J. Appl. Prob. **45,** 901–906 (2008)
Printed in England

THE NUMBER OF TWO CONSECUTIVE SUCCESSES IN A HOPPE–PÓLYA URN

LARS HOLST,* *Royal Institute of Technology*

Abstract

In a sequence of independent Bernoulli trials the probability of success in the kth trial is $p_k = a/(a + b + k - 1)$. An explicit formula for the binomial moments of the number of two consecutive successes in the first n trials is obtained and some consequences of it are derived.

Keywords: Bernoulli trial; binomial moment; Hoppe urn; overlapping indicators; Pólya urn; records; random permutation

2000 Mathematics Subject Classification: Primary 60C05
Secondary 60K99

1. Introduction

An urn initially contains one white ball and one black ball of weight $a > 0$ and $b \geq 0$, respectively. Balls are randomly drawn from the urn with probabilities proportional to weights. Every time a white ball or a black ball is drawn from the urn, it is replaced with a ball of weight 1 of a colour not already in the urn, otherwise a ball is replaced together with a copy of it. We call this drawing scheme a *Hoppe–Pólya urn*. If $b = 0$, there is no black ball, the so-called *Hoppe urn*. If all balls emanating from a draw of the white or black are coloured white or, respectively, black, we obtain the well-known *Pólya urn*.

Let the sequence of independent Bernoulli random variables $I_1, I_2, I_3, \dots$ indicate the drawings of the white ball, the 'successes' or 'records' in the Hoppe–Pólya urn. Obviously,

$$p_k = \mathrm{P}(I_k = 1) = 1 - \mathrm{P}(I_k = 0) = \frac{a}{a + b + k - 1}, \qquad k = 1, 2, \dots.$$

The number of successes in the first n trials can be written as

$$K_n = I_1 + I_2 + \cdots + I_n,$$

and the number of two consecutive successes can be written as

$$M_n = I_1 I_2 + I_2 I_3 + \cdots + I_{n-1} I_n.$$

An explicit formula for the binomial moments of M_n is the main result of this paper. Note that $0 \leq M_n \leq n - 1$.

For $p_k = a/(a + b + k - 1)$, the Borel–Cantelli lemma implies that

$$M_\infty = \sum_{k=1}^{\infty} I_k I_{k+1} < +\infty$$

Received 21 December 2007; revision received 8 May 2008.
* Postal address: Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden.
Email address: lholst@math.kth.se

with probability 1. For the case in which $a = 1$ and $b = 0$, i.e. $p_k = 1/k$, connected with record values and random permutations, Hahlin (1995) proved that M_∞ is Poisson distributed with mean 1. After that, an unpublished proof of the same result by Diaconis inspired a number of studies on the distribution of M_∞; see Chern *et al.* (2000), Mori (2001), Joffe *et al.* (2004), Sethuraman and Sethuraman (2004), Holst (2007), and the references therein. To the author's knowledge, the result in this paper on the distribution of M_n for finite n has not been obtained previously.

2. Notation and facts

Following Knuth (1992), we denote falling and rising factorials by

$$x^{\underline{n}} = x(x-1)\cdots(x-n+1),$$
$$x^{\overline{n}} = x(x+1)\cdots(x+n-1) = \sum_{j=1}^{n} \begin{bmatrix} n \\ j \end{bmatrix} x^j,$$

where $\begin{bmatrix} n \\ j \end{bmatrix}$ is a cycle number or signless Stirling number of the first kind. Recall the combinatorial interpretation: $\begin{bmatrix} n \\ j \end{bmatrix}$ is the number of permutations of $1, 2, \ldots, n$ with j cycles.

For K_n equals the number of successes in the first n trials, we have

$$\begin{aligned} \mathrm{E}(x^{K_n}) &= \prod_{k=1}^{n} \left(\frac{a}{a+b+k-1} x + 1 - \frac{a}{a+b+k-1} \right) \\ &= \frac{(ax+b)^{\overline{n}}}{(a+b)^{\overline{n}}} \\ &= \sum_{j=1}^{n} \begin{bmatrix} n \\ j \end{bmatrix} \frac{(ax+b)^j}{(a+b)^{\overline{n}}} \\ &= \sum_{i=0}^{n} x^i \sum_{j=i}^{n} \begin{bmatrix} n \\ j \end{bmatrix} \binom{j}{i} \frac{a^i\, b^{j-i}}{(a+b)^{\overline{n}}}. \end{aligned}$$

Hence, for $i = 0, 1, 2, \ldots n$,

$$\mathrm{P}(K_n = i) = \sum_{j=i}^{n} \begin{bmatrix} n \\ j \end{bmatrix} \frac{(a+b)^j}{(a+b)^{\overline{n}}} \binom{j}{i} \left(\frac{a}{a+b} \right)^i \left(\frac{b}{a+b} \right)^{j-i}.$$

In particular, for $b = 0$, i.e. Hoppe's urn, we obtain the *cycle distribution*,

$$\mathrm{P}(K_n = i) = \begin{bmatrix} n \\ i \end{bmatrix} \frac{a^i}{a^{\overline{n}}}, \qquad i = 1, 2, \ldots, n,$$

for an a-biased random permutation; see Arratia *et al.* (2003, p. 100).

The number of times the white ball, or balls emanating from it, is drawn in the first n trials, X_n, has the following *Pólya–Eggenberger distribution*:

$$\mathrm{P}(X_n = i) = \binom{n}{i} \frac{a^{\overline{i}}\, b^{\overline{n-i}}}{(a+b)^{\overline{n}}} = \mathrm{E}\left(\binom{n}{i} U^i (1-U)^{n-i} \right), \qquad i = 0, 1, 2, \ldots, n,$$

where U is a Beta(a, b) random variable with density

$$f_U(u) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u)^{b-1}, \qquad 0 < u < 1.$$

Using the binomial distribution, we obtain, for $r = 1, 2, \dots, n$,

$$\mathrm{E}\binom{X_n}{r} = \mathrm{E}\bigg(\sum_{i=r}^{n} \binom{i}{r}\binom{n}{i} U^i (1-U)^{n-i}\bigg) = \mathrm{E}\bigg(\binom{n}{r} U^r\bigg) = \binom{n}{r} \frac{a^{\overline{r}}}{(a+b)^{\overline{r}}}.$$

Recall that a random variable S with the hypergeometric distribution

$$\mathrm{P}(S = i) = \binom{c}{i}\binom{d}{n-i} \bigg/ \binom{c+d}{n}$$

has the binomial moment

$$\mathrm{E}\binom{S}{r} = \binom{n}{r} \frac{c^{\underline{r}}}{(c+d)^{\underline{r}}}.$$

For an integer-valued random variable $Z \geq 0$ having a probability generating function with a radius of convergence larger than 1, we have

$$\mathrm{E}(x^Z) = \mathrm{E}((1 + (x-1))^Z) = \sum_{r=0}^{\infty} \mathrm{E}\binom{Z}{r}(x-1)^r = \sum_{i=0}^{\infty} x^i \sum_{r=i}^{\infty} (-1)^{r-i} \binom{r}{i} \mathrm{E}\binom{Z}{r},$$

which gives the following probability function of Z expressed in binomial moments:

$$\mathrm{P}(Z = i) = \sum_{r=i}^{\infty} (-1)^{r-i} \binom{r}{i} \mathrm{E}\binom{Z}{r}, \qquad i = 0, 1, 2, \dots.$$

Note that if $0 \leq Z < n$ then $\mathrm{E}\binom{Z}{r} = 0$ for $r \geq n$.

3. The number of two consecutive successes

The following result implicitly gives the distribution of M_n.

Theorem 3.1. *For $p_k = a/(a+b+k-1)$ and $r = 1, 2, \dots, n-1$,*

$$\mathrm{E}\binom{M_n}{r} = \frac{a^r}{(a+b+n-1)^{\underline{r}}} \sum_{k=1}^{r} \binom{r-1}{r-k}\binom{n-r}{k} \frac{a^{\overline{k}}}{(a+b)^{\overline{k}}}.$$

Before proving the theorem we consider the special case $b = 0$, i.e. Hoppe's urn. A more general result is Proposition 3 of Holst (2007).

Lemma 3.1. *For $p_k = a/(a+k-1)$ and $r = 1, 2, \dots, n-1$,*

$$\mathrm{E}\binom{M_n}{r} = \binom{n-1}{r} \frac{a^r}{(a+n-1)^{\underline{r}}}.$$

https://doi.org/10.1239/jap/1222441836 Published online by Cambridge University Press

Proof. For $N_n = M_n + I_n$, we have

$$\mathrm{E}(t^{N_{n+1}}) = p_{n+1}\,\mathrm{E}(t^{N_n})t + (1 - p_{n+1})\,\mathrm{E}(t^{M_n}),$$

which implies that

$$\mathrm{E}\binom{N_{n+1}}{r} = p_{n+1}\Big(\mathrm{E}\binom{N_n}{r} + \mathrm{E}\binom{N_n}{r-1}\Big) + (1 - p_{n+1})\,\mathrm{E}\binom{M_n}{r}.$$

For $p_k = a/(a + k - 1)$, the random variable N_n has the same distribution as the number of fix-points in an a-biased random permutation of $1, 2, \dots, n$, and

$$\mathrm{E}\binom{N_n}{r} = \binom{n}{r}\frac{a^r}{(a+n-1)^{\underline{r}}};$$

see Arratia *et al.* (2003 pp. 95–96). Using this and the relation above, proves the assertion.

Proof of Theorem 3.1. Consider the Hoppe–Pólya urn and the random variable X_n introduced in Section 2. In the X_n 'white' drawings, the probability of obtaining the white ball in the jth trial is

$$p_j^* = \frac{a}{a + j - 1}.$$

Given $X_n = x$, the number of times the white ball is consecutively drawn in these 'white' drawings, M_x^*, is distributed as in Lemma 3.1.

Conditional on $X_n = x$, we can argue as follows. Among the x 'white' draws let W_1 denote a drawing that gives a white ball and let W_0 denote a drawing that gives a ball which emanates from a white ball. Let B denote a 'black' drawing. The result of the 'white' draws can be written as $W_1 W_{i_2} W_{i_3} \cdots W_{i_x}$, where $i_2, \dots, i_x$ are 0 or 1. For $M_x^* = y$, y of the pairs $W_1 W_{i_2}, \dots, W_{i_x - 1} W_{i_x}$ are of type $W_1 W_1$. For $M_n = z$ consecutive draws $W_1 W_1$ among the original n draws (with x Ws and $n - x$ Bs), there are z pairs of the y $W_1 W_1$-pairs among the 'white' draws which are intact and $y - z$ which are split by at least one B between $W_1 W_1$. The number of ways to choose the pairs to be intact is $\binom{y}{z}$. After such a splitting, there are $x - z$ 'free' Ws to combine with the $n - x - (y - z)$ 'free' Bs, and there are $\binom{n-y}{x-z}$ such combinations. As each combination of x Ws and $n - x$ Bs has the same probability, $1/\binom{n}{x}$, we obtain

$$\mathrm{P}(M_n = z \mid X_n = x) = \sum_y \mathrm{P}(M_x^* = y)\binom{y}{z}\binom{n-y}{x-z}\Big/\binom{n}{x}.$$

Thus, M_n's probability function can be written as

$$\mathrm{P}(M_n = z) = \sum_{x,y} \mathrm{P}(X_n = x)\,\mathrm{P}(M_x^* = y)\binom{y}{z}\binom{n-y}{x-z}\Big/\binom{n}{x}$$

with the binomial moment

$$\mathrm{E}\binom{M_n}{r} = \sum_{x,y} \mathrm{P}(X_n = x)\,\mathrm{P}(M_x^* = y)\sum_z \binom{z}{r}\binom{y}{z}\binom{n-y}{x-z}\Big/\binom{n}{x}.$$

Using the formula for the binomial moment of the hypergeometric distribution and Lemma 3.1, we obtain

$$\begin{aligned}
\mathrm{E}\binom{M_n}{r} &= \sum_{x,y} \mathrm{P}(X_n = x)\,\mathrm{P}(M_x^* = y)\binom{x}{r}\frac{y^{\underline{r}}}{n^{\underline{r}}} \\
&= \sum_x \binom{x}{r}\binom{n}{r}^{-1} \mathrm{P}(X_n = x) \sum_y \binom{y}{r} \mathrm{P}(M_x^* = y) \\
&= \sum_x \binom{x}{r}\binom{n}{r}^{-1} \binom{n}{x}\frac{a^{\overline{x}} b^{\overline{n-x}}}{(a+b)^{\overline{n}}}\binom{x-1}{r}\frac{a^r}{(a+x-1)^{\underline{r}}}.
\end{aligned}$$

Hence, the binomial moment of the Pólya–Eggenberger distribution gives

$$\begin{aligned}
\mathrm{E}\binom{M_n}{r} &= \frac{a^r}{(a+b+n-1)^{\underline{r}}} \sum_x \binom{n-r}{x-r}\frac{a^{\overline{x-r}} b^{\overline{n-r-(x-r)}}}{(a+b)^{\overline{n-r}}} \sum_{k=1}^{r}\binom{r-1}{r-k}\binom{x-r}{k} \\
&= \frac{a^r}{(a+b+n-1)^{\underline{r}}} \sum_{k=1}^{r}\binom{r-1}{r-k} \sum_t \binom{t}{k}\binom{n-r}{t}\frac{a^{\overline{t}}\, b^{\overline{n-r-t}}}{(a+b)^{\overline{n-r}}} \\
&= \frac{a^r}{(a+b+n-1)^{\underline{r}}} \sum_{k=1}^{r}\binom{r-1}{r-k} \mathrm{E}\binom{X_{n-r}}{k} \\
&= \frac{a^r}{(a+b+n-1)^{\underline{r}}} \sum_{k=1}^{r}\binom{r-1}{r-k}\binom{n-r}{k}\frac{a^{\overline{k}}}{(a+b)^{\overline{k}}},
\end{aligned}$$

which proves the assertion.

The distribution of M_∞ was obtained by Mori (2001). It is a special case of the distribution given in Theorem 1 of Holst (2007).

Corollary 3.1. *Conditional on a Beta(a,b) random variable U, M_∞ is Poisson distributed with mean aU.*

Proof. From Theorem 3.1, it follows that

$$\mathrm{E}\binom{M_n}{r} \to \frac{a^r}{r!}\frac{a^{\overline{r}}}{(a+b)^{\overline{r}}}, \qquad n \to \infty.$$

As $\mathrm{E}(U^r) = a^{\overline{r}}/(a+b)^{\overline{r}}$, we obtain, using the Poisson distribution,

$$\mathrm{E}\binom{M_\infty}{r} = \mathrm{E}\bigg(\mathrm{E}\bigg(\binom{M_\infty}{r} \,\bigg|\, U\bigg)\bigg) = \mathrm{E}\bigg(\frac{(aU)^r}{r!}\bigg) = \frac{a^r\,\mathrm{E}(U^r)}{r!} = \frac{a^r}{r!}\frac{a^{\overline{r}}}{(a+b)^{\overline{r}}}.$$

The assertion follows from the moment convergence.

The distribution of M_n for $p_k = p$ was studied by Hirano *et al.* (1991); see also the references therein. Letting $a, b \to \infty$ such that $a/(a+b) \to p$, we obtain their result.

Corollary 3.2. *For $p_k = p$ and $r = 1, 2, \ldots, n-1$,*

$$\mathrm{E}\binom{M_n}{r} = p^r \sum_{k=1}^{r}\binom{r-1}{r-k}\binom{n-r}{k}p^k.$$

https://doi.org/10.1239/jap/1222441836 Published online by Cambridge University Press

Finally, consider the Pólya urn starting with one white ball of weight a and one black ball of weight b. Every drawn ball is replaced together with one ball of the same colour and of weight 1. In n drawings, the number of times a white ball is drawn, X_n, has the Pólya–Eggenberger distribution. Let Y_n denote the number of times a white ball is consecutively drawn.

Corollary 3.3. *For the Pólya urn and* $r = 1, 2, \ldots, n-1$,

$$\mathrm{E}\binom{Y_n}{r} = \sum_{k=1}^{r} \binom{r-1}{r-k}\binom{n-r}{k} \frac{a^{\overline{r+k}}}{(a+b)^{\overline{r+k}}}.$$

Proof. Set $J_k = 1$ if the kth drawn ball is white, otherwise set $J_k = 0$. It is a well known, easily proved fact that, conditional on a Beta(a, b) random variable U, the random variables $J_1, J_2, \ldots$ are independent and Bernoulli distributed with success probability U. Thus, it follows from Corollary 3.2 that

$$\mathrm{E}\binom{Y_n}{r} = \mathrm{E}\biggl(U^r \sum_{k=1}^{r} \binom{r-1}{r-k}\binom{n-r}{k} U^k\biggr) = \sum_{k=1}^{r} \binom{r-1}{r-k}\binom{n-r}{k} \mathrm{E}(U^{r+k}),$$

which proves the assertion.

References

Arratia, R., Barbour, A. D. and Tavaré, S. (2003). *Logarithmic Combinatorial Structures: A Probabilistic Approach.* European Mathematical Society, Zürich.

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

Hahlin, L. O. (1995). Double records. Res. Rep. 1995:12, Department of Mathematics, Uppsala University.

Hirano, K., Aki, S., Kashiwagi, N. and Kuboki, H. (1991). On Ling's binomial and negative binomial distributions of order k. *Statist. Prob. Lett.* **11,** 503–509.

Holst, L. (2007). Counts of failure strings in certain Bernoulli sequences. *J. Appl. Prob.* **44,** 824–830.

Joffe, A., Marchand, E., Perron, F. and Popadiuk, P. (2004). On sums of products of Bernoulli variables and random permutations. *J. Theoret. Prob.* **17,** 285–292.

Knuth, D. (1992). Two notes on notation. *Amer. Math. Monthly* **99,** 403–422.

Mori, T. F. (2001). On the distribution of sums of overlapping products. *Acta Sci. Math.* **67,** 833–841.

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 of Mathematical Statistics, Beachwood, OH, pp. 140–152.