Preface

Extreme value theory developed from an interest in studying the behavior of the maximum or minimum (extremes) of i.i.d. random variables. Historically, the study of extremes can be dated back to Nicholas Bernoulli who studied the mean largest distance from the origin to \( n \) points scattered randomly on a straight line of some fixed length (Gumbel, 1958). Now extreme value theory has found applications in finance, risk management, telecommunication, environmental and pollution studies and many more fields. In this course we study the probabilistic approach to extreme value theory with emphasis on Multivariate Extreme Value Theory (often abbreviated as MEVT). The course is divided into the following broad parts:

1. A brief introduction to univariate extreme value theory along with the theory of regular variation of functions.
2. Multivariate extreme value theory with regular variation on cones.
3. Point Process techniques with applications to extreme value theory.

The course will cover most of its material from the books by Embrechts et al. (1997), de Haan and Ferreira (2006), Resnick (2007, 2008). I have also borrowed extensively from lecture notes of Sidney Resnick and Parthanil Roy. Further references will be provided as and when required.

Notations and concepts

Let us fix a probability space once and for all: \((\Omega, \mathcal{F}, P)\). All the random variables (elements) mentioned in this course will be defined on this space. In the univariate case we have a random variable \( X : \Omega \to \mathbb{R} \) to be a measurable map. The probability measure of \( X \) is defined as \( P_X^{-1} \). Suppose \( X_n \) and \( X \) are random variables with distribution functions \( F_n \) and \( F \) respectively (for all \( n \geq 1 \)).

- \( X_n \xrightarrow{P} X \) denotes \( X_n \) converges to \( X \) in probability as \( n \to \infty \).
- \( X_n \xrightarrow{w.p.1} X \) or \( X_n \to X \) a.s. denotes \( X_n \) converges to \( X \) with probability 1 or almost surely as \( n \to \infty \).
- \( F_n \Rightarrow F \) denotes \( F_n \) converges in distribution to \( F \) as \( n \to \infty \). We often write \( X_n \Rightarrow X \) or \( X_n \Rightarrow F \).

We also have the following notations:

- \( f(x) \sim g(x) \) means \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = 1 \).
- \( f(x) = o(g(x)) \) means \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = 0 \).

More to be added as we proceed.
# Contents

## Preface

1 Univariate Extreme Value Theory
- 1.1 Introduction ................................................. 1
  - 1.1.1 Domain of Attraction condition ...................... 6
  - 1.1.2 Uniform convergence .................................. 8
- 1.2 Regular Variation .......................................... 8
  - 1.2.1 Properties of regularly varying functions ......... 9
- 1.3 A condition for belonging to an extreme value domain .. 11
- 1.4 Applications ............................................. 12

2 Multivariate Extreme Value Theory ........................................ 13
- 2.1 Introduction ................................................ 13
- 2.2 Limit Distributions ......................................... 14
  - 2.2.1 Max-Infinitely Divisible Distributions ............... 14
  - 2.2.2 Max-stability ............................................ 16
  - 2.2.3 Some properties of Multivariate EVDs ................ 17
- 2.3 Standardization ............................................ 19
- 2.4 Exponent measure .......................................... 21
  - 2.4.1 One point uncompactification ......................... 22
  - 2.4.2 Some properties of \( \nu_0 \) .......................... 23
- 2.5 Spectral measure ........................................... 24
- 2.6 Asymptotic Independence ................................... 28
- 2.7 Vague convergence .......................................... 30
  - 2.7.1 Vague convergence in Extreme Value Theory ......... 33

3 Point processes .................................................. 35
- 3.1 Basics of point processes .................................. 35
- 3.2 Laplace Functionals ........................................ 38
- 3.3 Poisson Process or Poisson Random Measure ............... 39
  - 3.3.1 Transformation of a Poisson Process ................. 42
  - 3.3.2 Augmentation of a Poisson Process .................... 43
- 3.4 Weak Convergence of Point Processes and Random Measures 44
  - 3.4.1 Applications to Extreme Value Theory ............... 48
Chapter 1

Univariate Extreme Value Theory

1.1 Introduction

We have a fixed probability space $(Ω, F, P)$ where all our random variables are defined. Let us assume that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables from some non-degenerate distribution $F$ on $\mathbb{R}$. ($X$ is degenerate if $P(X = c) = 1$ for some constant $c \in \mathbb{R}$.) We are interested in the tail behavior of this data, for example, in a risk management context we are interested in the probability of an (extreme) risk set, or estimating an extreme (high or low) quantile. One way to approach this problem is to provide an empirical estimate of the probability. Quite often than not, it is found that we get a probability which is zero. This is undesirable, or more precisely uninformative for us. Let us explore a little bit. Denote by $M_n = \max\{X_1, X_2, \ldots, X_n\}$, $n \geq 1$.

Also denote the right end-point of $F$ by $x_F := \sup\{x : F(x) < 1\}$.

It is easy to check that $M_n \uparrow x_F$ in probability because,

$x < x_F : P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x)$

$= F^n(x) \to 0, \text{ as } n \to \infty.$

$x \geq x_F : P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x)$

$= F^n(x) \to 1, \text{ as } n \to \infty. \quad (if \ x_F < \infty)$

Also note that $M_n$ is increasing, hence $M_n \uparrow x_F$ a.s. This result is quite uninformative for our purposes and does not answer the basic questions in our mind. An analogous example would be the difference between the SLLN and the CLT. One provides almost sure convergence, the other distributional convergence, which can be used to calculate probabilities and quantiles.

Recall for $X_n \sim F_n, Y \sim F$, we have $F_n$ converging to $F$ in distribution or weakly if $F_n(x) \to F(x)$ for $x \in C(F)$, where for any non-decreasing function $h$,

$C(h) := \{x : h \text{ is continuous at } x\}.$

We write $F_n \Rightarrow F$ or $X_n \Rightarrow Y$ or $X_n \Rightarrow F$.

Remark 1.1.1. Let $X_1, X_2, \ldots$ be i.i.d, $F$, non-degenerate, with $E(X_1) = \mu$. Define $S_n = X_1 + X_2 + \ldots + X_n$.

First note that if we only assume $E|X_1| < \infty$ we have from SLLN $n^{-1}S_n \to \mu$ almost surely. With the additional assumption of $Var(X_1) = \sigma^2 < \infty$ we get the Central Limit Theorem: $S_n - n\mu$ $\sqrt{n\sigma}$ $\Rightarrow Z, \quad Z \sim N(0,1).$
Hence for large $n$, we can approximate $P(S_n \leq x) \approx P(Z \leq \frac{x-b_n}{a_n \sigma})$.

More generally suppose $X_1, X_2, \ldots$ be i.i.d. $F$, non-degenerate, and $S_n = X_1 + X_2 + \ldots + X_n$. We do not put any moment condition here.

- If there exists non-degenerate random variable $W$ with distribution $G$ and $\exists a_n > 0, b_n \in \mathbb{R}$ such that
  \[ \frac{S_n - b_n}{a_n} \Rightarrow W, \quad (1.1.1) \]
  then $F$ belongs to the (sum) domain of attraction of $G$, denoted $F \in D(G)$, or by abuse of notation $X_1 \in D(G)$.

- If $\exists a_n > 0, b_n \in \mathbb{R}$ such that $S_n = a_n X_1 + b_n$, then $F$ is a (sum) stable distribution.

- The class of all distributions $G$ appearing as a limit in (1.1.1) coincides with the class of (sum) stable distributions.

So following a similar logic as the previous remark where we dealt with sums of i.i.d. random variables we proceed to make more robust assumptions on our distribution function $F$ to deal with the maximum of i.i.d. random variables. We assume that $F$ belongs to the maximum domain of attraction of an extreme value distribution.

**Definition 1.1.1.** A univariate distribution function $F$, belongs to the maximum domain of attraction of a distribution function $G$ if

1. $G$ is a non-degenerate distribution.
2. There exist real-valued sequences $a_n > 0, b_n \in \mathbb{R}$, such that
   \[ P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F_n(a_n x + b_n) \Rightarrow G(x). \quad (1.1.2) \]

Hence for large $n$, we can approximate $P(M_n \leq x) \approx G\left(\frac{x-b_n}{a_n}\right)$. We denote $F \in D(G)$. We often ignore the term 'maximum' and abbreviate domain of attraction as DOA.

Now we are faced with certain questions.

1. Given any $F$, does there exist $G$ such that $F \in D(G)$ ?
2. Given any $F$, if $G$ exists, is it unique ?
3. Can we characterize the class of all possible limits $G$ according to Definition 1.1.1 ?
4. Given a limit $G$, what properties should $F$ have so that $F \in D(G)$ ?
5. How can we compute $a_n, b_n$ ?

The goal of this chapter is to answer some of these questions (without proof most of the time), and thus provide a glimpse into the world of Univariate Extreme Value Theory. Proofs of most of the results are available in Embrechts et al. (1997), Resnick (2008).

**Remark 1.1.2.** The theory is developed with maxima of i.i.d. random variables but can be done with minima too since,

\[ \min\{X_1, X_2, \ldots, X_n\} = -\max\{-X_1, -X_2, \ldots, -X_n\}. \]

Let us proceed to answer the above questions now. Consider the following example.
1.1. INTRODUCTION

Example 1.1.1. Let $X$ follow a standard exponential distribution. Therefore the d.f. of $X$ is given by:

$$F_X(x) = 1 - e^{-x}, \quad x > 0.$$ 

Now if we have $X_1, X_2, \ldots$ i.i.d. $F$, then

$$\mathbb{P}(M_n \leq x + \log n) = (1 - e^{-x - \log n})^n = (1 - e^{-x / n})^n \to \exp\{-e^{-x}\} =: \Lambda(x), \quad x \in \mathbb{R}.$$ 

The limit distribution is called the Gumbel distribution.

So we deduce that the Gumbel distribution is a possible limit distribution according to Definition 1.1.1.

The following proposition provides a partial answer to Question 1.

Proposition 1.1.1. If there exists $G$, such that $F \in D(G)$, then $F$ has to satisfy the following condition at the right end point $x_F$:

$$\lim_{x \uparrow x_F} \frac{F(x)}{F(x_-)} = 1. \quad (1.1.3)$$ 

Here $f(x) = \lim_{y \to x} f(y)$.

1. Discrete distributions such as Geometric, Poisson, Negative Binomial do not satisfy condition (1.1.3). Hence they are not in any (max) DOA.

2. All continuous distributions satisfy (1.1.3). But that does not mean they belong to a (max) DOA. e.g., log-Pareto does not belong to any DOA.

Definition 1.1.2. Distribution functions $F$ and $G$ are of the same type if

$$F(x) = G(Ax + B), A > 0, B \in \mathbb{R}.$$ 

So the distributions are same upto a location and a scale change.

Theorem 1.1.2 (Convergence to types theorem). Suppose $U(x)$ and $V(x)$ are two non-degenerate distribution functions. Suppose for $n \geq 1$, there exist distribution functions $F_n, a_n, \alpha_n > 0, b_n, \beta_n \in \mathbb{R}$ such that

$$F_n(a_n x + b_n) \Rightarrow U(x), \quad F_n(\alpha_n x + \beta_n) \Rightarrow V(x). \quad (1.1.4)$$

Then as $n \to \infty$,

$$\frac{\alpha_n}{a_n} \to A > 0, \quad \frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R} \quad \text{and} \quad V(x) = U(Ax + B). \quad (1.1.5)$$

This means $U(x)$ and $V(x)$ are of the same type, in other words, if $X \sim U(x)$ and $Y \sim V(x)$ then $X$ and $Y$ are of the same type.

We skip the proof of this theorem which is available in Resnick (2008).

The answer to Question 2 is quite clear from Theorem 1.1.2. If $F \in D(G_1)$ and $F \in D(G_2)$ then we must have $G_1$ and $G_2$ to be of the same type. Let us see what we can say about Question 3.

Definition 1.1.3. A non-degenerate random distribution function $F$ is max-stable if for $X_1, X_2, \ldots, X_n$ i.i.d. $F$ there exist $a_n > 0, b_n \in \mathbb{R}$ such that $M_n \overset{d}{=} a_n X_1 + b_n$. ($M_n = \max\{X_1, \ldots, X_n\}$)
For each \( n \), \( M_n \) and \( X_1 \) are of the same type here.

**Example 1.1.2.** Suppose \( F(x) = \exp(-\frac{1}{2}x), x > 0 \). This is the Unit Fréchet distribution. If we take \( a_n = n, b_n = 0 \), then it follows that

\[
\mathbb{P}\left( \frac{M_n - b_n}{a_n} \leq x \right) = F^n(nx) = \exp\left(-\frac{n}{nx}\right) = F(x).
\]

**Theorem 1.1.3.** The class of all max-stable distribution functions coincide with the class of all limit laws \( G \) as given in (1.1.1).

**Proof.** 1. One part is trivial. If \( X_1, X_2, \ldots \) are i.i.d. \( G \), \( G \) is max-stable and \( M_n = \bigvee_{i=1}^n X_i \), then

\[
M_n \overset{d}{=} a_n X_1 + b_n
\]

for some \( a_n > 0, b_n \in \mathbb{R} \). Then \( \forall x \in \mathbb{R} \)

\[
\lim_{n \to \infty} \mathbb{P}\left( \frac{M_n - b_n}{a_n} \leq x \right) = G(x).
\]

2. Now suppose \( H \) is non-degenerate and \( \exists a_n > 0, b_n \in \mathbb{R} \) such that

\[
\lim_{n \to \infty} F^n(a_n x + b_n) = H(x).
\]

We claim that \( H \) is max-stable. Observe that \( \forall k \in \mathbb{N} \),

\[
\lim_{n \to \infty} F^{nk}(a_n x + b_n) = H^k(x),
\]

and

\[
\lim_{n \to \infty} F^{nk}(a_{nk} x + b_{nk}) = H(x).
\]

Now apply Convergence to types theorem: there exist \( a_k^* > 0, b_k^* \in \mathbb{R} \) such that

\[
\lim_{n \to \infty} \frac{a_{nk}}{a_n} = a_k^*, \quad \lim_{n \to \infty} \frac{b_{nk} - b_n}{a_n} = b_k^*
\]

and

\[
H(x) = H^k(a_k^* x + b_k^*) \quad \forall k \in \mathbb{N}.
\]

Thus if \( Y_1, \ldots, Y_k \) are i.i.d. from \( H \) then for all \( k \in \mathbb{N} \),

\[
Y_1 \overset{d}{=} \frac{\bigvee_{i=1}^n Y_i - b_k^*}{a_k^*} \quad \text{which implies} \quad \bigvee_{i=1}^n Y_i \overset{d}{=} a_k^* Y_1 + b_k^*.
\]

\[\square\]

**Theorem 1.1.4** (Fisher-Tippett (1928), Gnedenko(1943)). Suppose there exist \( a_n > 0, b_n \in \mathbb{R}, n \geq 1 \) such that

\[
\frac{M_n - b_n}{a_n} \Rightarrow G
\]

where \( G \) is non-degenerate, then \( G \) is of one of the following three types:

1. **Type I, Gumbel**: \( \Lambda(x) = \exp\{-e^{-x}\}, x \in \mathbb{R} \).

2. **Type II, Fréchet**: \( \Phi_\alpha(x) = \begin{cases} 0 & \text{if } x < 0, \\ \exp\{-x^{-\alpha}\} & \text{if } x \geq 0, \end{cases} \) for some \( \alpha > 0 \).
1.1. INTRODUCTION

3. Type III, Weibull: $\Psi_\alpha(x) = \begin{cases} 
\exp\{-(x^\alpha)\} & \text{if } x < 0, \\
1 & \text{if } x \geq 0,
\end{cases}$

for some $\alpha > 0$.

These three types of distributions are together called the class of Extreme Value distributions (abbreviated EV distributions). Interesting probabilistic fact:

$$X \sim \Phi_\alpha \Leftrightarrow -\frac{1}{X} \sim \Psi_\alpha \Leftrightarrow \log X^\alpha \sim \Lambda.$$ 

But this does not help us in studying the tail behavior of all these together as each of them has a different tail behavior near their right end points.

**Remark 1.1.3.** The moral of the story is that

Class of EV distributions = Max-stable distributions = Distributions appearing as limits in Definition 1.1.1

Thus we have a characterization of the limit distributions appearing as limits in Definition 1.1.1, which answers Question 3.

**Definition 1.1.4** (Generalized Extreme Value Distributions). For any $\gamma \in \mathbb{R}$, define the distribution $G_\gamma(x) = \exp\{-1 + \gamma x\}^{1/\gamma}, \quad 1 + \gamma x > 0$

where for $\gamma = 0$ we have $G_0(x) = -\exp\{-e^{-x}\}$. The family of distributions $G_\gamma\left(\frac{x-\mu}{\sigma}\right)$, for $\mu \in \mathbb{R}, \sigma > 0, \gamma \in \mathbb{R}$ is called the family of generalized extreme value distributions under von Mises or von Mises-Jenkins parametrization.

1. $\gamma = 0$. This relates to the case of Gumbel distribution, $G_0(x) = \exp\{-e^{-x}\}, x \in \mathbb{R}$. Observe that

$$1 - G_0(x) = 1 - \exp\{-e^{-x}\} \sim e^{-x}, \quad x \to \infty.$$ 

All moments of this distribution exist, so it is not heavy-tailed.

2. $\gamma > 0 : 1 + \gamma x > 0$ implies $x > -\frac{1}{\gamma}$.

After a linear shift this is the same type as Fréchet:

$$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, x > 0 \quad \text{where } \alpha = \frac{1}{\gamma}.$$ 

Observe that

$$1 - \Phi_\alpha(x) = 1 - \exp\{-x^{-\alpha}\} \sim x^{-\alpha}, \quad x \to \infty.$$ 

This is a heavy-tailed distribution. Note that

$$\int_0^{\infty} x^p \Phi_\alpha(dx) \begin{cases} < \infty & \text{if } p < \alpha, \\
= \infty & \text{if } p \geq \alpha.
\end{cases}$$ 

3. $\gamma < 0 : 1 - |\gamma|x > 0$ implies $x < \frac{1}{|\gamma|}$.

This is the same type as Weibull (actually inverse-Weibull):

$$\Psi_\alpha(x) = \exp\{-(x^\alpha)\}, x < 0 \quad \text{where } \alpha = \frac{1}{|\gamma|}.$$ 

This distribution function has a finite right tail $x_F = \frac{1}{|\gamma|}$. Questions of being heavy-tailed or light-tailed do not arise.

**Remark 1.1.4.** We do not answer questions 4 in this course explicitly. Many sufficient conditions exist for $F$ to belong to the domain of attraction of some EVD $G_\gamma$. One popular criteria is the von Mises condition. For further reading consult de Haan and Ferreira (2006), Embrechts et al. (1997).
CHAPTER 1. UNIVARIATE EXTREME VALUE THEORY

1.1.1 Domain of Attraction condition

**Definition 1.1.5.** Suppose \( f : \mathbb{R} \rightarrow \mathbb{R} \) is a non-decreasing function. Then define the *generalized or left-continuous inverse* of \( f \) as:

\[
f^\leftarrow(x) = \inf\{y : f(y) \geq x\}
\]

with the convention \( \inf\{\mathbb{R}\} = -\infty, \inf\emptyset = \infty \).

**Lemma 1.1.5.** If \( g, f_n, n \geq 1 \) are real-valued non-decreasing functions and

\[
\forall x \in (a, b) \cap \mathcal{C}(g) : f_n(x) \rightarrow g(x),
\]

then

\[
\forall y \in (g(a), g(b)) \cap (\mathcal{C}(g^\leftarrow)) : f^\leftarrow_n(y) \rightarrow g^\leftarrow(y).
\]

**Remark 1.1.5.** The following are equivalent for all \( x \), s.t. \( 0 < G(x) < 1 \):

1. \( \exists a_n > 0, b_n \in \mathbb{R}, G \) non-degenerate such that,

\[
F^n(a_n x + b_n) \Rightarrow G(x), \quad \text{as } n \to \infty.
\]

2. \( \exists a_n > 0, b_n \in \mathbb{R}, G \) non-degenerate such that, \( \forall x \in \mathcal{C}(G) \),

\[
n(1 - F(a_n x + b_n)) \rightarrow -\log G(x), \quad \text{as } n \to \infty.
\]

3. \( \exists a(t) > 0, b(t) \in \mathbb{R}, G \) non-degenerate such that, for \( x \in \mathcal{C}(G) \),

\[
t(1 - F(a(t)x + b(t))) \rightarrow -\log G(x), \quad \text{as } t \to \infty.
\]

4. Define \( U = \left(\frac{1}{1-F}\right)^\leftarrow. \) Then \( \exists a(t) > 0, b(t) \in \mathbb{R}, G \) non-degenerate such that,

\[
x > 0 : \frac{U(tx) - b(t)}{a(t)} \rightarrow D(x) = G^\leftarrow(e^{-1/x}), \quad \text{as } n \to \infty.
\]

5. \( \exists a(t) > 0, b(t) \in \mathbb{R}, G \) non-degenerate such that,

\[
F^t(a(t)x + b(t)) \Rightarrow G(x), \quad \text{as } n \to \infty.
\]

**Proof.** We will prove some of the implications. Rest are left as exercises. Fix \( x \in \mathcal{C}(G), 0 < G(x) < 1 \).

1 \( \Leftrightarrow 2 \) Clearly, here in either case we must have

\[
F(a_n x + b_n) \overset{n \to \infty}{\longrightarrow} 1.
\]

Recall that using l'Hôpital's rule we have

\[
\lim_{z \to 1} \frac{\log z}{1 - z} = \lim_{z \to 1} \frac{-1/z}{-1} = 1. \quad (1.1.6)
\]

Therefore, as \( n \to \infty \)

\[
F^n(a_n x + b_n) \Rightarrow G(x).
\]

\[
\Leftrightarrow n \log F(a_n x + b_n) \Rightarrow -\log G(x). \quad \text{(Taking logarithms)}
\]

\[
\Leftrightarrow n(1 - F(a_n x + b_n)) \Rightarrow \log G(x). \quad \text{(Using (1.1.6))}
\]
1.1. INTRODUCTION

2 ⇒ 3 Let \( a(t) = a[t] \), \( b(t) = b[t] \) where for any \( x \in \mathbb{R} \), \([x]\) denotes the highest integer which has value less than or equal to \( x \). Then

\[
 t(1 - F(a(t)x + b(t))) \leq ([t] + 1)(1 - F(a[t]x + b[t])) = ([t] + 1) \times [t](1 - F(a[t]x + b[t]))
\]

\[
 t \to \infty 1 \times (-\log G(x)) = -\log G(x). \tag{1.1.7}
\]

Similarly we can show that

\[
 \lim_{t \to \infty} t(1 - F(a(t)x + b(t))) \geq -\log G(x). \tag{1.1.8}
\]

Hence the implication.

3 ⇒ 2 Straight forward.

3 ⇒ 4

\[
 t(1 - F(a(t)x + b(t))) \xrightarrow{t \to \infty} (-\log G(x))
\]

implies

\[
 \frac{1}{t} \left( \frac{1}{1 - F(a(t)x + b(t))} \right) \to \frac{1}{-\log G(x)}.
\]

Inverting the above convergence, we get

\[
 \frac{U(ty) - b(t)}{a(t)} \to \left( -\frac{1}{\log G} \right)^{\gamma}(y) = G^{-}(e^{-1/y}).
\]

4 ⇒ 3 Another inversion.

3 ⇔ 5 Easy.

\[ \square \]

Remark 1.1.6. Recall that \( G \) is parametrized by \( \gamma \in \mathbb{R} \) (Definition 1.1.4)

\[
 G_{\gamma}(x) = \exp(-1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0.
\]

Thus \( G_{\gamma}^{-}(y) = \frac{(-\log y)^{-\gamma - 1}}{\gamma} \). Therefore equivalence (4) in the previous remark implies

\[
 y > 0 : \frac{U(ty) - b(t)}{a(t)} \xrightarrow{t \to \infty} G^{-}(e^{-1/y}) = \frac{y^{\gamma} - 1}{\gamma}.
\]

\[
 y = 1 : \frac{U(t) - b(t)}{a(t)} \xrightarrow{t \to \infty} \frac{1}{\gamma} - 1 = 0.
\]

Therefore we have

\[
 y > 0 : \frac{U(ty) - U(t)}{a(t)} \xrightarrow{t \to \infty} \frac{y^{\gamma} - 1}{\gamma}.
\]

It is quite clear that we can take

\[
 b(t) = U(t) = \left( \frac{1}{1 - F} \right)^{-\gamma} = F^{-}\left( 1 - \frac{1}{t} \right), \quad t > 0.
\]

Note that \( U \) is like a quantile function and as \( t \to \infty \), \( U(t) \) is a very high quantile of the distribution \( F \). For \( t \geq 1 \), \( U(t) \) is precisely the \( (1 - \frac{1}{t}) \times 100^{th} \) percentile. This provides a partial answer to Question 5. More coming up.
CHAPTER 1. UNIVARIATE EXTREME VALUE THEORY

1.1.2 Uniform convergence

Definition 1.1.6 (Uniform convergence). Consider a sequence of functions \( f_n : \mathbb{R} \rightarrow \mathbb{R}, n \geq 0 \). Then \( f_n, n \geq 1 \) converges to \( f_0 \) uniformly on \( A \subset \mathbb{R} \), if

\[
\sup_{x \in A} |f_n(x) - f_0(x)| \rightarrow 0, \quad n \rightarrow \infty.
\]

Then \( f_n \) converges to \( f_0 \) locally uniformly if it converges uniformly on all compact sets, i.e., for any \( a < b \),

\[
\sup_{x \in [a,b]} |f_n(x) - f_0(x)| \rightarrow 0, \quad n \rightarrow \infty.
\]

Lemma 1.1.6. Suppose \( \{h_n : \mathbb{R} \rightarrow \mathbb{R}, n \geq 0\} \) are non-decreasing, \( h_0 \) is continuous and \( h_n(x) \rightarrow h_0(x) \) for all \( x \in \mathbb{R} \) then \( h_n \rightarrow h_0 \) locally uniformly.

For an idea about the proof see (Resnick, 2008, page 1).

Remark 1.1.7. A consequence of Lemma 1.1.6 is that the convergence in the domain of attraction condition given by (1.1.2) is locally uniform. This follows for the following three facts.

1. The limits in (1.1.2) are extreme value distributions. So the limit \( G = G_\gamma \), for some \( \gamma \in \mathbb{R} \) as defined in Definition 1.1.4 and hence are clearly continuous.
2. \( F_n \) and \( G \) are distribution functions and hence non-decreasing.
3. \( F_n(a_n x + b_n) \rightarrow G(x), \forall x \in C(G) \) according to (1.1.2). But \( G \) is continuous. Thus \( F_n(a_n x + b_n) \rightarrow G(x), \forall x \in \mathbb{R} \).

Now we can apply Lemma 1.1.6 to conclude the uniform convergence in (1.1.2).

1.2 Regular Variation

The concept of regular variation of functions is very intrinsically related to the study of extreme value theory. In this section we survey the very basic properties of a regularly varying function. The material covered can be found in (de Haan and Ferreira, 2006, Appendix A), Resnick (2007) and a nice monograph by Seneta (1976). \((\mathbb{R}^+ = (0, \infty))\)

Definition 1.2.1 (Regular variation). A measurable function \( U : \mathbb{R}^+ \mapsto \mathbb{R}^+ \) is regularly varying with index \( \rho \in \mathbb{R} \) (written \( U \in RV_\rho \)) if for \( x > 0 \),

\[
\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\rho. \quad (1.2.1)
\]

\( \rho \) is called the exponent of variation.

When \( \rho = 0 \), then \( U \) is called slowly varying. We often denote slowly varying functions by \( L(x) \). Observe that if \( U \in RV_\rho \), then \( L(x) := x^{-\rho} U(x) \in RV_0 \). Thus any regularly varying function \( U \) with index \( \rho \) can be represented by \( x^\rho L(x) \), where \( L \) is a slowly varying function.

Example 1.2.1. The generic example of a \( \rho \)-varying function is \( x^\rho \).

1. The functions \( \log(1 + x), \log \log(e + x), \exp(\sqrt{\log(1 + x)}) \) are all slowly varying functions. In fact any function \( U \) so that \( \lim_{x \rightarrow \infty} U(x) =: U(\infty) \) exists, positive and is finite is slowly varying.
2. On the other hand \( e^x, \sin(x + 2), \exp[\log x] \) are not regularly varying.
Example 1.2.2. For applications in probability, our interest is in distributions whose tails are regularly varying. Let us see some examples

1. Pareto distribution: \( F(x) = 1 - x^{-\alpha}, x \geq 1, \alpha > 0 \).

2. Fréchet distribution: \( \Phi_{\alpha}(x) = \exp\{-x^{-\alpha}\}, x \geq 0, \alpha > 0 \).

\( \Phi_{\alpha}(x) \) has the property that
\[ 1 - \Phi(x) \sim x^{-\alpha}, \quad \text{as} \quad x \to \infty. \]

3. Cauchy distribution: p.d.f. \( f(x) = \frac{1}{\pi (1 + x^2)} \) with distribution function \( F \). Here
\[ 1 - F(x) \sim (\pi x)^{-1}. \]

1.2.1 Properties of regularly varying functions

In this context let us also review some properties of regularly varying functions which will be helpful. These are quite useful.

1. \( U \in RV_{\rho}, \rho > 0 \) then \( U^{-} \in RV_{1/\rho} \).

2. \( U_1 \in RV_{\rho_1}, U_2 \in RV_{\rho_2} \). Then \( U_1 + U_2 \in RV_{\max(\rho_1, \rho_2)} \). Moreover if \( \lim_{x \to \infty} U_2(x) = \infty \), then
\[ U_1 \circ U_2 \in RV_{\rho_1 \rho_2}. \]

3. Karamata’s Theorem:

(a) If \( \rho \geq -1 \) then \( V \in RV_{\rho} \) implies \( \int_0^t V(s)ds \in RV_{\rho+1} \) and
\[ \lim_{t \to \infty} \frac{\int_0^t V(s)ds}{t} = \rho + 1. \]

If \( \rho < -1 \) then \( V \in RV_{\rho} \) implies \( \int_0^t V(s)ds < \infty \) and \( \int_0^t V(s)ds \in RV_{\rho+1} \) and
\[ \lim_{t \to \infty} \frac{\int_0^t V(s)ds}{t} = -\rho - 1. \]

(b) If \( V \) satisfies
\[ \lim_{t \to \infty} \frac{tV(t)}{\int_0^t V(s)ds} = \lambda \in (0, \infty), \]

then \( V \in RV_{-\lambda^{-1}} \). If \( \int_0^\infty V(s)ds < \infty \) and
\[ \lim_{t \to \infty} \frac{tV(t)}{\int_0^t V(s)ds} = \lambda \in (0, \infty), \]

then \( V \in RV_{-\lambda^{-1}} \). This is a partial converse of (a).
4. **Karamata Representation:** $V \in RV_\rho$ if and only if $V$ can be represented as

$$V(x) = c(x) \exp \left\{ \frac{1}{x} \int c(s) ds \right\}$$

where $c : (0, \infty) \to (0, \infty)$ and $c : (0, \infty) \to \mathbb{R}$ are functions such that

$$\lim_{x \to \infty} c(x) = c \in (0, \infty) \quad \text{and} \quad \lim_{s \to \infty} c(s) = \rho.$$

5. $U \in RV_\rho$. Then $\lim_{t \to \infty} \frac{U(tx)}{U(t)} = x^{\rho}$ locally uniformly.

6. (Potter bounds) If $U \in RV_\rho$ then $\forall \epsilon > 0$, there exists $t_0$ such that $\forall t > t_0$,

$$(1 - \epsilon)x^{\rho - \epsilon} \leq \frac{U(tx)}{U(t)} \leq (1 + \epsilon)x^{\rho + \epsilon}.$$  \hspace{1cm} (1.2.2)

for all $x \geq 1$.

**Theorem 1.2.1.** $F \in D(G_\gamma)$ if and only if

1. $\gamma > 0 : x_F = \infty$ and $F \in RV_{-1/\gamma}$ where $F(x) = 1 - F(x)$. In this case, $F_n(a_n x + b_n) \xrightarrow{n \to \infty} \Phi_\frac{1}{\gamma}(x)$ with $a_n = U(n), b_n \equiv 0$.

2. $\gamma < 0 : x_F < \infty$ and $\tilde{F}(x) = 1 - F(x_F - \frac{1}{x})$, $x > 0$. In this case, $F_n(a_n x + b_n) \xrightarrow{n \to \infty} \Psi_{-\frac{1}{\gamma}}(x)$ with $a_n = x_F - U(n), b_n \equiv x_F$.

3. $\gamma = 0 : x_F \leq \infty$ and

$$\lim_{t \to x_F} \frac{1 - F(t + xf(t))}{1 - F(t)} = e^{-x}, \quad \forall x \in \mathbb{R}.$$  \hspace{1cm} (1.2.3)

for a suitable positive function $f$. If (1.2.3) holds for some $f$ then $\int_1^{x_F} (1 - F(s)) ds < \infty$ for $t < x_F$ and $f$ can be chosen as

$$f(t) := \frac{\int_{1}^{t} (1 - F(s)) ds}{1 - F(t)}, \quad t < x_F.$$  \hspace{1cm} (1.2.4)

In this case $F^n(a_n x + b_n) \xrightarrow{t \to \infty} \Lambda(x)$ with $b_n = U(n)$ and $a_n = f \circ U(n)$.

**Example 1.2.3.** Suppose $X_1, X_2, \ldots$ are i.i.d. from the uniform distribution on $(0, 1)$. If we denote the distribution of $U(0, 1)$ by $F$, then

$$F(x) = \begin{cases} 
0 & \text{if } x \leq 0, \\
x & \text{if } 0 < x < 1, \\
1 & \text{if } x \geq 1.
\end{cases}$$

Note that $x_F = 1$. Therefore $\gamma \leq 0$, according to Theorem 1.2.1. Trivially $\tilde{F}(x) = 1$ for $0 < x < 1$. Now

$$\forall x \geq 1, \quad \tilde{F}(x) = 1 - F(x_F - \frac{1}{x}) = \frac{1}{x} \in RV_{-1}.$$  \hspace{1cm} (1.2.5)

Thus using Theorem 1.2.1, we have $a_n = x_F U(n) = \frac{1}{n}, b_n = x_F = 1$. Check that

$$F^n(\frac{1}{n} + 1) \to e^x = \Psi_1(x), \quad x \leq 0.$$
1.3 A condition for belonging to an extreme value domain

**Theorem 1.3.1.** Let $F$ be a distribution function with right end point $x_F$. Suppose $F''(x)$ exists and $F'(x) > 0$ in some left neighbourhood of $x_F$. If $F$ satisfies the von Mises condition, i.e.:

$$\lim_{t \uparrow x_F} \frac{(1 - F(t))F''(t)}{(F'(t))^2} = -\gamma - 1 \quad (1.3.1)$$

then $F \in D(G_\gamma)$.

For a proof of the Theorem see (de Haan and Ferreira, 2006, page 15). If (1.3.1) holds then $F^n(a_nx + b_n) \to G_\gamma(x)$, $\forall x$ with $a_n = U(n)$ and $b_n = U'(n) = 1/(nF'(b_n))$.

If we define $R(t) = -\log(1 - F(t))$, also called the integrated hazard function, then we have

$$r(t) = R'(t) \quad \text{hazard function}$$

$$= \frac{F'(t)}{1 - F(t)}.$$ 

Then if the von Mises conditions hold we have

$$\lim_{t \uparrow x_F} \left( \frac{1}{r(t)} \right)' = \lim_{t \uparrow x_F} \left( \frac{1 - F(t)}{F'(t)} \right)' = \lim_{t \uparrow x_F} \left( \frac{-(F'(t))^2 - (1 - F(t))F''(t)}{(F'(t))^2} \right)' \quad \text{(using } (1.3.1) \text{)}$$

$$= -1 - (-\gamma - 1) = \gamma.$$ 

So the von Mises condition can be also be written as $\lim_{t \uparrow x_F} \left( \frac{1}{r(t)} \right)' = \gamma$ which is often more convenient to check.

**Example 1.3.1.** Let $F(x) = e^{-xp}, x > 0, p > 0$. Then

$$R(t) = x^p.$$ 

$$r(t) = R'(t) = px^{p-1}.$$ 

$$\frac{1}{r(t)} = p^{-1}x^{1-p}.$$ 

$$\left( \frac{1}{r(t)} \right)' = \frac{(1-p)}{p}x^{-p} \to 0 \quad \text{as } n \to \infty.$$ 

Hence in this case $\gamma = 0$.

**Example 1.3.2.** Let $F(x) = x^{-\alpha}, x \geq 1, \alpha > 0$. Then $F(x) = x^{-\alpha} = e^{-\alpha \log x}$.

$$R(t) = \alpha \log x.$$ 

$$r(t) = R'(t) = \frac{\alpha}{x}.$$ 

$$\frac{1}{r(t)} = \frac{x}{\alpha}.$$ 

$$\left( \frac{1}{r(t)} \right)' = \frac{1}{\alpha}.$$ 

Hence in this case $\gamma = \frac{1}{\alpha}$. 
Example 1.3.3. Try the same with the following:

\[ F(x) = \frac{e}{x \log x}, \quad x > e. \]

Remark 1.3.1. A nice application of estimating the end-point of an extreme value distribution can be found in Einmahl and Magnus (2008).

1.4 Applications

We would conclude this chapter with a small comment on applying the techniques we learnt here. In applications, the following two approaches have been popular in univariate extreme value theory:

1. **Method of Block Maximas:** Measure maximum of observations over a block (say, yearly maxima): \( X_1, \ldots, X_n \). Hope (and hence assume)
   - (a) i.i.d. (or close to it)
   - (b) common distribution \( G(\frac{x-\mu}{\sigma}) \).

2. **Method of Exceedances:** From the entire data extract excesses relative to a threshold \( t \). Hope (and hence assume)
   - (a) i.i.d.
   - (b) common distribution approximately Generalized Pareto (de Haan and Ferreira, 2006, page 65).

We have not seen the Generalized Pareto distribution here but it appears quite naturally in these problems. See de Haan and Ferreira (2006) for further details.

There are other questions that arise in a statistical study. How to get \( t \)? How to quantify error in approximating actual distribution and limit distribution. We won’t study these here.
Chapter 2

Multivariate Extreme Value Theory

2.1 Introduction

Let us consider the following examples.

Example 2.1.1. This is the famous example with respect to flooding in the Netherlands. Engineers have been concerned about the safety of the town of Petten in Netherlands, which is protected by a seawall from the North Sea. The maximum wave height (H) and still water level (S) data near a seawall in the town of Petten in The Netherlands has been collected during 828 storms. After some study the scientists decided that the city of Petten is flooded by breaching this seawall when \((H, S)\) belongs to the failure set \(C\) defined as follows:

\[ C := \{(H, S) : 0.3H + S > 7.6 \}. \]

One way to estimate the failure probability, i.e., the probability of the set \(C\) can be calculated using empirical methods. But since there has been no breaching of the wall during this observation period, this approach is not viable here. The problem is very similar to the univariate problems we had seen earlier. The way out is EVT. See de Haan and de Ronde (1998) for further studies on this particular problem.

Example 2.1.2. Internet traffic data. It is often surmised (with empirical evidence) that network traffic data portrays heavy tail behavior. So Extreme Value Theory is a natural tool to use when concerned with multiple variables, important among which are size of file transferred in a network session, amount of time taken to do so and the average throughput rate. See Maulik et al. (2002).

Let us assume that \(X_1, X_2, \ldots, X_n\) are \(d\)-dimensional i.i.d. random vectors from some distribution \(F\). If we intend to study the extremes of the distribution \(F\), we must know what an extreme value really mean in the multivariate set-up. Of course, for \(d \geq 2\), there is no natural ordering of the random sample. Barnett (1976) talks about four kinds of ordering for multivariate data set which leads to different approaches for studying extremes (maxima or minima) in a multivariate set-up.

1. Component-wise maxima depending on Marginal ordering.

2. Maxima based on Reduced (Aggregate) ordering based on a single value computed from a multivariate observation through a function \(f : \mathbb{R}^d \mapsto \mathbb{R}\). Usually the function \(f\) is some measure of generalized distance, say, \(f(x) = (x - \alpha)^T \Sigma (x - \alpha)\).

3. Maxima based on Partial ordering, say, based on convex hulls.

Our approach for extreme value analysis will be based on the first one mentioned above. It turns out that the theory behind component-wise maxima is quite rich and provides answers to the questions we have. We will study multivariate extreme value theory in the case dimension, \( d = 2 \). Most of the results and definitions given henceforth will also hold for \( d > 2 \).

In univariate extreme value theory we found that the limit distributions of sample maxima can be characterized through a parametric family of generalised extreme value or generalised Pareto distribution. We will learn in this section that such is not the case with the multi-dimensional case. Most of the material here is drawn from Beirlant et al. (2004), de Haan and Ferreira (2006), Resnick (2008).

2.2 Limit Distributions

Assume \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) are i.i.d. from a distribution \( F(x, y) \). Since we are taking the component-wise maxima approach, let us denote

\[
M_n := \left( \bigvee_{i=1}^{n} X_i, \bigvee_{i=1}^{n} Y_i \right) = (M_n^{(1)}, M_n^{(2)}).
\]

First observe that \( M_n \) might not be a sample point. But still the rich theory developed through this method answers many questions. Secondly, as we had done in the univariate case we wouldn’t study the minima, since

\[
m_n := \left( \bigwedge_{i=1}^{n} X_i, \bigwedge_{i=1}^{n} Y_i \right) = -\left( \bigvee_{i=1}^{n} -X_i, \bigvee_{i=1}^{n} -Y_i \right).
\]

2.2.1 Max-Infinitely Divisible Distributions

Recall from the univariate theory that a univariate distribution function \( F \) is in the domain of attraction of a non-degenerate d.f. \( G \) if \( \exists a(t) > 0, b(t) \in \mathbb{R} \) such that as \( t \to \infty \),

\[
F(tx + b(t)) \Rightarrow G(x).
\]

It so happens that though \( F^t \) is a distribution function whenever \( F \) is for one-dimensional distribution such is not the case for dimension \( d > 1 \). Consider the following example.

**Example 2.2.1** (from Resnick (2008)). Let \( X_1, \ldots, X_n \) be iid random variables from a (continuous distribution) \( H \). We are interested in the range of \( H \) given by

\[
R_n := \bigvee_{i=1}^{n} X_i - \bigwedge_{i=1}^{n} X_i = \bigvee_{i=1}^{n} X_i + \bigvee_{i=1}^{n} (-X_i).
\]

Thus it is only natural to look at the joint distribution of \((X_i, -X_i)\), which has distribution \( F \), say. Then

\[
(M_n^{(1)}, M_n^{(2)}) = \left( \bigvee_{i=1}^{n} X_i, \bigvee_{i=1}^{n} (-X_i) \right) = \bigvee_{i=1}^{n} (X_i, -X_i)
\]

and \( R_n = M_n^{(1)} + M_n^{(2)} \). Clearly \( F \) is defined on \( \{(u_1, u_2) \in \mathbb{R}^2 : u_1 + u_2 = 0\} \). Now observe that

\[
F(0, 1) = F\{ (u_1, u_2) \in \mathbb{R}^2 : u_1 + u_2 = 0, 0 \leq u_2 \leq 1 \} =: p_1,
\]

\[
F(1, 0) = F\{ (u_1, u_2) \in \mathbb{R}^2 : u_1 + u_2 = 0, 0 \leq u_1 \leq 1 \} =: p_2,
\]

\[
F(1, 1) = p_1 + p_2, \quad F(0, 0) = 0.
\]
2.2. LIMIT DISTRIBUTIONS

For $F^*$ to be a distribution function $F^*(A) \geq 0$ for any Borel $A \subset \mathbb{R}^2$. Consider $A = (0,1]^2$.

$$F^*((0,1]^2) = F^*(1,1) - F^*(0,1) - F^*(1,0) + F^*(0,0) = (p_1 + p_2)^t - p_1^t - p_2^t.$$  

The last term might not be non-negative for $t < 1$. If $H$ is Uniform(-1,1), observe that it fails with $t = 1/2$.

**Definition 2.2.1.** A distribution function $F$ on $\mathbb{R}^d$ is **max-infinitely divisible or max-id** if for every $n$ there exists a distribution function $F_n$ on $\mathbb{R}^d$ such that

$$F = F_n.$$  

i.e., $F^{1/n}$ is a distribution. If $X \sim F$ we also call $X$ max-id.

**Proposition 2.2.1.** Suppose $F_n, n \geq 0$ are probability distribution functions on $\mathbb{R}^d$. If $F_n^n \Rightarrow F_0$ then $F_0$ is max-id. Consequently,

1. $F$ is max-id if and only if $F^*$ is a d.f. for all $t > 0$.
2. The class of max-id distributions is closed under weak convergence.

**Proof.** $F_n$’s are d.f.’s and $F_n^n \Rightarrow F_0$. Suppose $x$ is a continuity point of $F_0$. We show that $F_n^{[nt]}(x) \Rightarrow F_0^t(x)$ for all $t > 0$. Clearly $F_n^{[nt]}$ are distribution functions for any $n \geq 1, t > 0$ with $[nt] \geq 1$ and also $F_0^t$ is non-defective (proper), since $F_0$ is non-defective. Fix $t > 0$. If $F_0(x) > 0$ then, as $n \to \infty$,

$$F_n^{[nt]}(x) = (F_n^n)^{[nt]/n}(x) \to 0 = F_0^t(x).$$

If $F_0(x) > 0$, then $F_n(x) \to 1$, and as $n \to \infty$

$$-\log F_n^{[nt]}(x) = -\frac{[nt]}{n} \log F_n^n(x)$$

$$\to -t \log F_0(x) = -\log F_0^t(x).$$

Therefore $F_n^{[nt]}(x) \Rightarrow F_0^t(x), \forall t > 0$. So $F_0^t$ is a distribution function. Thus $F_0^{1/n}, n \geq 1$ is a d.f. and hence $F_0$ is max-id.

Consequence (1) follows easily from the above $F_n = (F_0^{1/n})^n$. Consequence (2) follows since if $G_n$ are max-id and $G_n \Rightarrow G_0$ then $G_0 = \lim_{n \to \infty} G_n = \lim_{n \to \infty} (G_n^n)^n$ and hence is max-id. \hfill $\Box$

**Proposition 2.2.2** (Characterizing max-id distributions). $F$ is max-id if and only if for some $l \in (-\infty, \infty)^d$, there exists an exponent measure $\mu$ on $\mathbb{E} = [l, \infty] \setminus \{l\}$ such that

$$F(y) = \begin{cases} \exp\{-\mu([-\infty, y]^c]\}, & y \geq l \\ 0 & \text{otherwise}. \end{cases}$$

Here $\mu$ is an exponent measure if it is Radon (on compact sets) and

1. $\mu\left(\bigcup_{i=1}^d \{y \in \mathbb{E} : y^{(i)} = \infty\}\right) = 0$.
2. Either $l > -\infty$ OR $x \geq l$ and $x^{(i)} = -\infty$ for some $1 \leq i \leq d$ implies $\mu([-\infty, x]^c) = \infty$.

**Definition 2.2.2** (Multivariate Domain of Attraction). A bivariate distribution function $F$ is said to be in the domain of attraction of a bivariate distribution function $G$ if

1. $G$ has non-degenerate marginal distributions $G_1$ and $G_2$.  

2. There exist sequences \( a_n, c_n > 0, b_n, d_n \in \mathbb{R} \), such that

\[
P\left( \frac{M_n^{(1)} - b_n}{a_n} \leq x, \frac{M_n^{(2)} - d_n}{c_n} \leq y \right) = F^n(a_n x + b_n, c_n y + d_n) \Rightarrow G(x, y).
\]

(2.2.1)

We denote \( F \in D(G) \).

Here \( G \) is a bivariate extreme-value distribution. As we had asked in the univariate case several questions come up now:

1. Given any \( F \), does there exist \( G \) such that \( F \in D(G) \) ?
2. If \( F \in D(G) \), is \( G \) unique ?
3. Can we characterize the class of all possible limits \( G \) according to Definition 2.2.2 ?
4. How can we compute \( a_n, b_n, c_n, d_n \) ?

First note that easy consequences of (2.2.1) are:

\[
\lim_{n \to \infty} P\left( M_n^{(1)} - b_n \leq x \right) = G(x, \infty) =: G_1(x),
\]

and

\[
\lim_{n \to \infty} P\left( M_n^{(2)} - d_n \leq y \right) = G(\infty, y) =: G_2(y).
\]

The same result could be reached using continuous mapping theorem and the two projection maps. Hence \( G_1 \) and \( G_2 \) are univariate extreme value distributions and the constants \( a_n, b_n, c_n, d_n \) can be computed from our knowledge of univariate extreme value theory. Thus there exists \( \gamma_1, \gamma_2 \in \mathbb{R} \) such that we can write:

\[
G_1(x) = \exp\left\{-(1 + \gamma_1 x)^{-1/\gamma_1} \right\}, \quad 1 + \gamma_1 x > 0,
\]

\[
G_2(y) = \exp\left\{-(1 + \gamma_2 y)^{-1/\gamma_2} \right\}, \quad 1 + \gamma_2 y > 0.
\]

Note that for \( \gamma_i = 0 \), we have \( G_i(x) = \exp\{e^{-x}\} \). Now note that though we can compute the marginals through our techniques this does not yield us \( G \). We need to know the dependence structure in \( G \) in order to combine \( G_1 \) and \( G_2 \) together. More on this later.

### 2.2.2 Max-stability

**Definition 2.2.3.** We say a distribution \( G \) is max-stable if for \( i = 1, \ldots, d \) and every \( t > 0 \), \( \exists \alpha_i(t) > 0, \beta_i(t) \in \mathbb{R} \) such that,

\[
G^t(x) = G((\alpha_1(t) x^{(1)} + \beta_1(t), \ldots, \alpha_d(t) x^{(d)} + \beta_d(t))).
\]

Clearly for all \( t > 0 \), \( G^t \) is a distribution function and hence all max-stable distributions are also max-id.

**Exercise:** Give an example of a distribution which is max-id but not max-stable.

**Proposition 2.2.3.** The class of multivariate extreme value distributions is precisely the class of max-stable distributions.

**Proof.** The proof mimics the proof for the univariate case and is left as an exercise.
2.2. LIMIT DISTRIBUTIONS

2.2.3 Some properties of Multivariate EVDs

Remark 2.2.1. Any bivariate extreme value distribution is continuous (in (x,y)). This follows from the subsequent lemma.

Lemma 2.2.4. If G is a bivariate distribution function such that its marginal distribution functions G1 and G2 are continuous, then G(x, y) is itself continuous in (x, y).

Proof. For 0 ≤ u, v ≤ 1 and 0 ≤ u', v' ≤ 1,
\[ |u v - u' v'| = |(u - u') v + u' (v - v')| \leq |u - u'| + |(v - v')|. \] (2.2.2)

Now for x, y, x', y ∈ ℝ, we know that as (x, y) → (x', y'), G1(x) → G1(x') and G2(y) → G2(y'). Suppose (X, Y) ~ G. Then
\[ |G(x, y) - G(x', y')| = |E[1\{X ≤ x\}, 1\{Y ≤ y\}] - E[1\{X ≤ x'\}, 1\{Y ≤ y'\}]| \]
\[ \leq E[1\{X ≤ x\}, 1\{Y ≤ y\}] - 1\{x ≤ x'\}, 1\{y ≤ y'\}] | \]
\[ \leq E[1\{X ≤ x\} - 1\{x ≤ x'\}] + E[1\{y ≤ y\} - 1\{y ≤ y'\}] \text{ (using (2.2.2))} \]
\[ = |G1(x) - G1(x')| + |G2(y) - G2(y')| \text{ (check)} \]
\[ \to 0, \text{ as (x, y) → (x', y').} \]

Thus since the univariate marginals of a bivariate extreme value distribution are continuous, it itself is continuous. This leads us to another result.

Proposition 2.2.5. For F ∈ D(G), the convergence
\[ F^n(a_n x + b_n, c_n y + d_n) \to G(x, y) \]

is locally uniform as n → ∞.

Proof. We provide hints to the proof here. The original analytical proof is due to Buchanan and Hildebrandt (1908).

1. From the fact that F_n, n ≥ 1 are all monotone non-decreasing functions we can show that G has to be non-decreasing.
2. Fix S = [a, b] × [c, d] ⊂ ℝ^2. We show that F^n(a_n x + b_n, c_n y + d_n) converges to G(x, y) uniformly in S.
3. G is continuous on ℝ^2. Hence G is uniformly continuous on S (compactness). Thus, given ε > 0, \( \exists \ δ > 0 \) such that \( d((x, y), (x', y')) < 2δ \) implies \( |G(x, y) - G(x', y')| < ε \). Consider an open cover of S with \( B((x, y), 0) = (x - δ, x + δ) \times (y - δ, y + δ) \). By Heine-Borel Theorem, there exists a finite subcover of \( B_i = (x_i - δ, x_i + δ) \times (y_i - δ, y_i + δ), i = 1, \ldots, k \) of S.
4. For all \( (x_i, y_i), i = 1, \ldots, k \), find \( M_i \) such that for \( n > M_i \), \( |F^n(a_n x_i + b_n, c_n y_i + d_n) - G(x_i, y_i)| < ε \).
5. Clearly for each \( i = 1, \ldots, k \)
\[ V_{n_i} := (F^n(a_n (x_i + δ) + b_n, c_n (y_i + δ) + d_n) - F^n(a_n (x_i - δ) + b_n, c_n (y_i - δ) + d_n)) \]
\[ \to \leq 2ε. \]

Therefore \( \exists N_i, i = 1, \ldots, k \) such that for \( n > N_i \) we have \( V_{n_i} < 3ε \).
6. Define $N = \max\{M_1, \ldots, M_k, N_1, \ldots, N_k\}$. Pick $(x, y) \in S$. $(x, y) \in B_i^*$ for some $i^*$. Easy to show that

$$|F^n(a_n x_n + b_n, c_n y_n + d_n) - G(x, y)|$$

$$\leq |F^n(a_n x_n + b_n, c_n y_n + d_n) - F^n(a_n x^*, b_n, c_n y^* + d_n)| + |F^n(a_n x^*, b_n, c_n y^* + d_n) - G(x^*, y^*)|$$

$$+ |G(x^*, y^*) - G(x, y)|$$

$$\leq V_{n^*} + \epsilon + \epsilon$$

$$\leq 3\epsilon + 2\epsilon = 5\epsilon.$$

Hence the statement is true. \( \square \)

A consequence of this fact is the following corollary.

**Corollary 2.2.6.** If \( \{x_n\} \) and \( \{y_n\} \) are real sequences such that \( x_n \to u, y_n \to v \), then

$$\lim_{n \to \infty} F^n(a_n x_n + b_n, c_n y_n + d_n) = G(u, v).$$

**Proof.** It is easy but we will still go over it. By the local uniform convergence from Proposition 2.2.5, we know that given \( \epsilon > 0 \), there exists \( \delta > 0 \) such that

$$\sup_{u - \delta \leq x \leq u + \delta} \sup_{v - \delta \leq y \leq v + \delta} |F^n(a_n x + b_n, c_n y + d_n) - G(x, y)| \leq \epsilon/2.$$

Also since \( G \) is continuous and \( (x_n, y_n) \to (u, v) \), there exists \( N_1 \geq 1 \), such that when \( n \geq N_1 \),

$$|G(x_n, y_n) - G(u, v)| < \epsilon/2.$$

Also there exists \( N_2 \geq 1 \) such that for \( n \geq N \), we have \( u - \delta \leq x_n \leq u + \delta, v - \delta \leq y_n \leq v + \delta \). Let \( N = N_1 \vee N_2 \). Therefore combining all the above we observe that, for \( n \geq N \),

$$|F_n(a_n x_n + b_n, c_n y_n + d_n) - G(u, v)|$$

$$\leq |F_n(a_n x_n + b_n, c_n y_n + d_n) - G(x_n, y_n)| + |G(x_n, y_n) - G(u, v)|$$

$$\leq \sup_{u - \delta \leq x \leq u + \delta} \sup_{v - \delta \leq y \leq v + \delta} |F^n(a_n x + b_n, c_n y + d_n) - G(x, y)| + |G(x_n, y_n) - G(u, v)|$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Hence we have the result. \( \square \)

Let us go back to the question of uniqueness.

**Theorem 2.2.7.** If \( F \in D(G) \) and \( F \in D(H) \), then there exist \( A, C > 0 \) and \( B, D \in \mathbb{R} \) such that,

$$H(x, y) = G(Ax + B, Cy + D).$$

**Proof.** Since \( F \in D(G) \), from (2.2.1) we know there exist \( a_n, c_n > 0, b_n, d_n \in \mathbb{R} \), such that

$$F^n(a_n x + b_n, c_n y + d_n) \Rightarrow G(x, y).$$

By taking marginals \( F_1, F_2 \) of \( F \) we get the following weak convergences:

$$F^n_1(a_n x + b_n) \Rightarrow G_1(x),$$

$$F^n_2(c_n x + d_n) \Rightarrow G_2(x).$$
2.3. STANDARDIZATION

Since we also have \( F \in D(H) \), there exist \( a'_n, c'_n > 0, b'_n, d'_n \in \mathbb{R} \), such that

\[
F^n(a'_n x + b'_n, c'_n y + d'_n) \Rightarrow H(x, y).
\]

By taking marginals \( F_1, F_2 \) of \( F \) we get the following weak convergences:

\[
F_1^n(a'_n x + b'_n) \Rightarrow H_1(x),
\]

\[
F_2^n(c'_n x + d'_n) \Rightarrow H_2(x).
\]

Thus from univariate convergence to types theorem (Theorem 1.1.2) we have

\[
\frac{a'_n}{a_n} \to A > 0, \quad \frac{b'_n - b_n}{a_n} \to B, \quad \frac{c'_n}{c_n} \to C > 0, \quad \frac{d'_n - d_n}{c_n} \to D.
\]

Using Corollary 2.2.6, it is easy to check that \( H(x, y) = G(Ax + B, Cy + D) \).

\[
\Box
\]

2.3 Standardization

Let us now proceed towards characterizing \( G \). We have seen that the marginals of a bivariate (multivariate) extreme-value distribution (EVD) are all univariate EVDs. To study the dependence structure of the distribution, it would be much more convenient if all the marginals were the same. For this we would actually perform a transformation of the marginals which we call standardization. There are multiple choices for this transformation leading to either Uniform or Gumbel or Weibull or Fréchet marginals. Each one has its own merit and has been explored in the literature. We would reduce to Fréchet(1) margins across all co-ordinates the merits of which will be evident soon.

Recall from the univariate theory the following notation: For \( i = 1, 2 \)

\[
U_i(t) = F_i^-(1 - \frac{1}{t}) = \left( \frac{1}{1 - F_i(t)} \right)^- (t), \quad t > 1.
\]

(2.3.1)

\( F_1, F_2 \) are the marginal distributions of \( F \). (Notational hazard) From univariate theory we know that

\[
F_1^n(a_n x + b_n) \to G_1(x) = \exp\{-(1 + \gamma_1 x)^{-1/\gamma_1}\}, \quad 1 + \gamma_1 x > 0,
\]

implies that

\[
\frac{U_1(tx) - b(t)}{a(t)} = \frac{x^{\gamma_1} - 1}{\gamma_1}
\]

for some functions \( a(t) > 0, b(t) \in \mathbb{R} \). This immediately leads to the fact that we can take \( b := U_1 \) and we can take \( b_n = U_1(n) \) and \( a_n = a(n) \). A similar result can be obtained for the other marginal. Hence we have for \( x, y > 0 \):

\[
x_n := \frac{U_1(nx) - b_n}{a_n} \to \frac{x^{\gamma_1} - 1}{\gamma_1} =: u, \quad (2.3.2)
\]

\[
y_n := \frac{U_2(ny) - d_n}{c_n} \to \frac{y^{\gamma_2} - 1}{\gamma_2} =: v. \quad (2.3.3)
\]
Therefore for all \( x, y > 0 \),
\[
\lim_{n \to \infty} F^n(U_1(nx), U_2(ny)) = G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right)
\]
This is a direct consequence of Corollary 2.2.6 with \( x_n, y_n \) defined as in (2.3.2)-(2.3.3). Thus what we have proved can be summarized as:

**Proposition 2.3.1 (Standardization).** Suppose there exist \( a_n, c_n > 0, b_n, d_n \in \mathbb{R} \) such that
\[
\lim_{n \to \infty} F^n(a_n x + b_n, c_n y + d_n) = G(x, y)
\]
weakly. Then
\[
\lim_{n \to \infty} F^n(U_1(nx), U_2(ny)) = G_0(x, y)
\]
for \( x, y > 0 \), where
\[
G_0(x, y) = G\left(\frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2}\right),
\]
with \( \frac{e-1}{\gamma} = \log x \), for \( \gamma = 0 \). Here \( G_1, G_2 \) are the marginals of \( G \) with respective EV parameters \( \gamma_1, \gamma_2 \) and \( U_i \) as defined in (2.3.1).

We can also write \( G(x, y) = G_0((1 + \gamma_1 x)^{1/\gamma_1}), (1 + \gamma_2 y)^{1/\gamma_2} \).

**Corollary 2.3.2.** (2.3.5) is equivalent to the following:

1. \( n(1 - F(U_1(nx), U_2(ny))) \to - \log G_0(x, y) \), as \( t \to \infty \),
2. \( t(1 - F(U_1(tx), U_2(ty))) \to - \log G_0(x, y) \) as \( t \to \infty \).

**Proof.** From (2.3.5) we have a \( n \to \infty \)
\[
F^n(U_1(nx), U_2(ny)) \to G_0(x, y).
\]
Taking logarithms on both sides we get as \( n \to \infty \)
\[
-n \log F(U_1(nx), U_2(ny)) \to - \log G_0(x, y).
\]
Now since \( \log(x) \sim x - 1 \) as \( x \to 1 \), we have as \( n \to \infty \),
\[
n(1 - F(U_1(nx), U_2(ny))) \to - \log G_0(x, y).
\]
We can easily check that (2.3.7) also holds for the integer \( n \) replaced by the real \( t \) using inequalities akin to (1.1.7).

We can connect to multivariate regular variation on cones here. \( C \subset \mathbb{R}^d \) is a cone if whenever \( x \in C \), then also \( tx \in C \). A measurable function \( h : C \to [0, \infty] \) is multivariate regularly varying on \( C \) with limit function \( \lambda \) if there exists \( V : (0, \infty) \to (0, \infty) \) with \( V \in RV_{\rho} \) for some \( \rho \in \mathbb{R} \) such that,
\[
\lim_{t \to \infty} \frac{h(tx)}{V(t)} = \lambda(x) \quad \forall x \in C.
\]
2.4. EXPONENT MEASURE

Well why do we call this standardization and what do we gain out of Proposition 2.3.1. Observe that

\[ F_{n}(U_{1}(nx), U_{2}(ny)) = \Pr \left[ \bigvee_{i=1}^{n} X_{i} \leq U_{1}(nx), \bigvee_{i=1}^{n} Y_{i} \leq U_{2}(ny) \right] = \Pr \left[ \bigvee_{i=1}^{n} \frac{U_{1}^{-}(X_{i})}{n} \leq x, \bigvee_{i=1}^{n} \frac{U_{2}^{-}(Y_{i})}{n} \leq y \right] \]  

(check)

Thus we have from (2.3.5) that

\[ \Pr \left[ \bigvee_{i=1}^{n} \frac{X_{i}^{*}}{n} \leq x, \bigvee_{i=1}^{n} \frac{Y_{i}^{*}}{n} \leq y \right] \rightarrow G_0(x, y) \]

weakly. Now let us find the distribution of \( X_{i}^{*} \) and \( Y_{i}^{*} \).

\[ \Pr(X_{i}^{*} \leq t) = \Pr(X_{i} \leq U_{i}(t)) = \Pr \left( X_{i} \leq F_{i}^{-}(1 - \frac{1}{t}) \right) = \Pr \left( \frac{1}{1 - F_{i}(X_{i})} \leq t \right). \]

Note that if \( F_{i} \) is continuous then \( F_{i}(X_{i}) \sim U(0, 1) \). Then

\[ \Pr(X_{i}^{*} \leq t) = \Pr \left( \frac{1}{1 - F_{i}(X_{i})} \leq t \right) = \Pr \left( F_{i}(X_{i}) \leq 1 - \frac{1}{t} \right) = 1 - \frac{1}{t}. \]

Thus \( X_{i}^{*} \) follows a standard Pareto distribution. The same is true for \( Y_{i}^{*} \). Hence the transformations \( U_{1}^{-} \) and \( U_{2}^{-} \) standardizes \( X_{i}, Y_{i} \) to standard Pareto distributions. Even when \( F \) is not continuous the standardized variables \( X_{i}^{*}, Y_{i}^{*} \) have asymptotically Pareto-like tails. Also note that the marginal \( G_{01} \) of \( G_{0} \) turns out to be:

\[ G_{01}(x) = G_{0}(x, \infty) = G \left( \frac{x^{\gamma_{1}-1}}{\gamma_{1}}, \infty \right) = G_{1} \left( \frac{x^{\gamma_{1}-1}}{\gamma_{1}} \right) = \exp \left\{ - (1 + \gamma_{1} \left( \frac{x^{\gamma_{1}}}{\gamma_{1}} \right)^{-1/\gamma_{1}} \right\} = e^{-1/x} = \Phi_{1}(x), x > 0, \]

which is the Fréchet (1) distribution. The same is true for the other marginal, \( G_{02}(y) = e^{-1/y}, y > 0 \). Thus now we have reduced the problem of characterizing all bivariate extreme value distributions to the the problem of characterizing all bivariate EVD’s with Fréchet(1) marginals.

2.4 Exponent measure

We have seen that all max-stable distributions are also max-\( \text{id} \). Recall that max-\( \text{id} \) distributions are characterized by their exponent measure (Proposition 2.2.2). See Balkema and Resnick (1977) for further reference.
This exponent measure may be uniquely chosen by fixing \( l = 0 \) for the space \( E = [l, \infty] \setminus \{l\} \). Suppose \( x, y > 0 \). Then using (2.3.6) we can check that as \( n \to \infty \),

\[
np \left[ \left( \frac{X_1}{n}, \frac{Y_1}{n} \right) \in [0, (x, y)]^c \right] = np(X_1 \leq U_1(nx), Y_1 \leq U_2(ny)) \\
= n(1 - F(U_1(nx), U_2(ny))) \\
\to - \log G_0(x, y) =: \nu_0([0, (x, y)]^c).
\]  

(2.4.1)

\( \nu_0 \) is called the exponent measure. Let us introduce some basics to see what kind of measure, what space, etc. we will be talking about next.

**Remark 2.4.1.** The measure \( \nu_0 \) defines a Radon measure on \( E := [0, \infty]^2 \setminus \{(0, 0)\} \).

For all practical purposes a Radon measure means that the measure is finite on compact sets of the space. We will prove (with hints) the statement in the remark in this section. Let us discuss what the topology is and how the measures we define behave in this topology.

### 2.4.1 One point uncompactification

Let \((T, \mathcal{T})\) be a topological space, \( T \) being the set and \( \mathcal{T} \) being the topology, that is, the collection of open subsets of \( T \), satisfying the following:

1. \( \emptyset \in \mathcal{T} \) and \( T \in \mathcal{T} \).
2. \( \mathcal{T} \) is closed under arbitrary union and finite intersection.

If \( D \subset T \), define \( T^# = T \cap D^c \) and give it the topology

\[
\mathcal{T}^# = \mathcal{T} \cap D^c.
\]

**Theorem 2.4.1.** If \( K(T) \) denotes the collection of compact subsets of \( T \) then the compact subsets of \( T^# \) are

\[
K(T^#) = \{ K \in K(T) : K \cap D = \emptyset \}.
\]

See (Resnick, 2007, Section 6.1.3) for further details and proofs. Now let us look at our space \( E = [0, \infty]^2 \setminus \{(0, 0)\} \).

1. \( E \) is a one-point uncompactification of the compact topological space \([0, \infty]^2\), with the origin, \((0, 0)\) removed.
2. \( E \) also inherits the topology of \([0, \infty]^2\). Clearly from the previous note, the compact sets of \( E \) are the compact sets of \([0, \infty]^2\) that are bound away from zero.
3. \( E \) also inherits the \( \sigma \)-field of \([0, \infty]^2\): \( B \) is measurable in \( E \) iff \( B \cup \{(0, 0)\} \) is measurable in \([0, \infty]^2\). This is the Borel \( \sigma \)-field, \( \mathcal{B}(E) \) which is generated by all compact subsets of \( E \).
4. Given two Radon measures \( \nu_1 \) and \( \nu_2 \) such that

\[
\nu_1([0, (x, y)]^c) = \nu_2([0, (x, y)]^c), \quad (x, y) \in E,
\]

we have \( \nu_1 \equiv \nu_2 \).

To show this, first we show that they are equal on rectangles contained in \( E \) (easy). The rectangles form a \( \pi \)-system, hence using Dynkin’s \( \pi \)-\( \lambda \) theorem they must equal on the entire Borel \( \sigma \)-field.
5. Now we define a measure on $E$ as follows:

$$\nu_0([0, (x, y)]^c) := -\log G_0(x, y).$$

This we have already done in (2.4.1). Now this can be easily extended to $B(E)$. First we extend it to the rectangles contained in $E$. Then we apply Carathéodory’s extension theorem to define $\nu_0$ on all Borel subsets of $E$. Thus we have shown the existence of $\nu_0$. We need to check $\nu_0(A) \geq 0$ for all $A \in B(E)$ which can be done through checking rectangles (use max-id property).

6. $\nu_0$ is Radon. Take any compact subset $K \subset E$. Then there exists $\delta > 0$ such that $D \subset [0, (\delta, \delta)]^c$. Therefore,

$$\nu_0(K) \leq \nu_0([0, (\delta, \delta)]^c) \leq \nu_0([0, \infty] \times (\delta, \infty]) + \nu_0((\delta, \infty] \times [0, \infty])$$

$$= -\log(G_{01}(\delta)) + -\log(G_{02}(\delta)) = -2\log(\exp(-1/\delta)) = 2/\delta < \infty.$$

7. Observe now that

$$G_0(x, y) = e^{-\nu_0([0, (x, y)]^c)}, \forall x, y > 0.$$ 

Hence $\nu_0$ is called the exponent measure. Characterizing all $G_0$’s now boil down to characterizing all $\nu_0$’s as defined above.

2.4.2 Some properties of $\nu_0$

1. $\nu_0$ is the unique Radon measure on $E$ such that

$$G_0(x, y) = e^{-\nu_0([0, (x, y)]^c)}, \forall x, y > 0.$$ 

This follows from property 5 for Radon measures in the previous subsection.

2. For any Borel set $A \subset E$, we have the following homogeneity property or scaling property:

$$\nu_0(cA) = c^{-1} \nu_0(A), \quad \forall c > 0,$$

(2.4.2)

where $cA = \{cx : x \in A\}, c > 0$.

Proof. It is enough to prove this for sets of the form $A = [0, (x, y)]^c$. (and refer to Dynkin’s $\pi$ - $\lambda$).

$$\nu_0(cA) = \lim_{t \to \infty} tP\left([X^*_t, Y^*_t] / t \in [0, (cx, cy)]^c\right)$$

$$= \lim_{t \to \infty} c^{-1} tP\left([X^*_t / c, Y^*_t / c] \in [0, (x, y)]^c\right)$$

$$= c^{-1} \nu_0(A).$$

3. $\nu_0$ puts 0-mass on lines through infinity.

$$\nu_0\left([\{\infty\} \times [0, \infty]] \cup ([0, \infty] \times \{\infty\})\right) = 0.$$ 

This easily follows from property 2.
CHAPTER 2. MULTIVARIATE EXTREME VALUE THEORY

4. If either \( x = 0 \) or \( y = 0 \), then \( \nu_0([0, (x, y)]^c) = \infty \). This follows from property 1.

Remark 2.4.2. A few interesting things to note about the homogeneity property:

1. Take a set \( A \in B(\mathbb{E}) \). \( k \times A \) has \( \frac{1}{k} \) times the \( \nu_0 \) measure of \( A \). As we move away from 0, the measures of the sets become sparser.
2. \( \nu_0(cA) = c^{-1}\nu_0(A) \) comes from the fact that the marginals are all standard Fréchet.
3. We can empirically estimate \( \nu_0(A) \) and calculate \( \nu_0(cA) \) for sets \( cA \) which has empirical measure zero.

Theorem 2.4.2. \( G_0 \) is a bivariate EVD with standard Fréchet type marginals if and only if there exists a unique Radon measure \( \nu_0 \) on \( \mathbb{E} \) satisfying (2.4.2) such that

\[
G_0(x, y) = \exp\{-\nu_0([0, (x, y)]^c)\}, \quad \forall x, y > 0.
\] (2.4.3)

Proof. We have seen that if \( G_0 \) is a bivariate EVD then (2.4.3) holds. Now suppose (2.4.3) holds along with the homogeneity property (2.4.2). We need to show.

1. \( G_0 \) is a bivariate c.d.f. Here the important thing to show is that every set has non-negative probability under \( G_0 \). The rest are easy. So if we show this for rectangle, that is good (using Caratheodary’s extension theorem). Consider the rectangle \((a, b] \times (c, d)\) in \( \mathbb{E} \). We need to show

\[
G_0(c, d) - G_0(c, b) - G_0(a, d) + G_0(a, b) \geq 0.
\]

This can be done. Alternatively we can go through point process techniques which becomes easier.

2. Next we show that \( G_0 \) is a bivariate EVD for any Radon measure \( \nu_0 \). Note that from homogeneity property (2.4.2) we have

\[
G_0^n(nx, ny) = G_0(x, y). \quad \forall n \geq 1.
\]

This implies \( G_0 \) is max-stable and \( G_0 \in D(G_0) \). Hence \( G_0 \) is a bivariate EVD.

3. Finally we show that \( G_0 \) has standard Fréchet type marginals. Check that

\[
G_{01}(x) = G_0(x, \infty) = e^{-\nu_0([0, (x, \infty)]^c)} = e^{-\nu_0(x \times [0, (1, \infty)]^c)} = e^{-x^{-1}\nu_0([0, (1, \infty)]^c)} = e^{-C/x},
\]

where \( C = \nu_0([0, (1, \infty)]^c) > 0 \). We can similarly show that \( G_{02}(y) = e^{-D/y} \) for some constant \( D > 0 \). Hence \( G_0 \) has standard Fréchet type marginals.

\[\square\]

2.5 Spectral measure

The fun in having characterized bivariate EVDs in terms of exponent measures is that now a (pseudo-)polar co-ordinate transformation of the co-ordinates would lead to a (spectral) decomposition of the co-ordinates. Let us explain further. We have a Radon measure \( \nu_0 \) defined on \( \mathbb{E} \) which has the nice homogeneity property (2.4.2).

Consider any norm \( || \cdot || : \mathbb{R}^d \mapsto [0, \infty) \). Recall, a function \( || \cdot || : \mathbb{R}^d \mapsto [0, \infty) \) is a norm if:

1. \( ||x|| = 0 \) if and only if \( x = 0 \).
2.5. SPECTRAL MEASURE

2. ||c|| = |c||x||, \forall x \in \mathbb{R}^d, c \in \mathbb{R}.

3. ||x + y|| \leq ||x|| + ||y|| (Triangle inequality).

We want to define a polar co-ordinate transformation for \( x \in \mathbb{R}^2 \):

\[ x \mapsto (||x||, \frac{x}{||x||}) =: (r, a). \]

This obviously causes difficulties when \( ||x|| = 0 \). For \( d=2 \) consider \( x = (x_1, x_2) \) and we define \( T : \mathbb{R}^2 \setminus \{(0, 0)\} \mapsto (0, \infty) \times \mathbb{R}_+ \) as

\[ T(x) = \left( ||(x_1, x_2)||, \frac{(x_1, x_2)}{||(x_1, x_2)||} \right) = (r, a) \text{ (say)}. \]

where \( \mathbb{R} = \{(x_1, x_2) : ||(x_1, x_2)|| = 1\} \), is the unit sphere in \( \mathbb{R}^2 \) with respect to the norm \(|\cdot||\). Naturally \( T, T^- \) are continuous bijections now with \( 0 \) excluded. One problem is that polar co-ordinate transformation is not defined on the lines through \( \infty \), whereas the exponent measure is. We need some restriction arguments to get around this. We now consider the mapping \( T \) restricted to \([0, \infty)^2 \setminus \{(0, 0)\} \mapsto (0, \infty) \times \mathbb{R}_+ \) where \( \mathbb{R}_+ = \mathbb{R} \cup [0, \infty)^2 \). Therefore

\[ \nu_0 \{ x : ||x|| > r, \frac{x}{||x||} \in \Lambda \} = \nu_0 \{ x : ||r^{-1}x|| > 1, \frac{r^{-1}x}{||r^{-1}x||} \in \Lambda \} = \nu_0 \{ ry : ||y|| > 1, \frac{y}{||y||} \in \Lambda \} =: r^{-1} \Psi(\Lambda) \quad \text{(homogeneity property)} \]

where \( \nu_1(r, \infty) = r^{-1}, r > 0 \). therefore we have \( \nu_0 \circ T^{-1} = \nu_1 \times \Psi \). Observe that

1. \( \Psi \) is a finite measure on subsets of \( \mathbb{R}_+ \).

2. The measure \( \Psi \) depends on the choice of the norm.

3.

\[ \nu_1((x, \infty)) = x^{-1} = \int_{(x, \infty)} r^{-2}dr, \quad x > 0 \]

which implies that

\[ \nu_1(A) = \int_A r^{-2}dr \quad \forall A \in \mathcal{B}((0, \infty)) \]

Remark 2.5.1. A few norms that are popularly used are:

1. \( L^2 \) norm: \( ||x|| = \sqrt{(x_1^2 + x_2^2)} \).

2. \( L^1 \) norm: \( ||x|| = x_1 + x_2 \).

3. \( L^\infty \) norm: \( ||x|| = x_1 \vee x_2 \).
CHAPTER 2. MULTIVARIATE EXTREME VALUE THEORY

Let us derive \( \nu_0 \) in terms of \( \Psi \) for the kinds of sets we have been looking at.

\[
- \log G_0(x) = \nu_0([0, x]^c) = \iint_{\{(r, a) : r > 0, a \in \mathbb{R}_+ \text{ or } r_{a_1} > x_1 \text{ or } r_{a_2} > x_2 \}} r^{-2} \, dr \Psi(da)
\]

\[
= \int_{a \in \mathbb{R}_+} \left[ \int_{r > \frac{a_1}{x_1}} r^{-2} \, dr \right] \Psi(da)
\]

\[
= \int_{a \in \mathbb{R}_+} \left( \frac{a_1}{x_1} \vee \frac{a_2}{x_2} \right) \Psi(da).
\]

Now observe that if the marginals of \( G_0 \) are \( \Phi_1 \), then

\[
\exp\{ -1/x_1 \} = \exp\{ -1 \int_{a \in \mathbb{R}_+} a_1 \Psi(da) \}.
\]

Hence we have

\[
1 = \int_{a \in \mathbb{R}_+} a_1 \Psi(da) = \int_{a \in \mathbb{R}_+} a_2 \Psi(da).
\]

Suppose now our choice of norm is the Euclidean norm: \( ||x|| = \sqrt{x_1^2 + x_2^2} \). Then

\[
\{ (\cos \theta, \sin \theta) : 0 \leq \theta \leq \frac{\pi}{2} \}.
\]

So we can define the spectral measure in terms of \( \theta \):

\[
\Psi([0, \theta]) = \Psi(\{ a \in \mathbb{R} : 0 \leq \tan^{-1} \frac{a_2}{a_1} \leq \theta \}).
\]

Hence

\[
G_0(x) = \exp\left\{ -\int_0^{\pi/2} \left( \frac{\cos \theta}{x_1} \vee \frac{\sin \theta}{x_2} \right) \Psi(d\theta) \right\}.
\]

Let us look at some examples:

**Example 2.5.1** (independence).

\[
G_0(x_1, x_2) = \Phi_1(x_1) \Phi_1(x_2)
\]

Then

\[
\nu_0([0, x]^c) = \frac{1}{x_1} + \frac{1}{x_2}.
\]

Naturally \( \nu_0(x_1, x_2) = 0 \). Put density \( x^{-2} \, dx \) on the axes. Then \( \Psi \) is the two point distribution on \((0, 1) \) and \((1, 0) \). This is known as asymptotic independence for any \( F \in D(G_0) \) where \( G_0(x_1, x_2) = \Phi_1(x_1) \Phi_1(x_2) \).

**Example 2.5.2** (total dependence).

Suppose \( \xi \) follows \( \Phi_1(x) \) and \((X, Y) = (\xi, \xi) \). Then

\[
P(X \leq x, Y \leq y) = P(\xi \leq x, \xi \leq y)
= \Phi(x \wedge y)
= e^{-\frac{1}{x} \vee \frac{1}{y}}.
\]

Hence \( \Psi \) concentrates on \((1, 1) \) on the unit sphere:

\[
\frac{1}{x} \vee \frac{1}{y} = \int_{a \in \mathbb{R}_+} \left( \frac{a_1}{x} \vee \frac{a_2}{y} \right) \Psi(da).
\]

\( \nu_0 \) concentrates density \( r^{-2} \, dr \) on the diagonal. So we have \( \Psi(\{ \pi/4 \}) = 1 \).
Example 2.5.3. Suppose we have a distribution
\[ G_0(x, y) = \exp\{-(x^{-2} + y^{-2})^{1/2}\}, \quad x > 0, y > 0. \]
Then
\[ \nu_0([0, (x, y)]^c) = \sqrt{x^{-2} + y^{-2}} = \frac{\sqrt{x^2 + y^2}}{xy}. \]
We want to find $\Psi$ or $\tilde{\Psi}$.


\[
\nu_0([0, (x, y)]^c) = \frac{\pi}{2} \int_0^{\pi/2} \cos \theta \, x \vee \sin \theta \, y \, d\theta
= \int_0^{\tan^{-1} \frac{y}{x}} \cos \theta \, d\theta + \int_{\tan^{-1} \frac{y}{x}}^{\frac{\pi}{2}} \frac{\sin \theta}{y} \, d\theta
\]
(since $\cos \theta \frac{y}{x} > \sin \theta \frac{y}{x}$ is equivalent to $\tan^{-1} \frac{y}{x} > \theta$)
\[
= \frac{1}{x} \sin(\tan^{-1} \frac{y}{x}) + \frac{1}{y} \cos(\tan^{-1} \frac{y}{x})
= \frac{\sqrt{x^2 + y^2}}{xy}.
\]

2. Method 2: Assume that a continuous density exists for $\nu_0$, that is, we have $\nu_0'$ such that
\[
\nu_0(A) = \iint_A \nu_0'(x, y) \, dx \, dy.
\]
Therefore it is easy to check that,
\[
\nu_0'(x, y) = -\frac{\partial}{\partial x} \frac{\partial}{\partial y} \nu_0([0, (x, y)]^c).
\]
\[
\nu_0'(x, y) := \frac{\partial \nu_0([0, (x, y)]^c)}{\partial x \partial y} = (\sqrt{x^2 + y^2})^{-3}.
\]
Changing to polar co-ordinates implies that the density is
\[
\nu_0'(x, y) \, dx \, dy = \nu_0'(r \cos \theta, r \sin \theta) \, r \, dr \, d\theta
= r^{-3} \, r \, dr \, d\theta.
\]
Therefore
\[
\tilde{\Psi}([0, \theta_0]) = \nu_0\{x : ||x|| > 1, \theta \leq \theta_0\}
= \iint_{r > 1, \theta \leq \theta_0} r^{-2} \, dr \, d\theta
= \theta_0.
\]
Example 2.5.4. This time we work with the $L^1$ norm. So $\mathbb{N}_+ = \{(w, 1 - w) : 0 \leq w \leq 1\}$. Recall the marginal conditions

$$1 = \int_{\mathbb{N}_+} a_1 \Psi(da) = 1, i = 1, 2.$$ 

Define $H(w) = \overline{\psi}([0, \sqrt{2}w]) = \psi\{(v, 1 - v) : 0 \leq v \leq w\}$. Therefore

$$1 = \int_0^1 w \overline{\psi}([0, \sqrt{2}w]) = \int_0^1 wH(dw) = \int_0^1 (1 - w)H(dw) = H(1) - \int_0^1 wH(dw).$$

Therefore

$$\int_0^1 wH(dw) = \frac{H(1)}{2} = 1, \text{ i.e., } H(1) = 2.$$ 

If we put $H(dw) = 2dw$ it is easy to check that

$$\nu_0([0, (x, y)]^c) = \frac{1}{x} + \frac{1}{y} - \frac{1}{x + y}.$$ 

Note that a Uniform distribution under the $L^1$-norm gives a different bivariate extreme value distribution than the one under $L^2$-norm.

2.6 Asymptotic Independence

Clearly the dependence among the components of a multivariate extreme value distribution is an important issue under consideration. Under independence any distribution $F$ is the product of its marginals. Recall we still have $d = 2$. We are concerned with the limit distribution $G$ being independent here where $F \in D(G)$ and $G$ has marginals $G_1$ and $G_2$. A bivariate distribution function $F$ is said to be asymptotically independent if $F \in D(G)$ and $G(x_1, x_2) = G_1(x_1)G_2(x_2), \forall x_1, x_2 \in \mathbb{R}$. We say random variables $X_1$ and $X_2$ are asymptotically independent if $F$ is asymptotically independent where $(X_1, X_2) \sim F$. The following theorem characterizes asymptotic independence in bivariate extreme value distributions.

**Theorem 2.6.1.** Suppose $(X, Y) \sim F$ with marginals $F_i \in D(G_i)$ for $i = 1, 2$ with respective standardizations $a_n > 0, b_n \in \mathbb{R}$ and $c_n > 0, d_n \in \mathbb{R}$. Also let $X^* = 1/(1 - F_1(X))$ and $Y^* = 1/(1 - F_2(Y))$. Then the following are equivalent

1. $F \in D(G)$ where $G(x) = G_1(x_1)G_2(x_2)$
2. $\lim_{t \to \infty} tP(X^* > t, Y^* > t) = 0$
3. $\lim_{n \to \infty} nP(X > a_nx + b_n, Y > c_ny + d_n) = 0$.

**Proof.** We provide hints to the proof of the Theorem.

$2 \leftrightarrow 3$: This can be shown using Proposition 2.3.1 on standardization of the marginals. It is left as an exercise to the reader.
2.6. ASYMPTOTIC INDEPENDENCE

1 → 2: Using Proposition 2.3.1, as \( t \to \infty \)

\[
\begin{align*}
  t \mathbb{P}(X^* > t, Y^* > t) &= t \mathbb{P}(X > U_1(t), Y > U_2(t)) \\
  &\to \nu_0([x, \infty] \times [y, \infty]) = 0.
\end{align*}
\]

The last equality is justified as follows. Since \( G_0(x, y) = \exp \left\{ - \left( \frac{\xi}{x} + \frac{\eta}{y} \right) \right\} \) for \( x, y > 0 \), thus \( \nu_0([0, (x, y)]^c) = \frac{1}{x} + \frac{1}{y} \). Therefore \( \nu_0((x, \infty] \times (y, \infty]) = 0 \).

2 → 1: We show that

\[
\nu_0([0, (x, y)]^c) = \frac{1}{x} + \frac{1}{y}, \quad x, y > 0
\]

which is equivalent to \( G(x, y) = G_1(x)G_2(y) \) (check). We know that

\[
\begin{align*}
  t \mathbb{P}\left( \left( \frac{X^*}{t}, \frac{Y^*}{t} \right) \in [0, (x, y)]^c \right) &\to \nu_0([0, (x, y)]^c).
\end{align*}
\]

Also note that

\[
\begin{align*}
  t \mathbb{P}\left( \left( \frac{X^*}{t}, \frac{Y^*}{t} \right) \in [0, (x, y)]^c \right) &= t \mathbb{P}(X^* > tx) + t \mathbb{P}(Y^* > ty) - t \mathbb{P}(X^* > tx, Y^* > ty) \\
  &\to \frac{1}{x} + \frac{1}{y} - 0.
\end{align*}
\]

Clearly \( t \mathbb{P}(X^* > tx, Y^* > ty) \leq t \mathbb{P}(X^* > tx \wedge y), Y^* > ty \wedge y) \leq (x \wedge y)^{-1} t^{-1} \mathbb{P}(X^* > s, Y^* > s) \to 0 \).

\[\Box\]

**Example 2.6.1.** Suppose \( \xi \sim U(0, 1) \). Define

\[
(X^*, Y^*) = \left( \frac{1}{\xi}, \frac{1}{1-\xi} \right).
\]

Clearly \( X^*, Y^* \) are both Pareto(1) and \( U_1(t) = U_2(t) = t \). Note that

\[
\begin{align*}
  t \mathbb{P}\left( \frac{1}{\xi} > t, \frac{1}{1-\xi} > t \right) &= t \mathbb{P}\left[ 1 - \frac{1}{t} < \xi < \frac{1}{t} \right] \\
  &= 0, \quad \text{if } t \geq 2.
\end{align*}
\]

Hence \( X^* \) and \( Y^* \) are asymptotically independent but they are not independent.

**Example 2.6.2.** Suppose \((X, Y) \sim N_2(0, \Sigma)\) where \( \text{Var}(X) = \text{Var}(Y) = 1, \text{Cov}(X, Y) = \rho < 1 \). Then \( X \) and \( Y \) are asymptotically independent. In fact any \( X \) which is multivariate normally distributed with all cross-correlation less than 1 is asymptotically independent over the co-ordinates.

We want to show that

\[
\lim_{t \to \infty} t \mathbb{P}(U_1^-(X) > t, U_2^-(Y) > t) = 0,
\]

or, alternatively,

\[
\lim_{t \to \infty} t \mathbb{P}(X > U_1(t), Y > U_2(t)) = 0.
\]
Note that
\[ U_1(t) = U_2(t) = \left( \frac{1}{1 - \Phi} \right)^{-1}(t) = \Phi^{-1}(t) \uparrow \infty \text{ as } t \to \infty. \]

Therefore we can show as \( s \to \infty \)
\[ P(X > s | Y > s) = \frac{P(X > s, Y > s)}{P(X > s)} \]
\[ = \frac{P(X > U_1(t), Y > U_2(t))}{P(Y > U_2(t))} \quad \text{with } \Phi(s) = 1 - \frac{1}{t} \]
\[ = tP(X > U_1(t), Y > U_2(t)) \to 0, \]
and that suffices. For \( \rho = -1 \), clearly \( X = -Y \) almost surely and hence \( P(X > s | Y > s) = 0 \) for all \( s > 0 \).

So assume now \( |\rho| < 1 \). Check that \( \frac{X + Y}{2} \overset{d}{=} cZ \) where \( 0 < c = \sqrt{\frac{1 + \rho}{2}} \) and \( Z \sim N(0,1) \). Now
\[ \lim_{s \to \infty} \frac{P(X > s | Y > s)}{P(X > s)} = \lim_{s \to \infty} \frac{P(X > s, Y > s)}{P(X > s)} \]
\[ \leq \lim_{s \to \infty} \frac{P(X + Y > 2s)}{P(X > s)} \]
\[ = \lim_{s \to \infty} \frac{P(Z > s/c)}{P(Z > s)} \]
\[ = \lim_{s \to \infty} \frac{1 - \Phi(s/c)}{1 - \Phi(s)} \]
\[ = \lim_{s \to \infty} \frac{\phi(s/c)}{c\phi(s)} \quad \text{(L'Hopital)} \]
\[ = \lim_{s \to \infty} \frac{1}{c} \exp\left\{ -\frac{s^2}{2} (d^2 - 1) \right\} \quad \text{where } d^2 = \frac{1}{c^2} > 1 \]
\[ = 0. \]

Remark 2.6.1. We can mention a couple of more statements equivalent to any of the statements (1)-(3) in Theorem 2.6.1.

4. The exponent measure \( \nu_0 \) concentrates on \( L = L_1 \cup L_2 \) where \( L_1 = (0,\infty] \times \{(0,0)\} \) and \( L_2 = \{(0,0)\} \times (0,\infty] \).

5. The spectral measure \( S \) on \( \mathbb{R}_+ \) concentrates on \( \{e_1, e_2\} \) where \( e_i = L_i \cap \mathbb{R}_+ \), \( i = 1, 2 \).

Remark 2.6.2. For \( d \geq 2 \), asymptotic independence is equivalent to pairwise asymptotic independence of all possible pair of components.

2.7 Vague convergence

Suppose \( E \) is a locally compact topological space with countable base. We can safely think of \( E = \mathbb{R}^d \) or \( \overline{\mathbb{R}^d} \) or subsets thereof. When we will talk about point processes we will see that our points live here. Let \( \mathcal{E} \) be the Borel \( \sigma \)-field of \( E \).

A measure \( \mu : \mathcal{E} \to [0,\infty] \) is a set function such that

1. \( \mu(\emptyset) = 0 \) and \( \mu(A) \geq 0 \) for \( A \in \mathcal{E} \).
2.7. VAGUE CONVERGENCE

If \{A_n, n \geq 1\} are mutually disjoint sets in \(\mathcal{E}\), then

\[
\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad (\sigma\text{-additivity})
\]

\(\mu\) is Radon if \(\mu K < \infty\) for all \(K\), compact in \(\mathcal{E}\). Now denote

\[M_+ (\mathcal{E}) = \{ \mu : \mu \text{ is a non-negative Radon measure on } \mathcal{E}\}.
\]

\(M_+ (\mathcal{E})\) can be made into a complete separable metric space with respect to the vague metric.

Now recall from weak convergence of probability measures (Billingsley, 1968), that we considered a class of continuous and bounded test functions \(f\) and if

\[P_n(f) := \int f(x) \mu_n(dx) \to \int f(x) \mu(dx) =: P(f),\]

then we said \(P_n\) converges weakly to \(P\). Note that the measures in \(M_+ (\mathcal{E})\) are potentially infinite. So if we want to follow the same route we need functions on compact support, where the measures are finite. So define

\[C^+_K (\mathcal{E}) := \{ f : \mathcal{E} \to [0, \infty) : f \text{ is continuous with compact support}\}.
\]

Let \(\mu_n \in M_+ (\mathcal{E}), \forall n \geq 0\). Then \(\mu_n\) converges to \(\mu_0\) vaguely if \(\forall f \in C^+_K (\mathcal{E})\),

\[\mu_n(f) := \int f(x) \mu_n(dx) \to \int f(x) \mu_0(dx) =: \mu_0(f).
\]

Write \(\mu_n \overset{v}{\to} \mu_0\).

**Example 2.7.1.** Let \(\mathcal{E} \subset \mathbb{R}^d\) with metric \(d\). For \(x \in \mathcal{E}\) and \(A \in \mathcal{E}\) define

\[\varepsilon_x (A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}\]

Then \(\mu_n := \varepsilon_{x_n} \overset{v}{\to} \mu_0 := \varepsilon_{x_0}\) in \(M_+ (\mathcal{E})\) if and only if \(\langle x_n, x_0 \rangle \to 0\) as \(n \to \infty\).

Suppose \(d(x_n, x_0) \to 0\) then for any \(f \in C^+_K (\mathcal{E})\),

\[\mu_n(f) = f(x_n) \to f(x_0) = \mu(f_0),\]

by continuity of \(f\).

Now suppose \(d(x_n, x_0) \to 0\). Then there exists \(\epsilon > 0\) and a subsequence \(\{n'\} \subset \{n\}\) s.t. \(d(x_{n'}, x) > \epsilon\) along the subsequence. Now define the function \(\phi\)

\[\phi(t) = \begin{cases} 1 & t < 0, \\ 1 - t & 0 \leq t \leq 1, \\ 0 & t \geq 1. \end{cases}\]

Now we define the function \(f_\epsilon(y) = \phi\left(\frac{d(x_0, y)}{\epsilon}\right)\). Clearly \(f \in C^+_K (\mathcal{E})\). Then

\[|f(x_n) - f(x_0)| = |\phi\left(\frac{d(x_0, x_{n'})}{\epsilon}\right) - \phi(0)| = 1 \to 0.
\]

We can define a point measure \(m\), which is an element of \(M_+ (\mathcal{E})\), s.t.

\[m = \sum_{i=1}^{\infty} \varepsilon_{x_i},\]

where \(x_i \in \mathcal{E}\) and \(m\) is Radon. This class of measures denoted by \(M_p(\mathcal{E}) \subset M_+ (\mathcal{E})\) will be further discussed in the following chapter on point processes.
**Topology on** $M_+(E)$

Since we will be working with compact sets and measures on $E$ we need to know the topology of $E$. The basic neighbourhoods of $\mu_0$ can be seen as either of the following:

1. $\{\mu : |\mu(f_i) - \mu_0(f_i)| < \epsilon_i, i = 1, \ldots, k\}$ for $\epsilon_i > 0, f_i \in C^+_K(E)$.
2. $\{\mu : |\mu(G_i) - \mu_0(G_i)| < \epsilon_i, i = 1, \ldots, k\}$ for $\epsilon_i > 0, G_i$ open and relatively compact.

The topology on $M_+(E)$ is the smallest topology making the maps $\mu \mapsto \mu(f)$ continuous from $M_+(E) \to \mathbb{R}_+$ for all $f \in C^+_K(E)$.

We denote by $\mathcal{M}(E)$, the $\sigma$-algebra on $M_+(E)$, the smallest $\sigma$-algebra of subsets of $M_+(E)$ that make the maps $m \mapsto m(f) = \int_E f dm$ from $M_+(E) \to \mathbb{R}$ measurable $\forall f \in C^+_K(E)$. In fact

$$\mathcal{M}_+(E) = \sigma\{\{m \in M_+(E) : m(f) \in B\}, f \in C^+_K(E), B \in \mathcal{B}([0, \infty])\}.$$ 

and also

$$\mathcal{M}_+(E) = \mathcal{B}(M_+(E)).$$

**Portmanteau Theorem**

Note that we sometimes use integrals of functions and sometimes measures of relatively compact sets while discussing vague convergence. The following variant of Urysohn’s Lemma actually leads us to the result, which we call the Portmanteau Theorem.

**Lemma 2.7.1.** 1. Let $K$ be compact in $E$. Then $\exists K_n, n \geq 1$ compact sets in $E$ with $K_n \uparrow K$ and $\exists f_n \in C^+_K(E)$ with $f_n \downarrow$ such that

$$1_K \leq f_n \leq 1_{K_n} \downarrow 1_K.$$ 

2. Let $G$ be open and relatively compact in $E$. Then $\exists G_n, n \geq 1$ open, relatively compact sets in $E$ with $G_n \uparrow G$ and $\exists f_n \in C^+_K(E)$ with $f_n \uparrow$ such that

$$1_G \geq f_n \geq 1_{G_n} \uparrow 1_G.$$ 

**Theorem 2.7.2** (Portmanteau Theorem). Let $\mu_n \in M_+(E), n \geq 0$. Then the following are equivalent:

1. $\mu_n \xrightarrow{v} \mu_0$.
2. $\mu_n(B) \to \mu_0(B)$ for all relatively compact $B$ for which $\mu(\partial B) = 0$.
3. For all compact $K$ and for all open relatively compact $G$, we have

$$\limsup_{n \to \infty} \mu_n(K) \leq \mu_0(K),$$

$$\liminf_{n \to \infty} \mu_n(G) \geq \mu_0(G).$$

**Proof.** Hints:

1 $\to$ 3: Suppose $K$ is compact. Then from Lemma 2.7.1, there exists $K_n$ compact and $f_n \in C^+_K(E)$ non-increasing such that

$$1_K \leq f_n \leq 1_{K_n} \downarrow K.$$
2.7. VAGUE CONVERGENCE

Fix an integer \( m \), then

\[
\limsup_{n \to \infty} \mu_n(K) \leq \limsup_{n \to \infty} \mu_n(f_m) \to \mu(f_m) \quad (n \to \infty).
\]

Since \( f_m \leq 1_{K_{m_0}}, \forall m \geq m_0 \), we can use DCT to see that

\[
\mu_0(f_m) \to \mu_0(K) \quad (m \to \infty).
\]

We can similarly show \( \liminf_{n \to \infty} \mu_n(G) \geq \mu_0(G) \) for any relatively compact open \( G \).

3 \to 2: Let \( B \) be relatively compact with \( \mu(\partial B) = 0 \), i.e., with \( B^0 \subset B \subset \overline{B} \),

\[
\mu(B^0) = \mu(B) = \mu(\overline{B}).
\]

Then

\[
\mu_0(B) = \mu_0(B^0) \leq \liminf \mu_n(B^0) \leq \liminf \mu_n(B) \leq \limsup \mu_n(B) \leq \mu_0(B) = \mu_0(B).
\]

2 \to 1 too technical, check (Resnick, 2008, Chapter 3.4).

\[\square\]

2.7.1 Vague convergence in Extreme Value Theory

**Theorem 2.7.3.** Suppose \( X_1 \) is a non-negative random variable with distribution function \( F \). Denote \( F(x) = 1 - F(x) \). Then the following are equivalent:

1. \( F \in RV_{-\alpha}, \alpha > 0 \).

2. There exists a sequence \( \{b_n\} \) with \( b_n \to \infty \) such that

\[
\lim_{n \to \infty} nF(b_n) = x^{-\alpha}, x > 0, \alpha > 0.
\]

3. There exists a sequence \( \{b_n\} \) with \( b_n \to \infty \) such that

\[
\mu_n(\cdot) := nP\left\{ \frac{X_1}{b_n} \in \cdot \right\} \to \nu_\alpha(\cdot)
\]

in \( M_+(0, \infty] \) where \( \nu_\alpha(x, \infty] = x^{-\alpha} \).

4. \( F \in D(\Phi_\alpha) \).

**Proof.** We provide some hints here.

1 \leftrightarrow 4: We have already done this in Theorem 1.2.1.

1 \leftrightarrow 2: Easy to check. Recall that we can choose \( b_n = b(n) = F^{-1}(1 - \frac{1}{n}) \).

3 \rightarrow 2: Trivial.
2 → 3: Let $f \in C^+_K((0, \infty])$. Now

$$
\mu_n(f) := nE f\left(\frac{X_1}{b_n}\right) = \int_0^\infty f(x) n P\left[\frac{X_1}{b_n} \in dx\right].
$$

Since $f$ has compact support, $\text{supp}(f) \subset (\delta, \infty]$ for some $\delta > 0$. On $(\delta, \infty]$ define

$$
P_n(\cdot) = \frac{\mu_n(\cdot)}{\mu_n(\delta, \infty]}.
$$

For $y \in (\delta, \infty]$, $P_n(\cdot)$ is a probability measure. Therefore $\forall y > \delta$

$$
P_n(y, \infty] = \frac{\mu_n(y, \infty]}{\mu_n(\delta, \infty]}
\xrightarrow{y^{-\alpha}} \frac{\nu_\alpha(y, \infty]}{\nu_\alpha(\delta, \infty]} = y^{-\alpha} P(y, \infty].
$$

Therefore $P_n \Rightarrow P$ on $(\delta, \infty]$. Thus $P_n(f) \to P(f)$. Now it is easy to see that

$$
\mu_n(f) = \int_0^\infty f(x) \mu_n(dx) = \mu_n(\delta, \infty] \int_0^\infty f(x) \frac{\mu_n(dx)}{\mu_n(\delta, \infty]}
\xrightarrow{\delta^{-\alpha}} \mu_n(\delta, \infty] P(f)
\xrightarrow{\delta^{-\alpha}} \nu_\alpha(\delta, \infty] \int_0^\infty f(x) \frac{\nu_\alpha(dx)}{\nu_\alpha(\delta, \infty]} = \nu_\alpha(f).
$$

This is true $\forall f \in C^+_K((0, \infty])$. Hence $\mu_n \xrightarrow{v} \nu_\alpha$. 

\[\square\]
Chapter 3

Point processes

Point process techniques have been used quite extensively in extreme value theory because of the host of
tools that can be used to prove various asymptotic results with elegance and relative ease. A point process
is a random distribution of points in a space. Usually this space is taken to be $\mathbb{R}, \mathbb{R}^d$ (their compactified
versions) or subsets thereof. What kinds of distributions of points are we interested in? A couple of examples
follows hence:

1. Arrival and departure times from a queue.
2. Location of trees in a forest.
3. Location of tanks in a battlefield.
4. Position and times of earthquakes/ hurricanes in the next 100 years.

3.1 Basics of point processes

Let us start with a standard set up. Let $\mathbb{E}$ be a 'nice' topological space. By this we mean that $\mathbb{E}$ is a
Hausdorff space which is locally compact and with countable base (alternatively, 'second countable'). This
is the assumption we had taken in earlier chapters too. For us most of the time $\mathbb{E}$ will be a subset of the
compactified Euclidean space of finite dimension, i.e., $\mathbb{E} \subseteq [-\infty, \infty]^d$. Also denote by $\mathcal{E}$ the Borel $\sigma$-algebra
on $\mathbb{E}$. Now define the measure $\varepsilon_x$ on $\mathcal{E}$ by

$$
\varepsilon_x(A) = \begin{cases} 
1, & x \in A \\
0, & x \notin A
\end{cases}, \quad A \in \mathcal{E}.
$$

A point measure on $\mathbb{E}$ is a measure $m$ of the following form: Let $\{x_i, i \geq 1\}$ be a countable collection of
points of $\mathbb{E}$ (which are not necessarily distinct). Then

$$m = \sum_{i=1}^{\infty} \varepsilon_{x_i},$$

and for $K \in \mathcal{E}$ compact we have $m(K) < \infty$. This means $m$ is Radon, it can take the value $\infty$, but NOT on
a compact set. Now let us define the following:

$$M_p(\mathbb{E}) := \{\mu : \mu \text{ is a point measure in } \mathbb{E}\}$$

$$\mathcal{M}_p(\mathbb{E}) := \sigma\text{-algebra generated by subsets of } M_p(\mathbb{E}).$$

= smallest $\sigma$-algebra containing all sets of the form
Alternatively, $\mathcal{M}(\mathbb{E})$ is the smallest $\sigma$-algebra making the evaluation maps $m \mapsto m(F)$ from $M_p(\mathbb{E}) \to [0, \infty]$ measurable $\forall F \in \mathcal{E}$.

**Definition 3.1.1.** A point process on $\mathbb{E}$ is a measurable map, say $N$, from a probability space $(\Omega, A, \mathbf{P}) \to (M_p(\mathbb{E}), \mathcal{M}_p(\mathbb{E}))$.

Essentially $N$ is a random element of $M_p(\mathbb{E})$. The probability law of $N$ is given by $\mathbf{P}_N$ where

$$\mathbf{P}_N = \mathbf{P} \circ N^{-1} = \mathbf{P}(N \in \cdot) \text{ on } \mathcal{M}_p(\mathbb{E}).$$

**Remark 3.1.1.** Let $S_m = \{x \in \mathbb{E} : m(\{x\}) \neq 0\}$, which is the set of points charged by $m$. This is the support of $m$, i.e., the smallest closed set $F$ such that $m(F^c) = 0$. $m(x)$ is the multiplicity of $x$. $m$ is called simple if $m(\{x\}) \leq 1$, $\forall x \in \mathbb{E}$. For our purposes we will be dealing mostly with simple point processes throughout this course.

**Remark 3.1.2.** A simple way to think about the point process $N$ is as follows: A point process on $\mathbb{E}$ is a map $N : \Omega \times \mathcal{E} \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ s.t.

$$N(\omega, A) = \sum_{i \in I} \varepsilon_{X_i(\omega)}(A),$$

for a countable index set $I$, and $\forall i \in I, X_i : \Omega \to \mathbb{E}$ is measurable with $\mathbf{P}(X_i(\omega) = X_j(\omega)) = 0, i \neq j$ and $N(\omega, K) < \infty$ for all $K$ compact in $\mathbb{E}$.

**Remark 3.1.3.** As a consequence of our definition of point process we have the following:

1. Fix $\omega \in \Omega$. Then $N(\omega, \cdot)$ is a point measure and $N(\omega, F)$ is the number of points in $F$ for the realization $\omega \in \Omega$.

2. Fix $A \in \mathcal{E}$. Then $\omega \mapsto N(\omega, A)$ is a measurable map from $(\Omega, A) \to ([0, \infty], \mathcal{B}([0, \infty]))$.

In fact we can state the following proposition in lieu of the fact that it often seems hard to check that $N$ is a point process just from its definition.

**Proposition 3.1.1.** $N$ is a point process if and only of the map $\omega \mapsto N(\omega, F)$ is measurable $(\Omega, A) \to ([0, \infty], \mathcal{B}([0, \infty])) \forall F \in \mathcal{E}$.

**Proof.** First assume that $N$ is a point process. Then $\omega \mapsto N(\omega, \cdot)$ is a measurable map from $(\Omega, A) \to (M_p(\mathbb{E}), \mathcal{M}_p(\mathbb{E}))$, by definition. Again by definition of $\mathcal{M}_p(\mathbb{E})$, $m \mapsto m(F)$ is a measurable map from $(M_p(\mathbb{E}), \mathcal{M}_p(\mathbb{E})) \to ([0, \infty], \mathcal{B}([0, \infty]))$. Since $\omega \mapsto N(\omega, F)$ is a composition of these two measurable maps it is measurable too.

Now assume that $\omega \mapsto N(\omega, F)$ is measurable $\forall F \in \mathcal{E}$. Define

$$\mathcal{G} = \{A \in \mathcal{M}(\mathbb{E}) : N^{-1}(A) \in A\} \subseteq \mathcal{M}(\mathbb{E}).$$

It is easy to check that $\mathcal{G}$ is a $\sigma$-algebra (check!!). Note that for $F \in \mathcal{E}$ and any $B \in \mathcal{B}([0, \infty])$,

$$N^{-1}\{m : m(F) \in B\} = \{\omega : N(\omega, F) \in B\} \in A.$$

Therefore $\mathcal{G}$ contains all sets of the form $\{m : m(F) \in B\}, B \in \mathcal{B}([0, \infty]), F \in \mathcal{E}$. Therefore

$$\mathcal{G} \subseteq \sigma\{\{m : m(F) \in B\}, B \in \mathcal{B}([0, \infty]), F \in \mathcal{E}\} = \mathcal{M}_p(\mathbb{E}).$$

\(\blacksquare\)
3.1. BASICS OF POINT PROCESSES

Recall that for stochastic processes, knowledge of all finite dimensional distributions is equivalent to knowing the distribution of the stochastic process. A similar idea is true for point processes. We state a result without proof. For further details see Resnick (2008).

**Proposition 3.1.2.** Let \( N \) be a point process on \( E \) defined on the probability space \((\Omega, \mathcal{A}, P)\). Then the law \( P_N \) of \( N \) is uniquely determined by the knowledge of

\[
P(N(A_1) = n_1, N(A_2) = n_2, \ldots, N(A_k) = n_k),
\]

for all \( A_i \in \mathcal{E}, n_i \in \{0, 1, 2, \ldots\} \cup \{\infty\}, 1 \leq i \leq k \) with \( k \geq 1 \).

In fact we can restrict the sets \( A_i \) to a \( \pi \)-system (collection of sets closed under finite intersection, see Billingsley (1995) for further details) and the same conclusion holds. For example the open intervals form a \( \pi \)-system in \( \mathbb{R} \). Now we can have a couple more definitions under this characterization.

**Definition 3.1.2** (Equality in law). Two point processes \( N_1 \) and \( N_2 \) on \( E \) (possibly defined on different probability spaces \((\Omega_1, \mathcal{A}_1)\) and \((\Omega_2, \mathcal{A}_2)\)) are same in law if for all \( k \geq 1 \),

\[
(N_1(A_1), N_1(A_2), \ldots, N_1(A_k)) \overset{d}{=} (N_2(A_1), N_2(A_2), \ldots, N_2(A_k)),
\]

\( \forall A_i \in \mathcal{E}, 1 \leq i \leq k \).

**Definition 3.1.3** (Independence). Let \((\Omega, \mathcal{A}, P)\) a probability space and \((\mathcal{E}_i, \mathcal{F}_i), i = 1, 2\) be state spaces. Suppose \( N_i : \Omega \to \mathcal{M}_P(E_i), i = 1, 2 \) are point processes. Then \( N_1 \) and \( N_2 \) are independent if the induced \( \sigma \)-algebras \( N_1^{-1}(\mathcal{M}_P(E_1)) \) and \( N_2^{-1}(\mathcal{M}_P(E_2)) \) are independent.

In fact we can show that \( N_1 \) and \( N_2 \) are independent if and only if \( (N_1(F_i); 1 \leq i \leq k) \) and \( (N_2(G_i; 1 \leq j \leq l)) \) are independent random vectors for any \( k, l \geq 1, F_i \in \mathcal{E}_1, G_i \in \mathcal{E}_2 \).

An important concept related to point processes which determines in some sense the rate of occurrence of points of the point process is the mean measure or intensity.

**Definition 3.1.4.** The intensity or mean measure of a point process \( N \) is the measure \( \mu \) defined on \( E \) as follows:

\[
\mu(F) := \mathbb{E}(N(F)) = \int \limits_{\Omega} N(\omega, F) P(d\omega) = \int \limits_{\mathcal{M}_P(E)} m(F) P_N(dm)
\]

for \( F \in \mathcal{E} \).

As an exercise check that \( \mu \) is indeed a measure. \( \mu \) may not be Radon. Try finding an example where \( \mu \) is not Radon. We will talk more about this soon.

Let \( f : (\mathbb{E}, \mathcal{E}) \to ([0, \infty], \mathcal{B}([0, \infty])) \) be a measurable map. Then define for all \( \omega \in \Omega \),

\[
N(\omega, f) := \int \limits_{\mathbb{E}} f(x) N(\omega, dx) \leq \infty.
\]

Note here that for \( A \in \mathcal{E} \) \( N(\omega, A) = N(\omega, 1_A) \) where \( \forall x \in \mathbb{E} \)

\[
1_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}
\]

**Proposition 3.1.3.** For any measurable \( f : (\mathbb{E}, \mathcal{E}) \to ([0, \infty], \mathcal{B}([0, \infty])) \), the map \( \omega \mapsto N(\omega, f) \) is a measurable map from \((\Omega, \mathcal{A}) \to ([0, \infty], \mathcal{B}([0, \infty]))\).

**Proof.** Clearly true for \( f = 1_A, A \in \mathcal{E} \). Can show for simple functions \( f = \sum_{i=1}^{k} c_i A_i \) for \( k \geq 1, A_i \in \mathcal{E}, c_i \geq 0 \). Now for any function \( f \) take \( f_n \) simple s.t. \( f_n \uparrow f \). By monotone convergence \( N(\omega, f_n) \uparrow N(\omega, f) \) and hence the statement. \( \square \)

A word on notation: We often use \( N(A) \) instead of \( N(\omega, A) \) and \( N(f) \) instead of \( N(\omega, f) \) as we similarly suppress the \( \omega \) for our random variables.

As an exercise check that \( \mu(f) := \mathbb{E}N(f) = \int \limits_{\mathbb{E}} f(x) \mu(dx) \).
CHAPTER 3. POINT PROCESSES

3.2 Laplace Functionals

Definition 3.2.1 (Laplace functional). Suppose $N$ is a point process on $E$. Then the Laplace functional $\psi_N$ is a map from all non-negative Borel functions $f : E \mapsto [0, \infty]$ to $[0, \infty)$ given by

$$\psi_N(f) := \mathbb{E}[\exp(-N(f))] \quad (3.2.1)$$

Remark 3.2.1. $\psi_N$ as defined above is the Laplace transform (see any general measure theoretic probability book for definition) of the distribution of the point process $N$.

$$\psi_N(f) := \mathbb{E}\left[\exp\left\{-N(f)\right\}\right]$$

$$= \int_{\Omega} \exp\left\{-\int_{E} f(x)m(dx)\right\} P_N(dm)$$

Proposition 3.2.1. The Laplace functional $\psi_N$ of $N$ uniquely determines the law of $N$. (Two point processes $N_1$ and $N_2$ have the same law iff they have the same Laplace functional, i.e., $\psi_{N_1}(f) = \psi_{N_2}(f), \forall$ measurable $f : E \mapsto [0, \infty]$).

Proof.

$$\psi_{N_1}(f) = \psi_{N_2}(f) \quad \forall \text{ measurable } f : E \mapsto [0, \infty].$$

Hence we have,

$$\psi_{N_1}(f) = \psi_{N_2}(f) \quad \forall \text{ simple measurable functions.}$$

Thus, $\forall k \geq 1, A_1, A_2, \ldots, A_k \in E$ and $\lambda_1, \ldots, \lambda_k \in [0, \infty)$

$$\psi_{N_1}(\sum_{i=1}^{k} \lambda_i 1_{A_i}) = \mathbb{E}[\exp\{-\sum_{i=1}^{k} \lambda_i N_1(A_i)\}] = \mathbb{E}[\exp\{-\sum_{i=1}^{k} \lambda_i N_2(A_i)\}] = \psi_{N_2}(\sum_{i=1}^{k} \lambda_i 1_{A_i}).$$

This means that the Laplace transforms of $(N_1(A_1), N_1(A_2), \ldots, N_1(A_k))$ and $(N_2(A_1), N_2(A_2), \ldots, N_2(A_k))$ are equal, which means that

$$(N_1(A_1), N_1(A_2), \ldots, N_1(A_k)) \overset{d}{=} (N_2(A_1), N_2(A_2), \ldots, N_2(A_k)), \forall A_i \in E, \lambda_i \geq 0, 1 \leq k, \forall k \geq 1.$$

Therefore the laws of $N_1$ and $N_2$ are equal.

Therefore it is enough to specify Laplace functionals of a point process when talking about its law/distribution. Laplace functionals are very useful in studying weak convergence of point processes. We will calculate some Laplace functionals as we study Poisson point processes. Also note that the moments of $N$ can also be determined from $\psi_N$. For example check (exercise) that for $f \geq 0,$ measurable

$$\mu(f) := \mathbb{E}N(f) = \int_{E} f(x)\mu(dx) = \lim_{t \downarrow 0} \frac{(1 - \psi_N(tf))}{t}.$$

To check this observe

$$\lim_{t \downarrow 0} \frac{1 - e^{-tN(f)}}{t} = N(f).$$
3.3. POISSON PROCESS OR POISSON RANDOM MEASURE

3.3 Poisson Process or Poisson Random Measure

Poisson processes are a very important class of point processes. They play a similar role for point processes as the role played by normal distributions for random variables. Suppose we want to model the location of trees in a forest. Let us see why a Poisson process would be a suitable candidate for finding the distribution of the trees. Let \( E \) denote the space of the forest. Let \( X_i, i \geq 1 \) denote the locations of the trees in the forest.

Suppose \( p_A \) be the probability that a tree, whose location is specified by \( X_i \), lies in \( A \). If we denote \( N(A) \) to be the number of trees in site \( A \), then we can write

\[
N(A) = \sum_{i \in X_i} A
\]

Suppose we know the location of a total of \( n \) trees in the forest, then considering the distribution of locations of trees being independent (we are trying to model naively, details will follow soon), we can see that

\[
\mathbb{P}(N(A) = k) = \binom{n}{k} p_A^k (1 - p_A)^{n-k}
\]

(Binomial with success rate \( p_A \))

\[
\approx \frac{(np_A)^k}{k!} \left(1 - \frac{np_A}{n}\right)^{n-k}
\]

(using Sterling’s formula, etc.)

\[
\approx \frac{\mu(A)}{k!} \exp^{-\mu(A)} \sim \text{Poisson}(\mu(A)),
\]

where \( \mu(A) \) denotes the average number of trees in the region \( A \) and \( \mu(A) \approx np_A \) for large \( n \). Thus the number of trees in \( A \) is distributed approximately as a Poisson random variable with parameter \( \mu(A) \).

We formalize this idea next in the form of a Poisson Process. First recall that a measure \( \mu \) on a space \((E, \mathcal{E})\) is \( \sigma \)-finite if \( E \) can be written as a countable union of sets each of which has a finite measure. For example the Lebesgue measure on \( \mathbb{R} \) or \( \mathbb{R}^d \) is a \( \sigma \)-finite measure.

**Definition 3.3.1 (Poisson Process).** Let \((E, \mathcal{E}, \mu)\) be a \( \sigma \)-finite measure space. A point process \( N \) defined on \((E, \mathcal{E})\) is a Poisson Process or a Poisson Random Measure with mean measure or intensity measure \( \mu \) if \( N \) satisfies:

1. For any \( F \in \mathcal{E}, k \geq 0 \), we have \( N(F) \sim \text{Poisson}(\mu(A)) \), i.e.,

\[
\mathbb{P}(N(F) = k) = \begin{cases} 
\exp\{-\mu(F)\}\frac{\mu(F)^k}{k!}, & \mu(F) < \infty, \\
0, & \mu(F) = \infty.
\end{cases}
\]

This means that \( \mu(F) = \infty \) implies \( \mathbb{P}(N(F) = \infty) = 1 \).

2. For any \( k \geq 1 \), if \( F_1, F_2, \ldots, F_k \) are disjoint in \( \mathcal{E} \), then \((N(F_1), N(F_2), \ldots, N(F_k))\) are independent random variables.

We assume \( E \) to be a ‘nice’ topological space. But Poisson Processes can be defined for more general spaces which we do not venture into for this course. Think \( E \subseteq [-\infty, \infty]^d \). \( \mu \) also need not be Radon, but we will have this for most of our examples, which enables us to use weak and vague convergence techniques. As an exercise one can try to find a \( \sigma \)-finite measure that is not Radon.

Notationally, if \( N \) is a Poisson Random Measure with mean measure \( \mu \) we write \( N \sim \text{PRM}(\mu) \).

**Remark 3.3.1.** Since we have a definition as above we would face some questions: Given a \( \sigma \)-finite measure \( \mu \), does \( N \) exist? If so is it uniquely defined? We try answering these hence.
1. First let us answer the question of existence. Given a $\sigma$-finite measure $\mu$ on $E$, can we construct $N \sim PRM(\mu)$.

(a) If $\mu \equiv 0$ then take $N \equiv 0$ which will satisfy the aforementioned properties of a $PRM(\mu)$.

(b) Suppose $\mu$ is finite, $\mu(E) < \infty$. Let $\tau, X_1, X_2, \ldots$ be independent random variables where $\tau$ follows a Poisson distribution with parameter $\mu(E)$ and $X_i$'s are distributed with distribution $\mu(\cdot)\mu(E)$. Now define
\[
N(\omega, A) = \begin{cases} \tau(\omega) \sum_{i=1}^{\tau(\omega)} \epsilon X_i(\omega)(A) & \text{on } \tau > 0, \\ 0, & \text{otherwise.} \end{cases}
\]
Check now that $N \sim PRM(\mu)$.

(c) Now suppose $\mu$ is $\sigma$-finite. Then decompose $\mu = \sum_{k=1}^{\infty} \mu_k$ by partitioning $E = \bigcup_{k=1}^{\infty} F_k$ where each $F_k$ is relatively compact with finite measure and defining $\mu_k(\cdot) = \mu(\cdot \cap F_k)$. We know how to construct a PRM for a finite measure. The rest is left as an exercise. Check (Resnick, 2008, page 133).

2. The Laplace functional of $N \sim PRM(\mu)$ is given by:
\[
\psi_N(f) = \exp\{-\int_E (1 - e^{-f(x)})\mu(dx)\} \quad (3.3.1)
\]
for any measurable $f \geq 0$. Conversely a point process with Laplace functional (3.3.1) must be $PRM(\mu)$.

If we prove the previous statement then since Laplace functionals uniquely determine the law of a point process, we would know that our definition of $PRM(\mu)$ is unique.

Proof. First check that if $Z$ follows a Poisson distribution with parameter $\lambda$ then the Laplace transform of $Z$ is given by
\[
E e^{-tZ} = \sum_{k=0}^{\infty} e^{-tk} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{-t})^k}{k!} = \exp\{-1 - e^{-t}\lambda\} \quad \forall t > 0.
\]

Let $N \sim PRM(\mu)$. If $c > 0, F \in \mathcal{E}$ and we take $f(x) = c1_F(x)$, then
\[
N(f) = \int_{E} f(x)N(dx) = cN(F).
\]
Therefore
\[
\psi_N(f) = E e^{-N(f)} = E e^{-cN(F)} = \exp\{-1 - e^{-c})\mu(F)\} \quad \text{as } N(F \sim Poisson(\mu(F))
\]
\[
= \exp\{-\int_{E} (1 - f(x))\mu(dx)\}.
\]
3.3. POISSON PROCESS OR POISSON RANDOM MEASURE

Now take $c_i \geq 0, F_i \in \mathcal{E}, 1 \leq i \leq k$ which are pairwise disjoint and define $f(x) = \sum_{i=1}^{k} c_i 1_{F_i}(x)$. First note that $N(f) = \sum_{i=1}^{k} c_i N(F_i)$. Now

$$\psi_N(f) = \mathbb{E}e^{-N(f)} = \prod_{i=1}^{k} \mathbb{E}e^{-c_i N(F_i)} \quad \text{(by independence)}$$

$$= \prod_{i=1}^{k} \exp \left\{ - \int_{\mathbb{R}} \left( 1 - e^{-c_i 1_{F_i}(x)} \right) \mu(dx) \right\}$$

$$= \exp \left\{ - \int_{\mathbb{R}} \sum_{i=1}^{k} \left( 1 - e^{-c_i 1_{F_i}(x)} \right) \mu(dx) \right\}$$

$$= \exp \left\{ - \int_{\mathbb{R}} \left( 1 - e^{-\sum_{i=1}^{k} c_i 1_{F_i}(x)} \right) \mu(dx) \right\} \quad \text{(easy to check)}$$

Now given $f \geq 0$, we can find simple functions $f_n$ as defined above such that $f_n \uparrow f$ as $n \to \infty$. Observe that this implies $N(f_n) \uparrow N(f) \forall \omega \in \Omega$. Also note that $e^{-N(g)} \leq 1$ for all $g \geq 0$ measurable. Therefore using Dominated Convergence Theorem

$$\psi_N(f) = \mathbb{E}e^{-N(f_n)} = \mathbb{E}e^{-N(f)} =: \psi_N(f). \quad (3.3.2)$$

On the otherhand, by Monotone Convergence Theorem,

$$\psi_N(f) = \int_{\mathbb{R}} \left( 1 - e^{-f(x)} \right) \mu(dx) \uparrow \int_{\mathbb{R}} \left( 1 - e^{-f(x)} \right) \mu(dx) \quad (3.3.3)$$

Therefore we have for any $f \geq 0$, measurable

$$\psi_N(f) = \exp \left\{ - \int_{\mathbb{R}} (1 - f(x)) \mu(dx) \right\}.$$

Conversely, suppose $N$ has Laplace Functional given by (3.3.1). For any $F \in \mathcal{E}$, let $f = \lambda 1_F$, for some $\lambda > 0$. Then

$$\mathbb{E}e^{-\lambda N(F)} = \mathbb{E}e^{-N(f)} = \exp \left\{ - (1 - e^{-\lambda}) \mu(F) \right\},$$

which is the Laplace Transform of $\text{Poisson}(\mu(F))$. Thus criterion (1) for $\text{PRM}(\mu)$ holds. Now for $k \geq 1$, let $F_1, \ldots, F_k$ be disjoint in $\mathcal{E}$. Take $f(x) = \sum_{i=1}^{k} \lambda_i 1_{F_i}(x)$ for some $\lambda_i \geq 0$. Then

$$\mathbb{E}e^{-\sum_{i=1}^{k} \lambda_i 1_{F_i}} = \mathbb{E}e^{-N(f)} = \exp \left\{ - \int_{\mathbb{R}} (1 - f(x)) \mu(dx) \right\}$$

$$= \exp \left\{ - \int_{\mathbb{R}} \left( 1 - e^{-\sum_{i=1}^{k} \lambda_i 1_{F_i}(x)} \right) \mu(dx) \right\}$$

$$= \exp \left\{ - \int_{\mathbb{R}} \sum_{i=1}^{k} \left( 1 - e^{-\lambda_i 1_{F_i}(x)} \right) \mu(dx) \right\}.$$
CHAPTER 3. POINT PROCESSES

\[ = \prod_{i=1}^{k} \mathbb{E}e^{-\lambda_i N(F_i)}. \]

Hence \((N(F_1), N(F_2), \ldots, N(F_k))\) are independent random variables and criterion (2) for \(PRM(\mu)\) is satisfied.

**Remark 3.3.2.** If \(\mu = \lambda \times \text{Leb}, \lambda > 0\) then \(N \sim PRM(\mu)\) is called a Homogeneous Poisson Process with rate \(\lambda\).

**Definition 3.3.2** (Rate function). Suppose \(E \subseteq [-\infty, \infty]^d, d \geq 1\) and \(N \sim PRM(\mu)\) such that

\[ \mu(A) = \int_A \lambda(x)dx \quad \forall A \in \mathcal{E}, \]

for some function \(\lambda : E \rightarrow \mathbb{R}\). Then \(\lambda(\cdot)\) is called the rate function or intensity function of the Poisson Process \(N\). For a Homogeneous Poisson Process (HPP), \(\lambda(x) \equiv \lambda\) for some \(\lambda > 0\).

**Proposition 3.3.1.** Suppose \(\{E_j; j \geq 1\}\) are i.i.d. \(\exp(1)\) random variables. Let \(\Gamma_n = \sum_{i=1}^{n} E_n\). If \(N := \sum_{i=1}^{\infty} \varepsilon_i\) then \(N \sim HPP\) on \([0, \infty)\) with rate \(\lambda = 1\).

### 3.3.1 Transformation of a Poisson Process

Here we see that under suitable transformation Poisson Point Processes still remain Poisson Point Processes.

**Proposition 3.3.2.** Let \(E_1\) and \(E_2\) be two ‘nice’ spaces and let \(\mathcal{E}_1, \mathcal{E}_2\) be their associated \(\sigma\)-algebras. Let \(T : (E_1, \mathcal{E}_1) \rightarrow (E_2, \mathcal{E}_2)\) be a measurable map. If

\[ N_1 = \sum_{i \in I} \varepsilon_{X_i} \sim PRM(\mu_1) \quad \text{on} \quad E_1 \]

then,

\[ N_2 := \sum_{i \in I} \varepsilon_{T(X_i)} \sim PRM(\mu_2) \quad \text{on} \quad E_2, \]

where \(\mu_2 = \mu_1 \circ T^{-1}\).

**Proof.** Let \(f_2 : E \rightarrow [0, \infty)\) be measurable. Then

\[ N_2(f_2) = \sum_{i \in I} \varepsilon_{T(X_i)}(f_2(B)) = \sum_{i \in I} \varepsilon_{X_i}(T^{-1}(f_2(B))) = N_1 \circ T^{-1}(f_2), \quad \forall B \in \mathcal{E}_2. \]

Now since,

\[ N_2(f_2) = \int_{E_2} f_2(x) N_2(dx) = \int_{E_2} f_2(x) N_1 \circ T^{-1}(dx) = \int_{E_1} f_2(T(y)) N_1(dy) = N_1(f_2 \circ T). \]

we have,

\[ \psi_{N_2}(f_2) = \mathbb{E}[\exp(-N_2(f_2))] = \mathbb{E}[\exp(-N_1(f_2 \circ T))]. \]
3.3. POISSON PROCESS OR POISSON RANDOM MEASURE

\[ = \exp \left\{ - \int_{\mathcal{E}_1} (1 - e^{-f_2 \circ T(y)}) \mu_1(dy) \right\} \]
\[ = \exp \left\{ - \int_{\mathcal{E}_2} (1 - e^{-f_2(x)}) \mu_1 \circ T^{-1}(dx) \right\} \]
\[ = \exp \left\{ - \int_{\mathcal{E}_2} (1 - e^{-f_2(x)}) \mu_2(dx) \right\}. \]

**Example 3.3.1.** Suppose \( N = \sum_{i=1}^{\infty} \varepsilon_{X_i} \) is a HPP on \([0, \infty)\) with rate 1. Then compute the rate function of the point process \( N^* = \sum_{i=1}^{\infty} \varepsilon_{X_i^2} \).

Note here that \( E_1 = E_2 = [0, \infty) \), \( T(x) = x^2 \) and \( \mu_1 = \text{Leb} \). Therefore using Proposition 3.3.2, \( N^* \sim \text{PRM}(\mu_2) \) where \( \mu_2 = \text{Leb} \circ T^{-1} \). For \( y > 0 \),

\[ \mu_2(0, y] = \text{Leb} \circ T^{-1}(0, y] = \text{Leb}(\{x > 0 : T(x) \in (0, y]\}) \]
\[ = \text{Leb}(\{x > 0 : T(x) \leq y\}) = \sqrt{y} = \int_0^\infty \frac{1}{2\sqrt{y}} dy. \]

So \( N^* \) is a Poisson Process on \((0, \infty)\) with rate function \( \lambda(y) = \frac{1}{2\sqrt{y}}, y > 0 \).

### 3.3.2 Augmentation of a Poisson Process

Here we attach one more number with each of the points of the point process. This process is called marking.

**Proposition 3.3.3.** Let \( E_1, E_2 \) be ‘nice’ spaces equipped with their \( \sigma \)-algebras \( E_1 \) and \( E_2 \). Suppose we have

\[ N = \sum_i \varepsilon_{X_i} \sim \text{PRM}(\mu_1) \quad \text{on} \quad E_1 \]

and \( \{J_n\} \) are i.i.d. \( E_2 \)-valued random elements with common distribution \( F \). Also assume that the point process \( N \) and the sequence \( \{J_n\} \) are independent. Note \( P(J_1 \in B) = F(B), \forall B \in E_2 \). Then the point process defined on \( E_1 \times E_2 \):

\[ N^* = \sum_i \varepsilon_{(X_i, J_i)} \]

is a PRM with mean measure \( \mu \otimes F \), i.e., for \( A_1 \in \mathcal{E}_1, i = 1, 2, \)

\[ \mu \otimes F(A_1 \times A_2) = \mu \otimes F(\{(e_1, e_2) : e_1 \in A_1, e_2 \in A_2\}) \]
\[ = \mu(A_1)F(A_2). \]


So we have given the mark \( J_n \) to the point \( X_n \). \( N^* \) is also called a Marked Poisson Process. Observe that

\[ E \sum_n \varepsilon_{(X_n, J_n)}(A_1 \times A_2) = \sum_n P((X_n, J_n) \in A_1 \times A_2) \]
Example 3.3.2. Suppose calls arrive at a telephone exchange following a PRM \( \mu \) at times \( \{ X_i \} \subseteq (-\infty, \infty) \). Again, length of the calls \( \{ J_i \} \) are i.i.d. from a distribution \( F \) and independent of the call arrivals \( \{ X_i \} \). Show that the call terminations form a Poisson Process. Find its mean measure.

Since \( N = \sum_i \varepsilon_{X_i} \sim PRM(\mu) \), by Proposition 3.3.3, we have

\[
N_1 := \sum_i \varepsilon_{(X_i, J_i)} \sim PRM(\mu \otimes F) \quad \text{on} \quad \mathbb{E}_1 = (-\infty, \infty) \times (0, \infty).
\]

Let \( \mathbb{E}_2 = (-\infty, \infty) \). Now define \( T : \mathbb{E}_1 \to \mathbb{E}_2 \) as \( T(x, y) = x + y \) and the point process

\[
N_2 := \sum_i \varepsilon_{X_i + J_i} \quad \text{on} \quad \mathbb{E}_2 = (-\infty, \infty).
\]

Then from Proposition 3.3.2 we have \( N_2 \sim PRM((\mu \otimes F) \circ T^{-1}) \). The mean measure of \( B \subseteq (-\infty, \infty) \), Borel is given by

\[
(\mu \otimes F) \circ T^{-1}(B) = (\mu \otimes F)(\{(x, y) : x + y \in B\}) = \mu \ast F(B) \quad \text{(convolution)}.
\]

Interesting fact: if \( \mu \) is Lebesgue, then so is \( \mu \ast F \), whatever the distribution of \( F \). Let \( -\infty < a < b < \infty \).

\[
\mu \ast F((a, b]) = \mu \otimes F\{(x, y) : x + y \in [a, b]\} = \int_{\{(x, y) : x + y \in [a, b]\}} \mu(dx)F(dy) = \int_0^\infty Leb((a-y, b-y])F(dy)
\]

\[
= (b-a) \int_0^\infty F(dy) = b - a.
\]

3.4 Weak Convergence of Point Processes and Random Measures

Let us recall a little bit of weak convergence before we start. Let \( \mathcal{S} \) be a complete separable metric space (c.s.m.s.) equipped with the Borel \( \sigma \)-algebra \( \mathcal{S} \) and metric \( d \). Let \((\Omega, \mathcal{A}, P)\) be a probability space. Suppose \( X \) denotes a random element in the space \( \mathcal{S} \).

<table>
<thead>
<tr>
<th>( \mathcal{S} )</th>
<th>( X )</th>
<th>( \mathbb{R} )</th>
<th>( \mathbb{R}^d )</th>
<th>( C[0, \infty) )</th>
<th>( M_p(\mathcal{E}) )</th>
<th>( M_{+}(\mathcal{E}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>space</td>
<td>random element</td>
<td>random variable</td>
<td>random vector</td>
<td>random processes with continuous paths</td>
<td>stochastic point processes</td>
<td>random measures</td>
</tr>
</tbody>
</table>
3.4. WEAK CONVERGENCE OF POINT PROCESSES AND RANDOM MEASURES

Suppose \( \{X_n; n \geq 0\} \) are random elements of \( \mathcal{S} \). The distribution of \( X_n \) is \( P_n = P \circ X_n^{-1} \) on \( \mathcal{S} \). Also \( X_n, n \geq 1 \) converges weakly to \( X_0 \) if for all \( f \in C(\mathcal{S}) \), the space of bounded continuous real-valued functions in \( \mathcal{S} \), \( Ef(X_n) \to Ef(X) \). We write \( X_n \Rightarrow X_0 \) or \( P_n \Rightarrow P_0 \). A Portmanteau Theorem gives alternative characterizations of weak convergence Billingsley (1968). Let us also state without proof here a very important result from weak convergence theory.

**Theorem 3.4.1** (Continuous Mapping Theorem). Let \( (\mathcal{S}_1, d_i), i = 1, 2 \) be two metric spaces where \( \{X_n, n \geq 0\} \) are random elements in \( (\mathcal{S}_1, \mathcal{S}_1) \) and \( X_n \Rightarrow X_0 \). Now if \( h : \mathcal{S}_1 \to \mathcal{S}_2 \) satisfies

\[
P(X_0 \in D_n) = P(X_0 \in \{ s_1 \in \mathcal{S}_1 : h \text{ is discontinuous at } s_1 \}) = 0,
\]

then \( h(X_n) \Rightarrow h(X_0) \) in \( \mathcal{S}_2 \).

We also know that weak convergence of random vectors can be checked by the point-wise convergence of their corresponding characteristics functions. Now when we want to talk about weak convergence of random measures, i.e., random elements of \( \mathcal{M}_+(\mathcal{E}) \), this can be checked by showing for each \( f \in C^+_K(\mathcal{E}) \) the convergence of the corresponding Laplace functionals.

**Theorem 3.4.2.** Suppose \( \{\eta_n; n \geq 1\} \) and \( \eta \) are random measures in \( \mathcal{M}_+(\mathcal{E}) \). Then

\[
\eta_n \Rightarrow \eta \quad \text{in} \quad \mathcal{M}_+(\mathcal{E})
\]

if and only if

\[
\psi_{\eta_n}(f) := Ee^{-\eta_n(f)} \to Ee^{-\eta(f)} = \psi_\eta(f), \quad \forall f \in C^+_K(\mathcal{E}).
\]

**Proof.** First assume \( \eta_n \Rightarrow \eta \) as \( n \to \infty \). For \( f \in C^+_K(\mathcal{E}) \) define the continuous functional \( T_f : \mathcal{M}_+(\mathcal{E}) \to [0, \infty) \) by \( T_f(\mu) = \mu(f) = \int f(x)\mu(dx) \) (check that \( T_f \) is indeed continuous under the vague metric). Therefore by Proposition 3.4.1

\[
T_f(\eta_n) = \eta_n(f) \Rightarrow \eta(f) = T_f(\eta).
\]

Also note that \( e^{-\mu(s)} \leq 1 \) for any \( \mu \in \mathcal{M}_+(\mathcal{E}), g \in C^+_K(\mathcal{E}) \). Therefore Dominated convergence theorem implies

\[
\psi_{\eta_n}(f) = Ee^{-\eta_n(f)} \to Ee^{-\eta(f)} = \psi_\eta(f).
\]

Converse: The converse essentially requires a tightness argument in terms of \( \eta_n(f) \), for further details see (Resnick, 2008, Chapter 3.5). We provide a simpler proof by assuming a different characterization of weak convergence in \( \mathcal{M}_+(\mathcal{E}) \). Suppose \( \psi_{\eta_n}(f) \to \psi_\eta(f) \) for all \( f \in C^+_K(\mathcal{E}) \). We prove \( \eta_n \Rightarrow \eta \) by assuming we know the following: \( \eta_n \Rightarrow \eta \) iff for any family \( \{h_j\}_{j \geq 1} \subset C^+_K(\mathcal{E}) \),

\[
(\eta_n(h_j); j \geq 1) \Rightarrow (\eta(h_j); j \geq 1) \quad \text{in} \quad \mathbb{R}^\infty.
\]

(3.4.1)

It in fact suffices to show this for a finite collection, i.e., for some \( d \geq 1 \),

\[
(\eta_n(h_j); 1 \leq j \leq d) \Rightarrow (\eta(h_j); 1 \leq j \leq d) \quad \text{in} \quad \mathbb{R}^d.
\]

(3.4.2)

We show that the Laplace transforms converge for (3.4.2). Let \( \lambda_i \geq 0, 1 \leq i \leq d \). Then

\[
E \exp\{-\sum_{i=1}^d \lambda_i \eta_n(h_i)\} = E \exp\{-\eta_n(\sum_{i=1}^d \lambda_i h_i)\}
\]

\[
\to E \exp\{-\eta(\sum_{i=1}^d \lambda_i h_i)\} \quad \text{since} \quad \sum_{i=1}^d \lambda_i h_i \in C^+_K(\mathcal{E})
\]

\[
= E \exp\{-\sum_{i=1}^d \lambda_i \eta(h_i)\}.
\]

Hence we have shown (3.4.2) and hence the converse.

\[\square\]
A good question to ask here is when does \( N_n \sim PRM(\mu_n) \) converge weakly to \( N \sim PRM(\mu) \)? Naturally it seems that \( N_n \Rightarrow N \) iff \( \mu_n \Rightarrow \mu \). This is true.

**Proposition 3.4.3.** Suppose \( \mu, \mu_n, n \geq 1 \) are Radon measures on \( \mathbb{E} \) and \( N_n \sim PRM(\mu_n) \) and \( N \sim PRM(\mu) \). Then \( N_n \Rightarrow N \) iff \( \mu_n \Rightarrow \mu \).

**Proof.** Suppose \( \mu_n \Rightarrow \mu \). Let \( f \in C^0_K(\mathbb{E}) \). Clearly then \( 1 - e^{-f} \in C^0_K(\mathbb{E}) \). Therefore

\[
\int_{\mathbb{E}} (1 - e^{-f(x)}) \mu_n(dx) \to \int_{\mathbb{E}} (1 - e^{-f(x)}) \mu(dx). \tag{3.4.3}
\]

Hence

\[
\psi_{N_n}(f) = \mathbb{E}(\exp(-N_n(f))) = \exp \{ - \int_{\mathbb{E}} (1 - e^{-f(x)}) \mu_n(dx) \} \to \exp \{ - \int_{\mathbb{E}} (1 - e^{-f(x)}) \mu(dx) \} = \psi_N(f).
\]

Hence \( N_n \Rightarrow N \).

**Converse:** Now assume \( N_n \Rightarrow N \). Then since Laplace functionals converge, so for all \( f \in C^0_K(\mathbb{E}) \),

\[
\int_{\mathbb{E}} (1 - e^{-f(x)}) \mu_n(dx) \to \int_{\mathbb{E}} (1 - e^{-f(x)}) \mu(dx).
\]

We will prove the following characterization of vague convergence, see Theorem 2.7.2 (Portmanteau theorem for vague convergence): \( \mu_n \Rightarrow \mu \) if and only if for all compact \( K \) and for all open relatively compact \( G \), we have

\[
\lim_{n \to \infty} \sup \mu_n(K) \leq \mu_0(K),
\]

\[
\lim_{n \to \infty} \inf \mu_n(G) \geq \mu_0(G).
\]

Suppose \( K \) is compact in \( \mathbb{E} \). Then according to the variant of Urysohn’s Lemma (Lemma (2.7.1)) we had seen earlier, there exists compact \( K_n \downarrow K \) and \( f_n \in C^0_K(\mathbb{E}) \), with \( f_n \) non-increasing, such that,

\[
1_K \leq f_n \leq 1_{K_n} \downarrow 1_K.
\]

which implies

\[
1 - e^{-1} \leq 1 - e^{-f_n} \leq 1 - e^{-1_{K_n}} \downarrow 1 - e^{-1}.
\]

Fix an integer \( m \).

\[
\lim_{n \to \infty} \sup (1 - e^{-1}) \mu_n(K) = \lim_{n \to \infty} \sup \int_{\mathbb{E}} (1 - e^{-1_{K_n}}(x)) \mu_n(dx)
\]

\[
\leq \lim_{n \to \infty} \int_{\mathbb{E}} (1 - e^{-f_m(x)}) \mu_n(dx)
\]

\[
\to \int_{\mathbb{E}} (1 - e^{-f_m(x)}) \mu(dx). \quad \text{(as } n \to \infty\text{)}
\]

Now \( f_m \leq 1_{K_{m_0}}, \forall m \geq m_0 \), i.e., \( 1 - e^{-f_m} \leq 1 - e^{-1_{K_{m_0}}}, \forall m \geq m_0 \). Therefore we use DCT to get,

\[
\lim_{n \to \infty} \sup (1 - e^{-1}) \mu_n(K) \leq \int_{\mathbb{E}} (1 - e^{-f_m(x)}) \mu(dx)
\]

\[
\to \int_{\mathbb{E}} (1 - e^{-1_{K}}(x)) \mu(dx) = (1 - e^{-1}) \mu(K).
\]

Similarly we can show for open and relatively compact sets \( G \) that \( \lim_{n \to \infty} \inf \mu_n(G) \geq \mu_0(G) \).

Hence \( \mu_n \Rightarrow \mu \).
Corollary 3.4.4. Suppose $\mu_n, n \geq 1, \mu \in M_+(E)$. $\mu_n \Rightarrow \mu$ iff $\forall f \in C^+_K(E)$,
\[ \int_E (1 - e^{-f(x)}) \mu_n(dx) \rightarrow \int_E (1 - e^{-f(x)}) \mu(dx). \]

Proof. It is clear from the proof of Theorem 3.4.3. See (3.4.3).

Finally we talk about convergence of random point measures to Poisson Random measures.

Proposition 3.4.5 (Convergence of Empirical measures). Suppose for each $n \geq 1$, \{X_{n,j}; 1 \leq j \leq n\} is a sequence of i.i.d. random elements of $(E, E)$. Let $N$ be PRM($\mu$) on $M_p(E)$. Then
\[ \sum_{j=1}^{n} \varepsilon_{X_{n,j}} \Rightarrow N \equiv \text{PRM}(\mu) \]
if and only if
\[ nP(X_{n,1} \in \cdot) = E\left( \sum_{j=1}^{n} \varepsilon_{X_{n,j}}(\cdot) \right) \Rightarrow \mu(\cdot) \text{ in } M_+(E). \]

Proof. Recall the following result from real analysis:

Lemma: \{x_n\} is a sequence of real numbers. Then $(1 - x_n^n) \rightarrow e^{-x}$ if and only if $x_n \rightarrow x$ as $n \rightarrow \infty$.

Now let $N_n := \sum_{j=1}^{n} \varepsilon_{X_{n,j}}$. Then for any $f \in C^+_K(E)$,
\[ \psi_{N_n}(f) = E\left( \exp\{-N_n(f)\}\right) \]
\[ = E\left( \exp\left\{ -\sum_{j=1}^{n} f(X_{n,j}) \right\}\right) \]
\[ = [E\left( \exp\{-f(X_{n,1})\}\right)]^n \]
\[ = \left[ \frac{1}{n} \int_E \exp\{-f(x)\} nP(X_{n,1} \in dx) \right]^n \]
\[ = \left( 1 - \frac{1}{n} \int_E \left( 1 - \exp\{-f(x)\}\right) nP(X_{n,1} \in dx) \right)^n \]
\[ \Rightarrow \exp\left\{ -\int_E \left( 1 - e^{-f(x)}\right) \mu(dx) \right\} = \psi_N(f). \]

The convergence in the previous line holds using the analysis result if and only if
\[ \int_E (1 - e^{-f(x)}) nP_n(X_{n,1} \in dx) \rightarrow \int_E (1 - e^{-f(x)}) \mu(dx), \forall f \in C^+_K(E) \]
which holds if and only if $nP(X_{n,1} \in \cdot) \Rightarrow \mu(\cdot)$ using Corollary 3.4.4.

This also has a nice and useful corollary applicable in extreme value theory.

Corollary 3.4.6. Suppose we have $\mathbb{R}^d$-valued i.i.d. random elements $X_1, X_2, \ldots$ from $E \subset [-\infty, \infty]^d, d \geq 1$ and a sequence of reals $b_n \rightarrow \infty$. Then
\[ N_n := \sum_{i=1}^{n} \varepsilon_{X_i} \Rightarrow \text{PRM}(\mu) \]
iff \( n\mathbb{P}\left( \frac{X_i}{b_n} \in \cdot \right) \xrightarrow{\text{w}} \mu(\cdot). \)

**Proof.** Put \( X_{n,j} = \frac{X_j}{b_n}, j \leq n \) in Proposition 3.4.5. \( \square \)

### 3.4.1 Applications to Extreme Value Theory

For univariate extreme value theory we provided an alternative characterization of \( F \in D(\Phi_\alpha) \) in terms of vague convergences in Theorem 2.7.3. Now we extend this to point process theory.

**Theorem 3.4.7.** Suppose \( X_1, X_2, \ldots \) are non-negative random variables with distribution function \( F \). Denote \( \overline{F}(x) = 1 - F(x) \). Then the following are equivalent:

1. \( F \in RV_{\alpha}, \alpha > 0 \).
2. There exists a sequence \( \{b_n\} \) with \( b_n \to \infty \) such that
   \[
   \lim_{n \to \infty} n\overline{F}(b_n) = x^{-\alpha}, x > 0, \alpha > 0.
   \]
3. There exists a sequence \( \{b_n\} \) with \( b_n \to \infty \) such that
   \[
   \mu_n(\cdot) := n\mathbb{P}\left\{ \frac{X_1}{b_n} \in \cdot \right\} \xrightarrow{\text{w}} \nu_\alpha(\cdot)
   \]
   in \( M_+(0, \infty] \) where \( \nu_\alpha(x, \infty] = x^{-\alpha} \).
4. \( F \in D(\Phi_\alpha) \).
5. There exists a sequence \( \{b_n\} \) with \( b_n \to \infty \) such that
   \[
   \sum_{i=1}^n \varepsilon X_i / b_n \Rightarrow \text{PRM}(\nu_\alpha).
   \]

**Proof.** (1)-(4) are equivalent using Theorem 2.7.3. (3) \( \leftrightarrow \) (5) from Corollary 3.4.6. \( \square \)

We can also make a similar statement for Multivariate Extreme Value Theory. Suppose \( (X_i, Y_i), i \geq 1 \) are i.i.d. \( F \) with margins \( F_1, F_2 \) and \( F \in D(G), F_1 \in D(G_{\gamma_1}) \) and \( F_2 \in D(G_{\gamma_2}) \) with \( E = [0, \infty]^2 \setminus \{0\} \). Let

\[
G_0(x, y) = G\left( \frac{x^{\gamma_1} - 1}{\gamma_1}, \frac{y^{\gamma_2} - 1}{\gamma_2} \right), x, y > 0.
\]

For \( i = 1, 2, \ldots \), we perform the usual standardization

\[
X_i^* = \frac{1}{1 - F_1(X_i)}, \quad Y_i^* = \frac{1}{1 - F_2(Y_i)}
\]

We know that \( \forall x, y \in [0, \infty)^2 \setminus \{0\}, \)

\[
n\mathbb{P}\left( \left( \frac{X_i^*}{\sqrt{n}}, \frac{Y_i^*}{\sqrt{n}} \right) \in [0, (x, y)]^c \right) \to -\log G_0(x, y)
\]

\[
=: \nu_0([0, (x, y)]^c).
\] (3.4.4)

As an exercise, show that implies

\[
\mu_n(\cdot) := n\mathbb{P}\left( \left( \frac{X_i^*}{\sqrt{n}}, \frac{Y_i^*}{\sqrt{n}} \right) \in \cdot \right) \xrightarrow{\text{w}} \nu_0(\cdot) \text{ in } M_+(E).
\] (3.4.5)

and conversely (3.4.5) implies (3.4.1) (which is trivial using Portmanteau Theorem). We need a couple of facts for the proof.
Definition 3.4.1. A subset $M \subseteq M_+ (\mathbb{E})$ is vaguely relatively compact if for every sequence $\{\mu_n\} \subset M$, there exists a subsequence $\{\mu_{n'}\} \subset \{\mu_n\}$ such that $\mu_{n'} \stackrel{v}{\rightarrow} \mu_0$, where $\mu_0 \in M_+ (\mathbb{E})$.

Lemma 3.4.8. $M \subseteq M_+ (\mathbb{E})$ is vaguely relatively compact iff $\sup_{\mu \in M} \mu (f) < \infty$, $\forall f \in C^+_K (\mathbb{E})$.

For a proof check (Resnick, 2007, Lemma 6.1).

Theorem 3.4.9. Suppose $(X_1, Y_1)$ is a bivariate random variable with distribution function $F$ and margins $F_1, F_2$. Then the following are equivalent:

1. $F \in D(G)$ with $F_1 \in D(G_{\gamma_1})$ and $F_2 \in D(G_{\gamma_2})$.

2. $\mu_n \stackrel{v}{\rightarrow} \nu_0$ with definitions as given in (3.4.5) nd (3.4.1).

3. $\sum_{i=1}^{n} \varepsilon \left( \frac{X_i}{n}, \frac{Y_i}{n} \right) \Rightarrow PRM (\nu_0)$. 

Bibliography


