On Generalizations of \((k_1, k_2)\)-runs

A. N. Kumar* and N. S. Upadhye**

Department of Mathematics, Indian Institute of Technology Madras, Chennai-600036, India.
*Email: amit.kumar2703@gmail.com
**Email: neelesh@iitm.ac.in

Abstract

The paper deals with three generalized dependent setups arising from an independent sequence of Bernoulli trials. Various distributional properties, such as probability generating function, probability mass function and moments are discussed for these setups and their waiting time. Also, explicit forms of probability generating function and probability mass function are obtained. Finally, an application to demonstrate the relevance of the results is given.

Keywords : \((k_1, k_2)\)-runs; waiting time; probability generating function; probability mass function; moments; Fibonacci words.

MSC 2010 Subject Classifications : Primary : 60E05, 62E15 ; Secondary : 60C05, 60E10.

1 Introduction

Runs and patterns play a crucial role in applied statistics and has numerous applications, for example, reliability theory (see Fu [9] and Fu and Hu [11]), nonparametric hypothesis testing (Balakrishnan and Koutras [5]), DNA sequence analysis (Fu et al. [10]), statistical testing (Balakrishnan et al. [7]), computer science (Sinha et al. [21]) and quality control (Moore [18]) among many others.

A run can be defined as an occurrence of specific patterns of failures or successes or both in a sequence of Bernoulli trials. In particular, a pattern of consecutive successes of length \(k\) is considered by Philippou et al. [19] and described geometric and negative binomial distribution of order \(k\). Also, Philippou and Makri [20] discussed binomial distribution of order \(k\). Later, Huang and Tsai [13] extended the pattern by observing at least \(k_1\) consecutive failures followed by at least \(k_2\) consecutive successes and studied a modified binomial distribution of order \(k\) or \((k_1, k_2)\)-runs. Recently, Dafnis et al. [8] also considered three types of \((k_1, k_2)\)-runs which include the pattern discussed in Huang and Tsai [13]. Though there have been several studies on this topic, still there are many problems which can not be studied based on the available literature. For example,
\( \text{(i) let us consider the quality control problem in which the system is said to be in control, whenever, (on the control chart) not more than two consecutive points exceed the control limits and at least three succeeding points are inside the control limits (see (T1) below with } \ell_1 = 1, k_1 = 2 \text{ and } \ell_2 = 3 \text{). Similarly, (ii) consider a climatology problem, in which, climatologist is interested in knowing the distribution of at least two consecutive rainy days followed by exactly five consecutive dry days (see (T2) below with } \ell_1 = 2 \text{ and } k_2 = 5 \text{). Also, there are several such problems that occur in brand switching, learning, reliability and queuing models. Hence, there is a need to generalize the results related to } (k_1, k_2) \text{-runs.}

\( \text{In this paper, we generalize } (k_1, k_2) \text{-runs to include the following patterns, for } 1 \leq \ell_1 \leq k_1 \text{ and } 1 \leq \ell_2 \leq k_2, \)

\( (T1) \text{ at least } \ell_1 \text{ and at most } k_1 \text{ consecutive 0's followed by at least } \ell_2 \text{ consecutive 1's.} \)

\( (T2) \text{ at least } \ell_1 \text{ consecutive 0's followed by at least } \ell_2 \text{ and at most } k_2 \text{ consecutive 1's.} \)

\( (T3) \text{ at least } \ell_1 \text{ and at most } k_1 \text{ consecutive 0's followed by at least } \ell_2 \text{ and at most } k_2 \text{ consecutive 1's.} \)

\( \text{Note that (T1), (T2) and (T3) contain various } (k_1, k_2) \text{-runs. For example,} \)

1. if \( \ell_1 = k_1 \) then (T1) leads to, exactly \( \ell_1 \) consecutive 0's followed by at least \( \ell_2 \) consecutive 1's,

2. if \( \ell_2 = k_2 \) then (T2) leads to, at least \( \ell_1 \) consecutive 0's followed by exactly \( \ell_2 \) consecutive 1's,

3. if \( \ell_1 = \ell_2 = 1 \) then (T3) leads to, at most \( k_1 \) consecutive 0's followed by at most \( k_2 \) consecutive 1's,

4. if \( \ell_1 = k_1 \) and \( \ell_2 = k_2 \) then (T3) leads to, exactly \( k_1 \) consecutive 0's followed by exactly \( k_2 \) consecutive 1's,

\( \text{and similarly, other special cases can be seen by choosing the values for } \ell_1 \text{ and } \ell_2, k_1 \text{ and } k_2 \text{ appropriately.} \)

Dafnis et al. \[ \text{considered two special cases of (T3), namely, (i) } \ell_1 = \ell_2 \text{ and (ii) } \ell_1 = k_1 \text{ and } \ell_2 = k_2. \]

\( \text{Next, let us define} \)

\[
I_s^{(m)} := \begin{cases} 
(1 - \zeta_m) \cdots (1 - \zeta_{m+\ell_1-1}) (1 - \zeta_{m+\ell_1}) \cdots (1 - \zeta_{m+s+\ell_1-1}) \zeta_{m+s+\ell_1} \cdots \zeta_{m+s+\ell_1+\ell_2-1}, & m = 1, \\
\zeta_m (1 - \zeta_{m+1}) \cdots (1 - \zeta_{m+\ell_1}) (1 - \zeta_{m+\ell_1+1}) \cdots (1 - \zeta_{m+s+\ell_1+1}) \zeta_{m+s+\ell_1+1} \cdots \zeta_{m+s+\ell_1+\ell_2}, & m \geq 2,
\end{cases}
\]

\[
J_t^{(m)} := (1 - \zeta_m) \cdots (1 - \zeta_{m+\ell_1-1}) \zeta_{m+\ell_1} \cdots \zeta_{m+\ell_1+\ell_2-1} \zeta_{m+\ell_1+\ell_2} \cdots \zeta_{m+t+\ell_1+\ell_2-1} (1 - \zeta_{m+t+\ell_1+\ell_2}),
\]

\[
K_{s,t}^{(m)} := \begin{cases} 
(1 - \zeta_m) \cdots (1 - \zeta_{m+\ell_1-1}) (1 - \zeta_{m+\ell_1}) \cdots (1 - \zeta_{m+s+\ell_1-1}) \zeta_{m+s+\ell_1}, & m = 1, \\
\zeta_m (1 - \zeta_{m+1}) \cdots (1 - \zeta_{m+\ell_1}) (1 - \zeta_{m+\ell_1+1}) \cdots (1 - \zeta_{m+s+\ell_1+1}) \zeta_{m+s+\ell_1+1}, & m \geq 2,
\end{cases}
\]

\[
I_m := \max_{0 \leq s \leq k_1-\ell_1} I_s^{(m)}, \quad J_m := \max_{0 \leq s \leq k_2-\ell_2} J_s^{(m)}, \quad K_m := \max_{0 \leq s \leq k_2-\ell_2} K_s^{(m)},
\]

where \( \zeta_1, \zeta_2, \ldots, \zeta_n \) be a finite sequence of independent Bernoulli trials with success (denoted by 1) probability \( p \) and failure (denoted by 0) probability \( q = 1 - p \). Also, Let \( H^n_{k_1,k_2}, H^n_{\ell_1,\ell_2,k_2} \) and \( H^n_{k_1,\ell_1,\ell_2,k_2} \) be the number.
of occurrences for \((T1)\), \((T2)\) and \((T3)\) type event, respectively. Then, random variable representation of \(H^n_{\ell_1,k_1,\ell_2}, H^n_{\ell_1,\ell_2,k_2}\) and \(H^n_{\ell_1,k_1,\ell_2,k_2}\) can be seen as follows:

\[
H^n_{\ell_1,k_1,\ell_2} = \sum_{m=1}^{n-\ell_1-\ell_2+1} I_m, \quad H^n_{\ell_1,\ell_2,k_2} = \sum_{m=1}^{n-\ell_1-\ell_2} J_m \quad \text{and} \quad H^n_{\ell_1,k_1,\ell_2,k_2} = \sum_{m=1}^{n-\ell_1-\ell_2} K_m.
\]

Now, let us consider a particular realization in a sequence of 20 Bernoulli trials given by

\[
001111101100010100011.
\]

Here, note that

\[(T1) \ H^{20}_{1,1,1} = 2, \ H^{20}_{1,2,2} = 2, \ H^{20}_{2,2,3} = 1 \text{ and } H^{20}_{1,2,1} = 3.\]

\[(T2) \ H^{20}_{1,1,2} = 3, \ H^{20}_{1,2,1} = 1, \ H^{20}_{2,2,2} = 0 \text{ and } H^{20}_{1,4,4} = 1.\]

\[(T3) \ H^{20}_{1,1,1,1} = 1, \ H^{20}_{1,2,2,2} = 1, \ H^{20}_{1,1,1,2} = 2 \text{ and } H^{20}_{1,2,1,2} = 2.\]

For more details about runs and patterns, we refer the reader to Aki [1], Aki et al. [2], Antzoulakos et al. [3], Antzoulakos and Chadjiconstantinidis [4], Balakrishnan and Koutras [5], Dafnis et al. [8], Fu and Koutras [12], Koutras [14, 15] and Makri et al. [17] and references therein.

This paper is organized as follows. In Section 2, we obtain the double probability generating function (PGF) and waiting time for \(H^n_{\ell_1,k_1,\ell_2}, H^n_{\ell_1,\ell_2,k_2}\) and \(H^n_{\ell_1,k_1,\ell_2,k_2}\). Next, using double PGF, we derive recursive relation in PGF, probability mass function (PMF) and moments and also derive an explicit form of PGF and PMF. Finally, using double PGF for waiting time, we obtain the PGF, recursive relations in PMF and moments. In Section 3, we demonstrate the relevance of the results with an interesting application to Fibonacci words.

## 2 Distributions Related to \(H^n_{\ell_1,k_1,\ell_2}, H^n_{\ell_1,\ell_2,k_2}\) and \(H^n_{\ell_1,k_1,\ell_2,k_2}\)

In this section, we discuss various distributional properties such as PGF, PMF and moments for \(H^n_{\ell_1,k_1,\ell_2}, H^n_{\ell_1,\ell_2,k_2}\) and \(H^n_{\ell_1,k_1,\ell_2,k_2}\) and their waiting time.

The method used, to derive the results is general, can be formulated in the following way. Let \(Y_n\) be a random variable related to \((k_1,k_2)\)-runs. Then, we can define a Markov chain \(\{Z_t, t \geq 0\}\) on discrete space \(\Omega\) (which can be partitioned into discrete subspaces \(\{0,1,2,\ldots\}\) of maximum length \(\varepsilon_n\) and containing one and only one \((k_1,k_2)\)-event) such that \((k_1,k_2)\)-runs has occurred \(v\) times if and only if Markov chain is in \(v\)-th discrete subspace (say \(E_v = \{E_{v,0}, E_{v,1}, \ldots, E_{v,v}\}\) such that \(\Omega = \cup_{v \geq 0} E_v\)). Now, assume \(A\) and \(B\) be \((r+1) \times (r+1)\) matrices when \((k_1,k_2)\)-runs are observed from \(v\) to \(v+1\) times, respectively. Let \(\phi_n(\cdot)\) and \(\Phi(\cdot,\cdot)\) be the single and double generating function of \(Y_n\) and \(H_j(\cdot)\) and \(H(\cdot,\cdot)\) be the single and double generating function of \(j\)-th waiting time for \(Y_n\). Then, the double generating function for \(Y_n\) and its waiting time is given
respectively, where \( \kappa_0 \) is the initial distribution, \( \vartheta(z,t) = I - z(A + tB) \) be \((r + 1) \times (r + 1)\) matrix, \( 1^t \) is the transpose of row matrix \((1,1,\ldots,1)\) with \((r + 1)\) entries and \( I \) is \((r + 1) \times (r + 1)\) identity matrix. For more details, we refer the reader to Antzoulakos et al. \cite{[8]} and Dafnis et al. \cite{[3]}.

Let us define some notations as

\[
a(p) := q^ip^{j}, \quad \ell := \ell_1 + \ell_2, \quad m_1 := k_1 - \ell_1 + 1, \quad m_2 := k_2 - \ell_2 + 1,
\]

\( \rho_r \) is the \( r \)-th waiting time for \((k_1, k_2)\)-runs, \( p_{,n} \) and \( g_r(\cdot) \) be the PMF of \((k_1, k_2)\)-runs and \( \rho_r \), respectively. Also, define \( \mu_{n,j} \) and \( \tilde{\mu}_{r,j} \) be the \( j \)-th (non-central) moment of \((k_1, k_2)\)-runs and \( \rho_r \), respectively, where \( n \) denotes the number of Bernoulli trials.

### 2.1 Distribution of \( H^n_{\ell_1,k_1,\ell_2} \) and its Waiting Time

Recall that \( H^n_{\ell_1,k_1,\ell_2} \) is the number of occurrences of \((at least \ell_1)\) at most \( k_1 \) consecutive 0’s followed by at least \( \ell_2 \) consecutive 1’s. Here, \( r = k_1 + \ell_2 + 1 \) and \( k_1^+ \) is the element after \( k_1 \) consecutive 0’s (if failures occur) in \( \{0,1,\ldots,k_1,k_1^+ = k_1 + 1,k_1 + 2,\ldots,k_1 + \ell_1 + 1\} \). It is easy to see that \( \mathbb{P}(H^n_{\ell_1,k_1,\ell_2} = 0) = 1 \) and \( \varepsilon_n := \sup \{ x : \mathbb{P}(H^n_{\ell_1,k_1,\ell_2} = x) > 0 \} = [n/\ell] \). Therefore, \( \kappa_0 = (1,0,\ldots,0)_{1 \times (k_1 + \ell_2 + 2)} \), \( A = [a_{i,j}]_{(k_1 + \ell_2 + 2) \times (k_1 + \ell_2 + 2)} \) with non-zero entries

- \( a_{1,i} = p \) and \( a_{i+1,i} = q \) for \( 1 \leq i \leq \ell_1 \),
- \( a_{i,k_1+3} = p \) and \( a_{i,i+1} = q \) for \( \ell_1 + 1 \leq i \leq k_1 + 1 \),
- \( a_{k_1+2,i} = p \) and \( a_{k_1+2,k_1+2} = q. \)
- \( a_{i,2} = q \) for \( k_1 + 3 \leq i \leq k_1 + \ell_2 + 2 \) and \( a_{i,i+1} = p \) for \( k_1 + 3 \leq i \leq k_1 + \ell_2 \).
- \( a_{k_1+\ell_2+2,k_1+\ell_2+2} = p \)

and \( B = [b_{i,j}]_{(k_1+\ell_2+2) \times (k_1+\ell_2+2)} \) is the matrix of non-zero entry \( b_{k_1+\ell_2+1,k_1+\ell_2+2} = p \). Hence, using (1), it can be easily verified that

\[
\Phi(t,z) = \sum_{n=0}^{\infty} \phi_n(t)z^n = \frac{1}{1 - z - (qz)^t (pz)^{2} (1 - (qz)^{k_1+\ell_1+1})} = \frac{1}{1 - z - a(p)z^t (t - 1) (1 - (qz)^{m_1})}, \tag{3}
\]

Now, using (3), we have the following results.
Theorem 2.1. The recursive relation in PGF, PMF and moments of $H_{t_1, k_1, \ell_2}^n$, for $n \geq \ell$, are given by

(i) $\phi_n(t) = \phi_{n-1}(t) + a(p)(t-1) [\phi_{n-\ell}(t) - q^{m_1} \phi_{n-\ell-m_1}(t)]$

with initial condition $\phi_n(t) = 1$, for $n \leq \ell - 1$.

(ii) $p_{m,n} = p_{m,n-1} + a(p) [p_{m-1,n-\ell} - p_{m,n-\ell} - q^{m_1} (p_{m-1,n-\ell-m_1} - p_{m,n-\ell-m_1})]$

with initial conditions $p_{0,n} = 1$ and $p_{m,n} = 0$, $m > 0$ for $n \leq \ell - 1$.

(iii) $\mu_{n,j} = \mu_{n-1,j} + a(p) \sum_{i=0}^{j-1} \binom{j}{k} [\mu_{n-\ell,k} - q^{m_1} \mu_{n-\ell-m_1,k}]$, for $j \geq 1$

with initial conditions $\mu_{n,0} = 1$ and $\mu_{n,j} = 0$ for all $j \geq 1$ and $n \leq \ell - 1$.

Proof. From (3), (i) follows and using the definition of PGF, (ii) follows. Substituting $t = e^z = \sum_{m=0}^{\infty} x^m/m!$ in (i) and comparing the coefficient of $x^m/m!$, (iii) follows. $\Box$

Next, we obtain an explicit form of PGF and PMF using Theorem 2.1.

Theorem 2.2. Assume the conditions of Theorem 2.1 hold, then PGF and PMF of $H_{t_1, k_1, \ell_2}^n$ are given by

(i) $\phi_n(t) = \sum_{u=0}^{n} \sum_{v=0}^{n-\ell} \binom{n-u}{u} \binom{n-v}{v} \left( n - u(\ell-1) - v(\ell+m_1) - 1 \right) (-1)^v q^{m_1} (a(p)(t-1))^{u+v}.$

(ii) $p_{m,n} = \sum_{u=0}^{n} \sum_{v=0}^{n-\ell} \binom{n-u}{u} \binom{n-v}{v} \left( n - u(\ell-1) - v(\ell+m_1) - 1 \right) \left( \frac{u+v}{m} \right) (-1)^{u-m} q^{m_1} a(p)^{u+v},$

where $\binom{n}{u_1 u_2 \ldots u_{\ell}} = \frac{n!}{u_1! u_2! \ldots u_{\ell}!}$.

Proof. (i) For $(t, z) \in \{|t| \leq 1, |z| < 1 \text{ and } |z + a(p)z^{\ell}(t-1)(1-(qz)^{m_1})| < 1\}$, (3) can be written as

$$\Phi(t, z) = \sum_{n=0}^{\infty} (z + a(p)z^{\ell}(t-1)(1-(qz)^{m_1}))^n.$$

Now, using binomial expansion and interchanging summations, we get the required result.

(ii) Following the steps similar to (i) with recursive relation (ii) of Theorem 2.1, the proof follows. $\Box$

Next, using (2) with some algebraic manipulations, it can be easily verified that

$$H(t, z) = 1 + \sum_{r=1}^{\infty} \left( \frac{a(p)t^{\ell}(1-(qt)^{m_1})}{1 - t + a(p)t^{\ell}(1-(qt)^{m_1})} \right)^r z^r. \quad (4)$$

Hence, using (4), we have the following theorem.

Theorem 2.3. Let $\delta_{i,j}$ denote Kronecker delta function. The PGF, PMF and moments of $\rho_r$, for $r \geq 1$, are given by

(i) $H_r(t) = \left( \frac{a(p)t^{\ell}(1-(qt)^{m_1})}{1 - t + a(p)t^{\ell}(1-(qt)^{m_1})} \right)^r.$

(ii) $g_r(m) = g_r(m-1) + a(p) \left[ g_{r-1}(m-\ell) - g_r(m-\ell) - q^{m_1} (g_{r-1}(m-\ell-m_1) - g_r(m-\ell-m_1)) \right]$

for $m \geq \ell r$ with initial condition $g_0(m) = \delta_{m,0}$, $g_r(m) = 0$ for $m \leq \ell r - 1$. 

Also, if 0 occurs after at least \( \ell_1 \) consecutive 0’s followed by \( \ell_2 \) at most \( k_2 \) consecutive 1’s. Here, \( r = \ell_1 + k_2, \Pr(H_{\ell_1,\ell_2,k_2}^0 = 0) = 1 \) and \( \varepsilon_n := \sup \{ x : \Pr(H_{\ell_1,\ell_2,k_2}^n = x) > 0 \} = \lfloor n/\ell \rfloor \).

Hence, using (1), it can be easily verified that

\[
\Phi(t, x) = \frac{1 - a(p)x(t - 1)\sum_{i=1}^{m_2} (px)^{i-1}}{1 - z - a(p)x(t - 1)(1 - (px)^{m_2})}. \tag{5}
\]

Now, using (5), the following theorem can be easily derived.

**Theorem 2.4.** The recursive relation in PGF, PMF and moments of \( H_{\ell_1,\ell_2,k_2}^n \), for \( n \geq \ell + 1 \), are given by

(i) \( \phi_n(t) = \phi_{n-1}(t) + a(p)(t - 1) [\phi_{n-\ell}(t) - p^{m_2}\phi_{n-\ell-m_2}(t)] - a(p)(t - 1)p^{n-\ell}1(\ell + 1 \leq n \leq \ell + m_2 - 1) \)

with initial condition \( \phi_1(t) = 1 \), for \( n \leq \ell \), where \( 1(A) \) denotes the indicator function of set \( A \).

(ii) \( p_{m,n} = p_{m,n-1} + a(p) \left[ p_{m-1,n-\ell} - p_{m,n-\ell} - p^{m_2}(p_{m-1,n-\ell-m_2} - p_{m,n-\ell-m_2}) \right] - a(p)p^{n-\ell}1(m = 1, \ell + 1 \leq n \leq \ell + m_2 - 1) - 1(m = 0, \ell + 1 \leq n \leq \ell + m_2 - 1) \)

with initial conditions \( p_{0,n} = 1, p_{m,n} = 0, m > 0 \) for \( n \leq \ell \).

(iii) \( \mu_{n,j} = \mu_{n-1,j} + a(p) \sum_{k=0}^{j-1} \binom{j}{k} [\mu_{n-\ell,k} - p^{m_2}\mu_{n-\ell-m_2,k}] - a(p)p^{n-\ell}1(\ell + 1 \leq n \leq \ell + m_2 - 1) \)

for \( j \geq 1 \) with initial conditions \( \mu_{n,0} = 1 \) and \( \mu_{n,j} = 0 \) for all \( j \geq 1 \) and \( n \leq \ell \).

Next, we obtain an explicit form for PGF and PMF using Theorem 2.4.
Theorem 2.5. Assume the conditions of Theorem 2.4 hold, then PGF and PMF of \( H^\ell_{\ell_1, \ell_2, k_2} \) are given by

(i) \( \phi_n(t) = \chi_n(t) - a(p)(t-1) \sum_{i=\ell}^{\ell+m_2-1} p^{i-\ell} \chi_{n-i}(t) \)

(ii) \( p_{m,n} = V_{m,n} - a(p) \sum_{i=\ell}^{\ell+m_2-1} p^{i-\ell} (V_{m-1,n-i} - V_{m,n-i}) \),

where

\[
\chi_n(t) = \sum_{u=0}^{n-1} \sum_{v=0}^{m-1} \left( n-u(\ell-1) - v(\ell + m_2 - 1) \right) (-1)^v p^{um_2} (a(p)(t-1))^{u+v} \]

and

\[
V_{m,n} = \sum_{u=0}^{n-1} \sum_{v=0}^{m-1} \left( n-u(\ell-1) - v(\ell + m_2 - 1) \right) \left( u+v \right) (-1)^um_2 a(p)^{u+v}. \]

Next, using (2), it can be easily verified that

\[
H(t, z) = 1 + \frac{c}{1 - pt} \sum_{r=1}^{\infty} \left( \frac{a(pt)(1 - (pt)^m_2)}{1 - t + a(pt)(1 - (pt)^m_2)} \right)^r z^r. \tag{6} \]

Hence, using (6), the following theorem can be easily derived.

Theorem 2.6. The PGF, PMF and moments of \( \rho_r \), for \( r \geq 1 \), are given by

(i) \( H_r(t) = \frac{c}{1 - pt} \left( \frac{a(pt)(1 - (pt)^m_2)}{1 - t + a(pt)(1 - (pt)^m_2)} \right)^r. \)

(ii) \( g_r(m) = g_r(m-1) + a(p) [g_{r-1}(m-\ell) - g_r(m-\ell) - p^{m_2} (g_{r-1}(m-\ell-m_2) - g_r(m-\ell-m_2))], \ r \geq 2 \)

with initial condition \( g_0(m) = \delta_{0,m} \) and

\[
g_1(m) = g_1(m-1) - a(p) [g_1(m-\ell) - p^{m_2} g_1(m-\ell-m_2)] + qa(p)p^{m-\ell-1} 1(\ell+1 \leq m \leq \ell + m_2), \]

for \( m \geq \ell r + 1 \), \( g_r(m) = 0 \) whenever \( m \leq \ell r \) and \( r \geq 1 \).

(iii) \( \bar{\mu}_{r,j} = \sum_{k=0}^{j} \binom{j}{k} \left[ \bar{\mu}_{r,k} + a(p) (\ell^{-k} - p^{m_2} \ell + m_2)^{-k} \right] \left( \bar{\mu}_{r-1,k} - \bar{\mu}_{r,k} \right), \ j \geq 1 \) and \( r \geq 2 \)

with initial condition \( \bar{\mu}_{0,i} = \delta_{0,i} \) and

\[
\bar{\mu}_{1,j} = \sum_{k=0}^{j} \binom{j}{k} \bar{\mu}_{1,k} \left[ 1 - a(p) (\ell^{-k} - p^{m_2} \ell + m_2)^{-k} \right] + qa(p) \sum_{k=\ell+1}^{\ell+m_2} k^j p^{k-\ell-1} \]

The proofs of Theorems 2.4 - 2.6 follow using steps similar to the proofs of Theorems 2.1 - 2.3.

2.3 Distribution of \( H^n_{\ell_1, k_1, \ell_2, k_2} \) and its Waiting Time

Recall that \( H^n_{\ell_1, k_1, \ell_2, k_2} \) is the number of occurrences of (at least \( \ell_1 \)) at most \( k_1 \) consecutive 0’s followed by (at least \( \ell_2 \)) at most \( k_2 \) consecutive 1’s. Here, \( r = k_1 + k_2 + 1 \) and \( k_1^+ \) is the element after \( k_1 \) consecutive 0’s (if failures occur) in \( \{0, 1, \ldots, k_1, k_1^+ = k_1 + 1, k_1 + 2, \ldots, k_1 + k_2 + 1\} \). It is easy to see that \( \mathbb{P}(H^n_{\ell_1, k_1, \ell_2, k_2} = 0) = 1 \) and \( \varepsilon = \sup \{ x : \mathbb{P}(H^n_{\ell_1, k_1, \ell_2, k_2} = x) > 0 \} = \lfloor n/\ell \rfloor \). Also, if 0 occurs after (at least \( \ell_1 \)) at most \( k_1 \) consecutive 0’s followed by (at least \( \ell_2 \)) at most \( k_2 \) consecutive 1’s then \( H^n_{\ell_1, k_1, \ell_2, k_2} \) moves \( v \) (any) to \( v + 1 \) times. Therefore, \( \kappa_0 = (1, 0, \ldots, 0)_1 \times (k_1 + k_2 + 2) \), \( A = [a_{i,j}]_{(k_1 + k_2 + 2) \times (k_1 + k_2 + 2)} \) with non-zero entries.
• $a_{i,1} = p$ and $a_{i,i+1} = q$ for $1 \leq i \leq \ell_1$,

• $a_{i,k_1+3} = p$ and $a_{i,i+1} = q$ for $\ell_1 + 1 \leq i \leq k_1 + 1$,

• $a_{k_1+2,1} = p$ and $a_{k_1+2,k_1+2} = q$,

• $a_{i,2} = q$ for $k_1 + 3 \leq i \leq k_1 + \ell_2 + 1$ and $a_{i,i+1} = p$ for $k_1 + 3 \leq i \leq k_1 + k_2 + 1$,

• $a_{k_1+k_2+2,1} = p$

and $B = [b_{i,j}](k_1+k_2+2) \times (k_1+k_2+2)$ is the matrix of non-zero entries $b_{i,2} = q$ for $k_1 + \ell_2 + 2 \leq i \leq k_1 + k_2 + 2$.

Hence, using (i), it can be easily verified that

$$
\phi(t, z) = \sum_{n=0}^{\infty} \phi_n(t) z^n = \frac{1 - a(p)z^\ell(t - 1) (1 - (qz)^{m_1}) \sum_{i=1}^{m_2} (pz)^i}{1 - a(p)z^\ell(t - 1) (1 - (qz)^{m_1}) (1 - (pz)^{m_2})},
\tag{7}
$$

Now, using (i), the following theorem can be easily derived.

**Theorem 2.7.** The recursive relations in PGF, PMF and moments of $H_{n,\ell_1,\ell_2,\ell_3}$, for $n \geq \ell + 1$, are given by

(i) $\phi_n(t) = \phi_{n-1}(t) + a(p)[t-1][\phi_{n-\ell}(t) - q^{m_1}\phi_{n-\ell-m_1}(t) - p^{m_2}\phi_{n-\ell-m_2}(t) + q^{m_1}p^{m_2}\phi_{n-\ell-m_1-m_2}(t)]$

$- a(p)(t-1)p^{n-\ell}\left[1(\ell + 1 \leq n \leq \ell + m_2 - 1) - q/p \right]^{m_1}_{m_1}(1(\ell + 1 \leq n \leq \ell + m_1 + m_2 - 1))$

with initial condition $\phi_0(t) = 1$, for $n \leq \ell$.

(ii) $p_{m,n} = p_{m,n-1} - a(p) p^{n-\ell}\left[1(m = 1, \ell + 1 \leq n \leq \ell + m_2 - 1) - 1(m = 0, \ell + 1 \leq n \leq \ell + m_2 - 1)\right]$\n
$- (q/p)^{m_1}\left[1(m = 1, \ell + 1 \leq n \leq \ell + m_1 + m_2 - 1) - 1(m = 0, \ell + m_1 + m_2 - 1)\right]$\n
$+ a(p)[p_{m-1,n-\ell} - p_{m,n-\ell} - q^{m_1}(p_{m-1,n-\ell-m_1} - p_{m,n-\ell-m_1}) - p^{m_2}(p_{m-1,n-\ell-m_2} - p_{m,n-\ell-m_2})]$\n
$+ q^{m_1}p^{m_2}(p_{m-1,n-\ell-m_1-m_2} - p_{m,n-\ell-m_1-m_2})$

with initial conditions $p_{0,n} = 1$ and $p_{m,n} = 0$, $m > 0$ for $n \leq \ell$.

(iii) $\mu_{n,j} = \mu_{n-1,j} + a(p) \sum_{k=0}^{j-1} \binom{j}{k}[\mu_{n-\ell,k} - q^{m_1}\mu_{n-\ell-m_1,k} - p^{m_2}\mu_{n-\ell-m_2,k} + q^{m_1}p^{m_2}\mu_{n-\ell-m_1-m_2,k}]$

$- a(p) p^{n-\ell}\left[1(\ell + 1 \leq n \leq \ell + m_2 - 1) - (q/p)^{m_1}\right]^{m_1}_{m_1}(1(\ell + 1 \leq n \leq \ell + m_1 + m_2 - 1))$, $j \geq 1$

with initial conditions $\mu_{n,0} = 1$ and $\mu_{n,j} = 0$ for all $j \geq 1$ and $n \leq \ell$.

Next, we obtain an explicit form for PGF and PMF using Theorem 2.7.

**Theorem 2.8.** Assume the conditions of Theorem 2.7 hold, then PGF and PMF of $H_{n,\ell_1,\ell_2,\ell_3}$ are given by

(i) $\phi_n(t) = \varphi_n(t) - a(p)(t-1)\left[\sum_{i=\ell}^{\ell+m_2-1} p^{i-\ell}\varphi_n(i) - \left(q/p\right)^{m_1}\sum_{i=\ell+m_1}^{\ell+m_1+m_2-1} p^{i-\ell}\varphi_n(i)\right]$
Next, using (2), it can be easily verified that

\[ H(t,z) = 1 + \frac{qt}{1 - pt} \sum_{r=1}^{\infty} \left( \frac{a(p)t^r(1 - (qt)^{m_1})(1 - (pt)^{m_2})}{1 - t + a(p)t^r(1 - (qt)^{m_1})(1 - (pt)^{m_2})} \right)^z. \] (8)

Hence, using (5), the following theorem can be easily derived.

**Theorem 2.9.** The PGF, PMF and moments of \( \rho_r \), for \( r \geq 1 \), are given by

(i) \( H_r(t) = \frac{qt}{1 - pt} \left( \frac{a(p)t^r(1 - (qt)^{m_1})(1 - (pt)^{m_2})}{1 - t + a(p)t^r(1 - (qt)^{m_1})(1 - (pt)^{m_2})} \right)^r. \)

(ii) \( g_r(m) = g_r(m-1) + a(p)g_{r-1}(m-\ell) - g_r(m-\ell) = q^{m_1}(g_{r-1}(m-\ell + m_1) - g_r(m-\ell - m_1)) \)
\[ - p^{m_2}(g_{r-1}(m-\ell-m_2) - g_r(m-\ell-m_2) + q^{m_1}p^{m_2}(g_{r-1}(m-\ell-m_1-m_2) - g_r(m-\ell-m_1-m_2))), \]
for \( r \geq 2 \) with initial condition \( g_0(m) = \delta_{m,0} \) and
\[ g_1(m) = g_1(m-1) + qa(p)p^{m-\ell-1} \left( 1 + \ell + m \leq m \leq \ell + m_2 \right) - \frac{m_1}{p} \left( 1 + \ell + m + 1 \leq m \leq \ell + m_1 + m_2 \right) \)
\[ - a(p)[g_1(m-\ell) - q^{m_1}g_1(m-\ell-m_1) - p^{m_2}g_1(m-\ell-m_2) + q^{m_1}p^{m_2}g_1(m-\ell-m_1-m_2)], \]
for \( m \geq \ell + 1 \), \( g_r(m) = 0 \) whenever \( m \leq \ell r \) and \( r \geq 1 \).

(iii) \( \tilde{\mu}_{r,j} = \sum_{k=0}^{j} \binom{j}{k} \left( \tilde{\mu}_{r+k} + a(p)(\ell^{j-k} - q^{m_1}(\ell + m_1)^{j-k} - p^{m_2}(\ell + m_2)^{j-k} \right. \)
\[ + q^{m_1}p^{m_2}(\ell + m_1 + m_2)^{j-k})] \]
\[ \left. (\tilde{\mu}_{r-1,k} - \tilde{\mu}_{r,k}), \quad j \geq 1, \text{ and } r \geq 2 \right. \]
with initial condition \( \tilde{\mu}_{0,i} = \delta_{0,0} \) and
\[ \tilde{\mu}_{1,j} = \sum_{k=0}^{j} \binom{j}{k} \tilde{\mu}_{1+k} \left[ 1 - a(p)(\ell^{j-k} - q^{m_1}(\ell + m_1)^{j-k} - p^{m_2}(\ell + m_2)^{j-k} \right. \)
\[ + q^{m_1}p^{m_2}(\ell + m_1 + m_2)^{j-k})] \]
\[ \left. + qa(p)\left( \sum_{k=\ell+1}^{\ell+2} k^j p^{k-\ell-1} - \frac{m_1}{p} \right) \right\} \]
\[ \left. \sum_{k=\ell+1}^{\ell+2} k^j p^{k-\ell-1} \right\}. \]

The proofs of Theorems 2.7 - 2.9 follow using steps similar to the proofs of Theorems 2.1 - 2.3.
Remarks 2.1.  
(i) It is important to note that the expression \( \sum_{i=1}^{n} (pz)_{i}^{2} = \sum_{i=1}^{n} (pz)_{i} - 1 \) appears in (5) and (7), as expected, since the pattern can be completed if a failure occurs after \( \ell_{2} + 1 \) (up to \( k_{2} \)) consecutive successes. Also, with the same justification, the expressions (4) and (8) have the term \( qt/(1 - pt) \). However, (3) and (4) are in easy form as the pattern is completed just after \( \ell_{2} \) consecutive successes.

(ii) The explicit form of PGF and PMF in Theorems 2.2, 2.5 and 2.8 can also be expressed in different forms as the binomial expansion can be written \( (a + b)^{n} = \sum_{u=0}^{n} \binom{n}{u} a^{u} b^{n-u} = \sum_{u=0}^{n} \binom{n}{u} a^{n-u} b^{u} \). It is up to the end-user to choose an appropriate form and modify the results.

(iii) The results derived in Section 2, are based on Markov chain approach (see Fu and Koutras [12] and Dafnis et al. [8]). However, the results can also be derived using combinatorial method similar to Huang and Tsai [13].

(iv) It can be easily verified that for \( \ell_{1} = k_{1} \) and \( \ell_{2} = k_{2} \), Theorems 2.7 - 2.9 are same as Theorems 3.1 - 3.8 of Kumar and Upadhye [16], as expected.

(v) For \( \ell_{1} = 1 = \ell_{2} \), \( H_{1,1,k_{1},k_{2}} = X^{(3)}_{n} \) of Dafnis et al. [8] (in their notation). Also, Dafnis et al. [8] in Theorem 4.7 proved that for \( r \geq 1 \), the PGF for waiting time for \( X^{(3)}_{n} \) is given by
\[
H_{r}(z) = \left( \frac{(qz)(pz)(1 - (qz)^{k_{1}})(1 - (pz)^{k_{2}})}{1 - z + (qz)(pz)(1 - (qz)^{k_{1}})(1 - (pz)^{k_{2}})} \right)^{r} (1 - (pz)^{k_{2}})^{-1}. \tag{9}
\]
But, observe that \( H_{r}(1) = 1/(1 - p^{k_{2}}) \neq 1 \) unless \( p = 0 \). Therefore, the expression (9) is incorrect and hence Theorems 4.8 and 4.9 of Dafnis et al. [8] are also incorrect. We correct and generalize these erroneous results in Theorem 2.9.

3 An Application to Fibonacci Words

Fibonacci words are particular sequences of binary numbers 0 and 1 (or two alphabets) and it is used to model physical systems with the aperiodic order such as quasi-crystals. Also, Fibonacci word studied widely in the field of combinatorics on words. Fibonacci words are formed in a similar way as Fibonacci numbers (repeated addition) and, in this process, \( n \)-th Fibonacci word depends on \((n - 1)\)-th and \((n - 2)\)-th Fibonacci words of 0’s and 1’s. The construction can be explained as follows:

\[ C_{0} = 0 \quad \text{and} \quad C_{1} = 01 \]

then \( n \)-th Fibonacci word is given by
\[ C_{n} = C_{n-1}C_{n-2}. \]

For example, 10-th element of Fibonacci words is given by
and the random variable representation is given by

\[(1 - \zeta_1)\zeta_2(1 - \zeta_3)(1 - \zeta_4)\zeta_5(1 - \zeta_6)\zeta_7(1 - \zeta_8)(1 - \zeta_9)\zeta_{10}(1 - \zeta_{11})(1 - \zeta_{12})\zeta_{13}(1 - \zeta_{14})\zeta_{15}(1 - \zeta_{16})(1 - \zeta_{17})\zeta_{18}(1 - \zeta_{19})\zeta_{20}(1 - \zeta_{21})(1 - \zeta_{22})\zeta_{23}(1 - \zeta_{24})(1 - \zeta_{25})\zeta_{26}(1 - \zeta_{27})\zeta_{28}(1 - \zeta_{29})(1 - \zeta_{30}) \ldots \]

Also, the sub-words “11” and “000” never occur in Fibonacci words and last two digits are “01” and “10”, alternately. For more details on Fibonacci words, we refer the reader to Berstel [6]. Now, observe that Fibonacci words can be seen as a pattern of either exactly one 1 followed by (at least one) at most two consecutive 0’s followed by exactly one 1 and hence the distribution of patterns adopted the distribution of either $H_{1,1,2}$ or $H_{1,2,1,1}$ respectively, for $n$-th Fibonacci word. For large values of $n$, the probabilities and moments of the distribution of these patterns can be calculated from the distribution of either $H_{1,1,1,2}$ or $H_{1,2,1,1}$. Next, we compute some probabilities and mean for $H_{1,2,1,1}$ and its waiting time for various values of $p$ and $n = 60$.

| $n$ | $m$ | $p = 0.35$ | $p = 0.36$ | $p = 0.37$ | $p = 0.38$ | $p = 0.39$ | $p = 0.40$ |
|-----|-----|-------------|-------------|-------------|-------------|-------------|-------------|
| 0   | 0.0081259 | 0.0073285 | 0.0066661 | 0.0061179 | 0.0056670 | 0.0052998 |
| 1   | 0.0363192 | 0.0335660 | 0.0312188 | 0.0292301 | 0.0275615 | 0.0261798 |
| 2   | 0.0844787 | 0.0798366 | 0.0757692 | 0.0722423 | 0.0692234 | 0.0666826 |
| 60  | 0.1353360 | 0.1305530 | 0.1262260 | 0.1223700 | 0.1189930 | 0.1160990 |
| 4   | 0.1669740 | 0.1641700 | 0.1614830 | 0.1589750 | 0.1566960 | 0.1546850 |
| 5   | 0.1683560 | 0.1684990 | 0.1684180 | 0.1681850 | 0.1678630 | 0.1675060 |
| $E(H_{1,2,1,1})$ | 5.07803 | 5.17016 | 5.25346 | 5.32777 | 5.39297 | 5.44896 |

| $r$ | $m$ | $p = 0.45$ | $p = 0.46$ | $p = 0.47$ | $p = 0.48$ | $p = 0.49$ | $p = 0.50$ |
|-----|-----|-------------|-------------|-------------|-------------|-------------|-------------|
| 3   | 0.1361250 | 0.1341360 | 0.1320230 | 0.1297920 | 0.1274990 | 0.1250000 |
| 4   | 0.1361250 | 0.1341360 | 0.1320230 | 0.1297920 | 0.1274990 | 0.1250000 |
| 5   | 0.0612563 | 0.0617026 | 0.0620508 | 0.0623002 | 0.0624500 | 0.0625000 |
| 6   | 0.0427262 | 0.0437101 | 0.0446207 | 0.0454542 | 0.0462068 | 0.0468750 |
| 7   | 0.0529177 | 0.0534260 | 0.0538587 | 0.0542141 | 0.0549408 | 0.0546875 |
| 8   | 0.0547707 | 0.0548654 | 0.0549045 | 0.0548879 | 0.0548157 | 0.0546875 |
| 9   | 0.0464322 | 0.0465889 | 0.0467123 | 0.0468019 | 0.0465865 | 0.0468750 |
| 10  | 0.0399053 | 0.0401752 | 0.0404228 | 0.0406466 | 0.0408449 | 0.0410156 |
| $E(\rho_1)$ | 2.17153 | 2.31385 | 2.45255 | 2.58869 | 2.72324 | 2.85714 |

Observe that the upper range of $m$ is $\lfloor n/\ell \rfloor = \lfloor 60/2 \rfloor = 30$, while we obtain the probabilities up to $m = 5$ and others can be computed in a similar way. Also, for waiting time distribution, it is known that $m \geq \ell r + 1 = 3$. So, we obtain probabilities by taking $m$ up to 10 in Table 2. Moment for $H_{1,2,1,1}$ and $\rho_1$ are obtained in Table 1 and Table 2 respectively.
References

[1] Aki, S. (1997). On sooner and later problems between success and failure runs. Advances in Combinatorial Methods and Applications to Probability and Statistics (Ed., N. Balakrishnan), 385-400, Borkhäuser, Boston.

[2] Aki, S., Kuboki, H. and Hirano, K. (1984). On discrete distributions of order $k$. Ann. Inst. Statist. Math., 36, 431-440.

[3] Antzoulakos, D. L., Bersimis, S., Koutras, M. V. (2003). Waiting times associated with the sum of success run lengths. In: Lindqvist, B., Doksum, K. (Eds.), Mathematical and Statistical Methods in Reliability, World Scientific, Singapore, 141-157.

[4] Antzoulakos, D. L. and Chadjiconstantinidis, S. (2001). Distributions of numbers of success runs of fixed length in Markov dependent trials. Ann. Inst. Statist. Math., 53, 599-619.

[5] Balakrishnan, N. and Koutras, M. V. (2002). Runs and Scans with Applications, John Wiley & Sons, New York.

[6] Berstel, J. (1986). Fibonacci words - a survey, in: Rozenberg, G., Salomaa, A. (Eds.), The Book of L, Springer, Berlin.

[7] Balakrishnan, N., Mohanty, S. G. and Aki, S. (1997). Start-up demonstration tests under Markov dependence model with corrective actions. Ann. Inst. Statist. Math., 49, 155-169.

[8] Dafnis, S. D., Antzoulakos, D. L. and Philippou, A. N. (2010). Distribution related to $(k_1, k_2)$ events. J. Stat. Plan. Inference, 140, 1691-1700.

[9] Fu, J. C. (1986). Reliability of consecutive-$k$-out-of-$n$:F system with $(k-1)$-step Markov dependence, IEEE Trans. Reliability, 35, 602-606.

[10] Fu, J. C., Lou, W. Y. W. and Chen, S. C. (1999). On the probability of pattern matching in nonaligned DNA sequences: a finite Markov chain imbedding approach, In Scan Statistics and Applications (Eds., J. Glaz and N. Balakrishnan), 287-302, Birkhäuser, Boston.

[11] Fu, J. C. and Hu, B. (1987). On reliability of a large consecutive-$k$-out-of-$n$:F system with $(k-1)$-step Markov dependence, IEEE Trans. Reliability, 36, 75-77.

[12] Fu, J. C. and Koutras, M. V. (1994). Distribution theory of runs: a Markov chain approach, J. Amer. Statist. Assoc., 89, 1050-1058.

[13] Huang, W. T. and Tsai, C. S. (1991). On a modified binomial distribution of order $k$. Statist. Prob. Lett., 11, 125-131.

[14] Koutras M.V. (1996). On a waiting time distribution in a sequence of Bernoulli trials, Ann. Inst. Statist. Math., 48, 789-806.

[15] Koutras, M. V. (1997). Waiting time distributions associated with runs of fixed length in two-state Markov chains, Ann. Inst. Statist. Math., 49, 123-139.

[16] Kumar, A. N. and Upadhye, N. S. (2017). On Discrete Gibbs Measure Approximation to Runs. arXiv preprint, arXiv:1701.08294.

[17] Makri, F. S., Philippou, A. N. and Psillakis, Z. M. (2007). Shortest and longest length of success runs in binary sequences. J. Statist. Plann. Inference, 137, 2220-2239.

[18] Moore, P. T. (1958). Some properties of runs in quality control procedures, Biometrika, 45, 89-95.

[19] Philippou, A. N., Georgiou, C. and Philippou, G. N. (1983). A generalized distribution and some of its properties. Statist. Prob. Lett., 1, 171-175.

[20] Philippou, A. N. and Makri, A. (1986). Success, runs and longest runs. Statist. Prob. Lett., 4, 211-215.

[21] Sinha, K., Sinha, B. P. and Datta, D. (2010). CNS: a new energy efficient transmission scheme for wireless sensor networks, Wireless Networks Journal (ACM / Springer), 16, 2087-2104.