J. Appl. Prob. **45,** 1181–1185 (2008)
Printed in England

A NOTE ON EMBEDDING CERTAIN BERNOULLI SEQUENCES IN MARKED POISSON PROCESSES

LARS HOLST,* *Royal Institute of Technology*

Abstract

A sequence of independent Bernoulli random variables with success probabilities $a/(a+b+k-1)$, $k = 1, 2, 3, \ldots$, is embedded in a marked Poisson process with intensity 1. Using this, conditional Poisson limits follow for counts of failure strings.

Keywords: Bernoulli trial; conditional limit theorem; Poisson limit; random permutation; record; sums of indicators

2000 Mathematics Subject Classification: Primary 60C05
Secondary 60K99

1. Introduction

Inspired by Huffer *et al.* (2008) we construct in this note an embedding in a marked Poisson process of a sequence of independent Bernoulli random variables with success probabilities $a/(a+b+k-1)$, $k = 1, 2, 3, \ldots$. From the embedding, conditional Poisson limit distributions follow for the number of d-strings, that is, subsequent successes interrupted by $d-1$ failures in the sequence. A special case is the Poisson limits for the number of small cycles in a random permutation biased by the number of cycles.

Other methods have previously been used to obtain such limits; see Arratia *et al.* (2003), Holst (2007), Holst (2008), Huffer *et al.* (2008), and the references therein. The embedding technique gives much more concise and transparent derivations and a better understanding of why the Poisson limits occur in such cases.

2. The embedding

Let $P, Z_1, Z_2, Z_3, \ldots$ be independent random variables, where the Zs are exponential with mean 1 and $0 < P \leq 1$. The waiting time for a Z to exceed $\log(1/P)$ is

$$L_0 = \min\left\{k\colon Z_k > \log\left(\frac{1}{P}\right)\right\},$$

having the following conditional geometric distribution:

$$\mathrm{P}(L_0 = \ell \mid P = p) = (1-p)^{\ell-1} p, \qquad \ell = 1, 2, \ldots.$$

By the lack of memory property of the exponential distribution, the excess $X_1 = Z_{L_0} - \log(1/P)$ is exponentially distributed with mean 1 and independent of (P, L_0). Set $T_1 = X_1$.

For $a > 0$, the waiting time

$$L_1 = \min\left\{k > L_0\colon Z_k > \log\left(\frac{1}{P}\right) + \frac{T_1}{a}\right\} - L_0$$

Received 27 August 2008; revision received 28 October 2008.
* Postal address: Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden.
Email address: lholst@math.kth.se

https://doi.org/10.1239/jap/1231340241 Published online by Cambridge University Press

has the conditional distribution

$$\mathrm{P}(L_1 = \ell \mid P = p,\ T_1 = t) = (1 - p\mathrm{e}^{-t/a})^{\ell-1} p\mathrm{e}^{-t/a}, \qquad \ell = 1, 2, \ldots.$$

The excess $X_2 = Z_{L_0+L_1} - \log(1/P) - T_1/a$ is exponentially distributed with mean 1 and independent of (P, L_0, L_1, T_1). Set $T_2 = T_1 + X_2$.

Analogously, the waiting time L_2 for the next Z to exceed $\log(1/P) + T_2/a$ is geometric as above and the excess X_3 is exponential with mean 1 and independent of $(P, L_0, L_1, L_2, T_1, T_2)$. Set $T_3 = T_2 + X_3$.

In the same way, define the waiting times $L_3, L_4, \ldots$, the excesses $X_4, X_5, \ldots$, and the random variables $T_4, T_5, \ldots$. The sequence of 'records', $T_1, T_2, T_3, \ldots$, is a Poisson process with intensity 1. Conditional on $P = p$, $\{(T_i, L_i),\ i = 1, 2, 3, \ldots\}$ is a marked Poisson process with the marking distribution

$$\mathrm{P}(L_i = \ell \mid P = p,\ T_i = t) = (1 - p\mathrm{e}^{-t/a})^{\ell-1} p\mathrm{e}^{-t/a}, \qquad \ell = 1, 2, \ldots.$$

To indicate the times for the records, we introduce the Bernoulli random variables $I_k = 1$ if $k \in \{L_0, L_0 + L_1, L_0 + L_1 + L_2, \ldots\}$, otherwise $I_k = 0$. For $P \equiv 1$ and $a = 1$, the Is indicate ordinary records among the Zs. Rényi's theorem shows that these indicators are independent with $\mathrm{P}(I_n = 1) = 1/n$. The theorem below generalizes this well-known result.

We say that a random variable P with density

$$f(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1}, \qquad 0 < p < 1,$$

is $\mathrm{Beta}(a, b)$, where $a > 0$ and $b > 0$; $\mathrm{Beta}(a, 0)$ is interpreted as $P \equiv 1$. Recall that

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

We use the notation $s^{\overline{n}} = s(s+1)\cdots(s+n-1)$ for rising factorials.

Theorem 2.1. *Let P be* $\mathrm{Beta}(a, b)$, *$a > 0$ and $b \geq 0$. Then the record indicators, $I_1, I_2, I_3, \ldots$, are independent random variables with* $\mathrm{P}(I_n = 1) = a/(a + b + n - 1)$.

Proof. We give a proof for the case in which $b > 0$. The proof is easily modified for $b = 0$, that is, for $P \equiv 1$.

Consider $I_1, I_2, \ldots, I_n$. We have

$$\mathrm{P}(I_1 = \cdots = I_n = 0) = \mathrm{P}(L_0 > n) = \mathrm{E}((1-P)^n) = \frac{b^{\overline{n}}}{(a+b)^{\overline{n}}}$$

and

$$\mathrm{P}(I_1 = \cdots = I_{n-1} = 0,\ I_n = 1) = \mathrm{P}(L_0 = n) = \mathrm{E}((1-P)^{n-1}P) = \frac{ab^{\overline{n-1}}}{(a+b)^{\overline{n}}}.$$

https://doi.org/10.1239/jap/1231340241 Published online by Cambridge University Press

Changing variables and integrating by parts we obtain, for $1 \le \ell < n$,

$$\begin{aligned} f_0(n,a,b,\ell) &= \mathrm{P}(L_0 = \ell,\ L_1 > n-\ell) \\ &= \mathrm{E}\biggl((1-P)^{\ell-1} P \int_0^\infty (1-P\mathrm{e}^{-x/a})^{n-\ell}\mathrm{e}^{-x}\,\mathrm{d}x\biggr) \\ &= \mathrm{E}\biggl((1-P)^{\ell-1} P \int_0^1 (1-Pu)^{n-\ell} a u^{a-1}\,\mathrm{d}u\biggr) \\ &= \mathrm{E}((1-P)^{n-1}P) + (n-\ell)\,\mathrm{E}\biggl((1-P)^{\ell-1}P^2 \int_0^1 (1-Pu)^{n-\ell-1}u^a\,\mathrm{d}u\biggr) \\ &= \frac{ab^{\overline{n-1}}}{(a+b)^{\overline{n}}} + \frac{a}{a+b}\frac{n-\ell}{a+1} f_0(n-1,a+1,b,\ell). \end{aligned}$$

Induction proves that

$$f_0(n,a,b,\ell) = \mathrm{P}(L_0 = \ell,\ L_1 > n-\ell) = \frac{ab^{\overline{n}}}{(a+b)^{\overline{n}}(b+\ell-1)}.$$

For $1 \le \ell_0,\dots,\ell_j,\ \ell_0+\cdots+\ell_j \le n$, set

$$f_j(n,a,b,\ell_0,\dots,\ell_j) = \mathrm{P}(I_k = 1 \text{ if } k \in \{\ell_0, \ell_0+\ell_1,\dots,\ell_0+\cdots+\ell_j\},\ \text{else } I_k = 0).$$

Changing variables we find that

$$\begin{aligned} &f_j(n,a,b,\ell_0,\dots,\ell_j) \\ &\quad = \mathrm{P}(L_0=\ell_0,\dots,L_j=\ell_j,\ L_{j+1} > n-\ell_0-\cdots-\ell_j) \\ &\quad = \mathrm{E}\biggl((1-P)^{\ell_0-1}P \\ &\qquad \times \int_0^\infty\cdots\int_0^\infty (1-P\mathrm{e}^{-x_1/a})^{\ell_1-1}P\mathrm{e}^{-x_1/a}\cdots(1-P\mathrm{e}^{-(x_1+\cdots+x_j)/a})^{\ell_j-1} \\ &\qquad\qquad \times P\mathrm{e}^{-(x_1+\cdots+x_j)/a}(1-P\mathrm{e}^{-(x_1+\cdots+x_{j+1})/a})^{n-\ell_0-\cdots-\ell_j} \\ &\qquad\qquad \times \mathrm{e}^{-(x_1+\cdots+x_{j+1})}\,\mathrm{d}x_1\cdots\mathrm{d}x_{j+1}\biggr) \\ &\quad = \mathrm{E}\biggl((1-P)^{\ell_0-1}Pa^{j+1} \\ &\qquad \times \int_0^1\cdots\int_0^1 (1-Pu_1)^{\ell_1-1}Pu_1\cdots(1-Pu_1\cdots u_j)^{\ell_j-1}Pu_1\cdots u_j \\ &\qquad\qquad \times (1-Pu_1\cdots u_{j+1})^{n-\ell_0-\cdots-\ell_j}u_1^{a-1}\cdots u_{j+1}^{a-1}\,\mathrm{d}u_1\cdots\mathrm{d}u_{j+1}\biggr) \\ &\quad = \mathrm{E}\biggl((1-P)^{\ell_0-1}Pa^{j} \\ &\qquad \times \int_0^1\cdots\int_0^1 (1-Pu_1)^{\ell_1-1}Pu_1^{a+j-1}\cdots(1-Pu_1\cdots u_j)^{\ell_j-1}Pu_j^{a} \\ &\qquad\qquad \times \biggl(\int_0^1 (1-Pu_1\cdots u_{j+1})^{n-\ell_0-\cdots-\ell_j}au_{j+1}^{a-1}\,\mathrm{d}u_{j+1}\biggr)\mathrm{d}u_1\cdots\mathrm{d}u_j\biggr). \end{aligned}$$

Integration by parts gives

$$\int_0^1 (1 - Pu_1 \cdots u_{j+1})^{n-\ell_0-\cdots-\ell_j} a u_{j+1}^{a-1} \, \mathrm{d}u_{j+1}$$
$$= (1 - Pu_1 \cdots u_j)^{n-\ell_0-\cdots-\ell_j}$$
$$+ (n - \ell_0 - \cdots - \ell_j) Pu_1 \cdots u_j \int_0^1 (1 - Pu_1 \cdots u_{j+1})^{n-\ell_0-\cdots-\ell_j-1} u_{j+1}^a \, \mathrm{d}u_{j+1},$$

implying the recursion

$$\begin{aligned}
&f_j(n, a, b, \ell_0, \dots, \ell_j) \\
&\quad = \frac{a}{a+b} \int_0^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} p^a (1-p)^{b-1} (1-p)^{\ell_0 - 1} p \\
&\qquad \times \biggl(a^j \int_0^1 \cdots \int_0^1 (1 - pu_1)^{\ell_1 - 1} pu_1 \cdots (1 - pu_1 \cdots u_{j-1})^{\ell_{j-1}-1} pu_1 \cdots u_{j-1} \\
&\qquad\qquad \times (1 - pu_1 \cdots u_j)^{n-1-\ell_0-\cdots-\ell_{j-1}} u_1^a \cdots u_j^a \, \mathrm{d}u_1 \cdots \mathrm{d}u_j \\
&\qquad + (n - \ell_0 - \cdots - \ell_j) a^j \int_0^1 \cdots \int_0^1 (1 - pu_1)^{\ell_1 - 1} pu_1 \cdots (1 - pu_1 \cdots u_j)^{\ell_j - 1} \\
&\qquad\qquad \times pu_1 \cdots u_j (1 - pu_1 \cdots u_{j+1})^{n-1-\ell_0-\cdots-\ell_j} u_1^a \cdots u_{j+1}^a \, \mathrm{d}u_1 \cdots \mathrm{d}u_{j+1} \biggr) \mathrm{d}p \\
&\quad = \frac{a}{a+b} \biggl(\frac{a^j}{(a+1)^j} f_{j-1}(n-1, a+1, b, \ell_0, \dots, \ell_{j-1}) \\
&\qquad + (n - \ell_0 - \cdots - \ell_j) \frac{a^j}{(a+1)^{j+1}} f_j(n-1, a+1, \ell_0, \dots, \ell_j) \biggr).
\end{aligned}$$

This is satisfied by

$$\begin{aligned}
&f_j(n, a, b, \ell_0, \dots, \ell_j) \\
&\quad = \mathrm{P}(L_0 = \ell_0, \dots, L_j = \ell_j,\ L_{j+1} > n - \ell_0 - \cdots - \ell_j) \\
&\quad = \frac{a^{j+1} b^{\overline{n}}}{(a+b)^{\overline{n}} (b + \ell_0 - 1)(b + \ell_0 + \ell_1 - 1) \cdots (b + \ell_0 + \cdots + \ell_j - 1)}.
\end{aligned}$$

From this, it follows that $I_1, I_2, I_3, \dots$ are independent Bernoulli random variables with $\mathrm{P}(I_k = 1) = a/(a+b+k-1)$.

3. Poisson limits

Conditional on $P = p$, the marking theorem in Kingman (1993, Section 5.2) shows that the sequences

$$\{(T_i, L_i = \ell),\ i = 1, 2, \dots\}, \qquad \ell = 1, 2, \dots,$$

are independent marked Poisson processes on the positive real line with intensities

$$\lambda_\ell(t) = (1 - p\mathrm{e}^{-t/a})^{\ell-1} p\mathrm{e}^{-t/a}, \qquad \ell = 1, 2, \dots .$$

Thus, the number of Ts marked with ℓ, N_ℓ, is Poisson with mean

$$\int_0^\infty \lambda_\ell(t)\,\mathrm{d}t = \frac{a}{\ell}(1-(1-p)^\ell)$$

and $N_1, N_2, \ldots$ are conditionally independent.

Let $I_1, I_2, I_3, \ldots$ be independent Bernoulli variables with success probabilities $a/(a+b+k-1)$, $k=1,2,\ldots$. By the above theorem, such a sequence can be considered as a record indicator in an embedding where P is $\text{Beta}(a,b)$. Consider the number of d-strings, that is,

$$M_d = \sum_{k=1}^\infty I_k(1-I_{k+1})\cdots(1-I_{k+d-1})I_{k+d},$$

which, by the embedding, can be identified by N_d. Hence, conditional on $P=p$, the random variables $M_1, M_2, \ldots$ are independent Poisson with means as above. This agrees with results in Holst (2007) and Huffer *et al.* (2008).

For $a=\theta>0$ and $b=0$, the Bernoulli variables above appear in connection with θ-biased random permutations; see Arratia *et al.* (2003, pp. 95, 96). The counts of different failure strings in $1I_2\cdots I_n1$ correspond to the number of cycles $C_1^{(n)}, C_2^{(n)}, \ldots, C_n^{(n)}$ of sizes $1,2,\ldots,n$ in a θ-biased random permutation of $1,2,\ldots,n$. The limit counts as $n\to\infty$ for the number of small cycles are given by independent Poisson random variables $M_1, M_2, \ldots$ with $\mathrm{E}(M_d)=\theta/d$; cf. Arratia *et al.* (2003, Theorem 5.1).

Finally, consider a sequence of independent indicators, $I_1\equiv 1, I_2, I_3, \ldots$, with $\mathrm{P}(I_k=1)=a/(a+b+k-2)$, $k=2,3,\ldots$, where $b\geq 1$. With Z exponential with mean 1 and independent of P, which is $\text{Beta}(a+1,b-1)$, we find that $P'=P\mathrm{e}^{-Z/a}$ is $\text{Beta}(a,b)$. Using P', we can generate, by the embedding, a sequence $I_1', I_2', \ldots$ with $\mathrm{P}(I_k'=1)=a/(a+b+k-1)$. For $k=2,3,\ldots$, set $I_k=I_{k-1}'$ with $\mathrm{P}(I_k=1)=a/(a+b+k-2)$. Conditional on $P=p$, the number of d-strings in $1I_2I_3\ldots$ is a Poisson random variable M_d with mean $a(1-(1-p)^d)/d$ and $M_1, M_2, \ldots$ are independent. This is in agreement with Huffer *et al.* (2008). For $b<1$, the distribution of M_d is not conditional Poisson; see Huffer *et al.* (2008).

Acknowledgement

I am grateful to an anonymous referee for suggesting an extension of the original embedding.

References

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

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

Holst, L. (2008). The number of two consecutive successes in a Hoppe–Pólya urn. *J. Appl. Prob.* **45,** 901–906.

Huffer, F., Sethuraman, J. and Sethuraman, S. (2008). A study of counts of Bernoulli strings via conditional Poisson processes. To appear in *Proc. Amer. Math. Soc.*

Kingman, J. F. C (1993). *Poisson Processes* (Oxford Studies Prob. **3**). Oxford University Press.

https://doi.org/10.1239/jap/1231340241 Published online by Cambridge University Press