CUMULATIVE DISTRIBUTION FUNCTION OF MARKOV
ORDER 2 GEOMETRIC DISTRIBUTION OF ORDER K

E. SHMERLING *

Abstract. Simple formulas for calculating the values of probability mass function, cumulative distribution function and moments about the origin of Markov order 2 geometric order \( k \) random variables are derived from a system of first-order difference equations for the probabilities of non-occurrence of a success run of length \( k \) in the first \( n \) trials. Results of experimental calculations that verified the obtained formulas are presented.

AMS(MOS) subject classification. 2000 Mathematics Subject Classification: 60J10; 97K60

Key Words. Markov chain; Geometric distribution of order \( k \); Difference equations

1. Introduction. The distribution of the number of trials until the first occurrence of consecutive \( k \) successes in Bernoulli trials with success probability \( p \) is called a geometric distribution of order \( k \). This definition is due to Philippou, Georghiou and Philippou (see [4]).

A lot of important results in studying the properties of this distribution and its application have been obtained by various researchers (see [2]-[4]).

Markov order \( m \) geometric distribution of order \( k \) designated \( M_mG_k \) is an extension of the geometric distribution of order \( k \) (see [1],[2]).

Let \( X_1, X_2, ... \) be a time homogeneous two-state Markov chain of second order with transition probabilities

\[
p_{ijl} = P(X_t = l | X_{t-1} = j, X_{t-2} = i), \quad t \geq 3, \quad 0 \leq i, j, l \leq 1.
\]

The distribution of waiting time for the first occurrence of a success run of length \( k \) in this case will be called Markov order 2 geometric distribution of order \( k \).

* Department of Computer Science and Mathematics, Ariel University Center of Samaria, Ariel 44837, Israel
An explicit expression for the probability generating function of $M_2G_k$ has been derived (see [1]). It has the form

\begin{equation}
\phi^{0,0} = p_{00} \phi^{0,0}(t) + p_{01} \phi^{0,1}(t) + p_{10} \phi^{1,0}(t) + p_{11} \phi^{1,1}(t),
\end{equation}

where

\begin{align*}
\phi^{0,1} &= \frac{p_{01} p_{11}^{k-2} t^{k-1}}{1 - (p_{10} t + p_{00} q_{10} t^2)(q_{01} t + \frac{1-p_{11} t^{k-2}}{1-p_{11} t} \cdot p_{01} q_{11} t^2)}, \\
\phi^{0,0} &= \frac{p_{01} p_{11}^{k-2} t^{k-1}}{1 - (p_{10} t + p_{00} q_{10} t^2)(q_{01} t + \frac{1-p_{11} t^{k-2}}{1-p_{11} t} \cdot p_{01} q_{11} t^2)}, \\
\phi^{1,0} &= \frac{p_{01} p_{11}^{k-2} t^{k-1} (1 - (p_{10} q_{01} t^2 + p_{00} q_{00} q_{10} t^3))}{1 - (p_{10} t + p_{00} q_{10} t^2)(q_{01} t + \frac{1-p_{11} t^{k-2}}{1-p_{11} t} \cdot p_{01} q_{11} t^2)}, \\
\phi^{1,1} &= \frac{p_{11}^{k-2} t^{k-2} (1 - (p_{10} q_{01} t^2 + p_{00} q_{00} q_{10} t^3))}{1 - (p_{10} t + p_{00} q_{10} t^2)(q_{01} t + \frac{1-p_{11} t^{k-2}}{1-p_{11} t} \cdot p_{01} q_{11} t^2)}.
\end{align*}

Obviously, formulas (2) are very complicated and unsuitable for calculating the values of the probability mass function (pmf), cumulative distribution function (cdf) and moments about the origin of $M_2G_k$.

In the presented article we derive simple formulas for calculating the values of cdf of $M_2G_k$. In a similar way analogous formulas for $M_mG_k$, $m > 2$, $k > m$, can be derived. These formulas are obtained from a system of first order difference equations described in section 2. Simple formulas for calculating the moments of higher order of $M_2G_k$ are also derived and presented in this section.

Matlab functions for calculating the values of cdf and moments about the origin of $M_2G_k$ via the presented formulas have been developed (they are presented in Appendices B, C, F, G). Results of experimental calculations carried out utilizing these functions, which proved the validity of the derived formulas are presented in section 3.

2. Simple formulas for defining the cdf of $M_2G_k$ and its moments about the origin. Let $NR(i)$, $i = 2, 3, ...$ designate the following vector of order $k + 1$:

\[ NR(i) = \begin{bmatrix} NR(1, i) \\ \vdots \\ NR(k + 1, i) \end{bmatrix}, \]
where \( NR(1, i) \) is the probability that no run of length \( k \) occurs in first \( i \) trials, \( i \)-th and \( (i-1) \)-th trials are failures, \( NR(2, i) \) is the probability that no run of length \( k \) occurs in first \( i \) trials, \( i \)-th trial is a failure, \( (i-1) \)-th trial is a success, \( NR(j, i) \), \( 3 \leq j \leq k + 1 \) is the probability that no run of length \( k \) occurs in first \( i \) trials and from those \( i \) trials the last \( j - 2 \) were successful, trial with index \( i - j + 2 \) was a failure provided that \( i - j + 2 > 0 \).

Obviously, the sum \( \sum NR(i) = \Sigma_{j=1}^{k+1} NR(j, i) \) is equal to the probability that no run of length \( k \) occurs during the first \( i \) trials, and the complement of this sum is the \( i \)-th value of \( cd f \) of \( M_2G_k \):

\[
cdf(i) = P(M_2G_k \leq i) = 1 - \sum NR(i).
\]

Luckily, the sequence \( NR(i), \ i = 2, 3, ... \) coincides with the solution of a system of first-order difference equations for the probabilities of non-occurrence of a success run of length \( k \) in first \( n \) trials. The following recursive equalities hold for any \( i \geq 2 \)

\[
\begin{align*}
NR(1, i + 1) &= NR(1, i) \cdot p_{000} + NR(2, i) \cdot p_{100} \\
NR(2, i + 1) &= NR(3, i) \cdot p_{010} + NR(4, i) \cdot p_{110} + \ldots + NR(k, i) \cdot p_{110} \\
NR(3, i + 1) &= NR(1, i) \cdot p_{001} + NR(2, i) \cdot p_{101} \\
NR(4, i + 1) &= NR(3, i) \cdot p_{011} \\
NR(5, i + 1) &= NR(4, i) \cdot p_{111} \\
\vdots
NR(k + 1, i + 1) &= NR(k, i) \cdot p_{111}.
\end{align*}
\]

In matrix form the system of equalities (3) can be written as

\[
NR(i + 1) = TRM \cdot NR(i), \ \ i \geq 2,
\]

where the transition probability matrix \( TRM \) has the form:

\[
TRM = \begin{bmatrix}
p_{000} & p_{100} & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & p_{010} & p_{110} & p_{110} & \cdots & p_{110} & p_{110} \\
p_{001} & p_{101} & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & p_{011} & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & p_{111} & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 0 & p_{111} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & p_{111} & 0 \\
\end{bmatrix}.
\]
Thus, we have the following formula for calculating $NR(i)$

$$NR(i) = TRM^{i-2} \cdot NR(2), \quad i \geq 2.$$ 

Obviously, the initial value $NR(2)$ has the form

$$NR(2) = \begin{bmatrix} p_{00} \\ p_{10} \\ p_{01} \\ p_{11} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$ 

Finally, we have obtained the following simple formula for calculating the values of cdf:

$$cdf(i) = 1 - \text{sum}(TRM^{i-2} \cdot NR(2)), \quad i > 2.$$ 

The values of the pmf of $M_2G_k$ can be calculated in the following way:

$$pmf(i) = P_{111} \cdot NR(i - 1), \quad i > 2, \quad pmf(1) \equiv 0, \quad pmf(2) \equiv 0.$$ 

Here $P_{111}$ designates the following vector of order $k+1$:

$$P_{111} = [0, 0, \ldots, 0, p_{111}]^\top.$$ 

In view of (4) we obtain the following formula

$$pmf(i) = P_{111} \cdot TRM^{i-3} \cdot NR(2), \quad i > 2, \quad pmf(1) \equiv 0, \quad pmf(2) \equiv 0.$$ 

Utilizing (5) enables one to derive simple formulas for calculating moments about the origin.

Let’s first derive the formula for the mean of $M_2G_k$.

$$M_1 = \sum_{i=3}^{\infty} i \cdot pmf(i) = P_{111} \cdot \left( \sum_{i=3}^{\infty} i \cdot TRM^i \right) \cdot TRM^{-3} \cdot NR(2).$$ 

We have the following well-known formula for calculating the sum of an infinite series $\sum_{i=3}^{\infty} i \cdot q^i$, where $q$ is some positive number less than 1:

$$\sum_{i=3}^{\infty} i \cdot q^i = q \cdot (1 - q)^{-2} - q - 2q^2.$$ 

Substituting $TRM$ instead of $q$ and the identity matrix of order $k+1$ instead of 1, we obtain the following formula

$$\sum_{i=3}^{\infty} i \cdot TRM^i = TRM \cdot (\text{eye}(k + 1) - TRM)^{-2} - TRM - 2 \cdot TRM^2.$$
Thus the formula for calculating $M_1$ via $TRM$ takes the form

$$M_1 = P_{111} \cdot (TRM \cdot (eye(k + 1) - TRM)^{-2} - TRM - 2 \cdot TRM^2 \cdot TRM^{-2} \cdot NR(2)).$$

Now let’s derive the formula for the second moment about the origin of $M_2 G_k$. We have $\sum_{i=3}^{\infty} i^2 \cdot q_i = q \cdot (1 - q)^{-3} + (1 - q)^{-3} - q - 4q^2$, from which follows the corresponding formula for the sum of matrix series and the formula for calculating $M_2$ :

$$M_2 = P_{111} \cdot (TRM \cdot (eye(k + 1) - TRM)^{-3} + (eye(k + 1) - TRM)^{-3} - TRM - 4 \cdot TRM^2) \cdot TRM^{-3} \cdot NR(2).$$

Analogously, simple formulas for calculating moments of $M_2 G_k$ of order greater than 2 can be derived.

3. Results of experimental calculations. Matlab functions for calculating specific values of $cdf$ and $pmf$ of $M_2 G_k$ based on the presented formulas designated $cdf_m2gk$ and $pmf_m2gk$ were developed and presented in Appendices B and D. Functions $pmf_m2gk_tab$ and $cdf_m2gk_tab$ that build tables of $pmf$ and $cdf$ distributions, respectively, and also plot the graphs of these functions, are presented in Appendices C and E. Functions $mean_m2gk$ for calculating the mean of $M_2 G_k$ and $smao_m2gk$ for calculating the second moment about the origin via the formulas described in section 2 are presented in Appendices F and G. In experimental calculations we assumed $k = 4$, $p_{00} = 0.1$, $p_{10} = 0.25$, $p_{01} = 0.3$, $p_{001} = 0.5$, $p_{011} = 0.55$, $p_{101} = 0.53$, $p_{111} = 0.75$ The results of experimental calculations are presented in tables 1 and 2.

<table>
<thead>
<tr>
<th>$MG_4$</th>
<th>$PMF_M2G_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>0.0523</td>
</tr>
<tr>
<td>9-16</td>
<td>0.0228</td>
</tr>
<tr>
<td>17-24</td>
<td>0.0099</td>
</tr>
<tr>
<td>25-32</td>
<td>0.0043</td>
</tr>
<tr>
<td>33-40</td>
<td>0.0013</td>
</tr>
<tr>
<td>41-48</td>
<td>0.0005</td>
</tr>
</tbody>
</table>
Table 2: Cdf distribution

<table>
<thead>
<tr>
<th>$MG_4$</th>
<th>$CDF_{M_2G_4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>0.0000 0.1968 0.2896 0.3461 0.4085 0.4686</td>
</tr>
<tr>
<td>9-16</td>
<td>0.5210 0.5678 0.6102 0.6486 0.6831 0.7142 0.7423 0.7676</td>
</tr>
<tr>
<td>17-24</td>
<td>0.7904 0.8110 0.8296 0.8463 0.8614 0.8750 0.8873 0.8984</td>
</tr>
<tr>
<td>25-32</td>
<td>0.9084 0.9174 0.9255 0.9328 0.9394 0.9453 0.9507 0.9555</td>
</tr>
<tr>
<td>33-40</td>
<td>0.9599 0.9638 0.9674 0.9706 0.9735 0.9761 0.9784 0.9805</td>
</tr>
<tr>
<td>41-48</td>
<td>0.9824 0.9842 0.9857 0.9871 0.9884 0.9895 0.9905 0.9915</td>
</tr>
</tbody>
</table>
The mean and second moment about the origin were calculated. All the values obtained in experimental calculations utilizing the described above functions were verified by comparing them to the calculated values obtained via alternative methods.

REFERENCES


A. TRM and NR(2) initialization function.

function [trm,nr2] = init_trm_nr2(k,p00,p10,p01,p001,p011,p101,p111)

% The function initializes the transition probability matrix and % the nr2 vector
p11=1-p00-p10-p01; p000=1-p001; p010=1-p011; p100=1-p101;
p110=1-p111; nr2(:,1)=[p00;p10;p01;p11;zeros(k-3,1)];
trm=zeros(k+1,k+1);
trm(1,1)=p000;
trm(1,2)=p100;
trm(2,3)=p010;
trm(2,4:end)=p110;
trm(3,1)=p001;
trm(3,2)=p101;
trm(4,3)=p011;
for i=5:k+1
    trm(i,i-1)=p111;
end;

B. Function for calculating a single cdf value.

function [cdf] = cdf_m2gk(k,p00,p10,p01,p001,p011,p101,p111,n)
% Output : cdf - probability that a run of length k occurs during first n % trials
[trm,nr2]=init_trm_nr2(k,p00,p10,p01,p001,p011,p101,p111);
C. Function for building cdf table and plotting its graph.
function [cdf_tab]=cdf_m2gk_tab(k,p00,p10,p01,p001,p011,p101,p111,eps)
    n=2; tmp=0; while (tmp<1-eps)
        tmp= cdf_m2gk(k,p00,p10,p01,p001,p011,p101,p111,n);
        cdf_tab(n)=tmp;
        n=n+1;
    end; cdf_tab=cdf_tab(1:end-1); plot(cdf_tab);

D. Function for calculating a single pmf value.
function p=pmf_m2gk(k,p00,p10,p01,p001,p011,p101,p111,n)
    [trm,nr2]=init_trm_nr2(k,p00,p10,p01,p001,p011,p101,p111);
    vec_p111=[zeros(1,k) p111]; if n>2 p=vec_p111*trm^(n-3)*nr2(:,1); else
        p=0;
    end;

E. Function for building pmf distribution table and plotting its graph.
function pmf_tab=pmf_m2gk_tab(k,p00,p10,p01,p001,p011,p101,p111,eps)
    n=k;
    tmp=1; while (tmp>eps)
        tmp= pmf_m2gk(k,p00,p10,p01,p001,p011,p101,p111,n);
        pmf_tab(n)=tmp;
        n=n+1;
    end; pmf_tab=pmf_tab(1:end-1); plot(pmf_tab);

F. Function for calculating the mean.
function mn = mean_m2gk(k,p00,p10,p01,p001,p011,p101,p111)
    [trm,nr2]=init_trm_nr2(k,p00,p10,p01,p001,p011,p101,p111);
    vec_p111=[zeros(1,k) p111];
    mn=vec_p111*(trm*(eye(k+1)-trm)^(-2)-trm-2*trm^2)*trm^(-3)*nr2(:,1);

G. Function for calculating second moment about the origin.
function ms = smao_m2gk(k,p1,p11,p01)
% Output: ms - the second moment about the origin of
% Markov-geometric random variable of order k
[trm,nr1]=init_trm_nr1(k,p1,p11,p01); vec_p11=[zeros(1,k-1) p11];
ms=vec_p11*([trm^2+trm]*(eye(k)-trm)^(-3)-trm)*trm^(-2)*nr1(:,1);