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

LARS HOLST

Department of Mathematics, Royal Institute of Technology
SE–100 44 Stockholm, Sweden

E-mail: lholst@math.kth.se

November 27, 2007

Abstract

In a sequence of independent Bernoulli trials the probability of success in the k:th 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 trials; binomial moments; Hoppe's urn; Overlapping indicators; Pólya's urn; Records; Random permutations

2000 Mathematics Subject Classification: Primary 60C05
Secondary 60K99

1 Introduction

An urn contains initially one white 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 the white or the black ball is drawn it is replaced into the urn together with a ball of weight one and with a colour not already in the urn, else 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's urn.* If all balls emanating from a draw of the white (black) are coloured white (black) we get the well known *Pólya's 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,

$$p_k = P(I_k = 1) = 1 - P(I_k = 0) = a/(a + b + k - 1), \quad k = 1, 2, \dots .$$

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

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

and the number of two consecutive successes is

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

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

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

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

with probability one. For the case $a = 1$ and $b = 0$, that is $p_k = 1/k$, 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 After that an unpublished proof of the same result by Diaconis inspired a number of studies on the distribution of M_∞, see Chern *et al.* (2000), Mori (2001), Joffe *et al.* (2004),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 our knowledge the result in this paper on the distribution of M_n for finite n has not been obtained previously.

2 Notations and facts

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

$$x^{\underline{n}} = x(x-1)\cdots(x-n+1),\ x^{\overline{n}} = x(x+1)\cdots(x+n-1) = \sum_{j=1}^{n} \left[{n \atop j}\right] x^j,$$

where $\left[{n \atop j}\right]$ is a cycle number or signless Stirling number of the first kind. Recall the combinatorial interpretation: $\left[{n \atop j}\right]$ is the number of permutations of $1, 2, \dots, n$ with j cycles.

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

$$E(x^{K_n}) = \prod_{k=1}^{n} \Big(\frac{a}{a+b+k-1} x + 1 - \frac{a}{a+b+k-1}\Big) = \frac{(ax+b)^{\overline{n}}}{(a+b)^{\overline{n}}}$$

$$= \sum_{j=1}^{n} \begin{bmatrix} n \\ j \end{bmatrix} \frac{(ax+b)^j}{(a+b)^{\overline{n}}} = \sum_{i=0}^{n} x^i \sum_{j=i}^{n} \begin{bmatrix} n \\ j \end{bmatrix} \binom{j}{i} \frac{a^i\, b^{j-i}}{(a+b)^{\overline{n}}}.$$

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

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

In particular for $b = 0$, that is Hoppe's urn, we get the *cycle distribution*

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

for an a-biased random permutations; see Arratia *et al.* (2003) page 100.

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

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

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

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

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

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

Recall that a random variable S with the hypergeometric distribution

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

has the binomial moment

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

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

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

which gives the probability function of Z expressed in binomial moments

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

Note that if $0\le Z<n$ then $E\binom{Z}{r}=0$ for $r\ge n$.

3 The number of two consecutive successes

The following result gives implicitely the distribution of M_n.

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

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

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

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

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

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

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

which implies

$$E\binom{N_{n+1}}{r}=p_{n+1}\left(E\binom{N_n}{r}+E\binom{N_n}{r-1}\right)+(1-p_{n+1})E\binom{M_n}{r}.$$

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

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

see Arratia *et al.* (2003) pages 95 and 96. Using this and the relation above proves the assertion. □

Proof of the theorem. Consider the Hoppe-Pólya urn and the random variable X_n introduced in Section 2. In the X_n 'white' drawings the probability of getting the white ball in the j:th trial is $p_j^* = a/(a+j-1)$. Given $X_n = x$, the number of times the white ball was consecutively drawn in these 'white' drawings, M_x^*, is distributed as in the lemma.

Conditional on $X_n = x$ we can argue as follows. Let among the x 'white' draws W_1 denote a drawing giving the white ball and W_0 giving a ball emanating from the white. B denotes a 'black' drawing. The result of the 'white' draws can be written $W_1 W_{i_2} W_{i_3} \dots W_{i_x}$ where $i_2, \dots, i_x$ are 0 or 1. For $M_x^* = y$ we have that 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 W's and $n-x$ B's), 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' W's to combine with $n-x-(y-z)$ 'free' B's and there are $\binom{n-y}{x-z}$ such combinations. As each combination of x W's and $n-x$ B's has the same probability $1/\binom{n}{x}$ we get

$$P(M_n = z | X_n = x) = \sum_y P(M_x^* = y) \frac{\binom{y}{z}\binom{n-y}{x-z}}{\binom{n}{x}}.$$

Thus M_n:s probability function can be written

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

with the binomial moment

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

Using the formula for the binomial moment of the hypergeometic distribution and the lemma we get

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

Hence the binomial moment of the Pólya-Eggenberger distribution gives

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

which proves the assertion. □

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

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

Proof. From the theorem it follows that

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

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

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

The assertion follows from the moment convergence. □

The distribution of M_n for $p_k = p$ is studied in Hirano *et al.* (1991) and 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 $r = 1, 2, \ldots, n-1$:*

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

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 one. In n drawings the number of times a white ball is drawn, X_n, has the Pólya-Eggenberger distribution. Let Y_n be the number of times a white ball is consecutively drawn.

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

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

Proof. Set $J_k = 1$, if the k:th drawn ball is white, else $J_k = 0$. 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 the previous corollary that

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

which proves the assertion. □

References

[1] Arratia, R., Barbour, A.D. and Tavaré, S. (2003). *Logarithmic Combinatorial Structures: a Probabilistic Approach.* European Mathematical Society Publishing House, ETH-Zentrum, Zürich.

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

[3] Hahlin, L.O. (1995). Double Records. Res. Rep. 1995:12, Department of Mathematics, Uppsala University.

[4] Hirano, K., Aki, S., Kashiwagi, N. and Kuboki, H. (1991). On Ling's binomial and negative binomial distributions of order k. *Statistics Probability Letters* **11**, 503–509.

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

[6] 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.

[7] Knuth, D. (1992). Two notes on notations. *Amer. Math. Monthly* **99**, 403–422.

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

[9] 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, Ohio, pp. 140–152.