Compound Poisson approximation
for the distribution of extremes
A.D.Barbour, S.Y.Novak and A.Xia

Compound Poisson approximation for the distribution of extremes

A.D.Barbour
Angewandte Mathematik
Universität Zürich
Winterthurerstrasse 190
8057 Zürich, Switzerland

S.Y.Novak
EURANDOM
P.O. Box 513, Eindhoven 5600 MB, The Netherlands

A.Xia*
Department of Statistics, School of Mathematics
The University of New South Wales
Sydney 2052, Australia
e-mail: xia@maths.unsw.edu.au

Abstract

We evaluate accuracy of compound Poisson approximation for the distribution of empirical point processes of extremes. The accuracy is assessed in terms of a Wasserstein metric, which is generally more suitable for the purpose than the total variation metric. The argument uses Bernstein blocks and Lindeberg's method of compositions.

Key words and phrases: extreme values, point processes, compound Poisson process, total variation metric, coupling.

AMS 1991 Subject Classification: Primary 60G70; secondary 60F05.

1 Introduction

Let $X, X_1, X_2, \ldots$ be a strictly stationary sequence of (dependent) random variables. We say that X_i is an *extreme value* if $X_i > u$, where $u \equiv u_n$ is "close" to the right end point of the distribution of the random variable X.

Extreme value theory has important applications to insurance and finance (when the k-th largest element $X_{n,k}$ of the sample $X_1, \ldots, X_n$ represents the k-th largest claim or the k-th largest gain (loss) of a stock in a certain period of time), in flood prediction and prevention in hydrology, and in network modelling, meteorology, etc. (see Embrechts et al. (1997) and references therein). The basic information about extremes in the sample is collected in the number $M_{n,u}$ of exceedances above the level u among the random variables $X_1, \ldots, X_n$:

$$M_{n,u} = \sum_{i=1}^{n} 1_{\{X_i > u\}} . \tag{1.1}$$

The random variables $X_{n,k}$ and $M_{n,u}$ are closely related, since the events $\{X_{n,k} \leq u\} = \{M_{n,u} < k\}$.

If one is interested in more information about the joint distribution of the large values $X_{n,k}$, then processes of exceedances must be introduced. A one–dimensional point process $N_{n,u}$ marks the indices where high level exceedances occur:

$$N_{n,u}(B) = \sum_{i=1}^{n} 1_{\{i/n \in B, X_i > u\}}, \tag{1.2}$$

for any Borel set $B \subset [0,1]$. A two–dimensional point process $\Xi_{n,f}$ contains in addition the information about the heights of exceedances:

$$\Xi_{n,f}(A) = \sum_{i=1}^{n} 1_{\{(i/n, f^{-1}(X_i)) \in A\}}, \tag{1.3}$$

for any Borel set $A \subset [0,1] \times [0, \infty)$, where f is a strictly decreasing function from $\mathbf{R}_+ = [0, \infty)$ to $\mathbf{R}$, and interest is mainly concentrated on $[0,1] \times [0, f^{-1}(u))$.

The limiting behaviour of extremes under various asymptotic régimes has been well studied, and the books by Leadbetter *et al.* (1983) and Embrechts *et al.* (1997) give good accounts of the theory. Results particularly relevant to this paper are those of Hsing *et al.* (1988), who showed that the only possible limit laws for $N_{n,u}$ are compound Poisson distributions, and of Novak (1998), who established necessary and sufficient conditions for the weak convergence of $\Xi_{n,f}$ to a compound Poisson point process.

In this paper, we move away from the idea of a limit, and instead consider finite samples: we investigate the distance between the distributions of the empirical processes of exceedances $N_{n,u}$ and $\Xi_{n,f}$ from natural compound Poisson approximations, for any fixed choices of n, u and f; the approximation of $M_{n,u}$ in this way was addressed in Novak (1998). Even in the case where sequences indexed by n are considered, and $u = u_n$ and $f = f_n$ are chosen to ensure non–trivial limiting behaviour, the distance between the empirical processes

2

and their corresponding limit laws is still important, since a limit theorem is useless if the distance from the limit is not "small".

Novak (1998) evaluated the total variation distance between $\mathcal{L}(M_{n,u})$ and an appropriate compound Poisson distribution, where the total variation distance between the distributions of random elements $\mathbf{P}$ and $\mathbf{Q}$ with common domain $\mathcal{B}$ is defined by

$$d_{TV}(\mathbf{P}, \mathbf{Q}) = \sup_{B \in \mathcal{B}} |\mathbf{P}(B) - \mathbf{Q}(B)|.$$

However, the points of the process $N_{n,u}$ are concentrated on a subset $\mathcal{R}$ of rational numbers in $[0, 1]$, whereas the points of any distributional limit N_∞ hit $\mathcal{R}$ with probability 0. This makes the total variation distance unsuitable for measuring the accuracy of the approximation $N_{n,u} \approx N_\infty$, since it would always be the case that $d_{TV}(\mathcal{L}(N_{n,u}), \mathcal{L}(N_\infty))$ took the value 1. Hence we need weaker metrics for measuring the distance between the distributions of point processes.

Let Γ be an interval $[0, a]$, and let $\mathcal{G}$ be the σ–field generated by open sets in Γ. Define

$$\mathcal{H} = \left\{ \sum_{i=1}^{n} c_i \delta_{t_i}(\cdot) : \ t_1, \cdots, t_n \in \Gamma, \ n \geq 1 \right\},$$

where $c_i \in \mathbf{N}$ and $\delta_x(\cdot)$ is the Dirac measure at x: $\delta_x(B) = 1_B(x)$. Then $\mathcal{H}$ is the space of finite, non–negative integer valued counting measures on $(\Gamma, \mathcal{G})$, and a realizaton of a point process on Γ is just an element ξ of $\mathcal{H}$: for $a = 1$, $N_{n,u}$ is an example. The interpretation is that $\xi = \sum_{i=1}^{n} c_i \delta_{t_i}$ consists of a configuration of a total of $k = \sum_{i=1}^{n} c_i$ points, with c_i points located at t_i, $1 \leq i \leq n$; we denote the list of points of ξ by $\tilde{\xi} = \{\tilde{t}_j, 1 \leq j \leq k\}$, where each of the t_i appears c_i times in the list $\tilde{\xi}$. More generally, one can take Γ to be a compact metric space with a metric d_0, and define $\mathcal{H}$ to be the family of all finite, non–negative integer valued counting measures on $(\Gamma, \mathcal{G})$. In Section 3, where we study the distribution of the process $\Xi_{n,f}$, Γ is a rectangle.

Now let X and Y be random elements of $\mathcal{H}$, and set $\mathbf{Q}_X = \mathcal{L}(X)$ and $\mathbf{Q}_Y = \mathcal{L}(Y)$. We use the class of Wasserstein metrics to measure the distance between the probability distributions $\mathbf{Q}_X$ and $\mathbf{Q}_Y$. These metrics are defined by setting

$$d^*(\mathbf{Q}_X, \mathbf{Q}_Y) \equiv d^*(\mathbf{Q}_X, \mathbf{Q}_Y | d) = \inf_{(X', Y')} \mathbb{E}\, d(X', Y'), \tag{1.4}$$

where d is any distance between elements in $\mathcal{H}$, and the infimum ranges over all pairs (X', Y') with values in $\mathcal{H}^2$ and marginal distributions $\mathcal{L}(X') = \mathbf{Q}_X$ and $\mathcal{L}(Y') = \mathbf{Q}_Y$. This leaves great freedom of choice, since the distance d has still to be chosen; we restrict ourselves to those of the form

$$d_1(\xi, \eta) = \begin{cases} 1, & \text{if } \xi(\Gamma) \neq \eta(\Gamma), \\ k^{-1} \min_\tau \sum_{l=1}^{k} d_0(\tilde{y}_l, \tilde{z}_{\tau(l)}), & \text{if } \xi(\Gamma) = \eta(\Gamma) = k > 0, \\ 0, & \text{if } \xi(\Gamma) = \eta(\Gamma) = 0, \end{cases}$$

where $\xi = \sum_{i=1}^{m'} c_i' \delta_{y_i}(\cdot)$ and $\eta = \sum_{i=1}^{m''} c_i'' \delta_{z_i}(\cdot)$ are elements of $\mathcal{H}$, d_0 is a metric on Γ, and the minimum is taken over all possible permutations τ of $(1, 2, ..., k)$. This distance minimises

3

the average d_0-distance between pairs of points, with respect to the choice of matching. Thus two configurations, one obtained from the other by a small shift in Γ, are at small d_1 distance one from another, whereas, with respect to total variation distance, they would be far apart. The metric d^* on $\mathcal{H}$ derived according to (1.4) from $d = d_1$ we refer to as d_2.

There remains the choice of metric d_0 on Γ, which itself needs to be carefully considered. It is most natural to use choices of d_0 which are based on Euclidean distance, but are also scale invariant, in the sense that expressing the locations of all points in new units should not change the distance of a configuration from a reference configuration; for reasons of robustness, we also require that $d_0(x, y) \le 1$. Scale invariance is achieved by implicitly taking typical configurations from the approximating compound Poisson process as references, and requiring that d_0 be chosen so that this process has unit intensity. Thus, if Γ is an interval $[0, a]$, we can take

$$d_0(x_1, x_2) = \min\{c|x_1 - x_2|, 1\},$$

where c is the intensity of the reference process with respect to Lebesgue measure (the average intensity, if c were not constant; but here we only consider stationary processes). In Section 2, we prefer to achieve this by scaling the point process $N_{n,u}$ to have unit intensity, so that then $c = 1$. For Γ a rectangle in $\mathbf{R}^2$, we take

$$d_0((x_1, y_1), (x_2, y_2)) = \min\{(c_1|x_1 - x_2| + c_2|y_1 - y_2|), 1\},$$

where $c_1 c_2$ is the (average) intensity of the reference process with respect to two dimensional Lebesgue measure, and the ratio c_1/c_2 can be chosen to reflect the relative importance of discrepancies in the x and y directions.

Although the Wasserstein metric d_2 is rather weaker than the total variation metric, a small value of $d_2(\mathbf{Q}_1, \mathbf{Q}_2)$ still implies that the $\mathbf{Q}_1$ and $\mathbf{Q}_2$ distributions of many functionals of the random measures are close to one another. One such functional is the total number of points; another, more sophisticated functional is the empirical distribution function of the inter-point distances. As a further example, suppose that the function g is bounded and Lipschitz on Γ. Then the functional

$$h(\xi) = \begin{cases} \|g\| & \text{if } \xi(\Gamma) = 0; \\ \int_\Gamma g(x)\xi(dx)\big/\xi(\Gamma), & \text{if } \xi(\Gamma) > 0, \end{cases}$$

is d_1-Lipschitz with constant $\max\{2\|g\|, \|g'\|\}$, so that, for random elements X, Y of $\mathcal{H}$,

$$|\mathbb{E}h(X) - \mathbb{E}h(Y)| \le \max\{2\|g\|, \|g'\|\}d_2(\mathcal{L}(X), \mathcal{L}(Y)).$$

Bounded functions of the pair $(h(X), X(\Gamma))$ which are Lipschitz in the first coordinate can also be considered. Thus the d_2 metric provides a useful measure of the rate of convergence; as is shown below, it is effective in the current situation, whereas total variation distance is not.

Our approximating distributions we define as follows. For a fixed interval $[0, a]$, we use $\mathbf{CP}(\lambda\mu, \nu)$ to denote the distribution of the compound Poisson process

$$\sum_{s:\, Y_s \in [0,a]} Z_s \delta_{Y_s} \tag{1.5}$$

4

with intensity measure $\lambda\mu$ and multiplicity distribution ν. Here μ denotes Lebesgue measure, λ is the intensity coefficient, $\{Z, Z_1, Z_2, ...\}$ are independent random variables (independent of the sequence $\{Y_s\}$) with $\mathcal{L}(Z_s) = \nu$, and $\{Y_s\}$ are the points of a Poisson process on $\mathbf{R}_+$ with intensity measure $\lambda\mu$. The mean measure of $\mathbf{CP}(\lambda\mu, \nu)$ is then given by $\lambda\mu\mathbb{E}Z$. Expression (1.5) can equivalently be written as $\sum_{s=1}^{\pi} Z_s\delta_{Y_s}$, where the random variable $\pi = \#\{s : 0 \le Y_s \le a\}$ has the Poisson $Po(a\lambda)$ distribution. Such processes, with continuous intensity measures, are the natural approximations for the stationary processes $N_{n,u}$.

For $\Xi_{n,f}$, we exchange Z_s for a finite random measure in $\mathbf{R}_+$, which is used to approximate not only the number of exceedances at indices i such that i/n is near a point $t \in [0, a]$, but also the (extreme) values $f^{-1}(X_i)$ taken there. We use the notation $PC(\lambda\mu, \nu^*)$ to denote the corresponding Poisson cluster process, having intensity measure $\lambda\mu$ on $[0, a]$ for the occurrence of clusters, and probability measure (multiplicity distribution) ν^* over the family of finite point measures in $\mathbf{R}_+$, which describes the distribution of the clusters. The two types of processes are linked, inasmuch as the measure ν^* induces the distribution ν of the number of points in a cluster:

$$\nu\{j\} := \nu^*\{\eta : \eta\{[0, f^{-1}(u))\} = j\}.$$

Hence the one–dimensional compound Poisson process $\mathbf{CP}(\lambda\mu, \nu)$ can be derived as a summary of the process $PC(\lambda\mu, \nu^*)$. The Poisson cluster process is a compound Poisson process on $[0, a] \times \mathbf{R}_+$ when the probability measure ν^* is concentrated on the set of point measures which consist of a single atom, and is a Poisson process when these atoms are restricted to having mass 1.

In the next two sections, we provide explicit bounds for the accuracy of compound Poisson approximation to the point processes $N_{n,u}$ and $\Xi_{n,f}$, in terms of the Wasserstein metric d_2.

2 Compound Poisson approximation to $N_{n,u}$

The main result of the section, Theorem 2.1, bounds the d_2–distance between the distribution of $N_{n,u}$ and a compound Poisson process, whose mean measure is proportional to Lebesgue measure μ. This latter stipulation is natural, in view of stationarity. In order to formulate the theorem, it is necessary to decide on the carrier space and on the metric d_0. The standard approach is to choose $\Gamma = [0, 1]$ and $d_0(x, y) = |x - y|$, the Euclidean distance, but, as discussed in the previous section, this is not scale invariant. Our choice is therefore to modify the definition of $N_{n,u}$, retaining the usual Euclidean scale, and defining

$$d_0(x, y) := \min\{|x - y|, 1\}, \quad x, y \in \mathbf{R}. \tag{2.1}$$

Thus, suppressing the indices n and u, we write

$$p = \mathbb{P}(X > u),$$

and set

$$\widetilde{N}(B) = \sum_{i=1}^{n} 1_{\{ip \in B, X_i > u\}} = \sum_{i=1}^{n} 1_{\{X_i > u\}}\delta_{ip}(B) \tag{2.2}$$

5

for any Borel set $B \subset \Gamma := [0, np]$.

In order to state the main theorem, and as a principal tool in the proof, we shall need the classical Bernstein's "blocks": see also Hsing *et al.* (1988). Fix any $r \in \mathbf{Z}_+$, and divide $\{1, ..., n\}$ into blocks of length r by putting

$$B_r(i) = \{(i-1)r + 1, ..., (ir) \wedge n\}.$$

Define

$$T_{ri} = \sum_{j \in B_r(i)} 1_{\{X_j > u\}} \qquad 1 \leq i \leq \lceil n/r \rceil,$$

and let $\widehat{T}_{ri}$, $1 \leq i \leq \lfloor n/r \rfloor$ be independent copies of T_{r1}, noting that the T_{ri} are also identically distributed for $1 \leq i \leq \lfloor n/r \rfloor$; the notation $\lfloor x \rfloor$ denotes the greatest integer $m \leq x$, $\lceil x \rceil$ the least integer $m \geq x$. Then let ν_r denote the conditional distribution of T_{r1} given that $T_{r1} \geq 1$, and put

$$\nu_r[a, b) = \nu_r\{[a, b)\}, \quad q_r = \mathbb{P}(T_{r1} \geq 1), \quad \theta_r = q_r/rp.$$

Note that q_r and θ_r are functions of the level u, and that $q_r \leq rp \wedge 1$. If the limit

$$\theta = \lim_{n \to \infty} \theta_{r_n},$$

exists for any sequences u_n and r_n such that np is bounded away from 0 and ∞ and $1 \ll r_n \ll n$, then it is called the *extremal index* of the sequence $\{X_i, i \geq 1\}$ (O'Brien (1974), Leadbetter *et al.* (1983, Chapter 3.7), Novak (1996)). The compound Poisson process $\mathbf{CP}(\theta_r \mu, \nu_r)$ on $[0, np]$ is the approximation that we use for the distribution of $\widetilde{N}$.

In the proof, the blocks are used essentially to show that the joint distribution of the T_{ri}, $1 \leq i \leq \lfloor n/r \rfloor$, is close to that of $\widehat{T}_{ri}$, $1 \leq i \leq \lfloor n/r \rfloor$, under suitable mixing conditions. We consider two such. Let $\mathcal{F}_{m,s}$ be the σ–algebra generated by the events $\{X_i > u\}$, $m \leq i \leq s$. Set

$$
\begin{aligned}
\varphi(l) &= \max_{1 \leq m \leq n} \sup_{A \in \mathcal{F}_{1,m}, B \in \mathcal{F}_{m+l,n}} |\mathbb{P}(B|A) - \mathbb{P}(B)|, \\
\alpha(l) &= \max_{1 \leq m \leq n} \sup_{A \in \mathcal{F}_{1,m}, B \in \mathcal{F}_{m+l,n}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.
\end{aligned}
$$

Then set

$$\varepsilon^* = 4rp + 2nr^{-1}lp + \varepsilon + \left(\frac{1.65}{\sqrt{1 - q_r}} + e^{q_r}\right)q_r,$$

where

$$\varepsilon \equiv \varepsilon(l, r, M) = \min\left\{nr^{-1}\varphi(l); \, Mnr^{-1}\alpha(l) + 2nr^{-1}\mathbb{P}(T_{r1} \geq M)\right\}.$$

Note that $\mathbb{P}(T_{r1} \geq M) = q_r \nu_r[M, \infty)$, and that

$$r^{-1}\mathbb{P}(T_{r1} \geq M) \leq r^{-1}\mathbb{E}T_{r1}/M = p/M. \qquad (2.3)$$

A better estimate is valid under additional assumptions. For instance, if $c_\varphi = \sum_i \varphi^{1/2}(2^i) < \infty$ and $M > rp$, then, using Utev's (1989) result, we get

$$\mathbb{P}(T_{r1} \geq M) \leq C_t(rp)^t(M - rp)^{-2t},$$

where the constant C_t depends only on c_φ and t.

We are now in a position to state our main theorem.

Theorem 2.1 *For any choices of $M \geq 1$ and l, r such that $1 \leq l < r \leq n$, we have*

$$d_2\left(\mathcal{L}\left(\widetilde{N}\right), \mathbf{CP}(\theta_r \mu, \nu_r)\right) \leq \varepsilon^*, \tag{2.4}$$

where both processes are restricted to the interval $[0, np]$.

In order to establish the theorem, we first prove the following lemma, which quantifies the approximate independence of the block processes. Write

$$D_r(n) = d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri}\right)\right).$$

Lemma 2.2 *For any choices $1 \leq l \leq r \leq n$ and $M \geq 1$, we have*

$$D_r(n) \leq 2nr^{-1}lp + \varepsilon(l, r, M). \tag{2.5}$$

Proof of Lemma 2.2. First of all, let $T_{ri}^{(l)} = \sum_{j \in B_r^{(l)}(i)} 1_{\{X_j > u\}}$, where $B_r^{(l)}(i) := \{(i-1)r + 1, ..., ir - l\}$ for $1 \leq i \leq \lfloor n/r \rfloor$, so that $B_r^{(l)}(i)$ is obtained by deleting a sub-block of length l at the right end of block $B_r(i)$, and $B_r^{(l)}(\lfloor n/r \rfloor + 1) = B_r(\lfloor n/r \rfloor + 1)$. Then, since

$$\mathbb{P}\left(\sum_{j \in B_r(i) \setminus B_r^{(l)}(i)} 1_{\{X_j > u\}} \neq 0\right) \leq lp,$$

it follows that

$$d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}^{(l)}\right)\right) \leq \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri} \neq \sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}^{(l)}\right) \leq \frac{n}{r} lp. \tag{2.6}$$

Let $(\widehat{T}_{r,i}, \widehat{T}_{r,i}^{(l)})$ be independent copies of $(T_{ri}, T_{ri}^{(l)})$, which are also independent of $X_1, ..., X_n$. Similarly to (2.6),

$$d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri}\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri}^{(l)}\right)\right) \leq nr^{-1}lp. \tag{2.7}$$

By Lindeberg's (1922) method of compositions (cf. Novak (1998) and Eberlein (1984)), we have

$$\mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}^{(l)} \in A\right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri}^{(l)} \in A\right) = \sum_{j=1}^{\lfloor n/r \rfloor} \Delta_j(A), \tag{2.8}$$

7

for any $A \subset \mathcal{H}$, where

$$\Delta_j(A) = \mathbb{P}\left(\sum_{i=1}^{j}\delta_{irp}T_{ri}^{(l)} + \sum_{i=j+1}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A\right) - \mathbb{P}\left(\sum_{i=1}^{j-1}\delta_{irp}T_{ri}^{(l)} + \sum_{i=j}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A\right). \quad (2.9)$$

It remains to estimate the individual terms in (2.8), using the mixing coefficients φ and α.

The atoms of the measure $\sum_{i=1}^{j-1}\delta_{irp}T_{ri}^{(l)}$ are at points in the set

$$\Gamma^* = \{rp, 2rp, ..., (\lfloor n/r \rfloor - 1)rp\}.$$

Thus the corresponding space $\mathcal{H} = \mathcal{H}_r$ is countable, so that we may write $\mathcal{H}_r = \{h_i, i \geq 1\}$. Then

$$
\begin{aligned}
|\Delta_j(A)| &\leq \sum_{m \geq 1}\mathbb{P}\left(\sum_{i=1}^{j-1}\delta_{irp}T_{ri}^{(l)} = h_m\right) \\
&\quad \times \left|\mathbb{P}\left(h_m + \delta_{jrp}T_{rj}^{(l)} + \sum_{i=j+1}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A \;\middle|\; \sum_{i=1}^{j-1}\delta_{irp}T_{ri}^{(l)} = h_m\right)\right. \\
&\qquad \left. - \mathbb{P}\left(h_m + \sum_{i=j}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A\right)\right| \\
&\leq \sum_{m \geq 1}\mathbb{P}\left(\sum_{i=1}^{j-1}\delta_{irp}T_{ri}^{(l)} = h_m\right)\varphi(l) = \varphi(l).
\end{aligned}
$$

Substituting this estimate into (2.8), we get

$$\left|\mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor}\delta_{irp}T_{ri}^{(l)} \in A\right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A\right)\right| \leq nr^{-1}\varphi(l). \quad (2.10)$$

Alternatively, for any set $C \subset \mathbf{Z}_+$, define

$$
\begin{aligned}
H_{st}(C) &= \left\{\sum_{i=s}^{t}\delta_{irp}m_i : (m_s, m_{s+1}, \ldots, m_t) \in C^{t-s+1}\right\}, \\
H_{st} &= H_{st}(\mathbf{Z}_+), \quad H_{st}^M = H_{st}([0, M-1]).
\end{aligned}
$$

Then, for any $M \in \mathbf{N}$ and $A \subset H_{1,\lfloor n/r \rfloor}^M$, we have

$$
\begin{aligned}
&\left|\mathbb{P}\left(\sum_{i=1}^{j}\delta_{irp}T_{ri}^{(l)} + \sum_{i=j+1}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A\right) - \mathbb{P}\left(\sum_{i=1}^{j-1}\delta_{irp}T_{ri}^{(l)} + \sum_{i=j}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} \in A\right)\right| \\
&\qquad \leq \sum_{h_2 \in H_{j+1,\lfloor n/r \rfloor}^M}\sum_{m=0}^{M-1}\mathbb{P}\left(\sum_{i=j+1}^{\lfloor n/r \rfloor}\delta_{irp}\widehat{T}_{ri}^{(l)} = h_2\right) \times
\end{aligned}
$$

8

$$\left| \mathbb{P}\left(\left\{ \sum_{i=1}^{j-1} \delta_{irp} T_{ri}^{(l)} + m\delta_{jrp} + h_2 \in A \right\} \cap \left\{ T_{rj}^{(l)} = m \right\} \right) \right.$$

$$\left. - \mathbb{P}\left(\sum_{i=1}^{j-1} \delta_{irp} T_{ri}^{(l)} + m\delta_{jrp} + h_2 \in A \right) \mathbb{P}\left(\widehat{T}_{rj}^{(l)} = m \right) \right|$$

$$\leq M\alpha(l),$$

and (2.8) implies that, for $A \subset H_{1,\lfloor n/r \rfloor}^M$,

$$\left| \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}^{(l)} \in A \right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri}^{(l)} \in A \right) \right| \leq Mnr^{-1}\alpha(l). \tag{2.11}$$

A standard argument now extends this to the bound

$$\left| \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri}^{(l)} \in A \right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri}^{(l)} \in A \right) \right|$$

$$\leq Mnr^{-1}\alpha(l) + \mathbb{P}\left(\max_{1 \leq j \leq \lfloor n/r \rfloor} \widehat{T}_{rj}^{(l)} \geq M \right) + \mathbb{P}\left(\max_{1 \leq j \leq \lfloor n/r \rfloor} T_{rj}^{(l)} \geq M \right)$$

$$\leq Mnr^{-1}\alpha(l) + 2nr^{-1}\mathbb{P}(T_{r1} \geq M), \tag{2.12}$$

valid for any $A \subset \mathcal{H}_r = H_{1,\lfloor n/r \rfloor}$. Combining these two bounds with (2.6) and (2.7), the lemma follows. $\blacksquare$

Proof of Theorem 2.1. The properties of d_2 yield

$$d_2\left(\mathcal{L}(\widetilde{N}), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri} \right) \right) \leq rp + q_r \leq 2rp, \tag{2.13}$$

the points of $\widetilde{N}$ each being moved at most a distance rp, and the last short block $B_r(\lfloor n/r \rfloor + 1)$ being omitted. By Lemma 2.2,

$$d_2\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} T_{ri} \right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri} \right) \right) \leq 2nr^{-1}lp + \varepsilon(l, r, M).$$

Then, setting $\pi_r = q_r \sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp}$, it follows that

$$d_2\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} \widehat{T}_{ri} \right), \mathrm{CP}(\pi_r, \nu_r) \right)$$

$$\leq d_2\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{irp} 1_{\{\widehat{T}_{ri} > 0\}} \right), \mathrm{Po}(\pi_r) \right) \leq \left(\frac{1.65}{\sqrt{1 - q_r}} + e^{q_r} \right) q_r, \tag{2.14}$$

9

where the last inequality is from Xia (1997). Now note that

$$d_2\left(\mathbf{CP}(\pi_r,\nu_r),\mathbf{CP}(\theta_r\mu,\nu_r)\right) \le d_2\left(\mathrm{Po}(\pi_r),\mathrm{Po}(\theta_r\mu)\right),$$

and, from Brown and Xia (1995, Formula (2.8)),

$$d_2\left(\mathrm{Po}(\pi_r),\mathrm{Po}(\theta_r\mu)\right) \le rp + q_r\lfloor n/r\rfloor - n/r| \le rp + q_r \le 2rp. \tag{2.15}$$

Combining (2.14)–(2.15) with Lemma 2.2, the theorem follows. ■

In order to use Theorem 2.1 for limit asymptotics as $n \to \infty$, it is traditional to suppose that $u = u_n$ is chosen so that $np_n \to t \in (0,\infty)$. A very weak mixing condition is then to suppose that $\alpha_n(l_n) \to 0$ for some sequence $l_n \to \infty$ such that $l_n = o(n)$. Choose M_n in such a way that $M_n \to \infty$ and that $M_n\alpha_n(l_n) \to 0$, and then choose $r_n = o(n)$ so that $l_n/r_n \to 0$ and that $M_n n r_n^{-1}\alpha_n(l_n) \to 0$. It then follows that the right hand side of (2.4) converges to 0.

Suppose now that the same mixing coefficients φ and α are valid for all $u = u_n$, and that $np_n \to t \in (0,\infty)$. Then, taking the φ–mixing estimate, one can choose $l = l_n$ so that $l/\varphi(l) \approx t^{-1}n$ and then $r = r_n$ so that $r = \lfloor n\sqrt{\varphi(l_n)/t}\rfloor$. This makes the bound in Theorem 2.1 of order $O(\sqrt{t\varphi(l_n)})$. So if $\varphi(l) \le l^{-\beta}$ for some $\beta > 0$, take $l_n = \lfloor(t^{-1}n)^{1/(1+\beta)}\rfloor$ to get a bound of order $O(n^{-\beta/2(1+\beta)})$; for an m–dependent sequence X_i, similar considerations give a rate of $O(n^{-1/2})$. However, the same choices can also be used when $np_n = t_n \to \infty$, in which case the carrier space $[0,t_n]$ for the point process becomes ever larger. For instance, if $\varphi(l) \le l^{-\beta}$ and $t_n = n^\eta$ for some $\eta > 0$, then the bound is of order $O(n^{-\delta/2})$, where $\delta = (\beta - \eta(1 + 2\beta))/(1 + \beta)$, and is useful if $\eta < \beta/(1 + 2\beta)$.

Finally, in order to obtain a limit, it should also be the case that $\nu_{r_n} \to \nu$ and $\theta_{r_n} = q_{r_n}/r_np_n \to \theta$. It is then easy to see that

$$\begin{aligned}
d_{TV}(\mathbf{CP}(\theta'\mu,\nu),\mathbf{CP}(\theta''\mu,\nu)) &\le d_{TV}(\mathrm{Po}(\theta'\mu(\Gamma)),\mathrm{Po}(\theta''\mu(\Gamma))) \\
&\le |\theta' - \theta''|\mu(\Gamma)\min\left\{1, 1/\sqrt{\mu(\Gamma)\max(\theta',\theta'')}\right\}
\end{aligned}$$

and that

$$d_{TV}(\mathbf{CP}(\lambda\mu,\nu'),\mathbf{CP}(\lambda\mu,\nu'')) \le \lambda\mu(\Gamma)d_{TV}(\nu',\nu'').$$

Then the simple estimate

$$d_2(\mathbf{CP}(\theta_r\mu,\nu_r),\mathbf{CP}(\theta\mu,\nu)) \le np\theta d_{TV}(\nu_r,\nu) + \theta^{-1/2}|\theta_r - \theta|\sqrt{np} \tag{2.16}$$

for the distance between the processes over the whole interval $[0,np]$ enables one to complete a bound for the entire approximation.

3 Compound Poisson approximation to $\Xi_{n,f}$

The point process $\Xi_{n,f}$ is defined on the two–dimensional space $[0,1] \times [0,\infty)$, and the choice of d_0 should now reflect the typical two–dimensional distance between points. Here,

10

for convenience, we keep the first coordinate as in (1.3), and rescale the second, by choice of f, so as to make the average number of points in a unit rectangle approximately 1; then we take

$$d_0((x_1, y_1), (x_2, y_2)) = 1 \wedge [|x_1 - x_2| + |y_1 - y_2|]. \tag{3.1}$$

This suggests taking f so that

$$n\mathbb{P}(X_i > f(t)) \asymp t, \quad 0 \le t \le K < n, \tag{3.2}$$

where $u = f(K)$ is the lower limit of X_i-values that are considered to be extreme. In particular, if the X_i's have a continuous, strictly monotone distribution function F, take

$$f(t) = F^{-1}(1 - t/n), \quad 0 \le t \le K. \tag{3.3}$$

We then define the point process of interest on $\Gamma = [0, 1] \times [0, K]$ by the measure

$$\tilde{\tilde{\Xi}} = \sum_{i=1}^{n} 1_{\{X_i > u\}} \delta_{(i/n, f^{-1}(X_i))}. \tag{3.4}$$

We need appropriate mixing conditions. Let $\mathcal{F}_{st}^M$ be the sigma–field generated by the events $\{mK/M < f^{-1}(X_j) \le (m+1)K/M\}$ for $0 \le m \le M - 1$ and $s \le j \le t$. Define

$$\varphi^{[M]}(l) := \max_{1 \le m \le n} \sup_{A \in \mathcal{F}_{1,m}^M, B \in \mathcal{F}_{m+l,n}^M} |\mathbb{P}(B|A) - \mathbb{P}(B)|;$$

$$\alpha^{[M]}(l) := \max_{1 \le m \le n} \sup_{A \in \mathcal{F}_{1,m}^M, B \in \mathcal{F}_{m+l,n}^M} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Then set

$$\tilde{\varepsilon}(l, r, M) := \max\{nr^{-1}\varphi^{[M]}(l), e^M nr^{-1}\alpha^{[M]}(l) + 2nr^{-1}\mathbb{P}(T_{r1} \ge M)\}, \tag{3.5}$$

$$\nu_r^* := \mathcal{L}\left(\sum_{j \in B_r(1)} 1_{\{X_j > u\}} \delta_{f^{-1}(X_j)} \,\middle|\, T_{r1} \ge 1\right).$$

Theorem 3.1 *For any choices of $M \ge 1$ and l, r such that $1 \le l < r \le n$, we have*

$$d_2\left(\mathcal{L}\left(\tilde{\tilde{\Xi}}\right), PC(nr^{-1}q_r\mu, \nu_r^*)\right) \tag{3.6}$$

$$\le 2(n^{-1}r + M^{-1}K) + 2nlr^{-1}p + \tilde{\varepsilon}(l, r, M) + \left(\frac{1.65}{\sqrt{1 - q_r}} + e^{q_r} + 2\right)q_r,$$

on the rectangle $[0, 1] \times [0, K]$.

If the mixing coefficients decay fast enough then the right–hand side of (3.6) is of order $\frac{r}{n} + \frac{K}{M} + \frac{npl}{r} + rp + \tilde{\varepsilon}(l, r, M) \asymp \left(\frac{1}{M} + \frac{l}{r} + \frac{r}{n}\right)K$, where K may depend on n. This suggests choosing $r = \sqrt{nl}$ and $M = \sqrt{n/l}$. If φ decays exponentially fast then we can put $l = C \ln n$ with a large enough constant C, and the right–hand side of (3.6) becomes $O\left(K\sqrt{n^{-1}\ln n}\right)$.

11

For any fixed $r, M \in \mathbf{Z}$ and $0 \le m \le M - 1$, define

$$T_{r,i;M,m} := \sum_{j \in B_r(i)} 1_{\{mK/M < f^{-1}(X_j) \le (m+1)K/M\}},$$

$$T_{r,i;M,m}^{(l)} := \sum_{j \in B_r^{(l)}(i)} 1_{\{mK/M < f^{-1}(X_j) \le (m+1)K/M\}}. \tag{3.7}$$

Let

$$\left((\widehat{T}_{r,i;M,0}, \widehat{T}_{r,i;M,1}, \ldots, \widehat{T}_{r,i;M,M-1}), (\widehat{T}_{r,i;M,0}^{(l)}, \widehat{T}_{r,i;M,1}^{(l)}, \ldots, \widehat{T}_{r,i;M,M-1}^{(l)}) \right)$$

be independent copies of the pairs of vectors

$$\left((T_{r,i;M,0}, T_{r,i;M,1}, \ldots, T_{r,i;M,M-1}), (T_{r,i;M,0}^{(l)}, T_{r,i;M,1}^{(l)}, \ldots, T_{r,i;M,M-1}^{(l)}) \right),$$

independent also of $X_1, \ldots, X_n$. Denote

$$\Xi_{r,i;M} := \sum_{m=0}^{M-1} \delta_{(ri/n,(m+1)K/M)} T_{r,i;M,m} \qquad (1 \le i \le \lfloor n/r \rfloor + 1), \tag{3.8}$$

and make analogous definitions also of $\Xi_{r,i;M}^{(l)}$, $\widehat{\Xi}_{r,i;M}$ and $\widehat{\Xi}_{r,i;M}^{(l)}$. Set

$$\nu_r^{[M]} = \mathcal{L}\left(\sum_{m=0}^{M-1} \delta_{(m+1)K/M} T_{r,i;M,m} \,\middle|\, T_{r,i} \ge 1 \right), \tag{3.9}$$

and let $q_r = \mathbb{P}(T_{r,i} \ge 1)$ be as before.

We can now formulate the analogue of the basic Lemma 2.2, relating the "blocks" process to the corresponding independent process.

Lemma 3.2 *For any choices $1 \le l < r \le n$ and $M \ge 1$, we have*

$$d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M} \right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{\Xi}_{r,i;M} \right) \right) \le 2nlr^{-1}p + \tilde{\varepsilon}(l, r, M). \tag{3.10}$$

Proof of Lemma 3.2. It follows as for (2.6) and (2.7) that

$$d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M} \right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M}^{(l)} \right) \right) \le nlr^{-1}p, \tag{3.11}$$

$$d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{\Xi}_{r,i;M} \right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{\Xi}_{r,i;M}^{(l)} \right) \right) \le nlr^{-1}p. \tag{3.12}$$

Using Lindeberg's device, we observe that

$$\mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M}^{(l)} \in B \right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \widehat{\Xi}_{r,i;M}^{(l)} \in B \right)$$

$$= \sum_{j=1}^{\lfloor n/r \rfloor} \left\{ \mathbb{P}\left(\sum_{i=1}^{j} \Xi_{r,i;M}^{(l)} + \sum_{i=j+1}^{\lfloor n/r \rfloor} \widehat{\Xi}_{r,i;M}^{(l)} \in B \right) - \mathbb{P}\left(\sum_{i=1}^{j-1} \Xi_{r,i;M}^{(l)} + \sum_{i=j}^{\lfloor n/r \rfloor} \widehat{\Xi}_{r,i;M}^{(l)} \in B \right) \right\}$$

12

for $B \subset \mathcal{H}$. So, arguing essentially as for (2.10), we deduce the bound

$$d_{TV}\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M}^{(l)}\right), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \hat{\hat{\Xi}}_{r,i;M}^{(l)}\right)\right) \leq \frac{n}{r}\varphi^{[M]}(l), \tag{3.13}$$

in terms of the coefficients $\varphi^{[M]}$. On the other hand, for a bound in terms of the coefficients $\alpha^{[M]}$, split

$$\mathbb{P}\left(\sum_{i=1}^{j} \Xi_{r,i;M}^{(l)} + \sum_{i=j+1}^{\lfloor n/r \rfloor} \hat{\hat{\Xi}}_{r,i;M}^{(l)} \in B\right) - \mathbb{P}\left(\sum_{i=1}^{j-1} \Xi_{r,i;M}^{(l)} + \sum_{i=j}^{\lfloor n/r \rfloor} \hat{\hat{\Xi}}_{r,i;M}^{(l)} \in B\right)$$

according to all the possible values of the vector $(T_{r,j;M,0}^{(l)}, T_{r,j;M,1}^{(l)}, \ldots, T_{r,j;M,M-1}^{(l)})$ which are consistent with $T_{r,j}^{(l)} \leq M - 1$, of which there are fewer than e^M, and argue as for (2.11), obtaining

$$\left|\mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M}^{(l)} \in B\right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \hat{\hat{\Xi}}_{r,i;M}^{(l)} \in B\right)\right| \leq e^M n r^{-1}\alpha^{[M]}(l) \tag{3.14}$$

for the corresponding events B. This implies that, for any $B \subset \mathcal{H}$,

$$\left|\mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M}^{(l)} \in B\right) - \mathbb{P}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \hat{\hat{\Xi}}_{r,i;M}^{(l)} \in B\right)\right|$$
$$\leq e^M n r^{-1}\alpha^{[M]}(l) + 2nr^{-1}q_r\nu_r\{[M,\infty)\}. \tag{3.15}$$

The proof is complete. ∎

Proof of Theorem 3.1. The theorem follows from Lemma 3.2 by much the same argument as is used to derive Theorem 2.1. Setting $\pi_r^* = q_r \sum_{i=1}^{\lfloor n/r \rfloor} \delta_{\frac{ir}{n}}$, it follows that

$$d_2\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \hat{\hat{\Xi}}_{r,i;M}\right), \mathrm{PC}(\pi_r^*, \nu_r^{[M]})\right)$$
$$\leq d_2\left(\mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \delta_{\frac{ir}{n}} 1_{\{\hat{T}_{ri} \geq 1\}}\right), \mathrm{Po}(\pi_r^*)\right) \leq \left(\frac{1.65}{\sqrt{1 - q_r}} + e^{q_r}\right) q_r, \tag{3.16}$$

where the last inequality is again from Xia (1997). The properties of d_2 yield

$$d_2\left(\mathcal{L}(\tilde{\tilde{\Xi}}), \mathcal{L}\left(\sum_{i=1}^{\lfloor n/r \rfloor} \Xi_{r,i;M}\right)\right) \leq rn^{-1} + KM^{-1} + q_r, \tag{3.17}$$

the points of $\tilde{\tilde{\Xi}}$ each being moved at most a distance $rn^{-1} + KM^{-1}$, and the last short block $B_r(\lfloor n/r \rfloor)$ being omitted, whereas, much as for (2.15),

$$d_2\left(\mathrm{PC}(\pi_r^*, \nu_r^{[M]}), \mathrm{PC}(nr^{-1}q_r\mu, \nu_r^*)\right) \leq rn^{-1} + KM^{-1} + q_r. \tag{3.18}$$

13

Combining (3.16)–(3.18) with Lemma 3.2, the theorem follows. ∎

There are other kinds of 'mixing' which could be exploited. One condition is that, for each $m \geq 1$, it is possible to construct a process $(X_{m+i}^{(m)}, \; i \geq 1)$ which is independent of $X_1, \ldots, X_m$, has the same distribution as $(X_{m+i}, \; i \geq 1)$, and satisfies

$$\mathbb{P}\left(\sup_{i \geq 1} \psi_1(i) |F(X_{m+i}) - F(X_{m+i}^{(m)})| > k\right) \leq 1/\psi_2(k), \tag{3.19}$$

for nondecreasing functions ψ_1 and ψ_2 such that $\lim_{i \to \infty} i^{-1}\psi_1(i) = \lim_{k \to \infty} \psi_2(k) = \infty$. Here, we assume that (3.3) is in force. Such a condition is typically satisfied, for instance, by the stochastic time reversal of the sequence of iterates of a uniformly expanding piecewise smooth map of $[0,1] \to [0,1]$: see Barbour, Gerrard and Reinert (1999). The bound (3.19) implies, using (3.3), that

$$\mathbb{P}\left(\sup_{i \geq l} |f^{-1}(X_{m+i}) - f^{-1}(X_{m+i}^{(m)})| > \frac{nk}{\psi_1(l)}\right) \leq 1/\psi_2(k). \tag{3.20}$$

Using the "blocks" argument, with the discretization

$$\Xi_{r,i} := \sum_{j \in B_r(i)} 1_{\{X_j > u\}} \delta_{(ri/n, f^{-1}(X_j))}; \quad \Xi_{r,i}^{(l)} := \sum_{j \in B_r^{(l)}(i)} 1_{\{X_j > u\}} \delta_{(ri/n, f^{-1}(X_j))},$$

it is easy to see that, for any function $g : \mathcal{H} \to \mathbf{R}$ such that $|g(\xi) - g(\eta)| \leq d_1(\xi, \eta)$, and for any $\xi_2 \in \mathcal{H}$, we have

$$\left| \mathbb{E}g\left(\sum_{i=1}^{j-1} \Xi_{r,i}^{(l)} + \Xi_{r,j}^{(l)} + \xi_2\right) - \mathbb{E}g\left(\sum_{i=1}^{j-1} \Xi_{r,i}^{(l)} + \widehat{\Xi}_{r,j}^{(l)} + \xi_2\right) \right|$$

$$\leq \frac{1}{\psi_2(k)} + \frac{nk}{\psi_1(l)} + \frac{2rk}{\psi_1(l)},$$

where $m = (j-1)r - l$ and

$$\widehat{\Xi}_{r,i}^{(l)} := \sum_{j \in B_r^{(l)}(i)} 1_{\{X_j^{(m)} > u\}} \delta_{(ri/n, f^{-1}(X_j^{(m)}))}.$$

In the upper bound, the first term arises when the event

$$G := \left\{\sup_{i \geq l} |f^{-1}(X_{m+i}) - f^{-1}(X_{m+i}^{(m)})| > nk/\psi_1(l)\right\}$$

occurs, the second when the event

$$G^c \cap \left\{\bigcap_{j' \in B_r^{(l)}(j)} \{1_{\{X_{j'} > u\}} = 1_{\{X_{j'}^{(m)} > u\}}\}\right\}$$

14

occurs, and the last term comes when the remaining event occurs, which entails

$$\cup_{j' \in B_r^{(l)}(j)} \{|F(X_{j'}) - F(u)| \le k/\psi_1(l)\},$$

because $F(X_{j'})$ has the uniform distribution on $[0,1]$. This replaces (3.13) in the proof of Lemma 3.2, leading to the following result.

Theorem 3.3 *Suppose that (3.3) and (3.19) hold. Then, taking any choices of $k \ge 1$ and l, r such that $1 \le l < r \le n$, we have*

$$d_2\left(\mathcal{L}\left(\widetilde{\Xi}\right), PC(nr^{-1}q_r\mu, \nu_r^*)\right) \tag{3.21}$$

$$\le 2n^{-1}r + 2nlr^{-1}p + \frac{n}{r\psi_2(k)} + \left(\frac{n}{r} + 2\right)\frac{nk}{\psi_1(l)} + \left(\frac{1.65}{\sqrt{1-q_r}} + e^{q_r} + 2\right)q_r,$$

for the processes on $[0,1] \times [0,K]$.

4 Applications

Example 1. As a first application, take $X_j := X(j)$, $j \in \mathbf{Z}$, where the stationary Markovian process X is Brownian motion reflected at zero, with drift $-c$, $c > 0$, and with infinitesimal variance σ^2. This process arises as a typical heavy traffic limit in the analysis of queueing models: see Harrison (1985). The stationary distribution F of X_0 is given by

$$1 - F(x) = e^{-\beta x}, \quad \text{where } \beta := 2c/\sigma^2. \tag{4.1}$$

Recalling (3.3), we take

$$f(t) := -\beta^{-1}\log(t/n), \quad 0 \le t \le K := n^\alpha, \tag{4.2}$$

say, for any fixed $0 \le \alpha < 1/2$, corresponding to taking $u = \beta^{-1}(1-\alpha)\log n$ as the lower limit of 'extreme' values among the X_j's, with an expected number of n^α extreme values on the interval $1 \le j \le n$.

In order to analyse the processes $N_{n,u}$ and $\Xi_{n,f}$ of extremes, we use a mixing condition of φ-type. This involves the future distribution of X conditional on *any* set $A \in \mathcal{F}_{1,m}$, which is easier to handle for the related bounded Markov process $\widehat{X} = \widehat{X}u_n(t)$ constructed by reflecting also at the upper boundary $3\beta^{-1}\log n$. The processes X_j and $\widehat{X}_j$ have almost identical distributions on $1 \le j \le n$, as can be seen by the following coupling construction. Start X and $\widehat{X}$ independently at time $-l_n$, where $l_n := \lceil 9(\sigma/c)^2 \log n \rceil$, and run them from then on with the *same innovations*. This coupling is monotone until the time τ_n at which processes first meet, which occurs before the initially larger of the processes first hits zero; thereafter, they remain coupled until X next exceeds $3\beta^{-1}\log n$. Now, for any $m \in \mathbf{Z}$ and $T \ge 1$, we have

$$\mathbb{P}\left(\max_{m \le j \le m+T-1} X_j > 3\beta^{-1}\log n\right) \le T(1 - F(3\beta^{-1}\log n)) \sim Tn^{-3}; \tag{4.3}$$

15

also, $\mathbb{P}\left(\min_{m \le j \le m+T-1} X_j > 0 \mid X_0 = x\right)$ is increasing in x, and satisfies

$$\mathbb{P}\left(\min_{m \le j \le m+T-1} X_j > 0 \mid X_0 = x\right) \le \mathbb{P}\left(\mathcal{N}(x - cT, T\sigma^2) > 0\right) \le e^{-(cT-x)^2/2T\sigma^2}, \quad (4.4)$$

so that taking $x = 3\beta^{-1} \log n$ and $T = -m = l_n$ gives

$$\mathbb{P}(\tau_n > 0) \le n^{-3}. \quad (4.5)$$

It thus follows, from (4.5) and from (4.3) with $m = -l_n$ and $T = n + 1 + l_n$, that

$$d_{TV}\left(\mathcal{L}\left((X_j)_{j=1}^n\right), \mathcal{L}\left((\widehat{X}_j)_{j=1}^n\right)\right) \le \mathbb{P}\left[\bigcup_{j=1}^n \{X_j \ne \widehat{X}_j\}\right] \le (n + 2 + l_n)n^{-3} \le 2n^{-2}, \quad (4.6)$$

whenever n is large enough that $n \ge 2 + l_n$. A similar coupling argument, now used with two realizations of $\widehat{X}$ starting at time m, one with the stationary distribution of $\widehat{X}_m$ (Harrison (1985, p. 90)) and the other with any arbitrarily chosen distribution, then shows that

$$\varphi_{\widehat{X}}(l_n) \le n^{-2}. \quad (4.7)$$

This enables Theorems 2.1 and 3.1 to be simply applied to the $\widehat{X}$-sequence.

It also follows from (4.3) and (4.5) that

$$d_{TV}(\mathcal{L}_r, \widehat{\mathcal{L}}_r) \le (r + 2 + l_n)n^{-3}, \quad (4.8)$$

where $\mathcal{L}_r := \mathcal{L}\left(\sum_{j \in B_r(1)} 1_{\{X_j > u\}} \delta_{f^{-1}(X_j)}\right)$, and the hat is used to denote quantities derived from the process $\widehat{X}$. This latter bound is useful for relating the approximations given in Theorems 2.1 and 3.1 for the $\widehat{X}$-sequence to those for the original X-sequence, because

$$\begin{aligned}
d_{TV}(\mathcal{L}_r, \widehat{\mathcal{L}}_r) &= \tfrac{1}{2}\{|q_r - \hat{q}_r| + \|q_r \nu_r^* - \hat{q}_r \hat{\nu}_r^*\|\} \\
&= \tfrac{1}{2}\{|q_r - \hat{q}_r| + \|q_r(\nu_r^* - \hat{\nu}_r^*) + (q_r - \hat{q}_r)\hat{\nu}_r^*\|\} \\
&\ge \tfrac{1}{2}\{|q_r - \hat{q}_r| + q_r\|(\nu_r^* - \hat{\nu}_r^*)\| - |q_r - \hat{q}_r|\|\hat{\nu}_r^*\|\} \\
&= q_r d_{TV}(\nu_r^*, \hat{\nu}_r^*) \ge q_r d_{TV}(\nu_r, \hat{\nu}_r);
\end{aligned} \quad (4.9)$$

furthermore, much as for (4.8), $|p - \hat{p}| \le (2 + l_n)n^{-3}$. Hence it follows that, on $[0, np]$,

$$\begin{aligned}
d_{TV}\left(\mathbf{CP}(\theta_r\mu, \nu_r), \mathbf{CP}(\hat{\theta}_r\mu, \hat{\nu}_r)\right) &\le np|\theta_r - \hat{\theta}_r| + n\theta_r p \, d_{TV}(\nu_r, \hat{\nu}_r) \\
&\le nr^{-1}|q_r - \hat{q}_r| + n|p - \hat{p}| + nr^{-1}d_{TV}(\mathcal{L}_r, \widehat{\mathcal{L}}_r) \\
&= O(n^{-2}(1 + r^{-1}\log n)),
\end{aligned} \quad (4.10)$$

with a further error of at most $n|p - \hat{p}|$ to account for the difference between the intervals $[0, np]$ and $[0, n\hat{p}]$; and, on $[0, 1]$,

$$d_{TV}\left(\mathrm{PC}(nr^{-1}q_r\mu, \nu_r^*), \mathrm{PC}(nr^{-1}\hat{q}_r\mu, \hat{\nu}_r^*)\right) = O(n^{-2}(1 + r^{-1}\log n)) \quad (4.11)$$

16

also.

So taking $l = l_n$ and $r = r_n = \lfloor (n \log n)^{1/2} \rfloor$, and noting that $p = 1 - F(u) = n^{\alpha-1}$, we can apply Theorem 2.1 to the $\widehat{X}$-sequence, with an error which is of order $O(n^{\alpha-1/2}\sqrt{\log n})$, and then convert the result to the approximation

$$d_2\left(\mathcal{L}\left(\widetilde{N}\right), \mathrm{CP}(\theta_r\mu, \nu_r)\right) = O\left(n^{\alpha-1/2}\sqrt{\log n}\right),$$

on $[0, np]$ for the X-sequence, since the additional error is of smaller order. In a similar way, we obtain

$$d_2\left(\mathcal{L}\left(\widetilde{\Xi}\right), \mathrm{PC}(nr^{-1}q_r\mu, \nu_r^*)\right) = O\left(n^{\alpha-1/2}\sqrt{\log n}\right),$$

on $[0,1] \times [0, n^\alpha]$ for the X-sequence, since M can be chosen to be arbitrarily large.

Example 2. Let the sequence $(X_j;\ 1 \le j \le n)$ be defined by the deterministic 'tent map' recursion $X_{j-1} = h_c(X_j)$ starting with $X_n \sim U[0,1]$, where, for some $0 < c < 1$,

$$h_c(x) = \begin{cases} x/c & \text{if } 0 \le x \le c; \\ (1-x)/(1-c) & \text{if } c \le x \le 1. \end{cases}$$

This process has the same joint distributions as the Markov chain on $[0,1]$ with transition probabilities

$$X_{j+1} = \begin{cases} cX_j & \text{with probability } c; \\ 1 - (1-c)X_j & \text{with probability } 1-c, \end{cases}$$

which has stationary distribution $F = U[0,1]$ and satisfies (3.19) with

$$\psi_1(i) = (1/\max\{c, 1-c\})^i$$

and $\psi_2(M) = \infty$ for $M > 1$ (Barbour et al. (1999)). In accordance with (3.3), define $f(t) = 1 - t/n$ for $0 \le t \le n^\alpha$, for any fixed $0 \le \alpha < 1/2$. Take

$$r = \lfloor \{n \log n\}^{1/2} \rfloor, \quad l = \lceil (3/\log[1/\max\{c, 1-c\}]) \log n \rceil,$$

and observe that $p = n^{\alpha-1}$ so $np = cn^\alpha \to \infty$; applying Theorem 3.3, it follows that

$$d_2\left(\mathcal{L}\left(\widetilde{\Xi}\right), \mathrm{PC}(nr^{-1}q_r\mu, \nu_r^*)\right) = O\left(n^{\alpha-1/2}\sqrt{\log n}\right).$$

In both examples, the approximation improves with increasing n for all values of $\alpha < 1/2$, so that the main problem that remains is to identify q_r and the distribution ν_r^*. This is usually no easy matter. In the latter case, the simple form of the recursion shows that, if $X_j > u$, then the s consecutive preceding values $X_{j'}$, $j - s \le j' \le j - 1$, have to satisfy $X_{j'} \le c$, where $s = \lfloor \{(1-\alpha)/\log(1/c)\} \log n \rfloor - 1$, and this event requires s consecutive choices of the first branch of h_c^{-1}, an event of probability at most $c^{-2}n^{\alpha-1} = O(p)$. Thus the approximation is actually a Poisson process approximation, to an extra error of order at most $O((n/r)rp^2) = O(n^{2\alpha-1})$, a relatively small adjustment. If, on the other hand, the same techniques were used for the small extremes, the process approximating $N_{n,u}$ would be a compound Poisson process with $\nu_r\{j\} = (1-c)c^{j-1}$, $j \ge 1$, a geometric random variable, to the same order of accuracy. In the former case, the distributions ν_r and ν_r^* involve excursion theory for Brownian motion with negative drift.

17

REFERENCES

Barbour, A. D. and Brown, T. C. (1992) *Stein's method and point process approximation,* Stochastic Processes and their Applications 43, 9-31.

Barbour, A. D., Gerrard, R. and Reinert, G. (1999) *Iterates of expanding maps,* Probab. Theory Related Fields (to appear).

Brown, T. C. and Xia, A. (1995) *On Metrics in Point Process Approximation,* Stochastics and Stochastics Reports 52, 247-263.

Eberlein, E. (1984) *Weak convergence of partial sums of absolutely regular sequences,* Statist. Prob. Letters **2**, 291-293.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer Verlag, Berlin.

Harrison, J. M. (1985) *Brownian motion and stochastic flow systems.* Wiley, New York.

Hsing, T., Hüsler, J. and Leadbetter, M. R. (1988) *On the exceedance point process for stationary sequence,* Probab. Theory Related Fields 78, 97-112.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* — New York: Springer Verlag, 366 pp.

Lindeberg, Y.W. (1922) *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung,* Math. Z., **15**, 221-225.

Novak, S. Y. (1996) *Extreme values in stationary sequences,* Siberian Adv. Math. **6** (3), 68-80.

Novak, S. Y. (1998) *On the limiting distribution of extremes,* Siberian Adv. Math. **8** (2), 70-95.

O'Brien, G. (1974) *Limit theorems for the maximum term of a stationary process,* Ann. Probab. **2**, 540-545.

Utev, S.A. (1989) *Sums of φ-mixing random variables,* Trudy Inst. Mat. (Novosibirsk) **13**, 78-100. (in Russian)

Xia, A. (1997) *On the rate of Poisson process approximation to a Bernoulli process,* J. Appl. Probab. **34**, 898-907.