236) it is proved that $f(t)=t^{k+1}-ct+c=0$ does not have a multiple root if $p<\frac{k}{k+1}$. Lemma 1 proves the statement not only for $p\in(0,\frac{k}{k+1})$ but also for $p\in(\frac{k}{k+1},1)$. When $p=\frac{k}{k+1}$, we see that the multiple root is $\frac{1}{p}=\frac{k+1}{k}$.

PROPOSITION 3. *For $0<p<1$, let $q=1-p$. We denote by $F(x)$ the cumulative probability at x of the geometric distribution of order k, where k is a positive integer. Then, all the roots $z_1,z_2,\dots,z_k$ of the polynomial $h(t)=1-qt-pqt^2-\cdots-p^{k-1}qt^k$ are simple. Further, for $i=0,1,2,\dots,k$, there exist complex* *numbers* $\rho_0, \rho_1, \rho_2, \ldots, \rho_k$ *such that*

$$F(x) = \sum_{i=0}^{k} \frac{\rho_i}{z_i^{x-k+1}},$$

where $z_0 = 1$, $V(t) = (1-t)h(t)$ *and for* $i = 0, 1, 2, \ldots, k$, $\rho_i = \frac{-p^k}{V'(z_i)}$.

Proof. From Remark 1 we see that the only one root $t = \frac{1}{p}$ of $f(t)$ may possibly be a multiple root which is also a root of the numerator of $\phi(t)$ and $\psi(t)$. Dividing by $(1-pt)$ the numerator and the denominator of $\phi(t)$ and $\psi(t)$, we have
$$\phi(t) = \frac{(pt)^k}{h(t)} \text{ and } \psi(t) = \frac{(pt)^k}{(1-t)h(t)}, \text{ where } h(t) = 1 - qt - pqt^2 - \cdots - p^{k-1}qt^k.$$
Then, Lemma 1 implies that $h(t)$ does not have a multiple root. Noting that $h(1) = p^k \neq 0$, we see that $(1-t)h(t)$ also does not have a multiple root. Here, we shall use Proposition 1. The generating function of the cumulative probabilities of the geometric distribution of order k, which is shifted so that the support begins with 0, is written as

$$\Phi(t) = \frac{p^k}{(1-t)h(t)}.$$

Let $z_1, z_2, \ldots, z_k$ be the roots of the polynomial $h(t)$. Setting $z_0 = 1$, for $i = 0, 1, \ldots, k$, we define

$$\rho_i = \frac{-p^k}{V'(z_i)},$$

where $V(t) = (1-t)h(t)$. Then, since the generating function meets the assumption of Proposition 1, we obtain the following partial fraction expansion

$$\begin{aligned}\Phi(z) &= \sum_{i=0}^{k} \frac{\rho_i}{(z_i - z)} \\ &= \sum_{i=0}^{k} \frac{\rho_i}{z_i} \frac{1}{1 - \frac{z}{z_i}} \\ &= \sum_{i=0}^{k} \frac{\rho_i}{z_i} \sum_{n=0}^{\infty} \left(\frac{z}{z_i}\right)^n.\end{aligned}$$

The coefficient of z^n of the generating function $\Phi(z)$ is the cumulative probability at n of the shifted geometric distribution of order k. Therefore, the cumulative probability at x of the usual geometric distribution of order k is written as

$$F(x) = \sum_{i=0}^{k} \frac{\rho_i}{z_i^{x-k+1}},$$

which completes the proof. □

EXAMPLE 1. By using the algorithm given in Proposition 3, we shall calculate some cumulative probabilities of the geometric distribution of order k using **R**, which is an open-source software environment for statistical computing and graphics [**11**]. In Appendix, we provide R source programs including functions `pgeometric(x, k, p)`, `dgeometric(x, k, p)`, `pgeo(x, k, p)`, `dgeo(x, k, p)`. Here, `pgeometric()` and `dgeometric()` are based on Proposition 3, and `pgeo()` and `dgeo()` are based on the recurrence relation given in Section 1. We have calculated the cumulative probabilities of $G_{10}(0.3)$ for $x = 100, 1000, 10000, 100000, 1000000, 10000000$.

As R's reply to each command

```
pgeometric(c(100,1000,10000,100000,1000000,10000000),10,0.3)
```

and

```
pgeo(c(100,1000,10000,100000,1000000,10000000),10,0.3) ,
```

the following same sequence

```
[1] 0.0003778577 0.0040897915 0.0404594110 0.3385530590
[5] 0.9839744984 1.0000000000
```

is returned. However, the elapsed time of the calculation only for $x = 10000000$ by the function `pgeo()` was over 30 seconds whereas the elapsed time was zero in seconds by the function `pgeometric()` in my personal computer. In fact, R's reply to the commands

```
system.time(pgeo(10000000,10,0.3))[3]
```

and

```
system.time(pgeometric(10000000,10,0.3))[3]
```

are 36.92 and 0, respectively.

4. Waiting time for the (k_1, k_2)-event

In this section we shall give an example of distributions whose generating functions of the probability and the cumulative probability are rational functions. Moreover, the numerators and the denominators of the functions do not have multiple roots.

Let k_1 and k_2 be integers greater than one. Let T be the waiting time in independent $\{0,1\}$-valued sequence for 1-run of length k_2 just after 0-run of length k_1 or more. The distribution of T is the waiting time for the (k_1, k_2)-event (see Huang and Tsai[**8**], Dafnis, Antzoulakos, and Philippou[**5**], and Stefanov and Manca[**13**]).

Let us derive the probability generating function $\phi(t)$ of T by using the method of conditional probability generating functions. Let $\phi_0(t)$ be the conditional probability generating function of the waiting time for the (k_1, k_2)-event from starting with a time at which a 0, to be precise a 0-run of length 1, is observed. Similarly, we define the conditional probability generating function $\phi_1(t)$ of the waiting time for the (k_1, k_2)-event from starting with a time at which a 1 is observed just after 0-run of length k_1 or more. Then, the following relations hold.

$$\begin{cases} \phi(t) &= \sum_{i=1}^{\infty} p^{i-1} q t^i \phi_0(t) \\ \phi_0(t) &= \sum_{i=1}^{k_1-1} q^{i-1} p t^i \phi(t) + \sum_{i=k_1}^{\infty} q^{i-1} p t^i \phi_1(t) \\ \phi_1(t) &= \sum_{i=1}^{k_2-1} p^{i-1} q t^i \phi_0(t) + p^{k_2-1} t^{k_2-1}. \end{cases}$$

We shall solve the above equations for obtaining $\phi(t)$. We can rewrite each equation as

$$\begin{cases} \phi(t) &= \frac{qt}{1-pt} \phi_0(t) \\ \phi_0 &= \frac{pt(1-(qt)^{k_1-1})}{1-qt} \phi(t) + \frac{ptq^{k_1-1}t^{k_1-1}}{1-pt} \phi_1(t) \\ \phi_1(t) &= \frac{qt(1-(pt)^{k_2-1})}{1-pt} \phi_0(t) + (pt)^{k_2-1}. \end{cases}$$

By deleting $\phi_1(t)$ from the last two equations, and substituting the first equation, we obtain an equation which includes only $\phi_0(t)$. Solving the equation we have

$$\phi_0(t) = \frac{(pt)^{k_2}(qt)^{k_1-1}(1-pt)}{1-t+(pt)^{k_2}(qt)^{k_1}}.$$

By multiplying $\frac{qt}{1-pt}$, we obtain

$$\phi(t) = \frac{(pt)^{k_2}(qt)^{k_1}}{1-t+(pt)^{k_2}(qt)^{k_1}}.$$

The roots of the denominator of $\phi(t)$ is the same as the roots of the polynomial

$f(t) = t^{k_1+k_2} - ct + c$, where $c = \frac{1}{p^{k_2} q^{k_1}}$.

PROPOSITION 4. *Let k_1 and k_2 be integers greater than one. Then, all the roots of $f(t)$ are simple.*

PROOF. By setting $k = k_1 + k_2$, we can write $f'(t) = kt^{k-1} - c$. Suppose that the equation $f(t) = 0$ has a multiple root α. Then, from $f'(\alpha) = 0$, we see that

$\alpha^{k-1} = \frac{c}{k}$ and hence there exists an integer $\ell \in \{0, 1, 2, \ldots, k-2\}$ such that

$$\alpha = \sqrt[k-1]{\frac{c}{k}} e^{i\frac{2\pi\ell}{k-1}}$$

holds. Since α is also a root of $f(t)$, it holds that

$$\frac{c}{k} \sqrt[k-1]{\frac{c}{k}} e^{i2\pi\ell} e^{i\frac{2\pi\ell}{k-1}} - c \sqrt[k-1]{\frac{c}{k}} e^{i\frac{2\pi\ell}{k-1}} + c = 0.$$

Dividing by c the both sides of the equation, we obtain

$$\sqrt[k-1]{\frac{c}{k}} e^{i\frac{2\pi\ell}{k-1}} \left(\frac{1}{k} e^{i2\pi\ell} - 1 \right) + 1 = 0.$$

Noting that $e^{i2\pi\ell} = 1$, we see that

$$\frac{k-1}{k} \sqrt[k-1]{\frac{c}{k}} e^{i\frac{2\pi\ell}{k-1}} = 1,$$

and hence we have $\ell = 0$. Therefore, we obtain $c = \frac{k^k}{(k-1)^{k-1}}$. Here, we examine whether some $p \in (0,1)$ exists satisfying

$$\frac{1}{c} = p^{k_2} q^{k_1} = \frac{(k-1)^{k-1}}{k^k} = \frac{(k_1+k_2-1)^{k_1+k_2-1}}{(k_1+k_2)^{k_1+k_2}}.$$

We set $g(p) = p^{k_2}(1-p)^{k_1}$ for $p \in (0,1)$. Then, $g(p)$ attains its maximum value at $p = \frac{k_2}{k_1+k_2}$ and the maximum value is

$$g\left(\frac{k_2}{k_1+k_2}\right) = \frac{k_1^{k_1} k_2^{k_2}}{(k_1+k_2)^{k_1+k_2}}.$$

But, since $k_1 > 1$ and $k_2 > 1$ hold, the next inequality holds

$$\frac{k_1^{k_1} k_2^{k_2}}{(k_1+k_2)^{k_1+k_2}} < \frac{(k_1+k_2-1)^{k_1+k_2-1}}{(k_1+k_2)^{k_1+k_2}},$$

and we see that such a $p \in (0,1)$ satisfying the above equation does not exist. This is a contradiction with the assumption that $f(t)$ has a multiple root. Finally, we shall prove the above inequality. Let $a > 1$ be a constant and let $h(x) = (x+a-1)^{x+a-1} - x^x a^a$. Since $h'(x) = (1+\log(x+a-1))(x+a-1)^{x+a-1} - (1+\log x)x^x > 0$ and $h(1) = 0$, $h(x) > 0$ holds for all $x > 1$. Thus, by setting $x = k_1$ and $a = k_2$, we obtain the above inequality. This completes the proof. □

EXAMPLE 2. By using the algorithm given in Proposition 4, we shall calculate some cumulative probabilities of the waiting time for (k_1, k_2)-event using **R**. In

Table 1. Values of cumulative probabilities of the waiting time for (k_1, k_2)-event with $k_1 = k_2 = 6$. These values are calculated based on the partial fraction expansion of the generating function of cumulative probabilities of the distribution given in Proposition 4.

p	$x = 100$	$x = 1000$	$x = 10000$	$x = 100000$
0.1	0.00005	0.00053	0.00529	0.05175
0.2	0.00149	0.01646	0.15432	0.81323
0.3	0.00761	0.08140	0.57581	0.99981
0.4	0.01690	0.17256	0.85238	1.00000
0.5	0.02155	0.21505	0.91333	1.00000
0.6	0.01690	0.17256	0.85238	1.00000
0.7	0.00761	0.08140	0.57581	0.99981
0.8	0.00149	0.01646	0.15432	0.81323
0.9	0.00005	0.00053	0.00529	0.05175

Appendix, we provide R functions `pk1k2(x, k1, k2, p)` and `dk1k2(x, k1, k2, p)`. Here, we have tabulated the cumulative probabilities of (6,6)-event for $x = 100, 1000, 10000$ and 100000 with $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$. Since $k_1 = k_2 = 6$, the cumulative probabilities are symmetric with respect to p about $p = \frac{1}{2}$. Every row of the table has been calculated by the following R's command

```
round(pk1k2(c(100,1000,10000,100000),6,6, p),digits=5) .
```

5. Some explicit expressions of waiting times

In the previous sections, we have seen that the approach from the partial fraction expansion is very useful for numerical calculations of probabilities and cumulative probabilities of waiting time distributions. However, when the roots of the denominator of a rational generating function are written explicitly, we can obtain explicit expressions of the probabilities or cumulative probabilities. As examples of such cases, we shall discuss explicit expressions of the geometric distributions of order 2, 3 and 4. Though some combinatorial explicit expressions of probabilities of the distributions are known, explicit expressions without multinomial or binomial coefficients may be of interest.

5.1. Geometric distribution of order 2

Let us provide an explicit expression of the probability function of the geometric distribution of order 2. The probability generating function of the shifted geometric distribution of order 2 so that its support begins with zero is written as

$$\phi(z) = \frac{p^2}{1 - qt - pqt^2} = \frac{-\frac{p}{q}}{(z_1 - z)(z_2 - z)},$$

where z_1 and z_2 are the roots of $pqz^2 + qz - 1 = 0$, that is,

$$z_1 = \frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p} \qquad \text{and} \qquad z_2 = -\frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p}.$$

The constants a and b satisfying $\phi(z) = \dfrac{a}{z_1 - z} + \dfrac{b}{z_2 - z}$ are $a = \dfrac{p}{q}\sqrt{\dfrac{p^2q}{4p+q}}$ and $b = -a$. Then, we see that

$$\begin{aligned}\phi(z) &= \frac{a}{z_1 - z} + \frac{b}{z_2 - z} \\ &= \frac{a}{z_1}\sum_{n=0}^{\infty}\left(\frac{z}{z_1}\right)^n - \frac{a}{z_2}\sum_{n=0}^{\infty}\left(\frac{z}{z_2}\right)^n.\end{aligned}$$

By extracting the coefficient of z^n in the probability generating function, we see that the value of the probability function at n is

$$\frac{a}{z_1}\frac{1}{z_1^n} - \frac{a}{z_2}\frac{1}{z_2^n} = \frac{a}{z_1^{n+1}} - \frac{a}{z_2^{n+1}}.$$

Therefore, since the probability at n of the usual geometric distribution of order 2 is the coefficient of z^{n-2} in the above expansion, if X follows the geometric distribution of order 2, the probability function of X can be written as

$$\begin{aligned}P(X = x) &= \frac{p}{q}\sqrt{\frac{p^2q}{4p+q}}\left(\frac{1}{\left(\frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p}\right)^{x-1}} - \frac{1}{\left(-\frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p}\right)^{x-1}}\right) \\ &= \frac{p^2}{\sqrt{4pq+q^2}}\left\{\left(\frac{\sqrt{4pq+q^2}+q}{2}\right)^{x-1} + (-1)^x\left(\frac{\sqrt{4pq+q^2}-q}{2}\right)^{x-1}\right\} \\ &= \frac{p^2}{\sqrt{q^2+4pq}}\left\{\left(\frac{q+\sqrt{q^2+4pq}}{2}\right)^{x-1} - \left(\frac{q-\sqrt{q^2+4pq}}{2}\right)^{x-1}\right\}.\end{aligned}$$

Therefore, we have obtained the next proposition.

PROPOSITION 5. *If X follows the geometric distribution of order 2 with parameter p, the probability function can be written as*

$$P(X = x) = \frac{p^2}{\sqrt{q^2+4pq}}\left\{\left(\frac{q+\sqrt{q^2+4pq}}{2}\right)^{x-1} - \left(\frac{q-\sqrt{q^2+4pq}}{2}\right)^{x-1}\right\},$$

for $x = 2, 3, \ldots$.

EXAMPLE 3. In order to check the above result numerically, we have made the following R function `dgeom2(x,p)` based on Proposition 5.

```
dgeom2<-function(x,p){
 q<-1-p
 a<-sqrt(4*p*q+q^2)
 z1<-(a+q)/2
 z2<-(q-a)/2
 p^2/a*(z1^(x-1)-z2^(x-1))}
```

To compare the values of probabilities of the geometric distribution of order 2 based on `dgeom2(x,p)` with the values calculated by `dgeo(x,k,p)` based on the recurrence relation given in Section 1, we have verified R's replies to the commands

```
round(dgeom2(1:20,0.7),digits=5)
round(dgeo(1:20,2,0.7),digits=5)
```

are the same as the following sequence

```
 [1] 0.00000 0.49000 0.14700 0.14700 0.07497 0.05336
 [7] 0.03175 0.02073 0.01289 0.00822 0.00517 0.00328
[13] 0.00207 0.00131 0.00083 0.00052 0.00033 0.00021
[19] 0.00013 0.00008 .
```

However, the values of elapsed time for the commands

```
dgeom2(100000,0.01)
```

and

```
dgeo(100000,2,0.01)
```

were 0 and 0.45 in seconds, respectively, whereas the values of the calculation were the same as `4.957091e-09`.

5.2. Geometric distribution of order 3

We can obtain an explicit expression of the probability function of the geometric distribution of order 3. The probability generating function of the shifted geometric distribution of order 3 can be represented as

$$\phi(z) = \frac{p^3}{1 - qz - pqz^2 - p^2qz^3} = \frac{-\frac{p}{q}}{(z_1 - z)(z_2 - z)(z_3 - z)},$$

where z_1, z_2 and z_3 are the roots of $p^2qz^3 + pqz^2 + qz - 1 = 0$. Setting

$$A = \sqrt[3]{\frac{\sqrt{27p^2 + 14pq + 3q^2}}{6\sqrt{3}p^3q} + \frac{27p + 7q}{54p^3q}},$$

the roots are represented as

$$
\begin{aligned}
z_1 &= A - \frac{2}{9p^2A} - \frac{1}{3p} \\
z_2 &= \left(-\frac{1}{2}A + \frac{1}{9p^2A} - \frac{1}{3p}\right) + \left(\frac{\sqrt{3}}{2}A + \frac{\sqrt{3}}{9p^2A}\right) i \\
z_3 &= \left(-\frac{1}{2}A + \frac{1}{9p^2A} - \frac{1}{3p}\right) - \left(\frac{\sqrt{3}}{2}A + \frac{\sqrt{3}}{9p^2A}\right) i, \text{ where } i = \sqrt{-1}.
\end{aligned}
$$

For $j = 1, 2, 3$, we set

$$\rho_j = \frac{\frac{p}{q}}{3{z_j}^2 + \frac{2}{p} z_j + \frac{1}{p^2}}.$$

Then, the probability generating function can be written as

$$\phi(z) = \frac{\rho_1}{z_1 - z} + \frac{\rho_2}{z_2 - z} + \frac{\rho_3}{z_3 - z}.$$

Therefore, if X follows the usual geometric distribution of order 3, it holds that

$$P(X = x) = \frac{\rho_1}{z_1^{x-2}} + \frac{\rho_2}{z_2^{x-2}} + \frac{\rho_3}{z_3^{x-2}}.$$

EXAMPLE 4. In order to check the above result numerically, we have made the following R function `dgeom3(x,p)` based on the above result.

```
dgeom3<-function(x,p){
 q<-1-p
 A<-sqrt(27*p^2+14*p*q+3*q^2)/6/sqrt(3)/p^3/q
 A<-(A+(27*p+7*q)/(54*p^3*q))^(1/3)
 z1<-A-2/(9*p^2*A)-1/3/p
 a<--A/2+1/(9*p^2*A)-1/3/p
 b<-sqrt(3)*A/2+sqrt(3)/(9*p^2*A)
 z2<-a+b*1i
 z3<-a-b*1i
 r1<-p/q/(3*z1^2+2*z1/p+1/p^2)
 r2<-p/q/(3*z2^2+2*z2/p+1/p^2)
 r3<-p/q/(3*z3^2+2*z3/p+1/p^2)
 prob<-r1/z1^(x-2)+r2/z2^(x-2)+r3/z3^(x-2)
 Re(prob)}
```

To compare the values of probabilities of the geometric distribution of order 3 based on `dgeom3(x,p)` with the values calculated by `dgeo(x,k,p)` based on the recurrence relation given in Section 1, we have verified R's replies to the commands

```
round(dgeom3(1:20,0.7),digits=5)
round(dgeo(1:20,3,0.7),digits=5)
```

are the same as the following sequence

```
 [1] 0.00000 0.00000 0.34300 0.10290 0.10290 0.10290
 [7] 0.06761 0.05702 0.04643 0.03584 0.02888 0.02302
[13] 0.01824 0.01455 0.01158 0.00921 0.00733 0.00584
[19] 0.00464 0.00370 .
```

However, the values of elapsed time for the commands

```
dgeom3(10000,0.1)
```

and

```
dgeo(10000,3,0.1)
```

were 0 and 0.08 in seconds, respectively, whereas the values of the calculation were the same as `1.085257e-07`.

5.3. Geometric distribution of order 4

Let us expand the probability genarating function of the geometric distribution of order 4. We modify the probability generating function of the shifted geometric distribution of order 4 as

$$\phi(z) = \frac{p^4}{1 - qz - pqz^2 - p^2qz^3 - p^3qz^4} = \frac{-\frac{p}{q}}{(z_1 - z)(z_2 - z)(z_3 - z)(z_4 - z)},$$

where z_1, z_2, z_3 and z_4 are the root of $p^3qz^4 + p^2qz^3 + pqz^2 + qz - 1 = 0$. Setting

$$A = \sqrt[3]{\frac{\sqrt{\frac{256p^3+203p^2q+88pq^2+16q^3}{q}}}{6\sqrt{3}p^6q} + \frac{45p + 20q}{54p^6q}}$$

and

$$B = \sqrt{\frac{-15p^2qA + 36p^4qA^2 - 8q - 48p}{qA}},$$

we can write the four roots of the denominator of the probability generating function as follows.

$$z_1 = \frac{\sqrt{\frac{15}{2pB} - A + \frac{12p+2q}{9p^4qA} - \frac{5}{6p^2}}}{2} - \frac{B}{12p^2} - \frac{1}{4p},$$

$$z_2 = -\frac{\sqrt{\frac{15}{2pB} - A + \frac{12p+2q}{9p^4qA} - \frac{5}{6p^2}}}{2} - \frac{B}{12p^2} - \frac{1}{4p},$$

$$z_3 = \frac{\sqrt{\frac{15}{2pB} + A - \frac{12p+2q}{9p^4qA} + \frac{5}{6p^2}}}{2}i + \frac{B}{12p^2} - \frac{1}{4p},$$

and

$$z_4 = -\frac{\sqrt{\frac{15}{2pB} + A - \frac{12p+2q}{9p^4qA} + \frac{5}{6p^2}}}{2}i + \frac{B}{12p^2} - \frac{1}{4p}, \text{ where } i = \sqrt{-1}.$$

Here, for $j = 1, 2, 3, 4$, defining

$$\rho_j = \frac{\frac{p}{q}}{4z_j^3 + \frac{3}{p}z_j^2 + \frac{2}{p^2}z_j + \frac{1}{p^3}},$$

we obtain

$$\phi(z) = \frac{\rho_1}{z_1 - z} + \frac{\rho_2}{z_2 - z} + \frac{\rho_3}{z_3 - z} + \frac{\rho_4}{z_4 - z}.$$

Therefore, if X follows the usual geometric distribution of order 4, the probability function can be written as

$$P(X = x) = \frac{\rho_1}{z_1^{x-3}} + \frac{\rho_2}{z_2^{x-3}} + \frac{\rho_3}{z_3^{x-3}} + \frac{\rho_4}{z_4^{x-3}}.$$

Here, z_1 is the unique positive root which has the minimum absolute value among the roots. The root z_2 is the unique negative root and z_3 and z_4 are imaginary root which are conjugate to each other.

EXAMPLE 5. In order to check the above result numerically, we have made the following R function `dgeom4(x,p)` based on the above result.

```
dgeom4<-function(x,p){
 q<-1-p
 A<-sqrt((256*p^3+203*p^2*q+88*p*q^2+16*q^3)/q)/(6*sqrt(3)*p^6*q)
 A<-A + (45*p+20*q)/(54*p^6*q)
 A<-A^(1/3)
 B<-sqrt((-15*p^2*q*A+36*p^4*q*A^2-8*q-48*p)/(q*A))
 C<--A+(12*p+2*q)/(9*p^4*q*A)-5/(6*p^2)
 z1<-sqrt(15/(2*p*B)+C)/2-B/(12*p^2)-1/(4*p)
 z2<--sqrt(15/(2*p*B)+C)/2-B/(12*p^2)-1/(4*p)
 z3<--sqrt(-(-15/(2*p*B)+C))*1i/2+B/(12*p^2)-1/(4*p)
```

```
z4<-sqrt(-(-15/(2*p*B)+C))*1i/2+B/(12*p^2)-1/(4*p)
r1<-p/q/(4*z1^3+3*z1^2/p+2*z1/p^2+1/p^3)
r2<-p/q/(4*z2^3+3*z2^2/p+2*z2/p^2+1/p^3)
r3<-p/q/(4*z3^3+3*z3^2/p+2*z3/p^2+1/p^3)
r4<-p/q/(4*z4^3+3*z4^2/p+2*z4/p^2+1/p^3)
prob<-r1/z1^(x-3)+r2/z2^(x-3)+r3/z3^(x-3)+r4/z4^(x-3)
Re(prob)}
```

To compare the values of probabilities of the geometric distribution of order 4 based on `dgeom4(x,p)` with the values calculated by `dgeo(x,k,p)` based on the recurrence relation given in Section 1, we have verified R's replies to the commands

```
round(dgeom4(1:20,0.7),digits=5)
round(dgeo(1:20,4,0.7),digits=5)
```

are the same as the following sequence

```
 [1] 0.00000 0.00000 0.00000 0.24010 0.07203 0.07203
 [7] 0.07203 0.07203 0.05474 0.04955 0.04436 0.03917
[13] 0.03398 0.03004 0.02647 0.02328 0.02045 0.01801
[19] 0.01584 0.01394 .
```

However, the values of elapsed time for the commands

```
dgeom4(10000,0.1)
```

and

```
dgeo(10000,4,0.1)
```

were 0 and 0.08 in seconds, respectively, whereas the values of the calculation were the same as `3.660392e-05`.

6. Explicit expressions of the probability and the cumulative probability functions of the negative binomial distribution of order 2

Though we can treat general order of the negative binomial distribution of order k as Shmuelli and Cohen[**12**], we shall provide an explicit expression of the probability function and the cumulative probability function of the negative binomial distribution of order 2.

The probability generating distribution of the shifted negative binomial distribution of order 2, to be denoted by $\overline{NB_2(r,p)}$, can be written as

$$\left(\frac{p^2}{1-qz-pqz^2}\right)^r.$$

Here, the shifted negative binomial distribution of order k is the distribution of $X - kr$, where X follows the usual negative binomial distribution of order k. Similarly as in the case of the geometric distribution of order 2, we obtain two distinct roots as

$$z_1 = \frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p},$$

and

$$z_2 = -\frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p}.$$

These roots are multiple roots of the denominator of the generating function with multiplicity r. Therefore, the probability function can be written as follows. If W follows $\overline{NB_2(r,p)}$, then it holds that

$$P(W = x) = \sum_{i=1}^{2}\sum_{j=1}^{r}(-1)^j a_{-j}^{(i)}\binom{j+x-1}{j-1}\frac{1}{z_i^{j+x}},$$

where

$$\begin{aligned} a_{-j}^{(1)} &= \frac{(-1)^r(\frac{p}{q})^r}{(r-j)!}\left(\frac{d^{r-j}}{dz^{r-j}}\frac{1}{(z-z_2)^r}\right)_{|z=z_1} \\ &= (-1)^{2r-j}(\frac{p}{q})^r\binom{2r-j-1}{r-j}\frac{1}{(z_1-z_2)^{2r-j}}, \\ a_{-j}^{(2)} &= (-1)^{2r-j}(\frac{p}{q})^r\binom{2r-j-1}{r-j}\frac{1}{(z_2-z_1)^{2r-j}}, \end{aligned}$$

for $j = 1, 2, \ldots, r$. Further calculation implies the next expression,

$$\begin{aligned} P(W = x) = {} & \frac{p^{3r}}{(4p+q)^r}\sum_{j=1}^{r}\binom{2r-j-1}{r-j}\binom{j+x-1}{j-1}\left(\frac{\sqrt{q^2+4pq}}{pq}\right)^j \\ & \times\left\{\left(\frac{\sqrt{q^2+4pq}+q}{2}\right)^{j+x} + (-1)^x\left(\frac{\sqrt{q^2+4pq}-q}{2}\right)^{j+x}\right\}. \end{aligned}$$

The generating function of the cumulative probabilities of $\overline{NB_2(r,p)}$ is given by

$$\begin{aligned} \phi(z) &= \frac{1}{1-z}\left(\frac{p^2}{1-qz-pqz^2}\right)^r \\ &= (-1)^{r+1}\left(\frac{p}{q}\right)^r\frac{1}{(z-1)(z-z_1)^r(z-z_2)^r} \end{aligned}$$

$$= (-1)^{r+1}\left(\frac{p}{q}\right)^r \left\{ \frac{b^{(0)}_{-1}}{z-1} + \sum_{i=1}^{2}\sum_{j=1}^{r} \frac{b^{(i)}_{-j}}{(z-z_i)^j} \right\},$$

where z_1 and z_2 are the same as the probability generating function,

$$z_1 = \frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p},$$

and

$$z_2 = -\frac{1}{2}\sqrt{\frac{4p+q}{p^2q}} - \frac{1}{2p}.$$

For $i = 1, 2$, $b^{(i)}_{-j}$ can be derived by the next formula

$$b^{(i)}_{-j} = \frac{1}{(r-j)!}\lim_{z\to z_i} \frac{d^{r-j}}{dz^{r-j}} \left\{ \frac{(z-z_i)^r}{(z-1)(z-z_1)^r(z-z_2)^r} \right\}.$$

Thus, we can write $b^{(i)}_{-j}$ for $i = 1, 2$ by the next formulas.

$$\begin{aligned} b^{(1)}_{-j} &= \frac{(-1)^{r-j}}{(r-j)!}\sum_{k=0}^{r-j}\binom{r-j}{k} k! \frac{(2r-j-k-1)!}{(r-1)!}(z_1-1)^{-(k+1)}(z_1-z_2)^{-(2r-j-k)} \\ &= (-1)^{r-j}\sum_{k=0}^{r-j}\binom{2r-j-k-1}{r-1}\frac{1}{(z_1-1)^{k+1}(z_1-z_2)^{2r-j-k}}, \end{aligned}$$

and

$$b^{(2)}_{-j} = (-1)^{r-j}\sum_{k=0}^{r-j}\binom{2r-j-k-1}{r-1}\frac{1}{(z_2-1)^{k+1}(z_2-z_1)^{2r-j-k}}.$$

Consequently, we obtain the next explicit expression of the cumulative distribution function.

$$P(W \le x) = 1 + (-1)^{r+1}\left(\frac{p}{q}\right)^r \sum_{j=1}^{r}\sum_{i=1}^{2}(-1)^j b^{(i)}_{-j}\binom{x+j-1}{j-1}\frac{1}{z_i^{j+x}}.$$

EXAMPLE 6. As a numerical application of the result in this section, we shall calculate exact cumulative probabilities of the negative binomial distribution of order 2 for large x. We have provided in Appendex R functions `dnb(x,r,p)` and `pnb(x,r,p)` besed on the above result. They are R functions for calculating the probability mass function and the cumulative probability function, respectively. When p is small and/or r is large, the r-th 1-run of length 2 does not occur soon. For example, let us evaluate numerically the cumulative probabilities of the negative binomial distribution with $r = 30$ and $p = 0.02$ for $x = 50000, 70000, 90000, 100000$. R's reply to the command

```
round(pnb2(c(50000,70000,90000,100000),30,0.02),digits=5)
```

was $(0.01751, 0.33958, 0.83630, 0.94517)$. Each cumulative probability agrees with the replies to the next commands,

```
round(sum(dnb2(1:50000,30,0.02)),digits=5)
round(sum(dnb2(1:70000,30,0.02)),digits=5)
round(sum(dnb2(1:90000,30,0.02)),digits=5)
round(sum(dnb2(1:100000,30,0.02)),digits=5),
```

which are appropriate sum of probabilities calculated by `dnb2(x,r,p)`.

Appendix: R source programs

We provide here the R source programs including functions used in examples of the manuscript. The functions `h()` and `h1()` are R functions for polynomials $f(t)$ and $f'(t)$, respectively, in Section 3. The functions `pgeometric()` and `dgeometric()` are the cumulative probability function and the probability mass function, respectively, based on the partial fraction expansion of the generating functions of the geometric distribution of order k in Section 3. The function `geom()` is the R function for the recurrence relation in Section 1. The R functions `pgeo()` and `dgeo()` are the cumulative probability function and the probability mass function, respectively, based on the recurrence relation.

The R functions `pk1k2(x, k1, k2, p)` and `dk1k2(x, k1, k2, p)` are the cumulative probability function and the probability mass function of the waiting time for (k_1, k_2)-events.

The R functions `dnb2(x,r,p)` and `pnb2(x,r,p)` are the probability mass function and the cumulative probability function, respectively, which are based on the results given in Section 6.

```
h<-function(t,k,p){
 q<-1-p
 s<-1
 for (i in 1:k){
  s<-s-q*p^(i-1)*t^i}
 s}

h1<-function(t,k,p){
 q<-1-p
 s<--q
 for (i in 1:(k-1)){
  s<-s-q*(i+1)*p^i*t^i}
```

```
 s}

# Cumulative distribution function
#  of the geometric distribution of order k
pgeometric<-function(x,k,p){
  q<-1-p
  poly<-1
  for (i in 1:k){poly<-c(poly,-p^(i-1)*q)}
  z<-polyroot(poly)
  z<-c(1,z)
  r<-c()
  for (i in 1:(k+1)){
   a<--p^k/(-h(z[i],k,p)+(1-z[i])*h1(z[i],k,p))
   r<-c(r,a)}
  prob<-0
  for (i in 1:(k+1)){
   prob<-prob+r[i]/z[i]^(x-k+1)}
  Re(prob)}

# PMF of the geometric distribution of order k
dgeometric<-function(x,k,p){
  q<-1-p
  poly<-1
  for (i in 1:k){poly<-c(poly,-p^(i-1)*q)}
  z<-polyroot(poly)
  r<-c()
  for (i in 1:k){
   a<--p^k/(h1(z[i],k,p))
   r<-c(r,a)}
  prob<-0
  for (i in 1:k){
   prob<-prob+r[i]/z[i]^(x-k+1)}
  Re(prob)}

# The recurrence relation of the pmf
#  of the geometric distribution of order k
geom <- function(k,p,n1=50){
 b<-rep(0,n1+1)
 a1<-p^k; a2<-a1*(1-p)
 b[k+1]<-a1
 for (i in (k+2):(2*k+1)){
  b[i] <- p^k*(1-p) }
 for (i in (2*k+2):(n1+1)){
   b[i] <- b[i-1]-a2*b[i-k-1] }
 return(b) }

dgeo<-function(x,k,p){
len<-length(x)
 m<-max(x)
```

```
 a<-geom(k,p)
 if (m>50){
 a<-geom(k,p,m+1)}
 return(a[x+1])}

pgeo<-function(x,k,p){
  a<-geom(k,p)
  m<-max(x)
  if (m>50){
  a<-geom(k,p,m+1)}
  cp<-cumsum(a)
  return(cp[x+1])}

# Cumulative probability of the waiting time for (k1,k2)-event
pk1k2<-function(x,k1,k2,p){
  q<-(1-p)
  k<-k1+k2
  a<-1/p^k2/q^k1
   V1<-function(t,k1,k2,p){
   q<-1-p
   a<-1/p^k2/q^k1
   k<-k1+k2
   v1<-(k+1)*t^k-k*t^(k-1)-2*a*t+2*a
  v1}
  if (k==2) { pol<-c(-a,2*a,-1,1)} else {
    pol<-c(-a,2*a,-a,rep(0,k-3),-1,1)}
  b<-polyroot(pol)
  r<-c()
  for (i in 1:(k+1)){
   a1<-1/(V1(b[i],k1,k2,p))
   r<-c(r,a1)}
  prob<-0
  for (i in 1:(k+1)){prob<-prob+r[i]/(b[i])^(x+1-k)}
  if (length(x)>1){
    for (j in 1:length(x)){
     if (x[j]<k ){ prob[j]<-0 }}}
  if (length(x)==1 && x<k) { prob<-0 }
  Re(prob)}

# PMF of the waiting time for (k1,k2)-event
dk1k2<-function(x,k1,k2,p){
  q<-(1-p)
  k<-k1+k2
  a<-1/p^k2/q^k1
   V1<-function(t,k1,k2,p){
   q<-1-p
   a<-1/p^k2/q^k1
   k<-k1+k2
   v1<-k*t^(k-1)-a
```

```
    v1}
   if (k==2) { pol<-c(a,-a,1)} else {
     pol<-c(a,-a,rep(0,k-2),1)}
   b<-polyroot(pol)
   r<-c()
   for (i in 1:k){
    a1<- -1/(V1(b[i],k1,k2,p))
    r<-c(r,a1)}
   prob<-0
   for (i in 1:k){prob<-prob+r[i]/(b[i])^(x+1-k)}
    if (length(x)>1){
     for (j in 1:length(x)){
     if (x[j]<k ){ prob[j]<-0 }}}
   if (length(x)==1 && x<k) { prob<-0 }
   Re(prob)}

# PMF of the negative binomial distribution of order 2
dnb2<-function(x,r,p){
  q<-1-p
  a<-sqrt(q^2+4*p*q)
  prob<-0
  for (j in 1:r){
   b<-choose((2*r-j-1),(r-j))*choose((j+x-1),(j-1))*(a/p/q)^j
   b<-b*(((a+q)/2)^(j+x)+(-1)^x*((a-q)/2)^(j+x))
   prob<-prob+b}
  prob<-prob*p^(3*r)/(4*p+q)^r
  prob}

# CDF of the negative binomial distribution of order 2
pnb2<-function(x,r,p){
  q<-1-p
  z<-sqrt((4*p+q)/(p^2*q))/2
  z1<--1/2/p+z
  z2<--1/2/p-z
  cprob<- 0
  for (j in 1:r){
   aj1<-0
   for (k in 0:(r-j)){
    w<-(-1)^(r-j)*choose(2*r-j-k-1,r-1)/(z1-1)^(k+1)/(z1-z2)^(2*r-j-k)
    aj1<-aj1 + w }
   aj2<-0
   for (k in 0:(r-j)){
    w<-(-1)^(r-j)*choose(2*r-j-k-1,r-1)/(z2-1)^(k+1)/(z2-z1)^(2*r-j-k)
    aj2<-aj2 + w }
   cprob<-cprob+(-1)^j*aj1*choose(x+j-1,j-1)/z1^(j+x)
   cprob<-cprob+(-1)^j*aj2*choose(x+j-1,j-1)/z2^(j+x)
  }
  cprob<-cprob*(-1)^(r+1)*(p/q)^r +1
  cprob}
```

Acknowledgments. I would like to thank the reviewers for helpful comments.

References

[1] Aki, S. (2012). Statistical modeling for discrete patterns in a sequence of exchangeable trials, *Annals of the Institute of Statistical Mathematics*, **64**, 633-655.

[2] Aki, S. and Hirano, K. (2013). Coupon collector's problems with statistical applications to rankings, *Annals of the Institute of Statistical Mathematics*, **65**, 571-587.

[3] Balakrishnan, N., Koutras, M. (2002). *Runs and Sscans with Applications*, Wiley, New York.

[4] Conway, J. B. (1973). *Functions of One Complex Variable*, 2nd edition, Springer-Verlag.

[5] Dafnis, S. D., Antzoulakos, D. L., Philippou, A. N. (2010). Distributions related (k_1, k_2) events, *Journal of Statistical Planning and Inference*, **140**, 1691-1700.

[6] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, Third edition, Wiley, New York.

[7] Flajolet, P., Sedgewick, R. (2010). *Analytic Combinatorics*, Cambridge University Press.

[8] Huang, W., Tsai, C. (1991). On modified binomial distribution of order k, *Statistics and Probability Letters*, 11, 125-131.

[9] Inoue, K., Aki, S. and Hirano, K. (2011). Distributions of simple patterns in some kinds of exchangeable sequences, *Journal of Statistical Planning and Inference*, **141**, 2532-2544.

[10] Pemantle, R., Wilson, M. C. (2013). *Analytic Combinatorics in Several Variables*, Cambridge University Press.

[11] R Core Team (2013). *A Language and Environment for Statistical Computing.* In R Foundation for Statistical Computing (Ed.), Vienna, Austria. URL http://www.R-project.org.

[12] Shmuelli, G. and Cohen, A. (2000). Run-related probability functions applied to sampling inspection", *Technometrics*, **42:2**, 188-202.

[13] Stefanov, V. T., Manca, R. (2013). Distributions associated with (k_1, k_2) events on semi-Markov binary trials, *Journal of Statistical Planning and Inference*, **143**, 1233-1243.

Sigeo Aki

Department of Mathematics

Faculty of Engineering Science

Kansai University

3-3-35 Yamate-cho, Suita-shi

Osaka 564-8680, Japan

E-mail:aki@kansai-u.ac.jp