

Some notes on Poisson limits for empirical point processes

André Dabrowski

B. Gail Ivanoff
givanoff@uottawa.ca

Rafał Kulik
rkulik@uottawa.ca

Department of Mathematics and Statistics
University of Ottawa
585 King Edward Av.
Ottawa, ON K1N 6N5, Canada

January 20, 2009

Abstract

Define the scaled empirical point process on an independent and identically distributed sequence $\{Y_i : i \leq n\}$ on \mathbb{R}^d as the random point measure with masses at $a_n^{-1}Y_i$. For suitable a_n we obtain the weak limit of these point processes through a novel use of a dimension-free method based on the convergence of compensators of multiparameter martingales. The method extends previous results in several directions. We obtain limits at points where the density of Y_i may be zero, but has regular variation. The joint limit of the empirical process evaluated at distinct points is given by independent Poisson processes. Applications are provided both to nearest-neighbour density estimation in high dimensions, and to the asymptotic behaviour of multivariate extremes such as those arising from bivariate normal copulas.

Keywords: multiparameter martingales; point processes; density estimation; multivariate extremes; local empirical processes

1 Introduction

Point processes and their limits arise naturally in many areas of statistics, and have a number of applications ranging from survival analysis to spatial statistics. Point processes also arise in probability theory as limits for extreme value processes, in studying limits of sums of stable non-Gaussian variables and in queuing models. Of course the Poisson process is a fundamental concept in

martingale theory. Weak convergence of the empirical point process underlies many applications, and this paper employs the relatively recent area of multiparameter martingales to establish a novel and unified approach to proving such limits for scaled empirical point processes. Although various elegant and powerful methods have been developed for particular classes of problems, the generalized martingale approach provides an extremely simple, dimension-free method of addressing a variety of old and new distributional questions.

Given a random sample of random vectors $\{Y_i : i \leq n\}$ in \mathbb{R}^d and a suitable class of sets $\{A\}$, the empirical point process is defined by

$$N_A^{(n)} = \sum_{i=1}^n \mathbb{I}_{\{Y_i \in A\}}.$$

As noted above, the weak convergence of $N_A^{(n)}$ has been extensively studied using a variety of methods. In particular, in the univariate case, a strong approximation approach can be used to establish weak convergence of the local empirical process (see Einmahl, 1997, and the references therein):

$$L_{n,x}(t) = \sum_{i=1}^n \mathbb{I}_{\{Y_i \in [x - ta_n, x + ta_n]\}}, \quad t \in [0, 1]. \quad (1)$$

If the sequence of constants, a_n , is appropriately chosen then the limit process is homogeneous Poisson. However, this strong Poisson approximation is difficult to implement (or at least cannot be extended directly) if one wants to study the joint behaviour of

$$(L_{n,x_1}(\cdot), \dots, L_{n,x_m}(\cdot)),$$

i.e. when estimating the density of Y_1 simultaneously at (x_1, \dots, x_m) (see Section 4.1). Even in the Gaussian case, where simultaneous approximation by independent Wiener processes is known, Deheuvels *et al.* (2000) points out that a major technical difficulty arises in proving independence of $L_{n,x_j}(\cdot)$ at separate x_j .

The aim of this paper is to illustrate how to apply the multiparameter martingale theory of Ivanoff and Merzbach (2000) to study *weak* Poisson limits for empirical point processes on \mathbb{R}^d . This approach requires only the simple computation of so-called *-compensators to identify Poisson limits for scaled empirical point processes. The compensator method exploited here is particularly attractive in that it is independent of the dimension of the underlying random vectors, and so easily generalizes results from the univariate to the multivariate case. In addition, the martingale approach allows one to handle the joint behaviour at multiple points with ease through a judicious definition of the associated history (filtration). In particular, we shall show that the asymptotic behaviour of the local empirical process at distinct points x_1, \dots, x_m can be described by independent Poisson processes, an intuitive but otherwise technically challenging result.

The method has additional benefits. First, only (multivariate) regular variation of the density f of Y_1 is required, and the limits are explicitly written

in terms of f . Indeed, we can discover the appropriate scaling constants even when f is regularly varying but $f(0) = 0$, i.e. a case with inhomogeneous Poisson limits excluded in Borisov (2000), or characterize the distributional behaviour of joint extremes for different bivariate copulas. This recovers Einmahl (1997, Corollaries 2.4 and 2.5) where the limit Poisson process has a product mean measure, but also extends to more complex cases (Corollary 4.4). In particular, this method identifies extreme value limits for copulas with asymptotically dependent multivariate extremes more simply than methods employing multivariate regular variation (c.f. Resnick, 1987).

The paper is structured as follows. The next section will review key elements of the theory of multi-parameter martingales and in particular, the use of *-compensators in proving weak convergence of a sequence of point processes. Explicit formulas are given for the *-compensators of empirical point processes. Section 3 defines the scaled empirical point process generated by a sample and establishes Poisson process limits for such processes at arbitrary quantiles. This proceeds in steps from the classical non-negative and univariate case (yielding limits similar to those for extreme value processes), to the multivariate and multidimensional cases. In each case the proof simplifies to the straightforward calculation of *-compensators, and highlights the universality of the martingale approach. Section 4 on applications illustrates the utility of our results by establishing for the first time weak limits for nearest-neighbour estimates of joint densities (again at several points simultaneously), and by providing new extreme value limits for multivariate copulas.

2 Notation and background: Point processes and martingale methods

We provide a brief introduction to point processes and martingale methods indexed by general Euclidean spaces using the set-indexed framework introduced in Ivanoff and Merzbach (2000). We need definitions mimicking those for martingales indexed by \mathbb{R}_+ .

Set $T = \mathbb{R}_+^d$, and $\mathcal{A} = \{A_t = [0, t] : t \in T\} \cup \{\emptyset\}$, where $[0, t] = [0, t_1] \times \dots \times [0, t_d]$ for $t = (t_1, \dots, t_d)$. If $T = \mathbb{R}$ and $t_i < 0$, then $[0, t_i]$ is interpreted as $[t_i, 0]$. Set-inclusion on \mathcal{A} induces a partial order, \preceq , on T : $s \preceq t$ if and only if $A_s \subseteq A_t$. This is not the usual partial order on \mathbb{R}^d : e.g. $\{0\}$ is the (unique) minimal element, and all quadrants are equipped with their own partial order. In particular, if $T = \mathbb{R}$, points with different signs are incomparable. This special structure permits us to define a 2^d -sided martingale theory.

The semi-algebra \mathcal{C} is the class of all subsets of T of the form

$$C = A \setminus B, \quad A \in \mathcal{A}, \quad B \in \mathcal{A}(u),$$

where $\mathcal{A}(u)$ denotes the class of sets which are finite unions of sets from \mathcal{A} . Let (Ω, \mathcal{F}, P) be any complete probability space.

Example 2.1 If $T = \mathbb{R}_+$, then $\mathcal{A}(u) = \mathcal{A}$, since a finite union of intervals $[0, t_i]$, $t_i > 0$, leads to another interval of this form. Consequently, \mathcal{C} is the set of intervals $(s, t]$, $0 \leq s < t$. If $T = \mathbb{R}_+^2$, then $\mathcal{A}(u) \neq \mathcal{A}$. The class \mathcal{C} includes (but is much richer than) the class of all rectangles of the form $(s_1, t_1] \times (s_2, t_2]$, $0 \leq s_1 < t_1$, $0 \leq s_2 < t_2$.

A filtration indexed by $\mathcal{A}(u)$ is a class $\{\mathcal{F}_A : A \in \mathcal{A}(u)\}$ of complete sub- σ -fields of \mathcal{F} (for details we refer to Ivanoff and Merzbach (2000)). For consistency, we define $\mathcal{F}_T = \mathcal{F}$. We may associate σ -algebras with sets in \mathcal{C} : for $C \in \mathcal{C} \setminus \mathcal{A}$, let $\mathcal{G}_C^* = \bigvee_{B \in \mathcal{A}(u), B \cap C = \emptyset} \mathcal{F}_B$, and for $A \in \mathcal{A}$, $A \neq \emptyset$, define $\mathcal{G}_A^* = \mathcal{F}_\emptyset$. A (\mathcal{A} -indexed) *stochastic process* $X = \{X_A : A \in \mathcal{A}\}$ is a collection of random variables indexed by \mathcal{A} , and is *adapted* if X_A is \mathcal{F}_A -measurable for every $A \in \mathcal{A}$. By convention, $X_{\{\emptyset\}} = 0$.

Example 2.2 If $T = \mathbb{R}_+$, take $C = (s, t] \in \mathcal{C}$. Then the only sets which belong to $\mathcal{A}(u) = \mathcal{A}$ and do not intersect with C are the intervals $[0, v]$, $v \leq s$. Then $\mathcal{G}_C^* = \bigvee_{[0, v], v \leq s} \mathcal{F}_{[0, v]} = \mathcal{F}_{[0, s]}$, which is the history up to time s .

If $T = \mathbb{R}_+^2$, take $C = (s_1, t_1] \times (s_2, t_2]$, $s_1 < t_1$, $s_2 < t_2$. Now, all of the sets $[0, u_1] \times [0, u_2] \in \mathcal{A}(u)$, where $u_1 \leq s_1$ and $u_2 \geq 0$, or $u_1 \geq 0$ and $u_2 \leq s_2$, do not intersect with C . Consequently,

$$\mathcal{G}_C^* = \mathcal{F}_{[0, s_1] \times [0, \infty)} \vee \mathcal{F}_{[0, \infty) \times [0, s_2]}.$$

In other words, \mathcal{G}_C^* is the extended or wide history at (s_1, s_2) , and contains all information which is not strictly in the future of (s_1, s_2) . Note that this is much larger than the strict history $\mathcal{F}_{[0, s_1] \times [0, s_2]}$; see Remark 2.5.

A process $X : \mathcal{A} \rightarrow \mathbb{R}$ is *increasing* if for every $\omega \in \Omega$, the function $X(\omega)$ can be extended to a finitely additive function on \mathcal{C} satisfying $X_{\{\emptyset\}}(\omega) = 0$ and $X_C(\omega) \geq 0$, $\forall C \in \mathcal{C}$, and such that if (A_n) is a decreasing sequence of sets in $\mathcal{A}(u)$ such that $\bigcap_n A_n \in \mathcal{A}(u)$, then $\lim_n X_{A_n}(\omega) = X_{\bigcap_n A_n}(\omega)$. A process $N = \{N_A, A \in \mathcal{A}\}$ is a *point process* if it is an increasing process taking its values in \mathbb{N} , and almost surely for any $t \in T$, $N_{\{t\}} = 0$ or 1 .

Example 2.3 Once again, if $T = \mathbb{R}_+$ then the process X is increasing if for each $[s, t] \in \mathcal{C}$, $s < t$, we have $X_C = X_t - X_s \geq 0$. If $T = \mathbb{R}_+^2$, then for $C = (s_1, t_1] \times (s_2, t_2]$, $s_1 < t_1$, $s_2 < t_2$ we have $X_C = X_{t_1, t_2} - X_{s_1, t_2} - X_{t_1, s_2} + X_{s_1, s_2}$.

Note that if N is a point process on $T = \mathbb{R}$, then $N_t := N_{[0, t]}$ (for t positive or negative) and not $N_{(-\infty, t]}$. A similar remark is true for point processes defined on \mathbb{R}^d .

Now, let Y be a random variable with values in T and let $N_C = 1_{\{Y \in C\}}$. Then clearly N is a single jump point process (a point process with only one jump).

Let Λ be a σ -finite measure on T . As usual, N is a Poisson process on T with mean measure Λ if N is a point process where $N_C \sim \text{Poisson}, \Lambda_C$, $\forall C \in \mathcal{C}$, and whenever $C_1, \dots, C_n \in \mathcal{C}$ are disjoint, N_{C_1}, \dots, N_{C_n} are independent. If Λ is

absolutely continuous with respect to Lebesgue measure, its density λ is called the intensity of the Poisson process.

An integrable process $M = \{M_A, A \in \mathcal{A}\}$ is called a *pseudo-strong martingale* if for any $C \in \mathcal{C}$, $E[M_C | \mathcal{G}_C^*] = 0$. The process \bar{X} is a **-compensator* of the increasing process X if it is increasing and the difference $X - \bar{X}$ is a pseudo-strong martingale. The asymptotic behaviour of a sequence of point processes may be determined by *-compensators as shown in the following lemma which generalizes Theorem 8.2.2 and Corollary 8.2.3 of Ivanoff and Merzbach (2000) to multivariate point processes on $T = \mathbb{R}^d$ or \mathbb{R}_+^d .

To state this result, we consider k point processes $N(1), \dots, N(k)$ all adapted to a common $\mathcal{A}(u)$ -indexed filtration $\{\mathcal{F}_A\}$ and so that with probability one, none of the processes have a jump point in common. The k -variate point process \vec{N} is defined by $\vec{N}_A = \langle N_A(1), \dots, N_A(k) \rangle$ and has (k -variate) *-compensator $\vec{\Lambda} = (\Lambda(1), \dots, \Lambda(k))$ if $\Lambda(i)$ is a *-compensator for $N(i)$ with respect to the common filtration $\{\mathcal{F}_A\}$.

In what follows, “ \longrightarrow_P ” denotes convergence in probability and “ $\longrightarrow_{\mathcal{D}}$ ” denotes convergence in both finite dimensional distribution and in distribution in the Skorokhod topology if $T = \mathbb{R}_+^d$ (identifying $N_t^{(n)}$ (respectively, N_t) with $N_{A_t}^{(n)}$ (respectively, N_{A_t})). We remark that the Skorokhod topology may be extended to all of the quadrants in \mathbb{R}^d on the space of “outer-continuous functions with inner limits”, and the convergence in the lemma below holds in this case as well. In the sequel, convergence in the Skorokhod topology will be interpreted in this way.

Lemma 2.4 *Let $(\vec{N}^{(n)})$ be a sequence of k -variate point processes on T adapted to a filtration $\{\mathcal{F}_A\}$ and $(\vec{\Lambda}^{(n)})$ a sequence of corresponding *-compensators. Suppose that for each $A \in \mathcal{A}$ and $i = 1, \dots, k$ the sequences $(N_A^{(n)}(i))$ and $(\Lambda_A^{(n)}(i))$ are uniformly integrable and that $\Lambda_A^{(n)}(i) \longrightarrow_P \Lambda_A(i)$ where $\Lambda(i)$ is a deterministic measure on T absolutely continuous with respect to Lebesgue measure. Then $\vec{N}^{(n)} \longrightarrow_{\mathcal{D}} \vec{N}$, where $\vec{N} = \langle N(1), \dots, N(k) \rangle$ and $N(1), \dots, N(k)$ are independent Poisson processes with mean measures $\Lambda(1), \dots, \Lambda(k)$, respectively.*

Remark 2.5 The above lemma shows the importance of the *-martingale concept (i.e. defining the martingale property with respect to the extended past \mathcal{G}_C^*). Generally speaking, a deterministic *-compensator characterizes a Poisson process. On the other hand, if we consider $T = \mathbb{R}_+^2$, we can still define the martingale property with respect to the strict past (see Example 2.2), but in this case it is possible to construct several adapted point processes which have the same compensator (see Ivanoff and Merzbach (2000) for the examples). In other words, working with the compensators with respect to the strict past would not permit us to conclude convergence to Poisson processes in Lemma 2.4.

Proof: The proof of this theorem is a straightforward generalization of Theorem 8.2.2 in Ivanoff and Merzbach (2000) along with an application of Watanabe’s

characterization of the k -variate Poisson process on \mathbb{R}_+ ; see Brémaud (1981, Theorem T6).

◊

We conclude this section by defining empirical point processes on T and computing their $*$ -compensators. Let Y be a T -valued random variable with continuous distribution function F . The *single jump* point process $J = \{J_A = \mathbb{I}_{\{Y \in A\}} : A \in \mathcal{A}\}$ has $*$ -compensator

$$\bar{J}_A = \int_A \mathbb{I}_{\{Y \in E_u\}} (F(E_u))^{-1} dF(u)$$

with respect to its minimal filtration, where $E_t = \{t' \in T : t \preceq t'\}$ (cf. Ivanoff and Merzbach (2000), equation (8.6)). Now, suppose that Y_1, \dots, Y_n are i.i.d. with distribution $F(t) = \mathbb{P}(Y_i \leq t)$ and let $\mathcal{F} = \bigvee_{i=1}^n \mathcal{F}^{(i)}$ where $\mathcal{F}^{(i)}$ is the minimal filtration generated by the single jump process associated with Y_i . Let A_Y be the random set $[0, Y]$. Then the empirical point process $N^{(n)}$ defined by

$$N_A^{(n)} = \sum_{i=1}^n \mathbb{I}_{\{Y_i \in A\}} \quad (2)$$

has $*$ -compensator $\Lambda^{(n)}$ where

$$\Lambda_A^{(n)} = \sum_{i=1}^n \int_A \mathbb{I}_{\{Y_i \in E_u\}} (F(E_u))^{-1} dF(u). \quad (3)$$

Example 2.6 If $T = \mathbb{R}_+$ then (3) reads as follows: $A = A_t = [0, t]$ for $t \geq 0$, \preceq is just \leq , the standard ordering, $E_u = [u, \infty)$. By noting that $F(E_u) = \mathbb{P}(Y_i \geq u) =: \bar{F}(u)$ we have the standard result:

$$\Lambda_t^{(n)} := \Lambda_{A_t}^{(n)} = \int_{A_t} \sum_{i=1}^n \mathbb{I}_{\{Y_i \in E_u\}} \frac{dF(u)}{F(E_u)} = \int_0^t \sum_{i=1}^n \mathbb{I}_{\{Y_i \geq u\}} \frac{dF(u)}{\bar{F}(u)}. \quad (4)$$

Example 2.7 If $T = \mathbb{R}$ then $A_t = [0, t]$ or $A_t = [t, 0]$ depending on the sign of t . We have $s \preceq t$ if $0 \leq s \leq t$, or $t \leq s \leq 0$, where \leq is the standard order on \mathbb{R} . Points with different signs are incomparable. The sets E_u will be either $[u, \infty)$ or $(-\infty, u]$ depending the sign of u . If $u > 0$, then as above $F(E_u) = \bar{F}(u)$; otherwise, if $u < 0$, then $F(E_u) = F(u)$. Now the $*$ -compensator is given by (4) if $t \geq 0$, and if $t < 0$,

$$\Lambda_t^{(n)} := \Lambda_{A_t}^{(n)} = \int_{A_t} \sum_{i=1}^n \mathbb{I}_{\{Y_i \in E_u\}} \frac{dF(u)}{F(E_u)} = \int_t^0 \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq u\}} \frac{dF(u)}{F(u)}. \quad (5)$$

Example 2.8 Consider $d = 2$. Let $Y_i = (Y_{i1}, Y_{i2})$, $i = 1, \dots, n$. If $T = \mathbb{R}_+^2$ then (3) reads as follows: $A = A_{\mathbf{t}} = [0, t_1] \times [0, t_2]$ for $\mathbf{t} = (t_1, t_2)$, $E_{\mathbf{u}} = \{\mathbf{t}' :$

$t'_i \geq u_i, i = 1, 2\}$, for $\mathbf{u} = (u_1, u_2)$. By noting that $F(E_{\mathbf{u}}) = \mathbf{P}(Y_{i1} \geq u_1, Y_{i2} \geq u_2) =: \bar{F}(\mathbf{u})$ we have

$$\Lambda_{\mathbf{t}}^{(n)} := \Lambda_{A_{\mathbf{t}}}^{(n)} = \int_{A_{\mathbf{t}}} \sum_{i=1}^n \mathbb{I}\{Y_{i1} \geq u_1, Y_{i2} \geq u_2\} \frac{dF(\mathbf{u})}{\bar{F}(\mathbf{u})} .$$

To extend this example to $T = \mathbb{R}^2$ we proceed as in Example 2.7, treating each quadrant separately.

3 Poisson limits at quantiles

A Poisson limit theorem for a scaled empirical point process is given in Theorem 8.3.3 of Ivanoff and Merzbach (2000). In particular, for a sequence $\{Y_i\}$ of \mathbb{R}^d -valued variables with distribution F , it defines the asymptotic behaviour of $N^{(n)}$ near the origin, provided that F has derivatives at 0 in all quadrants. The results of this section will be seen to extend and generalize this theorem in many directions.

3.1 Univariate case

We can use the preceding section to determine the limiting behaviour of empirical point processes at quantiles. Consider a sequence $\{Y_n\}$ of i.i.d. real-valued positive random variables with distribution F . Assume now that $F(0) = 0$ and that F is regularly varying at 0 with index $\alpha > 0$, i.e. for all $t \geq 0$,

$$\lim_{x \searrow 0} \frac{F(xt)}{F(x)} = t^\alpha , \quad (6)$$

see e.g. Resnick (1987). This implies that for x in a neighbourhood of 0, $F(x) = \ell(x)x^\alpha$. Here and in the sequel ℓ is a slowly varying function at 0 or at ∞ as required, and it can be different at each appearance.

Let a_n be such that $F(a_n) = n^{-1}$. This ensures that $a_n \sim n^{-1/\alpha}\ell(n)$ for some function ℓ slowly varying at ∞ . Henceforth, we write $c_n \sim d_n$ if $\lim_{n \rightarrow \infty} c_n/d_n = 1$.

Since $a_n \rightarrow 0$ we have

$$nF(a_nt) = \frac{F(a_nt)}{F(a_n)} \rightarrow t^\alpha . \quad (7)$$

Define

$$N^{(n)} = \sum_{i=1}^n \delta_{a_n^{-1}Y_i} .$$

We have by (3)

$$\Lambda_t^{(n)} = \int_0^t \sum_{i=1}^n \mathbb{I}\{a_n^{-1}Y_i \geq u\} \frac{dF(a_nu)}{\bar{F}(a_nu)} . \quad (8)$$

We first reprove the well-known result (see e.g. Resnick (1987, Proposition 3.21) concerning Poisson limits for empirical point processes. An elegant argument can be applied (see e.g. Borisov, 2000, and the references therein) where the law of $N^{(n)}$ is approximated (in the total variation sense and for each n separately) by $Poi(\nu_n)$, the Poisson random measure with $\nu_n(A) = nE(1_{\{n^{1/\alpha} X_i \in A\}})$, where the X_i 's are uniform on an appropriately chosen ball. If n is sufficiently large, strong approximation methods yield the coupling of the empirical point processes to a single Poisson random measure, and weak convergence follows. The alternative approach to the proof presented here illustrates the martingale technique. It will be seen subsequently that the proof easily generalizes to the multivariate context and to establishing simultaneous limits at interior quantiles of F .

Corollary 3.1 *Assume that $F(0) = 0$ and (6) holds. Then the sequence $(N^{(n)})$ converges in distribution to N in the Skorokhod topology on $D[0, \infty)$ where N is a Poisson process with mean measure $\Lambda_t = t^\alpha$ (intensity $\lambda(t) = \alpha t^{\alpha-1}$).*

Proof. Since $N^{(n)}$ is square integrable with bounded second moments (uniformly in n), the conditions of Lemma 2.4 will be satisfied if it is shown that the sequence $(\Lambda_t^{(n)})$ given by (8) converges in L_2 to t^α .

$$\begin{aligned} \mathbb{E}[\Lambda_t^{(n)}] &= \mathbb{E} \left[\int_0^t \sum_{i=1}^n \mathbb{I}\{a_n^{-1} Y_i \geq u\} \frac{dF(a_n u)}{\bar{F}(a_n u)} \right] \\ &= \sum_{i=1}^n \int_0^t \mathbb{P}(a_n^{-1} Y_i \geq u) \frac{dF(a_n u)}{\bar{F}(a_n u)} = \sum_{i=1}^n \int_0^t dF(a_n u) = nF(a_n t) \rightarrow t^\alpha \end{aligned}$$

by applying (7). Using the independence of the Y_i 's,

$$\begin{aligned} \mathbb{E}[(\Lambda_t^{(n)})^2] &= \int_0^t \int_0^t \sum_{i=1}^n \mathbb{E}[\mathbb{I}\{Y_i \geq a_n u, Y_i \geq a_n v\}] \frac{dF(a_n u) dF(a_n v)}{\bar{F}(a_n u) \bar{F}(a_n v)} \\ &\quad + 2 \int_0^t \int_0^t \sum_{i < j}^n \mathbb{E}[\mathbb{I}\{Y_i \geq a_n u, Y_j \geq a_n v\}] \frac{dF(a_n u) dF(a_n v)}{\bar{F}(a_n u) \bar{F}(a_n v)} \\ &= \int_0^t \int_0^t \sum_{i=1}^n \mathbb{P}(Y_i \geq a_n(u \vee v)) \frac{dF(a_n u) dF(a_n v)}{\bar{F}(a_n u) \bar{F}(a_n v)} \\ &\quad + n(n-1) \int_0^t \int_0^t dF(a_n u) dF(a_n v) \\ &= n \int_0^t \int_0^t \mathbb{P}(Y_1 \geq a_n(u \vee v)) \frac{dF(a_n u) dF(a_n v)}{\bar{F}(a_n u) \bar{F}(a_n v)} \\ &\quad + n(n-1)(F(a_n t))^2. \end{aligned}$$

Now, the first term converges to 0, because

$$n \int_0^t \int_0^t \mathbb{P}(Y_1 \geq a_n(u \vee v)) \frac{dF(a_n u) dF(a_n v)}{\bar{F}(a_n u) \bar{F}(a_n v)}$$

$$\leq n \int_0^t \int_0^t \frac{dF(a_n u) dF(a_n v)}{\bar{F}(a_n u)} \leq n [\bar{F}(a_n t)]^{-1} [F(a_n t)]^2,$$

and $\bar{F}(a_n t) \rightarrow 1$ and $nF^2(a_n t) \rightarrow 0$ as $n \rightarrow \infty$. Consequently, $(\Lambda_t^{(n)})$ converges to t^α in L_2 and therefore in probability. \odot

We may extend Corollary 3.1 to the entire line. Denote by P_F the probability measure associated with a distribution F . We say that F is *regularly varying on the right (left)* at u with index $\alpha > 0$ ($\beta > 0$) if for all $t > 0$, respectively,

$$\lim_{x \searrow 0} \frac{P_F((u, u + xt])}{P_F((u, u + x])} = t^\alpha \quad \left(\lim_{x \searrow 0} \frac{P_F((u - xt, u])}{P_F((u - x, u])} = t^\beta \right), \quad (9)$$

Clearly, if F has support $(0, \infty)$ then for $u = 0$ the above condition reduces to (6).

If F satisfies (9), we shall choose a_n and b_n so that

$$F(a_n + u) - F(u) = n^{-1}, \quad F(u) - F(-b_n + u) = n^{-1}. \quad (10)$$

Fix $q \in (0, 1)$ and set $x_q = F^{-1}(q)$ and assume that F satisfies (9) at $u = x_q$. Let

$$N^{(n)}(q) = \sum_{i=1}^n \delta_{a_n^{-1}[Y_i - x_q]} I[Y_i \geq x_q] + \sum_{i=1}^n \delta_{b_n^{-1}[Y_i - x_q]} I[Y_i < x_q]. \quad (11)$$

Corollary 3.2 *Assume (9). Then $N^{(n)}(q) \xrightarrow{\mathcal{D}} N$, where N is a Poisson process on \mathbb{R} with intensity*

$$\lambda_t = \begin{cases} \alpha t^{\alpha-1} & \text{if } t > 0 \\ \beta |t|^{\beta-1} & \text{if } t < 0 \end{cases}.$$

Proof. The argument in the preceding proof can now be repeated for $t > 0$ and $t < 0$ to prove that $\Lambda_t^{(n)}$ converges in L_2 to t^α if $t > 0$ and to $|t|^\beta$ if $t < 0$. Let G be the distribution of $Y_i - x_q$. Thus, $G(s) = F(s + x_q)$. To see that the norming sequences a_n and b_n are chosen appropriately, in view of (3) applied to the distribution G of $Y_i - x_q$ (cf. also (4)), we have for $t > 0$,

$$\Lambda_t^{(n)} = \int_0^t \sum_{i=1}^n \mathbb{I}\{a_n^{-1}(Y_i - x_q) \geq u\} \frac{dG(a_n u)}{\bar{G}(a_n u)}.$$

Thus, using the same calculation as in the proof of Corollary 3.1,

$$\begin{aligned} \mathbb{E} [\Lambda_t^{(n)}] &= n \int_0^t dG(a_n u) = n[F(a_n t + x_q) - F(x_q)] \\ &= \frac{F(a_n t + x_q) - F(x_q)}{F(a_n + x_q) - F(x_q)} \rightarrow t^\alpha \end{aligned}$$

by the first part of (9). Moreover, bearing in mind Example 2.7, we have for $t < 0$,

$$\begin{aligned}\mathbb{E} \left[\Lambda_t^{(n)} \right] &= n \int_t^0 dG(b_n u) = -n[F(b_n t + x_q) - F(x_q)] \\ &= \frac{F(b_n t + x_q) - F(x_q)}{F(-b_n + x_q) - F(x_q)} \rightarrow |t|^\beta\end{aligned}$$

by the second part of (9). The result follows from Lemma 2.4. ◊

The power of the martingale method can be seen when one wants to obtain the asymptotic joint distribution of several $N^{(n)}(q)$. The key is an appropriate definition of the filtration.

Theorem 3.3 *Let $0 \leq q_1 < q_2 < \dots < q_k \leq 1$ and assume that (9) holds for each $x_{q_i}, i = 1, \dots, k$, with α_i and β_i , respectively. Then*

$$\langle N^{(n)}(q_1), N^{(n)}(q_2), \dots, N^{(n)}(q_k) \rangle \longrightarrow_{\mathcal{D}} \langle N(1), \dots, N(k) \rangle,$$

where $\langle N(1), \dots, N(k) \rangle$ is a k -variate Poisson process on \mathbb{R} with independent components and marginal intensities $\lambda(i)$, $i = 1, \dots, k$, given by

$$\lambda_t(i) = \begin{cases} \alpha t^{\alpha_i-1} & \text{if } t > 0 \\ \beta |t|^{\beta_i-1} & \text{if } t < 0 \end{cases}.$$

Proof: For clarity, we will consider only the case $k = 2$ and verify the conditions of Lemma 2.4. The general result follows in a straightforward manner.

We begin by observing that it suffices to show joint convergence of

$$\langle N^{(n)}(q_1), N^{(n)}(q_2) \rangle$$

for all $t \in [-K, K]$ for any arbitrary finite constant K . Assume that F is regularly varying on the right and left of x_{q_i} with index α_i and β_i , respectively, $i = 1, 2$. As before, define $a_n^{(i)}$ and $b_n^{(i)}$, $i = 1, 2$ according to (10). Choose M large enough that for $n \geq M$, $[x_{q_1} - Kb_n^{(1)}, x_{q_1} + Ka_n^{(1)}]$ and $[x_{q_2} - Kb_n^{(2)}, x_{q_2} + Ka_n^{(2)}]$ do not intersect. This will ensure that those points Y_j which are jump points of $N^{(n)}(q_1)$ are not jump points of $N^{(n)}(q_2)$ and vice versa.

Consider for $i = 1, 2$, $1 \leq j \leq n$ the single jump point process

$$J^{(n,j)}(q_i) = \delta_{(a_n^{(i)})^{-1}[Y_j - x_{q_i}]} I[Y_j \geq x_{q_i}] + \delta_{(b_n^{(i)})^{-1}[Y_j - x_{q_i}]} I[Y_j < x_{q_i}].$$

It is adapted to $\mathcal{F} = \{\mathcal{F}_{A_t} : -K \leq t \leq K\}$, where

$$\mathcal{F}_{A_t} = \begin{cases} \sigma\{\mathbb{I}_{\{Y_j \in [x_{q_i}, x_{q_i} + ta_n^{(i)}]\}}, i = 1, 2; j = 1, \dots, n\} & \text{if } t \geq 0 \\ \sigma\{\mathbb{I}_{\{Y_j \in [x_{q_i} + tb_n^{(i)}, x_{q_i}]\}}, i = 1, 2; j = 1, \dots, n\} & \text{if } t \leq 0. \end{cases}$$

We will compute a *-compensator $\bar{J}^{(n,j)}(q_i)$ of the single jump process $J^{(n,j)}(q_i)$. We consider only $0 < t < K$ as the argument for $t < 0$ is similar. Let

$$U_t = [x_{q_1} - Kb_n^{(1)}, x_{q_1} + ta_n^{(1)}] \cup [x_{q_2} - Kb_n^{(2)}, x_{q_2} + ta_n^{(2)}].$$

Then for $C = (t, t'] \in \mathcal{C}$, it follows that $\mathbb{I}_{\{Y_j \in U_t\}} \in \mathcal{G}_C^*$ and so heuristically, the compensator $\bar{J}^{(n,j)}(q_i)$ satisfies

$$\bar{J}_{dt}^{(n,j)}(q_i) = \frac{\mathbb{I}_{\{Y_j \in U_t^c\}} dF(x_{q_i} + a_n^{(i)}t)}{1 - F(U_t)},$$

provided that $[x_{q_1} - Kb_n^{(1)}, x_{q_1} + Ka_n^{(1)}]$ and $[x_{q_2} - Kb_n^{(2)}, x_{q_2} + Ka_n^{(2)}]$ are disjoint intervals. Using (3) it is straightforward to verify that for $n \geq M$ the *-compensator $\Lambda^{(n)}(i)$ of $N^{(n)}(q_i)$ is

$$\Lambda_t^{(n)}(i) = \sum_{j=1}^n \int_0^t \frac{\mathbb{I}_{\{Y_j \in U_s^c\}} dF(x_{q_i} + a_n^{(i)}s)}{1 - F(U_s)}. \quad (12)$$

Exactly as in the comments leading to Corollary 3.2 we have $\mathbb{E}[\Lambda_t^{(n)}(i)] \sim t^{\alpha_i}$ for the appropriate constant α_i , since F is slowly varying on the right at x_{q_i} .

The argument that $\mathbb{E}[(\Lambda_t^{(n)}(i))^2] \rightarrow (t^{\alpha_i})^2$ is also similar to that used in the proof of Corollary 3.1. Also, $N^{(n)}(q_i)$ is square integrable with bounded second moments (uniformly in n). Thus the conditions of Lemma 2.4 have been satisfied and the result follows. \odot

3.2 Multivariate case

Let $\{Y_n\}_{n \geq 1}$ be a sequence of i.i.d. \mathbb{R}^d -valued random variables with continuous distribution F . Following the pattern of the preceding section, we may obtain a point process limit if the regular variation index at \mathbf{u} for F depends on the choice of orthant. To be precise, let O_k be the k th orthant and \mathbf{e}_k its associated unit vector, $k = 1, \dots, 2^d$. Let $\mathbf{u}_j \in \mathbb{R}^d$, $j = 1, \dots, m$. Then F is regularly varying at \mathbf{u}_j from orthant O_k , with index $\alpha_{k,j}$ and rate $W_{k,j}$ if for $\mathbf{t} \in O_k$

$$\lim_{x \searrow 0} \frac{P_F((\mathbf{u}_j, x\mathbf{t} + \mathbf{u}_j])}{P_F((\mathbf{u}_j, x\mathbf{e}_k + \mathbf{u}_j])} = W_{k,j}(\mathbf{t}). \quad (13)$$

From the above definition it follows that the function $W_{k,j}$ is homogeneous of order $\alpha_{k,j}$, i.e. $W_{k,j}(s\mathbf{t}) = s^{\alpha_{k,j}} W_{k,j}(\mathbf{t})$, $\alpha_{k,j} > 0$, see e.g. Resnick (1987).

Let $a_{n,k,j}$ be such that $P_F((\mathbf{u}_j, a_{n,k,j}\mathbf{e}_k + \mathbf{u}_j]) = n^{-1}$. We define in analogy to (11)

$$N^{(n)}(j) := N^{(n)}(\mathbf{u}_j) = \sum_{k=1}^{2^d} \sum_{i=1}^n \delta_{a_{n,k,j}^{-1}[Y_i - \mathbf{u}_j]} I[Y_i \in O'_{k,j}], \quad (14)$$

where $O'_{k,j} = O_k + \mathbf{u}_j$.

Theorem 3.4 *Assume that the orthant-wise regular variation conditions (13) are satisfied at $\mathbf{u}_j, j = 1, \dots, m, \mathbf{x}_j \in \mathbb{R}^d$. For each j , let $N^{(n)}(j)$ denote the \mathbb{R}^d -indexed point process of (14). Then*

$$\langle N^{(n)}(1), N^{(n)}(2), \dots, N^{(n)}(m) \rangle \longrightarrow_{\mathcal{D}} \langle N(1), N(2), \dots, N(m) \rangle$$

where $\langle N(1), N(2), \dots, N(m) \rangle$ is a vector of independent Poisson processes, where the j th component process is parameterized by \mathbb{R}^d and its mean measure is given orthant-wise by $W_{k,j}(\cdot)$.

Proof. Once again, the proof is carried out by showing that the relevant $*$ -compensators converge in L_2 to the appropriate deterministic limit. For ease of exposition, assume that $d = 2$. Consider the case $m = 1$ and assume without loss of generality that $\mathbf{u}_1 = \mathbf{0}$. Then we have $k = 4$ orthants. Consider, for $k = 1$, the orthant \mathbb{R}_+^2 and its associated unit vector $\mathbf{e}_1 = (1, 1) =: \mathbf{1}$. Write for simplicity $a_{n,1,1} = a_n$. Then, the $*$ -compensator is given by (cf. Example 2.8)

$$\int_{A_t} \sum_{i=1}^n \mathbb{I}\{a_n^{-1}Y_{i1} \geq v_1, a_n^{-1}Y_{i2} \geq v_2\} \frac{dF(a_n \mathbf{v})}{F(a_n \mathbf{v})}.$$

Taking expected value one obtains

$$nF(A_t) = \frac{P_F((0, a_n \mathbf{t}])}{P_F((0, a_n \mathbf{1}])} \rightarrow W_{1,1}(\mathbf{t}).$$

The L^2 convergence is treated in the same way as in Corollary 3.1.

The other 3 orthants are treated similarly. The extension from $m = 1$ to arbitrary m follows along the lines of Theorem 3.3.

◻

Examples of regularly varying distributions are readily constructed. One source of examples are distributions based on copulas as described, for example, in Nelsen (1999) and Section 4.3 below.

4 Applications

4.1 Local empirical processes

In the introduction we pointed out that one of our motivations was to obtain a weak limit for local empirical processes, as defined in (1). As a particular application of Theorem 3.3 we obtain the following result.

Let $0 < q_1 < q_2 < \dots < q_k < 1$ and $x_j = F^{-1}(q_j)$, $j = 1, \dots, k$ where F is a distribution on \mathbb{R} . Assume that

$$\lim_{x \searrow 0} \frac{P_F((x_j, x_j + xt])}{P_F((x_j, x_j + x])} = t^\alpha \quad \text{and} \quad \lim_{x \searrow 0} \frac{P_F((x_j - xt, x_j])}{P_F((x_j - x, x_j])} = t^\alpha,$$

for all $t > 0$. Since the regular variation behaviour is the same on either side of x_j , the scaling constant $a_n = a_{n,j}$ may be chosen so that $F(a_n + x_j) - F(x_j) =$

n^{-1} for arbitrary $j = 1, \dots, k$. Thus (cf. (11)), the point processes are defined as $N^{(n)}(j) = \sum_{i=1}^n \delta_{a_n^{-1}[Y_i - x_j]}$, $j = 1, \dots, k$. Note further, that $N_A^{(n)}(j)$ with $A = [-t, t]$ (cf. (2)) agrees with $L_{n,x_j}(t)$ (cf. (1)). Consequently, we obtain the following result.

Corollary 4.1 *The vector $(L_{n,x_1}(\cdot), \dots, L_{n,x_k}(\cdot))$ of the local empirical processes converges weakly to the vector of independent Poisson processes with the same intensities $2\alpha|t|^{\alpha-1}$.*

4.2 Local Density Estimation

Consider a sample $\{Y_1, Y_2, \dots, Y_n\}$ with common marginal differentiable distribution F on $[0, 1]$, and assume that its density f is positive on the range $[0, 1]$. Let F_n denote the empirical distribution and define $[t]_n^+ = Y_{(k+1)}$ and $[t]_n^- = Y_{(k)}$ if $Y_{(k)} \leq t < Y_{(k+1)}$. We put $[t]_n^- = 0$ if $t < Y_{(1)}$ and $[t]_n^+ = 1$ if $t > Y_{(n)}$. A naive nearest-neighbour estimator of the density at t is given by

$$\hat{f}(n, t) = \frac{1}{n([t]_n^+ - [t]_n^-)}. \quad (15)$$

Additional information on nearest-neighbour density estimates can be found in Härdle (1990) or Silverman (1992), including comments on performance, and modifications.

For $t \in (0, 1)$ the fact that F is differentiable and that f is positive (i.e. F is of regular variation index $\alpha = 1$ at t) allows us to write

$$\hat{f}(n, t)/f(t) = \frac{1}{f(t)([t]_n^+ - t) + n(t - [t]_n^-)} \xrightarrow{\mathcal{D}} \frac{1}{f(t)(E_1 + E_2)} \quad (16)$$

where E_1 and E_2 are independent exponential variables of mean $1/f(t)$. This convergence follows from Corollary 3.2 and the continuous mapping theorem, and follows the pattern set for extreme value processes as given in Resnick (1987). As each limiting Poisson process has a constant rate function equal to $f(t)$, the distance from t to the first point to the left or right has an exponential distribution with mean $1/f(t)$. Since such an exponential variable can be written as the product of $1/f(t)$ and an exponential of mean 1, and the sum of two independent mean 1 exponentials is a $\Gamma(2, 1)$ variable, we have identified the limiting distribution of $\hat{f}(n, t)/f(t)$ as Inverse Gamma, $\Gamma^{(-1)}(2, 1)$. The mode, mean and variance of an Inverse Gamma density of parameters (α, β) are $\beta/(\alpha + 1)$, $\beta/(\alpha - 1)$ (for $\alpha > 1$) and $\beta^2/((\alpha - 1)(\alpha - 2))$ (for $\alpha > 2$), respectively. Thus we see that this naive estimator of $f(t)$ has mode $f(t)/3$, mean $f(t)$ and infinite variance.

This development can be easily extended to estimators based on the k lower nearest neighbours and k upper nearest neighbours. As above, asymptotically the spacings between consecutive neighbours are independent exponential variables with mean $1/f(t)$. The asymptotic joint density is the product of $2k$ exponentials, and the sufficient statistic is just the total distance from the lower k th-nearest neighbour of t , $[t]_n^{-k}$, to the upper k th-nearest neighbour, $[t]_n^{+k}$.

Corollary 4.2 *The asymptotically uniformly minimum variance unbiased estimator based on k nearest neighbours ($k > 1$) is*

$$\hat{f}_k(n, t) = \frac{(2k-1)/n}{[t]_n^{+k} - [t]_n^{-k}},$$

and $\hat{f}_k(n, t)/f(t)$ has an asymptotic $\Gamma^{(-1)}(2k, 1)$ density. Moreover, the limiting distribution of

$$\left\langle \frac{\hat{f}_k(n, t_1)}{f(t_1)}, \frac{\hat{f}_k(n, t_2)}{f(t_2)}, \dots, \frac{\hat{f}_k(n, t_m)}{f(t_m)} \right\rangle$$

is given by a vector of m independent $\Gamma^{(-1)}(2k, 1)$ Inverse Gamma variables.

Consequently we can obtain the limiting distribution of expressions such as approximate integrals,

$$\hat{E}(g(Y)) = \sum_{i=1}^m g(t_i)(\hat{f}_k(n, t_i)),$$

even for arbitrary dimension (Theorem 3.4) with appropriate norming.

Remark 4.3 On the other hand, we see that $\hat{f}_k(n, t)/f(t)$ still has an Inverse Gamma distribution, but with finite variance for $k \geq 1$. It has asymptotic variance $1 + 1/(2k - 2)$, and so remains inherently random regardless of the fixed number of nearest neighbours used in the estimate. Nearest-neighbour methods have become popular in data mining, classification and computing applications, and rapid algorithms exist for finding the k nearest neighbours to a point t even in high dimensions. The above discussion shows that even in highly regular cases, the best k -nearest-neighbour density estimate will not converge in probability to the desired limit, and remains random.

4.3 Multivariate extremes

Let $\{(Y_{n1}, Y_{n2})\}_{n \geq 1}$ be an i.i.d. sequence of bivariate random vectors. To focus on the bivariate dependence structure rather than the marginal distributions, we assume that (Y_{11}, Y_{12}) has a copula C and standard uniform marginals (cf Nelsen (1999)). We want to characterize

$$\mathbb{P}(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2)$$

as $x \searrow 0$. If Y_{11} and Y_{12} are independent, then the above probability factors and we can apply standard extreme value methods (e.g. Resnick, 1987) to the marginals. However, if Y_{11} and Y_{12} are dependent but the maxima are asymptotically independent then the extreme value methods fail; see Fougères (2004) for a general discussion of this problem. For most known families of copulas which have the asymptotic independence property of maxima, we have (cf. Heffernan, 2000)

$$\mathbb{P}(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2) \sim cx^2. \quad (17)$$

Note that the above definition means that the distribution function of (Y_{11}, Y_{12}) is regularly varying at $(1, 1)$ from the orthant O_k , where O_k is associated with the unit vector $(-1, -1)$, see (13). By Theorem 3.4, the appropriate scaling to obtain a point process limit for the joint extremes is $a_n = n^{-1/2}$, and not the $a_n = n^{-1}$ that would be used to normalize the marginal variables individually. Note, moreover, that the methods of this paper are “dimension free”, and so we can address multivariate copulas of any dimension.

Further we can address the joint extreme value behaviour of copulas with the asymptotic independence property but where (17) is not satisfied. As an example, consider the case when C is the bivariate normal copula with correlation $\rho \in (0, 1]$ – i.e. $C(x, y)$ is given by a joint normal distribution function at $(\Phi^{-1}(x), \Phi^{-1}(y))$ with standard marginals and correlation ρ . We have

$$\mathbb{P}(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2) \sim x^{2/(1+\rho)} g(t_1, t_2)$$

for a function g as $x \searrow 0$, and so

$$\lim_{x \searrow 0} \frac{\mathbb{P}(Y_{11} > 1 - xt_1, Y_{12} > 1 - xt_2)}{\mathbb{P}(Y_{11} > 1 - x, Y_{12} > 1 - x)} = \frac{g(t_1, t_2)}{g(1, 1)},$$

for $t_1, t_2 \geq 0$. For $\mathbf{u} = (1, 1)$ and $t_1, t_2 \leq 0$, formula (13) is satisfied with

$$W(t_1, t_2) = \frac{g(-t_1, -t_2)}{g(1, 1)}.$$

Applying the results of Section 3.2, with $\mathbf{u} = (1, 1)$, $m = 1$, the third orthant \mathbb{R}_-^2 , its associated unit vector $\mathbf{e}_3 = (-1, -1)$, we define $a_n = a_{n,3,1}$ to be such that

$$P(Y_{11} > 1 - a_n, Y_{12} > 1 - a_n) = n^{-1},$$

so that $a_n = n^{-(1+\rho)/2} \ell(n)$, where $\ell(n)$ is slowly varying at infinity.

Corollary 4.4 *Assume that $\{\mathbf{Y}_n = (Y_{n1}, Y_{n2})\}_{n \geq 1}$ are i.i.d. with a common normal copula of parameter ρ and uniform marginals. Then for $\mathbf{u} = (1, 1)$, the process*

$$N^{(n)} = \sum_{i=1}^n \delta_{a_n^{-1}[\mathbf{Y}_i - \mathbf{u}]}$$

converges to a Poisson process on \mathbb{R}_-^2 with mean measure $W(\cdot, \cdot)$.

Acknowledgement.

The research of the first two authors was supported by grants from the Natural Sciences and Engineering Research Council of Canada. This research was conducted while Rafał Kulik was a postdoctoral fellow at the University of Ottawa. Unfortunately, our good friend and colleague, André Dabrowski, passed away before this article was completed.

References

- [1] Borisov, I. S. (2000). A note on Poisson approximation of rescaled set-indexed empirical processes. *Statist. Probab. Lett.* **46**, 101–103.
- [2] Borisov, I. S. (2003). A remark on a theorem of R. L. Dobrushin, and couplings in the Poisson approximation in abelian groups. (Russian) *Teor. Veroyatnost. i Primenen.* **48**, 576–583; translation in *Theory Probab. Appl.* **48** (2004), 521–528.
- [3] Brémaud, P. (1981). *Point Processes and Queues* Springer-Verlag, Berlin.
- [4] Deheuvels, P., Einmahl, U., and Mason, D.M. (2000). Asymptotic Independence of the Local Empirical Processes Indexed by Functions. *High Dimensional Probability II*, Giné, E., Mason, D.M. and Wellner, J.A., Eds. Birkhäuser.
- [5] Einmahl, J. H. J. (1997). Poisson and Gaussian approximation of weighted local empirical processes. *Stochastic Process. Appl.* **70**, 31–58.
- [6] Fougères, A.-L. (2004). Multivariate extremes. In: *Extreme Values in Finance, Telecommunications and the Environment.*, 373–388. Eds. B. Finkenstädt and H. Rootzén.
- [7] Härdle, W. (1990). *Applied Nonparametric Regression*. Econometric Society Monographs, Cambridge University Press, Cambridge.
- [8] Heffernan, J.E. (2000). A Directory of Coefficients of Tail Dependence. *Extremes* **3:3**, 279–290.
- [9] Ivanoff, B.G. and Merzbach, E. (1997). Poisson convergence for set-indexed empirical point processes. *Statist. Probab. Lett.* **32**, 81–86.
- [10] Ivanoff, B.G. and Merzbach, E. (2000) *Set-Indexed Martingales*. Chapman & Hall/CRC Press, Boca Raton.
- [11] Nelsen, R. B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics, 139. Springer-Verlag, New York.
- [12] Resnick, S. (1987). *Regular Variation, Extremes and Point Processes*. Springer.
- [13] Silverman, B. W. (1992). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability **26**, Chapman and Hall, London.