ON THE DISTRIBUTION OF SUMS OF OVERLAPPING PRODUCTS

Tamás F. Móri

Eötvös Loránd University, Budapest

Abstract. We consider a series of overlapping products of the form $X_1X_2+X_2X_3+X_3X_4+\cdots$ where $X_1, X_2, \ldots$ are independent Bernoulli random variables. We compute the exact distribution of every tail section for a particular choice of the X's, thus extending a result of Csörgő and Wu [2]. As a generalization, sums of multiple products are also studied.

1. Introduction

Let $X_1, X_2, \ldots$ be independent Bernoulli random variables with distribution $P(X_n = 1) = p_n = 1 - P(X_n = 0)$. For $i \ge 1$ let

$$N_i = \sum_{r=i}^{\infty} X_r X_{r+1}, \tag{1.1}$$

and let $g_i(s) = E\left(s^{N_i}\right)$ denote the probability generating function of N_i. We are interested in the exact distribution of N_i, particularly in the case where

$$p_n = \frac{\lambda}{\mu + n - 1}, \tag{1.2}$$

with $0 < \lambda \le \mu$. There exist certain results for particular choices of λ and μ. The distribution of N_i together with that of its finite sections was computed in [2] by Csörgő and Wu in the particular case where (1.2) holds with $\lambda = 1$. For $\lambda = \mu$ the distribution of N_1 is known to be Poisson with mean λ. This follows, for instance, from a more general result of Arratia, Barbour and Tavaré [1] on the limiting distribution of the cycle lengths in a random permutation under the Ewens sampling formula (see p.95 of [3]). In [1] purely combinatorial methods were used, while Csörgő and Wu found certain recursive formulas for the probabilities in question by probabilistic arguments. Since their method did not seem applicable when $\lambda \ne 1$, the case of general μ and λ remained open.

1991 *Mathematics Subject Classification.* Primary 60E05, Secondary 62E15,.

Key words and phrases. Ewens sampling formula, beta mixture of Poisson distribution, generating function.

Research supported by the Hungarian National Foundation for Scientific Research, Grant No. T-29621

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

In Section 2 of the present note we will show that, under (1.2), the distribution of N_i is identical with a beta mixture of Poisson distributions. Our approach aims at the probability generating functions instead of the probabilities themselves. In their paper [2] the authors also call for the study of series similar to (1.1) but with multiple products as summands. In Section 3 we derive some general formulas concerning the distribution of random variables

$$N_i^{(d)} = \sum_{r=i}^{\infty} X_r X_{r+1} \cdots X_{r+d},$$

with $d \geq 1$ (for $d = 1$ this coincides with (1.1)). In particular, the distribution of $N_i^{(2)}$ will be computed explicitly in the case (1.2).

2. Main results

In order that the series defining N_1 converge with probability 1 it is sufficient (and also necessary) that

$$\sum_{r=1}^{\infty} p_r p_{r+1} < \infty, \tag{2.1}$$

Indeed, let

$$N_o = \sum_{r=1}^{\infty} X_{2r-1} X_{2r}, \quad N_e = \sum_{r=1}^{\infty} X_{2r} X_{2r+1};$$

then the necessary and sufficient condition for the two series to converge is that the events $\{X_r X_{r+1} = 1\}$ occur only for finitely many values of r. Since these series have independent summands, both parts of the Borel–Cantelli lemma can be applied.

In the sequel we always suppose that (2.1) holds. Note that this condition even implies the finiteness of $g_1(s)$ for every positive s:

$$\begin{aligned} g_1(s) = E\left(s^{N_o} s^{N_e}\right) &\leq E^{1/2}\left(s^{2N_o}\right) E^{1/2}\left(s^{2N_e}\right) \\ = \prod_{r=1}^{\infty} \left[1 + (s^2 - 1) p_{2r-1} p_{2r}\right]^{1/2} &\prod_{r=1}^{\infty} \left[1 + (s^2 - 1) p_{2r} p_{2r+1}\right]^{1/2} \\ \leq \prod_{r=1}^{\infty} \exp\left(\tfrac{1}{2}(s^2 - 1) p_{2r-1} p_{2r}\right) &\prod_{r=1}^{\infty} \exp\left(\tfrac{1}{2}(s^2 - 1) p_{2r} p_{2r+1}\right) \\ = \exp\Big(\tfrac{1}{2}(s^2 - 1) &\sum_{r=1}^{\infty} p_r p_{r+1}\Big) < \infty. \end{aligned}$$

For $n \geq 0$ let $a_{n,i} = E\binom{N_i}{n}$. Then we have

$$g_i(s) = \sum_{n=0}^{\infty} a_{n,i} (s-1)^n.$$

The generating functions $g_i(s)$ and their coefficients $a_{n,i}$ are determined by the following recursion.

Lemma 2.1.

$$
\begin{aligned}
&g_i(s) = [1 + p_i(s-1)]\, g_{i+1}(s) - p_i\,(1 - p_{i+1})\,(s-1) g_{i+2}(s),\\
&a_{0,i} = 1,\\
&a_{n,i} - a_{n,i+1} = p_i\,(a_{n-1,i+1} - a_{n-1,i+2}) + p_i p_{i+1} a_{n-1,i+2},\ n \geq 1,\\
&a_{n,i} \to 0 \ \textit{as}\ i \to \infty,\ n \geq 1.
\end{aligned}
$$

Proof. Let us apply the theorem of total expectation to $g_i(s)$ according to the first $r \geq i$ for which $X_r = 0$. We obtain

$$
\begin{aligned}
g_i(s) &= (1-p_i)\, g_{i+1}(s) + \sum_{r=i+1}^{\infty} p_i \cdots p_{r-1}\,(1-p_r)\, s^{r-i-1} g_{r+1}(s)\\
&= (1-p_i)\, g_{i+1}(s) + p_i\,(1-p_{i+1})\, g_{i+2}(s) + p_i s\,[g_{i+1}(s) - (1-p_{i+1})\, g_{i+2}(s)]\\
&= [1 + p_i(s-1)]\, g_{i+1}(s) - p_i\,(1-p_{i+1})\,(s-1) g_{i+2}(s),
\end{aligned}
$$

which is just the first line of the lemma. By expanding the generating functions and comparing the corresponding coefficients on the two sides one arrives at the second and third lines. The last relation follows from the monotone convergence theorem.

As a corollary, it immediately follows that

$$
a_{1,i} = \sum_{j=i}^{\infty} p_j p_{j+1},
$$

which is otherwise due to the fact that $a_{1,i} = EN_i$. By continuing the iteration we can derive explicit formulas for $a_{n,i}$; but they become more and more complicated as n grows. Complicated formulas become simpler in the particular case (1.2). For the sake of brevity let us write

$$
\beta_{i,j} = p_i p_{i+1} \cdots p_{i+j},\ i \geq 1, j \geq 0, \quad c_i = 1 + \frac{\lambda}{i},\ i \geq 1.
$$

These quantities then satisfy the recursion

$$
\beta_{i,r} = \frac{\lambda}{r}\,(\beta_{i,r-1} - \beta_{i+1,r-1}) \tag{2.3}
$$

for arbitrary integers $r \geq 1$, $i \geq 1$.

Theorem 2.1.

$$
a_{n,i} - a_{n,i+1} = \beta_{i,n} c_1 c_2 \cdots c_{n-1}, \tag{2.4}
$$

$$
a_{n,i} = \frac{\lambda}{n} \beta_{i,n-1} c_1 \cdots c_{n-1} = \frac{\lambda^n}{n!} \prod_{r=0}^{n-1} \frac{\lambda + r}{\mu + i + r - 1}. \tag{2.5}
$$

Proof. First of all, notice that (2.5) follows from (2.4) by summation, since

$$
a_{n,i} - a_{n,i+1} = \frac{\lambda}{n}\,(\beta_{i,n-1} - \beta_{i+1,n-1})\, c_1 c_2 \cdots c_{n-1}
$$

by (2.3).

Now the proof can be carried by induction over n. For $n = 1$ the third line of Lemma 2.1 reads as $a_{1,i} - a_{1,i+1} = p_i p_{i+1} = \beta_{i,1}$. In order to prove (2.4) for $n+1$ with a general n let us use Lemma 2.1 again, then apply the induction hypothesis to $a_{n,i+1} - a_{n,i+2}$ and $a_{n,i+2}$. We obtain that

$$\begin{aligned} a_{n+1,i} - a_{n+1,i+1} &= p_i \beta_{i+1,n} c_1 c_2 \cdots c_{n-1} + p_i p_{i+1} \frac{\lambda}{n} \beta_{i+2,n-1} c_1 c_2 \cdots c_{n-1} \\ &= \beta_{i,n+1} c_1 c_2 \cdots c_{n-1} \left(1 + \frac{\lambda}{n}\right) \\ &= \beta_{i,n+1} c_1 c_2 \cdots c_n, \end{aligned}$$

as needed.

Thus the generating functions are

$$g_i(s) = \sum_{n=0}^{\infty} \frac{[\lambda(s-1)]^n}{n!} \prod_{r=0}^{n-1} \frac{\lambda + r}{\mu + i - 1 + r}. \tag{2.6}$$

This function is known as the confluent hypergeometric function

$${}_1F_1[\lambda, \mu + i - 1, \lambda(s-1)],$$

and the corresponding probability distribution is the so-called beta mixture of Poisson distribution, see p.330 of [4], which can be obtained in the following way. Let V be a random variable of beta distribution with parameters λ and $\mu - \lambda + i - 1$, and let the conditional distribution of ξ, supposed V is given, be Poisson with parameter λV; that is,

$$P(\xi = k | V) = \frac{(\lambda V)^k}{k!} e^{-\lambda V}, \quad k = 0, 1, \dots.$$

Then the probability generating function of ξ is just g_i.

By expanding (2.6) into power series we obtain the probability $P(N_i = k)$ as the coefficient of s^k.

Corollary 2.1.

$$\begin{aligned} P(N_i = k) &= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} \prod_{r=0}^{n-1} \frac{\lambda + r}{\mu + i - 1 + r} (-1)^{n-k} \binom{n}{k} \\ &= \frac{\lambda^k}{k!} \sum_{j=0}^{\infty} \frac{(-\lambda)^j}{j!} \prod_{r=0}^{j+k-1} \frac{\lambda + r}{\mu + i - 1 + r}. \end{aligned} \tag{2.7}$$

Transparently, N_1 possesses the Poisson(λ) distribution when $\lambda = \mu$.

3. Sums of multiple products

Let $d \geq 1$, and

$$N_i^{(d)} = \sum_{r=i}^{\infty} X_r X_{r+1} \cdots X_{r+d}, \quad g_i^{(d)}(s) = E\left(s^{N_i^{(d)}}\right). \tag{3.1}$$

Similarly to the case $d = 1$ it is easy to see that the convergence of the series $\beta_{1,d} + \beta_{2,d} + \cdots$ is sufficient for the generating functions $g_i^{(d)}(s)$ to be finite for arbitrary positive s. Again, let

$$g_i^{(d)}(s) = \sum_{n=0}^{\infty} a_{n,i}^{(d)} (s-1)^n.$$

Then the following analogue to Lemma 2.1 can be proved.

Lemma 3.1.

$$g_i^{(d)}(s) = g_{i+1}^{(d)}(s) + (s-1)\sum_{j=0}^{d-1} \beta_{i,j}\left[g_{i+j+1}^{(d)}(s) - g_{i+j+2}^{(d)}(s)\right] + (s-1)\beta_{i,d}\, g_{i+d+1}^{(d)}(s),$$

$$a_{0,i}^{(d)} = 1,$$

$$a_{n,i}^{(d)} - a_{n,i+1}^{(d)} = \sum_{j=0}^{d-1} \beta_{i,j}\left[a_{n-1,i+j+1}^{(d)} - a_{n-1,i+j+2}^{(d)}\right] + \beta_{i,d} a_{n-1,i+d+1}^{(d)}, \ n \geq 1,$$

$$a_{n,i}^{(d)} \to 0 \ \textit{as } i \to \infty, \ n \geq 1.$$

Proof. The recursion for the generating functions can be proved in the same way as it was done in Section 2 in the case of $d = 1$. Then the other recursion for the coefficients will immediately follow. By the theorem of total expectation we have

$$\begin{aligned}
g_i^{(d)}(s) &= (1-p_i)\, g_{i+1}^{(d)}(s) + \sum_{j=1}^{d-1} \beta_{i,j-1}\,(1-p_{i+j})\, g_{i+j+1}^{(d)}(s) + \\
&\qquad\qquad + \sum_{j=d}^{\infty} \beta_{i,j-1}\,(1-p_{i+j})\, s^{j+1-d} g_{i+j+1}^{(d)}(s) \\
&= (1-p_i)\, g_{i+1}^{(d)}(s) + \sum_{j=1}^{d-1} \beta_{i,j-1}\,(1-p_{i+j})\, g_{i+j+1}^{(d)}(s) + \\
&\qquad\qquad + s\Big[p_i g_{i+1}^{(d)}(s) - \sum_{j=1}^{d-1} \beta_{i,j-1}\,(1-p_{i+j})\, g_{i+j+1}^{(d)}(s)\Big] \\
&= g_{i+1}^{(d)}(s) + (s-1)p_i g_{i+1}^{(d)}(s) - (s-1)\sum_{j=1}^{d-1} \beta_{i,j-1}\,(1-p_{i+j})\, g_{i+j+1}^{(d)}(s) \\
&= g_{i+1}^{(d)}(s) + (s-1)\sum_{j=0}^{d-1} \beta_{i,j}\left[g_{i+j+1}^{(d)}(s) - g_{i+j+2}^{(d)}(s)\right] + (s-1)\beta_{i,d}\, g_{i+d+1}^{(d)}(s).
\end{aligned}$$

In the rest of the paper we are going to study the case $d = 2$ in detail. Then the coefficients satisfy

$$a_{n,i}^{(2)} - a_{n,i+1}^{(2)} = p_i \left[a_{n-1,i+1}^{(2)} - a_{n-1,i+2}^{(2)} \right] + \\ + p_i p_{i+1} \left[a_{n-1,i+2}^{(2)} - a_{n-1,i+3}^{(2)} \right] + p_i p_{i+1} p_{i+2} a_{n-1,i+3}^{(2)}. \quad (3.2)$$

Let us choose the probabilities according to (1.2). From (3.1) we first get

$$a_{1,i}^{(2)} - a_{1,i+1}^{(2)} = \beta_{i,2}, \quad a_{1,i}^{(2)} = \frac{\lambda}{2}\beta_{i,1},$$

and then by plugging back into (3.2) we obtain

$$\begin{aligned} a_{2,i}^{(2)} - a_{2,i+1}^{(2)} &= \beta_{i,3} + \beta_{i,4} + \frac{\lambda}{2}\beta_{i,4} = \beta_{i,3} + c_2\beta_{i,4}, \\ a_{2,i}^{(2)} &= \frac{\lambda}{3}\beta_{i,2} + \frac{\lambda}{4}c_2\beta_{i,3}, \\ a_{3,i}^{(2)} - a_{3,i+1}^{(2)} &= [\beta_{i,4} + c_2\beta_{i,5}] + [\beta_{i,5} + c_2\beta_{i,6}] + \left[\frac{\lambda}{3}\beta_{i,5} + \frac{\lambda}{4}c_2\beta_{i,6} \right] \\ &= \beta_{i,4} + (c_2 + c_3)\beta_{i,5} + c_2 c_4 \beta_{i,6}, \\ a_{3,i}^{(2)} &= \frac{\lambda}{4}\beta_{i,3} + \frac{\lambda}{5}(c_2 + c_3)\beta_{i,4} + \frac{\lambda}{6}c_2 c_4 \beta_{i,5}, \end{aligned}$$

and so on, one after another. For the general formula introduce $A_{0,r}^{(d)} = 1$, and

$$A_{j,r}^{(d)} = \sum_{d \le t_1,\ t_1 + d \le t_2,\ \cdots,\ t_{j-1} + d \le t_j,\ t_j + d \le r} c_{t_1} c_{t_2} \cdots c_{t_j}$$

for $j \ge 1$, $r \ge (j+1)d$. Obviously, $A_{j,r}^{(d)}$ is a polynomial of λ, with degree not greater than j.

Theorem 3.1.

$$a_{n,i}^{(2)} - a_{n,i+1}^{(2)} = \sum_{j=1}^{n} \beta_{i,n+j} A_{j-1,n+j}^{(2)}, \quad a_{n,i}^{(2)} = \sum_{j=1}^{n} \beta_{i,n+j-1} \frac{\lambda}{n+j} A_{j-1,n+j}^{(2)}. \quad (3.3)$$

Proof. This can be proved again by induction. Consider the recursion (3.2) and apply the induction hypothesis to the right-hand side.

$$\begin{aligned} a_{n+1,i}^{(2)} - a_{n+1,i+1}^{(2)} &= p_i \sum_{j=1}^{n} \beta_{i+1,n+j} A_{j-1,n+j}^{(2)} + p_i p_{i+1} \sum_{j=1}^{n} \beta_{i+2,n+j} A_{j-1,n+j}^{(2)} + \\ &\qquad + p_i p_{i+1} p_{i+2} \sum_{j=1}^{n} \beta_{i+3,n+j-1} \frac{\lambda}{n+j} A_{j-1,n+j}^{(2)} \\ &= \sum_{j=1}^{n} \left[\beta_{i,n+j+1} + \beta_{i,n+j+2} c_{n+j} \right] A_{j-1,n+j}^{(2)} \\ &= \beta_{i,n+2} A_{0,n+1}^{(2)} + \beta_{i,2n+2} c_{2n} A_{n-1,2n}^{(2)} + \\ &\qquad + \sum_{j=2}^{n} \beta_{i,n+j+1} \left[A_{j-1,n+j}^{(2)} + c_{n+j-1} A_{j-2,n+j-1}^{(2)} \right]. \end{aligned}$$

It is easy to see that $A^{(2)}_{0,n+1} = A^{(2)}_{0,n+2}$, $c_{2n}A^{(2)}_{n-1,2n} = A^{(2)}_{n,2n+2}$; furthermore

$$A^{(2)}_{j-1,n+j} + c_{n+j-1}A^{(2)}_{j-2,n+j-1} = A^{(2)}_{j-1,n+j+1}, \quad 2 \le j \le n.$$

Now, the first formula of (3.3) immediately follows, while the second one is obtained from the first one by summation.

As a corollary we can write down the generating functions.

$$\begin{aligned} g_i^{(2)}(s) &= 1 + \sum_{n=1}^{\infty}(s-1)^n \sum_{j=1}^{n} \beta_{i,n+j-1} \frac{\lambda}{n+j} A^{(2)}_{j-1,n+j} \\ &= 1 + \sum_{r=1}^{\infty} \beta_{i,r} \frac{\lambda}{r+1} \sum_{0\le j<r/2} A^{(2)}_{j,r+1}(s-1)^{r-j}. \end{aligned} \tag{3.4}$$

Here (3.4) better reflects how the generating function depends on μ: $g_i^{(2)}(s)$ depends on i and μ through $i+\mu$, and $i+\mu$ is only contained in the factors $\beta_{i,r}$.

Finally, the distribution of $N_i^{(2)}$ can be obtained from (3.4) by expanding it into power series.

Corollary 3.1. *For* $k = 0, 1, 2, \dots$

$$P\left(N_i^{(2)} = k\right) = \sum_{n=k}^{\infty}(-1)^{n-k}\binom{n}{k}\sum_{j=1}^{n}\beta_{i,n+j-1}\frac{\lambda}{n+j}A^{(2)}_{j-1,n+j}.$$

Acknowledgement

The author is indebted to Sándor Csörgő for calling his attention to the topic.

References

1. Arratia, R., Barbour, A. D. and Tavaré, S., *Poisson process approximations for the Ewens sampling formula*, Ann. Appl. Probab. **2** (1992), 519–535.
2. Csörgő, S. and Wu, W. B., *On sums of overlapping products of independent Bernoulli random variables*, Ukraïn'skiĭ Matematicheskiĭ Zhurnal **52** (2000), 1304–1309.
3. Ewens, W. J., *Mathematical Population Genetics*, Springer, Berlin, 1979.
4. Johnson, N. L., Kotz, S. and Kemp, A. W., *Univariate Discrete Distributions*, 2nd Ed., Wiley, New York, 1992.

Department of Probability Theory and Statistics,
Eötvös Loránd University,
Kecskeméti u. 10-12, Budapest, Hungary H-1053
E-mail address: moritamas@ludens.elte.hu