SONDERDRUCK AUS

OPERATIONS RESEARCH VERFAHREN
METHODS OF OPERATIONS RESEARCH

XXXIX

ASYMPTOTIC EXPANSIONS FOR THE MEAN AND VARIANCE OF LOGARITHMIC INTER-RECORD TIMES

Dietmar Pfeifer, Aachen

Abstract

For the inter-record times $\{\Delta_n\}_{n\in\mathbb{N}}$ of a sequence of i.i.d. random variables with continuous cumulative distribution function it is shown that

$$E(\ln\Delta_n) = n - C + \mathcal{O}(\frac{n}{2^n}),\ V(\ln\Delta_n) = n + \frac{\pi^2}{6} + \mathcal{O}(\frac{n^2}{2^n})\ (n\to\infty)$$

with C denoting Euler's constant. This gives rise to an improved approximation formula for the distribution of Δ_n (using the asymptotic normality of $\ln\Delta_n$), especially for n being small.

1. Introduction

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables (r.v.'s) defined on a probability space $(\Omega,\mathcal{A},P)$ with continuous cumulative distribution function (c.d.f.) F. The sequence $\{\Delta_n\}_{n\in\mathbb{N}}$ of inter-record times is recursively defined by

$$(1.1)\qquad \Delta_0 = 1,\ \Delta_{n+1} = \min\{k\in\mathbb{N} \mid X_{U_n+k} > X_{U_n}\}$$

$$\text{with } U_n = \sum_{k=0}^{n}\Delta_k \qquad (n\in\mathbb{Z}^+).$$

In order to be well defined let $\min(\emptyset) = X_\infty = \infty$. $\{U_n\}_{n\in\mathbb{N}}$ is the sequence of record times and $\{X_{U_n}\}_{n\in\mathbb{Z}^+}$ the sequence of record values. By the assumption of continuity we have $\Delta_n < \infty$

a.s. for all n, hence the sequence of record values is infinite a.s. (Shorrock [6]). It is also known that the conditional distribution of Δ_n given $X_{U_{n-1}}$ is geometric with

$$(1.2) \qquad P(\Delta_n = k \mid X_{U_{n-1}}) = \{1 - F(X_{U_{n-1}})\} F^{k-1}(X_{U_{n-1}}) \text{ a.s.} \qquad (n,k \in \mathbb{N}).$$

The distribution of X_{U_n} is of Erlang type (Karlin [4]) with

$$(1.3) \qquad P(X_{U_n} \leq t) = \int_0^{R(t)} \frac{s^n}{n!} e^{-s} ds = 1 - e^{-R(t)} \sum_{k=0}^{n} \frac{R^k(t)}{k!}$$

where $R(t) = -\ln(1-F(t))$ with $\ln(0) = -\infty$ $(t \in \mathbb{R})$. (1.2) and (1.3) imply that the distribution of Δ_n is independent of F, thus F may henceforth assumed to be the c.d.f. of an exponentially distributed r.v. with unit mean. In this case we have $R(t) = t$ $(t \geq 0)$, hence X_{U_n} is gamma - distributed with mean $n+1$ $(n \in \mathbb{Z}^+)$ (Neuts [3]).

2. Main Results

2.1 *Theorem.* Let $S_1(k) = \sum_{j=1}^{k-1} \frac{1}{j}$, $S_2(k) = \sum_{j=1}^{k-1} \frac{1}{j^2}$ $(k \in \mathbb{N})$.

Then we have

$$E(S_1(\Delta_n)) = n, \quad V(S_1(\Delta_n)) = n + E(S_2(\Delta_n))$$

$$\text{with } \frac{\pi^2}{6} - 2\,E\left(\frac{1}{\Delta_n}\right) \leq E(S_2(\Delta_n)) \leq \frac{\pi^2}{6} \qquad (n \in \mathbb{N}).$$

Proof: Let $Q_n = P^{X_{U_{n-1}}}$, $n \in \mathbb{N}$. Then by (1.2) and (1.3),

$$E(S_1(\Delta_n)) = \sum_{k=1}^{\infty} S_1(k) P(\Delta_n = k) = \int \sum_{k=1}^{\infty} S_1(k)\{F^{k-1}(t) - F^k(t)\}\, Q_n(dt)$$

$$= \int \sum_{k=1}^{\infty} \{S_1(k+1) - S_1(k)\} F^k(t) Q_n(dt)$$

$$= \int \sum_{k=1}^{\infty} \frac{1}{k} F^k(t) Q_n(dt) = \int -\ln(1-F(t)) Q_n(dt)$$

$$= \int R(t) Q_n(dt) = E(X_{U_{n-1}}) = n.$$

Correspondingly,

$$E(S_1^2(\Delta_n)) = \sum_{k=1}^{\infty} S_1^2(k) P(\Delta_n = k) = \int \sum_{k=1}^{\infty} \{S_1^2(k+1) - S_1^2(k)\} F^k(t) Q_n(dt)$$

$$= \int \sum_{k=1}^{\infty} \frac{1}{k} \{S_1(k+1) + S_1(k)\} F^k(t) Q_n(dt)$$

$$= 2 \int \sum_{k=1}^{\infty} \frac{1}{k} S_1(k) F^k(t) Q_n(dt) + \int \sum_{k=1}^{\infty} \frac{1}{k^2} F^k(t) Q_n(dt)$$

$$= \int \sum_{k=2}^{\infty} \frac{1}{k} \left(\sum_{j=1}^{k-1} \frac{1}{j} + \frac{1}{k-j} \right) F^k(t) Q_n(dt) + \int \sum_{k=1}^{\infty} \{S_2(k+1) - S_2(k)\} F^k(t) Q_n(dt)$$

$$= \int \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} \frac{1}{j(k-j)} F^k(t) Q_n(dt) + \int \sum_{k=1}^{\infty} S_2(k) \{F^{k-1}(t) - F^k(t)\} Q_n(dt)$$

$$= \int \left\{ \sum_{k=}^{\infty} \frac{1}{k} F^k(t) \right\}^2 Q_n(dt) + \sum_{k=1}^{\infty} S_2(k) P(\Delta_n = k)$$

$$= \int \ln^2(1-F(t)) Q_n(dt) + E(S_2(\Delta_n))$$

$$= \int R^2(t) Q_n(dt) + E(S_2(\Delta_n)) = E\left(X_{U_{n-1}}^{\ 2}\right) + E(S_2(\Delta_n)),$$

hence

$$V(S_1(\Delta_n)) = V(X_{U_{n-1}}) + E(S_2(\Delta_n)) = n + E(S_2(\Delta_n)).$$

Since $\sum_{j=k}^{\infty} \frac{1}{j^2} \leq 2\sum_{j=k}^{\infty} \frac{1}{j(j+1)} = 2\sum_{j=k}^{\infty}\left(\frac{1}{j} - \frac{1}{j+1}\right) = \frac{2}{k}$ $(k \in \mathbb{N})$ and $\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}$ we have

$$\frac{\pi^2}{6} - 2E(\frac{1}{\Delta_n}) \leq \frac{\pi^2}{6} - E\sum_{j=\Delta_n}^{\infty} \frac{1}{j^2} = E(S_2(\Delta_n)) \leq \frac{\pi^2}{6}. \square$$

The indicated result now follows from the close relationship between $S_1(\Delta_n)$ and $\ln\Delta_n$ which is given by the following result:

2.2 *Lemma.* For every $k \in \mathbb{N}$, we have

$$(2.1) \qquad S_1(k) \leq \ln k + C \leq S_1(k) + \frac{1}{k}$$

$$(2.2) \qquad S_1^2(k) \leq (\ln k + C)^2 \leq S_1^2(k) + \frac{2}{k}\ln k + \frac{3}{k}$$

with $C = 0.577216...$ denoting Euler's constant.

Proof: It is known that for $k \in \mathbb{N}$

$$(2.3) \qquad \ln k + C = S_1(k) + \frac{1}{2k} - \int_{k-1}^{\infty} \frac{t-\mathrm{Int}(t)-1/2}{(1+t)^2}\,dt$$

with Int(t) denoting the integer part of t (Erwe [2], p. 49). Since

$$\left| \int_{k-1}^{\infty} \frac{t-\mathrm{Int}(t)-1/2}{(1+t)^2}\,dt \right| \leq \frac{1}{2}\int_{k-1}^{\infty} \frac{1}{(1+t)^2}\,dt = \frac{1}{2k}$$

(2.1) is proved. (2.2) now follows by taking squares on both sides of (2.1) and applying (2.1) again. □

2.3 *Theorem.* For the mean and variance of $\ln\Delta_n$ we have

$$E(\ln\Delta_n) = n - C + \mathcal{O}(\frac{n}{2^n})$$

$$V(\ln\Delta_n) = n + \frac{\pi^2}{6} + \mathcal{O}(\frac{n^2}{2^n}) \qquad (n \to \infty).$$

Proof: The following inequalities will be needed:

(2.4) $$\sum_{k=1}^{\infty} \ln k \; t^{k-1} \le \frac{1}{1-t} \ln (\frac{1}{1-t}) \qquad (0 \le t < 1)$$

(2.5) $$\sum_{k=1}^{\infty} \frac{\ln k}{k} t^k \le \frac{1}{2} \ln^2(\frac{1}{1-t}) \qquad (0 \le t < 1)$$

(2.6) $$t \le \frac{t}{1-e^{-t}} \le t + 1 \qquad (t \ge 0)$$

(2.7) $$\frac{n}{2^{n+1}} \le E(\frac{1}{\Delta_n}) \le \frac{n+2}{2^{n+1}} \qquad (n \in \mathbb{N})$$

(2.8) $$E\left(\frac{\ln\Delta_n}{\Delta_n}\right) \le \frac{n(n+3)}{2^{n+3}} \qquad (n \in \mathbb{N})$$

The proof of these is as follows:

i) By (2.3) we have for $0 \le t < 1$

$$\sum_{k=1}^{\infty} \ln k \; t^{k-1} \le \sum_{k=2}^{\infty} \sum_{j=1}^{k-1} \frac{1}{j} t^{k-1} = \sum_{1 \le j < k} \frac{1}{j} t^{k-1}$$

$$= \sum_{j=1}^{\infty} \frac{1}{j} \sum_{k=j+1}^{\infty} t^{k-1} = \frac{1}{1-t} \sum_{j=1}^{\infty} \frac{t^j}{j} = \frac{1}{1-t} \ln(\frac{1}{1-t}).$$

This is (2.4).

ii) Let $f(t) = \sum_{k=1}^{\infty} \frac{\ln k}{k} t^k$ $(0 \le t < 1)$. Then by (2.4),

$$f'(t) = \sum_{k=1}^{\infty} \ln k \; t^{k-1} \le \frac{1}{1-t} \ln(\frac{1}{1-t}) \text{ with } f(0) = 0,$$

hence

$$f(t) = \int_0^t f'(s)ds \leq \int_0^t \frac{1}{1-s} \ln(\frac{1}{1-s})ds = \frac{1}{2} \ln^2(\frac{1}{1-t}).$$

This is (2.5).

iii) (2.6) immediately follows from the well-known inequality

$$e^t \geq t + 1 \qquad (t \geq 0).$$

iv) With the notation of the proof of Theorem 2.1 we have

$$E(\frac{1}{\Delta_n}) = \sum_{k=1}^{\infty} \frac{1}{k}P(\Delta_n=k) = \int \sum_{k=1}^{\infty} \frac{1}{k}(1-F(t))F^{k-1}(t)Q_n(dt)$$

$$= \int_0^{\infty} \frac{t^{n-1}}{(n-1)!} e^{-2t} \sum_{k=1}^{\infty} \frac{1}{k}(1-e^{-t})^{k-1}dt$$

$$= \int_0^{\infty} \frac{t^{n-1}}{(n-1)!} e^{-2t} \frac{t}{1-e^{-t}} dt \qquad (n \in \mathbb{N}).$$

This together with (2.6) implies (2.7).

v) Just as in iv), we have

$$E\left(\frac{\ln\Delta_n}{\Delta_n}\right) = \int_0^{\infty} \frac{t^{n-1}}{(n-1)!} e^{-2t} \sum_{k=1}^{\infty} \frac{\ln k}{k}(1-e^{-t})^{k-1}dt$$

$$\leq \frac{1}{2} \int_0^{\infty} \frac{t^{n-1}}{(n-1)!} e^{-2t} \frac{t^2}{1-e^{-t}} dt \qquad (n \in \mathbb{N})$$

by (2.5). This together with (2.6) implies (2.8).

From Theorem 2.1, Lemma 2.2 and (2.7) now follows

$$(2.9) \qquad n = E(S_1(\Delta_n)) \leq E(\ln\Delta_n) + C \leq E(S_1(\Delta_n)) + E(\frac{1}{\Delta_n})$$

$$\leq n + \frac{n+2}{2^{n+1}} \qquad (n \in \mathbb{N}).$$

This is the first part of Theorem 2.3. Since $V(\ln\Delta_n) = V(\ln\Delta_n + C)$ we have by Lemma 2.2

$$(2.10) \qquad V(S_1(\Delta_n)) - 2E(S_1(\Delta_n))E(\frac{1}{\Delta_n}) - E^2(\frac{1}{\Delta_n})$$

$$\leq V(\ln\Delta_n) \leq V(S_1(\Delta_n)) + 2E\left(\frac{\ln\Delta_n}{\Delta_n}\right) + 3E(\frac{1}{\Delta_n}) \quad (n \in \mathbb{N}).$$

Applying Theorem 2.1 again together with (2.7) and (2.8) the second part of the Theorem now follows from (2.10). □

As Rényi [5] has shown both $\frac{1}{\sqrt{n}}(\ln\Delta_n - n)$ and $\frac{1}{\sqrt{n}}(\ln U_n - n)$ are asymptotically normally distributed for $n \to \infty$. However, since for the record time and inter-record time sequences the same normalizing constants are chosen no suitable approximation of the distributions of Δ_n and U_n seems to be possible thus if n is small (Neuts [3]). As can be seen from the following table approximation of the distribution of Δ_n is very much improved if the asymptotic expressions for mean and variance (without the $\mathcal{O}$-terms) are used as normalizing constants (adapted from Chandler [1]):

n	k	$P(\Delta_n \leq k)$	$\Phi\left(\frac{\ln k - n}{\sqrt{n}}\right)$	$\Phi\left(\frac{\ln k - (n - C)}{\sqrt{n + \frac{\pi^2}{6}}}\right)$
2	1	0,2500	0,0786	0,2281
	2	0,3889	0,1777	0,3512
	3	0,4792	0,2619	0,4326
	4	0,5433	0,3322	0,4924
	5	0,5917	0,3912	0,5389
	10	0,7255	0,5847	0,6775
	20	0,8264	0,7593	0,7950
	50	0,9114	0,9118	0,9039
3	1	0,1250	0,0416	0,1305
	2	0,2126	0,0915	0,2111
	5	0,3755	0,2110	0,3529
	10	0,5147	0,3436	0,4778
	20	0,6455	0,4990	0,6048
	50	0,7839	0,7007	0,7552
	100	0,8582	0,8230	0,8444
4	1	0,0625	0,0228	0,0748
	5	0,2209	0,1160	0,2227
	10	0,3325	0,1980	0,3186
	20	0,4577	0,3078	0,4287
	50	0,6186	0,4825	0,5816
	100	0,7223	0,6189	0,6906
	500	0,8837	0,8659	0,8800
5	1	0,0313	0,0127	0,0431
	5	0,1234	0,0647	0,1376
	10	0,2002	0,1138	0,2054
	20	0,2992	0,1850	0,2899
	50	0,4494	0,3133	0,4215
	100	0,5631	0,4299	0,5282
	500	0,7782	0,7065	0,7565
	1000	0,8426	0,8032	0,8325

(Φ denotes the distribution function of the standard normal distribution)

References

[1] K.N. CHANDLER:
The distribution and frequency of record values, J. Roy. Statist. Soc. Ser. B 14 (1952), 220 - 228.

[2] F. ERWE:
Differential- und Integralrechnung, Bd. 2, Bibliographisches Institut, Mannheim 1972.

[3] M.F. NEUTS:
Waitingtimes between record observations, J. Appl. Prob. 4 (1967), 206 - 208.

[4] S. KARLIN:
A First Course in Stochastic Processes, Ac. Press, New York, 1966, 266 - 268.

[5] A. RENYI:
Théorie des éléments saillants d'une suite d'observations, Colloquium on Combinatorial Methods in Probability Theory (1962), 104 - 115, Mathematisk Institut, Aarhus Universitet, Denmark.

[6] R.W. SHORROCK:
A limit theorem for inter-record times, J. Appl. Prob. 9 (1972), 219 - 223.

Address:
Institut für Statistik und Wirtschaftsmathematik
RWTH Aachen
Wüllnerstr. 3
D-5100 Aachen

received: 8o-o7-o7