İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 6, No. 3, December 2013, pp. 110–115
ISSN 1300-4077 | 13 | 3 | 110 | 115

ON A GENERALIZATION OF LING'S BINOMIAL DISTRIBUTION

Femin YALCIN
Department of Mathematics, Faculty of Arts and Sciences
Izmir University of Economics, 35330, Balcova, Izmir, Turkey

Abstract: In a sequence of n binary trials, distribution of the random variable $M_{n,k}$, denoting the number of overlapping success runs of length exactly k, is called Ling's binomial distribution or Type II binomial distribution of order k. In this paper, we generalize Ling's binomial distribution to Ling's q-binomial distribution using Bernoulli trials with a geometrically varying success probability. An expression for the probability mass function of this distribution is derived. For $q = 1$, this distribution reduces to Ling's binomial distribution.

Key words: Binomial distribution of order k, Ling's binomial distribution, q-distributions, runs.
History: Submitted: 13 July 2013; Revised: 13 September 2013; Accepted: 1 October 2013

1. Introduction

Much attention has been paid to the distribution of the number of runs of fixed length, say k, in a sequence of n $(n \geq k)$ binary trials. Ling (1988) [4] introduced a binomial distribution of order k which is based on overlapping counting. This distribution is also called Type II binomial distribution of order k and differs from the Type I binomial distribution of order k, which was studied by Hirano (1986) [3] and Philippou and Makri (1986) [6]. The Type I binomial distribution of order k is based on nonoverlapping counting scheme. In a sequence of n binary trials, we use $N_{n,k}$ $(M_{n,k})$ to denote a random variable which has Type I (Type II) binomial distribution of order k. $N_{n,k}$ $(M_{n,k})$ is actually the number of nonoverlapping (overlapping) success runs of length exactly k in n trials. Consider a sequence of $n = 12$ trials 101110111110 assuming "1 " as a success and "0" as a failure. If $k = 3$, then $N_{12,3} = 2$ and $M_{12,3} = 4$.

Ling (1988) [4] obtained the following recursive and nonrecursive equations for the pmf of $M_{n,k}$ when the corresponding binary trials are independent and identically distributed with the probability of success p.

$$P\{M_{n,k} = x\} = \begin{cases} p^n & \text{if } x = n-k+1, \\ 2p^{n-1}q & \text{if } x = n-k\,(>0), \\ \sum\limits_{j=1}^{x+k} p^{j-1} q P\{M_{n-j,k} = x - \max(0, j-k)\} & \text{if } 0 \leq x < n-k \end{cases} \tag{1.1}$$

and

$$P\{M_{n,k} = x\} = \sum_{i=0}^{n} \sum_{\substack{x_1+2x_2+\cdots+nx_n+i=n \\ \max(0,i-k+1)+\sum\limits_{j=k+1}^{n}(j-k)x_j = x}} \binom{x_1+x_2+\cdots+x_n}{x_1, x_2, \ldots, x_n} p^n \left(\frac{q}{p}\right)^{\sum\limits_{i=1}^{n} x_i}.$$

Then, Godbole (1992) [2] derived a simpler formula

$$P\{M_{n,k}=x\}=\begin{cases} p^n & \text{if } x=n-k+1,\\ \sum\limits_{y=\lfloor \frac{n}{k}\rfloor}^{n} q^y p^{n-y} \sum\limits_{j=0}^{\lfloor \frac{n-y}{k}\rfloor} (-1)^j \binom{y+1}{j}\binom{n-jk}{y} & \text{if } x=0,\\ \sum\limits_{y} q^y p^{n-y} \sum\limits_{v} \binom{y}{v} & \\ \left\{\sum\limits_{j} (-1)^j \binom{y-v}{j}\binom{x-j(n-k)-1}{y-v-1} \sum\limits_{m} (-1)^m \binom{v+1}{m}\binom{n-x-k(y-v+m)}{v}\right. & \text{if } 1\le x\le n-k.\\ \left. \quad +\sum\limits_{j} (-1)^j \binom{y-v+1}{j}\binom{x-j(n-k)-1}{y-v} \sum\limits_{m} (-1)^m \binom{v}{m}\binom{n-x-k(y-v+1+m)}{v-1}\right\} & \end{cases} \tag{1.2}$$

An even simpler formula was obtained by Makri, Philippou, and Psillakis (2007) [5]. For $s=0$ and $l=k-1$, Theorem 2.1. of [5] gives

$$P\{M_{n,k}=x\}=\begin{cases} \sum\limits_{y=\lfloor \frac{n}{k}\rfloor}^{n} q^y p^{n-y} C(n-y,y+1,k-1) & \text{if } x=0,\\ \sum\limits_{y=\lfloor \frac{n+x(k-1)}{k}\rfloor -x}^{n-k-(x-1)} q^y p^{n-y} \sum\limits_{i=1}^{\lfloor \frac{n-y}{k}\rfloor} \binom{y+1}{i}\binom{x-1}{i-1} & \text{if } x\neq 0,\\ \times C(n-y-ik-(x-i);i,y+1-i;0,k-1) & \end{cases} \tag{1.3}$$

where

$$C(\alpha;i,r-i;m-1,n-1)=\sum_{j_1=0}^{\lfloor \frac{\alpha}{m}\rfloor}\sum_{j_2=0}^{\lfloor \frac{\alpha-mj_1}{n}\rfloor} (-1)^{j_1+j_2}\binom{i}{j_1}\binom{r-i}{j_2}\binom{\alpha-mj_1-nj_2+r-1}{r-1}$$

and

$$C(\alpha,r,m-1)=C(\alpha;i,r-i;m-1,m-1)=\sum_{j=0}^{\lfloor \frac{\alpha}{m}\rfloor}(-1)^j\binom{r}{j}\binom{\alpha-mj+r-1}{r-1}.$$

Charalambides (2010) [1] studied discrete q-distributions on Bernoulli trials with a geometrically varying success probability. Let us consider a sequence $X_1,\ldots,X_n$ of zero (failure)-one (success) Bernoulli trials such that the trials of the subsequences after the $(i-1)$st zero until the ith zero are independent with failure probability

$$q_i=1-\theta q^{i-1},\quad i=1,2,\ldots,\ 0<\theta<1, 0<q\le 1. \tag{1.4}$$

The probability mass function of the number Z_n of successes in n trials $X_1,\ldots,X_n$ is given by

$$P\{Z_n=r\}=\begin{bmatrix} n\\ r\end{bmatrix}_q \theta^r \prod_{i=1}^{n-r}(1-\theta q^{i-1}) \tag{1.5}$$

for $r=0,1,\ldots,n$, where

$$\begin{bmatrix} n\\ r\end{bmatrix}_q=\frac{[n]_{r,q}}{[r]_q!}$$

and $[x]_{k,q}=[x]_q[x-1]_q\cdots[x-k+1]_q$, $[x]_q=(1-q^x)/(1-q)$, $[x]_q!=[1]_q[2]_q\cdots[x]_q$ ([1]). The distribution given by 1.5 is called a q-binomial distribution.

Yalcin and Eryilmaz (2014) [7] obtained the distribution of $N_{n,k}$ for the model 1.4. The resulting distribution is the Type I q-binomial distribution of order k. In this paper, we study the distribution of $M_{n,k}$ under the model 1.4. The new distribution is called Type II q-binomial distribution of order k or Ling's q-binomial distribution.

Note that, throughout the paper, for integers n and m, and real number x, let $\binom{n}{m}$ and $\lfloor x \rfloor$ denote the binomial coefficients and the greatest integer less than or equal to x, respectively. We also assume for convenience that if $a > b$, then $\sum_{i=a}^{b} = 0$ and $\prod_{i=a}^{b} = 1$.

2. Type II q-binomial distribution of order k

We first note the following Lemma which will be useful in the sequel.

Lemma 1. *For $0 < q \leq 1$, define*

$$B_q(r,s,t) = \sum_{\substack{y_1+\cdots+y_r=s \\ I_1y_1+\cdots+I_ry_r-(k-1)(I_1+\cdots+I_r)=t \\ y_j\geq 0,\ j=1,2,\ldots,r}} \cdots \sum q^{y_2+2y_3+\cdots+(r-1)y_r},$$

where

$$I_j = \begin{cases} 1 & \text{if } y_j \geq k, \\ 0 & \text{otherwise} \end{cases}$$

and $y_j s$ are nonnegative integers, $j = 1,2,\ldots,r$. Then $B_q(r,s)$ obeys the following recurrence relation

$$B_q(r,s,t) = \begin{cases} \sum_{j=0}^{k-1} q^{(r-1)j} B_q(r-1,s-j,t) & \\ + \sum_{j=k}^{s} q^{(r-1)j} B_q(r-1,s-j,t-j+k-1) & \text{if } r>1,\ s\geq 0,\ \text{and } t\geq 0, \\ 1 & \text{if } (r=1,\ s\geq k \ \text{ and } t=s-k+1) \\ & \text{or } (r=1,\ 0\leq s<k \text{ and } t=0), \\ 0 & \text{otherwise.} \end{cases}$$

Proof. *Considering the values that y_r can take, we have*

$$\begin{aligned} B_q(r,s,t) = {} & \sum_{\substack{y_1+\cdots+y_{r-1}=s \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}} \\ & + q^{r-1} \sum_{\substack{y_1+\cdots+y_{r-1}=s-1 \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}} \\ & + q^{2(r-1)} \sum_{\substack{y_1+\cdots+y_{r-1}=s-2 \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}} \\ & + \cdots + q^{(k-1)(r-1)} \sum_{\substack{y_1+\cdots+y_{r-1}=s-k+1 \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}} \\ & + q^{k(r-1)} \sum_{\substack{y_1+\cdots+y_{r-1}=s-k \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t-1 \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}} \end{aligned}$$

$$+q^{(k+1)(r-1)} \sum_{\substack{y_1+\cdots+y_{r-1}=s-k-1 \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t-2 \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}}$$

$$+\cdots+q^{s(r-1)} \sum_{\substack{y_1+\cdots+y_{r-1}=0 \\ I_1y_1+\cdots+I_{r-1}y_{r-1}-(k-1)\left(I_1+\cdots+I_{r-1}\right)=t-s+k-1 \\ y_j\geq 0,\ j=1,2,\ldots,r-1}} \cdots \sum q^{y_2+2y_3+\cdots+(r-2)y_{r-1}}$$

$$\begin{aligned} &= B_q\left(r-1,s,t\right)+q^{r-1}B_q\left(r-1,s-1,t\right)+q^{2(r-1)}B_q\left(r-1,s-2,t\right) \\ &\quad +\cdots+q^{(k-1)(r-1)}B_q\left(r-1,s-k+1,t\right)+q^{k(r-1)}B_q\left(r-1,s-k,t-1\right) \\ &\quad +q^{(k+1)(r-1)}B_q\left(r-1,s-k-1,t-2\right)+\cdots+q^{s(r-1)}B_q\left(r-1,0,t-s+k-1\right) \end{aligned}$$

for $r>1$, $s\geq 0$, and $t\geq 0$. The other parts of the recurrence are obvious. □

THEOREM 1. *For $0<q\leq 1$, the probability mass function of the number of overlapping success runs of length k in n trials is given by*

$$P\left\{M_{n,k}=x\right\}=\sum_{i=0}^{n}\theta^{n-i}\prod_{j=1}^{i}\left(1-\theta q^{j-1}\right)B_q(i+1,n-i,x), \tag{2.1}$$

$x=0,1,\ldots,n-k+1.$

PROOF. *Let S_n denote the total number of zeros (failures) in n binary trials. Then*

$$P\left\{M_{n,k}=x\right\}=\sum_{i}P\left\{M_{n,k}=x,S_n=i\right\}.$$

The joint event $\{M_{n,k}=x,S_n=i\}$ can be described with the following binary sequence which consists of i zeros.

$$\underbrace{1\ldots1}_{y_1}0\underbrace{1\ldots1}_{y_2}0\ldots0\underbrace{1\ldots1}_{y_i}0\underbrace{1\ldots1}_{y_{i+1}},$$

where

$$\begin{aligned} & y_1+y_2+\cdots+y_{i+1}=n-i \\ & s.t. \\ & I_1\left(y_1-k+1\right)+I_2\left(y_2-k+1\right)+\cdots+I_{i+1}\left(y_{i+1}-k+1\right)=x \\ & y_j\geq 0 \text{ and } I_j=\begin{cases}1 \text{ if } y_j\geq k, \\ 0 \text{ otherwise,}\end{cases}\quad j=1,2,\ldots,i+1. \end{aligned}$$

Under the model 1.4,

$$\begin{aligned} P\left\{M_{n,k}=x\right\} &= \sum_{i} \sum_{\substack{y_1+\cdots+y_{i+1}=n-i \\ I_1y_1+\cdots+I_{i+1}y_{i+1}-(k-1)\left(I_1+\cdots+I_{i+1}\right)=x}}\cdots\sum \left(\theta q^0\right)^{y_1}\left(1-\theta q^0\right)\left(\theta q\right)^{y_2}\left(1-\theta q\right)\cdots\left(\theta q^{i-1}\right)^{y_i} \\ &\qquad \times\left(1-\theta q^{i-1}\right)\left(\theta q^i\right)^{y_{i+1}} \\ &= \sum_{i=0}^{n}\theta^{n-i}\prod_{j=1}^{i}\left(1-\theta q^{j-1}\right) \sum_{\substack{y_1+\cdots+y_{i+1}=n-i \\ I_1y_1+\cdots+I_{i+1}y_{i+1}-(k-1)\left(I_1+\cdots+I_{i+1}\right)=x \\ y_j\geq 0,\ j=1,2,\ldots,i+1}}\cdots\sum q^{y_2+2y_3+\cdots+(i-1)y_i+iy_{i+1}} \\ &= \sum_{i=0}^{n}\theta^{n-i}\prod_{j=1}^{i}\left(1-\theta q^{j-1}\right)B_q\left(i+1,n-i,x\right). \end{aligned}$$

Thus the proof is completed. □

REMARK 1. For $q=1$, the probability mass function of $M_{n,k}$ given in 2.1 is an alternative to 1.1, 1.2, and, 1.3 because the Type II 1-binomial distribution of order k is actually Ling's binomial distribution.

EXAMPLE 1. For $n=5$ and $k=2$, below we compute the pmf of $M_{n,k}$.

$$\begin{aligned}
P\{M_{5,2}=0\} &= \theta^3(1-\theta)(1-\theta q)q^3+\theta^2(1-\theta)(1-\theta q)\left(1-\theta q^2\right)\left(q+q^2+2q^3+q^4+q^5\right)\\
&\quad +\theta(1-\theta)(1-\theta q)\left(1-\theta q^2\right)\left(1-\theta q^3\right)\left(1+q+q^2+q^3+q^4\right)\\
&\quad +(1-\theta)(1-\theta q)\left(1-\theta q^2\right)\left(1-\theta q^3\right)\left(1-\theta q^4\right),\\
P\{M_{5,2}=1\} &= \theta^3(1-\theta)(1-\theta q)\left(q+2q^2+2q^4+q^5\right)\\
&\quad +\theta^2(1-\theta)(1-\theta q)\left(1-\theta q^2\right)\left(1+q^2+q^4+q^6\right),\\
P\{M_{5,2}=2\} &= \theta^4(1-\theta)\left(q+q^2+q^3\right)+\theta^3(1-\theta)(1-\theta q)\left(1+q^3+q^6\right),\\
P\{M_{5,2}=3\} &= \theta^4(1-\theta)\left(1+q^4\right),\\
P\{M_{5,2}=4\} &= \theta^5.
\end{aligned}$$

In Table 1 and Table 2, we respectively compute the probability mass function of $M_{10,2}$ for selected values of the parameters θ and q and the expected value of $M_{n,k}$ for different choices of k,n and the parameters θ and q. The numerical results indicate that $E(M_{n,k})$ is increasing in both θ and q.

x	$\theta=0.5, q=0.5$	$\theta=0.5, q=0.8$	$\theta=0.9, q=0.5$
0	0.68854	0.48212	0.14144
1	0.16216	0.24141	0.08809
2	0.07597	0.13410	0.07780
3	0.03699	0.07168	0.06976
4	0.01828	0.03682	0.06291
5	0.00913	0.01880	0.05812
6	0.00453	0.00833	0.05202
7	0.00244	0.00465	0.06236
8	0.00098	0.00111	0.03882
9	0.00098	0.00098	0.34868

Table 1. Probability mass function of $M_{10,2}$.

n	k	$\theta=0.5, q=0.5$	$\theta=0.5, q=0.8$	$\theta=0.9, q=0.5$
10	2	0.6048	1.0661	5.1925
	3	0.2721	0.4288	4.2369
20	2	0.6067	1.1416	7.3807
	3	0.2734	0.4485	6.3441
	5	0.0638	0.0866	4.8372

Table 2. Expected value of $M_{n,k}$.

Acknowledgements

The author would like to express her sincere thanks to the anonymous referee for careful reading of the first version of this paper and his/her useful comments and suggestions.

References

[1] Charalambides, C.A., 2010, The q-Bernstein Basis as a q-Binomial Distributions, Journal of Statistical Planning and Inference, 140, 2184-2190.

[2] Godbole, A.P., 1992 The Exact and Asymptotic Distribution of Overlapping Success Runs, Communications in Statistics-Theory and Methods, 21, 953-967.

[3] Hirano, K., 1986, Some Properties of the Distributions of Order k, Fibonacci Numbers and Their Applications (eds. G.E. Bergum, A.N. Philippou, and A.F. Horadam), 43-53.

[4] Ling, K.D., 1988, On Binomial Distribution of Order k, Statistics & Probability Letters, 6, 247-250.

[5] Makri, F. S., Philippo, A. N., and Psillakis, Z. M., 2007, Polya, Inverse Polya, and Circular Polya Distributions of Order k for l-Overlapping Success Runs, Communications in Statistics-Theory and Methods, 36, 657-668.

[6] Philippou, A.N. and Makri, F.S., 1986, Success Runs and Longest Runs, Statistics & Probability Letters, 4, 211-215.

[7] Yalcin, F. and Eryilmaz, S., 2014, q-Geometric and q-Binomial Distributions of Order k, Journal of Computational and Applied Mathematics, 271, 31-38.