A NOTE ON EMBEDDING CERTAIN BERNOULLI SEQUENCES IN MARKED POISSON PROCESSES

LARS HOLST

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

E-mail: lholst@math.kth.se

August 26, 2008

Abstract

A sequence of independent Bernoulli random variables with success probabilities $\theta/(\theta + k - 1)$, $k = 1, 2, 3, \ldots$, is embedded in a marked Poisson process of rate one. Using this, Poisson limits follow for counts of failure strings and for counts of small cycles in a θ-biased random permutation.

Keywords: Bernoulli trials; Poisson limits; records; random permutations; sums of indicators

2000 Mathematics Subject Classification: Primary 60C05
Secondary 60K99

Running title: Embedding in marked Poisson processes

1 Introduction

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

For the more general case with independent Bernoulli trials with success probabilities $a/(a + b + k - 1)$, $a > 0$, $b \geq 0$, $k = 1, 2, \ldots$, conditional Poisson limits are derived in Holst (2007). Introducing the embedding technique, a more general class of Bernoulli sequences is studied in Huffer *et al.* (2008). However, we think the special case considered in this note deserves special attention.

2 The embedding

Let $Z_1, Z_2, Z_3, \ldots$ be independent exponentially distributed random variables with mean one and set $T_1 = Z_1$. For $\theta > 0$ define the waiting time

$$L_1 = \min\{k > 1 : Z_k > T_1/\theta\} - 1,$$

having the following conditional geometric distribution

$$P(L_1 = \ell \mid T_1 = t) = \left(1 - e^{-t/\theta}\right)^{\ell-1} e^{-t/\theta}, \quad \ell = 1, 2, \ldots.$$

By the lack of memory property of the exponential distribution the excess $X_2 = Z_{1+L_1} - T_1/\theta$ is exponentially distributed with mean one and independent of T_1. Set $T_2 = T_1 + X_2$.

The waiting time for the next Z_k to exceed T_2/θ is

$$L_2 = \min\{k > 1 + L_1 : Z_k > T_2/\theta\} - 1 - L_1,$$

with the same conditional distribution

$$P(L_2 = \ell \mid T_2 = t) = \left(1 - e^{-t/\theta}\right)^{\ell-1} e^{-t/\theta}, \quad \ell = 1, 2, \ldots.$$

Analogously define the waiting times $L_3, L_4, \ldots$, the excesses $X_3, X_4, \ldots$, and the "θ-records" $T_3 = T_2 + X_3$, $T_4 = T_3 + X_4, \ldots$. The sequence $T_1, T_2, T_3, \ldots$ is a Poisson process of rate one and $\{(T_i, L_i), i = 1, 2, 3, \ldots\}$ is a marked Poisson process. To indicate the times for the θ-records introduce the Bernoulli random variables $I_k = 1$ if $k \in \{1, 1 + L_1, 1 + L_1 + L_2, \ldots\}$, else $I_k = 0$.

For $\theta = 1$ the random variable $I_n = 1$ if $Z_n > \max\{Z_1, \ldots, Z_{n-1}\}$, else $I_n = 0$, indicates an ordinary record among the Z's. Rényi's Theorem shows that these indicators are independent Bernoulli variables with $P(I_n = 1) = 1/n$. The following theorem generalizes this to general $\theta > 0$.

2

Theorem 2.1 *The indicators of θ-records $I_1, I_2, I_3, \ldots$ are independent Bernoulli random variables with $P(I_n = 1) = \theta/(\theta + n - 1)$.*

Proof. Note that $I_1 \equiv 1$. Consider $I_2, \ldots I_n$ and

$$f_0(n, \theta) = P(I_2 = \cdots = I_n = 0) = P(L_1 > n - 1) = E\left(\left(1 - e^{-T_1/\theta}\right)^{n-1}\right)$$

$$= \int_0^\infty (1 - e^{-x/\theta})^{n-1} e^{-x} dx = \theta \int_0^1 (1-u)^{n-1} u^{\theta-1} du = \frac{(n-1)!}{(\theta+1)(\theta+2)\cdots(\theta+n-1)}.$$

For $1 \le \ell_1, \ldots, \ell_j, 1 + \ell_1 + \cdots + \ell_j \le n$ set

$$f_j(n, \theta, \ell_1, \ldots, \ell_j) = P(I_k = 1 \text{ if } k \in \{1 + \ell_1, \ldots, 1 + \ell_1 + \cdots + \ell_j\}, \text{ else } I_k = 0).$$

We have

$$f_j(n, \theta, \ell_1, \ldots, \ell_j) = P(L_1 = \ell_1, \ldots, L_j = \ell_j, L_{j+1} > n - \ell_1 - \cdots - \ell_j - 1)$$

$$= \int_0^\infty \cdots \int_0^\infty (1 - e^{-x_1/\theta})^{\ell_1-1} e^{-x_1/\theta} \cdots (1 - e^{-(x_1+\cdots+x_j)/\theta})^{\ell_j-1} e^{-(x_1+\cdots+x_j)/\theta}$$

$$\times (1 - e^{-(x_1+\cdots+x_{j+1})/\theta})^{n-\ell_1-\cdots-\ell_j-1} e^{-(x_1+\cdots+x_{j+1})} dx_1 \cdots dx_{j+1}$$

$$= \theta^{j+1} \int_0^1 \cdots \int_0^1 (1 - u_1)^{\ell_1-1} u_1 \cdots (1 - u_1 \cdots u_j)^{\ell_j-1} u_1 \cdots u_j$$

$$\times (1 - u_1 \cdots u_{j+1})^{n-\ell_1-\cdots-\ell_j-1} u_1^{\theta-1} \cdots u_{j+1}^{\theta-1} du_1 \cdots du_{j+1}$$

$$= \theta^j \int_0^1 \cdots \int_0^1 (1 - u_1)^{\ell_1-1} u_1^{\theta+j-1} \cdots (1 - u_1 \cdots u_j)^{\ell_j-1} u_j^{\theta}$$

$$\times \left(\int_0^1 (1 - u_1 \cdots u_{j+1})^{n-\ell_1-\cdots-\ell_j-1} \theta u_{j+1}^{\theta-1} du_{j+1}\right) du_1 \cdots du_j.$$

Integration by parts gives

$$\int_0^1 (1 - u_1 \cdots u_{j+1})^{n-\ell_1-\cdots-\ell_j-1} \theta u_{j+1}^{\theta-1} du_{j+1} = (1 - u_1 \cdots u_j)^{n-\ell_1-\cdots-\ell_j-1}$$

$$+ (n - \ell_1 - \cdots - \ell_j - 1) u_1 \cdots u_j \int_0^1 (1 - u_1 \cdots u_{j+1})^{n-\ell_1-\cdots-\ell_j-2} u_{j+1}^{\theta} du_{j+1},$$

implying the recursion

$$f_j(n, \theta, \ell_1, \ldots, \ell_j) = \theta^j \left[\frac{1}{(\theta+1)^j} f_{j-1}(n-1, \theta+1, \ell_1, \ldots, \ell_{j-1}) \right.$$

$$\left. + (n - \ell_1 - \cdots - \ell_j - 1) \frac{1}{(\theta+1)^{j+1}} f_j(n-1, \theta+1, \ell_1, \ldots, \ell_j) \right].$$

This is satisfied by

$$f_j(n, \theta, \ell_1, \ldots, \ell_j) = \frac{(n-1)!\,\theta^j}{(\theta+1)\cdots(\theta+n-1)\ell_1(\ell_1+\ell_2)\cdots(\ell_1+\ell_2+\cdots+\ell_j)},$$

which proves that $I_1(\equiv 1), I_2, I_3, \ldots$ are independent Bernoulli random variables with $P(I_k = 1) = \theta/(\theta + k - 1)$. $\qquad\square$

3 Poisson limits

The Marking Theorem in Kingman (1993, Section 5.2) shows that the sequences

$$\{(T_i, L_i = \ell),\ i = 1, 2, \ldots\}, \quad \ell = 1, 2, \ldots,$$

give independent Poisson processes on the positive real line with intensities

$$\lambda_\ell(t) = \left(1 - e^{-t/\theta}\right)^{\ell-1} e^{-t/\theta}, \quad \ell = 1, 2, \ldots.$$

As $\int_0^\infty \lambda_\ell(t)dt = \theta/\ell$ the number of ℓ-marked θ-records, Z_ℓ, is Poisson distributed with mean θ/ℓ and $Z_1, Z_2, \ldots$ are independent random variables.

Let $I_1(\equiv 1), I_2, I_3, \ldots$ be independent Bernoulli variables with success probabilities $\theta/(\theta + k - 1)$, $k = 1, 2, \ldots$. Consider the number of d-strings, that is

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

By the embedding it follows that M_d can be identified by the number of points in $\{(T_i, L_i = d),\ i = 1, 2, \ldots\}$. Thus M_d is Poisson with mean θ/d and $M_1, M_2, \ldots$ are independent; *cf.* Holst (2007, Proposition 5).

The Bernoulli variables above appear in connection with θ-biased random permutations; see Arratia *et al.* (2003, p. 95, 96). The counts of different failure strings in $1I_2 \ldots I_n 1$ 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. The limit distribution as $n \to \infty$ for the number of small cycles is given by the independent Poisson random variables $M_1, M_2, \ldots$; *cf.* Arratia *et al.* (2003, Theorem 5.1).

4

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

[3] HUFFER, F., SETHURAMAN, J. AND SETHURAMAN, S. (2008). A study of counts of Bernoulli strings via conditional Poisson processes. *Preprint.*

[4] KINGMAN, J.F.C (1993). *Poisson processes.* Clarendon Press, Oxford.