Statistics and Probability Letters 99 (2015) 185–191

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

# Asymptotic independence of three statistics of maximal segmental scores

Aleksandar Mijatović, Martijn Pistorius *

*Department of Mathematics, Imperial College London, United Kingdom*

ARTICLE INFO

*Article history:*
Received 25 January 2014
Accepted 15 January 2015
Available online 28 January 2015

*MSC:*
60G51
60F05
60G17

*Keywords:*
Maximal segmental score
Asymptotic independence
Asymptotic overshoot
Random walk

ABSTRACT

Let $\xi_1, \xi_2, \ldots$ be an iid sequence with negative mean. The $(m, n)$-segment is the subsequence $\xi_{m+1}, \ldots, \xi_n$ and its *score* is given by $\max\{\sum_{m+1}^{n} \xi_i, 0\}$. Let $R_n$ be the largest score of any segment ending at time $n$, $R_n^*$ the largest score of any segment in the sequence $\xi_1, \ldots, \xi_n$, and $O_x$ the overshoot of the score over a level $x$ at the first epoch the score of such a size arises. We show that, under the Cramér assumption on $\xi_1$, asymptotic independence of the statistics $R_n$, $R_n^* - y$ and $O_{x+y}$ holds as $\min\{n, y, x\} \to \infty$. Furthermore, we establish a novel Spitzer-type identity characterising the limit law $O_\infty$ in terms of the laws of $(1, n)$-scores. As corollary we obtain: (1) a novel factorisation of the exponential distribution as a convolution of $O_\infty$ and the stationary distribution of $R$; (2) if $y = \gamma^{-1} \log n$ (where $\gamma$ is the Cramér coefficient), our results, together with the classical theorem of Iglehart (1972), yield the existence and explicit form of the joint weak limit of $(R_n, R_n^* - y, O_{x+y})$.


## 1. Introduction and the main result

Consider a sequence of iid random variables $\{\xi_i\}_{i\in\mathbb{N}}$ with negative mean and denote by $S = \{S_n\}_{n\in\mathbb{N}^*}$ the random walk corresponding to $\{\xi_i\}$: $S_0 \doteq 0$ and $S_n \doteq \sum_{i=1}^{n} \xi_i$. For any $m < n$ with $n \in \mathbb{N}$, $m \in \mathbb{N}^* \doteq \mathbb{N} \cup \{0\}$, the *segmental score* of the $(m, n)$-segment $\{\xi_i\}_{i=m+1}^{n}$ of $\{\xi_i\}$ is given by the maximum of the sum of the elements in the segment and zero (as usual we denote $x^+ \doteq \max\{x, 0\}$, $x \in \mathbb{R}$):

$$\left(\sum_{i=1+m}^{n} \xi_i\right)^+ = (S_n - S_m)^+.$$

The notion of segmental scores arises naturally in several areas of applied probability and statistics. For their application in the study of DNA sequences see *e.g.* Avery and Henderson (1999) and Karlin and Dembo (1992). Segmental scores also play an important role in sequential change point detection problems of mathematical statistics (*e.g.* CUSUM test), see Siegmund and Venkatraman (1995), Moustakides (2004) and Shiryaev (1996). Moreover, in sequential analysis in the context of abortion epidemiology, the maximal segmental score is proposed in Levin and Kline (1985) as a test statistic for the detection of a *one-sided epidemic alternative* for the increase in the mean of a sequence of independent random variables (see also Commenges et al., 1986 for related applications of the epidemic alternative in experimental neurophysiology). For the role of the segmental scores in queueing theory see Asmussen (1982, 2003) and Iglehart (1972). It is of interest in all of these applications to quantify the fluctuations of the segmental scores. In the recent paper Mikosch and Rackauskas (2010)

* Corresponding author.
*E-mail addresses:* a.mijatovic@imperial.ac.uk (A. Mijatović), m.pistorius@imperial.ac.uk (M. Pistorius).

http://dx.doi.org/10.1016/j.spl.2015.01.015
0167-7152/

(see also Mikosch and Moser, 2013) this problem is studied under heavy-tailed step-size distributions, where an appropriate scaling of segmental scores is necessary for the analysis. In the case of an exponentially thin positive tail, *i.e.* under the Cramér assumption, no scaling is required and the asymptotics of the fluctuations of the segmental scores can be analysed directly, which is the aim of the present paper.

Two natural statistics measuring the fluctuations of the segmental scores are $R_n$, the largest score of any $(m, n)$-segment (*i.e.* of any segment ending at time $n$), and $R^*_n$, the largest score of any of the segment in $\{\xi_i\}_{i=1}^n$ (i.e. the largest score that has arisen up to time $n$). More precisely, for $n \in \mathbb{N}$, we have

$$R_n \doteq \max_{m\in\{0,\dots,n-1\}} (S_n - S_m)^+ \quad \text{and} \quad R^*_n \doteq \max_{m,k\in\{0,\dots,n\},m<k} (S_k - S_m)^+.$$

A third statistic quantifying the fluctuations of segmental scores is the first segmental score larger than $x > 0$, which is given by $R_{H(x)}$ with $H(x)$ the first time an increment of the random walk larger than $x$ occurs:

$$H(x) \doteq \min\{k \in \mathbb{N} : \exists\, m < k \text{ such that } S_k - S_m > x\}.$$

The main contribution of this paper is to give sufficient conditions for the three statistics

$$R_n, \qquad Q_{n,x} \doteq R^*_n - x, \qquad O_x \doteq R_{H(x)} - x,$$

of the maximal increments of the walk $S$ to be asymptotically independent in the sense that the joint CDF is asymptotically equal to the product of the marginal CDFs of the statistics:

**Definition.** A family of random vectors $\{(U^1_z, \dots, U^d_z)\}_{z\in\mathcal{Z}}$ on a given probability space, indexed by $z \in \mathcal{Z} \subset [0, \infty)^l$, $d, l \in \mathbb{N}$, is *asymptotically independent* if the joint CDF is asymptotically equal to a product of the CDFs of the components: i.e. for any $a_i \in (-\infty, \infty]$, $i = 1, \dots, d$, it holds

$$P(U^1_z \le a_1, \dots, U^d_z \le a_d) = \prod_{i=1}^d P(U^i_z \le a_i) + o(1) \quad \text{as } \min\{z_1, \dots, z_l\} \to \infty.$$

Our result states that the asymptotic independence of the three statistics above essentially holds under the Cramér assumption on the step-size distribution (which in particular implies $E[\xi_1] < 0$):

**Assumption 1.** The distribution of $\xi_1$ has finite mean, is non-lattice and satisfies *Cramér's condition*, i.e. $E[e^{\gamma\xi_1}] = 1$ for some $\gamma \in (0, \infty)$, and $E[|\xi_1| e^{\gamma\xi_1}]$ is finite.

**Theorem 1.** *Under Assumption 1, the triplet $\{(R_n, Q_{n,y}, O_{y+x})\}_{n\in\mathbb{N},x,y\in\mathbb{R}_+}$ is asymptotically independent, where $\mathbb{R}_+ \doteq [0, \infty)$. Furthermore, the following limit in distribution holds: $O_x \xrightarrow{\mathcal{D}} O_\infty$ as $x \to \infty$, where $O_\infty$ is a non-negative distribution with the characteristic function*

$$E[e^{i\theta O_\infty}] = \frac{\gamma}{\gamma - i\theta} \cdot \exp\left\{\sum_{n=1}^\infty \frac{1}{n}\left(1 - E\left[e^{i\theta S_n^+}\right]\right)\right\}, \quad \textit{for all } \theta \in \mathbb{R}. \tag{1.1}$$

**Remarks.** (i) A classical time-reversal argument implies that $R_n$ and $\max_{m\in\{0,\dots,n\}} S_m$ have the same law for every $n \in \mathbb{N}$. Hence $R_n$ converges in distribution, as $n \uparrow \infty$, to $S^*_\infty \doteq \sup_{n\in\mathbb{N}} S_n$, which is finite (as $E[\xi_1] < 0$ by Assumption 1) and follows a distribution characterised by Spitzer's identity (see Asmussen, 2003, p. 230).

(ii) Note that Spitzer's identity (Asmussen, 2003, p. 230) and a time-reversal argument imply that the second factor in (1.1) is equal to $1/E[e^{i\theta R_\infty}]$. The asymptotic independence of Theorem 1 therefore yields the joint law of the weak limit $(R_\infty, O_\infty)$. In particular, the limit law $R_\infty + O_\infty$ of the sum $R_n + O_x$, as $\min\{x, n\} \to \infty$, is characterised by the identity

$$E\left[e^{i\theta(R_\infty+O_\infty)}\right] = \frac{\gamma}{\gamma - i\theta}, \quad \forall\theta \in \mathbb{R}, \text{ and hence } \gamma \cdot (R_\infty + O_\infty) \sim \text{Exp}(1).$$

This establishes a novel factorisation of the exponential distribution Exp(1) as the convolution of the distribution of the asymptotic overshoot and the stationary distribution of a reflected random walk with step-size distribution satisfying Assumption 1. Note further that, unlike in the Wiener–Hopf factorisation, here the supports of the factorising random variables are in general not disjoint.

(iii) Note that $Q_{n,x}$ does not admit a non-degenerate weak limit along any sequence $(n, x)$ with $\min\{n, x\} \to \infty$. A sufficient condition for the weak convergence of the statistic $Q_{n,x}$ is given in Iglehart (Iglehart, 1972, Thm. 2): if $x(n) = \gamma^{-1}\log(Kn)$, for a certain positive constant $K$, then $\gamma\, Q_{n,x(n)}$ converges weakly to a Gumbel distribution as $n \uparrow \infty$.

(iv) The main technical fact established in this paper is that asymptotically, as $\min\{x, y, n\} \to \infty$, the probability that $R$ crosses the level $x + y$ for the first time during the excursion of $R$ away from 0 straddling time $n$ vanishes (see Section 2.2). This fact, in conjunction with the independence of distinct excursions, essentially implies the asymptotic independence in Theorem 1.