A Festschrift for Herman Rubin
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 45 (2004) 140–152

On counts of Bernoulli strings and connections to rank orders and random permutations

Jayaram Sethuraman[1] and Sunder Sethuraman[2]

Florida State University and Iowa State University

Abstract: A sequence of independent random variables $\{X_1, X_2, \ldots\}$ is called a $B-$harmonic Bernoulli sequence if $P(X_i = 1) = 1 - P(X_i = 0) = 1/(i + B)$ $i = 1, 2, \ldots$, with $B \geq 0$. For $k \geq 1$, the count variable Z_k is the number of occurrences of the k-string $(1, \underbrace{0, \ldots, 0}_{k-1}, 1)$ in the Bernoulli sequence... This paper gives the joint distribution P_B of the count vector $\mathbf{Z} = (Z_1, Z_2, \ldots)$ of strings of all lengths in a $B-$harmonic Bernoulli sequence. This distribution can be described as follows. There is random variable V with a Beta$(B, 1)$ distribution, and given $V = v$, the conditional distribution of $\mathbf{Z}$ is that of independent Poissons with intensities $(1 - v)$, $(1 - v^2)/2$, $(1 - v^3)/3, \ldots$.

Around 1996, Persi Diaconis stated and proved that when $B = 0$, the distribution of Z_1 is Poisson with intensity 1. Emery gave an alternative proof a few months later. For the case $B = 0$, it was also recognized that $Z_1, Z_2, \ldots, Z_n$ are independent Poissons with intensities $1, \frac{1}{2}, \ldots, \frac{1}{n}$. Proofs up until this time made use of hard combinational techniques. A few years later, Joffe et al, obtained the marginal distribution of Z_1 as a Beta-Poisson mixture when $B \geq 0$. Their proof recognizes an underlying inhomogeneous Markov chain and uses moment generating functions.

In this note, we give a compact expression for the joint factorial moment of $(Z_1, \ldots, Z_N)$ which leads to the joint distribution given above. One might feel that if Z_1 is large, it will exhaust the number of 1's in the Bernoulli sequence $(X_1, X_2, \ldots)$ and this in turn would favor smaller values for Z_2 and introduce some negative dependence. We show that, on the contrary, the joint distribution of $\mathbf{Z}$ is positively associated or possesses the FKG property.

1. Introduction and summary

Let $\{X_i : i \geq 1\}$ be a sequence of independent Bernoulli random variables with success probabilities $p_i = P(X_i = 1) = 1 - P(X_i = 0)$ for $i \geq 1$. For integers $k \geq 1$, the sequence $(1, \underbrace{0, \ldots, 0}_{k-1}, 1)$ will be called a k-string. Such a k-string represents a wait of length k for an "event" to happen since the last time it happened, or a run of length $k-1$ of "non-events." Let Z_k be the count (which may possibly be infinite) of such k strings in the Bernoulli sequence $\{X_1, X_2, \ldots\}$. This paper is concerned with the joint distribution of the count vector $\mathbf{Z} \stackrel{def}{=} (Z_1, Z_2, \ldots)$ of all k-strings. Such problems appear in many areas such as random permutations, rank orders, genetics, abundance of species, etc.

[1]Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, USA. e-mail: sethu@stat.fsu.edu

[2]Department of Mathematics, 430 Carvel Hall, Iowa State University, Ames, IA 50011, USA. e-mail: sethuram@iastate.edu

Keywords and phrases: rank order statistics, records, random permutations, factorial moments, cycles, strings, Bernoulli.

AMS 2000 subject classifications: primary, 60C35; secondary 60E05.

Let $Y_{i,k}$ be the indicator variable that a k-string has occurred at time i,

$$Y_{i,k} \;=\; X_i \prod_{j=1}^{i+k-1} (1 - X_{i+j}) X_{i+k} \;=\; I\Big((X_i, X_{i+1}, \ldots, X_{i+k}) = (1, \underbrace{0, \ldots, 0}_{k-1}, 1) \Big), \tag{1}$$

for $i \geq 1, k \geq 1$, where as usual, an empty product is defined to be equal to 1. A simple expression for Z_k is then given by

$$Z_k \;=\; \sum_{i=1}^{\infty} Y_{i,k} \quad \text{for } k \geq 1. \tag{2}$$

While Z_k is not a sum of independent summands, it can be easily expressed as the sum of k series, each of which has independent summands. From this observation and Kolomogorov's three-series theorem we can state the following remark which gives a necessary and sufficient condition that the random variable Z_k be finite a.s.

Remark 1. The count random variable Z_k of k-strings is finite a.s. if and only if $E[Z_k] = \sum_{i \geq q} p_i \prod_{j=1}^{k-1} (1 - p_{i+j}) p_{i+k} < \infty$.

In this paper, we will concentrate exclusively on independent Bernoulli sequences, with a particular type of "harmonic" sequence for $\{p_i\}$, which allows for explicit computations and also, in some cases, connects the count vector $(Z_1, Z_2, \ldots)$ with the study of rank order statistics and random permutations. In fact, we will assume that $\{p_i\}$ satisfies

$$p_i(1 - p_{i+1}) = p_{i+1} \;\text{ or equivalently }\; p_i - p_{i+1} = p_i p_{i+1} \;\text{ for } n \geq 1. \tag{3}$$

We will avoid the case $p_1 = 0$, since then the only solution to (3) is the the trivial solution $p_i \equiv 0$. We will therefore assume, for the rest of this paper, that $p_1 = 1/(1 + B)$, with $B \geq 0$, so that from (3)

$$p_i = \frac{1}{i + B} \quad \text{for } i \geq 1. \tag{4}$$

We will refer to an independent Bernoulli sequence with $\{p_i\}$ given in (4) as a *$B-$harmonic Bernoulli sequence.* Occasionally, when we wish to emphasize the dependence on B, we will write $Z_{k,B}$ for the count variable Z_k, and $\mathbf{Z}_B$ for the count vector $\mathbf{Z}$. From Remark 1,

$$\begin{aligned} E[Z_{k,B}] \;&\leq\; \sum_{i \geq 1} p_i p_{i+k} \\ &=\; \sum_{i \geq 1} \frac{1}{(i + B)(i + k + B)} < \infty, \end{aligned}$$

and thus $Z_{k,B}$ is finite, for all $k \geq 1$, a.s.

When the counts $(Z_1, Z_2, \ldots)$ are almost-surely finite, their joint distribution becomes an object of interest, especially its dependence on the sequence of probabilities $\{p_i\}$. Around 1996, Persi Diaconis observed that, for $0-$harmonic Bernoulli sequences, the distribution of the count variable Z_1 is Poisson with intensity 1. A few months later [Emery (1996)] gave another proof in an unpublished manuscript. It is known that the count vector $(Z_1, \ldots, Z_k)$ of a $0-$harmonic Bernoulli

sequence can be thought of the limit of the vector $(C_1(n), \ldots, C_k(n))$ of numbers of cycles of different orders among permutations of $\{1, 2, \ldots, n\}$. (More details are given in the next section.) This fact coupled with the classical results (see [Arratia et al. (2003)], [Arratia (1992)]) establish that the joint distribution of the count vector $(Z_1, Z_2, \ldots, Z_k)$, from a $0-$harmonic Bernoulli sequence, is that of independent Poissons with intensities $(1, \frac{1}{2}, \ldots, \frac{1}{k})$, respectively. All these proofs mentioned are based on combinatorial methods.

[Joffe et al. (2002)] considered general $B-$harmonic Bernoulli sequences and obtained the moment generating function of of $Z_{1,B}$ by noticing that $\{(S_i, X_{i+1}),\ i = 1, 2, \ldots\}$ forms an inhomogeneous Markov chain, where $S_i = \sum_{m=1}^{i} X_m X_{m+1}$. From this they identified the distribution of Z_1 as a Generalized Hypergeometric Factorial (GHF) law which is more easily stated as a Beta-mixture of Poisson distributions.

In this paper we consider general $B-$harmonic Bernoulli sequences and obtain the joint distribution P_B of the count vector $\mathbf{Z}_B = (Z_{1,B}, Z_{2,B}, \ldots)$. With the addition of another random variable V, the joint distribution Q_B of $(V, \mathbf{Z}_B)$ can be described as follows: the distribution of V is Beta with parameters $(B, 1)$ and the conditional distribution $P_{B,v}$ of $\mathbf{Z}_B$ given $V = v$, is that of independent Poissons with intensities $(1 - v),\ (1 - v^2)/2,\ (1 - v^3)/3, \ldots$. These results are contained in Theorem 2.

We also compute the covariance of $Z_{k,B}$ and $Z_{m,B}$ for $k \leq m$ and note that it is positive for $B > 0$ in Corollary 2. We also show that P_B has the FKG or the positive association property in Theorem 3. There are intuitions for both positive and negative correlations between $Z_{k,B}$ and $Z_{m,B}$ and so this result is perhaps of interest. A plausible justification for positive correlations arises from the feeling that more completed k-strings allow one to "start over" more times in the Bernoulli sequence and so can lead to more strings of length m. Although with the interpretation of $Z_{k,B}$ as the number of cycles of length k among random permutations of $E_{n,B} = \{1, 2, \ldots, n + B\}$ when $B \geq 0$ is an integer (see the next section), the "age-dependent"-cycle count mapping gives perhaps the opposite interpretation. Namely, with more k-cycles formed, there should be less "room" for m-cycles to form in $E_{n,B}$, leading to negative association between $Z_{k,B}$ and $Z_{m,B}$. One may think, however, for fixed $k < m$ much smaller than $n \uparrow \infty$, that such "boundary" considerations are negligible and the first explanation is more reasonable given that the mixture distribution is of Beta type which has interpretations with respect to "reinforcement" dynamics (e.g. Polya urns). On the other hand, since the asymptotic joint distribution depends on B, we know that the "boundary" is not completely ignored in the limit, thereby confusing the matter once more. It would be of interest to have a better understanding of these dependence issues.

Our methods avoid the use of combinatorial techniques. We first show, in Lemma 2, that factorial powers of count variables $Z_{k,B}$, which are sums of indicator variables $Y_{i,k}$ (see (2)) can be expressed as simple sums of products of the $Y_{i,k}$'s. For $B-$harmonic Bernoulli sequences, many products of the form $Y_{i,k}Y_{j,k}$ vanish and there are some independence properties among the $Y_{i,k}$'s; see (6), (7) and (8). These are exploited in Lemma 1, Lemma 2 and Lemma 3 to obtain the joint factorial moments of $(Z_{1,B}, \ldots, Z_{n,B})$ in the main theorem (Theorem 1) which is further simplified in Theorem 2 by recognizing it as the sum of probabilities of inequalities among independent exponential variables. The joint distribution of $(Z_{1,B}, \ldots, Z_{n,B})$ can be deduced from this simplified expression for the factorial moments.

Even though the frequency of wait times between 1's of all orders are finite a.s., it is interesting to note that there are infinitely many 1's in the original Bernoulli

sequence (since $\sum_{i\geq 1} p_i = \sum_{i\geq 1} 1/(i+B) = \infty$). However, the events (i.e. 1's) are so sparse that the wait to the first event has infinite mean when $B > 0$. Let $N = \inf\{i \geq 1 : X_i = 1\}$ be the wait to the first event. Then $P(N = k) = B/[(k-1+B)(k+B)]$ when $B > 0$, and though $P(N < \infty) = 1$ we have $E[N] = \infty$. In a similar fashion, when $B = 0$, $X_1 = 1$ a.s. and the wait for the second event has infinite expectation. It is also not difficult to see that, no matter the value of $B \geq 0$, the number of 1's, $N_n = \sum_{i=1}^{n} X_i$, satisfies $N_n/\log n \to 1$ a.s., and $(N_n - \log n)/\sqrt{\log n} \xrightarrow{d} N(0,1)$ (cf. Example 4.6, Ch. 2 [Durrett (1995)]).

Finally, a statistician may ask whether the value of B can be consistently estimated from the count vector Z of all k-strings. To say that this can be done is the same as saying that P_B and $P_{B'}$ are mutually singular for $B \neq B'$. Let M_B be the joint distribution of a $B-$harmonic Bernoulli sequence $\{X_i, i = 1, 2, \ldots\}$. We show in Theorem 4, by use of Kakutani's criterion, that M_B and $M_{B'}$ are absolutely continuous with respect to each other for $B \neq B'$. This implies the same for P_B and $P_{B'}$, and thus B cannot be consistently estimated from $\mathbf{Z}$.

2. Related areas

Count vectors of k-strings as described above, apart from being objects of intrinsic research interest, have concrete interpretations with respect to combinatorics, genetics, ecology, statistics, and other areas (cf. [Arratia et al. (2003)], [Johnson et al. (1992)], and [Antzoulakos and Chadjiconstantinidis (2001)] and references therein). We will describe some connections to rank orders, record values and permutations for the case when $B \geq 0$ is an integer. In both situations, there is an embedded sequence of independent Bernoulli r.v.'s with respect to which the counts of k-strings have various interpretations.

Rank orders and record values. Let $\{\xi_n : n \geq 1\}$ be a sequence of i.i.d. r.v.'s with common continuous distribution function F. One might think of ξ_n as the amount of rainfall or the flood level in the nth year. Let $\xi_{1,n} < \xi_{2,n} < \cdots < \xi_{nn}$ be the ordered values of $\{\xi_i : 1 \leq i \leq n\}$ and define $R_n = j$ if $\xi_n = \xi_{j,n}$. It is a well known theorem of Renyi that $\{R_n : n \geq 1\}$ are independent and uniformly distributed on their respected ranges (cf. Example 6.2, Ch. 1 [Durrett (1995)]). Let $\{a_1, a_2, \ldots\}$ be a sequence of integers such that $1 \leq a_n \leq n$ and define $X_n = I(R_n = a_n)$. The sequence $\{X_n,\ n \geq 1\}$ is an example of a $0-$harmonic Bernoulli sequence, for any choice of the sequence $\{a_1, a_2, \ldots\}$. The sequence $\{X_{n,B} = X_{n+B}, n \geq 1\},\ n \geq 1\}$ is an example of a $B-$harmonic Bernoulli sequence when $B \geq 0$ is an integer.

In the special case $a_n = n$ for $n \geq 1$, the event $X_{n,B} = 1$ means that a record, with respect to the rainfall amounts in the first B years (which were lost or not properly recorded), was set during the year $n + B$. In this case, $Z_{k,B}$ is the number of times records were set after a wait of $k - 1$ years from a previous record.

Of course, by choosing $\{a_n\}$ differently, one can vary the interpretation of $Z_{n,B}$.

Random permutations. For $B \geq 0$ an integer, let $E_{n,B} = \{1, 2, \ldots, n + B\}$. We now describe the "Feller" algorithm which chooses a permutation $\pi : E_{n,B} \to E_{n,B}$ uniformly from the $(n+B)!$ possible permutations (cf. Section 4 [Joffe et al. (2002)], Chapter 1 of [Arratia et al. (2003)]).

1. Draw the first element uniformly from $E_{n,B}$ and call it $\pi(1)$. If $\pi(1) = 1$, a cycle of length 1 has been completed. If $\pi(1) = j \neq 1$, make a second draw uniformly from $E_{n,B} \setminus \{\pi(1)\}$ and call it $\pi(\pi(1)) = \pi(j)$. Continue drawing elements naming

them $\pi(\pi(j))), \pi((\pi(\pi(j)))), \ldots$ from the remaining numbers until 1 is drawn, at which point a cycle (of some length) is completed.

2. From the elements left after the first cycle is completed, $E_{n,B} \setminus \{\pi(1), \ldots, 1\}$, follow the process in step 1 with the smallest remaining number taking the role of "1." Repeat until all elements of $E_{n,B}$ are exhausted.

When $B = 0$, n such Feller draws produces a random permutation, $\pi : E_{n,0} \to E_{n,0}$. However, when $B > 0$, in n such Feller draws, $\pi : E_{n,B} \to E_{n,B}$ is only injective, and there may be the possibility that no cycle of any length is completed.

Let now $\{I_i^{(n)} : 1 \leq i \leq n\}$ be the indicators of when a cycle is completed at the ith drawing in n Feller draws from $E_{n,B}$. It is not difficult to see that $\{I_i^{(n)}\}$ are independent Bernoulli random variables with $P(I_i^{(n)} = 1) = 1/(n + B - i + 1)$, since at time i, independent of the past, there is exactly one choice among the remaining $n + B - i + 1$ members left in $E_{n,B}$ to complete the cycle (to paraphrase Example 5.4, Ch. 1 [Durrett (1995)]).

For $1 \leq k \leq n$, let $D_{k,B}^{(n)}$ be the number of cycles of length k in the first n Feller draws from $E_{n,B}$. It is easy to see that

$$D_{k,B}^{(n)} \xrightarrow{p} Z_{k,B} \quad \text{for } k \geq 1$$

and we give a quick proof below.

Indeed, since a cycle of length k is finished on the mth draw, for $m \geq k + 1$, exactly when $I_{m-k}^{(n)}(1 - I_{m-k+1}^{(n)}) \cdots (1 - I_{m-1}^{(n)}) I_m = 1$, and also since the first cycle is a k-cycle exactly when $(1 - I_1^{(n)})(1 - I_2^{(n)}) \cdots (1 - I_{k-1}^{(n)}) I_k^{(n)} = 1$, we have

$$D_{k,B}^{(n)} = (1 - I_1^{(n)})(1 - I_2^{(n)}) \cdots (1 - I_{k-1}^{(n)}) I_k^{(n)} + \sum_{i=1}^{n-k} I_i^{(n)} (1 - I_{i+1}^{(n)}) \cdots (1 - I_{i+k-1}^{(n)}) I_{i+k}^{(n)}.$$

Let $\{X_i : i \geq 1\}$ be independent Bernoulli random variables defined on a common space with $P(X_i = 1) = 1/(i + B)$, so that $X_i = I_{n-i+1}^{(n)}$ in law for $1 \leq i \leq n$. We can then write $D_{k,B}^{(n)}$ equivalently in distribution as

$$D_{k,B}^{(n)} \stackrel{d}{=} \sum_{i=1}^{n-k} X_i (1 - X_{i+1}) \cdots (1 - X_{i+k-1}) X_{i+k} \; + \; X_{n-k+1} \prod_{j=n-k+2}^{n} (1 - X_j).$$

As $\lim_{n\to\infty} X_{n-k+1}(1 - X_{n-k+2}) \cdots (1 - X_n) = 0$ in probability, we have

$$D_{k,B}^{(n)} \xrightarrow{p} \sum_{i \geq 1} X_i (1 - X_{i+1}) \cdots (1 - X_{i+k-1}) X_{i+k} \; = Z_{k,B}. \tag{5}$$

We see from this construction, that $Z_{k,B}$ represents the asymptotic number of "young" or "age-dependent" k-cycle numbers, that is, those formed in the first n Feller draws from sets of size $n + B$.

3. Preliminary lemmas

We will use the following standard definition of the factorial power of order r of an integer a:

$$a^{[r]} = \begin{cases} a(a-1) \cdots (a-r+1) & \text{when } a, r \geq 1 \\ 1 & \text{when } r = 0 \\ 0 & \text{when } a = 0. \end{cases}$$

Equation (2) gives a representation for the count variable Z_k of k-strings as a series of dependent summands $Y_{i,k}$, defined in (1) in terms of the $B-$harmonic Bernoulli sequence $\{X_i, i \geq 1\}$. The summands $\{Y_{i,k}, i \geq 1\}$ are indicator variables with the following useful properties

$$Y_{i,k}^2 = Y_{i,k}, \quad Y_{i,k}Y_{i,k'} = 0 \text{ if } k \neq k', \quad Y_{i,k}Y_{i',k'} = 0 \text{ for } i+1 \leq i' < i+k, \tag{6}$$

$$Y_{i,k} \text{ and } Y_{i+k+j,m} \text{ are independent for } j \geq 1, \tag{7}$$

$$\left.\begin{array}{l} E(Y_{i,k}) = \frac{1}{(i+k-1+B)(i+k+B)}, \text{ and} \\ E(Y_{i,k}Y_{i+k,m}) = \frac{1}{(i+k-1+B)(i+k+m-1+B)(i+k+m+B)}. \end{array}\right\} \tag{8}$$

These properties allow us to give simplified expressions for products of factorial powers of the count vector $(Z_1, \ldots, Z_n)$ in terms of $\{Y_{i,k}\}$.

The following lemma gives a representation for the factorial power of a sum of arbitrary indicator variables.

Lemma 1. *Let $(I_1, I_2, \ldots)$ be indicator variables, and let $Z = \sum_{i \geq 1} I_i$ be their sum. Then for integers $r \geq 1$, the factorial powers of Z have the following representation:*

$$Z^{[r]} = \sum_{\substack{i_1,\ldots,i_r \\ \text{distinct}}} I_{i_1} I_{i_2} \cdots I_{i_r} = r! \sum_{1 \leq i_1 < \ldots < i_r} I_{i_1} I_{i_2} \cdots I_{i_r}. \tag{9}$$

Proof. The proof is by induction. For $r = 1$, the identity in (9) is obvious. Now assume that the same identity holds for $r-1$, with $r \geq 2$. Write

$$\begin{aligned} Z^{[r]} &= (Z-(r-1)) \cdot Z^{[r-1]} \\ &= (Z-(r-1)) \cdot \sum_{\substack{i_1,\ldots,i_{r-1} \\ \text{distinct}}} I_{i_1} \cdots I_{i_{r-1}}. \end{aligned}$$

Since I_j is $0-1$ valued, $I_j^2 = I_j$ for all j, and we have

$$\begin{aligned} Z \sum_{\substack{i_1,\ldots,i_{r-1} \\ \text{distinct}}} I_{i_1} \cdots I_{i_{r-1}} &= \left[\sum_{i_r} I_{i_r}\right]\left[\sum_{\substack{i_1,\ldots,i_{r-1} \\ \text{distinct}}} I_{i_1} \cdots I_{i_{r-1}}\right] \\ &= (r-1) \sum_{\substack{i_1,\ldots,i_{r-1} \\ \text{distinct}}} I_{i_1} \cdots I_{i_{r-1}} + \sum_{\substack{i_1,\ldots,i_r \\ \text{distinct}}} I_{i_1} \cdots I_{i_{r-1}} I_{i_r}. \end{aligned}$$

Thus

$$Z^{[r]} = \sum_{\substack{i_1,\ldots,i_r \\ \text{distinct}}} I_{i_1} \cdots I_{i_r}.$$

This establishes the identity in (1) for r and completes the proof of Lemma 1. □

Lemma 1 can be used to obtain expressions of products of factorial powers of count vectors in a routine way. Lemma 2 will improve on this and give an alternative expression for such a product, by exploiting property (6) of $\{Y_{i,k}\}$. To state this result we will need the following notation.

Let $k_1, k_2, \ldots, k_n$ be distinct integers and let $r_1, r_2, \ldots, r_n$ be (not necessarily distinct) integers all of which are greater than or equal to 1. Let $R_0 = 0, R_m = \sum_1^n r_j, m = 1, \cdots, n$ and let $A_n = \{\lambda_l\}_{l=1}^{R_n} = \{\underbrace{k_1, \ldots, k_1}_{r_1}, \underbrace{k_2, \ldots, k_2}_{r_2}, \ldots, \underbrace{k_n, \ldots, k_n}_{r_n}\}$.

Let $\mathcal{S}_{A_n}$ be the $R_n!$ permutations of A_n, though there are only $\binom{R_n}{r_1,r_2,\cdots,r_n}$ distinct permutations. Finally, for $\pi \in \mathcal{S}_{A_n}$, let

$$S_m(\pi) \;=\; \sum_{j=1}^{m} \pi_j \text{ for } 1 \le m \le R_n. \tag{10}$$

Lemma 2. *For $n \ge 1$, let $k_1, \ldots, k_n \ge 1$ be distinct integers and $r_1, \ldots, r_n \ge 1$ be (not necessarily distinct) integers. Then,*

$$Z_{k_1}^{[r_1]} \cdots Z_{k_n}^{[r_n]} \;=\; \sum_{\pi \in \mathcal{S}_{A_n}} \sum_{1 \le i_1 < \cdots < i_{R_n}} Y_{i_1,\pi_1} Y_{i_2,\pi_2} \cdots Y_{i_{R_n},\pi_{R_n}}. \tag{11}$$

Proof. From Lemma 1 and (6), we get

$$\begin{aligned} Z_{k_1}^{[r_1]} \cdots Z_{k_n}^{[r_n]} &= \prod_{j=1}^{n} \sum_{\substack{i_{R_{j-1}+1}, \cdots, i_{R_j} \\ \text{distinct}}} Y_{i_{R_{j-1}+1},k_j} \cdots Y_{i_{R_j},k_j} \\ &= \sum_{\substack{i_1, \cdots, i_{R_n} \\ \text{distinct}}} Y_{i_1,k_1} \cdots Y_{i_{R_1},k_1} \cdots \;\cdots Y_{i_{R_{n-1}+1},k_n} \cdots Y_{i_{R_n},k_n} \\ &= \sum_{\pi \in \mathcal{S}_{A_n}} \sum_{1 \le i_1 < \cdots < i_{R_n}} Y_{i_1,\pi_1} Y_{i_2,\pi_2} \cdots Y_{i_{R_n},\pi_{R_n}}. \end{aligned}$$

This completes the proof of Lemma 2. □

For a vector of integers $\mathbf{k} = (k_1, k_2, \ldots)$ with $k_n \ge 1$ for all n, define $K_m = \sum_{j=1}^{m} k_j$ to be the partial sums, $k(r,s) = (k_r, k_{r+1}, \ldots, k_s)$ to be the segment from r to s. For $1 \le m \le n$ and $r \ge 1$, define

$$C(r : k(m,n)) \;=\; \sum_{r \le i_m < i_{m+1} < \cdots < i_n} Y_{i_m,k_m} Y_{i_{m+1},k_{m+1}} \cdots Y_{i_n,k_n}.$$

The following is a key lemma which gives two identities useful for the calculation of factorial moments of the count vector $(Z_{1,B}, \ldots, Z_{k,B})$.

Lemma 3. *For integers $r, n \ge 1$ and vectors $\mathbf{k}$ the following two identities hold:*

$$E[Y_{r,k_1} C(r+1; k(2,n+1)] \;=\; \prod_{m=1}^{n+1} \frac{1}{r-1+K_m+B} - \prod_{m=1}^{n+1} \frac{1}{r+K_m+B}, \tag{12}$$

and

$$E[C(r; k(1,n)] \;=\; \prod_{m=1}^{n} \frac{1}{r-1+K_m+B}. \tag{13}$$

Proof. The proof is by simultaneous induction for both (12) and (13) on n, the number of $Y_{i,k}$ factors in $C(r : k(l,m))$ where $m - l + 1 = n$. Throughout, we will rely heavily on the properties (6),(7) and (8) of $\{Y_{i,k}\}$.

We will now establish (12) for $n = 1$. Notice that

$$\begin{aligned} &E\big[Y_{r,k_1} C(r+1; k(2,2))\big] \\ &\quad = \sum_{i \ge r+1} E[Y_{r,k_1} Y_{i,k_2}] = \sum_{i \ge r+k_1} E[Y_{r,k_1} Y_{i,k_2}] \end{aligned}$$

$$
\begin{aligned}
&= E[Y_{r,k_1}Y_{r+k_1,k_2}] + \sum_{i\geq r+k_1+1} E[Y_{r,k_1}]E[Y_{i,k_2}] \\
&= \frac{1}{(r+k_1-1+B)(r+K_2-1+B)(r+K_2+B)} \\
&\quad + \frac{1}{(r+k_1-1+B)(r+k_1+B)} \sum_{i\geq r+k_1+1} \frac{1}{(i+k_2-1+B)(i+k_2+B)} \\
&= \frac{1}{(r+k_1-1+B)(r+K_2-1+B)(r+K_2+B)} \\
&\quad + \frac{1}{(r+k_1-1+B)(r+k_1+B)(r+K_2+B)} \\
&= \frac{1}{(r-1+k_1+B)(r-1+K_2+b)} - \frac{1}{(r+k_1+B)(r+K_2+B)}.
\end{aligned}
$$

This establishes (12) for $n = 1$.

Next,

$$
\begin{aligned}
E\big[C(r;k(1,1))\big] &= \sum_{i_1\geq r} E[Y_{i,k_1}] &&= \sum_{i\geq r} \frac{1}{(i+k_1-1+B)(i+k_1+B)} \\
&&&= \sum_{i\geq r} \left[\frac{1}{(i+k_1-1+B)} - \frac{1}{(i+k_1+B)}\right] \\
&&&= \frac{1}{r-1+k_1+B}
\end{aligned}
$$

which establishes (13) for $n = 1$.

For the induction step, let $N \geq 2$ and assume that (12) and (13) hold for $n = N-1$. We first establish (13) for $n = N$ by using the validity of (12) for $n = N-1$ as follows:

$$
\begin{aligned}
E\big[C(r;k(1,N))\big] &= E\left[\sum_{r\leq i_1<\cdots<i_N} Y_{i_1,k_1}\cdots Y_{i_N,k_N}\right] \\
&= E\left[\sum_{r\leq i} Y_{i,k_1}\left[\sum_{i+1\leq i_2<\cdots<i_N} Y_{i_2,k_2}\cdots Y_{i_N,k_N}\right]\right] \\
&= \sum_{r\leq i}\left[\prod_{m=1}^{N}\frac{1}{i+K_m-1+B} - \prod_{m=1}^{N+1}\frac{1}{i+K_m+B}\right] \\
&= \prod_{m=1}^{N}\frac{1}{r+K_m-1+B}.
\end{aligned}
$$

To finish the induction we now proceed to establish (12) for $n = N$, assuming that (12) holds for $n = N-1$ and (13) holds for $n = N$. Notice that

$$
\begin{aligned}
E[Y_{r,k_1}\mathcal{C}(r+1;k(2,N+1))] &= E[Y_{r,k_1}\mathcal{C}(r+k_1;k(2,N+1))] \\
&= E[Y_{r,k_1}Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))] \\
&\quad + E[Y_{r,k_1}]E[\mathcal{C}(r+k_1+1;k(2,N+1))].
\end{aligned}
$$

By conditioning on X_{r+k_1} and noting that many terms vanish when $X_{r+k_1} = 0$, the first term above simplifies as follows:

$$
\begin{aligned}
&E[Y_{r,k_1}Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))]\\
&\quad = E[Y_{r,k_1}E[Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))|X_r,\ldots,X_{r+k_1}]]\\
&\quad = E[Y_{r,k_1}E[Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))|X_{r+k_1}]]\\
&\quad = E[E[Y_{r,k_1}|X_{r+k_1}]E[Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))|X_{r+k_1}]]\\
&\quad = E[Y_{r,k_1}|X_{r+k_1}=1]\\
&\qquad \cdot E[Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))|X_{r+k_1}=1]P(X_{r+k_1}=1)\\
&\quad = E[Y_{r,k_1}|X_{r+k_1}=1]E[Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))].
\end{aligned}
$$

The assumption that (12) and (13) hold for $n = N-1$ yields

$$
\begin{aligned}
&E[Y_{r,k_1}\mathcal{C}(r+1;k(2,N+1))]\\
&\quad = E[Y_{r,k_1}|X_{r+k_1}=1]E[Y_{r+k_1,k_2}\mathcal{C}(r+K_2;k(3,N+1))]\\
&\qquad + E[Y_{r,k_1}]E[\mathcal{C}(r+k_1+1;k(2,N+1))]\\
&\quad = \frac{1}{r+k_1-1+B}\Big[\prod_{m=2}^{N+1}\frac{1}{r+K_m-1+B}-\prod_{m=2}^{N+1}\frac{1}{r+K_m+B}\Big]\\
&\qquad + \frac{1}{(r+k_1-1+B)(r+k_1+B)}\prod_{m=2}^{N+1}\frac{1}{r+K_m+B}\\
&\quad = \frac{1}{r+k_1-1+B}\Big[\prod_{m=2}^{N+1}\frac{1}{r+K_m-1+B}-\prod_{m=2}^{N+1}\frac{1}{r+K_m+B}\Big]\\
&\qquad + \frac{1}{r+k_1-1+B}\prod_{m=2}^{N+1}\frac{1}{r+K_m+B}-\prod_{m=1}^{N+1}\frac{1}{r+K_m+B}\\
&\quad = \prod_{m=1}^{N+1}\frac{1}{r+K_m-1+B}-\prod_{m=1}^{N+1}\frac{1}{r+K_m+B}.
\end{aligned}
$$

This establishes (12) for $n = N$ and completes the proof of the lemma. □

4. Main results and corollaries

Consider a $B-$harmonic Bernoulli sequence and the corresponding count vector $\mathbf{Z}_B$. For non-negative integers $s_1, s_2, \ldots, s_n$, define

$$\mu_B(s_1,\ldots,s_n) \;=\; E(Z_{1,B}^{[s_1]}Z_{2,B}^{[s_2]}\cdots Z_{n,B}^{[s_n]}).$$

The following theorem gives an explicit form for the factorial moments of this count vector which will be used to identify its joint distribution.

Theorem 1. *Let $\mathbf{Z}_B$ be the count vector arising from a $B-$harmonic Bernoulli sequence $\{X_i\}$. Let $k_1,\ldots,k_n$ be distinct integers and let $r_1,\ldots,r_n$ be not necessarily distinct integers, all greater than or equal to 1. Recall the notations $R_m, A_n, \mathcal{S}_{A_n}$ and $S_m(\pi)$ from just before (10). Then*

$$E\big[Z_{k_1,B}^{[r_1]}Z_{k_2,B}^{[r_2]}\cdots Z_{k_n,B}^{[r_n]}\big] \;=\; \sum_{\pi\in\mathcal{S}_{A_n}}\prod_{m=1}^{R_n}\frac{1}{S_m(\pi)+B} \tag{14}$$

Proof. From Lemmas 2 and 3, using the notation in (10),

$$\begin{aligned} E\big[Z_{k_1,B}^{[r_1]} Z_{k_2,B}^{[r_2]} \cdots Z_{k_n,B}^{[r_n]}\big] &= E\Big[\sum_{\pi\in\mathcal{S}_{A_n}} \sum_{1\le i_1<\cdots<i_{R_n}} Y_{i_1,\pi_1} Y_{i_2,\pi_2} \cdots Y_{i_{R_n},\pi_{R_n}}\Big] \\ &= E\Big[\sum_{\pi\in\mathcal{S}_{A_n}} C(1;\pi(1,R_n))\Big] \\ &= \sum_{\pi\in\mathcal{S}_{A_n}} \prod_{m=1}^{R_n} \frac{1}{S_m(\pi)+B}. \end{aligned}$$

This completes the proof of the theorem. □

The next theorem, which is the main result of this paper, gives the factorial moments of $(Z_{1,B},\cdots,Z_{N,B})$ for $B-$harmonic Bernoulli sequences and deduces the structure of the joint distribution of $\mathbf{Z}_B$.

Theorem 2. *For non-negative integers $s_1,\ldots,s_N$,*

$$\mu_B(s_1,\ldots,s_N) = \int_0^1 B v^{B-1} \prod_{j=1}^{N} \left(\frac{(1-v^j)}{j}\right)^{s_j}. \tag{15}$$

This implies that the joint distribution P_B of $\mathbf{Z}_B$ has the following structure: there is random variable V and the joint distribution Q_B of $(V,\mathbf{Z}_B)$ can be described as follows: V has a Beta$(B,1)$ distribution (which is the point mass at 0 when $B=0$) and given $V=v$, the conditional distribution $P_{B,v}$ of $(Z_{1,B},Z_{2,B},\ldots)$ is that of independent Poissons with intensities $1-v, \frac{1-v^2}{2},\ldots$ respectively.

Proof. First, let $B>0$ as the case $B=0$ is analogous or can be obtained by taking the limit $B\downarrow 0$. Second, to establish (15), we can assume that some $s_m>0$ for some m. In fact, let $(s_{k_1},\ldots,s_{k_n})$ be the vector formed from the non-zeros in $(s_1,s_2,\ldots,s_N)$, and let R_n, A_n, S_{A_n} and $S_m(\pi)$ for $\pi\in S_{A_n}$ be as defined near (10). Let also $W_0,W_1,W_2,\ldots,W_{R_n}$ be independent exponential r.v.'s with failure rates $B,\lambda_1,\ldots,\lambda_{R_n} \stackrel{def}{=} B,\underbrace{k_1,\ldots,k_1}_{r_1},\ldots,\underbrace{k_n,\ldots,k_n}_{r_n}$, respectively. Then, for any $\pi\in\mathcal{S}_{A_n}$

$$\prod_{m=1}^{R_n} \frac{\pi_m}{S_m(\pi)+B} = \prod_{m=1}^{R_n} \frac{\lambda_{\pi_m}}{S_m(\pi)+B} = P(W_{\pi_{R_n}} < W_{\pi_{R_n-1}} < \cdots < W_{\pi_1} < W_0). \tag{16}$$

From Theorem 1 and (16), we conclude

$$\begin{aligned} \Big(\prod_{j=1}^{N} j^{s_j}\Big)\cdot \mu_B(s_1,\ldots,s_N) &= \Big(\prod_{j=1}^{n} (k_j)^{s_{k_j}}\Big)\cdot E_B(Z_{k_1,B}^{[s_{k_1}]}\cdots Z_{k_n,B}^{[s_{k_n}]}) \\ &= \sum_{\pi\in S_{A_n}} \prod_{m=1}^{R_n} \frac{\pi_m}{S_m(\pi)+B} \\ &= \sum_{\pi\in S_{A_n}} P(W_{\pi_{R_n}} < \cdots < W_{\pi_1} < W_0) \\ &= P(\max(W_1,\ldots,W_{R_n}) < W_0) \end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty Be^{-By} \prod_{j=1}^{n} (1 - e^{-k_j y})^{s_{k_j}} \, dy \\
&= \int_0^1 Bv^{B-1} \prod_{j=1}^{N} (1 - v^j)^{s_j} \, dv \\
&= \left(\prod_{j=1}^{N} j^{s_j} \right) \cdot \int_0^1 Bv^{B-1} \prod_{j=1}^{N} E(Z_{j,v}^{[s_j]}) dv
\end{aligned}
$$

where, for each v, $Z_{1,v}, Z_{2,v}, \ldots$ are independent Poisson random variables with means $(1-v), (1-v^2)/2, \ldots$, respectively. This establishes the structure of P_B as desired. □

Remark 2. We now indicate an alternate argument to obtain Theorem 2. Consider the factorial moment generating function

$$\phi_B(t_1, \ldots, t_n) \stackrel{def}{=} \sum_{r_1, \ldots, r_n \geq 0} \mu_B(r_1, \ldots, r_n) \frac{{t_1}^{r_1} \ldots {t_n}^{r_n}}{r_1! \ldots r_n!}.$$

The denominator of the last factor in (14), $S_{R_n}(\pi) + B$, is the same for all values of π and equals $\sum_1^n r_j k_j + B$. Hence, we have the recurrence relation

$$\mu_B(r_1, \ldots, r_n) = \sum_1^n r_j \mu_B(r_1, \ldots, r_j - 1, \ldots, r_n)$$

which in turn leads to the partial differential equation

$$\sum_{j=1}^{n} j t_j \frac{\partial \phi_B}{\partial t_j} \;=\; (\sum_1^n t_j - B)\phi_B + B. \tag{17}$$

Also, the marginal factorial moment generating function $\phi_{j,B}(t_j)$ of $Z_{j,B}$ satisfies $jt_j \partial \phi_{j,B}(t_j)/\partial t_j = (t_j - B)\phi_{j,B}(t_j) + B$ with the boundary condition $\phi_{j,B}(0) = 1$. Its unique solution is $\phi_{j,B}(t_j) = \int_0^1 Bv^{B-1} \exp\{t_j(1-v^j)/j\} Bv^{B-1} dv$. Then, the boundary conditions for the equation in (17) are $\phi_B(0, \ldots, 0, t_j, 0, \ldots, 0) = \phi_{j,B}(t_j)$ for $1 \leq j \leq n$. It can be checked that equation (17) has a unique solution, namely

$$\phi_B(t_1, \ldots, t_n) \;=\; \int_0^1 Bv^{B-1} \exp\left\{ \sum_{j=1}^{n} \frac{t_j}{j}(1 - v^j) \right\} dv,$$

which immediately gives the description of the joint distribution of $\mathbf{Z}_B$ in Theorem 2.

We now give some corollaries of the main theorems. The first gives marginal factorial moments of the count $Z_{k,B}$.

Corollary 1. *For a $B-$harmonic Bernoulli sequence,*

$$E(Z_{k,B}^{[r]}) \;=\; \frac{r!}{(k+B)(2k+B)\cdots(rk+B)}$$

Proof. From Theorem 2,

$$E(Z_{k,B}^{[r]}) \;=\; \int_0^1 Bv^{B-1} \left(\frac{1 - v^k}{k} \right)^r = \frac{r!}{(k+B)(2k+B)\cdots(rk+B)}.$$

□

The second corollary computes the covariance between $Z_{k_1,B}$ and $Z_{k_2,B}$.

Corollary 2.

$$cov(Z_{k_1,B}, Z_{k_2,B}) \;=\; \frac{B}{(k_1+B)(k_2+B)(k_1+k_2+B)}.$$

Proof. From (14) in Theorem 1, we have

$$\begin{aligned} E(Z_{k_1,B})E(Z_{k_2,B}) &= \frac{1}{(k_1+B)(k_2+B)} \\ E(Z_{k_1,B}Z_{k_2,B}) &= \frac{1}{(k_1+B)(k_1+k_2+B)} + \frac{1}{(k_2+B)(k_1+k_2+B)} \\ &= \frac{1}{(k_1+B)(k_2+B)} + \frac{B}{(k_1+B)(k_2+B)(k_1+k_2+B)} \end{aligned}$$

This shows that $Z_{k_1,B}$ and $Z_{k_2,B}$ are positively correlated and

$$cov(Z_{k_1,B}, Z_{k_2,B}) \;=\; \frac{B}{(k_1+B)(k_2+B)(k_1+k_2+B)}.$$

□

The FKG or positive association property of P_B is now established.

Theorem 3. *The joint distribution P_B of $\mathbf{Z}$ possesses the FKG property.*

Proof. Let f, g be a bounded functions on R^∞ which are coordinate-wise increasing and are supported on a finite number of coordinates. We need to show that

$$\int f(\mathbf{Z})g(\mathbf{Z}))dP_B \;\geq\; \int f(\mathbf{Z})dP_B \int g(\mathbf{Z})dP_B. \tag{18}$$

It is well known that distributions on the real line and products of measures on the real line possess the FKG property [Liggett (1985)]. Since the Poisson distribution is stochastically increasing in its intensity parameter, the product measure $P_{v,B}$ (cf. Theorem 2) is stochastically decreasing in v. This means that for any bounded increasing function f, $\int f(\mathbf{z})dP_{v,B}$ is decreasing in v. Thus

$$\begin{aligned} \int f(\mathbf{Z})g(\mathbf{Z})dP_B &= \int_0^1 Bv^{B-1} \int f(\mathbf{Z})g(\mathbf{Z})dP_{v,B}\; dv \\ &\geq \int_0^1 Bv^{B-1} \int f(\mathbf{Z})dP_{v,B} \int g(\mathbf{Z})dP_{v,B}\; dv \\ &\qquad \text{since } P_{v,B} \text{ is a product measure} \\ &\geq \int_0^1 Bv^{B-1} \int f(\mathbf{Z})dP_{v,B}dv \cdot \int_0^1 Bv^{B-1} \int g(\mathbf{Z})dP_{v,B}\; dv \\ &\qquad \text{since } \int f(\mathbf{Z})dP_{v,B}, \int g(\mathbf{Z})dP_{v,B}, \text{ decreases in } v \\ &= E_B(f(\mathbf{Z}))E_B(g(\mathbf{Z})). \end{aligned}$$

This completes the proof of this theorem. □

Finally, in the introduction, we stated that the parameter B cannot be estimated from $\mathbf{Z}$. This is a consequence of the fact below.

Theorem 4. *Let M_B be the joint distribution of the $B-$harmonic Bernoulli sequence $\{X_i\}$. Then for $0 \leq B < B'$, the measures M_B and $M_{B'}$ are absolutely continuous with respect to one another.*

Proof. Since $M_B, M_{B'}$ are product measures, we compute the Kakutani dichotomy criterion

$$\prod_{k\geq 1}\left[\frac{1}{\sqrt{(k+B)(k+B')}}+\sqrt{1-\frac{1}{k+B}}\sqrt{1-\frac{1}{k+B'}}\right] = \prod_{k\geq 1}\left(1-\frac{1}{k^2}(1+o(1))\right) > 0.$$

Thus for $B \neq B'$, $M_B << M_{B'}$. This also implies that $P_B = M_B\mathbf{Z}^{-1} << P_{B'} = M_{B'}\mathbf{Z}^{-1}$. This proves this theorem. □

Acknowledgment

We thank Prof. K.B. Athreya who brought to our attention an initial version of the problem, Prof. A. Joffe who sent us [Joffe et al. (2002)], and Fred Huffer who gave the suggestion to use independent exponentials in Theorem 2.

This research was supported in part by grant NSF/DMS – 0071504.

References

[Antzoulakos and Chadjiconstantinidis (2001)] Antzoulakos, D., and Chadjiconstantinidis, S. (2001) Distributions of numbers of success runs of fixed length in Markov dependent trials. *Ann. Inst. Statist. Math.* **53** 599-619. MR1868894

[Arratia et al. (2003)] Arratia, R., Barbour, A. D., and Tavaré, S. (2003) *Logarithmic Combinatorial Structures: A Probabilistic Approach.* EMS Monographs in Mathematics, European Mathematical Society, Zürich. MR2032426

[Arratia (1992)] Arratia, R., and Tavare, S. (1992) The cycle structure of random permutations. *Ann. Probab.* **20** 1567-1591. MR1175278

[Durrett (1995)] Durrett, R. (1995) *Probability: Theory and Examples.* Duxbury, New York.

[Emery (1996)] Emery, M. (1998) *Sur un problème de Diaconis* – Unpublished manuscript.

[Joffe et al. (2002)] Joffe, A., Marchand, E., Perron, F., and Popadiuk, P. (2002) On sums of products of Bernoulli variables and random permutations. *pre-print.*

[Johnson et al. (1992)] Johnson, N. L., Kotz, S., and Kemp, A. W. (1992) *Univariate Discrete Distributions.* Second Edition, Wiley, New York. MR1224449

[Liggett (1985)] Liggett, T. M. (1985) *Interacting Particle Systems.* Springer-Verlag, New York. MR776231