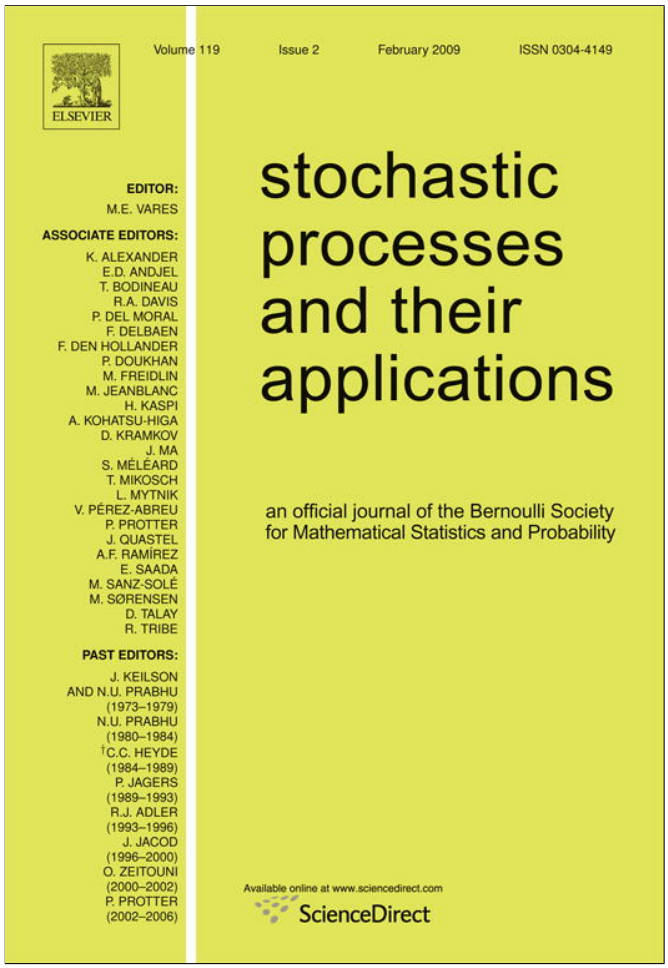
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 119 (2009) 468–490
stochastic processes and their applications
www.elsevier.com/locate/spa

Weak convergence of the tail empirical process for dependent sequences

Holger Rootzén*

Chalmers University of Technology, SE-419 62 Gothenburg, Sweden
Gothenburg University, SE-419 62 Gothenburg, Sweden

Received 2 July 2007; received in revised form 21 December 2007; accepted 4 March 2008
Available online 23 March 2008

Abstract

This paper proves weak convergence in D of the tail empirical process – the renormalized extreme tail of the empirical process – for a large class of stationary sequences. The conditions needed for convergence are (i) moment restrictions on the amount of clustering of extremes, (ii) restrictions on long range dependence (absolute regularity or strong mixing), and (iii) convergence of the covariance function. We further show how the limit process is changed if exceedances of a nonrandom level are replaced by exceedances of a high quantile of the observations. Weak convergence of the tail empirical process is one key to asymptotics for extreme value statistics and its wide range of applications, from geoscience to finance.
© 2008 Elsevier B.V. All rights reserved.

MSC: primary 60G70; secondary 60F17; 62G32

Keywords: Extremes; Clustering of extremes; Tail distribution function; Absolute regularity; Strong mixing

1. Introduction

This paper gives a number of convergence results for the tail empirical function for dependent stationary sequences. In addition to theoretical interest, the motivation comes from semiparametric methods for extremes, such as the Peaks over Thresholds (PoT) method. In these the statistical analysis only uses the part of the observations which exceed some suitably chosen high level. The methods, in particular with a generalized Pareto assumption for the tail

* Corresponding address: Mathematical Sciences, Chalmers, SE-41296 Gothenburg, Sweden. Tel.: +46 317723578; fax: +46 317723508.
E-mail address: rootzen@math.chalmers.se.

0304-4149/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spa.2008.03.003

Author's personal copy

distribution, are finding significant application, and are getting a firmer theoretical foundation. There are several recent books on the subject [7,23,26,27,2,34] and a large journal literature. The literature is complemented by a considerable body of software: for a review see [43].

Dependent observations are of basic interest in many classical application areas for extreme value statistics, e.g. geoscience and environmental science. Also, in the recent surge of interest in extreme values of financial data, dependence is the rule rather than the exception.

Our limit theorems provide a theoretical foundation for parts of dependent extreme value statistics. Earlier, unpublished and partial, versions of the present results for one of the cases (absolute regularity) treated here [39] has already found significant application in deriving asymptotic normality for PoT estimators; see [15–17]. The latter of these papers also shows that neglecting dependence can lead to severe underestimation of variability. Besides [37] which studies estimation for some specific heavy-tailed models, the only other extreme value estimation results for dependent sequences are aimed at Hill or Hill-like estimators [25,40,35,36,9,31,24].

To describe the results of this paper, let $\{\xi_i\}_{i=-\infty}^{\infty}$ be a stationary sequence with continuous marginal distribution function (d.f.) F. We throughout use the notation $\bar{F}(x) = 1 - F(x)$ for the tail d.f. Let $\{u_n\}_{n=1}^{\infty}$ be (high) levels and $\{\sigma_n > 0\}_{n=1}^{\infty}$ be norming constants. The tail function (or "conditional tail distribution function") is defined to be

$$T_n(x) = \frac{\bar{F}(u_n + x\sigma_n)}{\bar{F}(u_n)}, \quad x \geq 0, n = 1, 2, \ldots.$$

In much of this paper we assume that the tail function converges to a Generalized Pareto form, i.e.,

$$T_n(x) \to T(x) = \left(1 + \gamma \frac{x}{\sigma}\right)_+^{-1/\gamma}, \quad x \geq 0, n \to \infty, \tag{1.1}$$

where $\sigma > 0$ and $\gamma \in (-\infty, \infty)$ are parameters of the limit and the subscript $+$ denotes "positive part". For $\gamma = 0$ we interpret $T(x)$ to be the limit, $\mathrm{e}^{-x/\sigma}$, as $\gamma \to 0$. The special case of uniform distributions is considered separately in Section 5, for use as a technical tool here. It can also be useful for situations where the Generalized Pareto assumption isn't satisfied.

Let x_T be the right endpoint of the support of T, i.e., $x_T = \sup\{x; T(x) < 1\}$ so that $x_T = \infty$ for $\gamma \geq 0$, and $x_T = \sigma/|\gamma|$ for $\gamma < 0$. The tail empirical distribution function and the tail empirical process are defined as

$$\tilde{T}_n(x) = \frac{1}{n\bar{F}(u_n)} \sum_{i=1}^{n} 1_{\{\xi_i > u_n + x\sigma_n\}}$$

and

$$e(\tilde{T}_n)(x) = \sqrt{n\bar{F}(u_n)}(\tilde{T}_n(x) - T_n(x)),$$

respectively, with $1_{\{\}}$ denoting the indicator function which is one if the event in curly brackets occurs, and zero otherwise.

Let $D(I)$ be the space of real functions on the (finite or infinite) interval I which are right continuous and have left limits at each point, given the Lindvall–Stone extension of the Skorokhod J_1-topology; see [33].

The first results of this paper, Theorems 2.1 and 2.2, are that the tail empirical process converges in $D([0, x_T))$ to a continuous Gaussian process. The result requires three kinds of

Author's personal copy

assumptions beyond (1.1). We now briefly discuss these. A more detailed discussion is given in Section 4.

Often extremes in dependent sequences come in small clusters (in contrast to the case of independent sequences where extremes tend to be isolated from one another). The first kind of assumption restricts the size of clusters of large values by assuming that they have a suitably bounded p-th moment.

The second assumption makes clusters which are far apart asymptotically independent. For this long range dependence restriction we use either absolute regularity (sometimes also called β-mixing) or strong mixing. Absolute regularity is somewhat more restrictive, but still widely applicable. It applies e.g. in situations where coupling or regeneration holds; in particular for Markov chains, β-mixing with an exponential rate of decay is equivalent to geometric ergodicity ([6], Theorem 3.7), and it requires less by way of other conditions. Strong mixing holds more generally, but requires some strengthening of the other conditions.

Finally, we assume that the covariance function of $\{\tilde{T}_n(x); x \geq 0\}$ converges as $n \to \infty$. For standard empirical processes, this convergence may be obtained from more basic assumptions. However, this is not possible in the present tail context.

In applications, the "level" u_n has to be chosen using information from the sample. We study one way of doing this, by replacing u_n by $\xi_{[c_n]}$, the c_n-th largest of $\xi_1, \ldots, \xi_n$. Our results easily give convergence also in this situation. However, the limit is changed in an interesting way.

Related results in the literature concern tail empirical processes for independent sequences and ordinary empirical processes in dependent cases. Convergence and statistical application of the tail empirical process for i.i.d. sequences have been considered e.g. by Mason [29], Deheuvels and Mason [12], Einmahl [22], Csörgö et al. [8], Drees [14]; see also the review by Einmahl [21].

There is a large literature on convergence of standard empirical processes for dependent sequences. The problems in this case are of a rather different nature than for the tail empirical process. We refer the reader to [1,19,42] for sharp results and further information. A recent survey is [13]. β-mixing stems from Kolmogorov, and is used in [46]. Berbee [3] provides a sharp result. The form of it which we use is from [20].

We introduce some further notation, state the results for the case when (1.1) holds, and give a number of alternative conditions in Section 2. The more realistic case with exceedances over a random level $\xi_{[c_n]}$ is studied in Section 3. Section 4 discusses the conditions of the theorems and a further condition, $D'(u_n)$, which precludes clustering of extremes. It also lists a number of concrete models where the conditions are known to hold and mentions implications for probability, quantile, and mean excess plots. Section 5 contains the results for uniform marginal distributions, and Section 6 the proofs.

2. Convergence of the tail empirical process

In this section we state the result for the case when exceedances have a limiting Generalized Pareto distribution. We also introduce a number of variants of the conditions, which may be easier to check in specific cases. Sections 4 and 7 include some further discussion of the conditions.

Now, some notation which is needed to state the theorems. As in the introduction, let $\{\xi_i\}_{i=-\infty}^{\infty}$ be a stationary sequence with marginal d.f. F, let $\{\sigma_n > 0\}$ be a sequence of norming constants and let $\{u_n\}$ be a sequence of real numbers, the levels. Further, let $\mathcal{B}_{n,i}^{j} = \sigma\{\xi_i 1_{\{\xi_i > u_n\}}, \ldots, \xi_j 1_{\{\xi_j > u_n\}}\}$ be the σ-algebra generated by $\xi_i 1_{\{\xi_i > u_n\}}, \ldots, \xi_j 1_{\{\xi_j > u_n\}}$. With

Author's personal copy

some abuse of standard terminology we say that $\{\xi_i\}$ is β-mixing, or *absolutely regular*, if

$$\beta_n(k) = \sup_{\ell \geq 1} E \sup_{A \in \mathcal{B}^n_{n,\ell+k+1}} |P(A|\mathcal{B}^\ell_{n,1}) - P(A)| \to 0,$$

as $k = k_n \to \infty$, where by convention $\mathcal{B}^j_{n,i}$ is empty if $j < i$. The constants $\beta_n(k)$ are called the β-mixing constants for $\{\xi_n\}$.

Returning to standard terminology, let $\mathcal{B}^j_i$ be the σ-algebra generated by $\xi_i, \ldots, \xi_j$. Then, the sequence is strongly mixing if

$$\alpha(k) = \sup\{|P(AB) - P(A)P(B)|;\ A \in \mathcal{B}^\ell_1,\ B \in \mathcal{B}^\infty_{\ell+k+1},\ \ell \geq 1\} \to 0,$$

as $k \to \infty$. The constants $\alpha(k)$ are the strong-mixing (or α-mixing) constants for $\{\xi_n\}$.

Let $\ell_n \leq r_n \leq n$ be sequences of integers; r_n is the length of the big blocks and ℓ_n the lengths of the small blocks, in a "big blocks–small blocks" argument, and hence the number of blocks is of the order n/r_n. Further, we write

$$N_{n,i}(x, y) = \sum_{k=(i-1)r_n+1}^{ir_n} 1_{\{u_n + x\sigma_n < \xi_k \leq u_n + y\sigma_n\}}$$

for the number of normalized exceedance values in a block which fall between x and y, and often omit the index "i" when it is not needed. We will assume that the levels u_n tend to the right hand endpoint of the distribution as $n \to \infty$, but slowly enough to make the expected number of exceedances of u_n by $\xi_1, \ldots, \xi_n$ tend to infinity. Further, ℓ_n, the length of the small blocks which separate the big blocks, has to be big enough to make the big blocks asymptotically independent, but still small enough to make the small blocks asymptotically unimportant. Thus it will throughout, without further comment, be assumed that

$$\bar{F}(u_n) \to 0, \qquad n\bar{F}(u_n) \to \infty, \qquad \ell_n = o(r_n), \qquad r_n = o(n). \tag{2.1}$$

We will use the following conditions. For each $\theta < x_T$ and for $0 \leq x, y < \theta$ there is a constant c which only depends on θ such that

C1 $\quad E\{N_n(x, y)^p | N_n(x, y) \neq 0\} \leq c, \quad$ some $p \geq 2$,

C2 $\quad \beta_n(\ell_n) n/r_n \to 0$,

and

C3 $$\frac{1}{r_n \bar{F}(u_n)} C\left(\sum_{i=1}^{r_n} 1_{\{\xi_i > u_n + x\sigma_n\}}, \sum_{i=1}^{r_n} 1_{\{\xi_i > u_n + y\sigma_n\}}\right) \to r(x, y), \quad n \to \infty,$$

for some function $r(x, y)$. If $p = 2$ in C1 we in addition assume that

C4 $\quad r_n = o((n\bar{F}(u_n))^\nu), \quad$ for some $\nu < 1/2$.

We can now state the main results of this paper.

Theorem 2.1. *Suppose* (1.1) *and* C1–C3 *hold, and that if* $p = 2$ *also* C4 *is satisfied. Then*

$$e(\tilde{T}_n) \to e \ \text{ as } \ n \to \infty, \quad \text{in } D([0, x_T)),$$

where e is a centered continuous Gaussian process with covariance function $r(x, y)$.

Author's personal copy

For the strong-mixing case we also use the conditions

$$\text{D1} \quad \frac{1}{n\bar{F}(u_n)} \sum_{i,j=1}^{[n/r_n]} |C(N_{n,i}(x,y), N_{n,j}(x,y))| \le K|x-y|^{\mu},$$

and

$$\text{D2} \quad \alpha(\ell_n)n/r_n \to 0, \qquad n/r_n = o(n\bar{F}(v_n)^{\nu}), \qquad \alpha(n) = o(n^{\theta}),$$

for μ, ν and θ as specified below. We have the following result, corresponding to Theorem 5.1.

Theorem 2.2. *Suppose* (1.1) *and* C1, D1, D2, *and* C3 *hold with* $p/(p-2) < \theta$, $\nu < \theta/2 - (\theta+1)/(2(p-1))$, *and* $\mu > 2(\theta+p)/(p(\theta+1))$ *(the last condition in particular holds if* $\mu = 1$*). Then*

$$e(\tilde{T}_n) \xrightarrow{d} e \quad \text{as } n \to \infty, \text{ in } D([0, x_T)), \tag{2.2}$$

where e *is a continuous centered Gaussian process with covariance function* $r(x, y)$.

Remark 2.3. (i) For both theorems, if $\gamma \ge 0$, then $x_T = \infty$ and there is hence convergence in $D([0,\infty))$. For the case $\gamma < 0$, which entails $x_T < \infty$, it may be seen from the proof that if the conditions hold also for $\theta = x_T$, then there is convergence in $D([0,\infty))$. (But of course the limit is zero for $x > x_T$.)

(ii) For later use, we note that for both theorems there is convergence also in $D([-\epsilon, x_T))$ for an $\epsilon > 0$, provided the conditions are satisfied for u_n replaced by $u_n - \epsilon\sigma_n$. This can be seen by changing u_n to $u_n - \epsilon\sigma_n$ in the theorems.

(iii) The point of Theorem 2.2 is not so much to handle cases not covered by Theorem 2.1, although it seems likely such cases exist. It is instead that α-mixing sometimes can be much easier to handle; normal sequences providing a prominent example of this. □

The condition C1 is sometimes hard to prove, and we next consider alternative conditions which may be easier to check in concrete examples (for a discussion of the conditions, see Section 4). The conditions are assumed to hold for any $\theta \in (0, x_T)$ and $0 \le x, y \le \theta$ and a constant $c > 0$ which only depends on θ. The first one

$$\text{C1.1} \quad \frac{1}{r_n\bar{F}(u_n)} E(N_n(x,y)^p) \le c|x-y| \quad \text{some } p \ge 2$$

was proposed and extensively used by Drees for $p = 2$ and 4. (C1 may be written as $EN_n(x,y)^p \le cP(N_n(x,y) > 0)$. Essentially C1.1 amounts to replacing the bound $cP(N_n(x,y) > 0)$ by the larger and more tractable "Boole inequality" quantity $c\sum_{i=1}^{r_n} P(u_n + x\sigma_n < \xi_i \le u_n + y\sigma_n)$. This sometimes makes the condition easier to establish.)

Let

$$\tilde{N}_n(x,y) = \sum_{i=-r_n}^{r_n} 1_{\{u_n + x\sigma_n < \xi_i \le u_n + y\sigma_n\}}.$$

The next three conditions are

$$\text{C1.2} \quad E\{\tilde{N}_n(x,y)^p | u_n + x\sigma_n < \xi_0 \le u_n + y\sigma_n\} \le c, \quad \text{some } p \ge 2,$$

$$\text{C1.3} \quad E\{\tilde{N}_n(x,y)^p | \xi_0 = u_n + z\sigma_n\} \le c, \quad 0 \le z \le \theta, \quad \text{some } p \ge 2,$$

Author's personal copy

or that

C1.4 $E\{N_n(x, y)^p | \xi_0 = u_n + z\sigma_n\} \le c, \quad 0 \le z \le \theta, \{\xi_i\}$ is a Markov chain,

for some $p \ge 2$.

The condition C3 may, under an additional assumption, be simplified to

C3.1 $\bar{F}(u_n)^{-1} \sum_{|i|<r_n} (1 - |i|/r_n) P(\xi_0 > u_n + x\sigma_n, \xi_i > u_n + y\sigma_n) \to r(x, y).$

Corollary 2.4. (i) *The condition* C1 *of Theorems* 2.1 *and* 2.2 *may be replaced by any one of* C1.1–C1.4.
(ii) *If* $r_n \bar{F}(u_n) \to 0$ *then the condition* C3 *of Theorem* 2.1 *may be replaced by* C3.1.

3. Random levels

Let

$$c_n = n\bar{F}(u_n), \tag{3.1}$$

and let $\xi_{[c_n]}$ be the $[c_n]$-th largest of $\xi_1, \ldots, \xi_n$. In statistical applications c_n and u_n are not simultaneously known, since this entails knowledge of F. Hence, the level u_N is usually chosen as a function of the observations $\xi_1, \ldots, \xi_n$. Specifically u_n is often replaced by $\xi_{[c_n]}$. In this section we shall see that this changes the form of the limiting distribution in an interesting way, but that the previous results may be used in the derivation of the limit.

For this, let

$$\hat{T}_n(x) = \frac{1}{c_n} \sum_{i=1}^{n} 1_{\{\xi_i > \xi_{[c_n]} + x\sigma_n\}},$$

so that $\hat{T}_n$ is obtained by replacing u_n by $\xi_{[c_n]}$ in the definition of $\tilde{T}_n$. Further, let

$$e(\hat{T}_n)(x) = \sqrt{c_n} \left(\hat{T}_n(x) - \frac{\bar{F}(\xi_{[c_n]} + x\sigma_n)}{\bar{F}(\xi_{[c_n]})} \right) \tag{3.2}$$

be the normalized estimation error of $\hat{T}_n(x)$ considered as an estimator of the tail function at the level $\xi_{[c_n]}$.

We will assume, in correspondence to (2.1), that

$$c_n \to \infty, \qquad c_n/n \to 0, \tag{3.3}$$

that u_n is defined by (4.1), and that there is an $\epsilon > 0$ such that

$$T_n'(x) = \frac{\mathrm{d}}{\mathrm{d}x} \frac{\bar{F}(u_n + x\sigma_n)}{\bar{F}(u_n)} \to T'(x), \quad n \to \infty, \tag{3.4}$$

uniformly for $x \in [-\epsilon, \epsilon]$.

Theorem 3.1. *Suppose that* $e(\tilde{T}_n) \xrightarrow{d} e$ *in* $D([-\epsilon, \infty))$, *for some* $\epsilon > 0$ *and continuous process* e, *and that* (3.1)–(3.4) *hold. Then*

$$e_n(\hat{T}_n)(x) \xrightarrow{d} e(x) - T(x)e(0), \quad n \to \infty,$$

Author's personal copy

in $D([0, x_T))$. In particular, this convergence holds if the conditions of Theorems 2.1 *or* 2.2 *or of* Corollary 2.4 *hold, as modified in* Remark 2.3 *(ii), and* (3.4) *is satisfied.*

Finally, $|c_n - \sum_{i=1}^n 1_{\{\xi_i > \xi_{[c_n]}\}}| < 2$, and hence the result also holds if division by c_n is replaced by division by $\sum_{i=1}^n 1_{\{\xi_i > \xi_{[c_n]}\}}$ in the definition of $\hat{T}_n$. Equivalently, renormalizing $\hat{T}_n$ so that $\hat{T}_n(0) = 1$ does not change the limiting distribution. Similarly, $\sqrt{c_n}$ may be replaced by $\sqrt{\sum_{i=1}^n 1_{\{\xi_i > \xi_{[c_n]}\}}}$ in (3.2).

4. Comments on the conditions, examples, and applications

(i) *Condition* C1 is the main difference between the limit theory for tail empirical processes and standard empirical process theory. The appendix exhibits an example where C2, C3, and C4 hold, but where the result of Theorem 2.1 isn't true. In fact, in the example the β-mixing coefficients even decay exponentially fast. Hence, in general, some sort of condition like C1 is needed. Further, [16] argues that the variant C1.1 is rather close to being necessary for exponentially β-mixing variables, and it of course also is closely related to C3.

The intuitive content of C1 is that the lengths (or sizes) of clusters are suitably bounded, also conditionally. In particular this is brought out by the following somewhat more restrictive version of C1:

$$E(N_n(0, \infty)^p | N_n(x, y) > 0) \le c.$$

Here $N_n(0, \infty)$ is the number of exceedances in the first block, i.e. provided it is non-zero, $N_n(0, \infty)$ is the size of a cluster of exceedances.

Finally, it might be worth noting that C1 entails that $r_n \bar{F}(u_n)$ is bounded, so that the expected number of exceedances in a block of size r_n is bounded. (This claim is an easy consequence of C1.1 and Lyapunov's inequality, and C1.1 follows from C1.)

(ii) *Condition* C3 can also be seen to be needed; in fact, it is easy to construct examples of e.g. 2-dependent sequences where all the conditions of Theorem 2.1 are satisfied but where C3 doesn't hold and the tail empirical function doesn't converge. We leave this to the reader.

(iii) The norming of the empirical distribution function is in this paper by $\sqrt{n\bar{F}(u_n)}$, which means that the variation is assumed to be of the same order as if the ξ_i were independent. However, it would be straightforward to extend to more general norming sequences, for cases where dependence causes the variation to be of a larger order.

(iv) *Leadbetter's condition* $D'(v_n)$ holds for the sequence $\{v_n\}$ if

$$n \sum_{i=1}^{q_n} P(\xi_0 > v_n, \xi_i > v_n) \to 0, \quad m \to \infty,$$

for any sequence $q_n = o(n)$. If the condition holds for suitable sequences $\{v_n\}$ then it precludes clustering of extremes; see ([28], Section 3.4).

We next claim that if $D'(v_n)$ holds for any sequence $\{v_n\}$ such that $n\bar{F}(v_n) \to 1$ and if (2.1) and $r_n \bar{F}(u_n) \to 0$ also holds, then C3 follows, with

$$r(x, y) = T(x \vee y). \tag{4.1}$$

Thus, in this case if the remaining conditions of Theorem 2.1 also hold, then the limit has the same distribution as $W(T(x))$, for W a standard Brownian motion. In particular, this applies if the ξ_i's are i.i.d., as is well known; cf. [21].

Author's personal copy

To prove this claim put $m = m(n) = [1/\bar{F}(u_n)]$, $v_m = \inf\{u_n; [1/\bar{F}(u_n)] = m\}$ and $q_m = \sup\{r_n; [1/\bar{F}(u_n)] = m\}$. Then $m\bar{F}(v_m) \to 1$, and $q_m/m \to 0$ as $n \to \infty$ and hence

$$\begin{aligned} &F(u_n)^{-1} \sum_{0<|i|<r_n} P(\xi_0 > u_n + x\sigma_n, \xi_i > u_n + y\sigma_n) \\ &\quad \le F(u_n)^{-1} 2 \sum_{i=1}^{r_n} P(\xi_0 > u_n, \xi_i > u_n) \\ &\quad \le 2(m+1) \sum_{i=1}^{q_m} P(\xi_0 > v_m, \xi_i > v_m) \to 0, \quad n \to \infty. \end{aligned}$$

Further, by the definition of T_n,

$$\begin{aligned} &\bar{F}(u_n)^{-1} P(\xi_0 > u_n + x\sigma_n,\ \xi_0 > u_n + y\sigma_n) = T_n(x \vee y) \\ &\to T(x \vee y), \quad n \to \infty, \end{aligned}$$

by (1.1).

(v) *Examples:* Here we list a number of examples where the conclusion of Theorem 2.1 has been shown to hold. The results are taken from [39] and [15–17]. The list contains references to the proof of the result and the conditions which are needed for it in excess of (1.1) and (2.1).

- *k-dependent sequences* [39]. Condition:

$$\bar{F}(u_n)^{-1} \sum_{|i|<k} P(\xi_0 > u_n + x\sigma_n, \xi_i > u_n + y\sigma_n) \to r(x, y). \tag{4.2}$$

  $r(x, y)$ is the covariance of the limit process.
- *Finite moving averages* [39,16]. Conditions:

  The tails of the distributions are either asymptotic to $Cx^{\alpha}\mathrm{e}^{-x^p}$ or regularly varying.

  In the former case $\gamma = 0$ and if $p > 1$ the limiting covariance is given by (4.1), and otherwise it is given by (4.2) (and can be computed from the parameters α, p and the moving average weights). In the latter case $\gamma = 1/$(index of regular variation), and the limiting covariance is given in the references.
- *AR(1)-processes with α-stable innovations* [39,16]. Conditions:

$$u_n = O(n^{1/\alpha}(\log n)^{-\frac{2}{\alpha}-\epsilon}), \qquad (\log n)^{\frac{2}{\alpha}+\epsilon} = o(u_n).$$

  The limiting covariance is given in the references and $\gamma = 1/\alpha$.
- *Solutions to random difference equations* [15–17]. Here stationary solutions of $X_{i+1} = A_i X_i + B_i$ with (A_i, B_i) non-negative i.i.d. vectors are considered. Conditions:

  (i) there exist $\gamma, \xi > 0$ with $EA_1^{1/\gamma} = 1$, $EA_1^{1/\gamma+\xi} < \infty$, $EB_1^{1/\gamma+\xi} < \infty$,

  (ii) $\log^2 n \log^4(\log n)/n = o(\bar{F}(u_n))$, $\bar{F}(u_n) = o(n^{-\frac{1}{2\tau+1}})$, where $\tau > 0$ satisfies $\bar{F}(x) = x^{1/\gamma}(1 - O(x^{-\frac{\tau}{\gamma}}))$, and the existence of such a τ follows from (i).

  The limiting covariance is given in the references.
- *ARCH*(1)-processes [16,17]. If the innovations are symmetric and have finite variance, no further conditions are needed, except restrictions on u_n which may be derived from (ii) above. The limiting covariance cannot be computed analytically, but can be computed via simulation.
- *Causal linear time series with regularly varying innovations* ([17]; more general results in [9]). Conditions:

  The innovations have a Lipschitz continuous density, their tails are balanced, the weights decrease geometrically, the index of regular variation is less than 2, rather involved conditions on u_n given in [17].

Author's personal copy

The limiting covariance is given in the references and $\gamma = 1/(\text{index of regular variation})$.

(vi) *Application to diagnostic plotting:* In a pp-plot one plots $\hat{T}_n(x)$ on a suitable scale. The results from the previous section show that the deviations between this estimate and the true tail function $\frac{\bar{F}(\xi_{[c_n]}+x\sigma_n)}{\bar{F}(\xi_{[c_n]})}$ approximately is distributed as $1/\sqrt{c_n}$ times a continuous Gaussian process with covariance simply computed from the asymptotic covariance $r(x, y)$. By the remark after Theorem 3.1 this also holds if $\hat{T}_n$ is renormalized so that $\hat{T}_n(0) = 1$.

In a qq-plot one instead plots the inverse of $\hat{T}_n(x)$. It follows from Vervaat's Lemma [44] (see [16]) that the deviation from the inverse of the true tail functions has the same asymptotic distribution as the deviations in a pp-plot.

A mean excess plot shows the empirical mean of the observations over a high level, as a function of the level. Under suitable extra moment conditions our results also give approximations for the deviations from the true mean excess function; cf. [42]. However, to actually prove this would require further work.

The covariance function $r(x, y)$ is unknown and has to be estimated. One possibility is using the blocks method, where one first divides the "interval" $1, \ldots, n$ into "blocks" of length r_n, and also typically changes the scale by replacing x by $y = x\sigma_n$ to remove the unknown parameter σ_n. For each block one then computes a bivariate vector with first entry equal to the number of exceedances of $\xi_{[c_n]+x}$ in the block and second entry the number of exceedances of $\xi_{[c_n]+y}$ in the block. From these approximately n/r_n bivariate vectors one then computes an estimate of the correlation $r(x, y)$ in a standard way. The drawback of this procedure is the problem of how to choose the block length r_n. However the estimate may often be rather insensitive to this choice. [17] proposes very promising alternative ways to estimate quantities related to $r(x, y)$. It would be interesting to explore whether modifications of Drees' ideas could be used for the present problem.

For practical use it would often be desirable to replace $\bar{F}(\xi_{[c_n]} + x\sigma_n)/\bar{F}(\xi_{[c_n]})$ by the limiting generalized Pareto tail $T(x) = (1 + \gamma\frac{x}{\sigma})_+^{-1/\gamma}$ and the unknown parameter σ_n by an estimate. It is straightforward to see that the former is possible if (1.1) is strengthened to $\sqrt{c_n}|T_n(x) - T(x)| \to 0$ for $x > 0$. As for the latter, as is usual in this kind of problem, it would change the asymptotic variance. Furthermore the change could be different for different estimators of σ. However, it is likely that the changes would be moderate, and also that using the variance estimate just discussed is conservative.

5. Uniform marginal distributions

We now state versions of the results from Section 2 which instead of (1.1) assume uniform marginal distributions. Thus, let $\{\eta_i\}_{i=-\infty}^{\infty}$ be a stationary sequence with uniform marginal d.f.'s, $P(\eta_i \le x) = x$ for $0 \le x \le 1$. Let $\{v_n \in (0, 1)\}$ be a sequence of constants, and $\{r_n\}$, $\{\ell_n\}$ sequences of integers, which throughout are assumed to satisfy

$$v_n \to 0, \qquad nv_n \to \infty, \qquad \ell_n = o(r_n), \qquad r_n = o(n). \tag{5.1}$$

Further let

$$\bar{T}_n(x) = \frac{1}{nv_n}\sum_{i=1}^{n} 1_{\{\eta_i > 1 - v_n + v_n x\}},$$
$$e(\bar{T}_n(x)) = \sqrt{nv_n}(\bar{T}_n(x) - (1 - x)),$$

Author's personal copy

be the uniform tail and tail empirical functions, respectively, and write

$$\bar{N}_{n,i}(x, y) = \sum_{k=(i-1)r_n+1}^{ir_n} 1_{\{1-v_n+v_n x<\eta_k\le 1-v_n+v_n y\}},$$

for $0 \le x \le y \le 1$. As above, in this section a bar indicates quantities defined from the η_i's. We will use the following assumptions, corresponding to C1–C4. For each $\theta < 1$ and $0 \le x, y \le \theta$, there is a constant c which only depends on θ such that

$\bar{\text{C}}1$ $\quad E\{\bar{N}_n(x, y)^p | \bar{N}_n(x, y) \ne 0\} \le c, \quad$ some $p \ge 2$,

$\bar{\text{C}}2$ $\quad \bar{\beta}_n(\ell_n)n/r_n \to 0$,

where the $\bar{\beta}_n(k)$'s are the β-mixing coefficients for $\{\eta_i\}$, and

$$\bar{\text{C}}3 \quad \frac{1}{r_n v_n} C\left(\sum_{i=1}^{r_n} 1_{\{\eta_i>1-v_n+v_n x\}}, \sum_{i=1}^{r_n} 1_{\{\eta_i>1-v_n+v_n y\}}\right) \to \bar{r}(x, y), \quad n \to \infty,$$

for some function $\bar{r}(x, y)$. If $p = 2$ in $\bar{\text{C}}1$ we in addition assume that

$\bar{\text{C}}4$ $\quad r_n = o((nv_n)^\nu), \quad$ for some $\nu < 1/2$.

The first result gives convergence of the tail empirical function for uniformly distributed variables under absolute regularity.

Theorem 5.1. *Suppose that* $\bar{\text{C}}1$ *and* $\bar{\text{C}}2$ *are satisfied. Further, if* $p = 2$ *assume that* $\bar{\text{C}}4$ *holds. Then*

(i) $\{e(\bar{T}_n)(x); 0 \le x \le 1\}$ *is tight in* $D[0, 1)$*, and any distributional limit of a subsequence is continuous.*

(ii) *If in addition* $\bar{\text{C}}3$ *is satisfied, then*

$$e(\bar{T}_n) \xrightarrow{d} \bar{e} \quad \text{in } D[0, 1), \text{ as } n \to \infty, \tag{5.2}$$

where $\bar{e}$ *is a continuous centered Gaussian process with covariance function* $\bar{r}(x, y)$.

For the strong-mixing case, the new conditions are

$$\bar{\text{D}}1 \quad \frac{1}{nv_n} \sum_{i,j=1}^{[n/r_n]} |C(\bar{N}_{n,i}(x, y), \bar{N}_{n,j}(x, y))| \le K|x - y|^\mu,$$

and

$$\bar{\text{D}}2 \quad \bar{\alpha}(\ell_n)n/r_n \to 0, \qquad n/r_n = o((nv_n)^\nu), \qquad \bar{\alpha}(n) = o(n^\theta),$$

where $\bar{\alpha}$ are the strong-mixing constants for η_i. We then have the following result.

Theorem 5.2. *Suppose that* $\bar{\text{C}}1$, $\bar{\text{D}}1$, $\bar{\text{D}}2$, *and* $\bar{\text{C}}3$ *hold with* $p/(p-2) < \theta$, $\nu < \theta/2 - (\theta + 1)/(2(p-1))$, *and* $\mu > 2(\theta + p)/(p(\theta + 1))$ *(the last condition in particular holds if* $\mu = 1$*). Then*

$$e(\bar{T}_n) \xrightarrow{d} \bar{e} \quad \text{in } D[0, 1), \text{ as } n \to \infty, \tag{5.3}$$

where $\bar{e}$ *is a continuous centered Gaussian process with covariance function* $\bar{r}(x, y)$.

Author's personal copy

It of course is straightforward to translate C1.1–C1.4 and C3.1 to the present case. We leave this to the reader. Furthermore, Drees [15] has shown that $\bar{\text{C}}4$ may be slightly weakened, to $r_n(nv_n)^{-1/2}\log^4(nv_n) \to 0$. Drees [16,17] also gives some further conditions, which are more stringent but sometimes easier to check, and which also can replace $\bar{\text{C}}1$ and $\bar{\text{C}}3$.

6. Proofs

We first prove Theorem 5.1. For this, let $k_n = [n/(2r_n)]$ be the integer part of $n/(2r_n)$. In the proof of tightness, we split the "interval" $\{1, \ldots, n\}$ up into $2k_n$ blocks $I_1, \ldots, I_{2k_n}$ of r_n integers and a remaining block I_{2k_n+1} of length less than $2r_n$, and consider sums over even and odd blocks separately.

For $e_{n,i}(x) = (nv_n)^{-1/2}(1_{\{\eta_i > 1-v_n+v_n x\}} - v_n(1-x))$ let $\{f_{n,i}(x);\ 0 \le x \le 1,\ i \in I_{2j}\}$ have the same distribution as $\{e_{n,i}(x);\ 0 \le x \le 1,\ i \in I_{2j}\}$ for $j = 1, \ldots, k_n$, but let them be independent for different values of j. Thus the $f_{n,i}$'s have the same joint distribution as the $e_{n,i}$'s, except that $f_{n,i}$'s in separate even blocks are independent. Further, let

$$X_{n,j}(x) = \sum_{i \in I_{2j}} f_{n,i}(x)$$

be the sum over the j-th even block, so that the X_{nk}'s are mutually independent for $j = 1, \ldots, k_n$, and let

$$S_n(x) = \sum_{j=1}^{k_n} X_{n,j}(x)$$

be the sum of the even blocksums.

The next lemma shows that by β-mixing the $e_{n,i}$'s may be replaced by $f_{n,i}$'s in the proof of tightness of the sum over even blocks. Tightness of the sum over odd blocks may be proved in the same way as for the sum over even blocks. Finally, the sum over the last block, I_{2k_n+1}, is easily seen to tend to zero, and together these facts can be used to deduce tightness of $\bar{e}_n$.

Lemma 6.1. *Suppose* $\bar{\text{C}}2$ *holds and that* $\{S_n(x);\ 0 \le x \le 1\}$ *is tight in* $D[0,1)$*, with any subsequential distributional limit continuous. Then also the sum of the original variables over the even blocks,*

$$\left\{ \sum_{i=1}^{k_n} \sum_{i \in I_{2j}} e_{n,i}(x);\ 0 \le x \le 1 \right\}_{n=1}^{\infty},$$

is tight in $D[0,1)$*, and any limit in distribution of a subsequence is continuous.*

Proof. Using e.g. [20] the variation distance between $\{f_{n,i}(\cdot);\ i \in I_{2j},\ j = 1, \ldots, k_n\}$ and $\{e_{n,i}(\cdot);\ i \in I_{2j},\ j = 1, \ldots, k_n\}$ is bounded by $k_n\bar{\beta}_n(r_n) < \bar{\beta}_n(\ell_n)n/r_n \to 0$, by $\bar{\text{C}}2$, and the lemma follows. □

Next, $\bar{\text{C}}1$ gives a moment bound on the increments of the $X_{n,j}$'s. In the lemma, and later, K denotes a generic constant whose value may change from appearance to appearance, and which only depends on $\bar{\theta}$, where $\bar{\theta}$ is an arbitrary number in $(0, 1)$.

Author's personal copy

Lemma 6.2. *Suppose* $\bar{\mathrm{C}}1$ *holds. Then, for* $0 \le x, y \le \bar{\theta}$ *and* $\zeta \ge p \ge 2$,

$$E\{|X_{n,1}(y) - X_{n,1}(x)|^{\zeta}\} \le K k_n^{-1} \left(\frac{r_n^2}{n v_n}\right)^{(\zeta-p)/2} (n v_n)^{1-p/2} |y - x|, \tag{6.1}$$

for large n.

Proof. Assume $x \le y$ and recall the notation $\bar{N}_n(x, y) = \sum_{i=1}^{r_n} 1_{\{1-v_n+v_n x < \eta_i \le 1-v_n+v_n y\}}$. By the definition of $X_{n,1}$, since central moments are bounded by a constant times the moment of the variable itself,

$$E\{|X_{n,1}(y) - X_{n,1}(x)|^{\zeta}\} \le K (n v_n)^{-\zeta/2} E(\bar{N}_n(x, y)^{\zeta}).$$

Now, using $\bar{\mathrm{C}}1$ in the third step,

$$\begin{aligned} E(\bar{N}_n(x, y)^{\zeta}) &\le r_n^{\zeta-p} E(\bar{N}_n(x, y)^p) \\ &\le r_n^{\zeta-p} E(\bar{N}_n(x, y)^p | \bar{N}_n(x, y) \ne 0) P(\bar{N}_n(x, y) \ne 0) \\ &\le r_n^{\zeta-p} c r_n v_n |y - x|. \end{aligned} \tag{6.2}$$

Since $r_n \le n/k_n$, this proves the lemma. □

We can now control the variation of S_n on a mesh of size

$$\begin{aligned} \Delta = \Delta_n &= (n v_n)^{-1} \quad \text{for } p > 2 \\ &= \left(\frac{r_n^2}{n v_n}\right)^{1/(1/2-\nu)} \quad \text{for } p = 2. \end{aligned} \tag{6.3}$$

Let J_k^{δ} denote the set of integers i for which $0 \vee (k\delta - \Delta) \le i\Delta \le ((k+1)\delta + \Delta) \wedge \bar{\theta}$.

Lemma 6.3. *Suppose* $\bar{\mathrm{C}}1$ *holds. Then to* $\epsilon, \eta > 0$ *there is a* $\delta > 0$ *with*

$$P(\max_{i,j \in J_k^{\delta}} |S_n(j\Delta) - S_n(i\Delta)| > \epsilon) \le \eta\delta, \tag{6.4}$$

for $k \ge 0$ *and* n *large.*

Proof. First note that if $\bar{\mathrm{C}}1$ holds for some $p > 2$ then it also holds for $p = 2$, by Lyapunov's inequality. Now, let $\zeta \ge p \ge 2$ and for brevity write X for $X_{n,i}$ (for each n, the $\{X_{n,i} : i \ge 0\}$ form a stationary sequence). Then, by the Rosenthal inequality (see e.g. [32], p. 59) and Lemma 6.2,

$$\begin{aligned} E|S_n(x) - S_n(y)|^{\zeta} &\le K\{k_n E|X(y) - X(x)|^{\zeta} + (k_n E|X(y) - X(x)|^2)^{\zeta/2}\} \\ &\le K\left\{\left(\frac{r_n^2}{n v_n}\right)^{(\zeta-p)/2} (n v_n)^{1-p/2} |y - x| + |x - y|^{\zeta/2}\right\}. \end{aligned} \tag{6.5}$$

Now, assume $p > 2$ and take $\zeta = p$. It then follows from (6.5) that, for $|y - x| \ge (n v_n)^{-1} = \Delta$,

$$\begin{aligned} E|S_n(x) - S_n(y)|^p &\le K|y - x|^{p/2} \left\{\left(\frac{(n v_n)^{-1}}{|y - x|}\right)^{-1+p/2} + 1\right\} \\ &\le K|y - x|^{p/2}. \end{aligned}$$

Author's personal copy

Hence, by [5], Theorem 12.2, the left hand side of (6.4) is bounded by

$$K\frac{(\delta+2\Delta)^{p/2}}{\epsilon^p} = \left(K\frac{(\delta+2\Delta)^{p/2}}{\delta\epsilon^p}\right)\delta.$$

By taking δ sufficiently small and for n large $K(\delta+2\Delta)^{p/2}\delta^{-1}\epsilon^{-p} < \eta$ and (6.4) holds. This concludes the proof for the case $p > 2$.

Next, assume $p = 2$ and take $\zeta = 4$ in (6.5) to get that, for $|x-y| \geq (r_n^2/(nv_n))^{1/(1/2-\nu)} = \Delta$,

$$\begin{aligned} E|S_n(x)-S_n(y)|^4 &\leq K|y-x|^{1+(1/2-\nu)}\left\{\left(\frac{(r_n^2/nv_n)^{\frac{1}{1/2-\nu}}}{|y-x|}\right)^{(1/2-\nu)} + |x-y|^{(1/2+\nu)}\right\} \\ &\leq K|y-x|^{1+(1/2-\nu)}. \end{aligned}$$

The proof now can be completed in the same way as above. □

The next lemma, which uses monotonicity of $1_{\{\eta_i > 1-v_n+v_nx\}}$ and of $v_n(1-x)$ as functions of x, is classical, and its proof is hence left to the reader.

Lemma 6.4. *For* $0 \leq x, y \leq \bar{\theta}$,

$$\sup_{|x-y|\leq\delta} |S_n(x)-S_n(y)| \leq 6\max_{k\geq 0}\max_{i,j\in J_k^\delta}|S_n(i\Delta)-S_n(j\Delta)| + 8\sqrt{nv_n}\Delta.$$

Together, Lemmas 6.4 and 6.5 bound the variation of $\{S_n(t)\}$, by standard arguments.

Lemma 6.5. *Suppose that* $\bar{\mathrm{C}}1$–$\bar{\mathrm{C}}3$ *hold, and that if* $p = 2$ *also* $\bar{\mathrm{C}}4$ *is satisfied. Then*

(i) *for* $\epsilon, \eta > 0$ *there exists a* $\delta > 0$ *with*

$$\limsup_{n\to\infty} P\left(\sup_{|x-y|\leq\delta, 0\leq x,y\leq\bar{\theta}} |S_n(x)-S_n(y)| > \epsilon\right) \leq \eta,$$

and

(ii) *for* $e_{n,i}(x)$ *and* $I_{n,2k_n+1}$ *as defined after Theorem 5.1,*

$$\sup_{0\leq x\leq 1}\left|\sum_{i\in I_{2k_n+1}} e_{n,i}(x)\right| \xrightarrow{P} 0.$$

Proof. (i) With δ as above,

$$\begin{aligned} &\limsup_{n\to\infty} P(\max_{0\leq k}\max_{i,j\in J_k^\delta}|S_n(i\Delta)-S_n(j\Delta)| > \epsilon) \\ &\quad\leq \limsup_{n\to\infty}\sum_{0\leq k<1/\delta} P(\max_{i,j\in J_k^\delta}|S_n(i\Delta)-S_n(j\Delta)| > \epsilon) \leq \delta^{-1}\eta\delta = \eta. \end{aligned}$$

Further, by (6.3), $\sqrt{nv_n}\Delta \to 0$, as $n \to \infty$ and part (i) now follows from Lemma 6.4.

(ii) By definition, for $x \in [0, 1]$,

$$\left|\sum_{i\in I_{2k_n+1}} e_{n,i}(x)\right| \leq \frac{1}{\sqrt{nv_n}}\sum_{i\in I_{2k_n+1}}(1_{\{\eta_i>1-v_n\}} + v_n).$$

Author's personal copy

Here the right hand side does not depend on x. The length of I_{2k_n+1} is bounded by $2r_n$, and hence the expectation of the right hand side is bounded by $4(nv_n)^{-1/2}r_nv_n$. Further $nv_n \to \infty$ by assumption and r_nv_n is bounded (see Remark (i), Section 4) so that $4(nv_n)^{-1/2}r_nv_n \to 0$. This proves the second part of the lemma. □

Proof of Theorem 5.1. (i) By Lemma 6.5(i) and Theorem 15.5 of [5], for any $\bar\theta \in (0, 1)$ we have that $\{S_n(x); x \in [0, \bar\theta]\}$ is tight in $D[0, \bar\theta]$, and that any limit in distribution along a subsequence is continuous. By Lemma 6.2 the same holds for

$$\sum_{j=1}^{k_n}\sum_{i\in I_{2j}} e_{n,j}(x).$$

Using exactly the same reasoning, this conclusion also applies to

$$\sum_{j=1}^{k_n}\sum_{i\in I_{2j-1}} e_{n,j}(x).$$

Now, since

$$e(\tilde T_n(x)) = \sum_{j=1}^{k_n}\sum_{i\in I_{2j-1}} e_{n,i}(x) + \sum_{j=1}^{k_n}\sum_{i\in I_{2j}} e_{n,i}(x) + \sum_{i\in I_{2k+1}} e_{n,i}(x),$$

this together with Lemma 6.5(ii) proves the first part of Theorem 5.1.

(ii) Since tightness already is established, only convergence of finite-dimensional distributions remains to be proved. Further, using the Cramér–Wold device in a standard way, this is only notationally more complicated than proving one-dimensional convergence. (Note that the Lindeberg condition for Cramér–Wold is an easy consequence of the one-dimensional Lindeberg conditions.) Hence we only give the proof that $\bar e_n(x)$ converges in distribution to $\bar e(x)$, for x fixed.

To do this we show that the conditions of Corollary 4.2 of [41] (abbreviated to Corollary 4.2 in the following) are satisfied. Since strong-mixing coefficients are bounded by the β-mixing coefficients in $\bar{\mathrm{C}}2$, the mixing condition (2.1) of Corollary 4.2 is satisfied by assumption.

Next, we show that

$$\frac{1}{r_nv_n} V\left(\sum_{i=1}^{\ell_n} 1_{\{\eta_i > 1-v_n+v_nx\}}\right) \to 0. \tag{6.6}$$

Expanding the sum and using stationarity it can be seen that $E((\sum_1^{r_n} 1_{\{\eta_i>1-v_n+v_nx\}})^2) \geq (r_n/\ell_n - 1)E((\sum_1^{\ell_n} 1_{\{\eta_i>1-v_n+v_nx\}})^2)$. Combining this with (6.2), with $\zeta = p = 2$ and $y = 0$, we get that

$$\frac{1}{r_nv_n} E\left(\left(\sum_{i=1}^{\ell_n} 1_{\{\eta_i>1-v_n+v_nx\}}\right)^2\right) \leq cx/(r_n/\ell_n - 1) \to 0.$$

This implies (6.6), and thus the first part of condition (4.7) of Corollary 4.2 is satisfied. This argument is due to Holger Drees (personal communication.) The second part may be proved similarly to Lemma 6.5(ii).

Further, by Lemma 6.2 with $\zeta = p > 2$, if $\bar{\mathrm{C}}1$ holds then $k_nE|X_{n,i}(x)|^p \to 0$ and thus the Lindeberg condition is satisfied. If instead $\bar{\mathrm{C}}4$ is assumed to hold, then $|X_{n,i}(x)|$ is bounded

Author's personal copy

by $r_n(nv_n)^{-1/2} \to 0$, and again the Lindeberg condition holds. Thus, all the assumptions of Corollary 4.2 are shown to hold, and hence $\bar{e}_n(x) \xrightarrow{d} \bar{e}(x)$. □

Proof of Theorem 5.2. The proof of convergence of finite-dimensional distributions in Theorem 5.1 only requires strong mixing, so it is enough to prove that tightness follows from the present assumptions.

Choose the mesh size

$$\Delta = \Delta_n = \left(\frac{1}{nv_n}\right)^{1/2} (\log(nv_n))^{-1},$$

so that $(nv_n)^{1/2}\Delta_n \to 0$, choose $\theta' < \theta$ such that $p/(p-2) < \theta'$, $\nu < \theta'/2-(\theta'+1)/(2(p-1))$, and $\mu > 2(\theta'+p)/(p(\theta'+1))$ and put $\zeta = p(\theta'+1)/(p+\theta')$. It is straightforward to check that then $2 < \zeta < p$.

With the notation of Rio [38, p. 8 and 12], let $\bar{\alpha}^{-1}(u) = \sum_{i=0}^{\infty} 1_{\{u<\bar{\alpha}_i\}}$ and let Q be the right continuous inverse of the decreasing function $f(t) = P(|X(x) - X(y)| > t)$. To apply Theorem 6.3 of the cited reference we use Hölder's inequality to obtain that

$$\begin{aligned}\int_0^1 (\alpha^{-1}(u))^{\zeta-1} Q^{\zeta}(u)\mathrm{d}u &\le \left(\int_0^1 (\alpha^{-1}(u))^{p(\zeta-1)/(p-\zeta)}\right)^{(p-\zeta)/p} \left(\int_0^1 Q^p(u)\right)^{\zeta/p} \\ &\le K(E|X(x)-X(y)|^p)^{\zeta/p},\end{aligned}$$

since $p(\zeta-1)/(p-\zeta) < \theta$ so that the first integral on the right hand side is convergent. Further let $s_n^2 = \frac{1}{nv_n}\sum_{i,j=1}^{[n/r_n]} |C(\bar{N}_{n,i}(x,y), \bar{N}_{n,j}(x,y))|$, and with a minor change of notation, write $S_n(x) = \frac{1}{nv_n}\sum_{i=1}^{[n/r_n]}(N_{n,i}(x) - r_n v_n(1-x))$. Then by Rio [38, Theorem 6.3], using $\bar{\mathrm{D}}1$, Lemma 6.2 with $\zeta = p$, and easy computation in the second step, we obtain that for $|x-y| \ge \Delta$ and $\epsilon > 0$

$$\begin{aligned}E|S_n(x) - S_n(y)|^{\zeta} &\le K\{k_n(E|X(x)-X(y)|^p)^{\zeta/p} + s_n^{\zeta}\} \\ &\le K\{k_n^{1-\zeta/p}(nv_n)^{\zeta/p-\zeta/2}|x-y|^{\zeta/p} + |x-y|^{\mu\zeta/2}\} \\ &\le K\{k_n^{1-\zeta/p}(nv_n)^{\zeta/p-\zeta/2}\Delta^{\zeta/p-1-\epsilon}|x-y|^{1+\epsilon} + |x-y|^{\mu\zeta/2}\}.\end{aligned}$$

It can be checked that if ϵ is sufficiently small then $k_n^{1-\zeta/p}(nv_n)^{\zeta/p-\zeta/2}\Delta^{\zeta/p-1-\epsilon}$ is bounded, and furthermore that $\mu\zeta/2 > 1$. (The latter condition in particular is satisfied if $\mu = 1$.) Thus we have proved that if $\epsilon > 0$ is chosen sufficiently small then

$$E|S_n(x) - S_n(y)|^{\zeta} \le K|x-y|^{1+\epsilon}.$$

Inspection of the proof of Theorem 5.1 shows that tightness then follows along the lines of Lemmas 6.3–6.5. □

Theorems 2.1 and 2.2 follow from Theorems 5.1 and 5.2 by a simple "change of variables" argument. The proofs are the same, so we only prove Theorem 2.1.

Proof of Theorem 2.1. First recall the assumption (1.1) that

$$T_n(x) = \frac{\bar{F}(u_n + x\sigma_n)}{\bar{F}(u_n)} \to \left(1 + \gamma\frac{x}{\sigma}\right)^{-1/\gamma} = T(x), \quad x \ge 0, \tag{6.7}$$

as $n \to \infty$. Further, let

$$T_n^{\leftarrow}(y) = \sup\{x : T_n(x) \ge y\}, \quad 0 < y \le 1,$$

Author's personal copy

be the left continuous inverse of T_n. Since T is bounded, continuous and strictly decreasing, it follows that the convergence in (6.7) holds uniformly for $x \geq 0$ and that

$$T_n^{\leftarrow}(y) \to T^{-1}(y) = \frac{\sigma}{\gamma}(y^{-\gamma} - 1), \tag{6.8}$$

uniformly for $y \in [0, \tilde{\theta}]$ for any $\tilde{\theta} < T^{-1}(x_T)$. Since F is assumed to be continuous, $\eta_i = F(\xi_i)$ is uniformly distributed on [0, 1], and

$$\begin{aligned}\{\xi_i > u_n + x\sigma\} &\overset{\text{a.s.}}{=} \{\eta_i > F(u_n + x\sigma_n)\} \\ &= \{\eta_i > 1 - \bar{F}(u_n) + \bar{F}(u_n)(1 - T_n(x))\}.\end{aligned} \tag{6.9}$$

Hence, if we put $v_n = \bar{F}(u_n) \to 0$ and write $\bar{T}_n(x) = n^{-1}\sum_{i=1}^n 1_{\{\eta_i > 1 - v_n + v_n x\}}$, as before, then

$$\tilde{T}_n(x) \overset{\text{a.s.}}{=} \bar{T}_n(1 - T_n(x)). \tag{6.10}$$

We first show that $\{\eta_i\}$ satisfies the assumptions of Theorem 5.1. Clearly, (2.1) implies that (5.1) holds. Similarly as for (6.9),

$$\bar{N}(x, y) \overset{\text{a.s.}}{=} N_n(T_n^{\leftarrow}(1 - x), T_n^{\leftarrow}(1 - y)).$$

By (6.8), for any $\bar{\theta} \in (0, 1)$ there is a $\theta > 0$ such that $T_n^{\leftarrow}(1 - \bar{\theta}) \leq \theta < x_T$ for all n, and hence it follows from C1 that $\bar{\text{C}}1$ holds for $0 \leq x, y \leq \bar{\theta}$. Next, $\bar{\text{C}}2$ follows automatically from C2, and a straightforward argument shows that (6.8), C2, and C3 imply that

$$\begin{aligned}&\frac{1}{r_n v_n} C\left(\sum_{i=1}^{r_n} 1_{\{\eta_i > 1 - v_n + v_n x\}}, \sum_{i=1}^{r_n} 1_{\{\eta_i > 1 - v_n + v_n y\}}\right) \\ &= \frac{1}{r_n \bar{F}(u_n)} C\left(\sum_{i=1}^{r_n} 1_{\{\xi_i > u_n + T_n^{\leftarrow}(1-x)\sigma_n\}}, \sum_{i=1}^{r_n} 1_{\{\xi_i > u_n + T_n^{\leftarrow}(1-y)\sigma_n\}}\right) \\ &\to r(T^{-1}(1 - x), T^{-1}(1 - y)), \quad n \to \infty.\end{aligned}$$

Hence also $\bar{\text{C}}3$ holds. Similarly $\bar{\text{C}}4$ follows from C4. By Theorem 5.1 we hence have that

$$e(\bar{T}_n) \to \bar{e} \quad \text{in } D[0, \bar{\theta}], n \to \infty, \tag{6.11}$$

with $\bar{e}$ a centered continuous normal process with covariance
$\bar{r}(x, y) = r(T^{-1}(1 - x), T^{-1}(1 - y))$.

It now follows from (6.11), $T_n \to T$ uniformly and Theorem 3.1 of [45] that

$$e(\tilde{T}_n(x)) = e(\tilde{T}_n(1 - T_n(x))) \to \bar{e}(1 - T(x)) \quad \text{in } D[0, \theta], n \to \infty,$$

for any $\theta \in (0, x_T)$, since for such a θ there is a $\bar{\theta} \in (0, 1)$ with $1 - T_n(\theta) \leq \bar{\theta}$ for all sufficiently large n. Further, $\bar{e}(1 - T(x))$ is a centered continuous normal process with covariance function

$$\begin{aligned}\bar{r}(1 - T(x), 1 - T(y)) &= r(T^{-1}(1 - (1 - T(x))), T^{-1}(1 - (1 - T(y)))) \\ &= r(x, y),\end{aligned}$$

which completes the proof of Theorem 2.1. □

Proof of Corollary 2.3. (i) Inspection of the proof shows that C1 only is used (after transformation to uniform variables) in Lemma 6.2, and there only in the form corresponding to C1.1 for uniform variables. Hence we may as well just directly assume that C1.1 holds.

Author's personal copy

It is at once seen that C1.3 implies C1.2. We next show that the result of Theorem 2.1 holds also if C1 is replaced by C1.2. However, again, in the proof C1 is only used to establish $\bar{\mathrm{C}}1$, which in turn is only used in the proof of Lemma 6.2. Accordingly, we will prove that the result of Lemma 6.2 holds also under C1.2, transformed to uniform variables.

With notation as in the proof of Theorem 2.1, in particular with $\eta_i = F(\xi_i)$, $v_n = \bar{F}(u_n)$, we have, like for (6.10), that C1.2 implies that for $\bar{\theta} \in (0, 1)$ there is a c with

$$E\left\{\left(\sum_{-r_n}^{r_n} 1_{\{1-v_n+v_nx<\eta_i\le 1-v_n+v_nx\}}\right)^p \middle| 1 - v_n + v_nx < \eta_0 \le 1 - v_n + v_ny\right\} \le c,$$

for $0 \le x, y \le \bar{\theta}$. Using a straightforward stationarity argument, and with notation as in Lemma 6.2, we have that

$$\begin{aligned}
E\{\bar{N}_n(x, y)^\zeta\} &\le \sum_{i=1}^{r_n} E\{\bar{N}_n(x, y)^\zeta 1_{\{1-v_n+v_nx<\eta_i\le 1-v_n+v_ny\}}\} \\
&\le r_n E\left\{\left(\sum_{-r_n}^{r_n} 1_{\{1-v_n+v_nx<\eta_i\le 1-v_n+v_ny\}}\right)^\zeta 1_{\{1-v_n+v_nx<\eta_0\le 1-v_n+v_ny\}}\right\} \\
&= r_n^{1+\zeta-p} v_n|y-x| E\left\{\left(\sum_{-r_n}^{r_n} 1_{\{1-v_n+v_nx<\eta_i\le 1-v_n+v_ny\}}\right)^p \right. \\
&\qquad\qquad \left. \middle| 1 - v_n + v_nx < \eta_0 \le 1 - v_n + v_ny\right\} \\
&\le r v_n^{1+\zeta-p}|y-x|c.
\end{aligned}$$

Since this is the same bound as in (6.2) the result of Lemma 6.2 holds also if C1 is replaced by C1.2, as required.

Now instead assume that C1.4 is satisfied. By considerations similar to those above we have to bound $E\{\bar{N}_n(x, y)^\zeta\}$ using that for $0 \le z \le \bar{\theta} \in (0, 1)$,

$$E\{\bar{N}_n(x, y)|\eta_i = 1 - v_n + v_nz\} \le c. \tag{6.12}$$

Let $\nu = \min\{i \ge 1; \eta_i \in (1 - v_n + v_nx, 1 - v_n + v_ny]\}$ and write $S(k, \ell) = \sum_{i=k}^{\ell} 1_{\{1-v_n+v_nx<\eta_i\le 1-v_n+v_ny\}}$. Then, using the Markov property for the second step, non-negativity for the third step, and stationarity and (6.12) for the last step,

$$\begin{aligned}
E\{\bar{N}(x, y)^\zeta\} &= E\{S(\nu, r_n)^\zeta 1_{\{\nu\le r_n\}}\} \\
&= E\{E\{S(\nu, r_n)^\zeta|\eta_\nu\}1_{\{\nu\le r_n\}}\} \\
&\le E\{E\{S(\nu, \nu + r_n - 1)^\zeta|\eta_\nu\}1_{\{\nu\le r_n\}}\} \\
&\le r_n^{\zeta-p} E\{E\{S(\nu, \nu + r_n - 1)^p|\eta_\nu\}1_{\{\nu\le r_n\}}\} \\
&\le r_n^{\zeta-p} cE\{1_{\{\nu\le r_n\}}\} \le r_n^{1+\zeta-p} v_n(y-x)c,
\end{aligned}$$

which again establishes (6.2).

(ii) If $r_n\bar{F}(u_n) \to 0$ then

$$\begin{aligned}
&(r_n\bar{F}(u_n))^{-1} E\left(\sum_{i=1}^{r_n} 1_{\{\xi_i>u_n+x\sigma_n\}}\right) E\left(\sum_{i=1}^{r_n} 1_{\{\xi_i>u_n+y\sigma_n\}}\right) \\
&\quad \le (r_n\bar{F}(u_n))^{-1}(r_n\bar{F}(u_n))^2 = r_n\bar{F}(u_n) \to 0.
\end{aligned}$$

Author's personal copy

Hence, in C3 the left hand side may then be replaced by

$$(r_n \bar{F}(u_n))^{-1} E\left\{\sum_{i=1}^{r_n} 1_{\{\xi_i > u_n + y\sigma_n\}} \sum_{i=1}^{r_n} 1_{\{\xi_i > u_n + y\sigma_n\}}\right\}.$$

By a standard use of linearity and stationarity, this is seen to be equal to the left hand side of C3.1. □

Proof of Theorem 3.1. By Skorokhod's representation theorem we may replace convergence in distribution with almost sure convergence, and we may in fact also disregard the "almost", and assume that

$$e(\tilde{T}_n(x)) \to e(x), \quad n \to \infty, \tag{6.13}$$

uniformly on compact subintervals of $[-\epsilon, x_T)$, for $\epsilon > 0$ sufficiently small and as before with x_T the right hand endpoint of the support of T. We first show that this implies that

$$\sqrt{c_n} T_n'(0)(\xi_{[c_n]} - u_n)/\sigma_n \to -e(0). \tag{6.14}$$

Let $\tilde{T}_n^{\leftarrow}(y) = \sup\{x; \tilde{T}(x) > y\}$ be the right continuous inverse of $\tilde{T}_n$ and note that for all sufficiently large n, $T_n^{\leftarrow}(x)$ is well defined for $x \in [-\epsilon', \epsilon']$, for some suitable $\epsilon' > 0$, by (3.4). Theorem 3.1 of [45] and (6.13) imply that

$$\sqrt{c_n}(\tilde{T}_n(T_n^{\leftarrow}(x)) - x) \to e(T^{\leftarrow}(x)), \quad n \to \infty,$$

uniformly in $[-\epsilon', \epsilon']$. By a straightforward translation and extension argument, Lemma 1 of [44] can then be seen to imply that

$$\begin{aligned}\sqrt{c_n}(T_n(\tilde{T}_n^{\leftarrow}(x)) - x) &= \sqrt{c_n}((\tilde{T}_n \circ T_n^{\leftarrow})^{\leftarrow}(x) - x) \\ &\to e(T^{\leftarrow}(x)),\end{aligned} \tag{6.15}$$

uniformly on closed subintervals of $(-\epsilon', \epsilon')$.

It follows from (6.13) that $\tilde{T}_n^{\leftarrow}(1) - T_n^{\leftarrow}(1) \to 0$, and hence, by (3.4),

$$T_n(\tilde{T}_n^{\leftarrow}(1)) - 1 = T_n'(T_n^{\leftarrow}(1))(\tilde{T}_n^{\leftarrow}(1) - T_n^{\leftarrow}(1))(1 + o(1)), \quad n \to \infty.$$

Since $T_n^{\leftarrow}(1) = 0$, and since $\tilde{T}_n^{\leftarrow}(1) = (\xi_{[c_n]} - u_n)/\sigma_n$ by the definitions, this and (6.15) prove (6.14).

Next, again by the definitions,

$$e(\hat{T}_n(x)) = e(\tilde{T}_n)(x + (\xi_{[c_n]} - u_n)/\sigma_n) + \sqrt{c_n}\frac{\bar{F}(\xi_{[c_n]} + x\sigma_n)}{\bar{F}(\xi_{[c_n]})}\frac{\bar{F}(\xi_{[c_n]}) - \bar{F}(u_n)}{\bar{F}(u_n)} \tag{6.16}$$

Together, (6.13) and (6.14) ensure that

$$e(\tilde{T}_n)(x + (\xi_{[c_n]} - u_n)/\sigma_n) \to e(x), \quad n \to \infty,$$

uniformly on compact subintervals of $[0, x_T)$. Further, (3.4), (6.14) and $c_n \to \infty$, $c_n/n \to 0$ imply that

$$\frac{\bar{F}(\xi_{[c_n]} + x\sigma_n)}{\bar{F}(\xi_{[c_n]})} \to T(x), \quad n \to \infty, \tag{6.17}$$

Author's personal copy

for $x \geq 0$, and that

$$\frac{\bar{F}(\xi_{[c_n]}) - \bar{F}(u_n)}{\bar{F}(u_n)} = T_n'(0)(\xi_{[c_n]} - u_n)(1 + o(1))/\sigma_n. \tag{6.18}$$

It follows from (6.16)–(6.18) and (6.14) that

$$e(\hat{T}_n)(x) \to e(x) - T(x)e(0),$$

uniformly on compact subintervals of $[0, \infty)$. □

7. Epilogue: Other approaches

Although absolute regularity and strong mixing are widely applicable, there exist natural models where they are known not to hold or where they are not yet proved. Hence considerable effort has been spent on limit theory which uses weaker dependence restrictions; see e.g. [18, 10,11,30] and the references therein. It would be interesting to investigate how these ideas apply to tail empirical processes. However, this might require substantial new research, because of the different triangular array setup and multidimensional convergence here, and because summing indicator functions interferes with the linear structure exploited in some of the cited work.

As a final comment, it seems probable that it wouldn't require any substantially new ideas to extend the results to vector-valued ξ's, and that C3 is the only condition which would have to be strengthened for this.

Acknowledgements

I want to thank Holger Drees for many helpful suggestions, and for help with one ingredient in the proof of Theorem 5.1. Further I would like to thank two anonymous referees for useful comments — in particular the use of Rio's results for the strong-mixing case is in response to comments from one of the referees. This research was supported in part by the Swedish Foundation for Strategic Research.

Appendix. An example which satisfies C2, C3, and C4, but where the conclusion of Theorem 2.1 does not hold

The example uses the simplest possible discrete renewal process to group the integers into "cycles". Then, for each cycle, with probability 1/2 all the ξ_i's in the cycle are equal and equal to the cycle length, and with probability 1/2 they are i.i.d.

Formally, let $\{U_i\}_{-\infty}^{\infty}$ be i.i.d. r.v.'s with $P(U_i = 0) = P(U_i = 1) = 1/2$, let $\ldots, T_{-1} < T_0 \leq 0 < T_1 < T_2 < \cdots$ be the values of i where $U_i = 1$, let $T_i + 1, \ldots, T_{i+1}$ be the i-th cycle, and let $C_i = T_{i+1} - T_i$ be its length. Then, by straightforward computation,

$$P(C_i \geq x) = P(T_1 \geq x) = \frac{1}{2^{x-1}}, \quad x \geq 1, i \neq 0, \quad \text{and} \quad P(C_0 = x) = x\frac{1}{2^{x+1}}. \tag{A.1}$$

Next, let $\xi_i' = C_k$ for $T_k < i \leq T_{k+1}$, $k = 0, \pm 1, \ldots$, let $\{\xi_i''\}$ be continuous i.i.d. r.v.'s with $P(\xi_i'' > x) = 2x^2 2^{-x}$ for all x greater than some suitable x_0, let $\{I_k\}_{-\infty}^{\infty}$ be i.i.d. with $P(I_k = 1) = P(I_k = 0) = 1/2$ and define

$$\xi_i = I_k \xi_i' + (1 - I_k)\xi_i'' \quad \text{for } T_k \leq i < T_{k+1},\ k = 0, \pm, \ldots.$$

Author's personal copy

Thus, with probability 1/2 all values in a cycle are equal, and equal to the cycle length, and with probability 1/2 the values in the cycle are i.i.d.; and cycles are mutually independent. Then $\{\xi_i\}_{-\infty}^{\infty}$ is stationary, and (for $x > x_0$)

$$\begin{aligned}\bar{F}(x) &= P(\xi_0 > x) = \frac{1}{2}P(\xi_0' > x) + \frac{1}{2}P(\xi_0'' > x)\\ &= \frac{1}{2}P(C_0 > x) + \frac{1}{2}2x^2\frac{1}{2^x} = \frac{1}{2}\sum_{i=x+1}^{\infty} i\frac{1}{2^{i+1}} + x^2\frac{1}{2^x}\\ &\sim x^2\frac{1}{2^x}, \quad x \to \infty.\end{aligned} \tag{A.2}$$

Choose $0 < \ell < r < c\log 2 < 1$ and $\epsilon > 0$ such that $r < (1/2 - \epsilon)(1 - c\log 2)$ and put

$$u_n = [c\log n], \qquad r_n = [n^r], \qquad \ell_n = [n^\ell],$$

and set $\sigma_n = 1$ for $n = 1$.

We claim that the process $\{\xi_i\}$ satisfies all the conditions of Theorem 2.1, for $p = 2$, except C1, but the limit of $e(\tilde{T}_n)$ is not continuous. An easy modification would give an example where $\{e(\tilde{T}_n)\}_{n=1}^{\infty}$ in addition is not tight in the Skorokhod J_1 topology. This is left to the reader.

The following statements (to be proved below) substantiate the claim:

$$\begin{aligned}&\{\xi_i\} \text{ is beta-mixing at an exponential rate, i.e.,}\\ &\beta_n = O(\mathrm{e}^{-\beta n}) \quad \text{for some } \beta > 0.\end{aligned} \tag{A.3}$$

Further, setting $\gamma = 0, \sigma = 1/\log 2$, the conditions (1.1), (2.1), C3.1 and $r_n\bar{F}(u_n) \to 0$ (and hence also C3), and C4 hold, with

$$r(x, y) = \frac{1}{2^{x\vee y}} + \frac{1}{2^{\lceil x\vee y\rceil+1}}, \quad x, y \geq 0,$$

where $\lceil a\rceil$ denotes the smallest integer which is not strictly less than a. Finally,

$$e(\tilde{T}_n)(1-) - e(\tilde{T}_n)(1) \overset{P}{\nrightarrow} 0. \tag{A.4}$$

Sketch of Proof. The claim (A.3) follows by the same argument as in the proof of Theorem 6.1 of [4] since the C_i's have finite exponential moments.

Next, (1.1) easily follows from (A.2). Since $0 < r < c\log 2 < 1$, and defining $\theta_n \in [0, 1]$ by $u_n = c\log n - \theta_n$, we also get from (A.2) that

$$\bar{F}(u_n) \sim u_n^2\frac{1}{2^{u_n}} \sim c^2(\log n)^2 n^{-c\log 2}2^{\theta_n}, \tag{A.5}$$

and it then is straightforward to check (2.1). Moreover,

$$r_n(n\bar{F}(u_n))^{-1/2+\epsilon} = 0(n^r((\log n)^2 n^{1-c\log 2})^{-1/2+\epsilon}) \to 0, \quad n \to \infty,$$

and C2 and $r_n\bar{F}(u_n) \to 0$ are similarly seen to hold.

It is a bit more involved to establish C3.1. For $x, y \geq 0,\ i \geq 1$,

$$\begin{aligned}P(\xi_0 > u_n + x,\ \xi_i > u_n + y) \ =\ & \frac{1}{2}P(\xi_0' > u_n + x, \xi_i' > u_n + y, T_1 > i)\\ &+ \frac{1}{2}P(\xi_0'' > u_n + x, \xi_i'' > u_n + y, T_1 > i)\\ &+ P(\xi_0 > u_n + x, \xi_i > u_n + y, T_1 \leq i).\end{aligned} \tag{A.6}$$

Author's personal copy

Here, since u_n is an integer,

$$P(\xi_0' > u_n + x, \xi_i' > u_n + y, T_1 > i) = \sum_{k=i+1}^{\infty} P(\xi_0 > u_n + \lceil x \vee y \rceil | T_1 = k) P(T_1 = k). \tag{A.7}$$

Write $(a)_+$ for the positive part of a and set $u' = u_n + \lceil x \vee y \rceil + 1$. By (A.1) and the definition of the ξ_i, the last expression in (A.7) equals

$$\begin{aligned}\sum_{k=i+1}^{\infty} P(T_0 \le -(u' - k)) P(T_1 = k) &= \sum_{k=i+1}^{\infty} \left\{ \frac{1}{2^{u'-k}} 1_{\{k \le u'\}} + 1_{\{k > u'\}} \right\} \frac{1}{2^k} \\ &= \frac{1}{2^{u'}} (u' - i)_+ + \frac{1}{2^{u' \vee i}}.\end{aligned}$$

Thus, since $u'/r_n \to 0$, and $u' \to \infty$, as $n \to \infty$,

$$\begin{aligned}&\sum_{i=1}^{r_n - 1} \left(1 - \frac{i}{r_n}\right) P(\xi_0' > u_n + x, \xi_i' > u_n + y, T_1 > i) \\ &\quad = \sum_{i=1}^{r_n} \left(1 - \frac{i}{r_n}\right) \left\{ \frac{1}{2^{u'}} (u' - i)_+ + \frac{1}{2^{u' \vee i}} \right\} \sim \frac{1}{2} (u')^2 \frac{1}{2^{u'}}, \quad n \to \infty.\end{aligned} \tag{A.8}$$

By definition $P(\xi'' > u) \le 2\bar{F}(u_n)$ and hence, since $r_n \bar{F}(u_n) \to 0$,

$$\begin{aligned}&\bar{F}(u_n)^{-1} \sum_{i=1}^{r_n - 1} \left(1 - \frac{i}{r_n}\right) P(\xi_0'' > u_n + x, \xi_i'' > u_n + y, T_1 > i) \\ &\quad \le \bar{F}(u_n)^{-1} r_n P(\xi_0'' > u_n)^2 \to 0, \quad n \to \infty.\end{aligned} \tag{A.9}$$

Since $P(\xi_0 > u_n + x, \xi_i > u_n + x, T_1 \le i) \le \bar{F}(u_n)^2$, also

$$\bar{F}(u_n)^{-1} \sum_{i=1}^{r_n - 1} \left(1 - \frac{i}{r_n}\right) P(\xi_0 > u_n + x, \xi_i > u_n + x, T_1 \le i) \to 0. \tag{A.10}$$

The same arguments show that (A.8)–(A.10) hold also if the sum from 1 to $r_n - 1$ is replaced by a sum from $-(r_n - 1)$ to -1. Since in addition

$$P(\xi_0 > u_n + x, \xi_0 > u_n + y) = \bar{F}(u_n + x \vee y)$$

it follows from (A.6), (A.8)–(A.10) and (A.2)that

$$\begin{aligned}&\bar{F}(u_n)^{-1} \sum_{|i| < r_N} \left(1 - \frac{i}{r_n}\right) P(\xi_0 > u_n + x, \xi_i > u_n + y) \\ &\quad \sim \bar{F}(u_n)^{-1} \left\{ \bar{F}(u_n + x \vee y) + (u')^2 \frac{1}{2^{u'}} \right\} \\ &\quad \to \frac{1}{2^{x \vee y}} + \frac{1}{2^{\lceil x \vee y \rceil + 1}}, \quad n \to \infty,\end{aligned} \tag{A.11}$$

and hence C3.1 holds.

To prove (A.4), let $\eta_i = 1$ if i starts a sequence of $u_n + 1$ of ξ's which all equal $u_n + 1$, i.e. $\eta_i = 1$ if $\xi_{i-1}' \ne \xi_i' = \cdots = \xi_{i+u_n}' = u_n + 1$ and the corresponding $I_k = 1$, and let $\eta_i = 0$

Author's personal copy

otherwise. Then, writing $\Delta_n = n\bar{F}(u_n)\{\tilde{T}_n(1-) - \tilde{T}_n(1)\} = \sum_{i=1}^n 1_{\{\xi_i = u_n+1\}}$ and $S_n = \sum_{i=1}^n \eta_i$,

$$|\Delta_n - (u_n+1)S_n| \overset{\text{a.s.}}{\leq} u_n. \tag{A.12}$$

By straightforward computations, $E(\eta_i) = \frac{1}{2^{u_n+3}}$ and $V(\sum_{i=1}^{r_n} \eta_i) \sim r_n E\eta_1$. Using (A.3) and reasoning as for C3.1, S_n may be seen to satisfy the last set of conditions of Corollary 4.2 of [41], and hence

$$\frac{S_n - nE\eta_1}{\sqrt{nE\eta_1}} \xrightarrow{d} N(0,1), \quad n \to \infty. \tag{A.13}$$

Now, by (A.13) and $u_n/\sqrt{n\bar{F}(u_n)} \to 0$, and using that F is continuous, it follows that

$$\begin{aligned} e(\tilde{T}_n)(1-) - e(\tilde{T}_n(1)) &= \frac{\Delta_n}{\sqrt{n\bar{F}(u_n)}} \\ &= \frac{(u_n+1)(S_n - nE\eta_1)}{\sqrt{n\bar{F}(u_n)}} + E(\eta_1)\sqrt{\frac{n}{\bar{F}(u_n)}} + o(1) \\ &= \frac{S_n - nE\eta_1}{\sqrt{nE\eta_1}}\sqrt{\frac{(u_n+1)^2 E\eta_1}{\bar{F}(u_n)}} + m_n + 0(1). \end{aligned}$$

Since $(u_n+1)^2 E\eta_1/\bar{F}(u_n) \to 2^{-2}$, (A.4) follows. □

References

[1] M.A. Arcones, B. Yu, Central limit theorems for empirical and U-processes of stationary mixing sequences, J. Theoret. Probab. 7 (1994) 47–71.

[2] J. Beirlant, J. Segers, J. Teugels, Statistics of Extremes, Theory and Applications, Wiley, Chichester, 2005.

[3] H. Berbee, Random Walks with Stationary Increments and Renewal Theory Mathematical Centre Tracts, vol. 112, Mathematisch Centrum, Amsterdam, 1979.

[4] H. Berbee, Convergence rates in the strong law for bounded mixing sequences, Probab. Theory Related Fields 74 (1987) 255–270.

[5] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

[6] R.C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions, Probab. Surv. 2 (2005) 107–144.

[7] S.G. Coles, An Introduction to Statistical Modeling of Extreme Values, Springer, London, 2001.

[8] M. Csörgö, S. Csörgö, L. Horváth, D.M. Mason, Weighted empirical and quantile processes, Ann. Probab. 14 (1986) 31–85.

[9] S. Datta, W.P. McCormick, Inference for the tail parameters of a linear process with heavy tail innovations, Ann. Inst. Statist. Math. 50 (1998) 237–359.

[10] J. Dedecker, C. Prieur, New dependence coefficients. Examples and applications to statistics, Probab. Theory Related Fields 132 (2005) 203–236.

[11] J. Dedecker, P. Doukhan, A new covariance inequality and applications, Stochastic Process. Appl. 106 (2003) 63–80.

[12] P. Deheuvels, D.M. Mason, Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes, Ann. Probab. 18 (1990) 1693–1722.

[13] H. Dehling, T. Mikosch, M. Sörensen (Eds.), Empirical Process Techniques for Dependent Data, Birkhäuser, Boston, 2002.

[14] H. Drees, On smooth tail functionals, Scand J. Statist. 25 (1998) 187–210.

[15] H. Drees, Weighted approximations of tail processes for β-mixing random variables, Ann. Appl. Probab. 10 (2000) 1274–1301.

[16] H. Drees, Tail empirical processes under mixing conditions, in: H.G. Dehling, T. Mikosch, M. Sorensen (Eds.), Empirical Process Techniques for Dependent Data, Birkhäuser, Boston, 2002, pp. 325–342.

[17] H. Drees, Extreme quantile estimation for dependent data with applications to finance, Bernoulli 9 (2003) 617–657.

Author's personal copy

[18] P. Doukhan, S. Louhici, A new weak dependence condition and applications to moment inequalities, Stoch. Proc. Appl. 84 (1999) 313–342.
[19] P. Doukhan, P. Massart, E. Rio, Invariance principles for absolutely regular empirical processes, Ann. Inst. Henri Poincaré 31 (1995) 393–427.
[20] E. Eberlein, Weak convergence of partial sums of absolutely regular sequences, Statist. Probab. Lett. 2 (1984) 291–293.
[21] J.H.J. Einmahl, Limit theorems for tail processes with application to intermediate quantile estimation, J. Statist. Plan. Inference 32 (1990) 137–145.
[22] J.H.J. Einmahl, The empirical distribution function as a tail estimator, Statist. Neerlandica 44 (1990) 79–82.
[23] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events, Springer, New York, 1997.
[24] J.B. Hill, On tail index estimation using heterogeneous, dependent data. Working paper, Dept. of Economics, Florida International university, 2006.
[25] T. Hsing, On tail estimation using dependent data, Ann. Statist. 19 (1991) 1547–1569.
[26] S. Kotz, S. Nadarajah, Extreme Value Distributions: Theory and Applications, Imperial College Press, London, 2000.
[27] M. Kowaka, An Introduction to Life Prediction of Plant Materials. Application of Extreme Value Statistical Methods for Corrosion Analysis, Allerton Press, New York, 1994.
[28] M.R. Leadbetter, G. Lindgren, H. Rootzén, Extremes and Related Properties of Random Sequences and Processes, Springer, New York, 1983.
[29] D.M. Mason, A strong invariance principle for the tail empirical process, Ann. Inst. Henri Poincaré 24 (1988) 491–506.
[30] F. Merlevède, M. Peligrad, Recent advances in invariance principles for stationary sequences, Probab. Surv. 3 (2006) 136.
[31] S.Y. Novak, Inference on heavy tails from dependent data, Siberian Adv. Math. 12 (2002) 73–96.
[32] V.V. Petrov, Limit Theorems of Probability Theory, Sequences of Independent Variables, Oxford University Press, Oxford, 1995.
[33] D. Pollard, Convergence of Stochastic Processes, Springer, New York, 1984.
[34] R. Reiss, M. Thomas, Statistical Analysis of Extreme Values (for Insurance, Finance, Hydrology and Other Fields), 3rd revised edn, Birkhäuser, Basel, 2005.
[35] S. Resnick, C. Stărică, Consistency of Hill's estimator for dependent data, J. Appl. Probab. 32 (1995) 139–167.
[36] S. Resnick, C. Stărică, Asymptotic behavior of the Hill estimator for autoregressive data, Commun. Statist. - Stochastic Models 13 (1997) 703–731.
[37] S. Resnick, C. Stărică, Tail estimation for dependent data, Ann. Appl. Probab. 8 (1998) 1156–1183.
[38] E. Rio, Théorie Asymptotique des Processus Aléatoires Faiblement Dependents, Springer series Mathematiques & Applications, vol. 31, Paris, 2000.
[39] H. Rootzén, The tail empirical process for stationary sequences. Technical report, Department of Mathematics, Chalmers University, Sweden, 1995.
[40] H. Rootzén, M.R. Leadbetter, L. de Haan, Tail and quantile estimation for strongly mixing stationary processes. Report, Department of Statistics, University of North Carolina, 1992.
[41] H. Rootzén, M.R. Leadbetter, L. de Haan, On the distribution of tail array sums for strongly mixing stationary sequences, Ann. Appl. Probab. 8 (1998) 868–885.
[42] Q.M. Shao, H. Yu, Weak convergence for weighted empirical processes of dependent sequences, Ann. Probab. 24 (1996) 2098–2127.
[43] A. Stephenson, E. Gilleland, Software for the analysis of extreme events: the current status and future directions, Extremes 8 (3) (2005).
[44] W. Vervaat, Functional central limit theorems for processes with positive drift and their inverses, Z. Wahrscheinlichkeitsteorie verw. Geb. 23 (1972) 245–253.
[45] W. Whitt, Some useful functions for functional limit theorems, Math. Oper. Res. 5 (1980) 67–85.
[46] V.A. Volkonski, Yu.A. Rozanov, Some limit theorems for random functions, I and II, Theory Probab. Appl. 4 (1959) 178–197; 6 (1961) 186–198.