Notable Distributions in Insurance

Márk Frankli

Eötvös Loránd University, June 2025

Table of Contents

Table of Contents

The Pareto Distribution: Introduction

The Pareto distribution is a probability distribution that is used to describe social, scientific, and actuarial phenomena, among many others.

Several times, it is assumed that insurance claim amounts follow a Pareto distribution.

Our goal is to determine the parameters of the best-fitting Pareto distribution on existing insurance claim datasets.

The Pareto Distribution: Basic Properties

Definition (CDF of Pareto distribution)

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

Expected value and variance

$$\mathbb{E}[X] = \frac{\alpha}{\alpha - 1} c, \ \mathrm{Var}[X] = \frac{\alpha}{(\alpha - 1)^2 (\alpha - 2)} c^2$$

PDF of Pareto distribution

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

The Pareto Distribution: Estimating the Parameters

There are two parameters to estimate: α and c.

A natural way is to use Maximum Likelihood and the Method of Moments for parameter estimation. For now, let's focus on the former.

In practice, the parameter c is usually known in advance but can be chosen as $\hat{c} = \min_i X_i$ from the log-likelihood equation. Solving the log-likelihood equation for α we get the following:

Estimation for the parameter α

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

The Pareto Distribution: Estimating the Parameters

Now, we want to calculate the expected value and variance of $\hat{\alpha}$.

Remark

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

which can be seen from the density function transformation formula. Also,

Remark

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

Therefore, $T = \sum_{i=1}^{n} \log \frac{X_i}{c} \sim \Gamma(n, \alpha)$. $\hat{\alpha} = \frac{n}{T}$, so

Expected value of $\hat{\alpha}$

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

The Pareto Distribution: Estimating the Parameters

From this, we can calculate the variance:

Variance of $\hat{\alpha}$

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

We can improve this estimation by selecting $\alpha^* = \frac{n-1}{T}$, making the estimation unbiased. Next, let's use the Method of Moments to estimate the parameters. First, assume that c is given and $X \sim \mathrm{Pa}(\alpha^0, c)$. Solving $\overline{X} = \mathbb{E}[X]$ gives

First method of moments estimator

$$\alpha^0 = \frac{\bar{X}}{\bar{X} - c}$$

which is the first method of moments estimator.

The Pareto Distribution: Estimating the Parameters

Now, let's assume that both c and α are unknown. If $X \sim \mathrm{Pa}(\alpha, c)$ then

First two moments

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

are the first two moments.

A major difference between the maximum likelihood and the method of moments estimation is the way the parameter c is determined. In the former case, all sample points correspond to a nonzero value, while in the latter, some sample points $X_i < \hat{c}$ may result in an incorrect model. However, as c and n increase, the probability that all of $X_1, \dots, X_n$ are greater than c gets close to 1 (simulation).

The Pareto Distribution: Estimating the Parameters

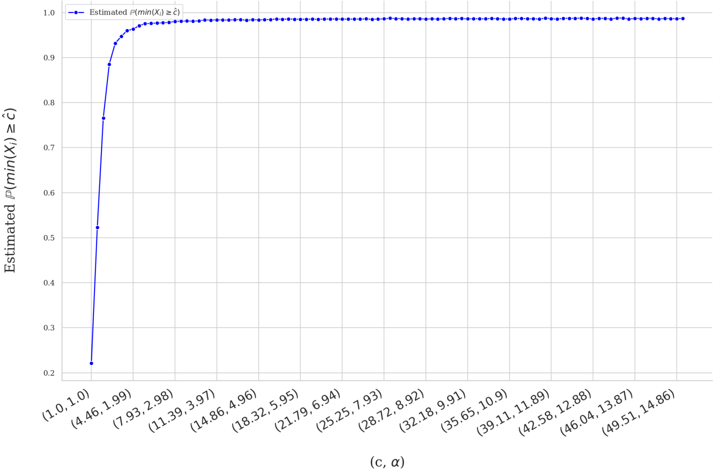

Figure: As c and n increase, the probability that all of $X_1, ..., X_n$ are greater than c gets close to 1.

Table of Contents

Extreme Value Theory: Introduction

The goal is to find 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, $\max(X_1, ..., X_n) \xrightarrow{p} x^*, \; n \to \infty$ since $\mathbb{P}(\max(X_1, ..., X_n) < x) = \mathbb{P}(X_1 < x, ..., X_n < x) = F^n(x)$ and if $\xi_n \xrightarrow{d} c \in \mathbb{R}$ then $\xi_n \xrightarrow{p} c$ for a sequence of random variables $\xi_1, \xi_2, \ldots$.

To avoid a degenerate limit distribution, we introduce real sequences $a_n > 0$ and b_n such that $\frac{\max(X_1,...,X_n)-b_n}{a_n}$ holds with $\lim\limits_{n\to\infty} F^n(a_n x + b_n) = G(x)$ for each continuity point of G, if G is a nondegenerate distribution function.

Definition (extreme value distribution)

G is called an extreme value distribution.

Definition (maximum domain of attraction)

The class of distribution functions F satisfying $\lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$ is called the maximum domain of attraction of G. Notation: $F \in \mathcal{D}(G)$.

Our goal is to find all limit distributions G for which $\lim_{n\to\infty} F^n(a_n x + b_n) = G(x)$ holds.

Extreme Value Theory: The Fisher-Tippet Theorem

The following important theorem classifies extreme value distributions:

Theorem (Fisher and Tippet, 1928)

The class of extreme value distributions is $G_\gamma(ax+b)$ where $a>0$ and

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

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

Furthermore, the following can also be shown to be true:

Lemma

There is a positive function f such that

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

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

Extreme Value Theory: GPD

Suppose that X is a random variable with distribution function F such that $F \in \mathcal{D}(G_\gamma)$. From the previous lemma, we have

Remark

$$\mathbb{P}\left(\frac{X-t}{f(t)} > x \,\middle|\, X > t\right) = \frac{\mathbb{P}\left(X > t \,\middle|\, \frac{X-t}{f(t)} > x\right)\mathbb{P}\left(\frac{X-t}{f(t)} > x\right)}{\mathbb{P}(X > t)}$$

$$= \frac{\mathbb{P}\left(\frac{X-t}{f(t)} > x\right)}{\mathbb{P}(X > t)} = \frac{1 - F(t + f(t)x)}{1 - F(t)}$$

That is, we found the conditional distribution of $\frac{X-t}{f(t)}$, which is

Conditional distribution of $(X-t)/f(t)$

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

Extreme Value Theory: GPD

The previous result motivates the following definition:

Definition

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

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

With this, 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 (approximately).

Table of Contents

POT: Introduction

It's 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.

The POT analysis depends highly on the value of u since, if the value is too large, only a few samples will exceed this threshold, increasing the estimator's variance. 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 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.

Table of Contents

Pareto-type Distributions: 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 $\alpha > 0$.

Definition (inverted Pareto coefficient)

For an income level $x > 0$, the inverted Pareto coefficient is

$$b^*(x) = \frac{1}{(1 - F(x))x} \int_x^\infty zf(z)dz, \ \ b(p) = \frac{1}{(1-p)Q(p)} \int_p^1 Q(z)dz$$

If the expected wealth above \$10 million is \$20 million, then $b^*(10^6) = 2$.

Definition (generalized Pareto curve)

We call the function $b : p \mapsto b(p)$ defined over $[\overline{p}, 1)$ a generalized Pareto curve, where $\overline{p} = F(c)$.

Pareto-type Distributions: Generalized Pareto Curves

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

Definition (asymptotic power law)

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 $\forall \lambda > 0$: $\lim\limits_{x \to \infty} \frac{L(\lambda x)}{L(x)} = 1$.

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

Generalized Pareto Curves can be used as a way to characterize and estimate income and wealth distributions more precisely than the standard Pareto model.