Some explicit formulae for the distributions of words *

Hayato Takahashi (Random Data Lab. Inc.)†

1 Introduction

The distributions of the number of the appearances of words (distributions of words for short) play important role in statistics, DNA analysis, information theory, see Balakrishnan et.al [1], Jacquet et.al [13], Lothire et.al [15], Robin et.al [21], Wald et.al [25], Waterman [26], and Zehavi et.al [27].

Generating functions of the distributions of words are given as rational functions, see Bassino et.al [2], Berthe et.al [3], Blom et.al [4], Chrysaphinou et.al [5], Feller [6], Flajolet et.al [7], Goulden et.al [10], Guibas et.al [11], and Régnier et.al [20]. From generating functions, we have approximations and recurrence formulae for the distributions of words. However except for simple cases, we neither expand rational functions into power series nor obtain their coefficients by differentiation, see Chapter 11 Section 4 pp. 275 Feller [6]. In other words, we cannot obtain explicit formulae for the distributions of words from rational generating functions in general.

In this article we show explicit formulae for 1. the joint distributions of nonoverlapping words for independent and identically distributed (i.i.d.) finite alphabet random variables and 2. the distributions of runs for i.i.d. binary random variables.

2 Joint distributions of nonoverlapping words

Let $\mathbf{N}(w_1,\ldots,w_l;X_1^n)$ be the number of the appearances of the words $w_1,\ldots,w_l$ in an arbitrary position of X_1^n, i.e.

$$\mathbf{N}(w_1,\ldots,w_l;X_1^n) := \Big(\sum_{i=1}^{n-|w_1|+1} I_{w_1}(X_i^n),\ldots,\sum_{i=1}^{n-|w_l|+1} I_{w_l}(X_i^n)\Big),$$

where $X_i^n = X_i\cdots X_n$ and $I_{w_j}(X_i^n)=1$ if $X_i\cdots X_{i+|w_j|-1}=w_j$ else 0 for all i,j.

For example $N(10,11;1011101)=(2,2)$. A word x is called overlapping if there is a word z such that x appears at least 2 times in z and $|z|<2|x|$ otherwise x is called nonoverlapping. A pair of words x,y is called overlapping if there is a word z such that x and y appear in z and $|z|<|x|+|y|$. A finite set of words S is called nonoverlapping if every pair (x,y) for $x,y\in S$ are not overlapping, otherwise, S is called overlapping. For example, sets of words, $\{11\}$, $\{10,01\}$, and $\{00,11\}$ are overlapping, and $\{10\}$ and $\{00111,00101\}$ are nonoverlapping.

*Parts of the paper have been presented in [23,24].

†Email: hayato.takahashi@ieee.org

Theorem 2.1 *Let $X_1 X_2 \cdots X_n$ be i.i.d. finite alphabet random variables. Let $w_1, \ldots, w_l$ be the set of nonoverlapping words. Let $m_i = |w_i|$ be the length of w_i and $P(w_i)$ the probability of w_i for $i = 1, \ldots, l$. Let*

$$
\begin{aligned}
A(k_1, \ldots, k_l) &= \binom{n - \sum_i m_i k_i + \sum_i k_i}{k_1, \ldots, k_l} \prod_{i=1}^{l} P^{k_i}(w_i), \\
B(k_1, \ldots, k_l) &= P(\sum_{i=1}^{n} I_{X_i^{i+m_i-1} = w_j} = k_j, \ j = 1, \ldots, l), \\
F_A(z_1, \ldots, z_l) &= \sum_{k_1, \ldots, k_l} A(k_1, \ldots, k_l) z^{k_1} \cdots z^{k_l}, \text{ and} \\
F_B(z_1, \ldots, z_l) &= \sum_{k_1, \ldots, k_l} B(k_1, \ldots, k_l) z^{k_1} \cdots z^{k_l}.
\end{aligned}
\tag{1}
$$

Then

$$F_A(z_1, z_2, \ldots, z_l) = F_B(z_1 + 1, z_2 + 1, \ldots, z_l + 1),$$

and

$$
\begin{aligned}
& P(N(w_1, \ldots, w_l; X^n) = (s_1, \ldots, s_l)) \\
&= \sum_{\substack{k_1, \ldots, k_l : \\ s_1 \le k_1, \ldots, s_l \le k_l \\ \sum_i m_i k_i \le n}} (-1)^{\sum_i k_i - s_i} \binom{n - \sum_i m_i k_i + \sum_i k_i}{s_1, \ldots, s_l, k_1 - s_1, \ldots k_l - s_l} \prod_{i=1}^{l} P^{k_i}(w_i).
\end{aligned}
\tag{2}
$$

Proof) For simplicity, we prove the theorem for $l = 1$. The proof of the general case is similar. Let $m = |w|$. Since w is nonoverlapping, the number of possible allocations such that w appears k-times in the string of length n is

$$\binom{n - mk + k}{k}.$$

This is because if we replace each w with additional extra symbol α in the string of length n then the problem reduces to choosing k α's among the string of length $n - mk + k$. Let

$$A(k) := \binom{n - mk + k}{k} P^k(w). \tag{3}$$

The function A is not the probability of k w's occurrences in the string, since we allow any letters in the remaining place except for the appearance of w. The function A may count the event that w appears more than k times. Let $B(t)$ be the probability that w appears k times. We have the following identity,

$$A(k) = \sum_{k \le t} B(t) \binom{t}{k}.$$

Let $F_A(z) := \sum_k A(k) z^k$ and $F_B(z) := \sum_k B(k) z^k$. Then

$$F_A(z) = \sum_k z^k \sum_{k \le t} B(t) \binom{t}{k}$$

$$
\begin{aligned}
&= \sum_t B(t) \sum_{k \le t} \binom{t}{k} z^k \\
&= \sum_t B(t)(z+1)^t \\
&= F_B(z+1).
\end{aligned}
$$

We have

$$
\begin{aligned}
F_B(z) &= F_A(z-1) \\
&= \sum_{k:\, mk \le n} \binom{n-mk+k}{k} (z-1)^k P^k(w) \\
&= \sum_{\substack{k,t:\, mk \le n \\ t \le k}} \binom{n-mk+k}{k} \binom{k}{t} z^t (-1)^{k-t} P^k(w) \\
&= \sum_t z^t \sum_{\substack{k:\, mk \le n \\ t \le k}} (-1)^{k-t} \binom{n-mk+k}{t, k-t} P^k(w),
\end{aligned}
$$

and (2) . ■

For the moments of the distributions of nonoverlapping word and the distributions of partial nonoverlapping words, see [22].

3 Runs

Words that consists of the same letter are called run. For example 111 and 00 are runs. In the following, we consider the distributions of runs of 0s for independent and identically distributed (i.i.d.) binary trials.

Let n be the sample size. Fu et.al [8] showed the distributions of the following five statistics of runs by Markov imbedding method.
For $x \in \{0,1\}^n$, let
(i) $E_{n,m}(x)$, the number of 0^m of size exactly m in x (Mood [17]),
(ii) $G_{n,m}(x)$, the number of 0^m of size greater than or equal to m in x (Makri et.al [16]),
(iii) $N_{n,m}(x)$, the number of nonoverlapping 0^m in x (Feller [6], Godbole [9], Hirano [12], Muselli [18], and Phillipou et.al [19]),
(iv) $M_{n,m}(x)$, the number of overlapping 0^m in x (Ling [14]), and
(v) $L_n(x)$, the size of the longest run of 0s in x (Makri et.al [16]).

For example, consider a run 00 in $x = 0010000100$. Then $n = 10, m = 2$ and $E_{10,2}(x) = 2$, $G_{10,2}(x) = 3$, $N_{10,2}(x) = 4$, $M_{10,2}(x) = 5$, and $L_{10}(x) = 4$.

An explicit formula for the distribution of L_n is given by that of G and

$$P(L_n = t) = P(G_{n,t+1} = 0) - P(G_{n,t} = 0),$$

see [8]. For other studies on runs see [1] and the references therein. In particular, explicit formulae for the distributions of $E_{n,m}(x)$ were not known before except for those given by Markov imbedding method [8]. In this article, we show new simple explicit formulae for the distributions of statistics (i)–(iv) by a unified manner.

3.1 Explicit formulae for the distributions of runs

Let $\{0,1\}^*$ be the set of finite binary strings and λ the empty word. Let $\bar{x} = 1w$ for $x = 0^t 1w$ where $w \in \{0,1\}^*$ and t is a non-negative integer. If $x = 0^n$ for some n then $\bar{x} = \lambda$. For $x \in \{0,1\}^n$, define $\bar{E}_{n,m}(x) := E_{|\bar{x}|,m}(\bar{x})$, $\bar{G}_{n,m}(x) := G_{|\bar{x}|,m}(\bar{x})$, $\bar{N}_{n,m}(x) := N_{|\bar{x}|,m}(\bar{x})$, and $\bar{M}_{n,m}(x) := M_{|\bar{x}|,m}(\bar{x})$. For example, $\bar{x} = 10000100$ if $x = 0010000100$ and $\bar{E}_{10,2}(x) = 1$, $\bar{G}_{10,2}(x) = 2$, $\bar{N}_{10,2}(x) = 3$, and $\bar{M}_{10,2}(x) = 4$.

To prove Theorem 3.1, we first enumerate $\bar{E}, \bar{G}, \bar{N}$, and $\bar{M}$ by inclusion-exclusion principles (Lemma 3.2) then we enumerate runs E, G, N, and M (Lemma 3.3).

Theorem 3.1 *Let $X_1, X_2, \ldots,$ be i.i.d. binary random variables from $P(X_i = 1) = P(1)$ and $P(X_i = 0) = P(0)$ for all i. Let $X_1^n = X_1 \cdots X_n$ for all n. Then for all t,*

(i)

$$P(\bar{E}_{n,m}(X_1^n) = t) = \sum_{\substack{k_1,k_2:\\ (m+1)k_1+(m+2)k_2 \le n,\\ t \le k_1+k_2}} (-1)^{k_1-t} \binom{n-(m+1)k_1-(m+2)k_2+k_1+k_2}{k_1, k_2} \times \binom{k_1+k_2}{t} P^{k_1}(10^m) P^{k_2}(10^{m+1}) \text{ and}$$

$$P(E_{n,m}(X_1^n) = t) = (P(\bar{E}_{n+1,m}(X_1^{n+1}) = t) - P(0)P(\bar{E}_{n,m}(X_1^n) = t))/P(1),$$

(ii) $$P(\bar{G}_{n,m}(X_1^n) = t) = \sum_{k:\ t \le k, (m+1)k \le n} (-1)^{k-t} \binom{n-(m+1)k+k}{t, k-t} P^k(10^m) \text{ and}$$

$$P(G_{n,m}(X_1^n) = t) = (P(\bar{G}_{n+1,m}(X_1^{n+1}) = t) - P(0)P(\bar{G}_{n,m}(X_1^n) = t))/P(1),$$

(iii) Let T be the maximum integer such that $Tm + 1 \le n$. Then

$$P(\bar{N}_{n,m}(X_1^n) = t) = \sum_{\substack{r,k_1,\ldots,k_T:\\ \sum_i (mi+1)k_i \le n,\ 0 \le r \le \sum_i k_i\\ t = \sum_i ik_i - r}} (-1)^r \binom{n-\sum_i (mi+1)k_i + \sum_i k_i}{k_1, \ldots, k_{n-m}} \binom{\sum_i k_i}{r} \times \prod_{i=1}^{T} P^{k_i}(10^{im}) \text{ and}$$

$$P(N_{n,m}(X_1^n) = t) = (P(\bar{N}_{n+1,m}(X_1^{n+1}) = t) - P(0)P(\bar{N}_{n,m}(X_1^n) = t))P^{-1}(1),$$

(iv)

$$P(\bar{M}_{n,m}(X_1^n)=t)=\sum_{\substack{r,k_1,\ldots,k_{n-m}:\\ \sum_i(m+i)k_i\le n,\ 0\le r\le\sum_i k_i\\ t=\sum ik_i-r}}(-1)^r\binom{n-\sum_i(m+i)k_i+\sum_i k_i}{k_1,\ldots,k_{n-m}}\binom{\sum_i k_i}{r}$$
$$\times\prod_{i=1}^{n-m}P^{k_i}(10^{m+i-1})\ \textit{and}$$

$$P(M_{n,m}(X_1^n)=t)=(P(\bar{M}_{n+1,m}(X_1^{n+1})=t)-P(0)P(\bar{M}_{n,m}(X_1^n)=t))P^{-1}(1).$$

To prove the theorem, we need some definitions and lemmas.

Let
$$\mathbf{N}'(w_1,\ldots,w_l;X_1^n):=(s_1-s_2,s_2-s_3,\ldots,s_l)$$
where $\mathbf{N}(w_1,\ldots,w_l;X_1^n)=(s_1,\ldots,s_l)$. For example $\mathbf{N}(100,1000;1010001)=(1,1)$ and $\mathbf{N}'(100,1000;1010001)=(0,1)$. Note that if w_1 is a prefix of w_2 and $(k_1,k_2)=\mathbf{N}(w_1,w_2;X_1^n)$ then $k_1\ge k_2$.

Lemma 3.2 *Let $X_1,X_2,\ldots,$ be i.i.d. binary random variables from $P(X_i=1)=P(1)$ and $P(X_i=0)=P(0)$ for all i. Let $w_1\sqsubset w_2\cdots\sqsubset w_l$ be an increasing sequence of nonoverlapping words. Let*

$$A(k_1,\ldots,k_l):=\binom{n-\sum_i m_ik_i+\sum_i k_i}{k_1,\ldots,k_l}\prod_{i=1}^{l}P^{k_i}(w_i),$$
$$B(k_1,\ldots,k_l):=P(\mathbf{N}'(w_1,\ldots,w_l;X_1^n)=(k_1,k_2,\ldots,k_l)),$$
$$F_A(z_1,\ldots,k_l):=\sum_{\substack{k_1,\ldots,k_l:\\ \sum_i m_ik_i\le n}}A(k_1,\ldots,k_l)z^{k_1}\cdots z^{k_l},\ \textit{and}$$
$$F_B(z_1,\ldots,z_l):=\sum_{\substack{k_1,\ldots,k_l:\\ \sum_i m_ik_i\le n}}B(k_1,\ldots,k_l)z^{k_1}\cdots z^{k_l}.$$

Then
$$F_A(z_1,\ldots,z_l)=F_B(z_1+1,z_1+z_2+1,\ldots,\sum_i z_i+1)\ \textit{and}^{1} \tag{4}$$
$$F_A(Y-1,(Y-1)Y,\ldots,(Y-1)Y^{l-1})=F_B(Y,Y^2,\ldots,Y^l). \tag{5}$$

Proof) We show (4) for $l=2$. The proof of the general case is similar. Observe that

$$A(k_1,k_2)=\sum_{k_2\le t_2,\ k_1+k_2\le t_1+t_2}B(t_1,t_2)\binom{t_2}{k_2}\sum_{0\le s\le t_2-k_2}\binom{t_2-k_2}{s}\binom{t_1}{k_1-s}. \tag{6}$$

[1] (4) is presented at Mathematical Society Japan, Okayama 2018.

Then

$$\begin{aligned}
F_A(z_1,z_2) =& \sum_{k_1,k_2} z_1^{k_1} z_2^{k_2} \sum_{k_2\le t_2,\ k_1+k_2\le t_1+t_2} B(t_1,t_2)\binom{t_2}{k_2} \sum_{0\le s\le t_2-k_2} \binom{t_2-k_2}{s}\binom{t_1}{k_1-s} \\
=& \sum_{t_1,t_2} B(t_1,t_2) \sum_{k_2\le t_2} \binom{t_2}{k_2} z_2^{k_2} \sum_{0\le s\le t_2-k_2,\ 0\le k_1-s\le t_1} \binom{t_2-k_2}{s}\binom{t_1}{k_1-s} z_1^{k_1} \\
=& \sum_{t_1,t_2} B(t_1,t_2) \sum_{k_2\le t_2} \binom{t_2}{k_2} z_2^{k_2}(z_1+1)^{t_1+t_2-k_2} \\
=& \sum_{t_1,t_2} B(t_1,t_2)(z_1+1)^{t_1+t_2}(\frac{z_2}{z_1+1}+1)^{t_2} \\
=& F_B(z_1+1, z_1+z_2+1).
\end{aligned}$$

Next set $z_1 = X, z_2 = X(X+1), \ldots, z_l = X(X+1)^{l-1}$ in (4). Then

$$F_A(X, X(X+1), \ldots, X(X+1)^{l-1}) = F_B(X+1, (X+1)^2, \ldots, (X+1)^l). \tag{7}$$

By setting $Y = X+1$ in (7), we have (5). ■

Lemma 3.3 *Let*

$$E_{n,m,t} = \{x \in \{0,1\}^n \mid E_{n,m}(x) = t\} \text{ and } \bar{E}_{n,m,t} = \{x \in \{0,1\}^n \mid \bar{E}_{n,m}(x) = t\}.$$

Then

$$P(\bar{E}_{n+1,m,t}) = P(0)P(\bar{E}_{n,m,t}) + P(1)P(E_{n,m,t}). \tag{8}$$

The sets $(G_{n,m,t}, \bar{G}_{n,m,t})$, $(N_{n,m,t}, \bar{N}_{n,m,t})$, and $(M_{n,m,t}, \bar{M}_{n,m,t})$ are defined by similar manner and (8) is true for them respectively.

Proof) Let $\bar{E}^0_{n+1,m,t} = \{0x \in \{0,1\}^{n+1} \mid \bar{E}_{n+1,m}(0x) = t\}$ and $\bar{E}^1_{n+1,m,t} := \{1x \in \{0,1\}^{n+1} \mid \bar{E}_{n+1,m}(1x) = t\}$. Then

$$\bar{E}^0_{n+1,m,t} = \{0x \in \{0,1\}^{n+1} \mid x \in \bar{E}_{n,m,t}\} \text{ and } \bar{E}^1_{n+1,m,t} = \{1x \in \{0,1\}^{n+1} \mid x \in E_{n,m,t}\}.$$

Since $\bar{E}_{n+1,m,t} = \bar{E}^0_{n+1,m,t} \cup \bar{E}^1_{n+1,m,t}$, we have

$$P(\bar{E}_{n+1,m,t}) = P(\bar{E}^0_{n+1,m,t}) + P(\bar{E}^1_{n+1,m,t}) = P(0)P(\bar{E}_{n,m,t}) + P(1)P(E_{n,m,t}).$$

The proof of the latter part is similar. ■

Proof of Theorem 3.1 (i). Let $l = 2$, $w_1 = 10^m$, and $w_2 = 10^{m+1}$ in Lemma 3.2. By (4), we have

$$F_A(z_1,z_2) = F_B(z_1+1, z_1+z_2+1). \tag{9}$$

Set $z_1 = x-1$ and $z_2 = 1-x$ in (9). We have

$$\begin{aligned}
F_A(x-1, 1-x) =& F_B(x,1) \\
=& \sum_{k_1,k_2:(m+1)k_1+(m+2)k_2\le n} P(\mathbf{N}'(w_1,w_2) = (k_1,k_2))x^{k_1}
\end{aligned}$$

$$= \sum_{k_1} \sum_{k_2:(m+1)k_1+(m+2)k_2 \le n} P(\mathbf{N}'(w_1, w_2) = (k_1, k_2))x^{k_1}$$
$$= \sum_{k_1} P(\bar{E}_{n,m} = k_1)x^{k_1}. \tag{10}$$

On the other hand,

$$F_A(x-1, 1-x) = \sum_{\substack{k_1,k_2: \\ (m+1)k_1+(m+2)k_2 \le n}} \binom{n-(m+1)k_1-(m+2)k_2+k_1+k_2}{k_1, k_2} P^{k_1}(w_1)P^{k_2}(w_2) \times (x-1)^{k_1}(1-x)^{k_2}$$

$$= \sum_{\substack{k_1,k_2: \\ (m+1)k_1+(m+2)k_2 \le n}} (-1)^{k_2} \binom{n-(m+1)k_1-(m+2)k_2+k_1+k_2}{k_1, k_2} P^{k_1}(w_1)P^{k_2}(w_2) \times (x-1)^{k_1+k_2}$$

$$= \sum_{\substack{k_1,k_2,t: \\ (m+1)k_1+(m+2)k_2 \le n \\ t \le k_1+k_2}} (-1)^{k_1+2k_2-t} \binom{n-(m+1)k_1-(m;2)k_2+k_1+k_2}{k_1, k_2} \binom{k_1+k_2}{t} \times P^{k_1}(w_1)P^{k_2}(w_2)x^t. \tag{11}$$

By (10) and (11), we have the first part of (i). The latter part of (i) follows from Lemma 3.3.
Proof of Theorem 3.1 (ii). Let $l = 1$, $w_1 = 10^m$ in Lemma 3.2. Then $F_A(z) = F_B(z+1)$.

$$F_B(z) = F_A(z-1)$$
$$= \sum_{k:(m+1)k \le n} \binom{n-(m+1)k+k}{k} P^k(w)(z-1)^k$$
$$= \sum_{k,t:(m+1)k \le n, t \le k} (-1)^{k-t} \binom{n-(m+1)k+k}{k} \binom{k}{t} P^k(w)z^t$$
$$= \sum_{k,t:(m+1)k \le n, t \le k} (-1)^{k-t} \binom{n-(m+1)k+k}{t, k-t} P^k(w)z^t.$$

On the other hand, $F_B(z) = \sum_k P(\bar{G}_{n,m} = k)z^k$ and we have the first part of (ii). The latter part of (ii) follows from Lemma 3.3.
Proof of Theorem 3.1 (iii). Let $w_1 = 10^m, w_2 = 10^{2m}, \ldots, w_T = 10^{Tm}$ where T is the maximum integer such that $|w_T| = Tm + 1 \le n$ in Lemma 3.2. Since

$$F_B(Y, Y^2, \ldots, Y^T) = \sum_{\substack{k_1,\ldots,k_T: \\ \sum_i (mi+1)k_i \le n}} B(k_1, \ldots, k_T)Y^{\sum ik_i},$$

$P(\bar{N}_{n,m}=t)=P(\sum ik_i=t)$ is the coefficient of Y^t in F_B. On the other hand, by expanding the left-hand-side of (5), we have

$$\begin{aligned}
&F_A(Y-1,(Y-1)Y,\ldots,(Y-1)Y^{l-1})\\
&=\sum_{k_1,\ldots,k_l}\binom{n-\sum|w_i|k_i+\sum k_i}{k_1,\ldots,k_l}(Y-1)^{\sum k_i}\prod Y^{(i-1)k_i}P^{k_i}(w_i)\\
&=\sum_{k_1,\ldots,k_l}\binom{n-\sum|w_i|k_i+\sum k_i}{k_1,\ldots,k_l}\prod P^{k_i}(w_i)\sum_r\binom{\sum k_i}{r}(-1)^rY^{\sum ik_i-r}. \qquad (12)
\end{aligned}$$

By setting $l=T$ and $|w_i|=mi+1$ for $i=1,\ldots,T$ in (12), we have the first part of (iii). The latter part of (iii) follows from Lemma 3.3.
Proof of Theorem 3.1 (iv). Let $w_1=10^m, w_2=10^{m+1},\ldots,w_{n-m}=10^{n-1}$ in Lemma 3.2. Since

$$F_B(Y,Y^2,\ldots,Y^{n-m})=\sum_{\substack{k_1,\ldots,k_{n-m}:\\ \sum_i(m+i)k_i\le n}}B(k_1,\ldots,k_{n-m})Y^{\sum ik_i},$$

$P(\bar{M}_{n,m}=t)=P(\sum ik_i=t)$ is the coefficient of Y^t in F_B. By setting $l=n-m$ and $|w_i|=m+i$ for $i=1,\ldots,n-m$ in (12), we have the first part of (iv). The latter part of (iv) follows from Lemma 3.3. ■

Remark 3.4 In theorem 3.1 (ii), $P(\bar{G}_{n,m}=t)$ is an explicit formula for the distribution of nonoverlapping word 10^m, which is a special case given in [22].

Remark 3.5 It is straightforward to extend Theorem 3.1 to i.i.d. random variables that take infinitely many values. Let p_j, $j=0,1,\ldots$ be a sequence of non-negative reals such that $\sum_j p_j=1$. Let $Y_1,Y_2,\ldots Y_n\in\{0,1,2,\ldots\}$ be i.i.d. trials from $Q(Y_i=j)=p_j$ for all i,j. Then the distributions of runs of zeros for infinitely many alphabets are given by $Q(E_{n,m}(Y_1^n)=t)=P(E_{n,m}(X_1^n)=t)$, $Q(G_{n,m}(Y_1^n)=t)=P(G_{n,m}(X_1^n)=t)$, $Q(N_{n,m}(Y_1^n)=t)=P(N_{n,m}(X_1^n)=t)$, and $Q(M_{n,m}(Y_1^n)=t)=P(M_{n,m}(X_1^n)=t)$ for all t, where $X_1,\ldots,X_n$ are binary i.i.d. trials with $P(X_i=1)=1-p_0$ and$P(X_i=0)=p_0$ for all i and $P(E_{n,m})$, $P(G_{n,m})$, $P(N_{n,m})$, and $P(M_{n,m})$ are given by Theorem 3.1 with P.

Acknowledgement

This work was supported by the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University. The author thanks Prof. Shigeki Akiyama (Tsukuba Univ.) for discussions.

References

[1] N. Balakrishnan and M. V. Koutras. *Runs and scans with applications*. John Wiley & Sons, 2002.

[2] F. Bassino, J. Clément, and P. Micodème. Counting occurrences for a finite set of words: combinatorial methods. *ACM Trans. Algorithms.*, 9(4):Article No. 31, 2010.

[3] V. Berthé and M. Rigo. *Combinatorics, words and symbolic dynamics*. Encyclopedia of Mathematics and Its Applications 159. Cambridge University Press, 2016.

[4] G. Blom and D. Thorburn. How many random digits are required until given sequences are obtained? *J. Appl. Probab.*, 19(3):518–531, 1982.

[5] O. Chrysaphinou and S. Papastavridis. A limit theorem on the number of overlapping appearances of a pattern in a sequence of independent trials. *Probab. Theory Related Fields*, 79:129–143, 1988.

[6] W. Feller. *An Introduction to probability theory and its applications Vol. 1*. Wiley, 3rd revised edition, 1970.

[7] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[8] J. C. Fu and M. V. Koutras. Distribution theory of runs: a Markov chain approach. *J. Amer. Statist. Assoc.*, 89(427):1050–1058, 1994.

[9] A. P. Godbole. Specific formulae for some success run distributions. *Statist. Probab. Lett.*, 10:119–124, 1990.

[10] I. Goulden and D. Jackson. *Combinatorial Enumeration*. John Wiley, 1983.

[11] L. Guibas and A. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, 30:183–208, 1981.

[12] K. Hirano. Some properties of the distributions of order k. pages 43–53, 1986. Fibonacci Numbers and their Applications, A. N. Phillipou, A. F. Horadam and G. E. Bergum eds, Reidel.

[13] P. Jacquet and W. Szpankowski. *Analytic Pattern Matching*. Cambridge University Press, 2015.

[14] K. D. Ling. On binomial distributions of order k. *Statist. Probab. Letters*, 6:247–250, 1988.

[15] M. Lothaire. *Applied Combinatorics on words*. Encyclopedia of Mathematics and Its Applications 105. Cambridge University Press, 2005.

[16] F. S. Makri, A. N. Philippou, and Z. M. Psillakis. Shortest and longest length of success runs in binary sequences. *J. Statist. Plan. Inference*, 137:2226–2239, 2007.

[17] A. M. Mood. The distribution theory of runs. *Ann. Math. Statist*, 11(4):367–392, 1940.

[18] M. Muselli. Simple expressions for success run distributions in Bernoulli trials. *Statist. Probab. Lett.*, 31:121–128, 1996.

[19] A. N. Phillipou and F. S. Makri. Successes, runs and longest runs. *Statist. Probab. Lett.*, 4:211–215, 1986.

[20] M. Régnier and W. Szpankowski. On pattern frequency occurrences in a Markovian sequence. *Algorithmica*, 22(4):631–649, 1998.

[21] S. Robin, F. Rodolphe, and S. Schbath. *DNA, words and models*. Cambridge University Press, english edition, 2005.

[22] H. Takahashi. The explicit formulae for the distributions of nonoverlapping words and its applications to statistical tests for pseudo random numbers. Arxiv 2105.05172.

[23] H. Takahashi. The explicit formula for the distributions of nonoverlapping words. *IEICE Technical Report IT2021-123*, 121(428):234–236, Mar 2022.

[24] H. Takahashi. Explicit formula for the distributions of runs. *IEICE Technical Report IT2022-65*, 122(355):208–210, Jan 2023.

[25] A. Wald and J. Wolfowitz. On a test whether two samples are from the same population. *Ann. Math. Statist*, 11(2):147–162, 1940.

[26] M. S. Waterman. *Introduction to computational biology*. Chapman & Hall, New York, 1995.

[27] E. Z. Zehavi and J. K. Wolf. On runlength codes. *IEEE Trans. Inform. Theory*, 34(1):45–53, 1988.