STATISTICS IN TRANSITION new series, September 2022
Vol. 23, No. 3, pp. 27–41, DOI 10.2478/stattrans-2022-0028
Received – 28.01.2021; accepted – 25.03.2022

A comparison of the method of moments estimator and maximum likelihood estimator for the success probability in the Fibonacci-type probability distribution

Yeil Kwon[1]

ABSTRACT

A Fibonacci-type probability distribution provides the probabilistic models for establishing stopping rules associated with the number of consecutive successes. It can be interpreted as a generalized version of a geometric distribution. In this article, after revisiting the Fibonacci-type probability distribution to explore its definition, moments and properties, we proposed numerical methods to obtain two estimators of the success probability: the method of moments estimator (MME) and maximum likelihood estimator (MLE). The ways both of them performed were compared in terms of the mean squared error. A numerical study demonsrated that the MLE tends to outperform the MME for most of the parameter space with various sample sizes.

Key words: Fibonacci probability distribution, generalized polynacci distribution, factorial moment generating function, method of moments, maximum likelihood estimator.

1. Introduction

A geometric random variable is defined by the number of independent Bernoulli trials until the first success with a success probability p. As a generalized version of the geometric random variable, a negative binomial random variable is defined by the number of independent Bernoulli trials until r successes. The negative binomial random variable does not require the r successes to be consecutive. It seems natural to be interested in the case in which we stop the Bernoulli trials after reaching r *consecutive* successes. For example, what is the probability that we need 10 independent Bernoulli trials to have three consecutive successes (Moivre, 1756)?

A Fibonacci-type probability distribution describes the behavior of a random variable N defined by the number of independent Bernoulli trials until the k-th consecutive success with a success probability p. Shane (1973) derived a probability mass function and distribution function of N using polynacci polynomials, and Turner (1979) approached the same problem with the Pascal-T triangle. Philippou et al. (1982, 1983) developed a new formula for the probability function for N in terms of the multinomial coefficient and Fibonacci polynomials of order k. Philippou also made a significant contribution to deriving the convolutions of Fibonacci-type polynomials (Philippou et al., 1985; Philippou & Makri, 1985; Philippou & Georghiou, 1989) and the distribution of the multivariate Fibonacci-type polynomials of order k (Philippou & Antzoulakos, 1990, 1991).

[1]University of Central Arkansas, USA. E-mail: ykwon1@uca.edu. ORCID: https://orcid.org/0000-0002-1663-5401.

A Fibonacci-type probability distribution potentially can be applied to numerous areas such as quality control, engineering, and transportation. For example, we can find the direct applications of negative binomial distribution in quality control (Das, 2003; Ma & Zhang,1996). Using a Fibonacci-type probability distribution can be an alternative way to improve the quality control process. Suppose a production line supervisor wants to make a stopping rule to control a defective rate. The supervisor can set a rule to stop the production line for inspection when three consecutive defectives are observed. In general, "consecutive defectives" indicate another type of hidden risk of the production line that cannot be captured by a stopping rule based on the negative binomial distribution. Thus, the stopping rule based on the Fibonacci-type probability distribution is an attractive method for multi-dimensional quality control. We can find similar applications: the number of flight operations until three successive accidents, the number of digital signals in specific data transmission devices until five consecutive missing signals. Needless to say, when we have the observations from the Fibonacci-type probability distribution, the precise estimation of the success probability p is one of the most critical procedures in data analysis.

This paper aims to find the estimators for the success probability p and examine their performances when we have observations from the Fibonacci-type probability distribution. We revisit the important results on the Fibonacci-type probability distribution in Section 2 and find the estimators for the success probability p using the moments of N and the likelihood function in Section 3. Although the moments of N are represented in an explicit function of p, the method of moment estimate for p is obtained using a numerical method because it is the solution to the k-th degree polynomial in p. Furthermore, since the Fibonacci-type probability distribution is defined as a recursive form, it is difficult to find the maximum likelihood function in an explicit form. We propose a numerical method to approximate the likelihood function to find the maximum likelihood estimate for p. In Section 4, we provide the numerical results, illustrating the performances of the two estimators in terms of the mean squared error (MSE). The simulation study demonstrates that the maximum likelihood estimator (MLE) has a smaller MSE than the method of moment estimator (MME) for $p > 1/2$ under various sample sizes.

2. Fibonacci-type Probability Distribution

2.1. Fibonacci Probability Distribution where $k = 2$ and $p = 1/2$

Let N be the number of coin flips until we have the first consecutive heads with $p = \Pr(H) = 1/2$. Examining a few cases, $\Pr(N=2) = \Pr(HH) = 1/4$, $\Pr(N=3) = \Pr(THH) = 1/8$, $\Pr(N=4) = \Pr(HTHH) + \Pr(TTHH) = 2/16$, $\Pr(N=5) = \Pr(HTTHH) + \Pr(THTHH) + \Pr(TTTHH) = 3/32$. In general, we can represent $\Pr(N=n)$, for a positive integer $n \geq 4$, with the following structure:

$$\Pr(N=n) = \Pr(n-3 \text{ flips with no consecutive heads}) \Pr(THH) = \frac{C_{n-3}}{2^n},$$

where C_{n-3} stands for the number of arrangements of $n-3$ coin flip results with no consecutive heads. It is evident that $C_{n-3} = C_{n-4} + C_{n-5}$. Let E be the event where $n-3$ flips occur with no consecutive heads. The event E can be split into two cases:

① $n-3$ flips with no consecutive heads with the last flip being tail (T), or

② $n-3$ flips with no consecutive heads with the second last flip being tail (T)

Then, the number of arrangements for case ① is C_{n-4}, and the number of arrangements for case ② is C_{n-5}, which implies that C_n forms the Fibonacci sequence. Therefore, the probability mass function of the random variable N can be provided by the following:

$$\Pr(N=n) = f_{n-1}\left(\frac{1}{2}\right)^n, \quad n = 2,3,4,\ldots, \tag{1}$$

where f_n is the n-th Fibonacci number with $f_0 = 0, f_1 = 1, f_2 = 1, f_3 = 2, f_4 = 3, f_5 = 5, f_6 = 8, \ldots$. Shane (1973) named the probability mass function (1) as Fibonacci probability distribution because (1) contains Fibonacci numbers. Let $s = \sum_{n=2}^{\infty} \frac{f_{n-1}}{2^n}$, which is the sum of all the probabilities of (1). Then, $s = 1$ due to the following:

$$s = \frac{f_1}{2^2} + \sum_{n=3}^{\infty}\left(\frac{f_{n-2}}{2^n} + \frac{f_{n-3}}{2^n}\right) = \frac{1}{4} + \frac{1}{2}\sum_{n=2}^{\infty}\frac{f_{n-1}}{2^n} + \frac{1}{4}\left(\frac{f_0}{2} + \sum_{n=2}^{\infty}\frac{f_{n-1}}{2^n}\right) = \frac{1}{4} + \frac{1}{2}s + \frac{1}{4}s,$$

implying that $s = 1$. The next problem we are interested in is $E(N)$, the expected value of N. Proposition (1) plays an important role in computing $E(N)$.

Proposition 1 *Let $m(y)$ be an infinite series of y defined by $m(y) = \sum_{n=2}^{\infty} f_{n-1}\, y^n$. Then, for $|y| < 1/\varphi$,*

$$m(y) = \frac{y^2}{1-y-y^2},$$

where $\varphi = \lim_{n\to\infty} f_n/f_{n-1}$.

Proof:

$$\begin{aligned}
m(y) &= y^2 + \sum_{n=3}^{\infty} f_{n-1}\, y^n = y^2 + \sum_{n=2}^{\infty} f_n\, y^{n+1} \\
&= y^2 + y\left(\sum_{n=2}^{\infty}(f_{n-2}+f_{n-1})\, y^n\right) = y^2 + y\left(\sum_{n=1}^{\infty} f_{n-1}\, y^{n+1} + \sum_{n=2}^{\infty} f_{n-1}\, y^n\right) \\
&= y^2 + y\left(y\sum_{n=2}^{\infty} f_{n-1}\, y^n + \sum_{n=2}^{\infty} f_{n-1}\, y^n\right) = y^2 + y^2 m(y) + y\, m(y).
\end{aligned}$$

The radius of convergence of $m(y)$ is given by $|y| < 1/\varphi$ since $m(y)$ converges when

$$\lim_{n\to\infty}\left|\frac{f_n\, y^{n+1}}{f_{n-1}\, y^n}\right| = \lim_{n\to\infty}\left|\frac{f_n}{f_{n-1}}\right| |y| = \varphi\, |y| < 1.$$

□

For a probability mass function of N in (1), we can also show $\sum_{n=2}^{\infty}\Pr(N=n)=1$ by using Proposition (1) because $\sum_{n=2}^{\infty}\Pr(N=n)=\sum_{n=2}^{\infty}\frac{f_{n-1}}{2^n},=m(1/2)=1$. Next, we find a way to calculate the mean and variance of N. From Proposition 1, we have

$$m'(y)=\sum_{n=2}^{\infty} n\, f_{n-1}\, y^{n-1}=\frac{y(2-y)}{(1-y-y^2)^2},$$

and

$$m''(y)=\sum_{n=2}^{\infty} n(n-1)\, f_{n-1}\, y^{n-2}=\frac{2(1+3y^2-y^3)}{(1-y-y^2)^3}.$$

As $y m'(y)=\sum_{n=2}^{\infty} n\, f_{n-1}\, y^n$, the expected value of N can be obtained by

$$E(N)=\sum_{n=2}^{\infty} n\Pr(N=n)=\sum_{n=2}^{\infty} n\, f_{n-1}\left(\frac{1}{2}\right)^n = y m'(y)\Big|_{y=1/2}=6.$$

Furthermore, by using simple algebra, we can find

$$y^2 m''(y)=\sum_{n=2}^{\infty} n^2 f_{n-1}\, y^n-\sum_{n=2}^{\infty} n\, f_{n-1}\, y^n,$$

resulting in

$$\sum_{n=2}^{\infty} n^2 f_{n-1}\, y^n = y^2 m''(y)+y m'(y). \tag{2}$$

Therefore, with $y=1/2$ in (2),

$$E(N^2)=\sum_{n=2}^{\infty} n^2 f_{n-1}\left(\frac{1}{2}\right)^n=\left(\frac{1}{2}\right)^2 m''\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right) m'\left(\frac{1}{2}\right)=52+6=58,$$

and

$$V(N)=E(N^2)-E^2(N)=58-6^2=22.$$

The factorial moment generating function of a random variable X, with a probability mass (or density) function $f(x)$, is defined by

$$g(t)=E\left(t^X\right),$$

if this expectation exists for all values of $t\in(1-h,1+h)$. One of the well-known properties of the factorial moment generating function is that it satisfies

$$g^{(r)}(t)\Big|_{t=1}=E\left[X(X-1)\cdots(X-r+1)\right],$$

giving us factorial moments as we can infer from its name. For example, $g'(1)=E(X)$, $g''(1)=E\left[X(X-1)\right]$, and $g^{(3)}(1)=E\left[X(X-1)(X-2)\right]$. In particular, $V(X)=E(X^2)-E^2(X)=g''(1)+g'(1)-[g'(1)]^2$.

From Proposition (1), the factorial moment generating function of random variable N can be calculated using $m(y)$ because

$$g(t) = E\left(t^N\right) = \sum_{n=2}^{\infty} t^n \Pr(N=n) = \sum_{n=2}^{\infty} f_{n-1}\left(\frac{t}{2}\right)^n = m\left(\frac{t}{2}\right) = \frac{t^2}{4-2t-t^2}.$$

Moreover, since

$$g'(t) = \frac{2t(4-t)}{(4-2t-t^2)^2} \qquad \text{and} \qquad g''(t) = \frac{4(8+6t-t^3)}{(4-2t-t^2)^3},$$

we obtain $E(N) = g'(1) = 6$, $E(N(N-1)) = g''(1) = 52$, and $V(N) = E(N(N-1)) + E(N) - E^2(N) = 22$, the same value of variance for N obtained by using (2).

2.2. Fibonacci Distribution Function where $k = 2$ and $p \neq 1/2$

Next, we consider the case in which $p = p(H) \neq 1/2$. Thus, in this subsection, N is defined as the number of coin flips until we have the first consecutive heads with $p = \Pr(H) \neq 1/2$. Let $q = \Pr(T) = 1-p$, then $\Pr(N=2) = \Pr(HH) = p^2$ and $\Pr(N=3) = \Pr(THH) = qp^2$. For a positive integer $n \geq 4$, $\Pr(N=n)$ can be described as follows:

$$\begin{aligned}\Pr(N=n) &= \Pr(n-3 \text{ flips with no consecutive heads})\Pr(THH)\\ &= \Pr(n-3 \text{ flips with no consecutive heads})\,qp^2.\end{aligned}$$

Let $G_n(p,q)$ denote $\Pr(N=n)$ and examine the first several cases of $\Pr(N=n)$. Then

$$\begin{aligned}
G_2(p,q) &= \Pr(N=2) = p^2,\\
G_3(p,q) &= \Pr(N=3) = p^2q(1),\\
G_4(p,q) &= \Pr(N=4) = p^2q(q+p),\\
G_5(p,q) &= \Pr(N=5) = p^2q(q^2+2pq),\\
G_6(p,q) &= \Pr(N=6) = p^2q(q^3+3pq^2+p^2q),\\
G_7(p,q) &= \Pr(N=7) = p^2q(q^4+4pq^3+3p^2q^2).
\end{aligned}$$

Unlike (1), each $\Pr(N=n)$ does not include a Fibonacci number. Shane (1973) used the polynacci numbers and the polynacci polynomials to find the probability function of N for $p \neq 1/2$. However, we propose an alternative representation of $G_n(p,q)$ by using Fibonacci-type polynomials with a similar idea applied to (1). The probability mass function of N with the success probability p, is obtained as follows:

$$f(n) = \Pr(N=n) = G_n(p,q), \qquad n = 2, 3, 4, \ldots. \tag{3}$$

where $G_n(p,q)$ is a Fibonacci-type polynomial defined as

$$G_n(p,q) = \begin{cases} p/q, & \text{if } n = 0, \\ 0, & \text{if } n = 1, \\ qG_{n-1}(p,q) + pqG_{n-2}(p,q), & \text{if } n \geq 2. \end{cases} \tag{4}$$

$G_n(p,q)$ in (3) is a valid probability mass function since $\sum_{n=2}^{\infty} \Pr(N = n) = 1$:

$$\begin{aligned}
\sum_{n=2}^{\infty} G_n(p,q) &= q\sum_{n=2}^{\infty} G_{n-1}(p,q) + pq\sum_{n=2}^{\infty} G_{n-2}(p,q) \\
&= q\sum_{k=1}^{\infty} G_k(p,q) + pq\sum_{l=0}^{\infty} G_l(p,q) \\
&= q\sum_{k=2}^{\infty} G_k(p,q) + qG_1(p,q) + pq\sum_{l=2}^{\infty} G_l(p,q) + pqG_0(p,q) + pqG_1(p,q) \\
&= q\sum_{k=2}^{\infty} G_k(p,q) + pq\sum_{l=2}^{\infty} G_l(p,q) + p^2.
\end{aligned}$$

Therefore,

$$\sum_{n=2}^{\infty} G_n(p,q) = \frac{p^2}{1-q-pq} = 1.$$

Theorem 1 *Let N denote the number of coin flips until we have the first consecutive heads with $p = \Pr(H) \neq 1/2$ and $q = \Pr(T) = 1-p$. The factorial moment generating function of N, $g(t)$, is given by*

$$g(t) = E\left(t^N\right) = \frac{p^2t^2}{1-qt-pqt^2},$$

where $G_n(p,q)$ is a Fibonacci-type polynomials defined in (4).

Proof:

$$\begin{aligned}
g(t) &= \sum_{n=2}^{\infty} \Big(qG_{n-1}(p,q) + pqG_{n-2}(p,q)\Big)t^n \\
&= \sum_{n=2}^{\infty} qG_{n-1}(p,q)t^n + \sum_{n=2}^{\infty} (pq)G_{n-2}(p,q)t^n \\
&= qt\sum_{n=2}^{\infty} G_{n-1}(p,q)t^{n-1} + pqt^2\sum_{n=2}^{\infty} G_{n-2}(p,q)t^{n-2} = qt\sum_{k=1}^{\infty} G_k(p,q)t^k + pqt^2\sum_{l=0}^{\infty} G_l(p,q)t^l \\
&= qt\left(\sum_{k=2}^{\infty} G_k(p,q)t^k + G_1(p,q)t\right) + pqt^2\left(\sum_{l=2}^{\infty} G_l(p,q)t^l + G_0(p,q)t^0 + G_1(p,q)t\right) \\
&= qtg(t) + pqt^2g(t) + pqt^2G_0(p,q) \qquad \Big(\because G_1(p,q) = 0\Big) \\
&= qtg(t) + pqt^2g(t) + p^2t^2.
\end{aligned}$$

Therefore, by solving the equation for $g(t)$, we have $g(t) = p^2t^2/(1-qt-pqt^2)$. □

In particular, when $p=q=1/2$, we have $g(t)=t^2/(4-2t-t^2)$, $g'(t)=(8-2t^2)/(4-2t-t^2)^2$, and $g''(t)=4\left(8+6t^2-t^3\right)/(4-2t-t^2)^3$. Thus, $E(N)=g'(1)=6$, and $V(N)=E\left(N(N-1)\right)+E\left(N\right)-E^2\left(N\right)=g''(1)+g'(1)-\left(g'(1)\right)^2=22$, which are the same results as we obtained in Section 2.1. From Theorem 1, we have the expected value and the variance of N as

$$E(N)=g'(1)=\left.\frac{p^2t(2-qt)}{(1-qt-pqt^2)^2}\right|_{t=1}=\frac{1+p}{p^2}.$$

and

$$V\left(N\right)=g''(1)+g'(1)-\left(g'(1)\right)^2=\frac{(1-p)(1+3p+p^2)}{p^4},$$

respectively, because $E\left(N(N-1)\right)$ can be obtained by

$$E\left(N(N-1)\right)=g''(1)=\left.\frac{2p^2(1+3pqt^2-pq^2t^3)}{(1-qt-pqt^2)^3}\right|_{t=1}=\frac{2(1+2p-p^2-p^3)}{p^4}.$$

2.3. Fibonacci-Type Probability Distribution with Order k

In this section, we discuss the most generalized version of the Fibonacci-type probability distribution. Let N denote the number of Bernoulli trials until we have the first k consecutive successes with the success probability p. It can be structured as follows:

$$\begin{aligned}\Pr(N=n)&=\Big(1-\Pr\big((n-k-1)\text{ flips with }k\text{ consecutive successes}\big)\Big)\Pr(F\underbrace{S\cdots S}_{k\text{ successes}})\\&=\Big(1-\Pr\big((n-k-1)\text{ flips with }k\text{ consecutive successes}\big)\Big)qp^k.\end{aligned}$$

Cleary, $\Pr(N=k)=p^k$. If $k+1\le n\le 2k$,
$\Pr\big((n-k-1)\text{ flips with }k\text{ consecutive successes}\big)=0$, since $n-k-1<k$.
Hence, $\Pr(N=n)=p^kq$, for $k+1\le n\le 2k$. We examine several cases for $n\ge 2k+1$,

$$\begin{aligned}&H_{2k+1}(p,q)=\Pr(N=2k+1)=p^kq(1-p^k),\\&H_{2k+2}(p,q)=\Pr(N=2k+2)=p^kq(1-2p^kq-p^{k+1})=p^kq(1-p^k(1+q)),\\&H_{2k+3}(p,q)=\Pr(N=2k+3)=p^kq(1-3p^kq^2-4p^{k+1}q-p^{k+2})=p^kq(1-p^k(1+2q)),\\&H_{2k+4}(p,q)=\Pr(N=2k+4)=p^kq(1-4p^kq^3-9p^{k+1}q^2-6p^{k+2}q-p^{k+3})=\\&\qquad p^kq(1-p^k(1+3q)).\end{aligned}$$

Philippou et al. (1983) found that the probability mass function for N is given by

$$f(n)=\Pr(N=n)=p^nF_{n+1-k}\left(\frac{q}{p}\right),\quad n=k,k+1,k+2,\ldots,\tag{5}$$

where

$$F_n(y)=\begin{cases}n, & \text{if } n=0,1,\\ y\sum_{i=1}^{n}F_{n-i}(y), & \text{if } 2\le n\le k,\\ y\sum_{i=1}^{k}F_{n-i}(y), & \text{if } n\ge k+1,\end{cases}\tag{6}$$

and (5) and (6) can be re-represented in a simpler form (Philippou & Makri, 1985) as

$$\Pr(N=n)=H_n(p,q)=\begin{cases} p^k, & \text{if } n=k,\\ p^k q, & \text{if } k+1\le n\le 2k,\\ H_{n-1}-p^k q H_{n-1-k}, & \text{if } n\ge 2k+1. \end{cases} \tag{7}$$

It is easy to show $\sum_{n=k}^{\infty}\Pr(N=n)=1$ based on (7) because

$$\begin{aligned}
\sum_{n=k}^{\infty}H_n(p,q) &= p^k+\sum_{n=k+1}^{2k}p^k q+\sum_{n=2k+1}^{\infty}\Big(H_{n-1}(p,q)-p^k q H_{n-1-k}(p,q)\Big)\\
&= p^k+kp^k q+\sum_{m=2k}^{\infty}H_m(p,q)-p^k q\sum_{l=k}^{\infty}H_l(p,q)\\
&= p^k+kp^k q+\sum_{m=2k}^{\infty}H_m(p,q)+(k-1)p^k q+p^k\\
&\quad -(k-1)p^k q-p^k-p^k q\sum_{l=k}^{\infty}H_l(p,q)\\
&= p^k q+\sum_{m=k}^{\infty}H_m(p,q)-p^k q\sum_{l=k}^{\infty}H_l(p,q),
\end{aligned}$$

which implies $\sum_{n=k}^{\infty}H_n(p,q)=1$.

Theorem 2 *Let N denote the number of Bernoulli trials until we have the first k consecutive successes with $0<p=\Pr(success)<1$, and $q=\Pr(fail)=1-p$. Then, the factorial moment generating function of N, $h(t)$, is given by*

$$h(t)=E\left(t^N\right)=\frac{p^k t^k(1-pt)}{1-t+p^k q t^{k+1}}, \tag{8}$$

where $H_n(p,q)$ is a Fibonacci-type polynomials defined in (7).

Proof:

$$\begin{aligned}
h(t) &= t^k p^k+\sum_{n=k+1}^{2k}t^n p^k q+\sum_{n=2k+1}^{\infty}t^n\Big(H_{n-1}(p,q)-p^k q H_{n-1-k}(p,q)\Big)\\
&= t^k p^k+p^k q\left(\frac{t^{k+1}(1-t^k)}{1-t}\right)+\sum_{m=2k}^{\infty}t^{m+1}H_m(p,q)-p^k q\sum_{l=k}^{\infty}t^{l+k+1}H_l(p,q)\\
&= t^k p^k+p^k q\left(\frac{t^{k+1}(1-t^k)}{1-t}\right)+t\left(\sum_{m=2k}^{\infty}t^m H_m(p,q)+t^k p^k+p^k q\left(\frac{t^{k+1}(1-t^{k-1})}{1-t}\right)\right)\\
&\quad -t\left(t^k p^k+p^k q\left(\frac{t^{k+1}(1-t^{k-1})}{1-t}\right)\right)-p^k q t^{k+1}\sum_{l=k}^{\infty}t^l H_l(p,q)\\
&= t^k p^k(1-t)+p^k q t^{k+1}+h(t)t-h(t)p^k q t^{k+1},
\end{aligned}$$

indicating

$$h(t) = \frac{p^k t^k (1-t+qt)}{1-t+p^k q t^{k+1}} = \frac{p^k t^k (1-pt)}{1-t+p^k q t^{k+1}}.$$

The expected value of N can be computed by using $h(t)$ in (8). Moreover, since

$$h'(t) = \frac{p^k t^k \left(k(t-1)(p-1/t)+q-p^k q t^k\right)}{(1-t+p^k q t^{k+1})^2},$$

the expected value of N is given by

$$E(N) = h'(t)\Big|_{t=1} = \frac{1-p^k}{p^k q}. \tag{9}$$

□

Corollary 1 *Let $N^{(k)}$ be the number of Bernoulli trials until we have the first k consecutive successes with $0 < p = \Pr(success) < 1$, and $q = \Pr(fail) = 1-p$. Then,*

$$E\left(N^{(k+1)}\right) = \frac{1}{p} E\left(N^{(k)}+1\right).$$

Proof: From (9), we have

$$E\left(N^{(k+1)}\right) = \frac{1-p^{k+1}}{p^{k+1}q} = \frac{1}{p}\left(\frac{1-p^k}{p^k q} + \frac{p^k - p^{k+1}}{p^k q}\right) = \frac{1}{p} E\left(N^{(k)}+1\right),$$

since $p^k - p^{k+1} = p^k(1-p) = p^k q$.

□

Corollary 1 indicates that the expected value of $N^{(k)}$ increases exponentially when k increases with a growth factor of $1/p$. Hence, it escalates dramatically as p is close to 0, for example, when $p = 1/2$, $E\left(N^{(2)}\right) = 6$, $E\left(N^{(3)}\right) = 14$ and $E\left(N^{(4)}\right) = 30$. For $p = 1/10$, $E\left(N^{(2)}\right) = 110$, $E\left(N^{(3)}\right) = 1110$ and $E\left(N^{(4)}\right) = 11110$.

3. Estimation Methods for p

3.1. Method of Moments Estimator (MME) for p

Assume a random sample with size m is given as

$$N_1, N_2, \ldots, N_m \overset{iid}{\sim} f(n), \tag{10}$$

where $f(n)$ is a pmf defined in (7). The first sample moment (the sample mean) is given by $\bar{N} = \sum_{i=1}^{m} N_i/m$. By replacing $E(N)$ with $\bar{N}$ in (9), we have the following equation:

$$\bar{N}p^{k+1} - (\bar{N}+1)\,p^k + 1 = (p-1)(\bar{N}p^k - p^{k-1} - \cdots - p - 1) = 0.$$

Thus, a positive root in $(0,1]$ of the equation $\bar{N}p^k - p^{k-1} - \cdots - p - 1 = 0$ becomes the MME for p. Proposition 2 indicates that a unique $\hat{p}_{mme}$ exists, and, in particular, if $\bar{N} = k$, the MME for p becomes 1. We have $\bar{N} = k$ only when all N_i's $(i = 1,\ldots,m)$ in (10) are equal to k, and this event occurs with the probability p^{km}. For example, if $p = 1/2$, $k = 2$ and $m = 10$, $\Pr(\bar{N} = k) = 1/2048$.

Proposition 2 *Let $\bar{N} = \sum_{i=1}^{m} n_i/m$ be a sample mean obtained from (10). Then, for $\bar{N} > k$,*

$$r(p) = \bar{N}p^k - p^{k-1} - \cdots - p - 1 \tag{11}$$

has only one zero in $(0,1)$. When $\bar{N} = k$, the solution of the equation $r(p) = 0$ is given by $p = 1$.

Proof:

First, suppose $\bar{N} > k$. We know $r(0) = -1 < 0$ and $r(1) = \bar{N} - k > 0$. Because $r(p)$ is a continuous function on $[0,1]$, by the intermediate value theorem, it has at least one zero in $(0,1)$. Furthermore, from the Descartes' rule (Albert, 1943), as there is one sign change in the coefficients of $r(p)$, the equation $r(p) = 0$ has exactly one positive root. Therefore, $r(p)$ has only one zero in $(0,1)$. In particular, when $\bar{N} = k$, $r(p)$ can be factored as

$$r(p) = (p-1)(kp^{k-1} + (k-1)p^{k-2} + (k-2)p^{k-3} + \cdots + 3p^2 + 2p + 1).$$

In addition, it turns out $kp^{k-1} + (k-1)p^{k-2} + (k-2)p^{k-3} + \cdots + 3p^2 + 2p + 1 = 0$ has no positive root by using the Descartes' rule again. Hence, the only root of $r(p) = 0$ on $[0,1]$ is $p = 1$.

□

For instance, when $k = 2$, the MME for p is the solution to the quadratic equation $\bar{N}p^2 - p - 1 = 0$, and it turns out to be

$$\hat{p}_{mme} = \frac{1 + \sqrt{1 + 4\bar{N}}}{2\bar{N}}.$$

When $p \geq 3$, $\hat{p}_{mme}$ can be obtained by finding the root of (11) with the numerical methods such as Newton's method and Halley's method.

3.2. Maximum Likelihood Estimator (MLE) for p

Under the same assumption of (10), the log-likelihood function $l(p)$ is

$$l(p) = \sum_{i=1}^{m} \ln\left[H_{n_i}(p,q)\right],$$

and, the maximum likelihood estimator of p is given by

$$\hat{p}_{mle} = \underset{0<p<1}{\arg\max} \sum_{i=1}^{m} \ln \left[H_{n_i}(p,q) \right].$$

However, because $H_{n_i}(p,q)$ is not provided in a closed form but in a recursive form, analytical derivation for the MLE of p is extremely challenging. Hence, a numerical method is proposed as follows:

- (Step 1) Discretize the values of $p \in (0,1)$, for example, $p_d = 0.01, 0.02, \ldots, 0.98, 0.99$.
- (Step 2) Given $n_1, \ldots, n_m$, and k, calculate $f(n_i) = H_{n_i}(p_d, q_d)$ based on (7) for all $i = 1, \ldots, m$ and all the discretized values of p_d.
- (Step 3) Approximate the log-likelihood function $l(p)$ by computing

$$l(p_d) = \sum_{i=1}^{m} \ln \left[H_{n_i}(p_d, q_d) \right].$$

- (Step 4) Find the optimal value of p_d^*, which maximizes $l(p_d)$.

4. Numerical Study

In this section, we compare the performance of the MME ($\hat{p}_{mme}$) and the computationally driven MLE ($\hat{p}_{mle}$) for the success probability p in terms of the MSE. In general, the MSE of an estimator $\hat{\theta}$ for a parameter θ is defined by $E(\hat{\theta} - \theta)^2$, and it can be decomposed as the sum of the variance of $\hat{\theta}$ and the squared bias of $\hat{\theta}$. In this simulation study, the MSE is estimated and decomposed by

$$\widehat{MSE}\left(\hat{\theta},\, \theta\right) = \frac{1}{R} \sum_{r=1}^{R} \left(\hat{\theta}_r - \theta \right)^2 = \widehat{Bias}^2(\hat{\theta}) + \widehat{Var}(\hat{\theta}),$$

where

$$\widehat{Bias}^2(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^{R} \left(\bar{\hat{\theta}} - \theta \right)^2, \quad \widehat{Var}(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^{R} \left(\hat{\theta}_r - \bar{\hat{\theta}} \right)^2, \quad \text{and} \quad \bar{\hat{\theta}} = \frac{1}{R} \sum_{r=1}^{R} \hat{\theta}_r,$$

with R the total number of simulations, and $\hat{\theta}_r$ the estimate for θ in the r-th repetition. Here, $\hat{\theta}$ represents both $\hat{p}_{mme}$ and $\hat{p}_{mle}$. We set the true success probability $p = 0.1, 0.3, 0.5, 0.7, 0.9$, the sample size $m = 5, 10, 20, 30, 50$, and the number of simulation $R = 500$.

Table 1 displays the results of the simulation when $k = 2$. Table 2 is in the same format as Table 1 but presents the results when $k = 4$. In other words, Table 1 illustrates the results of the numerical study with a random variable N defined by the number of Bernoulli trials until we have two consecutive successes. The estimated squared bias, variance, and MSE for each estimator are reported with units in 10^{-3}. Ratio columns display the ratios of the estimated squared bias, variance, and MSE for $\hat{p}_{mme}$ and $\hat{p}_{mle}$. Hence, the values of the ratio that are greater than 1 imply that $\hat{p}_{mle}$ outperforms $\hat{p}_{mme}$. For example, in Table 1, for $m = 10$ and $p = 0.7$, the value in the ratio column of $\widehat{Bias}^2$ is computed as

Table 1. Squared bias, variance, and mean squared error of $\hat{p}_{mme}$ and $\hat{p}_{mle}$ for $k=2$ (unit: 10^{-3}). The ratio columns represent the values of the squared bias (or variance) of $\hat{p}_{mme}$ divided by the squared bias (or variance) of $\hat{p}_{mle}$.

$k=2$		$\widehat{Bias^2}$			$\widehat{Var}$			$\widehat{MSE}$		
Sample Size	p	$\hat{p}_{mme}$	$\hat{p}_{mle}$	Ratio	$\hat{p}_{mme}$	$\hat{p}_{mle}$	Ratio	$\hat{p}_{mme}$	$\hat{p}_{mle}$	Ratio
$m=5$	0.1	0.0817	0.0819	0.997	0.800	0.804	0.994	0.884	0.886	0.997
	0.3	0.4885	0.4767	1.025	6.494	6.396	1.015	6.982	6.873	1.016
	0.5	1.9970	1.8523	1.078	14.11	13.84	1.019	16.11	15.70	1.026
	0.7	0.7481	0.5482	1.365	13.79	13.21	1.043	14.54	13.76	1.056
	0.9	0.0844	0.0514	1.642	6.963	6.676	1.043	7.047	6.728	1.047
$m=10$	0.1	0.0455	0.0457	0.997	0.366	0.366	1.000	0.412	0.412	1.000
	0.3	0.1191	0.1133	1.051	2.752	2.733	1.007	2.871	2.846	1.009
	0.5	0.2651	0.2239	1.184	6.578	6.464	1.018	6.843	6.688	1.023
	0.7	0.0467	0.0285	1.637	7.321	7.092	1.032	7.368	7.120	1.035
	0.9	0.0015	0.0001	11.28	3.900	3.783	1.031	3.902	3.783	1.031
$m=20$	0.1	0.0079	0.0079	1.012	0.141	0.141	0.999	0.149	0.149	1.000
	0.3	0.0211	0.0206	1.025	1.267	1.264	1.002	1.288	1.285	1.003
	0.5	0.0603	0.0561	1.075	2.703	2.684	1.007	2.763	2.741	1.008
	0.7	0.0678	0.0464	1.461	3.600	3.429	1.050	3.668	3.475	1.055
	0.9	0.0112	0.0086	1.294	2.134	2.036	1.048	2.145	2.045	1.049
$m=30$	0.1	0.0034	0.0034	1.001	0.099	0.098	1.003	0.102	0.102	1.003
	0.3	0.0034	0.0031	1.092	0.751	0.752	0.999	0.754	0.755	1.000
	0.5	0.0430	0.0419	1.027	1.832	1.826	1.003	1.875	1.868	1.004
	0.7	0.0294	0.0229	1.285	2.343	2.261	1.037	2.373	2.284	1.039
	0.9	0.0103	0.0046	2.222	1.321	1.230	1.074	1.332	1.235	1.079
$m=50$	0.1	0.0006	0.0006	1.003	0.060	0.059	1.002	0.060	0.060	1.002
	0.3	0.0000	0.0000	2.376	0.530	0.531	0.998	0.531	0.531	1.000
	0.5	0.0072	0.0074	0.972	1.167	1.165	1.002	1.175	1.172	1.002
	0.7	0.0055	0.0029	1.926	1.456	1.411	1.032	1.461	1.414	1.034
	0.9	0.0001	0.0013	0.109	0.803	0.752	1.068	0.803	0.753	1.066

$\left(0.0467\times10^{-3}\right)/\left(0.0285\times10^{-3}\right)=1.637$. This indicates the squared bias of $\hat{p}_{mme}$ is 63.7% greater than that of $\hat{p}_{mle}$ on average. We can interpret the numbers in the ratio columns of $\widehat{Var}$ and $\widehat{MSE}$ in the same manner. As for the decomposition of the MSE, Tables 1 and 2 show the variance (compared with the squared bias) explains a major portion of the MSE for both of the estimators. Except for the case with a small sample size m, and a small success probability p, more than 95% of the MSE is explained by the variance approximately. For the magnitude of the bias, the squared bias of $\hat{p}_{mle}$ is smaller than that of $\hat{p}_{mme}$ for most cases. In the variance comparison, although the values in the ratio column are not as large as the ratio values of the squared bias, the variance of $\hat{p}_{mle}$ is smaller than that of $\hat{p}_{mme}$ for most of the values of p and sample sizes. The MSE ratio of $\hat{p}_{mme}$ and $\hat{p}_{mle}$ exhibits a pattern similar to the variance ratio due to the substantial contribution of the variance to the MSE. When p is small, the MSE difference between $\hat{p}_{mme}$ and $\hat{p}_{mle}$ is not significantly large. However, for moderate and large values of p, the MSE of $\hat{p}_{mle}$ is smaller than that of $\hat{p}_{mme}$ for all sample sizes. The improvement caused by $\hat{p}_{mle}$ tends to be larger when p is closer to 1.

Table 2. Squared bias, variance, and mean squared error of $\hat{p}_{mme}$ and $\hat{p}_{mle}$ for $k = 4$ (unit: 10^{-3}). The ratio columns represent the values of the squared bias (or variance) of $\hat{p}_{mme}$ divided by the squared bias (or variance) of $\hat{p}_{mle}$.

$k = 4$		$\widehat{Bias^2}$			$\widehat{Var}$			$\widehat{MSE}$		
Sample Size	p	$\hat{p}_{mme}$	$\hat{p}_{mle}$	Ratio	$\hat{p}_{mme}$	$\hat{p}_{mle}$	Ratio	$\hat{p}_{mme}$	$\hat{p}_{mle}$	Ratio
$m = 5$	0.1	0.0166	0.0166	1.002	0.166	0.166	0.997	0.182	0.183	0.998
	0.3	0.1365	0.1347	1.013	1.717	1.709	1.005	1.854	1.844	1.006
	0.5	0.3298	0.3148	1.048	5.189	5.085	1.020	5.519	5.400	1.022
	0.7	0.3761	0.3158	1.191	6.316	6.066	1.041	6.692	6.382	1.049
	0.9	0.2024	0.1163	1.741	4.015	3.836	1.047	4.218	3.953	1.067
$m = 10$	0.1	0.0010	0.0010	0.993	0.075	0.075	1.005	0.076	0.076	1.005
	0.3	0.0455	0.0454	1.004	0.698	0.697	1.001	0.743	0.743	1.001
	0.5	0.0818	0.0791	1.034	2.133	2.119	1.007	2.215	2.198	1.008
	0.7	0.0340	0.0194	1.754	2.852	2.745	1.039	2.886	2.765	1.044
	0.9	0.0269	0.0117	2.311	2.022	1.908	1.060	2.049	1.920	1.067
$m = 20$	0.1	0.0011	0.0011	0.967	0.036	0.036	0.996	0.037	0.037	0.995
	0.3	0.0091	0.0092	0.995	0.356	0.356	1.000	0.365	0.365	1.000
	0.5	0.0136	0.0132	1.027	1.036	1.037	0.999	1.049	1.049	1.000
	0.7	0.0067	0.0051	1.313	1.609	1.582	1.017	1.616	1.587	1.018
	0.9	0.0014	0.0008	1.733	1.068	0.963	1.108	1.069	0.964	1.109
$m = 30$	0.1	0.0004	0.0004	0.994	0.022	0.023	0.991	0.023	0.023	0.991
	0.3	0.0014	0.0014	0.983	0.249	0.250	0.999	0.251	0.250	1.000
	0.5	0.0017	0.0017	0.996	0.622	0.622	1.001	0.624	0.623	1.001
	0.7	0.0145	0.0123	1.180	0.926	0.896	1.034	0.941	0.908	1.036
	0.9	0.0033	0.0016	2.033	0.638	0.577	1.107	0.642	0.578	1.110
$m = 50$	0.1	0.0001	0.0002	0.868	0.015	0.015	0.994	0.015	0.015	0.993
	0.3	0.0002	0.0002	1.013	0.140	0.140	1.000	0.140	0.140	1.000
	0.5	0.0013	0.0012	1.042	0.367	0.363	1.010	0.368	0.364	1.010
	0.7	0.0061	0.0059	1.035	0.594	0.592	1.004	0.600	0.598	1.003
	0.9	0.0013	0.0012	1.119	0.414	0.386	1.072	0.416	0.388	1.072

5. Conclusion

A Fibonacci-type probability distribution can be employed to determine the probabilistic behavior of a random variable N defined by the number of Bernoulli trials with a success probability p until we have k-consecutive successes. When $p = 1/2$, it can be expressed as an implicit form with the Fibonacci numbers. When $p \neq 1/2$, the Fibonacci-type probability distribution is represented in terms of Fibonacci-type polynomials recursively. We calculated the first and second moments of N by using the factorial moment generating function. In particular, the expected value of N increases exponentially with a growth factor of $1/p$ when the number of consecutive successes k increases by 1, while the expected value of a negative binomial random variable increases linearly for the unit increase of the number of successes. To compare MME with MLE, we used the computational methods to obtain the MLE by approximating the maximum likelihood function using the pmf of N defined recursively. The result of the simulation discloses that, for both MLE and MME, the biases are considerably smaller than the variances under all of the values of p and the sample sizes, indicating that the variance explains the majority of the MSE. Furthermore, we can see, in terms of the MSE, the MLE performs better than MME for a wide range of p, especially when p is greater than 1/2 for the various sample sizes.

References

Albert, G. E., (1943). An Inductive Proof of Descartes' Rule of Signs, *The American Mathematical Monthly*, 50(3), pp. 178–180.

Das, N., (2003). Study on Implementing Control Charts Assuming Negative Binomial Distribution with Varying Sample Size in a Software Industry. *Software Quality Professional*, 6, pp. 38–39.

Ma, Y., Zhang, Y., (1996). Q control charts for negative binomial distribution. *Computers & Industrial Engineering*, 31(3), pp. 813–816.

De Moivre, A. (1756). *The Doctrine of Chances*, 3rd ed., Chelsea Publishing Co., New York (reprint, 1967).

Philippou, A. N., Antzoulakos, D., (1990). Multivariate Fibonacci polynomials of order k and the multiparameter negative binomial distribution of the same order, G. E. Bergum (ed.) A. N. Philippou (ed.) A. F. Horadam (ed.), *Applications of Fibonacci Numbers* 3, Kluwer Academic Publishers, pp. 273–279.

Philippou, A. N., Antzoulakos, D., (1991).Generalized Multivariate Fibonacci Polynomials of Order k and the Multivariate Negative Binomial Distributions of the Same Order. *The Fibonacci Quarterly*, 29, 322–328.

Philippou, A. N., Georghiou, C., (1989). Convolutions of Fibonacci-type polynomials of order k and the negative binomial distributions of the same order. *The Fibonacci Quarterly*, 27(3), pp. 209–216.

Philippou, A. N., Georghiou, C., Philippou, G. N., (1983). Fibonacci polynomials of order k, multinomial expansions and probability. *International Journal of Mathematics and Mathematical Sciences*, 6(3), pp. 545–550.

Philippou, A. N., Georghiou, C., Philippou, G. N., (1985). Fibonacci-type polynomials of order k with probability applications. *The Fibonacci Quarterly*, 23(2), pp. 100–105.

Philippou, A. N., Makri, D. S., (1985). Longest Success Runs and Fibonacci-type Polynomials of Order k. *The Fibonacci Quarterly*, 23(4), pp. 338–346.

Philippou, A. N., Muwafi, A. A., (1982). Waiting for the k-th consecutive success and the Fibonacci sequence of order k. *The Fibonacci Quarterly*, 20(1), pp. 28–32.

Shane, H. D., (1973). A Fibonacci Probability Function. *The Fibonacci Quarterly*, 11(6), pp. 517–522.

Turner, S. J., (1979). Probability via the N-th Order Fibonacci-T Sequence. *The Fibonacci Quarterly*, 17(1), pp. 23–28.