Notable Distributions in Insurance

by

Márk Frankli

Thesis supervisor: **Miklós Arató**

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Bachelor's of Science
(Mathematics)

at Eötvös Lóránd University
2025

Contents

Introduction

In actuarial science and risk management, understanding the distribution of losses is essential for accurately modeling insurance claims. Among the most widely used distributions in insurance mathematics, the Pareto distribution plays a key role in characterizing heavy-tailed claim amounts. Additionally, Extreme Value Theory (EVT) provides a theoretical framework for analyzing extreme losses and shares deep connections with the Pareto distribution, making it a valuable tool in insurance risk assessment.

This thesis explores notable probability distributions in insurance, with a focus on Pareto-type distributions and their applications. The first chapter introduces the classical Pareto distribution, highlighting its key properties and methods for parameter estimation. Next, the thesis delves into the theoretical foundations of Extreme Value Theory, demonstrating its relationship with the Generalized Pareto Distribution. A dedicated chapter examines the Peaks Over Threshold method, a widely used approach for modeling extreme events, which is based on fundamental results from EVT. Finally, we discuss generalized Pareto curves, which provide a more refined modeling tool than the classical Pareto distribution.

Throughout this thesis, simulations using real-world insurance data are employed to illustrate the effectiveness of various estimation techniques. In the first chapter, a simulation is conducted to compare Maximum Likelihood Estimation and the Method of Moments in estimating the cutoff level of the Pareto distribution, offering insights into their practical performance.

Chapter 1

The Pareto Distribution

1.1 General Properties

The distribution function of a Pareto random variable X with parameters $\alpha > 0$ and $c > 0$ can be given by

$$F_X(x) = \begin{cases} 1 - \left(\frac{c}{x}\right)^\alpha & x \geq c \\ 0 & x < c \end{cases} \tag{1.1}$$

If X is Pareto distributed and $\alpha > 1$ its mean value can be given by the following formula:

$$\mathbb{E}[X] = \frac{\alpha}{\alpha - 1} c \tag{1.2}$$

Otherwise, if $\alpha \leq 1$ then $\mathbb{E}[X] = \infty$. For $\alpha > 2$, the variance of X is

$$\mathrm{Var}[X] = \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)} c^2 \tag{1.3}$$

If $\alpha \leq 2$ then $\mathrm{Var}[X] = \infty$. It follows from the distribution function F_X that the density function is

$$f_X(x) = \begin{cases} \frac{\alpha c^\alpha}{x^{\alpha+1}} & x \geq c \\ 0 & x < c \end{cases} \tag{1.4}$$

1.2 Estimating the parameters

The following section is based on the paper by Mette Rytgaard [8].

1.2.1 Maximum Likelihood Estimation

Let $X_1, ..., X_n$ be independent identically Pareto distributed random variables. Then its likelihood function is

$$\mathscr{L}(\alpha, c, x) = \prod_{i=1}^{n} \alpha c^{\alpha} X_i^{-\alpha-1} \mathbb{1}(X_i \geq c) = \alpha^n c^{\alpha n} \prod_{i=1}^{n} X_i^{-\alpha-1} \mathbb{1}(X_i \geq c) \tag{1.5}$$

Hence, its log-likelihood function is

$$\ell(\alpha, c, x) = n \log \alpha + \alpha n \log c - (\alpha + 1) \sum_{i=1}^{n} \log X_i \mathbb{1}(X_i \geq c) \tag{1.6}$$

When maximizing l a larger value of c will result in a larger value of l so the optimal value of c is

$$\hat{c} = \min_i X_i \tag{1.7}$$

while also keeping in mind that $\forall i : X_i \geq c$. From now on, we assume that the parameter c is known (as is very often the case in practice) or estimated using the above formula. For the estimation of α consider the partial derivative of l with respect to α:

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + n \log c - \sum_{i=1}^{n} \log X_i \mathbb{1}(X_i \geq c) \tag{1.8}$$

Making (1.8) equal to zero and expressing α we get the following:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log \frac{X_i}{c}} \mathbb{1}(X_i \geq c) \tag{1.9}$$

Remark. *Let* $Y = \log \frac{X}{c}$ *where* $X \sim \text{Pa}(\alpha, c)$. *Then* $Y \sim \exp(\alpha)$.

Proof. We can use the density function transformation formula to determine the density

function of Y.

$$f_Y(y) = f_X(h^{-1}(y)) \cdot \left| \frac{\partial h}{\partial y} h^{-1}(y) \right| \tag{1.10}$$

if $Y = h(X)$. Now, $h(y) = \log \frac{y}{c}$ so $h^{-1}(y) = ce^y$. Substitution gives us

$$f_Y(y) = \alpha c^\alpha (ce^y)^{-\alpha-1} ce^y = \alpha e^{-\alpha y} \tag{1.11}$$

□

Remark. *If $X, Y \sim \exp(\alpha)$ and independent, then $X + Y \sim \Gamma(2, \alpha)$.*

Proof. Using the convolution formula, we get the following:

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x-y) f_Y(y) dy = \int_0^{\infty} \alpha e^{-\alpha(x-y)} \alpha e^{-\alpha y} dy = \alpha^2 e^{-\alpha x} \int_0^x 1 dy = \alpha^2 x e^{-\alpha x} \tag{1.12}$$

□

It follows from the above that if $T = \sum_{i=1}^n \log \frac{X_i}{c}$ then $T \sim \Gamma(n, \alpha)$ with the density function

$$f_\Gamma(x) = \frac{\alpha^n}{(n-1)!} x^{n-1} e^{-\alpha x} \mathbb{1}(T \geq 0) \tag{1.13}$$

Since $\hat{\alpha} = \frac{n}{T}$, the expected value of $\hat{\alpha}$ is

$$\mathbb{E}[\hat{\alpha}] = \frac{n\alpha}{n-1} \int_0^{\infty} \frac{\alpha^{n-1}}{(n-2)!} x^{n-2} e^{-\alpha x} dx = \frac{n\alpha}{n-1} \tag{1.14}$$

because the density function of a random variable with $\Gamma(n-1, \alpha)$ distribution appears in the integral and integrating it on $\mathbb{R}$ gives 1. Next, $\mathbb{E}[\hat{\alpha}^2]$ can be calculated similarly:

$$\mathbb{E}[\hat{\alpha}^2] = \frac{n^2\alpha^2}{(n-1)(n-2)} \int_0^{\infty} \frac{\alpha^{n-2}}{(n-3)!} x^{n-3} e^{-\alpha x} dx = \frac{n^2\alpha^2}{(n-1)(n-2)} \tag{1.15}$$

Thus, the variance of $\hat{\alpha}$ is

$$\mathrm{Var}[\hat{\alpha}] = \mathbb{E}[\hat{\alpha}^2] - \mathbb{E}[\hat{\alpha}]^2 = \frac{n^2\alpha^2}{(n-1)^2(n-2)}. \tag{1.16}$$

An improved estimation of α could be $\alpha^* = \frac{n-1}{T}$ in a sense that $\mathbb{E}[\alpha^*] = \alpha$, making it unbiased. Furthermore, its variance is also smaller since

$$\mathrm{Var}[\alpha^*] = \frac{\alpha^2}{n-2} < \mathrm{Var}[\hat{\alpha}] \tag{1.17}$$

Since we assumed that $X_1, ..., X_n$ are independent Pareto distributed random variables with parameters (α, c) it follows that $Y_i = \log \frac{X_i}{c}$ are independent exponentially distributed random variables with parameter α. Now, we let

$$Z_n = \frac{1}{n-1} \sum_{i=1}^{n} Y_i \tag{1.18}$$

which is asymptotically normally distributed with parameters $\left(\frac{1}{\alpha}, \frac{1}{n\alpha^2}\right)$. To see why, note that we can use the central limit theorem since $Z_1, Z_2, ...$ is a sequence of i.i.d. random variables with finite mean and variance. Thus, as $n \to \infty$ the following holds:

$$F(Z_n < x) \approx \Phi\left(\frac{x - a_n}{b_n}\right) \tag{1.19}$$

By setting $a_n = \frac{1}{\alpha}$ and $b_n = \sqrt{\frac{1}{n\alpha^2}}$ we obtain the desired result.

Lemma 1. *Let $X_1, X_2, ...$ be i.i.d. random variables with mean μ and positive, finite variance σ^2. Let $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and suppose that f is a continuously differentiable function with $f'(\mu) \neq 0$. Then $\sqrt{n}(f(S_n) - f(\mu)) \xrightarrow{d} \mathcal{N}(0, (f'(\mu))^2\sigma^2)$.*

Proof. Throughout the proof we will use the Portmanteau lemma: $X_n \xrightarrow{d} X$ is equivalent to any of the following conditions

1. $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \quad \forall$ continuous and bounded function f
2. $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \quad \forall$ bounded and Lipschitz function f

It follows from the central limit theorem that $\sqrt{n}(\bar{X_n} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$. Since f is continuously differentiable $\exists \bar{X_n}' \in (\bar{X_n}, \mu)$ such that $f(\bar{X_n}) - f(\mu) = f'(\bar{X_n}')(\bar{X_n} - \mu)$ by the Lagrange mean value theorem. By applying the continuous mapping theorem and

the weak law of large numbers we get $f'(\bar{X_n}') \xrightarrow{p} f'(\mu)$. Let $Y_n = f'(\bar{X_n}')$, $c = f'(\mu)$, $X = \mathcal{N}(0, \sigma^2)$ and $X_n = \sqrt{n}(\bar{X_n} - \mu)$. With this notation, it suffices to only prove that $(X_n, Y_n) \xrightarrow{d} (X, c)$ since by letting $g(x, y) = xy$ (a continuous function) the continuous mapping theorem gives $g(X_n, Y_n) = X_n Y_n \xrightarrow{d} g(X, c) = \mathcal{N}(0, c^2\sigma^2)$.

Therefore, we focus on proving that if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c \in \mathbb{R}$ then $(X_n, Y_n) \xrightarrow{d} (X, c)$. For this notice that $(X_n, c) \xrightarrow{d} (X, c)$ because if f is any bounded and continuous function then $g(x) = f(x, c)$ is also bounded and continuous and by the Portmanteau lemma $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ from which $\mathbb{E}[f(X_n, c)] \to \mathbb{E}[f(X, c)]$ so $(X_n, c) \xrightarrow{d} (X, c)$ by again using the lemma. Furthermore, $|(X_n, Y_n) - (X_n, c)| = |Y_n - c| \xrightarrow{p} 0$.

Now let $Z_n = (X_n, Y_n)$, $W = (X, c)$ and $W_n = (X_n, c)$ be random variables. We claim that if $|Z_n - W_n| \xrightarrow{p} 0$ and $W_n \xrightarrow{d} W$ then $Z_n \xrightarrow{d} W$ which is exactly what we are trying to prove. For this let f be a bounded and Lipschitz function so $|f(x)| \leq M \; \forall \, x \in \mathbb{R}$ and $\exists K : |f(x) - f(y)| \leq K|x - y| \; \forall x, y \in \mathbb{R}$. Notice that

$$
\begin{aligned}
|\mathbb{E}[f(Z_n) - f(W_n)]| &\leq \mathbb{E}[|f(Z_n) - f(W_n)|] \\
&= \mathbb{E}[|f(Z_n) - f(W_n)|\mathbb{I}(|Z_n - W_n| < \epsilon)] \\
&+ \mathbb{E}[|f(Z_n) - f(W_n)|\mathbb{I}(|Z_n - W_n| \geq \epsilon)] \\
&\leq K\epsilon\mathbb{P}(|Z_n - W_n| < \epsilon) + 2M\mathbb{P}(|Z_n - W_n| \geq \epsilon) \\
&\leq K\epsilon + 2M\mathbb{P}(|Z_n - W_n| \geq \epsilon|)
\end{aligned}
$$

Also,

$$
\begin{aligned}
|\mathbb{E}[f(Z_n) - f(W)]| &\leq |\mathbb{E}[f(Z_n) - f(W_n)]| + |\mathbb{E}[f(W_n) - f(W)]| \\
&\leq K\epsilon + 2M\mathbb{P}(|Z_n - W_n| \geq \epsilon|) + |\mathbb{E}[f(W_n) - f(W)]|
\end{aligned}
$$

The first inequality uses the triangle inequality and the second one is true because of

the previous result. Now, if $\epsilon \to 0$ then $K\epsilon \to 0$ and $\mathbb{P}(|Z_n - W_n| \geq \epsilon|) \to 0$ since $|Z_n - W_n| \xrightarrow{p} 0$. Because $W_n \xrightarrow{d} W$, $|\mathbb{E}[f(W_n) - f(W)]| \to 0$ by the Portmanteau lemma. Therefore, $|\mathbb{E}[f(Z_n) - f(W)]| \to 0$ so $Z_n \xrightarrow{d} W$ by the lemma.

□

Consider the function $f(y) = \frac{1}{y}$ with $f'(\frac{1}{\alpha}) = -\alpha^2 \neq 0$. It follows from Lemma 1 that the sequence $a_n^* = f(Z_n)$ is asymptotically normally distributed with parameters $\left(\alpha, \frac{\alpha^2}{n}\right)$.

1.2.2 Method of Moments Estimation

First, we want to estimate α given that we know c. According to (1.2) $\mathbb{E}[X] = \frac{\alpha^0}{\alpha^0 - 1}c$ if $X \sim \text{Pa}(\alpha^0, c)$. Solving $\bar{X} = \mathbb{E}[X]$ gives $\alpha^0 = \frac{\bar{X}}{\bar{x} - c}$ which is the first method of moments estimator. Consider the independent random variables $X_1, ..., X_n$ where $X_i \sim \text{Pa}(\alpha, c)$. Then it follows from the central limit theorem that

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.20}$$

is asymptotically normally distributed with parameters $\left(\frac{\alpha}{\alpha-1}c, \frac{\alpha c^2}{(\alpha-1)^2(\alpha-2)n}\right)$ if $\alpha > 2$. Letting $f(y) = \frac{y}{y-c}$ with $f'\left(\frac{\alpha c}{\alpha-1}\right) \neq 0$ it follows from Lemma 1 that the sequence $a_n^0 = f(Y_n)$ is asymptotically distributed with parameters $\left(\alpha, \frac{\alpha(\alpha-1)^2}{n(\alpha-2)}\right)$. In practice, if we do not know every single loss amount (which is needed for the maximum likelihood estimation) only the total amount of losses and the number of losses exceeding c then we can only calculate α^0.

Now, we focus on estimating both α and c using the method of moments. For this, consider $X \sim \text{Pa}(\alpha, c)$ with its first two moments

$$\mu_1 = \mathbb{E}[X] = \frac{\alpha}{\alpha - 1}c = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \mu_2 = \mathbb{E}[X^2] = \frac{\alpha}{\alpha - 2}c^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{1.21}$$

where the samples $X_1, ..., X_n$ are independent identically Pareto distributed random variables. It follows from the first equation that $\hat{c} = \left(1 - \frac{1}{\alpha}\right)\bar{X}$ and substituting it back to the

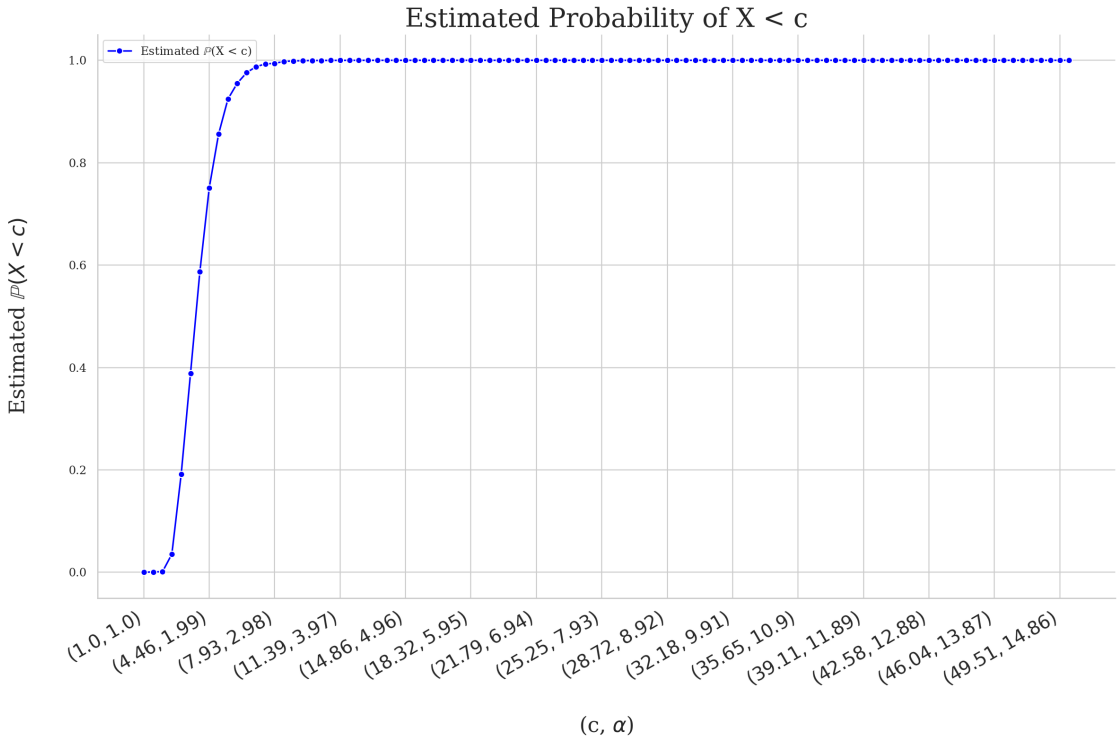

Figure 1.1: The probability that all sample points are greater than $\hat{c}$. The experiment was performed on a random sample size of 10^3 for different values of α and c.

second one gives $\hat{\alpha} = \left(\pm\sqrt{\bar{X^2}^2 - \bar{X^2}\bar{X^2}} + \bar{X^2} - \bar{X}^2 \right) \left(\bar{X^2} - \bar{X}^2 \right)^{-1}$ where $\bar{X^2} = \frac{1}{n} \sum_{i=1}^{n} X_i^2$.

A major difference between the maximum likelihood and the method of moment estimation is the way the parameter c is determined. In the former case, all sample points will correspond to a nonzero value, while in the latter there may be a sample point $X_i < \hat{c}$ such that the estimated model will return 0 for X_i. However, as c and n increase the probability that all of $X_1, ..., X_n$ is greater than c gets close to 1.

1.3 Simulations

Now we simulate some of the above illustrated methods to see how well they work in practice. We start with the methods described in Section 1.2.2. For this, we randomly selected parameters α and c from intervals $(2.5, 7)$ and $(2.8, 8)$, respectively and generated 10^3

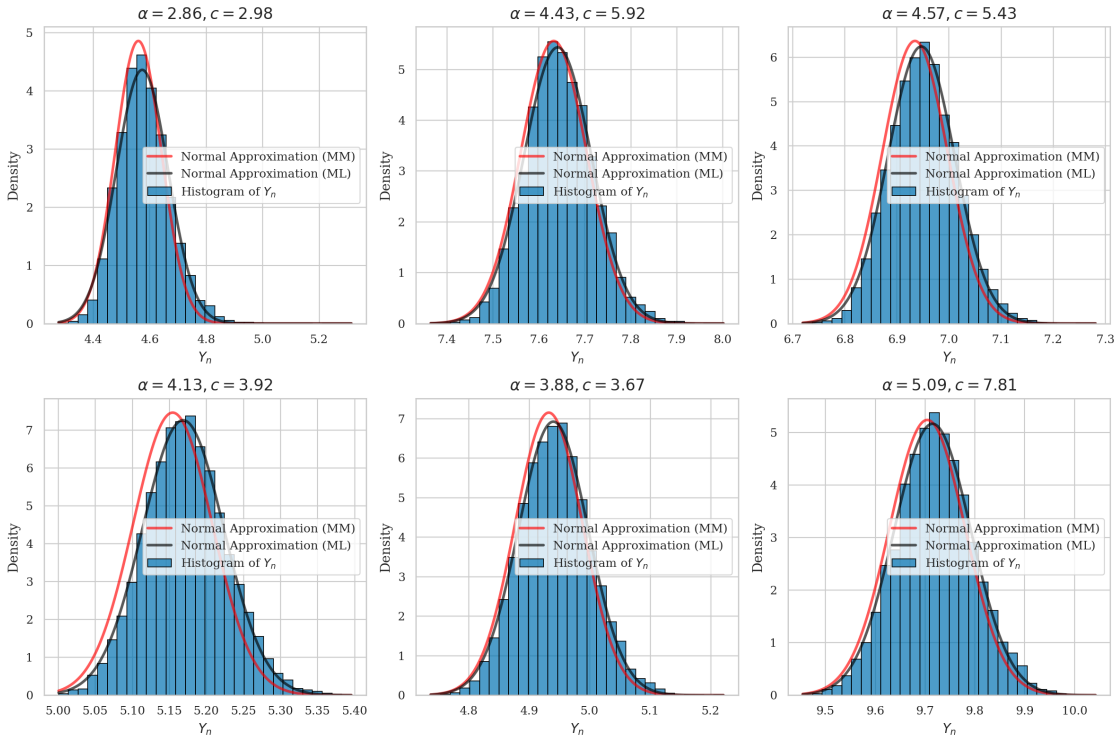

Figure 1.2: Six randomly generated Y_n samples that are asymptotically normally distributed. We can see that the maximum likelihood method gives a slightly better estimation than method of moments, which is a bit shifted to the left.

i.i.d. Pareto distributed random samples and repeated it 10^4 times to get Y_n. After this, we estimated the parameters α and c using maximum likelihood and method of moments.

Next, we tested these estimation techniques on a real-world dataset, containing 10,000 rows of claims insurance amounts with additional information about the customer and insurance policy. It is usually assumed that insurance claims follow a Pareto distribution allowing us use historical data to estimate the parameters of the distribution. Before fitting a Pareto distribution to the dataset only the claims with value greater than \$5,000 were kept. It is worth noting that when using method of moments the estimated $\hat{c}$ turned out to be greater than some of the claims amounts which can be a disadvantage compared to the maximum likelihood estimation.

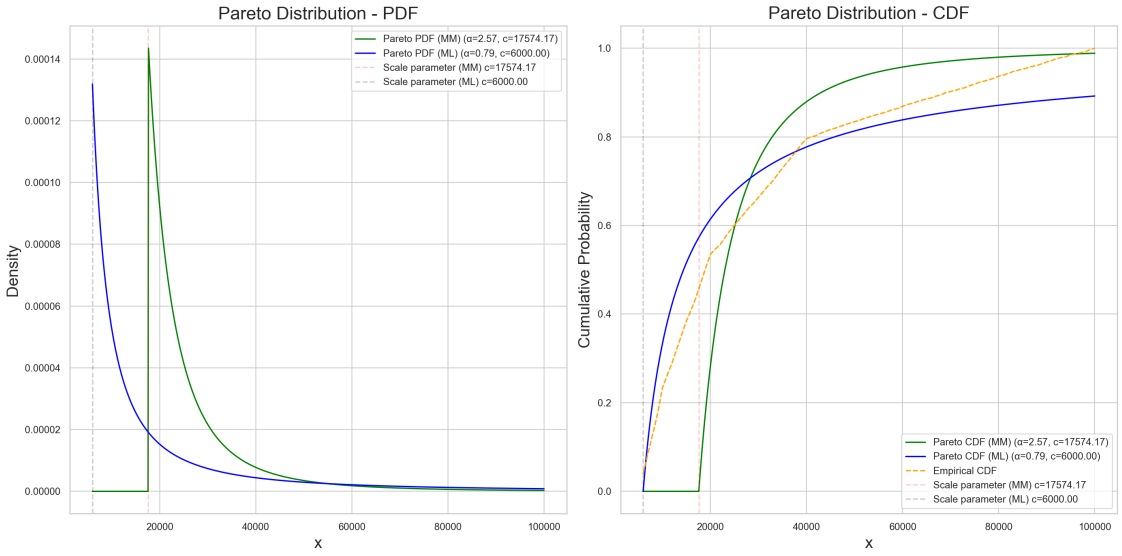

Figure 1.3: The PDF, CDF and ECDF of a Pareto distribution with parameters α and c determined by ML and MM with a cutoff level at 5,000.

Chapter 2

Extreme Value Theory

2.1 Introduction

Suppose X_1, X_2, ... are i.i.d. random variables. The central limit theorem is concerned with the limit of $X_1+X_2+...+X_n$ as $n \to \infty$, whereas in extreme value theory, our goal is to find out the limit behavior of $\max(X_1, ..., X_n)$ or $\min(X_1, ..., X_n) = -\max(-X_1, ..., -X_n)$.

Let F be the distribution function of X_i and $x^* = \sup\{x : F(x) < 1\}$. Then,

$$\max(X_1, ..., X_n) \xrightarrow{p} x^*, \ n \to \infty \tag{2.1}$$

since $\mathbb{P}(\max(X_1, ..., X_n) < x) = \mathbb{P}(X_1 < x, ..., X_n < x) = F^n(x)$. This expression converges to 0 if $x < x^*$ and to 1 if $x \geq x^*$. Therefore, to avoid a degenerate limit distribution, we introduce real sequences $a_n > 0$ and b_n such that

$$\frac{\max(X_1, ..., X_n) - a_n}{b_n} \tag{2.2}$$

holds with

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{2.3}$$

for each continuity point of G, if G is a nondegenerate distribution function.

Definition. *G is called an extreme value distribution*

Definition. *The class of distribution functions F satisfying (2.3) is called the maximum domain of attraction of G. Notation: $F \in \mathcal{D}(G)$.*

2.2 Alternative Formulations

In order to work with (2.3) easier, we will write it in a few different equivalent forms. By taking the logarithm of both sides, we get

$$\lim_{n\to\infty} n \log F(a_n x + b_n) = \log G(x) \tag{2.4}$$

Since for every fixed x the right hand side is constant it follows that $F(a_n x + b_n) \to 1$. So,

$$\lim_{n\to\infty} \frac{-\log F(a_n x + b_n)}{1 - F(a_n x + b_n)} = 1 \tag{2.5}$$

By using (2.5) and taking the reciprocal of both sides, we can rewrite (2.4) to

$$\lim_{n\to\infty} \frac{1}{n(1 - F(a_n x + b_n))} = -\frac{1}{\log G(x)} \tag{2.6}$$

2.3 Finding The Limit Distribution

Now, our goal is to find all limit distributions G for which (2.3) holds.

Definition. *For a nondecreasing function f let $f^{\leftarrow}(y) = \inf\{y : f(y) \geq x\}$ be its left-continuous inverse.*

The following lemma turns out to be useful in writing (2.6) in a more concise form.

Lemma 2. *Let f_n be a sequence of nondecreasing functions, g a nondecreasing function and $\forall x \in (a, b)$ for which x is a continuity point of g*

$$\lim_{n\to\infty} f_n(x) = g(x) \tag{2.7}$$

Then $\forall x \in (g(a), g(b))$ *that is a continuity point of* $g^{\leftarrow}$ *we have*

$$\lim_{n\to\infty} f_n^{\leftarrow}(x) = g^{\leftarrow}(x) \tag{2.8}$$

Proof. We want to show that $\forall \epsilon > 0\ \exists n_\epsilon\ \forall n \geq n_\epsilon$

$$f_n^{\leftarrow}(x) - \epsilon \leq g^{\leftarrow}(x) \leq f_n^{\leftarrow}(x) + \epsilon \tag{2.9}$$

if x is a continuity point of $g^{\leftarrow}$. Choose $0 < \epsilon_1 < \epsilon$, such that $g^{\leftarrow}(x) - \epsilon_1$ is a continuity point of g. Also, note that $g^{\leftarrow}$ is strictly monotonically increasing. Since g is nondecreasing it follows that $g^{\leftarrow}$ is also nondecreasing. For the sake of contradiction, if we assume that $\exists x, y : x < y$ and $g^{\leftarrow}(x) = g^{\leftarrow}(y)$ then g would take both x and y at $g^{\leftarrow}(x)$ which is impossible. From the definition of left-continuous inverse functions $g(g^{\leftarrow}(x)) \leq x$ but because g is continuous at $g^{\leftarrow}(x)-\epsilon_1$ and $g^{\leftarrow}$ is strictly monotonically increasing, $g(g^{\leftarrow}(x)-\epsilon_1) < x$ is also true. Next, choose $\delta < x-g(g^{\leftarrow}(x)-\epsilon_1)$. Since g is continuous at $g^{\leftarrow}(x)-\epsilon_1$, $\exists n_0\ \forall n \geq n_0 : f_n(g^{\leftarrow}(x) - \epsilon_1) < g(g^{\leftarrow}(x) - \epsilon_1) < x$. Applying $f_n^{\leftarrow}$ to both sides gives $g^{\leftarrow}(x) \leq f_n^{\leftarrow}(x) + \epsilon_1$ which proves the right inequality. The other direction can be proved similarly. □

Now, if we let $U^{\leftarrow}$ be the left-continuous inverse of $\frac{1}{1-F}$, then we can use Lemma 2 to write (2.6) as

$$\lim_{n\to\infty} \frac{U^{\leftarrow}(nx) - b_n}{a_n} = G^{\leftarrow}(e^{-\frac{1}{x}}) \coloneqq D(x) \tag{2.10}$$

since $\frac{U(a_n x+b_n)}{n} \Leftrightarrow U(a_n x + b_n) = yn \Leftrightarrow a_n x + b_n \geq U^{\leftarrow}(yn) \Leftrightarrow x \geq \frac{U^{\leftarrow}(yn)-b_n}{a_n}$. Similarly, $-\frac{1}{\log G(x)} \Leftrightarrow e^{-\frac{1}{y}} = G(x) \Leftrightarrow G^{\leftarrow}(e^{-\frac{1}{y}}) \leq x$ and applying Lemma 2 gives the desired result. This looks promising, as it is a much simpler expression than before and we will use it in the next lemma, which will be a stepping stone to the main theorem of this chapter.

Lemma 3. *Let* $a_n > 0$, b_n *be real sequences,* G *a nondegenerate distribution function,* $a(t) \coloneqq a_{\lfloor t \rfloor}$ *and* $b(t) \coloneqq b_{\lfloor t \rfloor}$. *The following statements are equivalent:*

1.

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x) \tag{2.11}$$

if x is a continuity point of G.

2.

$$\lim_{t\to\infty} t(1 - F(a(t)x + b(t))) = -\log G(x) \tag{2.12}$$

if x is a continuity point of G such that $0 < G(x) < 1$.

3.

$$\lim_{t\to\infty} \frac{U^{\leftarrow}(tx) - b(t)}{a(t)} = D(x) \tag{2.13}$$

if $x > 0$ is a continuity point of D.

Proof. (2) $\Leftrightarrow$ (3): Since (2) is just an alternative formulation of (2.3) the equivalence follows from (2.10) and Lemma 2.

(1) $\Leftrightarrow$ (3): We have already shown that (1) $\Leftrightarrow$ (2.10) so it is sufficient to show that (3) $\Leftrightarrow$ (2.10). For this, notice that if $t \geq 1$ and x is a continuity point of D then

$$\frac{U^{\leftarrow}(\lfloor t \rfloor\, x) - b(t)}{a(t)} \leq \frac{U^{\leftarrow}(tx) - b(t)}{a(t)} \leq \frac{U^{\leftarrow}\left(tx\left(1 + \frac{1}{\lfloor t \rfloor}\right)\right) - b(t)}{a(t)} \tag{2.14}$$

By letting $n = \lfloor t \rfloor$ the left-hand side converges to $D(x)$ due to (2.10). The right-hand side converges to $D(x') > D(x)$ if $x' > x$ for every continuity point x' and (3) follows from this. □

The following theorem identifies all nondegenerate distribution functions that can occur in (2.3) (in other words, the class of extreme value distributions).

Theorem 1 (Fisher and Tippet, 1928)**.** *The class of extreme value distributions is $G_\gamma(ax+b)$ where $a > 0$ and*

$$G_\gamma(x) = exp(-(1+\gamma x)^{-\frac{1}{\gamma}}),\ 1 + \gamma x > 0 \tag{2.15}$$

and if $\gamma = 0$ then the right-hand side is $exp(-e^{-x})$.

Proof. Consider the function D defined in Lemma 3 part (3) and suppose that 1 is a continuity point of D. Then,

$$\lim_{t\to\infty}\frac{U(tx)-U(t)}{a(t)}=\lim_{t\to\infty}\frac{U(tx)-b(t)-U(t)+b(t)}{a(t)}=D(x)-D(1)\coloneqq E(x) \tag{2.16}$$

for each continuity point $x>0$. If $y>0$,

$$\frac{U(txy)-U(t)}{a(t)}=\frac{U(txy)-U(ty)}{a(ty)}\cdot\frac{a(ty)}{a(t)}+\frac{U(ty)-U(t)}{a(t)} \tag{2.17}$$

Now, we want to show that the limits $\lim_{t\to\infty}\frac{U(ty)-U(t)}{a(t)}$ and $\lim_{t\to\infty}\frac{a(ty)}{a(t)}$ exist. We proceed with an indirect proof. Suppose $\exists A_1, A_2, B_1, B_2$ such that $A_1\neq A_2$ or $B_1\neq B_2$ where A_i and B_i are limit points of $\frac{a(ty)}{a(t)}$ and $\frac{U(ty)-U(t)}{a(t)}$, respecitvely, for $i=1,2$. Taking the limit of both sides in (2.17), we get

$$E(xy)=E(x)A_i+B_i \tag{2.18}$$

for $i=1,2$ and for every continuity point of $E(\cdot)$ and $E(\cdot y)$. For any x take a sequence of continuity points x_n such that $x_n\to x^-$. Since E is left-continuous (because D is the left-continuous inverse of G at $e^{-\frac{1}{x}}$), it follows that $E(x_ny)\to E(xy)$ so (2.18) holds for all $x,y>0$. Subtracting (2.18) for $i=1,2$ from each other gets

$$E(x)(A_1-A_2)=B_2-B_1 \tag{2.19}$$

If $A_1\neq A_2$ or $B_2\neq B_1$ then $E(x)$ would be a constant function but it is impossible since we already assumed that G is nondegenerate. Thus, $A_1=A_2$ and $B_1=B_2$ must hold. Therefore, the aforementioned limits exist and let

$$A(y)\coloneqq\lim_{t\to\infty}\frac{a(ty)}{a(t)} \tag{2.20}$$

for every $y>0$ and we can rewrite $E(xy)$ as

$$E(xy)=E(x)A(y)+E(y) \tag{2.21}$$

for every $x, y > 0$. Letting $s = \log x$, $t = \log y$ and $H(x) \coloneqq E(e^x)$, we have

$$H(t+s) = H(s)A(e^t) + H(t) \tag{2.22}$$

Since $H(0) = E(1) = D(1) - D(1) = 0$ we can rewrite (2.22) as

$$\frac{H(t+s) - H(t)}{s} = \frac{H(s) - H(0)}{s} A(e^t) \tag{2.23}$$

Since H is monotone (because e^x is monotone and the left-continuous inverse $G^{\leftarrow}(e^{-\frac{1}{x}})$ is monotone as well) $\exists t : \exists H'(t)$ so according to (2.23) H is differentiable everywhere. Thus, by taking the limit of both sides, as $s \to \infty$, we get

$$H'(t) = H'(0)A(e^t) \tag{2.24}$$

Now, if $H'(0) = 0$ was true then $H(t)$ would be constant but it contradicts with the fact that G is assumed to be nondegenerate. Hence, $Q(t) \coloneqq \frac{H(t)}{H'(0)} = A(e^t)$ is well-defined. It follows that $Q(0) = A(1) = 0$ and $Q'(0) = 1$. Then we can rewrite (2.22) as

$$Q(t+s) - Q(t) = Q(s)Q'(t),\ Q(s+t) - Q(s) = Q(t)Q'(s) \tag{2.25}$$

where in the last equation we exchanged the role of t and s. Subtracting these two from each other and dividing by s gets

$$Q(t)\frac{Q'(s) - 1}{s} = \frac{Q(s)}{s}(Q'(t) - 1) \tag{2.26}$$

so if $s \to \infty$ this becomes

$$Q(t)Q''(0) = Q'(t) - 1 \tag{2.27}$$

Since $Q'(t) = cH(t) + 1$ for a $c \in \mathbb{R}$ and we already showed that H is differentiable

everywhere it follows that Q is twice differentiable. Thus,

$$Q''(t) = Q'(t)Q''(0) \text{ and } \gamma \coloneqq (\log Q')'(t) = Q''(0) \tag{2.28}$$

and it follows that

$$Q'(t) = e^{\gamma t} \tag{2.29}$$

and it is well-defined for $t = 0$ as well since $Q'(0) = 1$. Moreover,

$$Q(t) = \int_0^t e^{\gamma t} = \frac{e^{\gamma t} - 1}{\gamma} \tag{2.30}$$

which is also well-defined since $Q(t) = 0$ at $t = 0$. Hence,

$$H(t) = H'(0)\frac{e^{\gamma t} - 1}{\gamma} \Leftrightarrow D(t) = D(1) + H'(0)\frac{t^{\gamma} - 1}{\gamma} \tag{2.31}$$

Also,

$$\begin{aligned} D(1) + H'(0)\frac{t^{\gamma} - 1}{\gamma} = x \Leftrightarrow t^{\gamma} - 1 = \frac{x - D(1)}{H'(0)}\gamma \Leftrightarrow t = \left(1 + \frac{x - D(1)}{H'(0)}\gamma\right)^{\frac{1}{\gamma}} \\ \Leftrightarrow D^{\leftarrow}(x) = \left(1 + \frac{x - D(1)}{H'(0)}\gamma\right)^{\frac{1}{\gamma}} \end{aligned} \tag{2.32}$$

By definition, $D(x) = G^{\leftarrow}(e^{-\frac{1}{x}})$, so

$$\begin{aligned} D^{\leftarrow}(x) = \left(1 + \frac{x - D(1)}{H'(0)}\gamma\right)^{\frac{1}{\gamma}} = -\frac{1}{\log G(x)} \Leftrightarrow G(x) &= exp\left(-\left(1 + \gamma\frac{x - D(1)}{H'(0)}\right)^{-\frac{1}{\gamma}}\right) \\ &= exp\left(-(1 + \gamma(ax + b))^{-\frac{1}{\gamma}}\right) \end{aligned} \tag{2.33}$$

for $a = \frac{1}{H'(0)}$ and $b = -\frac{D(1)}{H'(0)}$. If 1 is not a continuity point of D then choose an arbitrary continuity point x_0 of D and follow the proof with $U(tx_0)$. □

Theorem 2. *If $\gamma \in \mathbb{R}$ then the following are equivalent:*

1. *If* $1+\gamma x > 0$ *and* $a_n > 0$ *then*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = exp(-(1+\gamma x)^{-\frac{1}{\gamma}}) \tag{2.34}$$

2. *There is a positive function* a *such that*

$$\lim_{t\to\infty} \frac{U^{\leftarrow}(tx) - U^{\leftarrow}(t)}{a(t)} = D_\gamma(x) = \frac{x^\gamma - 1}{\gamma} \tag{2.35}$$

and if $\gamma = 0$ *the right-hand side is interpreted as* $\log x$.

3. *There is a positive function* a *such that*

$$\lim_{t\to\infty} t(1 - F(a(t)x + U(t))) = (1+\gamma x)^{-\frac{1}{\gamma}} \tag{2.36}$$

if $1+\gamma x > 0$.

4. *There is a positive function* f *such that*

$$\lim_{t\to x^{*-}} \frac{1 - F(t + f(t)x)}{1 - F(t)} = (1+\gamma x)^{-\frac{1}{\gamma}} \tag{2.37}$$

for all x *for which* $1+\gamma x > 0$ *and* $x^* = sup\{x : F(x) < 1\}$.

Proof. (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3): By letting $b(t) = U(t)$ the equivalence follows from Lemma 3 and Theorem 1.

(2) $\Leftrightarrow$ (4): Fix $\epsilon > 0$, let g be a nondecreasing function and $g^{\leftarrow}$ be its left-continuous inverse. It follows from the definition of $g^{\leftarrow}$ that $g(g^{\leftarrow}(t)) \le t$ and since g is nondecreasing $g(g^{\leftarrow}(t) - \epsilon) \le t$ and $g(g^{\leftarrow}(t) + \epsilon) \ge t$. It follows that

$$A = \frac{U^{\leftarrow}\left(\frac{1-\epsilon}{1-F(t)}\right) - U^{\leftarrow}\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} < \frac{t - U^{\leftarrow}\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} < \frac{U^{\leftarrow}\left(\frac{1+\epsilon}{1-F(t)}\right) - U^{\leftarrow}\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} = B \tag{2.38}$$

Because of (2),

$$\lim_{t\to x^{*-}} A = \frac{(1-\epsilon)^{\gamma}-1}{\gamma} \quad \text{and} \quad \lim_{t\to x^{*-}} B = \frac{(1+\epsilon)^{\gamma}-1}{\gamma} \tag{2.39}$$

Therefore,

$$\lim_{t\to x^{*-}} \frac{t - U^{\leftarrow}\left(\frac{1}{1-F(t)}\right)}{a\left(\frac{1}{1-F(t)}\right)} = 0 \tag{2.40}$$

Again because of (2), for every $x > 0$ the following holds

$$\lim_{t\to x^{*-}} \frac{U^{\leftarrow}\left(\frac{x}{1-F(t)}\right) - t}{a\left(\frac{1}{1-F(t)}\right)} = \frac{x^{\gamma}-1}{\gamma} \tag{2.41}$$

Hence,

$$\lim_{t\to x^{*-}} \frac{1-F(t)}{1-F\left(t + xa\left(\frac{1}{1-F(t)}\right)\right)} = (1+\gamma x)^{\frac{1}{\gamma}} \tag{2.42}$$

since

$$\begin{aligned} \frac{U^{\leftarrow}\left(\frac{x}{1-F(t)}\right) - t}{a\left(\frac{1}{1-F(t)}\right)} = x &\Leftrightarrow U^{\leftarrow}\left(\frac{x}{1-F(t)}\right) = a\left(\frac{1}{1-F(t)}\right)x + t \\ &\Leftrightarrow \frac{x}{1-F(t)} \geq \frac{1}{1-F\left(t + xa\left(\frac{1}{1-F(t)}\right)\right)} \\ &\Leftrightarrow x \geq \frac{1-F(t)}{1-F\left(t + xa\left(\frac{1}{1-F(t)}\right)\right)} \end{aligned} \tag{2.43}$$

and using Lemma 2 gives (2.42). □

2.4 The Generalized Pareto Distribution

Let's revisit part (4) of Theorem 2, which states that $\exists f : f > 0$ such that

$$\lim_{t\to x^{*-}} \frac{1-F(t+f(t)x)}{1-F(t)} = (1+\gamma x)^{-\frac{1}{\gamma}} \tag{2.44}$$

if $1 + \gamma x > 0$. Now, suppose that X is a random variable with distribution function F such that $F \in \mathcal{D}(G_{\gamma})$. Then the numerator and denominator can be rewritten as

$1 - \mathbb{P}(X < t + f(t)x) \Leftrightarrow \mathbb{P}\left(\frac{X-t}{f(t)} > x\right)$ and $\mathbb{P}(X > t)$, respectively. It follows that

$$\begin{aligned}\mathbb{P}\left(\frac{X-t}{f(t)} > x \,\middle|\, X > t\right) &= \frac{\mathbb{P}\left(X > t \,\middle|\, \frac{X-t}{f(t)} > x\right)\mathbb{P}\left(\frac{X-t}{f(t)} > x\right)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}\left(\frac{X-t}{f(t)} > x\right)}{\mathbb{P}(X > t)} \\ &= \frac{1 - F(t + f(t)x)}{1 - F(t)}\end{aligned} \tag{2.45}$$

where the second equality is true because $f, x > 0$. Hence,

$$\lim_{t \to x^{*-}} \mathbb{P}\left(\frac{X-t}{f(t)} > x \,\middle|\, X > t\right) = (1 + \gamma x)^{-\frac{1}{\gamma}} \tag{2.46}$$

if $0 < x < (\max(0, -\gamma))^{-1}$. That is, we found the conditional distribution of $\frac{X-t}{f(t)}$ given that $X > t$, which is

$$H_\gamma(x) \coloneqq 1 - (1 + \gamma x)^{-\frac{1}{\gamma}}, \quad 0 < x < (\max(0, -\gamma))^{-1} \tag{2.47}$$

Definition. *A random variable X is generalized Pareto distributed (GPD) if it has the following cumulative distribution function*

$$F_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-\frac{1}{\gamma}} & \gamma \neq 0 \\ 1 - e^{-x} & \gamma = 0 \end{cases} \tag{2.48}$$

So we essentially showed that $\frac{X-t}{f(t)}$ is approximately GPD from a given threshold t. This has many applications in insurance since we can set a threshold value for the claims amount and the exceedances will follow a generalized Pareto distribution. The only remaining step is to determine the parameters of such Pareto distribution, but this is fairly straightforward with Maximum Likelihood or Method of Moments estimation. The next chapter introduces more sophisticated techniques for modeling the behavior of exceedances.

Chapter 3

Peaks Over Threshold (POT)

3.1 Motivation

For a given dataset, extreme value analysis is used to model values that largely deviate from the mean. In the previous chapter, we showed that once a high (or low) enough threshold is set, the values above (or below) that level can be interpreted as Generalized Pareto Distributed variables. The Peaks Over Threshold (POT) method is one way to model extreme values, by first setting a threshold value u according to a given strategy, secluding all samples that exceed u and modeling those values using the tail of the exceedances.

3.2 Introduction

Let u be the threshold and X be an unbounded random variable with distribution function F. Then

$$\begin{aligned} F_u(x) = \mathbb{P}\left(X - u < x \mid X > u\right) &= \frac{\mathbb{P}\left(X > u \mid X < u + x\right)\mathbb{P}(X < u + x)}{\mathbb{P}(X > u)} \\ &= \frac{F(x+u) - F(u)}{\overline{F}(u)} \end{aligned} \tag{3.1}$$

since in the numerator we are looking for the probability of $u < X < u + x$. Further suppose that X is a random variable with i.i.d. samples $X_1, ..., X_n$. Let $K(u)$ be the set

of indices that exceed u and $N(u)$ be the number of exceedances in the sample. Now, we can define the excess Y_j for each $X_j > u$ as

$$\{Y_1, ..., Y_{N(u)}\} = \{X_i - u \mid i \in K(u)\} \tag{3.2}$$

It follows that samples Y_i are i.i.d. and if a large enough u is given, they are approximately GPD. These exceedances turn out to be useful for modeling $\overline{F}(x)$ but if we use the empirical distribution function for this task and x is very large then the EDF will depend on a few extreme values causing it to have a high variance. A better approach is to rewrite $\overline{F}(x)$ as

$$\overline{F}(x) = \overline{F}(u)\overline{F}_u(x-u) = \mathbb{P}(X > u)\mathbb{P}\left(X - u > x - u \mid X > u\right) \tag{3.3}$$

Replacing $\overline{F}(x)$ with the EDF and noting that $\overline{F}_u(x)$ is approximately GPD we get the so-called *POT estimator*, which is

$$\overline{F}(x) = \frac{N(u)}{n}\left(1 + \hat{\gamma}(x-u)\right)^{-\frac{1}{\hat{\gamma}}} \tag{3.4}$$

for every $u < x < \infty$ where $\hat{\gamma}$ is a parameter estimator.

3.3 Setting a Threshold

The POT analysis depends highly on the value of u since if the value is too large there will be only a few samples exceeding this threshold, making the estimator's variance fairly high. On the contrary, if u is too low, the exceedances will be less likely to follow a GPD, resulting in a higher bias. Thus, our goal is to find a sweet spot in order to balance variance and bias.

The simplest method for setting a threshold is to select the k-largest values. Common values that usually work well in practice are $k = \sqrt{n}$, $k = \frac{n^{2/3}}{\log\log n}$, or the 90th percentile. This rule of thumb method is the fastest way of finding the value of u, but due to its sim-

plicity, it may not work well for every dataset. Other, more sophisticated methods include the Mean Residual Life Plot and Parameter Stability Plot, which rely on the graphical interpretation of the data.

3.4 Parameter Estimation

Once a suitable threshold is set we can estimate the parameters of the generalized Pareto distribution using the maximum likelihood method. If $\gamma \neq 0$ then the density function is $f(x) = (1+\gamma x)^{-\frac{1}{\gamma}-1}$ and $f(x) = e^{-x}$ if $\gamma = 0$. Then the log-likelihood function, given the exceedances $Y_1, ..., Y_{N(u)}$ is

$$l(\gamma \mid Y_1, ..., Y_{N(u)}) = -\left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^{N(u)} \log(1 + \gamma Y_i) \tag{3.5}$$

if $\gamma \neq 0$ and

$$l(Y_1, ..., Y_{N(u)}) = \sum_{i=1}^{N(u)} Y_i \tag{3.6}$$

if $\gamma = 0$. Taking the derivative with respect to γ and setting it to zero will give the estimator $\hat{\gamma}$.

3.5 Simulations

We tested the POT analysis on the same real-world insurance dataset as in Section 1.3. The 90th percentile was used as the threshold value ($\sqrt{n}$ and $\frac{n^{2/3}}{\log \log n}$ were also tested, but there were only a few hundred data samples using such threshold levels, so we omitted their illustration) and the parameters were estimated using the maximum likelihood method.

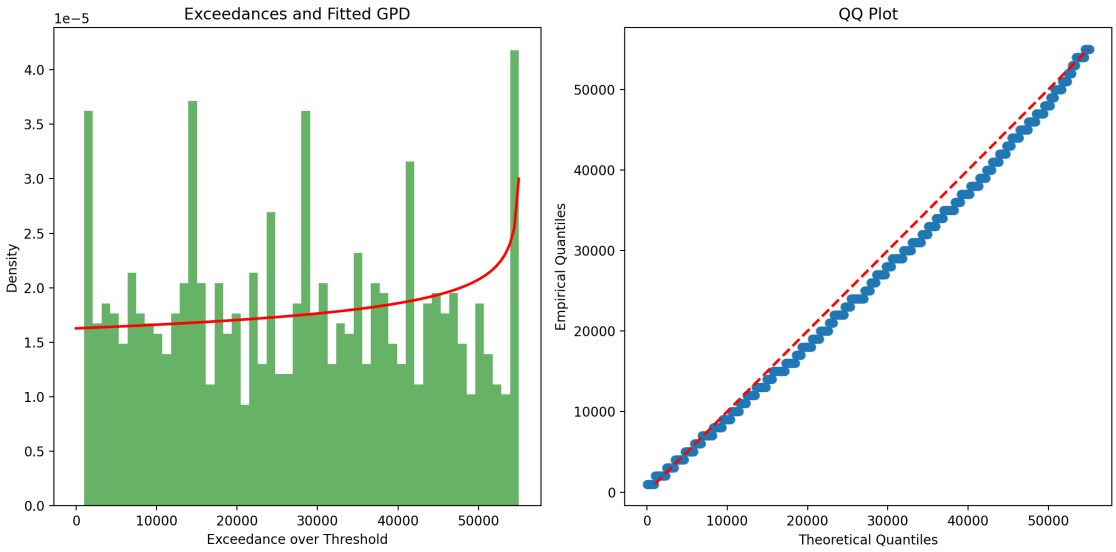

Figure 3.1: The histogram of exceedances and the fitted GPD using the 90th percentile as the threshold level. We can see it from the QQ plot that the extracted values closely follow a generalized Pareto distribution.

Chapter 4

Pareto-type Distributions

4.1 Introduction

Pareto-type distributions are a class of probability distributions that exhibit power-law behavior in the tail. More precisely, X has a Pareto-type distribution if $\mathbb{P}(X > x) \approx Cx^{-\alpha}$ with constant $C > 0$ and tail index $\alpha > 0$. It is easy to see that if $X \sim \text{Pareto}(\alpha, c)$, then X is also a Pareto-type distribution since $\mathbb{P}(X > x) = \left(\frac{c}{x}\right)^{\alpha}$, so we can let $C = c^{\alpha}$ with tail index α.

It is a well-known result that the upper tail of the income or wealth distributions can be modeled by a Pareto distribution (see [4]). However, this simple model turns out to be inaccurate in many cases, so our goal is to develop a new, more complex model that better captures the real world. This new model is the *generalized Pareto curves*.

4.2 Generalized Pareto Curves

Let X be a random variable that characterizes the distribution of income or wealth. Assume that X is integrable and its distribution function, F, is differentiable over $[a, \infty)$ or $\mathbb{R}$ and let f denote the density and Q the quantile function. Further assume that Q is invertible.

Definition. *For an income level $x > 0$, the inverted Pareto coefficient is $B(x) = \mathbb{E}(X \mid X > x)$ or*

$$b^*(x) = \frac{1}{(1 - F(x))x} \int_x^\infty z f(z) dz \tag{4.1}$$

The difference is that $b^*(x)$ is scale-invariant while $B(x)$ is not. For example, if our data suggests that the expected wealth is \$20 million above the threshold level \$10 million then $B(x) = 20 \cdot 10^6$ while $b^*(x) = 2$. Notice that we can rewrite $b^*(x)$ using the quantile function in the following form

$$b(p) = \frac{1}{(1 - p)Q(p)} \int_p^1 Q(z) dz \tag{4.2}$$

Note that if $X \sim \text{Pareto}(\alpha, c)$ then $b(p) = \frac{\alpha}{\alpha - 1}$. We call the function $b : p \mapsto b(p)$ defined over $[\overline{p}, 1)$ a *generalized Pareto curve*, where $\overline{p} = F(c)$. It is also reasonable to assume that $c > 0$ since negative thresholds do not have applications in practice and the definition has a singularity at 0.

Theorem 3. *If X satisfies the assumptions stated above then b is differentiable and $\forall p \in [\overline{p}, 1) : 1 - b(p) + (1 - p)b'(p) \leq 0$ and $b(p) \geq 1$.*

Proof. The fact that $b(p) \geq 1$ follows from the definition since $b(p) = b^*(x) = \frac{\mathbb{E}(X \mid X > x)}{x}$. Also, $\forall p \geq \overline{p}$:

$$(1 - p)Q(p)b(p) = \int_p^\infty Q(z) dz \tag{4.3}$$

Differentiating both sides with respect to p gives

$$(1 - p)Q(p)b'(p) + (1 - p)b(p)Q'(p) - b(p)Q(p) = -Q(p) \tag{4.4}$$

Since we assumed $c > 0$ it follows that $\forall p \geq \overline{p} : Q(p) > 0$. Then dividing both sides by $Q(p)$ gives

$$(1 - p)b'(p) + (1 - p)b(p)\frac{Q'(p)}{Q(p)} - b(p) = -1 \tag{4.5}$$

After rearranging:

$$(1-p)b(p)\frac{Q'(p)}{Q(p)} = b(p) - 1 - (1-p)b'(p) \tag{4.6}$$

Since the quantile function is increasing, the left-hand side is nonnegative, which concludes the proof. □

Theorem 4. *If X is defined for $x > c$, $F(c) = \overline{p}$ and the generalized Pareto curve is $b : [\overline{p}, 1) \mapsto \mathbb{R}$ then $\forall p \geq \overline{p}$, the p-th quantile is*

$$Q(p) = c\frac{(1-\overline{p})b(\overline{p})}{(1-p)b(p)} exp\left(-\int_{\overline{p}}^{p} \frac{1}{(1-p)b(u)}du\right) \tag{4.7}$$

Proof. From (4.6) we have

$$\frac{Q'(p)}{Q(p)} = \frac{1}{1-p} - \frac{1}{(1-p)b(p)} - \frac{b'(p)}{b(p)} \tag{4.8}$$

After integrating from $\overline{p}$ to p:

$$\int_{\overline{p}}^{p} \frac{Q'(u)}{Q(u)}du = \log Q(p) - \log Q(\overline{p}) = \int_{\overline{p}}^{p} \frac{1}{1-u}du - \int_{\overline{p}}^{p} \frac{1}{(1-u)b(u)}du - \int_{\overline{p}}^{p} \frac{b'(u)}{b(u)}du \tag{4.9}$$

From which

$$\begin{aligned} Q(p) &= Q(\overline{p})\exp\left(\int_{\overline{p}}^{p} \frac{1}{1-u}du - \int_{\overline{p}}^{p} \frac{1}{(1-u)b(u)}du - \int_{\overline{p}}^{p} \frac{b'(u)}{b(u)}du\right) \\ &= Q(\overline{p})\frac{(1-\overline{p})b(\overline{p})}{(1-p)b(p)}\exp\left(-\int_{\overline{p}}^{p} \frac{1}{(1-p)b(u)}du\right) \end{aligned}$$

and since $Q(\overline{p}) = c$ we are done. □

It can be verified that for power laws (e.g., the Pareto distribution) $b(p)$ is constant. However, pure power laws rarely exist in practice, so we can weaken the definition and only characterize distributions through their asymptotic behavior. This is the motivation for the following definition.

Definition. *X is an asymptotic power law if for some $\alpha > 0$, $1 - F(x) = L(x)x^{-\alpha}$, where $L(x)$ is a slowly varying function which means that $\forall \lambda > 0$: $\lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1$.*

We can also generalize the definition of slowly varying functions to allow arbitrary positive real numbers as the limit.

Definition. *L is a regularly varying function if $\forall \lambda > 0 : \lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} \in \mathbb{R}^+$*

In general, we will restrict ourselves to $\alpha > 1$ so that the means are finite. The following theorems are from Karamata, and they turn out to be useful in the proof of theorems regarding asymptotic power laws.

Theorem 5 (Karamata, direct half)**.** *Let f be a regularly varying function with index α, and be locally bounded on $[a, \infty)$. Then for any $\sigma < -(\alpha + 1)$*

$$\lim_{x\to\infty} \frac{x^{\sigma+1} f(x)}{\int_x^\infty t^\sigma f(t)dt} = -(\sigma + \alpha + 1) \tag{4.10}$$

Theorem 6 (Karamata, converse half)**.** *Let f be positive and locally integrable on $[a, \infty)$. If for some $\sigma < -(\alpha + 1)$,*

$$\lim_{x\to\infty} \frac{x^{\sigma+1} f(x)}{\int_x^\infty t^\sigma f(t)dt} = -(\sigma + \alpha + 1) \tag{4.11}$$

then f varies regularly with index α.

Lemma 4. *Let f be measurable and positive and assume that for some $\lambda_0 > 1$,*

$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} = \infty \quad (\lambda > \lambda_0) \tag{4.12}$$

Then (4.12) holds uniformly in λ over every interval $[\lambda_0, \infty)$ where $\lambda > \lambda_0^2$.

Corollary 1. *If $f \in R_\infty$ then Lemma 4 holds uniformly in λ over all intervals $(0, \lambda_0^{-1})$ and (λ, ∞) for every $\lambda_0 > 1$. Here, R_∞ denotes the set of regularly varying functions at ∞.*

The proofs for the above results can be found in the book *Regular Variation* [5]. For the next few proofs, the following lemma will be used, which states that we can rewrite $b^*(x)$ into a more convenient form.

Lemma 5.

$$b^*(x) = 1 + \frac{1}{(1-F(x))x}\int_x^\infty 1 - F(z)dz \tag{4.13}$$

Proof. Since

$$1 - F(x) = \int_x^\infty f(z)dz \tag{4.14}$$

we can use integration by parts:

$$\int_x^\infty zf(z)dz = \int_x^\infty (-z)(-f(z))dz = [z(1-F(z)]_{z=x}^\infty + \int_x^\infty 1 - F(z)dz \tag{4.15}$$

Because we assumed that X is integrable, we have $1 - F(x) = o\left(\frac{1}{x}\right)$ from Markov's inequality. This is because if $g(x) = \mathbb{P}(X > x)$ then

$$\mathbb{E}(|X|) = \int_0^\infty g(x)dx < \infty \tag{4.16}$$

and since g is nonincreasing, we have

$$\lim_{x\to\infty} xg(x) = 0 \tag{4.17}$$

Hence, the bracketed term converges to 0 if $x \to \infty$, so (4.15) becomes

$$x(1 - F(x)) + \int_x^\infty 1 - F(z)dz \tag{4.18}$$

and substituting it back to b^* concludes the proof. □

Theorem 7. *If $\alpha > 0$ then X is an asymptotic power law if and only if $\lim_{p\to 1} b(p) = \frac{\alpha}{\alpha-1}$*

Proof. $\Rightarrow$: Note that $\lim_{p\to 1} b(p) = \lim_{x\to\infty} b^*(x)$ and the assumption that L is slowly varying is equivalent to the assumption that $1 - F$ is regularly varying. Then, applying the direct

half of Karamata's theorem with $\sigma = 0$ to Lemma 5 we get

$$\lim_{x\to\infty} b^*(x) = 1 + \frac{1}{\alpha - 1} = \frac{\alpha}{\alpha - 1} \tag{4.19}$$

$\Leftarrow$: Since $\lim_{p\to 1} b(p) = \frac{\alpha}{\alpha-1}$ we have

$$\lim_{x\to\infty} \frac{1}{b^*(x) - 1} = \alpha - 1 \tag{4.20}$$

Applying the converse half of Karamata's theorem with $\sigma = 0$ we conclude that $1 - F$ is regularly varying with index $-\alpha$. □

It is worth mentioning that the previous theorem generalizes the fact that if $X \sim \text{Pareto}(\alpha, c)$ then $b(p) = \frac{\alpha}{\alpha-1}$ in an asymptotic sense.

Theorem 8. *$1 - F(x)$ is rapidly varying, meaning $\forall\lambda > 1 : \lim_{x\to\infty} \frac{1-F(\lambda x)}{1-F(x)} = 0$ if and only if $\lim_{p\to\infty} b(p) = 1$.*

Proof. $\Rightarrow$: After a change of variable to $z = tx$ in Lemma 5 we have

$$b^*(x) = 1 + \int_1^\infty \frac{1 - F(tx)}{1 - F(x)} dt = 1 + \int_1^K \frac{1 - F(tx)}{1 - F(x)} dt + \int_K^\infty \frac{1 - F(tx)}{1 - F(x)} dt \tag{4.21}$$

for $K > 1$. Since we assumed that F is differentiable, the function $t \mapsto \frac{1-F(tx)}{1-F(x)}$ is continuous on $[1, K]$, so it is bounded. Thus, from the dominated convergence theorem,

$$\lim_{x\to\infty} \left(\int_1^K \frac{1 - F(tx)}{1 - F(x)} dt \right) = \int_1^K \left(\lim_{x\to\infty} \frac{1 - F(tx)}{1 - F(x)} \right) dt = 0 \tag{4.22}$$

because we assumed that $1 - F(x)$ is rapidly varying. We also have

$$\lim_{x\to\infty} \frac{1 - F(xt)}{1 - F(x)} dt = 0 \tag{4.23}$$

uniformly for t on $[K, \infty)$ from Corollary 1. Thus, from the uniform convergence theorem,

$$\lim_{x\to\infty}\left(\int_K^\infty \frac{1-F(tx)}{1-F(x)}dt\right) = \int_K^\infty \left(\lim_{x\to\infty}\frac{1-F(tx)}{1-F(x)}\right)dt = 0 \tag{4.24}$$

Since both integrals converge to 0 in (4.21), $\lim_{x\to\infty} b^*(x) = 1$.

$\Leftarrow$: Since $\lim_{x\to\infty} b^*(x) = 1$,

$$\lim_{x\to\infty}\left(\int_1^\infty \frac{1-F(tx)}{1-F(x)}dt\right) = 0 \tag{4.25}$$

from Lemma 5. Let $\lambda > 1$ and $x > \overline{x}$. Because the function $t \mapsto \frac{1-F(tx)}{1-F(x)}$ is decreasing,

$$\frac{1-F(\lambda x)}{1-F(x)} < \frac{1-F(tx)}{1-F(x)} \quad \forall t < \lambda \tag{4.26}$$

After integrating both sides with respect to t from 1 to λ,

$$\frac{1-F(\lambda x)}{1-F(x)} < \frac{1}{\lambda-1}\int_1^\lambda \frac{1-F(tx)}{1-F(x)}dt < \frac{1}{\lambda-1}\int_1^\infty \frac{1-F(tx)}{1-F(x)}dt \tag{4.27}$$

because $\forall t : \frac{1-F(tx)}{1-F(x)} \geq 0$. Since the left-hand side is nonnegative, after taking the limit of both sides as $x \to \infty$ and using (4.25) we conclude that $1 - F$ is rapidly varying. □

4.3 Generalized Pareto Curves in Practice

Examples of rapidly varying functions include the normal or exponential distribution. Loosely speaking, such distributions have a thin tail and converge to zero faster than any power law. Since the generalized Pareto coefficient converges to one when $p \to 1$, it can characterize thin-tailed distributions from fat-tailed ones.

Theorem 7 and Theorem 8 divide the class of probability distributions into three categories based on the behavior of the generalized Pareto curve. There are strict power laws for which $b(p) \to c > 1$, thin-tailed distributions for which $b(p) \to 1$, and distributions that are not in the previous two categories. Those distributions may oscillate at an increasingly

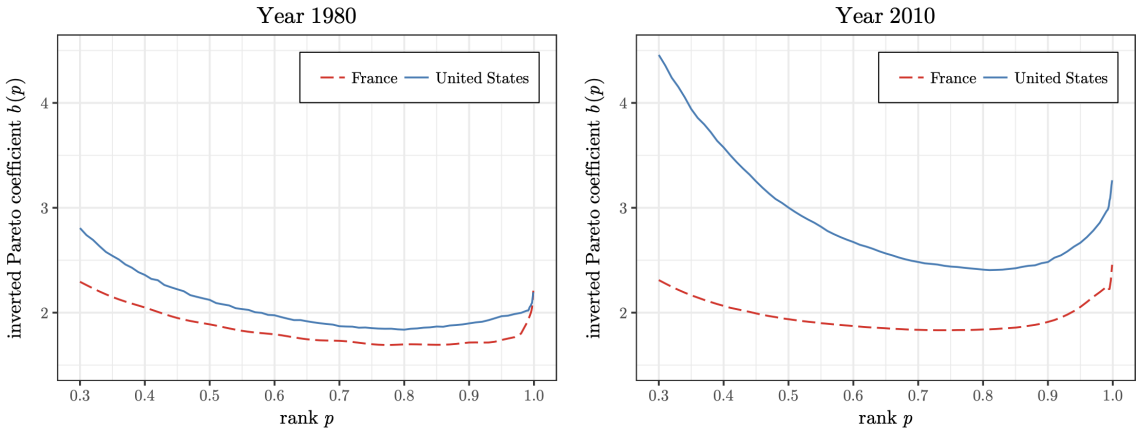

Figure 4.1: Generalized Pareto Curves of pre-tax income in France and the United States in 1980 and 2010, from the paper *Generalized Pareto Curves: Theory and Applications* [1]

fast rate without converging to anything (for example, $b(p) = 3 + \sin(\log(1-p))$ is such a function). However, from a practical point of view, the third category is not relevant, so we are left with two classes of distributions characterized by the limit behavior of $b(p)$.

When X is a strict power law, which means that $b(p)$ is constant, the level of inequality is the same as we move higher in the distribution. For example, the share of the top 20% of the population is the same as the top 2% among the top 20%, which is equal to the share of the top 0.2% among the top 2%. Deviations from this constant result in an unequal share as we move higher. If $b(p)$ increases near $p = 1$ then the top 0.2% among the top 2% gets a larger fraction of income than the top 2% among the top 20%. On the other hand, a decrease in $b(p)$ near $p = 1$ shows a reverse behavior.

As an example, take the levels of pre-tax income in France and the United States over the 1962-2014 period. The generalized Pareto curve has changed a lot more in the United States than in France, which reflects a well-known wealth distribution change in the US: the income distribution among the top is more unequal. In both countries, $b(p)$ converges to a value strictly greater than one, meaning that X is an asymptotic power law. However, the coefficients vary greatly even as p approaches one, so a simple Pareto distribution

could not have captured this information, since in this case $b(p)$ would be constant. Because $b(p)$ increases even as $p \to 1$, income levels are more skewed towards the very top in both countries than the simple Pareto distribution suggests. We can also observe that the curves are U-shaped. This is not specific to these two countries, as we can see a similar pattern in other countries as well. This fact is important since we know that $b(p)$ does not converge monotonically to a constant value as most other models suggest.

In practice, when trying to determine the generalized Pareto curves from a finite sample, a simple plug-in approach does not work. To see why, consider the sample $(X_1, X_2, ..., X_n)$ of i.i.d. copies of X and let X_i^* denote the i-th largest value in the sample (i.e., the i-th order statistic). Then, a natural reformulation of Section 4.2 is

$$\hat{b}_n(p) = \frac{1}{(n - \lfloor np \rfloor) X^*_{\lfloor np \rfloor + 1}} \sum_{k=\lfloor np \rfloor + 1}^{n} X_k^* \tag{4.28}$$

If $\frac{n-1}{n} \leq p < 1$ then $\hat{b}_n(p) = 1$ regardless of the sample. In general, as p gets close to one, the estimator is skewed towards one, which means that we cannot determine the asymptotic value of $b(p)$ solely from the sample. However, in *Generalized Pareto Curves: Theory and Applications* [1] the authors show a more sophisticated method called Pareto Interpolation for determining the generalized Pareto curve from a finite sample.

4.4 Other Pareto Coefficients

We can extend the original definition of $b(p)$ to generate an arbitrary number of Pareto coefficients that describe power law behavior. For this, notice that if $G(x) = 1 - F(x) = Cx^{-\alpha}$ then

$$a_n(x) = -\frac{xG^{(n)}(x)}{G^{(n-1)}(x)} - n + 1 = \alpha \quad \forall n > 0 \tag{4.29}$$

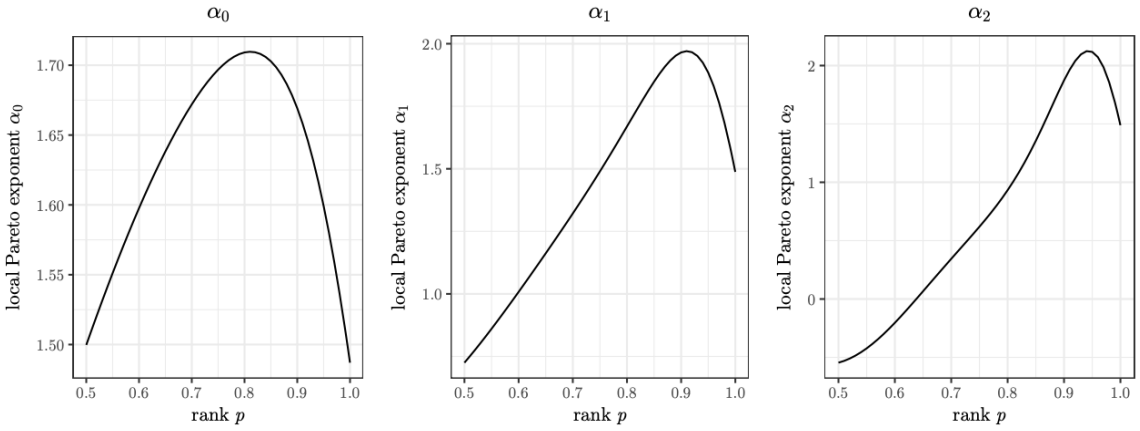

Figure 4.2: Distribution of pre-tax national income using α_0, α_1 and α_2 in the United States, 2010.

For example, $a_1(x) = \frac{xf(x)}{1-F(x)}$ for $n = 1$. We can also generalize this definition and allow negative derivatives, as long as $\alpha > -n + 1$:

$$G^{(n)}(x) = (-1)^n \underbrace{\int_x^\infty ... \int_{t_2}^\infty}_{|n| \text{ times}} G(t_1)dt_1 \; ... \; dt_{|n|} \tag{4.30}$$

We call $a_n(x)$ the *local Pareto coefficient* of order n. It follows from above that

$$a_0(x) = 1 + \frac{x(1 - F(x))}{\int_x^\infty 1 - F(t)dt} \tag{4.31}$$

Hence,

$$b(p) = \frac{a_0(x)}{a_0(x) - 1} \tag{4.32}$$

So we expressed $b(p) = b_0(p)$ in terms of the local Pareto coefficient of order 0. Similarly, we could have defined $b(p)$ as $b_n(p) = \frac{a_n(x)}{a_n(x)-1}$, but this involves estimating successive derivatives, which is difficult. Therefore, only a_0, a_1 and a_2 are used in practice. Figure 4.2 shows the curves of the aforementioned local Pareto coefficients. The traditional a_0 is more U-shaped than the other two, which are more skewed towards the left. However, no matter which one we choose, there is a change of slope as $p \to 1$, which is a major difference to simple power laws.

Conclusion

This thesis examined notable probability distributions in insurance, focusing on Pareto-type distributions and Extreme Value Theory (EVT). Beginning with the classical Pareto distribution, we explored its properties and parameter estimation techniques. In particular, we found that the Method of Moments may estimate a cutoff level larger than some observed data points, excluding part of the dataset. Simulations on real-world insurance data demonstrated the effectiveness of Pareto distributions in modeling claims, though some deviations suggest areas for improvement.

Building on this, we introduced Extreme Value Theory and established key results, including the Fisher-Tippett theorem, which characterizes the asymptotic distribution of extreme order statistics. We then examined the Peaks Over Threshold (POT) method, a widely used approach for modeling excess losses. Simulations confirmed that exceedances above a suitably chosen threshold closely follow a Generalized Pareto Distribution (GPD), reinforcing the practical applicability of EVT-based methods in insurance risk modeling.

The final chapter explored Pareto-type distributions and introduced generalized Pareto curves, which provide a more refined framework for modeling wealth, income, or insurance claims. They offer a better alternative to the classical Pareto model, capturing variations in tail behavior that standard approaches may overlook.

Bibliography

[1] Thomas Blanchet, Juliette Fournier, and Thomas Piketty. *Generalized Pareto Curves: Theory and Applications*. 2017. URL: https://www.amse-aixmarseille.fr/sites/default/files/events/telechargement.pdf.

[2] Mustafa Fatakdawala. *Insurance Claims Fraud Data*. URL: https://www.kaggle.com/datasets/mastmustu/insurance-claims-fraud-data/data?select=insurance_data.csv.

[3] Laurens De Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

[4] Charles I. Jones. "Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality". In: *Journal of Economic Perspectives* (2015). DOI: 10.1257/jep.29.1.29.

[5] Bingham N.H., C.M. Goldie, and J.M. Teugels. *Regular Variation*. Cambridge, 1987.

[6] James Pickands. *Statistical Inference Using Extreme Order Statistics*. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-3/issue-1/Statistical-Inference-Using-Extreme-Order-Statistics/10.1214/aos/1176343003.full.

[7] Max Rydman. *Application of the Peaks-Over-Threshold Method on Insurance Data*. URL: https://www.diva-portal.org/smash/get/diva2:1231783/FULLTEXT01.pdf.

[8] Mette Rytgaard. *Estimation in the Pareto distribution*. 1990. URL: https://www.casact.org/sites/default/files/database/astin_vol20no2_201.pdf.