Point Process Characterization of Extremes

Master Thesis

Author:
Csaba Hari
20-743-811

Supervisor:
Prof. Dr. Jean Bertoin

University of Zurich
Institute of Mathematics

June 3, 2026

Abstract

This thesis develops the point process characterization of extremes, unifying the block maxima and peaks-over-threshold methods of univariate extreme value theory.

We introduce Poisson point processes on Polish spaces, characterize their distributions via the Laplace functional, and establish convergence criteria through vague convergence of Radon measures.

The central result is the Poisson Limit Theorem for Extremes: for an i.i.d. sequence (X_i) with distribution F in the domain of attraction of a generalized extreme value distribution G_ξ, the normalized sample cloud $N_n = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}$ converges in distribution to $N \sim \text{PPP}(\mu)$, where $G(x) = \exp(-\mu((x, \infty)))$.

This single limiting object recovers both classical limits—the block maxima limit G_ξ and the peaks-over-threshold limit H_ξ—as distributional properties of N, naturally explaining why both approaches share the same shape parameter ξ.

Acknowledgements

First and foremost, I wish to thank Professor Jean Bertoin for giving me the opportunity to write this thesis on such a fascinating topic, and for the freedom he granted me in taking it in the direction I found most compelling. It was in his probability theory course that I first encountered the subject that would come to define the rest of my mathematical studies, and for that I am especially grateful.

I also wish to thank my friends at university, who went through every high and low of student life with me and made these years truly memorable.

To my family, who supported me unconditionally throughout my studies and were always there whenever I needed them — thank you.

Finally, I want to thank Eve, who patiently endured countless variations of the sentence: *"Imagine a square tile with some raindrops spread across..."* Her support and encouragement meant more than she knows.

Contents

Introduction **1**

1 Poisson Point Processes **3**

1.1 Poisson Random Variables 3

1.2 General Setup and Notation 5

1.3 Poisson Point Processes 7

1.4 Simulations and Examples 10

1.5 Properties of Poisson Point Processes 13

1.6 Distribution of Poisson Point Processes 18

2 Convergence **25**

2.1 Setup 26

2.2 Vague Convergence of Radon Measures (Deterministic Level) 30

2.3 Convergence in Distribution of Point Processes (Random Level) 32

3 Extreme Value Theory **37**

3.1 Maxima and Minima of Independent Random Variables 37

3.2 Block Maxima: Extreme Value Distributions and Max-Stability 40

3.3 Peaks Over Threshold: Generalized Pareto Distributions 44

4 Point Process Characterization of Extremes **53**

4.1 Setup: The Normalized Sample Cloud and the State Space 53

4.2 The Poisson Limit Theorem for Extremes 55

4.3 The Exponent Measure and its Classification 58

4.4 Connection to the Peaks-over-Threshold Approach 59

4.5 Worked Examples 61

Outlook **65**

List of Figures **67**

Bibliography **68**

Introduction

Extreme value theory is concerned with the statistical behaviour of unusually large or small observations. While classical probability theory focuses often on the average behaviour of large samples, extreme value theory asks a fundamentally different question: what can we say about the largest observation in a large sample? Such questions arise naturally in a wide range of applications. In hydrology, one is interested in the largest river discharge in a given year in order to design flood defences. In finance, one is interested in the largest daily loss of a portfolio in order to quantify tail risk. In insurance, one is interested in the largest claim in a given period in order to set appropriate reserves.

Two classical approaches have been developed to answer these questions.The *block maxima* approach studies the distribution of the sample maximum $M_n = \max(X_1, \ldots, X_n)$ of n independent and identically distributed observations, and asks whether the normalised maximum $(M_n - b_n)/a_n$ converges in distribution to a non-degenerate limit as $n \to \infty$. The answer is given by the Fisher–Tippett–Gnedenko theorem: the only possible limit laws are the generalised extreme value distributions G_ξ, parametrised by a single shape parameter $\xi \in \mathbb{R}$. The *peaks-over-threshold* approach instead retains all observations exceeding a high threshold u, and studies their conditional distribution above u. The Pickands–Balkema–de Haan theorem shows that these excess distributions converge to the generalised Pareto distribution H_ξ — with the same shape parameter ξ as in the block maxima approach.

The fact that both approaches yield the same shape parameter ξ is not a coincidence, but it is not immediately obvious why this should be the case from the classical theory alone. The goal of this thesis is to provide a natural and unified explanation, using the language of Poisson point processes. The key idea is to represent the entire collection of normalised observations as a random counting measure, the *normalised sample cloud*

$$N_n = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n},$$

and to study the limit of N_n as a point process rather than reducing it to a single statistic. The central result, the Poisson Limit Theorem for Extremes, states that N_n converges in distribution to a Poisson point process $N \sim \mathrm{PPP}(\mu)$, where the exponent measure μ satisfies $G(x) = \exp(-\mu((x, \infty]))$. Both the block maxima and peaks-over-threshold results then emerge as projections of this single limiting object: the block maxima limit G_ξ is the distribution of the rightmost point of N, and the peaks-over-threshold limit H_ξ is the conditional distribution of a point of N given that it exceeds a threshold. The shared shape parameter ξ is explained naturally: both arise from the same exponent measure μ.

Structure of the thesis

The thesis is organised as follows. **Chapter 1** introduces Poisson point processes on Polish spaces, establishing their fundamental properties and characterising their distributions via the Laplace functional. **Chapter 2** develops the necessary notion of convergence, introducing vague convergence of Radon measures and convergence in distribution of point processes, culminating in a practical criterion for convergence and its application to Poisson point processes. **Chapter 3** reviews the classical theory of univariate extreme value theory, covering the block maxima approach and the peaks-over-threshold approach, including the Fisher–Tippett–Gnedenko theorem and the Pickands–Balkema–de Haan theorem. **Chapter 4** presents the main results of the thesis: the Poisson Limit Theorem for Extremes, the classification of exponent measures, and the connection to the peaks-over-threshold approach, illustrated through three canonical worked examples.

References and structure

The overall structure of this thesis follows the approach of Balkema–Embrechts [Guu07], which serves as the primary reference. The theory of Poisson point processes and their introduction in Chapter 1 draws primarily on the lecture notes [**Tas25**]. The classical extreme value theory of Chapter 3 follows de Haan–Ferreira [**Lau07**] and Coles [**Col01**], which provide treatments of aspects not covered in [Guu07].

Prerequisites

The reader is assumed to be familiar with measure theory and basic probability theory. In particular, familiarity with σ-algebras, Borel measures, and some concepts of topology and functional analysis is assumed throughout.

1 Poisson Point Processes

1.1 Poisson Random Variables

Definition 1.1. [Tas25] Let $\lambda \in [0, \infty]$. A random variable N is a *Poisson random variable with parameter* λ if it takes values in $\mathbb{N}_0 \cup \{\infty\}$ and

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k \in \mathbb{N}_0 \cup \{\infty\}.$$

We write $N \sim \text{Poi}(\lambda)$.

Remark 1.2. (Basic properties)[Tas25] Recall the following facts from basic probability courses. Let $\lambda \in [0, \infty]$ and $N \sim \text{Poi}(\lambda)$.

- $\mathbb{E}(N) = \lambda$
- $\text{Var}(N) = \lambda$
- If $\lambda = 0$, then $N = 0$ a.s.
- If $\lambda = +\infty$ then $N = +\infty$ a.s. by convention.
- If $\lambda < \infty$, then $N < \infty$ a.s.
- The *Laplace transform* of N is

$$\mathbb{E}[e^{-tN}] = \exp\left(-\lambda(1 - e^{-t})\right), \qquad t \geq 0.$$

Proposition 1.3. *(Superposition) [Tas25] Let $(\lambda_i)_{i \geq 1}$ be a sequence of real nonnegative numbers. Let $(N_i)_{i \geq 1}$ be independent random variables such that $N_i \sim \text{Poi}(\lambda_i)$ for every $i \geq 1$. Then, the sum satisfies*

$$N := \sum_{i \geq 1} N_i \sim \text{Poi}(\lambda), \quad \textit{where } \lambda = \sum_{i \geq 1} \lambda_i.$$

Sketch of Proof. We proceed in two steps:

1. **Finite Case:** Let $k \geq 1$. Since the random variables N_i are independent, the distribution of the partial sum $S_k = \sum_{i=1}^{k} N_i$ is the convolution of the individual Poisson distributions. Standard computations show that:

$$S_k \sim \text{Poi}\left(\sum_{i=1}^{k} \lambda_i\right).$$

2. **Infinite Case:** Let S_k be the partial sum defined above. The sequence $(S_k)_{k\geq 1}$ is monotone increasing and converges almost surely to $N = \sum_{i=1}^{\infty} N_i$. If $\lambda = \sum_{i=1}^{\infty} \lambda_i < \infty$, we can take the limit $k \to \infty$ in the probability mass function of S_k:

$$\mathbb{P}(S_k = n) = e^{-\sum_{i=1}^{k} \lambda_i} \frac{(\sum_{i=1}^{k} \lambda_i)^n}{n!} \xrightarrow{k\to\infty} e^{-\lambda} \frac{\lambda^n}{n!}.$$

Thus, $N \sim \text{Poi}(\lambda)$. (The case where $\lambda = \infty$ leads to $N = \infty$ almost surely).

□

Proposition 1.4. *(Thinning)[Tas25] Let N be a Poisson random variable with parameter $\lambda > 0$. Let $(X_n)_{n\geq 1}$ be i.i.d. random variables in $\mathbb{N}$ independent of N. For every $i \in \mathbb{N}$, set $p_i = \mathbb{P}(X_1 = i)$ and define*

$$N_i = \sum_{n=1}^{N} \mathbf{1}_{X_n = i}.$$

The random variables $(N_i)_{i\geq 1}$ are independent with $N_i \sim \text{Poi}(\lambda p_i)$.

Sketch of Proof. We first consider the case where $X_1 \sim \text{Ber}(p)$, and let $p_0 = 1 - p$ and $p_1 = p$. Let $n_0, n_1 \in \mathbb{N}$ and set the total count $n = n_0 + n_1$. Conditioning on the total number of events $N = n$, the random variable N_1 (successes) follows a Binomial distribution $\text{Bin}(n, p_1)$. We compute the joint probability:

$$\begin{aligned}
\mathbb{P}(N_0 = n_0, N_1 = n_1) &= \mathbb{P}(N = n)\mathbb{P}(Y = n_1) \\
&= \frac{e^{-\lambda}\lambda^n}{n!} \cdot \binom{n}{n_1} p_1^{n_1} p_0^{n_0} \\
&= \left(\frac{e^{-\lambda p_0 n_0} (\lambda p_0)^{n_0}}{n_0!} \right) \left(\frac{e^{-\lambda p_1 n_1} (\lambda p_1)^{n_1}}{n_1!} \right).
\end{aligned}$$

This factorization shows that N_0 and N_1 are independent Poisson random variables with parameters λp_0 and λp_1. The general case for countably many types $(N_i)_{i\geq 1}$ follows by induction, effectively treating each N_k as a binary thinning of the remaining process $N_{\geq k}$.
□

Notice that Proposition 1.3 and Proposition 1.4 are inverse operations of one another. If we have for example wind claims and flood claims in an insurance company, Superposition tells us what happens when we combine the wind claims and flood claims datasets. Thinning tells us what happens if we filter a dataset e.g., "Let's only look at claims in Zürich," or "Let's only look at claims larger than 1 million Euros". Superposition will be an important tool later for connecting Poisson point processes to the advanced Extreme Value Theory.

1.2 General Setup and Notation

Standard probability theory provides us with a clear framework for modeling a single random location in space: the random variable. If we let X take values in a space S (e.g., $\mathbb{R}^d$), the behavior of this single point is fully captured by its probability distribution.

The challenge arises when we wish to model a random configuration of points, such as the epicenters of earthquakes, the arrival times of claims, the locations of stars or the locations of chocolate pieces in a cake. In these cases, not only are the locations random, but the total number of points may be random as well.

One might initially attempt to model such a system using a random sequence of vectors $(X_1, X_2, \dots)$. While intuitively appealing, this vector-based approach can become cumbersome, particularly when the number of points is random or when no natural ordering of the points exists. To avoid these difficulties, the theory of Point Processes adopts a functional perspective. Instead of treating the points as a list of coordinates, we view the entire configuration as a random integer-valued measure (a random counting measure).

This change in perspective from random variables to random measures, is the foundational step that allows us to rigorously apply tools from analysis and measure theory to the study of random scatter.

Notation Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and Let $(\mathcal{X}, d)$ be a nonempty Polish space (separable, complete, metric space). Let $\mathcal{B}$ denote the Borel σ-algebra on $\mathcal{X}$. Finally, let μ be a σ-finite measure on $(\mathcal{X}, \mathcal{B})$, meaning that there exists a countable partition

$$\mathcal{X} = \bigcup_{i \in \mathbb{N}} B_i$$

such that $\mu(B_i) < \infty$ for all i.

We denote by $\mathcal{M}$ the space of σ-finite, integer valued measures on $(\mathcal{X}, \mathcal{B})$. Formally

$$\mathcal{M} = \{\text{sigma-finite measures } \nu \text{ on } \mathcal{X} \text{ such that } \forall B \in \mathcal{B},\ \nu(B) \in \mathbb{N} \cup \{\infty\}\}.$$

The canonical σ-algebra on $\mathcal{M}$ is the smallest σ-algebra that makes all evaluation maps measurable. Formally

$$\mathcal{G} = \sigma\big(\{\nu \in \mathcal{M} : \nu(B) \in A\} :\ B \in \mathcal{B},\ A \subset \mathbb{N} \cup \{\infty\}\big).$$

Definition 1.5. [Tas25] A *point process* on $(\mathcal{X}, \mathcal{B})$ is a stochastic process

$$M = (M(B))_{B \in \mathcal{B}}$$

with values in $\mathbb{N} \cup \{\infty\}$, such that $M \in \mathcal{M}$ almost surely.

Interpretation To understand the intuition of a point process, one needs to expand the mindset about random variables. A point process is not a random variable taking values in some space (for example $\mathbb{R}$) in the usual sense, but it is a random measure $M_\omega(\cdot) \in \mathcal{M}$ over the space $(\mathcal{X}, \mathcal{B})$. The measure M_ω tells us how many points lie in each region $B \in \mathcal{B}$. Thus, the random integer $M_\omega(B)$ represents the number of points in B. Formally, a point process is a collection $M = (M_\omega(B))_{\omega \in \Omega, B \in \mathcal{B}}$ with values in $\mathbb{N} \cup \{\infty\}$ such that

- for every fixed B, $\omega \mapsto M_\omega(B)$ is measurable,
- for almost every $\omega \in \Omega$, the mapping $M_\omega : B \mapsto M_\omega(B)$ is an element of $\mathcal{M}$.

As usual in probability, the underlying parameter $\omega \in \Omega$ is implicit and as in the definition above, we will write M instead of M_ω from now on.

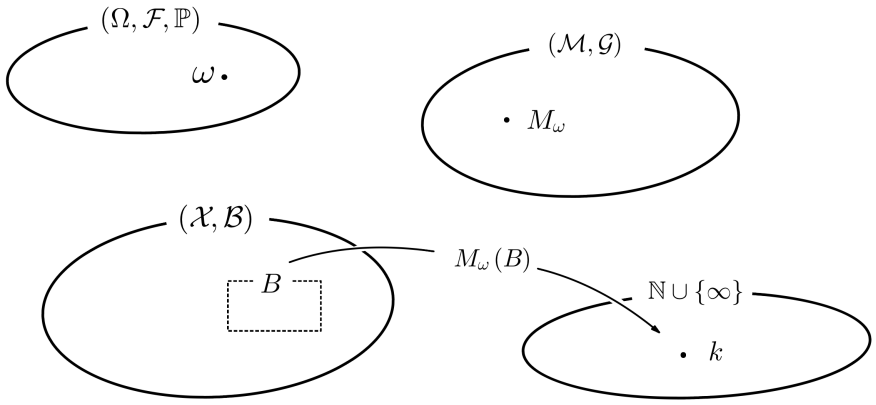

Figure 1: Overview of the spaces when working with point processes.

Example 1.6. Let $X_1, \ldots, X_n$ be i.i.d. random variables on $\mathcal{X} = [0, 1]$ and denote by $\mathcal{P} = \{X_1, \ldots, X_n\}$ the random set of points. Then the point process $M = \delta_{X_1} + \ldots + \delta_{X_n}$ corresponds to $\mathcal{P}$. ◊

Definition 1.7 (Mean/Intensity measure)**.** Let M be a point process on the Polish space $\mathcal{X}$. The *mean measure* or *intensity measure* of M is the measure ν on $(\mathcal{X}, \mathcal{B})$ defined by

$$\nu(B) := \mathbb{E}\big[N(B)\big], \qquad B \in \mathcal{B}.$$

Interpretation The mean measure of a point process plays the same role as the expectation of a random variable.

Proposition 1.8. *[Tas25](Representation as a finite Dirac sum) Let M be a point process on $(\mathcal{X}, \mathcal{B})$ such that $M(\mathcal{X}) < \infty$ almost surely. Then there exist $\sigma(M)$-measurable random variables $Z \in \mathbb{N}$ and $X_1, X_2, \cdots \in \mathcal{X}$ such that*

$$M = \sum_{i=1}^{Z} \delta_{X_i} \qquad a.s.$$

(In particular, $Z = M(\mathcal{X})$.)

Proof (sketch). Set $Z := M(\mathcal{X})$. For each ω, the measure M_ω is a finite integer-valued measure on $\mathcal{X}$, hence M_ω assigns its entire mass to finitely many points of $\mathcal{X}$, each counted with integer multiplicity, so it can be written as $M_\omega = \sum_{i=1}^{Z(\omega)} \delta_{x_i(\omega)}$ for suitable points $x_1(\omega), \ldots, x_{Z(\omega)}(\omega) \in \mathcal{X}$.

Since $\mathcal{X}$ is Polish, one can choose an enumeration of the atoms $x_i(\omega)$ in a $\sigma(M)$-measurable way (measurable enumeration of atoms of a finite counting measure). Defining $X_i(\omega) := x_i(\omega)$ gives the desired representation. □

This representation gives a correspondence between a random counting measure (the point process $M \in \mathcal{M}$) and a random set of points $\mathcal{P}_M = \{X_1, \ldots, X_Z\} \subset \mathcal{X}$. $M(B)$ corresponds to $|\mathcal{P}_M \cap B|$, the number of points of $\mathcal{P}_M$ in B (with multiplicity).

Definition 1.9 (Integral with respect to a point process)**.** Let M be a point process on $(\mathcal{X}, \mathcal{B})$ and let $f : \mathcal{X} \to [0, \infty]$ be a Borel-measurable function. Since M is integer-valued, for each fixed $\omega \in \Omega$ the measure M_ω is a deterministic counting measure, and the integral of f with respect to M is defined pathwise as the classical Lebesgue integral

$$\int_{\mathcal{X}} f(x)\, dM := \sum_{x \in \mathcal{P}_M} f(x),$$

where points are counted with multiplicity. For every $B \in \mathcal{B}$,

$$M(B) = \int_{\mathcal{X}} \mathbf{1}_B(x)\, dM.$$

1.3 Poisson Point Processes

Definition 1.10. (Poisson point process) A *Poisson point process* with mean measure μ on $(\mathcal{X}, \mathcal{B})$ is a point process M such that

1. For all $B_1, \ldots, B_k \subset \mathcal{X}$ measurable and disjoint, $M(B_1), \ldots, M(B_k)$ are independent,
2. For all $B \subset \mathcal{X}$ measurable, $M(B)$ has law $\text{Pois}(\mu(B))$.

We say $M \sim \mathrm{PPP}(\mu)$.

A Poisson point process with intensity measure μ may be interpreted as a model for points that are scattered randomly and independently in the space $\mathcal{X}$. For any measurable region $B \subset \mathcal{X}$, the random variable $M(B)$ counts the number of points falling in B. This number is Poisson distributed with $\mathbb{E}[M(B)] = \mu(B)$. Moreover, the independence of the counts on disjoint sets shows that there is no interaction between points in disjoint regions.

Intuitively, $\mu(B)$ represents the average density of points in the region B. Regions with large $\mu(B)$ are likely to contain many points, while regions with $\mu(B) = 0$ contain none almost surely. In this sense, the intensity measure μ fully determines the spatial structure of the point process.

As before, we have a representation as a proper process.

Theorem 1.11. *[Tas25] Let $M \sim \mathrm{PPP}(\mu)$ on $(\mathcal{X}, \mathcal{B})$. Let $Z := M(\mathcal{X})$ denote the total number of points in $\mathcal{X}$. Then there exist $\sigma(M)$-measurable random variables $X_n \in \mathcal{X}$, $n \geq 1$, such that*

$$M = \sum_{n=1}^{Z} \delta_{X_n} \qquad \textit{almost surely.}$$

Proof. Since μ is σ-finite, there exists a measurable partition $(B_i)_{i\in\mathbb{N}}$ of $\mathcal{X}$ such that $\mu(B_i) < \infty$ for all i. For each i, define the point process

$$M_i(B) := M(B \cap B_i), \qquad B \in \mathcal{B}.$$

Then $M_i(\mathcal{X}) = M(B_i) < \infty$ almost surely, so each M_i is a finite point process.

By Proposition 1.8 for finite point processes, there exist $\sigma(M_i)$-measurable random variables $Z^{(i)} \in \mathbb{N}$ and $X_1^{(i)}, X_2^{(i)}, \cdots \in B_i$ such that

$$M_i = \sum_{j=1}^{Z^{(i)}} \delta_{X_j^{(i)}} \qquad \text{almost surely.}$$

Since the sets (B_i) form a partition of $\mathcal{X}$, we have

$$M = \sum_{i=1}^{\infty} M_i.$$

Reordering the (countable) terms $\{X_j^{(i)}\}_{i,j}$ yields a sequence $(X_n)_{n\geq 1}$ such that

$$M = \sum_{n=1}^{Z} \delta_{X_n}, \qquad Z := M(\mathcal{X}),$$

which completes the proof. □

Proposition 1.12. *[Tas25](ppp existence on spaces with finite measure) Assume that $\mu(\mathcal{X}) < \infty$. Let Z and $(X_i)_{i\geq 1}$ be independent random variables such that*

$$Z \sim \mathrm{Pois}\big(\mu(\mathcal{X})\big), \qquad X_i \sim \frac{\mu(\cdot)}{\mu(\mathcal{X})}, \quad i \geq 1.$$

Then the random measure

$$M := \sum_{i=1}^{Z} \delta_{X_i}$$

is a Poisson point process with intensity measure μ on $(\mathcal{X}, \mathcal{B})$.

Proof. Let $k \geq 2$. Let $B_1, \ldots, B_{k-1} \in \mathcal{B}$ be disjoint, and set

$$B_k := \mathcal{X} \setminus \bigcup_{i=1}^{k-1} B_i.$$

Fix $n_1, \ldots, n_k \in \mathbb{N}$ and write $n := n_1 + \cdots + n_k$. Conditionally on $Z = n$, the points $X_1, \ldots, X_n$ are i.i.d. with distribution $\mu(\cdot)/\mu(\mathcal{X})$. Hence the vector $\big(M(B_1), \ldots, M(B_k)\big)$ has a multinomial distribution with parameters $\left(n; \frac{\mu(B_1)}{\mu(\mathcal{X})}, \ldots, \frac{\mu(B_k)}{\mu(\mathcal{X})}\right)$ independent of Z.

Therefore,

$$\mathbb{P}(M(B_1) = n_1, \ldots, M(B_k) = n_k) = \prod_{i=1}^{k} \frac{\mu(B_i)^{n_i}}{n_i!}\, e^{-\mu(B_i)}.$$

Summing over n_k yields

$$\mathbb{P}(M(B_1) = n_1, \ldots, M(B_{k-1}) = n_{k-1}) = \prod_{i=1}^{k-1} \frac{\mu(B_i)^{n_i}}{n_i!}\, e^{-\mu(B_i)},$$

which shows that $M(B_1), \ldots, M(B_{k-1})$ are independent Poisson random variables with parameters $\mu(B_1), \ldots, \mu(B_{k-1})$. □

Theorem 1.13. *[Tas25](ppp existence on spaces with infinite measure) Assume that μ is a σ-finite measure on $(\mathcal{X}, \mathcal{B})$, then there exists $M \sim \mathrm{PPP}(\mu)$ on $\mathcal{X}$.*

Proof. Let $\mathcal{X} = \bigcup_{i=1}^{\infty} A_i$ such that $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$. Let $\mu = \sum_{i=1}^{\infty} \mu_i$ where $\mu_i(B) := \mu(A_i \cap B)$ and thus $\mu_i(\mathcal{X}) < \infty$ for all $i \in \mathbb{N}$. Let (M_i) be independent $PPP(\mu_i)$. By superposition (Theorem 1.14), $M = \sum_{i=1}^{\infty} M_i$ is a $\mathrm{PPP}(\mu)$. □

1.4 Simulations and Examples

Proposition 1.12 provides a practical and flexible way to simulate Poisson point processes with a given intensity measure. In particular, when the intensity measure is finite, a Poisson point process can be constructed by first sampling a Poisson-distributed number of points and then placing them independently according to the normalized intensity measure.
Poisson point processes are frequently used as models for real-world data consisting of scattered points in space, often referred to as *sample clouds*. The simulation of Poisson point processes therefore serves both as a tool for building intuition and as a means to illustrate and verify theoretical results. In the following examples, we demonstrate how different choices of the intensity measure influence the geometric and probabilistic structure of the resulting point configurations.

One-dimensional Poisson process via exponential waiting times.

We begin by simulating a one-dimensional Poisson process (not a Poisson point process!) using its classical construction in terms of exponential waiting times. Starting at time 0, successive event times are generated iteratively by adding independent waiting times that follow an exponential distribution.

More precisely, if $T_1, T_2, \dots$ denote the event times, then

$$T_n = \sum_{i=1}^{n} W_i,$$

where $(W_i)_{i\geq 1}$ are independent and identically distributed exponential random variables with rate $\lambda > 0$. This construction reflects the defining property of the Poisson process that the waiting times between consecutive events are independent and exponentially distributed. For a detailed definition of Poisson processes, see [Tas25, Chapter 9].

In the implementation, the simulation is performed iteratively. Starting from an empty list of event times, a new waiting time is sampled and added to the most recent event time. This procedure is repeated until the current event time exceeds a prescribed observation window. The resulting list of event times represents a realization of a Poisson process on a finite time interval. As it turns out, the corresponding point configuration can be viewed as a Poisson point process on $\mathbb{R}_+$ with intensity measure $\lambda \,\mathrm{Leb}$.

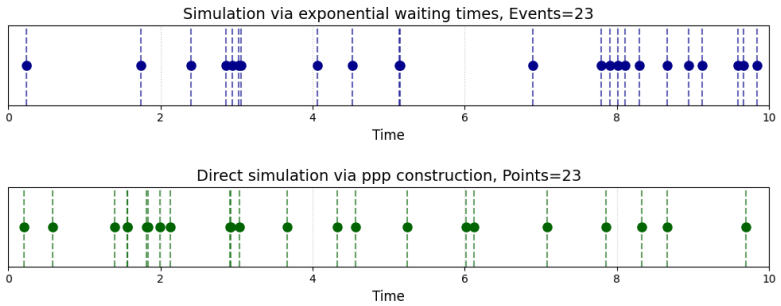

Figure 2: Comparison of two simulations of a one-dimensional Poisson point process on $[0, 10]$ with rate $\lambda = 2$: exponential waiting times (top) and direct Poisson point process construction (bottom). The observed number of points being 23 in both cases is purely coincidental.

Direct simulation via the Poisson point process construction.

We next simulate a one-dimensional Poisson point process using the general construction from Proposition 1.12. On a fixed interval $[0, T]$, we first sample the total number of points $Z \sim \text{Pois}(\lambda T)$ and then generate Z independent points uniformly on $[0, T]$.

This approach differs from the simulation via exponential waiting times in that the points are generated directly as a random configuration, rather than as the outcome of a sequential process evolving in time. Nevertheless, both procedures produce realizations of the same Poisson point process with intensity measure $\lambda \,\text{Leb}$ as proven in [Tas25, Chapter 9]. Accordingly, the resulting plot is statistically indistinguishable from the previous method as seen in figure 2.

Higher-dimensional Poisson point processes.

We now consider homogeneous Poisson point processes in two and three dimensions as shown in figure 3. This illustrates the scalability and geometric flexibility of the model. On a fixed Borel set $B \subset \mathbb{R}^d$, $d = 2, 3$, we first sample the total number of points $Z \sim \text{Pois}(\lambda \,\text{Leb}(B))$ and then generate Z independent points uniformly in B.

Inhomogeneous Poisson point processes.

Finally, we simulate a two-dimensional Poisson point process with a spatially varying intensity, illustrating how non-homogeneous intensities influence the distribution of points. We consider the domain $B = [0, 1]^2 \subset \mathbb{R}^2$ and define the intensity function

$$f(x, y) = 1 - (x^2 + y^2), \qquad (x, y) \in B.$$

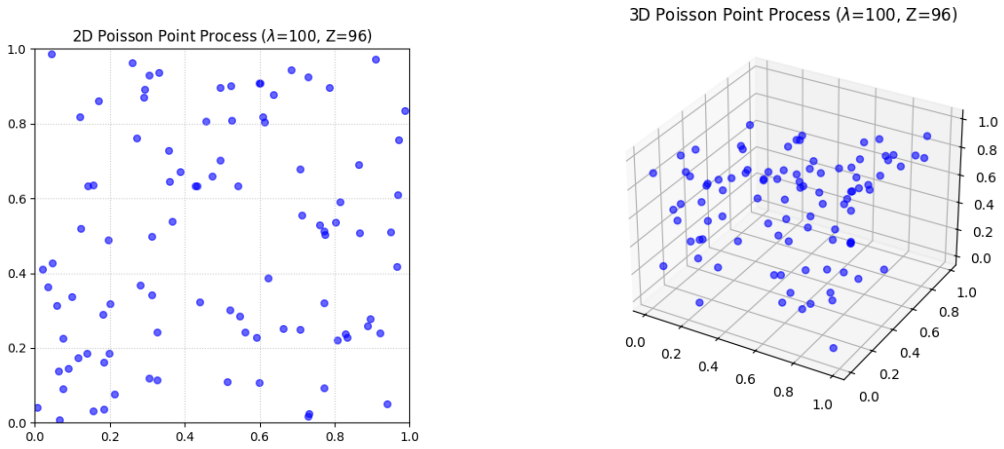

Figure 3: Two realizations of a homogeneous Poisson point process with intensity $\lambda = 100$ and intensity measure $\mu = \lambda \,\mathrm{Leb}$. Left: simulation on $[0,1]^2 \subset \mathbb{R}^2$. Right: simulation on $[0,1]^3 \subset \mathbb{R}^3$. As in the example before, in both cases the observed number of points is $Z = 96$, which is coincidental.

Note that $f \geq 0$ on B. Let $\lambda = 1000$ and define the intensity measure

$$\mu(A) := \lambda \int_A f(x,y)\,dx\,dy, \qquad A \in \mathcal{B},$$

corresponding to a high intensity in the lower left corner, decreasing gradually when moving away from the corner. The total intensity on B is

$$\mu(B) = 1000 \int_0^1 \int_0^1 f(x,y)\,dx\,dy = \frac{1000}{3}.$$

We first generate the total number of points

$$Z \sim \mathrm{Pois}(\mu(B)) = \mathrm{Pois}\left(\frac{1000}{3}\right),$$

and then sample the points using a rejection sampling procedure. Points $p = (x,y) \in B$ are proposed uniformly, the value $f(p)$ is computed, and the point is accepted with probability proportional to $f(p)$. This procedure is repeated until Z points are accepted.

As shown in figure 4, the resulting configuration exhibits a higher density of points near the lower left corner of the square and fewer points towards the upper right corner, reflecting the spatial variation of the intensity. This example illustrates how Poisson point processes with non-constant intensity measures can be simulated using the construction from Proposition 1.12.

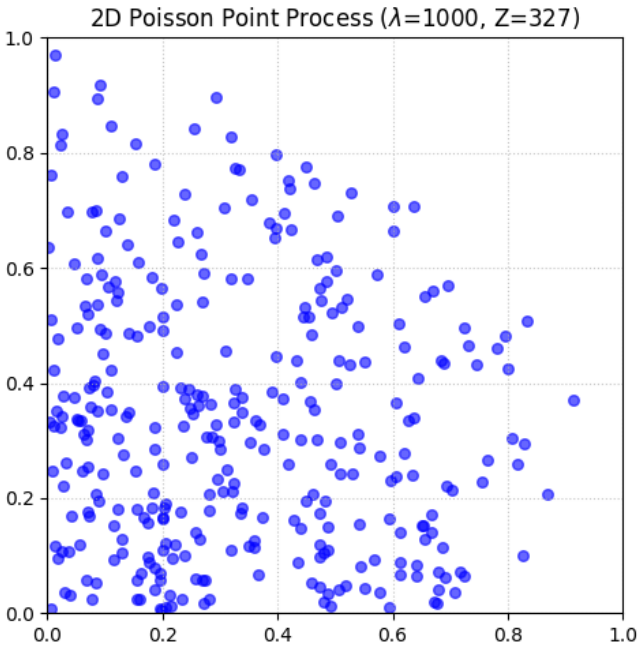

Figure 4: Simulation of an inhomogeneous Poisson point process on $[0,1]^2$ with intensity density $f(x,y) = 1-(x^2+y^2)$ and $\lambda = 1000$. The point density decreases as (x,y) moves away from the origin, reflecting the spatially varying intensity.

1.5 Properties of Poisson Point Processes

Superposition

As we have seen in Proposition 1.3, the sum of independent Poisson random variables is again Poisson. Since the mean measure of a Poisson point process determines its distribution, it is natural to ask how intensities behave under addition of independent processes. The next theorem shows that superposition extends from random variables to random measures and that intensities of disjoint regions add.

Theorem 1.14. *[Tas25](Superposition) Let $(M_i)_{i\geq 1}$ be independent Poisson point processes on $(\mathcal{X},\mathcal{B})$ with intensity measures $(\mu_i)_{i\geq 1}$. Assume that each μ_i is σ-finite and that $\mu := \sum_{i=1}^{\infty}\mu_i$ is σ-finite. Then the superposition*

$$M := \sum_{i=1}^{\infty} M_i, \qquad \text{i.e. } M(B) = \sum_{i=1}^{\infty} M_i(B) \quad \text{for } B \in \mathcal{B}.$$

is a $\mathrm{PPP}(\mu)$ *on* $(\mathcal{X},\mathcal{B})$.

Proof. Define $M(B) := \sum_{i\geq 1} M_i(B)$ for $B \in \mathcal{B}$. Since each M_i is a point process, $M(B)$ is

a well-defined $\mathbb{N}\cup\{\infty\}$-valued random variable and M is almost surely a measure. Because $\mu = \sum_{i\geq 1}\mu_i$ is σ-finite, there exists a partition $(E_n)_{n\geq 1}$ of $\mathcal{X}$ with $\mu(E_n) < \infty$, and

$$\mathbb{E}[M(E_n)] = \sum_{i\geq 1}\mathbb{E}[M_i(E_n)] = \sum_{i\geq 1}\mu_i(E_n) = \mu(E_n) < \infty,$$

which implies $M(E_n) < \infty$ almost surely. Hence M is a point process.

For any $B \in \mathcal{B}$, the random variables $M_i(B)$ are independent Poisson with parameters $\mu_i(B)$, so by superposition for Poisson random variables (Proposition 1.3),

$$M(B) = \sum_{i\geq 1} M_i(B) \sim \mathrm{Pois}\big(\mu(B)\big).$$

Finally, for disjoint $B_1, \ldots, B_k$, the families $\{M_i(B_j) : i \geq 1\}$ are independent for different j, which implies that $M(B_1), \ldots, M(B_k)$ are independent. Therefore M is a Poisson point process with intensity measure μ. □

Restriction

A central characteristic of Poisson point processes is that disjoint regions behave independently. The restriction theorem formalizes this intuition: cutting the space into pieces produces independent Poisson point processes on each part.

Theorem 1.15 (Restriction)**.** *Let μ be a σ-finite measure on $(\mathcal{X}, \mathcal{B})$ and let M be a Poisson point process on $\mathcal{X}$ with intensity measure μ.*

1. ***Restriction.*** *For any $C \in \mathcal{B}$, define the restricted random measure*

$$M_C(B) := M(B \cap C), \qquad B \in \mathcal{B}.$$

   Then M_C is a Poisson point process on $(\mathcal{X}, \mathcal{B})$ with intensity measure

$$\mu_C(B) := \mu(B \cap C), \qquad B \in \mathcal{B}.$$

2. ***Independence on disjoint sets.*** *If $C_1, C_2, \ldots, C_k \in \mathcal{B}$ are pairwise disjoint, then the restricted processes $M_{C_1}, \ldots, M_{C_k}$ are independent, and each satisfies $M_{C_i} \sim \mathrm{PPP}(\mu_{C_i})$.*

Proof. (1) Restriction. Fix $C \in \mathcal{B}$. Since $M \sim \mathrm{PPP}(\mu)$,

$$M_C(B) = M(B \cap C) \sim \mathrm{Poisson}(\mu(B \cap C)) = \mathrm{Poisson}(\mu_C(B)).$$

Now let $B_1, \dots, B_m \in \mathcal{B}$ be pairwise disjoint. Then the sets $B_1 \cap C, \dots, B_m \cap C$ are pairwise disjoint and since M is a Poisson point process, the random variables

$$M_C(B_j) = M(B_j \cap C), \qquad j = 1, \dots, m,$$

are independent. Hence $M_C \sim \mathrm{PPP}(\mu_C)$.

(2) Independence. Let $C_1, \dots, C_k \in \mathcal{B}$ be pairwise disjoint. We already know from (1) that each $M_{C_i} \sim \mathrm{PPP}(\mu_{C_i})$. To prove independence, take an arbitrary measurable sets $B \in \mathcal{B}$. Then the sets $B \cap C_i$ are pairwise disjoint, and thus the random variables

$$M_{C_i}(B) = M(B \cap C_i), \qquad i = 1, \dots, k,$$

are independent because M is a Poisson point process. Since $B \in \mathcal{B}$ was chosen arbitrarily, this is true for all $B \in \mathcal{B}$. This implies that the random measures $M_{C_1}, \dots, M_{C_k}$ are independent. □

First and Second Order Quantities

Unlike general point processes, Poisson point processes have no interaction between points. As a consequence, second-order quantities admit remarkably simple expressions. The following result shows that the variance of a Poisson stochastic integral depends only on the intensity measure.

Theorem 1.16 (First and second moment formula)**.** *Let $M \sim \mathrm{PPP}(\mu)$ and let $f : \mathcal{X} \to \mathbb{R}$ be measurable with $\int_{\mathcal{X}} |f| \, d\mu < \infty$. Then*

$$\mathbb{E}\left[\int_{\mathcal{X}} f \, dM\right] = \int_{\mathcal{X}} f \, d\mu, \qquad \mathrm{Var}\left(\int_{\mathcal{X}} f \, dM\right) = \int_{\mathcal{X}} f^2 \, d\mu.$$

In particular, if $\int f^2 \, d\mu < \infty$, then $\int f \, dM \in L^2$.

Proof. **Step 1: Step functions.** Let

$$f = \sum_{i=1}^{m} c_i \mathbf{1}_{B_i},$$

where the sets $B_1, \dots, B_m \in \mathcal{B}$ are pairwise disjoint with $\mu(B_i) < \infty$. Then

$$\int f \, dM = \sum_{i=1}^{m} c_i M(B_i).$$

Since $M(B_i) \sim \text{Poisson}(\mu(B_i))$ and the random variables $M(B_i)$ are independent,

$$\mathbb{E}\Big[\int f\,dM\Big] = \sum_{i=1}^{m} c_i \mu(B_i) = \int f\,d\mu,$$

and

$$\text{Var}\Big(\int f\,dM\Big) = \sum_{i=1}^{m} c_i^2 \mu(B_i) = \int f^2\,d\mu.$$

Step 2: Nonnegative measurable functions. If $f \geq 0$ and $\int f\,d\mu < \infty$, choose step functions $f_n \uparrow f$. By monotone convergence,

$$\int f_n\,dM \uparrow \int f\,dM, \quad \int f_n\,d\mu \uparrow \int f\,d\mu, \quad \int f_n^2\,d\mu \uparrow \int f^2\,d\mu.$$

Using Step 1 and passing to the limit yields the identities for f.

Step 3: General f. For general $f \in L^1(\mu)$ write $f = f^+ - f^-$. Since the supports $\{f > 0\}$ and $\{f \leq 0\}$ are disjoint, the integrals $\int f^+ dM$ and $\int f^- dM$ are independent (restriction property). Applying Step 2 to $f^\pm$ and using independence gives

$$\mathbb{E}\Big[\int f\,dM\Big] = \int f\,d\mu, \qquad \text{Var}\Big(\int f\,dM\Big) = \int f^2\,d\mu.$$

□

Corollary 1.17 (Covariance identity)**.** *Under the assumptions of the previous Theorem 1.16, let $f, g \in L^2(\mu)$. Then*

$$\text{Cov}\Big(\int f\,dM, \int g\,dM\Big) = \int fg\,d\mu.$$

Proof. Using

$$\text{Cov}(X,Y) = \frac{1}{2}\big(\text{Var}(X+Y) - \text{Var}(X) - \text{Var}(Y)\big)$$

and the variance formula of Theorem 1.16,

$$\text{Var}\Big(\int (f+g)\,dM\Big) = \int (f+g)^2\,d\mu,$$

the claim follows by expanding the square. □

Mapping

Consider a Polish space $(\mathcal{Y}, \mathcal{F})$ equipped with its Borel σ-algebra. Let $T : \mathcal{X} \to \mathcal{Y}$ be a measurable map. Given a measure μ on $\mathcal{X}$, we define the pushforward measure of μ under T as

$$T\#\mu(B) := \mu(T^{-1}(B)), \quad B \in \mathcal{F}.$$

The following theorem states, that Poisson point processes are preserved under measurable mappings, if the image of the mean measure is σ-finite.

Theorem 1.18 (Mapping theorem)**.** *Let $M \sim \mathrm{PPP}(\mu)$ on $\mathcal{X}$. Assume that the pushforward measure $T\#\mu$ is σ-finite on $(\mathcal{Y}, \mathcal{F})$. Define the image (pushforward) point process $T\#M$ on $(\mathcal{Y}, \mathcal{F})$ by*

$$(T\#M)(B) \;:=\; M\big(T^{-1}(B)\big), \qquad B \in \mathcal{F}.$$

Then $T\#M$ is a Poisson point process on $\mathcal{Y}$ with mean measure $T\#\mu$, i.e.$T\#M \sim \mathrm{PPP}(T\#\mu)$.

Proof. Fix $B \in \mathcal{F}$ such that $(T\#\mu)(B) < \infty$. Set $E := T^{-1}(B) \in \mathcal{B}$. Therefore, $T\#M(B)$ is a well defined random variable. Let $\mathcal{M}'$ be the space of σ-finite measures on $\mathcal{Y}$, taking values in $\mathbb{N} \cup \{\infty\}$. By definition of pushforward measures,

$$\mu(E) = \mu(T^{-1}(B)) = (T\#\mu)(B) < \infty,$$

hence $\nu \in \mathcal{M} \implies T\#\nu \in \mathcal{M}'$. Therefore, $M \in \mathcal{M}$ a.s. $\implies T\#M \in \mathcal{M}'$ a.s. and thus $T\#M$ is a point process.

Since $M \sim \mathrm{PPP}(\mu)$, we have

$$(T\#M)(B) = M(T^{-1}(B)) = M(E) \sim \mathrm{Poisson}(\mu(E)) = \mathrm{Poisson}\big((T\#\mu)(B)\big).$$

Next let $B_1, \ldots, B_m \in \mathcal{F}$ be pairwise disjoint with $(T\#\mu)(B_i) < \infty$. Then the preimages $E_i := T^{-1}(B_i)$ are pairwise disjoint measurable sets in $\mathcal{B}$. Hence the random variables $M(E_1), \ldots, M(E_m)$ are independent, because M is a Poisson point process. But $M(E_i) = (T\#M)(B_i)$, so

$$(T\#M)(B_1), \ldots, (T\#M)(B_m) \text{ are independent.}$$

Therefore, for all finite-measure sets $B \in \mathcal{F}$, the count $(T\#M)(B)$ is Poisson with mean $(T\#\mu)(B)$, and counts on disjoint sets are independent. Since $T\#\mu$ is σ-finite by assumption, this shows that $T\#M \sim \mathrm{PPP}(T\#\mu)$. □

1.6 Distribution of Poisson Point Processes

In the previous chapter we introduced Poisson point processes on a Polish space $(\mathcal{X}, \mathcal{B})$ through their defining properties, namely Poisson distributed counts on Borel sets and independence on disjoint sets. While this characterization is natural from a constructive viewpoint, it is not always the most convenient one for computations and for proving convergence. A point process $M = (M(B))_{B\in\mathcal{B}}$ is a single random object that lives in the space $\mathcal{M}$ of counting measures. Saying that two point processes M and M' are equal in law simply means

$$P_M = P_{M'},$$

i.e. they induce the same probability measure on $\mathcal{M}$. The aim of this chapter is to describe the distribution (law) of a point process in alternative ways, that are suitable for calculations and that can later be used as tools to show convergence of point processes.

Up to this point, point processes were defined as σ-finite integer-valued random measures on $(\mathcal{X}, \mathcal{B})$. For the study of distributions and, in particular, weak convergence, it is convenient to restrict attention to a smaller and more regular class of measures, namely to Radon point measures. From now on, unless stated otherwise, a point process on $(\mathcal{X}, \mathcal{B})$ will be understood as a measurable map

$$M : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow \mathcal{M}_p(\mathcal{X}),$$

where $\mathcal{M}_p(\mathcal{X})$ is the space of Radon point measures on $(\mathcal{X}, \mathcal{B})$ i.e. measures that are finite on compact subsets of $\mathcal{X}$.

The distribution

Definition 1.19. Let M and M' be two point processes on the Polish space $(\mathcal{X}, \mathcal{B})$. We say that M and M' have the same *distribution*, and write

$$M \stackrel{d}{=} M',$$

if their induced probability measure on $\mathcal{M}_p(\mathcal{X})$ coincide, i.e.

$$P_M = P_{M'}.$$

The next theorem gives us alternative equivalent definitions for equality in distribution, giving us more ways to show that two point proceses have the same distribution.

Theorem 1.20. *Let $(\mathcal{X}, d)$ be a Polish space and let M and M' be point processes on $(\mathcal{X}, \mathcal{B})$ which are almost surely finite on bounded Borel sets. The following statements are equivalent:*

1. $M \stackrel{d}{=} M'$.
2. *For all $k \geq 1$ and all bounded Borel sets $B_1, \ldots, B_k \in \mathcal{B}$,*
$$(M(B_1), \ldots, M(B_k)) \stackrel{d}{=} (M'(B_1), \ldots, M'(B_k)).$$
3. *For all bounded measurable functions f on $\mathcal{X}$ with bounded support,*
$$\int f \, dM \stackrel{d}{=} \int f \, dM'.$$

Proof. Throughout, let $\mathcal{B}_b := \{B \in \mathcal{B} : B \text{ bounded}\}$.

Step 1: (1) $\iff$ (2).

For $k \geq 1$ and $B_1, \ldots, B_k \in \mathcal{B}_b$, define the measurable map

$$\Pi_{B_1,\ldots,B_k} : \mathcal{M}_p(\mathcal{X}) \to \mathbb{N}_0^k, \qquad \Pi_{B_1,\ldots,B_k}(\mu) := (\mu(B_1), \ldots, \mu(B_k)).$$

Then sets of the form

$$\Pi^{-1}_{B_1,\ldots,B_k}(A), \qquad A \subseteq \mathbb{N}_0^k,$$

are in $\mathcal{G}$, and the collection of all such cylinder sets generates $\mathcal{G}$ (since $\mathcal{G}$ is the smallest σ-field making all evaluation maps $\mu \mapsto \mu(B)$ measurable).

If $P_M = P_{M'}$, then for every $k, B_1, \ldots, B_k$ and every $A \subseteq \mathbb{N}_0^k$,

$$\mathbb{P}\big((M(B_1), \ldots, M(B_k)) \in A\big) = \mathbb{P}\big(M \in \Pi^{-1}_{B_1,\ldots,B_k}(A)\big) = P_M\big(\Pi^{-1}_{B_1,\ldots,B_k}(A)\big)$$

$$= P_{M'}\big(\Pi^{-1}_{B_1,\ldots,B_k}(A)\big) = \mathbb{P}\big((M'(B_1), \ldots, M'(B_k)) \in A\big),$$

hence (2) holds. Conversely, if (2) holds, then P_M and $P_{M'}$ coincide on all cylinder sets and therefore on $\mathcal{G}$, so $P_M = P_{M'}$. This proves (1)$\iff$(2).

Step 2: (2) $\implies$ (3). First consider non-negative simple functions of the form

$$f = \sum_{i=1}^{m} c_i \mathbf{1}_{B_i}, \qquad c_i \geq 0, \quad B_i \in \mathcal{B}_b, \quad B_i \text{ disjoint.}$$

Then

$$\int f\,dM = \sum_{j=1}^{r} c_j M(B_j), \qquad \int f\,dM' = \sum_{j=1}^{r} c_j M'(B_j).$$

By (2), the vectors $(M(B_1),\dots,M(B_r))$ and $(M'(B_1),\dots,M'(B_r))$ have the same distribution; applying the measurable map $(x_1,\dots,x_r)\mapsto \sum_{j=1}^{r} a_j x_j$ yields

$$\int f\,dN \overset{d}{=} \int f\,dM$$

for all such simple f.

Now let f be an arbitrary bounded measurable function with bounded support, i.e. f is bounded and $f\equiv 0$ outside some bounded Borel set K. Choose simple functions f_n with support in K such that $f_n(x)\to f(x)$ pointwise and $\|f_n\|_\infty \le \|f\|_\infty$ (standard simple approximation). Since $M(K)<\infty$ a.s., the measure M charges only finitely many points in K, hence

$$\int f_n\,dM \longrightarrow \int f\,dM \quad \text{a.s.}$$

and similarly $\int f_n\,dM' \to \int f\,dM'$ a.s. For any $t\in\mathbb{R}$ we have $|e^{it\int f_n dM}|\le 1$, so by dominated convergence,

$$\mathbb{E}\Big[e^{it\int f\,dM}\Big] = \lim_{n\to\infty} \mathbb{E}\Big[e^{it\int f_n\,dM}\Big].$$

But $\int f_n\,dM \overset{d}{=} \int f_n\,dM'$ for each n, hence the characteristic functions coincide for every n, and therefore also in the limit:

$$\mathbb{E}\Big[e^{it\int f\,dM}\Big] = \mathbb{E}\Big[e^{it\int f\,dM'}\Big] \quad \text{for all } t\in\mathbb{R}.$$

Thus $\int f\,dM \overset{d}{=} \int f\,dM'$, proving (3).

Step 3: (3) $\Longrightarrow$ (2). Fix $k\ge 1$ and bounded Borel sets $B_1,\dots,B_k\in\mathcal{B}_b$. For any $t=(t_1,\dots,t_k)\in\mathbb{R}^k$, define the bounded measurable function

$$f_t := \sum_{i=1}^{k} t_i \mathbf{1}_{B_i}.$$

Its support is contained in $\bigcup_{i=1}^{k} B_i$, which is bounded, hence (3) implies

$$\sum_{i=1}^{k} t_i M(B_i) = \int f_t\,dM \overset{d}{=} \int f_t\,dM' = \sum_{i=1}^{k} t_i M'(B_i) \quad \text{for all } t\in\mathbb{R}^k.$$

By the Cramér–Wold device, this yields

$$(M(B_1), \dots, M(B_k)) \stackrel{d}{=} (M'(B_1), \dots, M'(B_k)).$$

Hence (2) holds.

Combining Steps 1–3 yields the equivalence of (1), (2), and (3). □

The Laplace functional

The Laplace functional uniquely determines the distribution of a point process. It plays the same role for point processes as the characteristic function does for real-valued random variables.

Definition 1.21 (Laplace functional)**.** Let $M \in \mathcal{M}$ (not necessarily Radon) be a point process on $(\mathcal{X}, \mathcal{B})$.

The *Laplace functional* of M is defined by

$$L_M(f) := \mathbb{E}\left[\exp\left(-\int_{\mathcal{X}} f(x)\, dM\right)\right],$$

for all non-negative measurable functions $f : \mathcal{X} \to \mathbb{R}_+$, with the convention that $e^{-\infty} = 0$.

Remark 1.22. Notice that for certain $\omega \in \Omega$ it can happen, that $\int f dM$ is infinite. This is why we need the convention that $e^{-\infty} = 0$. The larger the integral $\int f dM$, the smaller the laplace functional $L_M(f)$.

Theorem 1.23. *Let μ be a σ-finite measure on $(\mathcal{X}, \mathcal{B})$ and let $M \in \mathcal{M}$ be a point process on $(\mathcal{X}, \mathcal{B})$. Then $M \sim \mathrm{PPP}(\mu)$ if and only if for all measurable functions $f : \mathcal{X} \to [0, \infty)$,*

$$L_M(f) = \exp\left(-\int_{\mathcal{X}} \left(1 - e^{-f(x)}\right) \mu(dx)\right),$$

where $\exp(-\infty) = 0$.

Proof. "⇒" Assume $M \sim \mathrm{PPP}(\mu)$. First take a step function $f = \sum_{i=1}^{k} u_i \mathbf{1}_{B_i}$ with $u_i \geq 0$ and B_i disjoint. Then

$$\int f\, dM = \sum_{i=1}^{k} u_i M(B_i).$$

By independent increments, $M(B_i) \sim \text{Pois}(\mu(B_i))$ and the moment generating function of the Poisson random variable we have

$$L_M(f) = \prod_{i=1}^{k} \mathbb{E}[e^{u_i M(B_i)}] = \prod_{i=1}^{k} \exp\big(-\mu(B_i)(1-e^{-u_i})\big) = \exp\left(-\int (1-e^{-f})\, d\mu\right).$$

For general measurable $f \geq 0$, choose step functions $f_n \uparrow f$. Then $\int f_n\, dM \uparrow \int f\, dM$ and $1 - e^{-f_n} \uparrow 1 - e^{-f}$, hence by MCT

$$L_M(f) = \lim_{n\to\infty} L_M(f_n) = \lim_{n\to\infty} \exp\left(-\int (1-e^{-f_n})\, d\mu\right) = \exp\left(-\int (1-e^{-f})\, d\mu\right).$$

"$\Leftarrow$" Assume the Laplace functional identity holds for all $f \geq 0$. Let $B \in \mathcal{B}$ with $\mu(B) < \infty$ and take $f = t\mathbf{1}_B$, $t \geq 0$. Then

$$\mathbb{E}[e^{-tM(B)}] = L_M(t\mathbf{1}_B) = \exp\big(-\mu(B)(1-e^{-t})\big),$$

which is the moment generating function of a $\text{Pois}(\mu(B))$ random variable. Hence $M(B) \sim \text{Pois}(\mu(B))$.

Now let $B_1, \ldots, B_k$ be disjoint with $\mu(B_i) < \infty$ and take $f = \sum_{i=1}^{k} t_i \mathbf{1}_{B_i}$, $t_i \geq 0$. Then

$$\begin{aligned}\mathbb{E}\left[\exp\left(-\sum_{i=1}^{k} t_i M(B_i)\right)\right] = L_M(f) &= \exp\left(-\sum_{i=1}^{k} \mu(B_i)(1-e^{-t_i})\right) \\ &= \prod_{i=1}^{k} \exp\big(-\mu(B_i)(1-e^{-t_i})\big).\end{aligned}$$

Thus the joint Laplace transform of $(M(B_1), \ldots, M(B_k))$ factorizes into the product of the marginals, so $M(B_1), \ldots, M(B_k)$ are independent, and each $M(B_i) \sim \text{Pois}(\mu(B_i))$. Therefore M has independent Poisson increments with mean measure μ, i.e. $M \sim \text{PPP}(\mu)$. □

The following two exercises from [Guu07] serve as examples on how Theorem 1.20 and Theorem 1.23 can be applied.

Example 1.24. [Guu07, Chapter I, Exercise 3.5] Suppose $\mathcal{X}$ is the union of an increasing sequence of Borel sets B_n. Let M and N be point processes on $\mathcal{X}$. If the restrictions N^{B_n} and M^{B_n} have the same distribution for each n, then M and N have the same distribution.

Let $E_1, \ldots, E_m \in \mathcal{B}_b = \{B \subset \mathcal{X} : B \text{ bounded}\}$ be arbitrary. Since $\mathcal{X} = \bigcup_{i\geq 1} B_i$, we can find n large enough such that $E_k \subset B_n$ for all $k = 1, \ldots, m$, because E_k are bounded.

Since $M^{B_n} \stackrel{d}{=} N^{B_n}$, we have

$$M(E_k) = M(E_k \cap B_n) = M^{B_n}(E_k) = N^{B_n}(E_k) = N(E_k \cap B_n) = N(E_k).$$

Since this is true for any bounded Borel sets $E_1, \ldots, E_k$, we conclude by Theorem 1.20, that $M \stackrel{d}{=} N$. ◊

Example 1.25. [Guu07, Chapter I, Exercise 3.3] Let M be a point process on the Polish space $\mathcal{X}$ with mean measure μ which is finite on bounded sets. Suppose $L(f) = \exp(-\int 1 - e^{-f} d\mu)$ for all non-negative bounded continuous functions f on $\mathcal{X}$ with bounded support. Then M is a Poisson point process.

We want to use Theorem 1.20 and Theorem 1.23. Let $N \sim \mathrm{PPP}(\mu)$. We have

$$\mathbb{E}\left[e^{-\int f dM}\right] = e^{-\int (1-e^{-f}) d\mu} = \mathbb{E}\left[e^{-\int f dN}\right] \tag{1}$$

for all bounded continuous Borel functions f, vanishing outside a bounded set. In order to use Theorem 1.20 we need to extend this to all bounded Borel functions, not just continuous ones. Let $h \geq 0$ be an arbitrary bounded Borel function on $\mathcal{X}$ with bounded support.

Step 1: Approximate indicator functions.
Since μ is finite on bounded sets and $\mathcal{X}$ is Polish, μ is regular. Hence for every bounded Borel set E there exist compact sets $K_n \subset E$ with $K_n \uparrow E$ and $\mu(E \setminus K_n) \to 0$.
By Urysohn's lemma, for each K_n there exists a bounded continuous function g_n with bounded support such that $0 \leq g_n \leq 1$, $g_n = 1$ on K_n, and $g_n \uparrow 1_E$ μ-a.e.

Step 2: Step function approximations.
Approximate h by step functions $s_m = \sum_{j \geq 1} c_{m_j} \mathbf{1}_{E_{m_j}}$ with bounded steps E_{m_j} and $0 \leq s_m \uparrow h$ pointwise. Using the previous step for each $\mathbf{1}_{E_{m_j}}$, we approximate s_m by continuous bounded functions f_{m_k} with bounded support, such that $f_{m_k} \uparrow s_m$.

Step 3: Pass the limit twice to the Laplace functional via MCT.
Fix $m \in \mathbb{N}$. By (1) we have

$$\mathbb{E}\left[e^{-\int f_{m_k} dM}\right] = \mathbb{E}\left[e^{-\int f_{m_k} dN}\right] \quad \text{for all } k \in \mathbb{N},$$

and with $k \to \infty$ by MCT

$$\mathbb{E}\left[e^{-\int s_m dM}\right] = \mathbb{E}\left[e^{-\int s_m dN}\right] \quad \text{for all } m \in \mathbb{N}.$$

Now let $m \to \infty$. Since $0 \leq s_m \uparrow h$ pointwise and M is finite on bounded sets, we have

$$\int s_m \, dM \uparrow \int h \, dM \quad \text{a.s.}$$

Hence

$$e^{-\int s_m \, dM} \downarrow e^{-\int h \, dM} \quad \text{a.s.}$$

and by monotone convergence,

$$\mathbb{E}\big[e^{-\int h \, dM}\big] = \lim_{m \to \infty} \mathbb{E}\big[e^{-\int s_m \, dM}\big].$$

The same argument applies to N, which yields

$$\mathbb{E}\left[e^{-\int h dM}\right] = \mathbb{E}\left[e^{-\int h dN}\right]. \tag{2}$$

Since h was arbitrary we have (2) for all bounded Borel functions $h \geq 0$ with bounded support, thus by Theorem 1.20 we conclude that $M \stackrel{d}{=} N$. $\Diamond$

2 Convergence

The entire chapter answers one fundamental question: when does a sequence of point processes $(M_n)_{n\geq 1}$ converge to a limiting point process M, and how can this be verified in practice?

Up to now we have focused on identifying the law of a point process, for example with the Laplace functional. The next step is to develop a notion of convergence that is strong enough to imply convergence of the objects we care about (maxima and exceedances), but weak enough to handle the limits that occur in extreme value theory.

Why the topology matters. A point process is a random measure, so asking whether a sequence of point processes converges is really asking whether a sequence of random measures converges. The natural notion of convergence for random elements is convergence in distribution, and the question becomes: which topology should we put on the space of measures?

The answer is dictated by the setting. In extreme value theory, the limiting intensity measures are typically infinite globally. They assign infinite total mass to the whole space, while remaining finite on every compact region. A notion of convergence that tracks total mass, such as weak convergence, is therefore too strong: the sequences we care about will never converge in that sense. What we need instead is a *local* notion of convergence, one that only asks whether two measures assign approximately the same mass to every compact region, and ignores what happens far out in the tails. This is precisely vague convergence, and it is the natural topology for our purposes.

Two layers of convergence. There are two intertwined notions:

- *Vague convergence* of (deterministic) Radon measures, which is the natural mode for intensity/exponent measures arising in scaling limits.
- *Weak convergence* (i.e. convergence in distribution) of (random) point processes, meaning weak convergence of their laws on the space of Radon point measures equipped with the vague topology.

The payoff is that vague convergence of mean measures implies convergence in distribution of the associated Poisson point processes, giving a direct route from deterministic tail limits to stochastic process limits. [Guu07]

How it will be used later. The main application in this thesis is the point process characterization of extremes: we consider the normalized sample cloud $M_n = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}$

and prove $M_n \Rightarrow M$ for a Poisson point process M. This convergence is taken in the vague topology, and it leads directly to convergence of maxima via

$$G(x) = \exp\big(-\mu((x,\infty])\big),$$

where μ is the mean (exponent) measure of the limiting Poisson point process.

2.1 Setup

Section 2.1 collects the foundational definitions and results needed for the rest of Chapter 2. The theory of convergence of point processes involves several interlocking spaces, topologies, and notions of convergence, each of which requires careful setup before they can be used together. To avoid interrupting the main arguments later with digressions on basic definitions and theorems, we gather all of this preparatory material here in one place, so that Sections 2.2 and 2.3 can proceed without detours.

Convergence in distribution

Point processes are random elements of the space $M_p(\mathcal{X})$, so convergence in distribution of point processes is simply convergence in distribution of random elements in a metric space — a concept from classical probability theory. We therefore begin by recalling the relevant definitions and results in this general setting. The specific features of $M_p(\mathcal{X})$, namely the vague topology and its consequences, are then introduced in Sections 2.2 and 2.3 to make this general framework concrete and workable in our setting.

Recall the notion of weak convergence for probability measures. If S is a separable metric space, the *weak topology* on $\mathcal{P}(S)$ is the weakest topology for which

$$\mu \mapsto \int_S f \, d\mu$$

is continuous for every bounded continuous function $f \in C_b(S)$. Accordingly, a sequence $\mu_n \in \mathcal{P}(S)$ converges weakly to $\mu \in \mathcal{P}(S)$ if

$$\int_S f \, d\mu_n \longrightarrow \int_S f \, d\mu \qquad \text{for all } f \in C_b(S).$$

Definition 2.1 (Convergence in distribution)**.** Let (S,d) be a separable metric space and let $X, X_1, X_2, \ldots$ be S-valued random elements. We say that X_n *converges in distribution* to X, and write $X_n \Rightarrow X$, if

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)] \qquad \text{for all } f \in C_b(S),$$

where $C_b(S)$ denotes the space of bounded continuous functions $f : S \to \mathbb{R}$.

Remark 2.2. Let $\mathcal{L}(X_n)$ and $\mathcal{L}(X)$ denote the laws of X_n and X on $(S, \mathcal{B}(S))$. By the identity

$$\mathbb{E}[f(X_n)] = \int_S f(x)\, d\mathcal{L}(X_n)(x),$$

the above definition is equivalent to *weak convergence* of probability measures

$$\int_S f(x)\, d\mathcal{L}(X_n)(x) \to \int_S f(x)\, d\mathcal{L}(X)(x),$$

i.e. convergence tested against all bounded continuous functions.

Theorem 2.3 (Continuous Mapping Theorem)**.** *Let $X, X_1, X_2, \ldots$ be random elements of a separable metric space S, and $f : S \to V$ a continuous map into a separable metric space V. If $X_n \Rightarrow X$, then $f(X_n) \Rightarrow f(X)$.*

Proof. Let $f : S \to V$ be a continuous map. We need to show that

$$\mathbb{E}[\varphi(f(X_n))] \to \mathbb{E}[\varphi(f(X))] \qquad \text{for all } \varphi \in C_b(V).$$

Let $\varphi \in C_b(V)$ be arbitrary and let $\psi := \varphi \circ f : S \to \mathbb{R}$. Since $\varphi \in C_b(V) \Rightarrow \psi \in C_b(S)$ and therefore

$$\mathbb{E}[\varphi(f(X_n))] = \mathbb{E}[\psi(X_n)] \to \mathbb{E}[\psi(X)] = \mathbb{E}[\varphi(f(X))],$$

where we used that $X_n \Rightarrow X$. Since φ was arbitrary, this holds for all bounded continuous functions. □

Tightness

When studying convergence in distribution of a sequence of random elements (X_n), a fundamental obstruction can occur: mass may "escape to infinity." Concretely, even if X_n does not converge to any limit, it is always possible to extract a subsequence along which X_n converges — but only if the probability mass of the sequence stays concentrated in bounded regions and does not drift off to infinity. Tightness is precisely the condition that rules out this escape of mass. It requires that for any $\varepsilon > 0$, all the distributions $\mathcal{L}(X_n)$ can be simultaneously trapped inside a single compact set, up to a small error ε. Intuitively, a tight family is one where "the mass has nowhere to escape to." The importance of tightness in our setting is Prohorov's Theorem (Theorem 2.6): on a Polish space, tightness is equivalent to relative compactness in the weak topology, meaning every subsequence has a further convergent subsequence. This will be the key tool in the proof of Theorem 2.14, where we use tightness to extract convergent subsequences and then identify their limits.

Definition 2.4 (Tightness)**.** Let (S, d) be a metric space and let $(\mu_n)_{n\geq 1}$ be probability measures on $\big(S, \mathcal{B}(S)\big)$. The family (μ_n) is called *tight* if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset S$ such that

$$\mu_n(K_\varepsilon) \geq 1 - \varepsilon \qquad \text{for all } n \geq 1.$$

Equivalently, a family of S-valued random elements (X_n) is called *tight* if their laws $(\mathcal{L}(X_n))$ form a tight family of probability measures on S, i.e. for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset S$ such that

$$\mathbb{P}(X_n \in K_\varepsilon) \geq 1 - \varepsilon \qquad \text{for all } n \geq 1.$$

Theorem 2.5. *Let $(M_n)_{n\geq 1}$ be point processes on a locally compact separable metric space $\mathcal{X}$, viewed as random elements in $\mathcal{M}_p(\mathcal{X})$ endowed with the vague topology (Definition 2.7). Suppose that for every compact set $K \subset \mathcal{X}$, the family of random variables*

$$(M_n(K))_{n\geq 1}$$

is tight. Then $(M_n)_{n\geq 1}$ is tight in $\mathcal{M}_p(\mathcal{X})$.

Proof. Proven in [Guu07, theorem 4.23] □

Theorem 2.6 (Prohorov)**.** *Let (S, d) be a separable metric space and let $\{\mu_n : n \geq 1\} \subset \mathcal{P}(S)$ be a family of probability measures on $(S, \mathcal{B}(S))$, where $\mathcal{P}(S)$ is equipped with the weak topology.*

1. *If $\{\mu_n\}$ is tight, then it is relatively compact in $\mathcal{P}(S)$, i.e. every sequence (μ_n) contains a weakly convergent subsequence.*
2. *If, in addition, S is complete, then every relatively compact subset of $\mathcal{P}(S)$ is tight.*

In particular, if S is Polish (complete and separable), then a family in $\mathcal{P}(S)$ is tight if and only if it is relatively compact in the weak topology.

Proof. Proven in [Bil99, Section 5]. □

Spaces

Throughout this chapter, $\mathcal{X}$ denotes an *lcscH* space, i.e. a locally compact, second countable Hausdorff topological space. We write $\mathcal{B}(\mathcal{X})$ for its Borel σ-field. In our setting $\mathcal{X}$ is typically $\mathbb{R}^d$ or an open subset of $\mathbb{R}^d$. The lcscH assumption guarantees that Radon measures (finite on compact sets) form a natural state space and that the class $C_c(\mathcal{X})$

of compactly supported continuous functions is rich enough to define the vague topology. Moreover, with this framework the spaces of Radon measures and Radon point measures admit a well-behaved (metrizable) topology, which allows us to treat point processes as random elements in a standard way.

Throughout this chapter, $\mathcal{X}$ denotes an *lcscH* space, that is, a topological space which is locally compact, second countable, and Hausdorff. The canonical examples are $\mathbb{R}^d$ and open subsets of $\mathbb{R}^d$, which are the spaces that arise in the applications to extreme value theory in Chapters 3 and 4.

Each of the three conditions serves a specific purpose. *Local compactness* means that every point has a neighbourhood with compact closure. This makes Radon measures a natural and well-behaved class on $\mathcal{X}$, and it guarantees that the class $C_c(\mathcal{X})$ of compactly supported continuous functions is rich enough to separate measures and define the vague topology. *Second countability* provides a countable base for the topology, which implies separability and, crucially, allows the space of Radon measures $\mathcal{R}(\mathcal{X})$ to be equipped with a complete separable metric — making it a Polish space. This is what allows us to treat point processes as random elements of $M_p(\mathcal{X})$ in a standard probabilistic sense, and in particular to apply Prohorov's theorem and speak of convergence in distribution. *The Hausdorff condition* ensures that limits of sequences are unique and that the topology behaves as expected.

Together, these three conditions have a concrete consequence: one can always choose a metric on $\mathcal{X}$ such that the compact sets are precisely the bounded closed sets. Under this metric, Radon measures are exactly the measures that are finite on balls of any fixed radius. The space $\mathcal{R}(\mathcal{X})$ equipped with the vague topology is a Polish space [Guu07, Theorem 4.15], and since $M_p(\mathcal{X})$ is a closed subset of $\mathcal{R}(\mathcal{X})$, it inherits this Polish structure. The convergence theory developed in Sections 2.2 and 2.3 relies on this Polish structure.

A detailed treatment of the lcscH framework, including the construction of the metric and the proof that $\mathcal{R}(\mathcal{X})$ is Polish, can be found in Balkema–Embrechts [Guu07], Sections 4.2 and 4.4, which serve as the main reference for this chapter. Since developing this framework in full generality would take us too far from the main objectives of this thesis, we take the lcscH assumption as given and refer the interested reader there for the details.

Measure spaces. Let $\mathcal{R}(\mathcal{X})$ be the set of Radon measures on $\mathcal{X}$, i.e. Borel measures μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $\mu(K) < \infty$ for every compact set $K \subset \mathcal{X}$. Recall from the previous Chapter the space $\mathcal{M}_p(\mathcal{X}) \subset \mathcal{R}(\mathcal{X})$ of *Radon point measures*, i.e. Radon measures η satisfying $\eta(B) \in \mathbb{Z}_+ \cup \{\infty\}$ for all $B \in \mathcal{B}(\mathcal{X})$ (equivalently, $\eta(K) \in \mathbb{Z}_+$ for all compact K).

Before proceeding to the main results of this chapter, we summarise the three spaces, their

topologies, and the corresponding notions of convergence in Table 1. This overview is intended as a reference for the reader to keep in mind throughout Sections 2.2 and 2.3, where these spaces and convergence notions will be used simultaneously.

Space	**Nature**	**Topology**	**Convergence**
$(\mathcal{X}, d)$ lcscH space	**deterministic**	metric topology	—
$\mathcal{R}(\mathcal{X})$ Radon measures	**deterministic**	vague topology (Def. 2.7)	vague convergence $\mu_n \xrightarrow{v} \mu$ (Def. 2.8, Prop. 2.10)
$M_p(\mathcal{X})$ Radon point measures	**random** (random elements of $M_p(\mathcal{X})$)	vague topology induced from $\mathcal{R}(\mathcal{X})$ (subspace topology)	convergence in distribution $M_n \Rightarrow M$ (Def. 2.12, Thm. 2.14)

Table 1: Overview of the spaces, topologies, and notions of convergence used in Chapter 2. The chain $M_p(\mathcal{X}) \subset \mathcal{R}(\mathcal{X})$ reflects the fact that Radon point measures are a special class of Radon measures; the vague topology on $M_p(\mathcal{X})$ is inherited from $\mathcal{R}(\mathcal{X})$.

2.2 Vague Convergence of Radon Measures (Deterministic Level)

Definition 2.7 (Vague topology)**.** The *vague topology* on $\mathcal{R}(\mathcal{X})$ is the weakest topology for which

$$\mu \longmapsto \int_{\mathcal{X}} f \, d\mu$$

is continuous for every test function $f \in C_c(\mathcal{X})$ (continuous with compact support). We equip $\mathcal{M}_p(\mathcal{X})$ with the induced subspace topology.

Definition 2.8 (Vague convergence)**.** Let $\mathcal{X}$ be an lcscH space and let $\mu, \mu_1, \mu_2, \dots \in \mathcal{R}(\mathcal{X})$ be Radon measures. We say that (μ_n) *converges vaguely* to μ, and write $\mu_n \xrightarrow{v} \mu$, if

$$\int_{\mathcal{X}} f \, d\mu_n \longrightarrow \int_{\mathcal{X}} f \, d\mu \qquad \text{for every } f \in C_c(\mathcal{X}),$$

where $C_c(\mathcal{X})$ denotes the space of real-valued continuous functions on $\mathcal{X}$ with compact support.

Remark 2.9. Notice the difference between vague convergence and weak convergence, which

lies on the set of test functions. Weak convergence of Radon measures on an lcscH space $\mathcal{X}$ is defined by testing against functions in $C_b(\mathcal{X})$, whereas vague convergence is testing only against functions in $C_c(\mathcal{X})$, i.e. continuous functions with compact support.

If $\mathcal{X}$ is compact, then every continuous function on $\mathcal{X}$ has compact support, so

$$C_c(\mathcal{X}) = C(\mathcal{X}) = C_b(\mathcal{X}).$$

Hence on compact spaces vague convergence and weak convergence coincide.

If $\mathcal{X}$ is locally compact but not compact, then

$$C_c(\mathcal{X}) \subsetneq C_b(\mathcal{X}),$$

so vague convergence uses a smaller class of test functions than weak convergence. On a non-compact space, bounded continuous functions may detect mass escaping to infinity, whereas compactly supported test functions only probe the behaviour of the measures on compact regions. Therefore vague convergence is, in general, weaker than weak convergence on non-compact lcscH spaces.

Proposition 2.10 (Equivalent characterizations of vague convergence)**.** *Let $\mu, \mu_1, \mu_2, \ldots \in \mathcal{R}(\mathcal{X})$. The following are equivalent:*

1. $\mu_n \xrightarrow{v} \mu$;
2. $$\int_{\mathcal{X}} f \, d\mu_n \to \int_{\mathcal{X}} f \, d\mu \qquad \text{for all } f \in C_c(\mathcal{X}),\ f \geq 0;$$
3. $\mu_n(O) \to \mu(O)$ *for every relatively compact open set* $O \subset \mathcal{X}$ *with* $\mu(\partial O) = 0$.

Proof sketch. The equivalence (1) $\Leftrightarrow$ (2) is immediate by decomposing $f \in C_c(\mathcal{X})$ into its positive and negative parts.

For (2) $\Rightarrow$ (3), let $O \subset \mathcal{X}$ be relatively compact and open with $\mu(\partial O) = 0$. By Urysohn's lemma, for every compact $K \subset O$ there exists $g_K \in C_c(\mathcal{X})$ with $0 \leq g_K \leq 1$, $g_K = 1$ on K, and $\operatorname{supp}(g_K) \subset O$. Applying *2.* to g_K and using inner regularity of the Radon measure μ yields

$$\mu(O) \leq \liminf_n \mu_n(O).$$

A corresponding upper bound follows similarly, using $\mu(\partial O) = 0$, and hence $\mu_n(O) \to \mu(O)$.

For (3) $\Rightarrow$ (2), fix $f \in C_c(\mathcal{X})$, $f \geq 0$, and consider the open level sets $O_t = \{f > t\}$. They

are relatively compact. For all but countably many t, one has $\mu(\partial O_t) = 0$. Choosing step functions based on such levels gives lower and upper approximations of f, and *3.* implies convergence of their integrals. Letting the mesh go to zero yields

$$\int f \, d\mu_n \to \int f \, d\mu.$$

□

Example 2.11. [Guu07, Chapter I, Exercise 4.20] Let $\rho_n \xrightarrow{v} \rho$ in $\mathcal{R}(\mathcal{X})$, let $O \subset \mathcal{X}$ be open, and let $\psi : O \to [0, \infty)$ be continuous. Then $1_O \psi \rho_n \xrightarrow{v} 1_O \psi \rho$ in $\mathcal{R}(O)$.

Indeed, for $f \in C_c(O)$ the function $\psi f \in C_c(O)$, and its extension $\widetilde{\psi f}$ by 0 to $\mathcal{X}$ lies in $C_c(\mathcal{X})$. Thus

$$\int_O f \, d(1_O \psi \rho_n) = \int_{\mathcal{X}} \widetilde{\psi f} \, d\rho_n \to \int_{\mathcal{X}} \widetilde{\psi f} \, d\rho = \int_O f \, d(1_O \psi \rho).$$

◊

Intuitively, vague convergence controls the mass assigned to compact regions of the state space. For point processes, this is exactly the relevant local information, since a point process is a random counting measure that is finite on compact sets. We therefore use the vague topology on $\mathcal{M}_p(\mathcal{X})$ as the underlying topology when formulating convergence in distribution of point processes.

2.3 Convergence in Distribution of Point Processes (Random Level)

We now pass from deterministic Radon measures to random point measures. Since a point process on $\mathcal{X}$ is a random element of $\mathcal{M}_p(\mathcal{X})$, the natural notion of convergence is convergence in distribution of elements in the space $\mathcal{M}_p(\mathcal{X})$, which is equipped with the vague topology introduced above. The notion of convergence in distribution for point processes is analogous to the definitions given earlier, when we defined convergence in distribution for random variables.

Thus, if $M, M_1, M_2, \ldots$ are point processes on $\mathcal{X}$, we say that M_n converges in distribution to M if the laws $\mathcal{L}(M_n)$ converge weakly to $\mathcal{L}(M)$ as probability measures on $\mathcal{M}_p(\mathcal{X})$.

Definition 2.12. Let $M, M_1, M_2, \ldots$ be point processes on $\mathcal{X}$. We say that M_n converges in distribution to M, and write $M_n \Rightarrow M$, if

$$\mathbb{E}[F(M_n)] \to \mathbb{E}[F(M)] \qquad \text{for every } F \in C_b(\mathcal{M}_p(\mathcal{X})),$$

where $\mathcal{M}_p(\mathcal{X})$ is equipped with the vague topology.

Remark 2.13. Since M_n and M are $\mathcal{M}_p(\mathcal{X})$-valued random elements, $M_n \Rightarrow M$ is equivalently weak convergence of their laws $\mathcal{L}(M_n) \Rightarrow \mathcal{L}(M)$ as probability measures on $\mathcal{M}_p(\mathcal{X})$. In particular,

$$\mathbb{E}[F(M_n)] = \int_{\mathcal{M}_p(\mathcal{X})} F(\eta)\, d\mathcal{L}(M_n)(\eta),$$

and therefore $M_n \Rightarrow M$ if

$$\int_{\mathcal{M}_p(\mathcal{X})} F(\eta)\, d\mathcal{L}(M_n)(\eta) \longrightarrow \int_{\mathcal{M}_p(\mathcal{X})} F(\eta)\, d\mathcal{L}(M)(\eta) \qquad \text{for all } F \in C_b(\mathcal{M}_p(\mathcal{X})).$$

The next Theorem is the central Theorem of the random layer. It tells us that we don't need all $F \in C_b(\mathcal{M}_p(\mathcal{X}))$, it suffices to check a smaller, natural class.

Theorem 2.14 (Convergence in Distribution)**.** *Let $M, M_1, M_2, \ldots \in \mathcal{M}_p(\mathcal{X})$. The following are equivalent:*

1. *$M_n \Rightarrow M$;*
2. $$\int_{\mathcal{X}} f\, dM_n \Rightarrow \int_{\mathcal{X}} f\, dM \qquad \text{for all } f \in C_c(\mathcal{X}),\ f \geq 0;$$
3. *$(M_n(O_1), \ldots M_n(O_m)) \Rightarrow (M(O_1), \ldots, M(O_m)$ for relatively compact open sets $O_1, \ldots, O_m$ with $M(\partial O_i) = 0$ a.s. for $i = 1, \ldots, m$.*

Before we prove the Theorem, we unpack what it is really saying. We are working on the state space $\mathcal{M}_p(\mathcal{X})$, equipped with the vague topology. Statement *1.* is the definition of convergence in distribution.

Statement *2.* says: To obtain convergence in distribution of point processes, it suffices to have convergence in distribution of the real-valued random variables $\int_{\mathcal{X}} f\, dM_n$ for every $f \in C_c(\mathcal{X}),\ f \geq 0$. These maps generate the vague topology.

Statement *3.* says: It suffices to have convergence of the finite-dimensional distributions of counts on "nice" sets:

$$(M_n(O_1), \ldots M_n(O_m)),$$

for relatively compact open sets O_i whose boundaries contain almost surely no points.

Proof. (1)⇒(2): Assume $M_n \Rightarrow M$ as $\mathcal{M}_p(\mathcal{X})$-valued random elements. Fix $f \in C_c(\mathcal{X}),\ f \geq 0$. Consider the map

$$T_f : \mathcal{M}_p(\mathcal{X}) \to \mathbb{R}, \quad T_f(\mu) = \int f\, d\mu.$$

Notice that T_f is continuous by definition of the vague topology. Then by the continuous mapping Theorem (Theorem 2.3)

$$\int_{\mathcal{X}} f\, dM_n = T_f(M_n) \Rightarrow T_f(M) = \int_{\mathcal{X}} f\, dM.$$

(1) $\Rightarrow$ (3): Fix relatively compact open sets $O_1, \ldots, O_m \subset \mathcal{X}$ such that $M(\partial O_i) = 0$ a.s. for $i = 1, \ldots, m$. Consider one set O_k, $1 \leq k \leq m$. Define

$$\Psi_{O_k} : \mathcal{M}_p(\mathcal{X}) \to \mathbb{R}, \quad \Psi_{O_k}(\mu) = \mu(O_k).$$

Since $M(\partial O_k) = 0$ a.s., the map Ψ_{O_k} is continuous at M almost surely. Hence again by the continuous mapping Theorem (Theorem 2.3)

$$M_n(O_k) \Rightarrow M(O_k).$$

Now define

$$\Psi(\mu) := (\mu(O_1), \ldots, \mu(O_m)).$$

This map is continuous at every μ satisfying $\mu(\partial O_i) = 0$ for all i. Since this holds a.s. for $\mu = M$, the continuous mapping Theorem yields

$$(M_n(O_1),_n (O_m)) \Rightarrow (M(O_1), \ldots, M(O_m)$$

Showing (2) $\Rightarrow$ (1) and (3) $\Rightarrow$ (1) is more technical. We will proceed as follows:

- from the assumptions (2) or (3), show that the laws of M_n are tight;
- apply Prohorov Theorem (Theorem 2.6) to get a weakly convergent subsequence;
- identify that the subsequential limit has the same law as M;
- conclude that the whole sequence converges to M.

(2) $\Rightarrow$ (1): Assume (2). We want to show that for every compact $K \subset \mathcal{X}$, the family $\{M_n(K)\}$ is tight. Fix $K \subset \mathcal{X}$ compact. Let $f \in C_c(\mathcal{X})$ with $0 \leq f \leq 1$ and $f \equiv 1$ on K. Then

$$M_n(K) \leq \int f\, dM_n.$$

Let $\varepsilon > 0$ and fix $R \in \mathbb{R}$ such that $\mathbb{P}\left(\int f\, dM \geq R\right) < \varepsilon$. Since $\int f\, dM_n \Rightarrow \int f\, dM$, we have by the Portmanteau Theorem, that for every closed set F,

$$\limsup_{n\to\infty} \mathbb{P}\left(\int f\, dM_n \in F\right) \leq \mathbb{P}\left(\int f\, dM \in F\right).$$

Apply this to the set $F = \{x \in \mathbb{R} : |x| \geq R\}$ and so for all $\varepsilon > 0$ we get that

$$\limsup_{n\to\infty} \mathbb{P}\left(\int f\, dM_n > R\right) \leq \mathbb{P}\left(\int f\, dM > R\right) < \varepsilon.$$

Thus, there exists $n_0 \in \mathbb{N}$ such that $\mathbb{P}\left(\int f\, dM_n \geq R\right) < \varepsilon$ for all $n \geq n_0$. Since only finitely many indices $n \leq n_0$ remain, we can enlarge R to $R' \geq R$ so that $\mathbb{P}\left(\int f\, dM_n > R'\right) < \varepsilon$ for all $n \in \mathbb{N}$. Therefore $\left(\int f\, dM_n\right)$ is tight. Hence we have that for all $\varepsilon > 0$ there exists $R \in \mathbb{R}$ such that

$$\mathbb{P}\left(M_n(K) > R\right) \leq \mathbb{P}\left(\int f\, dM_n > R\right) < \varepsilon, \quad \text{for all } n \in \mathbb{N},$$

which shows by Theorem 2.5 that (M_n) is tight. Next by the Prohorov Theorem (Theorem 2.6)), every subsequence (M_{n_k}) has a convergent subsequence $(M_{n_{k_\ell}})$ such that $(M_{n_{k_\ell}}) \Rightarrow N$ for some point process $N \in \mathcal{M}_p(\mathcal{X})$. By (1) $\Rightarrow$ (2) applied to $(M_{n_{k_\ell}})$ we get

$$\int f\, dM_{n_{k_\ell}} \Rightarrow \int f\, dN \quad \text{for all } f \in C_c(\mathcal{X}),\ f \geq 0.$$

But by assumption

$$\int f\, dM_{n_{k_\ell}} \Rightarrow \int f\, dM.$$

Hence

$$\int f\, dN \stackrel{d}{=} \int f\, dM, \quad \text{for all } f \in C_c(\mathcal{X}),\ f \geq 0.$$

These integrals determine the law of a point process. Therefore,

$$N \stackrel{d}{=} M.$$

(3)$\Rightarrow$ (1): we proceed the same as above. We can show tightness, apply Prohorov Theorem and deduce that every subsequence has a convergent subsubsequence with $(M_{n_{k_\ell}}) \Rightarrow N$. Let (M_{n_k}) be an arbitrary subsequence. By tightness, it has a further subsequence $(M_{n_{k_\ell}})$ such that $M_{n_{k_\ell}} \Rightarrow N$ for some point process $N \in \mathcal{M}_p(\mathcal{X})$. Applying (1) $\Rightarrow$ (3) to this convergent subsubsequence yields

$$(N(O_1), \ldots, N(O_m)) \stackrel{d}{=} (M(O_1), \ldots, M(O_m)).$$

For relatively compact open sets $O_1, \ldots, O_m$ with $M(\partial O_i) = 0$ a.s. for $i = 1, \ldots, m$. These finite-dimensional distributions determine the law of a point process. Hence $N \stackrel{d}{=} M$, and we conclude that $M_n \Rightarrow M$. □

Theorem 2.14 enables us to state the following useful result. If we have vague convergence of

the mean measures of Poisson point processes, then we also have convergence in distribution of the Poisson point processes themselves.

Corollary 2.15 (Convergence of Poisson point processes)**.** *Let $\mathcal{X}$ be an lcscH space and let $M, M_1, M_2, \ldots \in \mathcal{M}_p(\mathcal{X})$ be Poisson point processes with mean measures $\mu, \mu_1, \mu_2, \ldots \in \mathcal{R}(\mathcal{X})$. If $\mu_n \xrightarrow{v} \mu$, then*

$$M_n \Rightarrow M$$

(where $\mathcal{M}_p(\mathcal{X})$ is equipped with the vague topology).

Proof. Recall that the Laplace functional of $N \sim \mathrm{PPP}(\rho)$ is explicitly given by

$$\mathbb{E}\left[e^{-\int f\,dN}\right] = \exp\left(-\int (1-e^{-f})d\rho\right), \quad f \geq 0.$$

If $f \in C_c^+(\mathcal{X})$ then $1-e^{-f}$ is again continuous, nonnegative and compactly supported. So vague convergence of $\mu_n \to \mu$ gives

$$\int (1-e^{-f})d\mu_n \to \int (1-e^{-f})d\mu \quad \text{for all } f \in C_c^+(\mathcal{X}).$$

Therefore

$$\mathbb{E}\left[e^{-\int f\,dM_n}\right] \to \mathbb{E}\left[e^{-\int f\,dM}\right].$$

Now replace f by tf with $t \geq 0$. Thus the Laplace transforms of the nonnegative random variables

$$X_n := \int f\,dM_n, \quad X := \int f\,dM$$

converge pointwise to the Laplace transform of X. Since Laplace transforms determine the distribution of nonnegative random variables, we obtain

$$\int f\,dM_n \Rightarrow \int f\,dM \quad \text{for all } f \in C_c^+(\mathcal{X}).$$

Applying Theorem 2.14, we conclude that $M_n \Rightarrow M$. □

In this chapter, we developed the notion of convergence in distribution for point processes and collected several criteria that make this convergence tractable in practice. In particular, for Poisson point processes, vague convergence of the mean measures already implies convergence in distribution of the processes themselves. These results provide the probabilistic framework needed in the sequel, when we connect extreme value theory with Poisson point processes.

3 Extreme Value Theory

Extreme value theory is concerned with the stochastic behaviour of unusually large observations. In contrast to classical limit theorems, where one studies sums or averages, the focus here lies on rare extreme events occurring in the upper tail of a distribution. Typical examples include extreme river discharges, large insurance losses, or exceptional movements in financial markets. In such situations, the main interest is not in the average behaviour, but rather in the most extreme outcomes.

Two natural approaches arise in this context. On the one hand, one may study the maximum of a sample, which describes the single largest observation among a collection of random variables. This could be the highest wind speed recorded over a year or the highest insurance claim in a year. On the other hand, one may consider exceedances over a high threshold, that is, all observations that are unusually large relative to a prescribed level. These two viewpoints form the basis of extreme value theory and will later be seen to be closely connected. The former is called block maxima approach (BM), and the latter is called peaks over threshold approach (POT).

In this chapter, we introduce the basic concepts and results of univariate extreme value theory that will be needed in the sequel. We begin with the classical Block maxima approach based on maxima of independent and identically distributed random variables, then turn to threshold exceedances.

The presentation in this chapter follows standard references from the literature, in particular [Guu07], [Lau07], [Res07], and [Col01]. Here, [Guu07] is the main reference of this thesis, [Lau07] serves as a general reference for extreme value theory, [Res07] provides a mathematically rigorous treatment including the point process perspective, and [Col01] is a readable introduction with a stronger emphasis on statistical aspects.

3.1 Maxima and Minima of Independent Random Variables

Let $X_1, X_2, \dots$ be independent and identically distributed random variables, and define the sample maximum by

$$M_n = \max(X_1, \dots, X_n), \qquad n \in \mathbb{N}.$$

A central question in extreme value theory is whether there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that the normalized maxima

$$\frac{M_n - b_n}{a_n}$$

converge in distribution to a non-degenerate limit law as $n \to \infty$.

Let F denote the distribution function of the random variables $X_1, X_2, \ldots$. Then, for every $x \in \mathbb{R}$,

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x).$$

Since $X_1, \ldots, X_n$ are independent and identically distributed, it follows that

$$\mathbb{P}(M_n \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = F^n(x).$$

Thus, the distribution function of the sample maximum M_n is given by F^n.

This simple identity shows that the behaviour of maxima is determined by the upper tail of the underlying distribution F. In many cases, however, the sequence M_n itself does not converge to a meaningful non-degenerate limit. Indeed, let $x_F := \sup\{x \in \mathbb{R} : F(x) \leq 1\}$ be the upper endpoint of F. If x is smaller than x_F, then typically $F(x) < 1$, and hence $F^n(x) \to 0$ as $n \to \infty$. This suggests that, in order to obtain interesting limit laws, one has to normalize the maxima appropriately.

We are therefore led to study whether there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = \mathbb{P}(M_n \leq a_n x + b_n) = F^n(a_n x + b_n)$$

converges to a non-degenerate distribution function as $n \to \infty$.

Definition 3.1 (Extreme value distribution)**.** Let F be a distribution function. We say that G is an *extreme value distribution* associated to F if G is non-degenerate and there exist sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ with $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{3}$$

for all continuity points x of G.

In the sequel, we restrict attention to sample maxima. This causes no loss of generality, since sample minima can be reduced to maxima by considering the reflected variables $-X_1, -X_2, \ldots$. Indeed,

$$\min(X_1, \ldots, X_n) = -\max(-X_1, \ldots, -X_n),$$

so that all results for minima follow directly from the corresponding results for maxima.

Example 3.2. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with

$$X_i \sim \mathrm{Exp}(\lambda), \qquad \lambda > 0,$$

and let

$$M_n^{\min} := \min(X_1, \dots, X_n), \qquad M_n^{\max} := \max(X_1, \dots, X_n).$$

We first consider the minimum. For $x \geq 0$,

$$\mathbb{P}(M_n^{\min} > x) = \mathbb{P}(X_1 > x, \dots, X_n > x).$$

By independence, this yields

$$\mathbb{P}(M_n^{\min} > x) = \prod_{i=1}^{n} \mathbb{P}(X_i > x) = (e^{-\lambda x})^n = e^{-n\lambda x}.$$

Hence,

$$\mathbb{P}(M_n^{\min} \leq x) = 1 - e^{-n\lambda x}, \qquad x \geq 0.$$

Thus, $M_n^{\min}$ is again exponentially distributed, now with rate $n\lambda$, that is,

$$M_n^{\min} \sim \mathrm{Exp}(n\lambda).$$

In particular,

$$\mathbb{E}[M_n^{\min}] = \frac{1}{n\lambda} \to 0 \qquad \text{as } n \to \infty.$$

Next, we consider the maximum. For $x \geq 0$,

$$\mathbb{P}(M_n^{\max} \leq x) = \mathbb{P}(X_1 \leq x, \dots, X_n \leq x).$$

Again by independence,

$$\mathbb{P}(M_n^{\max} \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = \left(1 - e^{-\lambda x}\right)^n.$$

Hence, the distribution function of $M_n^{\max}$ is given by

$$F_{M_n^{\max}}(x) = \left(1 - e^{-\lambda x}\right)^n, \qquad x \geq 0.$$

This shows that the maximum is no longer exponentially distributed. Moreover, as n increases, the maximum tends to larger values, which illustrates that maxima require a suitable normalization in order to obtain non-degenerate limit laws. $\Diamond$

The preceding example illustrates the basic phenomenon of extreme value theory: minima and maxima exhibit a fundamentally different behaviour from sums or averages. In particular, the maximum typically drifts to larger values as the sample size increases, so that a non-degenerate limit law can only be expected after a suitable normalization.

3.2 Block Maxima: Extreme Value Distributions and Max-Stability

We now turn to the central question of classical extreme value theory: which non-degenerate distribution functions G can arise as limits of normalized sample block maxima. A remarkable fact is that, up to affine normalization, there are only three possible types of such limit laws. These are called the extreme value distributions.

Definition 3.3. A non-degenerate distribution function G is called *max-stable* if for every $n \in \mathbb{N}$ there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$G^n(a_n x + b_n) = G(x), \qquad x \in \mathbb{R}.$$

Max-stability means that the maximum of n independent random variables with distribution function G, after a suitable affine normalization, again has distribution function G. Thus, max-stable distributions are precisely the possible fixed points of the operation of taking maxima.

The following theorem shows that every non-degenerate limit law of normalized sample maxima must be max-stable, and that there are only three possible types of such distributions.

Theorem 3.4 (Fisher–Tippett–Gnedenko)**.** *Let F be a distribution function. Suppose that there exist sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ with $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$F^n(a_n x + b_n) \to G(x), \quad \text{as } n \to \infty$$

where G is a non-degenerate distribution function. Then G is max-stable. Moreover, there exist constants $\alpha > 0$, $\mu \in \mathbb{R}$, and $\sigma > 0$ such that G is of one of the following three types:

1. ***Fréchet type:***
$$G(x) = \begin{cases} \exp\bigl(-\bigl(\frac{x-\mu}{\sigma}\bigr)^{-\alpha}\bigr), & x > \mu, \\ 0, & x \le \mu; \end{cases}$$

2. ***Weibull type:***
$$G(x) = \begin{cases} \exp\bigl(-\bigl(-\frac{x-\mu}{\sigma}\bigr)^{\alpha}\bigr), & x < \mu, \\ 1, & x \ge \mu; \end{cases}$$

3. ***Gumbel type:***
$$G(x) = \exp\bigl(-e^{-\frac{x-\mu}{\sigma}}\bigr), \qquad x \in \mathbb{R}.$$

Conversely, each distribution function of the above form is max-stable and may occur as the limit of suitably normalized sample maxima.

Theorem 3.4 shows that every non-degenerate limit law of suitably normalized sample maxima must belong to one of three types: Fréchet, Weibull, or Gumbel. Although these classes appear different at first sight, they can be written in a unified parametric form. This representation is particularly convenient, since it allows one to treat all three cases simultaneously through a single shape parameter ξ.

Theorem 3.5. *Let F be a distribution function. Suppose that there exist sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ with $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$F^n(a_n x + b_n) \to G(x), \quad \text{as } n \to \infty$$

where G is a non-degenerate distribution function. Then G can be written in the form

$$G(x) = \exp\left(-\left[1+\xi(\tfrac{x-\mu}{\sigma})\right]^{-1/\xi}\right), \qquad 1+\xi(\tfrac{x-\mu}{\sigma}) > 0,$$

for some parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$, with the convention

$$G(x) = \exp\left(-e^{-(x-\mu)/\sigma}\right) \qquad \text{if } \xi = 0.$$

The family of distributions of this form is called the generalized extreme value *(GEV) family.*

Remark 3.6. The cases $\xi > 0$, $\xi < 0$, and $\xi = 0$ correspond to the Fréchet, Weibull, and Gumbel type, respectively.

Proof. This is a standard result in EVT. A proof can be found in [Lau07, Chapter 1.1.3, Theorem 1.1.3]. □

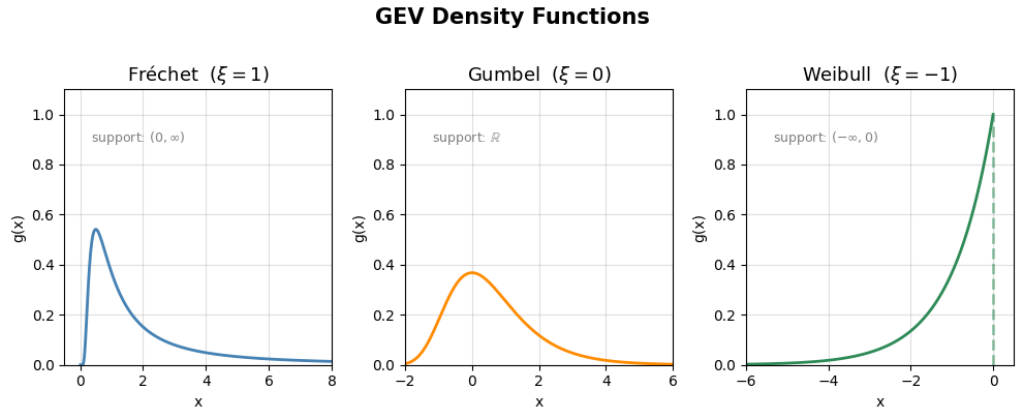

Figure 5: Density functions of the three standardised GEV distributions ($\mu = 0$, $\sigma = 1$), corresponding to the three cases of the Fisher–Tippett–Gnedenko Theorem (Theorem 3.4). Left: the Fréchet density $g(x) = x^{-2}e^{-1/x}$ on $(0, \infty)$, characterised by its heavy right tail ($\xi = 1$). Centre: the Gumbel density $g(x) = e^{-x}e^{-e^{-x}}$ on $\mathbb{R}$, with light exponential tails ($\xi = 0$). Right: the Weibull density $g(x) = e^{x}$ on $(-\infty, 0)$, with bounded support reflecting a finite upper endpoint ($\xi = -1$). The dashed vertical line at $x = 0$ marks the hard boundary of the Weibull support.

The distribution functions F satisfying (3) for a given non-degenerate distribution function G are said to belong to the *domain of attraction* of G. In this case, we write

$$F \in \mathcal{D}(G).$$

By Theorem 3.4, every non-degenerate limit distribution G belongs to one of the three extreme value types. Hence, the corresponding domains of attraction may be classified according to the shape parameter ξ of the generalized extreme value distribution. We can write $F \in \mathcal{D}(G_\xi)$ if F belongs to the domain of attraction of the extreme value distribution with shape parameter ξ.

Examples

We will now give three basic theoretical examples to illustrate how the normalizing sequences lead to a limit distribution within the GEV family. The examples can be found in [Col01, Chapter 3.1.5].

Example 3.7 (Exponential, $\xi = 0$)**.** Let $X_1, X_2, \ldots \sim \text{Exp}(1)$ i.i.d. random variables. The distribution function is given by $F(x) = 1 - e^{-x}$ for $x > 0$. Let $a_n = 1$ and $b_n = \log(n)$ for

all $n \in \mathbb{N}$. Then because of independence

$$\begin{aligned}\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) &= \mathbb{P}(X_1 \leq z + \log(n),\, X_2 \leq z + \log(n), \ldots) \\ &= F^n(z + \log(n)) \\ &= \left(1 - e^{-(z+\log(n))}\right)^n \\ &= \left(1 - n^{-1}e^{-z}\right)^n \to \exp(-e^{-z})\end{aligned}$$

as $n \to \infty$, for each fixed $z \in \mathbb{R}$. This is the Gumbel distribution corresponding to $\xi = 0$ in the GEV family.

$\Diamond$

Example 3.8 (Fréchet, $\xi = 1$)**.** Let $X_1, X_2, \ldots$ i.i.d. standard Fréchet random variables. The distribution function is given by $F(x) = e^{-1/x}$ for $x > 0$. Let $a_n = n$ and $b_n = 0$ for all $n \in \mathbb{N}$. Then

$$\begin{aligned}\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) &= F^n(nz) \\ &= \left(e^{-1/nz}\right)^n \\ &= \left(e^{-1/z}\right)\end{aligned}$$

as $n \to \infty$, for each fixed $z \in \mathbb{R}$. This is the Fréchet distribution corresponding to $\xi = 1$ in the GEV family.

$\Diamond$

Example 3.9 (Uniform, $\xi = -1$)**.** Let $X_1, X_2, \ldots$ i.i.d. standard uniform $\mathrm{U}(0, 1)$ random variables. The distribution function is given by $F(x) = x$ for $0 \leq x \leq 1$. Fix $z < 0$, suppose $n > -z$ and let $a_n = \frac{1}{n}$ and $b_n = 1$ for all $n > -z$. Then

$$\begin{aligned}\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) &= F^n\left(\frac{z}{n} + 1\right) \\ &= \left(\frac{z}{n} + 1\right)^n \to e^z\end{aligned}$$

as $n \to \infty$. This is the Weibull distribution corresponding to $\xi = -1$ in the GEV family.

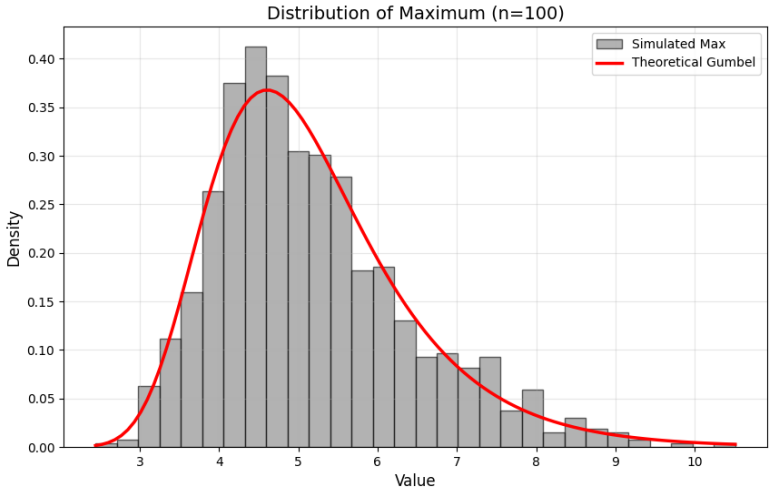

Figure 6: Distribution of the sample maximum $M_n = \max(X_1, \ldots, X_n)$ over $N = 10{,}000$ repetitions with $n = 100$ i.i.d. Exp(1) observations. The red curve is the theoretical Gumbel density with location $\mu = \log(100) \approx 4.605$ and scale $\sigma = 1$.

Remark 3.10 (Simulation of Example 3.7). Figure 6 provides a numerical illustration of Example 3.7. We draw $n = 100$ i.i.d. Exp(1) observations, record the maximum, and repeat this $N = 10{,}000$ times. By Example 3.7, the normalized maximum $(M_n - b_n)/a_n$ converges in distribution to the Gumbel law, with $b_n = \log n$ and $a_n = 1$. The theoretical overlay therefore uses location $\mu = b_n = \log(100)$ and scale $\sigma = a_n = 1$, with no free parameters to fit. The close agreement between the histogram and the theoretical curve confirms the convergence predicted by Theorem 3.4.

$\Diamond$

3.3 Peaks Over Threshold: Generalized Pareto Distributions

In addition to the block maxima approach, extreme value theory may also be developed from the perspective of exceedances over high thresholds. This is the idea underlying the peaks-over-threshold method. Instead of retaining only the single largest observation in a sample, one considers all observations that exceed a sufficiently high level and studies their behaviour relative to that threshold. This approach often makes more efficient use of the available data and leads to a complementary description of tail behaviour. A fundamental result in this direction shows that, for a large class of underlying distributions, the conditional distribution of exceedances over a high threshold converges to a generalized Pareto distribution. Moreover, this class coincides with the corresponding domain of attraction in the maxima approach, so that the two viewpoints are closely connected.

Let X be a random variable with distribution function F, and let u be a threshold such that $F(u) < 1$. The distribution of exceedances over the threshold u is described by the conditional excess distribution

$$F_u(x) := \mathbb{P}(X - u \leq x \mid X > u) = \frac{F(u+x) - F(u)}{1 - F(u)}, \qquad x \geq 0.$$

Thus, F_u gives the distribution of the overshoot above u, conditional on the event that X exceeds the threshold u. The central question in the peaks-over-threshold approach is whether these conditional excess distributions converge to a non-degenerate limit as the threshold u approaches the upper endpoint of F. The possible limit laws in this setting are the generalized Pareto distributions.

Definition 3.11. Let $\xi \in \mathbb{R}$. The *generalized Pareto distribution* with shape parameter ξ is the distribution function

$$H_\xi(x) := \begin{cases} 1 - (1+\xi x)^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-x}, & \xi = 0, \end{cases}$$

defined for all $x \geq 0$ such that $1 + \xi x > 0$. Thus, for $\xi \geq 0$ the support is $[0, \infty)$, whereas for $\xi < 0$ the support is $[0, -1/\xi]$.

Theorem 3.12 (Pickands–Balkema–de Haan)**.** *Let F be a distribution function with upper endpoint*

$$x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}.$$

Then $F \in \mathcal{D}(G_\xi)$ if and only if there exists a positive function $a(u)$ such that

$$1 - F_u(x) = \frac{1 - F(u + a(u)x)}{1 - F(u)} \to 1 - H_\xi(x) \qquad \text{as } u \uparrow x_F,$$

for every $x \geq 0$ such that $1 + \xi x > 0$, where H_ξ denotes the generalized Pareto distribution with shape parameter ξ.

The Pickands–Balkema–de Haan theorem provides the link between the block maxima and peaks-over-threshold approaches. It shows that a distribution function F belongs to the domain of attraction of a generalized extreme value distribution G_ξ if and only if the corresponding excess distributions over high thresholds converge to the generalized Pareto distribution H_ξ with the same shape parameter ξ.

Proof. Write $\bar{F} = 1 - F$ for the survival function throughout.

Direction 1: GPD convergence $\Rightarrow$ $F \in \mathcal{D}(G_\xi)$.

Assume there exists a positive function $a(u)$ such that for every x with $1+\xi x > 0$,

$$\frac{\bar{F}(u + a(u)x)}{\bar{F}(u)} \to (1+\xi x)^{-1/\xi} \qquad \text{as } u \uparrow x_F. \tag{4}$$

We construct normalizing sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that $F^n(a_n x + b_n) \to G_\xi(x)$.

Step 1: Choose the threshold sequence b_n. Since $\bar{F}(u) \to 0$ as $u \uparrow x_F$, and $\bar{F}$ is right-continuous and decreasing, we can choose a sequence $b_n \uparrow x_F$ such that

$$n\bar{F}(b_n) \to 1 \qquad \text{as } n \to \infty. \tag{5}$$

Indeed, for each n, the function $u \mapsto n\bar{F}(u)$ decreases continuously from ∞ to 0, so one can always find b_n with $n\bar{F}(b_n)$ close to 1. Moreover, $n\bar{F}(b_n) \to 1$ forces $\bar{F}(b_n) \sim 1/n \to 0$, confirming $b_n \uparrow x_F$.

Step 2: Set $a_n := a(b_n)$ and compute $n\bar{F}(b_n + a_n x)$. Applying the assumed convergence at $u = b_n$ gives

$$\frac{\bar{F}(b_n + a_n x)}{\bar{F}(b_n)} \to (1+\xi x)^{-1/\xi}.$$

Multiplying both sides by $n\bar{F}(b_n)$ and using (5),

$$n\bar{F}(b_n + a_n x) = \underbrace{n\bar{F}(b_n)}_{\to 1} \cdot \underbrace{\frac{\bar{F}(b_n + a_n x)}{\bar{F}(b_n)}}_{\to (1+\xi x)^{-1/\xi}} \to (1+\xi x)^{-1/\xi}. \tag{6}$$

Step 3: Conclude $F^n(b_n + a_n x) \to G_\xi(x)$. For any sequence $p_n \to c$, one has $\left(1 - \frac{p_n}{n}\right)^n \to e^{-c}$. Using (6) we get

$$\begin{aligned} F^n(b_n + a_n x) &= \big(1 - \bar{F}(b_n + a_n x)\big)^n \\ &= \left(1 - \frac{n\bar{F}(b_n + a_n x)}{n}\right)^n \\ &\to \exp\big(-(1+\xi x)^{-1/\xi}\big) = G_\xi(x), \end{aligned}$$

so $F \in \mathcal{D}(G_\xi)$.

Direction 2: $F \in \mathcal{D}(G_\xi) \Rightarrow$ GPD convergence.

Assume $F \in \mathcal{D}(G_\xi)$, i.e. there exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) \to G_\xi(x) = \exp\big(-(1+\xi x)^{-1/\xi}\big). \tag{7}$$

Step 1: Translate (7) *into a statement about* $\bar{F}$. For any sequence $p_n \to 0$, $(1-p_n)^n \to e^{-c}$ if and only if $np_n \to c$. Since $G_\xi(x) \in (0,1)$ and $\bar{F}(a_n x+b_n) \to 0$, applying this equivalence gives

$$n\bar{F}(a_n x + b_n) \to -\log G_\xi(x) = (1+\xi x)^{-1/\xi}. \tag{8}$$

Step 2: Introduce a threshold sequence. Fix y with $1+\xi y > 0$, and define $u_n := b_n + a_n y$. Applying (8) at $x = y$,

$$n\bar{F}(u_n) \to (1+\xi y)^{-1/\xi} > 0. \tag{9}$$

In particular, $\bar{F}(u_n) \sim (1+\xi y)^{-1/\xi}/n \to 0$, confirming $u_n \uparrow x_F$.

Step 3: Compute the tail ratio along u_n. Fix s with $1+\xi s > 0$ and consider the shifted argument

$$u_n + a_n(1+\xi y)s = b_n + a_n y + a_n(1+\xi y)s = b_n + a_n\big(y + (1+\xi y)s\big).$$

This is of the form $b_n + a_n z$ with $z := y + (1+\xi y)s$. We verify $1+\xi z > 0$:

$$1 + \xi z = 1 + \xi y + \xi(1+\xi y)s = (1+\xi y)(1+\xi s) > 0,$$

since both factors are positive by assumption. Applying (8) at $x = z$,

$$n\bar{F}\big(u_n + a_n(1+\xi y)s\big) \to (1+\xi z)^{-1/\xi} = \big((1+\xi y)(1+\xi s)\big)^{-1/\xi}. \tag{10}$$

Step 4: Divide to isolate the conditional excess tail. Dividing (10) by (9),

$$\frac{\bar{F}(u_n + a_n(1+\xi y)s)}{\bar{F}(u_n)} = \frac{n\bar{F}(u_n + a_n(1+\xi y)s)}{n\bar{F}(u_n)} \to \frac{(1+\xi y)^{-1/\xi}(1+\xi s)^{-1/\xi}}{(1+\xi y)^{-1/\xi}} = (1+\xi s)^{-1/\xi}.$$

Step 5: Identify the scaling function. Setting $a(u_n) := a_n(1+\xi y)$, the above reads

$$\frac{\bar{F}(u_n + a(u_n)s)}{\bar{F}(u_n)} \to (1+\xi s)^{-1/\xi} = 1 - H_\xi(s),$$

which is the desired GPD convergence along the sequence u_n.

Step 6: Extension to all $u \uparrow x_F$. The convergence holds along the sequence $u_n \uparrow x_F$. A standard monotonicity and interpolation argument extends this to all thresholds $u \uparrow x_F$: since $\bar{F}$ is monotone decreasing, any u between consecutive terms u_n and u_{n+1} can be sandwiched, and the convergence carries over in the limit. This yields a positive function

$a(u)$ defined for all u close to x_F such that

$$\frac{\bar{F}(u + a(u)x)}{\bar{F}(u)} \to (1 + \xi x)^{-1/\xi} = 1 - H_\xi(x) \qquad \text{as } u \uparrow x_F,$$

completing the proof. □

Examples

We reconsider the three examples from the block maxima chapter in terms of threshold exceedance models. The examples can be found in [Col01, Chapter 4.2.3].

Example 3.13 (Exponential, $\xi = 0$)**.** Let $X_1, X_2, \ldots \sim \text{Exp}(1)$ i.i.d. random variables with distribution function $F(x) = 1 - e^{-x}$ for $x > 0$. Notice that the upper endpoint is $x_F = \infty$. We verify that $F \in \mathcal{D}(H_\xi)$ with $\xi = 0$ by finding an appropriate scaling function $a(u) > 0$ such that the conditional excess distribution converges to the generalized Pareto distribution $H_0(x) = 1 - e^{-x}$.

Choice of scaling function. For large u, we compute

$$\frac{1 - F(u + a(u)x)}{1 - F(u)} = \frac{e^{-(u+a(u)x)}}{e^{-u}} = e^{-a(u)x}.$$

For this to converge to $e^{-x} = 1 - H_0(x)$, we require $a(u) = 1$.

Verification. With $a(u) = 1$, we compute the conditional excess distribution explicitly. For every $x \geq 0$,

$$\begin{aligned} 1 - F_u(x) &= \frac{1 - F(u + a(u)x)}{1 - F(u)} \\ &= \frac{e^{-(u+x)}}{e^{-u}} \\ &= e^{-x}. \end{aligned}$$

Conclusion. Since

$$\frac{1 - F(u + x)}{1 - F(u)} = e^{-x} = 1 - H_0(x) \qquad \text{for all } u > 0,$$

the exponential distribution F belongs to $\mathcal{D}(H_0)$ with scaling function $a(u) = 1$. Moreover, this is an exact result for all thresholds $u > 0$. The fact that $a(u) = 1$ remains constant reflects the memoryless property of the exponential distribution: the distribution

of exceedances above any threshold is identical to the original distribution, regardless of the threshold level. ◊

Example 3.14 (Fréchet, $\xi = 1$)**.** Let $X_1, X_2, \ldots$ i.i.d. standard Fréchet random variables with distribution function given by $F(x) = e^{-1/x}$ for $x > 0$. Notice that the upper endpoint is $x_F = \infty$. We verify that $F \in \mathcal{D}(H_\xi)$ with $\xi = 1$ by finding an appropriate scaling function $a(u) > 0$ such that the conditional excess distribution converges to the generalized Pareto distribution $H_1(x) = 1 - (1+x)^{-1}$.

Choice of scaling function. For large u, both $\frac{1}{u} \to 0$ and $\frac{1}{u+a(u)x} \to 0$, so the approximation $1 - e^{-t} \sim t$ as $t \to 0$ suggests

$$\frac{1 - F(u + a(u)x)}{1 - F(u)} \approx \frac{\frac{1}{u + a(u)x}}{\frac{1}{u}} = \frac{u}{u + a(u)x}.$$

For this to converge to $(1+x)^{-1}$, we require $a(u) = u$.

Verification. With $a(u) = u$, we compute the conditional excess distribution explicitly. For every $x \geq 0$,

$$\begin{aligned} 1 - F_u(x) &= \frac{1 - F(u + a(u)x)}{1 - F(u)} \\ &= \frac{1 - e^{-\frac{1}{u+ux}}}{1 - e^{-\frac{1}{u}}} \\ &= \frac{1 - e^{-\frac{1}{u(1+x)}}}{1 - e^{-\frac{1}{u}}}. \end{aligned}$$

As $u \to \infty$, both $\frac{1}{u(1+x)} \to 0$ and $\frac{1}{u} \to 0$. Applying the standard limit $\frac{1-e^{-t}}{t} \to 1$ as $t \to 0$ to numerator and denominator simultaneously gives

$$\frac{1 - e^{-\frac{1}{u(1+x)}}}{1 - e^{-\frac{1}{u}}} = \frac{\frac{1}{u(1+x)} \cdot \frac{1 - e^{-\frac{1}{u(1+x)}}}{\frac{1}{u(1+x)}}}{\frac{1}{u} \cdot \frac{1 - e^{-\frac{1}{u}}}{\frac{1}{u}}} \xrightarrow{u \to \infty} \frac{\frac{1}{u(1+x)} \cdot 1}{\frac{1}{u} \cdot 1} = \frac{1}{1+x} = 1 - H_1(x),$$

where the convergence of each ratio $\frac{1-e^{-t}}{t} \to 1$ as $t \to 0$ follows from L'Hôpital's rule, or equivalently from the definition of the derivative of e^{-t} at $t = 0$.

Conclusion. Since

$$\frac{1-F(u+ux)}{1-F(u)} \longrightarrow 1-H_1(x)=(1+x)^{-1} \qquad \text{as } u\uparrow\infty,$$

the Fréchet distribution F belongs to $\mathcal{D}(H_1)$ with scaling function $a(u)=u$. The fact that $a(u)=u$ grows proportionally to the threshold u reflects the heavy-tailed nature of the Fréchet distribution: exceedances above a high threshold remain of the same order of magnitude as the threshold itself.

$\Diamond$

Example 3.15 (Uniform, $\xi=-1$)**.** Let $X_1, X_2, \ldots \sim U(0,1)$ i.i.d. random variables with distribution function $F(x)=x$ for $x\in[0,1]$. Notice that the upper endpoint is $x_F=1$. We verify that $F\in\mathcal{D}(H_\xi)$ with $\xi=-1$ by finding an appropriate scaling function $a(u)>0$ such that the conditional excess distribution converges to the generalized Pareto distribution $H_{-1}(x)=1-(1+(-1)x)^{-1/(-1)}=1-(1-x)$, defined for $x\in[0,1]$.

Choice of scaling function. We compute

$$\frac{1-F(u+a(u)x)}{1-F(u)}=\frac{1-(u+a(u)x)}{1-u}=1-\frac{a(u)}{1-u}\,x.$$

For this to converge to $1-H_{-1}(x)=1-x$, we require $\dfrac{a(u)}{1-u}=1$, that is, $a(u)=1-u$.

Verification. With $a(u)=1-u$, we compute the conditional excess distribution explicitly. For every $x\in[0,1]$,

$$\begin{aligned}
1-F_u(x) &= \frac{1-F(u+a(u)x)}{1-F(u)} \\
&= \frac{1-\big(u+(1-u)x\big)}{1-u} \\
&= \frac{(1-u)(1-x)}{1-u} \\
&= 1-x.
\end{aligned}$$

Conclusion. Since

$$\frac{1-F(u+(1-u)x)}{1-F(u)}=1-x=1-H_{-1}(x) \qquad \text{for all } u\in(0,1),$$

the uniform distribution F belongs to $\mathcal{D}(H_{-1})$ with scaling function $a(u)=1-u$. As with the exponential case, this is an exact result holding for all thresholds $u\in(0,1)$, not merely

asymptotically. The fact that $a(u) = 1 - u \to 0$ as $u \uparrow 1$ reflects the bounded support of the uniform distribution: as the threshold approaches the upper endpoint, the window of possible exceedances shrinks to zero. $\Diamond$

4 Point Process Characterization of Extremes

4.1 Setup: The Normalized Sample Cloud and the State Space

In Chapter 3, we studied the asymptotic behaviour of extremes from two complementary perspectives. The block maxima approach retains only the single largest observation $M_n = \max(X_1, \ldots, X_n)$ from each sample, discarding all remaining information. The peaks-over-threshold approach is more efficient, retaining all observations that exceed a high level u, but still compresses the sample into a one-dimensional summary. Both approaches answer the question: *what does the upper tail of a large sample look like?* However, neither retains the full picture of where the large observations actually lie.

The point process approach takes a fundamentally different perspective. Instead of reducing the sample to a single statistic, we represent the entire collection of normalized observations as a random counting measure

$$N_n = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n},$$

placing a unit mass at the normalized location of every observation. This object, called the *normalized sample cloud*, is a point process in the sense of Chapter 1. The sample maximum M_n is no longer a separate object, it corresponds simply to the rightmost point of N_n:

$$\frac{M_n - b_n}{a_n} \leq x \iff N_n\big((x, \infty]\big) = 0.$$

This single observation is the key to the whole chapter. It says that the event that the normalized maximum does not exceed x is exactly the event that N_n places no points above x. As a consequence, once we understand the limiting behaviour of N_n as a point process, the limiting distribution of M_n follows immediately.

We now define the central object of this chapter precisely.

Definition 4.1 (Normalized Sample Cloud)**.** Let $X_1, X_2, \ldots$ be i.i.d. real valued random variables with distribution function F, and let $a_n > 0$ and $b_n \in \mathbb{R}$ be normalizing sequences. The *normalized sample cloud* is the point process on $\mathbb{R}$ defined by

$$N_n := \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}.$$

For a Borel set $A \subseteq \mathbb{R}$, the random variable $N_n(A)$ counts the number of normalized

observations falling in A:

$$N_n(A) = \#\left\{ i \in \{1, \dots, n\} : \frac{X_i - b_n}{a_n} \in A \right\}.$$

Interpretation. The normalized sample cloud N_n is the natural point process analogue of the normalized maximum $(M_n - b_n)/a_n$. Just as the normalized maximum asks where the single largest observation lands after centering and scaling, the sample cloud N_n records where *all* observations land. Regions of $\mathbb{R}$ where observations are dense correspond to high values of the density of F, while the upper tail will typically contain only finitely many points of N_n for large n.

Mean measure of N_n. Since the X_i are i.i.d., N_n is the superposition of n independent copies of $\delta_{(X_1 - b_n)/a_n}$. By linearity of expectation, the mean measure of N_n is

$$\mu_n(A) := \mathbb{E}[N_n(A)] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbf{1}_{\frac{X_i - b_n}{a_n} \in A} \right] = n \cdot \mathbb{P}\left(\frac{X_1 - b_n}{a_n} \in A \right) = n \cdot \mathbb{P}\big(X_1 \in a_n A + b_n\big).$$

In particular, for a half-line $(x, \infty]$,

$$\mu_n\big((x, \infty]\big) = n\bar{F}(a_n x + b_n),$$

where $\bar{F} = 1 - F$ is the survival function. We have already encountered this expression in the proof of Theorem 3.12: it is precisely the quantity that converges to $(1 + \xi x)^{-1/\xi}$ under the domain of attraction assumption.

Remark 4.2 (State space and topology). Two choices in the setup of the normalized sample cloud require justification: the state space E and the topology on it.

State space. The natural candidate for the state space of N_n is $\mathbb{R}$. However, for large n, the bulk of the observations satisfy $(X_i - b_n)/a_n \to -\infty$, since the normalizing sequences b_n and a_n are calibrated to the upper tail. These non-extreme observations collectively assign infinite mass to every neighbourhood of $-\infty$, so the limiting measure would fail to be locally finite (Radon) on $\mathbb{R}$.

The resolution is to work on the state space

$$E := (-\infty, \infty] \setminus \{-\infty\},$$

equipped with its natural topology as a locally compact space. In particular, sets of the form $[a, \infty]$ for finite $a \in \mathbb{R}$ are compact in E, and any set extending toward $-\infty$ is not, since $-\infty \notin E$. These upper tail regions $[a, \infty]$ are precisely the sets relevant to extreme value theory, as they capture the behaviour of the largest observations. By removing $-\infty$

from the state space, mass that drifts to $-\infty$ simply leaves E and is no longer tracked. The limiting point process is then Radon on E, since only finitely many observations remain in any compact region $[a, \infty]$ for large n. The point $+\infty$ is retained because the limiting process may accumulate mass there, particularly in the heavy-tailed case $\xi > 0$.

Topology. We equip $M_p(E)$ with the vague topology, as introduced in Chapter 2. Recall that vague convergence tests only against functions $f \in C_c(E)$, which have compact support in E and therefore vanish on a neighbourhood of $-\infty$. This means vague convergence automatically disregards the mass carried by the bulk observations toward $-\infty$ and probes only the behaviour of N_n on sets of the form $[a, \infty]$ for finite a — precisely the upper tail region relevant to extreme value theory.

Weak convergence, by contrast, tests against all bounded continuous functions $f \in C_b(E)$, which need not vanish near $-\infty$. Since the total mass $N_n(E) = n \to \infty$, the sequence (N_n) cannot converge weakly, and weak convergence is therefore the wrong notion here. Vague convergence is the natural and correct mode of convergence for point processes with infinite total mass, as argued in Chapter 2.

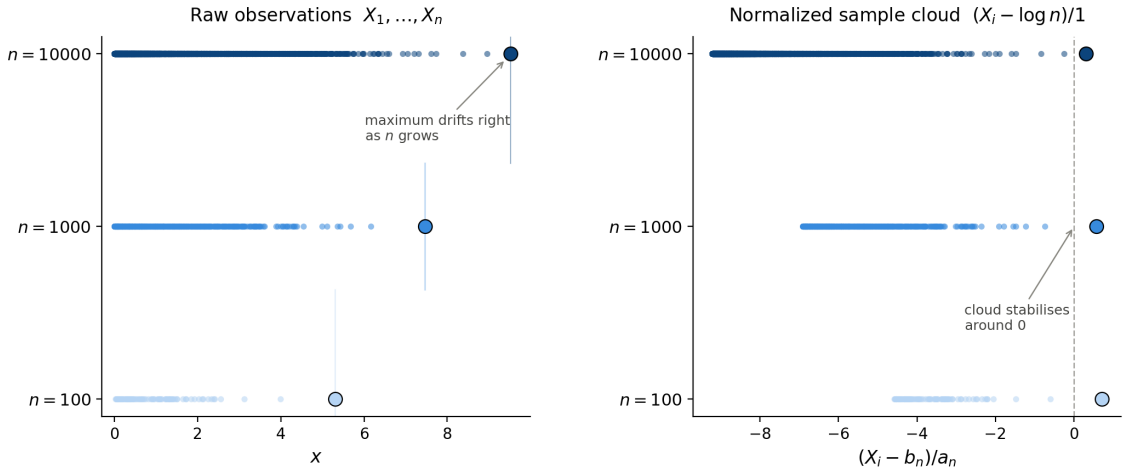

Figure 7: Left: raw Exp(1) observations for $n = 100, 1000, 10000$. The sample maximum (outlined marker) drifts to the right as n grows, illustrating that the unnormalized maximum has no non-degenerate limit. Right: the normalized sample cloud $(X_i - \log n)/1$ for the same samples. After centering by $b_n = \log n$ and scaling by $a_n = 1$, the cloud stabilises around zero and the maximum remains in a bounded region, consistent with convergence to the Gumbel distribution established in Example 3.7.

4.2 The Poisson Limit Theorem for Extremes

Theorem 4.3 (Poisson Limit Theorem for Extremes)**.** *Let $X_1, X_2, \ldots$ be i.i.d. real-valued random variables with distribution function F, and let $a_n > 0$, $b_n \in \mathbb{R}$ be normalizing*

sequences. Define the normalized sample cloud

$$N_n := \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}.$$

The following are equivalent:

1. (Convergence of maxima.) *The normalized sample maximum converges in distribution to a non-degenerate limit G:*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G, \qquad \textit{i.e.} \quad F^n(a_n x + b_n) \to G(x) \quad \textit{for all } x \in \mathbb{R}.$$

2. (Convergence of point processes.) *The normalized sample cloud converges in distribution with respect to the vague topology on $E = (-\infty, \infty] \setminus \{-\infty\}$:*

$$N_n \Rightarrow N,$$

where $N \sim \mathrm{PPP}(\mu)$ for some Radon measure μ on E satisfying $\mu(\{x\}) = 0$ for all $x \in E$.

Moreover, when either condition holds, G is necessarily a generalized extreme value distribution, and the limit measure μ and the limit distribution G are related by

$$G(x) = \exp\big(-\mu\big((x, \infty]\big)\big), \qquad x \in \mathbb{R}.$$

Proof. **Direction 1:** (2) $\Rightarrow$ (1)**.**

Assume $N_n \Rightarrow N$ where $N \sim \mathrm{PPP}(\mu)$. The key observation is that for every $x \in \mathbb{R}$,

$$\left\{ \frac{M_n - b_n}{a_n} \le x \right\} = \big\{ N_n\big((x, \infty]\big) = 0 \big\}, \tag{11}$$

since the normalized maximum does not exceed x if and only if none of the n normalized observations exceed x.

Fix x with $\mu(\{x\}) = 0$, which holds for all but countably many x since μ has no atoms. Then $(x, \infty]$ is a relatively compact open set in E with $N(\partial(x, \infty]) = N(\{x\}) = 0$ almost surely. By Theorem 2.14 statement (3),

$$N_n\big((x, \infty]\big) \Rightarrow N\big((x, \infty]\big).$$

Since $N \sim \mathrm{PPP}(\mu)$, the count $N((x,\infty])$ is Poisson with parameter $\mu((x,\infty])$, so

$$\mathbb{P}\big(N\big((x,\infty]\big)=0\big)=e^{-\mu((x,\infty])}.$$

Together with (11), this gives

$$\mathbb{P}\left(\frac{M_n-b_n}{a_n}\le x\right)=\mathbb{P}\big(N_n\big((x,\infty]\big)=0\big)\to e^{-\mu((x,\infty])}=:G(x),$$

establishing convergence of the normalized maxima.

Direction 2: (1) $\Rightarrow$ (2)**.**

Assume $F^n(a_nx+b_n)\to G(x)$ for all $x\in\mathbb{R}$.

Step 1: Mean measure of N_n. Since the X_i are i.i.d., by linearity of expectation the mean measure of N_n is

$$\mu_n(A)=\mathbb{E}[N_n(A)]=n\cdot\mathbb{P}\left(\frac{X_1-b_n}{a_n}\in A\right).$$

In particular, for a half-line $(x,\infty]$,

$$\mu_n\big((x,\infty]\big)=n\bar{F}(a_nx+b_n),$$

where $\bar{F}=1-F$.

Step 2: Vague convergence of mean measures. Define μ on E by $\mu((x,\infty]):=-\log G(x)$. We claim $\mu_n\xrightarrow{v}\mu$.

Let $p_n:=\bar{F}(a_nx+b_n)$. By assumption, $(1-p_n)^n\to G(x)\in(0,1)$, so $p_n\to 0$. Taking logarithms and using $\log(1-p_n)\sim -p_n$ as $p_n\to 0$,

$$np_n\sim -n\log(1-p_n)\to -\log G(x),$$

that is,

$$\mu_n\big((x,\infty]\big)=n\bar{F}(a_nx+b_n)\to -\log G(x)=\mu\big((x,\infty]\big).$$

This is pointwise convergence on half-lines $(x,\infty]$ with $\mu(\{x\})=0$. By Proposition 2.10 (equivalence (3) of vague convergence), we conclude $\mu_n\xrightarrow{v}\mu$.

Step 3: Convergence of Laplace functionals. We show directly that the Laplace functional of N_n converges to that of a $\mathrm{PPP}(\mu)$. Let $f\in C_c^+(E)$. By independence of the X_i,

$$\mathcal{L}_{N_n}(f)=\mathbb{E}\Big[e^{-\int f\,dN_n}\Big]=\Big(\mathbb{E}\Big[e^{-f((X_1-b_n)/a_n)}\Big]\Big)^n=\big(1-q_n\big)^n,$$

where $q_n := \int (1 - e^{-f})\, d\nu_n$ and $\nu_n(\cdot) = \mathbb{P}((X_1 - b_n)/a_n \in \cdot)$ is the one-point distribution. Since $\mu_n = n\nu_n$ and $\mu_n \xrightarrow{v} \mu$, and since $1 - e^{-f} \in C_c^+(E)$ whenever $f \in C_c^+(E)$,

$$nq_n = \int (1 - e^{-f})\, d\mu_n \to \int (1 - e^{-f})\, d\mu =: c.$$

Since $q_n \to 0$ and $nq_n \to c$, the standard limit $(1 - q_n)^n \to e^{-c}$ gives

$$\mathcal{L}_{N_n}(f) \to \exp\left(- \int (1 - e^{-f})\, d\mu \right),$$

which is exactly the Laplace functional of $\mathrm{PPP}(\mu)$ by Theorem 1.23. Since this holds for all $f \in C_c^+(E)$, Theorem 2.14 yields $N_n \Rightarrow N \sim \mathrm{PPP}(\mu)$.

Conclusion. When either condition holds, $G(x) = \exp(-\mu((x, \infty]))$ is a generalized extreme value distribution by Theorem 3.4, completing the proof. □

4.3 The Exponent Measure and its Classification

The mean measure μ of the limiting Poisson point process N in Theorem 4.3 is called the *exponent measure* of the limit distribution G. Since G must be a generalized extreme value distribution by Theorem 3.4, the exponent measure is not arbitrary. It must belong to one of three explicit families corresponding to the three GEV types.

To see this, recall that μ is determined by G via

$$\mu\big((x, \infty]\big) = -\log G(x), \qquad x \in E.$$

Differentiating with respect to x and using the three standardized forms of the GEV distribution from Section 3.2, We can compute the density of μ in each case.

Corollary 4.4 (Classification of Exponent Measures)**.** *Under the assumptions of Theorem 4.3, the exponent measure μ of the limiting Poisson point process has a density belonging to one of the following three families:*

$$\frac{d\mu}{dx} = \begin{cases} \dfrac{s}{x^{s+1}}, & x \in [1, \infty), \quad s = \dfrac{1}{\xi} > 0 \quad \textit{(Fréchet, } \xi > 0\textit{)}, \\ e^{-x}, & x \in \mathbb{R} \quad \textit{(Gumbel, } \xi = 0\textit{)}, \\ s|x|^{s-1}, & x \in [-1, 0), \quad s = -\dfrac{1}{\xi} > 0 \quad \textit{(Weibull, } \xi < 0\textit{)}. \end{cases} \tag{12}$$

Conversely, each of these families defines a valid exponent measure, and the corresponding limit distribution is a GEV distribution with shape parameter ξ. Moreover, apart from

a change of coordinates and a positive weight factor, (12) *contains all possible exponent measures.*

Proof. The Fréchet case follows by taking $G(x) = \exp(-x^{-s})$ for $x > 0$, so that $\mu((x,\infty]) = x^{-s}$, and differentiating. The Gumbel and Weibull cases follow analogously from the corresponding standardized forms. The classification result is [Guu07, Theorem 6.2]. □

Interpretation. The exponent measure μ is an infinite Radon measure on E, and its density $d\mu/dx$ is obtained by differentiating the survival function $\mu((x,\infty]) = -\log G(x)$. The parameter $s > 0$ is a reparametrization of the shape parameter: $s = 1/\xi$ in the Fréchet case and $s = -1/\xi$ in the Weibull case. In the Fréchet case the density s/x^{s+1} is a power law decaying slowly toward $+\infty$, encoding the heavy tail. In the Gumbel case the density e^{-x} decays exponentially. In the Weibull case the density $s|x|^{s-1}$ is supported on a bounded interval $[-1, 0)$, encoding the finite upper endpoint. Once μ is known, both the GEV distribution and the GPD are recovered via

$$G(x) = e^{-\mu((x,\infty])}, \qquad 1 - H_\xi(x) = \frac{\mu\big((u + a(u)x, \infty]\big)}{\mu\big((u,\infty]\big)},$$

so the exponent measure is the single object from which the entire extreme value theory of F can be read off.

4.4 Connection to the Peaks-over-Threshold Approach

The Poisson limit theorem established in Section 4.2 does more than recover the block maxima results of Chapter 3. It also provides a conceptual explanation for the peaks-over-threshold approach and the Pickands–Balkema–de Haan theorem. We now make this connection precise.

Recall the setting: $N_n \Rightarrow N$ where $N \sim \mathrm{PPP}(\mu)$ and the exponent measure μ satisfies $\mu((x,\infty]) = -\log G(x)$. The limiting point process N records the positions of all extreme observations in the limit. A natural question is: what does N look like above a high threshold u? The answer is given directly by the restriction Theorem (Theorem 1.15) from Chapter 1.

Proposition 4.5 (Threshold exceedances of the limiting PPP)**.** *Let $N \sim \mathrm{PPP}(\mu)$ be the limiting point process from Theorem 4.3, and let $u \in E$ be a threshold with $\mu(\{u\}) = 0$ and $\mu((u,\infty]) \in (0,\infty)$. Then:*

1. *The restricted process*

$$N^u(\,\cdot\,) := N(\,\cdot \cap (u,\infty])$$

is a $\mathrm{PPP}(\mu^u)$ *on* $(u, \infty]$*, where* $\mu^u(A) := \mu(A \cap (u, \infty])$*.*

2. *The normalized restricted measure*

$$\frac{\mu^u(\,\cdot\,)}{\mu((u,\infty])}$$

is a generalized Pareto distribution H_ξ *with the same shape parameter* ξ *as* G*.*

Proof. (1) follows immediately from the restriction Theorem (Theorem 1.15): restricting a $\mathrm{PPP}(\mu)$ to any measurable set C yields a $\mathrm{PPP}(\mu_C)$ where $\mu_C(\cdot) = \mu(\cdot \cap C)$.

(2) We verify this for each of the three GEV types using the exponent measure densities from Corollary 4.4. We work out the Gumbel case in detail; the Fréchet and Weibull cases are analogous.

For the Gumbel case ($\xi = 0$), the exponent measure has density e^{-x} on $[0, \infty)$. For a threshold $u \geq 0$ and $x > 0$,

$$\frac{\mu^u((u+x,\infty])}{\mu((u,\infty])} = \frac{\int_{u+x}^{\infty} e^{-t}\,dt}{\int_u^{\infty} e^{-t}\,dt} = \frac{e^{-(u+x)}}{e^{-u}} = e^{-x} = 1 - H_0(x),$$

which is the survival function of the generalized Pareto distribution with $\xi = 0$, as required. □

Unification of BM and POT. We can now see clearly why the block maxima and peaks-over-threshold approaches yield the same shape parameter ξ. Both are projections of the single limiting object $N \sim \mathrm{PPP}(\mu)$:

- The *block maxima approach* projects N onto its rightmost point:

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} \sup\{x : N(\{x\}) > 0\} \sim G_\xi.$$

- The *peaks-over-threshold approach* projects N onto the distribution of its points above a threshold u, normalized by $a(u)$:

$$\frac{X - u}{a(u)} \,\Big|\, X > u \xrightarrow{d} H_\xi.$$

Both G_ξ and H_ξ carry the same shape parameter ξ because they arise from the same exponent measure μ: one as its distribution function via $G(x) = e^{-\mu((x,\infty])}$, and the other as its normalized restriction above a threshold. The PPP framework can therefore be viewed as the natural probabilistic structure from which both classical EVT results emerge.

4.5 Worked Examples

We now complete the trilogy of examples begun in Example 3.7. The following examples illustrate Theorem 4.3 and Proposition 4.5 concretely by working through all three canonical distributions from Chapters 3. In each case we compute the normalized sample cloud, identify the limiting PPP, and recover both the block maxima and peaks-over-threshold results as projections of a single limiting object. Together, the three examples cover all three GEV types and all three families of exponent measures from Corollary 4.4.

Example 4.6 (Exponential, $\xi = 0$)**.** Let $X_1, X_2, \ldots \sim \text{Exp}(1)$ i.i.d. with distribution function $F(x) = 1 - e^{-x}$ for $x > 0$. Set $a_n = 1$ and $b_n = \log n$ as in Example 3.7.

The normalized sample cloud. The normalized sample cloud is

$$N_n = \sum_{i=1}^{n} \delta_{X_i - \log n}.$$

For a half-line $(x, \infty]$, the mean measure of N_n is

$$\mu_n\big((x,\infty]\big) = n\bar{F}(\log n + x) = n \cdot e^{-(\log n + x)} = n \cdot \frac{e^{-x}}{n} = e^{-x}.$$

The mean measure is already independent of n: $\mu_n = \mu$ for every n, where μ has density e^{-x} on $\mathbb{R}$. Vague convergence $\mu_n \xrightarrow{v} \mu$ is therefore immediate, and this is consistent with the memoryless property observed in Example 3.13: threshold exceedances are exact for all n, not merely asymptotic.

The limiting PPP. By Theorem 4.3, $N_n \Rightarrow N$ where $N \sim \text{PPP}(\mu)$ with exponent measure μ having density e^{-x}. This is the Gumbel exponent measure from Corollary 4.4 with $\xi = 0$. For any interval $(a, b]$,

$$N\big((a,b]\big) \sim \text{Poi}\Big(e^{-a} - e^{-b}\Big),$$

and the expected number of points above level x is e^{-x}, decaying exponentially — reflecting the light-tailed nature of the exponential distribution.

Recovery of the block maxima result. By the key identity (11),

$$\begin{aligned}
\mathbb{P}\left(\frac{M_n - \log n}{1} \le x\right) &= \mathbb{P}\big(N_n\big((x,\infty]\big) = 0\big) \\
&\to \mathbb{P}\big(N\big((x,\infty]\big) = 0\big) \\
&= e^{-\mu((x,\infty])} \\
&= e^{-e^{-x}} = G_0(x),
\end{aligned}$$

recovering the Gumbel limit of Example 3.7.

Recovery of the POT result. Fix a threshold $u \in \mathbb{R}$. By Proposition 4.5,

$$\frac{\mu^u\big((u+x,\infty]\big)}{\mu\big((u,\infty]\big)} = \frac{e^{-(u+x)}}{e^{-u}} = e^{-x} = 1 - H_0(x),$$

recovering the GPD result of Example 3.13 with $\xi = 0$ and $a(u) = 1$. ◊

Example 4.7 (Fréchet, $\xi = 1$)**.** Let $X_1, X_2, \ldots$ be i.i.d. standard Fréchet random variables with distribution function $F(x) = e^{-1/x}$ for $x > 0$. Set $a_n = n$ and $b_n = 0$ as in Example 3.8.

The normalized sample cloud. The normalized sample cloud is

$$N_n = \sum_{i=1}^{n} \delta_{X_i/n}.$$

For a half-line $(x,\infty]$ with $x > 0$, the mean measure of N_n is

$$\mu_n\big((x,\infty]\big) = n\bar{F}(nx) = n\left(1 - e^{-1/(nx)}\right).$$

As $n \to \infty$, using $1 - e^{-t} \sim t$ for small t,

$$\mu_n\big((x,\infty]\big) = n\left(1 - e^{-1/(nx)}\right) \sim n \cdot \frac{1}{nx} = \frac{1}{x}.$$

Hence $\mu_n \xrightarrow{v} \mu$ where μ is defined by $\mu((x,\infty]) = 1/x$ for $x > 0$.

The limiting PPP. By Theorem 4.3, $N_n \Rightarrow N$ where $N \sim \mathrm{PPP}(\mu)$ with exponent measure μ having density $1/x^2$ on $(0,\infty)$. This is the Fréchet exponent measure from Corollary 4.4 with $\xi = 1$ (i.e. $s = 1$). The power-law density $1/x^2$ reflects the heavy-tailed nature of the Fréchet distribution: the limiting PPP places significantly more mass near large values than in the Gumbel case.

Recovery of the block maxima result. By (11),

$$\mathbb{P}\left(\frac{M_n}{n} \le x\right) = \mathbb{P}\big(N_n\big((x,\infty]\big) = 0\big) \to \mathbb{P}\big(N\big((x,\infty]\big) = 0\big) = e^{-\mu((x,\infty])} = e^{-1/x} = G_1(x),$$

recovering the Fréchet limit of Example 3.8.

Recovery of the POT result. Fix a threshold $u > 0$. By Proposition 4.5,

$$\frac{\mu^u\big((u + a(u)x,\infty]\big)}{\mu\big((u,\infty]\big)} = \frac{1/(u+ux)}{1/u} = \frac{1}{1+x} = 1 - H_1(x),$$

where we set $a(u) = u$, recovering the GPD result of Example 3.14 with $\xi = 1$. The scaling function $a(u) = u$ growing proportionally with the threshold reflects the heavy-tailed nature of the Fréchet distribution: exceedances above a high threshold remain of the same order of magnitude as the threshold itself. ◊

Example 4.8 (Uniform, $\xi = -1$)**.** Let $X_1, X_2, \ldots \sim U(0,1)$ i.i.d. with distribution function $F(x) = x$ for $x \in [0,1]$. Set $a_n = 1/n$ and $b_n = 1$ as in Example 3.9.

The normalized sample cloud. The normalized sample cloud is

$$N_n = \sum_{i=1}^{n} \delta_{n(X_i - 1)}.$$

The normalized observations $n(X_i - 1)$ take values in $[-n, 0]$, so for large n the cloud spreads over $(-\infty, 0]$. For a half-line $(x, 0]$ with $x \in (-n, 0]$, the mean measure of N_n is

$$\mu_n\big((x,\infty]\big) = n\bar{F}\Big(\frac{x}{n} + 1\Big) = n\left(1 - \frac{x}{n} - 1\right) = n \cdot \left(-\frac{x}{n}\right) = -x.$$

Hence $\mu_n\big((x,\infty]\big) = -x$ for every n, so as in the exponential case the mean measure is already constant in n.

The limiting PPP. By Theorem 4.3, $N_n \Rightarrow N$ where $N \sim \mathrm{PPP}(\mu)$ with exponent measure μ defined by $\mu((x,\infty]) = -x$ for $x \in (-\infty, 0]$. Differentiating, μ has density 1 on $(-\infty, 0]$, or equivalently density $s|x|^{s-1}$ with $s = 1 = -1/\xi$, which is the Weibull exponent measure from Corollary 4.4 with $\xi = -1$. The support of μ is bounded above by 0, reflecting the finite upper endpoint $x_F = 1$ of the uniform distribution.

Recovery of the block maxima result. By (11), for $x < 0$,

$$\mathbb{P}(n(M_n - 1) \le x) = \mathbb{P}\big(N_n\big((x,\infty]\big) = 0\big) \to \mathbb{P}\big(N\big((x,\infty]\big) = 0\big) = e^{-\mu((x,\infty])} = e^x = G_{-1}(x),$$

recovering the Weibull limit of Example 3.9.

Recovery of the POT result. Fix a threshold $u \in (0,1)$, which corresponds to $v = n(u-1) \in (-n, 0)$ in normalized coordinates. By Proposition 4.5,

$$\frac{\mu^v\big((v + a(v)x, \infty]\big)}{\mu\big((v,\infty]\big)} = \frac{-(v + (-v)x)}{-v} = \frac{-v(1-x)}{-v} = 1 - x = 1 - H_{-1}(x),$$

where we set $a(v) = -v = 1 - u > 0$, recovering the GPD result of Example 3.15 with $\xi = -1$ and $a(u) = 1 - u$. As with the exponential case, this is an exact result holding for all thresholds $u \in (0,1)$. The scaling function $a(u) = 1 - u \to 0$ as $u \uparrow 1$ reflects the bounded support: as the threshold approaches the upper endpoint, the window of possible

exceedances shrinks to zero. $\lozenge$

The three examples together confirm the classification of Corollary 4.4 and illustrate how the PPP framework unifies the results of Chapter 3. We summarize the key quantities in Table 2.

Distribution	ξ	a_n	b_n	**Exponent density**	**GEV**	**POT**
Exp(1)	0	1	$\log n$	e^{-x} on $\mathbb{R}$	G_0	H_0, $a(u) = 1$
Fréchet	1	n	0	$1/x^2$ on $(0, \infty)$	G_1	H_1, $a(u) = u$
$U(0, 1)$	-1	$1/n$	1	1 on $(-\infty, 0]$	G_{-1}	H_{-1}, $a(u) = 1 - u$

Table 2: Summary of the three canonical examples. In each case the same shape parameter ξ governs the block maxima limit G_ξ, the peaks-over-threshold limit H_ξ, and the exponent measure of the limiting PPP(μ).

Outlook

Writing this thesis has been a genuinely enjoyable experience. The subject of point processes and extreme value theory is one I taught myself from scratch, assembling material from different sources, finding my own references, and trying to understand not just what is true but why each piece fits together the way it does. In writing it down, I set myself a specific goal: to produce a text that would have taught my past self exactly what I now know, and that would have spared me the confusion I encountered along the way. Although in mathematics, it is often the confusion itself that teaches most effectively. The first two directions below are ones I explored while working on this thesis and found compelling enough to want to share, but chose not to include in order to keep the scope manageable. The third is where I intend to turn next.

Multivariate extreme value theory

The most immediate continuation of this thesis is the extension to the multivariate setting, where observations take values in $\mathbb{R}^d$ for $d \geq 2$. The point process framework extends naturally: one considers the normalized sample cloud

$$N_n = \sum_{i=1}^{n} \delta_{(X_i - b_n)/a_n}$$

where $X_i \in \mathbb{R}^d$, and studies its limit as a Poisson point process on $(0, \infty]^d \setminus \{0\}$. The exponent measure μ now lives on a higher-dimensional space and encodes the dependence structure between the components, giving rise to the theory of multivariate max-stable distributions and multivariate generalised Pareto distributions. This is the natural continuation of the univariate theory developed here, and the reader is referred to Balkema–Embrechts [Guu07] for a thorough treatment. Including this extension in the present thesis would have made it considerably longer, and I have chosen instead to present the univariate theory in full detail.

The geometric approach

A second and very compelling direction, is the geometric approach to converging sample clouds. Rather than projecting the sample cloud onto coordinate axes and studying its extremes component by component, this approach asks a more fundamental question: what does the sample cloud itself look like geometrically as $n \to \infty$? The answer is studied through the convergence of the convex hull of the cloud, which is the smallest convex set containing all observations. The limit theory that emerges is different from the classical

univariate picture. We regard this as an interesting addition to the theory developed in this thesis, and again refer the reader to [Guu07] for details.

Statistical estimation

A third direction, and the one I intend to pursue next, is the statistical estimation of the parameters appearing in the theory. The present thesis is entirely theoretical: we have established that the normalized sample maximum converges to a GEV distribution G_ξ with shape parameter ξ, location parameter μ, and scale parameter σ, and that threshold exceedances converge to a GPD H_ξ. However, in practice, these parameters must be estimated from data and the choice of normalizing sequences a_n and b_n must be made concrete. The standard tools for this are maximum likelihood estimation for the GEV and GPD families, as well as dedicated tail index estimators such as the Hill estimator. The practical aspect of extreme value theory, that of fitting models to real data, assessing goodness of fit, and quantifying uncertainty in tail estimates, is developed in Coles [Col01], which serves as a natural starting point for the reader interested in applications. A more mathematically rigorous treatment of estimation in the context of regular variation can be found in de Haan–Ferreira [Lau07], and a comprehensive account of both theory and statistics of extremes is given in Embrechts, Klüppelberg & Mikosch [EKM97].

List of Figures

1 Overview of the spaces when working with point processes. 6

2 Comparison of two simulations of a one-dimensional Poisson point process on $[0, 10]$ with rate $\lambda = 2$: exponential waiting times (top) and direct Poisson point process construction (bottom). The observed number of points being 23 in both cases is purely coincidental. 11

3 Two realizations of a homogeneous Poisson point process with intensity $\lambda = 100$ and intensity measure $\mu = \lambda \,\mathrm{Leb}$. Left: simulation on $[0, 1]^2 \subset \mathbb{R}^2$. Right: simulation on $[0, 1]^3 \subset \mathbb{R}^3$. As in the example before, in both cases the observed number of points is $Z = 96$, which is coincidental. 12

4 Simulation of an inhomogeneous Poisson point process on $[0, 1]^2$ with intensity density $f(x, y) = 1-(x^2+y^2)$ and $\lambda = 1000$. The point density decreases as (x, y) moves away from the origin, reflecting the spatially varying intensity. 13

5 Density functions of the three standardised GEV distributions ($\mu = 0$, $\sigma = 1$), corresponding to the three cases of the Fisher–Tippett–Gnedenko Theorem (Theorem 3.4). Left: the Fréchet density $g(x) = x^{-2}e^{-1/x}$ on $(0, \infty)$, characterised by its heavy right tail ($\xi = 1$). Centre: the Gumbel density $g(x) = e^{-x}e^{-e^{-x}}$ on $\mathbb{R}$, with light exponential tails ($\xi = 0$). Right: the Weibull density $g(x) = e^x$ on $(-\infty, 0)$, with bounded support reflecting a finite upper endpoint ($\xi = -1$). The dashed vertical line at $x = 0$ marks the hard boundary of the Weibull support. 42

6 Distribution of the sample maximum $M_n = \max(X_1, \ldots, X_n)$ over $N = 10{,}000$ repetitions with $n = 100$ i.i.d. Exp(1) observations. The red curve is the theoretical Gumbel density with location $\mu = \log(100) \approx 4.605$ and scale $\sigma = 1$. 44

7 Left: raw Exp(1) observations for $n = 100, 1000, 10000$. The sample maximum (outlined marker) drifts to the right as n grows, illustrating that the unnormalized maximum has no non-degenerate limit. Right: the normalized sample cloud $(X_i - \log n)/1$ for the same samples. After centering by $b_n = \log n$ and scaling by $a_n = 1$, the cloud stabilises around zero and the maximum remains in a bounded region, consistent with convergence to the Gumbel distribution established in Example 3.7. 55

References

[Bil99] Patrick Billingsley. *Convergence of Probability Measures*. Chicago: John Wiley Sons, 1999. ISBN: 978-0-471-19745-4.

[Col01] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer, 2001. ISBN: 978-1-85233-459-8.

[EKM97] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events*. Berlin: Springer, 1997. ISBN: 978-3-540-60931-5.

[Guu07] Paul Embrechts Guus Balkema. *High Risk Scenarios and Extremes: A Geometric Approach*. Zürich: European Mathematical Society, 2007. ISBN: 978-3-03719-035-7.

[Lau07] Ana Ferreira Laurens de Haan. *Extreme Value Theory: An Introduction*. Springer, 2007. ISBN: 978-0-387-23946-0.

[Res07] Sidney I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Ithaca, NY: Springer, 2007. ISBN: 978-0-387-75952-4.

[Tas25] Vincent Tassion. *Applied Stochastic Processes, lecture notes*. Spring Semester 2025, `https://metaphor.ethz.ch/x/2025/fs/401-3602-00L/`. Zürich, 2025.

Declaration on the use of generative AI. I have used Claude Sonnet 4.6 (Anthropic, 2026) as an auxiliary tool in the writing process for improving the clarity and readability of the text. All content, arguments, and conclusions are entirely my own. I retain full responsibility for the accuracy and integrity of this thesis.