On waiting time distribution of runs of ones or zeroes in a Bernoulli sequence

Sungsu Kim  
(Worcester Polytechnic Institute)  
Chongjin Park  
(Department of Mathematics and Statistics, San Diego State University)  
Jungtaek Oh*  
(Department of Mathematics, Kyungpook National University)

June 16, 2014
Introduction

Definition of Runs

There are various definitions of runs
"We shall not give a general definition here, because new advances and applications of new criteria to new problems will probably soon render most definitions obsolete."
Wolfowitz(1943)

\[ W(k) \]

Denote the waiting time, the number of trials needed, to get either a consecutive k ones or k zeroes for the first time

Generalized Fibonacci sequence of order k

\[ F_k = \{ f_{i,k} | f_i = f_{i-1} + f_{i-2} + \cdots + f_{i-k}, f_1 = 1 \text{ and } f_i = 0 \text{ for } i \leq 0 \} \]
Introduction

Examples

\( F_1 \)

\[
F_1 = \{ f_i \mid f_i = f_{i-1} \} = \{ f_i \mid 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ldots \}
\]

Fibonacci numbers

\( F_2 \)

\[
F_2 = \{ f_{i_2} \mid f_i = f_{i-1} + f_{i-2} \} = \{ f_{i_2} \mid 1 \ 1 \ 2 \ 3 \ 5 \ 8 \ldots \}
\]

Tribonacci numbers

\( F_3 \)

\[
F_3 = \{ f_{i_3} \mid f_i = f_{i-1} + f_{i-2} + f_{i-3} \} = \{ f_{i_3} \mid 1 \ 1 \ 2 \ 4 \ 7 \ 13 \ldots \}
\]

Tetrunacci numbers

\( F_4 \)

\[
F_4 = \{ f_{i_4} \mid f_i = f_{i-1} + f_{i-2} + f_{i-3} + f_{i-4} \} = \{ f_{i_4} \mid 1 \ 1 \ 2 \ 4 \ 8 \ 15 \ldots \}
\]
Probability distribution function and probability generating function of $W(k)$

**Lemma 1.**

The number of cases, i.e. sequences, where a sequence of two ones occurred for the first time on the $w$th trial and no sequence of two zeroes occurred prior to the $w$th trial follows the Fibonacci sequence of order 1.

- A proof is easily established by noting that all the number of such cases is equal to 1.

**Corollary 2.**

The number of cases, i.e. sequences, where a sequence of two zeroes occurred for the first time on the $w$th trial and no sequence of two ones occurred prior to the $w$th trial follows the Fibonacci sequence of order 1.
Probability distribution function and probability generating function of $W(k)$

Lemma 3.

Let $h_{w-1}$ denote the probability of the first $w-2$ trials in a sequence of length $w$ ending with two ones without any two prior zeroes, and $h^*_{w-1}$ denote the probability of the first $w-2$ trials in a sequence of length $w$ ending with two zeroes without any two prior ones. Then, we have $h_w = qh^*_{w-1}$ and $h^*_w = ph_{w-1}$, for $w=2,3,\ldots$, with $h_1 = h^*_1 = 1$. 
Probability distribution function and probability generating function of $W(k)$

Proof.

Considering a sequence of length $w + 1$ ending with two ones without any prior sequence of two zeroes, the first $w - 2$ trials are equal to the first $w - 2$ trials of the sequence of length $w$ ending with two zeroes without any prior two ones. Now, since the $w - 1$th trial of the sequence of length $w + 1$ must be zero, we can write $h_w$ in terms of $h^*_{w-1}$ as $qh^*_{w-1}$. A proof can be made for $h^*_w = ph_{w-1}$ using a similar argument.

The pdf of $W(2)$ is given in the following theorem.
Probability distribution function and probability generating function of $W(k)$

**Theorem 4.**

For $k=2$, the pdf of $W(k)$ is

$$P(W(2) = w) = h_{w-1}p^2 + h_{w-1}^*q^2,$$

where $h_w = qh_{w-1}^*$ and $h_w^* = ph_{w-1}$, for $w=2,3,\ldots$, with $h_1 = 1$ and $h_1^* = 1$, and $h_i^*$ denotes the "complement" of $h_i$ for $i \geq 2$. 
Proof.
First, note that
\[ P(W(2) = 2) = pp + qq = p^2 + q^2 = h_1 p^2 + h_1^* q^2, \]
and
\[ P(W(2) = 3) = qpp + pqq = qp^2 + pq^2 = h_2 p^2 + h_2^* q^2. \]
Then, the theorem can be established using the above lemma. \(\square\)
Probability distribution function and probability generating function of $W(k)$

An alternative derivation of pdf of $W(2)$ is as follows.

\[ P(W(2) = 2) = pp + qq, \quad P(W(2) = 3) = pq \]
\[ P(W(2) = 4) = pq(pp + qq), \quad P(W(2) = 5) = (pq)^2 \]
\[ \vdots \]
\[ P(W(2) = w) = (pq)^{\frac{w+1}{2}-1} \text{ if } w(>2) \text{ is an odd number, and} \]
\[ P(W(2) = w) = (pp + qq)(pq)^{\frac{w-2}{2}} \text{ if } w(>1) \text{ is an even number.} \]

We present a closed form solution to probability generating function (pgf) of $W(2)$ in the following proposition.
Probability distribution function and probability generating function of $W(k)$

**Proposition 5.**

The pgf of $W(2)$ is

$$G_{W(2)}(z) = E(z^{W(2)}) = \frac{(p^2 + q^2)z^2 + pqz^3}{1 - pqz^2}.$$ 

**Proof.**

Note that the pgf of $W(2)$ is given by the infinite sum shown below.

$$G_{W(2)}(z) = E(z^{W(2)}) = \sum_{k=1}^{\infty} z^k P(W(2) = k)$$

$$= z^2(p^2 + q^2) + z^3 pq + z^4 pq(p^2 + q^2) + z^5(pq)^2 + \cdots \quad (1)$$

Then, applying the geometric sum formula to (1), we prove the theorem.
Probability distribution function and probability generating function of $W(k)$

The expected value of $W(2)$ is given by the following corollary.

**Corollary 6.**

Using Proposition 5, we obtain

$$E(W(2)) = \frac{2 + pq}{1 - pq}.$$ 

It can be proved by taking the first derivative of $G_{W(2)}(z)$ with respect to $z$, then setting $z = 1$.

The following propositions are stated and shown to be useful in proving the theorem for the pdf of $W(3)$. 
Probability distribution function and probability generating function of $W(k)$

**Lemma 7.**
Let $H(n)$ denote the number of cases where a sequence of three ones occurred for the first time on the $n$th trial and no sequence of three zeroes occurred prior to the $n$th trial. Then $H(n+1) = H(n) + H(n-1)$, with $n=4,5,\ldots$. That is, $H(n)$ follows the Fibonacci sequence of order 2.
Probability distribution function and probability generating function of $W(k)$

Proof.
$H(n)$ consists of sequences end like 111 on the $n$th trial, for $n=3,4,\ldots$ For example, $H(3)$ consists of 111, $H(4)$, 0111, $H(5)$, 00111 and 10111, and $H(6)$ 100111, 110111, 010111 etc.. Notice that the first three positions of the $H(6)$ are 100, 110, 010, all end with zero, and 100 is the complement of the first three digits of $H(4)$ and 110 and 010 are the complements of the first three digits of $H(5)$. Using the same argument, the first four positions of $H(7)$ can be constructed by the complements of the first four positions of $H(5)$, 1100, 0100 and the complements of first four positions of $H(6)$, 0110, 0010, 1010. Hence $H(7)$ consists of five possible sequences. A proof of Proposition 7 follows by inductively using the similar steps for $H(n+1)$.  \[\square\]
Probability distribution function and probability generating function of \( W(k) \)

corollary 8.

Let \( H^*_n \) denote the number of cases where a sequence of three zeroes occurred for the first time on the \( n \)th trial and no sequence of three ones occurred prior to the \( n \)th trial. Then \( H^*_{(n+1)} = H^*_n + H^*_n(n-1) \), with \( n=4,5,\ldots \). That is, \( H^*_n \) follows the Fibonacci sequence of order 2.

A proof follows similarly as done for lemma 7.
lemma 9.

Let $h_{w-2}$ denote the probability of the first $w$-3 trials in a sequence of length $w$ ending with three ones without any three prior zeroes, and $h^*_{w-2}$ denote the probability of the first $w$-3 trials in a sequence of length $w$ ending with three zeroes without any three prior ones. Then, $h_w = q h^*_{w-1} + q^2 h^*_{w-2}$ and $h^*_w = p h_{w-1} + p^2 h_{w-2}$, for $w=3,4,\ldots$, with $h_1 = 1$, $h_2 = q$ and $h^*_1 = 1$, $h^*_2 = p$. 
<table>
<thead>
<tr>
<th>$H(3)$</th>
<th>111</th>
<th>000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(4)$</td>
<td>0111</td>
<td>1000</td>
</tr>
<tr>
<td>$H(5)$</td>
<td>00111, 1011</td>
<td>11000, 01000</td>
</tr>
<tr>
<td>$H(6)$</td>
<td>100111, 11011, 01011</td>
<td>011000, 001000, 101000</td>
</tr>
</tbody>
</table>
Theorem 10.

For \( k = 3 \), the pdf of \( W(k) \) is

\[
P(W(3) = w) = h_{w-2}p^3 + h_{w-2}^*q^3,
\]

where \( h_w = qh_{w-1}^* + q^2h_{w-2}^* \) and \( h_w^* = ph_{w-1} + p^2h_{w-2} \), for \( w = 3, 4, \ldots \), with \( h_1 = 1 \), \( h_2 = q \), and \( h_1^* = 1 \) \( h_2^* = p \).

Now, the pdf of \( W(k) \) in the general case can be obtained recursively, following the similar argument used in the case of \( k = 3 \). Here, we present the following theorem without a proof.
Theorem 11.
The pdf of \( W(k) \) for \( k \in \{2, 3, 4, \ldots \} \) is

\[
P(W(k) = w) = h_{w-k+1}p^k + h^*_{w-k+1}q^k,
\]

where \( h_{w-k+1} = qh^{*}_{w-1} + q^2h^{*}_{w-2} + q^3h^{*}_{w-3} + \cdots + q^{k-1}h^{*}_{w-k+1} \)
and \( h^*_{w-k+1} = ph_{w-1} + p^2h_{w-2} + p^3h_{w-3} + \cdots + p^{k-1}h_{w-k+1} \),
for \( w=k,k+1,\ldots \), with \( h_w = h^*_w = 0 \) if \( w < 0 \), \( h_1 = 1 \), \( h_2 = q \), \( h_3 = qp + qq \), \ldots and \( h^*_1 = 1 \) \( h^*_2 = p \), \( h^*_3 = pq + pp \), \ldots.

Following the similar steps used for the pdf of \( W(k) \), the probability generating function (pgf) of \( W(k) \) is obtained.
Probability distribution function and probability generating function of $W(k)$

**Theorem 12.**

For $k=3$, the pgf of $W(k)$ can be written in a closed form as below.

$$G_{W(3)}(z) = \sum_{w=1}^{\infty} (h_w p^3 z^{w+2} + h_w^* q^3 z^{w+2}) = p^3 z^3 H(z) + q^3 z^3 H^*(z),$$

with $H(z) = (C + AC^*)/(1 - AA^*)$ and $H^*(z) = (C^* + A^* C)/(1 - AA^*)$, where $A = zp + z^2 p^2$, $A^* = zq + z^2 q^2$, $C = h_1 + zh_2 - zqh_1^*$, and $C^* = h_1^* + zh_2^* - zph_1$, and $h_1 = 1$, $h_2 = q$, $h_3 = qp + qq$ and $h_1^* = 1$ $h_2^* = p$, $h_3^* = pq + pp$, and for $n>2$, $h_n = qh_{n-1}^* + q^2 h_{n-2}^*$ and $h_n^* = ph_{n-1} + p^2 h_{n-2}$. 
Probability distribution function and probability generating function of $W(k)$

**Theorem 13.**

The pgf of $W(k)$ is written in a closed form as follows

$$G_{W(k)}(z) = \sum_{w=1}^{\infty} \left( h_w p^k z^{w+k-1} + h_w^* q^k z^{w+k-1} \right)$$

$$= p^k z^k H_w(z) + q^k z^k H_w^*(z),$$

with $H_w(z) = (C + A C^*)/(1 - A A^*)$ and $H_w^*(z) = (C^* + A^* C)/(1 - A A^*)$, where $A = z p + z^2 p^2 + \cdots + z^{k-1} p^{k-1}$, $A^* = z q + z^2 q^2 + \cdots + z^{k-1} q^{k-1}$, $C = h_1 + z h_2 - z q h_1^* - \cdots$, and $C^* = h_1^* + z h_2^* - z p h_1 - \cdots$, and $h_1 = 1$, $h_2 = q$, $h_3 = q p + q q$ and $h_1^* = 1$ $h_2^* = p$, $h_3^* = p q + p p$, and for $n > 3$, $h_n = q h_{n-1}^* + q^2 h_{n-2}^* + q^3 h_{n-3}^* + \cdots + q^{k-1} h_{n-k+1}^*$ and $h_n^* = p h_{n-1} + p^2 h_{n-2} + p^3 h_{n-3} + \cdots + p^{k-1} h_{n-k+1}$, where $h_n = h_n^* = 0$ for $n \leq 0$. 
Theorem 14.

For Fibonacci sequences of order \( k \), \( P(W(k) = w) \) is given by

\[
P(W(k) = w) = f_{w-k+1}(1/2)^{w-1} \quad \text{for } w = k, k + 1, k + 2, \ldots
\]

Note that \( f_k = 2^{k-2} \), for the Fibonacci sequence of order \( k \). For example, for \( k = 8 \),

\[
1 + 1 + 2 + 4 + 8 + 16 + 32 + 64 = 2^6 = f_8.
\]
Waiting time for either a consecutive k ones or k zeroes for the first time in a two state Markovian trial

Consider a sequence of Markov dependent Bernoulli variables $X_i$

$$P(X_i = 1) = p = 1 - P(X_i = 0) = 1 - q$$

Markov dependence

$$P(X_i = 1|X_{i-1} = 1) = a$$
$$P(X_i = 0|X_{i-1} = 0) = b.$$
Waiting time for either a consecutive k ones or k zeroes for the first time in a two state Markovian trial

**Theorem 15.**
For \(k=2\), the probability distribution function of \(W(2)\) is

\[
P(W(2) = 2n) = [(1 - a)(1 - b)]^{n-1}P(W(2) = 2) \\
P(W(2) = 2n + 1) = [(1 - a)(1 - b)]^{n-1}P(W(2) = 3)
\]

**Proof.**
Starting with

\[
P(W(2) = 2) = pa + qb,
\]
and

\[
P(W(2) = 3) = q(1 - b)a + p(1 - a)b,
\]
the pdf of \(W(2)\) can be calculated recursively.
Waiting time for either a consecutive $k$ ones or $k$ zeroes for the first time in a two state Markovian trial

**Corollary 16.**

From Theorem 15, we get

$$\sum_n P(W(2) = 2n) = \frac{(pa + qb)}{(1 - (1 - a)(1 - b))}$$

$$\sum_n P(W(2) = 2n + 1) = \frac{(qa(1 - b) + p(1 - a)b)}{(1 - (1 - a)(1 - b))}.$$

It follows that

$$\frac{(pa + qb + qa(1 - b) + p(1 - a)b)}{(1 - (1 - a)(1 - b))} = 1.$$
Waiting time for either a consecutive k ones or k zeroes for the first time in a two state Markovian trial

Suppose $pa + qb > q(1 - b)a + p(1 - a)b = qa + pb - ab \iff p(a - b) + q(b - a) + ab > 0$. In this case the probability of sequence terminating on even number of trials is larger than terminating on odd number of trials. In the symmetric case terminating on even number of trial is $\frac{2}{3}$. Therefore, $W(2)$ follows delayed geometric probability distribution.
Waiting time for either a consecutive \( k \) ones or \( k \) zeroes for the first time in a two state Markovian trial

**Theorem 17.**

The probability generating function of \( W(2) \) is

\[
G_{W(2)}(z) = \frac{((pa + qb)z^2 + (q(1 - b)a + p(1 - a)b)z^3)}{(1 - z^2(1 - a)(1 - b))}.
\]
Proof.

Multiplying $z^{2n}$ and $z^{2n+1}$ respectively on $P(W(2) = 2n)$ and $P(W(2) = 2n + 1)$, and summing over $n$, we can easily obtain the probability generating function of $W(2)$, that is,

$$\sum_{n} z^{2n} P(W(2) = 2n) = (pa + qb)z^2 \sum_{n} (z^2(1 - a)(1 - b))^{n-1}$$

$$= (pa + qb)z^2(1 - z^2(1 - a)(1 - b))^{-1},$$

$$\sum_{n} z^{n+1} P(W(2) = 2n + 1)$$

$$= (q(1 - b)a + p(1 - a)b)z^3 \sum_{n} (z^2(1 - a)(1 - b))^{n-1}$$

$$= (q(1 - b)a + p(1 - a)b)z^3(1 - z^2(1 - a)(1 - b))^{-1}.$$

Hence, Theorem 18 follows.
Waiting time for either a consecutive k ones or k zeroes for the first time in a two state Markovian trial

**Theorem 18.**

For $k=3$, the probability distribution function of $W(3)$ is

$$P(W(3) = 2 + n) = (1 - a)(1 - b)h_n aa + (1 - b)(1 - a)h^*_n bb,$$

$$(n > 4)$$

where $h_0 = 0$, $h_1 = p$, $h_2 = (1 - b)q$, $h_1^* = q$, $h_2^* = (1 - a)p$, $h_3 = (1 - a)p + bq$, $h_3^* = (1 - b)q + ap$, $h_4 = h_2^* b + h_3^*(1 - a)$, $h_4^* = h_2 a + h_3^*(1 - b)$, and $h_n = h_{n-1}^* + bh_{n-2}^*$, $h_n^* = h_{n-1} + ah_{n-2}$ for $n > 4$. Note that $h^*$ is "complement" of $h$, that is, $p$ and $a$ in $h$ are replaced by $q$ and $b$ respectively in $h^*$. 
Proof.
Since we have that

\[ P(W(3) = 3) = paa + qbb \]
\[ P(W(3) = 4) = q(1 - b)aa + p(1 - a)bb \]
\[ P(W(3) = 5) = (1 - a)p(1 - b)aa + bq(1 - b)aa+ \\
(1 - b)q(1 - a)bb + ap(1 - a)bb, \]

the pdf of \( W(3) \) can be calculated recursively.

Following the similar steps as in the case \( k = 3 \), the following theorem can be easily established.
Waiting time for either a consecutive \( k \) ones or \( k \) zeroes for the first time in a two state Markovian trial

**Theorem 19.**

The probability distribution of \( W(k) \) is

\[
P(W(k) = ((k - 1) + n) = (1 - a)(1 - b)h_n a^{k-1} + (1 - b)(1 - a)h_n^* b^{k-1}, \text{ for } n > k + 1
\]

\[
P(W(k) = k) = p a^{k-1} + q b^{k-1}
\]

\[
P(W(k) = k + 1) = q(1 - b)a^{k-1} + p(1 - a)b^{k-1}
\]

\[
P(W(k) = k + 2) = (1 - b)h_3 a^{k-1} + (1 - a)h_3^* b^{k-1},
\]

where \( h_0 = 0, h_1 = p, h_2 = (1 - b)q, h_1^* = q, h_2^* = (1 - a)p, h_3 = h_2^* + bh_1^*, h_3^* = h_2 + ah_1, h_4 = (1 - a)h_3^* + bh_2^* + bbh_1^*, h_4^* = (1 - b)h_3 + ah_2 + aah_1, \ldots, \) and

\[
h_n = h_{n-1}^* + bh_{n-2}^* + bbh_{n-3}^* + \cdots + b^{k-2} h_{n-k+1}^*, \quad h_n^* = h_{n-1} + ah_{n-2} + aah_{n-3} + \cdots + a^{k-2} h_{n-k+1},
\]

for \( n = k+2, k+3, \ldots \).
\[ h_{x} = \left( -a \right) h_{y} \]
Applications

• Waiting time to get a cluster of positive or negative response on certain treatment to a DNA sequence
• A streak of hits and no hits of baseball players
• An interesting application of the sooner waiting time problem with \( k = 2 \) can be found in Section 5 of Uchida and Aki (1995)
References


