The Pareto distribution: a Heavy-Tailed Distribution

- Pareto distribution

A random variable, X, has a **Pareto distribution** (Type I) if

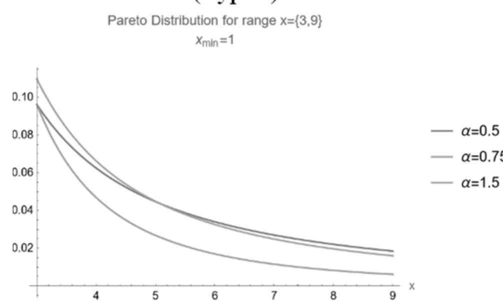
$$f_X(x;\alpha) = \frac{\alpha x_{\min}}{x^{\alpha+1}} \quad \text{for } x > x_{\min}$$
$$= 0 \quad \text{for } x < x_{\min}$$

$$E[X] = \infty \text{ for } \alpha \le 1$$

$$E[X] = \frac{\alpha x_{\min}}{\alpha - 1} \text{ for } \alpha > 1$$

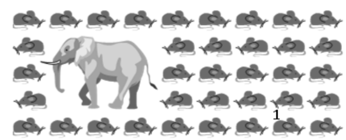
$$\text{Var[X]} = \infty \text{ for } \alpha \le 2$$

$$\text{Var[X]} = \frac{\alpha x_{\min}^2}{(\alpha-1)^2(\alpha-1)} \text{ for } \alpha > 2$$



Heavy-tailed property:

- Large Outliers: These distributions are more likely to produce extreme values compared to light-tailed distributions (like the normal distribution)
- For lower α we have a heavier tail.
- Higher α results in less heavy tail.



1

Bounded-Pareto Distribution

Empirical distributions are always bounded.

The Bounded-Pareto distribution has a Pareto shape but is finite.

<u>Definition</u>: $X \sim BoundedPareto(k, p, \alpha)$,

$$f_X(x) = C \cdot \alpha x^{-\alpha-1}, \qquad k \le x \le p$$

where $0 < \alpha < 2$ and where C is a normalizing constant.

BoundedPareto has similar properties to Pareto:

- near-infinite variance
- heavy-tailed property, assuming upper limit, p, is large.

2

Relationships between pdfs

Rayleigh PDF: The Rayleigh distribution describes the magnitude of a vector whose components are independent and identically distributed (i.i.d.) Gaussian random variables. If X and Y are i.i.d. Gaussian random variables with mean 0 and variance σ^2, the random variable $R = \sqrt{X^2 + Y^2}$ follows a Rayleigh distribution. The PDF of a Rayleigh distribution is given by:

$$f(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad \text{for } r \geq 0$$

Lognormal PDF: A lognormal distribution describes a continuous probability distribution of a random variable whose logarithm is normally distributed. If a random variable Y follows a normal distribution, then $X = e^Y$ follows a lognormal distribution. The PDF of a lognormal distribution is given by:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad \text{for } x > 0$$

Here, μ and σ^2 are the mean and variance of the natural logarithm of X, not X itself.

Chi-Squared Distribution

The chi-squared distribution is a continuous probability distribution that arises in statistics as the distribution of the sum of the squares of k independent standard normal random variables. The PDF of a chi-squared distribution with k degrees of freedom is given by:

$$f(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \quad \text{for } x \geq 0$$

where Γ is the gamma function.

Summation of Bernoulli Trials: The binomial distribution represents the sum of n independent Bernoulli trials. If X_i are n independent Bernoulli random variables each with success probability p, then the sum $Y = X_1 + X_2 + \cdots + X_n$ follows a binomial distribution with parameters n and p:

$$Y \sim \text{Binomial}(n, p)$$

Exponential Distribution

The exponential distribution is a continuous probability distribution that is often used to model the time between events in a Poisson process. The PDF of an exponential distribution with rate parameter λ (which is the inverse of the mean) is given by:

$$f(x; \lambda) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

When $\lambda = 1$, the distribution is called the standard exponential distribution.

Gamma Distribution

The gamma distribution is a continuous probability distribution that generalizes the exponential distribution. It is characterized by two parameters: the shape parameter α (also known as k when referring to it in the context of the sum of exponential random variables) and the rate parameter β (inverse of the scale parameter θ). The PDF of a gamma-distributed random variable X is given by:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad \text{for } x > 0$$

where $\Gamma(\alpha)$ is the gamma function, which generalizes the factorial function.

Sum of Exponential Random Variables

If X_i are n independent exponential random variables with the same rate parameter λ, their sum $Y = X_1 + X_2 + \cdots + X_n$ follows a gamma distribution with shape parameter $\alpha = n$ and rate parameter $\beta = \lambda$.

$$Y \sim \text{Gamma}(n, \lambda)$$