

## JOINT DISTRIBUTIONS OF COUNTS OF STRINGS IN FINITE BERNOULLI SEQUENCES

FRED W. HUFFER \* \*\* AND

JAYARAM SETHURAMAN,\* *Florida State University*

### Abstract

An infinite sequence  $(Y_1, Y_2, \dots)$  of independent Bernoulli random variables with  $P(Y_i = 1) = a/(a + b + i - 1)$ ,  $i = 1, 2, \dots$ , where  $a > 0$  and  $b \geq 0$ , will be called a  $\text{Bern}(a, b)$  sequence. Consider the counts  $Z_1, Z_2, Z_3, \dots$  of occurrences of patterns or strings of the form  $\{11\}, \{101\}, \{1001\}, \dots$ , respectively, in this sequence. The joint distribution of the counts  $Z_1, Z_2, \dots$  in the infinite  $\text{Bern}(a, b)$  sequence has been studied extensively. The counts from the initial finite sequence  $(Y_1, Y_2, \dots, Y_n)$  have been studied by Holst (2007), (2008b), who obtained the joint factorial moments for  $\text{Bern}(a, 0)$  and the factorial moments of  $Z_1$ , the count of the string  $\{1, 1\}$ , for a general  $\text{Bern}(a, b)$ . We consider stopping the Bernoulli sequence at a random time and describe the joint distribution of counts, which extends Holst's results. We show that the joint distribution of counts from a class of randomly stopped Bernoulli sequences possesses the *mixture of independent Poissons* property: there is a random vector conditioned on which the counts are independent Poissons. To obtain these results, we extend the *conditional marked Poisson process* technique introduced in Huffer, Sethuraman and Sethuraman (2009). Our results avoid previous combinatorial and induction methods which generally only yield factorial moments.

*Keywords:* Conditional marked Poisson process; Bernoulli sequence; counts of strings; random permutation; cycles; flaws and failures

2010 Mathematics Subject Classification: Primary 60C05

Secondary 60K99

### 1. Introduction

In many areas of study, we observe the occurrence or nonoccurrence of some special event over time or space. Thus, we may be observing the occurrence of record flows in a river, or flaws in a material being produced. In species tracking we want to know if the new catch is from a new species or not. In nonparametric Bayesian methods we want to know if the current observation is different from previous observations or not. In a random permutation  $\{\pi_1, \pi_2, \dots, \pi_N\}$  of  $\{1, 2, \dots, N\}$ , we would like to know if a 'cycle' ended with  $\pi_n$ ,  $n = 1, 2, \dots, N$ . Such observations can be represented by a Bernoulli sequence  $Y = (Y_1, Y_2, \dots)$ , where  $Y_n = 1$  or  $Y_n = 0$  stands for the occurrence or nonoccurrence of the event being observed.

We wish to study the distributions of counts of certain patterns of occurrence of such special events which have interesting interpretations. For instance, the count of the pattern  $\{1, 1\}$  in a random permutation is the number of cycles of length 1, and the count for  $\{1, 0, 1\}$  is the number of cycles of length 2, and so on. Similarly, each such pattern in the occurrence of flaws

Received 1 December 2011; revision received 30 January 2012.

\* Postal address: Department of Statistics, Florida State University, Tallahassee, FL 32306, USA.

\*\* Email address: huffer@stat.fsu.edu

will represent failures of a special type. In general, for  $k \geq 1$ , the pattern

$$\{1, \underbrace{0, 0, \dots, 0}_{k-1}, 1\}$$

will be called a string of order  $k$ . For a Bernoulli sequence  $\mathbf{Y}$ , let the count  $Z_k$ ,  $k \geq 1$ , be the number of strings of order  $k$  defined by

$$Z_k = \sum_{i=1}^{\infty} \mathbf{1}\{Y_i = Y_{i+k} = 1, Y_j = 0 \text{ for } i < j < i+k\}.$$

The purpose of this paper is to study the joint distribution of the count vector  $\mathbf{Z} = (Z_1, Z_2, \dots)$ . This has been the topic of several papers and we will summarize some of the results.

A well-studied Bernoulli sequence  $\mathbf{Y}$  is  $\text{Bern}(a, b)$ , which consists of independent random variables  $Y_1, Y_2, \dots$  with

$$P(Y_n = 1) = 1 - P(Y_n = 0) = \frac{a}{a + b + n - 1}, \quad n = 1, 2, \dots, \quad (1.1)$$

where  $a > 0$  and  $b \geq 0$ . Around 1996, Persi Diaconis noticed the surprising fact that, for the  $\text{Bern}(1, 0)$  sequence, the count  $Z_1$  is Poisson with mean 1! This result had been discovered earlier (see Kolchin (1971) and Hahlin (1995)), but Diaconis's rediscovery sparked a widespread interest in the topic, and very quickly this result was proved and extended by several people using different techniques. The joint distribution of the count vector  $\mathbf{Z}$  in  $\text{Bern}(a, 0)$  was obtained by using combinatorial methods in Arratia *et al.* (2003). For  $\text{Bern}(1, b)$ ,  $b > 0$ , Joffe *et al.* (2004) obtained the marginal distributions of  $Z_k$  and Sethuraman and Sethuraman (2004) obtained the joint distribution of  $\mathbf{Z}$  by using generating functions of factorial moments. Holst (2007) extended this result to  $\text{Bern}(a, b)$  by using a Pólya–Hoppe urn argument. Huffer *et al.* (2008), (2009) gave a new construction for Bernoulli sequences through *conditional marked Poisson processes* (CMPPs) which included all the Bernoulli sequences considered so far and several dependent Bernoulli sequences. For all such Bernoulli sequences, they showed that  $\mathbf{Z}$  possesses the

*mixture of independent Poissons* (MIP) property: there is a random variable (or vector)  $W$  conditioned on which the count variables  $Z_1, Z_2, \dots$  are independent Poisson.

This is a more transparent description of the distribution of  $\mathbf{Z}$ . Huffer *et al.* (2008) used beta random variables to construct a CMPP leading to the  $\text{Bern}(a, b)$  sequence. Holst (2008a), (2009), (2011) gave an interesting alternative CMPP construction using exponential random variables, and used this construction to study the distribution of the number of occurrences of  $r$  consecutive 1s in a  $\text{Bern}(a, b)$  sequence and the number of occurrences of two or more consecutive 1s in a  $\text{Bern}(1, 0)$  sequence.

Holst (2007), (2008b) studied finite Bernoulli sequences. Let  $\text{Bern}(a, b, n)$  denote the finite initial segment  $(Y_1, \dots, Y_n)$  of a  $\text{Bern}(a, b)$ . Holst obtained expressions for the joint factorial moments of  $(Z_1, Z_2, \dots, Z_k)$  from  $\text{Bern}(a, 0, n)$  and the factorial moments of  $Z_1$  from  $\text{Bern}(a, b, n)$ .

This paper was motivated by Holst's work on finite sequences. We consider initial segments  $(Y_1, Y_2, \dots, Y_\zeta)$  of  $\mathbf{Y} \sim \text{Bern}(a, b)$  which are finite but have a random length  $\zeta$ , and describe in Section 2 a general class of random variables  $\zeta$  for which the count vector  $\mathbf{Z}$  retains the

MIP property. The proof in Section 4 involves an extension of the CMPP construction in Huffer *et al.* (2009). For this class of  $\zeta$ , we give a general expression for the joint factorial moments of  $(Z_1, Z_2, \dots, Z_k)$ . The simplest instance of this class is when  $\zeta$  is a geometric random variable (with  $P(\zeta = k) = (1 - \xi)\xi^k$  for  $k \geq 0$ ) independent of  $\mathbf{Y}$ . By specializing our results to this case and using a power series expansion in  $\xi$ , we are able to deduce the conditional moments given  $\{\zeta = n\}$ , and in this way reproduce (in Sections 5 and 6) the results of Holst for  $\text{Bern}(a, b, n)$  and  $\text{Bern}(a, 0, n)$ , and, moreover, go beyond these results to obtain a general expression for the joint factorial moments of  $\text{Bern}(a, b, n)$ .

For the  $\text{Bern}(a, 0)$  sequence, the counts  $Z_1, Z_2, \dots$  are independent Poisson random variables (i.e. no mixture is required). In this case, there is a choice of  $\zeta$  which retains this property; in Section 7 we construct  $\zeta$  which is independent of the  $\text{Bern}(a, 0)$  sequence and such that the counts from  $(Y_1, \dots, Y_\zeta)$  are independent Poisson random variables.

### 2. Randomly stopped Bernoulli sequences

First we introduce some notation used throughout the paper. Any Bernoulli sequence  $\mathbf{Y}$  is determined by the positions of the 1s in the sequence which we denote by  $\mathbf{K} = (K_0, K_1, K_2, \dots)$ , where  $1 \leq K_0 < K_1 < K_2 < \dots$  and

$$\{K_0, K_1, K_2, \dots\} = \{i : Y_i = 1\}.$$

The sequence  $\mathbf{Y}$  may also be described by its waiting time sequence  $\mathbf{L} = (L_0, L_1, \dots)$  defined by  $L_0 = K_0$  and  $L_j = K_j - K_{j-1}$  for  $j \geq 1$ . The value  $L_j$  is the waiting time after  $K_{j-1}$  for the occurrence of the next 1. For  $j \geq 1$ , the event  $\{L_j = k\}$  represents the occurrence of a  $k$ -string so that

$$Z_k = \sum_{j \geq 1} \mathbf{1}\{L_j = k\}. \tag{2.1}$$

If the sequence  $\mathbf{Y}$  contains only finitely many 1s (say  $m$ ), we adopt the convention of setting  $K_j = L_j = \infty$  for  $j \geq m$ . A sequence with  $L_0 = \infty$  is identically 0.

Let  $\tau$  be a positive integer-valued random variable. Define the Bernoulli sequence  $\mathbf{Y}^* = (Y_1^*, Y_2^*, \dots)$  by

$$Y_i^* = Y_i \mathbf{1}\{i < \tau\},$$

so that  $Y_i^* = 0$  for all  $i \geq \tau$ . For the sequence  $\mathbf{Y}^*$ , we define the sequences  $\mathbf{K}^*$ ,  $\mathbf{L}^*$ , and  $\mathbf{Z}^*$  giving the positions of the 1s, the waiting times, and the string counts, respectively, in exactly the same way as we did for  $\mathbf{Y}$ . Our main interest will be in the joint distribution of the counts  $\mathbf{Z}^*$ . The count vector for  $\mathbf{Y}^*$  is clearly the same as that for the finite sequence  $Y_1, \dots, Y_\zeta$  with  $\zeta = \tau - 1$ , but it is often notationally more convenient to work with the infinite sequence  $\mathbf{Y}^*$ . We say that  $\mathbf{Y}^*$  is obtained by ‘killing’  $\mathbf{Y}$  at time  $\tau$ . (In our work the times  $\tau$  we construct will be stopping times for the sequence  $\mathbf{Y}$ , but  $\zeta = \tau - 1$  will not be. We introduce the term ‘kill’ to avoid possible confusion arising from this.) Since  $\mathbf{Y}^*$  contains only finitely many 1s, we have  $K_j^* = L_j^* = \infty$  for all sufficiently large  $j$ .

We now describe a general class of killing times  $\tau$ . To avoid minor complications, we assume that the sequence  $\mathbf{Y}$  contains infinitely many 1s almost surely. Let  $\phi = (\phi_1, \phi_2, \dots)$  and  $\psi = (\psi_1, \psi_2, \dots)$  be arbitrary sequences of constants in  $[0, 1]$  (with  $\phi_j \equiv \phi(j)$  and  $\psi_j \equiv \psi(j)$ ). We will kill the Bernoulli sequence  $\mathbf{Y}$  at a time  $\tau$  depending on  $\phi$  and  $\psi$  determined as follows. The values  $Y_1, Y_2, \dots$  are observed in sequence, and every time a ‘1’ is observed, a decision is made whether or not to kill the sequence at that point, with a probability

depending only on the waiting time since the previous ‘1’. If the first ‘1’ is observed at time  $K_0 = L_0 = j$ , the sequence is killed at time  $\tau = j$  with probability  $1 - \phi_j$ . For  $n \geq 1$ , if the sequence survives beyond time  $K_{n-1}$  then we kill it at time  $\tau = K_n$  with probability  $1 - \psi(L_n)$ . More formally, we define

$$\tau = K_N, \tag{2.2a}$$

where  $N \geq 0$  satisfies

$$P(N = 0 \mid \mathbf{L}) = 1 - \phi(L_0) \quad \text{and} \quad \frac{P(N = n \mid \mathbf{L})}{P(N \geq n \mid \mathbf{L})} = 1 - \psi(L_n) \quad \text{for } n \geq 1. \tag{2.2b}$$

This implies by elementary probability that

$$P(\tau > K_n \mid \mathbf{L}) = P(N > n \mid \mathbf{L}) = \phi(L_0) \prod_{i=1}^n \psi(L_i). \tag{2.3}$$

We allow  $N$  to take the value  $\infty$  with  $P(N = \infty \mid \mathbf{L}) = \phi(L_0) \prod_{i=1}^{\infty} \psi(L_i)$ , and define  $\tau = \infty$  whenever  $N = \infty$ .

For  $\tau$  satisfying (2.2), there is a simple relationship between the probabilities for  $\mathbf{Y}$  and  $\mathbf{Y}^*$ . Let  $\ell_0, \dots, \ell_n$  be any positive integers. Then

$$\begin{aligned} P(L_0^* = \ell_0, \dots, L_n^* = \ell_n) &= P(N > n, L_0 = \ell_0, \dots, L_n = \ell_n) \\ &= \phi(\ell_0) \prod_{j=1}^n \psi(\ell_j) P(L_0 = \ell_0, \dots, L_n = \ell_n), \end{aligned} \tag{2.4}$$

since  $L_j^* < \infty$  implies that both  $N > j$  and  $L_j = L_j^*$ .

We show in Section 4 that if  $\mathbf{Y}$  is  $\text{Bern}(a, b)$  and  $\tau$  satisfies (2.2) for some sequences  $\phi$  and  $\psi$ , then  $\mathbf{Z}^*$  satisfies the MIP property. However, for the moment, we keep the discussion general and make no restrictions on the sequence  $\mathbf{Y}$ .

An important special case is where

$$\phi_j = \psi_j = \xi^j \quad \text{for all } j \text{ with } 0 < \xi < 1. \tag{2.5}$$

This corresponds to killing  $\mathbf{Y}$  at a geometrically distributed time, for in this case (2.3) becomes  $P(\tau > K_n \mid \mathbf{L}) = \xi^{K_n}$  since  $K_n = L_0 + L_1 + \dots + L_n$ . To make this more precise, suppose that the geometric random variable  $\tau'$  is independent of  $\mathbf{Y}$  with  $P(\tau' = k) = (1 - \xi)\xi^{k-1}$  for  $k \geq 1$ , and define  $\tau = K_N$ , where  $N = \inf\{i : K_i \geq \tau'\}$ . Killing  $\mathbf{Y}$  at either  $\tau'$  or  $\tau$  produces the same sequence  $\mathbf{Y}^*$  (since  $Y_i = 0$  for  $\tau' \leq i < \tau$ ), and it is easily verified that  $N$  so defined satisfies the conditions in (2.2).

Let  $\mathbf{Z}^n$  denote the counts for the finite Bernoulli sequence  $(Y_1, Y_2, \dots, Y_n)$ . The importance of case (2.5) for us later lies in the fact that, because  $\mathbf{Y}$  and  $\tau'$  are independent, the distribution of  $\mathbf{Z}^*$  given  $\tau' = n + 1$  is identical to the distribution of  $\mathbf{Z}^n$ . Thus, we have the expansion

$$\mathcal{L}(\mathbf{Z}^*) = \sum_{n=0}^{\infty} \mathcal{L}(\mathbf{Z}^* \mid \tau' = n + 1) P(\tau' = n + 1) = (1 - \xi) \sum_{n=0}^{\infty} \mathcal{L}(\mathbf{Z}^n) \xi^n. \tag{2.6}$$

Applying this to  $\text{Bern}(a, b)$  will enable us to re-derive Holst’s results for  $\text{Bern}(a, b, n)$  and  $\text{Bern}(a, 0, n)$ . This is done in Sections 5 and 6.

We work chiefly with (2.5), but will briefly describe some other interesting examples. When  $\phi_i = \psi_i = \xi$  for all  $i$ , the sequence  $\mathbf{Y}$  is killed when a geometrically distributed number of 1s

have been observed, leading to a sequence  $Y^*$  for which  $S = \sum_{j=1}^{\infty} Y_j^*$  satisfies  $P(S = k) = (1 - \xi)\xi^k$  for  $k \geq 0$ . When  $\phi_i = \psi_i = 1$  for  $i \leq m$  and  $\phi_i = \psi_i = 0$  for  $i > m$ , the sequence  $Y$  is killed at the first occurrence of  $m$  or more consecutive 0s. When  $\phi_i = 1$  for all  $i$ ,  $\psi_m = 0$ , and  $\psi_i = 1$  for all  $i \neq m$ , the sequence  $Y$  is killed at the first occurrence of a '1' following exactly  $m$  trials after the previous '1'. In all of these cases, when  $Y$  is Bern( $a, b$ ), the resulting string counts  $Z^*$  satisfy the MIP property; see the discussion following Theorem 4.1.

### 3. Conditional marked Poisson processes

We will now describe the CMPP model introduced in Huffer *et al.* (2009). The properties of Poisson processes we use may be found in Resnick (1992) or Kingman (1993). In what follows,  $\mathbb{N} = \{1, 2, 3, \dots\}$ .

Suppose that

1.  $W$  has probability measure  $\nu$ ;
2. conditional on  $W = w$ , the sequence  $X = (X_1, X_2, X_3, \dots)$  of points in  $\mathbb{R}$  with  $0 < X_1 < X_2 < X_3 < \dots < 1$  is a nonhomogeneous Poisson process with intensity function  $\lambda_w(\cdot)$  on  $(0, 1)$ .

Furthermore, suppose that the sequence  $L = (L_0, L_1, L_2, \dots)$  satisfies

- 3  $P(L_0 = k \mid X, W = w) = r(w, k)$  for  $k \in \mathbb{N}$ ;
- 4 for  $n \geq 1$ ,  $P(L_n = k \mid W, X, L_0, \dots, L_{n-1}, X_n = x) = q(x, k)$  for  $k \in \mathbb{N}$ .

Then we say that  $(W, X, L)$  is the CMPP  $\mathcal{M}(\nu, \lambda, r, q)$ .

The intensity functions  $\lambda_w(\cdot)$  satisfy  $\lambda_w(x) \geq 0$ ,  $\int_0^x \lambda_w(u) du < \infty$  for  $0 < x < 1$ , and  $\int_0^1 \lambda_w(u) du = \infty$ . The functions  $r(x, k)$  and  $q(x, k)$  are mass functions in  $k$  for each value of  $x$ . Conditional on  $W = w$ , the pairs  $(X_n, L_n)$ ,  $n \geq 1$ , form a marked Poisson process. To each point  $X_n$ ,  $n \geq 1$ , is assigned a random positive integer mark  $L_n$ . A point  $X_n = x$  is assigned the mark value  $L_n = k$  with probability  $q(x, k)$ , independently of all other points and their marks. Since a marked Poisson process is just a Poisson process on a space which includes the mark values, an alternative description is that, conditional on  $W = w$ , the pairs  $(X_1, L_1), (X_2, L_2), \dots$  form a Poisson process with intensity function  $\lambda_w(x)q(x, k)$  on  $(0, 1) \times \mathbb{N}$ .

A CMPP  $(W, X, L)$  determines a Bernoulli sequence  $Y$  by taking  $L$  to be the sequence of waiting times for  $Y$ , that is, by defining  $K_n = \sum_{i=0}^n L_i$  for  $n \geq 0$  and taking  $Y_i = 1$  when  $i \in \{K_0, K_1, K_2, \dots\}$ . This sequence  $Y$  automatically satisfies the MIP property, which we may state explicitly as follows.

**MIP Property.** *Conditional on  $W = w$ , the counts  $Z_1, Z_2, \dots$  are independent Poisson random variables with*

$$E(Z_k \mid W = w) = \int_0^1 \lambda_w(x)q(x, k) dx. \quad (3.1)$$

This is Theorem 2.2 of Huffer *et al.* (2009). It follows from standard decomposition properties of Poisson processes: conditional on  $W = w$ , for all  $k \geq 1$ , the set  $\{X_n : L_n = k\}$  of points  $X_n$  assigned a mark value of  $k$  forms a Poisson process on  $(0, 1)$  with intensity function  $\lambda_w(x)q(x, k)$ , with these processes being independent for different values of  $k$ . The total number of points in  $\{X_n : L_n = k\}$ , which is  $Z_k$  according to (2.1), will thus have a Poisson distribution with the mean given in (3.1).

### 4. The Bernoulli sequence obtained by killing a CMPP

Let  $(W, X, L)$  be the CMPP  $\mathcal{M}(v, \lambda, r, q)$  described in the previous section, and let  $Y$  be its associated Bernoulli sequence. Let  $\phi = (\phi_1, \phi_2, \dots)$  and  $\psi = (\psi_1, \psi_2, \dots)$  be arbitrary sequences of constants in  $[0, 1]$  as in Section 2. We now describe a way to ‘kill’ the CMPP  $\mathcal{M}$  by adding to it a sequence of secondary marks  $D = (D_0, D_1, \dots)$  called ‘death’ marks, taking values in  $\{0, 1\}$  with distributions

$$\begin{aligned} P(D_0 = 0 \mid W, X, L, L_0 = \ell_0) &= \phi(\ell_0), \\ P(D_n = 0 \mid W, X, L, D_0, \dots, D_{n-1}, L_n = \ell_n) &= \psi(\ell_n) \quad \text{for } n \geq 1, \end{aligned} \tag{4.1}$$

that is, the marked point  $(X_n, L_n)$  is assigned the value  $D_n = 0$  with probability  $\psi(L_n)$ , independently of all other marked points. Define

$$N = \inf\{n : n \geq 0, D_n = 1\}, \quad \tau = X_N, \quad \text{and} \quad Y_i^* = Y_i \mathbf{1}\{i < \tau\} \quad \text{for } i \geq 1, \tag{4.2}$$

and let  $Y^* = (Y_1^*, Y_2^*, \dots)$ . The conditions in (4.1) clearly imply that  $N$  satisfies the requirements in (2.2). Thus,  $\tau$  belongs to the class of killing times studied in Section 2, so that  $Y$  and  $Y^*$  are related as in (2.4).

Introduce the fictitious point  $X_0 = 0$  (not considered part of the Poisson process), and define

$$T = X_N = \inf\{X_n : n \geq 0, D_n = 1\}. \tag{4.3}$$

The finite waiting times  $L_i^*$  for the sequence  $Y^*$  are exactly the set of marks  $L_i$  for which  $X_i < T$ . When  $T = 0$  (which occurs when  $D_0 = 1$ ), this set is empty and according to the convention in Section 2 we take  $L_i^* = \infty$  for all  $i \geq 0$ , corresponding to the sequence  $Y^*$  being identically 0. We think of the process  $\mathcal{M}$  as being killed at time  $T$ , and the sequence  $Y^*$  as arising from the marks  $L_i$  observed before this time.

We now show that the string counts  $Z^*$  for the sequence  $Y^*$  satisfy the MIP property.

**Theorem 4.1.** *Consider the CMPP  $\mathcal{M}(v, \lambda, r, q)$  and its corresponding Bernoulli sequence  $Y$ . Let  $D, Y^*$ , and  $T$  be as defined in (4.1), (4.2), and (4.3), and let  $Z^* = (Z_1^*, Z_2^*, \dots)$  be the string counts for  $Y^*$ .*

- (a) *Conditional on  $\{W = w, T = t\}$ , the counts  $Z_1^*, Z_2^*, \dots$  are independent Poisson random variables with means given by*

$$\mu_k(w, t) = E(Z_k^* \mid W = w, T = t) = \int_0^t \lambda_w(x)q(x, k)\psi(k) dx \quad \text{for } k = 1, 2, 3, \dots \tag{4.4}$$

- (b) *The random variable  $T$  satisfies*

$$P(T > t \mid W = w) = H(w) \exp\left(-\int_0^t \eta_w(x) dx\right) \quad \text{for } t \in (0, 1), \tag{4.5}$$

where

$$H(w) = \sum_{k \in \mathbb{N}} r(w, k)\phi(k), \tag{4.6}$$

$$\eta_w(x) = \sum_{k \in \mathbb{N}} \lambda_w(x)q(x, k)(1 - \psi(k)). \tag{4.7}$$

*Proof.* The equations in (4.1) imply that, conditional on  $W = w$ , the points in  $\{(X_n, L_n, D_n) : n \geq 1\}$  form a Poisson process on  $(0, 1) \times \mathbb{N} \times \{0, 1\}$  with intensity  $\lambda_w(x)q(x, k)\psi(k, d)$ , where we define  $\psi(k, 0) = \psi(k)$  and  $\psi(k, 1) = 1 - \psi(k)$ . Thus, conditional on  $W = w$ , the sets of points  $\mathcal{E}_0 = \{(X_n, L_n) : n \geq 1, D_n = 0\}$  and  $\mathcal{E}_1 = \{(X_n, L_n) : n \geq 1, D_n = 1\}$  are independent Poisson processes on  $(0, 1) \times \mathbb{N}$  with intensity functions  $\lambda_w(x)q(x, k)\psi(k)$  and  $\lambda_w(x)q(x, k)(1 - \psi(k))$ , respectively. Ignoring the marks  $L_n$  on the points in  $\mathcal{E}_1$ , we obtain the Poisson process  $\mathcal{E}'_1 = \{X_n : n \geq 1, D_n = 1\}$  with intensity  $\eta_w(x)$  as in (4.7), which is again conditionally independent of  $\mathcal{E}_0$  given  $W = w$ .

The random variable  $T$  in (4.3) is equal to 0 if  $D_0 = 1$ , and is equal to the first point in  $\mathcal{E}'_1$  otherwise. Since  $P(D_0 = 0 \mid W = w) = H(w)$  as in (4.6), and the probability given  $W = w$  that the interval  $(0, t)$  contains no points from  $\mathcal{E}'_1$  is  $\exp(-\int_0^t \eta_w(x) dx)$ , (4.5) is immediate. The conditional independence of  $\mathcal{E}_0$  and  $\mathcal{E}'_1$  implies that of  $\mathcal{E}_0$  and  $T$ . Thus, given  $\{W = w, T = t\}$ , the counts

$$Z_k^* = \sum_{\{i : i \geq 1, X_i < T\}} \mathbf{1}\{L_i = k\} = \sum_{\{i : (X_i, L_i) \in \mathcal{E}_0, X_i < T\}} \mathbf{1}\{L_i = k\}$$

are independent Poisson random variables with means as given in (4.4). This concludes the proof.

Since it was shown in Huffer *et al.* (2009) that the  $\text{Bern}(a, b)$  sequence arises from a CMPP, this verifies the assertion made in Section 2 that killing a  $\text{Bern}(a, b)$  sequence by any  $\tau$  satisfying (2.2) produces counts possessing the MIP property.

Finally, for use in the applications of Sections 5 and 6, we give a general expression for the joint factorial moments of the counts  $Z_1^*, \dots, Z_k^*$ .

**Corollary 4.1.** *Let  $Z_1^*, \dots, Z_k^*$  be the string counts in Theorem 4.1 obtained by killing the CMPP  $\mathcal{M}(v, \lambda, r, q)$ . Suppose that  $\eta_w(\cdot)$  in (4.7) satisfies*

$$v \left\{ w : \int_0^1 \eta_w(x) dx = \infty \right\} = 1, \tag{4.8}$$

and that  $(r_1, \dots, r_k) \neq (0, \dots, 0)$  are nonnegative integers. Then

$$\begin{aligned} E \left( \prod_{i=1}^k \binom{Z_i^*}{r_i} \right) &= \int \int_0^1 \left[ \prod_{i=1}^k \frac{(\mu_i(w, t))^{r_i}}{r_i!} \right] H(w) \eta_w(t) \\ &\quad \times \exp \left( - \int_0^t \eta_w(x) dx \right) dt dv(w). \end{aligned} \tag{4.9}$$

*Proof.* For a Poisson random variable  $Z$  with mean  $\theta$ , we know that

$$E \binom{Z}{r} = \frac{\theta^r}{r!} \quad \text{for integers } r \geq 0.$$

Thus, (4.4) implies that

$$E \left( \prod_{i=1}^k \binom{Z_i^*}{r_i} \right) = E E \left( \prod_{i=1}^k \binom{Z_i^*}{r_i} \mid W, T \right) = E \left( \prod_{i=1}^k \frac{(\mu_i(W, T))^{r_i}}{r_i!} \right). \tag{4.10}$$

The conditional distribution  $\mathcal{L}(T \mid W)$  is determined by (4.5). This distribution has an atom at  $T = 0$  (of size  $P(T = 0 \mid W = w) = 1 - H(w)$ ) which does not contribute to the expectation in

(4.10) since  $Z_i^* = 0$  for all  $i$  when  $T = 0$ . Assumption (4.8) guarantees that  $P(T < 1 \mid W) = 1$ . For  $0 < T < 1$ , the conditional distribution has a density given by

$$-\frac{\partial}{\partial t} P(T > t \mid W = w) = H(w)\eta_w(t) \exp\left(-\int_0^t \eta_w(x) dx\right),$$

which leads to the integral expression for (4.10) given in (4.9).

### 5. Joint moments of $Z^n$ from $\text{Bern}(a, b, n)$

In this section we apply the general results of Section 4 to the  $\text{Bern}(a, b)$  sequence  $\mathbf{Y}$  defined in (1.1). In particular, we calculate the joint factorial moments in (4.9) for the sequence  $\mathbf{Y}^*$  obtained by killing  $\mathbf{Y}$  by  $\tau$  satisfying (2.2) and (2.5), and use these to give a new derivation of the moment formulae in Holst (2007), (2008b) for the  $\text{Bern}(a, b, n)$  sequence  $(Y_1, \dots, Y_n)$ .

We begin by restating Proposition 3.1 of Huffer *et al.* (2009) which gives a CMPP model to generate  $\text{Bern}(a, b)$  when  $b > 0$ .

**Proposition 5.1.** (Huffer *et al.* (2009, Proposition 3.1).) *Consider the CMPP  $\mathcal{M}(v, \lambda, r, q)$  with*

$$v \sim \text{Beta}(b, a) \text{ with density } dv(w) = B(b, a)^{-1} w^{b-1} (1-w)^{a-1} dw, \quad 0 < w < 1, \quad (5.1)$$

where  $B(b, a) = \Gamma(b)\Gamma(a)/\Gamma(a+b)$ ,

$$\lambda_w(x) = \frac{a}{1-x} \mathbf{1}\{w < x < 1\}, \quad (5.2)$$

$$r(x, k) = q(x, k) = x^{k-1}(1-x), \quad k \in \mathbb{N}. \quad (5.3)$$

This CMPP determines a Bernoulli sequence  $\mathbf{Y}$  which has distribution  $\text{Bern}(a, b)$ .

Throughout this section,  $\mathbf{Y} = (Y_1, Y_2, \dots)$  denotes the  $\text{Bern}(a, b)$  sequence obtained via this proposition. For  $\alpha > 0$  and nonnegative integers  $r_1, \dots, r_k$ , define

$$\tilde{B}(\alpha; r_1, \dots, r_k) = \int_0^1 x^{\alpha-1} \prod_{i=1}^k (1-x^i)^{r_i} dx.$$

Let  $c^{\underline{p}} = c(c-1)\dots(c-p+1)$  and  $c^{\overline{p}} = c(c+1)\dots(c+p-1)$  denote descending and ascending factorial products, with  $c^{\underline{0}} = c^{\overline{0}} = 1$ . We now prove the following.

**Theorem 5.1.** *Let  $Z^n = (Z_1^n, Z_2^n, \dots)$  be the string counts for the  $\text{Bern}(a, b, n)$  sequence  $(Y_1, \dots, Y_n)$ , and let  $(r_1, \dots, r_k) \neq (0, \dots, 0)$  be nonnegative integers. Define  $R = \sum_{i=1}^k r_i$  and  $M = \sum_{i=1}^k i r_i$ . Then*

$$\begin{aligned} E\left(\prod_{i=1}^k \binom{Z_i^n}{r_i}\right) &= \mathbf{1}\{n \geq M+1\} \frac{a^{R+1}}{\prod_{i=1}^k r_i! i^{r_i}} \frac{B(n+b, a)}{B(b, a)} \\ &\quad \times \sum_{p=0}^{n-M-1} (-1)^p \frac{(a-1)^{\underline{p}}}{p!} \frac{(a+1)^{\overline{n-M-p-1}}}{(n-M-p-1)!} \tilde{B}(b+p; r_1, \dots, r_k). \end{aligned} \quad (5.4)$$

In particular,

$$E\binom{Z_1^n}{r} = \frac{a^{r+1}}{r!} \frac{B(b+n, a)}{B(b, a)} \sum_{p=0}^{n-r-1} (-1)^p \frac{(a-1)^{\underline{p}}}{p!} \frac{(a+1)^{\overline{n-r-p-1}}}{(n-r-p-1)!} B(b+p, r+1). \quad (5.5)$$



*Proof.* Let  $Y^*$  be the sequence obtained by killing  $Y$  by  $\tau$  satisfying (2.2) and (2.5), and let  $Z^*$  be the corresponding counts. We will calculate the factorial moments of  $Z^*$  using (4.9), expand these in a power series in  $\xi$ , and then exploit the relationship between  $Z^*$  and  $Z^n$  given in (2.6) to obtain the moments for  $Z^n$ .

Using (2.5), (4.4), (4.6), (4.7), (5.2), and (5.3), we find by simple calculations that

$$\mu_i(w, t) = \frac{a\xi^i}{i} (t^i - w^i) \mathbf{1}\{w < t\}, \tag{5.6a}$$

$$H(w) = \frac{(1-w)\xi}{1-\xi w}, \tag{5.6b}$$

$$\eta_w(x) = \left( \frac{a}{1-x} - \frac{a\xi}{1-\xi x} \right) \mathbf{1}\{w < x < 1\} = \frac{a(1-\xi)}{(1-x)(1-\xi x)} \mathbf{1}\{w < x < 1\}, \tag{5.6c}$$

$$\int_0^t \eta_w(x) dx = a \log \left[ \frac{(1-\xi t)(1-w)}{(1-t)(1-\xi w)} \right] \mathbf{1}\{w < t\}. \tag{5.6d}$$

Substituting these along with (5.1) into (4.9) and simplifying, we obtain

$$E\left(\prod_{i=1}^k \binom{Z_i^*}{r_i}\right) = Q\xi^{M+1}(1-\xi) \iint_A w^{b-1} \frac{(1-t)^{a-1}(1-\xi w)^{a-1}}{(1-\xi t)^{a+1}} \prod_{i=1}^k (t^i - w^i)^{r_i} dw dt, \tag{5.7}$$

where

$$Q = \frac{a^{R+1}}{B(b, a) \prod_{i=1}^k r_i! i^{r_i}} \quad \text{and} \quad A = \{0 < w < t < 1\}.$$

Upon substituting the expansions

$$(1-\xi w)^{a-1} = \sum_{p=0}^{\infty} (-1)^p \frac{(a-1)^{\underline{p}}}{p!} \xi^p w^p \quad \text{and} \quad \frac{1}{(1-\xi t)^{a+1}} = \sum_{q=0}^{\infty} \frac{(a+1)^{\overline{q}}}{q!} \xi^q t^q, \tag{5.8}$$

(5.7) becomes

$$Q\xi^{M+1}(1-\xi) \sum_{s=0}^{\infty} \xi^s \sum_{p=0}^s (-1)^p \frac{(a-1)^{\underline{p}}}{p!} \frac{(a+1)^{\overline{s-p}}}{(s-p)!} \\ \times \iint_A w^{b+p-1} t^{s-p} (1-t)^{a-1} \prod_{i=1}^k (t^i - w^i)^{r_i} dw dt.$$

We can separate the integration over  $w$  and  $t$  above with the transformations  $x = w/t$  and  $y = t$ , yielding

$$\iint_A w^{b+p-1} t^{s-p} (1-t)^{a-1} \prod_{i=1}^k (t^i - w^i)^{r_i} dw dt \\ = \int_0^1 t^{b+s+M} (1-t)^{a-1} dt \int_0^1 x^{b+p-1} \prod_{i=1}^k (1-x^i)^{r_i} dx \\ = B(b+s+M+1, a) \tilde{B}(b+p; r_1, \dots, r_k). \tag{5.9}$$

If we substitute this into the above and let  $n = s + M + 1$ , then (5.7) becomes

$$E\left(\prod_{i=1}^k \binom{Z_i^*}{r_i}\right) = (1 - \xi) \sum_{n=M+1}^{\infty} \xi^n Q B(b + n, a) \times \sum_{p=0}^{n-M-1} (-1)^p \frac{(a-1)^p}{p!} \frac{(a+1)^{\overline{n-M-1-p}}}{(N-M-1-p)!} \tilde{B}(b+p; r_1, \dots, r_k).$$

Since the coefficient of  $(1 - \xi)\xi^n$  in this series agrees with (5.4), using (2.6) completes the proof.

Our expression for the factorial moments of  $Z_1^n$  in (5.5) is very different from that given by Holst (2008b), i.e.

$$E\binom{Z_1^n}{r} = \frac{a^r}{(a+b+n-1)^r} \sum_{k=1}^r \binom{r-1}{r-k} \binom{n-r}{k} \frac{a^{\bar{k}}}{(a+b)^{\bar{k}}}, \tag{5.10}$$

but we can show that they are equal by the following argument. Write (5.7) with  $k = 1$  and  $r_1 = r$  explicitly as a double integral, i.e.

$$E\binom{Z_1^*}{r} = \frac{a^{r+1} \xi^{r+1} (1 - \xi)}{B(b, a) r!} \int_0^1 w^{b-1} \int_w^1 \frac{(1 - \xi w)^{a-1}}{(1 - \xi t)^{a+1}} (1 - t)^{a-1} (t - w)^r dt dw, \tag{5.11}$$

and manipulate the inner integral to obtain

$$\begin{aligned} & \int_w^1 \frac{(1 - \xi w)^{a-1}}{(1 - \xi t)^{a+1}} (1 - t)^{a-1} (t - w)^r dt \\ &= \frac{(1 - \xi w)^{r-1}}{1 - \xi} \int_w^1 (x - w)^{a-1} \left(\frac{1 - x}{1 - \xi x}\right)^r dx \\ &= \frac{r}{a} \int_w^1 (x - w)^a (1 - x)^{r-1} \frac{(1 - \xi w)^{r-1}}{(1 - \xi x)^{r+1}} dx, \end{aligned} \tag{5.12}$$

where the first equality is obtained by applying the substitution

$$x = w + (1 - \xi w) \left(\frac{1 - t}{1 - \xi t}\right)$$

and the second equality follows from integration by parts using

$$\frac{d}{dx} \left(\frac{1 - x}{1 - \xi x}\right)^r = -r(1 - \xi) \frac{(1 - x)^{r-1}}{(1 - \xi x)^{r+1}}.$$

Noting that

$$\frac{1 - \xi w}{1 - \xi x} = 1 + \frac{\xi(x - w)}{1 - \xi x},$$

we use a binomial expansion followed by expanding  $(1 - \xi x)^{-(k+1)}$  as in (5.8) to obtain

$$\begin{aligned} \frac{(1 - \xi w)^{r-1}}{(1 - \xi x)^{r+1}} &= \left(1 + \frac{\xi(x - w)}{1 - \xi x}\right)^{r-1} \frac{1}{(1 - \xi x)^2} \\ &= \sum_{k=1}^r \binom{r-1}{r-k} \frac{\xi^{k-1}(x - w)^{k-1}}{(1 - \xi x)^{k+1}} \\ &= \sum_{i=0}^{\infty} \xi^i \sum_{k=1}^r \binom{r-1}{r-k} \binom{i+1}{k} (x - w)^{k-1} x^{i-k+1}. \end{aligned}$$

Substituting this into (5.12) and placing it back into (5.11) leads to

$$\begin{aligned} E\left(\binom{Z_1^*}{r}\right) &= \frac{a^r(1 - \xi)}{B(b, a)(r - 1)!} \sum_{n=r+1}^{\infty} \xi^n \sum_{k=1}^r \binom{r-1}{r-k} \binom{n-r}{k} \\ &\quad \times \int \int_{0 < w < x < 1} x^{n-r-k} (1 - x)^{r-1} w^{b-1} (x - w)^{a+k-1} dx dw \\ &= \frac{a^r(1 - \xi)}{B(b, a)(r - 1)!} \sum_{n=r+1}^{\infty} \xi^n \sum_{k=1}^r \binom{r-1}{r-k} \binom{n-r}{k} B(a + b + n - r, r) B(b, a + k). \end{aligned}$$

In going from the first equality to the second, we have separated the integrals using the transformations  $u = w/x$  and  $v = x$  (as in (5.9)) and evaluated the resulting integrals in terms of beta functions. After simplifying the beta functions, we find that the coefficient of  $(1 - \xi)\xi^n$  in this expression agrees with (5.10), which (upon using (2.6)) completes the proof of (5.10).

**6. Joint moments of  $Z^n$  from  $Bern(a, 0, n)$**

The previous section required that  $b > 0$ . We now consider  $b = 0$ , which leads to some simplifications. We begin by describing a CMPP which generates the  $Bern(a, 0)$  sequence.

**Proposition 6.1.** *The CMPP  $\mathcal{M}(v, \lambda, r, q)$  with  $v = \delta_0$  and  $\lambda, r, q$  as in (5.2) and (5.3) determines a sequence  $Y$  which is  $Bern(a, 0)$ .*

This result is Remark 3.2 of Huffer *et al.* (2009, p. 2129). It may be proved by taking the limit  $b \downarrow 0$  in Proposition 3.1 of Huffer *et al.* (2009) or by direct calculation, following the proof of Proposition 3.1. The CMPP in this special case has  $(W, L_0) = (0, 1)$  with probability 1.

Using this CMPP, we will prove the following result, which is Proposition 3 of Holst (2007). We use the notation in Theorem 5.1.

**Theorem 6.1.** *The joint factorial moments of counts from  $Bern(a, 0, n)$  are given by*

$$E\left(\prod_{i=1}^k \binom{Z_i^n}{r_i}\right) = \mathbf{1}\{n \geq M + 1\} \frac{a^R}{\prod_{i=1}^k r_i! i^{r_i}} \frac{(n - 1)^{\underline{M}}}{(a + n - 1)^{\underline{M}}}. \tag{6.1}$$

*Proof.* We use the same argument as in Theorem 5.1. Let  $Y^*$  be the sequence obtained by killing  $Y$  by  $\tau$  satisfying (2.2) and (2.5), and let  $Z^*$  be the corresponding counts. We will calculate the factorial moments of  $Z^*$  using (4.9), expand in a power series in  $\xi$ , and then use (2.6).

Since  $v = \delta_0$ , the double integral in (4.9) reduces to a single integral over  $t$  with  $w$  set to zero everywhere in the integrand (i.e. all the quantities in (5.6) are evaluated at  $w = 0$ ). This leads to

$$E\left(\prod_{i=1}^k \binom{Z_i^*}{r_i}\right) = \frac{a^{R+1} \xi^{M+1} (1 - \xi)}{\prod_{i=1}^k r_i! i^{r_i}} \int_0^1 t^M \frac{(1 - t)^{a-1}}{(1 - \xi t)^{a+1}} dt.$$

Expanding  $(1 - \xi t)^{-(a+1)}$  as in (5.8) and then evaluating the resulting integrals in terms of beta functions and juggling the factorials yields

$$\frac{a^{R+1} \xi^{M+1} (1 - \xi)}{\prod_{i=1}^k r_i! i^{r_i}} \sum_{q=0}^{\infty} \xi^q \frac{(a + 1)^{\bar{q}}}{q!} B(m + q + 1, a) = \frac{a^R (1 - \xi)}{\prod_{i=1}^k r_i! i^{r_i}} \sum_{n=M+1}^{\infty} \xi^n \frac{(n - 1)^{\underline{M}}}{(a + n - 1)^{\underline{M}}}.$$

Since the coefficient of  $(1 - \xi)\xi^n$  in this series is (6.1), using (2.6) completes the proof.

### 7. Stopping a Bern( $a, 0$ ) sequence so that the counts are still independent Poisson

In this section we prove the following result.

**Theorem 7.1.** *Let  $Y$  be a Bern( $a, 0$ ) sequence, and let  $\tau$  be independent of  $Y$  and take values in  $\{2, 3, 4, \dots\}$  with distribution given by*

$$P(\tau > k) = B(k - 1, a + 1)^{-1} \int_0^U x^{k-2} (1 - x)^a dx \tag{7.1}$$

for  $k \geq 2$ . Here  $U$  is an arbitrary fixed value in  $(0, 1)$ . Then  $Y^*$  defined by  $Y_i^* = Y_i \mathbf{1}\{i < \tau\}$  is a Bernoulli sequence whose counts  $Z_k^*, k \geq 1$ , are independent Poisson random variables with means  $aU^k/k$ .

*Proof.* The sequence  $Y$  arises from the CMPP (call it  $\mathcal{M}$ ) in Proposition 6.1. Let  $K_0, K_1, K_2, \dots$  be the positions of the 1s in  $Y$ . For  $\mathcal{M}$ , we have  $P(W = 0) = 1$ , so that  $X_1, X_2, \dots$  are simply the points in a Poisson process with intensity  $\lambda(x) \equiv \lambda_0(x) = a/(1 - x)$  on  $(0, 1)$ .

Define

$$\tilde{\tau} = \inf\{K_n : X_n \geq U\},$$

and consider the sequence  $\tilde{Y}$  with  $\tilde{Y}_i = Y_i \mathbf{1}\{i < \tilde{\tau}\}$  which corresponds to observing the marks  $L_i$  only for points  $X_i$  in  $(0, U)$ . For this sequence, let  $\tilde{K}_i, \tilde{L}_i$ , and  $\tilde{Z}_i$  denote the positions of the 1s, waiting times, and counts, respectively. By the same argument used to prove (3.1) and (4.4), we see that the counts  $\tilde{Z}_k$  are independent Poisson random variables with means

$$E(\tilde{Z}_k) = \int_0^U \lambda(x) q(x, k) dx = \int_0^U \frac{a}{1 - x} x^{k-1} (1 - x) dx = \frac{aU^k}{k},$$

which is the property required in our theorem.

We now complete the proof by showing that  $\tilde{Y}$  has the same distribution as  $Y^*$  in the statement of Theorem 7.1 (that is, killing  $Y$  by  $\tau$  or by  $\tilde{\tau}$  produces sequences with the same distribution), so that  $(\tilde{Z}_1, \tilde{Z}_2, \dots)$  and  $(Z_1^*, Z_2^*, \dots)$  have the same distribution.

Choose arbitrary integers  $k_0, \dots, k_n$  satisfying  $1 = k_0 < k_1 < k_2 < \dots < k_n < \infty$ , and let  $\ell_i = k_i - k_{i-1}$  for  $i = 1, \dots, n$ . Then

$$\begin{aligned} P(\tilde{L}_1 = \ell_1, \dots, \tilde{L}_n = \ell_n) &= P(\tilde{K}_1 = k_1, \dots, \tilde{K}_n = k_n) \\ &= P(K_1 = k_1, \dots, K_n = k_n, \tilde{\tau} > k_n) \\ &= P(L_1 = \ell_1, \dots, L_n = \ell_n, \tilde{\tau} > k_n) \\ &= P(L_1 = \ell_1, \dots, L_n = \ell_n, X_n < U). \end{aligned} \tag{7.2}$$

Since the points  $(X_1, L_1), (X_2, L_2), \dots$  form a Poisson process on  $(0, 1) \times \mathbb{N}$  with intensity  $\lambda(x)q(x, k)$ , we have

$$\begin{aligned} P(L_1 = \ell_1, \dots, L_n = \ell_n, X_n < U) &= \int_{A_n(U)} P(L_1 = \ell_1, \dots, L_n = \ell_n, X_1 \in dx_1, \dots, X_n \in dx_n) \\ &= \int_{A_n(U)} \exp\left(-\int_0^{x_n} \lambda(u) du\right) \prod_{i=1}^n \lambda(x_i)q(x_i, \ell_i) dx_1 \cdots dx_n \\ &= \int_{A_n(U)} (1 - x_n)^a \prod_{i=1}^n ax_i^{\ell_i-1} dx_i, \end{aligned}$$

where  $A_n(U) = \{0 < X_1 < X_2 < \dots < X_n < U\}$  and  $(1 - x_n)^a = \exp(-\int_0^{x_n} \lambda(u) du)$  is the probability of no points other than  $x_1, \dots, x_n$  in  $[0, x_n]$ . By successively integrating over  $x_1$ , then  $x_2$ , etc., we obtain

$$\frac{a^n}{\ell_1(\ell_1 + \ell_2) \cdots (\ell_1 + \ell_2 + \dots + \ell_n)} \int_0^U x_n^{\ell_1+\ell_2+\dots+\ell_n-1} (1 - x_n)^a dx_n. \tag{7.3}$$

(This calculation is similar to that in Proposition 3.1 of Huffer *et al.* (2009).) The corresponding probability for the usual Bern( $a, 0$ ) sequence without any killing is obtained by setting  $U = 1$  in (7.3), so that taking a ratio and using (7.2) and  $k_n = 1 + \sum_{i=1}^n \ell_i$  leads to

$$\begin{aligned} \frac{P(\tilde{L}_1 = \ell_1, \dots, \tilde{L}_n = \ell_n)}{P(L_1 = \ell_1, \dots, L_n = \ell_n)} &= P(\tilde{\tau} > k_n \mid L_1 = \ell_1, \dots, L_n = \ell_n) \\ &= \frac{\int_0^U x^{k_n-2} (1-x)^a dx}{\int_0^1 x^{k_n-2} (1-x)^a dx} \\ &= P(\tau > k_n) \end{aligned} \tag{7.4}$$

by (7.1).

Now consider the sequence  $Y_1^*, Y_2^*, Y_3^*, \dots$  of Theorem 7.1, and let  $L_1^*, L_2^*, L_3^*, \dots$  be the waiting times for this sequence. Then, using (7.4), we obtain

$$\begin{aligned} P(L_1^* = \ell_1, \dots, L_n^* = \ell_n) &= P(L_1 = \ell_1, \dots, L_n = \ell_n, \tau > k_n) \\ &= P(L_1 = \ell_1, \dots, L_n = \ell_n) P(\tau > k_n) \\ &= P(L_1 = \ell_1, \dots, L_n = \ell_n) P(\tilde{\tau} > k_n \mid L_1 = \ell_1, \dots, L_n = \ell_n) \\ &= P(L_1 = \ell_1, \dots, L_n = \ell_n, \tilde{\tau} > k_n) \\ &= P(\tilde{L}_1 = \ell_1, \dots, \tilde{L}_n = \ell_n), \end{aligned}$$

so that  $Y^*$  and  $\tilde{Y}$  have the same distributions, completing the proof.

An interesting related fact is the following.

**Corollary 7.1.** *Let  $M$  be the position of the last ‘1’ in the sequence  $Y^*$  of Theorem 7.1, that is,  $M = \sup\{i : Y_i^* = 1\}$ . Then  $M - 1$  has a negative binomial distribution with probability mass function*

$$P(M - 1 = k) = \binom{a + k - 1}{k} (1 - U)^a U^k \quad \text{for } k = 0, 1, 2, \dots$$

and moment generating function

$$E \exp(s(M - 1)) = \left( \frac{1 - U}{1 - Ue^s} \right)^a. \quad (7.5)$$

*Proof.* We verify that  $M - 1$  has the moment generating function given in (7.5). Since  $Y^*$  and  $\tilde{Y}$  have the same distribution, we have

$$\begin{aligned} M &= \sup\{i < \tau : Y_i = 1\} \\ &\stackrel{D}{=} \sup\{i < \tilde{\tau} : Y_i = 1\} \\ &= \sup\{K_i : X_i < U\} \\ &= 1 + \sum_{\{i : X_i < U\}} L_i \\ &= 1 + \sum_{\{i : \tilde{L}_i < \infty\}} \tilde{L}_i \\ &= 1 + \sum_{k=1}^{\infty} k \tilde{Z}_k \end{aligned}$$

according to (2.1). The random variables  $\tilde{Z}_k$  are independent Poisson random variables with means  $aU^k/k$ , so that the moment generating function of  $M - 1$  is easily calculated using the fact that a Poisson random variable with mean  $\theta$  has moment generating function  $e^{\theta(e^s - 1)}$ . This calculation (which is omitted) leads to (7.5).

## References

- ARRATIA, R., BARBOUR, A. D. AND TAVARÉ, S. (1992). Poisson process approximations for the Ewens sampling formula. *Ann. Appl. Prob.* **2**, 519–535.
- ARRATIA, R., BARBOUR, A. D. AND TAVARÉ, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach*. European Mathematical Society, Zürich.
- HAHLIN, L.-O. (1995). Double records. Res. Rep. 1995:12, Department of Mathematics, Uppsala University.
- HOLST, L. (2007). Counts of failure strings in certain Bernoulli sequences. *J. Appl. Prob.* **44**, 824–830.
- HOLST, L. (2008a). A note on embedding certain Bernoulli sequences in marked Poisson processes. *J. Appl. Prob.* **45**, 1181–1185.
- HOLST, L. (2008b). The number of two consecutive successes in a Hoppe–Pólya urn. *J. Appl. Prob.* **45**, 901–906.
- HOLST, L. (2009). On consecutive records in certain Bernoulli sequences. *J. Appl. Prob.* **46**, 1201–1208.
- HOLST, L. (2011). A note on records in a random sequence. *Ark. Mat.* **49**, 351–356.
- HUFFER, F., SETHURAMAN, J. AND SETHURAMAN, S. (2008). A study of counts of Bernoulli strings via conditional Poisson processes. Preprint. Available at <http://arxiv.org/abs/0801.2115v1>.
- HUFFER, F., SETHURAMAN, J. AND SETHURAMAN, S. (2009). A study of counts of Bernoulli strings via conditional Poisson processes. *Proc. Amer. Math. Soc.* **137**, 2125–2134.
- JOFFE, A., MARCHAND, É., PERRON, F. AND POPADIUK, P. (2004). On sums of products of Bernoulli variables and random permutations. *J. Theoret. Prob.* **17**, 285–292.
- KINGMAN, J. F. C. (1993). *Poisson Processes*. Oxford University Press.

- KOLCHIN, V. F. (1971). A problem of the allocation of particles in cells and cycles of random permutations. *Theory Prob. Appl.* **16**, 74–90.
- RESNICK, S. I. (1992). *Adventures in Stochastic Processes*. Birkhäuser, Boston, MA.
- SETHURAMAN, J. AND SETHURAMAN, S. (2004). On counts of Bernoulli strings and connections to rank orders and random permutations. In *A festschrift for Herman Rubin* (IMS Lecture Notes Monogr. Ser. **45**), Institute for Mathematical Statistics, Beachwood, OH, pp. 140–152.