Z. Wahrscheinlichkeitstheorie verw. Gebiete
66, 293-296 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

A Note on Moments of Certain Record Statistics

D. Pfeifer

Institut für Statistik und Wirtschaftsmathematik, Technische Hochschule,
Wüllnerstr. 3, 5100 Aachen

Summary. Asymptotic expansions for the mean and variance of the logarithms of record and inter-record times from i.i.d. random variables are given, refining rough estimations from Rényi's and Neuts' Central Limit Theorem for these record statistics.

1. Introduction

Let $\{X_n; n \geqq 1\}$ be a sequence of i.i.d. random variables with continuous distribution function. The *inter-record times* $\{\Delta_n; n \geqq 0\}$ and *record times* $\{U_n; n \geqq 0\}$ are recursively defined by

$$\Delta_0 = 1, \quad \Delta_{n+1} = \min\{k;\, X_{U_n+k} > X_{U_n}\} \quad \text{with } U_n = \sum_{k=0}^{n} \Delta_k,\ n \geqq 0, \tag{1}$$

which by the continuity assumption above are a.s. well-defined. As has been proved by Rényi [5] and Neuts [3],

$$\frac{\log U_n - n}{\sqrt{n}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1), \qquad \frac{\log \Delta_n - n}{\sqrt{n}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1) \tag{2}$$

using the same normalizing constants for the essentially different sequences $\{U_n; n \geqq 0\}$ and $\{\Delta_n; n \geqq 0\}$. However, (2) suggests that

$$\begin{aligned} E(\log U_n) &\approx E(\log \Delta_n) \approx n \\ \operatorname{Var}(\log U_n) &\approx \operatorname{Var}(\log \Delta_n) \approx n \end{aligned} \tag{3}$$

should hold for large n. In fact, the sequences $\{\log U_n; n \geqq 0\}$ and $\{\log \Delta_n; n \geqq 0\}$ are closely related to the arrival-time sequence $\{T_n;\ n \geqq 1\}$ of a unit-rate Poisson process (cf. Shorrock [7], Resnick [6], Deheuvels [2]). For example, $\{\Delta_n;\ n \geqq 1\}$ is identically distributed with the sequence $\{[Y_n/-\log(1-\exp(-T_n)]+1;\ n \geqq 1\}$ where $\{Y_n;\ n \geqq 1\}$ is a sequence of i.i.d. exponentially distributed random variables with unit mean, independent of the underlying Poisson process, and [.] denotes the integer part. From this, we have by suitable construction (cf. Deheuvels [2])

$$\log \Delta_n = \log Y_n + T_n + o(1) \quad \text{a.s. } (n \to \infty). \tag{4}$$

Since $-\log Y_n$ follows a doubly-exponential distribution we have

$$\begin{aligned} E(\log \Delta_n) &= n - C + o(1) \\ \operatorname{Var}(\log \Delta_n) &= n + \frac{\pi^2}{6} + o(1) \quad (n \to \infty) \end{aligned} \tag{5}$$

where C denotes Euler's constant (cf. Pfeifer [4]).

In the present note, we will establish more elaborate expressions for these moments, from which the different growth behaviour of $\{U_n;\ n \geqq 0\}$ and $\{\Delta_n;\ n \geqq 0\}$ will become more apparent.

2. Main Result

Theorem. *For* $n \to \infty$,

$$E(\log \Delta_n) = n - C + \mathcal{O}\left(\frac{n}{2^n}\right), \qquad \operatorname{Var}(\log \Delta_n) = n + \frac{\pi^2}{6} + \mathcal{O}\left(\frac{n^2}{2^n}\right), \tag{6}$$

$$E(\log U_n) = n + 1 - C + \mathcal{O}\left(\frac{n^2}{2^n}\right), \quad \operatorname{Var}(\log U_n) = n + 1 - \frac{\pi^2}{6} + \mathcal{O}\left(\frac{n^3}{2^n}\right). \tag{7}$$

Proof. While (6) was proved in [4] approximating the log function by means of partial sums of the harmonic series (from which the error term essentially is coming up) we shall for the proof of (7) make use of William's [9] representation of record times (cf. also Westcott [8]). For this purpose, let $\{\xi_n;\ n \geqq 1\}$ be a sequence of i.i.d. random variables following an exponential distribution with unit mean. Define random variables $\{V_n;\ n \geqq 0\}$ and $\{D_n;\ n \geqq 0\}$ by

$$\begin{aligned} V_0 = D_0 = 1, \quad & V_n = [V_{n-1} e^{\xi_n}] + 1, \\ D_n = V_n - V_{n-1}, \quad & n \geqq 1. \end{aligned} \tag{8}$$

Then

$$\{(V_n, D_n);\ n \geqq 0\} \overset{\mathcal{D}}{=} \{(U_n, \Delta_n);\ n \geqq 0\}.$$

Since by (8),

$$\log(e^{\xi_n} - 1) < \log D_n - \log V_{n-1} \leqq \log\left(e^{\xi_n} - 1 + \frac{1}{V_{n-1}}\right), \quad n \geqq 1 \tag{9}$$

and

$$E\{\log(e^{\xi_n} - 1)\} = 0, \qquad E\left\{\log\left(e^{\xi_n} - 1 + \frac{1}{V_{n-1}}\right)\right\} = E\left(\frac{\log V_{n-1}}{V_{n-1} - 1}\right)$$

by the independence of ξ_n and V_{n-1}, we have

$$E(\log \Delta_n) - E\left(\frac{\log V_{n-1}}{V_{n-1} - 1}\right) \leqq E(\log U_{n-1}) \leqq E(\log \Delta_n). \tag{10}$$

But since $V_{n-1} \geqq D_{n-1}+1$, $E\left(\frac{\log V_{n-1}}{V_{n-1}-1}\right)$ is of no higher order than $E\left(\frac{\log \Delta_{n-1}}{\Delta_{n-1}}\right) = \mathcal{O}\left(\frac{n^2}{2^n}\right)$ $(n \to \infty)$ which was proved in [4].

Hence the first part of (7) follows from (6). For the second part, multiplying (9) with $\log D_n + \log V_{n-1}$, we obtain after some straightforward calculations similar to those above, using the independence of ξ_n and V_{n-1} and some techniques of [4],

$$E(\log^2 U_{n-1}) = E(\log^2 \Delta_n) - \frac{\pi^2}{3} + \mathcal{O}\left(\frac{n^3}{2^n}\right) \quad (n \to \infty). \tag{11}$$

From this and (6) the second part of (7) is obvious.

It should be noted that the error term for $E(\log \Delta_n)$, $\mathcal{O}\left(\frac{n}{2^n}\right)$, is best possible which e.g. follows from the refined representation of inter-record times given in Deheuvels [1] or by a further series expansion of (2.3) in [4]. On the other hand, an improvement for the error term of $E(\log U_n)$ could be achieved using a similar approach as in [4], giving

$$n+1-C-\frac{1}{2^n} \leqq E(\log U_n) \leqq n+1-C \tag{12}$$

by means of the Markov property of $\{U_n;\, n \geqq 0\}$.

Unfortunately, the latter approach does not yield an appropriate error bound for $\operatorname{Var}(\log U_n)$ such as in (7).

Acknowledgement. I should like to thank P. Deheuvels very much for some fruitful discussions and for pointing out his paper [1] to me which lead to an improvement of this note.

References

1. Deheuvels, P.: Strong approximation in extreme value theory and applications. Colloquia Mathematica Societatis Janos Bolyai 36, Limit Theorems in Probability and Statistics, Veszprém (Hungary), 1982. [To be published. Amsterdam: North Holland]
2. Deheuvels, P.: The complete characterization of the upper and lower class of the record and inter-record times of an i.i.d. sequence (1984). [To appear]
3. Neuts, M.F.: Waiting times between record observations. J. Appl. Probability **4**, 206–208 (1967)
4. Pfeifer, D.: Asymptotic expansions for the mean and variance of logarithmic inter-record times. Methods of Operations Research **39**, 113–121 (1981)
5. Rényi, A.: Théorie des éléments saillants d'une suite d'observations. Colloq. on Combinatorial Methods in Probability Theory, 104–115, Mathematisk Institut, Aarhus Universitet, Denmark (1962)
6. Resnick, S.I.: Extremal processes and record value times. J. Appl. Probability **10**, 864–868 (1973)
7. Shorrock, R.W.: A limit theorem for inter-record times. J. Appl. Probability **9**, 219–223 (1972)
8. Westcott, M.: A note on record times. J. Appl. Probability **14**, 637–639 (1977)
9. Williams, D.: On Rényi's 'record' problem and Engel's series. Bull. London Math. Soc. **5**, 235–237 (1973)

Received December 10, 1982; in revised form February 20, 1984

Note Added in Proof

The results above suggest that the a.s. convergence in (4) also might be of exponential order; this can in fact be proved by an appropriate Borel-Cantelli argument, giving a rate of convergence at least as good as

$$\log \Delta_n = \log Y_n + T_n + o\left(\frac{n^\beta}{2^n}\right) \text{ a.s. } (n \to \infty) \tag{13}$$

for every $\beta > 1$.

Addendum

Here we give details for the arguments outlined on p. 295 to prove relation (12).

By the Markov property of record times, we have

$$P\left(U_{n+1}=k|U_n=j\right)=\frac{j}{k(k-1),}\ \ k>j\in\mathbb{N}. \tag{14}$$

Let $a_{jk}:=\begin{cases}\dfrac{j}{k(k-1)}, & k>j\in\mathbb{N}\\ 0, & \text{otherwise}\end{cases}$ and $S(n):=\sum_{i=1}^{n}\frac{1}{i}.$ It follows that

$$E\left(S\left(U_{n+1}\right)|U_n=j\right)=\sum_{k=1}^{\infty}a_{jk}\sum_{i=1}^{k}\frac{1}{i}=\sum\sum_{1\le i\le k}\frac{a_{jk}}{i}=\sum_{i=1}^{\infty}\frac{1}{i}\sum_{k=i}^{\infty}a_{jk}=\sum_{i=1}^{j}\frac{1}{i}\sum_{k=i}^{\infty}a_{jk}+\sum_{i=j+1}^{\infty}\frac{1}{i}\sum_{k=i}^{\infty}\frac{j}{k(k-1)}$$

$$=\sum_{i=1}^{j}\frac{1}{i}\sum_{k=j+1}^{\infty}\frac{j}{k(k-1)}+\sum_{i=j+1}^{\infty}\frac{j}{i(i-1)}=\sum_{i=1}^{j}\frac{1}{i}+1, \tag{15}$$

hence

$$E\left(S\left(U_{n+1}\right)|U_n\right)=S\left(U_n\right)+1\ \text{ for } n\ge 0. \tag{16}$$

By induction, this yields

$$E\left(S\left(U_n\right)\right)=E\left[E\left(S\left(U_n\right)|U_{n-1}\right)\right]=E\left(S\left(U_{n-1}\right)\right)+1=\ldots=E\left(S\left(U_0\right)\right)+n=n+1 \text{ for } n\ge 0. \tag{17}$$

Now since

$$S(k)-\frac{1}{k}-C\le\log k\le S(k)-\mathrm{C}\ \text{ for } k\in\mathbb{N} \tag{18}$$

as was proved in [4], we finally get

$$n+1-C-E\left(\frac{1}{U_n}\right)=E\left(S\left(U_n\right)\right)-C-E\left(\frac{1}{U_n}\right)\le E\left[\log\left(U_n\right)\right]\le n+1-C\ \ \text{ for } n\ge 0. \tag{19}$$

It remains to prove that

$$E\left(\frac{1}{U_n}\right)\le\frac{1}{2^n}\ \text{ for } n\ge 0. \tag{20}$$

But this follows from (8), observing that $Z_n=e^{-\xi_n}$ is uniformly distributed over $(0,1)$ so that $\frac{1}{V_n}\le\frac{Z_n}{V_{n-1}}$ and hence

$$E\left(\frac{1}{V_n}\middle|V_{n-1}=j\right)\le\frac{E\left(Z_n\right)}{j}=\frac{1}{2j}\ \text{ or }\ E\left(\frac{1}{U_n}\middle|U_{n-1}\right)\le=\frac{1}{2U_{n-1}}\ \text{ for } n\in\mathbb{N} \tag{21}$$

which, by induction, implies (20).