**Counts of Bernoulli Success Strings in a Multivariate Framework** [1]


Djilali AIT AOUDIA[a], Éric MARCHAND[b,2], François PERRON[a]

*a Université de Montréal, Département de mathématiques et de statistique, Montréal, QC, CANADA, H3C 3J7*

*b Université de Sherbrooke, Département de mathématiques, Sherbrooke, QC, CANADA, J1K 2R1*



Summary

Distributional findings are obtained relative to Bernoulli arrays $\{X_{k,j}, k \geqslant 1, j = 1, \ldots, r+1\}$, where the rows $(X_{k,1}, \ldots, X_{k,r+1})$ are independently distributed as Multinomial$(1, p_{k,1}, \ldots, p_{k,r+1})$ for $k \geqslant 1$, with $p_{k,1} = \cdots = p_{k,r} = \frac{1}{b+k}$, $b \geqslant r-1$. The quantities of interest relate to the measure of the number of runs of length 2 and are $\underline{S}_n = (S_{n,1}, \ldots, S_{n,r})$, $\underline{S} = \lim_{n\to\infty} \underline{S}_n$, where $S_{n,j} = \sum_{k=1}^{n} X_{k,j} X_{k+1,j}$. We obtain the mixed binomial moments of both the components of $S_n$ and $\underline{S}$, as well as provide an explicit expression for the probability mass function of $\underline{S}$. Moreover, we derive the multivariate Poisson mixture representation: $S_i | V \sim^{\text{indep.}}$ Poisson$(v_i)$ with $V = (V_1, \ldots, V_r) \sim$ Dirichlet$(1, \ldots, 1, b-r+1)$, as well as establish the distinct mixture representation: $\underline{S}|\alpha \sim p_\alpha$, $\alpha \sim$ Beta$(1, b)$ with $p_\alpha$ a multivariate mass function with Poisson$(\alpha)$ marginals given by $p_\alpha(s_1, \ldots, s_r) = \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \frac{e^{-a} \alpha^{s_1+\ldots+s_r+j}}{(s_1+\ldots+s_r+j)!}$.



*AMS 2010 subject classifications:* 60C05, 60E05, 62E15

*Keywords and phrases*: Arrays; Bernoulli; Binomial moments; Dirichlet; Multinomial; Poisson distribution; Poisson mixtures; Runs.


# 1 Introduction

In many studies of data, one is concerned with the occurrence or non-occurrence of a special event over time. In the counting of species, one wishes to know whether a new catch is from a new species; in the study of records, one wishes to assess the likelihood of a new record or not; in classification, one wishes to know whether a current observation is from a different population than past observations. And so on. Such data reduces to Bernoulli strings of 0's and 1's and the counting of runs, or consecutive 1's has generated much interest and for a long time.

[1] November 5, 2015


[2] Corresponding author: eric.marchand@usherbrooke.ca

Multivariate versions are also of interest in situations, namely for situations with $r$ connected strings and where a motivating question is to describe the nature of the dependence. Bernoulli arrays were introduced by Ait Aoudia and Marchand (2010) and consist of entries $\{X_{k,j}, k \geqslant 1, j = 1, \ldots, r+1\}$, where the rows $\underline{X}_k = (X_{k,1}, \ldots, X_{k,r+1})$ are independently distributed as Multinomial$(1, p_{k,1}, \ldots, p_{k,r+1})$, $k \geqslant 1$. Such arrays arise in sampling one object at a time from an urn with $r+1$ colours, with replacement and reinforcement (i.e., adding balls as in a Polyà urn). Quantities of interest include the number of runs $S_{n,j} = \sum_{k=1}^{n} X_{k,j} X_{k+1,j}$ of length 2 in the $j^{\text{th}}$ column, their limits $S_j = \lim_{n\to\infty} S_{n,j}$, and namely the random vectors

$$\underline{S}_n = (S_{n,1}, \ldots, S_{n,r}), \text{ and } \underline{S} = \lim_{n\to\infty} \underline{S}_n . \tag{1.1}$$

For the univariate case, an elegant result consists of the Poisson mixture representation

$$S_1 | U \sim Poisson(aU) \text{ with } U \sim Beta(a,b), \text{ for } p_{k,1} = \frac{a}{a+b+k-1}, \, a > 0, b \geqslant 0, \tag{1.2}$$

as obtained by Mori (2001), as well as Holst (2008). This class of solutions contains degenerate cases with $b = 0$, i.e., $\mathbb{P}(U = 1) = 1$, and where $S_1$ is Poisson distributed with mean parameter $a$. Moreover, the $a = 1, b = 0$ Poisson(1) solution was recognized in the mid 1990's by Persi Diaconis, as well as Hahlin (1995), and earlier versions are due to Arratia, Barbour, Tavaré (1992), Kolchin (1971), and Goncharov (1944). This elegant result has inspired much interest and lead to various findings relative to the distributions of $S_{n,j}$ and $S_j$ for various other configurations of the $\{p_{k,j}\}$'s, relationships and implications for Pólya urns, records, matching problems, marked Poisson processes, etc, as witnessed by the work of Chern, Hwang and Yeh (2000), Csörgó and Wu (2000), Holst (2007,2008), Huffer, Sethuraman and Sethuraman (2009), Joffe et al. (2004, 2000), Mori (2001), Sethuraman and Sethuraman (2004), among others.

With such Poisson distributions and Poisson mixtures arising naturally in these univariate ($r = 1$) situations, it seems natural to investigate multivariate versions of such models. The findings of Ait Aoudia et al. (2014) are contributions in this direction with analysis for the totals $\sum_{j=1}^{r} S_{n,j}$ and $T = \sum_{j=1}^{r} S_j$, and bivariate representations for the distributions of $\underline{S}_n$ and $\underline{S}$. For instance, they obtain the following multivariate extension of (1.2):

$$T | U \sim Poisson(\frac{aU}{r}) \text{ with } U \sim Beta(a,b), \text{ for } p_{k,l} = \frac{a}{r(a+b+k-1)}, \, a > 0, b \geqslant 0, \tag{1.3}$$

which yields $T \sim$ Poisson$(\lambda)$ for the degenerate case $a = r\lambda$, $b = 0$.

The main findings here are presented in Section 3, are based on the assumption

$$p_{k,1} = \cdots = p_{k,r} = \frac{1}{b+k}\,, \quad \text{with } b \geqslant r-1\,, \tag{1.4}$$

and include: **(i)** explicit expressions (Theorems 3.1 and 3.2) for the mixed binomial moments of $S_{n,1}, S_{n,2}, \ldots, S_{n,r}$; for all $n, r$; and of $S_1, S_2, \ldots, S_r$, **(ii)** an explicit expression for the distribution function of $\underline{S}$ (Theorem 3.2), **(iii)** two mixture representations (Theorem 3.3) for the distribution of $\underline{S}$.

As recalled in Section 2, mixed binomial moments yield expressions for probability mass and generating functions. The most intricate and challenging technical achievement is the binomial moment result of Theorem 3.1, which solves a recursive system of equations involving binomial moments related to the $S_{n,j}$'s and auxiliary variables $W_{n,j}$, defined below in Section 3.1. Solutions to such a system of equations were provided by Ait Aoudia et al. (2014) in the bivariate case $r=2$, but our result is unified and general for $r \geqslant 2$. The mixture representations are fascinating. On one hand, the components $S_i$ of $\underline{S}$ are distributed as independent Poisson$(v_i)$ conditional on $V_1, \ldots, V_r$ which follows a Dirichlet distribution. This neatly summarizes the dependence structure of the $S_j$'s, as well as replicate the marginal Beta mixture representation (1.2). The second mixture representation is quite different as the mixing parameter is univariate Beta and the dependence is reflected otherwise through a multivariate discrete distribution with Poisson and non-independent marginals, and which connects for instance to Schur-constant distributions of Castañer et al. (2015).

# 2 Multivariate Poisson mixtures and mixed binomial moments

For non-negative integer valued random variables $Z_1, Z_2, \ldots, Z_r$ and constants $k_i \in \{0,1,2,\ldots\}$, $i = 1, \ldots, r$, the mixed binomial moments are defined as $\mathbb{E}\left(\prod_{i=1}^r \binom{Z_i}{k_i}\right)$, where $\binom{z}{u}$ is taken to be equal to 0 for $u > z$. Our main distributional findings involve the determination of mixed binomial moments, which imply associated expressions for the probability generating and mass functions. Indeed, the Taylor series expansion about 1 of the probability generating function $\psi_{Z_1}$ having radius of convergence greater than 1 can be expressed as

$$\psi_Z(t) = \mathbb{E}\left[t^{Z_1}\right] = \sum_{k\geqslant 0} \mathbb{E}\binom{Z_1}{k}(t-1)^k, t \in [0,1], \tag{2.1}$$

and the probability mass function (pmf) of $Z$ can be written as

$$\mathbb{P}\left(Z_1 = i\right) = \sum_{k=i}^{\infty}(-1)^{k-i}\binom{k}{i}\mathbb{E}\binom{Z_1}{k}. \tag{2.2}$$

For the multivariate case, corresponding expressions are given by

$$\mathbb{E}[\prod_{i=1}^{r} t_i^{Z_i}] = \sum_{k_1\geqslant 0,\ldots,k_r\geqslant 0} \mathbb{E}\left(\prod_{i=1}^{r}\binom{Z_i}{k_i}\right)\prod_{j=1}^{r}(t_j-1)^{k_i},\ (t_1,t_2,\ldots,t_r)\in[0,1]^r, \tag{2.3}$$

and

$$\mathbb{P}\left(\bigcap_{j=1}^{r}\{Z_j = x_j\}\right) = \sum_{x_1\leqslant k_1,\ldots,x_r\leqslant k_r} (-1)^{\sum_{i=1}^{r}(k_i-x_i)}\mathbb{E}\left(\prod_{i=1}^{r}\binom{Z_i}{k_i}\right)\prod_{i=1}^{r}\binom{k_i}{x_i}, \tag{2.4}$$

as long as the Taylor series expansion at $(t_1,t_2,\ldots,t_r) = (1,1,\ldots,1)$ of the probability generating function converge on an open set containing the origin.

Multivariate Poisson mixtures, defined as distributions of $U = (U_1,\ldots,U_r)$, such that $U_i|V \sim^{indep.}$ Poisson$(V_i)$ and $V = (V_1,\ldots,V_r)$ is a random vector on $[0,\infty)^r$ will play a key role below. For such mixtures, it is easy to verify that mixed binomial moments (assuming they exist) are given by

$$\mathbb{E}\left(\prod_{i=1}^{r}\binom{U_i}{k_i}\right) = \mathbb{E}\left(\prod_{i=1}^{r}\frac{V_i^{k_i}}{k_i!}\right), \tag{2.5}$$

linking binomial moments of the $U_i$'s with mixed moments of the $V_i$'s. The specific mixing densities for $V$ that arise below with our findings are Dirichlet $(a_1,a_2,\ldots,a_{r+1})$; $a_i > 0$; densities given by

$$\frac{\Gamma(\sum_{i=1}^{r+1}a_i)}{\prod_{i=1}^{r+1}\Gamma(a_i)}(1-\sum_{i=1}^{r}v_i)^{a_{r+1}-1}\prod_{j=1}^{r}v_j^{a_j-1}\ \mathbb{I}_D(v)\,,$$

with $D$ is the open $(r-1)$ dimensional simplex defined by $D = \{(v_1,\ldots,v_r) : v_1 \geqslant 0,\ldots,v_r \geqslant 0, v_1+\cdots+v_{r-1}\leqslant 1\}$, their limiting versions as $a_{r+1}\to 0$ which we denote Dirichlet$(a_1,\ldots,a_r,0)$. For multivariate Poisson mixtures with such mixing densities, we will make use of the lemma below. Hereafter, we denote $(\gamma)^{\underline{k}}$ and $(\gamma)_k$ as the descending and ascending factorial with $(\gamma)^{\underline{0}} = (\gamma)_0 = 1$, $(\gamma)^{\underline{k}} = \prod_{j=0}^{k-1}(\gamma-j)$, and $(\gamma)_k = \prod_{j=0}^{k-1}(\gamma+j)$ for $k = 1,2,\ldots$.

**Lemma 2.1.** *Consider a multivariate Poisson mixture distribution $U_i|V \sim^{indep.}$ Poisson$(V_i)$ with mixing variable $V\sim$ Dirichlet$(a_1,\ldots,a_{r+1})$, and $a_i>0, i=1,\ldots,r$. Then,*

**(a)** *$U = (U_1, \dots, U_r)$ has mixed binomial moments*

$$\mathbb{E}\left(\prod_{i=1}^{r}\binom{U_i}{k_i}\right) = \frac{\Gamma(\sum_{i=1}^{r+1} a_i)}{\Gamma(a_{r+1} + \sum_{i=1}^{r}(a_i + k_i))} \prod_{i=1}^{r} \frac{(a_i)_{k_i}}{k_i!},$$

*and, for the particular case with $a_1 = 1, \dots, a_r = 1$, $a_{r+1} = b - r + 1$, we have*

$$\mathbb{E}\left(\prod_{i=1}^{r}\binom{U_i}{k_i}\right) = \frac{1}{(1+b)_k}, \quad \text{with } k = \sum_{i=1}^{r} k_i. \tag{2.6}$$

**(b)** *For cases where $a = \cdots = a_r = 1$ the distribution of $U_1 + \cdots + U_r$ is:* **(i)** *a Poisson mixture with $U_1 + \cdots + U_r | W \sim Poisson(W)$, $W \sim Beta(r, a_{r+1})$ for $a_{r+1} > 0$; and* **(ii)** *$Poisson(1)$ for $a_{r+1} = 0$.*

**Proof.** **(a)** Expression (2.6) follows immediately from the general result, while the latter is obtained by integrating the rhs of (2.5) with respect to the Dirichlet density.
**(b)** From the definition of the $U_i$'s, we have that $U_1 + \cdots + U_r | V \sim \text{Poisson}(W)$ with $W =^d V_1 + \cdots + V_r$. Result **(i)** follows as one obtains readily that $W \sim \text{Beta}(r, a_{r+1})$ for $a_{r+1} > 0$, while result **(ii)** follows since $W$ is degenerate at 1 for $a_{r+1} = 0$. $\square$

# 3 Distributions of $\underline{S}_n$ and $\underline{S}$

## 3.1 Mixed binomial moments of $\underline{S}_n$

We set $S_{0,j} = 0$ for all $j$, and we define the auxiliary random vectors

$$\underline{W}_{n,j} = (S_{n-1,1}, S_{n-1,2}, \dots, S_{n-1,j-1}, W_{n,j}, S_{n-1,j+1}, \dots, S_{n-1,r}),$$

where $W_{n,1}, \dots, W_{n,r}$ are such that

$$W_{n,j} := S_{n-1,j} + X_{n,j}, n \geqslant 1, j \in \{1, \dots, r\}. \tag{3.1}$$

Our goal here is to find the mixed binomial moments of $S_{n,1}, S_{n,2}, \dots, S_{n,r}$ and the mixed binomial moments of $W_{n,i}, i = 1, \dots, r$ and $S_{n-1,j}, j = 1, \dots, r, j \neq i$ and derive explicitly the distribution of $\underline{S}_n$ and $\underline{S}$ under assumption (1.4).

**Theorem 3.1.** *Let $r \geqslant 1$, $k_j(r) \geqslant 0$, $j = 1, \dots, r$, $k(r) = \sum_{j=1}^{r} k_j(r)$, and assume that $k(r) > 0$. Let $r_+ = \sum_{j=1}^{r} \mathbb{1}(k_j(r) > 0)$, $\mathscr{R} = \{1, \dots, r\}$ and $\mathscr{R}(i) = \mathscr{R} \setminus \{i\}$, $i \in \mathscr{R}$. Then, under assumption (1.4), we have for all $r \geqslant 1$*

$$\mathbb{E}\left(\prod_{j \in \mathscr{R}}\binom{S_{n,j}}{k_j(r)}\right) = \begin{cases} \frac{1}{(1+b)_{k(r)}} \frac{(n+1-k(r))^{\overline{r_+}}}{(n+b+1)^{\overline{r_+}}} & n \geqslant k(r) + r_+ - 1, \\ 0 & n < k(r) + r_+ - 1, \end{cases} \tag{3.2}$$

*and*

$$\mathbb{E}\left(\binom{W_{n+1,i}}{k_i(r)}\prod_{j\in\mathscr{R}(i)}\binom{S_{n,j}}{k_j(r)}\right)=\begin{cases}\frac{1}{(1+b)_{k(r)}}\frac{(n+1-k(r))^{\underline{r_+-\mathbb{1}(k_i(r)>0)}}}{(n+1+b)^{\underline{r_+-\mathbb{1}(k_i(r)>0)}}} & n\geqslant k(r)+r_+-1-\mathbb{1}(k_i(r)>0),\\ 0 & n<k(r)+r_+-1-\mathbb{1}(k_i(r)>0),\end{cases}\tag{3.3}$$

*for all* $i\in\mathscr{R}$.

**Proof.** Let $\mathscr{N}=\{1,\ldots,n+1\}$. We have

$$\begin{aligned}X_{m,j} &\in \{0,1\}, \quad (m,j)\in\mathscr{N}\times\mathscr{R},\\ \sum_{j\in\mathscr{R}}X_{m,j} &\leqslant 1 \quad , \quad m\in\mathscr{N}.\end{aligned}$$

The proof will be developed in two steps called STEP ONE (all of the $k_j(r)$'s are greater or equal than 1) and STEP TWO (some of the $k_j(r)$'s are equal to 0).

**STEP ONE.** Consider, for the moment, that $k_j(r)\geqslant 1$, for all $j\in\mathscr{R}$.
If $S_{n,j}\geqslant k_j(r)$, for all $j\in\mathscr{R}$ then $\sum_{m\in\mathscr{N}}X_{m,j}\geqslant k_j(r)+1$, for all $j\in\mathscr{R}$. Therefore,

$$n<k(r)+r-1 \quad \text{implies that} \quad \prod_{j\in\mathscr{R}}\binom{S_{n,j}}{k_j(r)}=0\,.$$

If $i\in\mathscr{R}$, $S_{n,j}\geqslant k_j(r)$ for all $j\in\mathscr{R}(i)$, and $W_{n+1,i}\geqslant k_i(r)$, then $\sum_{m\in\mathscr{N}}X_{m,j}\geqslant k_j(r)+1$ for all $j\in\mathscr{R}(i)$, and $\sum_{m\in\mathscr{N}}X_{m,i}\geqslant k_i(r)$. Therefore,

$$n<k(r)+r-2 \quad \text{implies that} \quad \prod_{j\in\mathscr{R}(i)}\binom{S_{n,j}}{k_j(r)}\times\binom{W_{n+1,i}}{k_i(r)}=0\,.$$

If $n=k(r)+r-2$ and $i\in\mathscr{R}$ then $\prod_{j\in\mathscr{R}(i)}\binom{S_{n,j}}{k_j(r)}\times\binom{W_{n+1,i}}{k_i(r)}$ is a Bernoulli random variable.
The only way to obtain

$$\binom{W_{n+1,i}}{k_i(r)}\prod_{j\in\mathscr{R}(i)}\binom{S_{n,j}}{k_j(r)}=1$$

is by choosing $\{t_1,\ldots,t_{r-1}\}=\mathscr{R}(i)$ and having that

$$X_{m,j}=\begin{cases}1 & \text{if} \quad j=t_s,\ \sum_{\ell\in\mathscr{R}(i):\,\ell<s}(k_\ell(r)+1)<m\leqslant\sum_{\ell\in\mathscr{R}(i):\,\ell\leqslant s}(k_\ell(r)+1), \quad s\in\mathscr{R}(i),\\ 1 & \text{if} \quad j=i,\ k(r)+r-k_i\leqslant m\leqslant k(r)+r-1\\ 0 & \text{otherwise}\end{cases}$$

for $(m,j)\in\mathscr{N}\times\mathscr{R}$. Therefore, we have for $n=k(r)+r-2$,

$$\binom{W_{n+1,i}}{k_i(r)}\prod_{j\in\mathscr{R}(i)}\binom{S_{n,j}}{k_j(r)}\sim\text{Bernoulli}((r-1)!\prod_{m=1}^{n+1}p_m)=\text{Bernoulli}((r-1)!/(1+b)_{k(r)+r-1}).$$

This shows that (3.2) and (3.3) hold for $n \leqslant k(r)+r-2,\ k_1(r),\ldots,k_r(r)>0$. Now, starting with $r=1$, knowing that (3.2) and (3.3) hold for $r=1, n=k_1(1)-1$, proceed by induction assuming that these formulas are true for $n=k_1(1)-1,\ldots,m-1$. Condition on $\underline{X}_{m+1}$ to obtain

$$\begin{aligned}
\mathbb{E}\left(\binom{S_{m,1}}{k_1(1)}\right) &= p_{m+1}\mathbb{E}\left(\binom{W_{m,1}}{k_1(1)}\right)+(1-p_{m+1})\mathbb{E}\left(\binom{S_{m-1,1}}{k_1(1)}\right)\\
&= \frac{1}{(m+b+1)}\frac{1}{(1+b)_{k(1)}}+[1-\frac{1}{m+b+1}]\frac{1}{(1+b)_{k(1)}}\frac{(m-k(1))}{(m+b)}\\
&= \frac{1}{(1+b)_{k(1)}}\frac{(m+1-k(1))}{(m+b+1)}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}\left(\binom{W_{m+1,1}}{k_1(1)}\right)=p_{m+1}\mathbb{E}\left(\binom{W_{m,i}+1}{k_1(1)}\right)+(1-p_{m+1})\mathbb{E}\left(\binom{S_{m-1,1}}{k_1(1)}\right)\\
&= p_{m+1}\Big\{\mathbb{E}\left(\binom{W_{m,i}}{k_1(1)-1}\right)+\mathbb{E}\left(\binom{W_{m,i}}{k_1(1)}\right)\Big\}+(1-p_{m+1})\mathbb{E}\left(\binom{S_{m-1,1}}{k_1(1)}\right)\\
&= \frac{1}{(m+b+1)}\left\{\frac{1}{(1+b)_{k(1)-1}}+\frac{1}{(1+b)_{k(1)}}\right\}+[1-\frac{1}{m+b+1}]\frac{1}{(1+b)_{k(1)}}\frac{(m-k(1))}{(m+b)}\\
&= \frac{1}{(1+b)_{k(1)}}.
\end{aligned}$$

Now, let $\ell>1$. Knowing that formulas (3.2) and (3.3) hold for $r=\ell,\quad n=k(\ell)+\ell-2$, we asume that these formulas hold for $r=\ell,\quad n=k(\ell)+\ell-2,\ldots,m-1$, and seek to establish by induction that they also hold for $r=\ell,\quad n=m$. We condition again on $\underline{X}_{m+1}$ to obtain

$$\begin{aligned}
\mathbb{E}\left(\prod_{j\in\mathscr{R}}\binom{S_{m,j}}{k_j(\ell)}\right) &= p_{m+1}\sum_{i=1}^{\ell}\mathbb{E}\left(\binom{W_{m,i}}{k_i(\ell)}\prod_{j\in\mathscr{R}(i)}\binom{S_{m-1,j}}{k_j(\ell)}\right)+(1-\ell p_{m+1})\mathbb{E}\left(\prod_{j\in\mathscr{R}}\binom{S_{m-1,j}}{k_j(\ell)}\right)\\
&= \frac{\ell}{(m+b+1)}\frac{1}{(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}+[1-\frac{\ell}{m+b+1}]\frac{1}{(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell}}}{(m+b)^{\underline{\ell}}}\\
&= \frac{\ell}{(m+b+1)}\frac{1}{(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}+[\frac{m+b-\ell+1}{m+b+1}]\frac{1}{(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{r}}}{(m+b)^{\underline{\ell}}}\\
&= \frac{1}{(m+b+1)}\frac{1}{(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}[\ell+m-k(\ell)-\ell+1]\\
&= \frac{1}{(1+b)_{k(\ell)}}\frac{(m+1-k(\ell))^{\underline{\ell}}}{(m+b+1)^{\underline{\ell}}}.
\end{aligned}$$

Similarly,

$$\mathbb{E}\left(\binom{W_{m+1,i}}{k_i(\ell)}\prod_{j\in\mathscr{R}(i)}\binom{S_{m,j}}{k_j(\ell)}\right) = p_{m+1}\mathbb{E}\left(\binom{W_{m,i}+1}{k_i(\ell)}\prod_{j\in\mathscr{R}(i)}\binom{S_{m-1,j}}{k_j(\ell)}\right)$$

$$+p_{m+1}\sum_{j\in\mathscr{R}(i)}\mathbb{E}\left(\binom{W_{m,j}}{k_j(\ell)}\prod_{t\in\mathscr{R}(j)}\binom{S_{m-1,t}}{k_t(\ell)}\right)+(1-\ell p_{m+1})\mathbb{E}\left(\prod_{j\in\mathscr{R}}\binom{S_{n-1,j}}{k_j(\ell)}\right)$$

$$= \; p_{m+1}\sum_{j\in\mathscr{R}}\mathbb{E}\left(\binom{W_{m,j}}{k_j(\ell)}\prod_{t\in\mathscr{R}(j)}\binom{S_{m-1,t}}{k_t(\ell)}\right)+p_{m+1}\mathbb{E}\left(\binom{W_{m,i}}{k_i(\ell)-1}\prod_{j\in\mathscr{R}(i)}\binom{S_{m-1,j}}{k_j(\ell)}\right)$$

$$+(1-\ell p_{m+1})\mathbb{E}\left(\prod_{j\in\mathscr{R}}\binom{S_{n-1,j}}{k_j(\ell)}\right)$$

$$= \; \frac{\ell}{(m+b+1)(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}+\frac{1}{(m+b+1)(1+b)_{k(\ell)-1}}\frac{(m-k(\ell)+1)^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}$$

$$+[1-\frac{\ell}{m+b+1}]\frac{1}{(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell}}}{(m+b)^{\underline{\ell}}},$$

where we have made use of Pascal's binomial identity $\binom{d+1}{c}=\binom{d}{c}+\binom{d}{c-1}$ and the induction hypothesis. When $k_i(\ell)=1$ in the first equality above, we proceed differently by making a link with an analogous problem where $\ell$ is replaced by $\ell-1$ (also see **STEP TWO** for more details). Simplifying, we obtain

$$\mathbb{E}\left(\binom{W_{m+1,i}}{k_i(\ell)}\prod_{j\in\mathscr{R}(i)}\binom{S_{m,j}}{k_j(\ell)}\right) =$$

$$\frac{1}{(m+b+1)(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}[\ell+\frac{(b+k(\ell))(m-k(\ell)+1)}{m-k(\ell)-\ell+2}+(m-\ell-k(\ell)+1)]$$

$$=\frac{1}{(m+b+1)(1+b)_{k(\ell)}}\frac{(m-k(\ell))^{\underline{\ell-1}}}{(m+b)^{\underline{\ell-1}}}[\frac{(m-k(\ell)+1)(m+b-\ell+2)}{m-k(\ell)-\ell+2}]$$

$$=\frac{1}{(1+b)_{k(\ell)}}[\frac{(m+1-k(\ell))(m-k(\ell))^{\underline{\ell-1}}}{m-k(\ell)-\ell+2}][\frac{m+b-\ell+2}{(m+1+b)(m+b)^{\underline{\ell-1}}}]$$

$$=\frac{1}{(1+b)_{k(\ell}}\frac{(m+1-k(\ell))^{\underline{\ell-1}}}{(m+1+b)^{\underline{\ell-1}}},$$

which completes the proof for $k_1(\ell),\ldots,k_\ell(\ell)>0$, $\ell\geqslant 1$.

**STEP TWO.** Suppose that

$$k_j(r)>0 \quad \text{for all} \quad j\in\mathscr{C},\mathscr{C}\neq\varnothing,$$
$$k_j(r)=0 \quad \text{for all} \quad j\in\mathscr{R}\setminus\mathscr{C},\mathscr{C}\neq\mathscr{R}.$$

We wish to evaluate

$$\mathbb{E}\left(\prod_{j\in\mathscr{C}}\binom{S_{m,j}}{k_j(r)}\right) \quad \text{and} \quad \mathbb{E}\left(\binom{W_{m+1,i}}{k_i(\ell)}\prod_{j\in\mathscr{C}\setminus\{i\}}\binom{S_{m,j}}{k_j(r)}\right), \tag{3.4}$$

for all $i \in \mathscr{C}$. By a symmetry argument, we can assume without loss of generality that

$$k_1(r), \ldots, k_{r_+}(r) > 0.$$

Finally, linking the evaluation of (3.4) with **STEP ONE.** with $r$ replaced by $r_+$ leads to the desired result and completes the proof. □

## 3.2 Distribution of $\underline{S}$ and mixture representations

As a consequence of Theorem 3.1, we obtain the following for the asymptotic case when $n \to \infty$.

**Theorem 3.2.** *Under assumption (1.4),*

**(a)** *The mixed binomial moments of $S_1, S_2, \ldots, S_r$ are given by*

$$\mathbb{E}\left(\prod_{j=1}^{r}\binom{S_j}{k_j}\right) = \frac{1}{(1+b)_k}, \tag{3.5}$$

*with $k = \sum_{i=1}^{r} k_i$.*

**(b)** *The probability mass function of $\underline{S} = (S_1, \ldots, S_r)$ is given by*

$$\mathbb{P}\left(S_1 = s_1, \ldots, S_r = s_r\right) = \frac{1}{(1+b)_t}\,{}_1F_1(t+r; t+b+1; -1), \tag{3.6}$$

*with $t = \sum_{i=1}^{r} s_i$ and ${}_1F_1$ the confluent hypergeometric function given by ${}_1F_1(\gamma_1;\gamma_2;z) = \sum_{j=0}^{\infty}\frac{(\gamma_1)_j}{(\gamma_2)_j}\frac{z^j}{j!}$.*

**(c)** *For $m \in \mathbb{N}$, the distribution of $\underline{S} \mid \sum_{i=1}^{r} S_i = t$ is uniform on the simplex $M(t) = \{v = (v_1, v_2, \ldots, v_r) : v_i \geqslant 0, \sum_{i=1}^{r} v_i = t\}$.*

**Proof.** Part **(a)** follows directly from Theorem 3.1 and Lebesgue's monotone convergence theorem by taking $n \to \infty$. Part **(c)** follows from part **(b)** since $\mathbb{P}\left(S_1 = s_1, \ldots, S_r = s_r\right)$

is a function of $s_1 + \cdots + s_r$. Finally, for part **(b)**, use (2.4) to obtain

$$\begin{aligned}
\mathbb{P}(S_1 = s_1, \ldots, S_r = s_r) &= \sum_{\{(k_1,\ldots,k_r):\, k_i \geqslant s_i, i=1,\ldots,r\}} \frac{(-1)^{\sum_{i=1}^r (k_i - s_i)}}{(1+b)_k} \prod_{i=1}^r \binom{k_i}{s_i} \\
&= \sum_{m=t}^{\infty} \frac{(-1)^{m-t}}{(1+b)_m} \sum_{\{(k_1,\ldots,k_r):\, k_i \geqslant s_i, i=1,\ldots,r, \sum_{i=1}^r k_i = m\}} \prod_{i=1}^r \binom{k_i}{s_i} \\
&= \sum_{m=t}^{\infty} \frac{(-1)^{m-t}}{(1+b)_m} \binom{m+r-1}{t+r-1} \\
&= \sum_{m=0}^{\infty} \frac{(-1)^m}{(1+b)_t\,(1+b+t)_m} \frac{(t+r)_m}{m!},
\end{aligned}$$

which is (3.6) and completes the proof except the justification of the next to last equality. This equality can be justified by the following combinatorial argument for re-expressing the number of ways of choosing $t+r-1$ objects among $m+r_1-1$. Number the objects 1 through $m+r-1$ and fix positive integers $x_1, \ldots, x_r$ such that $x_1 + \cdots + x_r = t$. Condition of the selection of the $x_1$ th largest, $(x_1 + x_2)$ th largest, ..., $x_1 + \cdots + x_{r-1}$ th largest numbered objects. The identity is established as the number of such possible choices is equal to $\prod_{i=1}^r \binom{k_i}{x_i}$ with $k_i \geqslant x_i$ for $i = 1, \ldots, r$ and $k_1 + \cdots + k_r = m$. □

Our next result establishes two elegant mixture representations.

**Theorem 3.3.** *Under assumption (1.4),*

**(a)** *The distribution of $\underline{S}$ is a multivariate Poisson mixture with $S_i|V \sim^{indep.}$ Poisson$(V_i)$ for $i = 1, \ldots, r$, and $V = (V_1, \ldots, V_r) \sim$ Dirichlet$(a_1 = 1, \ldots, a_r = 1, a_{r+1} = b - r + 1)$;*

**(b)** *The distribution of $\underline{S}$ admits the representation*

$$\underline{S}|\alpha \sim p_\alpha\,, \;\; \alpha \sim \text{Beta}(1, b)\,, \tag{3.7}$$

*with $p_\alpha$ the multivariate probability mass function on $\mathbb{N}^r$ given by*

$$p_\alpha(s_1, \ldots, s_r) = \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \frac{e^{-\alpha} \alpha^{s_1 + \ldots + s_r + j}}{(s_1 + \ldots + s_r + j)!}\,,\; \alpha \in (0, 1]\,. \tag{3.8}$$

**Proof.**

**(a)** Part **(a)** is a direct consequence of Theorem 3.2 and Lemma 2.1.

**(b)** Set $t = \sum_i s_i$. In view of expression (3.6), it suffices to show that

$$(1+b)_t \int_0^1 b(1-\alpha)^{b-1}\, g_{t,r}(\alpha)\, d\alpha \;=\; {}_1F_1(t+r; t+b+1; -1)\,, \tag{3.9}$$

for all $r = 1, 2, \ldots$, $b \geqslant r-1$, $t = 0, 1, \ldots$, and with $g_{t,r}(\alpha) = p_\alpha(s_1, \ldots, s_r)$. We prove by induction on $r$ and we make use of the integral representation

$$\int_0^1 \alpha^{\gamma_1 - 1}(1-\alpha)^{\gamma_2-\gamma_1-1}\, e^{-\alpha}\, d\alpha \;=\; \frac{\Gamma(\gamma_1)\,\Gamma(\gamma_2-\gamma_1)}{\Gamma(\gamma_2)}\; {}_1F_1(\gamma_1;\gamma_2;-1)\;, \tag{3.10}$$

as well as the recurrence relation

$$\gamma_2\; {}_1F_1(\gamma_1+1;\gamma_2;-1) \;=\; \gamma_2\; {}_1F_1(\gamma_1;\gamma_2;-1) \;-\; {}_1F_1(\gamma_1+1;\gamma_2+1;-1)\,, \tag{3.11}$$

both valid for $\gamma_2 > \gamma_1 > 0$ (e.g., NIST Digital Library of Mathematical Functions). For $r = 1$, we have $g_{t,1}(\alpha) = \frac{e^{-\alpha}\,\alpha^t}{t!}$ and (3.9) is readily verified by using (3.10). Now, assume (3.9) holds for a given $r_0$ and observe that $g_{t,r_0+1}(\alpha) = g_{t,r_0}(\alpha) - g_{t+1,r_0}(\alpha)$ by making use of Pascal's identity and by re-arranging terms. We now obtain by exploiting this along with (3.11)

$$\begin{aligned} (1+b)_t \int_0^1 b(1-\alpha)^{b-1}\, g_{t,r_0+1}(\alpha)\, d\alpha \;&=\; {}_1F_1(t+r_0; t+b+1; -1) \\ &-\; \frac{1}{t+b+1}\, {}_1F_1(t+r_0+1; t+b+2; -1) \\ &=\; {}_1F_1(t+r_0+1; t+b+1; -1)\,, \end{aligned}$$

which is indeed (3.9). □

As a consequence of the above, we recover a Poisson mixture representation for the total number of runs of length 2, and as expressed in (1.3).

**Corollary 3.4.** *Under assumption (1.4), the total number of runs $T = \sum_{i=1}^r S_i$ of length 2 is Poisson$(1)$ distributed when $b = r-1$. Otherwise for $b > r-1$, $T$ admits the mixture representation*

$$T|L \sim \mathit{Poisson}(L)\,,\;\; L \sim \mathit{Beta}(r, b-r+1)\,.$$

**Proof.** The result is a direct application of Theorem 3.3 and Lemma 2.1. □

**Remark 3.5.** *The above results can also be obtained using directly Theorem 3.2. For instance with $b = r - 1$, we have using (2.3)*

$$\begin{aligned}
\mathbb{E}\left[t^{\sum_{i=1}^r S_i}\right] = \mathbb{E}\left[\prod_{i=1}^r t^{S_i}\right] &= \sum_{\{(k_1,\ldots,k_r):\, k_i \geqslant 0, i=1,\ldots,r\}} \frac{(t-1)^{\sum_i k_i}}{(r)_{\sum_i k_i}} \\
&= \sum_{k=0}^{\infty} \sum_{\{(k_1,\ldots,k_r):\, k_i \geqslant 0, i=1,\ldots r, \sum_{i=1}^r k_i = k\}} \frac{(t-1)^k}{(r)_k} \\
&= \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} \frac{(t-1)^k}{(r)_k} \\
&= e^{t-1},
\end{aligned}$$

*which establishes the Poisson$(1)$ distributional result for $T$ when $b = r - 1$.*
*The above corollary is already known and due to Ait Aoudia et al. (2014), but it is presented here for sake of completeness and since our derivation differs in how the joint distribution of $S_1, S_2, \ldots, S_r$ is used. The other findings above, which we now further discuss, generalize Ait Aoudia's bivariate case ($r = 2$) results.*

Theorem 3.2's multivariate Poisson mixture representation of $\underline{S}$ extends in a most interesting way the known marginal distribution representation for $S_i$, $i = 1, \ldots, r$, which is a Beta mixture of Poisson distributions. Here the components $S_i$ are clearly dependent, but the representation tells us that they are conditionally independent and the dependence is reflected through the dependence of the mixing components of the Dirichlet. In contrast to the Dirichlet mixture, the dependence in representation **(b)** of Theorem 3.2 is reflected through the conditional distributions of $\underline{S}$, and the mixing variable $\alpha$ is univariate. Furthermore, it is readily verified that the conditional marginal distributions of $S_i|\alpha$ are Poisson$(\alpha)$, which is consistent with the univariate Beta mixture result.

**Remark 3.6.** *The joint probability mass functions (pmf) $p_\alpha$ in (3.8), which has Poisson$(\alpha)$ marginals, are examples of the larger class of pmf's of the form*

$$\gamma(x_1, \ldots, x_r) = \sum_{j=0}^{r-1} \binom{r-1}{j} (-1)^j \, p(\sum_i x_i + j) \,, \quad x_1, \ldots, x_r \in \mathbb{N}, \tag{3.12}$$

*which were analyzed in the bivariate case ($r = 2$) and otherwise referred to by Ait Aoudia and Marchand (2014). In the above form, $p$ is a univariate pmf, $\gamma$ has marginals given by $p$, and $\gamma$ is a valid pmf as long as $p$ is $(r-1)$-monotone, i.e., $\sum_{j=0}^{r-1} \binom{r-1}{j} (-1)^j \, p(m+j) \geqslant 0$ for all $m$. These conditions indeed apply for $p \sim$ Poisson$(\alpha)$ with $\alpha < 1$ leading to $\gamma$ as in (3.8). For this and further properties of such distributions, we refer to Castañer et al. (2015).*

## Acknowledgments

Éric Marchand's and François Perron's gratefully acknowledge the support for research provided by NSERC of Canada. We are grateful to Chris Jones for useful discussions and insight on the distributions represented by (3.12).

# References

Ait Aoudia, Djilali, Marchand, É., Perron, F. & Ben Hadj Slimene, L. (2014). On runs, bivariate Poisson mixtures and distributions that arise in Bernoulli arrays. *Electronic Communications in Probability*, **19**, 1-12.

Ait Aoudia, D. and Marchand É. (2014). On a simple construction of bivariate probability functions with fixed marginals. *The American Statistician*, **68**, 170-173.

Ait Aoudia, D. and Marchand É. (2010). On the number of runs for Bernoulli arrays. *Journal of Applied Probability*, **47**, 367-377.

Arratia, R., Barbour, A.D., and Tavaré, S. (1992). Poisson process approximations for the Ewens sampling formula. *Annals of Applied Probability*, **2**, 519-535.

Castañer, A., Claramunt, M.M., Lefèvre, C. & Loisel, S. (2015). Discrete Schur-constant models. *Journal of Multivariate Analysis*, **140**, 343-362.

Chern, H.-H., Hwang, H.-K., and Yeh, Y.-N. (2000). Distribution of the number of consecutive records. *Random Structures Algorithms*, **17**, 169-196.

Csörgó, M. and Wu, B.W. (2000). On sums of overlapping products of independent Bernoulli random variables. *Ukrainian Mathematical Journal*, **52**, 1304-1309.

Goncharov, V. (1944). On the field of combinatory analysis. *Soviet Math. Izv., Ser. Math.*, **8**, 3-48. In Russian.

Hahlin, L.O. (1995). Double Records. *Research Report # 12*, Department of Mathematics, Uppsala University.

Holst, L. (2008). The number of two-consecutive successes in a Hoppe-Polyá urn. *Journal of Applied Probability*, **45**, 901-906.

Holst, L. (2007). Counts of failure strings in certain Bernoulli sequences. *Journal of Applied Probability*, **44**, 824-830.

Huffer, W. F., Sethuraman, J. and Sethuraman, S. (2009). A study of counts of Bernoulli strings via conditional Poisson processes. *Proceedings of the American Mathematical Society*, **137**, 2125-2134.

Joffe, A., Marchand É., Perron, F., and Popadiuk, P. (2004). On sums of products of Bernoulli variables and random permutations. *Journal of Theoretical Probability*, **17**, 285-292.

Joffe, A., Marchand É., Perron, F., and Popadiuk, P. (2000). On sums of products of Bernoulli variables and random permutations. *Research Report # 2686*, Centre de Recherches Mathématiques, Montréal, Canada.

Kolchin, V.F. (1971). A problem of the allocation of particles in cells and cycles of random permutations. *Theory of Probability and its Applications*, **16**, 74-90.

Mori, T.F. (2001). On the distribution of sums of overlapping products. *Acta Scientiarum Mathematica (Szeged)*, **67**, 833-841.

NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.9 of 2014-08-29. Online companion to [OLBC10].

Sethuraman, J. and Sethuraman, S. (2004). On counts of Bernoulli strings and connections to rank orders and random permutations. *A Festschrift for Herman Rubin. IMS Lecture Note Monograph Series*, **45**, Institute of Mathematical Statistics, pp. 140-152.