On a Particular Sum of Dependent Bernoulli and its Relationship to a Matching Type Problem

Anatole Joffe[*†] Éric Marchand[‡†]
François Perron[*†] Paul Popadiuk[§]

CRM2686

August 2000

[*]Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Qc, CANADA H3C 3J7

[†]Research supported by NSERC of Canada

[‡]Department of Mathematics and Statistics, University of New Brunswick, P.0. Box 4400, Fredericton, N.B. CANADA, E3B 5A3

[§]Department of Mathematics and Statistics, Concordia University, 7141 Sherbrooke St. West, Montréal, Qc, CANADA H4B 1R6

Summary

Let $S_n = \sum_{k=1}^{n} X_k X_{k+1}$ and $S = \lim_{n\to\infty} S_n$ where $\{X_k\}_{k=1}^{\infty}$ are independent Bernoulli random variables with mean p_k. For the particular case when $p_k = \frac{1}{k+B}$ with $B \geq 0$, we show that the distribution of S is a Beta mixture of Poisson distributions. We also give an interesting connection with a matching type problem.

Résumé

Considérons $\{X_k\}_{k=1}^{\infty}$ un suite de variables aléatoires Bernoulli indépendantes avec $\mathrm{P}(X_k = 1) = p_k$, et $\sum_{k\geq 1} p_k p_{k+1} < \infty$. Posons $S_n = \sum_{k=1}^{n} X_k X_{k+1}$ et $S = \lim_{n\to\infty} S_n$. Nous trouvons une forme explicite pour la fonction génératrice des probabilités de S_n. On montre que dans le cas particulier où $p_k = \frac{1}{k+B}$, où $B \geq 0$, la distribution de S devient un mélange de lois de Poisson.

1. INTRODUCTION

Let $\{X_k\}_{k=1}^{\infty}$ be a set of independent Bernoulli random variables with $\mathrm{P}(X_k = 1) = p_k$, and $\sum_{k\geq 1} p_k p_{k+1} < \infty$. We set $S_n = \sum_{k=1}^{n} X_k X_{k+1}$ and $S = \sum_{k=1}^{\infty} X_k X_{k+1}$. We are interested in the distribution of S.

The result according to which this limiting distribution is a Poisson distribution with mean 1 when $p_k = \frac{1}{k}$ is, to our knowledge, attributable to Diaconis. The proof, which is said to be "combinatorial" in nature, does not seem to lend itself easily to more general p_k's. Section 2 gives a novel and analytical proof of the above result, which does lend itself to extensions. We do this by establishing an explicit recurrence for the probability generating functions of S_n. For the particular case when $p_k = \frac{1}{k+B}$ with $B \geq 0$, we show that the distribution of S is a Beta mixture of Poisson distributions.

For the cases where $p_k = \frac{1}{k+B}$ with B being a non negative integer, we also establish in Section 3 a connection between the distribution of S_n and the distribution of the number of fixed points Z_n among the first n drawn points in a uniformly randomly generated permutation of $n + B$ points. We then obtain for the latter problem the probability generating function of Z_n thus yielding an alternate derivation of the distributions of S_n and S.

For the case $B = 0$, the distribution of the number of fixed points is a classical problem in probability and was first analyzed by Montmort (1713). It also is a problem which has been discussed extensively and which has led to numerous extensions (see Feller, 1968 or Johnson, Kemp and Kotz, 1992, chapter 10, among others, for a survey). For instance, the distribution of Z_n $(B = 0)$ is related to the number of cycles of length 1 in a uniform random permutation. More generally, Lengyel(1997) considers the moments of the number of cycles of length r, while Arratia, Goldstein, and Gordon (1990) study the joint distribution of the number of cycles of length 1,2,...f(n) and its proximity to its asymptotic distribution, that is the joint distribution of independent Poisson(1), Poisson(2),...Poisson(f(n)) random variables. Finally, we note that Michel Émery (1998) has also established some of the results of this paper with intricate but interesting methods of proof.

2. ANALYTICAL APPROACH

We define the random variables $\{S_n\}_{n\geq 0}$ as above with $S_0 = 0$. As well, we define the random variables $\{W_n\}_{n\geq 0}$ so that $W_0 = 0$, and $W_n = S_{n-1}+X_n$; $n \geq 1$. We exploit the relationship between the W_n's and S_n's, which share the same limiting distribution S, in setting up a recursive set of equations involving their probability generating functions $G_n(t) = E[t^{S_n}]$ and $H_n(t) = E[t^{W_n}]$; $n \geq 0$. This relationship is given by Lemma 1. Corollary 1 gives the recurrence in terms of the H_n's alone and offers a further specialization for certain types of p_k's, which includes the cases where $p_k = \frac{1}{k+B}$.

Lemma 1. *We have*
(a) $G_n(t) = p_{n+1}H_n(t) + (1 - p_{n+1})G_{n-1}(t)$; $n \geq 1$;
(b) $H_n(t) = (1 - p_n)G_{n-2}(t) + t\, p_n H_{n-1}(t)$; $n \geq 2$.

Proof. The result follows since $(S_n|X_{n+1} = 1) =^d W_n$, $(S_n|X_{n+1} = 0) =^d S_{n-1}$; $n \geq 1$;
$(W_n|X_n = 1) =^d 1 + W_{n-1}$, $(W_n|X_n = 0) =^d S_{n-2}$; $n \geq 2$.

Corollary 1. *(a) We have for $n \geq 1$,*

$$H_{n+1}(t) - H_n(t) = (t-1)[p_{n+1}H_n(t) - p_n(1-p_{n+1})H_{n-1}(t)].$$

(b) Whenever the p_k's satisfy the conditions $p_k(1-p_{k+1}) = \alpha p_{k+1}$ for some $\alpha \geq 1$,

$$H_{n+1}(t) - H_n(t) = (t-1)p_{n+1}[H_n(t) - \alpha H_{n-1}(t)].$$

Proof. Part (b) is immediate from part (a). For part (a), we will establish that

$$\begin{aligned} H_n(t) &= (t-1)p_nH_{n-1}(t) + G_{n-1}(t) \,, && (1) \\ H_n(t) &= (t-1)p_n(1-p_{n+1})H_{n-1}(t) + G_n(t) \,; && (2) \end{aligned}$$

which will lead to the desired result. From parts (b) and (a) of Lemma 1, $H_n(t) = (t-1)p_nH_{n-1}(t) + [p_nH_{n-1}(t) + (1-p_n)G_{n-2}(t)] = (t-1)p_nH_{n-1}(t) + G_{n-1}(t)$, which yields (1). Substituting (1) into part (a) of Lemma 1, we obtain $G_n(t) = (t-1)p_{n+1}p_nH_{n-1}(t) + p_{n+1}G_{n-1}(t) + (1-p_{n+1})G_{n-1}(t) = (t-1)p_{n+1}p_nH_{n-1}(t) + G_{n-1}(t)$. Finally, merging (1) with this last expression leads to (2) and concludes the proof.

Remark 1. *Here is an alternative proof of Corollary 1 which bypasses Lemma 1, and involves the W_k's only. Let K be the arrival time of the last success in the sequence $X_1, ..., X_n$ in other words*

$$K = \max\{\ell : 0 \leq \ell \leq n \text{ and } X_\ell = 1\}$$

with $X_0 = 1$. Let $H_j(t) = 1$ for $j = -1$. Since, for $n \geq 1$, $(W_n|K = n) =^d 1 + W_{n-1}$, and $(W_n|K = j) =^d W_{j-1}$ for $1 \leq j \leq n-1$; using the fact that $P_n[K = j] = (1-p_n)P_{n-1}[K = j]$ for $j = 0, 1, \ldots, n-1$, $n \geq 1$ we obtain

$$\begin{aligned} H_n(t) &= \sum_{j=0}^{n} P_n[K=j]E[t^{W_n}|K=j] \\ &= \sum_{j=0}^{n-1}(1-p_n)P_{n-1}[K=j]H_{j-1}(t) + p_ntH_{n-1}(t) \\ &= (1-p_n)[\{\sum_{j=0}^{n-2} P_{n-1}[K=j]H_{j-1}(t) + p_{n-1}tH_{n-2}(t)\} - p_{n-1}(t-1)H_{n-2}(t)] + p_ntH_{n-1}(t) \\ &= (1-p_n)[H_{n-1}(t) - p_{n-1}(t-1)H_{n-2}(t)] + p_ntH_{n-1}(t) \\ &= H_{n-1}(t) + (t-1)[p_nH_{n-1}(t) - (1-p_n)p_{n-1}H_{n-2}(t)], \end{aligned}$$

which leads directly to part (a) of Corollary 1.

Now, we obtain our main result by solving the recurrence of Corollary 1 for the cases where $\alpha = 1$, in other words $p_k = \frac{1}{k+B}$.

Theorem 1. *Whenever $p_k = \frac{1}{k+B}$; $B \geq 0$; we have*

(a) $H_n(t) = \sum_{k=0}^{n} \frac{(t-1)^k}{(1+B)_k}$ *and* $G_n(t) = 1 + \sum_{k=1}^{n} \frac{(t-1)^k}{(1+B)_k}(\frac{n+1-k}{n+B+1})$;

(b) The probability generating function of S is $E[t^S] = {}_1F_1(1; 1+B; t-1)$;

(c) *The distribution of S admits the following beta mixture of Poisson representation: $S|P =^d$ Poisson(P) with $P =^d$ Beta$(1, B)$. In particular, S is distributed as Poisson with mean 1 when $B = 0$.*

Proof. Part (b) of Corollary 1, when $\alpha = 1$, implies that $H_{n+1}(t) = H_n(t) + [H_1(t) - H_0(t)] \prod_{k=1}^{n}(t-1)p_{k+1} = H_n(t) + (t-1)^{n+1} \prod_{k=1}^{n+1} p_k$, since $H_1(t) = 1 + (t-1)p_1$ and $H_0(t) = 1$. The probability generating function for H_n is derived from this recurrence, and leads to that of G_n via Lemma 1. Parts (b) and (c) follow directly from part (a).

Remark 2. *We have $E(S) = \frac{1}{1+B}$, $Var(S) = \frac{1}{(1+B)}[1 + \frac{B}{(1+B)(2+B)}]$, and $P[S = x] = \frac{1}{(B+1)_x}\, {}_1F_1(x+1, x+B+1, -1)$; $x = 0, 1, 2, ...$; which may be derived from the probability generating function of S, or alternatively from the mixture representation. The distribution of S is of the GHF type (Generalized Hypergeometric Factorial), and the Beta mixture of Poisson representation of S is known (see Johnson, Kotz, and Kemp, 1992, chapter 2).*

3. COMBINATORIAL APPROACH

This approach relies on the connection between the distributions of W_n and Z_n where Z_n arises in the matching type problem that follows.

Assume that the elements 1,2,...n+B ($B \in \{0, 1, 2, ...\}$) are permuted according to a uniformly generated permutation π and drawn as follows. First, draw element $\pi(1)$. If $\pi(1) = 1$, a cycle of length 1 is completed and we start over with elements 2,...,n+B. If $\pi(1) = j$ with $j \neq 1$, draw $\pi(j), \pi(\pi(j)), \ldots$ and so on until element 1 is drawn completing a cycle. Then continue the process with the remaining elements (in increasing order). What is the distribution of the number of matches (or fixed points) Z_n among the first n drawn elements?

Equivalently, Z_n may be defined as the number of cycles of length 1 among the first n drawings. The next result establishes the equivalence between the distributions of Z_n and W_n. It is followed by a derivation of the probability generating function $M_n(t) = E[t^{Z_n}]$. [5]

Theorem 2. *For $n \geq 1$, the distribution of Z_n is identical to that of W_n with $p_k = \frac{1}{k+B}$.*

Proof. Let us define the Bernoulli random variables $\{Y_k\}_{k=1}^{n}$ as $Y_k = 1$ if and only if a cycle is completed at the kth drawing. It follows that a match occurs for the first drawing if and only if $Y_1 = 1$, and on the kth drawing if and only if $Y_{k-1}Y_k = 1$. Hence $Z_n =^d Y_1 + \sum_{k=2}^{n} Y_k Y_{k-1}$, with P$(Y_k = 1) = 1 - P(Y_k = 0) = \frac{1}{n+B-k+1}$. Finally, set $X_k = Y_{n-k+1}$; $k = 1, ..., n$; which implies $Z_n =^d X_n + \sum_{k=1}^{n-1} X_k X_{k+1} =^d W_n$, since the X_k's are also independent Bernoulli random variables with $p_k = P(X_k = 1) = P(Y_{n-k+1} = 1) = \frac{1}{B+k}$.

Theorem 3. *Setting Z_0=0, the probability generating function of Z_n is given by*

$$M_n(t) = \sum_{k=0}^{n} \frac{(t-1)^k}{(1+B)_k}.$$

[5] Since the case $B = 0$ is the classical matching case, the distributions of W_n and S given in Theorem 1 for $B = 0$ may be obtained directly from the well known distribution of Z_n.

Proof. We condition on the length τ of the first cycle. Since the length is uniformly distributed on $\{1, ..., n+B\}$, we have

$$M_n(t) = \sum_{j=1}^{n+B} E[t^{Z_n}|\tau = j]\frac{1}{n+B} = \frac{1}{n+B}[t\ M_{n-1}(t) + \sum_{j=2}^{n} M_{n-j}(t) + B].$$

Now, by considering the expression $(n+B+1)M_{n+1}(t) - (n+B)M_n(t)$, we obtain the recurrence

$$M_{n+1}(t) - M_n(t) = \frac{(t-1)}{n+B+1}[M_n(t) - M_{n-1}(t)],$$

which coincides with the recurrence given in part (b) of Corollary 1 and, consequently, the proof of Theorem 1 yields the result.

References

Arratia, R., Goldstein, L., and Gordon, L., 1990. Poisson approximation and the Chen-Stein method (with comments and a rejoinder by the authors). *Statistical Science*, 4, 403-434.

Émery, M., 1998. Sur un problème de Diaconis. Unpublished manuscript.

Feller, W., 1968. *An introduction to probability theory and its applications*, volume 1, 2nd ed., Wiley, New York.

Johnson, N.L., Kotz, S., and Kemp, A.W., 1992. *Univariate discrete distributions*, 2nd ed., Wiley, New York.

Lengyel, T., 1997. The moments of the number of cycles of a random permutation by simple enumeration. *Sankhyā, Series A*, 59, 133-137.

Montmort, P.R., 1713. *Essay d'analyse sur les jeux de hazard*, 2nd ed., Paris (reprint Chelsea House, New York, 1981.)