Ark. Mat., **49** (2011), 351–356
DOI: 10.1007/s11512-010-0131-3

A note on records in a random sequence

Lars Holst

Abstract. In an infinite sequence of independent identically distributed continuous random variables we study the number of strings of two subsequent records interrupted by a given number of non-records. By embedding in a marked Poisson process we prove that these counts are independent and Poisson distributed. Also the distribution of the number of uninterrupted strings of records is considered.

1. Introduction

The records in a given sequence of numbers are the elements in the sequence with values strictly larger than all previous ones. The literature on different aspects of records is extensive, see for example [3] and [15] and the numerous references therein.

In this paper we consider the record model where the sequence of numbers is generated by independent identically distributed continuous random variables. Set $I_k = 1$ if the kth element in the sequence is a record, set otherwise $I_k = 0$. By symmetry the random variable I_k has the Bernoulli distribution

$$P(I_k = 1) = 1 - P(I_k = 0) = \frac{1}{k},$$

and the *record indicators* $I_1, I_2, \dots$ are independent. From the Borel–Cantelli lemma, independence, and the divergence of the harmonic series

$$\sum_{k=1}^{\infty} P(I_k = 1) = \sum_{k=1}^{\infty} \frac{1}{k} = +\infty$$

it follows that with probability one infinitely many records occur. Note that the number of records grows very slowly in this model as the expected number of records in the first n trials is $\sum_{k=1}^{n} 1/k$. We will study how the rather few records occur among the many non-records.

In Section 2 we derive the distribution of the number of strings where two subsequent records are interrupted by a given number of non-records. The number of strings consisting only of records is studied in Section 3.

2. Counts of non-records between records

The number of double records, that is two records immediately after each other, can be written

$$M_1 = \sum_{k=1}^{\infty} I_k I_{k+1},$$

and, more generally, the number of strings with two subsequent records interrupted by $d-1$ non-records is

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

As

$$E(M_d) = \sum_{k=1}^{\infty} \frac{1}{k}\frac{k}{k+1}\dots\frac{k+d-2}{k+d-1}\frac{1}{k+d} = \sum_{k=1}^{\infty}\left(\frac{1}{k+d-1}-\frac{1}{k+d}\right) = \frac{1}{d}$$

the Borel–Cantelli lemma shows that M_d is finite with probability one.

Theorem 2.1. *The counts $M_1, M_2, \dots$ are independent Poisson random variables with expectations $E(M_d)=1/d$, $d=1,2,\dots$.*

Proof. Any continuous distribution can be used in the model we consider. Let the random sequence be given by $Z_1, Z_2, \dots$, where the Z_k's are independent exponential random variables all with mean 1, that is for, $k=1,2,\dots$,

$$P(Z_k > x) = e^{-x}, \quad x > 0.$$

Conditional on the first record value $T_1=Z_1$, the number of Z_k's until the second record, L_1, has the geometric distribution

$$P(L_1 = \ell \mid T_1) = (1-e^{-T_1})^{\ell-1}e^{-T_1}, \quad \ell = 1,2,\dots.$$

For the second record value $T_2=Z_{1+L_1}$ the lack of memory property of the exponential distribution shows that the excess T_2-T_1 is exponential with mean 1 and independent of T_1 and L_1.

Given the second record value T_2, the waiting time L_2 until the third record has

$$P(L_2 = \ell \mid T_2) = (1 - e^{-T_2})^{\ell-1} e^{-T_2}, \quad \ell = 1, 2, \dots .$$

The excess $T_3 - T_2 = Z_{1+L_1+L_2} - Z_{1+L_1}$ is exponential with mean 1 and independent of the past. Analogously we introduce the random variables L_3, T_4, L_4, T_5 etc.

The sequence $0 < T_1 < T_2 < T_3 < \dots$ is a marked Poisson process Π with intensity 1, where T_k is marked with the geometric random variable L_k. Consequences of this are, that the d-marked T_k's is a Poisson process Π_d with intensity

$$\lambda_d(t) = (1 - e^{-t})^{d-1} e^{-t}, \quad t > 0,$$

and that the processes $\Pi_1, \Pi_2, \dots$ are independent, see [13, Section 5.2]. The total number of points in Π_d is Poisson with mean $\int_0^\infty \lambda_d(t)\, dt = 1/d$.

The double records occur at the points marked 1, that is Π_1. Hence the number of double records M_1 is Poisson with mean 1. Similarly, M_d, the number of points in Π_d, has a Poisson distribution with mean $1/d$. As $\Pi_1, \Pi_2, \dots$ are independent the counts $M_1, M_2, \dots$ are also independent. $\square$

Hahlin [6] derived the distribution of M_1. In a number of papers M_1 and related random variables have been studied by different methods, see for example [4], [7], [8], [12], [14] and [16], and the references in these papers. The random variables $M_1, M_2, \dots$ also occur in connection with random permutations and Ewens Sampling Formula, see [1], [2] and [5]. To our knowledge the embedding approach used above is new. More general cases using embedding are considered in [9] and [11].

3. Strings of records

With probability one infinitely many records occur, but only finitely many double records. Hence most of the records are isolated from each other. How many strings of uninterrupted subsequent records occur? That is, what is the distribution of

$$N = \sum_{k=1}^{\infty} I_k I_{k+1} (1 - I_{k+2}).$$

Clearly N is at most equal to the number of double records, and

$$E(N) = \sum_{k=1}^{\infty} \frac{1}{k} \frac{1}{k+1} \frac{k+1}{k+2} = \frac{3}{4}.$$

As the number of double records is Poisson with mean 1 we have

$$P(N=0)=P(M_1=0)=e^{-1}=0.3679\dots .$$

To illustrate how the embedding above can be used we will calculate $P(N=1)$. We have

$$P(N=1)=\sum_{n=1}^{\infty}P(N=1,M_1=n)=P(M_1=1)+\sum_{n=2}^{\infty}P(N=1,M_1=n).$$

Recall that the points of the double records Π_1 occur with intensity e^{-t} and the points in $\Pi_2\cup\Pi_3\cup\dots$ with intensity $1-e^{-t}$. Conditional on that $n>1$ double records have occurred and at the positions $t_1<t_2<\dots<t_n$, the probability that none of the points in $\Pi_2\cup\Pi_3\cup\dots$ occur in the interval (t_1,t_n) is

$$\exp\left(-\int_{t_1}^{t_n}(1-e^{-t})\,dt\right)=e^{-t_n+t_1}\exp(e^{-t_1}-e^{-t_n}).$$

As M_1 is Poisson with mean 1 we get for $n>1$ the unconditional probability

$$\begin{aligned}&P(N=1,M_1=n)\\&\quad=\int\dots\int I(0<t_1<\dots<t_n)e^{-\sum_{j=1}^{n}t_j}e^{-1}e^{-t_n+t_1}\exp(e^{-t_1}-e^{-t_n})\,dt_1\dots dt_n.\end{aligned}$$

Changing variables $u_j=e^{-t_j}$ gives

$$\begin{aligned}P(N=1,M_1=n)&=\frac{1}{e}\int\dots\int I(0<u_n<\dots<u_1<1)\frac{u_n}{u_1}e^{u_1-u_n}\,du_1\dots du_n\\&=\frac{1}{e}\int_0^1\int_0^1 I(0<u_n<u_1<1)\frac{(u_1-u_n)^{n-2}}{(n-2)!}\frac{u_n}{u_1}e^{u_1-u_n}\,du_1\,du_n\\&=\frac{1}{e}\int_0^1\int_0^1\frac{v^{n-2}}{(n-2)!}(1-v)u^{n-1}e^{uv}\,du\,dv.\end{aligned}$$

Hence

$$\begin{aligned}\sum_{n=2}^{\infty}P(N=1,M_1=n)&=\frac{1}{e}\int_0^1\int_0^1(1-v)ue^{uv}\sum_{n=2}^{\infty}\frac{(uv)^{n-2}}{(n-2)!}\,du\,dv\\&=\frac{1}{e}\int_0^1\int_0^1(1-v)ue^{2uv}\,du\,dv\\&=\frac{1}{e}\sum_{\ell=0}^{\infty}\frac{1}{\ell!}\int_0^1\int_0^1(1-v)u(2uv)^{\ell}\,du\,dv\\&=\frac{1}{e}\sum_{\ell=0}^{\infty}\frac{2^{\ell}}{\ell!}\left(\frac{1}{\ell+1}-\frac{1}{\ell+2}\right)\frac{1}{\ell+2}.\end{aligned}$$

As $P(M_1\!=\!1)\!=\!e^{-1}$ we have shown that

$$P(N=1)=\frac{1}{e}\left(1+\sum_{\ell=0}^{\infty}\frac{2^{\ell}}{(\ell+2)!(\ell+2)}\right)=0.5227\dots.$$

By similar, but more involved calculations, one can obtain

$$P(N=2)=0.1012\dots, \quad P(N=3)=0.0078\dots \quad \text{and} \quad P(N=4)=0.0003\dots.$$

Further results on consecutive records can be found in [3] and [10].

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. Arratia, R., Barbour, A. D. and Tavaré, S., *Logarithmic Combinatorial Structures*: *A Probabilistic Approach*, European Mathematical Society, Zürich, 2003.
3. Chern, H.-H. and Hwang, H.-K., Limit distribution of the number of consecutive records, *Random Structures Algorithm* **26** (2005), 404–417.
4. Chern, H.-H., Hwang, H.-K. and Yeh, Y.-N., Distribution of the number of consecutive records, *Random Structures Algorithms* **17** (2000), 169–196.
5. Gnedin, A., Coherent random permutations with record statistics, in 2007 *Conference on Analysis of Algorithms*, *AofA* 07, Discrete Math. Theor. Comput. Sci. Proc. **AH**, pp. 147–158, Assoc. Discrete Math. Theor. Comput. Sci., Nancy, France, 2007.
6. Hahlin, L.-O., Double records, Research Report 1995:12, Department of Mathematics, Uppsala University, Uppsala, 1995.
7. Holst, L., Counts of failure strings in certain Bernoulli sequences, *J. Appl. Probab.* **44** (2007), 824–830.
8. Holst, L., The number of two consecutive successes in a Hoppe–Pólya urn, *J. Appl. Probab.* **45** (2008), 901–906.
9. Holst, L., A note on embedding certain Bernoulli sequences in marked Poisson processes, *J. Appl. Probab.* **45** (2008), 1181–1185.
10. Holst, L., On consecutive records in certain Bernoulli sequences, *J. Appl. Probab.* **46** (2009), 1201–1208.
11. Huffer, F., Sethuraman, J. and Sethuraman, S., A study of counts of Bernoulli strings via conditional Poisson processes, *Proc. Amer. Math. Soc.* **137** (2009), 2125–2134.
12. Joffe, A., Marchand, E., Perron, F. and Popadiuk, P., On sums of products of Bernoulli variables and random permutations, *J. Theoret. Probab.* **17** (2004), 285–292.
13. Kingman, J. F. C., *Poisson Processes*, Oxford Studies Probab. **3**, Oxford University Press, Oxford, 1993.
14. Móri, T. F., On the distribution of sums of overlapping products, *Acta Sci. Math.* (*Szeged*) **67** (2001), 833–841.

15. Nevzorov, V. B., *Records: Mathematical Theory*, Stokhastika **4**, Izdatel'stvo FAZIS, Moscow, 2000 (Russian). English transl.: Translations of Mathematical Monographs **194**, Amer. Math. Soc., Providence, RI, 2001.
16. Sethuraman, J. and Sethuraman, S., 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**, pp. 140–152, Institute of Mathematical Statistics, Beachwood, OH, 2004.

Lars Holst
Department of Mathematics
Royal Institute of Technology
SE-100 44 Stockholm
Sweden
lholst@math.kth.se

Received October 7, 2009
published online September 22, 2010