University of Zurich
Institute of Mathematics

Master Thesis

Beyond the Bell Curve: Heavy-Tailed Distributions and Their Limit Theorems

Author:
Teodora Krstic

Supervisor:
Prof. Dr. Jean Bertoin
Co-Supervisor:
Dr. Martin Minchev

June 2026

Acknowledgements

I would first like to express my deepest gratitude to Prof. Dr. Jean Bertoin for giving me the opportunity to write my thesis in the field of probability and for welcoming me under your guidance shortly before your retirement. Your passion for mathematics and teaching has been a true inspiration to me.

I am sincerely grateful to Dr. Martin Minchev for supervising this thesis, for your valuable inputs, your careful corrections, and especially for encouraging me to simply begin writing. Working with you has been both motivating and enriching.

I would also like to thank Dr. Alejandro Rosales Ortiz and Sophie Schmidhuber for their support throughout my studies, especially in Probability, Functional Analysis, and Chaotic Dynamical Systems, subjects that shaped my mathematical interests the most. Your enthusiasm, patience, and joy in teaching left a lasting impression on me, and I deeply appreciate not only your guidance, but also the people you are.

I owe heartfelt thanks to my family, who was always there when I needed support, and to Sascha, who stayed by my side throughout these years. To my friends, thank you for the encouragement, the patience, and the strength you gave me along the way. A special thank you to those who stood by me from the beginning, Kristina, Siria, and Oriana, and to those who became part of this journey along the way, Laura and Elia.

Without all of you, I would not be where I am today.

Thank you.
Teodora Krstic

Abstract

This thesis develops a rigorous treatment of heavy-tailed distributions and their limit theorems. Such distributions arise naturally in finance, network science, and risk modeling, where extreme events occur more frequently than predicted by classical Gaussian models. The thesis is organized into three parts. First, we establish the theoretical foundations through regular variation and subexponentiality, showing that large sums are dominated by single large summands. Second, we prove the generalized central limit theorem, where stable distributions arise as universal limits for normalized sums of independent and identically distributed random variables. Third, we study multiplicative and extremal processes, leading to the Fréchet, Weibull, and Gumbel distributions as the only possible limits of normalized maxima. Applications to financial returns, preferential attachment models, and inter-record times demonstrate the practical relevance of the theoretical framework.

Contents

1 Introduction **1**

2 Preliminaries **3**

3 Heavy-tailed distributions **5**
3.1 Examples of heavy-tailed distributions 8
3.2 Regular Variation 10
3.2.1 Scale invariance and asymptotic scale invariance 10
3.2.2 Regular varying functions 11
3.2.3 Properties of regular varying functions 12
3.3 Subexponentiality 19
3.3.1 Conspiracies and catastrophes 20
3.3.2 Subexponential distributions 20
3.3.3 Conspiracies and catastrophes in random walks 29

4 The Generalized Central Limit Theorem **41**
4.1 Classical Central Limit Theorem 41
4.2 Stable Distributions 44
4.2.1 Properties of Stable Distributions 48
4.3 Generalized Central Limit Theorem 51
4.4 Application: Financial Returns 56

5 Limit Theorems for Multiplicative and Extremal Processes **61**
5.1 Multiplicative processes 61
5.2 Variations on multiplicative processes 64
5.3 Application: Preferential attachment 67
5.4 Application: The Zipf-Mandelbrot mechanism 73
5.5 Extremal processes 75
5.5.1 Max-Stable Distributions 76
5.5.2 The Extremal Central Limit Theorem 80
5.5.3 An alternative approach: the Generalized Pareto Distribution . . 83
5.6 Application: The time between record-breaking events 85

6 Conclusions **89**

A Principle of the One Big Jump **91**

Declaration of Independence **99**

Bibliography **100**

Chapter 1

Introduction

Heavy-tailed distributions are a central object of study in modern probability theory, arising naturally in a wide range of phenomena: the distribution of wealth and city sizes, financial market fluctuations, earthquake magnitudes, network traffic, and the degree sequences of large graphs. What unites these seemingly disparate applications is a common statistical signature: extreme events occur far more frequently than classical models based on the Gaussian distribution would predict. Understanding this phenomenon, both mathematically and conceptually, is the goal of this thesis.

A natural starting point is the classical central limit theorem, which tells us that sums of independent and identically distributed (i.i.d.) random variables with finite variance converge, after appropriate normalization, to a Gaussian distribution. This result is one of the cornerstones of probability theory and provides a compelling explanation for the universality of the bell curve in nature. It also, however, carries an implicit message: that typical randomness is well-behaved, fluctuations average out, and extreme events are exponentially unlikely.

Many real-world systems violate this premise. In financial markets, large daily price movements occur far more often than a Gaussian model predicts. In network science, the degree distribution of the internet or of citation graphs follows a power law, with a small number of nodes accumulating a disproportionately large number of connections. In insurance and risk management, the bulk of losses stem from rare but catastrophic events. In all of these settings, the tail of the distribution, the probability of observing extreme values, is the primary object of interest, and light-tailed models systematically underestimate it.

This thesis develops the mathematical theory of heavy-tailed distributions in a unified and rigorous framework. We begin in Chapter 3 by giving a precise definition of heavy-tailedness and introducing the key structural properties that distinguish heavy-tailed from light-tailed distributions. Two central concepts are introduced here: regular variation, which provides a general and analytically tractable notion of power-law tail behavior, and subexponentiality, which formalizes the principle that large values of a sum are typically caused by a single exceptionally large summand, the so-called catastrophe principle. These two notions are compared, their relationship is made precise, and the

contrast with light-tailed behavior is illustrated through Cramér's theorem.

Chapter 4 addresses the question of limit theorems for additive processes in the heavy-tailed regime. We revisit the classical central limit theorem and identify its limitations when finite variance fails. This leads to the class of stable distributions, which are characterized by stability under addition and which naturally include heavy-tailed behavior. The generalized central limit theorem characterizes all possible non-degenerate limits of normalized sums of i.i.d. random variables, placing the Gaussian distribution as a special boundary case within a much richer family. An application to financial return data, motivated by the empirical work of Mandelbrot, illustrates how stable laws arise naturally in practice and why classical Gaussian assumptions are inadequate for modeling financial markets.

Chapter 5 turns to two further mechanisms that generate heavy-tailed behavior. Multiplicative processes, where successive random factors scale the current state of a system rather than shifting it, lead naturally to lognormal and, under more general conditions, to power-law limiting distributions. This mechanism underlies many economic and physical phenomena, from income distributions to the sizes of firms. We also analyze extremal processes, focusing on the running maximum of independent random variables. Here a distinct analogue of the central limit theorem emerges: the extremal central limit theorem, which shows that suitably normalized maxima converge to one of three universal distributions - the Fréchet, Weibull, or Gumbel distributions - collectively forming the generalized extreme value family. A notable feature of this framework is that heavy-tailed limits can arise even when the underlying variables are light-tailed, providing a qualitatively different mechanism for the emergence of extremes. This is further illustrated by an additional result: the heavy-tailed nature of inter-record times.

Throughout, the exposition aims to balance mathematical rigor with conceptual clarity. Proofs are given in full where they illuminate the structure of the theory, and key results are accompanied by examples and interpretations. The reader is assumed to have a working knowledge of graduate-level probability theory. Chapter 2 fixes the notation and terminology used throughout the thesis.

The exposition in this thesis follows the treatment of Nair et al. [2022], which serves as the primary reference and organizational framework throughout. Additional sources are cited where material is drawn from elsewhere.

Chapter 2

Preliminaries

Throughout this thesis, the reader is assumed to be familiar with the foundations of probability theory at the level of a graduate course. Standard references include Billingsley [2008] and Feller [1971]. Rather than providing a full introduction, this chapter fixes notation and collects the definitions that will be used freely throughout the thesis without further comment.

The central objects of study are sums, maxima, and products of random variables, and understanding their distributional limits requires careful control of tail behavior. The notation and concepts introduced here are chosen precisely with this goal in mind.

Definition 2.1. *Let X be a real-valued random variable with distribution F. The* survival function *(or* tail function*) of F is the function $\overline{F}(x) : \mathbb{R} \to [0,1]$ defined by*

$$\overline{F}(x) := \mathbb{P}(X > x) = 1 - F(x).$$

Throughout this thesis, unless stated otherwise, tail behavior refers to the right tail, i.e. the behavior of $\overline{F}(x) := \mathbb{P}(X > x)$ as $x \to \infty$.

For two positive functions f and g, we use the following asymptotic notation as $x \to \infty$:

- $f(x) \sim g(x)$ if $\dfrac{f(x)}{g(x)} \to 1$,
- $f(x) = o(g(x))$ if $\dfrac{f(x)}{g(x)} \to 0$,
- $f(x) = O(g(x))$ if $\limsup_{x\to\infty} \dfrac{f(x)}{g(x)} < \infty$.

These relations are used interchangeably and without further reference throughout.

Chapter 3

Heavy-tailed distributions

In order to study the phenomena associated with heavy-tailed behavior, it is essential to clarify what we mean by a *heavy-tailed distribution*. Although the term is widely used across probability theory, statistics, and applied fields, its meaning is not always consistent. Depending on the context, heavy-tailed may be used interchangeably with terms such as *power-law*, *scale-free*, or *subexponential*. However, these terms are not strictly equivalent, and this ambiguity often leads to confusion.

At an intuitive level, the concept of a heavy tail is inherently comparative. A distribution is said to have a heavier tail than another if it assigns more probability to extreme events. For instance, a Gaussian distribution has a heavier tail than a uniform distribution, while an exponential distribution has a heavier tail than a Gaussian. However, such comparisons alone do not provide a precise definition.

A commonly accepted mathematical benchmark for defining heavy-tailed behavior is the exponential distribution. Informally, a distribution is considered heavy-tailed if its tail decays more slowly than any exponential function. This idea can be understood in terms of the survival function $\overline{F}(x) = \mathbb{P}(X > x)$ which measures the probability that a random variable exceeds a large threshold x.

Indeed, for an exponential random variable with parameter $\lambda > 0$, one has

$$\overline{F}(x) = e^{-\lambda x}.$$

Thus its tail is exactly exponential, making it the canonical borderline case. A particularly intuitive way to understand this boundary is through the *hazard rate*. In general, the hazard rate is defined by

$$h(x) = \lim_{\Delta x \downarrow 0} \frac{\mathbb{P}(x \leq X < x + \Delta x \mid X \geq x)}{\Delta x},$$

whenever the limit exists. It measures the instantaneous rate at which failure occurs at time x, conditional on survival up to time x. In the absolutely continuous case, where X admits a density f and survival function $\overline{F}$, this reduces to

$$h(x) = \frac{f(x)}{\overline{F}(x)}.$$

In probabilistic terms, $h(x)\,dx$ is approximately the probability that X falls in the interval $[x, x+dx]$, given that $X > x$.

The survival function and the hazard rate are linked by the identity

$$\overline{F}(x) = \exp\left(-\int_0^x h(s)\,ds\right).$$

Thus the tail behavior of a distribution is governed by the cumulative hazard

$$H(x) = \int_0^x h(s)\,ds.$$

For an exponential distribution $h(x) = \lambda$ is constant. Therefore

$$H(x) = \lambda x,$$

and hence

$$\overline{F}(x) = e^{-\lambda x}.$$

This constant hazard rate reflects the memoryless property:

$$\mathbb{P}(X > x + t \mid X > x) = \mathbb{P}(X > t).$$

In words, once the variable has survived up to a large value x, its remaining behavior is statistically the same as it was at the beginning. Large values are not progressively penalized as x increases.

This is precisely why the exponential law is a natural boundary. If the hazard rate eventually becomes larger than a positive constant, then the cumulative hazard grows faster than linearly, and the survival function decays faster than exponentially. Such distributions are light-tailed. For example, a Weibull distribution with shape parameter $\alpha > 1$ has

$$\overline{F}(x) = e^{-x^\alpha},$$

so its hazard rate increases with x. The longer the variable has already survived, the greater the instantaneous pressure toward termination. Large values become increasingly unlikely. By contrast, for many heavy-tailed distributions, the hazard rate tends to zero. For example, a Pareto distribution satisfies

$$\overline{F}(x) \sim x^{-\alpha}.$$

Its hazard rate behaves like

$$h(x) \sim \frac{\alpha}{x},$$

which decreases to zero. Conditional on already having reached a large value, the instantaneous rate of termination is small. In this sense, very large observations are not strongly penalized; survival to a large level makes further survival comparatively easier.

The exponential distribution is therefore the point at which the cumulative hazard grows linearly. Faster-than-linear growth produces tails lighter than exponential, while slower-than-linear growth produces tails heavier than exponential. This makes the exponential law the natural dividing line between light-tailed and heavy-tailed behavior.

We now formalize the definition of a heavy tail.

Definition 3.1. *A distribution F is said to be* heavy-tailed *if and only if, for all $\mu > 0$,*

$$\limsup_{x\to\infty} \frac{\overline{F}(x)}{e^{-\mu x}} = \infty.$$

Otherwise, F is said to be light-tailed*. A random variable X is said to be* heavy-tailed (light-tailed) *if its distribution is* heavy-tailed (light-tailed)*.*

In other words, the tail of F decays more slowly than any exponential function: for every $\mu > 0$, the ratio $\frac{\overline{F}(x)}{e^{-\mu x}}$ becomes arbitrarily large along some sequence as $x \to \infty$. Thus, no exponential function provides a uniform upper bound for the tail decay.

This definition highlights the key feature of heavy-tailed distributions: the non-negligible probability of extreme events. Unlike light-tailed distributions (such as the Gaussian), where large deviations are exponentially unlikely, heavy-tailed distributions allow for much higher variability and the occasional occurrence of very large values. This distinction becomes evident when comparing the tail behavior on both linear and logarithmic scales, as illustrated in Figure 3.1.

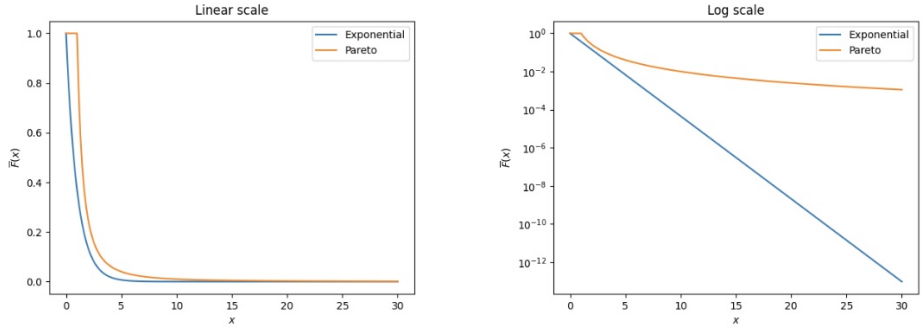

Figure 3.1: *Comparison between light-tailed and heavy-tailed distributions. The panels display the survival functions of an exponential distribution with mean* 1 *and a Pareto distribution with minimum value $x_m = 1$ and shape parameter $\alpha = 2$. On a linear scale (left), both tails appear to decay rapidly. On a logarithmic scale (right), the distinction becomes clear: the exponential distribution exhibits a much faster decay, whereas the Pareto distribution decreases more slowly, revealing its heavy-tailed nature.*

The definition above admits several equivalent formulations, which provide additional insight into the behavior of heavy-tailed distributions and highlight the limitations of classical analytical tools.

Lemma 3.2. *Let X be a random variable. The following are equivalent:*

1. *X is heavy-tailed;*
2. *$\mathbb{E}[e^{sX}] = \infty$ for all $s > 0$;*
3. *$\liminf_{x\to\infty} \frac{-\log \mathbb{P}(X>x)}{x} = 0$.*

Proof. See Lemma 1.2 in Nair et al. [2022]. □

This characterization provides important insight into the nature of heavy-tailed distributions. Condition (2) shows that heavy-tailed distributions do not admit finite exponential moments, in contrast with light-tailed distributions, for which the moment generating function plays a central role in analysis. In particular, the existence of exponential moments enables the use of exponential bounds to control tail probabilities and forms the basis of large deviation techniques and classical limit theorems. This indicates that many classical probabilistic techniques based on exponential moments are not applicable in the heavy-tailed setting. Condition (3) offers an alternative perspective: it shows that the tail of the distribution decays slower than any exponential rate. In other words, extreme events are significantly more likely than in light-tailed models.

3.1 Examples of heavy-tailed distributions

Before studying in greater detail the definition of heavy-tailed distributions and their properties, we introduce a number of representative examples that will serve as reference points throughout this thesis. We also include some classical light-tailed distributions for comparison, as they highlight the qualitative differences in tail behavior.

Light-tailed distributions.

Gaussian distribution. A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ has density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_+.$$

The Gaussian distribution is light-tailed, as its tail decays faster than any exponential.

Exponential distribution. A random variable $X \sim \text{Exp}(\lambda)$ has density

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \lambda \in \mathbb{R}_+,$$

and survival function

$$\overline{F}(x) = e^{-\lambda x}, \quad x \geq 0, \lambda \in \mathbb{R}_+.$$

The exponential distribution serves as the natural boundary between light-tailed and heavy-tailed distributions, as illustrated by the preceding heuristic discussion.

Gumbel distribution. A random variable $X \sim \text{Gumbel}(\mu, \beta)$ has density

$$f(x) = \frac{1}{\beta} \exp\left(-\frac{x-\mu}{\beta} - e^{-\frac{x-\mu}{\beta}}\right), \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \beta \in \mathbb{R}_+,$$

and survival function

$$\overline{F}(x) = 1 - \exp\left(-e^{-\frac{x-\mu}{\beta}}\right), \quad x \in \mathbb{R},\ \mu \in \mathbb{R},\ \beta \in \mathbb{R}_+.$$

The Gumbel distribution is a fundamental light-tailed distribution in extreme value theory, which will be discussed in Chapter 5.

Heavy-tailed distributions.

Pareto distribution. The Pareto distribution is a canonical example of a heavy-tailed distribution. It was originally introduced by Vilfredo Pareto in the study of wealth distribution, where he observed that a large fraction of wealth is concentrated in a small portion of the population (the so-called 80–20 principle). Since then, it has become a standard model for heavy-tailed phenomena in a wide range of applications, including finance, insurance, and network traffic.

A random variable $X \sim \text{Pareto}(x_m, \alpha)$ has density

$$f(x) = \alpha \frac{x_m^\alpha}{x^{\alpha+1}}, \quad x \geq x_m,$$

and survival function

$$\overline{F}(x) = \left(\frac{x}{x_m}\right)^{-\alpha}, \quad x \geq x_m > 0, \alpha \in \mathbb{R}_+.$$

The parameter $\alpha > 0$ is referred to as the *tail index* and determines the heaviness of the tail: smaller values of α correspond to heavier tails. The Pareto distribution exhibits polynomial decay, i.e. $\overline{F}(x) \sim x^{-\alpha}$ as $x \to \infty$. When only the asymptotic tail behavior is of interest, wewill write

$$\overline{F}(x) \sim x^{-\alpha},$$

since the scale parameter x_m contributes only a multiplicative constant and does not affect the polynomial decay rate.

Weibull distribution. A random variable $X \sim \text{Weibull}(\alpha, \beta)$ has density

$$f(x) = \alpha\beta(\beta x)^{\alpha-1} e^{-(\beta x)^\alpha}, \quad x \geq 0, \alpha, \beta \in \mathbb{R}_+,$$

and survival function

$$\overline{F}(x) = e^{-(\beta x)^\alpha}, \quad x \geq 0, \alpha, \beta \in \mathbb{R}_+.$$

The Weibull distribution provides a useful bridge between heavy-tailed and light-tailed behavior. In particular, it is heavy-tailed when $\alpha < 1$, light-tailed when $\alpha \geq 1$, and reduces to the exponential distribution when $\alpha = 1$.

Lognormal distribution. A random variable $X \sim \text{LogNormal}(\mu, \sigma^2)$ is defined via the relation

$$\log X \sim \mathcal{N}(\mu, \sigma^2),$$

and has density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \quad x > 0, \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+.$$

The Lognormal distribution is heavy-tailed, although its tail decays faster than a power law. It often arises in multiplicative processes, as will be discussed in Chapter 5.

Fréchet distribution. A random variable $X \sim \text{Fréchet}(\alpha, \beta, x_m)$ has density

$$f(x) = \alpha\beta(\beta(x - x_m))^{-(1+\alpha)} \exp\left(-(\beta(x - x_m))^{-\alpha}\right), \quad x > x_m, \alpha, \beta \in \mathbb{R}_+,$$

and cumulative distribution function

$$F(x) = \exp\left(-(\beta(x - x_m))^{-\alpha}\right), \quad x > x_m, \alpha, \beta \in \mathbb{R}_+,$$

The Fréchet distribution is a fundamental heavy-tailed distribution in extreme value theory, which will be discussed in Chapter 5.

These examples illustrate the diversity of heavy-tailed behavior. In contrast to light-tailed distributions such as the Gaussian and Exponential, heavy-tailed distributions decay more slowly, leading to a significantly higher probability of extreme events. This qualitative difference has profound implications for modeling and analysis, and will be explored in detail in the following chapters.

3.2 Regular Variation

Many phenomena encountered in practice exhibit a striking lack of a characteristic scale. For example, while quantities such as human height are concentrated around a typical value, other quantities, such as wealth or city sizes, may vary over several orders of magnitude. This qualitative difference is closely related to the notion of *scale invariance*, which lies at the heart of heavy-tailed behavior. Intuitively, a distribution is said to be scale invariant if its shape remains unchanged under rescaling. A canonical example is provided by the Pareto distribution, whose tail satisfies

$$\overline{F}(x) = x^{-\alpha},$$

so that

$$\overline{F}(\lambda x) = \lambda^{-\alpha}\overline{F}(x),$$

for all $\lambda > 0$. Thus, rescaling the argument results only in a multiplicative change, while preserving the overall structure of the tail. Scale invariance is a remarkable property that appears across a wide range of disciplines, from physics to economics and network science. In particular, it is closely related to the idea of *universality*, whereby systems of very different nature can exhibit the same macroscopic behavior. In this work, however, we focus on the mathematical implications of scale invariance in the context of probability theory.

3.2.1 Scale invariance and asymptotic scale invariance

We now formalize the notion of scale invariance.

Definition 3.3. *A distribution F is* scale invariant *if there exists an $x_0 > 0$ and a continuous positive function g such that for all x, λ satisfying $x, \lambda x \geq x_0$,*

$$\overline{F}(\lambda x) = g(\lambda)\overline{F}(x).$$

This definition captures the idea that changing the scale of observation does not alter the shape of the distribution, up to a multiplicative factor. Scale invariance is essentially equivalent to having a Pareto-type tail.

Theorem 3.4. *A distribution F is scale invariant if and only if F has a power law tail, i.e. there exists $x_0 > 0, c \geq 0$, and $\alpha > 0$ such that $\overline{F}(x) = cx^{-\alpha}$ for $x \geq x_0$.*

Proof. See Theorem 2.2 in Nair et al. [2022]. □

Although exact scale invariance provides a useful theoretical framework, it is rarely observed in empirical data. In many applications, distributions deviate substantially from a power law over most of their support, while exhibiting approximately power-law behavior only in the tail. This observation motivates the study of asymptotic scale invariance, which captures the tail behavior of the distribution.

Definition 3.5. *A distribution F is* asymptotically scale invariant *if there exists a continuous positive function g such that for any $\lambda > 0$,*

$$\lim_{x \to \infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = g(\lambda).$$

This condition requires scale invariance to hold only asymptotically. In particular, it allows the body of the distribution to behave arbitrarily, while retaining a regular structure in the tail.

3.2.2 Regular varying functions

The notion of asymptotic scale invariance leads naturally to the concept of regular variation, which provides a precise framework for describing tails that behave like power laws.

Definition 3.6. *A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is* regularly varying of index $\rho \in \mathbb{R}$, *denoted $f \in \mathcal{RV}(\rho)$, if for all $y > 0$,*

$$\lim_{x \to \infty} \frac{f(xy)}{f(x)} = y^{\rho}.$$

Further, for $\rho \leq 0$, a distribution F is regularly varying of index ρ, denoted as $F \in \mathcal{RV}(\rho)$, if $\overline{F}(x)$ is a regularly varying function of index ρ.

From now on, when we assume that f is a regularly varying function, we assume that its domain is $\mathbb{R}_+$. Also, when we talk about regularly varying distributions, it is assumed that $\rho \leq 0$.

Regular variation can be interpreted as a generalized form of power-law behavior: regularly varying functions behave asymptotically like x^{ρ} up to a multiplicative factor.

Theorem 3.7. *A distribution F is asymptotically scale invariant if and only if it is regularly varying.*

Proof. See Theorem 2.5 in Nair et al. [2022]. □

3.2.3 Properties of regular varying functions

The class of regularly varying functions is particularly appealing from an analytic perspective. Although more general than pure power laws, regularly varying functions behave asymptotically as if they were polynomials. As a result, many calculations that are straightforward for power-law functions can be extended to a much broader class of distributions.

To formalize this idea, we begin by introducing slowly varying functions.

Definition 3.8. *A function $L : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be* slowly varying *if* $\lim_{x\to\infty} \frac{L(xy)}{L(x)} = 1$ *for all* $y > 0$.

Slowly varying functions can be viewed as functions that vary asymptotically more slowly than any polynomial. Typical examples include $\log x$ and functions converging to a constant. The importance of slowly varying functions is highlighted by the following representation theorem.

Theorem 3.9. *A function f is regularly varying with index ρ if and only if*

$$f(x) = x^\rho L(x),$$

where $L(x)$ is a slowly varying function.

Proof. See Theorem 2.8 in Nair et al. [2022]. □

This representation shows that regularly varying functions are asymptotically equivalent to power laws, up to a slowly varying correction. In particular, the slowly varying component can often be treated as approximately constant in asymptotic calculations. Although regularly varying functions resemble power laws, it is not immediately clear from the definition that regularly varying distributions are heavy-tailed.

Theorem 3.10. *Every regularly varying distribution is heavy-tailed.*

Proof. See Theorem 2.9 in Nair et al. [2022]. □

This result follows from the representation $\overline{F}(x) = x^{-\alpha}L(x)$ together with the fact that slowly varying functions grow slower than any exponential. Thus, the tail of a regularly varying distribution decays more slowly than any exponential function, which characterizes heavy-tailed behavior.

Differentiation of regular varying functions. We now illustrate one of the key analytic properties of regularly varying functions, namely their behavior under differentiation.

Theorem 3.11 (Monotone Density Theorem)**.** *Suppose that the function f is absolutely continuous with derivative f'. If $f \in \mathcal{RV}(\rho)$ and f' is eventually monotone then*

$$x\, f'(x) \sim \rho f(x).$$

Moreover, if $\rho \neq 0$, then $|f'(x)| \in \mathcal{RV}(\rho - 1)$.

This result shows that differentiation of regularly varying functions behaves as in the polynomial case, where $f(x) = x^\rho$ implies $f'(x) = \rho x^{\rho-1}$.

Proof. Assume that f' is eventually non-decreasing (the non-increasing case is analogous) over $x \geq x_0$ for some fixed x_0.

Fix $0 < a < b$ and assume that x is sufficiently large such that f' is monotone on $[ax, bx]$. Then, as f is absolutely continuous

$$\frac{f(bx) - f(ax)}{f(x)} = \int_{ax}^{bx} \frac{f'(t)}{f(x)}\, dt.$$

Using monotonicity of f', we obtain the bounds

$$\frac{f'(ax)}{f(x)}(bx - ax) \;\leq\; \frac{f(bx) - f(ax)}{f(x)} \;\leq\; \frac{f'(bx)}{f(x)}(bx - ax). \tag{3.1}$$

Since f is regularly varying,

$$\frac{f(ax)}{f(x)} \to a^\rho \quad \text{and} \quad \frac{f(bx)}{f(x)} \to b^\rho \quad \text{as } x \to \infty,$$

which implies that

$$\frac{f(bx) - f(ax)}{f(x)} \;\longrightarrow\; b^\rho - a^\rho \quad \text{as } x \to \infty.$$

Thus the first inequality of (3.1) implies that

$$\limsup_{x\to\infty} \frac{x f'(ax)}{f(x)} \;\leq\; \frac{b^\rho - a^\rho}{b - a}.$$

Letting $b \downarrow a$, the right side converges to the derivative of u^ρ at $u = a$, that is,

$$\frac{d}{du} u^\rho \bigg|_{u=a} = \rho a^{\rho-1}.$$

Hence,

$$\limsup_{x\to\infty} \frac{x f'(ax)}{f(x)} \;\leq\; \rho a^{\rho-1}.$$

Similarly, using the second inequality in (3.1) and letting $a \uparrow b$, we obtain

$$\rho b^{\rho-1} \;\leq\; \liminf_{x\to\infty} \frac{x f'(bx)}{f(x)}.$$

Choosing $a = b = 1$ and combining the above bounds yields

$$\lim_{x\to\infty} \frac{x f'(x)}{f(x)} = \rho,$$

that is,

$$x f'(x) \sim \rho f(x) \qquad \text{as } x \to \infty.$$

Since $f \in \mathcal{RV}(\rho)$, it follows that

$$\frac{f(x)}{x} \in \mathcal{RV}(\rho - 1)$$

giving that

$$|f'(x)| \in \mathcal{RV}(\rho - 1).$$

□

The monotonicity assumption in Theorem 3.11 is essential. The following example shows that regular variation of f alone does not guarantee convergence.

Example 3.12. *Fix $\rho, e \in \mathbb{R}$ and choose $a \in (0,1)$. For $x > e$ define*

$$f(x) := x^{\rho}\big(1 + \varepsilon(x)\big), \qquad \varepsilon(x) := \frac{a \sin\big((\log x)^2\big)}{\log x}.$$

Extend f smoothly and positively on $(0, e]$ so that $f \in C^1((0,\infty))$.

To see that f is regularly varying, for any $c > 0$,

$$\frac{f(cx)}{f(x)} = c^{\rho}\,\frac{1 + \varepsilon(cx)}{1 + \varepsilon(x)} \longrightarrow c^{\rho} \qquad \text{as} \quad x \to \infty,$$

because $\varepsilon(x) \to 0$. Thus $f \in \mathcal{RV}(\rho)$.

But notice that the function oscillates. Differentiating f

$$f'(x) = \rho x^{\rho-1}\big(1 + \varepsilon(x)\big) + x^{\rho}\varepsilon'(x),$$

and therefore

$$\frac{x f'(x)}{f(x)} = \rho + \frac{x\varepsilon'(x)}{1 + \varepsilon(x)}. \tag{3.2}$$

Notice that $\varepsilon'(x)$ is given by

$$\varepsilon'(x) = \frac{2a\frac{1}{x}(\log x)^2 \cos\big((\log x)^2\big) - \frac{1}{x}\sin\big((\log x)^2\big)}{(\log x)^2},$$

so

$$x\varepsilon'(x) = a\left(2\cos\big((\log x)^2\big) - \frac{\sin((\log x)^2)}{(\log x)^2}\right).$$

Taking the limit as $x \to \infty$, the first term $2a\cos((\log x)^2)$ keeps oscillating between $-2a$ and $2a$, while the second term tends to 0. Since $1 + \varepsilon(x) \to 1$, dividing by $1 + \varepsilon(x)$ does not remove the oscillations. From equality (3.2), we can conclude that $\frac{xf'(x)}{f(x)}$ oscillates and hence does not converge as $x \to \infty$. In particular, the oscillatory term forces f' to change sign infinitely often, so f' is not eventually monotone.

Karamata's Representation Theorem. Although regularly varying functions can be expressed as $f(x) = x^\rho L(x)$ with L slowly varying, this representation does not fully reveal their structure. A more precise characterization is given by the following theorem.

Theorem 3.13 (Karamata's Representation Theorem)**.** *$f \in \mathcal{RV}(\rho)$ if and only if it can be represented as*

$$f(x) = c(x) \exp\left\{ \int_1^x \frac{\beta(t)}{t}\, dt \right\}, \qquad x > 0, \tag{3.3}$$

where

$$\lim_{x\to\infty} c(x) = c \in (0, \infty), \qquad \lim_{x\to\infty} \beta(x) = \rho.$$

Proof. See Theorem 2.12 in Nair et al. [2022]. □

Karamata's representation can be used to show some useful properties of regularly varying distributions. The first one is that regularly varying distributions, in particular its tail, with index α appear asymptotically linear with slope ρ when viewed on a log-log plot (see Figure 3.2).

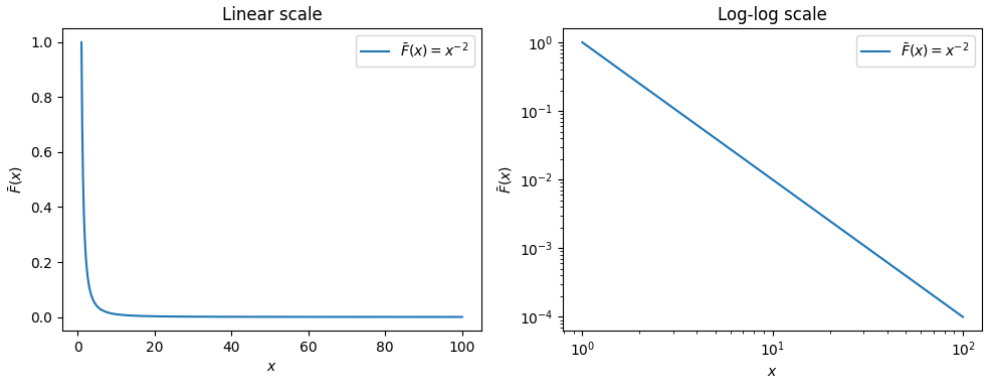

Figure 3.2: *Pareto survival function $\overline{F}(x) = x^{-2}$ shown on a linear scale (left) and a log-log scale (right). On the linear scale, the tail rapidly approaches zero, making its asymptotic behavior difficult to distinguish. On the log-log scale, the survival function becomes approximately linear, illustrating a fundamental property of regularly varying distributions.*

Proposition 3.14. *If $f \in \mathcal{RV}(\rho)$, then*

$$\lim_{x\to\infty} \frac{\log f(x)}{\log x} = \rho.$$

Proof. By the Karamata's representation in (3.3),

$$f(x) = c(x) \exp\left\{ \int_1^x \frac{\beta(t)}{t}\, dt \right\}.$$

Fix $\varepsilon > 0$. From the convergence $\lim_{x\to\infty} c(x) = c \in (0,\infty)$, there exists $x_0 > 1$ such that

$$\rho - \varepsilon < \beta(x) < \rho + \varepsilon \qquad \text{for} \quad x \geq x_0. \tag{3.4}$$

Taking the logarithm of f and dividing by $\log x$, we obtain

$$\frac{\log f(x)}{\log x} = \frac{\log c(x)}{\log x} + \frac{1}{\log x}\int_1^x \frac{\beta(t)}{t}\,dt. \tag{3.5}$$

We split the integral at x_0

$$\int_1^x \frac{\beta(t)}{t}\,dt = \int_1^{x_0} \frac{\beta(t)}{t}\,dt + \int_{x_0}^x \frac{\beta(t)}{t}\,dt.$$

Dividing by $\log x$, the first term satisfies

$$\frac{1}{\log x}\int_1^{x_0} \frac{\beta(t)}{t}\,dt \longrightarrow 0 \qquad \text{as} \quad x \to \infty.$$

For the second term, since $\beta(t) \to \rho$ and we have the bound $|\beta(t)| \leq \rho + \varepsilon$ for $t \geq x_0$, we obtain

$$\int_{x_0}^x \frac{\beta(t)}{t}\,dt \leq (\rho+\varepsilon)\big(\log x - \log x_0\big).$$

Hence combining the above in (3.5)

$$\limsup_{x\to\infty} \frac{\log f(x)}{\log x} \leq \rho + \varepsilon.$$

A similar argument gives

$$\liminf_{x\to\infty} \frac{\log f(x)}{\log x} \geq \rho - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that

$$\lim_{x\to\infty} \frac{\log f(x)}{\log x} = \rho.$$

□

Another important consequence of Karamata's representation theorem are Potter bounds. They formalize the heuristic that slowly varying functions grow more slowly than any power x^ε, for each $\varepsilon > 0$. In particular, they yield uniform polynomial control of regularly varying functions, which is useful, for example, in dominated convergence arguments and in tail estimates for heavy-tailed distributions. The following theorem and its proof follow from Proposition 2.6 in Resnick [2007].

Theorem 3.15 (Potter bounds)**.** *Let $f \in \mathcal{RV}(\rho)$. Then for every $\varepsilon > 0$ there exists $t_0 > 0$ such that for all $t \geq t_0$ and $x \geq 1$,*

$$(1-\varepsilon)x^{\rho-\varepsilon} \leq \frac{f(tx)}{f(t)} \leq (1+\varepsilon)x^{\rho+\varepsilon}. \tag{3.6}$$

Proof. Write $f(t) = t^\rho L(t)$, where L is slowly varying. Then

$$\frac{f(tx)}{f(t)} = x^\rho \frac{L(tx)}{L(t)}.$$

Apply Karamata's Representation Theorem 3.13 to L,

$$\frac{L(tx)}{L(t)} = \frac{c(tx)}{c(t)} \exp\left\{ \int_t^{tx} \frac{\beta(u)}{u}\, du \right\}. \tag{3.7}$$

Since $c(t) \to c > 0$, for every $\delta \in (0,1)$ there exists t_1 such that for $t \geq t_1$,

$$(1-\delta)c \leq c(t) \leq (1+\delta)c.$$

The same holds for $c(tx)$ for $x \geq 1$. Therefore

$$\frac{1-\delta}{1+\delta} \leq \frac{c(tx)}{c(t)} \leq \frac{1+\delta}{1-\delta}. \tag{3.8}$$

Since $\beta(u) \to 0$, for every $\varepsilon > 0$ there exists t_2 such that

$$|\beta(u)| \leq \varepsilon \quad \text{for all } u \geq t_2.$$

Thus for $t \geq t_2$ and $x \geq 1$,

$$-\varepsilon \int_t^{tx} \frac{1}{u}\, du \leq \int_t^{tx} \frac{\beta(u)}{u}\, du \leq \varepsilon \int_t^{tx} \frac{1}{u}\, du.$$

Taking the exponential, we obtain

$$x^{-\varepsilon} \leq \exp\left\{ \int_t^{tx} \frac{\beta(u)}{u}\, du \right\} \leq x^{\varepsilon}. \tag{3.9}$$

Combining the bounds (3.8), (3.9) in (3.7), for $t \geq t_0 := \max\{t_1, t_2\}$ and $x \geq 1$,

$$\frac{L(tx)}{L(t)} \in \left[\frac{1-\delta}{1+\delta} x^{-\varepsilon},\ \frac{1+\delta}{1-\delta} x^{\varepsilon} \right].$$

Multiplying by x^ρ gives

$$\frac{f(tx)}{f(t)} \in \left[\frac{1-\delta}{1+\delta} x^{\rho-\varepsilon},\ \frac{1+\delta}{1-\delta} x^{\rho+\varepsilon} \right].$$

Finally, choose $\delta > 0$ small enough such that

$$\frac{1-\delta}{1+\delta} \geq 1-\varepsilon, \qquad \frac{1+\delta}{1-\delta} \leq 1+\varepsilon.$$

This yields the desired inequality. □

Another property to recover from Karamata's Representation Theorem is the finiteness of moments of regularly varying distributions.

Proposition 3.16. *Let X be a non-negative random variable, which is in $\mathcal{RV}(-\rho)$. Then*

1. *For every $0 \leq i < \rho$, one has $\mathbb{E}[X^i] < \infty$.*
2. *For every $i > \rho$, one has $\mathbb{E}[X^i] = \infty$.*

Proof. Fix $i > 0$. Using Tonelli's Theorem

$$X^i = \int_0^X i x^{i-1}\, dx \quad \Longrightarrow \quad \mathbb{E}[X^i] = \mathbb{E}\left[\int_0^\infty i x^{i-1} \mathbf{1}_{\{X>x\}}\, dx\right] = \int_0^\infty i x^{i-1} \mathbb{P}(X > x)\, dx.$$

Hence

$$\mathbb{E}[X^i] = i \int_0^\infty x^{i-1} \overline{F}(x)\, dx. \tag{3.10}$$

By assumption $f \in \mathcal{RV}(-\rho)$, i.e. $\overline{F}(x) = x^{-\rho} L(x)$, so the integrand satisfies

$$x^{i-1}\overline{F}(x) = x^{i-1-\rho} L(x) \in \mathcal{RV}(i - 1 - \rho).$$

Set

$$u(x) := x^{i-1}\overline{F}(x) = x^{i-1-\rho} L(x)$$

and apply Karamata's Theorem distinguishing the three cases.

Case $i < \rho$. Then $i - 1 - \rho < -1$ and Karamata's theorem on integrals of regularly varying functions (see Theorem 1.5.11 in Bingham et al. [2013]), yields that

$$\int_0^\infty u(x)\, dx < \infty \quad \Leftrightarrow \quad \mathbb{E}[X^i] < \infty.$$

Case $i > \rho$. Then $i - 1 - \rho > -1$ and because $i > \rho$ we can fix $\varepsilon > 0$ such that $i - 1 - \rho - \varepsilon > -1$. By Potter bounds (3.6) applied to the slowly varying function L, there exist constants $x_0 > 0$ such that for all $x \geq x_0$,

$$L(x) \geq (1 - \varepsilon) x^{-\varepsilon}.$$

Hence for $x \geq x_0$,

$$u(x) = x^{i-1-\rho} L(x) \geq (1 - \varepsilon) x^{i-1-\rho-\varepsilon}.$$

Therefore,

$$\int_{x_0}^\infty u(x)\, dx \geq (1 - \varepsilon) \int_{x_0}^\infty x^{\,i-1-\rho-\varepsilon}\, dx.$$

Since $i - 1 - \rho - \varepsilon > -1$, the power integral on the right hand side diverges. Consequently,

$$\int_0^\infty x^{i-1} \bar{F}(x)\, dx = \infty,$$

and thus

$$\mathbb{E}[X^i] = \infty.$$

Case $i = \rho$. In this case $u(x) = x^{-1}L(x)$, so finiteness of $\mathbb{E}[X^i]$ depends on L. For instance:

$$L(x) \equiv 1 \;\Rightarrow\; \int_0^\infty \frac{1}{x}\,dx = \infty \;\Rightarrow\; \mathbb{E}[X^i] = \infty,$$

while

$$L(x) = \frac{1}{1+(\log x)^2} \;\Rightarrow\; \int_0^\infty \frac{1}{x(1+(\log x)^2)}\,dx < \infty \;\Rightarrow\; \mathbb{E}[X^i] < \infty.$$

□

To further illustrate the usefulness of regularly varying functions, we note their behavior under different operations such as taking minima, maxima, and sums.

Proposition 3.17. *Suppose that the random variables X and Y are independent and regularly varying of index $-\alpha_X$ and $-\alpha_Y$ respectively.*

1. $\min(X, Y)$ *is regularly varying with index* $-(\alpha_X + \alpha_Y)$.
2. $\max(X, Y)$ *is regularly varying with index* $-\min(\alpha_X, \alpha_Y)$.
3. $X + Y$ *is regularly varying with index* $-\min(\alpha_X, \alpha_Y)$.
   Moreover, $\mathbb{P}(X + Y > t) \sim \mathbb{P}(\max(X, Y) > t)$.

Proof. See Theorem 2.18 in Nair et al. [2022]. □

Although these results are intuitive, they are not trivial and do not generally hold for broader classes of heavy-tailed distributions. This highlights once more the analytical tractability and robustness of regularly varying distributions.

3.3 Subexponentiality

A central question when studying sums of random variables is the following: what is the most likely cause of an unusually large sum? At a heuristic level, there are two fundamentally different explanations. Either the sum is large because many values that are larger than average combine together, or because a single exceptionally large value dominates the sum. These two mechanisms correspond to what are commonly referred to as the *conspiracy principle* and the *catastrophe principle*, respectively.

The distinction between these two behaviors is closely tied to the tail of the distribution. Light-tailed distributions typically exhibit a conspiracy-type behavior, where large deviations are the result of many comparable contributions. In contrast, heavy-tailed distributions are characterized by a catastrophe-type behavior, where a single extreme observation dominates.

3.3.1 Conspiracies and catastrophes

We now formalize these two principles.

Definition 3.18 (Catastrophe principle)**.** *A distribution F on $\mathbb{R}_+$ is said to satisfy the* catastrophe principle *if, whenever $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with distribution F, then for every $n \geq 2$,*

$$\mathbb{P}(X_1 + \cdots + X_n > t) \sim \mathbb{P}(\max(X_1, \ldots, X_n) > t), \quad \text{as } t \to \infty.$$

This definition captures the idea that large values of the sum are typically caused by a single large observation.

Definition 3.19 (Conspiracy principle)**.** *A distribution F on $\mathbb{R}_+$ is said to satisfy the* conspiracy principle *if, whenever $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with distribution F, then for every $n \geq 2$,*

$$\mathbb{P}(\max(X_1, \ldots, X_n) > t) = o\big(\mathbb{P}(X_1 + \cdots + X_n > t)\big), \quad \text{as } t \to \infty.$$

In this case, large values of the sum arise from the combined contribution of many variables, none of which dominates individually.

3.3.2 Subexponential distributions

Since our focus is on heavy-tailed phenomena, we concentrate on distributions that exhibit the catastrophe principle. This leads to the class of subexponential distributions.

Definition 3.20. *A distribution F with support $\mathbb{R}_+$ is* subexponential *if, whenever $X_1, X_2, \ldots, X_n$ are i.i.d. random variables with distribution F, then for every $n \geq 2$,*

$$\mathbb{P}(X_1 + \cdots + X_n > t) \sim n\,\mathbb{P}(X_1 > t), \quad \text{as } t \to \infty.$$

At first glance, this definition may appear unrelated to the catastrophe principle. However, these two notions are in fact equivalent.

Proposition 3.21. *Consider $X_1, X_2, \ldots$ i.i.d. random variables with distribution F having support $\mathbb{R}_+$. The following statements are equivalent*

(i) F is subexponential, i.e.

$$\mathbb{P}(X_1 + X_2 + \cdots + X_n > t) \sim n\mathbb{P}(X_1 > t) \text{ for all } n \geq 2.$$

(ii) F satisfies the catastrophe principle, i.e.

$$\mathbb{P}(\max(X_1, X_2, \ldots, X_n) > t) \sim \mathbb{P}(X_1 + X_2 + \cdots + X_n > t) \text{ for all } n \geq 2.$$

Proof. Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution F. Recall that

$$\mathbb{P}\Big(\max(X_1, \ldots, X_n) \leq t\Big) = F(t)^n,$$

and therefore

$$\mathbb{P}\Big(\max(X_1,\dots,X_n)>t\Big)=1-F(t)^n=1-\big(1-\overline{F}(t)\big)^n.$$

Now compare this quantity to $\mathbb{P}(X_1>t)=\overline{F}(t)$. Using the binomial expansion, we obtain

$$1-\big(1-\overline{F}(t)\big)^n=1-\left(1-n\overline{F}(t)+\binom{n}{2}\overline{F}(t)^2-\cdots\right),$$

so that

$$1-\big(1-\overline{F}(t)\big)^n=n\overline{F}(t)-\binom{n}{2}\overline{F}(t)^2+O\big(\overline{F}(t)^3\big)=n\overline{F}(t)+o(\overline{F}(t)).$$

Dividing by $\overline{F}(t)$ gives

$$\frac{\mathbb{P}\big(\max(X_1,\dots,X_n)>t\big)}{\mathbb{P}(X_1>t)}=\frac{1-\big(1-\overline{F}(t)\big)^n}{\overline{F}(t)}=n+o(1).$$

Hence,

$$\mathbb{P}\big(\max(X_1,\dots,X_n)>t\big)\sim n\,\mathbb{P}(X_1>t). \tag{3.11}$$

Therefore, if F is subexponential, that is,

$$\mathbb{P}(X_1+\cdots+X_n>t)\sim n\,\mathbb{P}(X_1>t),$$

then necessarily

$$\mathbb{P}(X_1+\cdots+X_n>t)\sim\mathbb{P}\big(\max(X_1,\dots,X_n)>t\big).$$

Conversely, suppose that the catastrophe principle holds, that is,

$$\mathbb{P}\big(\max(X_1,\dots,X_n)>t\big)\sim\mathbb{P}(X_1+\cdots+X_n>t).$$

As shown above (3.11), independently of subexponentiality, we have

$$\mathbb{P}\big(\max(X_1,\dots,X_n)>t\big)\sim n\mathbb{P}(X_1>t).$$

Therefore,

$$\mathbb{P}(X_1+\cdots+X_n>t)\sim\mathbb{P}\big(\max(X_1,\dots,X_n)>t\big)\sim n\mathbb{P}(X_1>t).$$

Hence F is subexponential by transitivity.

This equivalence formalizes the idea that large sums are driven by a single large observation, which is precisely the catastrophe principle. □

In fact, it was shown in Proposition 3.17 that the catastrophe principle holds for regularly varying distributions. More generally, regularly varying tails form a subclass of subexponential distributions, and hence regular variation implies subexponentiality.

We now state a useful characterization of subexponential distributions, which allows one to verify subexponentiality through the asymptotic behavior of sums.

Proposition 3.22. *Let $X_1, X_2, \ldots, X_n$ be independent random variables with distribution F having support $\mathbb{R}_+$. The following statements are equivalent:*

(i) F *is subexponential, i.e.* $\mathbb{P}(X_1 + \cdots + X_n > t) \sim n\mathbb{P}(X_1 > t)$ *for all* $n \geq 2$.

(ii) $\mathbb{P}(X_1 + X_2 > t) \sim 2\mathbb{P}(X_1 > t)$.

(iii) $\mathbb{P}(X_1 + \cdots + X_n > t) \sim n\mathbb{P}(X_1 > t)$ *for some* $n \geq 2$.

We introduce the notation for the *Complementary Cumulative Distribution Function* of the sum of n i.i.d. random variables having distribution F as follows,

$$\overline{F^{n*}}(t) := \mathbb{P}(X_1 + X_2 + \cdots + X_n > t).$$

Proof. The proof proceeds by showing the chain of implications $(i) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$, which establishes the equivalence of the three statements.

$(i) \Rightarrow (iii)$: It follows immediately from the definition of subexponentiality.

$(iii) \Rightarrow (ii)$: The following argument is adapted from the proof of Lemma 1.3.4 in Embrechts et al. [1997]. Assume that for some $n \geq 2$,

$$\mathbb{P}(X_1 + \cdots + X_n > t) \sim n\,\mathbb{P}(X_1 > t),$$

that is,

$$\overline{F^{n*}}(t) \sim n\,\overline{F}(t) \qquad \text{as } t \to \infty.$$

So,

$$\limsup_{t\to\infty} \frac{\overline{F^{n*}}(t)}{\overline{F}(t)} = n. \tag{3.12}$$

By convolution,

$$\overline{F^{(n+1)*}}(t) = \overline{F^{n*}}(t) + \int_0^t \mathbb{P}(X_{n+1} > t - x)\, dF^{\,n*}(x). \tag{3.13}$$

Indeed, let $S_n := X_1 + \cdots + X_n$. Then

$$\overline{F^{(n+1)*}}(t) = \mathbb{P}(S_{n+1} > t) = \mathbb{P}(S_n + X_{n+1} > t).$$

Notice that

$$\{S_n + X_{n+1} > t\} = \{S_n > t\} \cup \{S_n \leq t,\ X_{n+1} > t - S_n\},$$

where the union is disjoint. Hence,

$$\mathbb{P}(S_n + X_{n+1} > t) = \mathbb{P}(S_n > t) + \mathbb{P}(S_n \leq t,\ X_{n+1} > t - S_n). \tag{3.14}$$

For the second term, we condition on the value of S_n. Using the independence of S_n and X_{n+1}, for $x \in [0, t]$, we get

$$\mathbb{P}(X_{n+1} > t - S_n \mid S_n = x) = \mathbb{P}(X_{n+1} > t - x).$$

Averaging over all possible values of S_n yields

$$\mathbb{P}(S_n \le t,\ X_{n+1} > t - S_n) = \int_0^t \mathbb{P}(X_{n+1} > t - x)\, dF^{\,n*}(x).$$

Substituting this into (3.14) gives

$$\overline{F^{(n+1)*}}(t) = \overline{F^{n*}}(t) + \int_0^t \mathbb{P}(X_{n+1} > t - x)\, dF^{\,n*}(x). \tag{3.15}$$

Now observe that

$$\int_0^t \overline{F}(t-x)\, dF^{\,n*}(x) \ \ge\ \overline{F}(t)\, F^{\,n*}(t). \tag{3.16}$$

Indeed, the tail function $\overline{F}$ is decreasing. Since $0 \le x \le t$, we have

$$t - x \le t \quad \Longrightarrow \quad \overline{F}(t-x) \ge \overline{F}(t).$$

Hence,

$$\int_0^t \overline{F}(t-x)\, dF^{\,n*}(x) \ \ge\ \overline{F}(t) \int_0^t dF^{\,n*}(x) = \overline{F}(t)\, F^{\,n*}(t).$$

Combining (3.15) and (3.16), we have

$$\overline{F^{(n+1)*}}(t) \ \ge\ \overline{F^{n*}}(t) + \overline{F}(t)\, F^{\,n*}(t).$$

Dividing by $\overline{F}(t)$ and taking the limsup as $t \to \infty$, we obtain

$$\limsup_{t\to\infty} \frac{\overline{F^{(n+1)*}}(t)}{\overline{F}(t)} \ \ge\ \limsup_{t\to\infty} \left(\frac{\overline{F^{n*}}(t)}{\overline{F}(t)} + 1 \right).$$

Denoting

$$L_n := \limsup_{t\to\infty} \frac{\overline{F^{n*}}(t)}{\overline{F}(t)},$$

we conclude that

$$L_{n+1} \ge L_n + 1. \tag{3.17}$$

Thus, each time one more summand is added, the lim sup of the tail ratio increases by at least one. In particular, by assumption (3.12) $L_n \le n$, then necessarily $L_2 \le 2$, which is what we want to prove. Indeed, by the recursive bound (3.17) above we obtain

$$\begin{aligned} L_{n-1} &\le L_n - 1 \le n - 1, \\ L_{n-2} &\le L_{n-1} - 1 \le n - 2, \\ &\dots \end{aligned}$$

and after $n-2$ steps,

$$L_2 \le n - (n-2) = 2,$$

which concludes the proof for $(iii) \Rightarrow (ii)$.

The proof of $(ii) \Rightarrow (i)$ is adapted from Chistyakov [1964], relying in particular on Lemmas 1 and 2 and Theorem 1 therein.

To relate the tail behavior of sums to that of a single summand, it is convenient to work with normalized tail ratios. These ratios measure how far the tail of the convolution deviates from linear growth. Specifically, for $t \geq 0$, we define

$$\gamma_n(t) := \frac{1 - \mathbb{P}(S_n \leq t)}{1 - \mathbb{P}(X_1 \leq t)} = \frac{1 - F_n(t)}{1 - F(t)},$$

which compares the tail of the sum S_n to that of a single summand.

The following lemma provides general upper and lower bounds on these ratios, independent of subexponentiality.

Lemma 3.23. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with distribution F, and let $S_n = X_1 + \cdots + X_n$. Then for all $n \geq 2$,*

$$n \;\leq\; \liminf_{t\to\infty} \gamma_n(t) \;\leq\; \limsup_{t\to\infty} \gamma_n(t) \;\leq\; n \limsup_{t\to\infty} \frac{1 - F(t/n)}{1 - F(t)}. \tag{3.18}$$

Proof of Lemma 3.23. First recall that using independence we know that

$$F^{n*}(t) = \mathbb{P}(S_n \leq t) \;\leq\; \mathbb{P}\Big(\max_{1\leq k\leq n} X_k \leq t\Big) = F(t)^n.$$

Hence,

$$1 - F^{n*}(t) \;\geq\; 1 - F(t)^n = (1 - F(t))\big(1 + F(t) + \cdots + F(t)^{n-1}\big).$$

Dividing by $1 - F(t)$ yields

$$\frac{1 - F^{n*}(t)}{1 - F(t)} \;\geq\; 1 + F(t) + \cdots + F(t)^{n-1} \;\xrightarrow[t\to\infty]{}\; n,$$

since $F(t) \to 1$ as $t \to \infty$. This proves the lower bound in (3.18).
For the upper bound notice that if $S_n > t$, then at least one summand must satisfy $X_k > \frac{t}{n}$, since otherwise $S_n < t$. Hence,

$$\{S_n > t\} \subset \bigcup_{k=1}^{n} \left\{ X_k > \frac{t}{n} \right\}.$$

Taking probabilities and using independence, we get

$$1 - F^{n*}(t) = \mathbb{P}(S_n > t) \leq \sum_{k=1}^{n} \mathbb{P}\left(X_k > \frac{t}{n} \right) = n\big(1 - F(t/n)\big).$$

Dividing by $1 - F(t)$ yields

$$\frac{1 - F^{n*}(t)}{1 - F(t)} \leq n \frac{1 - F(t/n)}{1 - F(t)}.$$

Taking the lim sup as $t \to \infty$, we obtain

$$\limsup_{t\to\infty} \frac{1-F_n(t)}{1-F(t)} \le n \limsup_{t\to\infty} \frac{1-F(t/n)}{1-F(t)}.$$

This proves the upper bound in (3.18) and concludes the proof. □

Lemma 3.23 shows that control of convolution tails necessarily involves stability properties of the underlying distribution. In particular, to prevent the upper bound from degenerating, the tail of F must be insensitive to bounded shifts. This property is formalized by long-tailedness, which we now derive from the asymptotic behavior of the two-fold convolution.

Lemma 3.24 (Long-tailedness)**.** *Let X, X_1, X_2 independent random variables with distribution F having support $\mathbb{R}_+$. Assume that*

$$\lim_{t\to\infty} \frac{\mathbb{P}(X_1 + X_2 > t)}{\mathbb{P}(X > t)} = 2. \tag{3.19}$$

Then F is long-tailed, which means that for every fixed $u > 0$,

$$\lim_{t\to\infty} \frac{\mathbb{P}(X > t-u)}{\mathbb{P}(X > t)} = 1. \tag{3.20}$$

Proof of Lemma 3.24. Since X_i is nonnegative for $i \in \{1, 2\}$, the tail of the convolution satisfies

$$\mathbb{P}(X_1 + X_2 > t) = \int_0^\infty \mathbb{P}(X_1 > t-s)\, dF(s). \tag{3.21}$$

Splitting the integral at t and using that $\mathbb{P}(X_1 > t-s) = 1$ for $s > t$ together with $F(t) \to 1$, we obtain

$$\begin{aligned}\mathbb{P}(X_1 + X_2 > t) &= \int_0^t \big(1 - F(t-s)\big)\, dF(s) + \int_t^\infty \big(1 - F(t-s)\big)\, dF(s) \\ &= \int_0^t \big(1 - F(t-s)\big)\, dF(s) + (1 - F(t)).\end{aligned}$$

Dividing by $1 - F(t)$ yields

$$\frac{\mathbb{P}(X_1 + X_2 > t)}{1-F(t)} = 1 + \int_0^t \frac{1-F(t-s)}{1-F(t)}\, dF(s). \tag{3.22}$$

By assumption (3.19), the left-hand side converges to 2. Hence,

$$\int_0^t \frac{1-F(t-s)}{1-F(t)}\, dF(s) \;\longrightarrow\; 1, \qquad t \to \infty. \tag{3.23}$$

Fix $u \ge 0$ and split the integral:

$$\int_0^t \frac{1-F(t-s)}{1-F(t)}\, dF(s) = \int_0^u \frac{1-F(t-s)}{1-F(t)}\, dF(s) + \int_u^t \frac{1-F(t-s)}{1-F(t)}\, dF(s).$$

Since $1 - F(\cdot)$ is non-increasing, we have for $s \in [0, u]$

$$\frac{1 - F(t-s)}{1 - F(t)} \geq 1,$$

and therefore

$$\int_0^u \frac{1 - F(t-s)}{1 - F(t)}\, dF(s) \geq F(u). \tag{3.24}$$

For $s \geq u$, we use $t - s \leq t - u$ to obtain

$$\int_u^t \frac{1 - F(t-s)}{1 - F(t)}\, dF(s) \geq \frac{1 - F(t-u)}{1 - F(t)} \big(F(t) - F(u)\big). \tag{3.25}$$

Combining (3.24) and (3.25) gives

$$\int_0^t \frac{1 - F(t-s)}{1 - F(t)}\, dF(s) \geq F(u) + \frac{1 - F(t-u)}{1 - F(t)} \big(F(t) - F(u)\big).$$

Rearranging,

$$\frac{1 - F(t-u)}{1 - F(t)} \leq \frac{\displaystyle\int_0^t \frac{1 - F(t-s)}{1 - F(t)}\, dF(s) - F(u)}{F(t) - F(u)}.$$

Letting $t \to \infty$ and using (3.23) together with $F(t) \to 1$, we obtain

$$\limsup_{t\to\infty} \frac{1 - F(t-u)}{1 - F(t)} \leq \frac{1 - F(u)}{1 - F(u)} = 1.$$

Since the ratio is always at least 1 by (3.24), it follows that

$$\frac{1 - F(t-u)}{1 - F(t)} \longrightarrow 1, \qquad t \to \infty.$$

This completes the proof. □

Under long-tailedness, the normalized tail ratios $\gamma_n(t)$, defined by

$$\gamma_n(t) := \frac{1 - \mathbb{P}(S_n \leq t)}{1 - \mathbb{P}(X_1 \leq t)} = \frac{1 - F_n(t)}{1 - F(t)},$$

admit a precise asymptotic behavior. The following theorem shows that this property alone is sufficient to recover the defining asymptotics of subexponential distributions.

Theorem 3.25. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with distribution F having support $\mathbb{R}_+$. Let $S_n := X_1 + \cdots + X_n$. Assume that F is long-tailed and $\lim_{t\to\infty} \gamma_2(t) = 2$. Then for every $n \in \mathbb{N}$,*

$$\lim_{t\to\infty} \frac{1 - F_n(t)}{1 - F(t)} =: \lim_{t\to\infty} \gamma_n(t) = n. \tag{3.26}$$

Proof of Theorem 3.25. We proceed by induction on n.
Base case. For $n = 1$ we have $\gamma_1(t) = 1$ for all t, and the claim holds. For $n = 2$, the assumption yields

$$\lim_{t\to\infty} \gamma_2(t) = 2,$$

which establishes the base of the induction.

Induction hypothesis. Assume that for some $n \geq 2$,

$$\gamma_n(t) \longrightarrow n \qquad \text{as } t \to \infty.$$

Induction step. Using the convolution formula and conditioning on X_{n+1}, we obtain

$$\begin{aligned}\mathbb{P}(S_{n+1} > t) &= \int_0^\infty \mathbb{P}(S_n > t-u)\, dF(u) \\ &= \overline{F}(t) + \int_0^t \mathbb{P}(S_n > t-u)\, dF(u). \end{aligned} \tag{3.27}$$

Dividing by $\overline{F}(t)$ yields

$$\gamma_{n+1}(t) = 1 + \int_0^t \gamma_n(t-u)\frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u). \tag{3.28}$$

Fix $x > 0$ and split the integral at x

$$\gamma_{n+1}(t) = 1 + \int_0^x \gamma_n(t-u)\frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) + \int_x^t \gamma_n(t-u)\frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u). \tag{3.29}$$

For the first term, define the local supremum

$$\widetilde{\gamma}_n(t,x) := \sup_{t-x\leq s\leq t} \gamma_n(s).$$

Then

$$\int_0^x \gamma_n(t-u)\frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \leq \widetilde{\gamma}_n(t,x)\frac{\overline{F}(t-x)}{\overline{F}(t)}\, F(x).$$

Since F is long-tailed and $\gamma_n(t) \to n$ by hypothesis, we obtain

$$\widetilde{\gamma}_n(t,x) \longrightarrow n, \qquad \frac{\overline{F}(t-x)}{\overline{F}(t)} \longrightarrow 1, \qquad \text{as } t \to \infty. \tag{3.30}$$

For the second term in (3.29), we use that, by the induction hypothesis, γ_n is bounded as $t \to \infty$. Together with boundedness on compact intervals, this implies that there exists a constant $K_n < \infty$ such that $\gamma_n(t) \leq K_n$ for all $t \geq 0$. Hence,

$$\int_x^t \gamma_n(t-u)\frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \leq K_n \int_x^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u).$$

We claim that

$$\int_x^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \longrightarrow 1 - F(x), \qquad t \to \infty. \tag{3.31}$$

Indeed, set

$$I(t) := \int_0^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u).$$

Since

$$\begin{aligned}\gamma_2(t) &= \frac{\mathbb{P}(X_1 + X_2 > t)}{\overline{F}(t)} = \frac{1}{\overline{F}(t)} \int_0^\infty \mathbb{P}(X_1 > t-u)\, dF(u) \\ &= \frac{1}{\overline{F}(t)} \left(\int_0^t \overline{F}(t-u)\, dF(u) + \overline{F}(t) \right) = 1 + \int_0^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u).\end{aligned}$$

the assumption $\gamma_2(t) \to 2$ implies $I(t) \to 1$.
Moreover, since $\overline{F}$ is decreasing, for every fixed u with $0 \le u \le t$,

$$\frac{\overline{F}(t-u)}{\overline{F}(t)} \ge 1.$$

Hence

$$\int_0^x \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \ge F(x).$$

On the other hand,

$$\int_x^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) = I(t) - \int_0^x \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \le I(t) - F(x).$$

Therefore,

$$\limsup_{t\to\infty} \int_x^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \le 1 - F(x).$$

For the reverse inequality, using again

$$\frac{\overline{F}(t-u)}{\overline{F}(t)} \ge 1$$

on $[x, t]$, we get

$$\int_x^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \ge F(t) - F(x).$$

Letting $t \to \infty$ gives

$$\liminf_{t\to\infty} \int_x^t \frac{\overline{F}(t-u)}{\overline{F}(t)}\, dF(u) \ge 1 - F(x).$$

Combining the upper and lower bounds proves (3.31).

Combining the estimates (3.30) and (3.31) in (3.29), we obtain

$$\limsup_{t\to\infty} \gamma_{n+1}(t) \leq 1 + nF(x) + K_n(1 - F(x)).$$

Letting $x \to \infty$ yields

$$\limsup_{t\to\infty} \gamma_{n+1}(t) \leq n + 1.$$

On the other hand, Lemma 3.23 implies the lower bound

$$\liminf_{t\to\infty} \gamma_{n+1}(t) \geq n + 1.$$

Therefore,

$$\lim_{t\to\infty} \gamma_{n+1}(t) = n + 1,$$

which completes the induction. □

We now complete the original proof of Proposition 3.22.

Proof of Proposition 3.22. In order to conclude the proof we will show $(ii) \Rightarrow (i)$.

$(ii) \Rightarrow (i)$: Let the assumptions of Proposition 3.22 hold. Lemma 3.23 and Lemma 3.24 are proved under these assumptions. By Lemma 3.24, condition (3.20) is satisfied. Theorem 3.25 applies under condition (3.20) and yields conclusion

$$\lim_{t\to\infty} \frac{1 - F_n(t)}{1 - F(t)} = \lim_{t\to\infty} \gamma_n(t) = n,$$

which is exactly the statement of Proposition 3.22. This completes the proof. □

The previous results show that subexponential distributions capture the idea that extreme events are driven by a single large observation. This behavior is inherently linked to heavy-tailed phenomena. In fact, this connection holds in general, as every subexponential distribution is necessarily heavy-tailed.

Lemma 3.26. *Every subexponential distribution is heavy-tailed.*

Proof. See Lemma 3.8 in Nair et al. [2022]. □

3.3.3 Conspiracies and catastrophes in random walks

A particularly natural setting in which to compare heavy-tailed and light-tailed phenomena is that of a *random walk*. In this setting, the catastrophe principle and the conspiracy principle become concrete mechanisms that explain how rare events occur. The goal is not merely to estimate the probability of a large deviation, but also to understand what such a rare event typically looks like.

This perspective is especially useful in the present context, as it highlights the fundamental difference between heavy-tailed and light-tailed distributions. Indeed, we shall see that the same rare event can arise through two entirely different mechanisms depending

on the tail behavior of the increments. Our focus is therefore on two complementary phenomena: the *principle of a single big jump*, which governs large deviations in the heavy-tailed setting, and *Cramér's theorem*, which describes the corresponding behavior in the light-tailed case.

We first recall the standard definition of a random walk, which will serve as the basic object of study throughout this section.

Definition 3.27. *Let $(X_i)_{i\geq 1}$ be i.i.d. real-valued random variables with finite mean μ and define the* random walk

$$S_n = X_1 + X_2 + \cdots + X_n, \qquad n \geq 1.$$

Thus S_n represents the position of a walker after n steps, where the i-th step has size X_i.

For a random walk, the typical behavior is described by classical limit theorems. By the law of large numbers,

$$\frac{S_n}{n} \to \mu \qquad \text{a.s.},$$

so typically S_n is close to μn for large n. Moreover, the central limit theorem states that, if the variance is finite, then the fluctuations of order $\sqrt{n}$ around the mean μn as

$$S_n \approx \mu n + \sigma\sqrt{n}\, Z, \quad \text{where } Z \sim \mathcal{N}(0,1).$$

Therefore, if $a > \mu$, the event

$$\{S_n > an\}$$

is atypical: the level an differs from the value μn by a quantity of order n, which is much larger than the $\sqrt{n}$-scale captured by the central limit theorem. Hence such deviations lie outside the regime of normal fluctuations. For this reason, the quantity

$$\mathbb{P}(S_n > an), \qquad a > \mu,$$

is the natural object to study when analyzing large deviations of the random walk. The key point is that the notion of a "large" deviation must be understood relative to the typical scale of S_n. Since S_n is of order n, comparing it to a fixed threshold t is not appropriate: for large n, one typically has $S_n \gg t$, and thus the event $\{S_n > t\}$ eventually becomes trivial.

Catastrophes and conspiracies as explanations of rare events. The catastrophe principle and the conspiracy principle are explanations of how a rare event is created.

In the heavy-tailed case, a rare event is typically caused by a single extraordinary observation. In the random walk setting, this means that if S_n is much larger than its typical value, then the most likely explanation is that one increment X_i is exceptionally large, while the remaining increments behave in an ordinary way. This is the *catastrophe* picture.

In the light-tailed case, the mechanism is very different. A large value of S_n is then typically not caused by one exceptional jump, but by the cumulative effect of many increments that are all somewhat larger than usual. In that sense, the increments "conspire" to produce the rare event. This is the *conspiracy* picture.

The principle of a single big jump. We now turn to the heavy-tailed case. Suppose first that the increments are subexponential. Then for each fixed n,

$$\mathbb{P}(S_n > t) \sim n\,\mathbb{P}(X_1 > t), \qquad t \to \infty.$$

This already shows that the tail of the sum is asymptotically generated by one summand: a large value of S_n is most likely due to one exceptionally large increment.

However, this asymptotic result is formulated for fixed n and $t \to \infty$, whereas in the context of random walks one is primarily interested in the behavior as $n \to \infty$. Therefore, as already seen, one should study

$$\mathbb{P}(S_n > an), \qquad a > \mu \qquad \text{as} \quad n \to \infty.$$

If the catastrophe principle remains valid in this regime, then a heuristic argument leads to the following approximation. In order for S_n to exceed an, one expects that one variable, say X_i, is of order n, while the remaining $n-1$ variables contribute approximately their mean μ. Thus one informally writes

$$S_n \approx X_i + \mu(n-1),$$

and so the event $\{S_n > an\}$ should require

$$X_i \approx an - \mu(n-1) \approx (a-\mu)n.$$

Since any one of the n increments could play the role of this big jump, this suggests the approximation

$$\mathbb{P}(S_n > an) \approx n\,\mathbb{P}(X_1 > (a-\mu)n),$$

which is exactly the content of the principle of a single big jump.

Another way to see the plausibility of this principle is to compare it with scenarios involving two or more large jumps. If two increments were both of order n, their contribution would typically have probability of order $\big(n\,\mathbb{P}(X_1 > n)\big)^2$, which is of smaller order than the one-jump term $n\,\mathbb{P}(X_1 > (a-\mu)n)$.

Hence, the contribution of scenarios involving two or more large jumps is negligible compared to that of a single large jump. This confirms that exceedances of high levels are typically generated by exactly one exceptionally large increment.

The above argument is purely heuristic. The following theorem shows that, under regular variation of the increment distribution, the probability of a large exceedance is asymptotically equivalent to the probability that one increment exceeds the required threshold. In this sense, high exceedances are generated by a single big jump.

Theorem 3.28 (The principle of a single big jump)**.** *Suppose X_i are i.i.d. with finite mean μ and follow a regularly varying distribution with index $-\alpha$, where $\alpha > 1$. Then for $a > \mu$, the random walk $S_n = X_1 + \cdots + X_n$ satisfies*

$$\lim_{n\to\infty} \frac{\mathbb{P}(S_n > an)}{n\,\mathbb{P}(X_1 > (a-\mu)n)} = 1.$$

Proof. We establish the claim by proving matching asymptotic lower and upper bounds for $\Pr(S_n > an)$. The main idea is that, for heavy-tailed increments, the event $S_n > an$ is typically generated by a single exceptionally large summand. The lower bound is obtained by explicitly constructing such a scenario, while the upper bound shows that configurations without a large jump contribute only negligibly to the probability.

Without loss of generality, assume that $\mu = 0$ and $a > 0$. If $\mu \neq 0$, one may replace X_i by $X_i - \mu$ and a by $a - \mu$. We prove the result only in the case $\alpha > 2$. The case $\alpha \in (1, 2]$ requires additional truncation estimates and is omitted.

Lower bound. We first prove that

$$\liminf_{n\to\infty} \frac{\Pr(S_n > an)}{n \Pr(X_1 > an)} \geq 1.$$

Fix a constant $\delta > a$. Consider the event that one of the increments exceeds δn, while the remaining increments collectively contribute enough to push the sum above the level an. For each $i \in 1, \ldots, n$ define

$$B_i^{(n)} = \{X_i > \delta n, \sum_{j\neq i} X_j > (a - \delta)n\}.$$

Clearly,

$$\bigcup_{i=1}^{n} B_i^{(n)} \subseteq S_n > an,$$

and therefore

$$\mathbb{P}\left(\bigcup_{i=1}^{n} B_i^{(n)}\right) \leq \mathbb{P}(S_n > an).$$

Applying the inclusion-exclusion principle gives

$$\mathbb{P}\left(\bigcup_{i=1}^{n} B_i^{(n)}\right) \geq \sum_{i=1}^{n} \mathbb{P}(B_i^{(n)}) - \sum_{i<j} \mathbb{P}(B_i^{(n)} \cap B_j^{(n)}).$$

Since the random variables are i.i.d., the probabilities $\mathbb{P}(B_i^{(n)})$ are identical, and hence

$$\mathbb{P}(S_n > an) \geq n\mathbb{P}(B_1^{(n)}) - \frac{n(n-1)}{2}\mathbb{P}(B_1^{(n)} \cap B_2^{(n)}). \tag{3.32}$$

The intersection term is easily controlled by

$$\mathbb{P}(B_1^{(n)} \cap B_2^{(n)}) \leq \mathbb{P}(X_1 > \delta n, X_2 > \delta n) = \mathbb{P}(X_1 > \delta n)^2. \tag{3.33}$$

Furthermore, by independence,

$$\mathbb{P}(B_1^{(n)}) = \mathbb{P}(X_1 > \delta n)\mathbb{P}\left(\sum_{j=2}^{n} X_j > (a - \delta)n\right). \tag{3.34}$$

Because $\delta > a$, we have $a - \delta < 0$ and since $\frac{1}{n}\sum_{j=2}^{n} X_j \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ by the weak law of large numbers, it follows that

$$\mathbb{P}\left(\sum_{j=2}^{n} X_j > (a-\delta)n\right) \longrightarrow 1$$

as $n \to \infty$. Combining (3.32), (3.33), and (3.34) yields

$$\mathbb{P}(S_n > an) \geq n\mathbb{P}(X_1 > \delta n)\mathbb{P}\left(\sum_{j=2}^{n} X_j > (a-\delta)n\right) - \frac{n(n-1)}{2}\mathbb{P}(X_1 > \delta n)^2.$$

Dividing by $n\mathbb{P}(X_1 > an)$ yields

$$\frac{\mathbb{P}(S_n > an)}{n\mathbb{P}(X_1 > an)} \geq \frac{\mathbb{P}(X_1 > \delta n)}{\mathbb{P}(X_1 > an)}\,\mathbb{P}\left(\sum_{j=2}^{n} X_j > (a-\delta)n\right) - \frac{n(n-1)\mathbb{P}(X_1 > \delta n)^2}{2n\mathbb{P}(X_1 > an)}.$$

Moreover,

$$\frac{n(n-1)\mathbb{P}(X_1 > \delta n)^2}{n\mathbb{P}(X_1 > an)} = (n-1)\mathbb{P}(X_1 > an)\left(\frac{\mathbb{P}(X_1 > \delta n)}{\mathbb{P}(X_1 > an)}\right)^2.$$

Since $\mathbb{P}(X_1 > x)$ is regularly varying with index $-\alpha$,

$$\frac{\mathbb{P}(X_1 > \delta n)}{\mathbb{P}(X_1 > an)} \longrightarrow \left(\frac{\delta}{a}\right)^{-\alpha}$$

as $n \to \infty$, while

$$n\mathbb{P}(X_1 > an) = a^{-\alpha} n^{1-\alpha} L(an) \longrightarrow 0$$

as $n \to \infty$, because $\alpha > 1$. Therefore

$$\frac{n(n-1)\mathbb{P}(X_1 > \delta n)^2}{n\mathbb{P}(X_1 > an)} \longrightarrow 0$$

as $n \to \infty$. Thus, passing to the limit inferior gives

$$\liminf_{n\to\infty} \frac{\mathbb{P}(S_n > an)}{n\mathbb{P}(X_1 > an)} \geq \lim_{n\to\infty} \frac{\mathbb{P}(X_1 > \delta n)}{\mathbb{P}(X_1 > an)} = \left(\frac{\delta}{a}\right)^{-\alpha}.$$

Since $\delta > a$ was arbitrary, letting $\delta \downarrow a$ gives

$$\liminf_{n\to\infty} \frac{\mathbb{P}(S_n > an)}{n\mathbb{P}(X_1 > an)} \geq 1.$$

Upper bound. We now prove the reverse inequality,

$$\limsup_{n\to\infty} \frac{\mathbb{P}(S_n > an)}{n\mathbb{P}(X_1 > an)} \leq 1.$$

Set

$$A_n = \{S_n > an\}.$$

The idea is to split this event according to whether at least one summand is of order n. Fix $\tau \in (0, a)$ and write

$$\mathbb{P}(A_n) = \mathbb{P}(A_n, \max_{1\leq i\leq n} X_i > \tau n) + \mathbb{P}(A_n, \max_{1\leq i\leq n} X_i \leq \tau n). =: \mathbb{P}(A_{1,n}) + \mathbb{P}(A_{2,n}). \quad (3.35)$$

We first estimate $A_{1,n}$. Choose $\delta \in (\tau, a)$. Then

$$\begin{aligned}\mathbb{P}(A_{1,n}) &\leq \mathbb{P}\left(\max_{1\leq i\leq n} X_i > \delta n\right) + \mathbb{P}\left(S_n > an, \tau n < \max_{1\leq i\leq n} X_i \leq \delta n\right) \\ &\leq n\mathbb{P}(X_1 > \delta n) + n\mathbb{P}(X_1 > \tau n)\mathbb{P}\left(\sum_{j=2}^{n} X_j > n(a-\delta)\right).\end{aligned}$$

Dividing by $n\mathbb{P}(X_1 > an)$ yields

$$\frac{\mathbb{P}(A_{1,n})}{n\mathbb{P}(X_1 > an)} \leq \frac{\mathbb{P}(X_1 > \delta n)}{\mathbb{P}(X_1 > an)} + \frac{\mathbb{P}(X_1 > \tau n)}{\mathbb{P}(X_1 > an)}\mathbb{P}\left(\sum_{j=2}^{n} X_j > n(a-\delta)\right).$$

Since $a - \delta > 0$, the weak law of large numbers implies

$$\mathbb{P}\left(\sum_{j=2}^{n} X_j > (a-\delta)n\right) \longrightarrow 0$$

as $n \to \infty$. Moreover, regular variation yields

$$\frac{\mathbb{P}(X_1 > \tau n)}{\mathbb{P}(X_1 > an)} \longrightarrow \left(\frac{\tau}{a}\right)^{-\alpha}$$

as $n \to \infty$ and hence this ratio remains bounded. Therefore,

$$\frac{\mathbb{P}(X_1 > \tau n)}{\mathbb{P}(X_1 > an)}\mathbb{P}\left(\sum_{j=2}^{n} X_j > (a-\delta)n\right) \longrightarrow 0$$

as $n \to \infty$. Taking the limit superior in the previous inequality, we obtain

$$\limsup_{n\to\infty} \frac{\mathbb{P}(A_{1,n})}{n\mathbb{P}(X_1 > an)} \leq \lim_{n\to\infty} \frac{\mathbb{P}(X_1 > \delta n)}{\mathbb{P}(X_1 > an)} = \left(\frac{\delta}{a}\right)^{-\alpha}.$$

Letting $\delta \uparrow a$ gives

$$\limsup_{n\to\infty} \frac{\mathbb{P}(A_{1,n})}{n\mathbb{P}(X_1 > an)} \leq 1.$$

It remains to show that the second term in (3.35) is negligible, namely that for a suitable choice of $\tau \in (0, a)$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(A_{2,n})}{n\mathbb{P}(X_1 > an)} = 0.$$

Recall that

$$A_{2,n} = \{S_n > an,\ \max_{1\leq i\leq n} X_i \leq \tau n\}.$$

Let $X_1', \ldots, X_n'$ be i.i.d. random variables with the conditional distribution

$$\mathbb{P}(X_1^{'} \leq x) = \mathbb{P}(X_1 \leq x \mid X_1 \leq \tau n).$$

Then

$$\mathbb{P}(A_{2,n}) \leq \mathbb{P}\left(\sum_{i=1}^{n} X_i^{'} > an\right).$$

Set

$$\mu_n := \mathbb{E}[X_1'] = \mathbb{E}[X_1 \mid X_1 \leq \tau n].$$

Since $\mathbb{E}[X_1] = 0$, we have

$$\mu_n = -\frac{\mathbb{E}[X_1 \mathbf{1}_{\{X_1 > \tau n\}}]}{\mathbb{P}(X_1 \leq \tau n)} \leq 0,$$

and moreover $\mu_n \to 0$. Hence, for all sufficiently large n,

$$X_i' - \mu_n = X_i' + |\mu_n| \leq 2\tau n.$$

Define $Y_i' = X_i' - \mu_n$. Then the Y_i' are independent, centered, and bounded above by $2\tau n$. Also, assume $\alpha > 2$ such that $\mathbb{E}[X_1^2] < \infty$, in order to obtain

$$\mathrm{Var}(Y_i') = \mathrm{Var}(X_i') \leq \mathbb{E}[(X_i^*)^2] = \frac{\mathbb{E}[X_1^2\, \mathbf{1}_{\{X_1 \leq \tau n\}}]}{\mathbb{P}(X_1 \leq \tau n)} \leq \frac{\mathbb{E}[X_1^2]}{\mathbb{P}(X_1 \leq \tau n)}.$$

Since $\mathbb{P}(X_1 \leq \tau n) \to 1$, there exists a constant $C > 0$, independent of n, such that for all sufficiently large n

$$\mathrm{Var}(Y_i') \leq C.$$

Since $\mu_n \leq 0$,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i' > an\right) \leq \mathbb{P}\left(\sum_{i=1}^{n} Y_i' > an\right).$$

We now apply Prokhorov's inequality to the centered variables Y_i'.

Recall that the Prokhorov's inequality (see Prokhorov [1959]) states that if $Z_1, \ldots, Z_n$ are independent random variables with

$$\mathbb{E}[Z_i] = 0, \qquad Z_i \leq c$$

almost surely for all i, where $c > 0$ and $T_n = \sum_{i=1}^n \mathrm{Var}(Z_i)$, then for every $x > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n Z_i > x\right) \leq \left(\frac{cx + T_n}{T_n}\right)^{-x/(2c)}.$$

Since $Y_i' \leq 2\tau n$ and

$$T_n = \sum_{i=1}^n \mathrm{Var}(Y_i') \leq Cn,$$

we may apply Prokhorov's inequality with $c = 2\tau n$ and $x = an$. This yields

$$\mathbb{P}\left(\sum_{i=1}^n Y_i' > an\right) \leq \left(\frac{2a\tau n^2 + Cn}{Cn}\right)^{-a/(4\tau)}.$$

Since

$$\frac{2a\tau n^2 + Cn}{Cn} = 1 + \frac{2a\tau}{C}n,$$

there exists a constant $C' > 0$ such that

$$\mathbb{P}(A_{2,n}) \leq C' n^{-a/(4\tau)}$$

for all sufficiently large n.

Now, since $\mathbb{P}(X_1 > x)$ is regularly varying with index $-\alpha$,

$$n\mathbb{P}(X_1 > an) = a^{-\alpha} n^{1-\alpha} L(an)$$

for some slowly varying function L. Hence $n\mathbb{P}(X_1 > an)$ is regularly varying with index $1 - \alpha$. Therefore

$$\frac{\mathbb{P}(A_{2,n})}{n\mathbb{P}(X_1 > an)} \leq C' a^\alpha \frac{n^{-a/(4\tau)}}{n^{1-\alpha} L(an)} = C' a^\alpha \frac{n^{(\alpha-1)-a/(4\tau)}}{L(an)}.$$

Choose $\tau > 0$ sufficiently small so that

$$\frac{a}{4\tau} > \alpha - 1 \iff (\alpha - 1) - \frac{a}{4\tau} < 0.$$

Therefore,

$$\frac{n^{(\alpha-1)-a/(4\tau)}}{L(an)} \longrightarrow 0$$

as $n \to \infty$. Consequently,

$$\frac{\mathbb{P}(A_{2,n})}{n\mathbb{P}(X_1 > an)} \longrightarrow 0$$

as $n \to \infty$. Combining the estimates for $A_{1,n}$ and $A_{2,n}$ yields

$$\limsup_{n\to\infty} \frac{\mathbb{P}(S_n > an)}{n\mathbb{P}(X_1 > an)} \leq 1.$$

Together with the lower bound, this proves the desired asymptotic equivalence in the case $\alpha > 2$.

The case $\alpha \in (1, 2]$ requires a more delicate analysis, since the variance may grow with n and the above application of Prokhorov's inequality is no longer sufficient. We omit the proof and refer to Zwart [2008] for the full argument. □

The one big jump principle extends beyond regularly varying distributions to the broader class of subexponential distributions. Since the general statement requires additional technical assumptions, we restrict our discussion here to regularly varying tails and refer the reader to Appendix A for the general subexponential framework.

Cramér's theorem. We now consider the light-tailed case. The appropriate refinement of the conspiracy principle is given by Cramér's theorem. In contrast to the principle of a single big jump, large exceedances of the random walk are no longer caused by a single exceptionally large increment. Instead, they arise from a collective upward fluctuation of many increments, each contributing moderately to the deviation.

As in the heavy-tailed case, our goal is to understand the asymptotic behavior of

$$\mathbb{P}(S_n > an), \qquad a > \mu,$$

but the resulting estimates will have a markedly different form. In particular, for light-tailed distributions, these probabilities decay exponentially fast in n, in contrast to the much slower decay observed in the heavy-tailed setting.

The starting point for this analysis is the *Chernoff bound*, which form the basis of Cramér's theorem. The Chernoff bound is obtained by applying Markov's inequality to the exponential transformation of the sum. Recall that for any nonnegative random variable X and any $t > 0$, *Markov's inequality* states that

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$

Applying this inequality to $Y = e^{sS_n}$ yields the Chernoff bound. More precisely, for any $s > 0$,

$$\mathbb{P}(S_n \geq an) = \mathbb{P}\big(e^{sS_n} \geq e^{san}\big) \leq e^{-ans}\,\mathbb{E}\big[e^{sS_n}\big].$$

Since the increments are independent, this yields

$$\mathbb{P}(S_n \geq an) \leq \exp\Big(-n\big(as - \log \mathbb{E}[e^{sX_1}]\big)\Big).$$

Optimizing over $s > 0$ leads to the Chernoff bound for $\mathbb{P}(S_n > an)$.
Thus, the exponential decay rate is determined by the moment generating function of X_1, and hence depends on the entire distribution of X_1.

What Cramér's theorem adds is that this exponential upper bound is asymptotically sharp. More precisely, it provides a matching lower bound, showing that the Chernoff bound captures the true exponential rate of decay of $\mathbb{P}(S_n > an)$. In this way, Cramér's

theorem identifies the exact asymptotic behavior of the probabilities on the exponential scale.

Conceptually, this result reveals the mechanism by which such rare events occur. Under the condition $\{S_n > an\}$, the increments behave as if they were sampled from an exponentially tilted distribution whose mean exceeds the original mean μ. Thus, the deviation is not caused by a single exceptional increment, but rather by a collective upward shift of many increments. This provides a rigorous formulation of the conspiracy principle.

Theorem 3.29 (Cramér's Theorem)**.** *Let $(X_i)_{i\geq 1}$ be i.i.d. light-tailed random variables and consider $S_n = X_1 + \cdots + X_n$. Then for every $a > \mathbb{E}[X_1]$,*

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(S_n > an) = \sup_{s>0}\left\{as - \log \mathbb{E}[e^{sX_1}]\right\}.$$

This result shows that in the light-tailed regime the probability of a linear deviation decays exponentially fast in n. More importantly, the collective nature of the deviation becomes apparent in the proof through the exponential tilting argument: under the tilted measure, the increments acquire mean a, so that the event $\{S_n > an\}$ is realized through many moderate upward deviations rather than a single exceptionally large jump.

Proof. The upper bound follows directly from Chernoff's inequality. Indeed, for any $s > 0$,

$$\mathbb{P}(S_n > an) \leq e^{-ans}\mathbb{E}[e^{sX_1}]^n,$$

and optimizing over s yields

$$\liminf_{n\to\infty} \frac{-\log \mathbb{P}(S_n > an)}{n} \geq \sup_{s>0}\left(sa - \log \mathbb{E}[e^{sX_1}]\right).$$

We now turn to the matching lower bound. We denote by $\Lambda(s) := \log \mathbb{E}[e^{sX_1}]$ the logarithmic moment generating function of X_1. We prove the result under the simplifying assumptions that X_1 admits a density f and that the moment generating function is steep. For a complete proof see Theorem 2.7.7 in Durrett [2019]. We say that the moment generating function is steep if $\Lambda'(s) \to \infty$ as s approaches the boundary of its domain. This condition ensures that the equation $\Lambda'(s) = a$ admits a solution for any $a > \mathbb{E}[X_1]$. Writing

$$\bar{s} := \sup\{s : \mathbb{E}[e^{sX_1}] < \infty\},$$

we assume $\mathbb{E}[e^{sX_1}] \to \infty$ as $s \uparrow \bar{s}$. Under this condition, the variational problem

$$\sup_{s>0}\left(as - \log \mathbb{E}[e^{sX_1}]\right)$$

admits a unique maximizer $s_a \in (0, \bar{s})$, which satisfies the stationarity condition

$$a = \frac{\mathbb{E}[X_1 e^{s_a X_1}]}{\mathbb{E}[e^{s_a X_1}]}.$$

The maximizer s_a is characterized by the first-order optimality condition

$$a = \Lambda'(s_a) = \frac{\mathbb{E}[X_1 e^{s_a X_1}]}{\mathbb{E}[e^{s_a X_1}]},$$

which corresponds to matching the mean under the tilted measure to the level a.

The idea of the proof is to tilt the original distribution so that the rare event becomes typical. Define a new probability density by

$$\tilde{f}_a(x) := \frac{e^{s_a x} f(x)}{\mathbb{E}[e^{s_a X_1}]}.$$

Let $\tilde{X}$ denote a random variable with density $\tilde{f}_a$. A direct computation shows that

$$\mathbb{E}[\tilde{X}] = a,$$

so under the tilted measure, the sum $\tilde{X}_1 + \cdots + \tilde{X}_n$ concentrates around the threshold an. Let $(\tilde{X}_i)_{i\geq 1}$ be i.i.d. random variables with density $\tilde{f}_a$. We rewrite the probability as

$$\mathbb{P}(S_n > an) = \int_{\{x_1+\cdots+x_n>an\}} \prod_{i=1}^{n} f(x_i)\, dx_i.$$

Multiplying and dividing by the exponential tilt, we obtain

$$\begin{aligned}\mathbb{P}(S_n > an) &= \int_{\{x_1+\cdots+x_n>an\}} e^{-s_a \sum_{i=1}^n x_i} \prod_{i=1}^{n} \left(e^{s_a x_i} f(x_i)\right)\, dx_i \\ &= \mathbb{E}[e^{s_a X_1}]^n \int_{\{x_1+\cdots+x_n>an\}} e^{-s_a \sum_{i=1}^n x_i} \prod_{i=1}^{n} \tilde{f}_a(x_i)\, dx_i.\end{aligned}$$

To obtain a lower bound, we restrict the integration to the narrower set

$$an < x_1 + \cdots + x_n < an + \sqrt{n},$$

on which $e^{-s_a \sum x_i} \geq e^{-s_a(an+\sqrt{n})}$. This gives

$$\mathbb{P}(S_n > an) \geq e^{-s_a(an+\sqrt{n})} \mathbb{E}[e^{s_a X_1}]^n\, \mathbb{P}\Big(an < \tilde{X}_1 + \cdots + \tilde{X}_n < an + \sqrt{n}\Big).$$

Since $\mathbb{E}[\tilde{X}_i] = a$, the central limit theorem implies that

$$\mathbb{P}\left(an < \sum_{i=1}^{n} \tilde{X}_i < an + \sqrt{n}\right) \to \mathbb{P}(0 < N < 1),$$

where N is a centered Gaussian random variable with variance $\mathrm{Var}(\tilde{X}_1)$. In particular, this probability converges to a strictly positive constant.

Taking logarithms, dividing by n, and letting $n \to \infty$, the contribution of the Gaussian term vanishes, and we obtain

$$\limsup_{n\to\infty} \frac{-\log \mathbb{P}(S_n > an)}{n} \leq s_a a - \log \mathbb{E}[e^{s_a X_1}].$$

Since s_a is the maximizer, this gives the desired lower bound and completes the proof. □

The comparison between the principle of a single big jump and Cramér's theorem provides a clear illustration of the qualitative difference between heavy-tailed and light-tailed distributions. In both cases one studies the same event,

$$\mathbb{P}(S_n > an), \qquad a > \mu,$$

but the mechanism producing the event is completely different. For heavy-tailed increments, the exceedance is caused by one exceptional jump; for light-tailed increments, it is caused by a coordinated fluctuation of many increments. This dichotomy is one of the most important conceptual consequences of the theory of heavy-tails and subexponentiality.

Chapter 4

The Generalized Central Limit Theorem

The classical central limit theorem provides a compelling explanation for the ubiquity of the Gaussian distribution: sums of many i.i.d. random variables tend to behave approximately normally. However, as discussed in the previous chapter, heavy-tailed distributions are also widespread in applications. This raises a fundamental question: *If additive processes lead to Gaussian limits, why are heavy-tailed distributions so common?*

One answer is that the classical central limit theorem applies only under specific assumptions, most notably the existence of finite mean and variance. When these assumptions are relaxed, the Gaussian distribution can no longer describe the asymptotic behavior of sums. This suggests that the classical central limit theorem provides only a partial view of a more general phenomenon.

In fact, there exists a broader framework in which additive processes do not necessarily converge to Gaussian distributions, but instead to a larger class of distributions known as *stable distributions*. These distributions extend the notion of stability under summation beyond the Gaussian case and naturally include heavy-tailed behavior.

The goal of this chapter is to develop this more general perspective. We begin by revisiting the classical central limit theorem and highlighting its limitations. We then introduce stable distributions and their fundamental properties, before presenting the generalized central limit theorem, which characterizes all possible limiting distributions of normalized sums.

4.1 Classical Central Limit Theorem

Throughout this chapter, we consider *additive processes* of the form

$$S_n = X_1 + X_2 + \cdots + X_n,$$

where $(X_n)_{n\geq 1}$ are i.i.d. random variables.

The central question is to understand the behavior of S_n as $n \to \infty$. The asymptotic behavior of S_n can be understood by interpreting the law of large numbers and the central limit theorem as first-and second- order approximations of the process S_n.

We first describe the first-order behavior of the additive process, which reflects the averaging effect of summing independent random variables.

Theorem 4.1 (Strong Law of Large Numbers.)**.** *Let $(X_n)_{n\geq 1}$ be an infinite sequence of i.i.d. random variables having finite $\mathbb{E}[X_1] = \mu$. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu \qquad \text{as } n \to \infty.$$

Equivalently, we obtain

$$S_n = n\,\mu + o(n) \quad \text{a.s.}$$

Proof. See Theorem 2.4.1 in Durrett [2019]. □

In other words, up to a first-order approximation, S_n is linear in n and fully determined by the mean. While the law of large numbers provides a deterministic approximation, it does not describe the random fluctuations of S_n around its mean. This is provided by the central limit theorem.

Theorem 4.2 (Central Limit Theorem)**.** *Let $(X_n)_{n\geq 1}$ be an infinite sequence of i.i.d. random variables with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}(X_1) = \sigma^2 < \infty$. Then*

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{d} Z \quad \text{as } n \to \infty,$$

where $Z \sim \mathcal{N}(0, \sigma^2)$.

This yields the second-order approximation

$$S_n \approx n\mu + \sqrt{n}\, Z.$$

The error of the first-order approximation is of order $\sqrt{n}$ and, when properly normalized, converges in distribution to a Gaussian random variable. Thus, the central limit theorem can be viewed as a correction term to the law of large numbers.

The proof of the central limit theorem is most naturally formulated in terms of characteristic functions. For a random variable X, the characteristic function is defined by

$$\varphi_X(t) = \mathbb{E}[e^{itX}].$$

Characteristic functions are particularly well suited for the analysis of additive processes. On the one hand, they uniquely determine the distribution of a random variable. On the other hand, they transform sums of independent random variables into products, namely

$$\varphi_{X_1+\cdots+X_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t).$$

This multiplicative structure makes them a natural tool for studying the asymptotic behavior of sums.

Proof. Let $Y_i := X_i - \mu$, so that $\mathbb{E}[Y_i] = 0$ and $\mathrm{Var}(Y_i) = \sigma^2$. Denote by φ_Y the characteristic function of Y_i. Then $\varphi_Y(0) = 1$ and $\varphi_Y'(0) = 0$. Since $\mathrm{Var}(Y_i) < \infty$, the characteristic function is twice differentiable at 0 with

$$\varphi_Y''(0) = -\mathbb{E}[Y_i^2] = -\sigma^2.$$

Thus, a Taylor expansion around 0 yields

$$\varphi_Y(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \quad \text{as } t \to 0.$$

Now consider the normalized sum

$$Z_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

Using independence, its characteristic function satisfies

$$\varphi_{Z_n}(t) = \left(\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n.$$

Taking the limit as $n \to \infty$, we obtain

$$\varphi_{Z_n}(t) \to e^{-\sigma^2 t^2/2}.$$

The limiting function is the characteristic function of a Gaussian random variable with mean 0 and variance σ^2. The result follows from Lévy's continuity theorem (see Theorem 3.3.17 in Durrett [2019]), which implies convergence in distribution. □

The central limit theorem applies even when the summands are heavy-tailed, provided their variance is finite. For instance if X_1 follows a Pareto distribution with index $\alpha > 2$, then $\mathrm{Var}(X_1) < \infty$ and

$$\frac{S_n - n\mathbb{E}[X_1]}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

At first glance, this may seem surprising. Indeed, as seen in Proposition 3.17, if X_1 is regularly varying with index $-\alpha$, then any finite sum $S_n = X_1 + \cdots + X_n$ is also regularly varying with the same index and hence heavy-tailed. Thus, for every n, the distribution of S_n is heavy-tailed, while the limiting distribution is light-tailed.

The resolution of this apparent contradiction is that a Gaussian distribution, being light-tailed, cannot approximate the full tail of a heavy-tailed distribution. For large but finite n, the Gaussian approximation is accurate only up to a certain range of the tail. As $n \to \infty$, this range extends further and further, which is sufficient for convergence in distribution, even though the extreme tail behavior is never fully captured.

Limitations of the Classical Central Limit Theorem. The classical central limit theorem relies on several key assumptions, namely that the random variables $(X_i)_{i\geq 1}$ are (i) identically distributed, (ii) independent, and possess (iii) finite mean and (iv) finite variance.

It turns out that assumptions (i) and (ii) are not essential in a strict sense. If the $(X_i)_{i\geq 1}$ are not identically distributed, versions of the central limit theorem still hold under appropriate moment conditions, such as in the Lyapunov or Lindeberg central limit theorems. Similarly, certain forms of dependence can be allowed, provided they are sufficiently weak or localized. We refer to Theorem 27.4 in Billingsley [2008] for precise statements and proofs of these generalizations.

In contrast, assumptions (iii) and (iv) are closely tied to the classical Gaussian limit. If the variance is infinite, the usual central limit theorem no longer applies in general, and one must consider more general normalization schemes and possible non-Gaussian limits.

Therefore, in seeking a genuine generalization of the central limit theorem, we focus on the case where assumptions (iii) and (iv) fail. This leads to the question of which distributions can arise as limits of normalized sums in the absence of finite moments.

4.2 Stable Distributions

In view of the limitations of the classical central limit theorem, we consider limits of the more general form

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} G, \tag{4.1}$$

for suitable sequences of positive scaling constants $(a_n)_{n\geq 1}$, centering constants $(b_n)_{n\geq 1}$, and some random variable G.

We consider the possible G arising in (4.1). We already know one such limiting distribution from the classical central limit theorem, namely the Gaussian distribution. Indeed, if X_1, X_2 are independent Gaussian random variables with mean μ and variance σ^2, then $X_1 + X_2$ is again Gaussian with mean 2μ and variance $2\sigma^2$. More generally, for any $a_1, a_2 > 0$,

$$a_1 X_1 + a_2 X_2$$

is Gaussian distributed. Thus, the Gaussian distribution is closed under addition (up to scaling and shifting), a property that makes it particularly natural in the context of limit theorems. Other important examples of stable distributions include the Cauchy and Lévy distributions. This observation motivates the following definition.

Definition 4.3. *A distribution F is called* stable *if for any $n \geq 2$ and i.i.d. random variables $X_1, \ldots, X_n$ with distribution F, there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that*

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} a_n X_1 + b_n.$$

To better understand this definition, we now characterize stable distributions in terms of their appearance as limits of normalized sums of i.i.d. random variables, as in (4.1).

Theorem 4.4. *A non-degenerate random variable Z has a stable distribution if and only if there exists a sequence of i.i.d. random variables $X_1, X_2, \dots$ and deterministic sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, with positive $a_n \in \mathbb{R}$ and $b_n \in \mathbb{R}$ for all n, such that*

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow{d} Z.$$

In particular, the Gaussian distribution is one member of a broader class of stable laws.

Proof. We prove both directions.

($\Rightarrow$) Suppose that F is a stable distribution and let $(X_i)_{i\geq 1}$ be an i.i.d. sequence with distribution F. By the definition of stability, for every $n \geq 2$ there exist constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$X_1 + \cdots + X_n \overset{d}{=} a_n X_1 + b_n.$$

Rewriting this, we obtain

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \overset{d}{=} X_1.$$

Trivially it follows that

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow{d} F \qquad \text{as } n \to \infty.$$

($\Leftarrow$) Conversely, suppose that, as $n \to \infty$,

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow{d} F,$$

where $(X_i)_{i\geq 1}$ are i.i.d. random variables and $a_n > 0$.
Fix an integer $k \geq 2$. For $m = jk$ with $j \in \mathbb{N}$, define

$$Y_m := \frac{X_1 + \cdots + X_m - b_m}{a_m}, \qquad Z_m := \frac{X_1 + \cdots + X_m - kb_j}{a_j}.$$

As $j \to \infty$, we have $m = jk \to \infty$, and therefore

$$Y_m \xrightarrow{d} F.$$

On the other hand, we can decompose the sum into k blocks:

$$X_1 + \cdots + X_m = \sum_{\ell=1}^{k} \big(X_{(\ell-1)j+1} + \cdots + X_{\ell j}\big).$$

Hence,

$$Z_m = \sum_{\ell=1}^{k} \frac{X_{(\ell-1)j+1} + \cdots + X_{\ell j} - b_j}{a_j}.$$

Each term in the sum has the same distribution as

$$\frac{X_1 + \cdots + X_j - b_j}{a_j},$$

and the k terms are independent. Let φ_j denote the characteristic function of

$$\frac{X_1 + \cdots + X_j - b_j}{a_j},$$

and let φ denote the characteristic function of F. Since

$$\frac{X_1 + \cdots + X_j - b_j}{a_j} \xrightarrow{d} F,$$

we have

$$\varphi_j(t) \to \varphi(t) \qquad \text{for all } t \in \mathbb{R}.$$

By independence,

$$\mathbb{E}(e^{itZ_m}) = \prod_{\ell=1}^{k} \mathbb{E}\left(e^{it\frac{X_{(\ell-1)j+1}+\cdots+X_{\ell j}-b_j}{a_j}} \right) = \big(\varphi_j(t)\big)^k.$$

Therefore,

$$\mathbb{E}(e^{itZ_m}) \to \varphi(t)^k.$$

Since $\varphi(t)^k$ is the characteristic function of the k-fold convolution F^{k*}, Lévy's continuity theorem (see Thm.3.3.17 in Durrett [2019]) implies that $Z_m \xrightarrow{d} F^{k*}$.

Now observe that Y_m and Z_m differ only by deterministic centering and scaling

$$Y_m = \frac{a_j}{a_m} Z_m + \frac{kb_j - b_m}{a_m}.$$

It follows from Theorem A1.5 (convergence to types) in Embrechts et al. [1997] that their limiting distributions can differ only by a linear transformation. So, F^{k*} is a scaled and shifted version of F. Since this holds for every $k \geq 2$, by definition F is a stable distribution, which completes the proof.

□

However, this characterization is not fully satisfactory, as it does not provide an explicit description of the distributions that can arise as limits, nor does it indicate how the limiting distribution depends on the underlying random variables. A more tractable characterization is provided via characteristic functions.

A fundamental result is that stable distributions admit an explicit representation in terms of their characteristic functions.

Theorem 4.5. *A non-degenerate random variable X is said to be α-stable if and only if it can be written in the form $X \overset{d}{=} aZ + b$, where $a > 0$, $b \in \mathbb{R}$, and the random variable Z has characteristic function*

$$\varphi_Z(t) = \exp\left\{-|t|^\alpha \left(1 - i\beta \operatorname{sign}(t)\, \gamma(t,\alpha)\right)\right\},$$

with parameters $\alpha \in (0,2)$ and $\beta \in [-1,1]$. The function $\gamma(t,\alpha)$ is given by

$$\gamma(t,\alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right), & \text{if } \alpha \neq 1, \\ -\frac{2}{\pi}\log|t|, & \text{if } \alpha = 1, \end{cases}$$

and $\operatorname{sign}(t)$ denotes the sign function.

Proof Sketch. The argument follows the derivation of stable characteristic functions in Chapter 17, Sections 3–5 in Feller [1971].

Let $\phi(t) = \mathbb{E}[e^{itX}]$ denote the characteristic function of the stable random variable X. By stability, for every $n \geq 1$,

$$X_1 + \cdots + X_n \overset{d}{=} a_n X + b_n,$$

where $X, X_1, \ldots, X_n$ are i.i.d. with $a_n > 0$ and $b_n \in \mathbb{R}$ for all $n \in \mathbb{N}$. Passing to characteristic functions gives

$$\phi(t)^n = \phi(a_n t) e^{ib_n t}.$$

Writing $\psi(t) = \log \phi(t)$, we obtain

$$n\psi(t) = \psi(a_n t) + ib_n t.$$

The functional equation forces a_n to satisfy a Cauchy-type scaling relation whose only continuous solutions are $a_n = n^{\frac{1}{\alpha}}$. The stability property forces the normalization constants to satisfy $a_n = n^{\frac{1}{\alpha}}$ for some $\alpha \in (0,2]$. Since stable laws are infinitely divisible, their characteristic functions admit the canonical representation developed in Chapter 17, Section 2 of Feller [1971]. The result is known in the modern literature as the Lévy–Khintchine theorem. Namely, if $\phi(t) = \exp\{\psi(t)\}$ is infinitely divisible, then its exponent can be written in the form

$$\psi(t) = i\gamma t + \int_{\mathbb{R}} \left(e^{itx} - 1 - \frac{itx}{1+x^2}\right) \frac{1+x^2}{x^2}\, \nu(dx),$$

where $\gamma \in \mathbb{R}$ and ν is the canonical measure with density

$$\nu(dx) = \left(c_+ x^{-1-\alpha} \mathbb{1}_{\{x>0\}} + c_- |x|^{-1-\alpha} \mathbb{1}_{\{x<0\}}\right) dx.$$

Note that this density is locally integrable at zero precisely for $\alpha \in (0,2)$, which accounts for the restriction on the stability index. Substituting this measure into the canonical representation and evaluating the resulting integrals yields, for $\alpha \neq 1$,

$$\psi(t) = -c|t|^\alpha \left(1 - i\beta \operatorname{sign}(t) \tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mu t,$$

while for $\alpha = 1$,

$$\psi(t) = -c|t| \left(1 - i\beta \frac{2}{\pi} \operatorname{sign}(t) \log |t|\right) + i\mu t.$$

Exponentiating gives the stated representation of the characteristic function. □

This representation provides a complete description of all stable distributions and shows that they form a two-parameter family, parameterized by α and β. The parameter α, often referred to as the *stability index*, governs the tail behavior of the distribution. Smaller values of α correspond to heavier tails, while $\alpha = 2$ yields the Gaussian distribution, which has exponentially decaying tails. In this sense, α determines the rate at which the distribution's tails decay. The parameter β controls the *skewness* of the distribution. When $\beta = 0$, the distribution is symmetric about zero. Positive values of β produce distributions skewed to the right, while negative values yield left-skewed distributions. In particular, $\beta = 0$ corresponds to a real-valued characteristic function, and hence a symmetric distribution.

Only in a few special cases do stable distributions admit closed-form expressions for their density or distribution functions. Among the most notable examples are the Gaussian distribution, which arises when $\alpha = 2$ and $\beta = 0$, the Cauchy distribution, corresponding to $\alpha = 1$ and $\beta = 0$, and the Lévy distribution, obtained when $\alpha = \frac{1}{2}$ and $\beta = 1$. In general, however, stable distributions are most conveniently described through their characteristic functions.

4.2.1 Properties of Stable Distributions

A key feature of stable distributions is their heavy-tailed behavior when $\alpha \in (0, 2)$. This property is captured by the following result.

Theorem 4.6. *If X is α-stable and $\alpha \in (0, 2)$, then there exist constants $p, q \geq 0$ with $p + q > 0$ such that, as $x \to \infty$,*

$$\mathbb{P}(X > x) = (p + o(1))x^{-\alpha}, \qquad \mathbb{P}(X < -x) = (q + o(1))x^{-\alpha}.$$

Thus, either the left tail, the right tail, or both tails of X are regularly varying with index $-\alpha$.

Among stable distributions, the Gaussian distribution is distinguished as the only non-degenerate member that is light-tailed and has finite variance. All other stable laws are heavy-tailed and therefore have infinite variance. This makes the Gaussian a boundary case within the class of stable distributions and explains its role as the limiting distribution in the classical central limit theorem.

The asymptotic behavior of the tails is closely connected to the behavior of the characteristic function near the origin. In fact, Tauberian theorems establish that, for $\alpha \in (0, 2)$, the tail behavior of a distribution is equivalent to a specific asymptotic form of its characteristic function as $t \to 0$.

There are several Tauberian theorems; the one used here, which forms the basis of the result stated below, follows from Theorems 8.1.10 and 8.1.11 Bingham et al. [2013].

Let X be a real-valued random variable with characteristic function

$$\varphi(t) = U(t) + iV(t),$$

where

$$U(t) = \Re(\varphi(t)), \qquad V(t) = \Im(\varphi(t)).$$

For $x > 0$, define

$$H(x) := \mathbb{P}(|X| > x) = \mathbb{P}(X > x) + \mathbb{P}(X < -x), \qquad K(x) := \mathbb{P}(X > x) - \mathbb{P}(X < -x).$$

Then

$$\mathbb{P}(X > x) = \frac{H(x) + K(x)}{2}, \qquad \mathbb{P}(X < -x) = \frac{H(x) - K(x)}{2}.$$

For $0 < \alpha < 2$, $\alpha \neq 1$, define

$$C_\alpha := \frac{\pi}{2\Gamma(\alpha)\,\sin\!\left(\frac{\pi\alpha}{2}\right)}.$$

Theorem 4.7 (Two-sided Tauberian theorem for characteristic functions)**.** *For $\alpha \in (0,2)\setminus\{1\}$, the following are equivalent. Assume that there exist a slowly varying function L and a constant $\beta \in [-1,1]$ such that, as $t \to 0$,*

$$1 - U(t) \sim C_\alpha\, t^\alpha L(1/t), \tag{4.2}$$

$$V(t) \sim \beta \tan\!\left(\frac{\pi\alpha}{2}\right) C_\alpha\, t^\alpha L(1/t). \tag{4.3}$$

Then, as $x \to \infty$,

$$H(x) \sim x^{-\alpha} L(x), \tag{4.4}$$

$$K(x) \sim \beta\, x^{-\alpha} L(x). \tag{4.5}$$

We now use this result to derive the tail behavior of stable distributions. The proof follows an adaptation of the argument in the Proof of Theorem 5.7 in Nair et al. [2022], taking into account the imaginary part of the characteristic function and using the corresponding Tauberian Theorem 4.7.

Proof of Theorem 4.6. Let X be α-stable. Then the Representation Theorem 4.5 gives, for $t > 0$,

$$\varphi_X(t) = \exp\!\left(-|t|^\alpha \left(1 - i\beta\,\gamma(t,\alpha)\,\mathrm{sgn}(t)\right)\right), \qquad t \in \mathbb{R},$$

with parameters $\alpha \in (0,2)$ and $\beta \in [-1,1]$, where

$$\gamma(t,\alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right), & \text{if } \alpha \neq 1, \\ -\frac{2}{\pi}\log(|t|), & \text{if } \alpha = 1. \end{cases}$$

We treat the case $\alpha \neq 1$; the case $\alpha = 1$ follows by analogous arguments with a logarithmic correction. Assume $\alpha \neq 1$. Here, taking $\delta := \beta \tan\!\left(\frac{\pi\alpha}{2}\right)$,

$$\varphi_X(t) = \exp(-|t|^\alpha\,(1 - i\delta\ \mathrm{sgn}(t)))\,, \qquad t \in \mathbb{R}.$$

Write $\varphi_X(t) = U(t) + iV(t)$, where $U(t) = \Re(\varphi_X(t))$ and $V(t) = \Im(\varphi_X(t))$. We first determine the behavior of $U(t)$ and $V(t)$ as $t \downarrow 0$. Recall the elementary expansions

$$e^{-x} = 1 - x + o(x), \qquad x \to 0 \qquad \text{and} \qquad e^{ix} = 1 + ix + o(x). \qquad x \to 0.$$

Apply these with $x = t^\alpha$. Since $|t|^\alpha \to 0$ as $t \to 0$, we obtain

$$\exp(-|t|^\alpha) = 1 - |t|^\alpha + o(|t|^\alpha),$$

and

$$\exp\big(i\delta|t|^\alpha \,\mathrm{sgn}(t)\big) = 1 + i\delta|t|^\alpha \,\mathrm{sgn}(t) + o(|t|^\alpha).$$

Multiplying the two expansions gives

$$\begin{aligned}\varphi_X(t) &= \big(1 - |t|^\alpha + o(|t|^\alpha)\big)\big(1 + i\delta|t|^\alpha \,\mathrm{sgn}(t) + o(|t|^\alpha)\big)\\ &= 1 - |t|^\alpha + i\delta|t|^\alpha \,\mathrm{sgn}(t) + o(|t|^\alpha).\end{aligned}$$

Taking real and imaginary parts, we get

$$U(t) = 1 - |t|^\alpha + o(|t|^\alpha), \qquad V(t) = \delta|t|^\alpha \,\mathrm{sgn}(t) + o(|t|^\alpha).$$

Hence,

$$1 - U(t) \sim |t|^\alpha, \qquad V(t) \sim \beta \tan\Big(\frac{\pi\alpha}{2}\Big)\,|t|^\alpha \,\mathrm{sgn}(t), \qquad t \to 0.$$

Comparing with (4.2)-(4.3), we see that the Tauberian theorem applies with slowly varying function $L \equiv C_\alpha^{-1}$ and we obtain

$$H(x) := \mathbb{P}(|X| > x) \sim Lx^{-\alpha}, \qquad K(x) := \mathbb{P}(X > x) - \mathbb{P}(X < -x) \sim \beta L x^{-\alpha}, \qquad x \to \infty.$$

Since

$$\mathbb{P}(X > x) = \frac{H(x) + K(x)}{2}, \qquad \mathbb{P}(X < -x) = \frac{H(x) - K(x)}{2},$$

we conclude that

$$\mathbb{P}(X > x) \sim \frac{1+\beta}{2}\,L\,x^{-\alpha}, \qquad \mathbb{P}(X < -x) \sim \frac{1-\beta}{2}\,L\,x^{-\alpha}.$$

Setting

$$p := \frac{1+\beta}{2}\,L, \qquad q := \frac{1-\beta}{2}\,L,$$

where $p, q \geq 0$ and $p + q = L > 0$. We conclude

$$\mathbb{P}(X > x) \sim p\,x^{-\alpha}, \qquad \mathbb{P}(X < -x) \sim q\,x^{-\alpha}, \qquad x \to \infty.$$

which completes the proof for the case $\alpha \neq 1$. □

4.3 Generalized Central Limit Theorem

After having introduced stable distributions, we now state the generalized central limit theorem, which characterizes all possible non-degenerate limits of normalized sums of i.i.d. random variables.

Combining the previous results, namely that the limiting distributions of additive processes are stable (Theorem 4.4), together with the characterization of stable distributions (Theorem 4.5), allows us to determine precisely which stable distributions arise as limiting distributions.

Theorem 4.8 (Generalized Central Limit Theorem)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with distribution F. Then there exist deterministic sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, with positive $a_n \in \mathbb{R}$ and $b_n \in \mathbb{R}$ for all n, such that*

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow{d} Z,$$

if and only if Z is an α-stable distribution for some $\alpha \in (0, 2]$. Further,

(i) Z is Gaussian, i.e. $\alpha = 2$, if and only if $\int_{-x}^{x} y^2\, dF(y)$ is slowly varying as $x \to \infty$;

(ii) $|Z| \in \mathcal{RV}(\alpha)$, i.e. $\alpha \in (0, 2)$, if and only if

$$\overline{F}(x) = (p + o(1))x^{-\alpha}L(x), \qquad F(-x) = (q + o(1))x^{-\alpha}L(x)$$

as $x \to \infty$, where $L(x)$ is slowly varying, and $p, q \geq 0$, $p + q > 0$.

The most striking aspect of the generalized central limit theorem is its contrast with the classical central limit theorem. In the classical setting, the Gaussian distribution appears as a universal limit. In the present framework, however, the Gaussian distribution is only one among many possible limiting distributions, and in fact corresponds to the boundary case $\alpha = 2$.

Moreover, the Gaussian distribution may still arise as the limiting distribution even in situations where the variance of the underlying distribution is infinite. This occurs, for instance, when the truncated second moment $\int_{-x}^{x} y^2\, dF(y)$ is slowly varying as $x \to \infty$. A concrete example is given by a symmetric distribution with tail behavior

$$\mathbb{P}(|X_1| > x) \sim \frac{1}{x^2 \log x}, \qquad x \to \infty,$$

for which the variance is infinite, yet the normalized sums still converge to a Gaussian distribution. Indeed, the variance diverges as

$$\mathbb{E}[X_1^2] = \int_0^\infty \mathbb{P}(X_1^2 > t)\, dt = \int_0^\infty 2x\, \mathbb{P}(|X_1| > x)\, dx$$

and

$$\mathbb{P}(|X_1| > x) \sim \frac{1}{x^2 \log x}.$$

Then there exists $x_0 > 0$ such that

$$\mathbb{P}(|X_1| > x) \geq \frac{1}{2x^2 \log x}, \qquad x \geq x_0.$$

Hence

$$\mathbb{E}[X_1^2] \geq \int_{x_0}^{\infty} \frac{1}{x \log x}\, dx = \infty.$$

On the other hand, using the identity

$$\mathbb{E}[X_1^2 \mathbf{1}_{\{|X_1| \leq x\}}] = 2 \int_0^x t\, \mathbb{P}(|X_1| > t)\, dt - x^2 \mathbb{P}(|X_1| > x).$$

Using the tail asymptotics,

$$t\, \mathbb{P}(|X_1| > t) \sim \frac{1}{t \log t}, \qquad x^2 \mathbb{P}(|X_1| > x) \sim \frac{1}{\log x}.$$

Hence

$$\mathbb{E}[X_1^2 \mathbb{1}_{\{|X_1| \leq x\}}] \sim 2 \int_1^x \frac{dt}{t \log t} = 2 \log \log x.$$

which is slowly varying at infinity. Hence X_1 still belongs to the Gaussian domain of attraction, even though its variance is infinite.

The most important feature of the generalized central limit theorem is that heavy-tailed stable distributions can emerge as limits of additive processes. In particular, if the underlying distribution has regularly varying tails with index $\alpha \in (0, 2)$, so that the variance is infinite, then the limiting distribution is again regularly varying. This is consistent with the closure property of regularly varying functions under summation mentioned earlier (see Proposition 3.17), showing that this stability persists even at the level of distributional limits.

Furthermore, when $\alpha < 1$, even the mean of the limiting distribution is infinite.
Indeed, let X be a nonnegative random variable whose tail is regularly varying with index $\alpha \in (0, 1)$, that is, $\mathbb{P}(X > x) \sim x^{-\alpha} L(x)$ as $x \to \infty$, for some slowly varying function L. By the tail integral representation,

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > x)\, dx \geq \frac{1}{2} \int_A^{\infty} x^{-\alpha} L(x)\, dx,$$

for $A > 0$ sufficiently large, since the integrand is non-negative.
Now fix $\varepsilon > 0$ such that $\alpha + \varepsilon < 1$. Since L is slowly varying, it follows from Potter's bounds (Theorem 3.15) that for every $\varepsilon > 0$ there exist $c > 0$ and x_0 such that

$$L(x) \geq c x^{-\varepsilon}, \quad x \geq x_0.$$

Therefore

$$x^{-\alpha} L(x) \geq c\, x^{-(\alpha+\varepsilon)}, \qquad x \geq x_0,$$

and consequently

$$\mathbb{E}[X] \geq \int_{x_0}^{\infty} x^{-\alpha} L(x)\, dx \geq c \int_{x_0}^{\infty} x^{-(\alpha+\varepsilon)}\, dx.$$

Since $\alpha + \varepsilon < 1$, the last integral diverges, and thus, we conclude that $\mathbb{E}[X] = \infty$.

The generalized central limit theorem not only specifies which limiting distributions may arise, but also characterizes when a given stable distribution appears as the limit, through the notion of the domain of attraction.

Definition 4.9 (Domain of attraction)**.** *A distribution F is said to belong to the* domain of attraction *of a stable distribution Z if there exist sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ with a_n strictly positive for all n such that, for i.i.d. random variables $X_1, X_2, \ldots$ with distribution F,*

$$\frac{X_1 + \cdots + X_n - b_n}{a_n} \xrightarrow{d} Z.$$

In the Gaussian case, where $\alpha = 2$, the domain of attraction consists of all distributions for which the truncated second moment $\int_{-x}^{x} y^2\, dF(y)$ is slowly varying as $x \to \infty$. This includes all finite-variance distributions, as well as certain distributions with infinite variance. In contrast, for $\alpha \in (0, 2)$, the domain of attraction of an α-stable law consists of distributions with regularly varying tails of index $-\alpha$, i.e.

$$\overline{F}(x) = (p + o(1))x^{-\alpha} L(x), \qquad F(-x) = (q + o(1))x^{-\alpha} L(x),$$

as $x \to \infty$, where $L(x)$ is slowly varying, $p, q \geq 0$, and $p + q > 0$.

The normalization sequences appearing in the definition of the domain of attraction and in the generalized central limit theorem can, in the case of attraction to an α-stable distribution, be specified more explicitly in terms of the tail behavior of the underlying distribution; see Section 5.4 in Nair et al. [2022]. In particular, for $\alpha \in (0, 2)$, the scaling constants a_n may be chosen such that

$$\lim_{n\to\infty} \frac{nL(a_n)}{a_n^{\alpha}} = c \in (0, \infty).$$

In the Gaussian case $\alpha = 2$, the appropriate scaling is of order $\sqrt{n}$, corresponding to the classical central limit theorem. For $0 < \alpha < 2$, the correct normalization is instead of order $n^{1/\alpha}$ (up to a slowly varying factor), reflecting the heavier tails of stable laws.

While the centering constants b_n can be taken as

$$b_n = \begin{cases} 0, & \text{if } 0 < \alpha < 1, \\ n\, a_n \displaystyle\int_{-\infty}^{\infty} \sin\Big(\frac{x}{a_n}\Big)\, dF(x), & \text{if } \alpha = 1, \\ n\, \mathbb{E}[X_1], & \text{if } 1 < \alpha < 2. \end{cases}$$

The case $\alpha = 2$ is more involved, and we refer to Chapter 2.2, in particular Proposition 2.2.13 and Proposition 2.2.14 in Embrechts et al. [1997] for a precise specification of the corresponding normalization sequences.

We now present the proof of the generalized central limit theorem, following Theorem 2(c) and 3 of Chapter 7.5 in Feller [1971]. While the argument is rather technical, we include it for the sake of completeness and consistency. We prove only the sufficiency direction in part (ii), namely that regularly varying tails imply convergence of the normalized sums to an α-stable distribution.

Proof of Theorem 4.8. By Theorem 4.4, every non-degenerate limit distribution in the generalized central limit theorem is necessarily α-stable. By the classical central limit theorem, the case $\alpha = 2$ corresponds to convergence to a Gaussian distribution and is already well understood. We therefore restrict attention to the non-Gaussian case $\alpha \in (0, 2)$ and prove part (ii), showing that it is characterized by regularly varying tails.

Assume that

$$\mathbb{P}(X_1 > x) \sim p x^{-\alpha} L(x), \qquad \mathbb{P}(X_1 < -x) \sim q x^{-\alpha} L(x), \qquad x \to \infty,$$

where $p, q \geq 0$, $p + q > 0$, and L is slowly varying. Recall the definition of the *tail distribution function* and the *skewness function*

$$H(x) := \mathbb{P}(|X_1| > x) = \mathbb{P}(X_1 > x) + \mathbb{P}(X_1 < -x),$$

and

$$K(x) := \mathbb{P}(X_1 > x) - \mathbb{P}(X_1 < -x).$$

Since by assumption F is regularly varying,

$$H(x) \sim (p + q) x^{-\alpha} L(x), \qquad K(x) \sim (p - q) x^{-\alpha} L(x), \qquad \text{as} \quad x \to \infty.$$

Hence, if we set

$$\beta := \frac{p - q}{p + q} \in [-1, 1],$$

then

$$K(x) \sim \beta\, x^{-\alpha} \widetilde{L}(x), \qquad H(x) \sim x^{-\alpha} \widetilde{L}(x),$$

where $\widetilde{L}(x) := (p + q) L(x)$ is still slowly varying.
We now apply the Tauberian theorem (Theorem 4.7). It yields, as $t \to 0$,

$$1 - U(t) \sim C_\alpha |t|^\alpha \widetilde{L}(1/|t|),$$

and

$$V(t) \sim \beta \tan\Big(\frac{\pi\alpha}{2}\Big) C_\alpha |t|^\alpha \widetilde{L}(1/|t|)\ \operatorname{sgn}(t),$$

where $\varphi_X(t) = U(t) + iV(t)$ is the characteristic function of X_1. Hence

$$\varphi_X(t) = 1 - C_\alpha |t|^\alpha \widetilde{L}(1/|t|)\Big(1 - i\beta \tan\Big(\frac{\pi\alpha}{2}\Big) \operatorname{sgn}(t)\Big) + o\big(|t|^\alpha \widetilde{L}(1/|t|)\big). \tag{4.6}$$

Let $(a_n)_{n \geq 1}$ be a sequence such that $n a_n^{-\alpha} L(a_n) \to 1$. Then $a_n \to \infty$, and by slow variation, $\widetilde{L}(a_n/t) \sim \widetilde{L}(a_n)$ for fixed t.

We first consider the case $0 < \alpha < 1$. No centering is needed, so $b_n = 0$. For fixed $t \in \mathbb{R}$,

$$\mathbb{E}\Big[e^{itS_n/a_n}\Big] = \big(\varphi_X(t/a_n)\big)^n.$$

Using (4.6) with t/a_n, we obtain

$$\varphi_X(t/a_n) = 1 - C_\alpha(p+q)|t|^\alpha a_n^{-\alpha} L(a_n)\Big(1 - i\beta\tan\Big(\frac{\pi\alpha}{2}\Big)\operatorname{sgn}(t)\Big) + o(a_n^{-\alpha}L(a_n)).$$

Thus

$$n\big(\varphi_X(t/a_n) - 1\big) \longrightarrow -C_\alpha(p+q)|t|^\alpha\Big(1 - i\beta\tan\Big(\frac{\pi\alpha}{2}\Big)\operatorname{sgn}(t)\Big).$$

Since $\varphi_X(t/a_n) \to 1$, we may pass to logarithms and obtain

$$\big(\varphi_X(t/a_n)\big)^n \longrightarrow \exp\Big(-C_\alpha(p+q)|t|^\alpha\Big(1 - i\beta\tan\Big(\frac{\pi\alpha}{2}\Big)\operatorname{sgn}(t)\Big)\Big).$$

This is the characteristic function of an α-stable distribution.

Now let $1 < \alpha < 2$. In this case $\mathbb{E}[X_1]$ exists, and we center by $b_n = n\mu$ and $\mu := \mathbb{E}[X_1]$. Set $Y_i := X_i - \mu$. Then the characteristic function of Y_i is

$$\varphi_Y(t) = e^{-it\mu}\varphi_X(t).$$

Since $e^{-it\mu} = 1 - it\mu + o(t)$ and $\alpha > 1$, multiplying by $e^{-it\mu}$ removes the linear term without changing the principal t^α-term. Hence φ_Y satisfies

$$\varphi_Y(t) = 1 - C_\alpha(p+q)t^\alpha L(1/t)(1 - i\beta\tan\Big(\frac{\pi\alpha}{2}\Big) + o(t^\alpha L(1/t))).$$

Applying the previous argument to the centered sums $\sum_{k=1}^n Y_k = S_n - n\mu = S_n - b_n$ gives

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} Z,$$

with the same limit distribution Z.
Finally, writing

$$C := C_\alpha(p+q),$$

the limiting characteristic function takes the standard stable form

$$\varphi_Z(t) = \exp\Big(-Ct^\alpha\Big(1 - i\beta\,\tan\Big(\frac{\pi\alpha}{2}\Big)\Big)\Big).$$

The result follows from Lévy's continuity theorem (see Theorem 3.3.17 in Durrett [2019]), which implies convergence in distribution. The case $\alpha = 1$ requires a separate centering argument and is omitted here. □

4.4 Application: Financial Returns

A natural setting in which heavy-tailed phenomena arise is financial return data. Financial returns measure the relative change in the value of an investment over time. Instead of considering absolute prices, one focuses on gains or losses relative to the previous value.

Let P_t denote the price of a financial asset at time t. The *simple return* from time $t-1$ to time t is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{4.7}$$

while the *log-return* is

$$r_t = \log P_t - \log P_{t-1} = \log(1 + R_t). \tag{4.8}$$

For small price changes, simple returns and log-returns are close, since

$$\log(1 + R_t) = R_t + O(R_t^2).$$

Log-returns are especially convenient because multi-period returns are additive, i.e.

$$r_t^{(k)} = r_t + r_{t-1} + \cdots + r_{t-k+1},$$

which is the reason, why log-returns are widely used in financial modeling. Multi-period returns are reduced to sums of single-period increments. This additive structure provides the natural setting for the application of limit theorems such as the classical and generalized central limit theorem.

Classical financial theory is built on the premise that asset returns are approximately normally distributed. This assumption is closely tied to the classical central limit theorem (CLT), which asserts that sums of i.i.d. random variables with finite variance converge, after appropriate normalization, to a Gaussian distribution. Within this framework, randomness is "well-behaved": extreme events are rare, fluctuations tend to cancel out over time, and risk can be summarized by the variance.

However, empirical evidence has long challenged this paradigm. In his paper, *The variation of certain speculative prices* (Mandelbrot [1963]), Benoît Mandelbrot demonstrated that speculative price changes - using cotton prices as a primary example - exhibit statistical properties that are fundamentally incompatible with the Gaussian distribution. Subsequent empirical studies confirmed these observations across many financial markets and time periods; see, for example, Cont's survey of the stylized facts of asset returns (Cont [2001]). In particular, financial returns display heavy-tails and a remarkable degree of scale invariance. These features suggest that extreme events occur far more frequently than predicted by the normal distribution and that aggregation over time does not lead to Gaussian behavior.

This section interprets Mandelbrot's findings through the lens of the generalized central limit theorem (GCLT). Unlike the classical CLT, the GCLT allows for distributions with infinite variance and identifies stable distributions as the natural limiting laws. This shift has implications for how we understand aggregation, risk, and the statistical structure of financial markets.

Mandelbrot's Empirical Observations. Mandelbrot's analysis of speculative prices initiated a large empirical literature on the statistical structure of financial returns. Many of his observations were later confirmed across a broad range of asset classes and markets and are now commonly referred to as the *stylized facts* of financial returns (see Cont [2001]).

First, return distributions exhibit *heavy-tails*, meaning that large price changes occur with significantly higher probability than under a normal distribution. This implies that extreme events—market crashes or sudden spikes—are not anomalies but intrinsic features of financial systems.

Second, returns display *scale invariance*. When one compares distributions of returns across different time horizons (e.g. daily versus weekly), the overall shape remains similar after appropriate rescaling. This suggests a form of self-similarity in financial data.

Third, Mandelbrot observed that *convergence toward Gaussian behavior* under aggregation *is significantly slower* than predicted by the classical i.i.d. framework. Even over longer time horizons, return distributions retain substantial tail risk and deviations from normality.

Another important feature, emphasized in Cont [2001], is *volatility clustering*: large market moves tend to be followed by large moves, and small fluctuations by small fluctuations, leading to persistent dependence in volatility despite weak autocorrelation in returns themselves.

To account for these phenomena, Mandelbrot proposed that returns follow *stable Paretian distributions*, characterized by a tail index $\alpha \in (0, 2]$. When $\alpha < 2$, the variance is infinite, and the Gaussian model, which corresponds to $\alpha = 2$, becomes inadequate. The theoretical foundation is provided by the GCLT (see Theorem 4.8). We now turn to its financial interpretation.

Financial Interpretation of Stability. In financial markets, empirical return distributions often follow power-law tails of the form

$$\mathbb{P}(|X| > x) \sim x^{-\alpha}.$$

This implies that extreme events decay polynomially rather than exponentially, making them far more likely than in Gaussian models. From this perspective, large market movements are not rare outliers but a fundamental component of the system. Consequently, risk cannot be adequately captured by variance alone, as higher-order tail behavior dominates the distribution.

Under the classical CLT, aggregation smooths fluctuations: as one sums returns over time, the distribution converges toward normality, and extreme events become negligible. In contrast, the GCLT implies that aggregation preserves the heavy-tailed structure. The limiting distribution remains stable with the same tail index α, meaning that extreme risks persist even at longer time horizons. This challenges the common intuition that time diversification reduces risk.

Scaling Laws and Self-Similarity. A defining property of stable distributions is their scaling behavior. Let X be a stable random variable with stability index $\alpha \in (0, 2]$. Then

$$X_1 + \cdots + X_n \stackrel{d}{=} n^{1/\alpha} X + b_n.$$

Equivalently,

$$\frac{X_1 + \cdots + X_n - b_n}{n^{1/\alpha}} \stackrel{d}{=} X.$$

This scaling property shows that stable distributions are closed under aggregation, meaning sums of returns stay in the same distribution class up to a deterministic scaling factor, underlying Mandelbrot's observation that heavy-tails persist under aggregation. Unlike the Gaussian case, where scaling follows $\sqrt{n}$, here volatility grows as $n^{1/\alpha}$ for $\alpha < 2$, so the usual square-root-of-time rule for risk scaling fails in heavy-tailed settings.

Implications for Financial Modeling. Many foundational models in finance, including the Black–Scholes framework (see Black and Scholes [1973]), rely on the assumption of finite variance and light-tails. Within the GCLT framework, these assumptions are violated. If returns follow heavy-tailed distributions, variance may be infinite, and extreme events dominate the dynamics. This leads to systematic underestimation of risk, particularly for measures such as Value-at-Risk and expected shortfall when computed under Gaussian assumptions.

Stable distributions provide a natural alternative. They accommodate heavy-tails, allow for skewness, and arise as limiting distributions under aggregation. In this sense, they can be interpreted as equilibrium distributions of cumulative market shocks. However, their practical use is complicated by the lack of closed-form density functions (except in special cases) and the presence of infinite variance when $\alpha < 2$.

In the classical mean-variance framework, risk is quantified by variance, and diversification reduces risk by averaging independent fluctuations. In a heavy-tailed setting, this logic weakens significantly. If variance does not exist, it cannot serve as a meaningful risk measure. Moreover, diversification becomes less effective because extreme events dominate portfolio outcomes. This motivates the use of alternative, tail-sensitive risk measures.

Economic Interpretation. Heavy-tails may emerge from underlying economic mechanisms such as large, infrequent trades, abrupt information arrivals, or heterogeneous behavior among market participants. From the perspective of the GCLT, market returns can be viewed as the aggregation of such heavy-tailed shocks. Mandelbrot's work also highlights the fractal nature of financial markets. The observed scale invariance suggests that similar statistical patterns persist across time scales. Stable distributions are consistent with this phenomenon, as they are invariant under aggregation up to a scaling factor.

Mandelbrot's contribution can be understood as a paradigm shift: financial returns should not be viewed through the lens of the classical CLT but rather through the broader framework of the GCLT. Within this perspective, heavy-tails are not anomalies but

structural features, aggregation does not eliminate extreme risk, and stable distributions emerge as the natural limiting laws. This leads to a fundamentally different understanding of risk and challenges the central role of Gaussian models in finance. Ultimately, the GCLT provides a more realistic foundation for modeling financial markets, emphasizing the importance of extreme events and the limitations of traditional assumptions.

Chapter 5

Limit Theorems for Multiplicative and Extremal Processes

5.1 Multiplicative processes

The classical central limit theorem suggests that many systems in nature are governed by additive dynamics. This viewpoint has led to the widespread perception that Gaussian behavior is universal. However, many real-world phenomena evolve not additively but *multiplicatively*, meaning that changes occur proportionally to the current size of the system. In such settings, growth is driven by random factors that scale the current state rather than shift it.

Typical examples include financial returns, where wealth evolves through successive percentage gains or losses, and population dynamics, where each generation grows proportionally to the size of the previous one. More generally, multiplicative dynamics often appear in systems exhibiting a *"rich get richer"* effect: the larger an entity becomes, the faster it tends to grow. This phenomenon is observed in a wide range of contexts, including the spread of information in networks, the growth of firms, and the accumulation of links in complex networks.

To model such behavior, one naturally considers the multiplicative process

$$P_n = Y_1 \cdot Y_2 \cdot \ldots \cdot Y_n,$$

where $P_0 = 1$ and $(Y_i)_{i\geq 1}$ are i.i.d. positive random variables.

There is a crucial distinction to what we have seen so far: while additive processes typically preserve heavy tails, multiplicative processes can *create* them.

To illustrate how multiplicative structure generates heavy-tails, consider the simple case $P_2 = Y_1 Y_2$ where Y_1, Y_2 are independent and exponentially distributed with rate λ. Then, for any $t > 0$,

$$\mathbb{P}(P_2 > t) \geq \mathbb{P}(Y_1 > \sqrt{t})\, \mathbb{P}(Y_2 > \sqrt{t}),$$

where $\mathbb{P}(Y_i > t) = e^{-\lambda t}$, and thus

$$\mathbb{P}(P_2 > t) \;\geq\; e^{-2\lambda\sqrt{t}}.$$

This lower bound decays more slowly than any exponential function e^{-ct}. Therefore P_2 is heavy-tailed.

These observations suggest that, as $n \to \infty$, multiplicative processes should naturally lead to heavy-tailed limiting behavior. We now make this intuition precise.

A key property is that multiplicative processes can be transformed into additive ones by taking logarithms. Defining $X_i = \log Y_i$, we obtain

$$\log P_n = \sum_{i=1}^{n} X_i,$$

so that the analysis of P_n reduces to that of an additive process. This connection provides the fundamental intuition behind their behavior.

In particular, if the classical central limit theorem applies to $(X_i)_{i\geq 1}$, then $\log P_n$ is approximately Gaussian, and hence P_n is approximately Lognormal. Thus, even when the underlying factors Y_i are light-tailed, multiplicative growth naturally produces heavy-tailed outcomes.

This logarithmic connection between multiplicative and additive processes will also allow us to extend the analysis beyond the classical central limit theorem. In particular, when the variance of $\log Y_i$ is infinite, generalized limit theorems lead to stable laws, yielding even heavier-tailed behavior after exponentiation.

To make this intuition even more concrete, we study first and second order approximations, analogous to the additive case, using the law of large numbers and the central limit theorem.

Assume first that $\mathbb{E}[X_1]$ exists and $\mathrm{Var}(X_1) = \sigma^2 < \infty$. By the law of large numbers,

$$\frac{1}{n} \log P_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mathbb{E}[X_1],$$

and hence

$$\log P_n \overset{a.s.}{=} \mathbb{E}[X_1]\, n + o(n).$$

Exponentiating yields the first-order approximation

$$P_n \overset{a.s.}{=} \exp\big(\mathbb{E}[X_1]\, n + o(n)\big) = \big(e^{\mathbb{E}[\log Y_1]}\big)^n e^{o(n)}$$

Thus, the first order approximation of a multiplicative process is exponential, which in itself is not surprising. However, a more subtle and important point concerns the base of this exponential growth. One might expect, by analogy with additive processes and the law of large numbers, that the dominant growth rate of P_n should be given by $\mathbb{E}[Y_1]^n$, since independence implies $\mathbb{E}[P_n] = \prod_{i=1}^{n} \mathbb{E}[Y_i]$.

Instead, the typical growth of the process is governed by

$$\mu = e^{\mathbb{E}[\log Y_1]},$$

which is in general strictly smaller than $\mathbb{E}[Y_1]$. This discrepancy follows from Jensen's inequality, since

$$\mathbb{E}[\log Y_1] \leq \log \mathbb{E}[Y_1].$$

Consequently, the mean $\mathbb{E}[P_n]$ does not accurately describe the typical behavior of the process. This mismatch between mean behavior and typical realizations is a characteristic feature of multiplicative systems and is closely related to the emergence of heavy-tailed distributions.

Now assume additionally that $\mathrm{Var}(X_1) = \sigma^2 < \infty$. Notably, this condition may hold even when $\mathrm{Var}(Y_i) = \infty$, making the multiplicative setting substantially broader than the classical additive one.

Then the central limit theorem yields the second order approximation

$$\log P_n \stackrel{d}{=} \mathbb{E}[X_1]\, n + \sigma\sqrt{n}\, Z + o(\sqrt{n}), \qquad Z \sim \mathcal{N}(0,1).$$

Exponentiating once more, we obtain

$$P_n \stackrel{d}{=} \exp\big(\mathbb{E}[X_1]\, n + \sigma\sqrt{n}\, Z + o(\sqrt{n})\big),$$

and therefore

$$P_n \approx \mu^n e^{\sigma\sqrt{n}Z}.$$

Equivalently,

$$\left(\frac{P_n}{\mu^n}\right)^{1/\sqrt{n}} \xrightarrow{d} e^{\sigma Z},$$

where the limit is lognormal. In particular, the lognormal distribution is a a heavy-tailed distribution. This is the multiplicative analogue of the classical central limit theorem.

The contrast with additive processes is striking. In the additive case, the first order approximation is linear and the second order correction is additive. In the multiplicative case, both the growth and the fluctuations are exponential once we return from logarithmic scale. For this reason, multiplicative dynamics naturally amplify variability and provide a robust mechanism for the emergence of heavy-tailed distributions.

These heuristic arguments can be made rigorous, yielding the following result, often referred to as *Gibrat's law*.

Theorem 5.1 (Multiplicative Central Limit Theorem)**.** *Suppose* $(Y_i)_{i\geq 1}$ *is a sequence of i.i.d. strictly positive random variables satisfying* $\mathrm{Var}(\log Y_i) = \sigma^2 < \infty$*, and define* $\mu = e^{\mathbb{E}[\log Y_i]}$*. Then*

$$\left(\frac{Y_1 \cdots Y_n}{\mu^n}\right)^{1/\sqrt{n}} \xrightarrow{d} H, \text{ where } H \sim \mathrm{LogNormal}(0, \sigma^2).$$

The multiplicative central limit theorem mirrors the classical central limit theorem, but with centering and scaling adapted to the multiplicative nature of the process rather than an additive one.

Notice that the assumption of finite variance applies only to $(\log Y_i)_{i\geq 1}$, rather than to $(Y_i)_{i\geq 1}$ itself. As a consequence, it is compatible with a broad class of heavy-tailed distributions, including all distributions with polynomially decaying tails, and in particular regularly varying distributions. Indeed, suppose that Y satisfies the tail bound

$$\mathbb{P}(Y > x) \leq Cx^{-\alpha}, \qquad x \geq x_0,$$

for some constants $C, \alpha, x_0 > 0$. Setting $X = \log Y$, we obtain

$$\mathbb{P}(X > t) = \mathbb{P}(Y > e^t) \leq C(e^t)^{-\alpha} = Ce^{-\alpha t}.$$

Hence X has an exponentially decaying tail. In particular, X possesses moments of every order, i.e.

$$\mathbb{E}[|X|^p] < \infty \qquad \text{for all } p > 0.$$

Therefore,

$$\mathrm{Var}(\log Y) < \infty.$$

Thus the multiplicative central limit theorem applies to a wide range of heavy-tailed models, and only extremely heavy-tailed distributions are excluded.

When the condition of finite variance of $(\log Y_i)_{i\geq 1}$ fails, the classical central limit theorem is replaced by the generalized central limit theorem.

If $X_i = \log Y_i$ belongs to the domain of attraction of a stable distribution with index $\alpha \in (0, 2)$, then

$$\sum_{i=1}^{n} X_i \approx a_n + b_n Z_\alpha,$$

where Z_α is α-stable. Exponentiating yields

$$P_n \approx e^{a_n} \cdot e^{b_n Z_\alpha}.$$

Unlike the finite-variance case, fluctuations are driven by rare but exceptionally large values of $\log Y_i$. The stable limit reflects the significant contribution of extreme observations, leading to much heavier-tailed behavior.

This marks a fundamental transition: while finite-variance multiplicative processes lead to lognormal limits, the infinite-variance case gives rise to much heavier tails. This observation further reinforces the link between multiplicative processes and heavy-tailed phenomena: in contrast to additive models, heavy-tailed behavior emerges naturally and under remarkably weak assumptions.

5.2 Variations on multiplicative processes

The multiplicative central limit theorem shows that, under mild conditions, multiplicative processes lead to Lognormal limiting behavior. This already highlights a strong connection between multiplicative dynamics and heavy-tailed distributions. However, this

connection is in fact much more robust than the specific appearance of the Lognormal distribution.

Indeed, small modifications of the multiplicative mechanism can drastically change the limiting distribution, while preserving heavy-tailed behavior. In particular, while the emergence of the Lognormal distribution is fragile, the emergence of heavy tails is remarkably stable.

In this section, we discuss two important variations of multiplicative processes. In both cases, heavy-tailed distributions still arise, but the limiting distributions are no longer Lognormal. Instead, distributions with approximately power-law tails emerge, closely related to regularly varying distributions.

Multiplicative processes with a lower barrier. A natural modification of the classical multiplicative process is to impose a strictly positive lower bound. Specifically, define

$$P_n = \max(P_{n-1}Y_n, \ell), \qquad \ell > 0,$$

where $(Y_n)_{n\geq 1}$ are i.i.d. strictly positive random variables and $P_0 = 1$.

This model captures situations where the process cannot fall below a certain threshold, such as minimum income levels or population sizes. The presence of a lower barrier fundamentally changes the long-term behavior of the process. Instead of diverging or vanishing, the process stabilizes and converges in distribution.

Theorem 5.2 (Multiplicative process with lower barrier)**.** *Let $(Y_n)_{n\geq 1}$ be i.i.d. strictly positive random variables and consider*

$$P_n = \max(P_{n-1}Y_n, \ell), \qquad \ell > 0.$$

Assume that:

- $\mu := e^{\mathbb{E}[\log Y_1]} < 1$,
- $\mathbb{P}(Y_1 > 1) > 0$,
- $\mathbb{E}[Y_1^s] < \infty$ *for some* $s > 0$.

Then $P_n \xrightarrow{d} H$, where H is heavy-tailed and its tail satisfies

$$\lim_{x\to\infty} \frac{-\log \mathbb{P}(H > x)}{\log x} = s^*, \qquad s^* = \sup\{s \geq 0 : \mathbb{E}[Y_1^s] \leq 1\}.$$

Proof. See Theorem 6.2 in Nair et al. [2022]. □

In particular, the limiting distribution is no longer Lognormal. Instead, the asymptotic behavior of the tail in the above theorem can be directly related to results from the theory of regular variation. In particular, we have

$$\lim_{x\to\infty} \frac{\log \mathbb{P}(H > x)}{\log x} = -s^*,$$

which shows that the tail of the limiting distribution is asymptotically of power-law type.

Such logarithmic asymptotics are closely connected to regularly varying distributions. While this condition does not in general imply regular variation, it captures the same leading-order behavior and ensures that the tail is asymptotically equivalent to a power law up to slowly varying corrections.

Multiplicative processes with additive noise. Another important variation introduces additive noise into the multiplicative dynamics. Define

$$P_n = P_{n-1}Y_n + Q_n,$$

where $(Y_n)_{n\geq 1}, (Q_n)_{n\geq 1}$ are i.i.d., $Y_n > 0$, $Q_n \geq 0$, and $P_0 = 1$.

This model captures situations where multiplicative growth is combined with external inputs, such as bonuses in income, immigration in population models, or measurement noise.

Theorem 5.3 (Noisy multiplicative process)**.** *Let $(Y_n)_{n\geq 1}, (Q_n)_{n\geq 1}$ be i.i.d. with $Y_n > 0$ and $Q_n \geq 0$. Assume:*

- $\mathbb{P}(Y_1 > 1) > 0$,
- $\mu := e^{\mathbb{E}[\log Y_1]} < 1$,
- $\mathbb{E}[Y_1^s] < \infty$ *for some* $s > 0$,
- Q_1 *is light-tailed and* $\mathbb{P}(Q_1 > 0) > 0$.

Then $P_n \xrightarrow{d} H$, where H is heavy-tailed and satisfies

$$\lim_{x\to\infty} \frac{-\log \mathbb{P}(H > x)}{\log x} = s^*, \qquad s^* = \sup\{s \geq 0 : \mathbb{E}[Y_1^s] \leq 1\}.$$

Proof. See Theorem 6.3 in Nair et al. [2022]. □

As in the case of a lower barrier, the limiting distribution is not Lognormal but exhibits power-law type tails. Intuitively, the additive noise acts as a "soft" lower bound, preventing the process from collapsing to zero and allowing occasional large multiplicative growth to dominate.

Further variations of multiplicative processes. Beyond the models discussed above, several important extensions of multiplicative processes also lead to heavy-tailed behavior.

A particularly important class is given by *stochastic recurrence equations* of the form

$$P_n = A_n P_{n-1} + B_n,$$

where $(A_n, B_n)_{n\geq 1}$ are i.i.d. random pairs with $A_n > 0$. These models, often referred to as *Kesten processes*, generalize the noisy multiplicative process discussed above. Under suitable conditions, they admit stationary distributions with regularly varying tails, determined implicitly by the equation $\mathbb{E}[A^\alpha] = 1$.

Heuristically, this behavior arises from a balance between multiplicative growth and occasional large fluctuations: while the process is typically contracted, rare sequences of favorable multipliers A_n can generate exceptionally large values. This mechanism leads to power-law tails and provides one of the fundamental explanations for the emergence of Pareto-type distributions. For a detailed analysis, see Kesten [1973].

Another natural variation arises when the multiplicative process is observed at a random time horizon. Let

$$P_T = \prod_{i=1}^{T} Y_i,$$

where T is an independent integer-valued random variable. In this case, even if the factors Y_i are light-tailed, the additional randomness in the time horizon can induce heavy-tailed behavior.

The intuition is that large values of P_T are driven by rare events where the process experiences an unusually long period of multiplicative growth. Thus, variability in the time scale itself acts as an additional source of extremal behavior. Such effects are well documented in the literature on heavy-tailed phenomena; see Section 2.5.1 in Embrechts et al. [1997].

Together, these results show that while the Lognormal distribution arises naturally in the classical multiplicative central limit theorem, it is not universal. In contrast, heavy-tailed behavior is a robust consequence of multiplicative dynamics and persists under a wide range of model variations.

5.3 Application: Preferential attachment

Since multiplicative processes often lead to heavy-tailed distributions, it is natural to present a concrete example in which a heavy-tailed power law emerges from an underlying multiplicative growth mechanism. One of the best-known examples comes from *network science*. As data on social, biological and technological networks became available in the late 1990s, it was repeatedly observed that the degree distribution is extremely right-skewed and often follows a power law (see Barabási and Albert [1999]). This "universality" led researchers to search for a common explanation. Many generative models have been proposed, and a remarkable number rely on a multiplicative mechanism at their core. In this section, we focus on a celebrated mechanism, *the preferential attachment (or Yule) model*, and show how it produces a power-law degree distribution.

The preferential attachment mechanism. The preferential attachment mechanism captures the idea of cumulative growth, often summarized as a *"rich-get-richer"* effect, in evolving networks. Historically, this idea can be traced back to Yule [1925], who

introduced a stochastic model for the distribution of species across genera. It was later adapted by Simon [1955] to describe wealth distributions and by Price [1976], who applied it to citation networks under the name *"cumulative advantage"*. The modern formulation in network science is due to Barabási and Albert [1999], who rediscovered this mechanism in the context of growing graphs.

We describe a discrete-time version of the model. Let $t = 0, 1, \dots, n$ denote the time index. At $t = 0$ one vertex with two self-loops is present, to avoid degeneracy. At each time $t > 0$ a new vertex arrives with one outgoing edge and attaches to one existing vertex. Consequently, the new vertex enters the network with out-degree 1, in-degree 0, and total degree 1. The attachment rule mixes random choice with preferential attachment: with probability $\alpha \in [0, 1)$ the new vertex selects an existing vertex uniformly at random, and with probability $1 - \alpha$ it selects a vertex with probability proportional to its current in-degree. If a vertex has in-degree k at time t, then the probability that the newcomer connects to it is

$$\frac{\alpha}{t} + \frac{(1-\alpha)k}{2t}.$$

Here, the factor $\frac{1}{t}$ arises from the uniform choice among the t existing vertices, while the factor $\frac{1}{2t}$ in the denominator reflects the total degree in the network at time t, since each of the t edges contributes two degree units.

The key feature of the preferential attachment mechanism is that the expected increase in the in-degree of a vertex is proportional to its current degree. Preferential attachment networks therefore exhibit strong heterogeneity in node degrees, with a small number of highly connected nodes and many nodes of low degree. Consequently, high-degree nodes tend to occupy central positions in the network. This gives rise to a characteristic core-periphery structure, in which older, high-degree nodes form a densely connected core, while newer, low-degree nodes populate a sparse periphery connected primarily to the core. This induces a multiplicative growth dynamic. Such reinforcement leads to increasing disparities over time and results in a heavy-tailed degree distribution.

The degree distribution under preferential attachment. We now make the above heuristic precise. In particular, the degree distribution that emerges under the preferential attachment mechanism is not only heavy-tailed, but also power law and scale invariant.

Define $F_t(k)$ as the distribution of the in–degree, that is, the probability that a uniformly chosen vertex has in-degree smaller or equal k just before the arrival of the $(t+1)$-st vertex at time t. The following result shows that the degree distribution converges to a limiting power–law.

Theorem 5.4. *Under the preferential attachment model described above,*

$$F_t \xrightarrow{d} F \quad \text{as } t \to \infty,$$

where the tail of F satisfies

$$\overline{F}(k) \sim \beta\, k^{-\frac{2}{1-\alpha}} \quad \text{for some constant } \beta > 0.$$

Consequently, the distribution of degrees has a power–law tail with exponent $\frac{2}{1-\alpha}$.

The exponent depends continuously on the probability of preferential attachment $1-\alpha$: as $\alpha \to 0$ the process becomes purely preferential and the tail becomes heavier, while as $\alpha \to 1$ the process approaches a uniform random attachment and the tail becomes lighter.

Proof. For each $k \geq 1$, let $m_t(k)$ denote the expected number of vertices of in-degree k at time t, and define

$$p_t(k) := \frac{m_t(k)}{t}.$$

Thus $p_t(k)$ is the probability that a uniformly chosen vertex at time t has in-degree k. We will prove that for every fixed $k \geq 1$, the limit

$$p(k) := \lim_{t\to\infty} p_t(k)$$

exists, and then identify the asymptotic behaviour of $p(k)$ as $k \to \infty$.

At time t, a new vertex arrives and creates one outgoing edge to a pre-existing vertex. A vertex of in-degree k is chosen with probability

$$\frac{\alpha}{t} + \frac{(1-\alpha)k}{2t}.$$

Hence, for $k \geq 2$, the expected number vertices of degree k at time $t+1$ is obtained as follows:

- we start with the $m_t(k)$ vertices of degree k already present;
- we gain one degree-k vertex whenever a vertex of degree $(k-1)$ is chosen;
- we lose one degree-k vertex whenever a degree-k vertex is chosen.

Therefore,

$$\begin{aligned} m_{t+1}(k) &= m_t(k) + m_t(k-1)\left(\frac{\alpha}{t} + \frac{(1-\alpha)(k-1)}{2t}\right) - m_t(k)\left(\frac{\alpha}{t} + \frac{(1-\alpha)k}{2t}\right) \\ &= m_t(k)\left(1 - \frac{2\alpha + (1-\alpha)k}{2t}\right) + m_t(k-1)\left(\frac{2\alpha + (1-\alpha)(k-1)}{2t}\right). \end{aligned} \tag{5.1}$$

For $k = 1$, the arriving vertex itself contributes one new degree-1 vertex, and we lose one degree-1 vertex whenever a degree-1 vertex is chosen. Hence

$$m_{t+1}(1) = m_t(1) + 1 - m_t(1)\left(\frac{\alpha}{t} + \frac{1-\alpha}{2t}\right) = m_t(1)\left(1 - \frac{1+\alpha}{2t}\right) + 1. \tag{5.2}$$

To prove the existence of the limits $p(k)$, we will make use of the following auxiliary lemma.

Lemma 5.5. *Suppose that a sequence of strictly positive constants $(a_t)_{t\geq 1}$ satisfies*

$$a_{t+1} = a_t\left(1 - \frac{b_t}{t}\right) + c_t,$$

where $b_t \to b > -1$ and $c_t \to c$ as $t \to \infty$. Then

$$\frac{a_t}{t} \longrightarrow \frac{c}{1+b}.$$

Proof of Lemma 5.5. Set

$$x_t := \frac{a_t}{t}.$$

Then

$$a_{t+1} = (t+1)x_{t+1} = tx_t\left(1 - \frac{b_t}{t}\right) + c_t = (t - b_t)x_t + c_t,$$

so

$$x_{t+1} = \frac{t - b_t}{t+1}\, x_t + \frac{c_t}{t+1}.$$

Let

$$L := \frac{c}{1+b}.$$

Subtracting L from both sides gives

$$x_{t+1} - L = \frac{t - b_t}{t+1}(x_t - L) + \frac{1}{t+1}\Big(c_t - (1+b_t)L\Big).$$

Since $b_t \to b$ and $c_t \to c$, we have

$$c_t - (1+b_t)L \longrightarrow c - (1+b)\frac{c}{1+b} = 0.$$

Thus for every $\varepsilon > 0$, there exists T such that for all $t \geq T$, $|c_t - (1+b_t)L| \leq \varepsilon$. Hence, for $t \geq T$,

$$|x_{t+1} - L| \leq \frac{t - b_t}{t+1}|x_t - L| + \frac{\varepsilon}{t+1}.$$

Because $b_t \to b > -1$, for all large t one has

$$\frac{t - b_t}{t+1} = 1 - \frac{1+b_t}{t+1} \leq 1 - \frac{\delta}{t+1}$$

for some $\delta > 0$. This recursive inequality implies that $|x_t - L| \to 0$ as $t \to \infty$. This implies that $(x_t)_{t\geq 1}$ converges to L. □

We now return to the analysis of $m_t(k)$ and continue the proof of the existence of the limits $p(k)$. Apply Lemma 5.5 to (5.2) with

$$a_t = m_t(1), \qquad b_t = \frac{1+\alpha}{2}, \qquad c_t = 1.$$

Since $b_t \to \frac{1+\alpha}{2}$ and $c_t \to 1$, we obtain

$$p(1) = \lim_{t\to\infty} \frac{m_t(1)}{t} = \frac{1}{1+\frac{1+\alpha}{2}} = \frac{2}{3+\alpha}.$$

For $k \geq 2$, we proceed by induction on k. Assume that for some $k \geq 2$ the limit

$$p(k-1) = \lim_{t\to\infty} \frac{m_t(k-1)}{t}$$

exists. Rewrite (5.1) as

$$m_{t+1}(k) = m_t(k)\left(1 - \frac{2\alpha + (1-\alpha)k}{2t}\right) + m_t(k-1)\left(\frac{2\alpha + (1-\alpha)(k-1)}{2t}\right).$$

Apply Lemma 5.5 with

$$a_t = m_t(k), \qquad b_t = \frac{2\alpha + (1-\alpha)k}{2},$$

and

$$c_t = m_t(k-1)\left(\frac{2\alpha + (1-\alpha)(k-1)}{2t}\right).$$

By the induction hypothesis,

$$\frac{m_t(k-1)}{t} \longrightarrow p(k-1),$$

so

$$c_t \longrightarrow p(k-1)\frac{2\alpha + (1-\alpha)(k-1)}{2}.$$

Therefore the lemma yields

$$p(k) = \lim_{t\to\infty} \frac{m_t(k)}{t} = \frac{p(k-1)\,\frac{2\alpha+(1-\alpha)(k-1)}{2}}{1+\frac{2\alpha+(1-\alpha)k}{2}}.$$

After simplifying,

$$\begin{aligned} p(k) &= p(k-1)\,\frac{2\alpha + (1-\alpha)(k-1)}{2 + 2\alpha + (1-\alpha)k} \\ &= p(k-1)\,\frac{\frac{2\alpha}{1-\alpha} + k - 1}{\frac{2+2\alpha}{1-\alpha} + k}. \end{aligned} \tag{5.3}$$

This proves by induction that $p(k)$ exists for every $k \geq 1$.

Iterating (5.3) gives

$$p(k) = p(1)\prod_{j=2}^{k} \frac{\frac{2\alpha}{1-\alpha} + j - 1}{\frac{2+2\alpha}{1-\alpha} + j}.$$

Equivalently,

$$p(k) = p(1)\,\frac{\left(\frac{2\alpha}{1-\alpha}+1\right)\left(\frac{2\alpha}{1-\alpha}+2\right)\cdots\left(\frac{2\alpha}{1-\alpha}+k-1\right)}{\left(\frac{2+2\alpha}{1-\alpha}+2\right)\left(\frac{2+2\alpha}{1-\alpha}+3\right)\cdots\left(\frac{2+2\alpha}{1-\alpha}+k\right)}.$$

Using the identity $\Gamma(x+1) = x\Gamma(x)$, we get

$$p(k) = p(1)\,\frac{\Gamma\left(\frac{2\alpha}{1-\alpha}+k\right)}{\Gamma\left(\frac{2\alpha}{1-\alpha}+1\right)}\cdot\frac{\Gamma\left(\frac{2+2\alpha}{1-\alpha}+2\right)}{\Gamma\left(\frac{2+2\alpha}{1-\alpha}+k+1\right)}.$$

All factors independent of k can be absorbed into a positive constant β', so

$$p(k) = \beta'\,\frac{\Gamma\left(\frac{2\alpha}{1-\alpha}+k\right)}{\Gamma\left(\frac{2+2\alpha}{1-\alpha}+k+1\right)}.$$

Now, using the asymptotic relation $\frac{\Gamma(x)}{\Gamma(x+y)} \sim x^{-y}$ as $x \to \infty$, for fixed y (see *Stirling's formula for ratios of Gamma functions*). Set

$$x = k + \frac{2\alpha}{1-\alpha}, \qquad y = 1 + \frac{2}{1-\alpha}.$$

Then

$$\frac{2+2\alpha}{1-\alpha} + k + 1 = \left(k + \frac{2\alpha}{1-\alpha}\right) + \left(1 + \frac{2}{1-\alpha}\right) = x + y.$$

Therefore,

$$p(k) \sim \beta'\left(k + \frac{2\alpha}{1-\alpha}\right)^{-\left(1+\frac{2}{1-\alpha}\right)} \sim \beta' k^{-\left(1+\frac{2}{1-\alpha}\right)}.$$

Thus the limiting probability mass function decays as a power law.

Since

$$p(k) \sim \beta' k^{-\gamma}, \qquad \gamma := 1 + \frac{2}{1-\alpha} > 1,$$

we may use the integral test to obtain

$$\sum_{j>k} p(j) \sim \beta' \sum_{j>k} j^{-\gamma} \sim \beta' \int_k^\infty x^{-\gamma}\,dx.$$

Hence

$$\overline{F}(k) = \sum_{j>k} p(j) \sim \beta' \frac{1-\alpha}{2}\, k^{-\frac{2}{1-\alpha}}.$$

If we set $\beta := \beta' \frac{1-\alpha}{2} > 0$, then

$$\overline{F}(k) \sim \beta\, k^{-\frac{2}{1-\alpha}},$$

which proves the claim. □

5.4 Application: The Zipf-Mandelbrot mechanism

An alternative explanation for the emergence of heavy-tails comes from models of efficient communication and optimal coding. Building on Zipf's empirical observations of word frequencies, namely that the frequency p_n of the n-th most common word scales approximately as $p_n \propto n^{-1}$ (see Zipf [1949]), Mandelbrot proposed a mechanism to explain this empirical pattern based on optimization in language use. He argued that power-law distributions can arise as the solution to an optimization problem that balances the cost of producing words with the information they convey (see Mandelbrot [1953]).

Production cost. A more common word should be less costly to produce. The cost of a word includes everything that contributes to producing and transmitting it, for instance articulating letters, inserting spaces, or punctuation. Suppose that words are constructed from an alphabet containing q symbols. Then the number of distinct code-words of length C is q^C. Since each of the first n ranked words must be assigned a distinct code-word, the coding system must provide at least n different encodings, that is, one needs code-words of length satisfying

$$q^C \approx n.$$

Solving for C yields

$$C \approx \log_q n.$$

Thus the required code length grows logarithmically with the rank of the word. Also, the cost of transmitting a word is assumed to be approximately proportional to its length. The underlying idea is that longer words require more resources to produce and communicate. For example, a long word typically requires more letters to be written, more time and effort to be articulated, and more symbols to be transmitted through a communication channel than a short word. Consequently, longer code-words are regarded as more costly than shorter ones.

Mandelbrot models the cost of the n-th ranked word by

$$C_n = a + b \log(n + m).$$

Here, $a > 0$ represents a fixed baseline cost independent of the rank, such as the cost of initiating a word or inserting a space, while $b > 0$ converts code length into cost units and determines how strongly long words are penalized. The shift parameter $m \geq 0$ ensures that the most frequent words have finite cost and improves the empirical fit for low ranks. The logarithmic form of the cost function is therefore a direct consequence of the exponential growth in the number of possible code-words with their length.

Information content. Besides being inexpensive to produce, a useful language should also convey a large amount of information. Following Shannon's information theory, the information content associated with an event of probability p is measured by

$$-\log p.$$

Hence rare words carry more information than highly frequent ones. If the n-th ranked word is used with frequency p_n, then the average information conveyed per word is given by the *Shannon entropy*

$$H = -\sum_{n\geq 1} p_n \log p_n.$$

Large entropy corresponds to a more diverse and informative language, whereas small entropy indicates a highly repetitive one. On the other hand, each word has an associated production cost C_n. Averaging over all words yields the mean cost per word

$$C = \sum_{n\geq 1} p_n C_n.$$

Mandelbrot proposed that an efficient language balances these two competing effects by minimizing the ratio

$$\frac{C}{H},$$

which may be interpreted as the average production cost per unit of information transmitted. Viewing the normalized word frequencies $(p_n)_{n\geq 1}$ as a probability distribution, Mandelbrot minimizes the ratio $\frac{C}{H}$ over all choices of $(p_n)_{n\geq 1}$ satisfying the normalization condition

$$\sum_{n\geq 1} p_n = 1.$$

Applying the method of Lagrange multipliers yields the relation

$$p_n \propto e^{-\beta C_n}, \tag{5.4}$$

where $\beta > 0$ is a Lagrange multiplier associated with the entropy constraint. Substituting the logarithmic cost function

$$C_n = a + b\log(n+m)$$

into the relation (5.4) yields

$$p_n \propto e^{-\beta(a+b\log(n+m))} = e^{-\beta a}(n+m)^{-\beta b}.$$

Since the factor $e^{-\beta a}$ is independent of n, it only affects the normalization constant and may therefore be absorbed into it. Consequently,

$$p_n \propto (n+m)^{-B}, \qquad \text{where } B = \beta b.$$

Thus the optimal rank–frequency distribution follows a *Zipf–Mandelbrot law*, which exhibits a power–law tail. In particular, for large n,

$$p_n \sim n^{-B},$$

so the frequency decays polynomially rather than exponentially with the rank.

This derivation shows that a power-law distribution can emerge from optimizing the efficiency of a static code. In contrast to preferential attachment, where heavy-tails arise from dynamic reinforcement, the Zipf-Mandelbrot mechanism attributes heavy-tails to an equilibrium between production cost and communicative efficiency. Thus, heavy-tailed behavior may arise not only from stochastic growth mechanisms, but also from principles of efficient communication and optimal coding.

5.5 Extremal processes

In many applications, the most important events are not typical observations, but rather extreme ones. Examples include large financial losses, catastrophic natural disasters such as earthquakes and floods, extreme loads on engineering systems, or unusually high network traffic in communication systems. Such events are rare, but they often dominate the overall risk and system behavior. Consequently, understanding the statistical structure of extremes is of central importance in probability theory, statistics, and applications.

A natural mathematical framework for studying such phenomena is provided by *extremal processes*. Given a sequence of random variables $(X_i)_{i\geq 1}$, we consider the maximum

$$M_n = \max(X_1, \dots, X_n).$$

This process captures the evolution of record values over time and provides a simple yet powerful model for extreme events.

At first glance, extremal processes may appear less rich than additive processes, since the maximum depends only on the largest observation. However, this intuition is misleading. Extremal processes exhibit a rich asymptotic structure and, remarkably, give rise to universal limiting distributions, in close analogy to the central limit theorem.

There is a strong conceptual parallel to the additive setting, which we have studied so far: just as stable distributions characterize limits of normalized sums, a special class of distributions will characterize limits of normalized maxima. However, there is also a crucial difference. In additive processes, heavy-tailed limits arise only if the underlying variables are already heavy-tailed, typically with infinite variance. In contrast, extremal processes can generate heavy-tailed limiting distributions even when the underlying variables are light-tailed or even bounded. Thus, extremal processes provide a fundamentally different mechanism for the emergence of heavy-tails, which is studied in the framework of *extreme value theory*.

First recall the definition of maximum

$$M_n := \max(X_1, \dots, X_n),$$

where $(X_i)_{i\geq 1}$ are i.i.d. with distribution F. Similarly, the minimum can be expressed as

$$\min(X_1, \dots, X_n) = -\max(-X_1, \dots, -X_n).$$

We aim to study the behavior of M_n as $n \to \infty$. In contrast to additive and multiplicative processes, the distribution of M_n admits a simple closed form:

$$\mathbb{P}(M_n \leq t) = F(t)^n.$$

Moreover, let $x_F := \sup\{x : F(x) < 1\}$ denote the right endpoint of the distribution. Then

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq t) = \begin{cases} 0, & t < x_F, \\ 1, & t = x_F. \end{cases}$$

Thus, M_n converges to the right endpoint of the distribution (or diverges if $x_F = \infty$), and therefore, without further rescaling, provides no meaningful information about its fluctuations. Indeed, since $F(t)^n \to 0$ for all $t < x_F$, the maximum M_n eventually exceeds any fixed threshold below the right endpoint, which shows that its asymptotic behavior is entirely governed by the upper tail of the distribution.

In contrast to the additive setting, where classical limit theorems provide natural centering and scaling, no such approximation is available a priori for extremal processes. Consequently, identifying an appropriate normalization is not only necessary to obtain a non-degenerate limit, but also constitutes a central part of the analysis.

5.5.1 Max-Stable Distributions

We seek constants $(a_n) > 0$ and $(b_n) \in \mathbb{R}$, for $n \geq 1$, such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} Z,$$

for some non-degenerate limiting random variable Z.

In contrast to the additive setting, where the Gaussian distribution arises naturally due to its stability under summation, it is not immediately clear which distributions should arise in the study of maxima. To gain intuition, we look for distributions that are stable under maxima, in analogy with stable laws in the additive case.

A first indication comes from simple examples. For instance, if $X_i \sim \text{Exp}(1)$, then one can show that

$$M_n - \log n \xrightarrow{d} e^{-e^{-x}},$$

where $\Lambda(x) = e^{-e^{-x}}$ is the Gumbel distribution. Here, we choose $a_n = 1$ and $b_n = \log n$. Notice that this gives $M_n \overset{d}{=} \log n + Z + o(1)$, suggesting that extremes tend to change quickly initially before eventually settling down.

Moreover, the Gumbel distribution is itself stable under maxima, in the sense that the maximum of i.i.d. Gumbel random variables is again Gumbel up to an affine transformation. Indeed, if $X_1, \ldots, X_n$ are i.i.d. random variables with Gumbel distribution and $M_n = \max\{X_1, \ldots, X_n\}$, then

$$\Pr(M_n - \log n \leq t) = F(t + \log n)^n = \left(e^{-e^{-t-\log n}}\right)^n = e^{-ne^{-t-\log n}} = e^{-e^{-t}}.$$

This suggests that such distributions should play a role analogous to stable distributions in the additive case. This motivates the following definition.

Definition 5.6. *A distribution F is said to be* max-stable *if for every $n \geq 2$ and i.i.d. random variables $X_1, \ldots, X_n$ with distribution F, there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that*

$$\max(X_1, \ldots, X_n) \overset{d}{=} c_n X_1 + d_n.$$

A random variable is said to be max-stable if its distribution is max-stable.

The following theorem shows that max-stable distributions play the same role for maxima as stable laws do for sums.

Theorem 5.7. *A random variable Z is max-stable if and only if there exist i.i.d. random variables $(X_i)_{i\geq 1}$ and sequences $(a_n)_{n\geq 1}$, $(b_n)_{n\geq 1}$ with $a_n > 0$ such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} Z.$$

The above result may be interpreted as a central limit theorem for extremal processes, as it characterizes precisely which non-degenerate distributions can arise as limits of suitably normalized maxima. However, in this general form, the theorem remains somewhat abstract, since it does not provide direct insight into the structure of the limiting distributions nor into the conditions under which different limits appear.

As in the additive case (see Theorem 4.4), the following result provides the explicit description of max-stable distributions. Since the proof follows the same line of argument as in the additive setting, we omit it here. For a full proof, see Theorem 3.2.2 in Embrechts et al. [1997].

While there are strong parallels between additive and extremal processes, there are also important differences. In the additive setting, the generalized central limit theorem ensures that, under mild conditions, a non-degenerate limiting distribution exists for a wide class of underlying distributions (after suitable normalization). In contrast, in the extremal setting, the situation is more delicate: there are several commonly encountered distributions for which no non-degenerate limit exists for the normalized maxima, regardless of the choice of normalization sequences. In this sense, limits of extremal processes are significantly more fragile than those arising in the additive case.

So far, we have characterized the possible limiting distributions of extremal processes through the class of max-stable distributions. However, beyond identifying the Gumbel distribution as one example, this characterization remains rather abstract. In order to formulate a more concrete extremal analogue of the central limit theorem, we first need a more explicit description of the class of max-stable distributions.

In contrast to the additive case, where stable distributions are naturally described via their characteristic functions, extremal processes admit a more direct representation in terms of distribution functions. Indeed, since the maximum of independent random variables is expressed through the product of their distribution functions, it is natural to expect a structural characterization in these terms.

The following theorem provides such a representation.

Theorem 5.8. *A non-degenerate random variable X is max-stable if and only if $X \stackrel{d}{=} aZ + b$, for some $a > 0$, $b \in \mathbb{R}$, where the distribution of Z is one of the following:*

- *Fréchet:* $\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}) & x > 0, \end{cases}$

- *Weibull:* $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x \le 0 \\ 1 & x > 0, \end{cases}$
- *Gumbel:* $\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R},$

for $\alpha > 0$.

Proof Sketch. The argument is adapted from Proposition 0.3 in Resnick [1987]. Suppose that there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ and a non-degenerate distribution G such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G,$$

or equivalently,

$$F^n(a_n x + b_n) \to G(x).$$

Fix $t > 0$. Since $[nt] \sim nt$, we may consider the subsequence indexed by $[nt]$. Then

$$F^{[nt]}(a_{[nt]} x + b_{[nt]}) \to G(x).$$

On the other hand,

$$F^{[nt]}(a_n x + b_n) = \big(F^n(a_n x + b_n)\big)^{[nt]/n} \to G^t(x).$$

Hence the same sequence of distribution functions converges under two different affine normalizations. By the convergence-to-types theorem (see Theorem A1.5 in Embrechts et al. [1997]), if

$$F_n(a_n x + b_n) \to G(x), \qquad F_n(\alpha_n x + \beta_n) \to H(x),$$

with non-degenerate limits G and H, then there exist constants $a > 0$ and $b \in \mathbb{R}$ such that

$$\frac{a_n}{\alpha_n} \to a, \qquad \frac{b_n - \beta_n}{\alpha_n} \to b,$$

and

$$H(x) = G(ax + b).$$

Applying this theorem yields measurable functions $\alpha(t) > 0$ and $\beta(t) \in \mathbb{R}$ satisfying

$$G^t(x) = G(\alpha(t)x + \beta(t)). \tag{5.5}$$

Now compute G^{ts} in two different ways

$$G^{ts}(x) = G(\alpha(ts)x + \beta(ts)), \tag{5.6}$$
$$G^{ts}(x) = (G^s(x))^t. \tag{5.7}$$

Using the identity (5.5) twice,

$$G^s(x) = G(\alpha(s)x + \beta(s)),$$

hence (5.7) gives

$$\begin{aligned}(G^s(x))^t &= G(\alpha(t)(\alpha(s)x + \beta(s)) + \beta(t)),\\ &= G(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)).\end{aligned} \tag{5.8}$$

Since G is non-degenerate, the affine arguments in (5.6) and (5.8) must coincide, and thus

$$\begin{aligned}\alpha(ts) &= \alpha(t)\alpha(s),\\ \beta(ts) &= \alpha(t)\beta(s) + \beta(t).\end{aligned}$$

The first equation is the *multiplicative Cauchy functional equation*. Under measurability assumptions, its solutions are of the form

$$\alpha(t) = t^{-\theta}, \qquad \theta \in \mathbb{R}.$$

Substituting this expression into the second equation determines the corresponding form of β. According to whether $\theta > 0$, $\theta < 0$, or $\theta = 0$, one obtains respectively the Fréchet, Weibull, and Gumbel distributions. Therefore every non-degenerate max-stable distribution belongs, up to affine transformations, to one of these three classes. □

The three distributions appearing in this representation are referred to as the *extreme value distributions*. Despite belonging to the same class, they exhibit substantially different behavior. The Fréchet distribution is heavy-tailed with non-negative support, the Weibull distribution has bounded non-positive support and can be either heavy-tailed or light-tailed, and the Gumbel distribution is supported on the entire real line and is light-tailed.

Although these distributions differ significantly in their tail behavior and support, they are closely related. In particular, the following lemma shows that the Weibull and Gumbel distributions can be obtained as simple transformations of the Fréchet distribution.

Lemma 5.9. *Let X be a random variable. The following statements are equivalent*

(i) X has Fréchet distribution Φ_α,

(ii) $-1/X$ has Weibull distribution Ψ_α,

(iii) $\log(X^\alpha)$ has Gumbel distribution Λ.

Proof. $(i) \Longrightarrow (ii)$: Assume $X \sim \Phi_\alpha$ and define $Y := -\frac{1}{X}$. Since $X > 0$ a.s., we have $Y \leq 0$ a.s. For $y \leq 0$,

$$F_Y(y) = \mathbb{P}\left(-\frac{1}{X} \leq y\right) = \mathbb{P}\left(X \leq -\frac{1}{y}\right) = \Phi_\alpha\left(-\frac{1}{y}\right) = \exp\big(-(-y)^\alpha\big).$$

Moreover, $F_Y(y) = 1$ for $y > 0$. Hence $Y \sim \Psi_\alpha$.
The converse $(ii) \Longrightarrow (i)$ follows by reversing the calculation.

$(i) \implies (iii)$: Assume $X \sim \Phi_\alpha$ and define $Z := \log(X^\alpha)$. Since $X > 0$ a.s., Z is well-defined. For $z \in \mathbb{R}$,

$$F_Z(z) = \mathbb{P}(\log(X^\alpha) \leq z) = \mathbb{P}(X \leq e^{z/\alpha}) = \Phi_\alpha(e^{z/\alpha}) = \exp(-e^{-z}).$$

Hence $Z \sim \Lambda$.
The converse $(iii) \implies (i)$ again follows by reversing the calculation.

Thus, the three statements are equivalent. □

This lemma highlights that the three extreme value distributions are not fundamentally distinct, but rather arise from one another through simple transformations. This structural connection plays a key role in the analysis of extremal limits. Finally, it is useful to note that all three families can be unified into a single parametric form. Specifically, the *generalized extreme value distribution* is given by

$$H(x) = \exp\left(-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right),$$

for parameters $\mu \in \mathbb{R}$, $\sigma > 0$, and $\xi \in \mathbb{R}$,with the constraint $1 + \xi(x-\mu)/\sigma > 0$. The cases $\xi > 0$, $\xi < 0$, and $\xi = 0$ correspond to the Fréchet, Weibull, and Gumbel distributions, respectively.

Having obtained an explicit characterization of max-stable distributions, we are now in a position to describe precisely when each of these distributions arises as a limit of normalized maxima. This is achieved by the extremal central limit theorem, which we state next.

5.5.2 The Extremal Central Limit Theorem

We can now state the extremal analogue of the generalized central limit theorem.

Theorem 5.10 (Extremal Central Limit Theorem)**.** *Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with distribution F, and let $M_n := \max\{X_1, \ldots, X_n\}$. Then there exist sequences of constants $(a_n)_{n\geq 1} > 0$ and $(b_n)_{n\geq 1}$ such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G,$$

if and only if G has a non-degenerate max-stable distribution. In this case, the distribution of G belongs to the generalized extreme value family. Furthermore,

(i) G is Fréchet distributed if and only if $\overline{F}(x) = x^{-\alpha}L(x)$, for some slowly varying function L;

(ii) G is Weibull distributed if and only if $\overline{F}(\, x_F - \frac{1}{x}\,) = x^{-\alpha}L(x)$, for a finite upper endpoint $x_F := \sup\{x : F(x) < 1\} < \infty$, $\alpha > 0$ and slowly varying function L;

(iii) *G is Gumbel distributed if and only if there exists $z < x_F$ such that, for $x \in (z, x_F)$, $\overline{F}(x) = c(x) \exp\left\{-\int_z^x \frac{\beta(t)}{g(t)}\, dt\right\}$, where $c(x) \to c \in (0,\infty)$, $\beta(x) \to 1$ as $x \uparrow x_F$, and g is a positive, absolutely continuous function satisfying $g'(x) \to 0$ as $x \uparrow x_F$.*

In case (iii) g plays the role of an auxiliary scaling function; it can be taken as the mean excess function of F near x_F.

We restrict attention to the first case, in which G is Fréchet distributed, as treated in Theorem 3.3.7 of Embrechts et al. [1997]. The Weibull and Gumbel cases are discussed in Sections 3.3.2 and 3.3.3 of Embrechts et al. [1997]. To proceed, we note that the centering constants b_n are asymptotically negligible relative to a_n; this follows from the convergence-of-types theorem (see Appendix A1.5 in Embrechts et al. [1997]).

Lemma 5.11. *If*

$$F^n(c_n x + d_n) \to \Phi_\alpha(x), \qquad \text{as } n \to \infty,$$

for $x > 0$, $c_n > 0$ and $d_n \in \mathbb{R}$, then the centering is asymptotically negligible:

$$\frac{d_n}{c_n} \to 0.$$

We now proceed to the proof of Theorem 5.10.

Proof of Theorem 5.10 (i). ($\Leftarrow$) Suppose that $\bar{F}(x) = x^{-\alpha} L(x)$ with L slowly varying. We define the normalising sequence via the quantile function

$$c_n := F^{-1}(1 - 1/n) := \inf\{x \in \mathbb{R} : F(x) \geq 1 - 1/n\}.$$

Then $c_n \to \infty$ and $\bar{F}(c_n) \sim 1/n$. For $x > 0$,

$$\bar{F}(c_n x) = (c_n x)^{-\alpha} L(c_n x) = c_n^{-\alpha} x^{-\alpha} L(c_n x).$$

By slow variation, $L(c_n x)/L(c_n) \to 1$, hence

$$\frac{\bar{F}(c_n x)}{\bar{F}(c_n)} \to x^{-\alpha}.$$

Since $n\bar{F}(c_n) \to 1$, it follows that

$$n\,\bar{F}(c_n x) = n\,\bar{F}(c_n)\,\frac{\bar{F}(c_n x)}{\bar{F}(c_n)} \longrightarrow x^{-\alpha}.$$

Therefore,

$$\mathbb{P}(M_n \leq c_n x) = F(c_n x)^n = \big(1 - \bar{F}(c_n x)\big)^n \longrightarrow \exp(-x^{-\alpha}),$$

using $(1 - u_n)^n \to e^{-t}$ when $n u_n \to t$. For $x \leq 0$, clearly $F(c_n x)^n \to 0$. Hence

$$\frac{M_n}{c_n} \xrightarrow{d} \Phi_\alpha.$$

($\Rightarrow$) Assume that there exist sequences $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$F^n(c_n x + d_n) \to \Phi_\alpha(x) := \exp(-x^{-\alpha}), \qquad x > 0.$$

Since $F^n(c_n x + d_n) \to e^{-x^{-\alpha}} \in (0,1)$, we must have $\bar{F}(c_n x + d_n) \to 0$. Set

$$u_n(x) := \bar{F}(c_n x + d_n).$$

Then

$$F^n(c_n x + d_n) = (1 - u_n(x))^n \to e^{-x^{-\alpha}}.$$

By the standard limit $(1 - u_n)^n \to e^{-t}$ if and only if $n u_n \to t$, we obtain

$$n\bar{F}(c_n x + d_n) \to x^{-\alpha}, \qquad x > 0.$$

In particular, setting $x = 1$,

$$n\bar{F}(c_n + d_n) \to 1. \tag{5.9}$$

To proceed, we need to show that the centering d_n is asymptotically negligible relative to c_n. This follows from Lemma 5.11, which gives $d_n/c_n \to 0$.
We now compare the tail at different values of x. For $x, y > 0$,

$$\frac{n\bar{F}(c_n x + d_n)}{n\bar{F}(c_n y + d_n)} \to \frac{x^{-\alpha}}{y^{-\alpha}}.$$

Fix $y = 1$. Then

$$\frac{\bar{F}(c_n x + d_n)}{\bar{F}(c_n + d_n)} \to x^{-\alpha}.$$

Since $d_n/c_n \to 0$, we have $c_n + d_n \sim c_n$ and $c_n x + d_n \sim c_n x$. Hence, for every $\varepsilon \in (0, x)$ and all sufficiently large n,

$$c_n(x - \varepsilon) \le c_n x + d_n \le c_n(x + \varepsilon),$$

and also

$$c_n(1 - \varepsilon) \le c_n + d_n \le c_n(1 + \varepsilon).$$

By monotonicity of $\bar{F}$ and a sandwich argument, letting $n \to \infty$ and then $\varepsilon \downarrow 0$, we obtain

$$\frac{\bar{F}(c_n x)}{\bar{F}(c_n)} \to x^{-\alpha}.$$

In particular, $n\bar{F}(c_n) \to 1$. For $x > 0$,

$$\frac{\bar{F}(c_n x)}{\bar{F}(c_n)} = \frac{n\bar{F}(c_n x)}{n\bar{F}(c_n)} \to x^{-\alpha}.$$

Since $c_n \to \infty$, this implies

$$\frac{\bar{F}(tx)}{\bar{F}(t)} \to x^{-\alpha} \qquad (t \to \infty).$$

This is precisely the statement that $\overline{F}$ is regularly varying with index $-\alpha$, i.e. $\overline{F} \in \mathcal{RV}(-\alpha)$, which completes the proof of case (i) with normalizing constants $a_n = c_n$ and $b_n = 0$. $\square$

The normalizing constants $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ are determined by the tail behaviour of the underlying distribution F. In contrast to the classical central limit theorem, where the normalization is essentially universal, the sequences a_n and b_n in extreme value theory depend strongly on the maximum domain of attraction of F. A natural choice for the centering sequence is given by the upper quantiles

$$b_n = F^{-1}(1 - 1/n),$$

which satisfy

$$n\overline{F}(b_n) \approx 1.$$

Hence b_n represents the typical location of the sample maximum M_n. The scaling constants a_n are then chosen such that fluctuations around this level remain of order one. Consequently, the growth of the normalizing constants reflects the tail behaviour of F: for example, b_n grows polynomially for Pareto-type tails, logarithmically for exponential tails, and of order $\sqrt{2\log n}$ for Gaussian tails. Thus, in extreme value theory the normalization itself contains important information about the underlying distribution. See Section 3.3 in Embrechts et al. [1997] for a detailed discussion.

5.5.3 An alternative approach: the Generalized Pareto Distribution

The Extremal Central Limit Theorem of the previous section (see Theorem 5.10) showed that, under suitable normalizations, the maxima of an i.i.d. sequence can have only a limited class of non-degenerate limiting distributions, namely the max-stable distributions appearing in the generalized extreme value family. Thus, the asymptotic behavior of extremes can be described through the limiting distribution of normalized maxima.

There is, however, another natural way to study extremes. Instead of looking at the maximum of a sample, one may focus on the observations that exceed a high threshold and study the size of these exceedances. This leads to a threshold-based description of tail behavior.

The fundamental result connecting these two viewpoints is the Pickands–Balkema–de Haan theorem: whenever the distribution of normalized maxima converges to a non-degenerate extreme value limit, the corresponding excess distribution over high thresholds is asymptotically described by a generalized Pareto distribution.

We now introduce this generalized Pareto distribution and explain its role in modeling threshold exceedances. First, we begin by formalizing the notion of exceedances over a high threshold.

Definition 5.12. *Let X be a real-valued random variable with distribution F, and let u be a threshold such that $\mathbb{P}(X > u) > 0$. The* excess distribution over the threshold u *is defined by*

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

This describes the distribution of the overshoot $X - u$, given that the threshold u has been exceeded.

The central idea of the peaks-over-threshold approach is that, for large values of u, the distribution F_u can be approximated by a simple parametric family, regardless of the behavior of F away from its tail.

The generalized Pareto distribution. The limiting family is given by the generalized Pareto distributions (see Definition 3.4.9 in Embrechts et al. [1997]).

Definition 5.13. *For parameters $\xi \in \mathbb{R}$ and $\sigma > 0$, the* generalized Pareto distribution *is defined by its distribution*

$$H_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0, \end{cases}$$

for all $y \geq 0$ such that $1 + \xi y/\sigma > 0$.
Equivalently, the support is given by

$$y \in \begin{cases} [0, \infty), & \xi \geq 0, \\ [0, -\sigma/\xi], & \xi < 0. \end{cases}$$

The parameter ξ determines the tail behavior: $\xi > 0$ corresponds to heavy-tailed distributions, $\xi = 0$ to exponential-type tails, and $\xi < 0$ to distributions with a finite upper endpoint.

We now state the fundamental result linking exceedances to extreme value limits.

Theorem 5.14 (Pickands–Balkema–de Haan)**.** *Let F be a distribution with right endpoint $x_F = \sup\{x : F(x) < 1\}$. Then F belongs to the domain of attraction of a generalized extreme value distribution with index $\xi \in \mathbb{R}$ if and only if there exists a positive function $\sigma(u)$ such that*

$$\sup_{0 \leq y < x_F - u} \left| F_u(y) - H_{\xi,\sigma(u)}(y) \right| \longrightarrow 0, \qquad \text{as } u \to x_F.$$

Proof. See Theorem 3.4.13 in Embrechts et al. [1997]. □

The Pickands–Balkema–de Haan theorem provides a precise asymptotic description of the behavior of large exceedances. It states that, for sufficiently high thresholds u, the conditional distribution of the excess $X - u$, given that $X > u$, can be approximated by a generalized Pareto distribution.

In other words, while the original distribution F may be complicated, its tail behavior beyond a high threshold is asymptotically universal and depends only on the extreme value index ξ. The specific form of F away from its upper tail becomes irrelevant in this regime.

Moreover, the theorem shows that the same parameter ξ that appears in the limiting distribution of normalized maxima also governs the distribution of threshold exceedances.

This establishes a strong link between the two approaches to extreme value theory: the so-called block maxima method from the previous section and the peaks-over-threshold method.

From a modeling perspective, this result justifies approximating the distribution of exceedances over a high threshold by a generalized Pareto distribution, even when the underlying distribution F is unknown. In particular, it allows one to model rare, large events using a simple parametric family determined by the single shape parameter ξ, which captures the tail heaviness.

5.6 Application: The time between record-breaking events

In many classical examples of extremal processes, such as world records in athletics or extreme environmental measurements, the sequence of record values exhibits a clear slowing down over time, suggesting a light-tailed limiting behavior. However, if instead of focusing on the *values* of the records we consider the *times between successive records*, a different phenomenon emerges. Empirically, one observes long periods with no new records interspersed with bursts of frequent record-breaking events. This irregular behavior suggests that the inter-record times exhibit heavy-tailed behavior.

Our goal in this section is to formalize this observation and show that, under very general assumptions, the waiting times between records are indeed heavy-tailed, more specifically regularly varying.

Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with continuous distribution F. Define the running maximum process $M_n := \max\{X_1, \dots, X_n\}$. Let $(L_k)_{k\geq 1}$ denote the sequence of record times:

$$L_1 := 1, \qquad L_{k+1} := \min\{n > L_k : X_n > M_{L_k}\}, \quad k \geq 1.$$

Define the inter-record times

$$T_k := L_{k+1} - L_k, \quad k \geq 1.$$

Theorem 5.15. *Suppose that F is continuous. Then for every $k \geq 1$, the inter-record time T_k is heavy-tailed. More precisely,*

$$\mathbb{P}(T_k > n) \sim \frac{(\log n)^{k-1}}{(k-1)!\, n}, \qquad n \to \infty.$$

In particular, T_k is regularly varying with index -1.

The asymptotic behavior above reflects the fact that later records become progressively harder to break. Indeed, after the k-th record has occurred, the record level is typically large, and the probability that a new observation exceeds it becomes correspondingly small. Thus the waiting times between records should increase with the record level itself, rather than merely by a fixed constant factor.

The corresponding statement in Nair et al. [2022] (Theorem 7.9) omits the logarithmic factor $(\log n)^{k-1}$. This comes from treating the passage from one record level to the next as contributing only a constant multiplicative correction, whereas the relevant record levels themselves grow on the order of $\log n$.

Proof. The first step is to reduce the problem to the case where the observations have an exponential distribution. Since F is continuous, the random variables

$$U_i := F(X_i), \qquad i \geq 1,$$

are i.i.d. uniformly distributed on $[0, 1]$. Applying the monotone transformation

$$E_i := -\log(1 - U_i),$$

we obtain i.i.d. exponential random variables with mean 1.
This transformation does not change the order of the observations. Hence it does not change which observations are records, nor the waiting times between records. Therefore, without loss of generality, we may assume from now on that

$$X_i \sim \mathrm{Exp}(1), \qquad i \geq 1.$$

Let M_k denote the value of the k-th record. We first show that

$$M_k \sim \Gamma(k, 1).$$

For $k = 1$, the first record is simply the first observation, so

$$M_1 = X_1 \sim \mathrm{Exp}(1),$$

which is precisely a Gamma distribution with shape parameter 1 and rate 1. Suppose now that

$$M_k \sim \Gamma(k, 1).$$

The next record value M_{k+1} is the first observation after the k-th record which exceeds M_k. Condition on the event $M_k = m$. If X is an exponential random variable and $X > m$, then by the memoryless property of the exponential distribution,

$$(X - m \,\big|\, X > m) \sim \mathrm{Exp}(1).$$

Indeed, for $x \geq 0$,

$$\mathbb{P}(X - m > x \mid X > m) = \frac{\mathbb{P}(X > m + x)}{\mathbb{P}(X > m)} = \frac{e^{-(m+x)}}{e^{-m}} = e^{-x}.$$

Thus the amount by which the next record exceeds the previous record is distributed as an exponential random variable. Hence, in distribution,

$$M_{k+1} \stackrel{d}{=} M_k + E,$$

where $E \sim \mathrm{Exp}(1)$ is independent of M_k. Therefore M_{k+1} is distributed as the sum of $k+1$ independent exponential random variables with rate 1, and hence

$$M_{k+1} \sim \Gamma(k+1, 1).$$

This proves the claim by induction. Consequently, M_k has density

$$f_{M_k}(x) = \frac{x^{k-1}}{(k-1)!} e^{-x}, \qquad x > 0.$$

Next, condition on the event $M_k = x$. A future observation exceeds the current record level x with probability

$$\mathbb{P}(X_i > x) = e^{-x}.$$

Hence, conditional on $M_k = x$, the waiting time T_k is geometrically distributed with success probability e^{-x}, and therefore

$$\mathbb{P}(T_k > n \mid M_k = x) = (1 - e^{-x})^n.$$

Integrating against the density of M_k gives the integral representation

$$\mathbb{P}(T_k > n) = \frac{1}{(k-1)!} \int_0^\infty x^{k-1} e^{-x} (1 - e^{-x})^n \, dx,$$

which coincides with the representation derived in Neuts [1967].

Now substitute

$$y = e^{-x}, \qquad x = -\log y, \qquad dy = -e^{-x} \, dx.$$

Then

$$\mathbb{P}(T_k > n) = \frac{1}{(k-1)!} \int_0^1 (-\log y)^{k-1} (1-y)^n \, dy.$$

We now analyze the asymptotic behavior as $n \to \infty$. Set $y = \frac{u}{n}$, then $dy = \frac{du}{n}$, and therefore

$$\mathbb{P}(T_k > n) = \frac{1}{(k-1)!\, n} \int_0^n (\log n - \log u)^{k-1} \left(1 - \frac{u}{n}\right)^n du.$$

For every fixed $u > 0$,

$$\left(1 - \frac{u}{n}\right)^n \to e^{-u},$$

and

$$(\log n - \log u)^{k-1} = (\log n)^{k-1} \left(1 - \frac{\log u}{\log n}\right)^{k-1}.$$

Hence, for fixed u,

$$\frac{(\log n - \log u)^{k-1}}{(\log n)^{k-1}} \to 1.$$

Consequently,

$$\mathbb{P}(T_k > n) \sim \frac{(\log n)^{k-1}}{(k-1)!\, n} \int_0^\infty e^{-u} \, du.$$

Since

$$\int_0^\infty e^{-u}\,du = 1,$$

we conclude that

$$\mathbb{P}(T_k > n) \sim \frac{(\log n)^{k-1}}{(k-1)!\,n}.$$

Therefore T_k is regularly varying with index -1. □

This result shows that even when the underlying observations are light-tailed, the waiting times between records are heavy-tailed. Thus, it explains the empirical presence of long periods without new records, punctuated by bursts of frequent record-breaking events.

Chapter 6

Conclusions

This thesis has developed a rigorous and unified treatment of heavy-tailed distributions, organized around three complementary perspectives: structural characterization, additive limit theorems, and the emergence of extremes from multiplicative and extremal processes.

Summary. Chapter 3 established the foundational theory of heavy-tailed distributions. We introduced the class of regularly varying distributions as the natural framework for power-law tail behavior, proved key analytic properties including the Karamata representation theorem, the monotone density theorem, Potter bounds, and that regularly varying tails imply subexponentiality. The subexponential class was characterized through the catastrophe principle: in sums of subexponential random variables, large values are dominated by a single large summand. This was made precise through the principle of a single big jump for subexponential distributions, and contrasted with Cramér's theorem in the light-tailed setting, where large deviations arise instead through a collective conspiracy of many increments.

Chapter 4 addressed additive limit theorems beyond the classical regime. We showed that stable distributions form the universal class of limits of normalized sums of i.i.d. random variables, and characterized them through their characteristic functions. The generalized central limit theorem established that the Gaussian distribution is merely the boundary case $\alpha = 2$ of a continuum of stable distributions and that distributions with regularly varying tails of index $\alpha \in (0, 2)$ lie in the domain of attraction of the corresponding α-stable distribution. The application to financial returns connected these theoretical results to Mandelbrot's empirical observations, showing that heavy-tailed stable distributions provide a more realistic model for financial data than classical Gaussian assumptions.

Chapter 5 examined two further mechanisms generating heavy-tailed behavior. Multiplicative processes, analyzed via logarithmic transformation and the central limit theorem, naturally lead to lognormal limits under finite log-variance, and to even heavier tails via stable distributions when this condition fails. Variations including lower barriers, additive noise, and stochastic recurrence equations all preserve heavy-tailed behavior, showing that this property is a robust consequence of multiplicative dynamics. The

preferential attachment model demonstrated concretely how a power-law degree distribution emerges from a multiplicative growth mechanism. Finally, extremal processes were analyzed through the lens of max-stability, culminating in the extremal central limit theorem, which identifies the Fréchet, Weibull, and Gumbel distributions as the only possible limits of normalized maxima and characterizes the domains of attraction of each. Two further results enriched this picture: the Pickands-Balkema-de Haan theorem established that the generalized Pareto distribution arises as the universal limit for threshold exceedances, providing an alternative and complementary description of extreme tail behavior through the peaks-over-threshold method; and the inter-record times of i.i.d. sequences were shown to be regularly varying, revealing that heavy-tailed waiting times between extremes emerge even from light-tailed observations.

Outlook. Several natural directions suggest themselves for further study. The theory developed in this thesis treats random variables that are independent and identically distributed. Many applications, such as time series in finance, network dynamics, and spatial extremes in environmental science, involve dependence structures or non-stationarity. Extending the results on subexponentiality, stable limits, and extreme value theory to dependent settings is an active area of research; we refer the reader to Embrechts et al. [1997] and Resnick [2007] for a detailed treatment.

The statistical question of estimating tail indices and identifying the domain of attraction from data remains both theoretically rich and practically important. Estimators such as the Hill estimator for the tail index α and goodness-of-fit tests for the generalized extreme value family provide the empirical counterpart to the theoretical framework developed here, and represent a natural next step for anyone wishing to apply this theory to real data.

Appendix A

Principle of the One Big Jump

In the main text we studied events of the form $\{S_n > an\}$, which are natural from the perspective of large deviations of random walks. The theory of subexponential distributions is usually formulated more generally in terms of exceedance probabilities

$$\mathbb{P}(S_n > x) \quad \text{as } x \to \infty.$$

In order to formulate the result, we introduce several auxiliary sequences that describe different aspects of the behavior of the random walk (see Denisov et al. [2008], Section 2).

Definition A.1 (Natural scale sequence)**.** *A sequence $(b_n)_{n\geq 1}$ of positive numbers is called a* natural scale sequence *if the sequence $(S_n/b_n)_{n\geq 1}$ is tight, that is, for every $\varepsilon > 0$ there exists $K > 0$ such that*

$$\sup_{n\geq 1} \mathbb{P}\left(\frac{|S_n|}{b_n} > K\right) < \varepsilon.$$

A natural scale sequence $(b_n)_{n\geq 1}$ describes the typical size of fluctuations of the random walk in the absence of exceptionally large increments. The assumption that (b_n) is a natural scale sequence ensures that S_n remains of order b_n when no large jumps occur.

Definition A.2 (b-insensitivity sequence)**.** *Let $(b_n)_{n\geq 1}$ be a sequence with $b_n \to \infty$. A sequence $(I_n)_{n\geq 1}$ is called a* b-insensitivity sequence *if*

$$\frac{I_n}{b_n} \to \infty$$

and

$$\sup_{x\geq I_n} \sup_{0\leq t\leq b_n} \left|\frac{\bar{F}(x-t)}{\bar{F}(x)} - 1\right| \to 0 \qquad \text{as } n \to \infty.$$

A b-insensitivity sequence $(I_n)_{n\geq 1}$ specifies a region in which the tail of the distribution is stable under perturbations of size b_n. The assumption implies that, for $x \geq I_n$, shifts of order b_n do not significantly affect the tail probability $\bar{F}(x)$.

Definition A.3 (b-truncation sequence)**.** *Let $(b_n)_{n\geq 1}$ be a sequence. A sequence $(h_n)_{n\geq 1}$ is called a* b-truncation sequence *if*

$$\lim_{K\to\infty} \limsup_{n\to\infty} \sup_{x\geq h_n} \frac{n\,\mathbb{P}\Big(\{X_1+X_2>x\}\cap\{X_1,X_2\in(-\infty,-Kb_n)\cup(h_n,\infty)\Big)}{\overline{F}(x)} = 0.$$

A b-truncation sequence $(h_n)_{n\geq 1}$ determines the threshold separating small and large increments. The truncation condition ensures that the probability that two or more extreme increments jointly produce a large deviation is negligible compared to the contribution of a single large increment.

Definition A.4 (h-small-steps sequence)**.** *Let $(h_n)_{n\geq 1}$ be a sequence. A sequence $(J_n)_{n\geq 1}$ is called an* h-small-steps sequence *if*

$$\sup_{x\geq J_n} \frac{\mathbb{P}(S_n>x,\ X_1\leq h_n,\dots,X_n\leq h_n)}{n\bar{F}(x)} \to 0 \qquad \text{as } n\to\infty.$$

An h-small-steps sequence $(J_n)_{n\geq 1}$ characterizes the region in which large deviations cannot be produced by increments that are all bounded by h_n. The assumption implies that, for $x\geq J_n$, the event $\{S_n>x\}$ cannot occur through the accumulation of many small increments.

We are now in a position to state the main result, which formalizes the principle of one big jump for subexponential distributions. The result is from Denisov et al. [2008].

Theorem A.5 (One big jump principle)**.** *Let $(X_i)_{i\geq 1}$ be i.i.d. random variables with distribution F, where F is subexponential. Let $(b_n)_{n\geq 1}$ be a natural scale sequence, $(I_n)_{n\geq 1}$ a b-insensitivity sequence, $(h_n)_{n\geq 1}$ a b-truncation sequence, and $(J_n)_{n\geq 1}$ an h-small-steps sequence such that*

$$h_n = O(b_n), \qquad h_n \leq J_n.$$

Then

$$\sup_{x\geq I_n+J_n} \left|\frac{\mathbb{P}(S_n>x)}{n\bar{F}(x)} - 1\right| \to 0 \qquad \text{as } n\to\infty.$$

The key idea is to decompose the event $\{S_n>x\}$ according to how many increments exceed a suitable truncation level h_n. This yields three contributions:

- **No large increments:** All increments are bounded by h_n. The small-steps property of $(J_n)_{n\geq 1}$ ensures that this contribution is negligible uniformly for $x\geq J_n$.
- **Exactly one large increment:** Suppose that $X_i>h_n$ while $X_j\leq h_n$ for $j\neq i$, and that this single large increment is responsible for the exceedance of the level x. By symmetry, the contribution is

$$n\,\mathbb{P}(S_n>x,\ X_1>h_n,\ X_2,\dots,X_n\leq h_n).$$

Using the insensitivity of the tail $\bar{F}$ and the tightness of the remaining sum, one shows that

$$\mathbb{P}(S_n > x,\ X_1 > h_n,\ X_2, \dots, X_n \le h_n) \sim \bar{F}(x).$$

Hence the total one-big-jump contribution is

$$n\bar{F}(x)(1 + o(1)).$$

- **Two or more large increments:** The truncation property of $(h_n)_{n\ge1}$ ensures that this contribution is $o(n\bar{F}(x))$ uniformly in x.

Combining these estimates shows that $\mathbb{P}(S_n > x) \sim n\bar{F}(x)$, uniformly over the range of x specified in the theorem. Thus the exceedance is produced by a single large jump.

Having outlined the main idea, we now turn to the rigorous proof of Theorem A.5.

Proof of Theorem A.5 . The proof is based on a decomposition of the probability according to the number of large increments. Using a truncation argument and the assumptions, one can show that the contributions of configurations with no large jumps and with at least two large jumps are negligible. More precisely, under the assumptions of the theorem, Lemma 2.4 and Lemma 2.5 in Denisov et al. [2008] imply that

$$\mathbb{P}(S_n > x) = n\,\mathbb{P}(S_n > x,\ X_1 > h_n,\ X_2, \dots, X_n \le h_n)\,(1 + o(1)) \tag{A.1}$$

uniformly for $x \ge J_n$. We therefore focus on estimating the contribution of exactly one large jump, and prove matching upper and lower bounds.

Lower bound. Fix $\varepsilon > 0$. Since $(S_n/b_n)_{n\ge1}$ is tight, we can choose $K > 0$ large such that

$$\liminf_{n\to\infty} \mathbb{P}\big(|S_{n-1}| \le Kb_n\big) \ge 1 - \varepsilon/2.$$

Moreover, the following estimate holds (see Denisov et al. [2008], proof of Lemma 2.1):

$$\lim_{K\to\infty} \liminf_{n\to\infty} \mathbb{P}\Big(S_{n-1} \in [-K^2 b_n, K^2 b_n],\ |X_1| \le Kb_n, \dots, |X_{n-1}| \le Kb_n\Big) = 1. \tag{A.2}$$

By replacing K with $\sqrt{K}$ in (A.2), we obtain

$$\liminf_{n\to\infty} \mathbb{P}\Big(S_{n-1} \in [-Kb_n, Kb_n],\ |X_1| \le \sqrt{K}b_n, \dots, |X_{n-1}| \le \sqrt{K}b_n\Big) \ge 1 - \varepsilon.$$

In particular,

$$\liminf_{n\to\infty} \mathbb{P}\Big(S_{n-1} > -Kb_n,\ |X_1| \le \sqrt{K}\,b_n, \dots, |X_{n-1}| \le \sqrt{K}\,b_n\Big) \ge 1 - \varepsilon. \tag{A.3}$$

Now let $x \ge 0$. On the event

$$\Big\{X_n > x + Kb_n,\ S_{n-1} > -Kb_n,\ |X_1| \le \sqrt{K}\,b_n, \dots, |X_{n-1}| \le \sqrt{K}\,b_n\Big\}$$

we have $S_n = S_{n-1} + X_n > x$, and therefore

$$\mathbb{P}(S_n > x) \geq n\,\mathbb{P}\Big(X_n > x + Kb_n,\ S_{n-1} > -Kb_n,\ |X_1| \leq \sqrt{K}\,b_n, \ldots, |X_{n-1}| \leq \sqrt{K}\,b_n\Big).$$

By independence,

$$\mathbb{P}(S_n > x) \geq n\,\bar{F}(x + Kb_n)\,\mathbb{P}\Big(S_{n-1} > -Kb_n,\ |X_1| \leq \sqrt{K}\,b_n, \ldots, |X_{n-1}| \leq \sqrt{K}\,b_n\Big).$$

Since $(I_n)_{n\geq 1}$ is an insensitivity sequence, for this fixed K we have

$$\sup_{x \geq I_n} \left| \frac{\bar{F}(x + Kb_n)}{\bar{F}(x)} - 1 \right| \to 0.$$

Hence, for all sufficiently large n and uniformly in $x \geq I_n$,

$$\bar{F}(x + Kb_n) \geq (1+\varepsilon)^{-1}\bar{F}(x).$$

Therefore, uniformly in $x \geq I_n$,

$$\frac{\mathbb{P}(S_n > x)}{n\bar{F}(x)} \geq (1+\varepsilon)^{-1}\mathbb{P}\Big(S_{n-1} > -Kb_n,\ |X_1| \leq \sqrt{K}\,b_n, \ldots, |X_{n-1}| \leq \sqrt{K}\,b_n\Big).$$

Taking $\liminf_{n\to\infty}$ and applying (A.3), we obtain

$$\liminf_{n\to\infty} \inf_{x \geq I_n} \frac{\mathbb{P}(S_n > x)}{n\bar{F}(x)} \geq (1+\varepsilon)^{-1}(1-\varepsilon).$$

Finally, letting $\varepsilon \to 0$, we conclude that

$$\liminf_{n\to\infty} \inf_{x \geq I_n} \frac{\mathbb{P}(S_n > x)}{n\bar{F}(x)} \geq 1.$$

Before analyzing the upper bound, we state two key lemmas from Denisov et al. [2008], which will be used in the proof.

Lemma A.6 (Truncated large deviations bound)**.** *There exist constants $C, c > 0$ such that for all n and all $y \geq 0$,*

$$\mathbb{P}\big(S_n > y,\ X_1 \leq cb_n, \ldots, X_n \leq cb_n\big) \ \leq\ Ce^{-y/(cb_n)}.$$

Lemma A.7 (Negligibility of multiple big jumps)**.** *Let $(h_n)_{n\geq 1}$ be a truncation sequence such that $n\bar{F}(h_n) \to 0$. Then, uniformly for $x \geq I_n + J_n$,*

$$\mathbb{P}\big(S_n > x,\ \textit{at least two of the } X_i \textit{ exceed } h_n\big) = o\big(n\bar{F}(x)\big).$$

Upper bound. Set

$$B_i := \{X_i \le h_n\}, \qquad \overline{B}_i := \{X_i > h_n\}.$$

By the reduction (A.1) we stated, it suffices to estimate

$$\mathbb{P}(S_n > x, \overline{B}_1, B_2, \dots, B_n).$$

Write $S_n = X_1 + (S_n - X_1)$. Conditioning on X_1, we obtain

$$\mathbb{P}(S_n > x, \overline{B}_1, B_2, \dots, B_n) = \int_{h_n}^{\infty} \mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1})\, dF(u).$$

Split the integral at $x - J_n$

$$\begin{aligned}\int_{h_n}^{x-J_n} \mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1})\, dF(u) + \int_{x-J_n}^{\infty} \mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1})\, dF(u)\\ =: I_{n,1}(x) + I_{n,2}(x). \end{aligned} \tag{A.4}$$

For the first part,

$$\begin{aligned} I_{n,1}(x) &= \int_{h_n}^{x-J_n} \mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1})\, dF(u) \\ &\le \eta_n\, n \int_{h_n}^{x-J_n} \bar{F}(x-u)\, dF(u) \\ &\le \eta_n\, n \int_{h_n}^{x-h_n} \bar{F}(x-u)\, dF(u) \\ &\le \eta_n\, n\, \mathbb{P}(S_2 > x,\ X_1 > h_n,\ X_2 > h_n) \\ &= o(\bar{F}(x)), \end{aligned} \tag{A.5}$$

where

$$\eta_n := \sup_{y \ge J_n} \frac{\mathbb{P}(S_{n-1} > y,\ B_1, \dots, B_{n-1})}{n\, \bar{F}(y)} \longrightarrow 0,$$

by the small-steps property. Indeed, for $u \le x - J_n$ we have $x - u \ge J_n$, hence

$$\mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1}) \le \eta_n\, n\, \bar{F}(x-u).$$

Moreover, in (A.5) we used that $J_n \ge h_n$. Finally, the last term is $o(\bar{F}(x))$ uniformly in $x \ge I_n + J_n$ by the truncation property.

We now estimate $I_{n,2}(x)$. Since

$$I_{n,2}(x) = \int_{x-J_n}^{\infty} \mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1})\, dF(u),$$

we first note that, by insensitivity, for all sufficiently large n and some fixed constant $K > 0$,

$$\sup_{u \ge I_n} \frac{\bar{F}(u)}{\bar{F}(u + b_n)} \le e^{1/K^2}. \tag{A.6}$$

We distinguish between two cases, since the behavior of the integral depends on whether the integration range is of order b_n, where insensitivity alone suffices, or much larger, where additional control via exponential bounds on the truncated sum is required.

Case 1: $J_n \le Kb_n$. Since $x \ge I_n + J_n$, we have $x - J_n \ge I_n$, so that the insensitivity estimate (A.6) applies. Moreover, using that $\mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1}) \le 1$, we obtain

$$I_{n,2}(x) \le \int_{x-J_n}^{\infty} dF(u) = \bar{F}(x - J_n).$$

To compare $\bar{F}(x - J_n)$ with $\bar{F}(x)$, we apply (A.6) iteratively over increments of size b_n. Since $J_n \le Kb_n$, this yields

$$\bar{F}(x - J_n) \le \left(e^{1/K^2}\right)^{J_n/b_n} \bar{F}(x) = e^{J_n/(K^2 b_n)} \bar{F}(x) \le e^{1/K} \bar{F}(x).$$

Hence,

$$I_{n,2}(x) \le e^{1/K} \bar{F}(x).$$

Case 2: $J_n > Kb_n$. We split the integral $I_{n,2}$ at $x - Kb_n$.
Since $x - Kb_n \ge I_n$, the same argument as above gives

$$\int_{x-Kb_n}^{\infty} \mathbb{P}(S_{n-1} > x - u,\ B_1, \dots, B_{n-1})\, dF(u) \le \bar{F}(x - Kb_n) \le e^{1/K} \bar{F}(x).$$

It remains to estimate the integral over $[x - J_n, x - Kb_n]$. Making the change of variable $y = x - u$, we have $x - u \in [Kb_n, J_n]$. Thus

$$\begin{aligned}
&\int_{x-J_n}^{x-Kb_n} \mathbb{P}(S_n - X_1 > x - u,\ B_2, \dots, B_n)\, F(du) \\
&\qquad \le \bar{F}(x - J_n)\, \mathbb{P}(S_{n-1} > J_n,\ B_1, \dots, B_{n-1}) \\
&\qquad + \int_{Kb_n}^{J_n} \mathbb{P}(S_{n-1} \in dy,\ B_1, \dots, B_{n-1})\, \bar{F}(x - y).
\end{aligned}$$

In order to bound the first term, choose $c > 0$ such that $h_n \le cb_n$. Then by Lemma A.6,

$$\mathbb{P}(S_{n-1} > J_n, B_1, \dots, B_{n-1}) \le Ce^{-J_n/(cb_n)}.$$

Since $x - J_n \ge I_n$, insensitivity yields

$$\bar{F}(x - J_n) \le e^{J_n/(K^2 b_n)} \bar{F}(x).$$

Thus

$$\bar{F}(x - J_n)\, \mathbb{P}(S_{n-1} > J_n,\ B_1, \dots, B_{n-1}) \le Ce^{J_n/(K^2 b_n) - J_n/(cb_n)} \bar{F}(x).$$

Since $K^2 > c$, we have $\frac{1}{K^2} - \frac{1}{c} < 0$, and thus

$$e^{J_n/(K^2 b_n) - J_n/(cb_n)} = e^{\frac{J_n}{b_n}\left(\frac{1}{K^2} - \frac{1}{c}\right)} \to 0 \qquad \text{as } K \to \infty,$$

since $J_n/b_n \geq K$ and K can be chosen arbitrarily large. Hence the term is $o(\bar{F}(x))$.
Now in order to estimate the second term

$$\int_{Kb_n}^{J_n} \mathbb{P}(S_{n-1} \in dy,\ B_1, \ldots, B_{n-1})\, \bar{F}(x-y),$$

partition $[Kb_n, J_n]$ into intervals $(kb_n, (k+1)b_n]$, for $k = K, \ldots, \lfloor J_n/b_n \rfloor$. Then

$$\begin{aligned}
&\int_{Kb_n}^{J_n} \mathbb{P}(S_{n-1} \in dy,\ B_1, \ldots, B_{n-1})\, \bar{F}(x-y) \\
&\quad \leq \sum_{k=K}^{\lfloor J_n/b_n \rfloor} \mathbb{P}\left(\frac{S_{n-1}}{b_n} \in (k, k+1],\ X_1 \leq h_n, \ldots, X_{n-1} \leq h_n \right) \bar{F}(x - kb_n) \\
&\quad \leq \sum_{k=K}^{\lfloor J_n/b_n \rfloor} \mathbb{P}(S_{n-1} > kb_n,\ X_1 \leq cb_n, \ldots, X_{n-1} \leq cb_n)\, \bar{F}(x - kb_n).
\end{aligned}$$

Using $h_n \leq cb_n$ and Lemma A.6,

$$\mathbb{P}\left(\frac{S_{n-1}}{b_n} \in (k, k+1], \ldots \right) \leq Ce^{-k/c}.$$

By repeated use of insensitivity,

$$\bar{F}(x - kb_n) \leq e^{k/K^2} \bar{F}(x).$$

Hence

$$\begin{aligned}
&\int_{Kb_n}^{J_n} \mathbb{P}(S_{n-1} \in dy,\ B_1, \ldots, B_{n-1})\, \bar{F}(x-y) \\
&\qquad \leq C \sum_{k=K}^{\lfloor J_n/b_n \rfloor} e^{-k/c} e^{k/K^2} \bar{F}(x) \leq C \frac{e^{-K/c+1/K}}{1 - e^{-1/c+1/K^2}} \bar{F}(x).
\end{aligned}$$

Combining the above estimates, for every fixed $K > 0$ with $K^2 > c$, we obtain

$$I_{n,2} \leq \left(e^{1/K} + C \frac{e^{-K/c+1/K}}{1 - e^{-1/c+1/K^2}} + o(1) \right) \bar{F}(x),$$

uniformly for $x \geq I_n + J_n$, where $o(1) \to 0$ as $n \to \infty$.
This gives

$$\limsup_{n\to\infty} \sup_{x \geq I_n + J_n} \frac{I_{n,2}}{\bar{F}(x)} \leq e^{1/K} + C \frac{e^{-K/c+1/K}}{1 - e^{-1/c+1/K^2}}.$$

Since $K > 0$ is arbitrary, letting $K \to \infty$ yields

$$\limsup_{n\to\infty} \sup_{x \geq I_n + J_n} \frac{I_{n,2}}{\bar{F}(x)} \leq 1.$$

Combining this with the estimate on $I_{n,1}$, we conclude that

$$\limsup_{n\to\infty} \sup_{x\geq I_n+J_n} \frac{I_{n,2}}{\bar{F}(x)} \leq 1.$$

This proves the required upper bound for the term $I_{n,2}$. Combining this with the estimate on $I_{n,1}(x)$ and substituting in (A.4), we conclude that

$$\limsup_{n\to\infty} \sup_{x\geq I_n+J_n} \frac{\mathbb{P}(S_n > x, \overline{B}_1, B_2, \ldots, B_n)}{\bar{F}(x)} \leq 1.$$

Combining upper and lower bounds yields the claim. □

Declaration of Independence

I hereby declare that I have composed this work independently and without the use of any aids other than those declared. Additionally, I acknowledge the use of generative AI (ChatGPT, GPT-4o) to correct and linguistically refine the text of my thesis.

Bibliography

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999.

P. Billingsley. *Probability and Measure*. Wiley, 3rd edition, 2008.

N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 2013.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

V. P. Chistyakov. A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory of Probability and Its Applications*, 9(4):640–648, 1964.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

D. Denisov, A. B. Dieker, and V. Shneer. Large deviations for random walks under subexponentiality: The big-jump domain. *Annals of Probability*, 2008.

R. Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2019.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: For Insurance and Finance*. Springer, 1997.

W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. II*. Wiley, 1971.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973.

B. Mandelbrot. An informational theory of the statistical structure of language. In *Communication Theory*. Butterworth, 1953.

B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36(4): 394–419, 1963.

J. Nair, A. Wierman, and B. Zwart. *The Fundamentals of Heavy Tails: Properties, Emergence, and Estimation*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2022.

M. F. Neuts. Waiting times between record observations. *Journal of Applied Probability*, 4(1):206–208, 1967.

D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5):292–306, 1976.

Y. V. Prokhorov. An extremal problem in probability theory. *Theory of Probability and Its Applications*, 4(2):201–203, 1959.

S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987.

S. I. Resnick. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, 2007.

H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. *Philosophical Transactions of the Royal Society B*, 213:21–87, 1925.

G. K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, 1949.

B. Zwart. Rare events and heavy tails in stochastic systems, 2008.