Shao's theorem on the maximum of standardized random walk increments for multidimensional arrays

Zakhar Kabluchko, Axel Munk

▶ **To cite this version:**

Zakhar Kabluchko, Axel Munk. Shao's theorem on the maximum of standardized random walk increments for multidimensional arrays. ESAIM: Probability and Statistics, 2009, 13, pp.409-416. 10.1051/ps:2008020 . hal-00520011

HAL Id: hal-00520011
https://hal.science/hal-00520011

Submitted on 22 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESAIM: Probability and Statistics Will be set by the publisher
URL: http://www.emath.fr/ps/

SHAO'S THEOREM ON THE MAXIMUM OF STANDARDIZED RANDOM WALK INCREMENTS FOR MULTIDIMENSIONAL ARRAYS

Zakhar Kabluchko[1] and Axel Munk[1]

Abstract. We generalize a theorem of Shao (1995, Proc. Am. Math. Soc. 123, 575-582) on the almost-sure limiting behavior of the maximum of standardized random walk increments to multidimensional arrays of i.i.d. random variables. The main difficulty is the absence of an appropriate strong approximation result in the multidimensional setting. The multiscale statistic under consideration was used recently for the selection of the regularization parameter in a number of statistical algorithms as well as for the multiscale signal detection.

1991 Mathematics Subject Classification. 60F15.

The dates will be set by the publisher.

1. Introduction

Let $\{X_i : i \in \mathbb{N}\}$ be a sequence of i.i.d. random variables satisfying $\mathbf{E}[X_1] = 0$, $\mathbf{E}[X_1^2] = 1$. Suppose that the logarithmic moment generating function $\varphi(\theta) = \log \mathbf{E}[e^{\theta X_1}]$ is finite for some $\theta > 0$. Define the Cramér-Chernoff and the Erdös-Rényi functions by

$$I(t) = \sup_{\theta \geq 0}(\theta t - \varphi(\theta)),\ t \geq 0 \quad \text{and} \quad \alpha(c) = \sup\{t \geq 0 : I(t) \leq 1/c\},\ c > 0. \tag{1}$$

Let $S_n = X_1 + \ldots + X_n$, $S_0 = 0$ be the random walk generated by $\{X_i\}$ and let

$$L_n = \max_{0 \leq i < j \leq n} \frac{S_j - S_i}{\sqrt{j-i}}. \tag{2}$$

be the maximum of standardized random walk increments.
In the first edition of his book [15] Révész gave upper and lower bounds for the almost-sure limiting behavior of L_n in the case of Bernoulli-distributed summands [15, Th. 14.16] and conjectured that in this case $\lim_{n\to\infty} L_n/\sqrt{2\log n} = 1$ a.s. Establishing a general version of Révész conjecture, Shao [18] proved that

$$\lim_{n\to\infty} L_n/\sqrt{2\log n} = \alpha^* \in [1, \infty] \quad \text{a.s.}, \tag{3}$$

Keywords and phrases: standardized increments, Lévy's continuity modulus, almost sure limit theorem, Erdös-Rényi law, multidimensional i.i.d. array, statistical multiscale parameter selection, scan statistics

[1] Institut für Mathematische Stochastik, Georg-August-Universität Göttingen, Maschmühlenweg 8-10, D-37073 Göttingen, Germany; e-mail: kabluch@math.uni-goettingen.de, munk@math.uni-goettingen.de

© EDP Sciences, SMAI 1999

This provisional PDF is the accepted version. The article should be cited as: ESAIM: PS, doi: 10.1051/ps:2008020

where $\alpha^* = \sup_{c>0} \alpha(c)\sqrt{c/2}$. If X_i are normal or Bernoulli distributed, then $\alpha^* = 1$. Shao showed that α^* is finite if and only if $\mathbf{E}\exp(sX_1^{+2}) < \infty$ for some $s > 0$, where $X_1^+ = \max(X_1, 0)$. It follows from [18, Lemma 1] that there are distributions with $1 < \alpha^* < \infty$.
Shao's proof was considerably simplified by Steinebach [20], who deduced Shao's theorem from the Erdös-Renyi law of large numbers and the following auxiliary result:

$$\lim_{n\to\infty} \max_{\substack{0\le i<j\le n \\ k:=j-i}} \frac{S_j - S_i}{k\alpha(k/\log n)} = 1 \qquad \text{a.s.} \tag{4}$$

Our goal is to extend (3) and (4) to multidimensional arrays of i.i.d. random variables. Both Shao and Steinebach use in their proofs results of Hanson and Russo [11], which are based on the classical strong approximation theorem [13]. It is well known that extending strong approximation theorems to the case of multidimensional indices is non-trivial and in many cases only possible with an approximation rate which is weaker than the rate in the one-dimensional case. For example, the approximation rate given in Theorem 4 of [10] is too weak for our purposes, whereas Theorem 1.2 of [16], which gives a sufficiently strong rate, is true only in two dimensions. We use a completely elementary method which does not rely on strong approximation.
To state our results let $\{X_{\mathbf{n}} : \mathbf{n} = (n_1, \dots, n_d) \in \mathbb{N}^d\}$ be a d-dimensionally indexed array of i.i.d. random variables satisfying the following conditions

i) $\mathbf{E}[X_{\mathbf{n}}] = 0$, $\mathbf{E}[X_{\mathbf{n}}^2] = 1$
ii) the logarithmic moment generating function $\varphi(\theta) = \log \mathbf{E}[e^{\theta X_{\mathbf{n}}}]$ is finite for some $\theta > 0$.

We denote multi-indices by $\mathbf{m} = (m_1, \dots, m_d)$, $\mathbf{n} = (n_1, \dots, n_d)$ etc. If $m_i \le n_i$ for all $i = 1, \dots, d$, we write $\mathbf{m} \le \mathbf{n}$. Define $|\mathbf{n}| = \prod_{i=1}^d n_i$. We say that $\mathbf{n} \to \infty$ if $|\mathbf{n}| \to \infty$.
A set of the form

$$R(\mathbf{m}, \mathbf{n}) = \{m_1, \dots, n_1\} \times \dots \times \{m_d, \dots, n_d\}, \qquad \mathbf{m} \le \mathbf{n},$$

is called *discrete rectangle*. In order to simplify the notation we write $R(\mathbf{n})$ for the discrete rectangle

$$R((1, \dots, 1), \mathbf{n}) = \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_d\}.$$

Let $\mathcal{R} = \{R(\mathbf{m}, \mathbf{n}) \mid \mathbf{m} \le \mathbf{n}, \mathbf{m}, \mathbf{n} \in \mathbb{N}^d\}$ be the collection of all discrete rectangles and $\mathcal{R}(\mathbf{n})$ the collection of all discrete rectangles contained in $R(\mathbf{n})$. For a discrete rectangle R define

$$S_R = \sum_{\mathbf{n}\in R} X_{\mathbf{n}}.$$

The multidimensional analogue the maximum of standardized random walk increments (2) is then defined by

$$L(\mathbf{n}) = \max_{R\in\mathcal{R}(\mathbf{n})} \frac{S_R}{\sqrt{|R|}},$$

where $|A|$ stands for the number of elements of a finite set A.
Our goal is to prove the following theorem which is a higher-dimensional generalization of (3).

Theorem 1.1. *Let $\{X_{\boldsymbol{n}} : \boldsymbol{n} \in \mathbb{N}^d\}$ be a $d-$dimensional array of i.i.d. random variables satisfying conditions i) and ii). Then, with the above notation,*

$$\lim_{\boldsymbol{n}\to\infty} L(\boldsymbol{n})/\sqrt{2\log|\boldsymbol{n}|} = \alpha^* \quad \text{a.s.}$$

The following theorem is a multidimensional analogue of (4).

Theorem 1.2. *Let $\{X_{\boldsymbol{n}} : \boldsymbol{n} \in \mathbb{N}^d\}$ be a $d-$dimensional array of i.i.d. random variables satisfying conditions i) and ii). Then*

$$\lim_{\boldsymbol{n}\to\infty} \max_{R\in\mathcal{R}(\boldsymbol{n})} \frac{S_R}{|R|\alpha\left(\frac{|R|}{\log|\boldsymbol{n}|}\right)} = 1 \quad a.s.$$

Some statistical applications of Theorems 1.1 and 1.2 will be briefly discussed in the next section. In Section 3 we will prove Theorem 1.1. Theorem 1.2 as well as the subsequent Theorem 2.1 can be proved by the same methods. Note finally that all results remain true with the family of discrete rectangles replaced by the family of discrete circles or cubes, the proofs remaining essentially the same.

2. Applications to statistical multiscale analysis

Statistical methods which aim for the simultaneous control of a family of statistics on different scales (here indexed by the rectangles R) are sometimes denoted as statistical multiscale methods. We will briefly discuss some potential applications of the preceding theorems to this situation.

Selection of the stopping index of the EM-algorithm in PET. Recently [1] introduced a multiscale statistic similar to that in Theorem 1.2 for the proper selection of the stopping index of the expectation-maximisation (EM) algorithm in positron emission tomography (PET). It has been known for a long time that proper stopping of this algorithm is essential for a valid image reconstruction, albeit rigorous theoretical foundation of this empirical fact is still missing. In PET from an inhomogeneous Poisson process (photon counts) an image has to be recovered and the EM algorithm (and variants thereof) is one of the most prominent reconstruction methods. In [1] it has been advocated to select the stopping index such that the corresponding residuals (instead of $X_{\mathbf{n}}$) satisfy an analogue to Theorem 1.2. This has been investigated numerically and its superiority to some frequently used methods has been demonstrated.

Selection of the regularization parameter in penalized regression. Theorem 1.1 allows to extend the method of Davies and Kovac [3] for the proper selection of the threshold parameter in nonparametric function estimation to the multidimensional situation. More precisely, we will discuss the particular case of multivariate TV-norm penalization [9], [12]. To this end we consider a d-variate regression model

$$Y(\mathbf{k}/n) = f(\mathbf{k}/n) + \varepsilon_{\mathbf{k}},$$

where $\{\varepsilon_{\mathbf{k}}, \mathbf{k} \le \mathbf{n}\}$ is a rectangular array of i.i.d. random variables satisfying i) and ii), $\mathbf{n} = (n, \dots, n) \in \mathbb{N}^d$, n is a positive integer and $f : [0,1]^d \to \mathbb{R}$ is a function of bounded variation. Let the multivariate estimate $\hat{f}_{n,\lambda}$, $\lambda > 0$ of f be the minimizer of

$$\Delta_n(g) = \sum_{\mathbf{k}\le\mathbf{n}} (Y(\mathbf{k}/n) - g(\mathbf{k}/n))^2 + \lambda \mathrm{TV}(g),\ g \in \mathrm{TV}([0,1]^d),$$

see [9]. Then the analogous to [3] choice of the penalty parameter λ would be such that the residuals $r_{\mathbf{k}}(\lambda) = Y(\mathbf{k}/n) - \hat{f}_{n,\lambda}(\mathbf{k}/n)$ satisfy Theorem 1.1 with $X_{\mathbf{k}}$ replaced by $r_{\mathbf{k}}(\lambda)$.

Multiscale signal detection. We would like to mention a further extension of (3) and (4), which is very natural from a statistical viewpoint. Let F be a probability distribution function which is for simplicity supposed to have finite logarithmic moment generating function $\varphi(\theta) = \log \int e^{\theta x} dF(x)$ for all $\theta \in \mathbb{R}$. Suppose we are given a finite rectangular array of observations $\{X_{\mathbf{k}} : \mathbf{k} \le \mathbf{n}\}$ and would like to test the hypothesis

$$H_0 : \text{the observations } \{X_{\mathbf{k}} : \mathbf{k} \le \mathbf{n}\} \text{ are i.i.d. with distribution } dF$$

against the alternative

$$H_1 : \text{there is } \lambda \in \mathbb{R} \text{ and a discrete rectangle } R \in \mathcal{R}(\mathbf{n}) \text{ such that } \{X_{\mathbf{k}} : \mathbf{k} \in R\} \text{ are independent with distribution } dF \text{ if } \mathbf{k} \notin R \text{ and with distribution } e^{\theta x - \varphi(\theta)} dF(x) \text{ if } \mathbf{k} \in R.$$

Thus, the zero hypothesis H_0 says that the observations $\{X_{\mathbf{k}} : \mathbf{k} \le \mathbf{n}\}$ form an i.i.d. noise, whereas the alternative H_1 says that there is a rectangle where the observations are not distributed accordingly to F but rather to some distribution from the exponential family generated by F, which may be interpreted as a presence of a signal concentrated in the rectangle R. The likelihood ratio test rejects H_0 if the value of the test statistic

$$\max_{R \in \mathcal{R}(\mathbf{n})} |R| I\left(\frac{S_R}{|R|}\right) \tag{5}$$

is large, where $I(t) = \sup_{\theta \in \mathbb{R}}(\theta t - \varphi(\theta))$ denotes the Cramér-Chernoff information function.
The above statistic was introduced in [17] where its large deviations were studied. Similar but different statistics were considered in [5], [6]. The next theorem describes the almost-sure limiting behavior of the statistic (5).

Theorem 2.1. *Under the above assumptions we have*

$$\lim_{\boldsymbol{n} \to \infty} \frac{1}{\log |\boldsymbol{n}|} \max_{R \in \mathcal{R}(\boldsymbol{n})} |R| I\left(\frac{S_R}{|R|}\right) = 1 \quad a.s.$$

The information function I is non-negative and has a unique zero at $\mathbf{E}X_1$, the expectation of X_1. Thus, contrary to the statistics from Theorems 1.1 and 1.2, the statistic (5) from Theorem 2.1 controls both positive and negative deviations of the rectangular sums from their mean. The corresponding one-sided version of Theorem 2.1 may be easily formulated.

3. Proof of the main theorem

In this section we prove Theorem 1.1. We need the following multidimensional generalization of the Erdös-Rényi law of large numbers, see [4], [19].

Theorem 3.1. *Let $c > 0$ and define $c_{\boldsymbol{n}} = [c \log |\boldsymbol{n}|]$ for $\boldsymbol{n} \in \mathbb{N}^d$. Then*

$$\lim_{\boldsymbol{n} \to \infty} \max_{\substack{R \in \mathcal{R}(\boldsymbol{n}) \\ |R| = c_{\boldsymbol{n}}}} \frac{S_R}{c_{\boldsymbol{n}}} = \alpha(c) \quad a.s.$$

The maximum is taken over all discrete rectangles in $\mathcal{R}(\boldsymbol{n})$ containing $c_{\boldsymbol{n}}$ points.

Proof of Theorem 1.1. To prove the lower bound the method of [18] can be used. We have for every $c > 0$

$$\liminf_{\mathbf{n} \to \infty} \frac{1}{\sqrt{2 \log |\mathbf{n}|}} \max_{R \in \mathcal{R}(\mathbf{n})} \frac{S_R}{\sqrt{|R|}} \ge \liminf_{\mathbf{n} \to \infty} \sqrt{\frac{c}{2}} \max_{\substack{R \in \mathcal{R}(\mathbf{n}) \\ |R| = [c \log |\mathbf{n}|]}} \frac{S_R}{|R|},$$

which is a.s. equal to $\sqrt{\frac{c}{2}}\alpha(c)$ by Theorem 3.1. It follows that

$$\liminf_{\mathbf{n} \to \infty} L(\mathbf{n}) / \sqrt{2 \log |\mathbf{n}|} \ge \sup_{c > 0} \sqrt{\frac{c}{2}} \alpha(c) = \alpha^* \quad \text{a.s.}$$

The main difficulty is to prove the upper bound

$$\limsup_{\mathbf{n} \to \infty} \frac{1}{\sqrt{2 \log |\mathbf{n}|}} \max_{R \in \mathcal{R}(\mathbf{n})} \frac{S_R}{\sqrt{|R|}} \le \alpha^* \quad \text{a.s.} \tag{6}$$

We may suppose that α^* is finite since otherwise there is nothing to prove.
Write $\mathcal{R}^*(\mathbf{n})$ for the collection of all discrete rectangles whose top-right corner is $\mathbf{n}$. That is, $\mathcal{R}^*(\mathbf{n})$ consists of all rectangles of the form $R(\mathbf{m},\mathbf{n})$, where $\mathbf{m}\le\mathbf{n}$. We split the set $\mathcal{R}^*(\mathbf{n})$ into "small" and "big" rectangles. A rectangle is small if it contains not more than $k_{\mathbf{n}}$ points, where the multidimensionally indexed sequence $\{k_{\mathbf{n}}\}_{\mathbf{n}\in\mathbb{N}^d}$ is defined by $k_{\mathbf{n}}=[\log|\mathbf{n}|]^r$ for some $r>3$. The upper bound (6) follows via the Borel-Cantelli lemma from Lemmas 3.1 and 3.3 below, which estimate the maximum over small and big rectangles respectively.

Lemma 3.1. *Let $\varepsilon>0$. Then, for some $\delta>0$ and all sufficiently large $\boldsymbol{n}\in\mathbb{N}^d$,*

$$\mathbf{P}\left[\frac{1}{\sqrt{2\log|\boldsymbol{n}|}}\max_{\substack{R\in\mathcal{R}^*(\boldsymbol{n})\\ |R|<k_n}}\frac{S_R}{\sqrt{|R|}}>(1+\varepsilon)\alpha^*\right]<\frac{1}{|\boldsymbol{n}|^{1+\delta}}. \tag{7}$$

Proof. It follows from the Markov inequality that $\mathbf{P}[S_R/|R|>a]\le\exp\{-|R|I(a)\}$ for each $a>0$. Using this as well as the convexity of the function I we obtain

$$\begin{aligned}\mathbf{P}[S_R/\sqrt{|R|}>(1+\varepsilon)\alpha^*\sqrt{2\log|\mathbf{n}|}]&=\mathbf{P}[S_R/|R|>(1+\varepsilon)\alpha^*\sqrt{2\log|\mathbf{n}|/|R|}]\\&\le\exp\{-|R|I((1+\varepsilon)\alpha^*\sqrt{2\log|\mathbf{n}|/|R|})\}\\&\le\exp\{-(1+\varepsilon)|R|I(\alpha^*\sqrt{2\log|\mathbf{n}|/|R|})\}.\end{aligned}$$

It follows from the definition of α^* that $I(\alpha^*\sqrt{2t})\ge t$ for every $t\ge0$ and, consequently,

$$I(\alpha^*\sqrt{2\log|\mathbf{n}|/|R|})\ge\log|\mathbf{n}|/|R|.$$

It follows

$$\mathbf{P}[S_R/\sqrt{|R|}>(1+\varepsilon)\alpha^*\sqrt{2\log|\mathbf{n}|}]\le\exp\{-(1+\varepsilon)|R|\log|\mathbf{n}|/|R|\}=|\mathbf{n}|^{-(1+\varepsilon)}.$$

To finish the proof note that the number of rectangles $R\in\mathcal{R}^*(\mathbf{n})$ satisfying $|R|<k_{\mathbf{n}}$ is certainly smaller than $k_{\mathbf{n}}^d$ and thus the left-hand side of (7) may be estimated from above by $k_{\mathbf{n}}^d|\mathbf{n}|^{-(1+\varepsilon)}$ which is smaller than $|\mathbf{n}|^{-(1+\delta)}$ for $\delta<\varepsilon$ and $\mathbf{n}$ sufficiently large. □

Now we are going to treat the maximum over big rectangles. In the one-dimensional case, both Shao [18] and Steinebach [20] use to this end the results of Hanson-Russo [11]. However, as discussed above, the strong approximation results needed in [11] are not known for $d>2$. Thus, we need a different method not relying on strong approximation. We use the following classical moderate deviations theorem, see e.g. [8, Chapter XVI, §7].

Theorem 3.2. *Let $\{X_i:i\in\mathbb{N}\}$ be i.i.d. random variables satisfying $\mathbf{E}[X_1]=0$, $\mathbf{E}[X_1^2]=1$ and let S_n be the sequence of their partial sums. Suppose that $\mathbf{E}[e^{tX_1}]$ is finite for some $t>0$. Let $\{x(n)\}_{n\in\mathbb{N}}$ be a sequence satisfying $x(n)=o(n^{1/6})$, $n\to\infty$. Then*

$$\mathbf{P}\left[S_n/\sqrt{n}>x(n)\right]\sim\mathbf{P}[N>x(n)],\quad n\to\infty,$$

where N is a standard normal variable.

We also need the following elementary lemma.

Lemma 3.2. *Let $\{X_i:i\in\mathbb{N}\}$ be i.i.d. random variables satisfying $\mathbf{E}[X_1]=0$, $\mathbf{E}[X_1^2]=1$. Then there is constant C depending only on the distribution of X_i such that for every $k,n\in\mathbb{N}$, $k\le n$ and every $x>0$ we have*

$$\mathbf{P}[S_k>x]\le C\mathbf{P}[S_n>x].$$

Proof. Since $\lim_{i\to\infty}\mathbf{P}[S_i \geq 0] = 1/2$ by the central limit theorem, there is $c_1 > 0$ such that $\mathbf{P}[S_i \geq 0] > c_1$ for every $i \in \mathbb{N}$. It follows that

$$\mathbf{P}[S_n > x] \geq \mathbf{P}[S_k > x]\mathbf{P}[S_n - S_k \geq 0] = \mathbf{P}[S_k > x]\mathbf{P}[S_{n-k} \geq 0] > c_1\mathbf{P}[S_k > x]$$

and the statement of the lemma follows with $C = 1/c_1$. □

Now we are ready to estimate the maximum over big rectangles.

Lemma 3.3. *Let $\varepsilon > 0$ be given. Then, for some $\delta > 0$ and all sufficiently large $\boldsymbol{n} \in \mathbb{N}^d$,*

$$\mathbf{P}\left[\frac{1}{\sqrt{2\log|\boldsymbol{n}|}}\max_{\substack{R\in\mathcal{R}^*(\boldsymbol{n})\\ |R|>k_{\boldsymbol{n}}}}\frac{S_R}{\sqrt{|R|}} > 1+\varepsilon\right] < \frac{1}{|\boldsymbol{n}|^{1+\delta}}.$$

Proof. The statement of the lemma may be written in the following equivalent form

$$\mathbf{P}\left[\frac{1}{\sqrt{2\log|\mathbf{n}|}}\max_{\substack{\mathbf{m}\leq\mathbf{n}\\ |\mathbf{m}|>k_{\mathbf{n}}}}\frac{S_{R(\mathbf{m})}}{\sqrt{|\mathbf{m}|}} > 1+\varepsilon\right] < \frac{1}{|\mathbf{n}|^{1+\delta}}.$$

Take $\sigma > 1$ so close to 1 that $(\sigma-1)^{-1/2}\frac{\varepsilon}{2} > 2$. We may suppose that $\sigma^k \notin \mathbb{N}, k \in \mathbb{N}$. Let

$$K(\mathbf{n}) = K(\mathbf{n},\sigma) = \{\mathbf{k} = (k_1,\ldots,k_d) \in \mathbb{N}_0^d \quad | \quad \sigma^{k_i} < n_i, i = 1,\ldots,d\}.$$

For $\mathbf{k} \in K(\mathbf{n})$ define discrete rectangles $Q(\mathbf{k}) = Q(k_1,\ldots,k_d)$ and $S(\mathbf{k}) = S(k_1,\ldots,k_d)$ by

$$Q(\mathbf{k}) = \{1,\ldots,\lceil\sigma^{k_1}\rceil\}\times\ldots\times\{1,\ldots,\lceil\sigma^{k_d}\rceil\}$$

and

$$S(\mathbf{k}) = \mathbb{N}^d \cap \left([\sigma^{k_1},\sigma^{k_1+1}]\times\ldots\times[\sigma^{k_d},\sigma^{k_d+1}]\right).$$

For each $\mathbf{m} = (m_1,\ldots,m_d) \leq \mathbf{n}$ one can find uniquely defined $\mathbf{k} = \mathbf{k}(\mathbf{m})$ with $\mathbf{m} \in S(\mathbf{k})$. Let $\mathbf{m}_* = (\lceil\sigma^{k_1}\rceil,\ldots,\lceil\sigma^{k_d}\rceil)$ and $\mathbf{m}^* = (\lfloor\sigma^{k_1+1}\rfloor,\ldots,\lfloor\sigma^{k_d+1}\rfloor)$ be the bottom-left and top-right corners of the rectangle $S(\mathbf{k}(\mathbf{m}))$. Note that $R(\mathbf{m}_*) = Q(\mathbf{k}(\mathbf{m}))$. Define $P(\mathbf{m}) = R(\mathbf{m})\backslash R(\mathbf{m}_*)$. Then $R(\mathbf{m})$ is a disjoint union of $R(\mathbf{m}_*)$ and $P(\mathbf{m})$ and it follows that

$$\frac{S_{R(\mathbf{m})}}{\sqrt{|\mathbf{m}|}} \leq \frac{S_{R(\mathbf{m}_*)}}{\sqrt{|\mathbf{m}_*|}} + \frac{S_{P(\mathbf{m})}}{\sqrt{|\mathbf{m}_*|}}.$$

Thus

$$\max_{\substack{\mathbf{m}\leq\mathbf{n}\\ |\mathbf{m}|>k_{\mathbf{n}}}}\frac{S_{R(\mathbf{m})}}{\sqrt{|\mathbf{m}|}} \leq \max_{\substack{\mathbf{k}\in K(\mathbf{n})\\ |Q(\mathbf{k})|>k_{\mathbf{n}}}}\frac{S_{Q(\mathbf{k})}}{\sqrt{|Q(\mathbf{k})|}} + \max_{\substack{\mathbf{m}\leq\mathbf{n}\\ |\mathbf{m}|>k_{\mathbf{n}}}}\frac{S_{P(\mathbf{m})}}{\sqrt{|\mathbf{m}_*|}}.$$

Thus, in order to prove the lemma, it suffices to prove that for some $\delta > 0$

$$\mathbf{P}\left[\frac{1}{\sqrt{2\log|\mathbf{n}|}}\max_{\substack{\mathbf{k}\in K(\mathbf{n})\\ |Q(\mathbf{k})|>k_{\mathbf{n}}}}\frac{S_{Q(\mathbf{k})}}{\sqrt{|Q(\mathbf{k})|}} > 1+\frac{\varepsilon}{2}\right] < \frac{1}{|\mathbf{n}|^{1+\delta}} \tag{8}$$

and

$$\mathbf{P}\left[\frac{1}{\sqrt{2\log|\mathbf{n}|}}\max_{\substack{\mathbf{m}\leq\mathbf{n}\\ |\mathbf{m}|>k_{\mathbf{n}}}}\frac{S_{P(\mathbf{m})}}{\sqrt{|\mathbf{m}_*|}} > \frac{\varepsilon}{2}\right] < \frac{1}{|\mathbf{n}|^{1+\delta}}. \tag{9}$$

In order to prove (8) note that the left-hand side of (8) is not greater than

$$\sum_{\substack{\mathbf{k}\in K(\mathbf{n})\\ |Q(\mathbf{k})|>k_{\mathbf{n}}}} \mathbf{P}\left[\frac{S_{Q(\mathbf{k})}}{\sqrt{|Q(\mathbf{k})|}} > \left(1+\frac{\varepsilon}{2}\right)\sqrt{2\log|\mathbf{n}|}\right]$$

It follows from $|Q(\mathbf{k})| > k_{\mathbf{n}}$ and $\lim_{\mathbf{n}\to\infty} k_{\mathbf{n}}/(\log|\mathbf{n}|)^3 = \infty$ that $\sqrt{2\log|\mathbf{n}|} = o(|Q(\mathbf{k})|^{1/6})$ as $\mathbf{n}\to\infty$ and $|Q(\mathbf{k})| > k_{\mathbf{n}}$. Thus, we may use Theorem 3.2 to estimate the above sum by

$$C|K(\mathbf{n})|\,\mathbf{P}\left[N > (1+\varepsilon/2)\sqrt{2\log|\mathbf{n}|}\right] \le 2C|K(\mathbf{n})|\frac{1}{|\mathbf{n}|^{(1+\varepsilon/2)^2}}.$$

Here, N is a standard normal variable and we have used the well-known fact that $\mathbf{P}[N>x] \le 2e^{-x^2/2}$ as $x>0$. Finally, it is easy to see that $|K(\mathbf{n})| = O(\log n_1 \dots \log n_d) = o(|\mathbf{n}|^{\theta})$, $\mathbf{n}\to\infty$ for every $\theta>0$. This proves (8). To prove (9) note that by Lemma 3.2

$$\mathbf{P}\left[\frac{S_{P(\mathbf{m})}}{\sqrt{|\mathbf{m}_*|}} > \frac{\varepsilon}{2}\sqrt{2\log|\mathbf{n}|}\right] \le C\mathbf{P}\left[\frac{S_{P(\mathbf{m}^*)}}{\sqrt{|\mathbf{m}_*|}} > \frac{\varepsilon}{2}\sqrt{2\log|\mathbf{n}|}\right] \le C\mathbf{P}\left[\frac{S_{P(\mathbf{m}^*)}}{\sqrt{|P(\mathbf{m}^*)|}} > \frac{\varepsilon}{2}(\sigma-1)^{-1/2}\sqrt{2\log|\mathbf{n}|}\right],$$

since $|P(\mathbf{m}^*)|/|\mathbf{m}_*| \le \sigma-1$. Recalling that $(\sigma-1)^{-1/2}\frac{\varepsilon}{2} > 2$ and using Theorem 3.2, we may estimate the right-hand side from above by $O(1/|\mathbf{n}|^4)$. Thus, the left-hand side of (9) is not greater than

$$\sum_{\substack{\mathbf{m}\le\mathbf{n}\\ |\mathbf{m}|>k_{\mathbf{n}}}} O\left(\frac{1}{|\mathbf{n}|^4}\right) \le O\left(\frac{1}{|\mathbf{n}|^3}\right) < \frac{1}{|\mathbf{n}|^{1+\delta}}.$$

This finishes the proof of (9), Lemma 3.3 and Theorem 1.1. □

Acknowledgements. A. Munk and Z. Kabluchko acknowledge support from SFB 755 "Nanoscale Photonic Imaging" and from DFG FOR 916 "Statistical Regularisation". The authors are grateful to U. Einmahl for remarks concerning the strong approximation in the case of multidimensional arrays.

References

[1] Bissantz, N., Mair, B. Munk, A., 2008. A stochastic multiscale selection of the stopping iteration for MLEM reconstructions in PET. IEEE Trans. Med. Imag., tentatively accepted.
[2] Csörgö, M., Révész, P., 1981. Strong approximations in probability and statistics. Probability and Mathematical Statistics. New York-San Francisco-London, Academic Press.
[3] Davies, P.L., Kovac, A., 2001. Local extremes, runs, strings and multiresolution. (With discussion). Ann. Stat. 29, 1-65.
[4] Deheuvels, P., 1985. On the Erdös-Rényi theorem for random fields and sequences and its relationships with the theory of runs and spacings. Z. Wahrscheinlichkeitstheorie verw. Gebiete 70, 91-115.
[5] Dümbgen, L., Spokoiny, V.G., 2001. Multiscale testing of qualitative hypotheses. Ann. Stat. 29, 124-152.
[6] Dümbgen, L., Walther,G., 2006. Multiscale inference about a density. Preprint (Extended version: Technical report 56, Univ. of Bern), to appear Ann. Stat.
[7] Erdös, P., Rényi, A., 1970. On a new law of large numbers. J. Anal. Math. 23, 103-111.
[8] Feller, W., 1971. An introduction to probability theory and its applications. Vol II, 2nd ed., Wiley Series in Probability and Mathematical Statistics. New York etc.: John Wiley and Sons.
[9] van de Geer, S., Mammen, E., 2001. Discussion of "Local extremes, strings and multiresolution." Ann. Stat. 29, 56-59.
[10] Massart, P., 1989. Strong approximation for multivariate empirical and related processes, via KMT constructions. Ann. Probab. 17, 266-291.
[11] Hanson, D.L., Russo, R.P., 1983. Some results on increments of the Wiener process with applications to lag sums of i.i.d. random variables. Ann. Probab. 11, 609-623.

[12] Hinterberger, W., Hintermüller, M., Kunisch, K., von Oehsen, M., Scherzer, O., 2003. Tube methods for BV regularization. J. Math. Imag. Vision 19, 219-235.
[13] Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent RV's, and the sample DF. I. Z. Wahrscheinlichkeitstheor. Verw. Geb. 32, 111-131.
[14] Lanzinger, H., Stadtmüller, U., 2000. Maxima of increments of partial sums for certain subexponential distributions. Stochastic Processes Appl. 86, 307-322.
[15] Révész, P., 1990. Random walk in random and non-random environments. 1st ed., 1990; 2nd ed., 2005, World Scientific.
[16] Rio, E., 1997. Strong approximation for set-indexed partial sum processes via KMT constructions III. ESAIM: Probab. Stat. 1, 319-338.
[17] Siegmund, D., Yakir, B., 2000. Tail probabilities for the null distribution of scanning statistics. Bernoulli, 6, Number 2, 191-213.
[18] Shao, Q.-M., 1995. On a conjecture of Révész. Proc. Am. Math. Soc. 123, 575-582.
[19] Steinebach, J., 1983. On the increments of partial sum processes with multidimensional indices. Z. Wahrscheinlichkeitstheor. Verw. Geb. 63, 59-70.
[20] Steinebach, J., 1997. On a conjecture of Révész and its analogue for renewal processes. In: Szyszkowicz, Barbara (ed.), Asymptotic methods in probability and statistics. A volume in honour of Miklós Csörgö. ICAMPS '97, an international conference at Carleton Univ., Ottawa, Ontario, Canada. Amsterdam: North-Holland/ Elsevier.