Extremes, Sums, Lévy processes, and ageing

Lectures given in 2010 at the Technion, Haifa

Anton Bovier

Institut für Angewandte Mathematik
Rheinische Friedrich-Wilhelms-Universität Bonn
Endenicher Allee 60
53115 Bonn, Germany
bovier@uni-bonn.de

Contents

Preface

These notes grew out of various older versions of lecture notes on ageing and on extreme value theory. After working for a number of years on problems related to so-called *ageing* in the dynamics of disordered systems, mainly spin glasses, it became increasingly clear that in studying all these different models, that there was a operational mode at work in all of them, the interplay between extreme values and averaging in the theory of sums of random variables. The general theory of sums of independent random variables goes back at least to Gnedenko and Kolmogorov [26] from about 1950. Later, also situations with dependent random variables were studied intensely, under, as Durrett and Resnick [19] put it, *"a bewildering assortment of assumptions"*. Durrett and Resnick in 1978 give a clear picture of the situation, emphasizing the connection between convergence of an extremal process to a Poisson point process and the convergence of the corresponding sum process to the associated Lévy process, and formulating a clear and distinct set of conditions for these things to happen. It appears that these very beautiful results are not very widely known in the community at large.

The purpose of these lecture notes is thus twofold. On the one hand, I want to give a concise presentation of the connection between extreme values theory and Poisson point processes on the one hand, and pure jump Lévy processes and limit theorems for sums of random variables on the other. We will see, in particular, very clearly, that the dependence problem enters only at the level of the question whether the extremal process converges to a Poisson process. Many conditions for this to hold are known.

The second aspect will be applications of these observations to disordered systems. This will concern partition functions of disordered systems, such as the random energy model, but more than that the scaling properties of Markov jump processes in random environments, viz. *ageing*. It has long emerged that the *clock process* associated to these models plays a major part and, being a sum of in general dependent random variables, it is natural that the general limit theorems for such sums play a major rôle here. In general, there is still no free lunch, and verifying the hypothesis of the abstract theorems in concrete models is not immediate, but as I will explain in a variety of illustrative examples, there is a general strategy at work here, too.

I hope that these notes will help to advertise a central part of classical probability and demonstrate that amazing results can be wrought out of it in modern applications.

The insights I have gained over the years in this subject came from my collaborators, Gérard Ben Arous, Jiří Černý and Véronique Gayrard. With Gérard and Jiří [2] we first learned how to view the asymptotic clock as a sum of the extremes of a correlated process. Véronique, in her seminal series of papers [23, 24, 15], made the clear connection to the work of Durrett and Resnick and showed how to apply this in the context of a frozen random environment. I also learned a lot from Irina Kurkova when studying the problem of local energy statistics in random media, and from her and Mathias Löwe when analysing the fluctuations of the partition function of the REM (where in fact we rediscovered a lot of general wisdom from the theory of sums of triangular arrays). I wish to thank all of them for the pleasure of working together and exploring the fascinating world of probability theory.

These notes were written to a large part while I was holding a Lady Davies Visiting Professorship at the Technion, Haifa, and was staying at the William-Davidson Faculty of Industrial Engineering and Management. I would like to thank Dima Ioffe and the other colleagues at the Technion for their kind hospitality and the friendly atmosphere there. I thank the Lady Davies foundation for the financial support that made this stay possible.

Haifa, July 2010 Anton Bovier

1

Introduction

These lectures are motivated by recent research on certain properties of the dynamics of highly disordered systems, in particular spin glasses. More than 20 years ago, physicist began to investigate such systems and discovered some peculiar universal properties that they called *ageing*. About ten years ago, a small group of mathematicians became interested in this phenomenon and a systematic investigation of a number of models was started. While many deep probles remain unanswered, some understanding of the reasons begind what was going on emerged. Interestingly, this links back to some of the most classical parts of probability, *extreme value theory* and the fundamental limit theorems for sums of random variables, and hence *Lévy processes*. In these lectures I will tell this story backwards, exposing first the classical theory of Poisson convergence and stable Lévy processes and in the end explaining their appearance in the dynamics of disordered systems.

1.1 Characterisation of ageing

The term *ageing* refers to properties of a dynamical system out of equilibrium. In principle, this property refers to *real (physical) systems*. In the widest sense we can describe it as follows. Assume a systems is prepared (produced) at some initial time t_0. Then the system is left to itself. After some time t_w (called *waiting time*), an experimentalist may perform some measurement on the system. The question is, whether the experimentalist will be able to deduce the elapsed waiting time from his observation. If the answer is yes, we will say that the system ages, otherwise it does not.

Of course, this is a very general characterization and we will be interested in more specific situations.

In these lectures we will be concerned with mathematical models that correspond to this behavior. Again, one could look at very general dynamical systems, but we will confine our interest exclusively to *Markov processes* in random environments.

Let us introduce some notation.

1.2 Markov jump processes in random environments

The models we will ben interested in general are Markov jump processes in random environments. We construct them with a particular twist that will suit our purposes.

Our arena will be a sequence of loop-free graphs, $G_n(\mathcal{V}_n, \mathcal{E}_n)$ with set of vertices $\mathcal{V}_n$ and set of edges $\mathcal{E}_n$. Not that this sequence may consist of a single element, if the cardinality of $\mathcal{V}_n$ is infinite.

A *random environment* will be a family of positive random variable, $\tau_n(x), x \in \mathcal{V}_n$, defined on some abstract probability space, $(\Omega, \mathcal{F}, \mathbb{P})$. We will denote the sigma-algebra generated by these variables by $\mathcal{F}^\tau$, and their joint distribution by $\mathbb{P}$. Note that we do not assume independence.

Next we define discrete time Markov processes, J_n, with state space $\mathcal{V}_n$ and non-zero transition probabilities along the edges, $\mathcal{E}_n$. We denote by μ_n its initial distribution and by $p_n(x, y)$ the elements of its transition matrix. Note that the p_n may be random variables on the space $(\Omega, \mathcal{F}, \mathbb{P})$ (i.e. depend on the variables $\tau_n(x)$. We will assume that the process J_n admits a unique invariant measure π_n. We will often refer to J_n as the *fast chain*.

We will construct our process of interest, X_n, as a time change of J_n. To this end we set

$$\lambda_n(x) \equiv \pi_n(x)/\tau_n(x), \tag{1.1}$$

and define the *clock process*

$$S_n(k) = \sum_{i=0}^{k} \lambda_n^{-1}(J_n(i)) e_{n,i}\,, \quad k \in \mathbb{N}\,, \tag{1.2}$$

where $(e_{n,i}\,, n \in \mathbb{N}, i \in \mathbb{N})$ is a family of independent mean one exponential[1] random variables, independent of J_n.

[1] We may consider more general situations when $e_{n,i}$ have different distributions as well.

We now define our continuous time process of interest, X_n, as

$$X_n(t) = J_n(i), \quad \text{if } S_n(i) \leq t < S_n(i+1) \quad \text{for some } i. \tag{1.3}$$

One can readily verify that X_n is a continuous time Markov process with infinitesimal generator λ_n, whose elements are

$$\lambda_n(x,y) = \lambda_n(x) p_n(x,y) \tag{1.4}$$

and invariant measure

$$\pi_n(x)\lambda_n^{-1}(x) = \tau_n(x). \tag{1.5}$$

Note that the numbers $\lambda_n^{-1}(x)$ play the rôle of the mean holding time of the process X_n in a site x.

To avoid pathologies, we will always assume that our processes are such that, almost surely, for all but finitely many n,

$$\lambda_n(x) < \infty, \forall x \in \mathcal{V}_n. \tag{1.6}$$

Note that all Markov jump processes may be constructed in this way. For future reference, we will refer to the σ-algebra generated by the variables J_n and X_n as $\mathcal{F}^J$ and $\mathcal{F}^X$, respectively. We will write $P_{n,\omega} \equiv P_\omega \equiv P_n$ for the law of the process J_n, conditional on the σ-algebra $\mathcal{F}^\tau$, i.e. for fixed realisations of the random environment. Likewise we will call $\mathcal{P}_{n,\omega} \equiv \mathcal{P}_\omega \equiv \mathcal{P}_n$ the law of X_n conditional on $\mathcal{F}^\tau$.

This construction brings out the crucial rôle played by the clock process. If the chain J_n is rather quickly mixing, convergence to equilibrium can only be slowed through an erratic behaviour of the clock process. This process, on the other hand, is a sum of positive random variables, albeit in general dependent ones.

In connection with ageing, we do of course have some particular cases in mind.

1.2.1 Trap models

The best studied models for aging are the so-called *trap models*, introduced essentially by Bouchaud and Dean [13, 14]. These models were introduced as caricatures of more realistic models, but they teach us something about how one would like to think about ageing systems. The setting is like the one above, but we make some specific assumptions on the random environment:

(i) A random environment is given by variables $\tau_n(i) \equiv \tau_i, i \in \mathcal{V}_n$ which is a family of positive, independent, and identically distributed random variables that are in the domain of attraction of an α-stable distribution with $\alpha < 1$, i.e. $\lim_{t\uparrow\infty} t^\alpha \mathbb{P}[\tau_i > t] = 1$. In particular $\mathbb{E}\tau_i = +\infty$.

(ii) For any realization of the random variables τ_i, the fast chain, J_n, is a discrete time Markov process with transition probabilities proportional to

$$p_n(i,j) = \begin{cases} \frac{\tau_j^a}{\sum_{(i,k)\in\mathcal{E}} \tau_k^a}, & \text{if } (i,j) \in \mathcal{E}, \\ 0, & \text{, else} \end{cases} \tag{1.7}$$

for some parameter $0 \leq a \leq 1$.

(iii) The invariant measure of the chain J is uniform and will be chosen un-normalize as $\pi(i) = 1$.

We will in these lectures mainly consider the case $a = 0$, which is the original choice of Bouchaud. In that case the dynamics has a simple description: starting in some site, i, the process waits an exponential time with mean τ_i, and then moves on uniformly to one of its neighbors in the graph $\mathcal{G}$. So the random environment does not affect the paths of the process, but only the time spent along a path. We think of the process of interest as a *random time change* of a *fast* process. The general case is treated nicely in [23].

Now the random variables τ_i, i.e. the trapping times, have a very heavy-tailed distribution, so that as the process wanders about, it can find ever deeper traps, i.e. sites where it will wait longer and longer. So if it is the case that the process, by time T is with large probability in a trap whose waiting time is of order $g(T)$, then we can indeed determine the age of the process by studying its current typical sojourn times. The nice feature of trap models in that respect is that the state space has site by site a temporal characteristic, a feature that more complicated models do not immediately show.

There has been a considerable amount of work done on trap models in the case when $\mathcal{G} = \mathcal{G}_N$ is the complete graph and when $\mathcal{G} = \mathbb{Z}^d$, mostly by Ben Arous and Černý [8, 5, 6, 7].

To quantify ageing properties, one usually refers to the behaviour of certain time-time-correlation functions. This corresponds to the idea that an aging experiment corresponds to making an observation of the system during a time interval $[t_w, t_w + t$ and to record the outcome of the measurement as a function $R(t_w, t)$.

When studying trap models, the most commonly used *correlation functions* are

$$R[t_w, t] \equiv \mathbb{P}[X(t_w + t) = X(t_w)], \tag{1.8}$$

respectively its *quenched* version

$$R_\omega[t_w, t] \equiv \mathcal{P}_\omega[X(t_w + t) = X(t_w)]. \tag{1.9}$$

Another correlation function is

$$\Pi[t_w, t] \equiv \mathbb{P}[X(t_w + s) = X(t_w), \forall 0 \leq s \leq t], \tag{1.10}$$

respectively

$$\Pi_\omega[t_w, t] \equiv \mathcal{P}_\omega[X(t_w + t) = X(t_w), \forall 0 \leq s \leq t]. \tag{1.11}$$

One could of course, instead of just asking that $X(s) = X(t_w)$ ask for a milder version, like $\operatorname{dist}(X(s), X(t_w))$, for some distance, or one might ask for the distribution of such a distance. However, as again we will see later, it is in the spirit of the trap model to use the strict definition above: for large times, very deep traps are quite isolated, and so the right thing to realize the event is for the process to be in the same deep trap (most of) all the time.

One now speaks about ageing systems, if these functions, as t and t_w become large, do not become independent of t_w. In fact, in these notes we will not stress the role of correlation functions too much.

1.2.2 Glauber dynamics

Trap models may reproduce ageing behavior, but they are in some sense ad hoc models, that are not motivated by microscopic physical models. In particular, they have two features that seem artificial built in: the independence of the traps and the heavy tails of the distribution of the traps.

Models that are a step closer to reality are Glauber dynamics of (random) spin systems. Here we consider as state space the hypercube $\mathcal{S}_n \equiv \{-1, 1\}^n$ (we could also be more general), and defined on this an energy function (Hamiltonian) $H_n(\sigma)$ which may depend on a random parameter, i.e. may be considered as a random process indexed by $\mathcal{S}_n$. The examples we will be concerned with here are so-called *mean-field* spin glasses, where H_n is a centered Gaussian process with some covariance

$$\operatorname{cov}(H_n(\sigma), H_n(\sigma')) = n f(\operatorname{dist}_n(\sigma, \sigma')),$$

for some function f such that $f(0) = 1$ and dist_n a normalized distance. The most prominent examples are the p-spin interaction Sherrington-Kirkpatrick models, where

$$\text{cov}(H_n(\sigma), H_n(\sigma')) = nR_n(\sigma, \sigma')^p, \tag{1.12}$$

with $R_n(\sigma, \sigma') \equiv n^{-1} \sum_{i=1}^n \sigma_i \sigma'_i$. Given such a Hamiltonian, one constructs a *Gibbs measure*

$$\mu_{\beta,n}(\sigma) \equiv \frac{2^{-n} \exp(-H_n(\sigma))}{Z_{\beta,n}}, \tag{1.13}$$

where $Z_{\beta,n}$ is such that $\mu_{\beta,n}$ is a probability.

A *Glauber dynamics* is then a (discrete or continuous time) Markov chain that is reversible with respect to this measure. In most cases, one assumes also that only transitions are allowed in which a single spin is flipped at a time. Popular rates are:

Metropolis rates:

$$p(\sigma, \sigma') = \exp\left(-\beta[H_n(\sigma') - H_n(\sigma)]_+\right), \quad \text{if } |\sigma - \sigma'| = 2 \tag{1.14}$$

and zero else; these rates are rather difficult to handle because in our setting, the fast chain is not a simple object. First results are announced in the forthcoming paper [25].

Nicer to to handle are

random time change rates:

$$p(\sigma, \sigma') = \exp\left(\beta H_n(\sigma)\right), \quad \text{if } |\sigma - \sigma'| = 2 \tag{1.15}$$

This is easily cast in the above form with the fast chain given by simple random walk on the hypercube and $\lambda_n(\sigma) \equiv \exp\left(\beta H_n(\sigma)\right)$. Results on this type of dynamics were obtained in [3, 4, 7, 2, 15].

We see that in these dynamics, neither independence nor heavy tails appear. Nonetheless, one expects that under suitable conditions, trap model dynamics emerges as appropriate description of the long time behavior of these models (when $N \uparrow \infty$). To understand how this happens will be the main theme of these lectures.

Ackowledgement. These notes were written for the most part while I was a Lady Davies Visiting Professor at the Technion in Haifa. I am grateful to the Lady Davies foundation for their financial support and to Dima Ioffe and the William Davidson Faculty of Industrial Engeneering and Management for their kind hospitality.

2

Extreme value theory and Poisson point processes

Extreme value theory is an enormously rich subject with an inexhausible litterature. Two very good reference, from which much of this chapter is taken, are the textbooks [31] and [32].

2.1 Independent random variables

Our first building block will be the theory of extremes of random variables. The setting here is the following. Let X_i, $i \in \mathbb{N}$, be a family of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let us say that the X_i have all the same distribution, say, F, i.e. $\mathbb{P}(X_i \leq x) = F(x)$. Much of the material of this section is taken from [31] and [32].

2.1.1 Extremal distributions

The first question we can ask is whether something can be said about the distribution of

$$M_n \equiv \max_{i \leq n} X_i. \tag{2.1}$$

The first context in which one can say something is when the X_i are *independent* random variables. In that case, obviously,

$$\mathbb{P}(M_n \leq u) = \mathbb{P}\left(X_i \leq u, \forall i \leq n\right) = F(u)^n. \tag{2.2}$$

As probability is about limit theorems, to get an interesting result we must choose $u = u_n$ as a function of n and try to do this in such a way that

$$F(u_n)^n \to G \in (0, 1). \tag{2.3}$$

Even more interesting, we may try to find a family, $u_n(x)$, $x \in \mathbb{R}$, such that

$$F(u_n(x))^n \to G(x), \tag{2.4}$$

where $G(x)$ is a probability distribution function! Here and throughout $u_n(x)$ will be assumed to be a strictly monotone continuous function. Such a result could then be interpreted by saying that $u_n^{-1}(M_n)$ converges in law to a random variable with distribution function G.

One of the remarkable theorems in probability theory states that if $u_n(x)$ is chosen as an affine function, e.g. $u_n(x) = a_n x + b_n$, we have a full classification of the possible limit laws.

Theorem 2.1.1 *Let X_i, $i \in \mathbb{N}$ be a sequence of i.i.d. random variables. If there exist sequences $a_n > 0$, $b_n \in \mathbb{R}$, and a non-degenerate probability distribution function, G, such that*

$$\mathbb{P}\left[M_n \leq a_n x + b_n\right] \stackrel{w}{\to} G(x) \tag{2.5}$$

then $G(x)$ is of the same type as one of the three extremal-type distributions, namely:

(I) The Gumbel distribution,

$$G(x) = e^{-e^{-x}} \tag{2.6}$$

(II) The Fréchet distribution with parameter $\alpha > 0$,

$$G(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-x^{-\alpha}}, & \text{if } x > 0 \end{cases} \tag{2.7}$$

(III) The Weibull distribution with parameter $\alpha > 0$,

$$G(x) = \begin{cases} e^{-(-x)^{\alpha}}, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases} \tag{2.8}$$

Here we say that two functions, G, H, are of the same type, if there are constants, a, b, such that $G(x) = H(ax + b)$.

Proof. The proof of this theorem relies on the fact that the distributions we are looking for arise as limits of the form

$$F^n(a_n x + b_n) \to G(x) \tag{2.9}$$

This implies a certain invariance property of G, called *max-stability* .

Definition 2.1.1 A non-degenerate probability distribution function, G, is called *max-stable*, if for all $n \in \mathbb{N}$, there exists $a_n > 0, b_n \in \mathbb{R}$, such that, for all $x \in \mathbb{R}$,

$$G^n(a_n^{-1}x + b_n) = G(x) \tag{2.10}$$

The justification of this name is contained in the following proposition.

Proposition 2.1.2 *(i) A probability distribution, G, is max-stable, if and only if there exists probability distributions F_n and constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, for all $k \in \mathbb{N}$,*

$$F_n(a_{nk}^{-1}x + b_{nk}) \stackrel{w}{\to} G^{1/k}(x) \tag{2.11}$$

(ii) G is max-stable, if and only if there exists a probability distribution function, F, and constants $a_n > 0$, $b_n \in \mathbb{R}$, such that

$$F^n(a_n^{-1}x + b_n) \stackrel{w}{\to} G(x) \tag{2.12}$$

Proof. The proof of this proposition is slightly technical and will be omitted. It can be found in [31]. □

We do cite, however, an important result that is instrumental for the proof:

The next theorem is known as Khintchine's theorem:

Theorem 2.1.3 *Let F_n, $n \in \mathbb{N}$, be distribution functions, and let G be a non-degenerate distribution function. Let $a_n > 0$, and $b_n \in \mathbb{R}$ be sequences such that*

$$F_n(a_n x + b_n) \stackrel{w}{\to} G(x) \tag{2.13}$$

Then it holds that there are constants $\alpha_n > 0$, and $\beta_n \in \mathbb{R}$, and a non-degenerate distribution function G_*, such that

$$F_n(\alpha_n x + \beta_n) \stackrel{w}{\to} G_*(x) \tag{2.14}$$

if and only if

$$a_n^{-1}\alpha_n \to a, \quad (\beta_n - b_n)/a_n \to b \tag{2.15}$$

and

$$G_*(x) = G(ax + b) \tag{2.16}$$

There is a slight extension to this result.

Corollary 2.1.4 *If G is max-stable, then there exist functions $a(s) > 0, b(s) \in \mathbb{R}$, $s \in \mathbb{R}_+$, such that*

$$G^s(a(s)x + b(s)) = G(x) \tag{2.17}$$

Proof. This follows essentially by interpolation. We have that

$$G^{[ns]}(a_{[ns]}x + b_{[ns]}) = G(x) \tag{2.18}$$

But

$$\begin{aligned} G^n(a_{[ns]}x + b_{[ns]} &= G^{[ns]/s}(a_{[ns]}x + b_{[ns]})G^{n-[ns]/s}(a_{[ns]}x + b_{[ns]}) \\ &= G^{1/s}(x)G^{n-[ns]/s}(a_{[ns]}x + b_{[ns]}) \end{aligned}$$

As $n \uparrow \infty$, the last factor tends to one (as the exponent remains bounded), and so

$$G^n(a_{[ns]}x + b_{[ns]}) \stackrel{w}{\to} G^{1/s}(x) \tag{2.19}$$

and

$$G^n(a_n x + b_n) \stackrel{w}{\to} G(x) \tag{2.20}$$

Thus by Khintchine's theorem,

$$a_{[ns]}/a_n \to a(s), \quad (b_n - b_{[ns]})/a_n \to b(s) \tag{2.21}$$

and

$$G^{1/s}(x) = G(a(s)x + b(s)) \tag{2.22}$$

□

Theorem 2.1.1 is thus an immediate consequence of the following.

Theorem 2.1.5 *Any max-stable distribution is of the same type as one of the three distributions given in Theorem 2.1.1.*

Proof. Let us check that the three types are indeed max-stable. For the Gumbel distribution this is already obvious as it appears as extremal distribution in the Gaussian case. In the case of the Fréchet distribution, note that

$$\begin{aligned} G^n(x) &= \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-nx^{-\alpha}} = e^{-(n^{-1/\alpha}x)^{-\alpha}}, & \text{if } x > 0 \end{cases} \\ &= G(n^{-1/\alpha}x) \end{aligned} \tag{2.23}$$

which proves max-stability. The Weibull case follows in exactly the same way.

To prove that the three types are the only possible cases, we use Corollary 2.1.4. Taking the logarithm, it implies that, if G is max-stable, then there must be $a(s), b(s)$, such that

$$-s \ln(G(a(s)x + b(s))) = -\ln G(x) \tag{2.24}$$

One more logarithm leads us to

$$
\begin{aligned}
&-\ln\left[-s\ln\left(G(a(s)x+b(s))\right)\right] \\
&= -\ln\left[-\ln\left(G(a(s)x+b(s))\right)\right] - \ln s \stackrel{!}{=} -\ln\left[-\ln G(x)\right] \equiv \psi(x) \quad (2.25)
\end{aligned}
$$

or equivalently

$$\psi(a(s)x+b(s)) - \ln s = \psi(x) \tag{2.26}$$

Now ψ is an increasing function such that $\inf_x \psi(x) = -\infty$, $\sup_x \psi(x) = +\infty$. We can define the inverse $\psi^{-1}(y) \equiv U(y)$. Using (iv) Lemma **??**, we get that

$$\frac{U(y+\ln s) - b(s)}{a(s)} = U(y) \tag{2.27}$$

and subtracting the same equation for $y = 0$,

$$\frac{U(y+\ln s) - U(\ln s)}{a(s)} = U(y) - U(0) \tag{2.28}$$

Setting $\ln s = z$, this gives

$$U(y+z) - U(z) = \left[U(y) - U(0)\right] a(e^z) \tag{2.29}$$

To continue, we distinguish the case $a(s) \equiv 1$ and $a(s) \neq 1$ for some s.
Case 1. If $a(s) \equiv 1$, then

$$U(y+z) - U(z) = U(y) - U(0) \tag{2.30}$$

whose only solutions are

$$U(y) = \rho y + b \tag{2.31}$$

with $\rho > 0$, $b \in \mathbb{R}$. To see this, let $x_1 < x_2$ be any two points and let $\bar{x}$ be the middle point of $[x_1, x_2]$. Then (2.30) implies that

$$U(x_2) - U(\bar{x}) = U(x_2 - \bar{x}) - U(0) = U(\bar{x}) - U(x_1), \tag{2.32}$$

and thus $U(\bar{x}) = \left(U(x_2) - U(x_1)\right)/2$. Iterating this proceedure implies readily that on all points of the form $x_k^{(n)} x_1 + k2^{-n}(X_2 - x_1)$ we have that $U(x_k) = U(x_1) + k2^{-n}(U(x_2) - U(x_1))$; that is, on a dense set of points (2.31) holds. But since U is also monotonous, it is completely determined by its values on a dense set, so U is a linear function.
But then $\psi(x) = \rho^{-1}x - b$, and

$$G(x) = \exp\left(-\exp\left(-\rho^{-1}x - b\right)\right) \tag{2.33}$$

which is of the same type as the Gumbel distribution.
Case 2. Set $\widetilde{U}(y) \equiv U(y) - U(0)$, Then subtract from (2.29) the same equation with y and z exchanged. This gives

$$-\widetilde{U}(z)+\widetilde{U}(y)=a(e^z)\widetilde{U}(y)-a(e^y)\widetilde{U}(z) \tag{2.34}$$

or

$$\widetilde{U}(z)\,(1-a(e^y))=\widetilde{U}(y)\,(1-a(e^z)) \tag{2.35}$$

Now chose z such that $a(e^z)\neq 1$. Then

$$\widetilde{U}(y)=\widetilde{U}(z)\frac{1-a(e^y)}{1-a(e^z)}\equiv c(z)(1-a(e^y)) \tag{2.36}$$

Now we insert this result again into (2.29). We get

$$\begin{aligned}\widetilde{U}(y+z)&=c(z)\left(1-a(e^{y+z})\right) &(2.37)\\ &=\widetilde{U}(z)+\widetilde{U}(y)a(e^z) &(2.38)\\ &=c(z)\,(1-a(e^z))+c(z)\,(1-a(e^y))\,a(e^z) &(2.39)\end{aligned}$$

which yields an equation for a, namely,

$$a(e^{y+z})=a(e^y)a(e^z) \tag{2.40}$$

The only functions satisfying this equation are the powers, $a(x)=x^\rho$. Therefore,

$$U(y)=U(0)+c(1-e^{\rho y}) \tag{2.41}$$

Setting $U(0)=\nu$, going back to G this gives

$$G(x)=\exp\left(-\left(1-\frac{x-\nu}{c}\right)^{-1/\rho}\right) \tag{2.42}$$

for those x where the right-hand side is <1.
To conclude the proof, it suffices to discuss the two cases $-1/\rho\equiv\alpha>0$ and $-1/\rho\equiv-\alpha<0$, which yield the Fréchet, resp. Weibull types. □

□

Importantly, we can also identify which probability distributions give rise to the three types.

In the following theorem we set $x_F\equiv\sup\{x:F(x)<1\}$.

Theorem 2.1.6 *The following conditions are necessary and sufficient for a distribution function, F, to belong to the domain of attraction of the three extremal types:*

(I) Fréchet: $x_F=+\infty$,

$$\lim_{t\uparrow\infty}\frac{1-F(tx)}{1-F(t)}=x^{-\alpha},\quad \forall_{x>0},\alpha>0 \tag{2.43}$$

(II) Weibull: $x_F < +\infty$,

$$\lim_{h\downarrow 0} \frac{1-F(x_F - xh)}{1-F(x_F - h)} = x^\alpha, \quad \forall_{x>0}, \alpha > 0 \tag{2.44}$$

(III) Gumbel: $\exists g(t) > 0$,

$$\lim_{t\uparrow x_F} \frac{1-F(t+xg(t))}{1-F(t)} = e^{-x}, \quad \forall_x \tag{2.45}$$

Proof. We will only prove the sufficiency of the criteria. The following lemma contains an elementary, but important statement:

Lemma 2.1.7 *The statements*

$$n(1-F(a_n x + b_n)) \to g(x) \tag{2.46}$$

and

$$F^n(a_n x + b_n) \to e^{-g(x)} \tag{2.47}$$

are equivalent.

Proof. Assume first that (2.47) holds. Then

$$n \ln F(a_n x + b_n) \to -g(x). \tag{2.48}$$

But

$$n \ln(F(a_n x + b_n) = n \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \left(1 - F(a_n x + b_n)\right)^k. \tag{2.49}$$

It is easy to see that the only way the right hand side can coverge is when the $k = 1$-term in the sum,

$$-n(1-F(a_n x + b_n)), \tag{2.50}$$

tends to $-g(x)$, which then implies that all other terms tend to zero. Thus (2.46) must hold.
Conversely, if (2.46) holds, (2.47) follows since

$$\lim_{n\uparrow\infty} F^n(a_n x + b_n) = \lim_{n\uparrow\infty} \left(1 - (1-F(a_n x + b_n))\right)^n = e^{-\lim_{n\uparrow\infty} n(1-F(a_n x + b_n))} \tag{2.51}$$

□

Thus we only have to check when (2.46) holds with which $g(x)$.

Let us assume that there is a sequence, γ_n, such that

$$n(1-F(\gamma_n)) \to 1. \tag{2.52}$$

Since necessarily $F(\gamma_n) \to 1$, $\gamma_n \to x_F$, and we may choose $\gamma_n < x_F$, for all n. We now turn to the three cases.

Fréchet: We know that (for $x > 0$),

$$\frac{1-F(\gamma_n x)}{1-F(\gamma_n)} \to x^{-\alpha} \tag{2.53}$$

while $n(1-F(\gamma_n)) \to 1$. Thus,

$$n(1-F(\gamma_n x)) \to x^{-\alpha} \tag{2.54}$$

and so, for $x > 0$,

$$F^n(\gamma_n x) \to e^{-x^{-\alpha}}. \tag{2.55}$$

Since $\lim_{x\downarrow 0} e^{-x^{-\alpha}} = 0$, it must be true that, for $x \leq 0$,

$$F^n(\gamma_n x) \to 0 \tag{2.56}$$

which concludes the argument.

Weibull: Let now $h_n = x_F - \gamma_n$. By the same argument as above, we get, for $x > 0$

$$n(1-F(x_F - h_n x)) \to x^{\alpha} \tag{2.57}$$

and so

$$F^n(x_F - x(x_F - \gamma_n)) \to e^{-x^{\alpha}} \tag{2.58}$$

or equivalently, for $x < 0$,

$$F^n((x_F - \gamma_n)x + x_F) \to e^{-(-x)^{\alpha}} \tag{2.59}$$

Since, for $x \uparrow 0$, the right-hand side tends to 1, it follows that, for $x \geq 0$,

$$F^n(x(x_F - \gamma_n) - x_F) \to 1 \tag{2.60}$$

Gumbel: In exactly the same way we conclude that

$$n(1 - F(\gamma_n + xg(\gamma_n))) \to e^{-x} \tag{2.61}$$

from which the conclusion is now obvious, with $a_n = 1/g(\gamma_n)$, $b_n = \gamma_n$.

We are left with proving the existence of γ_n with the desired property. If F had no jumps, we could choose γ_n simply such that $F(\gamma_n) = 1 - 1/n$ and we would be done. The problem becomes more subtle since we want to allow for more general distribution functions. The best approximation seems to be

$$\gamma_n \equiv F^{-1}(1-1/n) = \inf\{x : F(x) \geq 1 - 1/n)\} \tag{2.62}$$

Then we get immediately that

$$\limsup n(1 - F(\gamma_n)) \leq 1. \tag{2.63}$$

But for $x < \gamma_n$, $F(x) \leq 1 - 1/n$, and so $n(1 - F(\gamma_n^-)) \geq 1$. Thus we may just show that

$$\liminf_n \frac{1 - F(\gamma_n)}{1 - F(\gamma_n^-)} \geq 1. \tag{2.64}$$

This, however, follows in all cases from the hypotheses on the functions F, e.g.

$$\frac{1 - F(x\gamma_n)}{1 - F(\gamma_n)} \to x^{-\alpha} \tag{2.65}$$

which tends to 1 as $x \uparrow 1$. This concludes the proof of the sufficiency in the theorem. □

We mention the following fact that characterises the class of distribution for which there is any possibility to get a limiting distribution for a rescaled maximum.

Theorem 2.1.8 *Let F be a distribution function. Then there exists a sequence, γ_n, such that*

$$n(1 - F(\gamma_n)) \to \tau, \quad 0 < \tau < \infty, \tag{2.66}$$

if and only if

$$\lim_{x \uparrow x_F} \frac{1 - F(x)}{1 - F(x^-)} = 1 \tag{2.67}$$

Remark 2.1.1 To see what is at issue, note that

$$\frac{1 - F(x)}{1 - F(x^-)} = 1 + \frac{p(x)}{1 - F(x^-)}, \tag{2.68}$$

where $p(x)$ is the probability of the "atom" at x, i.e. the size of the jump of F at x. Thus, (2.67) says that the size of jumps of F should diminish faster, as x approaches the upper boundary of the support of F, than the total mass beyond x.

Proof. Assume that (2.66) holds, but

$$\frac{p(x)}{1 - F(x^-)} \not\to 0. \tag{2.69}$$

Then there exists $\epsilon > 0$ and a sequence, $x_j \uparrow x_F$, such that

$$p(x_j) \geq 2\epsilon(1 - F(x_j^-)). \tag{2.70}$$

Now chose n_j such that

$$1 - \frac{\tau}{n_j} \leq \frac{F(x_j^-) + F(x_j)}{2} \leq 1 - \frac{\tau}{n_j + 1}. \tag{2.71}$$

The gist of the argument (given in detail below) is as follows: Since the upper and lower limit in (2.71) differ by only $O(1/n_j^2)$, the term in the middle must equal, up to that error, $F(\gamma_{n_j})$; but $F(x_j)$ and $F(x_j^-)$ differ (by hypothesis) by ϵ/n_j, and since F takes no value between these two, it is impossible that $\frac{F(x_j^-)+F(x_j)}{2} = F(\gamma_{n_j})$ to the precision required. Thus (2.67) must hold.

Let us formalize this argument. Now it must be true that either

(i) $\gamma_{n_j} < x_j$ i.o., or
(ii) $\gamma_{n_j} \geq x_j$ i.o.

In case (i), it holds that for these j,

$$n_j(1 - F(\gamma_{n_j})) > n_j(1 - F(x_j^-)). \tag{2.72}$$

Now replace in the right-hand side

$$F(x_j^-) = \frac{F(x_j^-) + F(x_j) - p(x_j)}{2} \tag{2.73}$$

and write

$$1 = \tau/n_j + 1 - \tau/n_j \tag{2.74}$$

to get

$$\begin{aligned} n_j(1 - F(x_j^-)) &= \tau + n_j \left(1 - \frac{\tau}{n_j} - \frac{F(x_j^-) + F(x_j) - p(x_j)}{2}\right) \\ &\geq \tau + \frac{n_j p(x_j)}{2} - n_j \left(\frac{\tau}{n_j} - \frac{\tau}{n_j + 1}\right) \\ &\geq \tau + \epsilon n_j (1 - F(x_j^-)) - \frac{\tau}{n_j + 1}. \end{aligned}$$

Thus

$$n_j(1 - F(x_j^-)) \geq \tau \frac{1 - 1/(n_j + 1)}{1 - \epsilon}. \tag{2.75}$$

For large enough j, the right-hand side will be strictly larger than τ, so that

$$\limsup_j n_j(1 - F(x_j^-)) > \tau, \tag{2.76}$$

and in view of (2.72), a fortiori

$$\limsup_j n_j(1 - F(\gamma_j^-)) > \tau, \tag{2.77}$$

in contradiction with the assumption.

In case (ii), we repeat the same argument mutando mutandis, to conclude that

$$\liminf_j n_j(1 - F(\gamma_j^-)) < \tau, \tag{2.78}$$

To prove the converse assertion, choose

$$\gamma_n \equiv F^{-1}(1 - \tau/n). \tag{2.79}$$

Using (2.67), one deduces (2.66) exactly as in the special case $\tau = 1$ in the proof of Theorem 2.1.6. □

2.1.2 Distribution of exceedances

The scale function $u_n(x)$ indicates the order of the largest of the first n random variables X_i. The next natural question is *how many* variables are bigger than such an extreme level u_n. We define the number, $S_n(u)$, of exceedances of a level u,

$$S_n(u) \equiv \#\{i \leq n : X_i > u\}. \tag{2.80}$$

Unsurprisingly, the Poisson distribution makes its first entrance:

Theorem 2.1.9 *Let X_i be iid random variables with common distribution F. If u_n is such that*

$$n(1 - F(u_n)) \to \tau, \quad 0 < \tau < \infty, \tag{2.81}$$

then

$$\mathbb{P}[S_n(u_n) = k] \to \frac{\tau^k}{k!} e^{-\tau} \tag{2.82}$$

Proof. The proof of this lemma is quite simple, but the result is important enough for us to justify giving it. We just need to consider all possible ways to realise the event $\{S_n(u_n) = k\}$. Namely

$$\begin{aligned}
\mathbb{P}[S_n(u_n) = k] &= \sum_{\{i_1,\dots,i_k\} \subset \{1,\dots,n\}} \prod_{\ell=1}^{k} \mathbb{P}[X_{i_\ell} > u_n] \prod_{j \notin \{1_i,\dots,i_k\}} \mathbb{P}[X_j \leq u_n] \\
&= \binom{n}{k} (1 - F(u_n))^k F(u_n)^{n-k} \\
&= \frac{1}{k!} \frac{n!}{n^k (n-k)!} \left[n(1 - F(u_n))\right]^k \left[F^n(u_n)\right]^{1-k/n}.
\end{aligned}$$

But, for any k fixed, $n(1 - F(u_n)) \to \tau$, $F^n(u_n) \to e^{-\tau}$, $k/n \to 0$, and $\frac{n!}{n^k(n-k)!} \to 1$. Thus (2.82) holds. □

Using very much the same sort of reasoning, one can generalise the question answered above to that of the numbers of exceedances of several extremal levels.

Theorem 2.1.10 *Let* $u_n^1 > n_n^2 \cdots > u_n^r$ *such that*

$$n(1 - F(u_n^\ell)) \to \tau_\ell,$$

with

$$0 < \tau_1 < \tau_2 < \ldots, < \tau_r < \infty.$$

Then, under the assumptions of the preceding theorem, with $S_n^i \equiv S_n(u_n^i)$,

$$\mathbb{P}\left[S_n^1 = k_1, S_n^2 - S_n^1 = k_2, \ldots, S_n^r - S_n^{r-1} = k_r\right] \to \frac{\tau_1^{k_1}}{k_1!} \frac{(\tau_2 - \tau_1)^{k_2}}{k_2!} \cdots \frac{(\tau_r - \tau_{r-1})^{k_r}}{k_r!} e^{-\tau_r} \tag{2.83}$$

Proof. Again, we just have to count the number of arrangements that will place the desired number of variables in the respective intervals. Then

$$\begin{aligned}
&\mathbb{P}\left[S_n^1 = k_1, S_n^2 - S_n^1 = k_2, \ldots, S_n^r - S_n^{r-1} = k_r\right] \\
&= \binom{n}{k_1, \ldots, k_r} \mathbb{P}\big[X_1, \ldots, X_{k_1} > u_n^1 \geq X_{k_1+1}, \ldots, X_{k_1+k_2} > u_n^2, \ldots \\
&\ldots, u_n^{r-1} \geq X_{k_1+\cdots+k_{r-1}+1}, \ldots, X_{k_1+\cdots+k_r} > u_n^r \geq X_{k_1+\cdots+k_r+1}, \ldots, X_n\big] \\
&= \binom{n}{k_1, \ldots, k_r} (1 - F(u_n^1))^{k_1} \left[F(u_n^1) - F(u_n^2)\right]^{k_2} \ldots \left[F(u_n^{r-1}) - F(u_n^r)\right]^{k_r} \\
&\qquad \times F^{n-k_1-\cdots-k_r}(u_n^r)
\end{aligned}$$

Now we write

$$\left[F(u_n^{\ell-1}) - F(u_n^\ell)\right] = \frac{1}{n}\left[n(1 - F(u_n^\ell)) - n(1 - F(u_n^{\ell-1}))\right]$$

and use that $\left[n(1 - F(u_n^\ell)) - n(1 - F(u_n^{\ell-1}))\right] \to \tau_\ell - \tau_{\ell-1}$. Proceeding otherwise as in the proof of Theorem 2.1.9, we arrive at (2.83) □

Note that the theorem states that the number of variables that fall into the different intervals are independent and Poisson distributed with Parameters $\tau_k - \tau_{k-1}$.

2.1.3 Poisson point processes

We want to reformulate the result of the previous theorem in an even more appealing way. Suppose that we are in the situation when F is in the domain of attraction of one of the three extremal distributions, and hence

$$n\mathbb{P}\left[X_i > u_n(x)\right] \to g(x), \tag{2.84}$$

with $g(x)$ monotone (decreasing) function. Then the preceding theorem says that the number of exceedances of the level $u_n(x)$ is Poisson with parameter $g(x)$, for any x. We may want to look at the collection of points, $(u_n^{-1}(X_i), i \leq n)$. The argument from above can easily be extended to show that for any two disjoint intervals $I, J \subset \mathbb{R}$, the number of points, $u_n^{-1}(X_i)$, that fall into the intervalls I and J, resp., converge to mutually independent Poisson random variables with parameters $-\int_I dg(x)$ and $-\int_J dg(x)$, respectively. This brings us to the formulation in terms of Poisson point processes. Let us recall some basic definitions and facts.

Point processes. Point processes are designed to describe the probabilistic structure of point sets in metric spaces, for our purposes $\mathbb{R}^d$. For reasons that may not be obvious immediately, a convenient way to represent a collection of points x_i in $\mathbb{R}^d$ is by associating to them a *point measure*.

Let us first consider a single point x. We consider the usual Borel-sigma algebra, $\mathcal{B} \equiv \mathcal{B}(\mathbb{R}^d)$, of $\mathbb{R}^d$, that is generated form the open sets in the open sets in the Euclidean topology of $\mathbb{R}^d$. Given $x \in \mathbb{R}^d$, we define the Dirac measure, δ_x, such that, for any Borel set $A \in \mathcal{B}$,

$$\delta_x(A) = \begin{cases} 1, & \text{if}\, x \in A \\ 0, & \text{if}\, x \notin A. \end{cases} \tag{2.85}$$

A *point measure* is a measure, μ, on $\mathbb{R}^d$, such that there exists a countable collection of points , $\{x_i \in \mathbb{R}^d, i \in \mathbb{N}\}$, such that

$$\mu = \sum_{i=1}^{\infty} \delta_{x_i} \tag{2.86}$$

and, if K is compact, then $\mu(K) < \infty$.

Note that the points x_i need not be all distinct. The set $S_\mu \equiv \{x \in \mathbb{R}^d : \mu(x) \neq 0\}$ is called the support of μ. A point measure such that for all $x \in \mathbb{R}^d$, $\mu(x) \leq 1$ is called simple.

We denote by $M_p(\mathbb{R}^d)$ the set of all point measures on $\mathbb{R}^d$. We equip this set with th sigma-algebra $\mathcal{M}_p(\mathbb{R}^d)$, the smallest sigma algebra that contains all subsets of $M_p(\mathbb{R}^d)$ of the form $\{\mu \in M_p(\mathbb{R}^d) : \mu(F) \in B\}$, where $F \in \mathcal{B}(\mathbb{R}^d)$ and $B \in \mathcal{B}([0,\infty))$. $\mathcal{M}_p(\mathbb{R}^d)$ is also characterized by saying that it is the smallest sigma-algebra that makes the evaluation maps, $\mu \to \mu(F)$, measurable for all Borel sets $F \in \mathcal{B}(\mathbb{R}^d)$.

Definition 2.1.2 A *point process*, N, is a random variable taking values in $M_p(\mathbb{R}^d)$, i.e. a measurable map, $N : (\Omega, \mathcal{F}) \to (M_p(\mathbb{R}^d), \mathcal{M}_p(\mathbb{R}^d))$, from a probability space to the space of point measures.

An important characteristic of a point process is its *intensity measure*, λ, defined as

$$\lambda(F) \equiv \mathbb{E}N(F) \tag{2.87}$$

for $F \in \mathcal{B}$.

If Q is a probability measure on $(M_p, \mathcal{M}_p)$, the *Laplace transform* of Q is a map, ψ from non-negative Borel functions on $\mathbb{R}^d$ to $\mathbb{R}_+$, defined as

$$\psi(f) \equiv \int_{M_p} \exp\left(-\int_{R^d} f(x)\mu(dx)\right) Q(d\mu). \tag{2.88}$$

If N is a point process, the *Laplace functional* of N is

$$\begin{aligned} \psi_N(f) \equiv \mathbb{E}e^{-N(f)} &= \int e^{-N(\omega,f)} \mathbb{P}(d\omega) \\ &= \int_{M_p} \exp\left(-\int_{R^d} f(x)\mu(dx)\right) P_N(d\mu), \end{aligned} \tag{2.89}$$

where we denote by P_N the law of N.

Proposition 2.1.11 *The Laplace functional, ψ_N, of a point process, N, determines N uniquely.*

Poisson point process. The most important class of point processes for our purposes will be Poisson point processes. One may characterize it as the "most random" point process with given intensity measure.

Definition 2.1.3 Let λ be a σ-finite, positive measure on $\mathbb{R}^d$. Then a point process, N, is called a Poisson point process with intensity measure λ ($PPP(\lambda)$), if

(i) For any $F \in \mathcal{B}(\mathbb{R}^d)$, and $k \in \mathbb{N}$,

$$\mathbb{P}[N(F) = k] = \begin{cases} e^{-\lambda(F)} \frac{(\lambda(F))^k}{k!}, & \text{if } \lambda(F) < \infty \\ 0, & \text{if } \lambda(F) = \infty, \end{cases} \tag{2.90}$$

(ii) If $F, G \in \mathcal{B}$ are disjoint sets, then $N(F)$ and $N(G)$ are independent random variables.

In the next theorem we will assert the existence of a Poisson point process with any desired intensity measure. We will give the proof, because it contains an explicit construction of Poisson point processes in terms of independent random variables which is extremely useful.

Proposition 2.1.12*(i) $PPP(\lambda)$ exists, and its law is uniquely determined by the requirements of the definition.*

(ii) The Laplace functional of $PPP(\lambda)$ is given, for $f \geq 0$, by

$$\Psi_N(f) = \exp\left(-\int_{\mathbb{R}^d} (1 - e^{-f(x)})\lambda(dx)\right) \tag{2.91}$$

Proof. Since we know that the Laplace functional determines a point process, in order to prove that the conditions of the definition uniquely determine the $PPP(\lambda)$, we show that they determine the form (2.91) of the Laplace functional. Thus suppose that N is a $PPP(\lambda)$. Let $f = c\mathbb{1}_F$. Then

$$\begin{aligned}\Psi_N(f) &= \mathbb{E}\exp(-N(f)) = \mathbb{E}\exp(-cN(F)) \\ &= \sum_{k=0}^{\infty} e^{-ck} e^{-\lambda(F)} \frac{(\lambda(F))^k}{k!} = e^{(e^{-c}-1)\lambda(F)} \\ &= \exp\left(-\int (1 - e^{-f(x)})\lambda(dx)\right),\end{aligned} \tag{2.92}$$

which is the desired form. Next, if F_i are disjoint, and $f = \sum_{i=1}^k c_i \mathbb{1}_{F_i}$, it is straightforward to

$$\Psi_N(f) = \mathbb{E}\exp\left(-\sum_{i=1}^k c_i N(F_i)\right) = \prod_{i=^1}^k \mathbb{E}\exp(-c_i N(F_i)) \tag{2.93}$$

due to the independence assumption (ii); a simple calculations shows that this yields again the desired form. Finally, for general f, we can choose a sequence, f_n, of the form considered, such that $f_n \uparrow f$. By monotone convergence then $N(f_n) \uparrow N(f)$. On the other hand, since $e^{-N(g)} \leq 1$, we get from dominated convergence that

$$\Psi_N(f_n) = \mathbb{E}e^{-N(f_n)} \to \mathbb{E}e^{-N(f)} = \Psi_N(f). \tag{2.94}$$

But, since $1 - e^{-f_n(x)} \uparrow 1 - e^{-f(x)}$, and monotone convergence gives once more

$$\Psi_N(f_n) = \exp\left(\int (1 - e^{-f_n(x)})\lambda(dx)\right) \uparrow \exp\left(\int (1 - e^{-f(x)})\lambda(dx)\right) \tag{2.95}$$

On the other hand, given the form of the Laplace functional, it is trivial to verify that the conditions of the definition hold, by choosing suitable functions f.

Finally we turn to the *construction* of $PPP(\lambda)$. Let us first consider the case $\lambda(\mathbb{R}^d) < \infty$. Then construct

(i) A Poisson random variable, τ, of parameter $\lambda(\mathbb{R}^d)$.
(ii) A family, X_i, $i \in \mathbb{N}$, of independent, $\mathbb{R}^d$-valued random variables with common distribution $\lambda/\lambda(R^d)$. This family is independent of τ.

Then set

$$N^* \equiv \sum_{i=1}^{\tau} \delta_{X_i} \tag{2.96}$$

It is not very hard to verify that N^* is a $PPP(\lambda)$.

To deal with the case when $\lambda(\mathbb{R}^d)$ is infinite, decompose λ into a countable sum of finite measures, λ_k, that are just the restriction of λ to a finite set F_k, where the F_k form a partition of $\mathbb{R}^d$. Then N^* is just the sum of independent $PPP(\lambda_k)$ N_k^*. □

2.1.4 Poisson processes of extremes

We can now formulate our first theorem on Poisson convergence.

Theorem 2.1.13 *Let X_i be iid random variables with common distribution function F, and assume $n(1-F(u_n(x))) \to g(x)$ for some monotone decreasing function $g : \Gamma \to \mathbb{R}_+$. Then the point process*

$$\mathcal{P}_n \equiv \sum_{i=1}^{n} \delta_{u_n^{-1}(X_i)} \tag{2.97}$$

converges weakly to the Poisson Point process on Γ with intensity measure $-dg(x)$.

To understand the notion of convergence used here, we need to discuss briefly some topological notions.

We will say that a sequence of measures, $\mu_n \in M_+(\mathbb{R}^d)$ converges *vaguely* to a measure $\mu \in M_+(\mathbb{R}^d)$, if, for all $f \in C_0^+(\mathbb{R}^d)$,

$$\mu_n(f) \to \mu(f) \tag{2.98}$$

Note that for this topology, typical open neighborhoods are of the form

$$B_{f_1,\dots,f_k,\epsilon}(\mu) \equiv \{\nu \in M_+(\mathbb{R}^d) : \forall_{i=1}^k |\nu(f_i) - \mu(f_i)| < \epsilon\}, \tag{2.99}$$

i.e. to test the closeness of two measures, we test it on their expectations on finite collections of continuous, positive functions with compact support. It is a fact that we shall not prove here that the vague topology is metrizable and turns $M_+(\mathbb{R}^d)$ into a Polish space. Given this topology, on can define the corresponding Borel sigma algebra, $\mathcal{B}(M_+(\mathbb{R}^d))$, which (fortunately) turns out to coincide with the sigma algebra $\mathcal{M}_+(\mathbb{R}^d)$ introduced before.

Having established the space of σ-finite measures as a complete, separable metric space, we can think of weak convergence of probability measures on this space just as if we were working on an Euclidean space.

Vague convergence of point measures is a very strong notion in the sense that it implies the convergence of the points of the supports. The following proposition makes this precise.

Proposition 2.1.14 *Let μ_n, $n \in \mathbb{N}$, and μ be in $M_p(\mathbb{R}^d)$, and $\mu_n \stackrel{v}{\to} \mu$. Let K be a compact set with $\mu(\partial K) = 0$. Then we have a labeling of the points of μ_n, for $n \geq n(K)$ large enough, such that*

$$\mu_n(\cdot \cap K) = \sum_{i=1}^p \delta_{x_i^n}, \quad \mu(\cdot \cap K) = \sum_{i=1}^p \delta_{x_i},$$

such that $(x_1^n, \dots, x_p^n) \to (x_1, \dots, x_p)$.

A particularly useful criterion for convergence of point processes is provided by *Kallenberg's theorem* [29].

Theorem 2.1.15 *Assume that ξ is a simple point process on a metric space E, and $\mathcal{T}$ is a Π-system of relatively compact open sets, and that for $I \in \mathcal{T}$,*

$$\mathbb{P}[\xi(\partial I) = 0] = 1. \tag{2.100}$$

If ξ_n, $n \in \mathbb{N}$ are point processes on E, and for all $I \in \mathcal{T}$,

$$\lim_{n\uparrow\infty} \mathbb{P}[\xi_n(I) = 0] = \mathbb{P}[\xi(I) = 0], \tag{2.101}$$

and

$$\lim_{n\uparrow\infty} \mathbb{E}\xi_n(I) = \mathbb{E}\xi(I) < \infty, \tag{2.102}$$

then

$$\xi_n \stackrel{w}{\to} \xi \tag{2.103}$$

Remark 2.1.2 The Π-system, $\mathcal{T}$, can be chosen, in the case $E = \mathbb{R}^d$, as finite unions of bounded rectangles.

We will give a proof of Kallenberg's theorem following Resnick's book [32] in Chapter 5.

Kallenberg's theorem gives us the tool to prove Theorem 2.1.13.

Proof. Note first that for any interval $(a, b] \subset \mathbb{R}_+$,

$$\mathbb{E}\mathcal{P}_n((a,b]) = n\mathbb{P}\left[u_n^{-1}(X_1) \in (a,b]\right] = n(F(u_n(b)) - F(u_n(a))) \to g(a) - g(b). \tag{2.104}$$

Next we consider the probability that $\mathcal{P}_n((a,b]) = 0$. Clearly, from what we have said at the beginning of the subsection, $\mathcal{P}_n((a,b])$ converges to a Poisson random variable with parameter $g(a) - g(b)$,

$$\mathbb{P}\left[\mathcal{P}_n((a,b]) = 0\right] \to \exp(g(b) - g(a)). \tag{2.105}$$

The corresponding result for finite collections of intervals follows in the same way, and since the limits correspond to the Poisson point process with intensity measure $-dg$, we are done. □

We can do better: suppose we want to know something about the places where large values are attained. Then it makes sense to define

$$\mathcal{N}_n \equiv \sum_{i=1}^{\infty} \delta_{(i/n, u_n^{-1}(X_i))} \tag{2.106}$$

as a point process on $\mathbb{R}^2$ (or more precisely, on $\Gamma \times \mathbb{R}_+$).

Theorem 2.1.16 *Under the assumptions of Theorem 2.1.13, the point process $\mathcal{N}_n$ converges to the Poisson point process, $\mathcal{N}$, on $\Gamma \times \mathbb{R}_+$ with intensity measure $(-dg)(x) \times dt$.*

Proof. The proof is the same as before, just taking into account the independence of the X_i! □

This last result gives the most complete description of the asymptotic structure of extremes of iid sequences of random variables.

2.2 Extensions to dependent sequences and triangular arrays

The type of results on extremes we have exposed above is fortunately not limited to iid sequences. In fact, there is a rather remarkable degree of universality present that allows to push these observations to fairly strongly correlated sequences. For our purposes it will also be important that we will be able to deal with random variables whose distribution changes as the sample size changes, i.e. we want to deal with *triangular arrays* of random variables. There are several settings in which creteria are known. A rather general one is the following.

2.2.1 The inclusion-exclusion principle.

Going through the proofs in the iid setting, one will find that a key observation was that

$$n(1 - F(u_n)) \to \tau \quad \Leftrightarrow \quad \mathbb{P}\left[M_n \leq u_n\right] \to e^{-\tau}. \tag{2.107}$$

This relation was instrumental for the Poisson distribution of the number of crossings of extreme levels. The key to this relation was the fact that in the iid case,

$$\mathbb{P}\left[M_n \leq u_n\right] = F^n(u_n) = \left(1 - \frac{n(1 - F(u_n))}{n}\right)^n \tag{2.108}$$

which converges to $e^{-\tau}$. The first equality fails of course in the dependent case. However, this equation is also far from necessary.

The following simple lemma gives a much weaker, and, as we will see, useful, criterium for convergence to the exponential function.

Lemma 2.2.17 *Assume that a sequence A_n satisfies, for any $s \in \mathbb{N}$, the bounds*

$$A_n \leq \sum_{\ell=0}^{2s} \frac{(-1)^\ell}{\ell!} a_\ell(n) \tag{2.109}$$

$$A_n \geq \sum_{\ell=0}^{2s+1} \frac{(-1)^\ell}{\ell!} a_\ell(n) \tag{2.110}$$

and, for any $\ell \in \mathbb{N}$,

$$\lim_{n \uparrow \infty} a_\ell(n) = a^\ell \tag{2.111}$$

Then

$$\lim_{n \uparrow \infty} A_n = e^{-a} \tag{2.112}$$

Proof. Obviously the hypothesis of the lemma imply that, for all $s \in \mathbb{N}$,

$$\limsup_{n\uparrow\infty} A_n \le \sum_{\ell=0}^{2s} \frac{(-a)^\ell}{\ell!} \tag{2.113}$$

$$\liminf_{n\uparrow\infty} A_n \ge \sum_{\ell=0}^{2s+1} \frac{(-a)^\ell}{\ell!} \tag{2.114}$$

But the upper and lower bounds are the partial series of the exponential function e^{-a}, which are absolutely convergent, and this implies convergence of A_n to these values. □

The reason that one may expect $\mathbb{P}[M_n \le u_n]$ to satisfy bounds of this form lies in the *inclusion-exclusion principle:*

Theorem 2.2.18 *Let $\mathcal{B}_i$, $i \in \mathbb{N}$ be a sequence of events, and let $\mathbb{1}_{\mathcal{B}}$ denote the indicator function of $\mathcal{B}$. Then, for all $s \in \mathbb{N}$,*

$$\mathbb{1}_{\cap_{i=1}^n \mathcal{B}_i} \le \sum_{\ell=0}^{2s} (-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{1}_{\cap_{r=1}^\ell \mathcal{B}_{j_r}^c} \tag{2.115}$$

$$\mathbb{1}_{\cap_{i=1}^n \mathcal{B}_i} \ge \sum_{\ell=0}^{2s+1} (-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{1}_{\cap_{r=1}^\ell \mathcal{B}_{j_r}^c} \tag{2.116}$$

Note that terms with $\ell > n$ are treated as zero.

Remark 2.2.1 Note that the sum over subsets $\{i_1,\dots,i_\ell\}$ is over all ordered subsets, i.e., $1 \le i_1 < i_2 < \cdots < i_\ell \le n$.

Proof. We write first

$$\mathbb{1}_{\cap_{i=1}^n \mathcal{B}_i} = 1 - \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} \tag{2.117}$$

We will prove the theorem by induction over n. The key observation is that

$$\begin{aligned} \mathbb{1}_{\cup_{i=1}^{n+1} \mathcal{B}_i^c} &= \mathbb{1}_{\mathcal{B}_{n+1}^c} + \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} \mathbb{1}_{\mathcal{B}_{n+1}} \\ &= \mathbb{1}_{\mathcal{B}_{n+1}^c} + \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} - \mathbb{1}_{\cup_{i=1}^n \mathcal{B}_i^c} \mathbb{1}_{\mathcal{B}_{n+1}^c} \end{aligned} \tag{2.118}$$

To prove an upper bound of some $2s+1$, we now insert an upper bound of that order in the second term, and a lower bound of order $2s$ in the third term. It is a simple matter of inspection that this reproduces exactly the desired bounds for $n+1$. □

The inclusion-exclusion principle has an obvious corollary.

Corollary 2.2.19 *Let X_i be any sequence of random variables. Then*

$$\mathbb{P}\left[M_n \leq u\right] \leq \sum_{\ell=0}^{2s}(-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r} > u\right] \tag{2.119}$$

$$\mathbb{P}\left[M_n \leq u\right] \geq \sum_{\ell=0}^{2s+1}(-1)^\ell \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r} > u\right] \tag{2.120}$$

Proof. The proof is straightforward. □

Note that the statement of the corollary is a statement for any fixed n. Thus we may as well consider families of random variables whose distribution changes with n. Hence, combining Lemma 2.2.17 and Corollary 2.2.19, we obtain a quite general criterion for triangular arrays of random variables [16].

Theorem 2.2.20 *Let X_i^n, $n \in \mathbb{N}$, $i \in \{1,\dots,n\}$ be a triangular array of random variables. Assume that, for any ℓ,*

$$\lim_{n\uparrow\infty} \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_{j_r}^n > u_n\right] = \frac{\tau^\ell}{\ell!} \tag{2.121}$$

Then, with $M_n \equiv \max_{i=1}^n X_i^n$,

$$\lim_{n\uparrow\infty} \mathbb{P}\left[M_n \leq u_n\right] = e^{-\tau} \tag{2.122}$$

Proof. The proof of the theorem is straightforward. □

Remark 2.2.2 In the iid case, (2.121) does of course hold, since here

$$\sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,n\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X_r > u_n\right] = \binom{n}{\ell} n^{-\ell}\left(n(1-F(u_n))\right)^\ell \tag{2.123}$$

A special case where Theorem 2.2.20 gives an easily verifiable criterion is the case of *exchangeable random variables.*

Corollary 2.2.21 *Assume that X_i^n is a triangular array of random variables such that, for any n, the joint distribution of $X_1^n,\dots,X_n^n$ is invariant under permutation of the indices $i,\dots,n$. If, for any $\ell \in \mathbb{N}$,*

$$\lim_{n\uparrow\infty} n^\ell \mathbb{P}\left[\forall_{r=1}^{\ell} X_r^n > u_n\right] = \tau^\ell \tag{2.124}$$

Then,

$$\lim_{n\uparrow\infty} \mathbb{P}\left[M_n \leq u_n\right] = e^{-\tau} \tag{2.125}$$

Proof. Again straightforward. □

Theorem 2.2.20 and its corollary have an obvious extension to the distribution the number of exceedances of extremal levels.

Theorem 2.2.22 *Let $u_n^1 > u_n^2 \cdots > u_n^r$, and let X_i^n, $n \in \mathbb{N}$, $i \in \{1, \dots, n\}$ be a triangular array of random variables. Assume that, for any $\ell \in \mathbb{N}$, and any $1 \leq s \leq r$,*

$$\lim_{n \uparrow \infty} \sum_{\{j_1, \dots, j_\ell\} \subset \{1, \dots, n\}} \mathbb{P}\left[\forall_{m=1}^{\ell} X_{j_m}^n > u_n^s\right] = \frac{\tau_s^\ell}{\ell!} \tag{2.126}$$

with

$$0 < \tau_1 < \tau_2 \dots, < \tau_r < \infty. \tag{2.127}$$

Then,

$$\begin{aligned} &\lim_{n \uparrow \infty} \mathbb{P}\left[S_n^1 = k_1, S_n^2 - S_n^1 = k_2, \dots, S_n^r - S_n^{r-1} = k_r\right] \\ &= \frac{\tau_1^{k_1}}{k_1!} \frac{(\tau_2 - \tau_1)^{k_2}}{k_2!} \dots \frac{(\tau_r - \tau_{r-1})^{k_r}}{k_r!} e^{-\tau_r} \end{aligned} \tag{2.128}$$

Proof. The general case is notationally very involved and we will not write the details down. The main idea can be seen in the case $r = 1$. As in the iid case, we write first

$$\mathbb{P}\left(S_n(u_n) = k\right) = \sum_{\{j_1, \dots, j_k\}} \mathbb{P}\left(\forall_{m=1}^{k} X_{j_m} > u_n, \forall_{i \notin \{j_1, \dots, j_r\}} X_i \leq u_n\right). \tag{2.129}$$

Then use the inclusion-exclusion principle on the indicator function

$$\mathbb{1}_{\cap_{i \notin \{j_1, \dots, j_r\}} \{X_i \leq u_n\}}. \tag{2.130}$$

and insert the resulting inequalities into (2.129). It then follows in complete analogy to the proof of 2.2.20, that

$$\mathbb{P}\left((S_n(u_n) = k\right) \to \frac{\tau^k}{k!} e^{-\tau}. \tag{2.131}$$

□

The results on the convergence to Poissonprocesses from the iid case also carry over under the assumptions above. We only need a slight strengthening of the hypothesis.

Theorem 2.2.23 *Let X_i^n, $n \in \mathbb{N}$, $i \in \{1, \dots, n\}$ be a triangular array of random variables, and assume that there is a non-increasing function, $u_n(\tau)$, such that, for any interval $I \subset \mathbb{R}_+$,*

$$\lim_{n\uparrow\infty} \sum_{\{j_1,\dots,j_\ell\}} \mathbb{1}_{j_i/n\in I, \forall i=1,\dots,\ell} \mathbb{P}\left[\forall_{m=1}^{\ell} X_{j_m}^n > u_n(\tau)\right] = \frac{|I|^\ell \tau^\ell}{\ell!}. \quad (2.132)$$

Then,

$$\sum_{i\in\mathbb{N}} \delta_{(i/n, u_n^{-1}(X_i))} \to \mathcal{P}, \quad (2.133)$$

where $\mathcal{P}$ is the Poisson point process on $\mathbb{R}_+^2$ with intensity measure the Lebesgue measure.

Remark 2.2.3 The hypothesis of this theorem are almost necessary in the following sense: if for some ℓ (2.132) fails to hold while the limit is finite, then we cannot have convergence to a Poisson process.

Another useful criterion was established by Durrett and Resnick [19]. It applies well if there is some Markovian structure in the data.

Theorem 2.2.24 *Let X_i^n be a triangular array of random variables with support in $\mathbb{R}_+$. Let ν be a sigma-finite measure on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, such that, for some $x_0 \in [0,\infty)$, $\nu([x_0,\infty)) < \infty$. Assume that, for a given sequence, a_n, at all non-atoms of ν, for all $t < T < \infty$, in probability,*

$$\sum_{i=1}^{[a_n t]} \mathbb{P}\left(X_i^n > x | \mathcal{F}_{n,i-1}\right) = t\nu((x,\infty)), \quad (2.134)$$

and

$$\sum_{i=1}^{[a_n t]} \left[\mathbb{P}\left(X_i^n > x | \mathcal{F}_{n,i-1}\right)\right]^2 = 0, \quad (2.135)$$

then

$$\sum_{i\in\mathbb{N}} \delta_{(X_i^n, i/a_n)} \to \mathcal{P}_\nu. \quad (2.136)$$

Proof. The prove of this theorem is not very hard and in fact similar to the margingale central limit theorem. The key step as usual is to show that $\mathbb{P}\left(\max_{i=1}^{a_n} X_i^n \leq x\right) \to e^{-\nu(x,\infty}$. One starts by writing

$$1 = \mathbb{E}\left(\prod_{i=1}^{a_n} \mathbb{1}_{X_i^n \leq x} e^{\ln \mathbb{P}(X_i^n \leq x | \mathcal{F}_{n,i-1})}\right), \quad (2.137)$$

which follows by computing succesively the conditional expectations. Then it follows from the hypothesis of the theorem that

$$\exp\left(\sum_{i=1}^{a_n} \ln \mathbb{P}\left(X_i^n \leq x | \mathcal{F}_{n,i-1}\right)\right) \rightarrow \exp\left(\nu(x,\infty)\right), \tag{2.138}$$

in probability. Pulling this term out of the expectation in (2.137) yields the claimed estimate. We leave the rest of the proof to the reeder. □

Remark 2.2.4 Note the slightly different formulations of the result: you can pass from the first to the second by inserting $\tau = \nu((x,\infty))$ and then replacing X_i^n by the random variable $(u_n \circ \nu)^{-1}(X_i^n)$. Notice that in the case of triangular arrays, the extremal types theorem does not apply. Therefore we do not have an a-priori choice for the measures ν.

3

Lévy processes

Poisson point processes are very intimately related to Lévy processes. This will be the content of this section. An excellent presentation of these *Lévy processes* was given by Kiyosi Itô in his Aarhus lectures [27]. Another good reference is Bertoin's book [9].

3.1 Definition and classification

There are several way we can approach the concept of Lévy processes. In a way we want to see them as continuous time analogs of sums of independent random variables, and hence as candidates for stochastic process limits of sums of random variables.

Definition 3.1.1 A stochastic process $(X_t, t \in \mathbb{R}_+)$ with values in $\mathbb{R}^d$ is called a Lévy process, if:

(i) X_t is a càdlàg process;
(ii) For any collection $0 = t_0 < t_1 < t_2 \cdots < t_k < \infty$, the family of random variables

$$Y_i \equiv X_{t_i} - X_{t_{i-1}}, \quad i = 1, \ldots, k$$

is independent;
(iii) For any $h > 0$, the law of $X_{t+h} - X_t$ is independent of t.

There are two facts that we will be interested in. First, there is a full classification of all Lévy processes. This is directly linked to the theory of infinitely divisible laws, and we will briefly discuss this issue below. The second is an explicit construction of all discontinuous Lévy processes via Poisson point processes. This is, in fact our main concern, as it will link the theory of sums of random variables to the theory of extremes.

We start with some basic facts about infinitely divisible laws.

Definition 3.1.2 A probability measure on $\mathbb{R}^d$ is called *infinitely divisible* if, for each n, there exists a probability measure, μ_n, on $\mathbb{R}^d$, such that, if V_i are independent random variables with law μ_n, then the law of $\sum_{i=1}^n V_i$ is μ.

The connection with Lévy processes is apparent, since clearly the law of X_t is infinitely divisible, being the law of the sum of iid random variables $Y_i \equiv X_{it/n} - X_{(i-1)t/n}$. Note that the Gaussian distribution is infinitely divisible, and that Brownian motion is the corresponding Lévy process.

The following famous *Lévy-Khintchine theorem* gives a complete characterization of infinitely divisible laws. We will state it without proof, but give the proof in a special case.

Theorem 3.1.1 *For each $b \in \mathbb{R}^d$, and non-negative definite matrix M, and each measure, ν, on $\mathbb{R}^d \backslash \{0\}$, that satisfies*

$$\int \min(|x|^2, 1)\nu(dx) < \infty, \tag{3.1}$$

the function

$$\phi(\theta) \equiv \exp(\psi(\theta)),$$

where

$$\psi(\theta) \equiv i(b, \theta) - \frac{1}{2}(\theta, M\theta) + \int (e^{i(\theta,x)} - 1 - i(\theta, x)\mathbb{1}_{|x|\leq 1})\nu(dx), \tag{3.2}$$

is the characteristic function of an infinitely divisible distribution. Moreover, the characteristic function of any infinitely divisible law can be written in this form with uniquely determined (b, M, ν).

Note that it is easy to see that any law of the form given above is infinitely divisible. Namely, for any $n \in \mathbb{N}$, consider the function

$$\psi_n(\theta) \equiv \frac{1}{n}\psi(\theta).$$

Then ϕ_n corresponds to a Lévy triple $(b/n, M/n, \nu/n)$, and if X_i are iid with characteristic function $\exp(\psi_n(\theta))$, then $\sum_{i=1}^n X_i$ has the characteristic function ϕ.

In the case of distributions that take values on the positive reals only, one has the following alternative result.

Theorem 3.1.2 *Let F be a distribution function on $\mathbb{R}_+$. Then F is the distribution function of an infinitely divisible law, iff, for $\lambda \geq 0$,*

$$\int_0^\infty e^{-\lambda x} F(dx) = \exp\left[-c\lambda - \int_0^\infty \left(1 - e^{-x\lambda}\right) \mu(dx)\right], \tag{3.3}$$

where $c \geq 0$ and μ is a measure on $(0, \infty)$ such that

$$\int_0^\infty (x \wedge 1)\mu(dx) < \infty. \tag{3.4}$$

The proofs of the Lévy-Khintchine theorems are purely analytic and will not be of interest to us here.

The description of infinitely divisible laws in terms of the (Lévy) triplets (b, M, ν) is called the Lévy-Khintchine representation, ν is called the Lévy measure, and ψ the characteristic exponent.

We now use the Lévy-Khintchine representation to study Lévy processes. Since $X_t = \sum_{i=1}^t Y_i$ where Y_i has the same law as X_1 (assume $t \in \mathbb{N}$ for a moment), we should expect that

$$\mathbb{E} \exp\left(i(\theta, X_t)\right) = \exp\left(t\psi(\theta)\right) \tag{3.5}$$

where ψ is the characteristic exponent of the distribution of X_1. In fact, for any infinitely divisible law, (3.5) provides a characteristic function of a process with independent and stationary increments.

The Lévy-Khintchine formula implies that any Lévy process is the sum of three independent process: a deterministic drift, a Brownian motion, and a process related to the measure ν; the latter will be most interesting for us. We will see that this is a pure jump process.

3.2 Lévy processes from Poisson counting processes

To start, let note that an important example of Lévy processes can be constructed from Poisson counting processes. Let N_t be a Poisson counting process, and let $Y_i, i \in \mathbb{N}$ be iid real random variables with distribution function F. Then define

$$X_t \equiv \sum_{i=1}^{N_t} Y_i.$$

Clearly X has càdlàg paths and independent increments (both the increments of N_t and the accumulated $Y's$ are independent). Moreover, it is easy to compute the characteristic function of $X_{t+s} - X_t$:

$$\mathbb{E}e^{i(\theta,X_{t+s}-X_t)} = \sum_{n=0}^{\infty} \frac{s^n e^{-s}}{n!} \left(\int e^{i(\theta,x)} F(dx) \right)^n \tag{3.6}$$
$$= \exp\left(s \int \left(e^{i(\theta,x)} - 1 \right) F(dx) \right)$$
$$= \exp\Bigg(si \left(\theta, \int_{|x|\leq 1} xF(dx) \right)$$
$$+ s \int \left(e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\leq 1} \right) F(dx) \Bigg)$$

Thus X is a Lévy process, called a *compound Poisson process* with Lévy triple $(\int_{\|x\|\leq 1} xF(dx), 0, F)$, where the Lévy measure is finite.

Compound Poisson processes are of course pure jump processes, i.e. the only points of change are discontinuities. We will, as an application, show that a non-trivial Lévy measure always makes a Lévy process discontinuous, i.e. produces jumps. This is the content of Lévy's theorem:

Theorem 3.2.3 *If X is a Lévy process with continuous paths, then it's Lévy triple is of the form $(b, M, 0)$, i.e.*

$$X_t = MB_t + bt,$$

where B_t is Brownian motion.

Proof. Let X_t be a Lévy process with triple (b, M, ν). Fix $\epsilon \in (0,1)$ and construct an independent Lévy process with characteristic exponent

$$\psi_\epsilon(\theta) \equiv i(b,\theta) - \frac{1}{2}(\theta, M\theta) + \int_{|x|\leq\epsilon} (e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\leq 1})\nu(dx).$$

Finally set $\psi^\epsilon(\theta) \equiv \psi(\theta) - \psi_\epsilon(\theta)$, i.e.

$$\psi^\epsilon(\theta) = \int_{|x|>\epsilon} (e^{i(\theta,x)} - 1 - i(\theta,x)\mathbb{1}_{|x|\leq 1})\nu(dx).$$

Due to the integrability assumption of Lévy measures, $\int_{|x|>\epsilon} \nu(dx) < \infty$, and therefore, the process Y^ϵ with characteristic exponent ψ_ϵ is a compound Poisson process, and as such has only finitely many jumps on any compact interval. If X^ϵ is the process with exponent ψ^ϵ, independent of Y^ϵ, then $X^\epsilon + Y^\epsilon$ have the same law as X. Now X^ϵ has only countably many jumps, that occur at times independent of the process Y^ϵ. But this means that, with probability one, all the jumps of Y^ϵ occur at times when there is no jump of X^ϵ, and whence X jumps whenever Y^ϵ jumps. But this means that X cannot be continuous, unless the process Y^ϵ never jumps, which is only the case if $\nu = 0$. This proves the theorem.

□

3.3 Lévy processes and Poisson processes

A slightly different look at the construction of compound Poisson processes will provide us with the means to construct general Lévy processes with pure jump part. For notational simplicity we consider only the case of Lévy processes with values in $\mathbb{R}$. To this end, let ν be any measure on $\mathbb{R}$ that satisfies the integrability condition (3.1). For $\epsilon > 0$, set $\nu^\epsilon(dx) \equiv \nu(dx)\mathbb{1}_{|x|>\epsilon}$. Then ν^ϵ is a finite measure. Define the measures on $\lambda_\epsilon(dx, dt) \equiv \nu^\epsilon(dx)dt$ be a measure on $\mathbb{R}^2$. Then we can associate to λ_ϵ a Poisson process, $\mathcal{P}_\epsilon$, on $\mathbb{R}^2$ with intensity measure λ_ϵ. Clearly, for any $\epsilon > 0$, and any $t < \infty$, $\nu^\epsilon((0, t] \times \mathbb{R}) < \infty$. Thus we can define the functions

$$X^\epsilon(t) \equiv \int_0^t \int x\mathcal{P}^\epsilon(ds, dx). \tag{3.7}$$

Note that this is nothing but a random finite sum, and in fact, up to a time change, a compound Poisson process (with Y distributed according to the normalization of the measure ν^ϵ). Now we may ask whether the limit $\epsilon \downarrow 0$ of these processes exists as a Lévy process. To do this, we would like to argue that

$$\int_0^t \int x\mathcal{P}(ds, dx) = \int_0^t \int x\mathcal{P}^\epsilon(ds, dx) + \int_0^t \int_{|x|<\epsilon} x\mathcal{P}(ds, dx)$$

and that the second integral tends to zero as $\epsilon \downarrow 0$. A small problem with this is that we cannot be sure under our conditions on ν that the second term is well defined, since we do not assume that

$$\mathbb{E}\left|\int_0^t \int_{|x|<\epsilon} x\mathcal{P}(ds, dx)\right| \leq \int_0^t \int_{|x|<\epsilon} |x|\lambda(ds, dx) = t\int_{|x|<\epsilon} |x|\nu(dx)$$

is finite. To remedy this problem, we modify the definition of our target process and set

$$X(t) \equiv ct + \int_0^t \int x \left(\mathcal{P}(ds, dx) - \mathbb{1}_{|x|\leq 1}\nu(dx)ds\right). \tag{3.8}$$

Note that this is the same as the original process, if $\int_{|x|\leq 1} x\nu(dx) = c$ is defined; this can indeed be decomposed as (for $0 < \epsilon < 1$)

$$X(t) = ct + \int_0^t \int_{|x|>\epsilon} x \left(\mathcal{P}(ds, dx) - \mathbb{1}_{|x|\leq 1}\nu(dx)ds\right) \tag{3.9}$$
$$+ \int_0^t \int_{|x|\leq\epsilon} x \left(\mathcal{P}(ds, dx) - \nu(dx)ds\right).$$

The first line is well defined. The second line satisfies

$$\mathbb{E} \int_0^t \int_{|x|\leq\epsilon} x \left(\mathcal{P}(ds, dx) - \nu(dx)ds\right) = 0, \tag{3.10}$$

and

$$\mathbb{E} \left(\int_0^t \int_{|x|\leq\epsilon} x \left(\mathcal{P}(ds, dx) - \nu(dx)ds\right) \right)^2 \tag{3.11}$$
$$= \int_0^t \int_{|x|\leq\epsilon} x^2 \lambda(ds, dx) = t \int_{|x|\leq\epsilon} x^2 \nu(dx)$$

The last expression[1] is finite, and hence the second line in (3.9) is a square integrable martingale. Moreover, the right-hand side of (3.10) tends to zero[2] as $\epsilon \downarrow 0$, and hence the second line in (3.9) tends to zero in probability as $\epsilon \downarrow 0$.

Since ϵ is arbitray, we see that $X(t)$ is a finite random variable (with possibly infinite variance), and that $X(t)$ is the limit of the càdlàg processes given by the first line of (3.9).

To conclude that $X(t)$ is a Lévy process we still need to show that it has a càdlàg version. But this follows from the fact that the martingale part is stochastically continuous using Doob's regularity theorem. The decomposition above with $\epsilon = 1$ is also known as *Lévy-Itô decomposition*.

Subordinators. Note that the process constructed above will in general not be an increasing process, also when ν is supported on the positive real numbers, except if in this case c is chosen as $c \geq \int_{x\leq 1} x\nu(dx)$, and of course subject to the condition that the latter integral is finite, which is the condition on the Lévy measure for subordinators from Theorem 3.1.2. The constuction of subordinators is actually much simpler due to nice monotonicity properties.

[1] It is a good exercise to verify this fomula using the explicit construction of the Poisson process given in the last section.

[2] This fact follows from Lebesgue's dominated convergence theorem.

4

The classical limit theorems for sums

In this chapter we will discuss the classical limit theorems for sums of independent random variables with a particular emphasis on the case of heavy-tailed random variables. We will begin for simplicity with the standard case of iid random variables.

4.1 Independent random variables

Every students in probability theory learns in the first course the following two facts about sums of iid rv's:

Theorem 4.1.1 (LLN) *Let X_i be iid random variables such that $\mathbb{E}X_1 = \mu$ exists, then*

$$\lim_{n\uparrow\infty} n^{-1}\sum_{i=1}^{n} X_i = \mu, \text{ a.s.}. \tag{4.1}$$

Moreover:

Theorem 4.1.2 (CLT) *Let X_i be iid random variables such that $\mathbb{E}X_1 = \mu$ and $\mathbb{E}(X_1-\mu)^2 = \sigma^2$ exist, then*

$$\lim_{n\uparrow\infty} \frac{1}{\sqrt{\sigma^2 n}}\sum_{i=1}^{n}(X_i-\mu) = Z, \tag{4.2}$$

where convergence is in distribution and Z is a standard normal Gauassian random variable.

A little later one is told that corresponding *functional limit laws* also hold:

Theorem 4.1.3 (LLN) *Let X_i be iid random variables such that $\mathbb{E}X_1 = \mu$ exists, then*

$$\lim_{n\uparrow\infty} n^{-1} \sum_{i=1}^{[nt]} X_i = \mu t, \text{ a.s.}, \tag{4.3}$$

uniformly on compact intervals.

and

Theorem 4.1.4 (IP) *Let X_i be iid random variables such that $\mathbb{E}X_1 = \mu$ and $\mathbb{E}(X_1 - \mu)^2 = \sigma^2$ exist, then*

$$\lim_{n\uparrow\infty} \frac{1}{\sqrt{\sigma^2 n}} \sum_{i=1}^{[nt]} (X_i - \mu) = B_t, \tag{4.4}$$

where convergence is in distribution on Wiener space equipped with the topology of uniform convergence on compact intervals and B_t is Brownian motion.

What happens when the conditions of these two theorems are seriously violated? Naturally, this will bring extremes and Lévy processes into the game.

The following are classical results, see [21, 26]:

Theorem 4.1.5 *Let X_i be iid random variables and assume that*

(i) $\mathbb{E}X_1 = \mu$ *exists finitely, and*
(ii) there exists $\alpha \in (1, 2)$ *such that*

$$n\mathbb{P}\left[X_1 > n^{1/\alpha} x\right] \to c_+ x^{-\alpha} \tag{4.5}$$

$$n\mathbb{P}\left[X_1 < -n^{1/\alpha} x\right] \to c_- x^{-\alpha} \tag{4.6}$$

with $c_+ + c_- > 0$,

then

$$n^{-1/\alpha} \sum_{i=1}^{[tn]} (X_i - \mathbb{E}X_i \mathbb{1}_{|X_i| \leq n^{1/\alpha}}) \to V_{\alpha,c_+,c_-}(t) \tag{4.7}$$

where $V_{\alpha,c+,c_-}$ is a stable Lévy process with Lévy triple $(0, 0, \nu_{\alpha,c_+,c_-})$, where

$$\nu_{\alpha,c_+,c_-}(dx) = c_+ \alpha x^{-\alpha-1} \mathbb{1}_{x>0} dx + c_- \alpha (-x)^{-\alpha-1} \mathbb{1}_{x<0} dx. \tag{4.8}$$

Convergence is in law with respect to the Skorokhod (J_1)-topology [1]

[1] I mention the topologies of convergence, but I postpone the discussion of all topological issues to Chapter 6.

Remark 4.1.1 Note that a drift appears unless $c_+ = c_-$, even though we centered! This is of course to compensate the unbalance of the large jumps.

Remark 4.1.2 The particular choice of the assumptions (4.5) is almost dictated by Theorem 2.1.6. One can in fact show, that, whenever $n\mathbb{P}(X > u_n x) \to x^{-a}$, with $\alpha \in (0,2]$, then $u_n = n^{1/\alpha} L(n)$ where L is a *slowly varying function* , that is, for any $t > 0$, $L(tn)/L(n) \to 1$, as $n \uparrow \infty$. Here I consider for simplicity only the case when $L(n) \equiv 1$, but the generalisations are quite straightforward and nothing fundamentally more interesting happens.

Proof. The proof of this theorem is very instructive and exhibits a strategy that will be used recurrently. Then we decompose, for $\epsilon > 0$ fixed,

$$\begin{aligned}\sum_{i=1}^{[nt]} X_i - \mathbb{E}X_i \mathbb{1}_{|X_i| \le n^{1/\alpha}} &= \sum_{i=1}^{[nt]} \left(X_i \mathbb{1}_{|X_i| \le n^{1/\alpha}\epsilon} - \mathbb{E}X_i \mathbb{1}_{|X_i| \le n^{1/\alpha}\epsilon} \right) \quad (4.9)\\ &+ \sum_{i=1}^{[nt]} \left(X_i \mathbb{1}_{1 \ge n^{-1/\alpha} X_i > \epsilon} - \mathbb{E}X_i \mathbb{1}_{X_i > n^{1/\alpha}\epsilon} \right)\\ &+ \sum_{i=1}^{[nt]} \left(X_i \mathbb{1}_{-1 \le n^{-1/\alpha} X_i < -\epsilon} - \mathbb{E}X_i \mathbb{1}_{X_i < -n^{1/\alpha}\epsilon} \right)\\ &\equiv Z_{nt}^{\le} + (Z_{nt}^{+} - \mathbb{E}Z_{nt}^{+,t}) + (Z_{nt}^{-} - \mathbb{E}Z_{nt}^{-,t}).\end{aligned}$$

Now, clearly, by the choice of the truncation, the terms $Z_{nt}^{\pm}$ will be of order $n^{1/\alpha}$ and they increase as ϵ tends to zero. Thus we have to control the size of the term $Z_{nt}^{\le}$. To do so, we will use a second moment estimates. For this it was quite crucial to recenter the truncated random variables in the three terms.

$$\begin{aligned}\mathbb{E}\left(Z_{nt}^{\le}\right)^2 &= \sum_{i=1}^{nt} \mathbb{E}\left(X_i \mathbb{1}_{|X_i| \le n^{1/\alpha}\epsilon} - \mathbb{E}X_i \mathbb{1}_{|X_i| \le n^{1/\alpha}\epsilon} \right)^2 \quad (4.10)\\ &\le nt\mathbb{E}X_1^2 \mathbb{1}_{|X_1| \le n^{1/\alpha}\epsilon}\end{aligned}$$

Using integration by parts, we see that

$$\begin{aligned} &\mathbb{E}X_1^2 \mathbb{1}_{|X_1| \leq n^{1/\alpha}\epsilon} \\ &= -n^{2/\alpha}\epsilon^2 \left(\mathbb{P}\left(X_1 > n^{1/\alpha}\epsilon \right) + \mathbb{P}\left(X_1 < -n^{1/\alpha}\epsilon \right) \right) \\ &+ 2\int_0^{n^{1/\alpha}\epsilon} x\mathbb{P}[X_1 > x]dx + 2\int_0^{n^{1/\alpha}\epsilon} x\mathbb{P}[X_1 < -x]dx. \end{aligned} \tag{4.11}$$

Changing variables and using the asymptotics of the law of X_1, we see that

$$\epsilon^2 n^{2/\alpha}\mathbb{P}\left(X_1 > n^{1/\alpha}\epsilon \right) \sim c_+ n^{2/\alpha-1}\epsilon^{2-\alpha}, \tag{4.12}$$

and

$$\begin{aligned} \int_0^{n^{1/\alpha}\epsilon} x\mathbb{P}[X_1 > x]dx &= n^{2/\alpha}\int_0^{\epsilon} y\mathbb{P}[X_1 > n^{1/\alpha}y]dy \\ &\sim n^{2/\alpha-1}\int_0^{\epsilon} y^{1-\alpha}dy = n^{2/\alpha-1}\frac{c_+}{2-\alpha}\epsilon^{2-\alpha}. \end{aligned} \tag{4.13}$$

Of course we still have to justify the passage to the limit under the integral. This will be postponed to Lemma 4.1.6.

The remaining terms are dealt with analogously, and we see that

$$\limsup_{n\uparrow\infty} \mathbb{E}\left(n^{-1/\alpha} Z_{nt}^{\leq} \right)^2 \leq \frac{\alpha(c_+ + c_-)}{2-\alpha}\epsilon^{2-\alpha} \tag{4.14}$$

which tends to zero, as $\epsilon \downarrow 0$. Thus we get readily that $Z_{nt}^{\leq}$ tends to zero as $n \uparrow \infty$ and $\epsilon \downarrow 0$.

Next we deal with Z_{nt}^{+}. For fixed ϵ, by the convergence Poisson convergence theorem,

$$n^{-1/\alpha} Z_{nt}^{+} = \int_{y>\epsilon}\int_0^t y\mathcal{P}_n(dy, ds), \tag{4.15}$$

where we define the point process

$$\mathcal{P}_n \equiv \sum_{i=1}^{\infty} \delta_{n^{-1/\alpha}X_i, i/n}. \tag{4.16}$$

Now $\mathcal{P}_n$ converges to the Poisson point process with intensity measure $c_+\alpha y^{-\alpha-1}dyds$, and for any $\epsilon > 0$ and $T < \infty$, on $(\epsilon, \infty) \times (0, T]$ the latter has finite intensity. The same applies to the term Z_{nt}^{-}. Thus, if we set

$$\nu(dx) \equiv c_+\alpha x^{-\alpha-1}dx\mathbb{1}_{x>0} + c_-\alpha(-x)^{-\alpha-1}dx\mathbb{1}_{x<0}, \tag{4.17}$$

we get that

$$n^{-1/\alpha}(Z_{nt}^{+} + Z_{nt}^{-}) \to \int_{|x|>\epsilon} \int_0^t x\mathcal{P}(dx, ds) \tag{4.18}$$

where $\mathcal{P}$ is the Poisson point process with intensity measure $\nu(dx)ds$. Moreover,

$$\begin{aligned} & n^{-1/\alpha} \sum_{i=1}^{[nt]} \mathbb{E}\left(X_i \mathbb{1}_{1\geq n^{-1/\alpha}X_i>\epsilon}\right) \qquad (4.19)\\ &= n^{1-1/\alpha} t \mathbb{E} X_1 \mathbb{1}_{1\geq n^{-1/\alpha}X_i>\epsilon} \\ &= t\int_\epsilon^1 n\mathbb{P}[X_1 > yn^{1/\alpha}]dy + \epsilon n\mathbb{P}[X_1 > \epsilon n^{1/\alpha}] - n\mathbb{P}[X_1 > n^{1/\alpha}] \\ &\to c_+ \int_\epsilon^1 y^{-\alpha} dy + c_+\epsilon^{1-\alpha} - c_+ \\ &= tc_+\alpha \int_\epsilon^1 x x^{-\alpha-1} dx = t\int_\epsilon^1 x\nu(dx) \end{aligned}$$

Combining this, we get that

$$n^{-1/\alpha}(Z_{nt}^{+} - \mathbb{E}Z_{nt}^{+}) \to \int_\epsilon^\infty \int_0^t x\left(\mathcal{P}(dx, ds) - \mathbb{1}_{x\leq 1}\nu(dx)dt\right), \tag{4.20}$$

where we used integration by parts backwards to arrive at the last line. Applying the same reasoning to Z_{nt}^{-}, we find that altogether,

$$\begin{aligned} & n^{-1/\alpha}(Z_{nt}^{+} - \mathbb{E}Z_{nt}^{+} + Z_{nt}^{-} - \mathbb{E}Z_{nt}^{-}) \qquad (4.21)\\ &\to \int_{|x|>\epsilon} \int_0^t x\left(\mathcal{P}(dx, ds) - \mathbb{1}_{|x|\leq 1}\nu(dx)dt\right). \end{aligned}$$

But we have seen in Chapter 3 that, as $\epsilon \downarrow 0$

$$\int_{|x|>\epsilon} \int_0^t x\left(\mathcal{P}(dx, ds) - \mathbb{1}_{|x|\leq 1}\nu(dx)dt\right) \to V_\nu(t), \tag{4.22}$$

where V_ν is the Lévy process with Lévy triple $(0, 0, \nu)$, provided $\int_0^\epsilon x^2\nu(dx) \downarrow 0$, as $\epsilon \downarrow 0$, which here is the case as one can see from (4.13). This proves the theorem (modulo tightness, do be done later). Notice how crucial the precise choice of the decomposition (4.9) was. In particular, the seperate centering of all terms is necessary to obtain expressions that converge as $\epsilon \downarrow 0$. □

Let us now show how the passage to the limit in (4.13) can be justified.

Lemma 4.1.6 *Let $0 \leq g \leq 1$ be a function such that $mg(m^{1/\alpha}x) \to x^{-\alpha}$, uniformly on compact intervals.*

Then, if $\alpha < s+1$,

$$\int_0^1 nx^s g(n^{1/\alpha}x)dx \to \int_0^1 x^{s-\alpha}dx. \tag{4.23}$$

Proof. Let us first note that

$$\int_0^1 nx^s g(n^{1/\alpha}x)dx = n^{-(s+1)/\alpha+1} \int_0^{n^{1/\alpha}} y^s g(y)dy. \tag{4.24}$$

Let us fix a number ρ_n such that $\rho_n \uparrow \infty$ and $\rho_n^{s+1} n^{-(s+1)/\alpha+1} \downarrow 0$. Then set $r_k \equiv \rho_n 2^k$ and write, with k_n such that $r_{k_n} = n^{1/\alpha}$,

$$n^{-(s+1)/\alpha+1} \int_0^{n^{1/\alpha}} y^s g(y)dy = n^{-(s+1)/\alpha+1} \int_0^{r_0} y^s g(y)dy + \sum_{k=0}^{k_n} n^{-(s+1)/\alpha+1} \int_{r_k}^{r_{k+1}} y^s g(y)dy. \tag{4.25}$$

The first term on the right-hand side is bounded as

$$\begin{aligned} & n^{-(s+1)/\alpha+1} \int_0^{r_0} y^s g(y)dy \\ & \leq n^{-(s+1)/\alpha+1} \int_0^{r_0} y^s dy n^{-(s+1)/\alpha+1} r_0^{s+1}/(s+1) \\ & = \rho_n^{s+1} n^{-(s+1)/\alpha+1}/(s+1), \end{aligned} \tag{4.26}$$

which tends to zero by hypothesis. Now

$$\int_{r_k}^{r_{k+1}} y^s g(y)dy = r_k^{s+1} \int_1^2 x^s g(xr_k)dy = r_k^{s+1-\alpha} \int_1^2 x^s r_k^\alpha g(r_k x)dx. \tag{4.27}$$

Now $r_k^\alpha g(r_k x) \to x^{-\alpha}$, uniformly in $x \in [1,2]$ and in $k \geq 1$, i.e. for any $\epsilon > 0$, there exists $n_0 < \infty$, s.t. for all $n \geq n_0$, and for all $x \in [0,1], k \geq 1$, $|r_k^\alpha g(r_k x - x^{-\alpha}| \leq \epsilon$. Hence, for such n,

$$\begin{aligned} & n^{-(s+1)/\alpha+1} \left| \sum_{k=0}^{k_n} \int_{r_k}^{r_k+1} y^s g(y)dy - \int_{r_0}^{n^{1/\alpha}} y^{s-\alpha}dy \right| \\ & \leq \epsilon \sum_{k=0}^{k_n} n^{-(s+1)/\alpha+1} r_k^{s+1-\alpha} \leq \epsilon n^{-(s+1)/\alpha+1} C r_{k_n}^{s+1-\alpha} \leq C\epsilon, \end{aligned} \tag{4.28}$$

for some constant C depeding on $s+1-\alpha$. This, together with the vanishing of the first term in (4.25) yields the claim of the lemma. □

This lemma will be used in the sequel whenever it is applicable without further mention.

In the case when X_1 has no mean, the analog of the functional LLN is the following theorem. This is the case that we will be mostly be concerned with in our applications.

Theorem 4.1.7 *Let X_i be iid random variables with support in $\mathbb{R}_+$ and assume that there exists $\alpha \in (0,1)$ such that*

$$n\mathbb{P}\left[X_1 > n^{1/\alpha}x\right] \to cx^{-\alpha} \tag{4.29}$$

with $c > 0$. Then

$$S_n(t) \equiv n^{-1/\alpha}\sum_{i=1}^{[tn]} X_i \to V_{\alpha,c}(t) \tag{4.30}$$

where $V_{\alpha,c}$ is a stable Lévy subordinator with Lévy triple $(0,0,\nu_{\alpha,c})$, where

$$\nu_{\alpha,c}(dx) = c\alpha x^{-\alpha-1}\mathbb{1}_{x>0}dx \tag{4.31}$$

Convergence is in law with respect to the Skorokhod (J_1)-topology.

Remark 4.1.3 If X_i has no mean and takes both positive and negative values, then one may simply decompose it into its positive and negative part and consider the sums of each part separately.

Proof. The proof is fully equal in spirit to the proof of the previous theorem, but considerably simpler. We thus leave it as an exercise. □

We have left open the three special cases $\alpha = 2, 1, 0$. They require some extra care. Let us first look at the case $\alpha = 2$.

Theorem 4.1.8 *Assume that the hypothesis of Theorem 4.1.5 are satisfied but $\alpha = 2$. Then*

$$\frac{1}{\sqrt{\frac{c_+ + c_-}{2} n\ln n}}\sum_{i=1}^{[nt]}(X_i - \mu) \to B_t, \tag{4.32}$$

where B_t is standard Brownian motion.

Proof. The key point is to understand that the extra logarithm in the normalisation of the variance comes from the fact that

$$\mathbb{E}X_1^2\mathbb{1}_{|X_1|\leq \epsilon n^{1/2}} \sim \ln n, \tag{4.33}$$

with the leading term independent of ϵ. To see this, note that the computation in (4.12) does not work in the case $\alpha = 2$, since the right-hand side is infinite. Thus we have to be more careful. For $0 \ll b_n \ll \sqrt{n}$ to be chosen later, set

$$\int_0^{\sqrt{n}\epsilon} x\mathbb{P}(X_1 > x)\,dx = \int_0^{b_n} x\mathbb{P}(X_1 > x)\,dx + \int_{b_n}^{\sqrt{n}\epsilon} x\mathbb{P}(X_1 > x)\,dx. \tag{4.34}$$

The first term is trivially bounded by $\frac{1}{2}b_n^2$. For the second, we get

$$\begin{aligned}\int_{b_n}^{\sqrt{n}\epsilon} x\mathbb{P}(X_1 > x)\,dx &= b_n^2 \int_1^{\sqrt{n}\epsilon/b_n} x\mathbb{P}(X_1 > xb_n)\,dx \\ &\to \int_1^{\sqrt{n}\epsilon/b_n} c_+ x^{-1} dx \\ &= \frac{1}{2}\ln n + \ln\epsilon + \ln b_n.\end{aligned} \tag{4.35}$$

If we chose $b_n = (\ln n)^{1/4}$, this term dominates the term of order b_n^2. Thus, we have to consider $(n\ln n)^{-1/2}Z_{nt}$ if we want to expect a limit. But this over-normalises the extreme part, that therefore tends to zero. Also, one can without difficulty compute all moments of the central part, and sees that these converge to the moments of a normal distribution. (This is because all higher truncated moments of X_1 beyond the second one do not get the "extra $\ln n$"). Thus here we get convergence to Brownian motion. See e.g. [34]. □

The case $\alpha = 1$ is even easier:

Theorem 4.1.9 *Assume that the hypothesis of Theorem 4.1.7 are satisfied but $\alpha = 1$. Then*

$$\frac{1}{cn\ln n}\sum_{i=1}^{[nt]} X_i \to t \tag{4.36}$$

Remark 4.1.4 Again we treated only the spacial case when the slowly varying function is equal to one. In the general case, the $\ln n$ in the normalisations is replaced by another slowly varying function that may diverge or converge, depending on the case in question.

Proof. The proof is quite similar to the previous one. Using the same refined analysis as for the truncated second moment in the case $\alpha = 2$, we see that

$$\mathbb{E}X_1 \mathbb{1}_{X \le \epsilon n} = \epsilon n \mathbb{P}(X_1 > \epsilon n) + \int_0^{\epsilon n} \mathbb{P}(X_1 > x)\, dx. \tag{4.37}$$

The first term is of order one, while the second

$$\int_0^{\epsilon b_n} \mathbb{P}(X_1 > x)\, dx + \int_{b_n}^{\epsilon n} \mathbb{P}(X_1 > x)\, dx, \tag{4.38}$$

where again the first term is bounded by b_n, while the second behaves like

$$\int_{b_n}^{\epsilon n} \mathbb{P}(X_1 > x)\, dx \to c \ln n + \ln \epsilon - \ln b_n. \tag{4.39}$$

With $b_n = \sqrt{\ln n}$, we see that to leading order $\mathbb{E}X_1 \mathbb{1}_{X \le \epsilon n} \sim c \ln n$. But the higher moments of the truncated variables behave like $\mathbb{E}X_1^k \mathbb{1}_{X \le \epsilon n} \sim (\epsilon n)^{k-1}$, and thus one can easily see that in the normalized sums, all higher moments vanish. Moreover, This this normalisation, the terms corresponding to $Z_n^>$ tend to zero. This yields the claimed result. □

4.2 Triangular arrays and non-heavy tailed variables

Inspecting the proof above reveals that there is lots of space towards generalisations. The first that comes to mind is an extension to triangular arrays. We will formulate this for the $\alpha < 1$ case, but it is obvious that the same can be done for the finite mean case. The following theorem does not use the most economical notation, but is convenient for our applications.

Theorem 4.2.10 *Let X_i^n, $n \in \mathbb{N}$, $i \in \mathbb{N}$ be a family of random variables such that, for each $n \in \mathbb{N}$ fixed, the family $X_i^n, i \in \mathbb{N}$ is iid with support in $\mathbb{R}_+$. Assume that there exists $\alpha \in (0,1)$ and sequences, c_n, a_n, such that*

$$a_n \mathbb{P}[X_1^n > c_n x] \to x^{-\alpha}. \tag{4.40}$$

If, moreover,

$$\lim_{e \downarrow 0} \limsup_{n \uparrow \infty} c_n^{-1} a_n \mathbb{E} \mathbb{1}_{X_1^n \le c_n \epsilon} X_1^n = 0, \tag{4.41}$$

then

$$S_n(t) \equiv c_n^{-1} \sum_{i=1}^{[t a_n]} X_i^n \to V_\alpha(t) \tag{4.42}$$

where V_α is a stable Lévy subordinator with Lévy triple $(0, 0, \nu_\alpha)$, where

$$\nu_\alpha(dx) = \alpha x^{-\alpha-1}\mathbb{I}_{x>0}dx \tag{4.43}$$

Convergence is in law with respect to the Skorokhod (J_1)-topology.

Proof. We again decompose $Z_n \equiv \sum_{i=1}^{[a_n t]} X_i^n$ into the central and extreme parts,

$$Z_n \equiv \sum_{i=1}^{[a_n t]} X_i^n \mathbb{I}_{X_i^n \leq \epsilon c_n} + \sum_{i=1}^{[a_n t]} X_i^n \mathbb{I}_{X_i^n > \epsilon c_n} \equiv Z_n^{\leq} + Z_n^{+}. \tag{4.44}$$

It is clear that

$$c_n^{-1} Z_{nt}^{+} \to \int_\epsilon^\infty \int_0^\infty x\mathcal{P}(dx, ds) \tag{4.45}$$

where $\mathcal{P}$ is the Poisson point process with intensity $c\alpha x^{-\alpha-1}dxds$, and therefore

$$\lim_{\epsilon\downarrow 0}\lim_{n\uparrow\infty} c_n^{-1} Z_n^{+} = V_\alpha. \tag{4.46}$$

The control of the $Z_n^{\leq}$ term is now given by assumption due to (4.41). □

Remark 4.2.1 One might hope that condition (4.41) is "automatic" from condition (4.40) and the fact that $\alpha < 1$. Indeed, one would like to argue that

$$\begin{aligned} c_n^{-1}\mathbb{E}Z_n^{\leq} &= a_n t c_n^{-1}\mathbb{E}X_i^n \mathbb{I}_{X_i^n \leq \epsilon c_n} \\ &= t\epsilon a_n \mathbb{P}\left(X_1^n > \epsilon c_n\right) + t\int_0^\epsilon a_n \mathbb{P}\left(X_1^n > y c_n\right) dy \\ &\to tc\epsilon^{1-\alpha} + tc\int_0^\epsilon y^{-\alpha} dy = ct\frac{2-\alpha}{1-\alpha}\epsilon^{1-\alpha} \end{aligned} \tag{4.47}$$

which indeed tends to zero, as desired. However, unlike in the iid case, we cannot immediately deduce the convergence of the integral from the convergence of the integrand. Of course, any condition that would allow to use Lebesgue's dominated convergence theorem would be fine. For instance, a sufficient condition for obtaining (4.41) is that

$$x^\alpha a_n \mathbb{P}\left(X_1^n > x c_n\right) \to 1, \tag{4.48}$$

uniformly in $x \in (0, 1]$.

Of course we also have the analogous result for the case when $\alpha \in (1, 2)$.

Theorem 4.2.11 *Let X_i^n, $n \in \mathbb{N}$, $i \in \mathbb{N}$ be a family of random variables such that, for each $n \in \mathbb{N}$ fixed, the family $X_i^n, i \in \mathbb{N}$ is iid with finite mean. Assume that there exists $\alpha \in (1,2)$ and sequences, c_n, a_n, such that*

$$a_n \mathbb{P}\left[X_1^n > c_n x\right] \to c_+ x^{-\alpha} \tag{4.49}$$

$$a_n \mathbb{P}\left[X_1^n < -c_n x\right] \to c_- x^{-\alpha}. \tag{4.50}$$

Assume furthermore that

$$\lim_{\epsilon \downarrow 0} \limsup_{n \uparrow \infty} a_n c_n^{-2} \mathbb{E} 1\!\!1_{|X_1^n - \mathbb{E}X_1^n| \leq \epsilon c_n} \left(X_1^n - \mathbb{E}X_1^n\right)^2 = 0. \tag{4.51}$$

Then

$$S_n(t) \equiv c_n^{-1} \sum_{i=1}^{[ta_n]} \left(X_i^n - \mathbb{E}X_i^n 1\!\!1_{|X_i^n| \leq c_n}\right) \to V_\alpha(t) \tag{4.52}$$

where V_α is a stable Lévy process with Lévy triple $(0,0,\nu_\alpha)$, where

$$\nu_\alpha(dx) = c_+ \alpha x^{-\alpha-1} 1\!\!1_{x>0} dx + c_- \alpha (-x)^{-\alpha-1} 1\!\!1_{x<0} dx. \tag{4.53}$$

Convergence is in law with respect to the Skorokhod (J_1)-topology.

Remark 4.2.2 Note that we have only partially centered S_n. Whether the remaining centering terms converge to what one would expect, namely $-\int_{|x|>1} x \nu_\alpha(dx)$ depends on additional uniformity conditions in the convergence of the law of X_i^n.

Proof. The proof follows step by step that of Theorem 4.1.5, using the assumption (4.51) to control the terms $Z_n^{\leq}$. □

Remark 4.2.3 The same remark as in the preceding theorem applies here. An appropriate assumption on the uniformity of the convergence in (4.49) and (4.50) implies condition (4.51), which is thus "natural".

Also the borderline cases $\alpha = 1, 2$ carry over without change to triangular arrays, provided again we assume convergences are such that we can pass them into integrals.

Theorem 4.2.12 *Let X_i^n, $n \in \mathbb{N}$, $i \in \mathbb{N}$ be a family of random variables such that, for each $n \in \mathbb{N}$ fixed, the family $X_i^n, i \in \mathbb{N}$ is iid with finite mean. Assume that there exists sequences, c_n, a_n, such that*

$$a_n \mathbb{P}\left[X_1^n > c_n x\right] \to c_+ x^{-2} \tag{4.54}$$

$$a_n \mathbb{P}\left[X_1^n < -c_n x\right] \to c_- x^{-2} \tag{4.55}$$

Then

$$\frac{1}{c_n\sqrt{(c_+ + c_-)\ln c_n}} \sum_{i=1}^{[ta_n]} (X_i^n - \mathbb{E}X_i^n) \to B_t \tag{4.56}$$

where B_t is Brownian motion. If $X_i^n \geq$ and there exist sequences a_n, c_n, such that

$$a_n \mathbb{P}[X_1^n > c_n x] \to x^{-1} \tag{4.57}$$

Then

$$\frac{1}{c_n \ln c_n} \sum_{i=1}^{[ta_n]} X_i^n \to t. \tag{4.58}$$

Proof. The proof is again the same as in the iid case. □

4.3 Applications

The results for the convergence of sums of triangular arrays already have a number of interesting applications. We give a few illustrations.

4.3.1 Application 1: the REM-like trap model

The REM-like trap model is the simplest model for a spin-glass dynamic that was proposed by Bouchaud and Dean [14]. We will construct it as follows.

The *state space* of the model is the complete graph on $\mathcal{V}_n \equiv \{1, \ldots, n\}$. To this we associate a *random environment* given by family of iid random variables, $\tau(i), \in \mathbb{N}$, whose distribution is in the domain of attraction of an α-stable distribution with $\alpha \in (0,1)$, i.e.

$$m\mathbb{P}\left(\tau(1) > m^{1/\alpha}x\right) \to x^{-\alpha}. \tag{4.59}$$

$\tau(i)i$ will be the mean trapping time of the trap i. The *fast chain*, $J_n(k)$, here is simply a sequance of iid random variables, uniformly distributed on $\mathcal{V}_n$, i.e.

$$\mathbb{P}[J_n(k) = i] = 1/n, \tag{4.60}$$

for each $i \in \mathcal{V}_n$ and each $k \in \mathbb{N}$. Let furthermore e_i be a family of iid exponential random variables with mean one. The families of the τ, Y, and e are mutually independent.

Now define the *clock processes*

$$S_n(k) \equiv \sum_{i=0}^{k-1} e_i \tau(J_n(i)). \tag{4.61}$$

Then Bouchaud's process is

$$X_n(t) \equiv J_n(S_n^{-1}(t)). \tag{4.62}$$

We will see that the crucial object to study is the clock process. Let us condition on the sigma-algebra generated by the environment, $\mathcal{F} \equiv \sigma(\tau(i), i \in \mathbb{N}$. Henceforth we call the conditional law of everything else $\mathcal{P}_\omega$. The point is that under $\mathcal{P}_\omega$, the random variables $Z_i^n(\omega) \equiv \tau(J_n(i))[\omega]$ are independent and identically distributed (for fixed n), and hence also the random variables $e_i Z_i^n(\omega)$ are independent. It is easy to compute their law: clearly

$$\mathcal{P}_\omega(Z_i^n > x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\tau(i)>x}, \tag{4.63}$$

and whence

$$\mathcal{P}_\omega(e_i Z_i^n > x) = \frac{1}{n}\sum_{i=1}^n e^{-x/\tau(i)}. \tag{4.64}$$

It is clear that

$$\frac{1}{n}\sum_{i=1}^n e^{-x/\tau(i)} \to \mathbb{E}e^{-x/\tau(1)} = \int_0^\infty e^{-x/t}\frac{x}{t^2}\mathcal{P}_\omega(\tau(1)>t)dt,\ \mathbb{P}-\text{ a.s.}. \tag{4.65}$$

In particular it is true that

$$\lim_{n\uparrow\infty}\frac{1}{n}\sum_{i=1}^n e^{-xm/\tau(i)} = \mathbb{E}e^{-xm/\tau(1)},\ \mathbb{P}-\text{ a.s.}, \tag{4.66}$$

and that

$$\begin{aligned} m\mathbb{E}e^{-xm^{1/\alpha}/\tau(1)} &= m\int_0^\infty e^{-xm^{1/\alpha}/t}\frac{xm^{1/\alpha}}{t^2}\mathbb{P}(\tau(1)>t)dt \qquad (4.67)\\ &= \int_0^\infty e^{-x/t}\frac{x}{t^2}m\mathbb{P}(\tau(1)>tm^{1/\alpha})dt \\ &\to \int_0^\infty e^{-x/t}\frac{x}{t^2}t^{-\alpha}dt = \Gamma(1+\alpha)x^{-\alpha}. \end{aligned}$$

as $m\uparrow\infty$. Hence we can readily conclude from Theorem 4.2.10 that for almost all environments,

$$\lim_{m\uparrow\infty}\lim_{n\uparrow\infty} m^{-1/\alpha}S_n(mt) \to V_{\alpha,\Gamma(1+\alpha)}(t). \tag{4.68}$$

However, more is true. Note that in our system we have a parameter n relating to the "size" of the system. Thus we can think of timescales, m, that depend on the system size n. For example, we may wish to consider timescales $m(n) = n^\beta$. Then we should ask under what circumstances it still holds true that

$$n^\beta \mathcal{P}_\omega(e_i Z_i^n > x n^{\beta/\alpha}) \to c x^{-\alpha}? \tag{4.69}$$

The non-trivial step is to check whether we can replace $\frac{1}{n}\sum_{i=1}^n e^{-xn^{\beta/\alpha}/\tau(i)}$ by $\mathbb{E}e^{-xn^{\beta/\alpha}/\tau(1)}$.

More precisely, we must check whether

$$\frac{1}{n}\sum_{i=1}^n n^\beta e^{-xn^{\beta/\alpha}/\tau(i)} \sim \mathbb{E} n^\beta e^{-xn^{\beta/\alpha}/\tau(i)} \to \Gamma(1+\alpha) x^{-\alpha}, \tag{4.70}$$

which is again about the validity of the strong law large numbers for triangular arrays. In our case it is fairly easy to show, using exponential Chebeychev inequalities, that the this is the case provided $\beta < 1$. For $\beta = 1$, it is quite clear that such a result will not hold and one must expect rather different behaviour (see [23]).

Lemma 4.3.13 *For all* $0 < \beta < 1$,

$$\frac{1}{n}\sum_{i=1}^n n^\beta e^{-xn^{\beta/\alpha}/\tau(i)} - \mathbb{E} n^\beta e^{-xn^{\beta/\alpha}/\tau(i)} \to 0, \ \mathbb{P}-\text{ a.s.}. \tag{4.71}$$

Proof. We will prove the result using an exponential Chebychev inequality. To do this we use the following bound on the Laplace transform of the variables $n^\beta e^{-xn^{\beta/\alpha}/\tau(1)}$:

$$\mathbb{E} e^{\lambda n^\beta e^{-xn^{\beta/\alpha}/\tau(i)}} = 1 + \int_0^\infty \mathbb{P}\left(\tau(1) > y n^{\beta/\alpha}\right) e^{\lambda n^\beta e^{-x/y}} n^\beta \lambda e^{-x/y} \frac{x}{y^2} dy \tag{4.72}$$

and so, with $\lambda = \gamma n^{-\beta}$,

$$\mathbb{E}\exp\left(\lambda\left(n^{\beta}e^{-xn^{\beta/\alpha}/\tau(i)}-\mathbb{E}n^{\beta}e^{-xn^{\beta/\alpha}/\tau(i)}\right)\right) \tag{4.73}$$
$$=\left(1+\int_0^\infty \mathbb{P}\left(\tau(1)>n^{\beta/\alpha}y\right)e^{-\gamma e^{-x/y}}\gamma e^{-x/y}\frac{x}{y^2}dy\right)$$
$$\times\exp\left(-\gamma\int_0^\infty \mathbb{P}\left(\tau(1)>n^{\beta/\alpha}y\right)e^{-x/y}\frac{x}{y^2}dy\right)$$
$$\sim 1+\int_0^\infty \mathbb{P}\left(\tau(1)>n^{\beta/\alpha}y\right)\left((e^{-\gamma e^{-x/y}}-1\right)\gamma e^{-x/y}\frac{x}{y^2}dy$$
$$\sim 1+n^{-\beta}\int_0^\infty\left(e^{-\gamma e^{-x/y}}-1\right)\gamma e^{-x/y}\frac{x}{y^{2+\alpha}}dy$$
$$\leq \exp\left(n^{-\beta}\int_0^\infty\left(e^{-\gamma e^{-x/y}}-1\right)\gamma e^{-x/y}\frac{x}{y^{2+\alpha}}dy\right)$$

The important fact is that the exponent is proportional to γ^2 for small γ. Thus Chebychev yields that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}n^{\beta}e^{-xn^{\beta/\alpha}/\tau(i)}-\mathbb{E}n^{\beta}e^{-xn^{\beta/\alpha}/\tau(i)}>r\right) \tag{4.74}$$
$$\leq \exp\left(-r\gamma n^{1-\beta}+n^{1-\beta}\int_0^\infty\left(e^{-\gamma e^{-x/y}}-1\right)\gamma e^{-x/y}\frac{x}{y^{2+\alpha}}dy\right)$$

Clearly, for any $r>0$, we can find $\gamma>0$ such that the right-hand side is of order $\exp\left(-n^{1-\beta}O(1)\right)$, which is summable. A corresponding bound for lower deviations yields the same result, and this gives the strong law of large numbers, as desired.

Finally, we will verify condition (4.41). First, the estimate (4.74) is good enough to show that we can replace the empirical probability in the formula for the moment by its mean. Then it remains to show that

$$\int_0^\epsilon \mathbb{E}n^{\beta}e^{-xn^{\beta/\alpha}/\tau(1)}dx\to\Gamma(1+\alpha)\int_0^\epsilon x^{-\alpha}dx. \tag{4.75}$$

But this can be done in the same way as in the proof of Lemma 4.1.6. □

Remark 4.3.1 In fact, we need a slightly stronger result, namely almost sure convergence uniform in x. This can be done using simple chaining due to the good continuity properties of our random variables in x. We will skip the details here.

We can now reap the consequence of these observations.

Proposition 4.3.14 *In the REM-like trap model, for any $0 < \beta < 1$, the clock process $S_n(k)$ satisfies, for almost all realisations of the random environment,*

$$n^{-\beta/\alpha} S_n \left(tn^{\beta}\right) \to V_{\alpha,\Gamma(1+\alpha)}(t). \tag{4.76}$$

Convergence is in distribution with respect to the Skorokhod J_1-topology.

4.3.2 Application 2: Sums of exponentials of random variables

The following application is a generalisation of the equilibrium of the REM that was studied by Ben Arous, Bogachev, and Molchanov [1] some years ago. It concerns the special case of triangular arrays when X_i^n takes the form

$$X_i^n = \exp\left(\alpha^{-1} b_n Z_i\right), \tag{4.77}$$

where Z_i are iid random variables. We will formulate their theorem as follows[1].

Theorem 4.3.15 *Let Z_i be iid random variables in the domain of attraction of the Gumbel distribution. Then there exists sequences b_n and c_n, defined by the requirement*

$$n\mathbb{P}\left(Z > b_n^{-1}\left(\ln c_n + z\right)\right) \to e^{-z}. \tag{4.78}$$

Assume that moreover

$$\int_{-\infty}^{0} e^{sz} n\mathbb{P}\left(Z > b_n^{-1}\left(\ln c_n + \alpha z\right)\right) dz \to \int_{-\infty}^{0} e^{(s-\alpha)x} dx, \tag{4.79}$$

for $s > \alpha$. Then the following hold, with X_i^n given by (4.77):

(i) *For $\alpha \in (1,2)$,*

$$c_n^{-1/\alpha} \sum_{i=1}^{[tn]} \left(X_i^n - \mathbb{E} X_i^n \mathbb{1}_{X_i^n \leq c_n^{1/\alpha}}\right) \to V_\alpha, \tag{4.80}$$

where V_α is the α-stable Lévy process with Lévy triple $(0, 0, \nu_\alpha)$, where

$$\nu_\alpha(dx) = \alpha x^{-1-\alpha} dx \mathbb{1}_{x>0}. \tag{4.81}$$

(ii) *For $\alpha \in (0,1)$,*

$$c_n^{-1/\alpha} \sum_{i=1}^{[tn]} X_i^n \to V_\alpha, \tag{4.82}$$

[1] In my view, the exposition in [1] obscures this simple theorem by stating conditions in terms of Laplace transforms etc. .

where V_α is the α-stable Lévy subordinator with Lévy triple $(0, 0, \nu_\alpha)$, *where*

$$\nu_\alpha(dx) = \alpha x^{-1-\alpha} dx \mathbb{1}_{x>0}. \tag{4.83}$$

(iii) If $\alpha = 2$, then

$$\frac{1}{\sqrt{\frac{1}{2} c_n \ln c_n}} \sum_{i=1}^{n} \left(X_i^n - \mathbb{E} X_i^n \mathbb{1}_{X_i^n \leq c_n^{1/2}} \right) \to B_t, \tag{4.84}$$

where B_t is Brownian motion.

(iv) If $\alpha = 1$, then

$$\frac{1}{c_n \ln c_n} \sum_{i=1}^{n} e^{b_n Z_i} \to t. \tag{4.85}$$

Proof. The proof is of course nothing more than the verification of the conditions (4.40) respectively (4.49). Now

$$n\mathbb{P}\left(\exp\left(\alpha^{-1} b_n Z_1\right) > c_n x\right) = n\mathbb{P}\left(\alpha^{-1} Z_1 > b_n^{-1} \left(\ln c_n + \ln x\right)\right). \tag{4.86}$$

Set first $\alpha = 1$ and $\ln x = z$. Then by definition of Z_1 being in the domain of attraction of the Gumbel distribution, there exits sequences b_n, c_n, such that

$$n\mathbb{P}\left(Z_1 > b_n^{-1} \left(\ln c_n + z\right)\right) \to e^{-z}. \tag{4.87}$$

But then

$$\begin{aligned} n\mathbb{P}\left(1/\alpha Z_1 > b_n^{-1}\left(\ln c_n^{1/\alpha} + z\right)\right) &= n\mathbb{P}\left(Z_1 > b_n^{-1} \left(\ln c_n + \alpha z\right)\right) \\ &\to e^{-\alpha z} = x^{-\alpha}. \end{aligned} \tag{4.88}$$

Finally, one uses condition (4.79) to verify that in all computations, limits can be passed through integrals, which allows to control the terms corresponding to the $Z_n^{\leq}$ parts of the sums. □

The particular case when the random variables are Gaussian corresponds to the computation of the partition function of the random energy model (REM). This case has been worked out in detail in [17].

As an illustrative example and in view of later appearances of exponentials of Gaussian random variables, the following lemma shows that the condition (4.79) always holds in the Gaussian case.

Lemma 4.3.16 *Let Z be a normal random variable and assume that b_n and c_n are such that, for all $z \in \mathbb{R}$,*

$$n\mathbb{P}\left(Z > b_n^{-1} \left(\ln c_n + \alpha z\right)\right) \to e^{-\alpha z}, \tag{4.89}$$

with $\alpha \in (0,2]$. Then condition (4.79) holds.

Proof. We will in fact proceed as in the remark following Theorem 4.1.7. We use the classical upper bound for Gaussian tail distributions, valid for all $u > 0$.

$$\mathbb{P}(Z > u) \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-u^2/2}}{u}. \tag{4.90}$$

For our case this gives

$$n\mathbb{P}\left(Z > b_n^{-1}(\ln c_n + \alpha z)\right) \leq n \frac{\exp\left(-\frac{(\ln c_n)^2}{2b_n^2} - \alpha z b_n^{-2} \ln c_n - \frac{(\alpha z)^2}{2b_n^2}\right)}{\sqrt{2\pi} b_n^{-1}(\ln c_n + \alpha z)}. \tag{4.91}$$

For $\alpha z > -(1-\epsilon)\ln c_n$, this quantity is bounded from above by

$$n \frac{\exp\left(-\frac{(\ln c_n)^2}{2b_n^2} - \alpha z b_n^{-2} \ln c_n\right)}{\sqrt{2\pi}\epsilon b_n^{-1} \ln c_n} = n \frac{\exp\left(-\frac{(\ln c_n)^2}{2b_n^2}\right)}{\sqrt{2\pi}\epsilon b_n^{-1} \ln c_n} e^{-\alpha z b_n^{-2} \ln c_n}. \tag{4.92}$$

Since by assumption $b_n^{-2} \ln c_n \to 1$, and

$$n \frac{\exp\left(-\frac{(\ln c_n)^2}{2b_n^2}\right)}{\sqrt{2\pi} b_n^{-1} \ln c_n} \to 1, \tag{4.93}$$

for any $\epsilon > 0$, we can find $n_0 < \infty$, independent of z, such that for all $n \geq n_0$,

$$n\mathbb{P}\left(Z > b_n^{-1}(\ln c_n + \alpha z)\right) \leq \epsilon^{-1}(1+\epsilon) e^{-z(\alpha - \epsilon\,\mathrm{sign}\,(\epsilon))} \equiv g(z). \tag{4.94}$$

But

$$\int_{-\infty}^{0} e^{sz} g(z) dz = \epsilon^{-1}(1+\epsilon) \int_{-\infty}^{0} e^{(s-\alpha-\epsilon)z} dz, \tag{4.95}$$

which is finite whenever $s - \alpha - \epsilon < 0$. But this can always be achieved by choosing, e.g. $\epsilon = (s-\alpha)/2$. Therefore, the sequence of functions

$$n\mathbb{P}\left(Z > b_n^{-1}(\ln c_n + \alpha z)\right) 1\!\!1_{z > -\ln c_n(1-\epsilon)} \tag{4.96}$$

are bounded from above by an integrable function $g(z)$ and converge to $e^{-\alpha z}$ for almost all z. Finally,

$$\int_{-\infty}^{-\alpha^{-1}\ln c_n(1-\epsilon)} dz e^{sz} n\mathbb{P}\left(Z > b_n^{-1}(\ln c_n + \alpha z)\right)$$
$$\leq n \int_{-\infty}^{-\alpha^{-1}\ln c_n(1-\epsilon)} e^{sz} dz = s^{-1} n c_n^{-s\alpha^{-1}(1-\epsilon)}. \tag{4.97}$$

One easily checks that the assumptions imply for the constants that,

$$nc_n^{-1/2}b_n^{-1} \to 1, \tag{4.98}$$

and so

$$nc_n^{-s\alpha^{-1}(1-\epsilon)} \sim n^{1-2s\alpha^{-1}(1-\epsilon)}, \tag{4.99}$$

which tends to zero since $s > \alpha$ if ϵ is small enough. □

Remark 4.3.2 Recently, Janssen [28] has shown that in the gerneral case of random variables in the domain of attraction of the Gumbel distribution there are always subsequences such that condition (4.28) holds along those.

4.3.3 Application 3: A Gaussian trap model

The simplest dynamics to impose on the REM is to imitate the REM-like trap model and to take as the fast chain iid uniform jumps on the hypercube. While this is a rather stupid model, some otherwise instructive computations are involved that makes it worthwhile to go through it. So here our state space is $\mathcal{V}_n = \{-1,1\}^n$. The mean holding times at each site, $\sigma \in \mathcal{V}_n$, are

$$\tau_n(\sigma) \equiv e^{\beta\sqrt{n}H_\sigma},$$

where H_σ are iid standard normal random variables. Then the clock process is

$$S_n(k) \equiv \sum_{i=0}^{k-1} e_i \tau_n(J_n(i)),$$

and $Z_i^n \equiv e_i \tau_n(J_n(i))$ are iid random variables under the quenched law.

We now look for time scales, a_n, such that for suitable normalisation constants, c_n,

$$\frac{1}{c_n} \sum_{i=0}^{[ta_n]-1} X_i^n \to V_\alpha(t). \tag{4.100}$$

By Theorem 4.2.10, this amounts to check whether

$$a_n \mathcal{P}_\omega \left[e_i e^{\beta\sqrt{n}H_\sigma} > c_n s \right] \to s^{-\alpha}. \tag{4.101}$$

By a simple computation,

$$a_n \mathcal{P}_\omega \left[e_i e^{\beta\sqrt{n}H_\sigma} > c_n s \right] = \frac{a_n}{2^n} \sum_{\sigma \in \mathcal{V}_n} \exp\left(-sc_n e^{-\beta\sqrt{n}H_\sigma} \right). \tag{4.102}$$

Again we expect that this converges to its mean under suitable conditions. We first compute this mean, which is given by

$$a_n \int \frac{dx}{\sqrt{2\pi}} e^{-\frac{x^2}{2} - s c_n e^{-\beta\sqrt{n}x}} \tag{4.103}$$

It is obviously suitable to change variables via

$$x = \frac{\ln c_n}{\beta\sqrt{n}} + \frac{z}{\beta\sqrt{n}}. \tag{4.104}$$

Then (4.103) can be written as

$$\begin{aligned} &a_n \frac{1}{\beta\sqrt{2\pi n}} e^{-\frac{(\ln c_n)^2}{2\beta^2 n}} \int e^{-\frac{\ln c_n}{\beta^2 n} z - s e^{-z} - \frac{z^2}{2\beta^2 n}} dz \\ &\sim a_n \frac{1}{\beta\sqrt{2\pi n}} e^{-\frac{(\ln c_n)^2}{2\beta^2 n}} \int e^{-\frac{\ln c_n}{\beta^2 n} z - s e^{-z}} dz \\ &= a_n \frac{1}{\beta\sqrt{2\pi n}} e^{-\frac{(\ln c_n)^2}{2\beta^2 n}} \int_0^\infty r^{\frac{\ln c_n}{\beta^2 n} - 1} e^{-sr} dr \\ &= a_n \frac{1}{\beta\sqrt{2\pi n}} e^{-\frac{(\ln c_n)^2}{2\beta^2 n}} s^{-\frac{\ln c_n}{\beta^2 n}} \Gamma(1 + \ln c_n/(\beta^2 n)). \end{aligned} \tag{4.105}$$

Note that in the passage to the second line we assume that s is a constant independent of n. Thus if we set $c_n \equiv e^{\gamma n}$ we must choose

$$a_n = \beta\sqrt{2\pi n} e^{\frac{\gamma^2}{2\beta^2} n} / \Gamma(1 + \alpha) \tag{4.106}$$

with $\alpha = \frac{\gamma}{\beta^2}$ to get (4.101). Thinking of c_n as the *time scale*, we see that to get a subordinator, the condition $\alpha < 1$ reads $\beta^2 > \gamma$, i.e. this is a condition that the temperature be low enough (depending on the time scale).

A second condition will emerge from the requirement that (4.102) converges to its average. This goes analogously to the REM-like trap model and requires

$$a_n \ll 2^n \tag{4.107}$$

that is $\frac{\gamma^2}{\beta^2} < 2\ln 2$.

5

Dependent random variables

We will now restrict us to the case of sums of non-negative random variables in the domain of attaction of an α-stable law with $\alpha \in (0,1)$. We will try to do away with the independence assumption. Having in mind the proofs we gave so far, it is amply clear that all what is needed is the convergence of the extremal process to the Poisson process.

5.1 Convergence of sums of dependent random variables, $\alpha < 1$

Theorem 5.1.1 *Let X_i^n be a triangular array of random variables taking values in $\mathbb{R}_+$. Assume that for some sequences, a_n, c_n, and any $t > 0$ fixed,*

$$\sum_{i=1}^{[a_n t]} \mathbb{P}\left(X_i^n > c_n x\right) \to t x^{-\alpha}, \tag{5.1}$$

with $\alpha \in (0,1)$. Assume further that

$$\sum_{i \in \mathbb{N}} \delta_{(i/a_n, c_n^{-1} X_i^n)} \to \mathcal{P}_\alpha, \tag{5.2}$$

where $\mathcal{P}_\alpha$ is the Poisson process on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity measures $\alpha x^{-\alpha-1} dx ds$. If, moreover,

$$\lim_{e \downarrow 0} \limsup_{n \uparrow \infty} c_n^{-1} \sum_{i=1}^{[t a_n]} \mathbb{E} \mathbb{1}_{X_i^n \leq c_n \epsilon} X_i^n = 0, \tag{5.3}$$

then

$$c_n^{-1} \sum_{i=1}^{[a_n t]} X_i^n \to V_\alpha(t). \tag{5.4}$$

Proof. Inspecting the proof of Theorem (4.2.10) that the convergence of the extreme part the sum, Z^+_{nt}, is assured by hypothesis (5.2), while the irrelevance of the central part, $Z^{\leq}_{nt}$, follows from just (5.4). Here there is no difference to the independent case, since we only used a first moment estimate that does not feel any dependence effects. □

Thus, we are in excellent shape: any mixing conditions that ensure Poisson convergence of the extremes will give convergence of the sum to a Lévy subordinator. For instance, we can work with the criteria of Section 2.2.

This gives the following two usefull corollaries.

Corollary 5.1.2 *Let X^n_i, $n \in \mathbb{N}$, $i \in \{1, \dots, n\}$ be a triangular array of non-negative random variables, and assume that there exists sequences a_n, c_N and a σ-finite measure, ν, satisfying the assumptions of a Lévy measure, such that, for any $t \in \mathbb{R}_+$,*

$$\lim_{n\uparrow\infty} \sum_{\{j_1,\dots,j_\ell\}\subset\{1,\dots,[a_n t]\}} \mathbb{P}\left[\forall_{r=1}^{\ell} X^n_{j_r} > c_n s\right] = t^\ell s^{-\alpha\ell}. \tag{5.5}$$

If in addition (5.3) holds, then

$$c_n^{-1} \sum_{i=1}^{[ta_n]} X^n_i \to V_\alpha, \tag{5.6}$$

where V_ν is the Lévy subordinator with Lévy measure $\nu(dx) = \alpha x^{-\alpha-1} \mathbb{1}_{x>0} dx$ and zero drift.

Using the criteria of Durrett and Resnick [19], we obtain the alternative version:

Corollary 5.1.3 *In the same setting as in Corollary 5.1.2, the assumptions (5.5) can be replaced by:*

$$\lim_{n\uparrow\infty} \sum_{i=1}^{[a_n t]} \mathbb{P}\left(X^n_i > x c_n | \mathcal{F}_{n,i-1}\right) = t x^{-\alpha}, \tag{5.7}$$

and

$$\lim_{n\uparrow\infty} \sum_{i=1}^{[a_n t]} \left[\mathbb{P}\left(X^n_i > c_n x | \mathcal{F}_{n,i-1}\right)\right]^2 = 0, \tag{5.8}$$

with convergence in probability and $\mathcal{F}_{n,i} = \sigma(X_j, j \leq i)$.

Remark 5.1.1 Since in the triangular array case we are not necessarily falling into the case of a stable process, there may be situations where other measures then $x^{-\alpha}dx$ appear. Obviously the theorem and its corollaries remain then true provided we add an addition that ensures the vanishing of the contributions $S_n^{\leq}$. See [19].

A similar general theorem can be stated for the case $\alpha \in (1,2)$, with additional mixing conditions that ensure the appropriate bound on the variance of the central part of the sum (see e.g. [19]).

5.2 Applications to ageing

The general theorem above is a key tool to study ageing in more complicated dyanamics. This has been investigated in some detail in a series of recent papers by V. Gayrard [23, 24]. Here we consider some simple situations as an illustration.

5.2.1 The REM-like trap model, averaged

In the previous section we looked at the REM like trap model under the quenched law. If we were interested in results under the joint measure, $\mathbb{P}$, what changes is that the random variables $\tau(J_n(i))$ are no longer independent, since with some probability $J_n(i) = J_n(j)$, for $i \neq j$. On the other hand, the marginal distribution of $\tau(J_n(i))$ is just the law of $\tau(1)$, and so

$$m\mathbb{P}\left[e_i\tau(J_n(i)) > sm^{1/\alpha}\right] \to \Gamma(1+\alpha)s^{-\alpha}. \tag{5.9}$$

Thus we see that Hypothesis (5.1) is satisfied with $m = n^\beta$ as in (4.69). To check that Assumption (5.2) is satisfied, we may either Theorem 2.2.23 or Theorem 2.2.24. For the former, we need to compute

$$\mathbb{P}\left[e_i\tau(J_n(i)) > sn^{\beta/\alpha}, \forall_{i\leq\ell}\right] \tag{5.10}$$

This probability depends only on the number of distinct values taken by the $J_n(i), i \leq \ell$. Hence

$$\mathbb{P}\left[e_i\tau(J_n(i)) > sn^{\beta/\alpha}, \forall_{i\leq\ell}\right] = \sum_{k=0}^{\ell-1}\mathbb{P}\left[e_i\tau_i > sn^{\beta/\alpha}, \forall_{i\leq\ell-k}\right] \times\mathbb{P}\left[\#\{J_n(i), 0\leq i<\ell\} = \ell-k\right] \tag{5.11}$$

But for $k \geq 1$,

$$\mathbb{P}\left[\#\{J_n(i), 0\leq i<\ell\} = \ell-k\right] = O(n^{-k}) \tag{5.12}$$

(note that no uniformity in ℓ is required!), so that

$$n^{\beta\ell}\sum_{k=0}^{\ell-1}\mathbb{P}\left[e_i\tau_i > sn^{\beta/\alpha}, \forall_{i\leq \ell-k}\right]\mathbb{P}\left[\#\{J_n(i), 0\leq i<\ell\} = \ell-k\right] \tag{5.13}$$
$$\sim \Gamma(1+\alpha)^{\ell}s^{-\alpha\ell}\left(1+\sum_{k=1}^{\ell-1}c_k n^{k(\beta-1)}\right)\to \Gamma(1+\alpha)^{\ell}s^{-\alpha\ell}.$$

This is very nice and we see that the same conditions as in the quenched regime apply.

The very same argument can of course also be used in the case of the Gaussian trap model.

5.2.2 The full REM, averaged

We are now also ready to treat the REM with the fast process chosen as the simple random walk on the hypercube, at least under the joint law $\mathbb{P}$.

The key point is that again we prove that the criterion from Theorem 2.2.23 is satisfied. Since the fast process is now simple random walk rather than an iid sequence, we cannot use exchangability. We get (I set all the s equal to one to get things that are easier to write down, but all works in the general case)

$$\sum_{k_1<k_2<\cdots<k_\ell}\mathbb{P}\left(\tau_n(J_n(k_i))\geq c_n, \forall i\leq \ell\right) \tag{5.14}$$
$$=\sum_{k_1<k_2<\cdots<k_\ell}\sum_{r=1}^{\ell}a_n^{-r}\mathbb{P}\left(\#\{J_n(k_i), i=1,\dots,\ell\}=r\right).$$

So clearly all boils down to establish some random walk properties on the hypercube. This involves a bit of combinatorics, but to see why things go well, consider the case $\ell = 2$. We have to consider only one quantity, namely

$$
\begin{aligned}
\sum_{k_1<k_2} \mathbb{P}\left(\#\{J_n(k_i), i=1,2\}=1\right) &= \sum_{k_1<k_2} \mathbb{P}\left(J_n(k_2)) = J_n(k_1)\right) \\
&= \sum_{k_1<k_2} \mathbb{P}\left(J_n(k_2-k_1) = J_n(0)\right) \\
&\leq a_n \sum_{k=1}^{a_n} \mathbb{P}\left(J_n(k) = J_n(0)\right) \\
&= a_n \frac{\mathbb{P}\left(T_{J_n(0)} \leq a_n\right)}{1-\mathbb{P}\left(T_{J_n(0)} \leq a_n\right)}, \qquad (5.15)
\end{aligned}
$$

where $T_\sigma \equiv \inf\{k > 0 : J_n(k) = \sigma\}$. The probability that the simple random walk on the hypercobe returns to its starting point in time less than a_n, for $a_n \ll 2^n$, is of order $1/n$ (in fact the most likely way to realise this is to return in the second step, which has probability $1/n$). Thus we see that the contribution with $r = 1$ is by a factor $1/n$ smaller than the dominant contribution corresponding to $r = 2$. Thus

$$
\begin{aligned}
&\sum_{k_1<k_2} \mathbb{P}\left(\tau_n(J_n(k_i)) \geq c_n,\, i=1,2\right) \qquad (5.16) \\
&= \sum_{k_1<k_2} \left(a_n^{-2} + \mathbb{P}\left(J_n(k_1) = J_n(k_2)\right)\left(a_n^{-1} - a_n^{-2}\right)\right) = \frac{1}{2} + O(n^{-1}).
\end{aligned}
$$

To deal with the general case is quite similar. We will leave the proof to the reader. As a result, we see that the REM under $\mathbb{P}$ behaves exactly like the Gaussian trap model we dealt with before.

6

Blocking and applications

Up to now we have been dealing exclusively with situations when the processes we considered converged, on the extreme scales, to Poisson point processes. If correlations are too strong, this will necessarily have to fail. The best known situation of this kind occurs in the theory of continuous time stochastic processes, where paths may e.g. be continuous. In such a situation one may still expect a Poissonian picture at large distances, but a more complex local structure of clusters replacing the single points of the Poisson process.

To establish the global Poissonian nature of the sum, a good way to proceed is to use suitable blocking. Introduce a new scale θ_n and blocked random variables with

$$Z_{n,i} \equiv \sum_{j=\theta_n(i-1)+1}^{i\theta_n} X_j^n, \tag{6.1}$$

in the hope that the new variables verify the hypothesis of our theorems on Poisson convergence.

In this chapter we show how this idea can be applied in the context of Markov jump processes as defined in Chapter 1. This approach was developed in [15] in order to obtain almost sure results in the p-spin SK model. Here we give a more pedagogical application in the context of the REM, where blocking is not really necessary [24].

6.1 Convergence of blocked clock processes

We will use the notation of Subsection 1.1. That is, we will consider $X_i^n = \lambda_n^{-1}(J_n(i))e_{n.i}$. We will be interested in studying the clock process (1.2) under the law of the fast chain, J_n, for fixed random environments. If J_n is rapidly mixing, we can hope to choose $\theta_n \ll a_n$ such that the random variables $J_n(\theta_n i), i \in \mathbb{N}$ are close to independent and distributed according to the invariant distribution π_n. But then, under the law $\mathcal{P}_{\mu_n}$, also the random variables $Z_{n,i}$, are close to independent and uniformly distributed (although with a complicated distribution that is a random variable depending on the random environment). In this context, it will be most convenient to use the conditions, of Durrett and Resnick [19], i.e. Corollary 5.1.3.

Let us now look at this in more detail.

For $y \in \mathcal{V}_n$ and $u > 0$, let

$$Q_n^u(y) \equiv \mathcal{P}_y\left(\sum_{j=0}^{\theta_n - 1} \lambda_n^{-1}(J_n(j))e_{n,j} > c_n u\right) \tag{6.2}$$

be the tail distribution of the aggregated jumps when X_n starts in y. Note that $Q_n^u(y)$, $y \in \mathcal{V}_n$, is a random function on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and so is the function $F_n^u(y)$, $y \in \mathcal{V}_n$ defined through

$$F_n^u(y) \equiv \sum_{x \in \mathcal{V}_n} p_n(y,x) Q_n^u(x)\,. \tag{6.3}$$

Writing $k_n(t) \equiv \lfloor \lfloor a_n t \rfloor / \theta_n \rfloor$, we further define

$$\nu_n^{J,t}(u,\infty) \equiv \sum_{i=0}^{k_n(t)-1} F_n^u\big(J_n(\theta_n(i))\big)\,, \tag{6.4}$$

$$(\sigma_n^{J,t})^2(u,\infty) \equiv \sum_{i=0}^{k_n(t)-1} \left[F_n^u\big(J_n(\theta_n(i))\big)\right]^2\,. \tag{6.5}$$

Finally, we set

$$\bar{S}_n(k) \equiv \sum_{i=1}^{k}\left(\sum_{j=\theta_n(i-1)+1}^{\theta_n i} c_n^{-1}\lambda_n^{-1}(J_n(j))e_{n,j}\right) + c_n^{-1}\lambda_n^{-1}(J_n(0))e_{n,0}. \tag{6.6}$$

and

$$S_n^b(t) \equiv \bar{S}_n(k_n(t)). \tag{6.7}$$

We now formulate four conditions for the sequence S_n^b to converge to a subordinator. Note that these conditions refer to given sequences of numbers a_n, c_n, and θ_n as well as a given realisation of the random environment.

Condition (A1). There exists a σ-finite measure ν on $(0,\infty)$ satisfying the hypothesis stated in Theorem 2.2.24, and such that, for all $t > 0$ and all $u > 0$,

$$P_{\mu_n}\Big(\big|\nu_n^{J,t}(u,\infty) - t\nu(u,\infty)\big| < \epsilon\Big) = 1 - o(1)\,, \quad \forall \epsilon > 0\,. \tag{6.8}$$

Condition (A2). For all $u > 0$ and all $t > 0$,

$$P_{\mu_n}\Big((\sigma_n^{J,t})^2(u,\infty) < \epsilon\Big) = 1 - o(1)\,, \quad \forall \epsilon > 0\,. \tag{6.9}$$

Condition (A3). For all $t > 0$,

$$\lim_{\epsilon\downarrow 0}\limsup_{n\uparrow\infty} \mathcal{E}_{\mu_n} \sum_{i=1}^{\lfloor a_n t\rfloor} \mathbb{1}_{\{\lambda_n^{-1}(J_n(i))e_i \le c_n\epsilon\}} c_n^{-1}\lambda_n^{-1}(J_n(i))e_i = 0. \tag{6.10}$$

Condition (A0'). For all $v > 0$,

$$\sum_{x\in\mathcal{V}_n} \mu_n(x) e^{-vc_n\lambda_n(x)} = o(1)\,. \tag{6.11}$$

Theorem 6.1.1 *For all sequences of initial distributions μ_n and all sequences a_n, c_n, and $1 \le \theta_n \ll a_n$ for which Conditions (A0'), (A1), (A2), and (A3) are verified, either $\mathbb{P}$-almost surely or in $\mathbb{P}$-probability, the following holds w.r.t. the same convergence mode:*

$$S_n^b(\cdot) \Rightarrow S_\nu(\cdot), \tag{6.12}$$

where S_ν is the Lévy subordinator with Lévy measure ν and zero drift. Convergence holds weakly on the space $D([0,\infty))$ equipped with the Skorokhod J_1-topology.

Remark 6.1.1 Note that Condition (A0') is there to ensure that last term in (6.6) converges to zero in the limit $n \uparrow \infty$.

Remark 6.1.2 The result of this theorem is stated for the *blocked* process $S_n^b(t)$. It implies immediately that under the same hypothesis, the original process $S_n(t)$ (defined in (**??**)) converges to S_ν in the weaker M_1-topology (see [33] for a detailed discussion of Skorokhod topologies). However, the statement of the theorem is strictly stronger than just convergence in M_1, and it is this form that is useful in applications.

Remark 6.1.3 To extract detailed information on the process X_n, e.g. the behaviour of correlation functions, from the convergence of the blocked clock process, one needs further information on the typical behaviour of the process during the θ_n steps of a single block. This is a model dependent issue and we will examplify how this can be done in the context of the p-psin SK model.

Proof. Throughout we fix a realisation $\omega \in \Omega$ of the random environment but do not make this explicit in the notation. We set

$$\widehat{S}_n^b(t) \equiv S_n^b(t) - c_n^{-1}\lambda_n^{-1}(J_n(0))e_{n,0}. \tag{6.13}$$

Condition (A0') ensures that $S_n^b - \widehat{S}_n^b$ converges to zero, uniformly. Thus we must show that under Conditions (A1) and (A2),

$$\widehat{S}_n^b(\cdot) \Rightarrow S_\nu(\cdot)\,. \tag{6.14}$$

This will be a simple corollary of Theorem 2.2.24. Recall that

$$k_n(t) \equiv \lfloor \lfloor a_n t \rfloor / \theta_n \rfloor, \tag{6.15}$$

and, for $i \geq 1$, define

$$Z_{n,i} \equiv \sum_{j=\theta_n(i-1)+1}^{\theta_n i} c_n^{-1}\lambda_n^{-1}(J_n(j))e_{n,j}\,. \tag{6.16}$$

By (6.7) and (6.13), $\widehat{S}_n^b(t) = \sum_{i=1}^{k_n(t)} Z_{n,i}$. We now want to apply Theorem 2.2.24 to the latter partial sum process. For this let $\{\mathcal{F}_{n,i}, n \geq 1, i \geq 0\}$ be the array of sub-sigma fields of $\mathcal{F}^X$ defined by (with obvious notations) $\mathcal{F}_{n,i} = \sigma\left(\cup_{j\leq \theta_n i}\{J_n(j), e_{n,j}\}\right)$, for $i \geq 0$. Clearly, for each n and $i \geq 1$, $Z_{n,i}$ is $\mathcal{F}_{n,i}$ measurable and $\mathcal{F}_{n,i-1} \subset \mathcal{F}_{n,i}$. Next observe that

$$\mathcal{P}_{\mu_n}\left(Z_{n,i} > z \,\middle|\, \mathcal{F}_{n,i-1}\right) = \sum_{x\in\mathcal{V}_n} \mathcal{P}_{\mu_n}\Big(J_n\big(\theta_n(i-1)+1\big) = x, Z_{n,i} > z \,\Big|\, \mathcal{F}_{n,i-1}\Big), \tag{6.17}$$

where

$$\begin{aligned} &\mathcal{P}_{\mu_n}\Big(J_n\big(\theta_n(i-1)+1\big) = x, Z_{n,i} > z \,\Big|\, \mathcal{F}_{n,i-1}\Big) \\ &= \mathcal{P}_{\mu_n}\Big(J_n\big(\theta_n(i-1)+1\big) = x, Z_{n,i} > z \,\Big|\, J_n\big(\theta_n(i-1)\big)\Big). \end{aligned} \tag{6.18}$$

Using Bayes' Theorem and the Markov property, the last line can be written as

$$p_n\big(J_n(\theta_n(i-1)), x\big)\mathcal{P}_{\mu_n}\left(\sum_{j=1}^{\theta_n} c_n^{-1}\lambda_n^{-1}(J_n(j-1))e_{n,j-1} > z \,\middle|\, J_n(0) = x\right). \tag{6.19}$$

Thus, in view of (6.2), (6.3), (6.4), and (6.5), it follows from (6.17), (6.18), and (6.19) that

$$\sum_{i=1}^{k_n(t)} \mathcal{P}_{\mu_n}\left(Z_{n,i} > z \mid \mathcal{F}_{n,i-1}\right) = \sum_{i=1}^{k_n(t)} \sum_{x\in\mathcal{V}_n} p_n\big(J_n(\theta_n(i-1)), x\big) Q_n^u(x)$$

$$= \sum_{i=1}^{k_n(t)} F_n^u\big(J_n(\theta_n(i-1))\big)$$
$$= \nu_n^{J,t}(u,\infty)\,. \tag{6.20}$$

Similarly we get

$$\sum_{i=1}^{k_n(t)} \left[\mathcal{P}_{\mu_n}\left(Z_{n,i} > \epsilon \mid \mathcal{F}_{n,i-1}\right)\right]^2 = \sum_{i=1}^{k_n(t)} \left[F_n^u\big(J_n(\theta_n(i-1))\big)\right]^2 = (\sigma_n^{J,t})^2(u,\infty)\,. \tag{6.21}$$

From (6.20) and (6.21) it follows that Conditions (A2) and (A1) of Theorem 6.1.1 are exactly the conditions from Theorem 2.2.24. Similarly Condition (A3) is Condition 5.3. Therefore the conditions of Theorem 2.2.24 are verified, and so $\widehat{S}_n^b \Rightarrow S_\nu$ in $D([0,\infty))$ where S_ν is a subordinator with Lévy measure ν and zero drift. □

We now come to the key step in our argument. This consists in reducing Conditions (A1) and (A2) of Theorem 6.1.1 to (i) a *mixing condition* for the chain J_n, and (ii) a *law of large numbers* for the random variables Q_n.

Again we formulate three conditions for given sequences a_n, c_n and a given realisation of the random environment.

Condition (A1-1). Let J_n is a periodic Markov chain with period q. There exists an integer sequence $\ell_n \in \mathbb{N}$, and a positive decreasing sequence ρ_n, satisfying $\rho_n \downarrow 0$ as $n \uparrow \infty$, such that, for all pairs $x, y \in \mathcal{V}_n$, and all $i \geq 0$,

$$\sum_{k=0}^{q-1} P_{\pi_n}\left(J_n(i+\ell_n+k) = y, J_n(0) = x\right) \leq (1+\rho_n)\pi_n(x)\pi_n(y)\,. \tag{6.22}$$

Condition (A2-1) There exists a measure ν as in Condition (A1) such that

$$\nu_n^t(u,\infty) \equiv k_n(t) \sum_{x\in\mathcal{V}_n} \pi_n(x) Q_n^u(x) \to t\nu(u,\infty)\,, \tag{6.23}$$

and

$$(\sigma_n^t)^2(u,\infty) \equiv k_n(t) \sum_{x\in\mathcal{V}_n} \sum_{x'\in\mathcal{V}_n} \pi_n(x) p_n^{(2)}(x,x') Q_n^u(x) Q_n^u(x') \to 0\,. \tag{6.24}$$

Condition (A3-1) For all $t > 0$,

$$\lim_{\epsilon\downarrow 0} \limsup_{n\uparrow\infty} k_n(t) \mathcal{E}_{\pi_n} \mathbb{1}_{\{\lambda_n^{-1}(J_n(0))e_0 \leq c_n\epsilon\}} c_n^{-1} \lambda_n^{-1}(J_n(0)) e_0 = 0. \tag{6.25}$$

Remark 6.1.4 The limiting measure ν may be deterministic or random.

Theorem 6.1.2 *Assume that for $\mu_n = \pi_n$ and for constants a_n, c_n, θ_n, Conditions (A1-1), (A2-1), (A3-1) and (A0') hold $\mathbb{P}$-a.s., resp. in $\mathbb{P}$-probability. Then the sequence of random stochastic process S_n^b converges to the process S_ν, weakly on the Skorokhod space $D[0,\infty)$ equipped with the J_1-topology, $\mathbb{P}$-almost surely, resp. in $\mathbb{P}$-probability.*

Proof. The proof of Theorem 6.1.2 comes in two steps. In the first we use the ergodic properties of the chain J_n to pass from sums along a chain J_n to averages with respect to the invariant measure of J_n.

We assume from now on that the initial distribution μ_n is the invariant measure π_n of the jump chain J_n.

Proposition 6.1.3 *Let $\mu_n = \pi_n$. Assume that Condition (A1-1) is satisfied. Then, choosing $\theta_n \geq \ell_n$, the following holds: for all $t > 0$ and all $u > 0$ we have that, for all $\epsilon > 0$,*

$$P_{\pi_n}\left(\left|\nu_n^{J,t}(u,\infty) - \nu_n^t(u,\infty)\right| \geq \epsilon\right) \leq \epsilon^{-2}\left[\rho_n\left(\nu_n^t(u,\infty)\right)^2 + (\sigma_n^t)^2(u,\infty)\right], \tag{6.26}$$

and

$$P_{\pi_n}\left((\sigma_n^{J,t})^2(u,\infty) \geq \epsilon\right) \leq \epsilon^{-1}(\sigma_n^t)^2(u,\infty). \tag{6.27}$$

Proof. To simplify notation, we only give the proof for the case when the chain J_n is aperiodic, i.e. $q = 1$. Details of how to deal with the general periodic case can be found in the proof of Proposition 4.1. of [23].

Let us first establish that

$$E_{\pi_n}\left[\nu_n^{J,t}(y)\right] = \nu_n^t(u,\infty), \tag{6.28}$$

$$E_{\pi_n}\left[(\sigma_n^{J,t})^2(u,\infty)\right] = (\sigma_n^t)^2(u,\infty). \tag{6.29}$$

To this end set

$$\pi_n^{J,t}(x) = k_n^{-1}(t)\sum_{j=1}^{k_n(t)} \mathbb{1}_{\{J_n(\theta_n(j-1))=x\}}, \quad x \in \mathcal{V}_n. \tag{6.30}$$

Then, Eqs. (6.4) and (6.5) may be rewritten as

$$\nu_n^{J,t}(u,\infty) = k_n(t) \sum_{y\in\mathcal{V}_n} \pi_n^{J,t}(y) F_n^u(y)\,, \tag{6.31}$$

$$(\sigma_n^{J,t})^2(u,\infty) = k_n(t) \sum_{y\in\mathcal{V}_n} \pi_n^{J,t}(y) \left(F_n^u(y)\right)^2 \,. \tag{6.32}$$

Since by assumption the initial distribution is the invariant measure π_n of J_n, the chain variables $(J_n(j), j \geq 1)$ satisfy $P_{\pi_n}(J_n(j) = x) = \pi_n(x)$ for all $x \in \mathcal{V}_n$, and all $j \geq 1$. Hence

$$E_{\pi_n}\left[\pi_n^{J,t}(y)\right] = \pi_n(y)\,, \tag{6.33}$$

$$E_{\pi_n}\left[\nu_n^{J,t}(u,\infty)\right] = k_n(t) \sum_{x\in\mathcal{V}_n} \pi_n(x) F_n^u(x)\,, \tag{6.34}$$

$$E_{\pi_n}\left[(\sigma_n^{J,t})^2(u,\infty)\right] = k_n(t) \sum_{x\in\mathcal{V}_n} \pi_n(x) \left(F_n^u(x)\right)^2 \,, \tag{6.35}$$

Eqs. (6.28) and (6.29) now follow readily from these identities. Indeed, inserting (6.3) into (6.34) and using that π_n is the invariant measure of J_n, we get,

$$E_{\pi_n}\left[\nu_n^{J,t}(u,\infty)\right] = k_n(t) \sum_{y\in\mathcal{V}_n} \sum_{x\in\mathcal{V}_n} \pi_n(x) p_n(x,y) Q_n^u(y)\,, \tag{6.36}$$

$$= k_n(t) \sum_{y\in\mathcal{V}_n} \pi_n(y) Q_n^u(y)\,, \tag{6.37}$$

which proves (6.28). Similarly, inserting (6.3) into (6.35) yields

$$E_{\pi_n}\left[(\sigma_n^{J,t})^2(u,\infty)\right] = k_n(t) \sum_{x\in\mathcal{V}_n} \pi_n(x) \left(\sum_{y\in\mathcal{V}_n} p_n(x,y) Q_n^u(y) \right)^2 \tag{6.38}$$

which gives (6.29) once observed that, by reversibility,

$$\sum_{x\in\mathcal{V}_n} \pi_n(x) p_n(x,y) p_n(x,y') = \pi_n(y) \sum_{x\in\mathcal{V}_n} p_n(y,x) p_n(x,y') = \pi_n(y) p_n^{(2)}(y,y'). \tag{6.39}$$

We are now ready to prove the proposition. In view of (6.29), (6.27) is nothing but a first order Chebychev inequality. To establish (6.26) set

$$\Delta_{ij}(x,y) = P_{\pi_n}\left(J_n(\theta_n(i-1)) = x, J_n(\theta_n(j-1)) = y\right) - \pi_n(x)\pi_n(y)\,. \tag{6.40}$$

A second order Chebychev inequality together with the expressions (6.34) for $E_{\pi_n}\left[\nu_n^{J,t}(u,\infty)\right]$ yields

$$P_{\pi_n}\left(\left|\nu_n^{J,t}(u,\infty)-E_{\pi_n}\left[\nu_n^{J,t}(u,\infty)\right]\right|\geq\epsilon\right) \tag{6.41}$$
$$\leq\epsilon^{-2}E_{\pi_n}\Big[k_n(t)\sum_{y\in\mathcal{V}_n}\left(\pi_n^{J,t}(y)-\pi_n(y)\right)F_n^u(y)\Big]^2$$
$$=\epsilon^{-2}\sum_{x\in\mathcal{V}_n}\sum_{y\in\mathcal{V}_n}F_n^u(x)F_n^u(y)\sum_{i=1}^{k_n(t)}\sum_{j=1}^{k_n(t)}\Delta_{ij}(x,y)\,.$$

Now $\sum_{i=1}^{k_n(t)}\sum_{j=1}^{k_n(t)}\Delta_{ij}(x,y)=(\overline{I})+(\overline{II})$ where

$$(\overline{I})\equiv\sum_{i=1}^{k_n(t)}\sum_{j=1}^{k_n(t)}\Delta_{ij}(x,y)1\!\!1_{\{j\neq i\}}\leq\rho_nk_n^2(t)\pi_n(x)\pi_n(y)\,, \tag{6.42}$$

as follows from Assumption (A1-1), choosing $\theta_n\geq\ell_n$, and

$$(\overline{II})\leq\sum_{1\leq i\leq k_n(t)}\Delta_{ii}(x,x)1\!\!1_{\{x=y\}} \tag{6.43}$$
$$=k_n(t)\left[P_{\pi_n}\left(J_n(\theta_n(i-1))=x\right)-\pi_n^2(x)\right]1\!\!1_{\{x=y\}}$$
$$=k_n(t)\pi_n(x)(1-\pi_n(x))1\!\!1_{\{x=y\}}\leq k_n(t)\pi_n(x)1\!\!1_{x=y}.$$

Inserting (6.43) and (6.42) in (6.41) we obtain, using again (6.29) and (6.33), that

$$P_{\pi_n}\left(\left|\nu_n^{J,t}(u,\infty)-E_{\pi_n}\left[\nu_n^{J,t}(u,\infty)\right]\right|\geq\epsilon\right)$$
$$\leq\epsilon^{-2}\left[\rho_n\left(\nu_n^t(u,\infty)\right)^2+(\sigma_n^t)^2(u,\infty)\right]\,. \tag{6.44}$$

Proposition 6.1.3 is proven. □

The proof of Theorem 6.1.2 is now immediate: combine the conclusions of Proposition 6.1.3 with Condition (A2-1) to get both conditions (A1) and (A2). Finally, Condition (A3) is Condition (A3-1), since we are starting from the invariant measure. □

6.2 Application to quenched ageing in the REM

We will now show how Theorem 6.1.2 can be applied in the REM to obtain quenched results. i.e. results for fixed environments. There are other ways to get these results, see [24], and so this section is mainly of pedagogical interest.

We recall that we live on the hypercube $\mathcal{V}_n=\Sigma_n=\{-1,1\}^n$. The random environment is given by iid normal variables $H_n(x)$ and the mean holding times, $\tau(x)$, are given by

$$\tau(x) \equiv \exp(\beta H_n(x)), \tag{6.45}$$

with $\beta \in \mathbb{R}$ the inverse temperature. The Markov chain J_n will again be the simple random walk on Σ_n, i.e.

$$p_n(x, x') = \begin{cases} \frac{1}{n}, & \text{if } \|x - x'\|_2^2 = 2, \\ 0, & \text{else}, \end{cases} \tag{6.46}$$

Theorem 6.2.4 *For all γ satisfying*

$$0 < \gamma < \beta^2 \wedge 2\ln 2, \tag{6.47}$$

the law of the stochastic process

$$S_n^b(t) \equiv e^{-\gamma n} S_n \left(\theta_n \left[t \tfrac{\beta\sqrt{2\pi}}{\Gamma(1+\gamma^2/\beta)} n^{1/2} e^{n\gamma^2/2\beta^2} \theta_n^{-1} \right] \right), \qquad t \geq 0, \tag{6.48}$$

with $\theta_n = \frac{3\ln 2}{2} n^2$, defined on the space of càdlàg functions equipped with the Skorokhod J_1-topology, converges to the law of γ/β^2-stable subordinator $V_{\gamma/\beta^2}(t), t \geq 0$. Convergence holds $\mathbb{P}$-almost surely.

Proof. To prove the theorem, we just need to see how the Conditions (A1-1), (A2-1), and (A3-1) can be verified.

This requires four steps, two if which are quite immediate:

Conditions (A1-1) for simple random walk has been established e.g. in [2] and [24]. The following lemma is taken from Proposition 3.12 of [24].

Lemma 6.2.5 *Let P_{π_n} be the law of the simple random walk on the hypercube Σ_n started in the uniform distribution. Let $\theta_n = \frac{3\ln 2}{2} n^2$. Then, for any $x, y \in \Sigma_n$ and any $i \geq 0$,*

$$\left| \sum_{k=0}^{1} P_{\pi_n} \left(J_n(\theta_n + i + k) = y, J_n(0) = x \right) - 2^{-2n+1} \right| \leq 2^{-3n+1}. \tag{6.49}$$

Clearly this implies that Condition (A1-1) holds.

Next, the second part of Condition (A2-1) will follow rather easily, once we have proven the first.

Thus, we need to show that

$$\nu_n^t(u, \infty) \to \nu^t(u, \infty) = \Gamma(1+\alpha) t u^{-\gamma/\beta^2}, \tag{6.50}$$

almost surely, as $n \uparrow \infty$.

Laplace transforms. Instead of proving the convergence of the distribution functions ν_n^t directly, we will pass to their Laplace transforms, prove their convergence and then use Feller's continuity lemma to deduce convergence of the original objects.

For $v > 0$, consider the Laplace transforms

$$\hat{\nu}_n^t(v) = \int_0^\infty du e^{-uv} \nu_n^t(u,\infty) \tag{6.51}$$
$$\hat{\nu}^t(v) = \int_0^\infty du e^{-uv} \nu^t(u,\infty).$$

With $Z_n \equiv \sum_{j=0}^{\theta_n - 1} c_n^{-1} \lambda_n^{-1}(J_n(j)) e_j$, we have, by definition of $\nu_n^t(u,\infty)$,

$$\nu_n^t(u,\infty) = k_n(t) \sum_{x \in \mathcal{V}_n} \pi_n(x) Q_n^u(x) = k_n(t) \mathcal{P}_{\pi_n}(Z_n > u).$$

Hence

$$\begin{aligned} \hat{\nu}_n^t(v) &= \int_0^\infty du e^{-uv} \nu_n^t(u,\infty) \\ &= k_n(t) \int_0^\infty du e^{-uv} \mathcal{P}_{\pi_n}(Z_n > u) \\ &= k_n(t) \frac{1 - \mathcal{E}_{\pi_n}\left(e^{-vZ_n}\right)}{v}, \end{aligned} \tag{6.52}$$

where the last equality follows by integration by parts.

Convergence of $\mathbb{E}\hat{\nu}_n^t(v)$.

Lemma 6.2.6 *Let $c_n = e^{\gamma n}$, $a_n = n^{1/2} e^{n\gamma^2/2\beta^2}$. For any $\beta, \gamma > 0$ such that $\gamma/\beta^2 \in (0,1)$, for any $v > 0$,*

$$\lim_{n \uparrow \infty} k_n(t) \mathbb{E}\left[1 - \mathcal{E}_{\pi_n}\left(e^{-vZ_n}\right)\right] = d_1 t v^{\gamma/\beta^2}. \tag{6.53}$$

for a constant $d_1 = \Gamma(1-\alpha)\Gamma(1+\alpha)$.

Proof. We will set $U_i \equiv H_n(J_n(i))$. The Laplace transforms we are after can be written, after integrating out the exponential variables e_i, as

$$\mathbb{E}\mathcal{E}_{\pi_n} e^{-vZ_n} = E_{\pi_n} \mathbb{E} G(U, v, \theta_n), \tag{6.54}$$

where

$$G(U, v, k) \equiv \prod_{i=0}^{k-1} \frac{1}{1 + v c_n^{-1} e^{\beta\sqrt{n} U_i}}, \tag{6.55}$$

Let us denote by $n_r(J)$ the number of points that the chain J visits r times up to time θ_n. Clearly, $\sum_{r=1}^{\theta_n} rn_r = \theta_n$. Clearly, the expectation of $G(U.v,\theta_n)$ depends only on J only though these numbers. In fact,

$$E_{\pi_n}\mathbb{E}G(U,v,\theta_n) = \sum_{\{n_r\}} P_{\pi_n}\left(n_r(J) = n_r, \forall r\right) \times \prod_{r=1}^{\theta_n}\left[\mathbb{E}e^{-rg\left(vc_n^{-1}e^{\beta\sqrt{n}U}\right)}\right]^{n_r}. \tag{6.56}$$

Next we compute, using partial integration [1],

$$\begin{aligned}\mathbb{E}e^{rg\left(vc_n^{-1}e^{\beta\sqrt{n}U}\right)} &= 1 - \int_0^\infty r\left(\frac{1}{1+x}\right)^{r+1}\mathbb{P}\left(vc_n^{-1}e^{\beta\sqrt{n}U} > x\right)dx \\ &\sim 1 - a_n^{-1}v^\alpha r\int_0^\infty\left(\frac{1}{1+x}\right)^{r+1}x^{-\alpha}dx \\ &= 1 - a_n^{-1}v^\alpha r d_r,\end{aligned} \tag{6.57}$$

where

$$d_r = \Gamma(1-\alpha)\frac{\Gamma(r+\alpha)}{\Gamma(r+1)}, \tag{6.58}$$

is a decreasing sequence. Hence we get readily that

$$\begin{aligned}&k_n(t)\left(1 - E_\pi\mathbb{E}G(U.v,\theta_n)\right) \\ &= k_n(t)\sum_{\{n_r\}}P_{\pi_n}\left(n_r(J) = n_r,\forall r\right)\left(1 - \prod_{r=1}^{\theta_n}\left(1 - a_n^{-1}v^\alpha r d_r\right)^{n_r}\right) \\ &= v^\alpha t\sum_{\{n_r\}}P_{\pi_n}\left(n_r(J) = n_r,\forall r\right)\theta_n^{-1}\sum_{r=1}^{\theta_n} rn_r d_r + O(a_n^{-1}) \\ &= v^\alpha t d_1 \\ &+ v^\alpha\sum_{\{n_r\}}P_{\pi_n}\left(n_r(J) = n_r,\forall r\right)\theta_n^{-1}\sum_{r=1}^{\theta_n} rn_r(d_r - d_1) + O(a_n^{-1})\end{aligned} \tag{6.59}$$

To conclude, we only need to show that the probability under P_{π_n} that say, $n_1 \leq \theta_n(1-n^{-1/2})$, tends to zero.

Lemma 6.2.7 *For the SRW, J, on Σ_n, we have that $\mathbb{E}n_1(J) \geq \theta_n(1-2/n)$. Consequently, $\mathbb{P}\left(n_1(J) < 1-n^{-1/2}\right) \leq n^{-1/2}$.*

[1] The passage to the limit in the equation can easily be justified using Lebesgue's dominated convergence and the very explicit control over the asymptotics of the Gaussian probability appearing.

Proof. Clearly,

$$\mathbb{E}n_1(J) \geq \sum_{k=1}^{\theta_n} \mathbb{I}_{J_n(\ell)\neq J_n(k), \forall k<\ell\leq\theta_n} \tag{6.60}$$

$$\geq \theta_n - \sum_{k=1}^{\theta_n} \mathcal{P}_{J_n(k)}\left(\tau_{J_n(k)} < \theta_n\right)$$

$$= \theta_n \left(1 - \mathbb{P}_{\mathbf{1}}\left(\tau_{\mathbf{1}} < \theta_n\right)\right). \tag{6.61}$$

The last probability can be computed in the one-dimensional chain $m_n(k) = \frac{1}{n}\sum_{i=1}^n J_{n,i}(k)$, the classical Ehrenfest chain on the state space $\{-1, -1+2/n, \ldots, 1-2/n, 1\}$. Then

$$\mathbb{P}_{\mathbf{1}}\left(\tau_{\mathbf{1}} < \theta_n\right) = \mathbb{P}_1\left(\tau_1 < \theta_n\right) \tag{6.62}$$

$$\leq \mathcal{P}_1\left(\tau_1 = 2\right) + \mathbb{P}_{1-4/n}\left(\tau_1 < \theta_n\right).$$

The first probability is equal to $1/n$. The second can be decomposed as

$$\mathbb{P}_{1-4/n}\left(\tau_1 < \theta_n\right) \leq \mathbb{P}_{1-4/n}\left(\tau_1 < \tau_0\right) + \mathbb{P}_{1-4/n}\left(\tau_1 < \theta_n, \theta_0 < \tau_1\right) \tag{6.63}$$

$$\leq \mathbb{P}_{1-4/n}\left(\tau_1 < \tau_0\right) + \mathbb{P}_0\left(\tau_1 < \theta_n\right).$$

The first probability can be computed exactly and yields

$$\mathbb{P}_{1-4/n}\left(\tau_1 < \tau_0\right) \leq cn^2,$$

whereas the second can be estimated a

$$\mathbb{P}_0\left(\tau_1 < \theta_n\right) \leq \theta_n / \mathbb{E}_0\tau_1 \leq \theta_n 2^{-n}.$$

This yields the claimed estimate on the expectation of $n_1(J)$. Chebeychev does the rest. □

□

Concentration of ν_n^t. To conclude the proof, we need to control the fluctuations of ν_n^t.

Lemma 6.2.8 *Under the same hypothesis as in Lemma 6.2.6,*

$$\mathbb{E}\left(\hat{\nu}_n^t(v) - \mathbb{E}\hat{\nu}_n^t(v)\right)^2 \leq Ct^2\theta_n a_n 2^{-n} v^\alpha. \tag{6.64}$$

Proof. We have to compute

$$\mathbb{E}\left(\mathcal{E}_{\pi_n} e^{-vZ_n} - \mathbb{E}\mathcal{E}_{\pi_n} e^{-vZ_n}\right)^2 \tag{6.65}$$

$$= \mathbb{E}\mathbb{E}' E_{\pi_n} E'_{\pi_n}\left(e^{-vZ_n(J)} e^{-vZ_n(J')} - e^{-vZ_n(J)} e^{-vZ'_n(J')}\right). \tag{6.66}$$

where we made the reference to the two walks explicit and where Z'_n depends on an independent copy, H'_n, of the random variables H_n. Now it is clear that the expectations of the two terms differ only if the two random walks J and J' intersect. On the other hand, using the computations from the preceding proof, it is obvious that we have the uniform bound

$$\left|\mathbb{E}\mathbb{E}' e^{-vZ_n(J)} e^{-vZ_n(J')} - 1\right| \leq d_1 2\theta_n a_n^{-1} v^{\alpha}(1 + (o1)). \tag{6.67}$$

Hence we obtain that

$$\begin{aligned} & k_n(t)^2 \mathbb{E}\left(\mathcal{E}_{\pi_n} e^{-vZ_n} - \mathbb{E}\mathcal{E}_{\pi_n} e^{-vZ_n}\right)^2 \\ & \leq t^2 d_1 2\theta_n^{-1} a_n v^{\alpha}(1+(o1)) P_{\pi_n} \otimes P'_{\pi_n}\left(\{J_n(i)\}_{i=0}^{\theta_n} \cap \{J_n(i)\}_{i=0}^{\theta_n} \neq \emptyset\right). \end{aligned} \tag{6.68}$$

But

$$\begin{aligned} & P_{\pi_n} \otimes P'_{\pi_n}\left(\{J_n(i)\}_{i=0}^{\theta_n} \cap \{J'_n(i)\}_{i=0}^{\theta_n} \neq \emptyset\right) \\ & \leq \sum_{i,j=0}^{\theta_n} P_{\pi_n} \otimes P'_{\pi_n}\left(J_n(i) = J'_n(j)\right) = \theta_n^2 2^{-n}. \end{aligned} \tag{6.69}$$

Here we used that both chains are in equilibrium. Inserting this estimate into (6.68) yields the claim of the lemma. □

Conclusion of the proof. Lemmata 6.2.6 and 6.2.8, together with Chebychev's inequality and the Borel-Cantelli lemma, establish that, for each $v > 0$,

$$\lim_{n\to\infty} \hat{\nu}_n^t(v) = \hat{\nu}^t(v) = d_1 t v^{\gamma/\beta^2 - 1}, \quad \mathbb{P} - \text{a.s.} \tag{6.70}$$

Together with the monotonicity of $\hat{\nu}_n^t(v)$ and the continuity of the limiting function $\hat{\nu}^t(v)$, this implies that there exists a subset $\Omega_1^\tau \subset \Omega^\tau$ of the sample space Ω^τ of the τ's with the property that $\mathbb{P}(\Omega_1^\tau) = 1$, and such that, on Ω_1^τ,

$$\lim_{n\to\infty} \hat{\nu}_n^t(v) = \hat{\nu}^t(v), \quad \forall\, v > 0. \tag{6.71}$$

Finally, applying Feller's Extended Continuity Theorem for Laplace transforms of (not necessarily bounded) positive measures (see [22], Theorem 2a, Section XIII.1, p. 433) we conclude that, on Ω_1^τ,

$$\lim_{n\to\infty} \nu_n^t(u,\infty) = \nu^t(u,\infty) = \Gamma(1+\alpha) t u^{-\gamma/\beta^2}, \quad \forall\, u > 0. \tag{6.72}$$

Thus we have established Conditions (A1-1) and (A2-1) under the stated conditions on the parameters γ, β.

It remains to show that Condition (A3-1) holds. We first show that $a_n c_n^{-1} \mathbb{E}\mathcal{E}_{\pi_n} \lambda_n^{-1}(\sigma) e_1 \mathbb{1}_{\lambda_n^{-1}(\sigma)e_1 \leq \epsilon c_n}$ is bounded by $\epsilon^{1-\alpha}$ in average, and then prove a concentration estimate .

Lemma 6.2.9 *Under the Assumptions of the theorem, there is a constant $K < \infty$, such that*

$$\limsup_{n\uparrow\infty} a_n c_n^{-1} \mathbb{E}\mathcal{E}_{\pi_n} \lambda_n^{-1}(\sigma) e_1 \mathbb{1}_{\lambda_n^{-1}(\sigma)e_1 \leq \epsilon c_n} \leq K\epsilon^{1-\alpha}. \tag{6.73}$$

Proof. The proof is through explicit estimates. We must control the integral

$$\begin{aligned} &\int_0^\infty e^{-x} dx \int_{-\infty}^\infty e^{-\frac{z^2}{2}} \mathbb{1}_{xe^{\beta\sqrt{n}z} \leq ec_n} e^{\beta\sqrt{n}z} dz \\ &= \int_0^\infty e^{-x} dx \left[\int_{-\infty}^{\frac{\ln c_n + \ln(\epsilon/x)}{\beta\sqrt{n}}} e^{-\frac{z^2}{2}+\beta\sqrt{n}z} dz \right] \\ &= \int_0^\infty e^{-x} dx \left[e^{\beta^2 n/2} \int_{-\infty}^{\frac{\ln c_n + \ln(\epsilon/x)}{\beta\sqrt{n}} - \beta\sqrt{n}} e^{-\frac{z^2}{2}} dz \right] \end{aligned} \tag{6.74}$$

Now for our choice $c_n = \exp(\gamma n)$, the upper integration limit in the z-integral is

$$\frac{\ln c_n + \ln(\epsilon/x)}{\beta\sqrt{n}} - \beta\sqrt{n} = \sqrt{n}\left(\frac{\gamma}{\beta} - \beta\right) + \frac{\ln\epsilon - \ln x}{\beta\sqrt{n}}. \tag{6.75}$$

Thus, for any $\gamma < \beta^2$, this tends to $-\infty$ uniformly for, say, all $x \leq n^2$. We therefore decompose the x-integral in the domain $x \leq n^2$ and its complement, and use first that

$$\int_{n^2}^\infty e^{-x} dx e^{-\frac{z^2}{2}} \mathbb{1}_{xe^{\beta\sqrt{n}z} \leq ec_n} e^{\beta\sqrt{n}z} dz \leq \epsilon c_n e^{-n^2}, \tag{6.76}$$

which tend to zero, as $n \uparrow \infty$. For the remainder we use the well-known bound, for $u > 0$,

$$\int_u^\infty e^{-z^2/2} \leq \frac{1}{u} e^{-u^2/2}. \tag{6.77}$$

This yields

$$e^{\beta^2 n/2} \int_{-\infty}^{\frac{\ln c_n + \ln(\epsilon/x)}{\beta\sqrt{n}} - \beta\sqrt{n}} e^{-\frac{z^2}{2}} dz \tag{6.78}$$
$$\leq e^{\beta^2 n/2} \frac{\exp\left(-\frac{1}{2}\left(\sqrt{n}\left(\frac{\gamma}{\beta} - \beta\right) + \frac{\ln\epsilon - \ln x}{\beta\sqrt{n}}\right)^2\right)}{(\beta^{-1}\gamma - \beta)\sqrt{n} + \frac{\ln\epsilon - \ln x}{\beta\sqrt{n}}}$$
$$= \frac{\exp\left(-n\frac{\gamma^2}{2\beta^2} + n\gamma\right)}{\sqrt{n}(\gamma/\beta - \beta) + o(1)} \exp\left(-(\gamma/\beta^2 - 1)\ln(\epsilon/x) + O(n^{-1/2}\right).$$

Hence

$$\limsup_{n\uparrow\infty} a_n c_n^{-1} \int_0^\infty e^{-x} dx \int_{-\infty}^\infty e^{-\frac{z^2}{2}} \mathbb{1}_{x e^{\beta\sqrt{n}z} \leq \epsilon c_n} e^{\beta\sqrt{n}z} dz \tag{6.79}$$
$$\leq \frac{1}{\gamma/\beta - \beta} \epsilon^{1-\alpha} \int_0^\infty x^{\alpha-1} e^{-x} dx,$$

which implies the assertion of the lemma since the last integral is finite. □

Lemma 6.2.10 *Under the Assumptions of the theorem, there is a constant $K < \infty$, such that*

$$a_n^2 c_n^{-2} \mathbb{E}\left(\mathcal{E}_{\pi_n} \lambda_n^{-1}(\sigma) e_1 \mathbb{1}_{\lambda_n^{-1}(\sigma)e_1 \leq \epsilon c_n} - \mathbb{E}\mathcal{E}_{\pi_n} \lambda_n^{-1}(\sigma) e_1 \mathbb{1}_{\lambda_n^{-1}(\sigma)e_1 \leq \epsilon c_n}\right)$$
$$\leq K\epsilon^{1-\alpha} a_n 2^{-n}. \tag{6.80}$$

Proof. The proof of this lemma is very similar to that of Lemma 6.2.8 and will be left to the reader. □

Now Theorem 6.2.4 follows from Theorem 6.1.2. □

6.2.1 Consequences for correlation functions

Finally, one must ask whether the convergence of the clock process in the form obtained here is useful for deriving ageing information in the sense that we can control the behaviour of certain correlation functions. One may be worried that a jump in limit of the coarse-grained clock process refers to a period of time during which the process still may make n^2 steps, and our limit result tells us nothing about how the process moves during that time. We will, however, show that essentially all this time is spent in a single site.

This allows to prove

Theorem 6.2.11 *Let $A_n(t,s)$ be the event defined by*

$$A_n(t,s) = \{\sigma_n(te^{\gamma n}) = \sigma_n((t+s)e^{\gamma n})\,. \tag{6.81}$$

Then, under the hypothesis of Theorem 6.2.4, for all $t > 0$ and $s > 0$,

$$\lim_{N\to\infty} \mathcal{P}_{\pi_n}(A_n(t,s)] = \frac{\sin\alpha\pi}{\pi}\int_0^{t/(t+s)} u^{\alpha-1}(1-u)^{-\alpha}\,du, \mathcal{P}-\text{a.s.}. \tag{6.82}$$

Proof. The proof of this theorem relies on the following simple estimate. Let us denote by $\mathcal{R}_n$ the range of the coarse grained and rescaled clock process S_n^b. The argument of [2] in the proof of Theorem 1.2 that the event $A_n(s,t)\cap\{\mathcal{R}_n\cap(s,t)\neq\emptyset\}$ has vanishing probability carries over unaltered.

What we have to show is that if the process spends the whole time from s to t within one block, then almost all of this time is spent, without interruption, in a single site.

The following lemma is quite obvious.

Lemma 6.2.12 *Let $\mathcal{M}_n \subset \Sigma_n$ be arbitrary. Then*

$$\begin{aligned}&\mathbb{P}\left(\exists_{\sigma\neq\sigma'\in\mathcal{M}_n} : H_\sigma \geq an \wedge H_\sigma \geq an/2\right)\\ &\leq |\mathcal{M}_n|^2 e^{-na^2/2}e^{-na^2/8}\end{aligned} \tag{6.83}$$

This lemma implies that even if in a set of size say n^2, there is a point where $H_\sigma \geq na$, then there will not be a second point of comparable size in that set, with overwhelming probability.

This means the following: within a block of θ_n steps of the chain J_n, that gives a contribution to a jump, there is only one site that contributes to the time. It remains to show that these contributions come in one "block", i.e. the process will not return to this site once it left it within θ_n steps. But this is an elementary property of the random walk on the hypercube.

Let us make this precise. As remarked above,

$$\begin{aligned}\mathcal{P}_{\pi_n}(A_n(s,t)) &= \mathcal{P}_{\pi_n}(A_n(s,t)\cap\{\mathcal{R}_n\cap(s,t)=\emptyset\})\\ &\quad+\mathcal{P}_{\pi_n}(A_n(s,t)\cap\{\mathcal{R}_n\cap(s,t)\neq\emptyset\})\,,\end{aligned} \tag{6.84}$$

where the second term tends to zero. Next we observe that

$$\begin{aligned}&\mathcal{P}_{\pi_n}(A_n(s,t)\cap\{\mathcal{R}_n\cap(s,t)=\emptyset\})\\ &=\mathcal{P}_{\pi_n}(\mathcal{R}_n\cap(s,t)=\emptyset)-\mathcal{P}_{\pi_n}((A_n(s,t))^c\cap\{\mathcal{R}_n\cap(s,t)=\emptyset\})\end{aligned} \tag{6.85}$$

Here the first term is what we want. To show that the second term tends to zero, we proceed as follows.

For any $N < \infty$, we clearly have

$$\begin{aligned} &\mathcal{P}_{\pi_n}\left((A_n(s,t))^c \cap \{\mathcal{R}_n \cap (s,t) = \emptyset\}\right) \qquad (6.86)\\ &= \sum_{k=0}^{k_n(N)-1} \mathcal{P}_{\pi_n}\left(((A_n(s,t))^c) \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}\right)\\ &+ \sum_{k=k_n(N)}^{\infty} \mathcal{P}_{\pi_n}\left(((A_n(s,t))^c) \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}\right). \end{aligned}$$

The second term is bounded by

$$\begin{aligned} &\sum_{k=k_n(N)}^{\infty} \mathcal{P}_{\pi_n}\left(((A_n(s,t))^c) \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}\right)\\ &\leq \mathcal{P}_{\pi_n}\left(S_n^b(N) \leq s\right) \to \mathcal{P}\left(V_\alpha(N) \leq s\right), \qquad (6.87) \end{aligned}$$

where convergence is almost sure with respect to the environment. The last probability can be made as small as desired by choosing N sufficiently large. It remains to deal with the first sum on the right-hand side of (6.86).

Define the event

$$\begin{aligned} &\mathcal{G}(k) \equiv \qquad (6.88)\\ &\bigcup_{\substack{k\theta_n \leq i<j<(k+1)\theta_n\\ J_n(i)\neq J_n(j)}} \left\{\lambda_n^{-1}(J_n(i))e_i \geq c_n(t-s)\theta_n^{-1}\right\} \cap \left\{\lambda_n^{-1}(J_n(j))e_j \geq \sqrt{c_n}\right\}. \end{aligned}$$

Using Eq. (4.106), we see that

$$\begin{aligned} \mathbb{E}\mathcal{P}_{\pi_n}\left(\mathcal{G}(k)\right) &\leq \theta_n^2 \mathbb{E}\exp\left(-c_n(t-s)\theta_n^{-1}e^{\beta\sqrt{n}H_\sigma}\right)\mathbb{E}\exp\left(-\sqrt{c_n}\theta_n^{-1}e^{\beta\sqrt{n}H_\sigma}\right)\\ &\sim \frac{(t-s)^{-\alpha}}{\beta^2 2\pi n}\theta_n^2\Gamma(1+\alpha)\Gamma(1+\alpha/2)e^{-n5\gamma/4} \leq \theta_n^2 a_n^{-1}e^{-n\gamma/4}. \qquad (6.89) \end{aligned}$$

Thus even

$$\mathbb{E}\mathcal{P}_{\pi_n}\left(\bigcup_{k=1}^{k_n(t)}\mathcal{G}(k)\right) \leq t\theta_n e^{-n\gamma/4}, \qquad (6.90)$$

which tends to zero exponentially fast. This implies that $\mathcal{P}_{\pi_n}\left(\bigcup_{k=1}^{k_n(t)}\mathcal{G}(k)\right)$ tends to zero $\mathbb{P}$-almost surely.

On the other hand, on the event $\mathcal{G}(k)^c \cap (A_n(s,t))^c \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}$, the following must be true: First, there still must exist some i such that $\lambda_n^{-1}(J_n(i))e_i \geq c_n(t-s)$, and second, the random walk must return to this site after visiting it.

Since obviously, all the probabilities of all events are independent of k, we consider in the sequel $k=0$ only, to simplify notations. In fact we decompose disjointly

$$\begin{aligned}
&\mathcal{G}(0)^c\left(A_n(s,t)\right)^c \cap \left\{(s,t)\subset \left(\bar{S}_n(0),\bar{S}_n(1)\right)\right\} = \qquad (6.91)\\
&\cup_{j=0}^{\theta_n-1}\cup_{\ell=2}^{\theta_n-i}\left\{\sum_{i=0}^{\theta_n-1} e_i\lambda_n^{-1}(J_n(i)) > c_n(t-s)\right\}\\
&\cap\left\{\lambda_n^{-1}(J_n(j)) > c_n(t-s)/\theta_n\right\}\cap\left\{\#\{i : J_n(i)=J_n(j)\}=\ell\}\right\}\cap\mathcal{G}(0)^c
\end{aligned}$$

Now

$$\begin{aligned}
&\mathbb{P}\otimes\mathcal{P}\left(\left\{\sum_{i=0}^{\theta_n-1} e_i\lambda_n^{-1}(J_n(i)) > c_n(t-s)\right\}\right. \qquad (6.92)\\
&\qquad\cap\left\{\lambda_n^{-1}(J_n(j)) > c_n(t-s)/\theta_n\right\}\\
&\qquad\left.\cap\left\{\#\{i : J_n(i)=J_n(j)\}=\ell\}\right\}\cap\mathcal{G}(0)^c\right)\\
&\leq \mathbb{P}\left(\sum_{r=2}^{\ell} e_r\lambda_n^{-1}(J_n(j)) > c_n(t-s)-\theta_n n^2\sqrt{c_n}\right)\\
&\qquad\times P_{\pi_n}\left(\#\{i : J_n(i)=J_n(j)\}=\ell\}\right) + O(\exp(-n^2)),
\end{aligned}$$

where the last term accounts for the probability one one of the exponential variables encountered could be larger that n^2. Now the random walk probability is, as we already know, bounded by

$$P_{\pi_n}\left(\#\{i : J_n(i)=J_n(j)\}=\ell\}\right) \leq Cn^{-\ell+1}. \qquad (6.93)$$

Moreover, the simplest estimate shows that

$$\begin{aligned}
&\mathbb{P}\left(\sum_{r=2}^{\ell} e_r\lambda_n^{-1}(J_n(j)) > c_n(t-s)-\theta_n n^2\sqrt{c_n}\right) \qquad (6.94)\\
&\leq \ell\mathbb{P}\left(e_1\lambda_n^{-1}(J_n(j)) > \frac{c_n(t-s)-\theta_n n^2\sqrt{c_n}}{\ell}\right)\\
&\leq a_n^{-1}(t-s)^{-\alpha}\ell^{1+\alpha}.
\end{aligned}$$

Combining this, we see that

$$\mathbb{E}\left(\sum_{k=0}^{k_n(N)} \mathcal{P}_{\pi_n}\left(\left(A_n(s,t)\right)^c \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}\right)\right) \leq CNn^{-1} \tag{6.95}$$

But this estimate implies that the term (6.135) converges to zero $\mathbb{P}$-almost surely, for any choice of N. Hence the result is obvious from the J_1 convergence of $\bar{S}_n$. □

6.3 Almost sure convergence in the p-spin model

The method explained above for the REM was originally introduced in [15] to treat the p-spin SK model under the quenched law. In that case, no other method was available. All the basic ideas of the proof for the REM above carry over, but the control of the convergence of $\hat{\nu}_n$ is much harder and requires the use of Gaussian interpolation techniques.

In this model, the underlying graphs $\mathcal{V}_n$ are the hypercubes $\Sigma_n = \{-1,1\}^n$. The random environment is given by a Gaussian process, H, indexed by Σ_n with zero mean and covariance

$$\mathbb{E}H_n(x)H_n(x') = nR_n(x,x')^p, \tag{6.96}$$

where $R_n(x,x') \equiv \frac{1}{n}\sum_{i=1}^n x_i x_i'$. The mean holding times, $\tau(x)$, are given again by (6.45), and the fast chain, J_n, is simple random walk. The result we obtain in this case is similar to that in the REM, but there is an extra limitation on the parameters.

Theorem 6.3.13 *For any $p \geq 3$, there exists a constant $K_p > 0$ that depends on β and γ, and a function $\zeta(p)$ such that for all γ satisfying*

$$0 < \gamma < \min\left(\beta^2, \zeta(p)\beta\right), \tag{6.97}$$

the law of the stochastic process

$$S_n^b(t) \equiv e^{-\gamma N} S_n\left(\theta_n \lfloor tK_\beta \boldsymbol{n}^{1/2} e^{n\gamma^2/2\beta^2}\theta_n^{-1}\rfloor\right), \qquad t \geq 0, \tag{6.98}$$

with $\theta_n = \frac{3\ln 2}{2}n^2$, and $K_\beta = \beta\sqrt{2\pi}/\Gamma(1+\gamma/\beta^2)$, defined on the space of càdlàg functions equipped with the Skorokhod J_1-topology, converges to the law of the stable subordinator $V_{\gamma/\beta^2}(t), t \geq 0$, of Lévy measure $K_p(\gamma/\beta^2)x^{-\gamma/\beta^2-1}dx$. Convergence holds $\mathbb{P}$-a.s. if $p > 4$, and in $\mathbb{P}$-probability, if $p = 3, 4$.

The function $\zeta(p)$ is increasing and it satisfies

$$\zeta(3) \simeq 1.0291 \qquad \textit{and} \qquad \lim_{p\to\infty}\zeta(p) = \sqrt{2\log 2}. \tag{6.99}$$

In [2] an analogous result is proven, with the same constants $\zeta(p)$ and K_p, but convergence there is law with respect to the random environment (and almost sure with respect to the trajectories J_n). Being able to obtain convergence under the law of the trajectories for fixed environments, as we do here, is a considerable conceptual improvement.

As in the REM, we can again draw conclusions for correlation functions. However, we must make a different choice, since traps are no more isolated points.

In this way we prove the almost-sure (or in probability) version of Theorem 1.2 of [2].

Theorem 6.3.14 *Let $A_n^\varepsilon(t,s)$ be the event defined by*

$$A_n^\varepsilon(t,s) = \left\{R_n\big(X_n\big(te^{\gamma n}\big), X_n\big((t+s)e^{\gamma n}\big)\big) \geq 1-\varepsilon\right\}. \tag{6.100}$$

Then, under the hypothesis of Theorem 6.3.13, for all $\varepsilon \in (0,1)$, $t>0$ and $s>0$,

$$\lim_{N\to\infty} \mathcal{P}_{\pi_n}\left(A_n^\varepsilon(t,s)\right] = \frac{\sin\alpha\pi}{\pi}\int_0^{t/(t+s)} u^{\alpha-1}(1-u)^{-\alpha}\,du. \tag{6.101}$$

Convergence holds $\mathbb{P}$-a.s. if $p>4$, and in $\mathbb{P}$-probability, if $p=3,4$.

6.3.1 Verification of Conditions A1-1 and A2-1

As J_n is exactly the same as in the REM, Condition (A1-1) is already verified. Also, Condition (A3) is exactly the same as in the REM, since correlations do not effect it.

Again, the second part of Condition (A2-1) will follow from the first just as in the REM.

Thus, all what is left to do is to show that

$$\nu_n^t(u,\infty) \to \nu^t(u,\infty) = K_p u^{-\gamma/\beta^2}, \tag{6.102}$$

almost surely, resp. in probability, as $n\uparrow\infty$.

Convergence of $\mathbb{E}\hat{\nu}_n^t(v)$ We will prove convergence of ν^t via the convergence of its Laplace transform as in the REM.

The following Lemma is an easy consequence of the results of [2]:

Lemma 6.3.15 *Let $c_n = e^{\gamma n}$, $a_n = K_\beta n^{1/2} e^{n\gamma^2/2\beta^2}$. For any $p \geq 3$, and $\beta,\gamma>0$ such that $\gamma/\beta^2 \in (0,1)$, there exists a finite positive constant, K_p, such that, for any $v>0$,*

$$\lim_{n\uparrow\infty} k_n(t)\mathbb{E}\left[1-\mathcal{E}_{\pi_n}\left(e^{-vZ_n}\right)\right] = K_p t v^{\gamma/\beta^2}. \tag{6.103}$$

Proof. In [2], the Laplace transforms $\mathbb{E}e^{-vZ_n}$ were computed even for $\theta_n = a_n t$. We just recall the key ideas and the main steps.

The point in [2] is to first fix a realisation of the chain J_n, and to define, for a given realisation, the one-dimensional normal Gaussian process

$$U^0(i) \equiv n^{-1/2} H_n(J_n(i)), \tag{6.104}$$

with covariance

$$\Lambda^0_{ij} = n^{-1}\mathbb{E}H_n(J_n(i))H_n(J_n(j)) = n^{-1}R_n(J_n(i), J_n(j))^p. \tag{6.105}$$

Moreover, they define a comparison process, U^1, as follows. Let ν be an integer of order N^ρ, with $\rho \in (1/2, 1)$. Then U^1 has covariance matrix

$$\Lambda^1_{ij} = \begin{cases} 1 - 2pN^{-1}|i-j|, & \text{if } [i/\nu] = [j/\nu] \\ 0, & \text{else.} \end{cases} \tag{6.106}$$

Finally they define the interpolating family of processes, for $h \in [0,1]$,

$$U^h(i) \equiv \sqrt{h}U^1(i) + \sqrt{1-h}U^0(i). \tag{6.107}$$

For any normal Gaussian process, U, indexed by $\mathbb{N}$, define the function

$$\mathcal{E}_{\pi_n}\left(F(U,v,k) \mid \mathcal{F}^J\right) \equiv G(U,v,k) = \exp\left(-\sum_{i=0}^{k-1} g\left(vc_n^{-1}e^{\beta\sqrt{n}U_i}\right)\right), \tag{6.108}$$

with $g(x) = \ln(1+x)$.

Then the Laplace transforms we are after can be written as

$$\begin{aligned} \mathbb{E}\mathcal{E}_{\pi_n} e^{-vZ_n} &= \mathbb{E}\mathcal{E}_{\pi_n}\left(\mathcal{E}_{\pi_n}\left(e^{-vZ_n} \mid \mathcal{F}^J\right)\right) \\ &= E_{\pi_n}\mathbb{E}G(U^0, v, \theta_n). \end{aligned} \tag{6.109}$$

Here we used that the conditional expectation, given $\mathcal{F}^J$, is just the expectation with respect to the variables $e_{n,i}$ which can be computed explicitly and gives rise to the function G.

The idea is now that U^1 is a good enough approximation to U^0, for most realisation of the chain J, that we can replace U^0 by U^1 in the last line above.

More precisely, we have the following estimate.

Lemma 6.3.16 *With the notation above we have that, for all* $p \geq 3$

$$k_n(t)E_{\pi_n}\left|\mathbb{E}G(U^0, v, \theta_n) - \mathbb{E}G(U^1, v, \theta_n)\right| \leq tCN^{1/2}/\nu. \tag{6.110}$$

Remark 6.3.1 [2] (see (Proposition 3.1)) prove that E_{π_n}-almost surely,

$$\mathbb{E}G(U^0, v, [a_n t]) - \mathbb{E}G(U^1, v, [a_n t]) \to 0. \tag{6.111}$$

This result would not be expected for our expression, but we do not need this. The proof of Proposition 3.1., however, directly implies our Lemma 6.3.16.

The computation of the expression involving the comparison process U^1 is fairly easy. First, note that by independence,

$$\begin{aligned}\mathbb{E}G(U^1, v, \theta_n) &= \left[\mathbb{E}G(U^1, v, \nu)\right]^{\theta_n/\nu} \\ &= \left[1 - \left(1 - \mathbb{E}G(U^1, v, \nu)\right)\right]^{\theta_n/\nu}\end{aligned} \tag{6.112}$$

But in [2], Proposition 2.1., it is shown that

$$a_n \nu^{-1} \left(1 - \mathbb{E}G(U^1, v, \nu)\right) \to K_p v^{\gamma/\beta^2}. \tag{6.113}$$

This implies immediately that

$$k_n(t) \left[1 - \left(1 - \mathbb{E}F(U^1, v, \nu)\right)\right]^{\theta_n/\nu} \to K_p v^{\gamma/\beta^2} t, \tag{6.114}$$

as desired. Combining this with Lemma 6.3.16, the assertion of Lemma 6.3.15 follows. □

Concentration of ν_n^t. To conclude the proof, we need to control the fluctuations of ν_n^t.

Lemma 6.3.17 *Under the same hypothesis as in Lemma 6.3.15, there exists an increasing function, $\zeta(p)$, such that for all $p \geq 3$, $\zeta(p) > 1$, and $\zeta(p) \uparrow \sqrt{2 \ln 2}$, such that, if $\gamma/\beta^2 < \min(1, \zeta(p)/\beta)$,*

$$\mathbb{E}\left(\hat{\nu}_n^t(v) - \mathbb{E}\hat{\nu}_n(v)\right)^2 \leq C n^{1-p/2}. \tag{6.115}$$

Proof. The proof is again very similar to the proof of Proposition 3.1 in [2]. We have to compute

$$\mathbb{E}\left(\mathcal{E}_{\pi_n} e^{-vZ_n}\right)^2 = \mathbb{E}\mathcal{E}_{\pi_n}\mathcal{E}'_{\pi_n}\left(e^{-v(Z_n+Z'_n)} \mid \mathcal{F}^J \times \mathcal{F}^{J'}\right) \tag{6.116}$$

To express this as in the previous proof, we introduce the Gaussian process V^0 by

$$V^0(i) \equiv \begin{cases} n^{-1/2} H_n(J_n(i)), & \text{if } 0 \leq i \leq \theta_n - 1 \\ n^{-1/2} H_n(J'_n(i)), & \text{if } \theta_n \leq i \leq 2\theta_n - 1. \end{cases} \tag{6.117}$$

Then, with the notation of (6.108)

$$\mathcal{E}_{\pi_n}\mathcal{E}'_{\pi_n}\left(e^{-v(Z_n+Z'_n)} \mid \mathcal{F}^J \times \mathcal{F}^{J'}\right) = G(V^0, v, 2\theta_n) \tag{6.118}$$

Next we define the comparison process V^1 with covariance matrix

$$\bar{\Lambda}^2_{ij} \equiv \begin{cases} \bar{\Lambda}^0_{ij}, & \text{if } i \wedge j < \theta_n \text{or } \ i \vee j \geq \theta_n, \\ 0, & \text{else.} \end{cases} \tag{6.119}$$

The point is that

$$E_{\pi_n} E'_{\pi_n} \mathbb{E} G(V^1, v, 2\theta_n) = \left(E_{\pi_n} \mathbb{E} G(V^0, v, \theta_n)\right)^2 = \left(\mathbb{E}\mathcal{E}_{\pi_n} e^{-vZ_n}\right)^2 . \tag{6.120}$$

On the other hand, using the standard Gaussian interpolation formula, we obtain the representation

$$\begin{aligned} &\mathbb{E}G(V^0, v, 2\theta) - \mathbb{E}G(V^0, v, \theta) \\ &= \frac{1}{2} \int_0^1 \sum_{\substack{0 \leq i < \theta_n \\ \theta_n \leq j < 2\theta_n}} \bar{\Lambda}^1_{ij} \mathbb{E} \frac{\partial^2 G(V^h, v, 2\theta_n)}{\partial v_i \partial v_j} dh + (i \leftrightarrow j) \, . \end{aligned} \tag{6.121}$$

The second derivatives of G were computed and bounded in [2], (see Eq. (3.7. and Lemma 3.2.

We recall these bounds:

Lemma 6.3.18 *With the notation above and the assumptions of Lemma 6.3.15,*

$$\begin{aligned} \left| \frac{\partial^2 G(V^h, v, 2\theta_n)}{\partial v_i \partial v_j} \right| &\leq v^2 c_n^{-2} \beta^2 N e^{\beta\sqrt{n}\left(V^h(i) + V^h(j)\right)} \\ &\times \exp\left(-2g\left(c_n^{-1} v e^{\beta\sqrt{n} V^h(i)}\right) - 2g\left(c_n^{-1} v e^{\beta\sqrt{n} V^h(i)}\right)\right) \\ &\equiv \Xi_n(\bar{\Lambda}^h_{ij}). \end{aligned} \tag{6.122}$$

Moreover, for $\lambda > 0$ small enough,

$$\Xi_n(c) \leq \bar{\Xi}_n(c) \tag{6.123}$$

$$= \begin{cases} C\left((1-c)^{-1/2} \wedge \sqrt{n}\right) e^{-\frac{\gamma^2 n}{\beta^2(1+c)}}, & \text{if } 1 > c > \gamma/\beta^2 + \lambda - 1, \\ CNe^{-n(\beta^2(1+c) - 2\gamma)}, & \text{if } c \leq (\gamma/\beta^2) + \lambda - 1, \end{cases}$$

where $C(\gamma, \beta, u, v, \lambda)$ is a suitably chosen constant independent of n and c. Then

Remark 6.3.2 Notice that, since $\gamma/\beta^2 < 1$ under our hypothesis, we can always choose λ such that the top line in (6.123) covers the case $c \geq 0$.

Note that, for $c \geq 0$, (see Eq. (3.25) in [2]; note that there is trivial misprint in the last inequality there)

$$\int_0^1 \Xi_n((1-h)c)dh \leq 2C \exp\left(-\frac{\gamma^2 n}{\beta^2(1+c)}\right). \tag{6.124}$$

The terms with negative correlation are in principle smaller than those with positive one, but some thought will reveal that one cannot really gain substantially over the bound

$$\int_0^1 \Xi_n((1-h)c)dh \leq C \exp\left(-\frac{\gamma^2 n}{\beta^2}\right), \tag{6.125}$$

that is used in [2] (See Eq. 3.24)).

Next we must compute the probability that $\bar{\Lambda}^1_{ij}$ takes on a specific value. But since $\bar{\Lambda}_{ij}$ is a function of $R_n(J_n(i), J'_n(j))$, this turns out to be very easy, namely, since both chains start in the invariant distribution:

$$\begin{aligned} &\mathcal{E}_{\pi_n}\mathcal{E}'_{\pi_n}\mathbb{1}_{nR_n(J_n(i),J'_n(j)=m} \\ &= t^2 \sum_{x,y\in\mathcal{S}^n} \mathcal{P}_{\pi_n}(J_n(i)=x)\mathcal{P}'_{\pi_n}(J'_n(i)=y)\mathbb{1}_{nR_n(x,y)=m} \qquad (6.126) \\ &= 2^{-n}\sum_{x\in\mathcal{S}^n}\mathbb{1}_{nR_n(x,1)=m} = 2^{-n}\binom{n}{(n-m)/2}. \qquad (6.127) \end{aligned}$$

Putting all things together, we arrive at the bound

$$\begin{aligned} &k_n(t)^2 \left|\mathbb{E}G(V^0,v,2\theta) - \mathbb{E}G(V^0,v,\theta)\right| \\ &\leq \sum_{m=0}^n 2^{-n}\binom{n}{(n-m)/2}\left(\frac{m}{n}\right)^p ne^{n\gamma^2/\beta^2} 2C\exp\left(-\frac{n\gamma^2}{\beta^2(+(m/n)^p)}\right) \\ &+ \sum_{m=0}^n 2^{-n}\binom{n}{(n-m)/2}\left(\frac{m}{n}\right)^p ne^{n\gamma^2/\beta^2} 2C\exp\left(-\frac{n\gamma^2}{\beta^2}\right), \qquad (6.128) \end{aligned}$$

where we did use that $k_n(t)\theta_n = t\sqrt{n}e^{n\gamma^2/\beta^2}$. Clearly the second term is smaller than the first, so we only need to worry about the latter. But this term is exactly the term (3.28) in [2], where it is shown that this is smaller than

$$C' t^2 n^{1-p/2}, \tag{6.129}$$

provided $\gamma < \zeta(p)$. This provides the assertion of our Lemma 6.3.17 and concludes its proof. □

Remark 6.3.3 The estimate on the second moment we get here allows to get almost sure convergence only if $p > 4$. It is not quite clear whether this is natural. We were tempted to estimate higher moments to get improved estimates on the convergence speed. However, any straightforward application of the comparison methods used here do get the same order for all higher moments. We have not been able to think of a tractable way to improve this result.

Conclusion of the proof. The proof is concluded in the same way as in the case of the REM, with Lemmata 6.3.15 and 6.3.17 replacing Lemmata 6.2.6 and **??**.

In the cases $p = 3, 4$, where our estimates give only convergence in probability, we obtain convergence of $\nu_n^t(u,\infty)$ in probability, e.g. by using sub-sequences.

Thus we have established Conditions (A1-1) and (A2-1) under the stated conditions on the parameters γ, β, p, and Theorem 6.3.13 follows from Theorem 6.1.2.

6.3.2 Consequences for correlation functions

We will now turn to the proof of Theorem 6.3.14.

The proof of this theorem relies on the following simple estimate. Let us denote by $\mathcal{R}_n$ the range of the coarse grained and rescaled clock process S_n^b. The argument of [2] in the proof of Theorem 1.2 that the event $A_n^\epsilon(s,t) \cap \{\mathcal{R}_n \cap (s,t) \neq \emptyset\}$ has vanishing probability carries over unaltered. However, while in their case, $A_n^\epsilon(s,t) \subset \{\mathcal{R}_n \cap (s,t) = \emptyset\}$, was obvious due to the fact that the coarse graining was done on a scale $o(n)$, this is not immediately clear in our case, where the number of steps within a block is of order n^2. What we have to show is that if the process spends the whole time from s to t within one bloc, then almost all of this time is spent, without interruption, within small ball of radius ϵn.

To do this, we need some simple facts about correlated Gaussian processes.

Lemma 6.3.19 *Let X, Y be standard Gaussian variables with covariance* $\operatorname{cov}(X,Y) = 1 - c$, $0 < c < 1/4$. *Then for $a \gg 1$,*

$$\mathbb{P}\left(X > a, Y > a(1-c/4)\right) \leq \frac{1}{a^2 2\pi c} e^{-a^2/2} \left(e^{-ca^2/32} + e^{-3ca^2/8}\right). \tag{6.130}$$

Proof. Note that the variables X, Y have the joint density

$$\frac{1}{2\pi(2c - c^2)} e^{-\frac{x^2}{2} - \frac{(y-(1-c)x)^2}{4c-2c^2}}. \tag{6.131}$$

Next,

$$\mathbb{P}\left(X > a, Y > a(1-c/2)\right) \leq \mathbb{P}\left(X > a, |Y-(1-c)X| > ac/4)\right) + \mathbb{P}\left(X > a\tfrac{1-c/2}{1-c}\right). \tag{6.132}$$

The result is now a trivial application of the standard tail estimates for Gaussian integrals. □

This lemma has the following corollary:

Corollary 6.3.20 *Let $H_n(\sigma)$ be the Gaussian process defined in (6.96). Let $\mathcal{M}_n \subset \Sigma_n$ be arbitrary. Then*

$$\begin{aligned}&\mathbb{P}\left(\exists_{\sigma,\sigma'\in\mathcal{M}_n} : R_n(\sigma,\sigma') < 1-\epsilon \,\text{and}\, H_n(\sigma) \geq an \wedge H_n(\sigma) \geq an(1-p\epsilon/4))\right)\\ &\leq |\mathcal{M}_n|^2 e^{-na^2/2} e^{-na^2 p\epsilon/40}\end{aligned} \tag{6.133}$$

This lemma implies that even if in a set of size say n^2, there is a point where $H_n(\sigma) > na$, then with overwhelming probability, all point of this size will be within a small ball of radius $o(n)$. (This is even crude: we may even chose ϵ depending on n). All points for which $H_n(\sigma) \leq na(1-c/4)$ will not give a perceptible contribution to the total time.

This means the following: within a block of θ_n steps of the chain J_n, that gives a contribution to a jump, there is only a very small ball which contributes to the time. It remains to show that these contributions come in one "block", i.e. the process will not return to this region once it left it within θ_n steps. But this is an elementary property of the random walk on the hypercube..

Let us make this precise. As remarked above,

$$\begin{aligned}\mathcal{P}_{\pi_n}\left(A_n^\epsilon(s,t)\right) &= \mathcal{P}_{\pi_n}\left(A_n^\epsilon(s,t) \cap \{\mathcal{R}_n \cap (s,t) = \emptyset\}\right)\\ &\quad + \mathcal{P}_{\pi_n}\left(A_n^\epsilon(s,t) \cap \{\mathcal{R}_n \cap (s,t) \neq \emptyset\}\right),\end{aligned} \tag{6.134}$$

where the second term tends to zero. Next we observe that

$$\begin{aligned}&\mathcal{P}_{\pi_n}\left(A_n^\epsilon(s,t) \cap \{\mathcal{R}_n \cap (s,t) = \emptyset\}\right)\\ &= \mathcal{P}_{\pi_n}\left(\mathcal{R}_n \cap (s,t) = \emptyset\right) - \mathcal{P}_{\pi_n}\left((A_n^\epsilon(s,t))^c \cap \{\mathcal{R}_n \cap (s,t) = \emptyset\}\right)\end{aligned} \tag{6.135}$$

Here the first term is what we want. To show that the second term tends to zero, we proceed as follows.

For any $N < \infty$, we clearly have

$$\begin{aligned}
&\mathcal{P}_{\pi_n}\left((A_n^\epsilon(s,t))^c\cap\{\mathcal{R}_n\cap(s,t)=\emptyset\}\right) \qquad (6.136)\\
&=\sum_{k=0}^{k_n(N)-1}\mathcal{P}_{\pi_n}\left(((A_n^\epsilon(s,t))^c)\cap\left\{(s,t)\subset\left(\bar{S}_n(k),\bar{S}_n(k+1)\right)\right\}\right)\\
&+\sum_{k=k_n(N)}^{\infty}\mathcal{P}_{\pi_n}\left(((A_n^\epsilon(s,t))^c)\cap\left\{(s,t)\subset\left(\bar{S}_n(k),\bar{S}_n(k+1)\right)\right\}\right).
\end{aligned}$$

The second term is bounded by

$$\begin{aligned}
&\sum_{k=k_n(N)}^{\infty}\mathcal{P}_{\pi_n}\left(((A_n^\epsilon(s,t))^c)\cap\left\{(s,t)\subset\left(\bar{S}_n(k),\bar{S}_n(k+1)\right)\right\}\right)\\
&\le\mathcal{P}_{\pi_n}\left(S_n^b(N)\le s\right)\to\mathcal{P}\left(V_\alpha(N)>s\right), \qquad (6.137)
\end{aligned}$$

where convergence is almost sure with respect to the environment. The last probability can be made as small as desired by choosing N sufficiently large. It remains to deal with the first sum on the right-hand side of (6.86).

Define the event

$$\begin{aligned}
\mathcal{G}_\rho(k)\equiv&\bigcup_{\substack{k\theta_n\le i<j<(k+1)\theta_n\\ R_n(J_n(i),J_n(j))\le 1-\rho}}\left\{\lambda_n^{-1}(J_n(i))e_{n,i}\ge c_n(t-s)\theta_n^{-1}\right\} \qquad (6.138)\\
&\cap\left\{\lambda_n^{-1}(J_n(i))e_{n,i}\ge c_n n^{-1}\theta_n^{-1}\right\}.
\end{aligned}$$

Note that Corollary 6.3.20 implies that the probability of this event with respect to the law $\mathbb{P}$ is bounded nicely uniformly in the variables J.

On the other hand, on the event $\mathcal{G}_\rho(k)^c\cap(A_n^\epsilon(s,t))^c\cap\left\{(s,t)\subset\left(\bar{S}_n(k),\bar{S}_n(k+1)\right)\right\}$, the following must be true: First, there still must exist some i such that $\lambda_n^{-1}(J_n(i))e_{n,i}\ge c_n(t-s)\theta_n^{-1}$, and second, the random walk must make a loop, i.e. it the event

$$\begin{aligned}
&\mathcal{W}_{\rho,\epsilon}(k) \qquad (6.139)\\
&\equiv\bigcup_{k\theta_n\le i<j<\ell<(k+1)\theta_n}\left\{R_n(J_n(i),J_n(j))>1-\epsilon\wedge R_n(J_n(i),J_n(\ell))\le 1-\rho\right\}.
\end{aligned}$$

The probability of this event is generously bounded by

$$P_{\pi_n}\left(\mathcal{W}_{\rho,\epsilon}(k)\right)\le n^4e^{-n(I(1-\rho)-I(1-\epsilon))}, \qquad (6.140)$$

where I is Cramèr's rate function.

By these considerations, we have the bound

$$\mathbb{E}\left(\sum_{k=0}^{k_n(N)} \mathcal{P}_{\pi_n}\left((A_n^\epsilon(s,t))^c \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}\right)\right) \tag{6.141}$$
$$\leq \sum_{k=0}^{k_n(N)} \mathbb{E}\Bigg(\mathcal{P}_{\pi_n}\left(\mathcal{G}_\rho(k)\right)$$
$$+\mathcal{P}_{\pi_n}\left(\left\{\exists_{k\theta_n \leq i < (k+1)\theta_n} \lambda_n^{-1}(J_n(i))e_{n_i} > c_n\theta_n^{-2}\right\} \cap \mathcal{W}_{\rho,\epsilon}(k)\right)\Bigg)$$

Now

$$\mathbb{E}\mathcal{P}_{\pi_n}\left(\mathcal{G}_\rho(k)\right) \leq a_n^{-1}e^{-\delta n}, \tag{6.142}$$

for some $\delta > 0$ depending on the choice of ρ. The simplest way to see this is to use that the probability that one of the $e_{n,i}$ is larger than n^2 is smaller than $\exp(-n^2)$, and then use the bound from Corollary 6.3.20.

Finally, the two events in $\left\{\exists_{k\theta_n \leq i < (k+1)\theta_n} \lambda_n^{-1}(J_n(i))e_{n_i} > c_n\theta_n^{-2}\right\} \cap \mathcal{W}_{\rho,\epsilon}(k)$ are independent, and hence

$$\begin{aligned}
&\mathbb{E}\mathcal{P}_{\pi_n}\left(\left\{\exists_{k\theta_n \leq i < (k+1)\theta_n} \lambda_n^{-1}(J_n(i))e_{n_i} > c_n\theta_n^{-2}\right\} \cap \mathcal{W}_{\rho,\epsilon}(k)\right)\\
&= \mathbb{P}\left(\exists_{k\theta_n \leq i < (k+1)\theta_n} \lambda_n^{-1}(J_n(i))e_{n_i} > c_n\theta_n^{-2}\right) P_{\pi_n}\left(\mathcal{W}_{\rho,\epsilon}(k)\right)\\
&\leq \theta_n^2\mathbb{P}\left(e^{\beta H_n(x)} > c_n n^{-4}\right) e^{-n(I(1-\rho)-I(1-\epsilon))} + \theta_n e^{-n^2}\\
&\leq \theta_n^4 a_n^{-1} n^{-\gamma\beta^{-2}-1/2} e^{-n(I(1-\rho)-I(1-\epsilon))} + \theta_n e^{-n^2}
\end{aligned} \tag{6.143}$$

Combining this, we see that

$$\mathbb{E}\left(\sum_{k=0}^{k_n(N)} \mathcal{P}_{\pi_n}\left((A_n^\epsilon(s,t))^c \cap \left\{(s,t) \subset \left(\bar{S}_n(k), \bar{S}_n(k+1)\right)\right\}\right)\right) \leq CNe^{-\delta n}, \tag{6.144}$$

for some positive δ, whatever the choice of ϵ. But this estimate implies that the term (6.135) converges to zero $\mathbb{P}$- almost surely, for any choice of N. Hence the result is obvious from the J_1 convergence of $\bar{S}_n$.

7

Convergence and topological issues

In the preceeding chapters I have deliberately slipped over topological issues related to process convergence in order to focus the discussion on the basic mechanisms. In this section I will collect the most important issues related convergence and the topologies that should be used.

7.1 Convergence on the space of point measures

In this section we complement the discussion from Section 2.

7.1.1 The vague topology

In Chapter 3 we have already introduced the notion of vague convergence on the space of point measures. Here we will elaborate this notion a little more.

The following properties of vague convergence are useful.

Proposition 7.1.1 *Let μ_n, $n \in \mathbb{N}$ be in $M_+(\mathbb{R}^d)$. Then the following statements are equivalent:*

(i) μ_n *converges vaguely to* μ, $\mu_n \stackrel{v}{\to} \mu$.
(ii) $\mu_n(B) \to \mu(B)$ *for all relatively compact sets,* B, *such that* $\mu(\partial B) = 0$.
(iii) $\limsup_{n\uparrow\infty} \mu_n(K) \leq \mu(K)$ *and* $\limsup_{n\uparrow\infty} \mu_n(G) \geq \mu(G)$, *for all compact* K, *and all open, relatively compact* G.

In the case of point measures, we would of course like to see that the point where the sequence of vaguely convergent measures are located converge. The following proposition tells us that this is true.

Proposition 7.1.2 *Let μ_n, $n \in \mathbb{N}$, and μ be in $M_p(\mathbb{R}^d)$, and $\mu_n \stackrel{v}{\to} \mu$.*

Let K be a compact set with $\mu(\partial K) = 0$. Then we have a labeling of the points of μ_n, for $n \geq n(K)$ large enough, such that

$$\mu_n(\cdot \cap K) = \sum_{i=1}^{p} \delta_{x_i^n}, \quad \mu(\cdot \cap K) = \sum_{i=1}^{p} \delta_{x_i},$$

such that $(x_1^n, \ldots, x_p^n) \to (x_1, \ldots, x_p)$.

Another useful and unsurprising fact is that

Proposition 7.1.3 *The set $M_p(\mathbb{R}^d)$ is vaguely closed in $M_+(\mathbb{R}^d)$.*

Thus, in particular, the limit of a sequence of point measures, will, if it exists as a σ-finite measure, be again a point measure.

Proposition 7.1.4 *The topology of vague convergence can be metrized and turns M_+ into a complete, separable metric space.*

Although we will not use the corresponding metric directly, it may be nice to see how this can be constructed. We therefore give a proof of the proposition that constructs such a metric.

Proof. The idea is to first find a countable collection of functions, $h_i \in \mathbb{C}_0^+(\mathbb{R}^d)$, such that $\mu_n \stackrel{v}{\to} \mu$ if and only if, for all $i \in \mathbb{N}$, $\mu_n(h_i) \to \mu(h_i)$. The construction below is form [29]. Take a family $G_i, i \in \mathbb{N}$, that form a base of relatively compact sets, and assume it to be closed under finite unions and finite intersections. One can find (by Uryson's theorem), families of functions $f_{i,n}, g_{i,n} \in C_0^+$, such that

$$f_{i,n} \uparrow \mathbb{1}_{G_i}, \quad g_{i,n} \downarrow \mathbb{1}_{G_i}$$

Take the countable set of functions $g_{i,n}, f_{i,n}$ as the collection h_i. Now $\mu \in M_+$ is determined by its values on the h_j. For, first of all, $\mu(G_i)$ is determined by these values, since

$$\mu(f_{i,n}) \uparrow \mu(G_i) \quad \text{and} \mu(g_{i,n}) \downarrow \mu(G_i)$$

But the family G_i is a Π-system that generates the sigma-algebra $\mathcal{B}(\mathbb{R}^d)$, and so the values $\mu(G_i)$ determine μ.

Now, $\mu_n \stackrel{v}{\to} \mu$, iff and only if, for all h_i, $\mu_n(h_i) \to c_i = \mu(h_i)$.

From here the idea is simple: Define

$$d(\mu, \nu) \equiv \sum_{i=1}^{\infty} 2^{-i} \left(1 - e^{-|\mu(h_i) - \nu(h_i)|}\right) \tag{7.1}$$

Indeed, if $D(\mu_n, \mu) \downarrow 0$, then for each ℓ, $|\mu_n(h_\ell) - \mu(h_\ell)| \downarrow 0$, and conversely. □

It is not very difficult to verify that this metric is complete and separable.

7.1.2 Weak convergence

Having established the space of σ-finite measures as a complete, separable metric space, we can think of weak convergence of probability measures on this space just as if we were working on an Euclidean space.

One very useful fact about weak convergence is Skorohod's theorem, that relates weak convergence to *almost sure convergence.*

Theorem 7.1.5 *Let $X_n, n = 0, 1, \dots$ be a sequence of random variables on a complete separable metric space. Then X_n converges weakly to a random variable X_0, iff and only if there exists a family of random variables X_n^*, defined on the probability space $([0,1], \mathcal{B}([0,1]), m)$, where m is the Lebesgue measure, such that*

(i) For each n, $X_n \stackrel{\mathcal{D}}{=} X_n^*$, and
(ii) $X_n^* \to X_0^*$, almost surely.

(for a proof, see [12]). While weak convergence usually means that the actual realisation of the sequence of random variables do not converge at all and oscillate widely, Skorohod's theorem says that it is possible to find an "equally likely" sequence of random variables, X_n^*, that do themselves converge, with probability one. Such a construction is easy in the case when the random variables take values in $\mathbb{R}$. In that case, we associate with the random variable X_n (whose distribution function is F_n, that form simplicity we may assume strictly increasing), the random variable $X_n^*(t) \equiv F_n^{-1}(t)$. It is easy to see that

$$m(X_n^* \leq x) = \int_0^1 \mathbb{1}_{F_n^{-1}(t) \leq x} dt = F_n(x) = \mathbb{P}(X_n \leq x)$$

On the other hand, if $\mathbb{P}[X_n \leq x] \to \mathbb{P}[X_0 \leq x]$, for all points of continuity of F_0, that means that for Lebesgue almost all t, $F_n^{-1}(t) \to F_0^{-1}(t)$, i.e. $X_n^* \to X_0^*$, m-almost surely.

Skorohod's theorem is very useful to extract important consequences from weak convergence. In particular,it allows to prove the convergence of certain functionals of sequences of weakly convergent random variables, which otherwise would not be obvious.

A particularly useful criterion for convergence of point processes is provided by *Kallenberg's theorem*, Theorem 2.1.15, which we will prove now.

Proof. The key observation needed to prove the theorem is that simple point processes are uniquely determined by their avoidance function. This seems rather intuitive, in particular in the case $E = \mathbb{R}$: if we know the probability that in an interval there is no point, we know the distribution of the gape between points, and thus the distribution of the points.

Let us note that we can write a point measure, μ, as

$$\mu = \sum_{y \in S} c_y \delta_y,$$

where S is the support of the point measure and c_y are integers. We can associate to μ the simple point measure

$$T^* \mu = \mu^* = \sum_{y \in S} \delta_y,$$

Then it is true that the map T^* is measurable, and that, if ξ_1 and ξ_2 are point measures such that, for all $I \in \mathcal{T}$,

$$\mathbb{P}\left[\xi_1(I) = 0\right] = \mathbb{P}\left[\xi_2(I) = 0\right], \tag{7.2}$$

then

$$\xi_1^* \stackrel{\mathcal{D}}{=} \xi_2^*.$$

To see this, let

$$\mathcal{C} \equiv \left\{ \{\mu \in M_p(E) : \mu(I) = 0\}, I \in \mathcal{T} \right\}.$$

The set $\mathcal{C}$ is easily seen to be a Π-system. Thus, since by assumption the laws, $\mathbb{P}_i$, of the point processes ξ_i coincide on this Π-system, they coincide on the sigma-algebra generated by it. We must now check that T^* is measurable as a map from $(M_p, \sigma(\mathcal{C}))$ to $(M_p, \mathcal{M}_p)$, which will hold, if for each I, the map $T_1^* : \mu \to \mu^*(I)$ is measurable form $(M_p, \sigma(\mathcal{C})) \to \{0, 1, 2, \dots\}$. Now introduce a family of finite coverings of (the relatively compact set) I, $A_{n,j}$, with $A_{n,j}$'s whose diameter is less than $1/n$. We will chose the family such that for each j, $A_{n+1,j} \subset A_{n,i}$, for some i. Then

$$T_1^* \mu = \mu^*(I) = \lim_{n \uparrow \infty} \sum_{j=1}^{k_n} \mu(A_{n,j}) \wedge 1,$$

since eventually, no $A_{n,j}$ will contain more than one point of μ. Now set $T_2^*\mu = (\mu(A_{n,j}) \wedge 1)$. Clearly,

$$(T_2^*)^{-1}\{0\} = \{\mu : \mu(A_{n,j}) = 0\} \subset \sigma(\mathcal{C}),$$

and so T_2^* is measurable as desired, and so is T_1^*, being a monotone limit of finite sums of measurable maps. But now

$$\mathbb{P}[\xi_1^* \in B] = \mathbb{P}[T^*\xi_1 \in B] = \mathbb{P}\left[\xi_1 \in (T^*)^{-1}(B)\right] = \mathbb{P}_1\left[(T^*)^{-1}(B)\right].$$

But since $(T^*)^{-1}(B) \in \sigma(\mathcal{C})$, by hypothesis, $\mathbb{P}_1\left[(T^*)^{-1}(B)\right] = \mathbb{P}_2\left[(T^*)^{-1}(B)\right]$, which is also equal to $\mathbb{P}[\xi_1^* \in B]$, which proves (7.2).

Now, as we have already mentioned, (2.102) implies uniform tightness of the sequence ξ_n. Thus, for any subsequence n', there exist a sub-sub-sequence, n'', such that $\xi_{n''}$ converges weakly to a limit, η. By compactness of M_p, this is a point process. Let us assume for a moment that (a) η is simple, and (b), for any relatively compact A,

$$\mathbb{P}[\xi(\partial A) = 0] \quad \Rightarrow \quad \mathbb{P}[\eta(\partial A) = 0]. \tag{7.3}$$

Then, the map $\mu \to \mu(I)$ is a.s. continuous with respect to η, and therefore, if $\xi_{n'} \stackrel{w}{\to} \eta$, then

$$\mathbb{P}[\xi_{n'}(I) = 0] \to \mathbb{P}[\eta(I) = 0].$$

But we assumed that

$$\mathbb{P}[\xi_n(I) = 0] \to \mathbb{P}[\xi(I) = 0],$$

so that, by the foregoing observation, and the fact that both η and ξ are simple, $\xi = \eta$.

It remains to check simplicity of η and (7.3).

To verify the latter, we will show that for any compact set, K,

$$\mathbb{P}[\eta(K) = 0] \geq \mathbb{P}[\xi(K) = 0]. \tag{7.4}$$

We use that for any such K, there exist sequences of functions, $f_j \in C_0^+(\mathbb{R}^d)$, and compact sets, K_j, such that

$$\mathbb{1}_K \leq f_j \leq 1_{K_j},$$

and $\mathbb{1}_{K_j} \downarrow \mathbb{1}_K$. Thus,

$$\mathbb{P}[\eta(K) = 0] \geq \mathbb{P}[\eta(f_j) = 0] = \mathbb{P}[\eta(f_j) \leq 0]$$

But $\xi_{n'}(f_j)$ converges to $\eta(f_j)$, and so

$$\mathbb{P}[\eta(f_j) \leq 0] \geq \limsup_{n'} \mathbb{P}[\xi_{n'}(f_j) \leq 0] \geq \mathbb{P}[\xi_{n'}(K_j) \leq 0].$$

Finally, we can approximate K_j by elements $I_j \in \mathcal{T}$, such that $K_j \subset I_j \downarrow K$, so that

$$\mathbb{P}\left[\xi_{n'}(K_j) \leq 0\right] \geq \limsup_{n'} \mathbb{P}\left[\xi_{n'}(I_j) \leq 0\right] = \mathbb{P}\left[\xi(K_j) \leq 0\right],$$

so that (7.4) follows.

Finally, to show simplicity, we take $I \in \mathcal{T}$ and show that the η has multiple points in I with zero probability. Now

$$\mathbb{P}\left[\eta(I) > \eta^*(I)\right] = \mathbb{P}\left[\eta(I) - \eta^*(I) < 1/2\right] \leq 2\left(\mathbb{E}\eta(I) - \mathbb{E}\eta^*(I)\right))$$

□

The latter, however, is zero, due to the assumption of convergence of the intensity measures.

Remark 7.1.1 The main requirement in the theorem is the convergence of the so-called *avoidance function*, $\mathbb{P}\left[\xi_n(I) = 0\right]$, (2.102). The convergence of the mean (the intensity measure) provides tightness, and ensures that all limit points are simple. It is only a sufficient, but not a necessary condition. It may be replaced by the tightness criterion that for all $I \in \mathcal{T}$, and any $\epsilon > 0$, one can find $R \in \mathbb{N}$, such that, for all n large enough,

$$\mathbb{P}\left[\xi_n(I) > R\right] \leq \epsilon, \tag{7.5}$$

if one can show that all limit points are simple (see [18]). Note that, by Chebeychev's inequality, (2.102) implies, of course, (7.5), but but vice versa. There are examples when (2.101) and (7.5) hold, but (2.102) fails.

7.2 Skorokhod topologies on the space of càdlàg functions

The second topological space we are concerned with is the space of càdlàg function, which is where our stochastic processes will live.

7.2.1 The càdlàg space $D_E[0, \infty)$

It will be important that we can treat the space of càdlàg functions with values in a metric space as a Polish space much like the space of continuous functions. The material from this section is taken from [20] where omitted proofs and further details can be found.

7.2.2 A Skorokhod metric

We will now construct a metric on càdlàg space which will turn this space into a complete metric space. This was first don by Skorokhod. In fact, there are various different metrics one may put on this space which will give rise to different convergence properties. This is mostly related to the question whether each jump in the limiting function is associated to one, several, or no jumps in approximating functions. A detailed discussion of these issues can be found in [33]. We will come back to this point later and begin with the strong and most popular topology, called J_1-topology.

Definition 7.2.1 Let Λ denote the set of all strictly increasing maps $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$, such that λ is Lipshitz continuous and

$$\gamma(\lambda) \equiv \sup_{0 \leq t < s} \left| \ln \frac{\lambda(s) - \lambda(t)}{s - t} \right| < \infty. \tag{7.6}$$

For $x, y \in D_E[0, \infty)$, $u \in \mathbb{R}_+$, and $\lambda \in \Lambda$, set

$$d(x, y, \lambda, u) \equiv \sup_{t \geq 0} \rho\left(x(t \wedge u), y(\lambda(t) \wedge u)\right). \tag{7.7}$$

Finally, the *Skorohod metric* on $D_E[0, \infty)$ is given as

$$d(x, y) \equiv \inf_{\lambda \in \Lambda} \left(\gamma(\lambda) \vee \int_0^\infty e^{-u} d(x, y, u, \lambda) du \right). \tag{7.8}$$

To get the idea behind this definition, note that with λ the identity, this is just the metric on the space of continuous functions. The rôle of the λ is to make the distance of two functions that look much the same, except that they jump at two points very close to each other by sizable amount, small. E.g., we clearly want the functions

$$x_n(t) = \mathbb{1}_{[1/n, \infty]}(t)$$

to converge to the function

$$x_\infty(t) = 1_{[0, \infty]}(t).$$

This is wrong under the sup-norm, since $\sup_t \|x_n(t) - x_\infty(t)\| = 1$, but it will be true under the metric d (Exercise!).

Lemma 7.2.6 *d as defined a above is a metric on $D_E[0, \infty)$.*

Proof. We first show that $d(x, y) = 0$ implies $y = x$. Note that for $d(x, y) = 0$, it must be true that there exists a sequence λ_n such that

$\gamma(\lambda_0) \downarrow 0$ and $\lim_{n\uparrow\infty} d(x, y, \lambda_n, u) = 0$; one easily checks that then

$$\lim_{n\uparrow\infty} \sup_{0\leq t\leq T} |\lambda_n(t) - t| = 0,$$

and hence $x(t) = y(t)$ at all continuity points of x. But since x and y are càdlàg , this implies $x = y$.

Symmetry follow from the fact that $d(x, y, \lambda, u) = d(y, x, \lambda^{-1}, u)$ and that $\gamma(\lambda) = \gamma(\lambda^{-1})$.

Finally we need to prove the triangle inequality. A simple calculation shows that

$$d(x, z, \lambda_2 \circ \lambda_1, u) \leq d(x, y, \lambda_1, u) + d(y, z, \lambda_2, u).$$

Finally $\gamma(\lambda_1 \circ \lambda_2) \leq \gamma(\lambda_1) + \gamma(\lambda_2)$, and putting this together one derives $d(x, z) \leq d(x, y) + d(y, z)$. □

Exercise: Fill in the details of the proof of the triangle inequality.

The next theorem completes our task.

Theorem 7.2.7 *If E is separable, then $D_E[0,\infty)$ is separable, and if E is complete, then $D_E[0,\infty)$ is complete.*

Proof. The proof of the first statement is similar to the proof of the separability of $C(J)$ (Theorem **??**) and is left to the reader. To prove completeness, we only need to show that every Cauchy sequence converges. Thus let $x_n \in D_E[0,\infty)$ be Cauchy. Then, for any constant $C > 1$, and any $k \in \mathbb{N}$, there exist values n_k, such that for all $n, m \geq n_k$, $d(x_n, x_m) \leq C^{-k}$. Then we can select sequences u_k, and λ_k, such that

$$\gamma(\lambda_k) \vee d(x_{n_k}, x_{n_{k+1}}, \lambda_k, u_k) \leq 2^{-k}.$$

Then, in particular,

$$\mu_k \equiv \lim_{m\uparrow\infty} \lambda_{k+m} \circ \lambda_{k+m-1} \circ \cdots \circ \lambda_{k+1} \circ \lambda_k$$

exists and satisfies

$$\gamma(\mu_k) \leq \sum_{m=k^\infty} \gamma(\lambda_m) \leq 2^{-k+1}.$$

Now

$$\begin{aligned}
&\sup_{t\geq 0}\rho\left(x_{n_k}(\mu_k^{-1}(t)\wedge u_k), x_{n_{k+1}}(\mu_{k+1}^{-1}(t)\wedge u_k)\right)\\
&=\sup_{t\geq 0}\rho\left(x_{n_k}(\mu_k^{-1}(t)\wedge u_k), x_{n_{k+1}}(\lambda_k(\mu_{k+1}^{-1}(t))\wedge u_k)\right)\\
&=\sup_{t\geq 0}\rho\left(x_{n_k}(t\wedge u_k), x_{n_{k+1}}(\lambda_k^{-1}(t)\wedge u_k)\right)\\
&\leq 2^{-k}.
\end{aligned}$$

Therefore, by the completeness of E, the sequence of functions $z_k \equiv x_{n_k}(\mu_k^{-1}(t))$ converges uniformly on compact intervals to a function z. Each z_k being càdlàg implies that z is also càdlàg . Since $\gamma(\mu_k) \to 0$, it follows that

$$\lim_{k\uparrow\infty}\sup_{0\leq t\leq T}\rho(x_{n_k}(\mu_k^{-1}(t)), z(t)) = 0,$$

for all T, and hence $d(x_{n_k}, z) \to 0$. Since a Cauchy sequence that contains a convergent subsequence converges, the proof is complete. □

To use Prohorov's theorem for proving convergence of probability measures on the space $D_E[0,\infty)$, we need first a characterisation of compact sets.

The first lemma states that the clusure of the space of step functions that are uniformly bounded and where the distance between steps is uniformly bounded from below is compact:

Lemma 7.2.8 *Let $\Gamma \subset E$ be compact and $\delta > 0$ be fixed. Let $A(\Gamma,\delta)$ denote the set of step functions, x, in $D_E[0,\infty)$ such that*

(i) *$x(t) \in \Gamma$, for all $\tau \in [0,\infty)$, and*
(ii) *$s_k(x) - s_{k-1}(x) > d$, for all $k \in \mathbb{N}$,*

where

$$s_k(x) \equiv \inf\{t > s_{k-1}(x) : x(t) \neq x(t-)\}.$$

Then the closure of $A(\Gamma,\delta)$ is compact.

We leave the prove as an exercise.

The analog of the modulus of continuity in the Arzelà-Ascoli theorem on càdlàg space is the following: For $x \in D_E[0,\infty)$, $\delta > 0$, and $T < \infty$, set

$$w(x,\delta,T) \equiv \inf_{t_i}\max_i \sup_{s,t\in[t_{i-1},t_i)}\rho(x(s),x(t)), \tag{7.9}$$

where the first infimum is over all collections $0 = t_0 < t_1 < \cdots < t_{n-1} < T < t_n$, with $t_i - t_{i-1} > \delta$, for all i.

The following theorem is the analog of the Arzelà-Ascoli theorem:

Theorem 7.2.9 *Let E be a complete metric space. Then the closure of a set $A \subset D_E([0,\infty)$ is compact, if and only if,*

(i) For every rational $t \geq 0$, there exists a compact set $\Gamma_t \subset E$, such that for all $x \in A$; $x(t) \in \Gamma_t$.
(ii) For each $T < \infty$,

$$\lim_{\delta \downarrow 0} \sup_{x \in A} w(x, \delta, T) = 0. \tag{7.10}$$

A proof of this result can be found, e.g. in [20]. s

Based on this theorem, we now get the crucial tightness criterion:

Theorem 7.2.10 *Let E be complete and separable, and let X_α be a family of processes with càdlàg paths. Then the family of probability laws, μ_α, of X_α, is conditionally compact if and only if the following holds:*

(i) For every $\eta > 0$ and rational $t \geq 0$, there exists a compact set, $\Gamma_{\eta,t} \subset E$, such that

$$\inf_\alpha \mu_\alpha \left(x(t) \in \Gamma_{\eta,t} \right) \geq 1 - \eta, \tag{7.11}$$

and

(ii) For every $\eta > 0$ and $T < \infty$, there exists $\delta > 0$, such that

$$\sup_\alpha \mu_\alpha \left(w(x, \delta, T) \geq \eta \right) \leq \eta. \tag{7.12}$$

7.2.3 Incomplete metrics and tightness criteria

For practical work it is often convenient to work with simpler metrics on càdlàg space that yield tightness criteria that are easier to verify. They fail to be complete, but this is not a practial handicap. We consider first the J_1 topology.

J_1 **topology** . Skorohod did not introduce the metric we defined in Definition 7.2.1; this was only done later by Billingsley [10]. Skorokhod used a simpler metric: for $f, g \in D$

$$d_{J_1}(f, g) = \inf_{\lambda \in \Lambda} \{ \|f \circ \lambda - g\|_\infty \vee \|\lambda - e\|_\infty \}, \tag{7.13}$$

where Λ is the set of strictly increasing functions mapping $[0,T]$ onto itself such that both λ and its inverse are continuous, and e is the identity map on $[0,T]$. It was shown by Kolmogorov [30] that the topology defined by this metric turns D into a complete space.

There is also a slightly more convenient continuity-module that we will use for this topology. We set

$$\begin{aligned} w_f(\delta,T) &= \sup_{t_1\le t\le t_2\le T, t_2-t_1\le\delta} \left\{\min\left(|f(t)-f(t_1)|, |f(t_2)-f(t)|\right)\right\}, \\ v_f(t,\delta,T) &= \sup_{:t_1,t_2\in[0,T]\cap(t-\delta,t+\delta)} \{|f(t_1)-f(t_2)|\}. \end{aligned} \tag{7.14}$$

The following result is a restatement of Theorem 12.12.3 of [33] and Theorem 15.3 of [11].

Theorem 7.2.11 *The sequence of probability measures $\{P_n\}$ is tight in the J_1-topology if*

(i) For each positive ε there exist c such that

$$P_n[f : \|f\|_\infty > c] \le \varepsilon, \qquad n \ge 1. \tag{7.15}$$

(ii) For each $\varepsilon > 0$ and $\eta > 0$, there exist a δ, $0 < \delta < T$, and an integer n_0 such that

$$P_n[f : w_f(\delta) \ge \eta] \le \varepsilon, \qquad n \ge n_0, \tag{7.16}$$

(iii) and

$$P_n[f : v_f(0,\delta) \ge \eta] \le \varepsilon \text{ and } P_n[f : v_f(T,\delta) \ge \eta] \le \varepsilon, \qquad n \ge n_0. \tag{7.17}$$

M_1-topology. In many natural cases, the J_1 topology is simply to fine to yield convergence. For instance, whenever we have a sequence of continuous functions that develops jumps in the limit, we cannot have convergence in the J_1 topology.

This excludes situations when (a) a number of small jumps converge to the same location and form a big one, or when a family of continuous curves develops a jump. In such cases one may still want to say that in a suitable sense, convergence takes place.

To this end Skorokhod introduced another (incomplete) metric, the M_1-metric. To define this, let, for $f \in D$, Γ_f be its completed graph,

$$\Gamma_f = \{(z,t) \in \mathbb{R}\times[0,T] : z = \alpha f(t-) + (1-\alpha) f(t), \alpha \in [0,1]\}. \tag{7.18}$$

A parametric representation of the completed graph Γ_f (or of f) is a

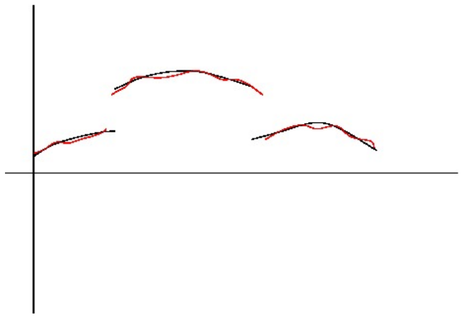

Fig. 7.1. Red curve is close to black curve in J_1

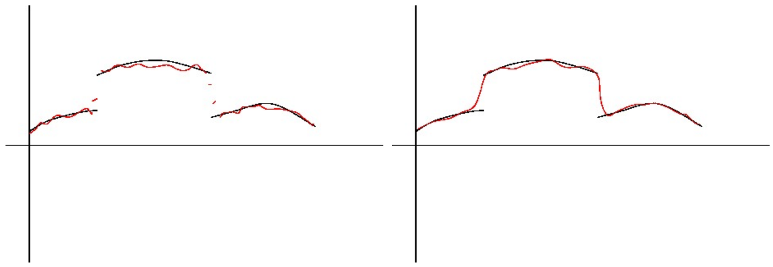

Fig. 7.2. Red curves are *not* close to black curves in J_1, but are in M_1

continuous bijective mapping $\phi(s) = (\phi_1(s), \phi_2(s))$, $[0,1] \mapsto \Gamma_f$ whose first coordinate ϕ_1 is increasing. If $\Pi(f)$ is set of all parametric representation of f, then the M_1-metric is defined by

$$d_{M_1}(f,g) = \inf\{\|\phi_1 - \psi_1\|_\infty \vee \|\phi_2 - \psi_2\|_\infty : \phi \in \Pi(f), \psi \in \Pi(g)\}. \tag{7.19}$$

Again the topology associated with this metric is complete, although the metric is not.

To prove tightness in this space, one has a theorem completely analogous to Theorem 7.2.10. The only difference is that the modulus of continuity, $w(x,\delta,T)$, used there is replaced by

$$w'(x,\delta,T) = \sup_{t_1 \leq t \leq t_2 \leq T, t_2 - t_1 \leq \delta} \left\{ \inf_{\alpha \in [0,1]} |x(t) - (\alpha x(t_1) + (1-\alpha)x(t_2))| \right\}. \tag{7.20}$$

One notices that this criterion no longer forbids accumulations of jumps large jumps at a single point.

Remark 7.2.1 There is also a simple metric for the J_1 topology that is not complete. It is given by

$$d_{J_1}(f,g) = \inf_{\lambda\in\Lambda}\{\|f\circ\lambda - g\|_\infty \vee \|\lambda - e\|_\infty\}, \tag{7.21}$$

where Λ is the set of strictly increasing functions mapping $[0,T]$ onto itself such that both λ and its inverse are continuous, and e is the identity map on $[0,T]$.

7.3 Implications for the convergence of sums

Let us see how we con use our tightness criteria in the proofs of the convergence of sums of random variables.

Let us consider the example of Theorem 4.2.10.

Let us check that the criteria of Theorem 7.2.11 are verified in our case. (i) was already checked when we considered the convergence of the one-dimensional marginals, since our process is monotone increasing (otherwise, we would need to use maximum inequalities). Condition (iii) amounts to checking that there is no jump at 0 and at T. In fact, using that all our processes are increasing,

$$\begin{aligned}\mathbb{P}\left[v_{S_n}(0,\delta) > \eta\right] &= \mathbb{P}\left[S_n(\delta) > \eta\right] \\ &\leq \mathbb{P}\left[Z^{\leq}_{n\delta} > K_n\eta/2\right] + \mathbb{P}\left[Z^{>}_{n\delta}K_n\eta/2\right] \\ &\leq 2\mathbb{E}Z^{\leq}_{n\delta}/(K_n\eta) + \delta n\mathbb{P}\left[X_i > K_n\eta/2\right] \\ &\leq 2c\delta\eta^{-1}\frac{2-\alpha}{1-\alpha}\epsilon^{1-\alpha} + \delta c(\eta/2)^{-\alpha}.\end{aligned} \tag{7.22}$$

Clearly, for any $\eta > 0$ and $\epsilon > 0$, the right-hand side of (7.22) can be made smaller than ϵ by an appropriate choice of δ.

The task to check (ii) is not much harder. We may check this condition again for $S^{\leq}_n$ and $S^{>}_n$ separately. For the former, we need a second moment estimate (with calculations we can borough from Chapter 4),

$$\mathbb{E}\left(S^{\leq}_n(\delta)\right)^2 \leq \text{const.}\left(\delta\epsilon^{2-\alpha} + \delta^2 c^{2-2\alpha}\right), \tag{7.23}$$

and then a standard partitioning argument tells us that

$$\begin{aligned}\mathbb{P}\left[w_{S^{\leq}_n}(d) > \eta/2\right] &\leq \mathbb{P}\left[\exists_{k\leq T/\delta} : S^{\leq}_n((k+1)\delta) - S^{\leq}_n(k\delta) \geq \eta/2\right] \\ &\leq \text{const.}T\eta^{-2}\left(\epsilon^{2-\alpha} + \delta\epsilon^{2-2\alpha}\right).\end{aligned} \tag{7.24}$$

which can be made small for any η and ϵ be making δ small enough.

For $S^{>}_n 0$, choose $\epsilon < \eta/2$. Then the event $\{w_{S^{>}_n}(\delta) > \eta/2\}$ can only occur if *two* atoms of the point process that are bigger than ϵ have distance smaller than 2δ. The probability of this to happen is controlled by

$$T\delta n^2 \mathbb{P}\left[X_i > K_n \epsilon\right]^2 \leq T c^2 \delta \epsilon^{-2\alpha}. \tag{7.25}$$

Now take δ such that $\delta\epsilon^{-2\alpha}$ is smaller that $c^2\rho/T$, and then ϵ so small that the bound in (7.24) is also smaller than ρ. Then our tightness criterion is satisfied.

The same general strategie works in other cases, too.

7.4 Implications for correlation functions

The main advantage of having convergence in the J_1-topology is that it ensures convergence of the jumps: If the limiting subordinator has a jump of given size, then the approximants had jumps converging to the same size, and it cannot be the case that there were many small jumps of the approximants that merged together to produce that of the limit.

But the jumps of the clock process S_n are closely linked to the correlation function. Indeed, if This implies in particular that $X_(s)$ remains constant on the interval $[t_w, t_w+t]$, if and only if the clock process jumps over this interval, i.e. if (t_w, t_w+t) is not in the range of S_N. Combining these observations, we get the following fact:

Lemma 7.4.12 *The correlation function* Π_N *satisfies*

$$\lim_{t_w\uparrow\infty} \lim_{N\uparrow\infty} \Pi_N(t_w, \theta t_w) = \mathbb{P}\left[(1, 1+\theta) \mathrm{int}\, \mathit{range}(V_\alpha) = \emptyset\right] \tag{7.26}$$

$$= \frac{\sin \alpha\pi}{\pi} \int_0^{1/(1+\theta)} u^{\alpha-1}(1-u)^{-\alpha} du.$$

Bibliography

[1] G. Ben Arous, L. V. Bogachev, and S. A. Molchanov. Limit theorems for sums of random exponentials. *Probab. Theory Related Fields*, 132(4):579–612, 2005.

[2] G. Ben Arous, A. Bovier, and J. Černý. Universality of the REM for dynamics of mean-field spin glasses. *Commun. Math. Phys.*, 282(3):663–695, 2008.

[3] G. Ben Arous, A. Bovier, and V. Gayrard. Glauber dynamics of the random energy model. I. Metastable motion on the extreme states. *Commun. Math. Phys.*, 235(3):379–425, 2003.

[4] G. Ben Arous, A. Bovier, and V. Gayrard. Glauber dynamics of the random energy model. II. Aging below the critical temperature. *Commun. Math. Phys.*, 236(1):1–54, 2003.

[5] G. Ben Arous and J. Černý. Bouchaud's model exhibits two aging regimes in dimension one. *Ann. Appl. Probab.*, 15(2):1161–1192, 2005.

[6] G. Ben Arous and J. Černý. Dynamics of trap models. In *École d'Été de Physique des Houches, Session LXXXIII "Mathematical Statistical Physics"*, pages 331–394. Elsevier, 2006.

[7] G. Ben Arous and J. Černý. The arcsine law as a universal aging scheme for trap models. *Commun. Pure Appl. Math.*, 61(3):289–329, 2008.

[8] G. Ben Arous, J. Černý, and T. Mountford. Aging in two-dimensional Bouchaud's model. *Probab. Theory Related Fields*, 134(1):1–43, 2006.

[9] J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[10] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[11] P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[12] P. Billingsley. *Weak convergence of measures: Applications in probability.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.

[13] J.-P. Bouchaud. Weak ergodicity breaking and aging in disordered systems. *J. Phys. I (France)*, 2:1705–1713, 1992.

[14] J.-P. Bouchaud and D. S. Dean. Aging on Parisi's tree. *J. Phys I(France)*, 5:265, 1995.

[15] A. Bovier and V. Gayrard. Convergence of clock processes in random environments and ageing in the p-spin sk model. preprint, Bonn University, SFB 611, 2010.

[16] A. Bovier and I. Kurkova. Poisson convergence in the restricted k-partitioning problem. *Random Structures Algorithms*, 30(4):505–531, 2007.

[17] A. Bovier, I. Kurkova, and M. Löwe. Fluctuations of the free energy in the REM and the p-spin SK models. *Ann. Probab.*, 30(2):605–651, 2002.

[18] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes*. Springer Series in Statistics. Springer-Verlag, New York, 1988.

[19] R. Durrett and S. I. Resnick. Functional limit theorems for dependent variables. *Ann. Probab.*, 6(5):829–846, 1978.

[20] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[21] W. Feller. A limit theorem for random variables with infinite moments. *Amer. J. Math.*, 68:257–262, 1946.

[22] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[23] V. Gayrard. Aging in reversible dynamics of disordered systems. I. Emergence of the arcsine law in Bouchaud's asymmetric trap model on the complete graph. preprint, LAPT, Univ. Marseille, 2010.

[24] V. Gayrard. Aging in reversible dynamics of disordered systems. II. Emergence of the arcsine law in the random hopping time dynamics of the REM. preprint, LAPT, Univ. Marseille, 2010.

[25] V. Gayrard. Aging in reversible dynamics of disordered systems. III. Emergence of the arcsine law in the Metropolis dynamics of the REM. preprint, LAPT, Marseille, 2010.

[26] B. V. Gnedenko and A. N. Kolmogorov. *Predel′nye raspredeleniya dlya summ nezavisimyh slučaĭnyh veličin.* Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad,], 1949.

[27] K. Itô. *Stochastic processes.* Springer-Verlag, Berlin, 2004. Lectures given at Aarhus University, Reprint of the 1969 original, Edited and with a foreword by Ole E. Barndorff-Nielsen and Ken-iti Sato.

[28] A. Janßen. Limit laws for power sums and norms of i.i.d. samples. *Probab. Theory Related Fields*, 146(3-4):515–533, 2010.

[29] O. Kallenberg. *Random measures.* Akademie-Verlag, Berlin, 1983.

[30] A. N. Kolmogorov. On the Skorohod convergence. *Teor. Veroyatnost. i Primenen.*, 1:239–247, 1956.

[31] M. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, 1983.

[32] S. Resnick. *Extreme values, regular variation, and point processes*, vol-

ume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York, 1987.

[33] W. Whitt. *Stochastic-process limits.* Springer Series in Operations Research. Springer-Verlag, New York, 2002.

[34] M. J. Wichura. Functional laws of the iterated logarithm for the partial sums of I.I.D. random variables in the domain of attraction of a completely asymmetric stable law. *Ann. Probability*, 2:1108–1138, 1974.

Index