Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 124 (2014) 2824–2867

stochastic processes and their applications

www.elsevier.com/locate/spa

Limiting distribution for the maximal standardized increment of a random walk

Zakhar Kabluchko[a,*], Yizao Wang[b]

[a] *Institute of Stochastics, Ulm University, Helmholtzstr. 18, 89069 Ulm, Germany*
[b] *Department of Mathematical Sciences, University of Cincinnati, 4302 French Hall, 2815 Commons Way, Cincinnati, OH, 45221-0025, United States*

Received 30 September 2013; received in revised form 10 March 2014; accepted 31 March 2014
Available online 15 April 2014

Abstract

Let $X_1, X_2, \dots$ be independent identically distributed (i.i.d.) random variables with $\mathbb{E}X_k = 0$, $\mathrm{Var}X_k = 1$. Suppose that $\varphi(t) := \log \mathbb{E}e^{tX_k} < \infty$ for all $t > -\sigma_0$ and some $\sigma_0 > 0$. Let $S_k = X_1 + \cdots + X_k$ and $S_0 = 0$. We are interested in the limiting distribution of the *multiscale scan statistic*

$$\mathbf{M}_n = \max_{0 \le i < j \le n} \frac{S_j - S_i}{\sqrt{j - i}}.$$

We prove that for an appropriate normalizing sequence a_n, the random variable $\mathbf{M}_n^2 - a_n$ converges to the Gumbel extreme-value law $\exp\{-e^{-cx}\}$. The behavior of $\mathbf{M}_n$ depends strongly on the distribution of the X_k's. We distinguish between four cases. In the *superlogarithmic* case we assume that $\varphi(t) < t^2/2$ for every $t > 0$. In this case, we show that the main contribution to $\mathbf{M}_n$ comes from the intervals (i, j) having length $l := j - i$ of order $a(\log n)^p$, $a > 0$, where $p = q/(q-2)$ and $q \in \{3, 4, \dots\}$ is the order of the first non-vanishing cumulant of X_1 (not counting the variance). In the *logarithmic* case we assume that the function $\psi(t) := 2\varphi(t)/t^2$ attains its maximum $m_* > 1$ at some unique point $t = t_* \in (0, \infty)$. In this case, we show that the main contribution to $\mathbf{M}_n$ comes from the intervals (i, j) of length $d_* \log n + a\sqrt{\log n}$, $a \in \mathbb{R}$, where $d_* = 1/\varphi(t_*) > 0$. In the *sublogarithmic* case we assume that the tail of X_k is heavier than $\exp\{-x^{2-\varepsilon}\}$, for some $\varepsilon > 0$. In this case, the main contribution to $\mathbf{M}_n$ comes from the intervals of length $o(\log n)$ and in fact, under regularity assumptions, from the intervals of length 1. In the remaining, fourth case, the X_k's are *Gaussian*. This case has been studied earlier in the literature. The main contribution

* Corresponding author. Tel.: +49 07315023527.
E-mail addresses: sachar.k@gmx.net, zakhar.kabluchko@uni-ulm.de (Z. Kabluchko), yizao.wang@uc.edu (Y. Wang).

http://dx.doi.org/10.1016/j.spa.2014.03.015

comes from intervals of length $a \log n$, $a > 0$. We argue that our results cover most interesting distributions with light tails. The proofs are based on the precise asymptotic estimates for large and moderate deviation probabilities for sums of i.i.d. random variables due to Cramér, Bahadur, Ranga Rao, Petrov and others, and a careful extreme value analysis of the random field of standardized increments by the double sum method.

MSC: 60G50; 60G70; 60F10; 60F05

Keywords: Extreme value theory; Increments of random walks; Erdős–Rényi law; Large deviations; Moderate deviations; Multiscale scan statistic; Cramér series; Gumbel distribution; Double sum method; Subgaussian distributions; Change-point detection

1. Introduction and statement of results

1.1. Introduction

Suppose we are given a long sequence of observations. The observations are assumed to be independent identically distributed (i.i.d.) random variables with zero mean and unit variance, except, possibly, for a short interval, where the observations have positive mean. This interval may be interpreted as a signal in an i.i.d. noise. The question is how to decide whether a signal is present and if yes, how to locate it. A natural approach is to build a *multiscale scan statistic*. For every interval we compute the sum of the observations in this interval divided by the square root of the length of the interval. Large values of this normalized sum indicate the presence of a signal. Since no a priori knowledge about the location and length of the interval containing the signal is available, we take the maximum of such normalized sums over all possible intervals of all possible lengths. Scan statistics with windows of *fixed size* have been much studied; see, e.g., [16,17]. A large class of limit theorems dealing with fixed window size are the Erdös–Rényi–Shepp laws; see, e.g., [10–14,6]. The scan statistic we are interested in is built using windows of *all possible sizes*. In order to use this statistic for testing purposes we need to know its asymptotic distribution under the null hypothesis.

We arrive at the following problem. Let $X_1, X_2, \dots$ be i.i.d. non-degenerate random variables with $\mathbb{E}X_k = 0$, $\mathrm{Var}X_k = 1$. Consider a random walk given by $S_k = X_1 + \cdots + X_k$, $k \in \mathbb{N}$, and $S_0 = 0$. For $n \in \mathbb{N}$ define the multiscale scan statistic $\mathbf{M}_n$ by

$$\mathbf{M}_n = \max_{0 \le i < j \le n} \frac{S_j - S_i}{\sqrt{j - i}}. \tag{1}$$

Following results on the asymptotic behavior of $\mathbf{M}_n$ as $n \to \infty$ are known. For random variables with finite exponential moments, Shao [39], confirming and extending a conjecture of Révész [38], proved that

$$\lim_{n \to \infty} \frac{\mathbf{M}_n}{\sqrt{2 \log n}} = \sqrt{m_*} \quad \text{a.s.} \tag{2}$$

Here, $m_* \in [1, \infty]$ is a constant determined explicitly in terms of the distribution of X_1. Shao's proof has been considerably simplified by Steinebach [43]; see also [23] for a multidimensional generalization. This describes the a.s. rate of growth of $\mathbf{M}_n$. But what about the limiting distribution? In the case when $X_1, X_2, \dots$ are i.i.d. standard Gaussian, Siegmund and Venkatraman [40]