Point process convergence for symmetric functions of high-dimensional random vectors

Johannes Heiny[1] · **Carolin Kleemann[2]**

Abstract

The convergence of a sequence of point processes with dependent points, defined by a symmetric function of iid high-dimensional random vectors, to a Poisson random measure is proved. This also implies the convergence of the joint distribution of a fixed number of upper order statistics. As applications of the result a generalization of maximum convergence to point process convergence is given for simple linear rank statistics, rank-type U-statistics and the entries of sample covariance matrices.

Keywords Point process convergence · Extreme value theory · Poisson process · Gumbel distribution · High-dimensional data · U-statistics · Kendall's tau · Spearman's rho

AMS 2000 Subject Classifications Primary 60G55 · Secondary 60G70 · 60B12

1 Introduction

In classical extreme value theory the asymptotic distribution of the maximum of random points plays a central role. Maximum type statistics build popular tests on the dependency structure of high-dimensional data. Especially, against sparse alternatives those tests possess good power properties (see Han et al. 2017; Drton et al. 2020; Zhou et al. 2019). Closely related to the maxima of random points are point processes, which play an important role in stochastic geometry and data analysis. They have applications

✉ Johannes Heiny
johannes.heiny@math.su.se

Carolin Kleemann
carolin.kleemann@rub.de

[1] Department of Mathematics, Stockholm University, Albano hus 1, 10691 Stockholm, Sweden

[2] Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstrasse 150, D-44801 Bochum, Germany

in statistical ecology, astrostatistics and spatial epidemiology (Baddeley 2007). For a sequence $(Y_i)_i$ of real-valued random variables, we set

$$\widetilde{M}_p := \sum_{i=1}^{p} \varepsilon_{(i/p, Y_i)},$$

where ε_x is the Dirac measure in x. Let $K := (0, 1) \times (u, \infty)$ with $u \in \mathbb{R}$. Then, $\widetilde{M}_p(K)$ counts the number of exceedances of the threshold u by the random variables $Y_1, \ldots, Y_p$. If $Y^{(k)}$ denotes the k-th upper order statistic of $Y_1, \ldots, Y_p$, it holds that $\{\widetilde{M}_p(K) < k\} = \{Y^{(k)} \leq u\}$, and in particular $\{\widetilde{M}_p(K) = 0\} = \{\max_{i=1,\ldots,p} Y_i \leq u\}$. Therefore, the weak convergence of a sequence of point processes gives information about the joint asymptotic distribution of a fixed number of upper order statistics. If the sequence $(Y_i)_i$ consists of independent and identically distributed (iid) random variables, maximum convergence and point process convergence are equivalent, but if the random variables exhibit dependency, this equivalence does not necessarily hold anymore. In this sense, point process convergence is a substantial generalization of the maximum convergence. Additionally, the time components i/p deliver valuable information of the random time points when a record occurs, i.e., the time points when $Y_j > \max_{i=1,\ldots,j-1} Y_i$.

Our main motivation comes from statistical inference for high-dimensional data, where the asymptotic distribution of the maximum of dependent random variables has found several applications in recent years (see for example Han et al. 2017; Drton et al. 2020; Zhou et al. 2019; Cai and Jiang 2011; Cai et al. 2013; Cai 2017; Cai and Liu 2011; Gösmann et al. 2022). The objective of this paper is to provide the methodology to extend meaningful results with reference to the convergence of the maximum of dependent random variables, to point process convergence.

To this end, we consider dependent points $T_{\mathbf{i}} := g_{n,p}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_m})$, where the index $\mathbf{i} = (i_1, i_2, \ldots, i_m) \in \{1, \ldots, p\}^m$. The random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are iid on $\mathbb{R}^n$ and $g_{n,p} : \mathbb{R}^{mn} \to \mathbb{R}$ is a measurable, symmetric function. Important examples include U-statistics, simple linear rank statistics, rank-type U-statistics, the entries of sample covariance matrices or interpoint distances.

Additionally, we assume that the dimension of the points n is growing with the number of points p. Over the last decades the environment and therefore the requirements for statistical methods have changed fundamentally. Due to the huge improvement of computing power and data acquisition technologies one is confronted with large data sets, where the dimension of observations is as large or even larger than the sample size. These high-dimensional data occur naturally in online networks, genomics, financial engineering, wireless communication or image analysis (see Johnstone and Titterington 2009; Clarke et al. 2008; Donoho 2000). Hence, the analysis of high-dimensional data has developed as a meaningful and active research area.

We will show that the corresponding point process of the points $T_{\mathbf{i}}$ converges to a Poisson random measure (PRM) with a mean measure that involves the m-dimensional Lebesgue measure and an additional measure μ. If we replace the points $T_{\mathbf{i}}$ with iid random variables with the same distribution, the (non-degenerate) limiting distribution of the maximum will necessarily be an extreme value distribution of the

form $\exp(-\mu(x))$. Moreover, the convergence of the corresponding point process will be equivalent to the condition

$$\binom{p}{m} \mathbb{P}(g_{n,p}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) > x) \to \mu(x), \qquad p \to \infty. \tag{1.1}$$

However, since the random points $T_{\mathbf{i}}$ are not independent, we additionally need the following assumption on the dependence structure

$$\mathbb{P}(g_{n,p}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) > x, g_{n,p}(\mathbf{x}_{m-l+1}, \ldots, \mathbf{x}_{2m-l}) > x) = o(p^{-(2m-l)}), \qquad p \to \infty, \tag{1.2}$$

where $l = 1, \ldots, m - 1$.

In the finite-dimensional case where n is fixed, several results about point process convergence are available in similar settings. In Silverman and Brown (1978), Silverman and Brown showed point process convergence for $m = 2$, $n = 2$ and $g_{2,p}(\mathbf{x}_i, \mathbf{x}_j) = a_p \|\mathbf{x}_i - \mathbf{x}_j\|_2^2$, where the $\mathbf{x}_i$ have a bounded and almost everywhere continuous density, a_p is a suitable scaling sequence and $\| \cdot \|_2$ is the Euclidean norm on $\mathbb{R}^2$. In the Weibull case $\mu(x) = x^\alpha$ for $x, \alpha > 0$, Dabrowski et al. (2002) proved a generalization to points with a fixed dimension and $g_{n,p}(\mathbf{x}_i, \mathbf{x}_j) = a_p h(\mathbf{x}_i, \mathbf{x}_j)$, where h is a measurable, symmetric function and a_p is a suitable scaling sequence.

Also in the finite-dimensional case, under similar assumptions as in (1.1) with $\mu(x) = \beta x^\alpha$ for $x, \alpha > 0$, $\beta \in \mathbb{R}$ and under condition (1.2), Schulte and Thäle (2012) showed convergence in distribution of point processes towards a Weibull process. The points of these point processes are obtained by applying a symmetric function $g_{n,p}$ to all m-tuples of distinct points of a Poisson process on a standard Borel space. In Schulte and Thäle (2016), this result was extended to more general functions μ and to binomial processes so that other PRMs were possible limit processes. In Decreusefond et al. (2016), Decreusefond, Schulte and Thäle provided an upper bound of the Kantorovich-Rubinstein distance between a PRM and the point process induced in the aforementioned way by a Poisson or a binomial process on an abstract state space. Notice that convergence in Kantorovich-Rubinstein distance implies convergence in distribution (see Panaretos and Zemel 2020, Theorem 2.2.1 or Decreusefond et al. 2016, p. 2149). In Chenavier et al. (2022) another point process result in a similar setting is given for the number of nearest neighbor balls in fixed dimension. Moreover, Basrak and Planinić (2021) presents a general framework for Poisson approximation of point processes on Polish spaces.

1.1 Structure of this paper

The remainder of this paper is structured as follows. In Section 2 we prove weak point process convergence for the dependent points $T_{\mathbf{i}}$ in the high-dimensional case as tool for the generalization of the convergence of the maximum (Theorem 2.1). We provide popular representations of the limiting process in terms of the transformed points of a homogeneous Poisson process. Moreover, we derive point process convergence for the record times. In Section 3 these tools are applied to study statistics based on relative ranks like simple linear rank statistics or rank-type U-statistics. We

also prove convergence of the point processes of the off-diagonal entries of large sample covariance matrices. The technical proofs are deferred to Section 4.

1.2 Notation

Convergence in distribution (resp. probability) is denoted by $\xrightarrow{d}$ (resp. $\xrightarrow{\mathbb{P}}$) and unless explicitly stated otherwise all limits are for $n \to \infty$. For sequences $(a_n)_n$ and $(b_n)_n$ we write $a_n = O(b_n)$ if $a_n/b_n \leq C$ for some constant $C > 0$ and every $n \in \mathbb{N}$, and $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. Additionally, we use the notation $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$ and $a_n \lesssim b_n$ if a_n is smaller than or equal to b_n up to a positive universal constant. We further write $a \wedge b := \min\{a, b\}$ for $a, b \in \mathbb{R}$ and for a set A we denote $|A|$ as the number of elements in A.

2 Point process convergence

We introduce the model that was briefly described in the introduction. Let $\mathbf{x}_1, \dots, \mathbf{x}_p$ be iid $\mathbb{R}^n$-valued random vectors with $\mathbf{x}_i = (X_{i1}, \dots, X_{in})^\top, i = 1, \dots, p$, where $p = p_n$ is some positive integer sequence tending to infinity as $n \to \infty$.

We consider the random points

$$T_{\mathbf{i}} := g_{n,p}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_m}),$$

where $\mathbf{i} = (i_1, i_2, \dots, i_m) \in \{1, \dots, p\}^m$ and $g_n = g_{n,p} : \mathbb{R}^{mn} \to \mathbb{R}$ is a measurable and symmetric function, where symmetric means $g_n(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m) = g_n(\mathbf{y}_{\pi(1)}, \mathbf{y}_{\pi(2)}, \dots, \mathbf{y}_{\pi(m)})$ for all $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m \in \mathbb{R}^n$ and all permutations π on $\{1, 2, \dots, m\}$. We are interested in the limit behavior of the point processes M_n towards a PRM M,

$$M_n = \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq p} \varepsilon_{(\mathbf{i}/p, T_{\mathbf{i}})} \xrightarrow{d} M, \qquad n \to \infty,$$

where $\mathbf{i}/p = (i_1/p, \dots, i_m/p)$. The limit M is a PRM with mean measure

$$\eta\left(\bigotimes_{l=1}^{m+1} (r_l, s_l) \right) = m! \lambda_m\left(\bigotimes_{l=1}^{m} (r_l, s_l) \right) \mu(r_{m+1}, s_{m+1}),$$

where λ_m is the Lebesgue measure on $\mathbb{R}^m$. For an interval (a, b) with $a < b \in \mathbb{R}$ we write $\mu(a, b) := \mu((a, b)) := \mu(a) - \mu(b)$ and $\mu : (v, w) \to \mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$ is a function satisfying $\lim_{x\to v} \mu(x) = \infty$ and $\lim_{x\to w} \mu(x) = 0$ for $v, w \in \bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ and $v < w$. Furthermore, we set $\eta_n(\cdot) := \mathbb{E}[M_n(\cdot)]$. We consider the M_n's and M as random measures on the state space

$$S = S_1 \times (v, w) = \{(z_1, z_2, \dots, z_m) : 0 < z_1 < z_2 < \dots < z_m \leq 1\} \times (v, w)$$

with values in $\mathcal{M}(S)$ the space of point measures on S, endowed with the vague topology (see Resnick 2008). The following result studies the convergence $M_n \xrightarrow{d} M$, which denotes the convergence in distribution in $\mathcal{M}(S)$.

Theorem 2.1 *Let* $\mathbf{x}_1, \ldots, \mathbf{x}_p$ *be n-dimensional, independent and identically distributed random vectors and* $p = p_n$ *is some sequence of positive integers tending to infinity as* $n \to \infty$. *Additionally, let* $g = g_n : \mathbb{R}^{mn} \to (v, w)$ *be a measurable and symmetric function, where* $v, w \in \bar{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$ *and* $v < w$. *Assume that there exists a function* $\mu : (v, w) \to \mathbb{R}^+$ *with* $\lim_{x \to v} \mu(x) = \infty$ *and* $\lim_{x \to w} \mu(x) = 0$ *such that, for* $x \in (v, w)$ *and* $n \to \infty$,

(A1) $\binom{p}{m} \mathbb{P}(g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) > x) \to \mu(x)$ and

(A2) $\quad \mathbb{P}(g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) > x, g_n(\mathbf{x}_{m-l+1}, \ldots, \mathbf{x}_{2m-l}) > x) = o(p^{-(2m-l)}) \qquad$ for $l = 1, \ldots, m-1$.

Then we have $M_n \xrightarrow{d} M$.

Note that (A1) ensures the correct specification of the mean measure, while (A2) is an anti-clustering condition. Both conditions are standard in extreme value theory. It is worth mentioning that

$$\lim_{n \to \infty} \mathbb{P}\left(\max_{1 \le i_1 < i_2 < \ldots < i_m \le p} T_{\mathbf{i}} \le x \right) = \exp(-\mu(x)) =: H(x), \qquad x \in \mathbb{R},$$

where we use the conventions $\mu(x) = 0$ if $x > w$, $\mu(x) = \infty$ if $x < v$, and $\exp(-\infty) = 0$. The typical distribution functions H are the Fréchet, Weibull and Gumbel distributions. In these cases, the limiting process M has a representation in terms of the transformed points of a homogeneous Poisson process. Let $(U_i)_i$ be an iid sequence of random vectors uniformly distributed on S_1 and $\Gamma_i = E_1 + \ldots + E_i$, where $(E_i)_i$ is an iid sequence of standard exponentially distributed random variables, independent of $(U_i)_i$.

It is well–known that $N_\Gamma := \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i}$ is a homogeneous Poisson process and hence it holds for every $A \subset (0, \infty)$ that $N_\Gamma(A)$ is Poisson distributed with parameter $\lambda_1(A)$ (see for example Embrechts et al. 1997, Example 5.1.10). For the mean measure η of M we get for a product of intervals $\bigotimes_{l=1}^{m}(r_l, s_l] \subset S_1$

$$\eta\left(\bigotimes_{l=1}^{m+1} (r_l, s_l] \right) = m! \lambda_m\left(\bigotimes_{l=1}^{m} (r_l, s_l] \right)(\mu(r_{m+1}) - \mu(s_{m+1}))$$

$$= \sum_{i=1}^{\infty} \mathbb{P}\left(U_i \in \bigotimes_{l=1}^{m}(r_l, s_l] \right) \mathbb{E}[\varepsilon_{\Gamma_i}(\mu(s_{m+1}), \mu(r_{m+1})]$$

$$= \mathbb{E}\left[\sum_{i=1}^{\infty} \varepsilon_{(U_i, \Gamma_i)}\left(\bigotimes_{l=1}^{m} (r_l, s_l] \times (\mu(s_{m+1}), \mu(r_{m+1})] \right) \right],$$

where we used in the second line that

$$\mathbb{P}\left(U_i \in \bigotimes_{l=1}^{m}(r_l, s_l] \right) = m! \prod_{l=1}^{m}(s_l - r_l)$$

as U_i is uniformly distributed on S_1 for every i and

$$\mathbb{E}[N_\Gamma(\mu(s_{m+1}), \mu(r_{m+1}))] = \lambda_1(\mu(s_{m+1}), \mu(r_{m+1})).$$

We get the following representations for the limiting processes M.

- Fréchet case: For $\alpha > 0$ the Fréchet distribution is given by $\Phi_\alpha(x) = \exp(-x^{-\alpha})$, $x > 0$. For $0 < r < s < \infty$ we have $\mu_{\Phi_\alpha}(r, s] = r^{-\alpha} - s^{-\alpha}$ and therefore, we can write

$$M = M_{\Phi_\alpha} = \sum_{i=1}^{\infty} \varepsilon_{(U_i, \Gamma_i^{-1/\alpha})}.$$

- Weibull case: For $\alpha > 0$ the Weibull distribution is given by $\Psi_\alpha(x) = \exp(-|x|^\alpha)$, $x < 0$. For $-\infty < r < s < 0$ we have $\mu_{\Psi_\alpha}(r, s] = |r|^\alpha - |s|^\alpha$ and

$$M = M_{\Psi_\alpha} = \sum_{i=1}^{\infty} \varepsilon_{(U_i, -\Gamma_i^{1/\alpha})}.$$

- Gumbel case: The Gumbel distribution is given by $\Lambda(x) = \exp(-e^{-x})$ for all $x \in \mathbb{R}$. For $-\infty < r < s < \infty$ we have $\mu_\Lambda(r, s] = e^{-r} - e^{-s}$ and

$$M = M_\Lambda = \sum_{i=1}^{\infty} \varepsilon_{(U_i, -\log \Gamma_i)}.$$

Besides the points $T_\mathbf{i}$, time components $\mathbf{i}/p = (i_1/p, \ldots, i_m/p)$ with $1 \leq i_1 \leq \ldots \leq i_m \leq p$ are considered in the definition of the point process M_n. Whenever we do not need the time components in the following, we will use the shorthand notation

$$N_n(\cdot) := M_n(S_1 \times \cdot) = \sum_{1 \leq i_1 \leq \ldots \leq i_m \leq p} \varepsilon_{T_\mathbf{i}}(\cdot). \tag{2.3}$$

Under the conditions of Theorem 2.1, N_n converges in distribution to $N(\cdot) := M(S_1 \times \cdot)$ which is a PRM with mean measure μ.

A direct consequence of the point process convergence is the convergence of the joint distribution of a fixed number of upper order statistics. In the Fréchet, Weibull and Gumbel cases the limit function can be described as the joint distribution function of transformations of the points Γ_i.

Corollary 2.2 *Let $G_{n,(j)}$ be the j-th upper order statistic of the random variables $(g_n(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_m}))$, where $1 \leq i_1 < i_2 < \ldots < i_m \leq p$. Under the conditions of Theorem 2.1 and for a fixed $k \geq 1$ the distribution function*

$$\mathbb{P}(G_{n,(1)} \leq x_1, \ldots, G_{n,(k)} \leq x_k),$$

where $x_k < \ldots < x_1 \in (v, w)$, converges to

$$\mathbb{P}\Big(N(x_1, w) = 0, N(x_2, w) \leq 1 \ldots, N(x_k, w) \leq k - 1\Big),$$

as $n \to \infty$. *In particular, in the Fréchet, Weibull and Gumbel cases, it holds that*

$$\mathbb{P}\Big(N(x_1, w) = 0, N(x_2, w) \leq 1 \dots, N(x_k, w) \leq k - 1\Big)$$

$$= \begin{cases} \mathbb{P}(\Gamma_1^{-1/\alpha} \leq x_1, \dots, \Gamma_k^{-1/\alpha} \leq x_k), & \text{if } \mu = \mu_{\Phi_\alpha}, \\ \mathbb{P}(-\Gamma_1^{1/\alpha} \leq x_1, \dots, -\Gamma_k^{1/\alpha} \leq x_k), & \text{if } \mu = \mu_{\Psi_\alpha}, \\ \mathbb{P}(-\log\Gamma_1 \leq x_1, \dots, -\log\Gamma_k \leq x_k), & \text{if } \mu = \mu_\Lambda. \end{cases}$$

Proof Since $N_n(x, w)$ is the number of vectors $\mathbf{i} = (i_1, \dots, i_m)$ with $1 \leq i_1 < i_2 < \dots < i_m \leq p$, for which $g_n(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_m}) \in (x, w)$, we get by Theorem 2.1 as $n \to \infty$

$$\mathbb{P}(G_{n,(1)} \leq x_1, \dots, G_{n,(k)} \leq x_k) = \mathbb{P}\Big(N_n(x_1, w) = 0, N_n(x_2, w) \leq 1, \dots, N_n(x_k, w) \leq k - 1\Big)$$

$$\to \mathbb{P}\Big(N(x_1, w) = 0, N(x_2, w) \leq 1 \dots, N(x_k, w) \leq k - 1\Big). \tag{2.4}$$

By the representation of the limiting point process in the Fréchet, Weibull and Gumbel cases, (2.4) is equal to one of the three distribution functions in the corollary. $\square$

One field, where point processes find many applications, is stochastic geometry. The paper Schulte and Thäle (2012), for example, considers order statistics for Poisson k-flats in $\mathbb{R}^d$, Poisson polytopes on the unit sphere and random geometric graphs.

Setting $k = 1$ in Corollary 2.2 we obtain the convergence in distribution of the maximum of the points $T_{\mathbf{i}}$.

Corollary 2.3 *Under the conditions of Theorem* 2.1 *we get*

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{1 \leq i_1 < i_2 < \dots < i_m \leq p} g_n(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_m}) \leq x\Big) = \exp(-\mu(x)), \qquad x \in \mathbb{R}.$$

Example 2.4 (Interpoint distances) Let $\mathbf{x}_i = (X_{i1}, \dots, X_{in})^\top$, $i = 1, \dots, p$ be n-dimensional random vectors, whose components $(X_{it})_{i,t \geq 1}$ are independent and identically distributed random variables with zero mean and variance 1. We are interested in the asymptotic behavior of the largest interpoint distances

$$D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 = \sum_{t=1}^n (X_{it} - X_{jt})^2, \qquad 1 \leq i < j \leq p,$$

where $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^n$. Figure 1 shows the four largest interpoint distances of 500 points on $\mathbb{R}^2$ with independent standard normal distributed components. Note that three of the largest four distances involve the same outlying vector $\mathbf{x}_i$.

We assume that there exists $s > 2$ such that $\mathbb{E}[|X_{11}|^{2s}(\log(|X_{11}|))^{s/2}] < \infty$ and $\mathbb{E}[X_{11}^4] \leq 5$ and that $p = p_n \to \infty$ satisfies $p = O(n^{(s-2)/4})$. Additionally, we let $(b_n)_n$ and $(c_n)_n$ be sequences given by

Fig. 1 Four largest distances between 500 normal distributed points
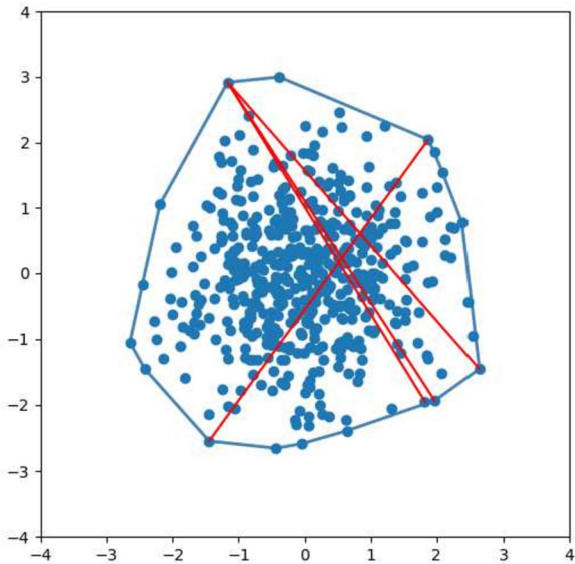
$$b_n = 2n + \sqrt{2n(\mathbb{E}[X^4] + 1)}d_n \quad \text{and} \quad c_n = \frac{d_n}{\sqrt{2n(\mathbb{E}[X^4] + 1)}},$$

where $d_n = \sqrt{2\log\tilde{p}} - \frac{\log\log\tilde{p} + \log 4\pi}{2(2\log\tilde{p})^{1/2}}$ with $\tilde{p} = p(p-1)/2$. For $x \in \mathbb{R}$ one can check that

$$\tilde{p}\,\mathbb{P}\big(c_n(D_{12} - b_n) > x\big) \to e^{-x} \quad \text{and} \quad \mathbb{P}\big(c_n(D_{12} - b_n) > x, c_n(D_{23} - b_n) > x\big) = o(p^{-3})$$

as $n \to \infty$ (see Heiny and Kleemann (2023) for details). Therefore, the conditions (A1) and (A2) in Theorem 2.1 hold for $m = 2$, $g_n(\mathbf{x}_i, \mathbf{x}_j) = c_n(D_{ij} - b_n)$ and $\mu(x) = e^{-x}$. By virtue of Theorem 2.1 we have

$$\sum_{1 \le i < j \le p} \varepsilon_{c_n(D_{ij} - b_n)} \xrightarrow{d} N_\Lambda = \sum_{i=1}^{\infty} \varepsilon_{-\log\Gamma_i}.$$

Finally Corollary 2.2 yields for a fixed $k \ge 1$ that

$$(D_{n,(1)}, \dots, D_{n,(k)}) \xrightarrow{d} (-\log\Gamma_1, \dots, -\log\Gamma_k),$$

where $D_{n,(\ell)}$ is the ℓ-th upper order statistic of the random variables $c_n(D_{ij} - b_n)$ for $1 \le i < j \le p$.

2.1 Record times

In Theorem 2.1 we showed convergence of point processes including time components. Therefore, we can additionally derive results for the record times $L(k), k \ge 1$

of the running maxima of the points $T_{\mathbf{i}} = g_n(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_m})$ for $\mathbf{i} = (i_1, \ldots, i_m)$, which are recursively defined as follows:

$$L(1) = 1,$$
$$L(k+1) = \inf\{j > L(k) : \max_{1 \le i_1 < \ldots < i_m \le j} T_{\mathbf{i}} > \max_{1 \le i_1 < \ldots < i_m \le L(k)} T_{\mathbf{i}}\}, \qquad k \in \mathbb{N},$$

(c.f. Sections 5.4.3 and 5.4.4 of Embrechts et al. 1997). To prove point process convergence for the record times we need the convergence in distribution of the sequence of processes $(Y_n(t), 0 < t \le 1)$ in $D(0, 1]$, the space of right continuous functions on $(0, 1]$ with finite limits existing from the left, defined by

$$Y_n(t) = \begin{cases} \max\limits_{1 \le i_1 < i_2 < \ldots < i_m \le \lfloor pt \rfloor} g_n(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_m}), & \frac{m}{p} \le t \le 1 \\ g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m), & \text{otherwise,} \end{cases}$$

where $\lfloor x \rfloor = \max\{y \in \mathbb{Z} : y \le x\}$ for $x \in \mathbb{R}$, towards an extremal process. We call $Y = (Y(t))_{t>0}$ an extremal process generated by the distribution function H, if the finite-dimensional distributions are given by

$$\mathbb{P}(Y(t_1) \le x_1, \ldots, Y(t_k) \le x_k) = H^{t_1}\Big(\bigwedge_{i=1}^{k} x_i\Big) H^{t_2-t_1}\Big(\bigwedge_{i=2}^{k} x_i\Big) \ldots H^{t_k-t_{k-1}}(x_k), \tag{2.5}$$

where $k \ge 1$, $0 < t_1 < \ldots < t_k$, $x_i \in \mathbb{R}$ and $1 \le i \le k$ (see Embrechts et al. 1997, Definition 5.4.3). To define convergence in distribution in $D(0, 1]$ we first need to introduce a metric $\mathcal{D}$ on $D(0, 1]$. To this end, let $\Lambda_{[0,1]}$ be a set of homeomorphisms

$$\Lambda_{[0,1]} = \{\lambda : [0, 1] \to [0, 1] : \lambda(0) = 0, \lambda(1) = 1, \lambda \text{ is continuous and strictly increasing}\}.$$

Then for $f, g \in D[0, 1]$ the Skorohod metric $\tilde{\mathcal{D}}$ is defined by (see Billingsley 1999, Section 12)

$$\tilde{\mathcal{D}}(f, g) := \inf\{\epsilon > 0 : \text{ there exists a } \lambda \in \Lambda_{[0,1]} \text{ such that}$$
$$\sup_{0 \le t \le 1} |\lambda(t) - t| \le \epsilon, \sup_{0 \le t \le 1} |f(t) - g(\lambda(t))| \le \epsilon\}.$$

Now set

$$\mathcal{D}(f, g) := \tilde{\mathcal{D}}(\tilde{f}, \tilde{g}), \quad f, g \in D(0, 1],$$

where $\tilde{f}$ and $\tilde{g}$ are the right continuous extensions of f and g on $[0, 1]$. The space of functions $D[0, 1]$ and therefore $D(0, 1]$ is separable under the Skorohod metric but not complete. However, one can find an equivalent metric, i.e., a metric which generates the same Skorohod topology, under which $D[0, 1]$ is complete (see Billingsley 1999, Theorem 12.2). In particular, the Skorohod metric and the equivalent metric generate the same open sets and thus the σ-algebras of the Borel sets, which are generated by these open sets, are the same. Therefore, a sequence of probability measures on $D(0, 1]$ is relatively compact if and only if it is tight (Billingsley 1999, Section 13). Hence, for every tight sequence of probability measures on $D(0, 1]$ the convergence

of the finite dimensional distributions on all continuity points of the limit distribution implies convergence in distribution (Billingsley 1999, Theorem 13.1).

For the PRM $M = \sum_{i=1}^{\infty} \varepsilon_{(U_i, \Delta_i)}$, where $(U_i)_i$ is an iid sequence of random vectors uniformly distributed on S_1 and

$$\Delta_i = \begin{cases} -\log(\Gamma_i) & \text{if } H = \Lambda, \\ \Gamma_i^{-1/\alpha} & \text{if } H = \Phi_\alpha, \\ -\Gamma_i^{1/\alpha} & \text{if } H = \Psi_\alpha, \end{cases}$$

we set

$$Y(t) = \sup\{\Delta_i : U_i^{(m)} \le t, i \ge 1\}, \qquad t \in (0, 1],$$

where $U_i^{(m)}$ is the m-th component of U_i. Then the process Y has the finite dimensional distributions in (2.5) for $k \ge 1$, $0 < t_i \le 1$, $x_i \in \mathbb{R}$ and $1 \le i \le k$. Therefore, Y is an extremal process generated by H restricted to the interval $(0, 1]$. For these processes we can show the following invariance principle by application of the continuous mapping theorem (see Billingsley 1999, Theorem 2.7 or Resnick 2008, p. 152).

Proposition 2.5 *Under the conditions of Theorem* 2.1 *and if* $H(\cdot) = \exp(-\mu(\cdot))$ *is an extreme value distribution it holds that*

$$Y_n \xrightarrow{d} Y, \qquad n \to \infty,$$

in $D(0, 1]$ *with respect to the metric* $\mathcal{D}$.

Since Y is a nondecreasing function, which is constant between isolated jumps, it has only countably many discontinuity points. Now let $(\tau_n)_n$ be the sequence of these discontinuity points of Y. Notice that by Embrechts et al. (1997), Theorem 5.4.7 the point process $\sum_{k=1}^{\infty} \varepsilon_{\tau_k}$ is a PRM with mean measure $v(a, b) = \log(b/a)$ for $0 < a < b \le 1$. We are ready to state our result for the point process of record times.

Theorem 2.6 *Under the conditions of Theorem* 2.1 *and if* $H(\cdot) = \exp(-\mu(\cdot))$ *is an extreme value distribution it holds that*

$$J_n := \sum_{k=1}^{p} \varepsilon_{p^{-1}L(k)} \xrightarrow{d} J := \sum_{k=1}^{\infty} \varepsilon_{\tau_k},$$

in $\mathcal{M}(0, 1]$, *the space of point measures on* $(0, 1]$.

Based on Theorem 2.6 we can make statements about the time points of the last and second last record at or before p.

Corollary 2.7 *Assume the conditions of Theorem* 2.6 *and let* $\zeta(p)$ *be the number of records among the random variables*

$$\max_{1 \leq i_1, \ldots, i_m \leq m} T_{\mathbf{i}}, \ldots, \max_{1 \leq i_1, \ldots, i_m \leq p} T_{\mathbf{i}}.$$

Then the following statements hold for $x, y \in (0, 1]$ *as* $n \to \infty$.

(1) $\mathbb{P}(p^{-1} L(\zeta(p)) \leq x) = \mathbb{P}(J_n(x, 1] = 0) \to \mathbb{P}(J(x, 1] = 0) = x$.
(2) $\mathbb{P}(p^{-1} L(\zeta(p)) \leq x, p^{-1} L(\zeta(p) - 1) \leq y) \to y + y \log(x/y)$ *for* $x > y$.
(3) $\mathbb{P}(p^{-1}(L(\zeta(p)) - L(\zeta(p) - 1)) \leq x) \to x(1 - \log(x))$.

Proof Let $0 < y < x \leq 1$. Part (1) is a direct consequence of the definitions of ζ and L. Part (2) follows by

$$\mathbb{P}(p^{-1} L(\zeta(p)) \leq x, p^{-1} L(\zeta(p) - 1) \leq y) = \mathbb{P}(J_n(x, 1] = 0, J_n(y, 1] \leq 1)$$
$$\to \mathbb{P}(J(x, 1] = 0, J(y, 1] \leq 1)$$

as $n \to \infty$ and

$$\mathbb{P}(J(x, 1] = 0, J(y, 1] \leq 1) = \mathbb{P}(J(x, 1] = 0)\mathbb{P}(J(y, x] \leq 1) = y + y \log(x/y).$$

To prove part (3) we assume that $\tau^{(1)}$ and $\tau^{(2)}$ are the first and the second upper order statistics of $(\tau_n)_n$. These upper order statistics exist since for every $a > 0$ there are only finitely many $\tau_n \in [a, 1]$. Then, we know by part (2) that

$$\mathbb{P}(\tau^{(1)} \leq x, \tau^{(2)} \leq y) = \mathbb{P}(J(x, 1] = 0, J(y, 1] \leq 1) = \begin{cases} y + y \log(x/y), & x > y \\ x, & \text{otherwise.} \end{cases} \tag{2.6}$$

Since

$$\lim_{n \to \infty} \mathbb{P}(p^{-1}(L(\zeta(p)) - L(\zeta(p) - 1)) \leq x) = \mathbb{P}(\tau^{(1)} - \tau^{(2)} \leq x)$$

we need to calculate $\mathbb{P}(\tau^{(1)} - \tau^{(2)} \leq x)$. The joint density of $\tau^{(1)}$ and $\tau^{(2)}$ can be deduced from (2.6), it is

$$f_{\tau^{(1)} \tau^{(2)}}(u, v) = \begin{cases} 1/u & u > v \\ 0, & \text{otherwise.} \end{cases}$$

Hence, we get the following distribution function of $\tau^{(1)} - \tau^{(2)}$

$$\mathbb{P}(\tau^{(1)} - \tau^{(2)} \leq x) = \int_0^x \int_0^{1-w} f_{\tau^{(1)} \tau^{(2)}}(w + v, v) dv\, dw$$
$$= \int_0^x \int_0^{1-w} 1/(w + v) dv\, dw = \int_0^x \log(1/w) dw = x(1 - \log(x)),$$

which completes the proof.

3 Applications

3.1 Relative ranks

In recent years, maximum-type tests based on the convergence in distribution of the maximum of rank statistics of a data set have gained significant interest for statistical testing (Han et al. 2017). Let $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be p-dimensional iid random vectors with $\mathbf{y}_t = (X_{1t}, \ldots, X_{pt})$ following a continuous distribution to avoid ties. We write Q_{it} for the rank of X_{it} among $X_{i1}, \ldots, X_{in}$. Additionally, let $R_{ij}^{(t)}$ be the relative rank of the j-th entry compared to the i-th entry; that is $R_{ij}^{(t)} = Q_{jt'}$ with t' such that $Q_{it'} = t$ for $t = 1, \ldots, n$.

A simpler explanation of $R_{ij}^{(t)}$ is that we look at the j-th and i-th rows of (Q_{it}) and find the location of t in the i-th row. Then we choose the value in the j-th row at this location.

Many important statistics are based on (relative) ranks; we consider two classes of such statistics in this section. First, we introduce the so–called simple linear rank statistics, which are of the form

$$V_{ij} = \sum_{t=1}^{n} c_{nt} \, g(R_{ij}^{(t)}/(n+1)), \qquad 1 \le i < j \le p,$$

where g is a Lipschitz function (also called score function), and (c_{nt}) with $c_{nt} = n^{-1} f(t/(n+1))$ for a Lipschitz function f and $\sum_{t=1}^{n} c_{nt}^2 > 0$ are called the regression constants. An example of such a simple linear rank statistic is Spearman's ρ, which will be discussed in detail in Section 3.1.2. For $1 \le i < j \le p$ the relative ranks $(R_{ij}^{(t)})_{t=1}^{n}$ depend on the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$, where $\mathbf{x}_k = (X_{k1}, \ldots, X_{kn})$ for $1 \le k \le p$. We assume that the vectors $\mathbf{x}_1, \ldots \mathbf{x}_p$ are independent. It is worth mentioning that the ranks (Q_{it}) remain the same if we transform the marginal distributions to the (say) standard uniform distribution. Thus, the joint distribution of $(R_{ij}^{(t)})_{t=1}^{n}$, and thereby the distribution of V_{ij}, does not depend on the distribution of $\mathbf{x}_i$ or $\mathbf{x}_j$. Therefore, we may assume without loss of generality that the random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are identically distributed. We can write $V_{ij} = g_{n,V}(\mathbf{x}_i, \mathbf{x}_j)$ for a measurable function $g_{n,V} : \mathbb{R}^{2n} \to \mathbb{R}$.

Next, we consider rank-type U-statistics of order $m < n$ of the form

$$U_{ij} = \frac{1}{n(n-1)\cdots(n-m+1)} \sum_{1 \le t_1 \ne \cdots \ne t_m \le n} h((X_{it_1}, X_{jt_1}), \ldots, (X_{it_m}, X_{jt_m})),$$

where the symmetric kernel h is such that U_{ij} depends only on $(R_{ij}^{(t)})_{t=1}^{n}$. An important example of a rank-type U- statistic is Kendall's τ, which will be studied in Section 3.1.1. For more examples we refer to Han et al. (2017) and references therein. As for simple linear rank statistics, we are able to write $U_{ij} = g_{n,U}(\mathbf{x}_i, \mathbf{x}_j)$, where $g_{n,U} : \mathbb{R}^{2n} \to \mathbb{R}$ is a measurable function and $\mathbf{x}_1, \ldots \mathbf{x}_p$ are iid random vectors.

An interesting property of rank-based statistics is the following pairwise independence. We also note that they are generally not mutually independent.

Lemma 3.1 (Lemma C4 in Han et al. 2017) *For $1 \leq i < j \leq p$, let Ψ_{ij} be a function of the relative ranks $\{R_{ij}^{(t)}, t = 1, \dots, n\}$. Assume $\mathbf{x}_1, \dots, \mathbf{x}_p$ are independent. Then for any $(i,j) \neq (k,l), i < j, k < l$, the random variables Ψ_{ij} and Ψ_{kl} are independent.*

As an immediate consequence we obtain pairwise independence of (U_{ij}) and (V_{ij}), respectively.

Lemma 3.2 *For any $(i,j) \neq (k,l)$, $i < j, k < l$, the random variables V_{ij} and V_{kl} are independent and identically distributed. Moreover, U_{ij} and U_{kl} are independent and identically distributed.*

We now want to standardize U_{ij} and V_{ij}. By independence of (X_{it}), we have

$$\mathbb{E}[V_{ij}] = \overline{g}_n \sum_{t=1}^{n} c_{nt}, \quad \text{Var}(V_{ij}) = \frac{1}{n-1} \sum_{t=1}^{n} (g(t/(n+1)) - \overline{g}_n)^2 \sum_{s=1}^{n} (c_{ns} - \overline{c}_n)^2,$$

where $\overline{g}_n = n^{-1} \sum_{t=1}^{n} g(t/(n+1))$ is the sample mean of $g(Q_{11}/(n+1)), \dots, g(Q_{1n}/(n+1))$ and $\overline{c}_n = \sum_{t=1}^{n} c_{nt}$. Expectation and variance of U_{ij} can also be calculated analytically. We set

$$\mu_V = \mathbb{E}[V_{12}], \sigma_V^2 = \text{Var}(V_{12}) \quad \text{and} \quad \mu_U = \mathbb{E}[U_{12}], \sigma_U^2 = \text{Var}(U_{12}),$$

and define the standardized versions of U_{ij} and V_{ij} by

$$\widetilde{V}_{ij} = (V_{ij} - \mu_V)/\sigma_V \quad \text{and} \quad \widetilde{U}_{ij} = (U_{ij} - \mu_U)/\sigma_U, \quad 1 \leq i < j \leq p.$$

It is well–known that $\widetilde{V}_{ij}$ and $\widetilde{U}_{ij}$ are asymptotically standard normal and the following lemma provides a complementary large deviation result.

Lemma 3.3 (Kallenberg 1982, p.404-405) *Suppose that the kernel function h is bounded and non-degenerate. Then we have for $x = o(n^{1/6})$ that*

$$\mathbb{P}(\widetilde{U}_{12} > x) = \overline{\Phi}(x)(1 + o(1)), \qquad n \to \infty.$$

Assume that the score function g is differentiable with bounded Lipschitz constant and that the constants $(c_{nt})_t$ satisfy

$$\max_{1 \leq t \leq n} |c_{nt} - \overline{c}_n|^2 \leq \frac{C^2}{n^{2/3}} \sum_{t=1}^{n} (c_{nt} - \overline{c}_n)^2, \quad \left| \sum_{t=1}^{n} (c_{nt} - \overline{c}_n)^3 \right|^2 \leq \frac{C^2}{n} \left| \sum_{t=1}^{n} (c_{nt} - \overline{c}_n)^2 \right|^3,$$
(3.1)

where C is some constant. Then it holds for $x = o(n^{1/6})$

$$\mathbb{P}(\widetilde{V}_{12} > x) = \overline{\Phi}(x)(1 + o(1)), \qquad n \to \infty.$$

For a discussion of (3.1), see (Kallenberg 1982, p.405). To proceed we need to find a suitable scaling and centering sequences for $\widetilde{V}_{ij}$ and $\widetilde{U}_{ij}$, respectively, such that

the conditions of Theorem 2.1 are fulfilled. For an iid standard normal sequence (X_i) it is known that

$$\lim_{p \to \infty} \mathbb{P}\left(\widetilde{d}_p\left(\max_{i=1,\dots,p} X_i - \widetilde{d}_p\right) \le x\right) = \exp(-\mathrm{e}^{-x}) = \Lambda(x), \qquad x \in \mathbb{R},$$

where $\widetilde{d}_p = \sqrt{2 \log p} - \frac{\log \log p + \log 4\pi}{2(2 \log p)^{1/2}}$; see Embrechts et al. (1997, Example 3.3.29). Since we are dealing with $p(p-1)/2$ random variables (V_{ij}) and (U_{ij}), respectively, which are asymptotically standard normal, $d_p = \widetilde{d}_{p(p-1)/2}$ seems like a reasonable choice for scaling and centering sequences.

Our main result for rank-statistics is the following.

Theorem 3.4

(a) *Suppose that the kernel function h is bounded and non-degenerate. If $p = \exp(o(n^{1/3}))$, the following point process convergence holds*

$$N_n^U := \sum_{1 \le i < j \le p} \varepsilon_{d_p(\widetilde{U}_{ij} - d_p)} \xrightarrow{d} N := \sum_{i=1}^{\infty} \varepsilon_{-\log \Gamma_i}, \qquad n \to \infty, \tag{3.2}$$

where $\Gamma_i = E_1 + \cdots + E_i, i \ge 1$, and (E_i) are iid standard exponential, i.e., N is a Poisson random measure with mean measure $\mu(x, \infty) = \mathrm{e}^{-x}, x \in \mathbb{R}$.

(b) *Assume that the score function g is differentiable with bounded Lipschitz constant and that the constants $(c_{nt})_t$ satisfy (3.1). Then if $p = \exp(o(n^{1/3}))$, it holds that*

$$N_n^V := \sum_{1 \le i < j \le p} \varepsilon_{d_p(\widetilde{V}_{ij} - d_p)} \xrightarrow{d} N, \qquad n \to \infty. \tag{3.3}$$

Proof We start with the proof of (3.3) for which we will use Theorem 2.1, as $\mathbf{x}_1, \dots \mathbf{x}_p$ are iid and $g_{n,V}$ is a measurable function. Therefore, we only have to show that for $x \in \mathbb{R}$ it holds

(1) $\frac{p(p-1)}{2} \mathbb{P}(\widetilde{V}_{12} > x_p) \to \mathrm{e}^{-x}$ as $n \to \infty$,
(2) $p^3 \mathbb{P}(\widetilde{V}_{12} > x_p, \widetilde{V}_{13} > x_p) \to 0$ as $n \to \infty$,

where $x_p = x/d_p + d_p$. We will begin with the proof of (1). Since $x_p \sim d_p = o(n^{1/6})$ we get by Lemma 3.3

$$\frac{p(p-1)}{2} \mathbb{P}(\widetilde{V}_{12} > x_p) = \frac{p(p-1)}{2} \bar{\Phi}(x_p)(1 + o(1))$$

and by Mill's ratio we have (writing $\tilde{p} = \frac{p(p-1)}{2}$)

$$\tilde{p}\,\bar{\Phi}(x_p) \sim \tilde{p}\frac{1}{\sqrt{2\pi}x_p}\,\mathrm{e}^{-x_p^2/2} \sim \tilde{p}\frac{1}{\sqrt{2\pi}\sqrt{2\log\tilde{p}}}\,\mathrm{e}^{-\log\tilde{p}+(\log\log\tilde{p})/2+(\log(4\pi))/2}\,\mathrm{e}^{-x} = \mathrm{e}^{-x}.$$

Regarding (2), we note that, by Lemma 3.2, $\widetilde{V}_{12}$ and $\widetilde{V}_{13}$ are independent. Thus, we get

$$p^3\mathbb{P}(\widetilde{V}_{12} > x_p, \widetilde{V}_{13} > x_p) = p^3\mathbb{P}(\widetilde{V}_{12} > x_p)^2 = p^3(\bar{\Phi}(x_p)(1 + o(1)))^2 \to 0, \qquad n \to \infty,$$

where we used Lemma 3.3 and Mill's ratio in the last two steps. That completes the proof of (3.3). The proof of (3.2) follows by analogous arguments. $\qquad\square$

Remark 3.5 Theorem 3.4 is a generalization of Theorems 1 and 2 in Han et al. (2017) who proved under the conditions of Theorem 3.4 and if $p = \exp(o(n^{1/3}))$ that

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{1\le i<j\le p} \widetilde{V}_{ij}^2 - 4\log p + \log\log p \le x \Big) = \exp\Big(-\tfrac{1}{\sqrt{8\pi}}\,e^{-x/2} \Big), \quad x \in \mathbb{R}$$

and

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{1\le i<j\le p} \widetilde{U}_{ij}^2 - 4\log p + \log\log p \le x \Big) = \exp\Big(-\tfrac{1}{\sqrt{8\pi}}\,e^{-x/2} \Big), \quad x \in \mathbb{R}.$$

As in Theorem 2.6, we additionally conclude point process convergence for the record times of the maxima of V_{ij} and U_{ij}. To this end, we investigate the sequence $(\max_{1\le i<j\le k} U_{ij})_{k\ge 1}$. This sequence jumps at time k if one of the random variables $U_{1k}, \dots, U_{k-1,k}$ is larger than every U_{ij} for $1 \le i < j \le k - 1$. Between these jump (or record) times the sequence is constant.

Let L^U be this sequence of record times defined by

$$L^U(1) = 1,$$
$$L^U(k + 1) = \inf\{\ell > L^U(k) : \max_{1\le i<j\le \ell} U_{ij} > \max_{1\le i<j\le L^U(k)} U_{ij}\}, \qquad k \in \mathbb{N},$$

and let L^V be constructed analogously.

Theorem 3.6 *Under the conditions of Theorem* 3.4 *it holds that*

$$\sum_{k=1}^{p} \varepsilon_{p^{-1}L^V(k)} \xrightarrow{d} J \quad \text{and} \quad \sum_{k=1}^{p} \varepsilon_{p^{-1}L^U(k)} \xrightarrow{d} J,$$

in $\mathcal{M}(0, 1]$, *the space of point measures on* $(0, 1]$, *where J is a Poisson random measure with mean measure* $v(a, b) = \log(b/a)$ *for* $0 < a < b \le 1$.

As in Corollary 2.7, we can draw conclusions on the index of the last and second last jump before or at p. Let $\zeta^U(p)$ be the number of records among $\max_{1\le i<j\le 2} U_{ij}, \dots, \max_{1\le i<j\le p} U_{ij}$. Then, as $n \to \infty$, we have for $x, y \in (0, 1]$

(1) $\mathbb{P}(p^{-1}L^U(\zeta^U(p)) \le x) \to \mathbb{P}(J(x, 1] = 0) = x,$
(2) $\mathbb{P}(p^{-1}L^U(\zeta^U(p)) \le x, p^{-1}L^U(\zeta^U(p) - 1) \le y) \to y + y\log(x/y)$ for $x > y$,
(3) $\mathbb{P}(p^{-1}(L^U(\zeta^U(p)) - L^U(\zeta^U(p) - 1)) \le x) \to x(1 - \log(x)),$

where (3) gives information about how much time elapses between the second last and the last jump of $(\max_{1 \leq i < j \leq k} U_{ij})_{k \geq 1}$ before or at p.

3.1.1 Kendall's tau

Kendall's tau is an example of a rank-type U-statistic with bounded kernel. For $i \neq j$ Kendall's tau τ_{ij} measures the ordinal association between the two sequences $(X_{i1}, \ldots, X_{in})$ and $(X_{j1}, \ldots, X_{jn})$. It is defined by

$$\tau_{ij} = \frac{2}{n(n-1)} \sum_{1 \leq t_1 < t_2 \leq n} \text{sign}(X_{it_1} - X_{it_2})\text{sign}(X_{jt_1} - X_{jt_2})$$

$$= \frac{2}{n(n-1)} \sum_{1 \leq t_1 < t_2 \leq n} \text{sign}(R_{ij}^{(t_2)} - R_{ij}^{(t_1)}),$$

where the function sign $: \mathbb{R} \to \{1, 0, -1\}$ is given by $\text{sign}(x) = x/|x|$ for $x \neq 0$ and $\text{sign}(0) = 0$. An interesting property of Kendall's tau is that there exists a representation as a sum of independent random variables. We could not find this representation in the literature. Therefore, we state it here. The proof can be found in Section 4.

Proposition 3.7 *We have*

$$\tau_{12} \stackrel{d}{=} \frac{4}{n(n-1)} \sum_{i=1}^{n-1} D_i \,,$$

where $(D_i)_{i \geq 1}$ are independent random variables with D_i being uniformly distributed on the numbers $-i/2, -i/2 + 1, \ldots, i/2$.

From Proposition 3.7 we deduce $\mathbb{E}[\tau_{ij}] = 0$ and $\text{Var}(\tau_{ij}) = \frac{2(2n+5)}{9n(n-1)}$. The next result is a corollary of Theorem 3.4.

Corollary 3.8 *Under the conditions of Theorem 3.4 we have*

$$N_n^\tau := \sum_{1 \leq i < j \leq p} \varepsilon_{d_p(\tau_{ij}/\sqrt{\text{Var}(\tau_{ij})} - d_p)} \stackrel{d}{\to} N = \sum_{i=1}^{\infty} \varepsilon_{-\log \Gamma_i} \,, \qquad n \to \infty \,.$$

3.1.2 Spearman's rho

An example of a simple linear rank statistic is Spearman's rho, which is a measure of rank correlation that assesses how well the relationship between two variables can be described using a monotonic function. Recall that Q_{ik} and Q_{jk} are the ranks of X_{ik} and X_{jk} among $\{X_{i1}, \ldots, X_{in}\}$ and $\{X_{j1}, \ldots, X_{jn}\}$, respectively, and

write $q_n = (n+1)/2$ for the average rank. Then for $1 \le i \ne j \le p$ Spearman's rho is defined by

$$\rho_{ij} = \frac{\sum_{k=1}^{n}(Q_{ik} - q_n)(Q_{jk} - q_n)}{\left(\sum_{k=1}^{n}(Q_{ik} - q_n)^2 \sum_{k=1}^{n}(Q_{jk} - q_n)^2\right)^{1/2}}$$

$$= \frac{12}{n(n^2 - 1)} \sum_{k=1}^{n} \left(k - \frac{n+1}{2}\right)\left(R_{ij}^{(k)} - \frac{n+1}{2}\right).$$

For mean and variance we get

$$\mathbb{E}[\rho_{ij}] = 0 \quad \text{and} \quad \text{Var}(\rho_{ij}) = 1/(n-1). \tag{3.4}$$

Therefore, we obtain the following corollary of Theorem 3.4.

Corollary 3.9 *Under the conditions of Theorem 3.4 it holds that*

$$N_n^\rho := \sum_{1 \le i < j \le p} \varepsilon_{d_p(\rho_{ij}/\sqrt{\text{Var}(\rho_{ij})} - d_p)} \xrightarrow{d} N.$$

The next auxiliary result allows us to transfer the weak convergence of a sequence of point processes to a another sequence of point processes, provided that the maximum distance between their points tends to zero in probability.

Proposition 3.10 *For arrays $(X_{i,n})_{i,n \ge 1}$ and $(Y_{i,n})_{i,n \ge 1}$ of real-valued random variables, let $N_n^X = \sum_{i=1}^{p} \varepsilon_{X_{i,n}}$ and assume that $N_n^X \xrightarrow{d} N$. Consider a point process $N_n^Y = \sum_{i=1}^{p} \varepsilon_{Y_{i,n}}$. If*

$$\max_{i=1,\ldots,p} |X_{i,n} - Y_{i,n}| \xrightarrow{\mathbb{P}} 0,$$

then $N_n^Y \xrightarrow{d} N$.

Example 3.11 It turns out that there is an interesting connection between Spearman's rho and Kendall's tau. By Hoeffding ([1948](), p.318) we can write Spearman's rho as

$$\rho_{ij} = \frac{n-2}{n+1} r_{ij} + \frac{3\tau_{ij}}{n+1}, \qquad 1 \le i \ne j \le p, \tag{3.5}$$

where

$$r_{ij} = \frac{3}{n(n-1)(n-2)} \sum_{1 \le t_1 \ne t_2 \ne t_3 \le n} \text{sign}(X_{it_1} - X_{it_2})\text{sign}(X_{jt_1} - X_{jt_3})$$

is the major part of Spearman's rho. Therefore, r_{ij} is a U-statistic of degree three with an asymmetric bounded kernel and with

$$\mathbb{E}[r_{ij}] = 0 \quad \text{and} \quad \text{Var}(r_{ij}) = \frac{n^2 - 3}{n(n-1)(n-2)}, \qquad 1 \le i \ne j \le p. \tag{3.6}$$

We now use Proposition 3.10 and Corollary 3.9 to show that

$$N_n^r := \sum_{1 \le i < j \le p} \varepsilon_{d_p(r_{ij}/\sqrt{\mathrm{Var}(r_{ij})} - d_p)} \xrightarrow{d} N := \sum_{i=1}^{\infty} \varepsilon_{-\log \Gamma_i}, \qquad n \to \infty. \tag{3.7}$$

For this purpose we consider the following difference

$$d_p\left(\frac{\rho_{ij}}{\sqrt{\mathrm{Var}(\rho_{ij})}} - d_p\right) - d_p\left(\frac{r_{ij}}{\sqrt{\mathrm{Var}(r_{ij})}} - d_p\right) = d_p\left(\frac{\rho_{ij}}{\sqrt{\mathrm{Var}(\rho_{ij})}} - \frac{r_{ij}}{\sqrt{\mathrm{Var}(r_{ij})}}\right).$$

By (3.4), (3.6) and (3.5) this expression is asymptotically equal to

$$\frac{d_p}{\sqrt{n}}(\rho_{ij} - r_{ij}) = \frac{3d_p}{\sqrt{n(n+1)}}(\tau_{ij} - r_{ij}).$$

Since $|\tau_{ij}|$ and $|r_{ij}|$ are bounded above by constants, we deduce that

$$\max_{1 \le i < j \le p} \left| \frac{3d_p}{\sqrt{n(n+1)}}(\tau_{ij} - r_{ij}) \right| \xrightarrow{\mathbb{P}} 0, \qquad n \to \infty,$$

which verifies the condition in Proposition 3.10. Since $N_n^\rho \xrightarrow{d} N$ by Corollary 3.9, we conclude the desired (3.7).

3.2 Sample covariances

An important field of current research is the estimation and testing of high-dimensional covariance structures. It finds application in genomics, social science and financial economics; see Cai (2017) for a detailed review and more references. Under quite general assumptions (Xiao and Wu 2013) investigated the maximum off-diagonal entry of a high-dimensional sample covariance matrix. We impose the same model assumptions (compare Xiao and Wu 2013, p. 2901-2903), but instead of the maximum we study the point process of off-diagonal entries.

We start by describing the model and spelling out the required assumptions. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be p-dimensional iid random vectors with $\mathbf{x}_i = (X_{1i}, \ldots, X_{pi})$, where $\mathbb{E}[X_{ji}] = 0$ for $1 \le j \le p$ and $\bar{X}_j := \frac{1}{n}\sum_{k=1}^{n} X_{jk}$. Denote $\Sigma = (\sigma_{i,j})_{1 \le i,j \le p}$ as the covariance matrix of the vector $\mathbf{x}_1$ and assume $\sigma_{i,i} = 1$ for $1 \le i \le p$. The empirical covariance matrix $(\hat{\sigma}_{i,j})_{1 \le i,j \le p}$ is given by

$$\hat{\sigma}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j), \qquad 1 \le i,j \le p.$$

A fundamental problem in high-dimensional inference is to derive the asymptotic distribution of $\max_{1 \le i < j \le p} |\hat{\sigma}_{i,j} - \sigma_{i,j}|$. Since the $\hat{\sigma}_{i,j}$'s might have different variances we need to standardize $\hat{\sigma}_{i,j}$ by $\theta_{i,j} = \mathrm{Var}(X_{i1}X_{j1})$, which can be estimated by

$$\hat{\theta}_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \left[(X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j) - \hat{\sigma}_{i,j} \right]^2.$$

We are interested in the points

$$M_{i,j} := \frac{|\hat{\sigma}_{i,j} - \sigma_{i,j}|}{\sqrt{\hat{\theta}_{i,j}}}, \qquad 1 \le i < j \le p.$$

Let $\mathcal{I}_n = \{(i,j) : 1 \le i < j \le p\}$ be an index set. We use the following notations to formulate the required conditions:

$$\mathcal{K}_n(t, r) = \sup_{1 \le i \le p} \mathbb{E}[\exp(t|X_{i1}|^r)],$$

$$\mathcal{M}_n(r) = \sup_{1 \le i \le p} \mathbb{E}[|X_{i1}|^r],$$

$$\theta_n = \inf_{1 \le i < j \le p} \theta_{i,j},$$

$$\gamma_n = \sup_{\substack{\alpha, \beta \in \mathcal{I}_n \\ \alpha \ne \beta}} |\mathrm{Cor}(X_{i1}X_{j1}, X_{k1}X_{l1})|, \quad \text{for } \alpha = (i,j), \ \beta = (k,l),$$

$$\gamma_n(b) = \sup_{\alpha \in \mathcal{I}_n} \sup_{\substack{A \subset \mathcal{I}_n \\ |A| = b}} \inf_{\beta \in A} |\mathrm{Cor}(X_{i1}X_{j1}, X_{k1}X_{l1})| \quad \text{for } \alpha = (i,j), \ \beta = (k,l).$$

Now, we can draft the following conditions.

(B1) $\liminf_{n\to\infty} \theta_n > 0$.

(B2) $\limsup_{n\to\infty} \gamma_n < 1$.

(B3) $\gamma_n(b_n) \log(b_n) = o(1)$ for any sequence (b_n) such that $b_n \to \infty$.

(B3') $\gamma_n(b_n) = o(1)$ for any sequence (b_n) such that $b_n \to \infty$ and for some $\varepsilon > 0$

$$\sum_{\alpha, \beta \in \mathcal{I}_n} (\mathrm{Cov}(X_{i1}X_{j1}, X_{k1}X_{l1}))^2 = O(p^{4-\varepsilon}) \quad \text{for } \alpha = (i,j), \ \beta = (k,l).$$

(B4) For some constants $t > 0$ and $0 < r \le 2$, $\limsup_{n\to\infty} \mathcal{K}_n(t, r) < \infty$, and

$$\log p = \begin{cases} o(n^{r/(4+r)}), & \text{if } 0 < r < 2, \\ o(n^{1/3}(\log n)^{-2/3}), & \text{if } r = 2. \end{cases}$$

(B4') $\log p = o(n^{r/(4+3r)})$, $\limsup_{n\to\infty} \mathcal{K}_n(t, r) < \infty$ for some constants $t > 0$ and $r > 0$.

(B4'') $p = O(n^q)$ and $\limsup_{n\to\infty} \mathcal{M}_n(4q + 4 + \delta) < \infty$ for some constants $q > 0$ and $\delta > 0$.

To be able to adopt parts of the proof of Theorem 2 in Xiao and Wu (2013) we consider (instead of $(M_{i,j})$) the transformed points $(W_{i,j})$ given by

$$W_{i,j} := \frac{1}{2}(n\, M_{i,j}^2 - 4\log p + \log\log p + \log 8\pi), \qquad 1 \le i < j \le p,$$

and we define the point processes

$$N_n^{(W)} := \sum_{1 \le i < j \le p} \varepsilon_{W_{i,j}}.$$

Theorem 3.12 *Let $\mathbb{E}[\mathbf{x}_1] = 0$ and $\sigma_{i,i} = 1$ for all i, and assume (B1) and (B2). Then under any one of the following conditions*:

  (i)   *(B3) and (B4),*
  (ii) *(B3') and (B4'),*
  (iii) *(B3) and (B4"),*
  (iv) *(B3') and (B4"),*

it holds, that

$$N_n^{(W)} \xrightarrow{d} N = \sum_{i=1}^{\infty} \varepsilon_{-\log\Gamma_i}, \qquad n \to \infty,$$

where $\Gamma_i = E_1 + \cdots + E_i$, $i \ge 1$, and (E_i) are iid standard exponential, i.e., N is a Poisson random measure with mean measure $\mu(x, \infty) = e^{-x}$, $x \in \mathbb{R}$.

Proof Under condition (i) set $\mathcal{E}_n = n^{-(2-r)/(4(r+4))}$ if $0 < r < 2$, and $\mathcal{E}_n = n^{-1/6}(\log n)^{1/3}(\log p)^{1/2}$ if $r = 2$. Under condition (ii) let $\mathcal{E}_n = (\log p)^{1/2}n^{-r/(6r+8)}$. Under (i) or (ii) we set

$$\tilde{X}_{ik} = X_{ik}\mathbb{1}_{\{|X_{ik}|\le T_n\}} - \mathbb{E}[X_{ik}\mathbb{1}_{\{|X_{ik}|\le T_n\}}], \qquad 1 \le i \le p, \ 1 \le k \le n,$$

where $T_n = \mathcal{E}_n(n/(\log p)^3)^{1/4}$. Under conditions (iii) and (iv) we set

$$\tilde{X}_{ik} = X_{ik}\mathbb{1}_{\{|X_{ik}|\le n^{1/4}/\log n\}}, \qquad 1 \le i \le p, \ 1 \le k \le n.$$

Additionally, we define $\tilde{\sigma}_{i,j} = \mathbb{E}[\tilde{X}_{i1}\tilde{X}_{j1}]$ and $\tilde{\theta}_{i,j} = \mathrm{Var}[\tilde{X}_{i1}\tilde{X}_{j1}]$. We consider

$$M_{1;i,j} = \frac{1}{\sqrt{\tilde{\theta}_{i,j}}}\left|\frac{1}{n}\sum_{k=1}^{n}\tilde{X}_{ik}\tilde{X}_{jk} - \tilde{\sigma}_{ij}\right|$$

and the transformed points

$$W_{1;i,j} = \frac{1}{2}(n\, M_{1;i,j}^2 - 4\log p + \log\log p + \log 8\pi).$$

We will show that $N_n^{(W_1)} := \sum_{1 \le i < j \le p} \varepsilon_{W_{1;i,j}} \xrightarrow{d} N$ and thus by Proposition 3.10 $N_n^{(W)} \xrightarrow{d} N$.

Therefore, we first apply Kallenberg's Theorem as in the proof of Theorem 2.1. We set

$$B = \bigcup_{k=1}^{q} B_k \subset \mathbb{R}$$

with disjoint intervals $B_k = (r_k, s_k]$ and show

(1) $\lim\limits_{n \to \infty} \mu_n^{(W_1)}(B) = \mu(B),$

(2) $\lim\limits_{n \to \infty} \mathbb{P}(N_n^{(W_1)}(B) = 0) = e^{-\mu(B)},$

where $\mu_n^{(W_1)}(B) = \mathbb{E}[N_n^{(W_1)}(B)]$ and μ is defined by $\mu(B_k) = e^{-r_k} - e^{-s_k}$.

From the proof of Theorem 2 of Xiao and Wu (2013, p. 2910, 2913-2914) we know that the conditions of Xiao and Wu (2013, Lemma 6) are satisfied. Furthermore, from the proof of Lemma 6 (Xiao and Wu 2013, p. 2909-2910) we get that for $z \in \mathbb{R}$ and

$$z_n = (4 \log p - \log \log p - \log 8\pi + 2z)^{1/2}$$

and $d \in \mathbb{N}$

$$\lim_{n \to \infty} \sum_{\substack{A \subset \mathcal{I}_n \\ |A| = d}} \mathbb{P}(\sqrt{n} M_{1;i_1 j_1} > z_n, \dots, \sqrt{n} M_{1;i_d j_d} > z_n) = \frac{e^{-dz}}{d!},$$

which is equivalent to

$$\lim_{n \to \infty} \sum_{\substack{A \subset \mathcal{I}_n \\ |A| = d}} \mathbb{P}(W_{1;i_1 j_1} > z, \dots, W_{1;i_d j_d} > z) = \frac{e^{-dz}}{d!}, \tag{3.8}$$

where $A = \{(i_1, j_1), \dots, (i_d, j_d)\}$. Therefore, we get for $d = 1$

$$\lim_{n \to \infty} \mu_n^{(W_1)}(B) = \lim_{n \to \infty} \sum_{k=1}^{q} \sum_{(i,j) \in \mathcal{I}_n} \mathbb{P}(W_{1;i,j} \in B_k) = \sum_{k=1}^{q} (e^{-r_k} - e^{-s_k}) = \mu(B).$$

which proves (1). Regarding (2), we use that $1 - \mathbb{P}(N_n^{(W_1)}(B) = 0) = \mathbb{P}\left(\bigcup_{1 \le i < j \le p} A_{i,j}\right),$ where $A_{i,j} = \{W_{1;i,j} \in B\}$. By Bonferroni's inequality we have for every $k \ge 1$,

$$\sum_{d=1}^{2k} (-1)^{d-1} \sum_{\substack{A \subset \mathcal{I}_n \\ |A| = d}} P_{A,B} \le \mathbb{P}\left(\bigcup_{1 \le i < j \le p} A_{i,j}\right) \le \sum_{d=1}^{2k-1} (-1)^{d-1} \sum_{\substack{A \subset \mathcal{I}_n \\ |A| = d}} P_{A,B}, \tag{3.9}$$

where $A = \{(i_1, j_1), \ldots, (i_d, j_d)\}$ and $P_{A,B} = \mathbb{P}(W_{1;i_1 j_1} \in B, \ldots, W_{1;i_d j_d} \in B)$. First letting $n \to \infty$ and then $k \to \infty$, we deduce from (3.8) and (3.9) that

$$\lim_{n \to \infty} \mathbb{P}(N_n^{(W_1)}(B) = 0) = 1 - \sum_{d=1}^{\infty} (-1)^{d-1} \frac{(\mu(B))^d}{d!} = \sum_{d=0}^{\infty} (-1)^d \frac{(\mu(B))^d}{d!} = e^{-\mu(B)}.$$

This proves (2) and we get $N_n^{(W_1)} \xrightarrow{d} N$. By Proposition 3.10 it remains to show

$$\max_{1 \le i < j \le p} |W_{1;i,j} - W_{i,j}| = \frac{n}{2} \max_{1 \le i < j \le p} |M_{1;i,j}^2 - M_{i,j}^2| \xrightarrow{\mathbb{P}} 0.$$

Fortunately, this is shown in the course of the proof of Theorem 2 of Xiao and Wu (2013, p. 2911-2916). □

The following examples are motivated by Xiao and Wu (2013, p. 2903-2905).

Example 3.13 (Physical dependence). Assume that $\mathbf{x}_1 = (X_{11}, \ldots, X_{p1})$ is distributed as a stationary process of the following form. For a measurable function g and a sequence of iid random variables $(\epsilon_i)_{i \in \mathbb{Z}}$ we set $\mathbf{x}_1 = (X_{11}, \ldots, X_{p1})$ with

$$X_{i1} = g(\epsilon_i, \epsilon_{i-1}, \ldots), \qquad i \ge 1,$$

and let $\mathbf{x}_k$, $2 \le k \le n$, be iid copies of $\mathbf{x}_1$. Moreover, for an iid copy $(\epsilon_i')_{i \in \mathbb{Z}}$ of $(\epsilon_i)_{i \in \mathbb{Z}}$ and

$$X_{i1}' = g(\epsilon_i, \ldots, \epsilon_1, \epsilon_0', \epsilon_{-1}, \ldots)$$

we define the *physical dependence measure* of order q (see Wu (2005))

$$\delta_q(i) = \mathbb{E}\big[|X_{i1} - X_{i1}'|^q\big]^{1/q} \quad \text{and} \quad \Psi_q(k) = \Big[\sum_{i=k}^{\infty} (\delta_q(i))^2\Big]^{1/2}.$$

Then, we conclude from Lemma 3 of Xiao and Wu (2013) and Theorem 3.12 the following statement.

Assume that $0 < \Psi_4(0) < \infty$ and $\mathrm{Var}(X_{i1}X_{j1}) > 0$ for all $i, j \in \mathbb{Z}$ and $|\mathrm{Cor}(X_{i1}X_{j1}, X_{k1}X_{l1})| < 1$ for all i, j, k, l, such that they are not all the same. Then, if either one of the conditions

(i) $\Psi_q(k) = o(1/\log k)$ as $k \to \infty$ and one of the assumptions (B4) and (B4') or
(ii) $\sum_{j=0}^{p} (\Psi_4(j))^2 = O(p^{1-\delta})$ for some $\delta > 0$ and one of the assumptions (B4') or (B4'')

is satisfied, we have

$$N_n^{(W)} \xrightarrow{d} N, \qquad n \to \infty.$$

As a special case we consider the linear process $X_{i1} = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}$, where the ϵ_j are iid with $\mathbb{E}[\epsilon_j] = 0$ and $\mathbb{E}[|\epsilon_j|^q] < \infty$ with $q \geq 4$ and for $a_j \in \mathbb{R}$ it holds that $\sum_{j=0}^{\infty} a_j^2 \in (0, \infty)$. Then the physical dependence measure is given by $\delta_q(j) = |a_j| \, \mathbb{E}[|\epsilon_0 - \epsilon_0'|^q]^{1/q}$. Moreover, the conditions $0 < \Psi_4(0) < \infty$ and $\mathrm{Var}(X_{i1} X_{j1}) > 0$ for all $i, j \in \mathbb{Z}$ and $|\mathrm{Cor}(X_{i1} X_{j1}, X_{k1} X_{l1})| < 1$ for all i, j, k, l, such that they are not all the same, are fulfilled. If $a_j = j^{-\beta} \ell(j)$, where $1/2 < \beta < 1$ and ℓ is a slowly varying function, then (X_{i1}) is a long memory process. The smaller the value of β, the stronger is the dependence between the (X_{i1}). If one of the assumptions (B4) or (B4') is satisfied, then condition (i) is fulfilled for every $\beta \in (1/2, 1)$.

Example 3.14 (Non-stationary linear processes). As in the previous example, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are iid random vectors. Now $\mathbf{x}_1 = (X_{11}, \ldots, X_{p1})$ is given by

$$X_{i1} = \sum_{t \in \mathbb{Z}} f_{i,t} \epsilon_{i-t}, \qquad i \geq 1,$$

where $(\epsilon_i)_{i \in \mathbb{Z}}$ is a sequence of iid random variables with mean zero, variance one and finite fourth moment and the sequences $(f_{i,t})_{t \in \mathbb{Z}}$ satisfy $\sum_{t \in \mathbb{Z}} f_{i,t}^2 = 1$. Let κ_4 be the fourth cumulant of ϵ_0 and

$$h_n(k) = \sup_{1 \leq i \leq p} \left(\sum_{|t| = \lfloor k/2 \rfloor}^{\infty} f_{i,t}^2 \right)^{1/2}.$$

Assume that $\kappa_4 > -2$ and

$$\limsup_{n \to \infty} \sup_{1 \leq i < j \leq p} \left| \sum_{t \in \mathbb{Z}} f_{i,i-t} f_{j,j-t} \right| < 1. \tag{3.10}$$

By Section 3.2 of Xiao and Wu (2013, p. 2904-2905) and Theorem 3.12 we get the following result. If either

(i) $h_n(k_n) \log k_n = o(1)$ for any positive sequence k_n such that $k_n \to \infty$ as $n \to \infty$ and one of the assumptions (B4) and (B4') or
(ii) $\sum_{k=1}^{p} (h_n(k))^2 = O(p^{1-\delta})$ for some $\delta > 0$ and one of the assumptions (B4') or (B4'')

holds, then we have $N_n^{(W)} \xrightarrow{d} N$ as $n \to \infty$.

To illustrate these assumptions we consider the special case $\mathbf{x}_1 := (\epsilon_1, \ldots, \epsilon_p) A_n$, where $A_n \in \mathbb{R}^{p \times p}$ is a deterministic, symmetric matrix with $(A_n)_{i,j} = a_{ij}$ for $1 \leq i, j \leq p$. We assume that $\sum_{t=1}^{p} a_{it}^2 = 1$ for every $1 \leq i \leq p$.

The covariance matrix of $\mathbf{x}_1$ is given by $\mathrm{Cov}(\mathbf{x}_1) = A_n A_n^T$ with $(A_n A_n^T)_{ij} = \sum_{t=1}^{p} a_{it} a_{jt}$. Observe that the diagonal entries are equal to 1. To satisfy assumption (3.10) we have to assume that the entries except for the diagonal are asymptotically smaller than 1, i.e.

$$\limsup_{n\to\infty} \ \sup_{1\le i<j\le p} \Big| \sum_{t=1}^{p} a_{it}a_{jt} \Big| < 1.$$

We set

$$h_n(k) = \sup_{1\le i\le p} \Big(\sum_{t=1}^{i-\lfloor k/2\rfloor} a_{it}^2 + \sum_{t=\lfloor k/2\rfloor+i}^{p} a_{it}^2 \Big)^{1/2}$$

as a measure of how close the matrices A_n are to diagonal matrices. For the point process convergence either (i) or (ii) has to be satisfied for h_n.

4 Proofs of the results

4.1 Proofs of the results in Section 2

Proof of Theorem 2.1 . We will follow the lines of the proof of Theorem 2.1 in Dabrowski et al. (2002). Since the mean measure η has a density, the limit process M is simple and we can apply Kallenberg's Theorem (see for instance Embrechts et al. (1997), p.233, Theorem 5.2.2) or Kallenberg 1983, p.35, Theorem 4.7). Therefore, it suffices to prove that for any finite union of bounded rectangles

$$R = \bigcup_{k=1}^{q} A_k \times B_k \subset S, \qquad \text{with} \qquad A_k = \bigotimes_{l=1}^{m}(r_k^{(l)}, s_k^{(l)}], \quad B_k = (r_k^{(m+1)}, s_k^{(m+1)}],$$

it holds that

(1) $\lim\limits_{n\to\infty} \eta_n(R) = \eta(R),$
(2) $\lim\limits_{n\to\infty} \mathbb{P}(M_n(R) = 0) = e^{-\eta(R)}.$

Without loss of generality we can assume that the A_k's are chosen to be disjoint. First we will show (1). Set $T := T_{(1,2,\dots,m)} = g_n(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)$. If $q = 1$ we get

$$\eta_n(R) = \mathbb{E}[M_n(A_1 \times B_1)] = \sum_{\mathbf{i}:\mathbf{i}/p\in A_1} \mathbb{P}(T_{\mathbf{i}} \in B_1)$$

$$\sim p^m \prod_{l=1}^{m}(s_1^{(l)} - r_1^{(l)})\mathbb{P}(T \in B_1).$$

Since assumption (A1) implies $p^m/(m!)\,\mathbb{P}(T \in B_1) \to \mu(B_1)$, we obtain the convergence $\eta_n(R) \to \eta(R)$ as $n \to \infty$. The case $q \ge 1$ follows by

$$\eta_n(R) = \sum_{k=1}^{q} \eta_n(A_k \times B_k) \rightarrow \sum_{k=1}^{q} \eta(A_k \times B_k) = \eta(R), \qquad n \rightarrow \infty.$$

To show (2), we let P_n be the probability mass function of the Poisson distribution with mean $\eta_n(R)$. Then we have

$$|\mathbb{P}(M_n(R) = 0) - \mathbb{P}(M(R) = 0)| \leq |\mathbb{P}(M_n(R) = 0) - P_n(0)| + |P_n(0) - \mathbb{P}(M(R) = 0)|$$
$$= |\mathbb{P}(M_n(R) = 0) - P_n(0)| + o(1),$$

where the last equality holds by (1). Therefore, we only have to estimate $|\mathbb{P}(M_n(R) = 0) - P_n(0)|$. For this we employ the Stein-Chen method (see Barbour et al. (1992) for a discussion). The Stein equation for the Poisson distribution P_n with mean $\eta_n(R)$ is given by

$$\eta_n(R)x(j + 1) - jx(j) = \mathbb{1}_{\{j=0\}} - P_n(0), \qquad j \geq 0. \tag{4.11}$$

This equation is solved by the function

$$x(0) = 0$$
$$x(j + 1) = \frac{j!}{\eta_n(R)^{j+1}} e^{\eta_n(R)} (P_n(\{0\}) - P_n(\{0\})P_n(\{0, \dots, j\})), \quad j = 0, 1, \dots$$

By (4.11) we see that

$$|\mathbb{P}(M_n(R) = 0) - P_n(0)| = |\mathbb{E}[\eta_n(R)x(M_n(R) + 1) - M_n(R)x(M_n(R))]|. \tag{4.12}$$

Therefore, we only have to estimate the right hand side of (4.12) and to this end we set

$$D := \{\mathbf{k} : \mathbf{k} = (k_1, k_2, \dots, k_m), 1 \leq k_1 < k_2 < \dots < k_m \leq p\},$$
$$I_{\mathbf{k}} := \sum_{i=1}^{q} \mathbb{1}_{A_i}(\mathbf{k}/p)\mathbb{1}_{B_i}(T_{\mathbf{k}}),$$
$$\eta_{\mathbf{k}} := \mathbb{E}[I_{\mathbf{k}}].$$

For $\mathbf{k} \in D$ let

$$D_{1\mathbf{k}} := \{\boldsymbol{\ell} \in D : \ell_i \neq k_j, i, j = 1, 2, \dots, m\} \quad \text{and}$$
$$D_{2\mathbf{k}} := \{\boldsymbol{\ell} \in D : \boldsymbol{\ell} \neq \mathbf{k}, \ell_i = k_j \text{ for some } i, j = 1, 2, \dots, m\}.$$

Then we have the disjoint union $D = D_{1\mathbf{k}} \,\dot{\cup}\, D_{2\mathbf{k}} \,\dot{\cup}\, \{\mathbf{k}\}$, and therefore,

$$M_n(R) = \sum_{\boldsymbol{\ell} \in D} I_{\boldsymbol{\ell}} = \sum_{\boldsymbol{\ell} \in D_{1\mathbf{k}}} I_{\boldsymbol{\ell}} + \left(I_{\mathbf{k}} + \sum_{\boldsymbol{\ell} \in D_{2\mathbf{k}}} I_{\boldsymbol{\ell}}\right) =: M_n^{(1)}(\mathbf{k}) + M_n^{(2)}(\mathbf{k}).$$

Now, we bound (4.12) by

$$\left| \sum_{\mathbf{k} \in D} \mathbb{E}[\eta_{\mathbf{k}} x(M_n(R) + 1) - I_{\mathbf{k}} x(M_n(R))] \right|$$

$$\leq \left| \sum_{\mathbf{k} \in D} \eta_{\mathbf{k}} \mathbb{E}[x(M_n(R) + 1) - x(M_n^{(1)}(\mathbf{k}) + 1)] \right| + \left| \sum_{\mathbf{k} \in D} [\mathbb{E}[I_{\mathbf{k}} x(M_n(R))] - \eta_{\mathbf{k}} \mathbb{E}[x(M_n^{(1)}(\mathbf{k}) + 1)]] \right|.$$

$$(4.13)$$

It suffices to show that both terms in (4.13) tend to zero as $n \to \infty$. From Barbour and Eagleson (1984, p.400) we have the following bound for the increments of the solution of Stein's equation

$$\Delta x := \sup_{j \in \mathbb{N}_0} |x(j + 1) - x(j)| \leq \min(1, 1/\eta_n(R)). \qquad (4.14)$$

Using (4.14) the first term of (4.13) is bounded above by

$$\sum_{\mathbf{k} \in D} \eta_{\mathbf{k}} |\mathbb{E}[x(M_n^{(1)}(\mathbf{k}) + M_n^{(2)}(\mathbf{k}) + 1) - x(M_n^{(1)}(\mathbf{k}) + 1)]| \leq \sum_{\mathbf{k} \in D} \eta_{\mathbf{k}} \mathbb{E}[M_n^{(2)}(\mathbf{k})].$$

Using the definitions of $\eta_{\mathbf{k}}$ and $M_n^{(2)}(\mathbf{k})$, we get

$$\sum_{\mathbf{k} \in D} \eta_{\mathbf{k}} \mathbb{E}[M_n^{(2)}(\mathbf{k})]$$

$$= \sum_{\mathbf{k} \in D} \left(\sum_{i=1}^{q} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p}) \mathbb{P}(T \in B_i) \right) \left(\sum_{\ell \in D} \sum_{j=1}^{q} \mathbb{1}_{A_j}(\tfrac{\ell}{p}) \mathbb{P}(T \in B_j) - \sum_{\ell \in D_{1\mathbf{k}}} \sum_{j=1}^{q} \mathbb{1}_{A_j}(\tfrac{\ell}{p}) \mathbb{P}(T \in B_j) \right)$$

$$= \sum_{i=1}^{q} \sum_{j=1}^{q} \mathbb{P}(T \in B_i) \mathbb{P}(T \in B_j) \sum_{\mathbf{k} \in D} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p}) \sum_{\ell \in D} \mathbb{1}_{A_j}(\tfrac{\ell}{p})$$

$$- \sum_{i=1}^{q} \sum_{j=1}^{q} \mathbb{P}(T \in B_i) \mathbb{P}(T \in B_j) \sum_{\mathbf{k} \in D} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p}) \sum_{\ell \in D_{1\mathbf{k}}} \mathbb{1}_{A_j}(\tfrac{\ell}{p}).$$

$$(4.15)$$

Since by assumption (A1) it holds that $p^m \mathbb{P}(T \in B_i) \to m!(\mu(r_i^{(m+1)}) - \mu(s_i^{(m+1)}))$ as $n \to \infty$, and

$$\frac{1}{p^m} \sum_{\mathbf{k} \in D} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p}) \to \lambda_m(A_i), \qquad n \to \infty$$

$$\frac{1}{p^{2m}} \sum_{\mathbf{k} \in D} \sum_{\ell \in D_{1\mathbf{k}}} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p}) \mathbb{1}_{A_j}(\tfrac{\ell}{p}) \to \lambda_m(A_i) \lambda_m(A_j) \qquad n \to \infty,$$

(4.15) and thus the first term of (4.13) tend to zero as $n \to \infty$. As every $I_{\mathbf{k}}$ only depends on $T_{\mathbf{k}}$ and because $D_{1\ell}$ only contains elements which have no component in common with ℓ, $M_n^{(1)}(\ell)$ and I_ℓ are independent. Therefore, the second term of (4.13) equals

$$\left| \sum_{\mathbf{k} \in D} \mathbb{E}[I_{\mathbf{k}}(x(M_n(R)) - x(M_n^{(1)}(\mathbf{k}) + 1)] \right| \leq \Delta x \sum_{\mathbf{k} \in D} \mathbb{E}[I_{\mathbf{k}}(M_n^{(2)}(\mathbf{k}) - 1)], \quad (4.16)$$

where the last inequality follows from (4.14). Since $I_{\mathbf{k}} \leq 1$ because the A_i are disjoint, the right-hand side in (4.16) is bounded above by

$$\Delta x \sum_{\mathbf{k} \in D} \mathbb{E}[I_{\mathbf{k}} \sum_{\ell \in D_{2\mathbf{k}}} I_{\ell}]$$

$$= \Delta x \sum_{\mathbf{k} \in D} \sum_{\ell \in D_{2\mathbf{k}}} \mathbb{E}\Big[\Big(\sum_{i=1}^{q} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p})\mathbb{1}_{B_i}(T_{\mathbf{k}})\Big)\Big(\sum_{j=1}^{q} \mathbb{1}_{A_j}(\tfrac{\ell}{p})\mathbb{1}_{B_j}(T_{\ell})\Big)\Big] \qquad (4.17)$$

$$= \Delta x \sum_{i=1}^{q} \sum_{j=1}^{q} \sum_{\mathbf{k} \in D} \sum_{\ell \in D_{2\mathbf{k}}} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p})\mathbb{1}_{A_j}(\tfrac{\ell}{p})\mathbb{P}(T_{\mathbf{k}} \in B_i, T_{\ell} \in B_j)$$

We set $D_{2\mathbf{k},r} := \{\ell \in D : |\{\ell_1, \ldots, \ell_m, k_1, \ldots, k_m\}| = 2m - r\}$. Notice that $\bigcup_{r=1}^{m-1} D_{2\mathbf{k},r} = D_{2\mathbf{k}}$. Therefore, (4.17) is equal to

$$\Delta x \sum_{i=1}^{q} \sum_{j=1}^{q} \sum_{r=1}^{m-1} \sum_{\mathbf{k} \in D} \sum_{\ell \in D_{2\mathbf{k},r}} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p})\mathbb{1}_{A_j}(\tfrac{\ell}{p})\mathbb{P}(T_{\mathbf{k}} \in B_i, T_{\ell} \in B_j).$$

By assumption (A2), we have $p^{2m-r}\mathbb{P}(T_{\mathbf{k}} \in B_i, T_{\ell} \in B_j) \to 0$ for $r = 1, \ldots m - 1$ as $n \to \infty$. Additionally, it holds that

$$\frac{1}{p^{2m-r}} \sum_{\mathbf{k} \in D} \sum_{\ell \in D_{2\mathbf{k},r}} \mathbb{1}_{A_i}(\tfrac{\mathbf{k}}{p})\mathbb{1}_{A_j}(\tfrac{\ell}{p}) = O(1), \qquad n \to \infty.$$

Consequently the second term of (4.13) tends to zero as $n \to \infty$. This completes the proof. $\qquad\square$

Proof of Proposition 2.5 We proceed similarly to the proof of Proposition 4.20 of Resnick (2008). We want to show that $Y_n \xrightarrow{d} Y$. Therefore, we define a map from the space of point measures $\mathcal{M}(S)$ to $D(0, 1]$, the space of right continuous functions on $(0, 1]$ with finite limits existing from the left, and show that this map is continuous. Then, the Proposition follows by the continuous mapping theorem.

To this end, for a point measure $\mathbf{m} = \sum_{k=1}^{\infty} \varepsilon_{(t_k, y_k)} \in \mathcal{M}(S)$ we define $V_1 : \mathcal{M}(S)$ $V_1 : \mathcal{M}(S) \to D(0, 1]$ through

$$V_1(\mathbf{m}) = V_1\Big(\sum_{k=1}^{\infty} \varepsilon_{(t_k, y_k)}\Big) = \begin{cases} v_{\mathbf{m}} : (0, 1] \to (v, w) \\ v_{\mathbf{m}}(t) = \begin{cases} \bigwedge_{k:t_k \le t} y_k, & \mathbf{m}(((0, 1]^{m-1} \times (0, t] \times (v, w)) \cap S) > 0 \\ \bigwedge_{k:t_k = t^*} y_k, & \text{otherwise,} \end{cases} \end{cases}$$

where $t^* = \sup\{s > 0 : \mathbf{m}(((0, 1]^{m-1} \times (0, s] \times (v, w)) \cap S) = 0\}$. V_1 is well-defined except at $\mathbf{m} \equiv 0$. Recalling the definition of N_n in (2.3), we note that $V_1(N_n)(t) = Y_n(t)$ and $V_1(N)(t) = Y(t)$ for $0 < t \le 1$.

We will start by proving the continuity of V_1 in the case, where $\mu(x) = -\log(H(x))$ and H is the Gumbel distribution. In this case, N has a.s. the following properties

$$N(((0, 1]^{m-1} \times \{1\} \times (-\infty, \infty)) \cap S) = 0,$$
$$N(((0, 1]^{m-1} \times (0, t] \times (x, \infty)) \cap S) < \infty,$$
$$N(((0, 1]^{m-1} \times [s, t] \times (-\infty, x)) \cap S) = \infty,$$

for any $0 < s < t < 1$ and $x \in \mathbb{R}$. Therefore, we only have to show continuity at $\mathbf{m} \in \mathcal{M}(S)$ with these properties. Let $(\mathbf{m}_n)_n$ be a sequence of point measures in $\mathcal{M}(S)$, which converges vaguely to $\mathbf{m}$ ($\mathbf{m}_n \to \mathbf{m}$) as $n \to \infty$ (see Resnick 2008, p. 140). Since $V_1(\mathbf{m})$ is right continuous there exists a right continuous extension on $[0, 1]$, which we denote with $\widetilde{V_1(\mathbf{m})}$. Now choose $\beta < \widetilde{V_1(\mathbf{m})}(0)$ such that $\mathbf{m}(S_1 \times \{\beta\}) = 0$. As $\mathbf{m}_n \overset{v}{\to} \mathbf{m}$, we can conclude from Resnick (2008, Proposition 3.12) that there exists a $1 \leq q < \infty$ such that for n large enough

$$\mathbf{m}_n(S_1 \times (\beta, \infty)) = \mathbf{m}(S_1 \times (\beta, \infty)) = q.$$

We enumerate and designate the q points in the following way $((t_i^{(n)}, j_i^{(n)}), 1 \leq i \leq q)$ with $0 < t_{1,m}^{(n)} < \ldots < t_{q,m}^{(n)} < 1$, where $t_{i,m}^{(n)}$ is the m-th component of $t_i^{(n)}$, such that by Resnick (2008, Proposition 3.13)

$$\lim_{n \to \infty} ((t_i^{(n)}, j_i^{(n)}), 1 \leq i \leq q) = ((t_i, j_i), 1 \leq i \leq q),$$

where $((t_i, j_i), 1 \leq i \leq q)$ is the analogous enumeration of points of $\mathbf{m}$ in $S_1 \times (\beta, \infty)$. Now choose

$$\delta < \frac{1}{2} \min_{1 \leq i, j \leq q} \|t_i - t_j\|_2$$

small enough so that the δ-spheres of the distinct points of the set $\{(t_i, j_i)\}$ are disjoint and in $S_1 \times [\beta, \infty)$. Pick n so large that every δ-sphere contains only one point of $\mathbf{m}_n$. Then set $\lambda_n : [0, 1] \to [0, 1]$ with $\lambda_n(0) = 0$, $\lambda_n(1) = 1$, $\lambda_n(t_{i,m}) = t_{i,m}^{(n)}$ and λ_n is linearly interpolated elsewhere on $[0, 1]$. For this λ_n it holds that

$$\sup_{0 \leq t \leq 1} |\widetilde{V_1(\mathbf{m}_n)}(t) - \widetilde{V_1(\mathbf{m})}(\lambda_n(t))| < \delta \quad \text{and} \quad \sup_{0 \leq t \leq 1} |\lambda_n(t) - t| < \delta.$$

Thereby, we get

$$\tilde{\mathcal{D}}(\widetilde{V_1(\mathbf{m}_n)}, \widetilde{V_1(\mathbf{m})}) = \mathcal{D}(V_1(\mathbf{m}_n), V_1(\mathbf{m})) < \delta,$$

which finishes the proof. The Fréchet and the Weibull case follow by similar arguments. $\square$

Proof of Theorem 2.6 We will proceed similarly as in (Resnick 2008, p. 217-218) using the continuous mapping theorem again. Since Y is the restriction to $(0, 1]$ of an extremal process (see Resnick 2008, Section 4.3), it is a nondecreasing function, which is constant between isolated jumps. Let $D^\uparrow(0, 1]$ be the subset of $D(0, 1]$ that contains all functions with this property. Set

$$V_2 : D^\uparrow(0, 1] \to \mathcal{M}(0, 1]$$

$$x \mapsto \sum_{i=1}^{\infty} \varepsilon_{t_i},$$

where $\{t_i\}$ are the discontinuity points of x. Then $V_2(Y_n) = \sum_{k=1}^{p} \varepsilon_{p^{-1}L(k)}$ and $V_2(Y) = \sum_{k=1}^{\infty} \varepsilon_{\tau_k}$, where $(\tau_k)_k$ is the sequence of discontinuity points of the extremal process generated by the Gumbel distribution $H = \Lambda$, c.f. above Theorem 2.6. By (Embrechts et al. [1997], Theorem 5.4.7) the point process $\sum_{k=1}^{\infty} \varepsilon_{\tau_k}$ is a PRM with mean measure $v(a, b) = \log(b/a)$ for $0 < a < b \le 1$. According to Proposition 2.5, it suffices to show that V_2 is continuous. Let $(x_n)_n$ be a sequence of functions in $D^\uparrow(0, 1]$ with $\mathcal{D}(x_n, x) \to 0$ as $n \to \infty$ for an $x \in D^\uparrow(0, 1]$. Then there exist $\lambda_n \in \Lambda_{[0,1]}$ such that

$$\sup_{0 \le t \le 1} |\tilde{x}_n(\lambda_n(t)) - \tilde{x}(t)| \to 0 \quad \text{and} \tag{4.18}$$

$$\sup_{0 \le t \le 1} |\lambda_n(t) - t| \to 0, \qquad n \to \infty, \tag{4.19}$$

where $\tilde{x}_n$ and $\tilde{x}$ are the right continuous extensions of x_n and x on $[0, 1]$. We want to prove the vague convergence

$$V_2(x_n) = \sum_{i=1}^{\infty} \varepsilon_{t_i^{(n)}} \xrightarrow{v} V_2(x) = \sum_{i=1}^{\infty} \varepsilon_{t_i},$$

where $\{t_i^{(n)}\}$ and $\{t_i\}$ are the discontinuity points of x_n and x, respectively. Consider an arbitrary continuous function f on $(0, 1]$ with compact support contained in an interval $[a, b]$ with $0 < a < b \le 1$, and x is continuous at a and b. It suffices to show that

$$\lim_{n \to \infty} \sum_{i=1}^{\infty} f(t_i^{(n)}) \mathbb{1}_{[a,b]}(t_i^{(n)}) = \sum_{i=1}^{\infty} f(t_i) \mathbb{1}_{[a,b]}(t_i). \tag{4.20}$$

The functions $x_n, x \in D^\uparrow(0, 1]$ have only finitely many discontinuity points in $[a, b]$. Therefore, only a finite number of terms in the sums are not equal to zero. Because of (4.18) and (4.19) the jump times on $[a, b]$ of x_n are close to those of x, which proves (4.20). Hence, V_2 is continuous, which finishes the proof.

4.2 Proof of Proposition 3.7

Let π denote the permutation of $\{1, \dots, n\}$ induced by the order statistics of $X_{21}, \dots, X_{2n}$, i.e.,

$$X_{2\pi(1)} > X_{2\pi(2)} > \cdots > X_{2\pi(n)} \quad \text{a.s.},$$

where the continuity of the distribution of X was used to avoid ties. We can rewrite τ_{12} as

$$
\begin{aligned}
\tau_{12} &= \frac{2}{n(n-1)} \sum_{1 \leq s < t \leq n} \text{sign}(X_{1s} - X_{1t})\text{sign}(X_{2s} - X_{2t}) \\
&= \frac{2}{n(n-1)} \sum_{1 \leq s < t \leq n} \text{sign}(X_{1\pi(s)} - X_{1\pi(t)}) \underbrace{\text{sign}(X_{2\pi(s)} - X_{2\pi(t)})}_{=1 \quad \text{a.s.}} \\
&\overset{d}{=} \frac{2}{n(n-1)} \sum_{1 \leq s < t \leq n} \text{sign}(X_{1s} - X_{1t}).
\end{aligned}
\tag{4.21}
$$

Let $\mathbf{q}_n = (q_1, \ldots, q_n)$ be a permutation of the set $\{1, \ldots, n\}$. If $i < j$ and $q_i > q_j$, we call the pair (q_i, q_j) an inversion of the permutation $\mathbf{q}_n$.

Since the $X_{11}, \ldots, X_{1n}$ are iid, the permutation

$$
\mathbf{q}_n = (Q_{11}, Q_{12}, \ldots, Q_{1n})
$$

consisting of the ranks is uniformly distributed on the set of the $n!$ permutations of $\{1, \ldots, n\}$. By I_n we denote the number of inversions of $\mathbf{q}_n$. For $s < t$, we have

$$
\text{sign}(X_{1s} - X_{1t}) = \begin{cases} 1, & \text{if } Q_{1s} < Q_{1t}, \\ -1, & \text{if } Q_{1s} > Q_{1t} \Leftrightarrow \text{ inversion at } (s, t). \end{cases}
$$

In view of (4.21), this implies

$$
\tau_{12} \overset{d}{=} \binom{n}{2}^{-1} \sum_{1 \leq s < t \leq n} \text{sign}(X_{1s} - X_{1t}) = \binom{n}{2}^{-1} \left[\binom{n}{2} - 2 I_n \right] = 1 - \frac{4}{n(n-1)} I_n.
$$

By Kendall and Stuart (1973, p. 479) or Margolius (2001, p. 3) (see also Sachkov 1997) the moment generating function of I_n is

$$
\mathbb{E}\left[e^{t I_n} \right] = \prod_{j=1}^{n} \frac{1 - e^{jt}}{j(1 - e^t)}, \quad t \in \mathbb{R}.
$$

We recognize that $\frac{1 - e^{jt}}{j(1 - e^t)}$ is the moment generating function of a uniform distribution on the integers $0, 1, \ldots, j - 1$. Let $(U_i)_{i \geq 1}$ be a sequence of independent random variables such that U_i is uniformly distributed on the integers $0, 1, \ldots, i$. We get

$$
1 - \frac{4}{n(n-1)} I_n \overset{d}{=} 1 - \frac{4}{n(n-1)} \sum_{i=1}^{n-1} U_i \overset{d}{=} \frac{4}{n(n-1)} \sum_{i=1}^{n-1} (U_i - i/2),
$$

establishing the desired result.

4.3 Proof of Proposition 3.10

Our idea is to transfer the convergence of N_n^X onto N_n^Y. To this end, it suffices to show (see Kallenberg 1983, Theorem 4.2) that for any continuous function f on $\mathbb{R}$ with compact support,

$$\int f \, dN_n^Y - \int f \, dN_n^X \xrightarrow{\mathbb{P}} 0, \quad n \to \infty.$$

Suppose the compact support of f is contained in $[K + \gamma_0, \infty)$ for some $\gamma_0 > 0$ and $K \in \mathbb{R}$. Since f is uniformly continuous, $\omega(\gamma) := \sup\{|f(x) - f(y)| : x, y \in \mathbb{R}, |x - y| \le \gamma\}$ tends to zero as $\gamma \to 0$. We have to show that for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(\left| \sum_{i=1}^{p} \left(f(Y_{i,n}) - f(X_{i,n}) \right) \right| > \varepsilon \right) = 0. \tag{4.22}$$

On the sets

$$A_{n,\gamma} = \left\{ \max_{i=1,\ldots,p} |Y_{i,n} - X_{i,n}| \le \gamma \right\}, \quad \gamma \in (0, \gamma_0),$$

we have

$$\left| f(Y_{i,n}) - f(X_{i,n}) \right| \le \omega(\gamma) \, \mathbb{1}_{\{X_{i,n} > K\}}.$$

Therefore, we see that, for $\gamma \in (0, \gamma_0)$,

$$\mathbb{P}\left(\left| \sum_{i=1}^{p} \left(f(Y_{i,n}) - f(X_{i,n}) \right) \right| > \varepsilon, A_{n,\gamma} \right)$$

$$\le \mathbb{P}\left(\omega(\gamma) \# \{ 1 \le i \le p : X_{i,n} > K \} > \varepsilon \right)$$

$$\le \frac{\omega(\gamma)}{\varepsilon} \mathbb{E}\left[\# \{ 1 \le i \le p : X_{i,n} > K \} \right]$$

$$= \frac{\omega(\gamma)}{\varepsilon} \mathbb{E} N_n^X((K, \infty])$$

$$\to \frac{\omega(\gamma)}{\varepsilon} \mathbb{E} N((K, \infty]), \quad n \to \infty.$$

By assumption, it holds $\lim_{n \to \infty} \mathbb{P}(A_{n,\gamma}^c) = 0$. Thus, letting $\gamma \to 0$ establishes (4.22).

Declarations

Ethical approval Not applicable.

Conflict of interest None.

References

Baddeley, A.: Spatial point processes and their applications. In Stochastic geometry, vol. 1892 of Lecture Notes in Math. Springer, Berlin, pp. 1–75 (2007)

Barbour, A.D., Eagleson, G.K.: Poisson convergence for dissociated statistics. J. Roy. Statist. Soc. Ser. B **46**(3), 397–402 (1984)

Barbour, A.D., Holst, L., Janson, S.: Poisson approximation, vol. 2 of Oxford Studies in Probability. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications (1992)

Basrak, B., Planinić, H.: Compound Poisson approximation for regularly varying fields with application to sequence alignment. Bernoulli **27**(2), 1371–1408 (2021)

Billingsley, P.: Convergence of probability measures, second ed. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication (1999)

Cai, T., Liu, W.: Adaptive thresholding for sparse covariance matrix estimation. J. Amer. Statist. Assoc. **106**(494), 672–684 (2011)

Cai, T., Liu, W., Xia, Y.: Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. J. Amer. Statist. Assoc. **108**(501), 265–277 (2013)

Cai, T.T.: Global testing and large-scale multiple testing for high-dimensional covariance structures. Annu. Rev. Stat. Appl. **4**, 423–446 (2017)

Cai, T.T., Jiang, T.: Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. Ann. Statist. **39**(3), 1496–1525 (2011)

Chenavier, N., Henze, N., Otto, M.: Limit laws for large kth-nearest neighbor balls. J. Appl. Probab. **59**(3), 880–894 (2022)

Clarke, R., Ressom, H.W., Wang, A., Xuan, J., Liu, M.C., Gehan, E.A., Wang, Y.: The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat. Rev. Cancer **8**(1), 37–49 (2008)

Dabrowski, A.R., Dehling, H.G., Mikosch, T., Sharipov, O.: Poisson limits for U-statistics. Stochastic Process. Appl. **99**(1), 137–157 (2002)

Decreusefond, L., Schulte, M., Thäle, C.: Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. Ann. Probab. **44**(3), 2147–2197 (2016)

Donoho, D.: High-dimensional data analysis: the curses and blessings of dimensionality. Technical Report, Stanford University (2000)

Drton, M., Han, F., Shi, H.: High-dimensional consistent independence testing with maxima of rank correlations. Ann. Statist. **48**(6), 3206–3227 (2020)

Embrechts, P., Klüppelberg, C., Mikosch, T.: 'Modelling Extremal Events for Insurance and Finance'. Applications of Mathematics (New York), vol. 33. Springer, Berlin (1997)

Gösmann, J., Stoehr, C., Heiny, J., Dette, H.: Sequential change point detection in high dimensional time series. Electron. J. Stat. **16**(1), 3608–3671 (2022)

Han, F., Chen, S., Liu, H.: Distribution-free tests of independence in high dimensions. Biometrika **104**(4), 813–828 (2017)

Heiny, J., Kleemann, C.: Maximum interpoint distance of high-dimensional random vectors. arXiv preprint arXiv:2302.06965 (2023)

Hoeffding, W.: A class of statistics with asymptotically normal distribution. Ann. Math. Statistics **19**, 293–325 (1948)

Johnstone, I.M., Titterington, D.M.: Statistical challenges of high-dimensional data. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **367**(1906)4237–4253 (2009)

Kallenberg, O.: Random measures, third ed. Akademie-Verlag, Berlin; Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London (1983)

Kallenberg, W.C.M.: Cramér type large deviations for simple linear rank statistics. Z. Wahrsch. Verw. Gebiete **60**(3), 403–409 (1982)

Kendall, M.G., Stuart, A.: The advanced theory of statistics. Vol. 2, third ed. Hafner Publishing Co., New York. Inference and relationship (1973)

Margolius, B.H.: Permutations with inversions. J. Integer Seq. 4, 2, Article 01.2.4, 13 (2001)

Panaretos, V.M., Zemel, Y.: An invitation to statistics in Wasserstein space. SpringerBriefs in Probability and Mathematical Statistics. Springer, Cham (2020)

Resnick, S.I.: Extreme Values, Regular Variation and Point Processes. Springer Series in Operations Research and Financial Engineering. Springer, New York. Reprint of the 1987 original (2008)

Sachkov, V.N.: Probabilistic methods in combinatorial analysis, vol. 56 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge. Translated from the Russian, Revised by the author (1997)

Schulte, M., Thäle, C.: The scaling limit of Poisson-driven order statistics with applications in geometric probability. Stochastic Process. Appl. **122**(12), 4096–4120 (2012)

Schulte, M., Thäle, C.: Poisson point process convergence and extreme values in stochastic geometry. In Stochastic analysis for Poisson point processes, vol. 7 of Bocconi Springer Ser. Bocconi Univ. Press, pp. 255–294 (2016)

Silverman, B., Brown, T.: Short distances, flat triangles and Poisson limits. J. Appl. Probab. **15**(4), 815–825 (1978)

Wu, W.B.: Nonlinear system theory: another look at dependence. Proc. Natl. Acad. Sci. USA **102**(40), 14150–14154 (2005)

Xiao, H., Wu, W.B.: Asymptotic theory for maximum deviations of sample covariance matrix estimates. Stochastic Process. Appl. **123**(7), 2899–2920 (2013)

Zhou, C., Han, F., Zhang, X.-S., Liu, H.: An extreme-value approach for testing the equality of large U-statistic based correlation matrices. Bernoulli **25**(2), 1472–1503 (2019)