Stockholms universitet

Tail Asymptotics: Large Deviation Theory and the Principle of a Single Big Jump

Erik Lind

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2026:15
Matematisk statistik
Maj 2026

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2026:15**
http://www.math.su.se

Tail Asymptotics: Large Deviation Theory and the Principle of a Single Big Jump

Erik Lind*

May 2026

Abstract

In insurance mathematics and risk modelling, the law of large numbers and the central limit theorem describe only the average case and do not quantify the probability of extreme aggregate losses. Letting $S(t)$ denote the total claim payout up to time t, this concern arises in two forms: a long-run question, the decay rate of $P(S(t) \geq xt)$ as $t \to \infty$, and a fixed-horizon question, the behaviour of $P(S(t) > x)$ as $x \to \infty$ for fixed t. Each question requires a different class of claim size distributions and is addressed by a different framework: the long-run question by large deviation theory, the fixed-horizon question by the Principle of a Single Big Jump.

Large deviation theory requires the claim sizes to be light-tailed. Its central object is the rate function $I(x)$, which characterises the exponential rate at which $P(S(t) \geq xt)$ decays. The Principle of a Single Big Jump instead requires claims to be subexponential, a subclass of heavy-tailed distributions, under which $P(S(t) > x) \sim \lambda t \cdot P(X > x)$ as $x \to \infty$.

These results are verified through simulation, and each framework is shown to break down when applied outside its conditions. Despite combining a random number of claims with random claim sizes, the compound aggregate loss process reduces to a single factor times a single tail under the right distributional conditions.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: erik.lind22@gmail.com. Supervisor: Daniel Ahlberg, Maria Deijfen.

Acknowledgements

I would like to thank my supervisors Daniel Ahlberg and Maria Deijfen for their guidance and feedback throughout this project.

I would also like to thank Lars Lidvall for helpful discussions, in particular regarding the proof in Section 2.3.

Claude (Anthropic) was used as an AI assistant during the writing process, primarily for discussing mathematical ideas, code optimization, suggesting structure, and refining language.

Contents

Chapter 1

Introduction

Every insurance company collects premiums (money in) and pays out claims (money out). The total payout, for any given year, may be modest, or in extreme scenarios, large enough that the company cannot meet its obligations. Two central questions in insurance mathematics are therefore: how likely are extreme scenarios, and how much money needs to be set aside to fulfil the company's obligations?

If claim sizes follow a light-tailed distribution, such as the exponential, large deviation theory provides an analytical framework for quantifying these probabilities. This gives rise to two practically useful directions. Given a time span, say five years, and a maximum acceptable probability of being unable to pay, one can approximate how much capital the insurer must hold. Conversely, given a fixed capital reserve, one can estimate the probability that aggregate losses will exceed it over any chosen time span. Both reduce to a single mathematical object: the rate function, which determines how unlikely extreme total payouts are for any given time span.

This framework relies on a crucial requirement: the tail of the claim size distribution must decay fast enough. However, many realistic loss scenarios are better described by distributions where the decay is not fast enough. These distributions have substantial probability mass far out in the tail, meaning massive individual claims are far more likely than in the case of faster tail decay. For such distributions the large deviation framework breaks down entirely.

For heavy-tailed claims the focus shifts. Rather than tracking how aggregate losses accumulate over time, the concern is what can happen within a single year. This is precisely what Solvency II requires: the European Union's regulatory framework for insurance capital, mandating that each insurer holds enough capital to cover aggregate losses at the 99.5th per-

centile [4, Art. 101(3)], that is, a loss level exceeded in only one out of two hundred years. The appropriate framework is the Principle of a Single Big Jump: when claim sizes are heavy-tailed, a catastrophic aggregate loss is driven not by many moderate claims acting together, but by a single extreme claim. This thesis develops both frameworks, examines when each applies, and illustrates the consequences for capital through simulation.

Chapter 2

Theoretical Foundation

The total payout up to time t, denoted $S(t)$, is the central object of this thesis. The introduction identified two practically useful questions: given a time horizon and an acceptable risk level, how much capital suffices, and given a fixed capital reserve (the amount of money the insurance company sets aside to pay for claims), how likely is it to be exceeded. Both questions relate to the tail of $S(t)$, and translating them into mathematics requires two distinct asymptotic frameworks.

The model used throughout is defined in Section 2.1. The number of claims is modelled by a Poisson process, where λ denotes the rate of claim arrivals per unit time. The claim sizes X_i are i.i.d. random variables drawn from a distribution that can be either light-tailed or heavy-tailed. Section 2.2 addresses the case of light-tailed claim sizes. By the law of large numbers, the aggregate loss per unit time converges to $\lambda E[X]$, the expected number of claims per unit time multiplied by the expected claim size. A large deviation is precisely an outcome that persistently exceeds this average: we fix a level $x > \lambda E[X]$ and ask how the probability $P(S(t) \geq xt)$ behaves as t grows. Note that the condition $x > \lambda E[X]$ only classifies what counts as a large deviation, an outcome that consistently exceeds the long-run average. It says nothing about how often such outcomes occur or how quickly their probability decays, and this is precisely what the law of large numbers cannot answer. Large deviation theory gives the answer: this probability is approximately $e^{-tI(x)}$, where $I(x)$ is the rate function.

Section 2.3 identifies the critical requirement behind this result: the moment generating function of the claim sizes must be finite for some $\theta > 0$. For many practically relevant claim types, this assumption fails. Individual losses can be so extreme that no exponential function can bound the tail. Such distributions are called heavy-tailed, characterised by tail decay slower than any exponential. As the canonical heavy-tailed distribution we

use the Pareto distribution and show that the rate function $I(x)$ cannot be computed for Pareto claim sizes, causing the large deviation theory (LDT) framework to break down entirely.

Section 2.4 develops the second framework. For heavy-tailed claim sizes the long-run question can no longer be answered by the LDT framework: extreme outcomes do not become negligible as time grows. The natural question is therefore fixed in time: given a fixed time horizon, what is the probability that the total payout exceeds a catastrophic threshold? Fixing t and asking how $P(S(t) > x)$ behaves as x grows, the Principle of a Single Big Jump gives the answer: $P(S(t) > x) \sim \lambda t \cdot P(X > x)$, meaning that extreme aggregate losses are driven by a single catastrophic claim. This is the appropriate framework for the Solvency II question: what capital x is needed such that $P(S(1) > x)$ is at most 0.5%?

2.1 The Poisson Process and the Aggregate Loss Model

The Poisson process is the standard model for claim arrivals: events occur at a constant rate, independently of one another. We model the total payout as follows.

Definition 2.1.1 (The Aggregate Loss Process [1, §2.2])**.** *The aggregate loss process is defined as*

$$S(t) = \sum_{i=1}^{N(t)} X_i,$$

where $N(t) \sim \text{Poisson}(\lambda t)$ *is the number of claims up to time* t*, and* $X_1, X_2, \ldots$ *are i.i.d. claim sizes independent of* $N(t)$*.*

To make this concrete, consider an insurer whose claims arrive at rate λ per year. Each claim has a random payout X_i, drawn independently from the same distribution, and $S(t)$ is the total amount paid out after t years. The capital questions from the introduction are precisely questions about the tail of $S(t)$.

The choice of distribution for X_i reflects how claim costs are spread across the possible outcomes. Most claims are routine and modest, but occasionally a claim is extreme. How extreme these rare events can be is what the tail of the distribution captures, and different choices lead to fundamentally different behaviour. A light-tailed distribution, such as the exponential, models a world where rare claims are simply unlikely to be very large: the tail decays fast. A heavy-tailed distribution, such as the Pareto, models a

world where rare claims can be catastrophically large: the tail decays so slowly that a single extreme claim can dominate the entire aggregate.

Therefore knowing the mean and variance is not enough for capital calculations. These quantities describe what a typical year looks like, not an extreme one. The mean gives the expected total payout, and the variance yields how much it usually varies. Neither says anything about how likely it is that the total payout is twice or three times the average. That is precisely the question that matters for capital, and large deviation theory is the tool that answers it. For heavy-tailed distributions the situation is worse still, as the framework breaks down entirely, something made explicit in Section 2.3. Instead, in Section 2.4, we develop the Principle of a Single Big Jump, a framework specifically for heavy-tailed distributions. This framework is the tool to answer the Solvency II question introduced in the introduction. The theory in the following sections is developed for general claim sizes. Simulations and concrete computations use the exponential distribution for the light-tailed case and the Pareto distribution for the heavy-tailed case.

2.2 The Chernoff Bound and the Rate Function

To answer the capital questions from the introduction, we need to bound $P(S(t) \geq xt)$, the probability that aggregate losses exceed a threshold that grows with the time horizon, where t is the time horizon in years and xt is the total capital threshold. As discussed, the mean and variance cannot quantify this probability. The key tool is the moment generating function (MGF) and will allow us to turn this tail probability into a quantity we can derive and optimize over.

One of the standard ways of bounding objects in probability theory is by Markov's inequality,

$$P(X \geq a) \leq \frac{E[X]}{a} \quad \text{for any } a > 0 \text{ and } X \geq 0.$$

Applying Markov directly gives $P(S(t) \geq xt) \leq E[S(t)]/(xt) = \lambda E[X]/x$, a constant that does not depend on t. This is useless for the capital calculation: it gives the same bound whether the time horizon is one year or a hundred, and tells us nothing about how the probability changes as t grows. Therefore, we need a bound that shrinks as t grows. By applying Markov's inequality to $e^{\theta S(t)}$ we get a bound that decays exponentially in t. This bound depends on a free parameter θ. Optimizing over θ gives rise to the rate function $I(x)$, a central object in large deviation theory.

By the monotonicity of the exponential, for any $\theta > 0$, the events $\{S(t) \geq xt\}$, $\{e^{S(t)} \geq e^{xt}\}$ and $\{e^{\theta S(t)} \geq e^{\theta xt}\}$ are identical.

Applying Markov's with $X = e^{\theta S(t)}$ and $a = e^{\theta xt}$ we have

$$P(e^{\theta S(t)} \geq e^{\theta xt}) \leq E[e^{\theta S(t)}] \cdot e^{-\theta xt}.$$

This result is precisely what is referred to as the Chernoff bound (or exponential Markov).

Proposition 2.2.1 (Chernoff Bound [1, §2.4.1])**.**

$$P(S(t) \geq xt) \leq E[e^{\theta S(t)}] \cdot e^{-\theta xt} \quad \textit{for any } \theta > 0.$$

This bound is the starting point for the capital calculation. The goal is now to simplify the right-hand side into something we can optimize over θ.

To apply Proposition 2.2.1, we need to compute $E[e^{\theta S(t)}]$ explicitly. Once we have it, the bound takes a clean exponential form in t, and optimizing over θ gives the tightest possible bound. That optimal bound is $e^{-tI(x)}$, where $I(x)$ is the rate function to be defined below. This is what lets us answer the long-run question from the introduction.

The counting variable $N(t)$ is random, so we cannot directly evaluate $E[e^{\theta S(t)}]$. We therefore condition on $N(t) = n$:

$$E[e^{\theta S(t)} | N(t) = n] = E[e^{\theta \sum_{i=1}^{n} X_i}] = E[e^{\theta X_1} \cdot e^{\theta X_2} \cdot \cdots \cdot e^{\theta X_n}].$$

The X_i are i.i.d. and independent of $N(t)$, so the expectations factor:

$$E[e^{\theta X_1}] \cdot E[e^{\theta X_2}] \cdot \cdots \cdot E[e^{\theta X_n}] = E[e^{\theta X}]^n = \Psi_X(\theta)^n.$$

Because $\Psi_X(\theta)$ does not depend on $N(t)$, the law of total expectation, which states that $E[Y] = E[E[Y \mid Z]]$ for any random variables Y and Z, gives

$$E[e^{\theta S(t)}] = E\big[E[e^{\theta S(t)} \mid N(t)]\big] = E[\Psi_X(\theta)^{N(t)}].$$

For fixed θ, the value $\Psi_X(\theta) = E[e^{\theta X}]$ is a positive real number, not a random variable. The only randomness in $\Psi_X(\theta)^{N(t)}$ therefore lies in the exponent $N(t)$. Rewriting

$$\Psi_X(\theta)^{N(t)} = e^{N(t) \ln \Psi_X(\theta)},$$

we obtain an expression that matches the MGF of $N(t)$. For $N(t) \sim$ Poisson(λt) this MGF is $\Psi_{N(t)}(s) = e^{\lambda t(e^s - 1)}$, and substituting $s = \ln \Psi_X(\theta)$ gives

$$E[e^{\theta S(t)}] = E[\Psi_X(\theta)^{N(t)}] = e^{\lambda t(e^{\ln \Psi_X(\theta)} - 1)} = e^{\lambda t(\Psi_X(\theta) - 1)}.$$

The MGF of the aggregate loss process is:

Proposition 2.2.2 (The MGF of the Aggregate Loss Process)**.**

$$\Psi_{S(t)}(\theta) = e^{\lambda t(\Psi_X(\theta)-1)}.$$

Substituting $E[e^{\theta S(t)}] = e^{\lambda t(\Psi_X(\theta)-1)}$ into Proposition 2.2.1 gives:

$$P(S(t) \geq xt) \leq \frac{e^{\lambda t(\Psi_X(\theta)-1)}}{e^{\theta xt}} = e^{-t(\theta x - \lambda(\Psi_X(\theta)-1))}.$$

This shows that the bound decays exponentially in t provided the MGF $\Psi_X(\theta)$ is finite for some $\theta > 0$.

The derived bound $P(S(t) \geq xt) \leq e^{-t(\theta x - \lambda(\Psi_X(\theta)-1))}$ holds for every $\theta > 0$, so we are free to choose whichever θ gives the tightest bound. The bound $e^{-t(\theta x - \lambda(\Psi_X(\theta)-1))}$ is smallest when $\theta x - \lambda(\Psi_X(\theta) - 1)$ is largest. As $t > 0$, the choice of θ that maximizes the expression does not depend on t, and the optimization reduces to

$$\max_{\theta>0}(\theta x - \lambda(\Psi_X(\theta) - 1)).$$

This brings us to the rate function.

Definition 2.2.1 (The Rate Function of the Compound Poisson Process)**.** *The rate function of the compound Poisson process is defined as*

$$I(x) = \max_{\theta>0}(\theta x - \lambda(\Psi_X(\theta) - 1))$$

Note that $I(x)$ is well-defined only when $\Psi_X(\theta)$ is finite for some $\theta > 0$: if the MGF is infinite for all $\theta > 0$, the exponent equals $+\infty$ for every θ and the maximum is not meaningful. We will return to this point in Section 2.3.

The bound $P(S(t) \geq xt) \leq e^{-tI(x)}$ is in fact tight: a matching lower bound holds, establishing that $e^{-tI(x)}$ captures the correct exponential rate of decay. Proving the lower bound is technical and beyond the scope of this thesis, but the tightness is verified numerically in Section 3.1. For the full derivation see [1, § 2.4.2].

With this, the large deviation framework is complete. We began with the capital questions from the introduction: given a time horizon and an acceptable probability, how much capital is needed, and given a fixed reserve, how likely is it to be exceeded? Both reduce to knowing $I(x)$. Now that we have a formula for it, Section 3.1 computes $I(x)$ explicitly for exponential claim sizes and uses it to solve both problems directly.

2.3 Heavy-Tailed Distributions

Section 2.1 introduced the aggregate loss model and noted that the mean and variance may not adequately describe extreme outcomes for heavy-tailed claim distributions. This section makes the distinction between light-tailed and heavy-tailed distributions precise.

The central condition is whether $E[e^{\theta X}]$ is finite, which is the same as asking whether the integral

$$\int_0^\infty e^{\theta x} f(x)\,\mathrm{d}x$$

converges. The factor $e^{\theta x}$ diverges, so the answer depends on how quickly the density $f(x)$ tends to zero. For $X \sim \mathrm{Exp}(\mu)$, the integral converges for $\theta < \mu$. For a Pareto random variable (defined below), the density decays polynomially. For any $\theta > 0$ the exponential factor $e^{\theta x}$ eventually dominates and the integral diverges. This difference motivates the following definition.

Definition 2.3.1 (Heavy-Tailed Distribution)**.** *A distribution is defined to be heavy-tailed if and only if [2, Def. 2.2]*

$$\Psi_X(\theta) = E[e^{\theta X}] = \infty \quad \textit{for all } \theta > 0.$$

A distribution that is not heavy-tailed is called light-tailed [2, Def. 2.3].

While the definition is stated in terms of the moment generating function, Figure 2.1 illustrates why the two classes carry the names they do: light-tailed distributions have tails that quickly become negligible, while heavy-tailed distributions retain substantial probability mass far out in the tail.

We verify this analytically for the exponential distribution. For $X \sim \mathrm{Exp}(\mu)$, a direct calculation gives $E[e^{\theta X}] = \int_0^\infty \mu e^{x(\theta-\mu)}\,\mathrm{d}x$, which converges if and only if $\theta < \mu$. Thus the exponential distribution is light-tailed.

We now turn to the heavy-tailed case.

Definition 2.3.2 (The Pareto Distribution)**.** *A random variable X is said to be Pareto distributed with parameters $\alpha > 0$ and $\kappa > 0$ if it has probability density function*

$$f_X(x) = \frac{\alpha\kappa^\alpha}{(x+\kappa)^{\alpha+1}}, \quad x \geq 0,$$

with expected value $E[X] = \kappa/(\alpha - 1)$ for $\alpha > 1$, and $E[X] = \infty$ otherwise.

Proposition 2.3.1 (The Pareto Distribution is Heavy-Tailed)**.** *If $X \sim$ Pareto(α, κ) then $\Psi_X(\theta) = \infty$ for all $\theta > 0$.*

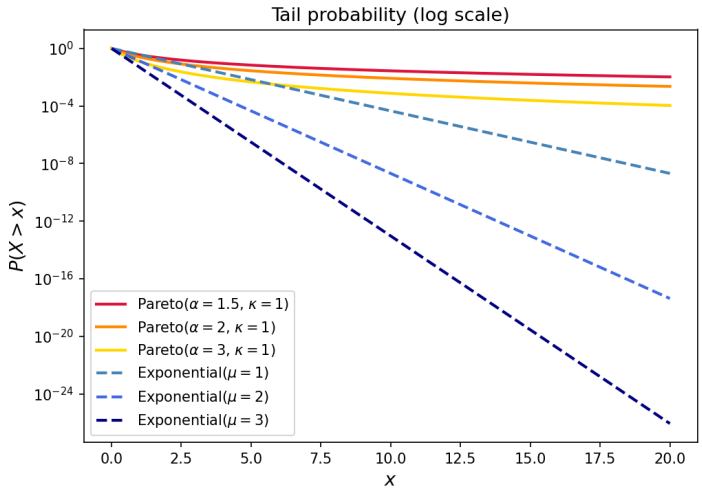

Figure 2.1: Tail probability $P(X > x)$ on a log scale for three Pareto distributions (solid lines) and three exponential distributions (dashed lines). The Pareto tails remain substantial across the full range while the exponential tails quickly become negligible, illustrating the fundamental difference between heavy-tailed and light-tailed behaviour.

Proof. We want to show that $E[e^{\theta X}] = \infty$ for all $\theta > 0$. Writing out the expectation using the Pareto density gives

$$E\left[e^{\theta X}\right] = \alpha\kappa^{\alpha} \int_0^{\infty} \frac{e^{\theta x}}{(x+\kappa)^{\alpha+1}}\, \mathrm{d}x.$$

Since $\alpha\kappa^{\alpha} > 0$, it suffices to show that the integral diverges. Here

$$f(x) = \frac{e^{\theta x}}{(x+\kappa)^{\alpha+1}},$$

and since exponentials dominate polynomials, $f(x) \to \infty$ as $x \to \infty$, so the integral over any tail is unbounded and the integral diverges. Hence $E[e^{\theta X}] = \infty$. □

The heavy-tail property is essential for what follows: subexponential distributions, introduced in Section 2.4, are by definition a subclass of heavy-tailed distributions. Establishing that the Pareto distribution is heavy-tailed is therefore a prerequisite for the next section, which culminates in a surprisingly clean result: the tail probability of the entire aggregate loss process reduces to that of a single claim.

2.4 Subexponential Distributions and the Principle of a Single Big Jump

This section develops the second asymptotic framework briefly introduced in the introduction. It is built around a subclass of heavy-tailed distributions called subexponential distributions. These distributions have a specific asymptotic property: when the sum of two independent subexponential random variables produces a large value, the probability is driven almost entirely by one of them being large. In the light-tailed case the opposite holds: both terms contribute, and the probability of one dominating vanishes. The same principle extends from two variables to n, and even to a random number of variables, yielding a remarkably simple two-factor expression for the tail probability of the entire aggregate loss process.

Throughout this section, X denotes a non-negative random variable representing an individual claim size, and $P(X > x)$ its tail probability. Unless stated otherwise, $X \sim \text{Pareto}(\alpha, \kappa)$.

Before we formally define subexponential distributions, a few properties of distributions on the positive real number line will be introduced. For any distribution on the positive real number line, the following bound is always true [2, eq. (2.7)]

$$\liminf_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X_1 > x)} \geq 2.$$

For heavy-tailed distributions the bound tightens further [2, Thm. 2.12],

$$\liminf_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X_1 > x)} = 2.$$

Finally, for heavy-tailed distributions for which the ratio converges, Foss states that the limit must equal 2 [2, p. 44], that is

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X_1 > x)} = 2.$$

This motivates the definition of subexponentiality, formally introduced below.

Definition 2.4.1 (Subexponential Distribution)**.** *A non-negative random variable X is called subexponential if, for X_1, X_2 i.i.d. copies of X, [2, Def. 3.1]*

$$P(X_1 + X_2 > x) \sim 2P(X > x) \quad \text{as } x \to \infty.$$

Here $f(x) \sim g(x)$ means $f(x)/g(x) \to 1$ as $x \to \infty$.

The definition is equivalent to two more intuitive formulations [2, p. 44]. The first states that given the sum is large, each variable is equally likely to be the one responsible:

$$P(X_1 > x \mid X_1 + X_2 > x) \to \frac{1}{2} \quad \text{as } x \to \infty.$$

The second states that the sum and its maximum have asymptotically equal tail probabilities:

$$P(X_1 + X_2 > x) \sim P(\max(X_1, X_2) > x) \quad \text{as } x \to \infty.$$

The first can be shown to be equivalent by noting that $X_2 \geq 0$ and $X_1 > x$ together imply $X_1 + X_2 > x$, so the joint probability simplifies to $P(X_1 > x)$. By the definition of conditional probability,

$$P(X_1 > x \mid X_1 + X_2 > x) = \frac{P(X_1 > x, X_1 + X_2 > x)}{P(X_1 + X_2 > x)} = \frac{P(X_1 > x)}{P(X_1 + X_2 > x)}.$$

By the definition of subexponentiality, this implies that

$$\frac{P(X_1 > x)}{P(X_1 + X_2 > x)} \sim \frac{P(X > x)}{2P(X > x)} = \frac{1}{2}.$$

For the second equivalence, we use the inclusion-exclusion principle. Recall that

$$\begin{aligned} P(\max(X_1, X_2) > x) &= P(X_1 > x \cup X_2 > x) \\ &= P(X_1 > x) + P(X_2 > x) - P(X_1 > x, X_2 > x) \\ &= 2P(X > x) - P(X > x)^2, \end{aligned}$$

where the last line uses the fact that X_i are i.i.d. Asymptotically the square term decays faster than the linear term and is therefore negligible. Thus, using the definition of subexponentiality,

$$P(\max(X_1, X_2) > x) \sim 2P(X > x) \sim P(X_1 + X_2 > x).$$

Both formulations capture the same idea: when the sum is large, one term is responsible. The first shows this through symmetry: each term is equally likely to be the one causing it. The second shows it directly: the sum exceeds x when and only when one of the terms does. This is the so-called *Principle of a Single Big Jump* for sums of subexponentially distributed random variables [2, p. 44].

To see what this definition captures, we simulate $n = 500{,}000$ i.i.d. pairs (X_1, X_2) and retain only those for which $X_1 + X_2 > x$. For each retained

pair we ask: how did the sum exceed x? Either one variable alone exceeded x (a single big jump), or neither did but together they do (cooperation). Figure 2.2 plots every retained pair for Pareto($\alpha = 2, \kappa = 1$) at $x = 50$ and Exponential($\mu = 1$) at $x = 5$, [1] coloured by mechanism. For Pareto, cooperation accounts for only 4% of cases. For the exponential, cooperation accounts for 66%. The contrast illustrates the fundamental difference between heavy-tailed and light-tailed behaviour.

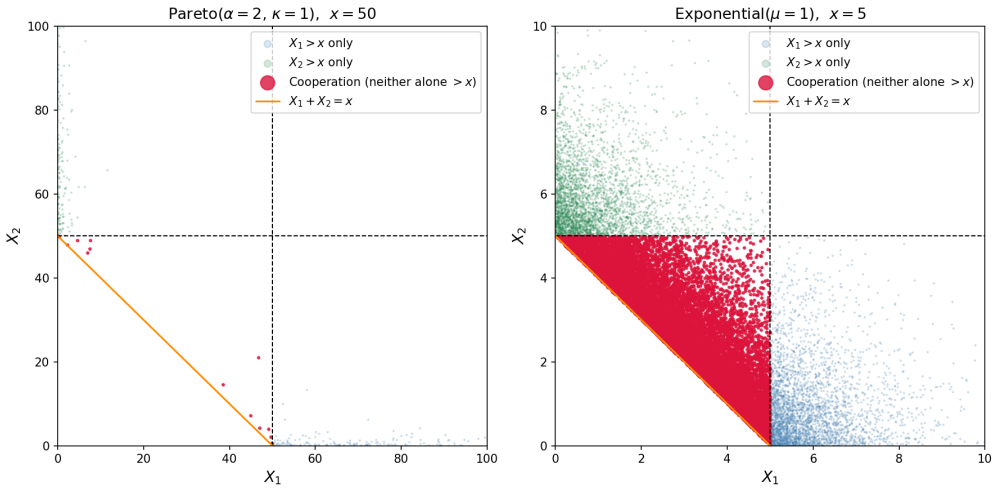

Figure 2.2: Simulated pairs (X_1, X_2) conditioned on $X_1 + X_2 > x$, coloured by mechanism. Blue: $X_1 > x$ alone, green: $X_2 > x$ alone, red: cooperation (neither alone exceeds x). Left: Pareto($\alpha = 2, \kappa = 1$) at $x = 50$, cooperation accounts for only 4%. Right: Exponential($\mu = 1$) at $x = 5$, cooperation accounts for 66%. Each panel uses $n = 500{,}000$ simulated pairs (before conditioning on $X_1 + X_2 > x$).

The cooperation fraction at threshold x is the probability that neither X_1 nor X_2 alone caused the sum to exceed x, given that the sum did. We call this cooperation because both terms had to contribute: neither was large enough on its own.

$$\begin{aligned} \text{cooperation fraction at } x &= P(X_1 \le x,\ X_2 \le x \mid X_1 + X_2 > x) \\ &= \frac{P(X_1 \le x,\ X_2 \le x,\ X_1 + X_2 > x)}{P(X_1 + X_2 > x)}. \end{aligned}$$

[1] The thresholds differ for practical reasons: for $X \sim \text{Exp}(\mu = 1)$, the sum $X_1 + X_2$ follows a Gamma$(2, \mu)$ distribution with tail $P(X_1 + X_2 > x) = (1 + \mu x)e^{-\mu x}$, giving $P(X_1 + X_2 > 50) = 51e^{-50} \approx 9.8 \times 10^{-21}$. Illustrating the same point at $x = 50$ would require an impractical number of simulations. Instead, $x = 5$ gives $P(X_1 + X_2 > 5) \approx 4\%$ and yields sufficient samples. The principle is the same regardless of the choice of x.

Figure 2.3 shows how this fraction evolves as x grows. For Pareto claims the fraction decreases toward zero: as the threshold rises, one single term increasingly explains the entire sum. For exponential claims the opposite holds: cooperation becomes the only mechanism. The scatter plot in Figure 2.2 captures one moment in this picture. This figure confirms that the behaviour persists as $x \to \infty$.

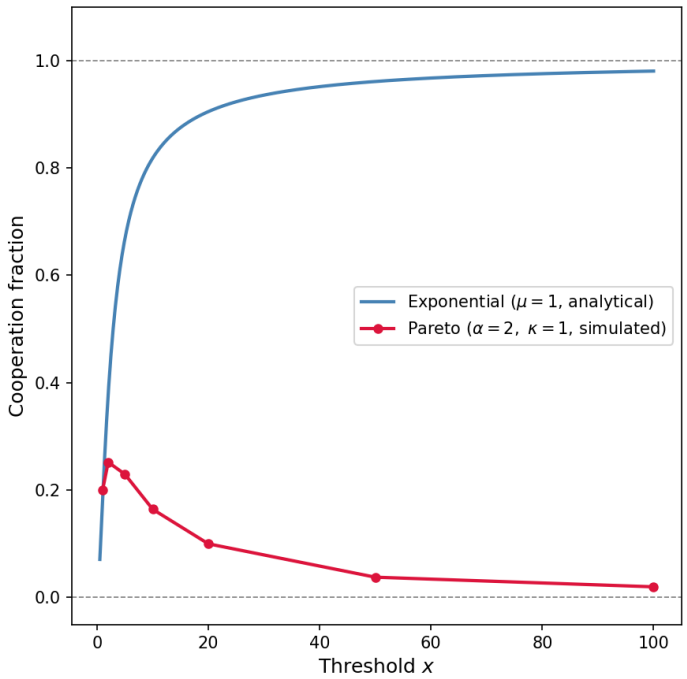

Figure 2.3: Cooperation fraction as a function of threshold x for Pareto($\alpha = 2, \kappa = 1$) (simulated, $n = 10{,}000{,}000$) and Exponential($\mu = 1$) (analytical). The two curves diverge in opposite directions: the Pareto fraction converges to zero while the exponential fraction converges to one.

The figures already suggest that the single big jump property fails for light-tailed distributions: cooperation dominates for exponential claims. This is not a coincidence. Subexponentiality requires that one term drives the sum, which is only possible when the tail is heavy. The following proposition confirms this formally.

Proposition 2.4.1 (Subexponential Implies Heavy-Tailed [2, Lem. 3.2])**.** *If X is subexponential then X is heavy-tailed.*

As we now show, the Pareto distribution belongs to this class.

To show that the Pareto distribution is subexponential, we verify two properties of its tail. The first, long-tailedness, captures the idea that the tail

decays so slowly that shifting the threshold by any fixed amount y has no asymptotic effect.

Definition 2.4.2 (Long-Tailed [2, Def. 2.21])**.** *A non-negative random variable X is called* long-tailed *if*

$$\frac{P(X > x + y)}{P(X > x)} \to 1 \quad \text{as } x \to \infty, \quad \text{for every fixed } y > 0.$$

Lemma 2.4.1 (The Pareto Distribution is Long-Tailed)**.** *If $X \sim \text{Pareto}(\alpha, \kappa)$ with $\alpha > 0$ and $\kappa > 0$, then X is long-tailed.*

Proof. Fix any $y > 0$. For $X \sim \text{Pareto}(\alpha, \kappa)$ the long-tailed ratio is:

$$\frac{P(X > x + y)}{P(X > x)} = \frac{\left(\frac{\kappa}{x + y + \kappa}\right)^{\alpha}}{\left(\frac{\kappa}{x + \kappa}\right)^{\alpha}} = \left(\frac{x + \kappa}{x + y + \kappa}\right)^{\alpha} = \left(1 + \frac{y}{x + \kappa}\right)^{-\alpha}.$$

As $x \to \infty$, the fraction $y/(x + \kappa) \to 0$, so the entire expression tends to 1. Thus X is long-tailed. □

The second condition, dominated-varying, ensures the tail does not thin out too quickly under scaling. Even if you double the threshold the tail is still at least a fixed fraction of what it was originally.

Definition 2.4.3 (Dominated-Varying [2, p. 11])**.** *A non-negative random variable X is called* dominated-varying *if there exists a constant $c > 0$ such that*

$$P(X > 2x) \geq c\, P(X > x) \quad \text{for all } x \geq 0.$$

Lemma 2.4.2 (The Pareto Distribution is Dominated-varying)**.** *If $X \sim \text{Pareto}(\alpha, \kappa)$ with $\alpha > 0$ and $\kappa > 0$, then X is dominated-varying.*

Proof. For $X \sim \text{Pareto}(\alpha, \kappa)$, compute the ratio $P(X > 2x)/P(X > x)$ explicitly. Since $P(X > x) = (\kappa/(x + \kappa))^{\alpha}$, the κ^{α} terms cancel and we obtain:

$$\frac{P(X > 2x)}{P(X > x)} = \frac{\left(\frac{\kappa}{2x + \kappa}\right)^{\alpha}}{\left(\frac{\kappa}{x + \kappa}\right)^{\alpha}} = \left(\frac{x + \kappa}{2x + \kappa}\right)^{\alpha}.$$

We claim the base satisfies $(x + \kappa)/(2x + \kappa) \geq 1/2$ for all $x \geq 0$ and $\kappa > 0$. Cross-multiplying, this is equivalent to $2(x + \kappa) \geq 2x + \kappa$, i.e. $\kappa \geq 0$, which always holds since $\kappa > 0$ by assumption. Since $\alpha > 0$, the map $t \mapsto t^{\alpha}$

is increasing on $t > 0$, so the inequality is preserved when raising to the power α:

$$\frac{P(X > 2x)}{P(X > x)} = \left(\frac{x+\kappa}{2x+\kappa}\right)^{\alpha} \geq \left(\frac{1}{2}\right)^{\alpha} = c > 0.$$

Since this holds for all $x \geq 0$, X is dominated-varying. □

Figure 2.4 illustrates both conditions for Pareto($\alpha = 2, \kappa = 1$) and Exponential($\mu = 1$): the Pareto ratios converge to their respective positive limits while the exponential ratios do not.

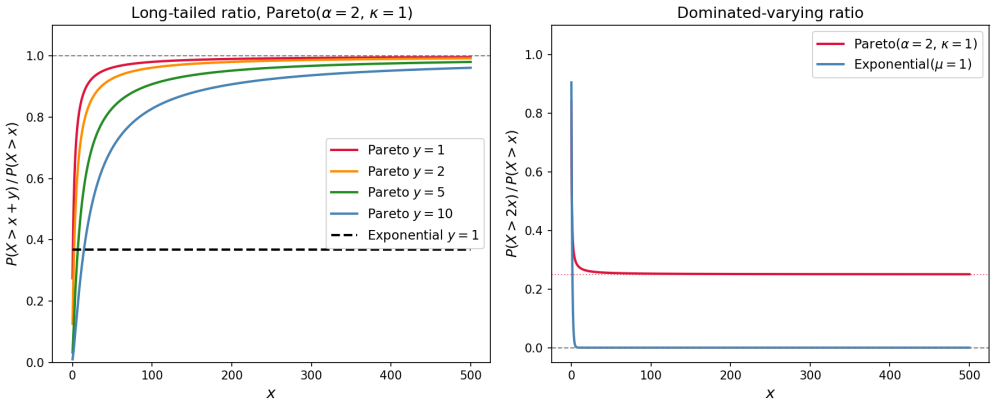

Figure 2.4: Left: long-tailed ratio $P(X > x+y)/P(X > x)$ for Pareto($\alpha = 2, \kappa = 1$) with $y = 1, 2, 5, 10$ and Exponential($\mu = 1$) with $y = 1$. All Pareto curves converge to 1 regardless of y, confirming long-tailedness: no fixed shift changes the tail asymptotically. The exponential ratio is the constant $e^{-1} \approx 0.368$, showing it is not long-tailed. Right: dominated-varying ratio $P(X > 2x)/P(X > x)$. The Pareto ratio stabilises at $(1/2)^{\alpha} = 0.25 > 0$, confirming that the tail remains a fixed fraction even when the threshold doubles. The exponential ratio decays to 0, failing this condition.

Putting these two criteria together, a tail that is insensitive to fixed shifts and does not thin out too quickly under scaling is sufficient to guarantee subexponentiality, which is formalised in the following theorem.

Theorem 2.4.1 ([2, Thm. 3.29])**.** *Let X be a non-negative random variable that is long-tailed and dominated-varying. Then X is subexponential.*

The proof is technical and falls outside the scope of this thesis.

Proposition 2.4.2 (The Pareto Distribution is Subexponential)**.** *If $X \sim$ Pareto(α, κ) with $\alpha > 1$, then X is subexponential.*

Proof. By Lemma 2.4.1, X is long-tailed. By Lemma 2.4.2, X is dominated-varying. Theorem 2.4.1 therefore applies, and X is subexponential. □

The concepts of heavy-tailedness and subexponentiality are not equivalent. There exist heavy-tailed distributions that are not subexponential. In practice, however, all commonly used heavy-tailed distributions are subexponential, including the Pareto, Burr, Cauchy, lognormal, and Weibull (with shape parameter $\alpha < 1$) distributions [2, p. 54].

The Principle of a Single Big Jump, as stated so far, applies to a fixed finite number of terms. In the aggregate loss model, however, the number of claims $N(t)$ is itself random. The following theorem extends the result to this setting: when claim sizes are subexponential and the number of claims is sufficiently well-behaved, the tail probability of the total payout is proportional to the expected number of claims times the tail of a single claim.

Theorem 2.4.2 ([2, Thm. 3.37])**.** *Let $X_1, X_2, \ldots$ be i.i.d. non-negative subexponential random variables, and let τ be a counting variable independent of $X_1, X_2, \ldots$ Suppose that $E[\tau] < \infty$ and $E[(1+\delta)^\tau] < \infty$ for some $\delta > 0$. Then*

$$P\left(\sum_{i=1}^{\tau} X_i > x\right) \sim E[\tau] \cdot P(X > x) \quad \text{as } x \to \infty.$$

The theorem has three conditions. The first requires subexponentiality of the claim sizes. The second requires the counting variable to have finite mean, which ensures the expected number of terms is finite. The third condition, $E[(1+\delta)^\tau] < \infty$, is less immediate. Setting $c = 1 + \delta$ gives

$$E[(1+\delta)^\tau] = E[c^\tau] = E[e^{\tau \ln c}] = \Psi_\tau(\ln c) < \infty,$$

which is exactly the condition that the MGF of τ is finite for some positive argument: the light-tailed condition. A counting variable failing this could itself produce large aggregate losses through volume alone, breaking the single big jump principle.

Proposition 2.4.3 (Principle of a Single Big Jump for Compound Poisson)**.** *Let $S(t) = \sum_{i=1}^{N(t)} X_i$ be the compound Poisson aggregate loss process with $X_i \sim$ Pareto(α, κ), $\alpha > 1$, i.i.d. and independent of $N(t) \sim$ Poisson(λt). Then*

$$P(S(t) > x) \sim \lambda t \cdot P(X > x) \quad \text{as } x \to \infty.$$

Proof. We verify the conditions of Theorem 2.4.2 with $\tau = N(t)$.

- $E[N(t)] = \lambda t < \infty$.
- For any $\delta > 0$, using that the MGF of Poisson(λt) is $\Psi_{N(t)}(s) = e^{\lambda t(e^s - 1)}$:

$$\begin{aligned} E[(1+\delta)^{N(t)}] = E[e^{N(t)\ln(1+\delta)}] &= \Psi_{N(t)}(\ln(1+\delta)) \\ &= e^{\lambda t(e^{\ln(1+\delta)} - 1)} = e^{\lambda t \delta} < \infty. \end{aligned}$$

- By Proposition 2.4.2, X is subexponential.

All conditions are satisfied. Thus the theorem is applicable for the aggregate loss model. □

With this result, the Solvency II question from the introduction has a concrete answer for Pareto claim sizes. Setting $t = 1$ and requiring $P(S(1) > x) \leq 0.005$, the approximation $P(S(1) > x) \approx \lambda \cdot P(X > x)$ reduces the problem to solving $\lambda \cdot P(X > x) = 0.005$ for x. The entire aggregate loss process, involving a random number of random claims, is captured by a simple two-factor expression: the tail probability of one claim scaled by the expected number of claims.

Chapter 3

Simulations and Illustrations

This chapter illustrates the two asymptotic frameworks developed in Chapter 2 through simulation. Section 3.1 derives the rate function for exponential claim sizes analytically and confirms the LDT result numerically. Section 3.2 shows that the same analysis fails for Pareto claim sizes, and uses the Principle of a Single Big Jump (PSBJ) result to explain why. Section 3.3 demonstrates the PSBJ approximation for Pareto claim sizes numerically. Finally, Section 3.4 shows that PSBJ does not hold for exponential claim sizes, completing the picture.

3.1 Large Deviation Rate Function for Exponential Claim Sizes

Before plotting anything we must solve the optimisation problem for the rate function when $X \sim \text{Exp}(\mu)$.

Proposition 3.1.1. *The rate function for the exponential distribution is*

$$I(x) = (\sqrt{\mu x} - \sqrt{\lambda})^2.$$

Proof. We solve $\max_{\theta>0}\{\theta x - \lambda(\Psi_X(\theta) - 1)\}$ by setting the derivative to zero. The MGF of $\text{Exp}(\mu)$ is $\Psi_X(\theta) = \mu/(\mu - \theta)$ for $\theta < \mu$, so

$$\frac{\partial}{\partial\theta}\left(\theta x - \lambda\frac{\mu}{\mu-\theta} + \lambda\right) = 0 \quad \Rightarrow \quad x = \frac{\lambda\mu}{(\mu-\theta)^2}$$

$$(\mu-\theta)^2 = \frac{\lambda\mu}{x} \quad \Rightarrow \quad \theta^* = \mu - \sqrt{\frac{\lambda\mu}{x}}$$

Substituting θ^* into the rate function gives

$$\begin{aligned} I(x) &= \left(\mu - \sqrt{\frac{\lambda\mu}{x}}\right) x - \lambda\left(\frac{\sqrt{x}\mu}{\sqrt{\lambda\mu}} - 1\right) \\ &= \mu x - \sqrt{\lambda\mu x} - \sqrt{\lambda\mu x} + \lambda \\ &= \left(\sqrt{\mu x} - \sqrt{\lambda}\right)^2. \end{aligned}$$

□

Next we aim to illustrate that the result from Section 2.2 holds. For this we will choose the parameters $\lambda = 1$, $\mu = 2$ and $x = 2$. This is motivated by the fact that we will have an expected value of $1/2$, thus with $x = 2$ we have a large deviation, 4 times larger than the mean, but not so large that we have issues simulating. These parameters are also chosen because they yield a particularly clean rate function. Specifically we have

$$I(x = 2) = \left(\sqrt{2 \cdot 2} - \sqrt{1}\right)^2 = 1.$$

Section 2.2 states that $P(S(t) \geq xt) \approx e^{-tI(x)}$, or equivalently,

$$\lim_{t\to\infty} \frac{1}{t} \ln P(S(t) \geq xt) = -I(x).$$

Rather than plotting $P(S(t) \geq xt)$ directly, which decays toward zero so rapidly that convergence is difficult to assess visually, we plot the log-scaled quantity $(1/t) \ln P(S(t) \geq xt)$, which converges to the constant $-I(x)$. Figure 3.1 provides numerical support for this.

We estimate $P(S(t) \geq xt)$ by Monte Carlo simulation. For each t, we draw the claim count $N(t) \sim \text{Poisson}(\lambda t)$ and use the fact that the sum of n i.i.d. $\text{Exp}(\mu)$ random variables is $\text{Gamma}(n, 1/\mu)$ distributed. Conditioning on $N(t) = n$, we have $S(t) \sim \text{Gamma}(n, 1/\mu)$.

The quantity $(1/t) \ln P(S(t) \geq xt)$ indicates convergence toward -1 as t grows, in agreement with $I(x) = 1$ computed above.

Remark 3.1.1. *For large t, the probability $P(S(t) \geq xt)$ becomes so small that none of the 5×10^6 simulations produce the event. For example, at $t = 15$ the approximation gives $P(S(15) \geq 30) \approx e^{-15} \approx 3 \times 10^{-7}$, meaning the event is expected to occur only 1.5 times among 5×10^6 simulations. Reliable estimation becomes impossible, which is why Figure 3.1 does not extend further.*

Finally, let us apply the rate function to the two capital problems from the introduction. Under the parameters $\lambda = 1$ claim per year and $\mu = 2$, each

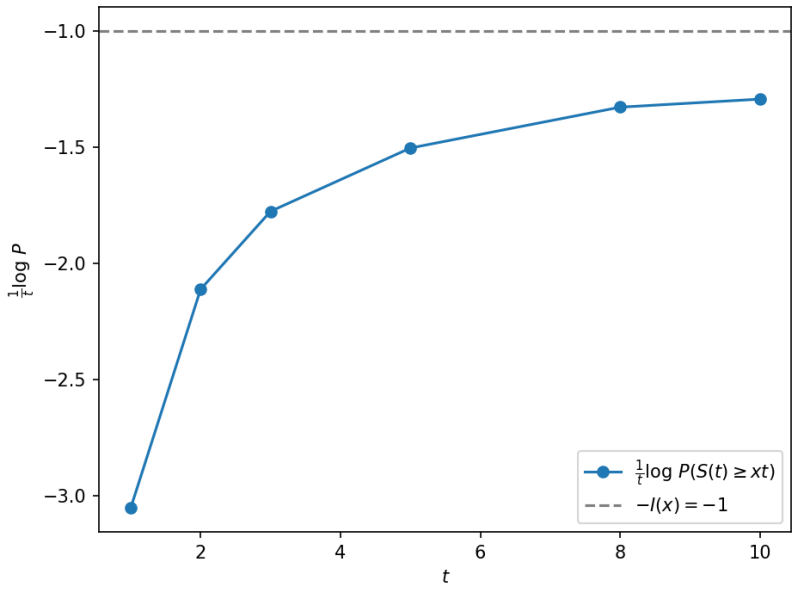

Figure 3.1: Simulated decay rate $(1/t)\ln P(S(t) \geq xt)$ converging to $-I(x) = -1$ (dashed line) as $t \to \infty$. Parameters: $\lambda = 1$, $\mu = 2$, $x = 2$, 5×10^6 simulations per t.

claim has expected size $E[X] = 1/\mu = 1/2$, so the expected total payout over five years is $\lambda \cdot E[X] \cdot 5 = 2.5$ (in the same units as the claim sizes).

First, suppose the insurer requires that the probability of aggregate losses exceeding the capital reserve over five years is at most $p = 0.01$. That is, find x such that $P(S(5) \geq 5x) \leq 0.01$, where xt is the total capital reserve. Setting $e^{-5I(x)} = 0.01$ gives

$$I(x) = \frac{\ln 100}{5} \approx 0.921.$$

Solving $(\sqrt{2x} - 1)^2 = 0.921$ for x yields

$$x = \frac{(1 + \sqrt{0.921})^2}{2} \approx 1.92,$$

so the required capital reserve is $xt = 5 \times 1.92 \approx 9.60$. This is roughly four times the expected total payout of 2.5, reflecting the cost of the $p = 0.01$ requirement: keeping the exceedance probability this low demands holding a substantial buffer above what the insurer would expect to pay in a typical scenario.

Now suppose instead that the insurer has set aside a fixed capital reserve $xt = 10$, so $x = 2$. Since $I(2) = 1$ was computed above, the probability that aggregate losses exceed the reserve over five years is

$$P(S(5) \geq 10) \approx e^{-5 \cdot 1} = e^{-5} \approx 0.0067,$$

approximately 0.67%.

The parameters are chosen for mathematical convenience rather than realism.

3.2 Failure of the Large Deviation Principle for Pareto Claim Sizes

We now repeat the analysis of Section 3.1 with Pareto claim sizes, with the goal of showing that the LDT framework breaks down in the heavy-tailed case. We keep $\lambda = 1$ and $x = 2$, and choose $\alpha = 2$ and $\kappa = 1$. The choice $\alpha = 2$ satisfies $\alpha > 1$, giving a finite mean, while keeping the tail decay slow.

The tail probability follows directly from the density in Definition 2.3.2:

$$\begin{aligned} P(X > x) &= \int_x^\infty \frac{\alpha\kappa^\alpha}{(u+\kappa)^{\alpha+1}}\,\mathrm{d}u = \left[\frac{-\kappa^\alpha}{(u+\kappa)^\alpha}\right]_x^\infty \\ &= \left(\frac{\kappa}{x+\kappa}\right)^\alpha \approx x^{-\alpha} \quad \text{for large } x. \end{aligned}$$

We now repeat the simulation procedure of Section 3.1 with Pareto claim sizes and plot the same quantity $(1/t)\ln P(S(t) \geq xt)$.

Figure 3.2 shows that the quantity converges to 0, unlike the exponential case where we had convergence to -1. As established in Section 2.3, the Pareto MGF is infinite for all $\theta > 0$, and therefore the rate function cannot be computed. A natural attempt to still characterize $P(S(t) \geq xt)$ is to apply the PSBJ result from Section 2.4:

$$P(S(t) \geq xt) \approx \lambda t \cdot P(X > xt) = \lambda t \cdot \left(\frac{\kappa}{xt+\kappa}\right)^\alpha \approx \frac{\lambda\kappa^\alpha}{x^\alpha} \cdot t^{1-\alpha}$$

With $\alpha = 2$ we have

$$P(S(t) \geq xt) \sim \frac{C}{t} \quad \text{where } C \text{ is some constant.}$$

In other words, polynomial decay. Taking the log and dividing by t gives

$$\frac{1}{t}\ln P(S(t) \geq xt) \approx \frac{1}{t}\ln\frac{C}{t} \to 0 \quad \text{as } t \to \infty$$

In the exponential case the same quantity converged to -1, reflecting exponential decay. Here the limit is 0 because the aggregate loss process $S(t)$ inherits the polynomial tail of its Pareto claims through the PSBJ approximation, as shown above. Since the LDT framework requires an exponential decay rate, heavy-tailed claim sizes cause the LDT framework to fail.

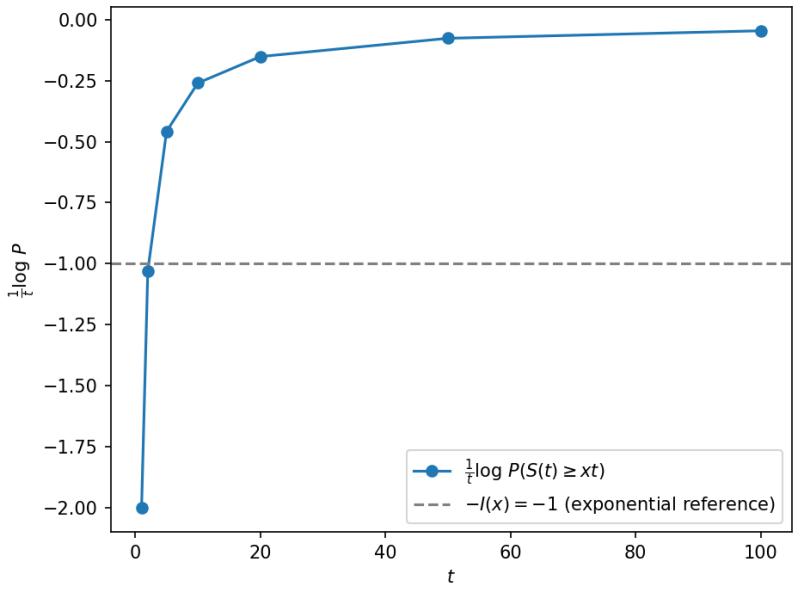

Figure 3.2: Simulated decay rate $(1/t)\ln P(S(t) \geq xt)$ for Pareto($\alpha = 2, \kappa = 1$) claim sizes. Unlike the exponential case, the quantity does not stabilize. It converges to 0 rather than a finite negative value, confirming that no useful rate function $I(x)$ exists. The dashed line at -1 is shown as a reference from Section 3.1. Parameters: $\lambda = 1$, $x = 2$, 10^6 simulations per t.

3.3 Subexponential Asymptotics for Pareto Claim Sizes

Section 2.4 established that for Pareto(α, κ) claim sizes the aggregate loss process satisfies

$$P(S(t) > x) \sim \lambda t \cdot P(X > x) \quad \text{as } x \to \infty.$$

This section verifies the approximation numerically and applies it to the Solvency II capital question from the introduction.

Unlike the exponential case in Section 3.1, the tail probability $P(S(t) > x)$ does not admit a closed-form expression for Pareto claim sizes. Conditioning on $N(t) = n$ reduces the problem to computing the tail $P(X_1 + \cdots + X_n > x)$. Already for $n = 2$, the convolution does not reduce to any standard distribution. For $n = 3$, this result must then be convolved with another Pareto density, producing an even more complex expression. The complexity grows dramatically at each step, and the expressions quickly become analytically unworkable. For the exponential distribution no such problem arises: a sum of n independent Exp(μ) variables follows

a Gamma$(n, 1/\mu)$ distribution, whose CDF is known explicitly, and this is the property used in Section 3.1. For Pareto claims no such expression exists and we must instead draw the individual claim sizes and sum them.

Monte Carlo simulation estimates $P(S(t) > x)$ by running a large number of independent realisations of $S(t)$ and computing the fraction that exceed x. With $n = 1{,}000{,}000{,}000$ simulations, the estimate is reliable for moderate x, where many realisations exceed the threshold. For large x, however, $P(S(t) > x)$ becomes very small, meaning only a handful of simulations produce $S(t) > x$. Since the estimated probability is simply the count divided by n, a small count means large relative error, causing the visible noise in the ratio plot for large x.

Figure 3.3 compares the simulated tail probability with the PSBJ approximation $\lambda t \cdot P(X > x)$, using Pareto$(\alpha = 2, \kappa = 1)$ claim sizes with $\lambda = 1$ and $t = 10$. The left panel plots both quantities against x. For small x there is a visible discrepancy, which is expected: the PSBJ result is asymptotic and makes no claim about finite x. As x grows the two curves approach each other, and for large x the approximation tracks the simulated probability closely. The right panel makes this convergence precise by plotting the ratio $P(S(t) > x)/(\lambda t \cdot P(X > x))$, which converges toward 1 as $x \to \infty$, confirming the asymptotic statement numerically. The ratio approaches but does not reach 1 within the plotted range, consistent with the fact that convergence is asymptotic.

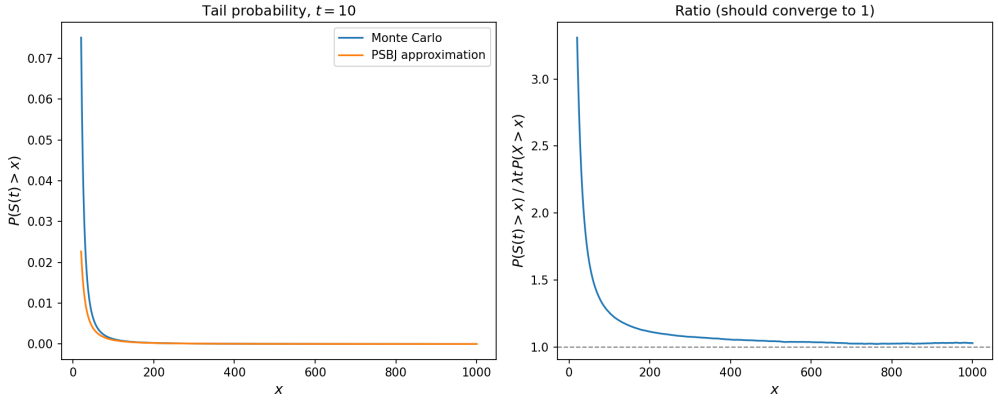

Figure 3.3: Left: simulated tail probability $P(S(t) > x)$ against the PSBJ approximation $\lambda t \cdot P(X > x)$ for Pareto$(\alpha = 2, \kappa = 1)$ claim sizes with $\lambda = 1$, $t = 10$, $n = 1{,}000{,}000{,}000$ simulations. Right: ratio of simulated probability to PSBJ approximation, converging toward 1 as $x \to \infty$.

Finally, we apply the PSBJ approximation to the Solvency II capital question from the introduction. Recall the parameters $\lambda = 1$ claim per year

and Pareto($\alpha = 2, \kappa = 1$) claim sizes, so that $E[X] = \kappa/(\alpha - 1) = 1$ and the expected total payout over one year is $\lambda \cdot E[X] \cdot 1 = 1$.

First, suppose the insurer requires that the probability of aggregate losses exceeding the capital reserve over one year is at most $p = 0.005$, the Solvency II standard. By the PSBJ approximation, $P(S(1) > x) \approx \lambda \cdot P(X > x) = \left(\frac{1}{x+1}\right)^2$. Setting this equal to p and solving for x:

$$\left(\frac{1}{x+1}\right)^2 = 0.005 \implies x = \frac{1}{\sqrt{0.005}} - 1 = \sqrt{200} - 1 \approx 13.14.$$

The required capital reserve is thus $x \approx 13.14$, roughly thirteen times the expected annual payout of 1, reflecting the cost of the heavy tail: even a moderate claim rate of $\lambda = 1$ per year requires holding capital far above the expected loss to meet a 0.5% exceedance probability.

Now suppose instead that the available capital reserve is $x = 10$. The probability that aggregate losses exceed it is

$$P(S(1) > 10) \approx \left(\frac{1}{11}\right)^2 = \frac{1}{121} \approx 0.0083,$$

approximately 0.83%, which exceeds the permitted 0.5%: a reserve of $x = 10$ is insufficient under Solvency II.

The parameters are chosen to match the simulation in this section rather than insurance realism. The method applies directly once realistic values for λ, α, κ, and target probability p are substituted.

3.4 Failure of the Principle of a Single Big Jump for Exponential Claim Sizes

We now repeat the analysis of Section 3.3 with exponential claim sizes, replacing Pareto($\alpha = 2, \kappa = 1$) claims with Exp($\mu = 1$) claims. The parameters $\lambda = 1$ and $t = 10$ are kept the same, giving the same expected total payout $\lambda t \cdot E[X] = 10$ in both cases, so any difference in behaviour is due to the tail, not the scale.

Figure 3.4 shows the ratio $P(S(t) > x)/(\lambda t \cdot P(X > x))$ alongside the two tail probabilities. Unlike the Pareto case in Section 3.3, the ratio does not converge to 1: it diverges, growing without bound as x increases. The PSBJ approximation becomes increasingly inaccurate, systematically underestimating the true tail probability.

Two features of the left panel are worth noting. First, the PSBJ approximation $\lambda t \cdot e^{-\mu x}$ appears as a straight line on the log scale, since

$\ln(\lambda t \cdot e^{-\mu x}) = \ln(\lambda t) - \mu x$ is linear in x. This is precisely the exponential decay that characterises light-tailed distributions. Second, the true tail $P(S(t) > x)$ decays noticeably slower than the approximation, reflecting the fact that aggregate losses are driven by cooperation between many claims rather than by a single large one. This is consistent with Figure 2.3, which showed the cooperation fraction converging to 1 for exponential claims.

The correct framework for exponential claim sizes is therefore not the PSBJ but the large deviation framework from Section 3.1: the true tail decays at the rate $e^{-tI(x/t)}$, where I is the rate function derived there.

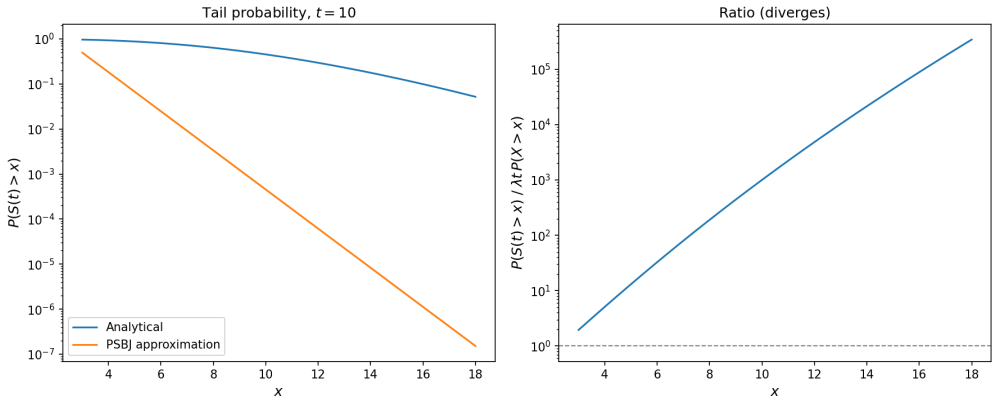

Figure 3.4: Left: tail probability $P(S(t) > x)$ and PSBJ approximation $\lambda t \cdot P(X > x)$ for $\text{Exp}(\mu = 1)$ claim sizes with $\lambda = 1$, $t = 10$, $n = 5 \times 10^6$ simulations. Both axes are log-scaled, the PSBJ approximation is a straight line since the exponential tail decays at a constant rate. Right: ratio of true tail to PSBJ approximation, diverging as x increases, confirming that the single big jump mechanism does not hold for light-tailed claims.

Chapter 4

Discussion

4.1 Distribution Choice in Practice

The frameworks developed in this thesis require different classes of claim size distributions: the large deviation framework requires a finite moment generating function, while the subexponential framework requires an infinite moment generating function for all θ. In practice, the choice of distribution is done by picking the one that best fits reality. However, concessions have to be made as real data rarely matches theory one-to-one. This section will briefly discuss how the process of choosing distribution is done in practice.

Before discussing the pipeline we will mention a brief practical fact: named distributions are named for a reason. The named distributions are the ones that are commonly used in practice. It should be noted that this does not inherently mean that reality can be encapsulated only by named distributions, but rather that named distributions are good enough, and probably more importantly: are easy to work with.

The process of distribution selection generally follows from first plotting the data as either a histogram or empirical distribution function, followed by trying to fit known densities or mass functions where the fit can be either confirmed visually or through different statistical tests, for example, hypothesis test, Chi-square goodness-of-fit test or likelihood ratio test. The parameters can then be found through statistical inference with for example maximum likelihood estimation. Finally, the results are compared and the overall best fit is chosen. For a systematic treatment of this see [5, Ch. 13].

This is not the only thing to keep in mind when choosing models as regulatory frameworks like Solvency II dictate specific requirements for insurance modelling. For example, modelling aggregate losses requires that the en-

tire portfolio follows a lognormal distribution [3, Annex XVII, Sect. B, point (2)(g)(iii)]. The lognormal distribution is subexponential, like the Pareto distribution, but does not yield equally convenient results. Since this is a portfolio level constraint, motivating the choice of the counting variable and claim size distribution would need to follow a different procedure entirely which does not match the goal of the thesis. Furthermore, the theorem used to show that the Pareto distribution is subexponential does not work for the lognormal distribution, which instead needs a different approach. Specifically the lognormal distribution fails the condition of dominated-varying and instead needs a proof involving the hazard function $R(x) = -\ln P(X > x)$[2, Def. 2.5]. For the specific theorem see [2, Thm. 3.30].

The motivation for using the exponential distribution and the Pareto distribution is rooted in intuition. As the large deviation framework requires at least exponential decay, choosing a distribution which decays precisely exponentially aids the intuition. Furthermore, the rate function for the exponential distribution yields a clean expression making derivation easier to follow. Similar arguments hold for the Pareto distribution, polynomial decay and a relatively simple proof path make the treatment of the theory as intuitive as possible.

While choosing the most realistic distributions would be beneficial for the connection to practical application, we chose distributions that instead focus on intuition rather than practical applicability. Regardless, the main results hold all the same.

4.2 Limitations of the Large Deviation Framework

In this section we will briefly cover some of the limitations of the large deviation framework. The first and most obvious one is that the moment generating function has to be finite for some $\theta > 0$. In practice, there are situations where a heavy-tailed distribution would be the best fit. Some natural examples are gambling modelling and insurance modelling where claim sizes can be so large that integrating over the full support yields a substantial contribution from the tail. This is something that light-tailed distributions cannot capture as the decay rate of the tail is too fast.

A second limitation concerns the rate function $I(x)$. It is characterised by the exponential rate at which $P(S(t) \geq xt)$ decays as t grows. This does not yield a precise value of the probability, but rather a trend indicator as the result is approximative. Therefore answering questions regarding specific

time horizons is never precise, unlike the subexponential framework which produces required capital or a probability given a capital for a specific year. This extends further still, the result is asymptotic, meaning that looking at a short time span, the results have most likely not yet converged, yielding results that might be misleading.

4.3 Limitations of the Subexponential Framework

Like the large deviation framework, the subexponential framework suffers from the same limitation of claim size distribution. Light-tailed distributions, just like heavy-tailed distributions, also arise from the modelling process. Furthermore, Theorem 2.4.2 requires the counting variable to have a finite moment generating function. In practice this means that we cannot answer questions regarding clustered events, for example natural disasters where we would need a counting variable that could handle highly varying time of arrivals.

Another limitation is the fact that the results are asymptotic in x. Figure 3.3 confirmed this visually, requiring a substantial threshold before convergence is reached. Practical applicability is thus limited as we need a large enough threshold before we can be certain the result holds.

Chapter 5

Conclusion

To solve problems that arise naturally in insurance and risk analysis we moved beyond the common tools of mean and variance and their applications: the central limit theorem and the law of large numbers. We did this by exploring the decay rate of the aggregate loss process. Through this we derived a tight bound when the tails decay exponentially, to answer the practical long-run question. However, this required a substantial assumption, that the claim sizes are light-tailed.

For the heavy-tailed case the question became how large reserves must be to satisfy a chosen exceedance probability. This is an important question for the actuarial field, which is governed by Solvency II regulations. To analyze the behavior of heavy-tailed distributions we introduced the class of subexponential distributions whose convolution exhibits a specific asymptotic behavior. We showed that this behavior could be generalized to a random sum of i.i.d. random variables. Chapter 3 confirmed both pictures through simulation: each framework tracked the true tail probability for its own claim distribution and broke down when applied to the other.

While the underlying mathematics proved intricate, the final results proved simple and elegant, allowing us to answer the practical questions from the introduction. This shows how the compound aggregate loss process, despite combining a random number of claims with random claim sizes, can in some cases reduce to a single factor and a single tail probability.

Bibliography

[1] Martin-Löf, A. and Sköllermo, A. *Notes on Risk Theory*, Mathematical Statistics, Department of Mathematics, Stockholm University, 2011.

[2] Foss, S., Korshunov, D. and Zachary, S. *An Introduction to Heavy-Tailed and Subexponential Distributions*, 2nd ed., Springer Series in Operations Research and Financial Engineering, Springer, 2013.

[3] European Commission, *Commission Delegated Regulation (EU) 2015/35* (Solvency II), Official Journal L 12, 2015.

[4] European Parliament and Council, *Directive 2009/138/EC* (Solvency II), Official Journal L 335, 2009.

[5] Klugman, S. A., Panjer, H. H. and Willmot, G. E. *Loss Models: From Data to Decisions*, 2nd ed., Wiley Series in Probability and Statistics, John Wiley & Sons, 2004.