LIMIT THEOREMS FOR EMPIRICAL PROCESSES OF CLUSTER FUNCTIONALS[1]

By Holger Drees and Holger Rootzén

University of Hamburg, and Chalmers University and Gothenburg University

Let $(X_{n,i})_{1 \le i \le n, n \in \mathbb{N}}$ be a triangular array of row-wise stationary $\mathbb{R}^d$-valued random variables. We use a "blocks method" to define clusters of extreme values: the rows of $(X_{n,i})$ are divided into m_n blocks $(Y_{n,j})$, and if a block contains at least one extreme value, the block is considered to contain a cluster. The cluster starts at the first extreme value in the block and ends at the last one. The main results are uniform central limit theorems for empirical processes $Z_n(f) := \frac{1}{\sqrt{n v_n}} \sum_{j=1}^{m_n} (f(Y_{n,j}) - Ef(Y_{n,j}))$, for $v_n = P\{X_{n,i} \neq 0\}$ and f belonging to classes of cluster functionals, that is, functions of the blocks $Y_{n,j}$ which only depend on the cluster values and which are equal to 0 if $Y_{n,j}$ does not contain a cluster. Conditions for finite-dimensional convergence include β-mixing, suitable Lindeberg conditions and convergence of covariances. To obtain full uniform convergence, we use either "bracketing entropy" or bounds on covering numbers with respect to a random semi-metric. The latter makes it possible to bring the powerful Vapnik–Červonenkis theory to bear. Applications include multivariate tail empirical processes and empirical processes of cluster values and of order statistics in clusters. Although our main field of applications is the analysis of extreme values, the theory can be applied more generally to rare events occurring, for example, in nonparametric curve estimation.

1. Introduction. The next challenge for extreme value statistics is modeling and estimation of the structure of clusters of extreme values. As one concrete example, the Europe 2003 heat wave may have killed around 60,000 persons. There has been a substantial discussion of whether it could be attributed to global warming. The Nature paper [Stott, Stone and Allen (2004)] uses extreme value methods with average summer temperature as a proxy for a heat wave to try to answer this question. However, the health effects are in reality linked to clusters of extremely high temperatures over much shorter time periods, and the fluctuations of temperature during this period determine risks.

Similarly, river flooding may be caused by not just one extreme rainfall event, but also by the ground already being saturated with water due to high precipitation

during the preceding 5–10 days. This was, for example, the case for the large flood which occurred in Northern Sweden on July 26, 2000. Thus, again, an entire sequence of large values are at the center of interest.

This paper develops an empirical limit theory for clusters of extremes in stationary sequences. It provides a unified basis for asymptotic analysis of statistical methods which aim at answering questions such as the ones above. Results include limit theorems for tail array sums, in particular, for multivariate tail empirical processes, and for joint survival functions of the values and order statistics in a cluster. More special examples, such as upcrossings, compound insurance claims, kernel density and bootstrap estimators, are also studied.

Estimation of the extremal index (roughly, the inverse of the expected clusters length) has received substantial attention in the extreme value statistics literature. The results of this paper can be used to prove asymptotic normality for a general type of estimator based on blocks of exceedances; see Drees (2010). There are also a few papers [e.g., Bortot and Tawn (1998), Sisson and Coles (2003)] on Markov chain modeling of clusters of extreme values. However, a major part of the work to develop useful statistical methods for the structure of clusters of extremes still remains to be done. Our goal is that this paper will be useful for the analysis of existing methods, and that it will spur development of new methods.

More specifically, we consider triangular arrays of row-wise stationary sequences of random variables. The variables are assumed to take their values in some set $E \subset \mathbb{R}^d$, with $E = \mathbb{R}$ and $E = \mathbb{R}^d$ as the standard examples. Clusters of extremes are defined through a "blocks" method. The variables in each row of the array are divided up into blocks, and a cluster of extremes starts with the first "extreme" value in a block, if there is such a value, and ends with the last one. Such a cluster is termed the "core" of the block. A function which maps a block into a real number is called a "cluster functional" if it only depends on the core of the block and if it equals 0 for blocks without extremes. In contrast to standard uniform central limit theorems, cores (i.e., clusters of extremes) consist of a random number of variables, and, hence, cluster functionals have to be defined on a space of vectors of arbitrary lengths.

The aim is to prove uniform central limit theorems for interesting classes of cluster functionals. We throughout use β-mixing (or, with another name, absolute regularity) as the basic dependence restriction. It is very widely applicable and makes it possible to transfer calculations from dependent blocks to easier calculations with independent blocks. Finite-dimensional convergence of the cluster functionals in addition requires Lindeberg conditions and convergence of covariances. We use suitable formulations of "bracketing entropy" to give conditions for asymptotic tightness, and bounds on covering numbers with respect to a random semimetric to prove asymptotic equicontinuity. The latter, in particular, makes it possible to use the Vapnik–Červonenkis theory to prove asymptotic equicontinuity. As usual, uniform central limit theorems follow from finite-dimensional convergence together with asymptotic tightness, or together with asymptotic equicontinuity.

In the important context of estimation for panel count data, two articles by Wellner and Zhang (2000, 2007) use uniform central limit theory for vectors of random lengths. These articles are aimed at the specific application and not at general theory. Hence, they use special properties (such as monotonicity) of the classes of functions, do not consider triangular arrays, assume that the vectors are independent, and, in the second paper, also assume that the lengths of the vectors are uniformly bounded. However, the basic tools to prove tightness, that is, random covering numbers for the general case, and bracketing entropy for the uniformly bounded case are the same as in the present paper. We have not found any other references on uniform central limit theory for random vectors with random lengths.

One application of the theory of this paper is to multivariate tail empirical processes for stationary time series. Let $(X_i)_{i \in \mathbb{N}}$ be a time series with marginal survival function $\bar{H} = 1 - H$. The univariate tail empirical process is defined as

$$e_n(x) := \frac{1}{\sqrt{n v_n}} \sum_{i=1}^{n} (\mathbf{1}_{\{X_{n,i} > x\}} - \bar{H}(u_n + a_n x)), \qquad x \in [0, \infty),$$

where

$$(1.1) \qquad X_{n,i} := \left(\frac{X_i - u_n}{a_n} \right)_+ = \max\left(\frac{X_i - u_n}{a_n}, 0 \right), \qquad 1 \le i \le n.$$

The multivariate tail empirical process is defined analogously; see Examples 3.1 and 3.8 below. In the definition $(u_n)_{n \in \mathbb{N}}$ is an increasing sequence of thresholds such that $v_n := P\{X_1 > u_n\} \to 0$, and $(a_n)_{n \in \mathbb{N}}$ is a sequence of positive normalizing constants such that the conditional distribution of $X_{n,1}$ given that $X_{n,1} > 0$ converges weakly to some nondegenerate limit. [In particular, the distribution function (df) of X_1 then belongs to the domain of attraction of some extreme value distribution.] Rootzén (1995, 2009) proved weak convergence of e_n to a Gaussian process; see Example 3.8 for details. Such limit theorems have proved quite useful for semi-parametric statistical analysis of the marginal tail behavior [Drees (2000, 2002, 2003)]. The present paper extends convergence to multivariate tail empirical processes and makes a small improvement of the results in Rootzén (2009).

Tail empirical processes do not capture information on location in the extreme clusters, and hence do not catch the serial extremal dependence structures which are at the center of interest in connection with, for example, heat waves or river floods. A second class of applications of our main theorems is to joint survival functions and joint distributions of the order statistic of the values within an extreme cluster.

The paper is organized as follows. In Section 2 we first introduce empirical processes of cluster functionals. This generalizes concepts first introduced by Yun (2000) and developed further by Segers (2003). We then derive uniform central limit theorems for these empirical processes under quite general abstract conditions. Sections 3 contains applications to tail array sums, with the multivariate tail

empirical process as a prominent example. In Section 4 we consider empirical processes of indicator variables, and, in particular, joint distributions of variables and of the order statistics in the clusters of extreme values. Proofs are given in Section 5.

2. Limit theorems for general empirical cluster processes. This section first sets out the basic definitions and assumptions which are used throughout the paper and then, in Section 2.1, gives conditions for finite-dimensional convergence of the empirical processes $(Z_n(f))_{f \in \mathcal{F}}$ (defined below). The following subsections consider asymptotic tightness and asymptotic equicontinuity of these empirical processes. As usual, finite-dimensional convergence together with either asymptotic tightness or asymptotic equicontinuity gives convergence of Z_n in the space $\ell^\infty(\mathcal{F})$ of bounded functions indexed by $\mathcal{F}$.

For some $d \in \mathbb{N}$, let E be a measurable subset of $\mathbb{R}^d$ containing 0 and let $(X_{n,i})_{1 \le i \le n, n \in \mathbb{N}}$ be a triangular array of row-wise stationary random variables (r.v.'s) with values in E. Typically the $(X_{n,i})$ have been obtained by "renormalization" of some other process, where the renormalization maps all nonextreme values to 0. A generic example (cf. the Introduction) is $E = \mathbb{R}$ and $X_{n,i} = (\frac{X_i - u_n}{a_n})_+$, where $(X_i)_{i \in \mathbb{N}}$ is a stationary univariate time series. Here u_n tends to the right endpoint of the support of X_i, so that $X_{n,i}$ is 0 unless X_i is "large," that is, unless $X_i > u_n$.

The "empirical process Z_n of cluster functionals" is defined as

$$Z_n(f) := \frac{1}{\sqrt{n v_n}} \sum_{j=1}^{m_n} (f(Y_{n,j}) - E f(Y_{n,j})), \qquad f \in \mathcal{F}.$$

Here $Y_{n,j}$ is the jth block of r_n consecutive values of the nth row of $(X_{n,i})$. Thus, there are $m_n := \lfloor n/r_n \rfloor := \max\{j \in \mathbb{N}_0 \mid j \le n/r_n\}$ blocks

$$Y_{n,j} := (X_{n,i})_{(j-1)r_n + 1 \le i \le j r_n}, \qquad 1 \le j \le m_n,$$

of length r_n. We write Y_n for a "generic block" so that $Y_n \overset{d}{=} Y_{n,1}$. The block lengths r_n tend to infinity, but slower than n, and

$$v_n := P\{X_{n,1} \neq 0\} \to 0.$$

Further, $\mathcal{F}$ is a class of "cluster functionals," that is, functions which only depend on the part of the block which contains all nonvanishing observations; see below.

In the univariate case $E = \mathbb{R}$, cluster functionals have been introduced by Yun (2000) and Segers (2003). The definition is as follows:

DEFINITION 2.1. (i) The set $E_\cup := \bigcup_{l \in \mathbb{N}} E^l$ of vectors of arbitrary length is equipped with the σ-field $\mathbb{E}_\cup$ that is induced by the Borel-σ-fields on E^l, $l \in \mathbb{N}$.

(ii) For an arbitrary $k \in \mathbb{N}$ and $x = (x_1, \ldots, x_k) \in E^k$ the *core* $x^c \in E_\cup$ of x is defined by

$$x^c := \begin{cases} (x_l)_{l_1 \leq l \leq l_2}, & \text{if } x \neq (0, \ldots, 0), \\ 0, & \text{otherwise,} \end{cases}$$

where

$$l_1 := \min\{i \in \{1, \ldots, k\} \mid x_i \neq 0\},$$
$$l_2 := \max\{i \in \{1, \ldots, k\} \mid x_i \neq 0\}.$$

The *length of the core* of x is defined as $L(x) := l_2 - l_1 + 1$ if $x^c \neq 0$ and $L(x) = 0$ if $x^c = 0$.

(iii) A measurable map $f : (E_\cup, \mathbb{E}_\cup) \to (\mathbb{R}, \mathbb{B})$ is called a *cluster functional* if $f(x) = f(x^c)$ for all $x \in E_\cup$, and $f(0) = 0$.

Typical examples are functionals of the type

$$f(x_1, \ldots, x_k) := \sum_{l=1}^{k} \phi(x_l),$$

where $\phi : E \to \mathbb{R}$ satisfies $\phi(0) = 0$, which are related to so-called tail array sums, and, in the case $E = [0, \infty)$,

$$f(x_1, \ldots, x_k) := \max_{1 \leq i \leq k} x_i,$$

which corresponds to the (componentwise) maximum of a cluster. Many more examples will be discussed in Sections 3 and 4.

The proofs below will use the well-known "big blocks, small blocks" technique together with a β-mixing condition to boil down convergence to convergence of sums over i.i.d. blocks. The β-mixing coefficients (also called the coefficients of absolute regularity) for $(X_{n,i})_{1 \leq i \leq n}$ are defined by

$$\beta_{n,k} := \sup_{1 \leq l \leq n-k-1} E\left(\sup_{B \in \mathcal{B}_{n,l+k+1}^n} |P(B|\mathcal{B}_{n,1}^l) - P(B)| \right),$$

where $\mathcal{B}_{n,i}^j$ denotes the σ-field generated by $(X_{n,l})_{i \leq l \leq j}$. Since the $X_{n,i}$ take values in a Polish space, the supremum can be taken over a countable set of B's, and hence is measurable. [On general spaces "sup" has to be replaced by "ess-sup," which is defined as a measurable function which is a.s. larger than or equal to $|P(B|\mathcal{B}_{n,1}^l) - P(B)|$ for all $B \in \mathcal{B}_{n,l+k+1}^n$ and a.s. smaller than or equal to all other measurable functions with this property.] In addition to the β-mixing coefficients and the lengths r_n of the big blocks, the "big blocks, small blocks" technique uses an intermediate sequence ℓ_n of integers, the lengths of small blocks which are used to separate the big blocks in the proofs.

Throughout we will use the following basic assumptions:

(B1) The rows $(X_{n,i})_{1 \le i \le n}$ are stationary, $\ell_n = o(r_n), \ell_n \to \infty, r_n = o(n)$, $r_n v_n \to 0, n v_n \to \infty$,
and
(B2) $\beta_{n,\ell_n} \frac{n}{r_n} \to 0$.
Sometimes we will also use the assumption
(B3) $\lim_{m \to \infty} \lim \sup_{n \to \infty} \beta_{n,m} = 0$.

It follows from $r_n v_n \to 0$ that $v_n \to 0$ and hence that nonzero values of $X_{n,i}$ are rare events. The most important example we have in mind are the standardized excesses given in (1.1). However, other examples occur in the context of nonparametric density estimation or nonparametric regression in a natural way (cf. Example 3.5). Since $n v_n$ is the expected number of nonzero values of $(X_{n,i})_{1 \le i \le n}$, the assumption $n v_n \to \infty$ seems necessary if one wants to obtain normally distributed limits.

More specifically, the assumption $r_n v_n \to 0$ means that the probability of a block being nonzero tends to zero. In particular, it implies that if the row variables are i.i.d., then, asymptotically, cores—or, equivalently, clusters of "extremes"—will have length one, as they intuitively should have. To see this, note that if the variables in a row are independent, then asymptotically the number of nonzero values in a block of length r_n has a Poisson distribution with mean $r_n v_n$ and that then the conditional probability that there are more than one nonzero value in a block, given that there is at least one nonzero value, is (approximately) $(1 - e^{-r_n v_n} - r_n v_n e^{-r_n v_n})/(1 - e^{-r_n v_n})$. This tends to zero if and only if $r_n v_n \to 0$.

For a given sequence $(r_n)_{n \in \mathbb{N}}$, assumption (B2) requires a minimum rate at which the mixing coefficients $\beta_{n,l}$ tend to 0 as $l \to \infty$. The condition (B3), for example, holds if the $X_{n,i}$ are obtained by renormalizing a single absolutely regular process.

REMARK 2.2. (i) The proofs of Theorems 2.3 and 2.8, of Lemma 2.5(ii) and (iii), and of Lemma 5.1 below, in fact, do not use the assumption $r_n v_n \to 0$ of (B1), but only that $v_n \to 0$. The same remark applies to Theorem 2.10 if one replaces condition (D5) below by the following slightly stronger version: For all $\delta > 0, n \in \mathbb{N}, l \in \{0, 1\}, (e_i)_{1 \le i \le \lfloor m_n/2 \rfloor + 1} \in \{-1, 0, 1\}^{\lfloor m_n/2 \rfloor + 1}$ and $k \in \{1, 2\}$, the map $\sup_{f,g \in \mathcal{F}, \rho(f,g) < \delta} \sum_{j=1}^{\lfloor m_n/2 \rfloor + l} e_j (f(Y_{n,j}^*) - g(Y_{n,j}^*))^k$ is measurable.

Hence, these results hold also if the assumption $r_n v_n \to 0$ is replaced by the weaker $v_n \to 0$.

(ii) It is not essential that E is a subset of $\mathbb{R}^d$. Indeed, one may assume that $X_{n,i}$ takes on values in an arbitrary set E. Then one chooses some special element $e_0 \in E$ which takes over the role of 0. In this more general setting, a cluster functional is defined as a functional on $\bigcup_{l \in \mathbb{N}} E^l$ whose value is not changed if e_0 is added at the beginning or at the end of some vector in $\bigcup_{l \in \mathbb{N}} E^l$.

2.1. *Convergence of fidis.* We first give a general result on the convergence of the finite-dimensional marginal distributions (fidis), and then introduce simpler, but more restrictive assumptions, which also are sufficient for convergence. Proofs are deferred to Section 5.

We will use the notation $x^{(k)}$ for the vector $(x_1, \ldots, x_k)$ made up by the first k components in the vector x, if x has at least k components, and otherwise $x^{(k)} = x$. Similarly, we write $x^{(\ell;k)} = (x_\ell, \ldots, x_k)$ for the vector consisting of components number ℓ to number k in x, if x has at least k components, and otherwise $x^{(\ell;k)}$ starts at component no. ℓ and ends at the end of x (if x is shorter than ℓ, then $x^{(\ell;k)} = 0$). As before, let $\mathcal{F}$ be a class of cluster functionals, and recall that $Y_n \overset{d}{=} Y_{n,1}$, where $Y_{n,1}$ is the first block in the nth row. For $f \in \mathcal{F}$ write

$$\Delta_n(f) := f(Y_n) - f\big(Y_n^{(r_n - \ell_n)}\big)$$

for the difference between f evaluated at the r_n components of the entire block and f evaluated at the first $r_n - \ell_n$ components of the block. The general "convergence conditions" are as follows:

(C1)
$$E\big((\Delta_n(f) - E\Delta_n(f))^2 \mathbf{1}_{\{|\Delta_n(f) - E\Delta_n(f)| \le \sqrt{nv_n}\}}\big) = o(r_n v_n),$$
$$P\big\{|\Delta_n(f) - E\Delta_n(f)| > \sqrt{nv_n}\big\} = o(r_n/n)$$

for all $f \in \mathcal{F}$.

(C2)
$$E\big((f(Y_n) - Ef(Y_n))^2 \mathbf{1}_{\{|f(Y_n) - Ef(Y_n)| > \varepsilon\sqrt{nv_n}\}}\big) = o(r_n v_n)$$
$$\forall \varepsilon > 0, f \in \mathcal{F}.$$

(C3)
$$\frac{1}{r_n v_n} \text{Cov}(f(Y_n), g(Y_n)) \to c(f, g) \qquad \forall f, g \in \mathcal{F}.$$

The block $Y_n^{(r_n - \ell_n)}$ is obtained from Y_n by omitting a small block of l_n observations at the end. Accordingly, (C1) means that asymptotically this omission does not influence the fidis of the empirical process of cluster functionals (see the proof of Lemma 5.1). By the definition of cluster functionals, this is usually fulfilled if, with high probability, there are few or no nonzero observations in the omitted short blocks. Specifically, if components number $r_n - l_n + 1 \le i \le r_n$ all are zero, then Y_n and $Y_n^{(r_n - \ell_n)}$ have the same core, and, thus, $\Delta_n(f) = 0$.

Assumption (C2) is the standard Lindeberg condition. The assumption of convergence of covariances, (C3), is the final ingredient needed to ensure finite-dimensional convergence in the present triangular array setup.

THEOREM 2.3. *Suppose the basic assumptions* (B1) *and* (B2) *hold, and that* (C1)–(C3) *are satisfied. Then the fidis of the empirical process* $(Z_n(f))_{f \in \mathcal{F}}$ *of cluster functionals converge to the fidis of a Gaussian process* $(Z(f))_{f \in \mathcal{F}}$ *with covariance function* c.

In general, the convergence (C3) of the covariance function must be verified directly. However, we also give additional sufficient conditions which are simpler to verify in some situations. A first very simple version, (C3′), requires convergence only after "truncation" to a fixed (but arbitrary) length. Before stating it, we recall the notation $L(Y_n)$ for the length of the core of Y_n:

(C3′) For $f \in \mathcal{F}$ it holds that

$$(2.1) \qquad \lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{r_n v_n} E\big(f(Y_n)^2 \mathbf{1}_{\{L(Y_n) > k\}}\big) = 0,$$

and for $f, g \in \mathcal{F}$ there is a sequence $R_{n,k}$ with $\lim_{k \to \infty} \limsup_{n \to \infty} |R_{n,k}| = 0$ such that

$$(2.2) \qquad \lim_{n \to \infty} \frac{1}{r_n v_n} E\big(f(Y_n) g(Y_n) \mathbf{1}_{\{L(Y_n) \le k\}}\big) + R_{n,k} = c_k(f, g).$$

A typical situation when (2.1) holds is when the cluster lengths $(L(Y_n))_{n=1}^{\infty}$ are tight under $P(\cdot | Y_n \ne 0)$ and $(f(Y_n)^2)_{n \in \mathbb{N}}$ is uniformly integrable under $P(\cdot \mid Y_n \ne 0)$, for $f \in \mathcal{F}$. This follows from the observation that $\frac{1}{r_n v_n} |E(\cdot)| \le |E(\cdot \mid Y_n \ne 0)|$, which in turn follows from $P(Y_n \ne 0) \le r_n v_n$.

In a second assumption (C3″) we generalize the powerful results of Segers (2003) to the present abstract setting. In doing this, we do not aim at the greatest possible generality, but give versions which suit our purposes best. It may be noted that, unlike in the situation considered by Segers, in general weak convergence of the indicators $\mathbf{1}_{\{0\}}(X_{n,i})$ does not follow from weak convergence of $X_{n,i}$. In the statement of the condition we use that the value of a cluster functional f applied to a sequence $(x_i)_{i \in \mathbb{N}}$ with $m_x := \sup\{i \in \mathbb{N} \mid x_i \ne 0\} < \infty$ can be defined in a natural way as $f((x_i)_{1 \le i \le m_x})$. The conditions are as follows:

(C3″)

(C3.1″) There is a sequence $W = (W_i)_{i \in \mathbb{N}}$ of E-valued r.v.'s such that, for all $k \in \mathbb{N}$, the joint conditional distribution $P^{(X_{n,i}, \mathbf{1}_{\{0\}}(X_{n,i}))_{1 \le i \le k} | X_{n,1} \ne 0}$ converges weakly to $P^{(W_i, \mathbf{1}_{\{0\}}(W_i))_{1 \le i \le k}}$, and all $f \in \mathcal{F}$ are a.s. continuous with respect to the distributions of $W^{(k)}$ and $W^{(2;k)}$, for all k, that is,

$$(2.3) \qquad \begin{aligned} & P\big\{W^{(2;k)} \in D_{f,k-1}, W_i = 0 \ \forall i > k\big\} \\ & \qquad = P\big\{W^{(k)} \in D_{f,k}, W_i = 0 \ \forall i > k\big\} = 0 \end{aligned}$$

with $D_{f,k}$ denoting the set of discontinuity points of $f_{|E^k}$.

(C3.2″) For all $f \in \mathcal{F}$ the sequence $(f(Y_n)^2)_{n \in \mathbb{N}}$ is uniformly integrable under $P(\cdot)/(r_n v_n)$.

Again, (C3.2″) is implied by the perhaps more intuitive condition that $(f(Y_n)^2)_{n \in \mathbb{N}}$ is uniformly integrable under $P(\cdot \mid Y_n \ne 0)$.

In the proof of the next two results we will, in fact, use a slightly weaker (but instead more complicated) version of (2.3); see Remark 2.6 below.

COROLLARY 2.4. *Suppose that* (B1), (B2) *and* (C1) *are satisfied. If, furthermore, either* (C2) *and* (C3′) *or else* (B3) *and* (C3″) *hold, then the fidis of the empirical process* $(Z_n(f))_{f \in \mathcal{F}}$ *of cluster functionals converge to the fidis of a Gaussian process* $(Z(f))_{f \in \mathcal{F}}$. *Specifically,* (C3′) *implies that* (C3) *holds and that the covariance function* c *of* Z *is obtained as*

$$c(f, g) = \lim_{k \to \infty} c_k(f, g).$$

If (C3″) *holds, then*

(2.4) $$c(f, g) = E\big((fg)(W) - (fg)(W^{(2;\infty)})\big).$$

Equation (2.4) is explained in Lemma 2.5 below. It generalizes the most important results of Segers (2003) to the present more abstract setting.

LEMMA 2.5. (i) *If* (B1) *and* (B3) *hold, then*

(2.5) $$E\big(f(Y_n) \mid Y_n \neq 0\big) = \frac{1}{\theta_n} E\big(f(X_n^{(r_n)}) - f(X_n^{(2,r_n)}) \mid X_{n,1} \neq 0\big) + o(1),$$

where the term o(1) *tends to* 0 *as* n *tends to* ∞ *uniformly for all cluster functionals* f *such that* $\|f\|_\infty \leq C$, *for any* $C \in \mathbb{R}$, *and*

$$\theta_n := \frac{P\{Y_n \neq 0\}}{r_n v_n} = P\big(X_n^{(2;r_n)} = 0 \mid X_{n,1} \neq 0\big)\big(1 + o(1)\big).$$

(ii) *If* (B1), (B3) *and the assumption of* (C3.1″) *all are satisfied, then*

(2.6) $$m_W = \sup\{i \geq 1 \mid W_i \neq 0\} < \infty$$

and

$$\lim_{n \to \infty} \theta_n = \theta := P\{W_i = 0 \, \forall i \geq 2\} = P\{m_W = 1\} > 0.$$

(iii) *If* (B1), (B3) *and* (C3.1″) *hold, then the conditional distribution* $P^{f(Y_n)|Y_n \neq 0}$ *converges weakly to the probability measure*

$$\mu_{f,W} := \frac{1}{\theta}\big(P\{f(W) \in \cdot\} - P\{f(W^{(2;\infty)}) \in \cdot, m_W \geq 2\}\big).$$

Note that $\mu_{f,W}(\mathbb{R}) = 1$ by (ii). However, it is not so obvious that $\mu_{f,W}$ is indeed a positive (and hence a probability) measure.

REMARK 2.6. We will prove Corollary 2.4 and Lemma 2.5 under the following weaker version of the continuity assumption (2.3):
For $k \in \mathbb{N}$ and $I \subset \{1, \ldots, k\}$ let $N_{k,I} := \{x \in E^k \mid x_i = 0, \forall i \in I, x_i \neq 0, \forall i \notin I\}$ and denote by $D_{f,k,I}$ the set of discontinuity points of $f|_{N_{k,I}}$. Then we assume

(2.7) $$P\{W^{(k)} \in D_{f,k,I}, W^{(k+1,\infty)} = 0\} = 0 \qquad \forall k \in \mathbb{N}, I \subset \{1, \ldots, k\},$$

(2.8) $$P\{W^{(2;k)} \in D_{f,k-1,I}, W^{(k+1,\infty)} = 0\} = 0 \qquad \forall k \geq 2, I \subset \{1, \ldots, k-1\}.$$

This version can be used in some examples where (2.3) is not satisfied, because the boundary of $[0, \infty)^k$ belongs to the discontinuity sets $D_{f,k}$ and, according to Lemma 2.5(ii), the r.v. W_i equals 0 with positive probability for $i > 1$.

In the situation considered by Segers (2003) [i.e., with $X_{n,i}$ defined by (1.1) for a stationary time series whose finite-dimensional marginal distributions all belong to the domain of attraction of some extreme value distribution], the sequence $(W_i)_{i \in \mathbb{N}}$ is related to the so-called tail sequence (or tail chain) $(U_i)_{i \in \mathbb{N}}$ [cf. Segers (2003), Theorem 2] via $W_i = \max(U_i, 0)$. Then (C3″) is automatically satisfied, for example, for bounded cluster functionals if $D_{f,m}$ is a Lebesgue null subset of $(0, \infty)^m$ for all m and $f \in \mathcal{F}$, because the r.v.'s U_i are continuous.

Further simpler, but more restrictive, sufficient conditions are given in Lemma 5.2 below. In particular, for bounded cluster functionals one obtains the following:

COROLLARY 2.7.   *If* $\|f\|_\infty = \sup_{x \in E_\cup} |f(x)| < \infty$ *for all* $f \in \mathcal{F}$ *and the conditions* (B1), (B2), (B3) *and* (C3.1″) *hold, then the fidis of the empirical process* $(Z_n(f))_{f \in \mathcal{F}}$ *of cluster functionals converge to the fidis of a Gaussian process* $(Z(f))_{f \in \mathcal{F}}$ *with covariance function* c *defined by* (2.4).

2.2. *Asymptotic tightness.*   In this subsection we give conditions which ensure asymptotic tightness of Z_n in the space $\ell^\infty(\mathcal{F})$. As a consequence, uniform central limit theorems for Z_n hold if in addition the conditions of Theorem 2.3 are satisfied. The alternative route via asymptotic equicontinuity is considered in the next subsection.

In general, the supremum of $Z_n(f)$ taken over uncountably many cluster functionals f need not be measurable. Hence, in some instances, one has to work with outer probabilities and expectations, denoted by P^* and E^* in the following; see van der Vaart and Wellner (1996), Section 1.2, for details. The sequence $(Z_n)_{n \in \mathbb{N}}$ is asymptotically tight if to any $\epsilon > 0$ there is a compact set $K \subset \ell^\infty(\mathcal{F})$ such that

$$\limsup_{n \to \infty} P^*(Z_n \notin K^\delta) < \epsilon \qquad \text{for any } \delta > 0.$$

Here K^δ is the set of elements in $\ell^\infty(\mathcal{F})$ which are at most a distance δ away from K.

We will use the assumptions (D1)–(D4) below to prove tightness. The first two assumptions in various ways restrict the sizes of the functions in $\mathcal{F}$. In particular, (D1) ensures that sample paths of Z_n belong to the space $\ell^\infty(\mathcal{F})$ of bounded functions on $\mathcal{F}$. Assumption (D3) is an asymptotic continuity condition on the covariance function which is needed to ensure that the limiting process has continuous sample paths. The most crucial condition, (D4), restricts the complexity of the index set $\mathcal{F}$ via the so-called bracketing entropy. To state this assumption, the following concept is needed.

The *bracketing number* $N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2^n)$ here is defined as the smallest number N_ε such that for each $n \in \mathbb{N}$ there exists a partition $(\mathcal{F}_{n,k}^\varepsilon)_{1 \leq k \leq N_\varepsilon}$ of $\mathcal{F}$ such that

$$(2.9) \qquad E^* \sup_{f,g \in \mathcal{F}_{n,k}^\varepsilon} (f(Y_n) - g(Y_n))^2 \leq \varepsilon^2 r_n v_n \qquad \forall 1 \leq k \leq N_\varepsilon.$$

The assumptions are as follows:

(D1) The index set $\mathcal{F}$ consists of cluster functionals f such that $E(f(Y_n)^2)$ is finite for all $n \geq 1$ and such that the envelope function

$$F(x) := \sup_{f \in \mathcal{F}} |f(x)|$$

is finite for all $x \in E_\cup$.

(D2)

$$E^*\big(F(Y_n)\mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}\big) = o\big(r_n\sqrt{v_n/n}\big) \qquad \forall \varepsilon > 0.$$

(D3) There exists a semi-metric ρ on $\mathcal{F}$ such that $\mathcal{F}$ is totally bounded (i.e., for all $\varepsilon > 0$ the set $\mathcal{F}$ can be covered by finitely many balls with radius ε w.r.t. ρ) such that

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{f,g \in \mathcal{F}, \rho(f,g) < \delta} \frac{1}{r_n v_n} E\big(f(Y_n) - g(Y_n)\big)^2 = 0.$$

(D4)

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \int_0^\delta \sqrt{\log N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2^n)}\, d\varepsilon = 0.$$

THEOREM 2.8. *If the basic assumptions* (B1) *and* (B2) *hold and* (D1)–(D4) *are satisfied, then the process Z_n is asymptotically tight in $\ell^\infty(\mathcal{F})$. If in addition the finite-dimensional distributions converge* [*which, in particular, hold if* (C1)–(C3) *also are satisfied*], *then Z_n converges to a Gaussian process Z with covariance function c.*

We collect a number of comments and variations of the conditions of the theorem in the following remark. In particular, we consider a strengthened version (D2′) of (D2):

(D2′) $$E^*\big(F^2(Y_n)\mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}\big) = o(r_n v_n) \qquad \forall \varepsilon > 0.$$

The proof of part (ii) of the remark is given in Section 5.

REMARK 2.9. (i) *If, for all $\varepsilon > 0$, there exists a partition $(\mathcal{F}_k^\varepsilon)_{1 \leq k \leq N_\varepsilon}$ of $\mathcal{F}$ which does not depend on n and which satisfies*

$$E^* \sup_{f,g \in \mathcal{F}_k^\varepsilon} (f(Y_n) - g(Y_n))^2 \leq \varepsilon^2 r_n v_n \qquad \forall 1 \leq k \leq N_\varepsilon,$$

then (D3) *and* (D4) *can be replaced with the simpler condition*

$$\int_0^\delta \sqrt{\log N_\varepsilon}\, d\varepsilon < \infty$$

for some $\delta > 0$ [*cf. Theorem* 2.11.9 *of* van der Vaart and Wellner (1996)].

(ii) *If $F(Y_n)$ satisfies the Lindeberg condition* (D2′), *then* (C2) *and* (D2) *are satisfied. In particular, this holds if $n v_n \to \infty$ and*

$$(2.10) \qquad\qquad E^* F(Y_n)^{2+\delta} = O(r_n v_n) \qquad \text{for some } \delta > 0.$$

(iii) *Thus*, *if* (B1), (B2), (C3), (D1), (D3) *and* (D4) *hold with a bounded envelope function F, then the empirical processes Z_n converge to a centered Gaussian process with covariance function c.*

2.3. *Asymptotic equicontinuity.*   Like tightness, the asymptotic equicontinuity of Z_n w.r.t. ρ, that is,

$$\forall \varepsilon, \eta > 0 \qquad \exists \delta > 0 : \limsup_{n\to\infty} P^* \left\{ \sup_{f,g\in\mathcal{F},\rho(f,g)<\delta} |Z_n(f) - Z_n(g)| > \varepsilon \right\} < \eta$$

is necessary and sufficient for the convergence of Z_n, provided all fidis of Z_n converge.

To prove asymptotic equicontinuity, we need a technical measurability condition, condition (D5) below, and, crucially, suitable bounds (D6) or (D6′) on the rate of increase of covering numbers. The condition (D5), in particular, is satisfied if the processes $(f(Y_n))_{f\in\mathcal{F}}$ are separable. The condition (D6) is stated in terms of a "random entropy," while (D6′), which implies (D6), is phrased in terms of uniform entropy. To state the assumptions, we need the following definitions: for a given semi-metric d on $\mathcal{F}$, the (random) *covering number* $N(\varepsilon, \mathcal{F}, d)$ is the minimum number of balls with radius ε w.r.t. d needed to cover $\mathcal{F}$. The condition (D6) bounds the rate of increase of $N(\varepsilon, \mathcal{F}, d_n)$ as ε tends to 0 for the random semi-metric

$$d_n(f, g) := \left(\frac{1}{n v_n} \sum_{j=1}^{m_n} (f(Y_{n,j}^*) - g(Y_{n,j}^*))^2 \right)^{1/2},$$

that is, the L_2-semi-metric w.r.t. empirical measure $(n v_n)^{-1} \sum_{j=1}^{m_n} \varepsilon_{Y_{n,j}^*}$, where $Y_{n,j}^*$, $1 \le j \le m_n$, are i.i.d. copies of $Y_{n,1}$. In (D6′) we instead use the supremum of all covering numbers $N(\varepsilon, \mathcal{F}, d_Q)$, where $d_Q(f, g) := (\int (f - g)^2 \, dQ)^{1/2}$ and Q ranges over the set of discrete probability measures $\mathcal{Q}$. With this notation, the conditions are as follows:

(D5) For all $\delta > 0, n \in \mathbb{N}, (e_i)_{1\le i\le \lfloor m_n/2 \rfloor} \in \{-1, 0, 1\}^{\lfloor m_n/2 \rfloor}$ and $k \in \{1, 2\}$, the map $\sup_{f,g\in\mathcal{F},\rho(f,g)<\delta} \sum_{j=1}^{\lfloor m_n/2 \rfloor} e_j (f(Y_{n,j}^*) - g(Y_{n,j}^*))^k$ is measurable.

(D6) $\qquad \lim_{\delta\downarrow 0} \limsup_{n\to\infty} P^* \left\{ \int_0^\delta \sqrt{\log N(\varepsilon, \mathcal{F}, d_n)}\, d\varepsilon > \tau \right\} = 0 \qquad \forall \tau > 0.$

(D6$'$) The envelope function F is measurable with $E(F(Y_n)^2) = O(r_n v_n)$ and

$$\int_0^1 \sup_{Q \in \mathcal{Q}} \sqrt{\log N\left(\varepsilon \left(\int F^2 \, dQ\right)^{1/2}, \mathcal{F}, d_Q\right)} \, d\varepsilon < \infty.$$

THEOREM 2.10. *Suppose the basic assumptions* (B1) *and* (B2) *hold and that* (D1), (D2$'$), (D3) *and* (D5) *are satisfied. Then if also* (D6) [*or, more restrictively,* (D6$'$)] *holds, it follows that Z_n is asymptotically equicontinuous. Further, if in addition the finite-dimensional distributions converge* [*which, in particular, holds if* (C1) *and* (C3) *also are satisfied*], *then Z_n converges to a Gaussian process with covariance function c.*

REMARK 2.11. In view of (D6$'$), one can apply the powerful Vapnik–Červonenkis theory to verify asymptotic equicontinuity. In particular, (D6$'$) is satisfied if $\mathcal{F}$ is a so-called VC-class or, more generally, a VC-hull class. We refer to Section 2.6 of van der Vaart and Wellner (1996) for an outline of the most important uniform bounds on covering numbers $N(\varepsilon(\int F^2 \, dQ)^{1/2}, \mathcal{F}, d_Q)$.

3. Generalized tail array sums. Generalizing the tail empirical process $e_n(x)$ (for some fixed $x \geq 0$), Rootzén, Leadbetter and de Haan (1990) considered so-called tail array sums

$$(3.1) \qquad \qquad \sum_{i=1}^n \phi(X_{n,i})$$

for functions $\phi : \mathbb{R} \to \mathbb{R}$ satisfying $\phi(0) = 0$ and $X_{n,i}$ defined by (1.1); see also Leadbetter and Rootzén (1993), Leadbetter (1995) and Rootzén, Leadbetter and de Haan (1998).

Like the tail empirical process, these tail array sums do not allow inference about the extremal dependence structure, as the summands $\phi(X_{n,i})$ depend on just one observation. However, if $X_{n,i}$ denotes the vector of d consecutive standardized excesses, that is,

$$(3.2) \quad X_{n,i} := \left(\left(\frac{X_i - u_n}{a_n}\right)_+, \left(\frac{X_{i+1} - u_n}{a_n}\right)_+, \ldots, \left(\frac{X_{i+d-1} - u_n}{a_n}\right)_+\right),$$

then the statistic (3.1) with $\phi : (E, \mathcal{B}(E)) \to (\mathbb{R}, \mathbb{B})$ (and $E = \mathbb{R}^d$) contains information on the extremal dependence structure.

Therefore, in the general setting of a row-wise stationary triangular array $(X_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq n}$ used in Section 2, the *generalized* (*standardized*) *tail array sum* (*tail array sum* for short) given by a measurable function $\phi : (E, \mathcal{B}(E)) \to (\mathbb{R}, \mathbb{B})$ with $\phi(0) = 0$ is defined as

$$(3.3) \qquad \qquad \tilde{Z}_n(\phi) := \frac{1}{\sqrt{n v_n}} \sum_{i=1}^n (\phi(X_{n,i}) - E\phi(X_{n,i})).$$

The tail array sum (3.3) can be obtained as the empirical process Z_n evaluated at the cluster functional

$$g_\phi : E_\cup \to \mathbb{R}, \qquad x = (x_1, \ldots, x_k) \mapsto \sum_{i=1}^{k} \phi(x_i),$$

if n is a multiple of r_n. In general,

$$\tilde{Z}_n(\phi) - Z_n(g_\phi) = (nv_n)^{-1/2} \sum_{i=r_n m_n + 1}^{n} (\phi(X_{n,i}) - E\phi(X_{n,i})),$$

which is asymptotically negligible under weak conditions specified in Corollary 3.6 below.

For the remainder of this section, we assume that a family Φ of functions ϕ of the above type is given, and assume it is totally bounded w.r.t. a semi-metric ρ_Φ and has a finite envelope function $\phi_{\max} := \sup_{\phi \in \Phi} |\phi|$.

EXAMPLE 3.1 (Multivariate tail empirical processes).   If $X_{n,i}$ is defined as in (3.2) and $\Phi := \{\mathbf{1}_{(x,\infty)} \mid x \in [0,\infty)^d\}$, then $(Z_n(g_\phi))_{\phi \in \Phi}$ is the (reparametrized) multivariate tail empirical process. In particular, if $d = 1$, then $(Z_n(g_\phi))_{\phi \in \Phi}$ is a reparametrization of the tail empirical process e_n discussed in the Introduction.

For simplicity, we will assume that the X_i are uniformly distributed; the general case can be easily obtained by a marginal quantile transformation [cf. Rootzén (2009) for details]. Then one chooses $a_n = 1 - u_n = v_n$ for a sequence of thresholds u_n tending to 1, so that the conditional distribution of the standardized excesses $X_{n,i} = (X_i - u_n)/a_n$, given that they are strictly positive is also uniform. Thus, it suffices to consider $\Phi := \{\mathbf{1}_{(x,1]} \mid x \in [0,1]^d\}$ with envelope function $\phi_{\max} = \mathbf{1}_{(0,1]^d}$ and metric $\rho_\Phi(\mathbf{1}_{(x,1]}, \mathbf{1}_{(y,1]}) := \max_{1 \le l \le d} |x_l - y_l|, x, y \in [0,1]^d$.

EXAMPLE 3.2 (Upcrossings).   If one is interested in upcrossings of a univariate time series over intervals $[x, y]$, then one may define $X_{n,i}$ as in Example 3.1 with $d = 2$ and consider $\Phi := \{\mathbf{1}_{[0,x) \times (y,1]} \mid x, y \in [0,1], x \le y\}$ with envelope function $\mathbf{1}_{\{(x,y) \in [0,1]^2 | x < y\}}$.

EXAMPLE 3.3 (Compound insurance claim).   If X_i denotes the ith claim of an insurance portfolio with deductible $u_n + a_n t$ and $X_{n,i}$ as in (1.1), then $\phi_t : \mathbb{R} \to [0,\infty)$ given by $\phi_t(x) = (x - t)\mathbf{1}_{(t,\infty)}(x)$ is the standardized total claimed amount. Thus, the empirical process $(Z_n(g_{\phi_t}))_{t \ge 0}$ corresponding to $\Phi := \{(x - t)\mathbf{1}_{(t,\infty)}(x) \mid t \ge 0\}$ describes the influence of the deductible on the random amount the insurance has to pay.

EXAMPLE 3.4 (Bootstrapping the Hill estimator).   A stationary time series $(X_i)_{i \in \mathbb{N}}$ has extreme value index $\gamma > 0$ if its marginal survival function $\bar{F}$ is regularly varying with index $-1/\gamma$, that is, if $\lim_{t \to \infty} \bar{F}(tx)/\bar{F}(t) =$

$x^{-1/\gamma}$. Let $X_{n,i} := X_i/u_n \mathbf{1}_{\{X_i > u_n\}}$, $\phi_1(x) = \log(x)\mathbf{1}_{\{x>1\}}$ and $\phi_2(x) = \mathbf{1}_{\{x>1\}}$ so that $E\phi_2(X_{n,1}) = v_n$ and $\gamma_n = E\phi_1(X_{n,1})/E\phi_2(X_{n,1}) = E\phi_1(X_{n,1})/v_n = E(\log(X_1/u_n)|X_1 > u_n) \to \gamma$ [cf. de Haan and Ferreira (2006), Theorem 1.2.1 and Remark 1.2.3]. Then the Hill estimator $\hat{\gamma}_n$ of γ may be written as

$$(3.4) \qquad \hat{\gamma}_n := \frac{\sum_{i=1}^{n} \log(X_i/u_n)\mathbf{1}_{\{X_i > u_n\}}}{\sum_{i=1}^{n} \mathbf{1}_{\{X_i > u_n\}}} = \frac{\gamma_n + \tilde{Z}_n(\phi_1)/\sqrt{nv_n}}{1 + \tilde{Z}_n(\phi_2)/\sqrt{nv_n}}.$$

Write $g_k := g_{\phi_k}$, $k \in \{1, 2\}$, and suppose we draw independent blocks $Y_i^{(n)}$ from the empirical distribution of $Y_{n,i}$, $1 \le i \le m_n$. Then a bootstrap version of the Hill estimator is obtained as

$$\hat{\gamma}_n^* := \frac{\sum_{i=1}^{m_n} g_1(Y_i^{(n)})}{\sum_{i=1}^{m_n} g_2(Y_i^{(n)})}.$$

EXAMPLE 3.5 (Kernel density estimators). In this simple example we demonstrate that applications of the theory presented in Section 2 are not restricted to extreme value theory. Further examples may be obtained from the literature on "local empirical processes." For the analysis of such processes for i.i.d. data we refer to Einmahl (1997), Giné, Mason and Zaitsev (2003) and Giné and Mason (2008) and to the lists of references in these papers.

Suppose that $(X_i)_{i \in \mathbb{N}}$ is a univariate stationary time series whose marginal df H has a Lebesgue density h. Kernel estimators of the type

$$\hat{h}_n(x_0) := \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{X_i - x_0}{b_n}\right)$$

are probably the most widely used nonparametric estimators for $h(x_0)$ ($x_0 \in \mathbb{R}$). Here K denotes a suitable kernel, for example, a probability density with support $[-1, 1]$, and $(b_n)_{n \in \mathbb{N}}$ is a sequence of bandwidths tending to 0. Let

$$X_{n,i} := \left(2 + \frac{X_i - x_0}{b_n}\right)\mathbf{1}_{[x_0-b_n, x_0+b_n]}(X_i), \qquad 1 \le i \le n,$$

where the constant 2 has been inserted to ensure $X_{n,i} > 0$ for $X_i \in [x_0 - b_n, x_0 + b_n]$. Let $\hat{H}_n$ be the corresponding empirical df. Then integration by parts yields

$$\hat{h}_n(x_0) = \frac{1}{b_n} \int K(y - 2)\hat{H}_n(dy)$$

$$= \frac{1}{b_n} \int (1 - \hat{H}_n(y + 2))K(dy)$$

$$= \frac{1}{nb_n} \int \sum_{i=1}^{n} \mathbf{1}_{(y+2,\infty)}(X_{n,i})K(dy),$$

provided that K has bounded variation. Hence, for $\bar{Z}_n(y) = \tilde{Z}_n(\mathbf{1}_{(y+2,\infty)})$, $y \in [-1, 1]$, and $n = r_n m_n$, we have that

$$\int \bar{Z}_n(y) K(dy) = \sqrt{\frac{n}{v_n}} b_n \big(\hat{h}_n(x_0) - E\hat{h}_n(x_0)\big),$$

where $\sqrt{n/v_n} b_n \sim \sqrt{n/(2h(x_0)b_n)} b_n = \sqrt{nb_n/(2h(x_0))}$ as $n \to \infty$, if h is continuous and positive at x_0. Thus, one obtains the asymptotic normality of $\hat{h}_n(x_0)$ from the convergence of $\bar{Z}_n$ (or $\tilde{Z}_n$) toward a Gaussian process. Indeed, this way it is not difficult to derive normal approximations for $\hat{h}_n$ uniformly over families of kernels with compact support.

To obtain conditions for weak convergence of tail array sums, we first focus on families Φ such that the envelope function $\phi_{\max}$ is bounded, which is true in the Examples 3.1, 3.2 and 3.5, but not in Example 3.3 (unless the support of $X_{n,i}$ is uniformly bounded). We let $\mathcal{F} := \{g_\phi \mid \phi \in \Phi\}$ be equipped with the semi-metric $\rho(g_\phi, g_\psi) = \rho_\Phi(\phi, \psi)$.

COROLLARY 3.6. *Suppose that $\phi_{\max} = \sup_{\phi \in \Phi} |\phi|$ is bounded and measurable, that Φ is totally bounded w.r.t. ρ_Φ, that* (B1) *and* (B2) *hold, and that $r_n = o(\sqrt{nv_n})$. Further assume that*

$$(3.5) \qquad E\left(\sum_{i=1}^{r_n} \mathbf{1}_{\{X_{n,i} \neq 0\}}\right)^2 = O(r_n v_n).$$

Then the conditions (C1), (D1) *and* (D2′) *hold, and thus also* (C2) *and* (D2) *are satisfied. Moreover,*

$$(3.6) \qquad \sup_{\phi \in \Phi} |\tilde{Z}_n(\phi) - Z_n(g_\phi)| \to 0 \qquad \text{in outer probability.}$$

If, in addition, (C3) *holds and one of the following two sets of conditions,*

(i) (D4) *with a partition of $\mathcal{F}$ independent of n, or*
(ii) (D3), (D5) *and* (D6),

are satisfied, then $(\tilde{Z}_n(\phi))_{\phi \in \Phi}$, and the empirical processes $(Z_n(g_\phi))_{\phi \in \Phi}$ of cluster functionals, converge weakly to a Gaussian process with covariance function c.

REMARK 3.7. (i) *It is possible to replace* (C3) *in the corollary by more basic assumptions. Specifically, assume that the cluster lengths $L(Y_n)$ satisfy*

$$(3.7) \qquad \lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{r_n v_n} P\{L(Y_n) > k\} = 0$$

[*which by Lemma* 5.2(vii) *holds if* (B3) *is satisfied*], *that there exist functions $d_j : \Phi^2 \to \mathbb{R}$ such that, for $k \in \mathbb{N}$ and $\phi, \psi \in \Phi$,*

$$(3.8) \qquad \frac{1}{v_n} E(\phi(X_{n,1})\psi(X_{n,k})) \to d_{k-1}(\phi, \psi) \qquad \text{as } n \to \infty,$$

and that

$$E\left(\sum_{i=1}^{r_n}\mathbf{1}_{\{X_{n,i}\neq 0\}}\right)^{2+\delta} = O(r_n v_n) \tag{3.9}$$

for some $\delta > 0$. Then (C3′), *and hence, by Corollary* 2.4, *also* (C3) *hold with*

$$c(g_\phi, g_\psi) = d_0(\phi, \psi) + \sum_{i=1}^{\infty}(d_i(\phi, \psi) + d_i(\psi, \phi)). \tag{3.10}$$

The proof is given in Section 5.

(ii) *Suppose that the following simpler version of* (C3″) *is satisfied, viz. that there exists a sequence $(W_i)_{i\in\mathbb{N}}$ of E-valued random variables such that, for all $k \in \mathbb{N}$, $P^{(X_{n,1},X_{n,k})|X_{n,1}\neq 0} \to P^{(W_1,W_k)}$ weakly, with $P\{W_k \in D_\phi \setminus \{0\}\} = 0$ for all $\phi \in \Phi$, $k \in \mathbb{N}$, where D_ϕ is the discontinuity set of ϕ. Then, in view of Lemma* 2.5, *Remark* 2.6 *and the boundedness of ϕ and ψ,*

$$\frac{1}{v_n}E\phi(X_{n,1})\psi(X_{n,k}) = E\big(\phi(X_{n,1})\psi(X_{n,k})|X_{n,1}\neq 0\big)$$

$$\to E\phi(W_1)\psi(W_k) =: d_{k-1}(\phi, \psi),$$

so that equation (3.8) *holds.*

EXAMPLE 3.8 (Multivariate tail empirical processes, ctd.). In this example we give a set of conditions for the convergence of the multivariate tail empirical process from Example 3.1 for uniformly distributed r.v.'s X_i. We then discuss how the condition (C3) on convergence of covariances may be checked in the present situation. Finally, we show that the central condition (3.11) may be weakened in the univariate case to condition (3.13). This improves earlier results in the literature.

Thus, we first show that if $r_n = o(\sqrt{nv_n})$, (B1), (B2) and (C3) are satisfied, and there exist a constant K and a $\delta > 0$ such that, for all sufficiently large n,

$$E\left(\sum_{i=1}^{r_n}\mathbf{1}_{(x,y]}\left(\frac{X_i - u_n}{a_n}\right)\right)^2 \leq K|\log(y-x)|^{-(1+\delta)}r_n v_n$$

$$\forall 0 \leq x < y \leq 1, y - x \leq 1/2, \tag{3.11}$$

then the multivariate tail empirical process

$$\left(\frac{1}{\sqrt{nv_n}}\sum_{i=1}^{n}\big(\mathbf{1}_{(x,1]}(X_{n,i}) - P\big(X_{n,i} \in (x, 1]\big)\big)\right)_{x\in[0,1]^d}$$

converges weakly to a Gaussian process with covariance function c.

Clearly, (3.11) implies (3.5). By Corollary 3.6, it is hence enough to show that condition (i) of the corollary is satisfied. Now, to each $\varepsilon > 0$, let $\eta = \eta_\varepsilon := \exp(-(K^{-1}d^{-3}\varepsilon^2)^{-1/(1+\delta)})$ and define sets

$$\Phi^\varepsilon_{(i_1,\ldots,i_d)} := \left\{ \mathbf{1}_{\times^d_{l=1}(x_l,1]} \mid (i_l - 1)\eta \leq x_l \leq \min(i_l\eta, 1) \, \forall 1 \leq l \leq d \right\},$$

$$i_1, \ldots, i_d \in \{1, \ldots, \lceil 1/\eta \rceil\},$$

such that $\bigcup_{i_1,\ldots,i_d \in \{1,\ldots,\lceil 1/\eta \rceil\}} \Phi^\varepsilon_{(i_1,\ldots,i_d)} = \Phi$. Since, by (B1) and (3.11),

$$E \sup_{\phi,\psi \in \Phi^\varepsilon_{(i_1,\ldots,i_d)}} |g_\phi(Y_n) - g_\psi(Y_n)|^2$$

$$= E\left(\sum_{i=1}^{r_n} \mathbf{1}_{\times^d_{l=1}((i_l-1)\eta,1] \setminus \times^d_{l=1}(i_l\eta,1]}(X_{n,i}) \right)^2$$

$$\leq E\left(\sum_{i=1}^{r_n} \sum_{l=1}^{d} \mathbf{1}_{((i_l-1)\eta,i_l\eta]}\left(\frac{X_{i+l-1} - u_n}{a_n} \right) \right)^2$$

$$\leq d^2 E \max_{1 \leq l \leq d} \left(\sum_{i=1}^{r_n} \mathbf{1}_{((i_l-1)\eta,i_l\eta]}\left(\frac{X_{i+l-1} - u_n}{a_n} \right) \right)^2$$

$$\leq d^3 K |\log \eta|^{-(1+\delta)} r_n v_n$$

$$= \varepsilon^2 r_n v_n,$$

it follows that

$$\log N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2^n) \leq \log(\lceil 1/\eta \rceil^d) = O(\varepsilon^{-2/(1+\delta)})$$

as $\varepsilon \downarrow 0$. Hence, the condition (D4) on entropy with bracketing holds with a partition independent of n, as required to prove the claim.

The convergence (C3) of covariance functions which was used above may sometimes be replaced by simpler conditions. Specifically, Remark 3.7 gives sufficient conditions for (C3) to hold, for general $d \in \mathbb{N}$. Assume, for example, that all bivariate distributions (X_1, X_m) belong to the domain of attraction of some bivariate extreme value distribution. Then, since the limiting random variables W_i are continuous on $(0, \infty)$, the assumptions of Remark 3.7(ii) are satisfied, and, hence, (3.8) holds [cf. Segers (2003), Theorem 2]. Further, condition (3.9) holds if and only if for some $\delta > 0$

$$(3.12) \qquad E\left(\sum_{i=1}^{r_n} \mathbf{1}_{(u_n,1]}(X_i) \right)^{2+\delta} = O(r_n v_n).$$

For the case $d = 1$, the condition (3.11) can be weakened, to the requirement that

$$(3.13) \qquad E\left(\sum_{i=1}^{r_n} \mathbf{1}_{(x,y]}\left(\frac{X_i - u_n}{a_n}\right)\right)^2 \leq h(y-x)r_n v_n \qquad \forall 0 \leq x < y \leq 1,$$

for some function $h : (0, \infty) \to (0, \infty)$ satisfying $\lim_{t\downarrow 0} h(t) = 0$. To see this, note that the functions $\phi_x = \mathbf{1}_{(x,1]}$, $x \in [0, 1]$, are linearly ordered, and hence so are the corresponding cluster functionals g_{ϕ_x}, $x \in [0, 1]$. Hence, $\mathcal{F} = \{g_{\phi_x} \mid x \in [0, 1]\}$ is a VC class of functions [van der Vaart and Wellner (1996), Section 2.6]. Thus, according to Remark 2.11, (D6′) [and hence also (D6)] is satisfied. The measurability condition (D5) holds, since all processes occurring in this setting are separable. Moreover, (D3) is satisfied for the metric $\rho(g_{\phi_x}, g_{\phi_y}) := |y - x|$:

$$\limsup_{n\to\infty} \frac{1}{r_n v_n} \sup_{x,y\in[0,1],|y-x|<\delta} E\big(g_{\phi_x}(Y_n) - g_{\phi_y}(Y_n)\big)^2$$

$$= \limsup_{n\to\infty} \frac{1}{r_n v_n} \sup_{x,y\in[0,1],|y-x|<\delta} E\left(\sum_{i=1}^{r_n} \mathbf{1}_{(x,y]}(X_{n,i})\right)^2$$

$$\leq \sup_{0<t\leq\delta} h(t)$$

$$\to 0$$

as $\delta \downarrow 0$ by (3.13), so that version (ii) of Corollary 3.6 applies. This proves the claim that (3.11) may be weakened to (3.13) in the univariate case.

If we could assume that $\{X_i; 1 \leq i \leq n\}$ could be split up into consecutive *independent* blocks of length r_n, then (3.13) would be seen to be the same as to assume that $E(Z_n(g_{\phi_y}) - Z_n(g_{\phi_x}))^2 \leq h(|y - x|)$, for some h with properties as above. This is the same as to assume that Z_n is uniformly mean square continuous. However, in the proofs in Section 5 we use mixing to translate to cases where this independence assumption in fact can be made, and, accordingly, (3.13) seems quite minimal. In fact, in view of the counterexamples in Hahn (1977), it may even be surprising that this condition is sufficient.

Rootzén (1995, 2009) proved convergence of the univariate tail empirical process e_n using a more restrictive version of (3.11) and the stronger condition that $r_n = o((nv_n)^{1/2-\varepsilon})$ for some $\varepsilon > 0$. In Drees (2000) Rootzén's conditions were slightly weakened to the requirements that $r_n = o((nv_n)^{1/2} \log^{-2}(nv_n))$ and that

$$(3.14) \qquad E\left(\sum_{i=1}^{r_n} \mathbf{1}_{(x,y]}\left(\frac{X_i - u_n}{a_n}\right)\right)^2 \leq K(y-x)r_n v_n \qquad \forall 0 \leq x < y \leq 1,$$

instead of (3.11). Condition (3.14) is much more restrictive than (3.11) for small $y - x$. In many specific time series models, it was condition (3.14) (for small $y - x$)

that turned out to be most difficult to verify; see, for example, the discussion of the solutions of a stochastic recurrence equation in Drees (2000), Section 4. Therefore, it might be useful that the bound in (3.11) converges to 0 much more slowly as $y - x$ tends to 0.

It is possible to deal with Examples 3.2 and 3.5 in a similar fashion.

As already mentioned, Example 3.3 does not fit into the framework of Corollary 3.6 if the underlying df belongs to the domain of attraction of an extreme value distribution with nonnegative extreme value index, because then the support is not bounded. In that case, condition (3.5) must be strengthened.

COROLLARY 3.9. *In the setting of Corollary* 3.6 *the assertions remain true if* $\phi_{\max}$ *is measurable but not necessarily bounded, provided* (3.5) *is replaced with*

$$(3.15) \qquad E\left(\sum_{i=1}^{r_n} \phi_{\max}(X_{n,i})\right)^{2+\delta} = O(r_n v_n) \qquad \text{for some } \delta > 0.$$

EXAMPLE 3.10 (Compound insurance claim, ctd.).   In the setting of Example 3.3, uniform convergence of the empirical process of cluster functionals can be expected only if the deductible t is restricted to some bounded set. Therefore, we consider the set $\Phi_T := \{\phi_t \mid t \in [0, T]\}$ for an arbitrary $T \in (0, \infty)$. This set is totally bounded w.r.t. the metric $d_\Phi(\phi_s, \phi_t) := |s - t|$. The envelope function is $\phi_{\max}(x) = \phi_0(x) = x_+$.

Suppose conditions (B1), (B2), (C3), (3.5) and

$$(3.16) \qquad\qquad E\left(\sum_{i=1}^{r_n} X_{n,i}\right)^{2+\delta} = O(r_n v_n)$$

for some $\delta > 0$, are satisfied. Then the empirical process $(Z_n(g_{\phi_t}))_{0 \le t \le T}$ converges weakly to a Gaussian process.

To see this, first observe that the functions ϕ_t are monotonically decreasing in t. Hence, Φ_T is a VC class of functions, so that (D6) holds (see Remark 2.10). Since all sample paths are continuous, the measurability condition (D5) trivially holds.

To prove (D3), check that

$$\sup_{0 \le s \le t \le T, |t-s| < \delta} \frac{1}{r_n v_n} E\left(\sum_{i=1}^{r_n} ((X_{n,i} - s)_+ - (X_{n,i} - t)_+)\right)^2$$

$$\le \sup_{0 \le s \le t \le T, |t-s| < \delta} \frac{1}{r_n v_n} E\left(\sum_{i=1}^{r_n} (t - s)\mathbf{1}_{(s,\infty)}(X_{n,i})\right)^2$$

$$\le \delta^2 \frac{1}{r_n v_n} E\left(\sum_{i=1}^{r_n} \mathbf{1}_{(0,\infty)}(X_{n,i})\right)^2.$$

By (3.5), the lim sup of the right-hand side (as n tends to ∞) is bounded by a multiple of δ^2, which yields (D3). Further, (3.16) is just a reformulation of (3.15) to the present setting. Hence, all the conditions of Corollary 3.9 have been verified, and thus the result follows.

By Corollary 2.4, the condition (C3) in turn follows if, in addition, one assumes that all finite-dimensional marginal distributions of the time series $(X_i)_{i\in\mathbb{N}}$ belong to the domain of attraction of some extreme value distributions and that the normalizing constants u_n and a_n are chosen accordingly. Then (C3.1$''$) holds [cf. Segers (2003), Theorem 2], and (C3.2$''$) also follows from (3.15) and Lemma 5.2(vi).

EXAMPLE 3.11 (Bootstrapping the Hill estimator, ctd.). Continuing Example 3.4, we now sketch proofs of asymptotic normality of the Hill estimator and of consistency of the block bootstrap. Full process convergence may also be obtained and is useful if, for example, u_n is replaced by the k_nth largest order statistic, for some suitable sequence k_n. We use asymptotic normality to show consistency of the block bootstrap—but the hope is that the bootstrap has better small-sample properties than the normal approximation with estimated variance.

For this we assume that (B1) and (B2) and, with the notation of Example 3.4, that for $k, l \in \{1, 2\}$

$$(3.17) \qquad E\left(\sum_{i=1}^{r_n} \phi_k(X_{n,i})\right)^4 = O(r_n v_n),$$

$$\lim_{n\to\infty} \frac{1}{r_n v_n} \sum_{i=1}^{r_n} \sum_{j=1}^{r_n} E(\phi_k(X_{n,i})\phi_l(X_{n,j})) = \sigma_{kl}.$$

Then, in a similar way as in the proofs of Corollaries 3.6 and 3.9, it can be seen that $(\tilde{Z}_n(\phi_k))_{1\le k\le 2}$ converges to a centered normal distribution with covariance matrix $(\sigma_{kl})_{1\le k,l\le 2}$. It follows that

$$(3.18) \qquad \hat{\gamma}_n = \gamma_n + (nv_n)^{-1/2}(\tilde{Z}_n(\phi_1) - \gamma \tilde{Z}_n(\phi_2)) + o_p((nv_n)^{-1/2}),$$

and thus that

$$(3.19) \qquad \sqrt{nv_n}(\hat{\gamma}_n - \gamma_n) \longrightarrow \mathcal{N}_{(0,\sigma_{11}+\gamma^2\sigma_{22}-2\gamma\sigma_{12})} \qquad \text{in distribution.}$$

Writing $X^{(n)} := (X_i)_{1\le i\le n}$ for the original data, we next show that

$$(3.20) \qquad \sup_{t\in\mathbb{R}}\left|P\left(\sqrt{nv_n}(\hat{\gamma}_n^* - \hat{\gamma}_n) \le t \mid X^{(n)}\right) - P\{\sqrt{nv_n}(\hat{\gamma}_n - \gamma_n) \le t\}\right| = o_P(1),$$

that is, consistency of the block bootstrap estimator. With the notation from Example 3.4,

$$\frac{E(g_1(Y_i^{(n)})|X^{(n)})}{E(g_2(Y_i^{(n)})|X^{(n)})} = \frac{m_n^{-1}\sum_{i=1}^{m_n} g_1(Y_{n,i})}{m_n^{-1}\sum_{i=1}^{m_n} g_2(Y_{n,i})} = \hat{\gamma}_n.$$

From arguments as in the proof of Lemma 5.1 below [in particular, (5.4)], it follows that if condition (3.17) holds, then $Z_n(g_k g_l) = O_P(1)$. Hence, for $k, l \in \{1, 2\}$,

$$\frac{1}{r_n v_n} \mathrm{Cov}\big(g_k\big(Y_1^{(n)}\big) g_l\big(Y_1^{(n)}\big) \mid X^{(n)}\big)$$

$$= \frac{1}{r_n v_n}\left(\frac{1}{m_n}\sum_{i=1}^{m_n} g_k(Y_{n,i}) g_l(Y_{n,i}) - \frac{1}{m_n}\sum_{i=1}^{m_n} g_k(Y_{n,i}) \cdot \frac{1}{m_n}\sum_{i=1}^{m_n} g_l(Y_{n,i})\right)$$

$$= \frac{1}{r_n v_n} \mathrm{Cov}(g_k(Y_{n,1}), g_l(Y_{n,1})) - \frac{1}{m_n} Z_n(g_k) Z_n(g_l)$$

$$+ \frac{1}{\sqrt{n v_n}}\big(Z_n(g_k g_l) - E(g_l(Y_{n,1})) Z_n(g_k) - E(g_k(Y_{n,1})) Z_n(g_l)\big)$$

$$\to \sigma_{kl}$$

in probability. Similarly, as in (3.18), we have that

$$\hat{\gamma}_n^* = \hat{\gamma}_n + (n v_n)^{-1}\sum_{i=1}^{m_n}\big(g_1\big(Y_i^{(n)}\big) - \gamma g_2\big(Y_i^{(n)}\big) - E\big(g_1\big(Y_i^{(n)}\big) - \gamma g_2\big(Y_i^{(n)}\big) \mid X^{(n)}\big)\big)$$

$$+ o_p((n v_n)^{-1}).$$

Moreover, one can conclude from (3.17) that

$$m_n E\left(\left(\frac{g_k(Y_1^{(n)}) - E(g_k(Y_1^{(n)}) | X^{(n)})}{\sqrt{n v_n}}\right)^3 \Big| X^{(n)}\right) = O_P(m_n (n v_n)^{-3/2} r_n v_n)$$

$$= O_P((n v_n)^{-1/2}),$$

and, thus, the Berry–Esséen inequality yields

$$\sup_{t \in \mathbb{R}}\left|P\left((n v_n)^{-1/2}\sum_{i=1}^{m_n}(g_1\big(Y_i^{(n)}\big) - \gamma g_2\big(Y_i^{(n)}\big)\right.\right.$$

$$\left.\left. - E\big(g_1\big(Y_i^{(n)}\big) - \gamma g_2\big(Y_i^{(n)}\big) \mid X^{(n)}\big)\big) \le t \mid X^{(n)}\right)\right.$$

$$\left. - \Phi\big((\sigma_{11} + \gamma^2 \sigma_{22} - 2\gamma \sigma_{12})^{-1/2} t\big)\right| = o_P(1).$$

In view of (3.19), this proves (3.20).

4. Indicator functionals. Another important class of cluster functionals are indicator functions. Notice that by definition these indicator functions are applied to whole clusters, while in the Examples 3.1, 3.2 and 3.5 above indicator functions of single observations $X_{n,i}$ were summed up. For $C \subset E_\cup$ the indicator function $\mathbf{1}_C$ is a cluster functional if and only if the set satisfies the following two conditions:

- $x = (x_1, \ldots, x_\ell) \in C \iff (0, x_1, \ldots, x_\ell) \in C \iff (x_1, \ldots, x_\ell, 0) \in C$ for all $x \in E_\cup$;
- $0 \notin C$.

In this section we study situations where the set of cluster functionals is of the form $\{\mathcal{F} = \{\mathbf{1}_C \mid C \in \mathcal{C}\}$ for some family $\mathcal{C} \subset 2^{E_\cup}$ of such sets.

EXAMPLE 4.1 (Joint survival function of cluster values). The conditional joint survival function of the first k observations in a cluster core Y_n^c, given that the core has length greater than or equal to k, can be estimated by

$$\frac{\sum_{j=1}^{m_n} \mathbf{1}_{C_{t_1,\ldots,t_k}}(Y_{n,j})}{\sum_{j=1}^{m_n} \mathbf{1}_{C_{0,\ldots,0}}(Y_{n,j})}$$

with

$$C_{t_1,\ldots,t_k} := \{x \in E_\cup \mid \exists j : x_i = 0 \; \forall 1 \leq i \leq j, \, x_{j+i} > t_i \; \forall 1 \leq i \leq k\}.$$

Obviously, a limit theorem for the empirical process

$$\tilde{Z}_n(t_1, \ldots, t_k) := Z_n(\mathbf{1}_{C_{t_1,\ldots,t_k}}), \qquad t_1, \ldots, t_k \in [0, 1],$$

is useful for the asymptotic analysis of the above estimator.

EXAMPLE 4.2 (Order statistics of cluster values). Let

$$D_{t_1,\ldots,t_k} := \bigcap_{j=1}^{k} E_{j,t_j}$$

with

$$E_{j,t_j} := \left\{ (x_1, \ldots, x_m) \in E_\cup \;\middle|\; m \in \mathbb{N}, \sum_{i=1}^{m} \mathbf{1}_{(t_j,1]}(x_i) \geq j \right\},$$

that is, $D_{t_1,\ldots,t_k}$ contains all vectors of arbitrary length such that the jth largest value exceeds t_j for all $1 \leq j \leq k$. Then the empirical process $\tilde{Z}_n(t_1, \ldots, t_k) = Z_n(\mathbf{1}_{D_{t_1,\ldots,t_k}})$ describes the standardized joint empirical survival function of the k largest order statistics of the cluster cores.

Next we discuss the conditions imposed in Theorem 2.10 to ensure convergence of the empirical processes considered in this section.

The conditions (D1) and (D2′) are trivial, and condition (C1) holds by Lemma 5.2(ii).

If $r_n v_n \to 0$ [which is a part of assumption (B1)], then (C3) is equivalent to

(4.1) $$\frac{1}{r_n v_n} P\{Y_{n,1} \in C \cap D\} \to c(\mathbf{1}_C, \mathbf{1}_D),$$

since $\mathrm{Cov}(\mathbf{1}_C(Y_n), \mathbf{1}_D(Y_n)) = P\{Y_n \in C \cap D\} - P\{Y_n \in C\} \cdot P\{Y_n \in D\}$ and since $P\{Y_n \in C\} \cdot P\{Y_n \in D\} = O((r_n v_n)^2) = o(r_n v_n)$.

Similarly, condition (D3) can be reformulated as

$$(4.2) \qquad \lim_{\delta \downarrow 0} \limsup_{n \to \infty} \sup_{C,D \in \mathcal{C}, \rho_{\mathcal{C}}(C,D) < \delta} \frac{1}{r_n v_n} P\{Y_n \in C \triangle D\} = 0,$$

where $C \triangle D = (C \setminus D) \cup (D \setminus C)$ denotes the symmetric difference between C and D and $\rho_{\mathcal{C}}$ is a semi-metric on $\mathcal{C}$ that induces a semi-metric ρ on $\mathcal{F}$ via $\rho(\mathbf{1}_C, \mathbf{1}_D) := \rho_{\mathcal{C}}(C, D)$.

If (C3″) holds, then

$$\frac{1}{r_n v_n} P\{Y_n \in C \triangle D\} \longrightarrow P\{(W_i)_{i \geq 1} \in C \triangle D\} - P\{(W_i)_{i \geq 2} \in C \triangle D\},$$

where $(W_i)_{i \geq 1} \in C \triangle D$ is interpreted as $(W_i)_{1 \leq i \leq m} \in C \triangle D$ for some $m \geq m_W$, that is, $W_i = 0$ for all $i > m$. If the following continuity property holds

$$\lim_{\delta \downarrow 0} \sup_{C,D \in \mathcal{C}, \rho_{\mathcal{C}}(C,D) < \delta} P\{(W_i)_{i \geq 1} \in C \triangle D\} - P\{(W_i)_{i \geq 2} \in C \triangle D\} = 0,$$

then results by Fabian (1970) may help to conclude (D3). However, in the examples of this section we will verify (D3) in a more direct way.

Finally, if $\mathcal{C}$ is a VC-class, then condition (D6′) is fulfilled (cf. Remark 2.11).

The following result gives conditions for the convergence of the empirical processes in Examples 4.1 and 4.2. Here we assume that the random variables $X_{n,i}$ are $[0, 1]$-valued so that it suffices to consider the processes $\tilde{Z}_n$ with index set $[0, 1]^k$. If the r.v.'s $X_{n,i}$ are standardized excesses defined in (1.1) (as we assume in the second part of the following corollary), then this can be achieved by a simple quantile transformation (cf. Example 3.1).

COROLLARY 4.3. (i) *Let $\tilde{Z}_n(t_1, \ldots, t_k)$ be as in Examples 4.1 or 4.2, with $t_i \in [0, 1]$, $i = 1, \ldots, k$, and suppose* (B1), (B2), (B3), (C3.1″) *and* (D3) *hold with $\rho(\mathbf{1}_{C_{s_1,\ldots,s_k}}, \mathbf{1}_{C_{t_1,\ldots,t_k}}) := \sum_{i=1}^k |s_i - t_i|$, respectively, $\rho(\mathbf{1}_{D_{s_1,\ldots,s_k}}, \mathbf{1}_{D_{t_1,\ldots,t_k}}) := \sum_{i=1}^k |s_i - t_i|$. Then $\tilde{Z}_n$ converges to a continuous Gaussian process. If $\tilde{Z}_n$ is as in Example 4.1, then the covariance function of the process is*

$$(4.3) \quad \begin{aligned} &\tilde{c}((s_1, \ldots, s_k), (t_1, \ldots, t_k)) \\ &= P\big\{(W_i)_{i \geq 1} \in C_{\max(s_1,t_1),\ldots,\max(s_k,t_k)}\big\} \\ &\quad - P\big\{(W_i)_{i \geq 2} \in C_{\max(s_1,t_1),\ldots,\max(s_k,t_k)}\big\}, \end{aligned}$$

and if $\tilde{Z}_n$ is as in Example 4.2, then the covariance function of the process is

$$(4.4) \quad \begin{aligned} &\tilde{c}((s_1, \ldots, s_k), (t_1, \ldots, t_k)) \\ &= P\bigg\{(W_i)_{i \geq 1} \in \bigcap_{j=1}^k E_{j,\max(s_j,t_j)}\bigg\} - P\bigg\{(W_i)_{i \geq 2} \in \bigcap_{j=1}^k E_{j,\max(s_j,t_j)}\bigg\}. \end{aligned}$$

(ii) *More specifically, assume that the r.v.'s $X_{n,i}$ are standardized excesses of a uniformly distributed univariate stationary time series (as in Example* 3.1) *and that all finite-dimensional marginal distributions belong to the domain of attraction of some extreme value distribution. Then the assertions of part* (i) *hold true if the conditions* (B1), (B2) *and* (B3) *are satisfied.*

In Example 4.1 we only considered the first k "extremes" in each cluster, where k is a fixed number. Since for most time series the cluster size is not bounded, the resulting empirical process does not give a full picture of the stochastic behavior of the clusters. To overcome this drawback, in the final example we define and analyze an empirical process of cluster functionals that takes *all* values of each cluster into account. As the cluster length is random, this requires work with a quite complex index set.

EXAMPLE 4.4 (Joint distribution of all cluster values).   Recalling the notation $L(x)$ for the length, say, j, of the core $x^c = (x_1^c, \ldots, x_j^c)$ of a vector x, we set

$$C_{j,t_1,\ldots,t_j} := \{x \in E_\cup \mid L(x) = j, x_i^c \in [0, t_i], \forall 1 \le i \le j\}.$$

Then the empirical process $\tilde{Z}_n(j, t_1, \ldots, t_j) := Z_n(\mathbf{1}_{C_{j,t_1,\ldots,t_j}})$, $j \in \mathbb{N}$, $t_i \ge 0$, describes the joint distribution of all the values in a cluster.

Like in Corollary 4.3(ii), for simplicity, we focus on the case that the clusters are based on standardized exceedances $X_{n,i}$ of a uniformly distributed stationary time series $(X_i)_{i \in \mathbb{N}}$, such that all finite-dimensional marginal distributions belong to the domain of attraction of some extreme value distribution. However, it is not difficult to generalize this result to a slightly more general setting which is analog to the one considered in Corollary 4.3(i).

Suppose that (B1), (B2) and (B3) hold, and that

$$(4.5) \qquad E\big(L(Y_n)^{1+\zeta} \mid Y_n \ne 0\big) = O_p(1) \qquad \text{some } \zeta > 0.$$

Then $\tilde{Z}_n$ converges weakly to a continuous Gaussian process with covariance function

$$
\begin{aligned}
&c((j, s_1, \ldots, s_j), (k, t_1, \ldots, t_k)) \\
(4.6) \quad &= \delta_{j,k}\big(P\{L(W) = k, W_i \le s_i \wedge t_i, \forall 1 \le i \le k\} \\
&\qquad - P\{L\big(W^{(2;\infty)}\big) = k, \big((W^{(2;\infty)})^c\big)_i \le t_i, \forall 1 \le i \le k\}\big),
\end{aligned}
$$

where $\delta_{j,k}$ is one if $j = k$ and zero otherwise.

The proof of this uniform central limit theorem is given in Section 5.

5. Proofs. In this section we prove the results from Sections 2–4. We start with fidi convergence, then consider asymptotic tightness and asymptotic equicontinuity, and finally prove the corollaries from Sections 3 and 4.

The first step in the proof of fidi convergence is to use mixing to bring the problem back to classical limit theory for i.i.d. variables. Let $Y_{n,j}^*$ denote i.i.d. copies of the original blocks $Y_{n,j}$ (which are identically distributed, but are not assumed to be independent—and which in interesting cases typically are dependent).

LEMMA 5.1. *Suppose* (B1), (B2) *and* (C1) *are satisfied. Then the fidis of* $(Z_n(f))_{f \in \mathcal{F}}$ *converge weakly if and only if the fidis of the sums of independent blocks*

$$Z_n^*(f) := \frac{1}{\sqrt{n v_n}} \sum_{j=1}^{m_n} (f(Y_{n,j}^*) - Ef(Y_{n,j}^*)), \qquad f \in \mathcal{F},$$

converge weakly. In this case the limit distributions are the same.

PROOF. Let

$$\Delta_{n,j}^*(f) := f(Y_{n,j}^*) - f((Y_{n,j}^*)^{(r_n - l_n)}), \qquad 1 \le j \le m_n,$$

and let $\Delta_{n,j}(f)$ be defined in the same way, but instead based on the original (dependent) blocks, so that $\Delta_{n,j}^*(f) \overset{d}{=} \Delta_{n,j}(f) \overset{d}{=} \Delta_n(f)$ for each j, with $\Delta_n(f)$ as in (C1). By Theorem 1 in Petrov [(1975), Section IX.1] applied to the i.i.d. random variables $X_{nk} := (n v_n)^{-1/2} \Delta_{n,k}^*(f)$, condition (C1) implies that

$$(5.1) \qquad \frac{1}{\sqrt{n v_n}} \sum_{j=1}^{m_n} (\Delta_{n,j}^*(f) - E\Delta_{n,j}^*(f)) = o_P(1) \qquad \forall f \in \mathcal{F}.$$

We next prove the analogous convergence for the dependent random variables, that is, that

$$(5.2) \qquad \frac{1}{\sqrt{n v_n}} \sum_{j=1}^{m_n} (\Delta_{n,j}(f) - E\Delta_{n,j}(f)) = o_P(1) \qquad \forall f \in \mathcal{F}.$$

Using Theorem 1 in Petrov [(1975), Section IX.1] again, it also follows from (C1) that the convergence analogous to (5.1) holds for the sums of the even numbered blocks

$$(5.3) \qquad \frac{1}{\sqrt{n v_n}} \sum_{j=1}^{\lfloor m_n/2 \rfloor} (\Delta_{n,2j}^*(f) - E\Delta_{n,2j}^*(f)) = o_P(1).$$

Since the even numbered blocks $Y_{n,j}$ are separated by r_n observations, a well-known inequality for the total variation distance [cf. Eberlein (1984)] between the joint distributions of dependent observations and independent copies yields

$$(5.4) \qquad \big\| P^{(Y_{n,2j})_{1 \le j \le \lfloor m_n/2 \rfloor}} - P^{(Y_{n,2j}^*)_{1 \le j \le \lfloor m_n/2 \rfloor}} \big\|_{TV} \le \lfloor m_n/2 \rfloor \beta_{n,r_n} \to 0$$

by (B2). Combining (5.3) with (5.4), we arrive at

$$\frac{1}{\sqrt{nv_n}} \sum_{j=1}^{\lfloor m_n/2 \rfloor} (\Delta_{n,2j}(f) - E\Delta_{n,2j}(f)) = o_P(1).$$

Together with the analogous convergence for the sum over the odd numbered blocks, this proves (5.2).

Thus, the fidis of Z_n converge if and only if the fidis of

$$\bar{Z}_n(f) := Z_n(f) - \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} (\Delta_{n,j}(f) - E\Delta_{n,j}(f))$$

$$= \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{m_n} (f(Y_{n,j}^{(r_n-\ell_n)}) - Ef(Y_{n,j}^{(r_n-\ell_n)})), \qquad f \in \mathcal{F},$$

converge, and in this case the limiting distributions are the same. Similarly, by (5.1), the corresponding assertion holds for the sums over the independent blocks, and then the lemma follows from the inequality for the total variation distance, since it implies that

$$\left\| P^{(Y_{n,j}^{(r_n-\ell_n)})_{1\le j\le m_n}} - P^{((Y_{n,j}^*)^{(r_n-\ell_n)})_{1\le j\le m_n}} \right\|_{\mathrm{TV}} \le m_n \beta_{n,l_n} \to 0$$

by (B2), since the shortened blocks $Y_{n,j}^{(r_n-\ell_n)}$ are separated by l_n observations. □

PROOF OF THEOREM 2.3.   The assertion follows from Lemma 5.1 and and the multivariate central limit theorem for triangular arrays of row-wise independent random vectors applied to $(Z_n^*(f_1), \ldots, Z_n^*(f_k))$. □

Next we present a useful technical lemma. It makes it possible to replace some of the assumptions of Theorem 2.3 by sufficient conditions which are more restrictive but often simpler to verify.

LEMMA 5.2.   (i) *If* $\mathrm{Var}(\Delta_n(f)) = o(r_n v_n)$, *then* (C1) *holds.*
(ii) *If* $nv_n \to \infty$ *and* $\|f\|_\infty := \sup_{x\in E_\cup} |f(x)| < \infty$, *then* (C1) *and* (C2) *hold.*
(iii) *If* $r_n v_n \to 0$ *and*

$$(5.5) \qquad \frac{1}{r_n v_n} E(f(Y_n)g(Y_n)) \to c(f,g) \qquad \forall f, g \in \mathcal{F},$$

then (C3) *holds.*
(iv) *If*

$$(5.6) \qquad E\big(f(Y_n)^2 \mathbf{1}_{\{|f(Y_n)|>\varepsilon\sqrt{nv_n}\}}\big) = o(r_n v_n) \qquad \forall \varepsilon > 0, f \in \mathcal{F},$$

then (C2) *holds.*

(v) *If $nv_n \to \infty$ and $(f(Y_n)^2)_{n \in \mathbb{N}}$ is uniformly integrable under $P(\cdot)/(r_n v_n)$ for all $f \in \mathcal{F}$, then* (C2) *holds.*

(vi) *If $E(f(Y_n)^{2+\delta}) = O(r_n v_n)$ for some $\delta > 0$ and all $f \in \mathcal{F}$, then $(f(Y_n)^2)_{n=1}^{\infty}$ is uniformly integrable under $P(\cdot)/(r_n v_n)$ for all $f \in \mathcal{F}$.*

(vii) *If* (B1) *and* (B3) *hold, then* $\lim_{k \to \infty} \limsup_{n \to \infty} \frac{1}{r_n v_n} P\{L(Y_n) > k\} = 0$ *and the cluster lengths $(L(Y_n))_{n \in \mathbb{N}}$ are tight under $P(\cdot | Y_n \neq 0)$.*

PROOF. (i) The first equation in (C1) follows at once, and the second one by using Chebyshev's inequality.

(ii) Under these conditions, (C2) obviously holds. Moreover, (C1) follows by (i), since $|\Delta_n(f)| \leq 2\|f\|_\infty \mathbf{1}_{\{\Delta_n(f) \neq 0\}}$ implies

$$
\begin{aligned}
\mathrm{Var}(\Delta_n(f)) &\leq E \Delta_n^2(f) \\
&\leq 4\|f\|_\infty^2 P\{\Delta_n(f) \neq 0\} \\
&= O(P\{X_{n,i} \neq 0 \text{ for some } r_n - l_n + 1 \leq i \leq r_n\}) \\
&= O(l_n v_n) \\
&= o(r_n v_n).
\end{aligned}
$$

(iii) By (5.5), $P\{Y_n \neq 0\} \leq r_n v_n \to 0$ and the Cauchy–Schwarz inequality, we have that

(5.7)
$$
\begin{aligned}
\frac{1}{\sqrt{r_n v_n}} E|f(Y_n)| &= \frac{1}{\sqrt{r_n v_n}} E\big(|f(Y_n)| \mathbf{1}_{\{Y_n \neq 0\}}\big) \\
&\leq \left(\frac{1}{r_n v_n} E(f(Y_n)^2) P\{Y_n \neq 0\} \right)^{1/2} \to 0
\end{aligned}
$$

for $f \in \mathcal{F}$. (C3) then follows readily from (5.5).

(iv) By (5.6), for any $\epsilon > 0$,

$$
\begin{aligned}
E\left(\left(\frac{|f(Y_n)|}{\sqrt{n v_n}} \right)^2 \right) &\leq \epsilon^2 + \frac{1}{n v_n} E\big(f(Y_n)^2 \mathbf{1}_{\{|f(Y_n)| > \epsilon \sqrt{n v_n}\}}\big) = \epsilon^2 + o\left(\frac{r_n v_n}{n v_n} \right) \\
&= \epsilon^2 + o(1).
\end{aligned}
$$

Hence, $E f(Y_n) = o(\sqrt{n v_n})$, and (C2) then follows from (5.6) by standard reasoning.

(v) By uniform integrability, $n/r_n \to \infty$ and Chebyshev's inequality,

$$
P\{|f(Y_n)| > \epsilon \sqrt{n v_n}\} \leq \frac{E(f(Y_n)^2)/(r_n v_n)}{\epsilon^2 n/r_n} \to 0.
$$

Using uniform integrability again, it follows that $E(f(Y_n)^2 \mathbf{1}_{\{|f(Y_n)| > \epsilon \sqrt{n v_n}\}})/(r_n v_n) \to 0$, so that (5.6) is satisfied. The result then follows from part (iv).

(vi) This is a well-known fact.

(vii) Let $M_{n,s}^t := \sum_{i=s+1}^t \mathbf{1}_{\{X_{n,i}\neq 0\}}$ be the number of nonvanishing observations in the time interval from $s+1$ to t and write $F_{n,i} = \{X_{n,1} = \cdots = X_{n,i-1} = 0, X_{n,i} \neq 0\}$, $i \geq 2$, and $F_{n,1} = \{X_{n,1} \neq 0\}$ for the events that the first nonzero value in row n occurs at position i. Then

$$P\{L(Y_n) > k\} = \sum_{i=1}^{r_n-k} P(L(Y_n) > k \mid F_{n,i}) P(F_{n,i})$$

$$= \sum_{i=1}^{r_n-k} P(M_{n,i+k}^{r_n} \neq 0 \mid F_{n,i}) P(F_{n,i})$$

$$\leq \sum_{i=1}^{r_n-k} (\beta_{n,k} + P\{M_{n,i+k}^{r_n} \neq 0\}) P(F_{n,i})$$

$$\leq (\beta_{n,k} + r_n v_n) P\{Y_n \neq 0\}$$

$$\leq (\beta_{n,k} + r_n v_n) r_n v_n.$$

The result then follows from (B3) and $r_n v_n \to 0$. $\quad\square$

PROOF OF COROLLARY 2.4. The first assertion follows if we prove that (C3$'$) implies (C3). However, using that $|E(f(Y_n)g(Y_n)\mathbf{1}_{\{L(Y_n)>k\}})| \leq (E(f(Y_n)^2 \times \mathbf{1}_{\{L(Y_n)>k\}})E(g(Y_n)^2\mathbf{1}_{\{L(Y_n)>k\}}))^{1/2}$, it follows from (2.1) and (2.2) that

$$\frac{1}{r_n v_n} E(f(Y_n)g(Y_n)) = \frac{1}{r_n v_n} E(f(Y_n)g(Y_n)\mathbf{1}_{\{L(Y_n)\leq k\}})$$

$$+ \frac{1}{r_n v_n} E(f(Y_n)g(Y_n)\mathbf{1}_{\{L(Y_n)>k\}})$$

$$= c_k(f,g) + R_{n,k}'$$

with $\lim_{k\to\infty} \limsup_{n\to\infty} R_{n,k}' = 0$. A standard subsequence argument then shows that $c(f,g) := \lim_{k\to\infty} c_k(f,g)$ exits, and that

$$\lim_{n\to\infty} \frac{1}{r_n v_n} E(f(Y_n)g(Y_n)) = c(f,g).$$

By Lemma 5.2(iii), it then follows that (C3) holds.

Now suppose instead that (B1), (B2), (B3), (C1) and (C3$''$) hold. Assumption (C2) then follows from Lemma 5.2(v), and, hence, only (C3) remains to be established. By Lemma 2.5(ii) and (iii), $\theta_n = P\{Y_n \neq 0\}/(r_n v_n) \to \theta > 0$ and $P^{(fg)(Y_n)|Y_n\neq 0}$ converges weakly to $\mu_{fg,W}$. Thus, the uniform integrability of $(fg)(Y_n)$ under $P(\cdot)/(r_n v_n)$ is equivalent to the uniform integrability under $P(Y_n \neq 0)$ so that

$$\frac{1}{r_n v_n} E(f(Y_n)g(Y_n)) = \frac{P(Y_n \neq 0)}{r_n v_n} E(f(Y_n)g(Y_n) \mid Y_n \neq 0) \to \theta \int x \mu_{fg,W}(dx)$$

$$= E((fg)(W) - (fg)(W^{(2;\infty)})).$$

It then follows from Lemma 5.2(iii) that (C3) holds with $c(f, g)$ given by (2.4).
□

PROOF OF LEMMA 2.5.   Again let $M_{n,s}^t := \sum_{i=s+1}^t \mathbf{1}_{\{X_{n,i} \neq 0\}}$ denote the number of nonvanishing observations in the time interval from $s + 1$ to t. Then

$$(5.8) \qquad \limsup_{n \to \infty} P(M_{n,l}^{r_n} \neq 0 \mid X_{n,1} \neq 0) \leq \limsup_{n \to \infty} (\beta_{n,l} + r_n v_n) \to 0$$

as $l \to \infty$, by (B3) and $r_n v_n \to 0$. Hence, the analog to condition (2) of Segers (2003) holds and one may conclude the assertions (i) and (ii) by essentially the same arguments as given for the proofs of Theorem 1 (with $t_n = r_n$), Corollary 2 and Theorem 3(i) there.

The proof of (iii) also follows the ideas used in the proof of Theorem 3(ii) in that paper. Nevertheless, we give more details, since we want to avoid working with the space $\mathbb{A}$ of sequences with almost all terms equal to 0 that was introduced by Segers (2003). Moreover, in this proof we replace assumption (2.3) in condition (C3.1″) by the weaker assumptions (2.7) and (2.8).

We first consider a bounded cluster functional g such that $D_{g,m,I} \subset D_{f,m,I}$ for all $m \in \mathbb{N}$ and $I \subset \{1, \ldots, m\}$. The result for f itself will then follow easily. Let $k \in \mathbb{N}$ be arbitrary and, as before, let $\| \cdot - \cdot \|_{\mathrm{TV}}$ denote the total variation distance between two measures. By (5.8), for all $\varepsilon > 0$ there exists $l > k$ such that for sufficiently large n and $X_n^{(k)} = (X_{n,i})_{1 \leq i \leq k}$

$$\| P(X_n^{(k)} \in \cdot, M_{n,k}^{r_n} = 0 \mid X_{n,1} \neq 0) - P(X_n^{(k)} \in \cdot, M_{n,k}^l = 0 \mid X_{n,1} \neq 0) \|_{\mathrm{TV}}$$

$$(5.9) \qquad \leq P(M_{n,l}^{r_n} \neq 0 \mid X_{n,1} \neq 0)$$

$$\leq \varepsilon$$

and, by (2.6),

$$\| P\{W^{(k)} \in \cdot, W^{(k+1;\infty)} = 0\} - P\{W^{(k)} \in \cdot, W^{(k+1;l)} = 0\} \|_{\mathrm{TV}}$$

$$(5.10) \qquad \leq P\{W_i \neq 0 \text{ for some } i > l\}$$

$$\leq \varepsilon.$$

Recall the definition of the sets $N_{k,I}$ for $I \subset \{1, \ldots, k\}$ from Remark 2.6. Since, according to assumption (C3.1″), the substochastic measures $P(X_n^{(k)} \in \cdot, X_n^{(k)} \in N_{k,I}, M_{n,k}^l = 0 \mid X_{n,1} \neq 0)$ converge weakly to the substochastic measure $P\{W^{(k)} \in \cdot, W^{(k)} \in N_{k,I}, W^{(k+1;l)} = 0\}$, it follows from (5.9) and (5.10) that, for all $k \in \mathbb{N}$, and all subsets $I \subset \{1, \ldots, k\}$,

$$(5.11) \qquad \begin{aligned} &P(X_n^{(k)} \in \cdot, X_n^{(k)} \in N_{k,I}, M_{n,k}^{r_n} = 0 \mid X_{n,1} \neq 0) \\ &\qquad \to P\{W^{(k)} \in \cdot, W^{(k)} \in N_{k,I}, W^{(k+1;\infty)} = 0\} \end{aligned}$$

weakly.

By assertion (i), we have

$$(5.12) \qquad E\big(g(Y_n) \mid Y_n \neq 0\big) = \frac{1}{\theta_n} E\big(g(X_n^{(r_n)}) - g(X_n^{(2;r_n)}) \mid X_{n,1} \neq 0\big) + o(1).$$

Again by (5.9) and the definition of a cluster functional,

$$(5.13) \qquad \begin{aligned} &\big|E\big(g(X_n^{(r_n)}) - g(X_n^{(2;r_n)}) \mid X_{n,1} \neq 0\big) \\ &\quad - E\big((g(X_n^{(l)}) - g(X_n^{(2;l)}))\mathbf{1}_{\{M_{n,l}^{r_n}=0\}} \mid X_{n,1} \neq 0\big)\big| \leq 2\varepsilon \|g\|_\infty. \end{aligned}$$

In view of (5.11) (with $k = l$), for all $I \subset \{1, \ldots, l\}$, the continuous mapping theorem yields

$$E\big(g(X_n^{(l)})\mathbf{1}_{\{X_n^{(l)} \in N_{l,I}\}}\mathbf{1}_{\{M_{n,l}^{r_n}=0\}} \mid X_{n,1} \neq 0\big) \to E\big(g(W^{(l)})\mathbf{1}_{\{W^{(l)} \in N_{l,I}\}}\mathbf{1}_{\{W^{(l+1;\infty)}=0\}}\big),$$

because the function $g|_{N_{l,I}}$ is bounded and continuous on the complement of the set $D_{f,l,I}$, which by (2.7) is a null set under the limit measure in (5.11). Sum up these equations for all $I \subset \{1, \ldots, l\}$ and combine this with an analogous result for $g(X_n^{(2;l)})$ to obtain

$$(5.14) \qquad \begin{aligned} &E\big((g(X_n^{(l)}) - g(X_n^{(2;l)}))\mathbf{1}_{\{M_{n,l}^{r_n}=0\}} \mid X_{n,1} \neq 0\big) \\ &\qquad \to E\big((g(W^{(l)}) - g(W^{(2;l)}))\mathbf{1}_{\{W^{(l+1;\infty)}=0\}}\big). \end{aligned}$$

Combining (5.10), (5.12)–(5.14) and $\theta_n \to \theta > 0$, one arrives at

$$(5.15) \qquad E\big(g(Y_n) \mid Y_n \neq 0\big) \to \frac{1}{\theta} E\big(g(W) - g(W^{(2;\infty)})\big).$$

Now, if f is an arbitrary cluster functional satisfying the conditions of the proposition and $h : \mathbb{R} \to \mathbb{R}$ is continuous and bounded, then an application of (5.15) with $g = h \circ f$ yields assertion (iii). $\quad\square$

PROOF OF COROLLARY 2.7. This is immediate from Corollary 2.4 and Lemma 5.2(ii). $\quad\square$

PROOF OF THEOREM 2.8. The processes Z_n are asymptotically tight if the analogous sums over the even numbered and over the odd numbered blocks

$$(5.16) \qquad \begin{aligned} &\frac{1}{\sqrt{nv_n}} \sum_{j=1}^{\lfloor m_n/2 \rfloor} \big(f(Y_{n,2j}) - Ef(Y_{n,2j})\big) \quad \text{and} \\ &\frac{1}{\sqrt{nv_n}} \sum_{j=1}^{\lceil m_n/2 \rceil} \big(f(Y_{n,2j-1}) - Ef(Y_{n,2j-1})\big) \end{aligned}$$

H. DREES AND H. ROOTZÉN

are asymptotically tight. In view of (5.4), the first expression is asymptotically tight if and only if the analogous expression with independent blocks, that is,

$$(5.17) \qquad \frac{1}{\sqrt{nv_n}} \sum_{j=1}^{\lfloor m_n/2 \rfloor} \big(f(Y_{n,2j}^*) - Ef(Y_{n,2j}^*)\big)$$

is asymptotically tight, which follows from Theorem 2.11.9 of van der Vaart and Wellner (1996) applied with $Z_{ni}(f) = f(Y_{n,2i})$ (and m_n replaced with $\lfloor m_n/2 \rfloor$). Observe that for a sequence of monotonically increasing positive functions $T_n(\delta)$ the convergence of $T_n(\delta_n)$ to 0 for all sequences $\delta_n \downarrow 0$ is equivalent to $\lim_{\delta \downarrow 0} \limsup_{n \to \infty} T_n(\delta) = 0$, so that the last two displayed conditions in Theorem 2.11.9 of van der Vaart and Wellner (1996) can be reformulated as (D3) and (D4), respectively. The proof of tightness of the sum over the blocks with odd numbers is the same. $\square$

PROOF OF REMARK 2.9(ii). By the Cauchy–Schwarz inequality,

$$E^*\big(F(Y_n)\mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}\big)$$
$$\leq \big(E^*(F^2(Y_n)\mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}) \cdot E^* \mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}\big)^{1/2}$$
$$\leq \left(\frac{(E^*(F^2(Y_n)\mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}))^2}{\varepsilon^2 nv_n}\right)^{1/2}$$
$$= o\left(\frac{(r_nv_n)^2}{nv_n}\right)^{1/2}$$
$$= o\big(r_n\sqrt{v_n/n}\big),$$

so (D2) holds. Further, (D2$'$) implies (5.6), and, hence, (C2) follows from Lemma 5.2(iv).

Next, suppose $E^*F^{2+\delta}(Y_n) = O(r_nv_n)$ and $nv_n \to \infty$. Then

$$E^*\big(F^2(Y_n)\mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}\big)$$
$$\leq (E^*F^{2+\delta}(Y_n))^{2/(2+\delta)} \cdot \big(E^* \mathbf{1}_{\{F(Y_n) > \varepsilon \sqrt{nv_n}\}}\big)^{1-2/(2+\delta)}$$
$$= O\big((r_nv_n)^{2/(2+\delta)}\big) \cdot \left(\frac{E^*F^{2+\delta}(Y_n)}{(\varepsilon \sqrt{nv_n})^{2+\delta}}\right)^{1-2/(2+\delta)}$$
$$= O\big(r_nv_n(nv_n)^{-\delta}\big)$$
$$= o(r_nv_n),$$

so that (D2$'$) holds. $\square$

PROOF OF THEOREM 2.10. First assume (D6) holds. Using the triangle inequality, it is easily seen that Z_n is asymptotically equicontinuous if both terms

given in (5.16) are asymptotically equicontinuous. Further, by (5.4), the first term is asymptotically equicontinuous if and only if (5.17) is asymptotically equicontinuous. However, asymptotic equicontinuity of (5.17) follows from Theorem 2.11.1 of van der Vaart and Wellner (1996). To see this, note that (D6) implies the analogous random entropy condition for the sums over the even numbered blocks, because the corresponding random semi-metric is smaller for these sums.

If m_n is even, then the second term in (5.16) has the same distribution as the first one, while for m_n odd with probability greater than or equal to $1 - r_n v_n \to 1$, the additional summand $(n v_n)^{-1/2}(f(Y_{n,m_n}) - E f(Y_{n,m_n}))$ equals $-(n v_n)^{-1/2} E f(Y_{n,m_n})$, which tends to 0 uniformly for $f \in \mathcal{F}$ [cf. (5.7)]. This proves the first assertion of the theorem. Theorem 2.3 then yields the convergence of Z_n, because the Lindeberg condition (C2) follows from (D2) [see Remark 2.9(ii)].

Next, to see that (D6$'$) implies (D6), check that the random semi-metric d_n can be represented as $d_n = (m_n/(n v_n))^{1/2} \cdot d_Q$ with the (random) probability measure $Q = m_n^{-1} \sum_{j=1}^{m_n} \varepsilon_{Y_{n,j}^*}$, and, hence, $N(\varepsilon, \mathcal{F}, d_n) = N(\varepsilon(n v_n/m_n)^{1/2}, \mathcal{F}, d_Q)$. If $\int F^2 \, dQ = 0$, then $d_n(f, g) = 0$ for all $f, g \in \mathcal{F}$ and the integral in (D6$'$) vanishes. Otherwise, for all $\eta > 0$ there exists a $\tau > 0$ such that, for sufficiently large n,

$$P\left\{\left(\int F^2 \, dQ\right)^{1/2} > \tau(n v_n/m_n)^{1/2}\right\} \le \frac{E F^2(Y_{n,1})}{\tau^2 n v_n/m_n} \le \eta,$$

since $E F^2(Y_n) = O(r_n v_n)$, and thus with probability larger than $1 - \eta$,

$$\int_0^\delta \sqrt{\log N(\varepsilon, \mathcal{F}, d_n)} \, d\varepsilon = \tau \int_0^{\delta/\tau} \sqrt{\log N(\varepsilon\tau, \mathcal{F}, d_n)} \, d\varepsilon$$

$$\le \tau \int_0^{\delta/\tau} \sup_{Q \in \mathcal{Q}} \sqrt{\log N\left(\varepsilon\left(\int F^2 \, dQ\right)^{1/2}, \mathcal{F}, d_Q\right)} \, d\varepsilon$$

$$\to 0$$

as $\delta \downarrow 0$, under (D6$'$). $\square$

PROOF OF COROLLARY 3.6. Condition (D1) is satisfied since $F(x_1, \ldots, x_k) \le \sum_{i=1}^k \phi_{\max}(x_i)$ and since $\phi_{\max}$ is assumed to be measurable and bounded. Similarly, condition (D2$'$) follows from $F(Y_n) \le r_n \|\phi_{\max}\|_\infty$, since $r_n = o(\sqrt{n v_n})$ by assumption.

By Lemma 5.2(i), assumption (C1) follows if we show that $\mathrm{Var}(\Delta_n(f)) = o(r_n v_n)$. Now,

$$E\left(\sum_{i=1}^{r_n} \mathbf{1}_{\{X_{n,i} \ne 0\}}\right)^2 \ge E \sum_{j=1}^{\lfloor r_n/l_n \rfloor} \left(\sum_{i=1}^{l_n} \mathbf{1}_{\{X_{n,(j-1)l_n+i} \ne 0\}}\right)^2$$

$$= \lfloor r_n/l_n \rfloor E\left(\sum_{i=1}^{l_n} \mathbf{1}_{\{X_{n,i} \ne 0\}}\right)^2$$

by the row-wise stationarity, and, consequently, by (3.5) and $l_n = o(r_n)$,

$$E(\Delta_{n,1}^2(f)) \le E\left(\sum_{i=1}^{l_n} \phi_{\max}(X_{n,i})\right)^2$$

$$\le \|\phi_{\max}\|_\infty^2 E\left(\sum_{i=1}^{l_n} \mathbf{1}_{\{X_{n,i}\neq 0\}}\right)^2$$

$$= O\left(\frac{l_n}{r_n} r_n v_n\right) = o(r_n v_n).$$

Further, (3.6) follows from

$$E^*\left(\sup_{\phi\in\Phi} \frac{1}{\sqrt{n v_n}}\left|\sum_{i=r_n m_n+1}^{n} (\phi(X_{n,i}) - E\phi(X_{n,i}))\right|\right)^2$$

$$\le E\left(\frac{2}{\sqrt{n v_n}}\|\phi_{\max}\|_\infty \sum_{i=r_n m_n+1}^{n} \mathbf{1}_{\{X_{n,i}\neq 0\}}\right)^2$$

$$= \frac{4\|\phi_{\max}\|_\infty^2}{n v_n} \cdot r_n v_n \to 0.$$

Therefore, the remaining assertions follow from Theorems 2.8 and 2.10 and Remark 2.9(i) and (ii). □

PROOF OF REMARK 3.7(i).   Since

$$\frac{1}{r_n v_n} E\big(g_\phi(Y_n)^2 \mathbf{1}_{\{L(Y_n)>k\}}\big)$$

$$\le \|\phi\|_\infty \frac{1}{r_n v_n} E\left(\left(\sum_{i=1}^{r_n} \mathbf{1}_{\{X_{n,i}\neq 0\}}\right)^2 \mathbf{1}_{\{L(Y_n)>k\}}\right)$$

$$\le \|\phi\|_\infty \left(\frac{1}{r_n v_n} E\left(\left(\sum_{i=1}^{r_n} \mathbf{1}_{\{X_{n,i}\neq 0\}}\right)^{2+\delta}\right)\right)^{2/(2+\delta)} \left(\frac{1}{r_n v_n} P\{L(Y_n) > k\}\right)^{\delta/(2+\delta)},$$

the first part (2.1) of (C3′) follows from (3.7) and (3.9), since ϕ is assumed to be bounded. Next,

$$\frac{1}{r_n v_n} E\big(g_\phi(Y_n) g_\psi(Y_n) \mathbf{1}_{\{L(Y_n)\le k\}}\big)$$

(5.18)
$$= \frac{1}{r_n v_n} \sum_{i,j\in\{1,\ldots,r_n\},|i-j|\le k-1} E\big(\phi(X_{n,i})\psi(X_{n,j})\mathbf{1}_{\{L(Y_n)\le k\}}\big)$$

$$= \frac{1}{v_n} E(\phi(X_{n,1})\psi(X_{n,1}))$$

$$+ \sum_{i=1}^{k-1} \frac{r_n - i}{r_n} \frac{1}{v_n} \big(E(\phi(X_{n,1})\psi(X_{n,i+1}))$$

$$+ E(\psi(X_{n,1})\phi(X_{n,i+1})) \big) + R_{n,k},$$

with

$$|R_{n,k}| = \frac{1}{r_n v_n} \left| \sum_{i,j \in \{1,\ldots,r_n\}, |i-j| \le k-1} E\big(\phi(X_{n,i})\psi(X_{n,j})\mathbf{1}_{\{L(Y_n)>k\}}\big) \right|$$

$$\le \|\phi\|_\infty \|\psi\|_\infty \frac{1}{r_n v_n} E\left(\left(\sum_{i=1}^{r_n} \mathbf{1}_{\{X_{n,i}\neq 0\}} \right)^2 \mathbf{1}_{\{L(Y_n)>k\}} \right).$$

It then follows as above that $\lim_{k\to\infty} \limsup_{n\to\infty} |R_{n,k}| = 0$, and, hence, the assumption (2.2) of (C3′) can be seen to be satisfied, with c given by (3.10). □

PROOF OF COROLLARY 3.9. Clearly, (3.15) implies (2.10) and hence also (D2′). Moreover, (3.15) implies that

$$E\left(\sum_{i=1}^{r_n} \phi_{\max}(X_{n,i}) \right)^2 \le E\left(\sum_{i=1}^{r_n} \phi_{\max}(X_{n,i}) \right)^{2+\delta} + P\left\{ 0 < \sum_{i=1}^{r_n} \phi_{\max}(X_{n,i}) \le 1 \right\}$$

$$= O(r_n v_n).$$

Hence, similar arguments as used in the proof of Corollary 3.6 show that $(Z_n(g_\phi))_{\phi\in\Phi}$ converges weakly to a Gaussian process. Finally, (3.6) and thus the convergence of $(\tilde{Z}_n(\phi))_{\phi\in\Phi}$ follows from

$$E^*\left(\sup_{\phi\in\Phi} \frac{1}{\sqrt{nv_n}} \left| \sum_{i=r_n m_n+1}^{n} \big(\phi(X_{n,i}) - E\phi(X_{n,i})\big) \right| \right)^2$$

$$\le E\left(\frac{1}{\sqrt{nv_n}} \sum_{i=1}^{r_n} \big(\phi_{\max}(X_{n,i}) + E\phi_{\max}(X_{n,i})\big) \right)^2$$

$$\le \frac{4}{nv_n} E\left(\sum_{i=1}^{r_n} \phi_{\max}(X_{n,i}) \right)^2$$

$$= O(r_n/n) \to 0. \qquad \square$$

PROOF OF COROLLARY 4.3. (i) The index set $\mathcal{C} := \{C_{t_1,\ldots,t_k} \mid t_1,\ldots,t_k \in [0,1]\}$ equipped with the metric $\rho_{\mathcal{C}}(\mathbf{1}_{C_{s_1,\ldots,s_k}}, \mathbf{1}_{C_{t_1,\ldots,t_k}}) := \max_{1\le l\le k} |s_l - t_l|$ is totally bounded. The same holds for $\mathcal{D} := \{D_{t_1,\ldots,t_k} \mid t_1,\ldots,t_k \in [0,1]\}$.

In view of the discussion preceding Corollary 4.3, the assertions follow from Theorem 2.10 combined with Corollary 2.7 if we verify condition (D5) and that the

index sets $\mathcal{C}$ and $\mathcal{D}$ are VC-classes. Condition (D5) is satisfied since all processes under consideration are separable.

That $\mathcal{C}$ is a VC-class may be established by observing that $C_{t_1,\ldots,t_k} = \psi^{-1}(\bigtimes_{l=1}^{k}(t_l,\infty))$ with

$$\psi : \mathbb{R}_\cup \to \mathbb{R}^k, \qquad (x_1,\ldots,x_m) \mapsto \begin{cases} (x_j,\ldots,x_{j+k-1}), \\ \qquad \text{if } j = \min\{i \mid x_i \neq 0\} \leq m-k+1, \\ (0,\ldots,0), \\ \qquad \text{else.} \end{cases}$$

Since $\{\bigtimes_{l=1}^{k}(t_l,\infty) \mid t_1,\ldots,t_k \geq 0\}$ is known to be a VC-class [cf. van der Vaart and Wellner (1996), Example 2.6.1], $\mathcal{C}$ is a VC-class, too [van der Vaart and Wellner (1996), Lemma 2.6.17(v)].

The sets $\mathcal{D}_j := \{E_{j,t} \mid t \geq 0\}$ are linearly ordered (i.e., $E_{j,s} \subset E_{j,t}$ if $s > t$) and, hence, they are VC-classes, and hence so is

$$\mathcal{D} = \mathcal{D}_1 \sqcap \mathcal{D}_2 \sqcap \cdots \sqcap \mathcal{D}_k = \left\{ \bigcap_{j=1}^{k} E_j \;\Big|\; E_j \in \mathcal{D}_j \right\}$$

[van der Vaart and Wellner (1996), Lemma 2.6.17(ii)].

(ii) By the results of Segers (2003), condition (C3.1″) is satisfied in the weaker version discussed in Remark 2.6, because the limit r.v.'s are continuous on $(0,\infty)$ and the discontinuity sets have Lebesgue measure 0. Hence, the assertions follow by part (i), if the asymptotic equicontinuity condition (D3) can be shown.

For this, first note that $C_{s_1,\ldots,s_k} \triangle C_{t_1,\ldots,t_k} \subset \{(x_1,\ldots,x_m) \in E_\cup \mid m \in \mathbb{N}, \exists 0 \leq j \leq m-k, 1 \leq l \leq k : x_i = 0, \forall 1 \leq i \leq j, x_{j+l} \in (\min(s_l,t_l), \max(s_l,t_l)]\}$. Thus, Lemma 2.5(i) and (ii) yield that

$$\frac{1}{r_n v_n} P\{Y_n \in C_{s_1,\ldots,s_k} \triangle C_{t_1,\ldots,t_k}\}$$

$$\leq \frac{1}{r_n v_n \theta_n} P\big(X_n^{(r_n)} \in C_{s_1,\ldots,s_k} \triangle C_{t_1,\ldots,t_k} \mid X_{n,1} \neq 0\big) \cdot P\{Y_n \neq 0\}$$

$$\qquad + o\bigg(\frac{P\{Y_n \neq 0\}}{r_n v_n}\bigg)$$

$$= P\big(X_n^{(r_n)} \in C_{s_1,\ldots,s_k} \triangle C_{t_1,\ldots,t_k} \mid X_{n,1} \neq 0\big) + o(1)$$

$$\leq \sum_{l=1}^{k} P\big(X_{n,l} \in (\min(s_l,t_l), \max(s_l,t_l)] \mid X_{n,1} \neq 0\big) + o(1)$$

$$\leq \sum_{l=1}^{k} P\big(X_{n,l} \in (\min(s_l,t_l), \max(s_l,t_l)] \mid X_{n,l} \neq 0\big) \cdot \frac{P\{X_{n,l} \neq 0\}}{P\{X_{n,1} \neq 0\}} + o(1)$$

$$= \sum_{l=1}^{k} |t_l - s_l| + o(1),$$

where the term $o(1)$ tends to 0 uniformly for all $s_1, \ldots, s_k, t_1, \ldots, t_k \in [0, 1]$. Now, (D3) follows immediately from the definition of ρ_C.

To verify condition (D3) for the indicator functions describing the largest order statistics in a cluster, note that

$$
\bigcap_{j=1}^{k} E_{j,s_j} \triangle \bigcap_{j=1}^{k} E_{j,t_j}
$$

$$
\subset \left\{ (x_1, \ldots, x_m) \in E_\cup \mid m \in \mathbb{N}, \right.
$$

$$
\sum_{i=1}^{m} \mathbf{1}_{(\min(s_j,t_j),1]}(x_i) \geq j, \sum_{i=1}^{m} \mathbf{1}_{(\max(s_j,t_j),1]}(x_i) < j \text{ for some } 1 \leq j \leq k \Big\}
$$

$$
\subset \{ (x_1, \ldots, x_m) \in E_\cup \mid m \in \mathbb{N},
$$

$$
x_i \in (\min(s_j, t_j), \max(s_j, t_j)] \text{ for some } 1 \leq j \leq k, 1 \leq i \leq m \}.
$$

This implies

$$
\frac{1}{r_n v_n} P \left\{ Y_n \in \bigcap_{j=1}^{k} E_{j,s_j} \triangle \bigcap_{j=1}^{k} E_{j,t_j} \right\}
$$

$$
\leq \sum_{j=1}^{k} P \big(X_{n,1} \in (\min(s_j, t_j), \max(s_j, t_j)] \mid X_{n,1} \neq 0 \big)
$$

$$
= \sum_{j=1}^{k} |t_j - s_j|
$$

from which (D3) follows. □

PROOF OF THE RESULT IN EXAMPLE 4.4. The convergence of the fidis of $\tilde{Z}_n$ to those of a Gaussian process with covariance function (4.6) follows from Corollary 2.7 by the same arguments as in the proof of Corollary 4.3(ii).

In view of the discussion before Corollary 4.3, the proof will be completed by showing that conditions (D3), (D5) and (D6) of the asymptotic equicontinuity Theorem 2.10 also are satisfied. The measurability condition (D5) holds since, for fixed k, the processes $(\mathbf{1}_{C_{k,t_1,\ldots,t_k}})_{(t_1,\ldots,t_k) \in [0,1]^k}$ are separable and a supremum of countably many suprema of separable processes are measurable.

We will use (4.2) to verify that (D3) is satisfied for the semi-metric

$$
\rho(\mathbf{1}_{C_{j,s_1,\ldots,s_j}}, \mathbf{1}_{C_{k,t_1,\ldots,t_k}})
$$

$$
:= \begin{cases} P\{L(W) \in \{j, k\}\}, & \text{if } j \neq k, \\ P\{L(W) = k, W_i \in (s_i \wedge t_i, s_i \vee t_i] \text{ for some } 1 \leq i \leq k\}, & \text{if } j = k. \end{cases}
$$

Now, $\mathcal{F} = \{\mathbf{1}_{C_{k,t_1,\dots,t_k}} \mid k \geq 1, t_1, t_2, \dots \in [0,1]\}$ is totally bounded with respect to ρ. To see this, for $\epsilon > 0$ given, choose $0 = a_{i,0} < a_{i,1} < \cdots < a_{i,m_i} = 1$ such that $P\{W_i \in (a_{i,j-1}, a_{i,j}]\} \leq \epsilon/k_\epsilon$ for $1 \leq i \leq k_\epsilon$ and $1 \leq j \leq m_i$, with k_ϵ chosen large enough to make $P\{L(W) \geq k_\epsilon\} < \epsilon/2$. Then

$$\{\mathbf{1}_{C_{k,t_1,\dots,t_k}} \mid k \geq k_\epsilon\}, \qquad \{\mathbf{1}_{C_{j,t_1,\dots,t_j}} \mid t_i \in [a_{i,\ell_i-1}, a_{i,\ell_i}], \forall 1 \leq i \leq j\},$$

for $1 \leq j \leq k_\epsilon$, $1 \leq \ell_i \leq m_i$, is a finite cover of $\mathcal{F}$ with diameter at most ϵ.

By Lemma 2.5,

$$(5.19) \qquad P\big(L(Y_n) = k \mid Y_n \neq 0\big) \to \frac{1}{\theta}\big(P\{L(W) = k\} - P\{L(W^{(2;\infty)}) = k\}\big),$$

and, by Sheffe's lemma, the convergence is uniform in $k \in \mathbb{N}$. (Note that, for $k \leq l$, the cluster functional $\mathbf{1}_{\{k\}} \circ L$ is constant on all sets $N_{l,I}$ defined in Remark 2.6.) Similarly,

$$P\big(L(Y_n) = k, (Y_n^c)_1 \leq t_1, \dots, (Y_n^c)_k \leq t_k \mid Y_n \neq 0\big)$$

$$(5.20) \qquad \to \frac{1}{\theta}\big(P\{L(W) = k, W_1 \leq t_1, \dots, W_k \leq t_k\}$$

$$- P\{L(W^{(2;\infty)}) = k, ((W^{(2;\infty)})^c)_i \leq t_i, \forall 1 \leq i \leq k\}\big),$$

and the convergence is uniform in $t_1, \dots, t_k$ for each fixed k, because the right-hand side defines a continuous function.

For $\epsilon > 0$ let $\delta = \epsilon/2$ and consider $j, t_1, \dots, t_j, k, t_1, \dots, t_k$ such that $\rho(\mathbf{1}_{C_{j,s_1,\dots,s_j}}, \mathbf{1}_{C_{k,t_1,\dots,t_k}}) < \delta$. Then for $j \neq k$ and n large,

$$\frac{1}{r_n v_n} P\{Y_n \in C_{j,s_1,\dots,s_j} \Delta C_{k,t_1,\dots,t_k}\} \leq \frac{1}{r_n v_n} P\{L(Y_n) \in \{j,k\}\}$$

$$= \theta_n P\big(L(Y_n) \in \{j,k\} \mid Y_n \neq 0\big) \leq \epsilon$$

by (5.19), Lemma 2.5(ii) and the definition of ρ.

If instead $j = k \leq k_\epsilon$, then using (5.20), for large n,

$$\frac{1}{r_n v_n} P\{Y_n \in C_{j,s_1,\dots,s_j} \Delta C_{k,t_1,\dots,t_k}\}$$

$$= \theta_n P\big(L(Y_n) = k, (Y_n^c)_i \in (s_i \wedge t_i, s_i \vee t_i] \text{ for some } 1 \leq i \leq k \mid Y_n \neq 0\big)$$

$$\leq \theta_n \Big(\frac{1}{\theta} P\{L(W) = k, W_i \in (s_i \wedge t_i, s_i \vee t_i] \text{ for some } 1 \leq i \leq k\} + \frac{\epsilon}{4}\Big) \leq \epsilon,$$

again by Lemma 2.5 and the definition of ρ.

Finally, if $j = k > k_\epsilon$, then for large n

$$\frac{1}{r_n v_n} P(Y_n \in C_{j,s_1,\dots,s_j} \Delta C_{k,t_1,\dots,t_k}) \leq P\big(L(Y_n) = k \mid Y_n \neq 0\big)$$

$$\leq 2P\big(L(W) > k_\epsilon\big) < \epsilon.$$

This concludes the proof of (4.2), and hence also the proof of (D3).

For the proof of (D6), let $\mathcal{C}_k = \{C_{j,t_1,\ldots,t_j} \mid 1 \le j \le k, t_1, \ldots, t_j \in [0,1]\}$ and $\mathcal{F}_k = \{\mathbf{1}_C \mid C \in \mathcal{C}_k\}$ so that $\mathcal{F} = \bigcup_{k=1}^{\infty} \mathcal{F}_k$. Define ψ_k as the function which maps $x \in E_{\cup}$ to the vector $(1, \ldots, 1)$ in $\mathbb{R}^{2k}$ if $L(x) > k$ or $L(x) = 0$ and which maps x to the vector

$$(1, \ldots, 1, 0, 1, \ldots, 1, x_1^c, \ldots, x_j^c, 0, \ldots, 0) \in \mathbb{R}^{2k},$$

if $1 \le L(x) := j \le k$. Here the first row of ones has $j - 1$ entries and the second row has $k - j$ entries, and, hence, the vector ends with $k - j$ zeros, so that the first k components encode the length of the cluster core. With this definition, it follows that

$$C_{j,t_1,\ldots,t_j} = \psi_k^{-1}\left(\mathbb{R}^{j-1} \times (-\infty, 0] \times \mathbb{R}^{k-j} \times \mathop{\times}_{i=1}^{j} (-\infty, t_i] \times \mathbb{R}^{k-j}\right).$$

The left orthants $\times_{i=1}^{2k}(-\infty, x_i]$ form a VC-class with index bounded by $2k + 1$ [van der Vaart and Wellner (1996), Example 2.6.1] and, hence, also $\mathcal{C}_k$ is a VC-class with index bounded by $2k + 1$ [Dudley (1999), Theorem 4.2.3]. By van der Vaart and Wellner [(1996), Theorem 2.6.7] for all sufficiently small ϵ and all $k \in \mathbb{N}$, $\mathcal{F}_k$ satisfies the metric entropy bound

$$\begin{aligned}
N\left(\epsilon\left(\int F^2 \, dQ\right)^{1/2}, \mathcal{F}_k, d_Q\right) &\le C(2k+1)(16e)^{2k+1}\epsilon^{-(4k+1)} \\
&\le \epsilon^{-(6k+2)},
\end{aligned}$$

(5.21)

with C denoting a universal constant that does not depend on k or ϵ.

Let $L_{n,1} > L_{n,2} > \cdots > L_{n,m_n}$ be the order statistics in descending order of the independent cluster lengths $(L(Y_{n,j}^*))_{j=1}^{m_n}$. Since the empirical L_2-semi-metric d_n satisfies

$$\sup_{i,j>k} d_n^2(\mathbf{1}_{C_{i,t_1,\ldots,t_i}}, \mathbf{1}_{C_{j,s_1,\ldots,s_j}}) \le \frac{1}{nv_n} \sum_{j=1}^{m_n} \mathbf{1}_{\{L(Y_{n,j}^*)>k\}},$$

it follows that the squared diameter of the set

$$\{C_{j,t_1,\ldots,t_j} \mid j > L_{n,\lfloor \epsilon^2 nv_n \rfloor}, t_1, \ldots, t_j \in [0,1]\}$$

w.r.t. d_n is bounded by

$$\frac{1}{nv_n} \sum_{j=1}^{m_n} \mathbf{1}_{\{L(Y_{n,j}^*)>L_{n,\lfloor \epsilon^2 nv_n \rfloor}\}} \le \frac{\lfloor \epsilon^2 nv_n \rfloor}{nv_n} \le \epsilon^2.$$

Reasoning as in the last part of the proof of Theorem 2.10, this together with (5.21) shows that (D6) follows if we prove that

(5.22) $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P\left\{\int_0^{\delta} \sqrt{\log \epsilon^{-(6L_{n,\lfloor \epsilon^2 nv_n \rfloor}+2)}} \, d\epsilon > \tau\right\} = 0$$

for all $\tau > 0$. By a change of variables and Hölder's inequality,

$$\int_0^\delta \sqrt{\log \epsilon^{-(6L_{n,\lfloor \epsilon^2 n v_n \rfloor} + 2)}} \, d\epsilon$$

$$\leq \sum_{j=1}^{\lceil \delta n v_n \rceil} \sqrt{8 L_{n,j}} \int_{(j/(n v_n))^{1/2}}^{((j+1)/(n v_n))^{1/2}} \sqrt{|\log \epsilon|} \, d\epsilon$$

$$\leq \frac{2}{n v_n} \sum_{j=1}^{\lceil \delta n v_n \rceil} \sqrt{L_{n,j}} \cdot n v_n \int_{j/(n v_n)}^{(j+1)/(n v_n)} \sqrt{\log \eta^{-1/2}} \eta^{-1/2} \, d\eta$$

$$\leq \left(\frac{1}{n v_n} \sum_{j=1}^{\lceil \delta n v_n \rceil} L_{n,j}^{1+\zeta} \right)^{1/(2+2\zeta)}$$

$$\times \left(\frac{1}{n v_n} \sum_{j=1}^{\lceil \delta n v_n \rceil} \left(n v_n \int_{j/(n v_n)}^{(j+1)/(n v_n)} \sqrt{\log \eta^{-1/2}} \right.\right.$$

$$\left.\left. \times \eta^{-1/2} \, d\eta \right)^{(2+2\zeta)/(1+2\zeta)} \right)^{(1+2\zeta)/(2+2\zeta)}.$$

Now,

$$E \left(\frac{1}{n v_n} \sum_{j=1}^{\lceil \delta n v_n \rceil} L_{n,j}^{1+\zeta} \right) \leq E \left(\frac{1}{n v_n} \sum_{j=1}^{m_n} L_n(Y_{n,j}^*)^{1+\zeta} \right) \leq E\bigl(L(Y_n)^{1+\zeta} \mid Y_n \neq 0 \bigr),$$

which is bounded by (4.5). Furthermore, applying Liapunov's inequality to the individual summands,

$$\frac{1}{n v_n} \sum_{j=1}^{\lceil \delta n v_n \rceil} \left(n v_n \int_{j/(n v_n)}^{(j+1)/(n v_n)} \sqrt{\log \eta^{-1/2}} \eta^{-1/2} \, d\eta \right)^{(2+2\zeta)/(1+2\zeta)}$$

$$\leq \frac{1}{n v_n} \sum_{j=1}^{\lceil \delta n v_n \rceil} n v_n \int_{j/(n v_n)}^{(j+1)/(n v_n)} \left(\frac{|\log \eta|}{\eta} \right)^{(1+\zeta)/(1+2\zeta)} \, d\eta$$

$$\leq \int_0^{2\delta} \left(\frac{|\log \eta|}{\eta} \right)^{(1+\zeta)/(1+2\zeta)} \, d\eta \to 0$$

as $\delta \to 0$. Hence, we have verified (5.22). This concludes the proof of (D6). $\square$

REFERENCES

BORTOT, P. and TAWN, J. A. (1998). Models for the extremes of Markov chains. *Biometrika* **85** 851–867. MR1666683

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*. Springer, New York. MR2234156

DREES, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. MR1810875

DREES, H. (2002). Tail empirical processes under mixing conditions. In *Empirical Process Techniques for Dependent Data* (H. G. Dehling, T. Mikosch and M. Sørensen, eds.) 325–342. Birkhäuser, Boston. MR1958788

DREES, H. (2003). Extreme quantile estimation for dependent data with applications to finance. *Bernoulli* **9** 617–657. MR1996273

DREES, H. (2010). Bias correction for blocks estimators of the extremal index. Univ. Hamburg. Preprint.

DUDLEY, R. (1999). *Uniform Central Limit Theorems*. Cambridge Univ. Press, Cambridge. MR1720712

EBERLEIN, E. (1984). Weak convergence of partial sums of absolutely regular sequences. *Statist. Probab. Lett.* **2** 291–293. MR0777842

EINMAHL, J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Process. Appl.* **70** 31–58. MR1472958

FABIAN, V. (1970). On uniform convergence of measures. *Z. Wahrsch. Verw. Gebiete* **15** 139–143. MR0277679

GINÉ, E. and MASON, D. M. (2008). Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.* **35** 739–761. MR2468873

GINÉ, E., MASON, D. M. and ZAITSEV, A. Y. (2003). The L_1-norm density estimator process. *Ann. Probab.* **31** 719–768. MR1964947

HAHN, M. G. (1977). Conditions for sample-continuity and the central limit theorem. *Ann. Probab.* **5** 351–360. MR0440679

LEADBETTER, M. R. (1995). On high level exceedance modeling and tail inference. *J. Statist. Plann. Inference* **45** 247–260. MR1342098

LEADBETTER, M. R. and ROOTZÉN, H. (1993). On central limit theory for families of strongly mixing additive random functions. In *Stochastic Processes: A Festschrift in Honour of Gopinath Kallianpur* (S. Cambanis et al., eds.) 211–223. Springer, New York. MR1427317

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin. MR0388499

ROOTZÉN, H. (1995). The tail empirical process for stationary sequences. Preprint, Chalmers Univ.

ROOTZÉN, H. (2009). Weak convergence of the tail empirical function for dependent sequences. *Stochastic Process. Appl.* **119** 468–490. MR2494000

ROOTZÉN, H., LEADBETTER, M. R. and DE HAAN, L. (1990). Tail and quantile estimators for strongly mixing stationary processes. Report, Dept. Statistics, Univ. North Carolina.

ROOTZÉN, H., LEADBETTER, M. R. and DE HAAN, L. (1998). On the distribution of tail array sums for strongly mixing stationary sequences. *Ann. Appl. Probab.* **8** 868–885. MR1627799

SEGERS, J. (2003). Functionals of clusters of extremes. *Adv. in Appl. Probab.* **35** 1028–1045. MR2014268

SISSON, S. and COLES, S. (2003). Modelling dependence uncertainty in the extremes of Markov chains. *Extremes* **6** 283–300. MR2125494

STOTT, P. A., STONE, D. A. and ALLEN, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature* **432** 610–613.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

WELLNER, J. A. and ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814. MR1792787

WELLNER, J. A. and ZHANG, Y. (2007). Two likelihood-bsed semiparametric estimation methods for panel count data with covariates *Ann. Statist.* **35** 2106–2142. MR2363965

YUN, S. (2000). The distribution of cluster functionals of extreme events in a dth-order Markov chain. *J. Appl. Probab.* **37** 29–44. MR1761659

UNIVERSITY OF HAMBURG
DEPARTMENT OF MATHEMATICS, SPST
BUNDESSTR. 55
20146 HAMBURG
GERMANY
E-MAIL: holger.drees@uni-hamburg.de

CHALMERS UNIVERSITY
DEPARTMENT OF MATHEMATICAL STATISTICS
412 96 GÖTEBORG
SWEDEN
AND
GOTHENBURG UNIVERSITY
DEPARTMENT OF MATHEMATICAL STATISTICS
412 96 GÖTEBORG
SWEDEN
E-MAIL: rootzen@math.chalmers.se