Analysis Of Statistical Distributions
Abstract

Anjali Susan Oommen

This paper contains an analysis of notable statistical distributions with major focus given to the Pareto Distribution as it is a very commonly used fat tailed distribution. This analysis was undertaken in the course of doing my honours project, "What are statistical fat tails and their consequences?". This project is being undertaken in the final year of BSc Mathematics, Azim Premji University, Bangalore, India and is being done under the guidance of Professor Shailaja D Sharma.

Analysis Of Statistical Distributions

Anjali Susan Oommen

This analysis of statistical distributions was written as part of my honours project, "What are statistical fat tails and their consequences?". According to Nassim Nicholas Taleb, distributions can be divided into 3 categories on the basis of their tails : Fat Tails, Semi Thick Tails and Thin Tails. (Taleb N.N ,2021)

An important example of fat tailed distribution is the power law family. In my research done so far, I have found 5 characterising properties for fat tails.

1. A fat-tailed distribution is a probability distribution that exhibits a large skewness or kurtosis, relative to that of either a normal distribution or an exponential distribution. (Wikipedia)

2. In a fat-tailed distribution, a small number of observations in a given dataset will represent the bulk of the statistical properties. Remote events happen less often but they command much greater respect. (Taleb N.N ,2021)

3. A distribution is said to be fat tailed if the ratio of standard deviation to mean absolute deviation is greater than 1.5. (Taleb N.N ,2021)

4. Some authors reserve the term "fat tail" to mean the subclass of heavy tailed distributions that exhibit power law decay behavior as well as infinite variance. Heavy tailed distributions are those that have heavier tails than the exponential distribution. (Glen S)

5. Fat tailed distributions either don't follow the Law of Large Numbers or follow it much slower than the Gaussian distribution does. (Gaussian taken as the standard). (Taleb N.N ,2021)

Semi thick tails are those subexponential distributions which are not power laws. A subexponential distribution is one whose tail decreases slower than any exponential tail. An example of such a distribution is $f(x) = e^{-x^{1/2}}$. It is shown in Fig 1. All subexponential distributions can be found to lie above the exponential distribution, $f(x) = e^{-x}$ on the graph preasymptotically.
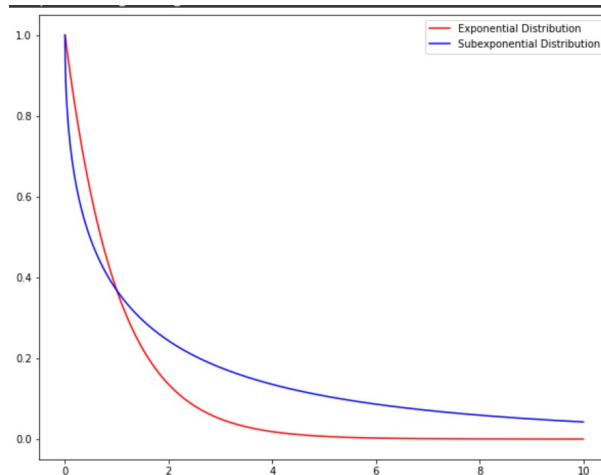

Fig 1

The thin tailed distributions are the Normal Distribution, Bernoulli Distribution and Binomial Distribution. They all have almost 0 probability in their tails.

In order to better understand what these categories mean, I have put together a few important distributions and looked at their properties (such as moments which I derived in a few cases). The focus of the project will be fat tailed distributions so a more detailed study has been done on the Pareto Distribution which is one the most widely used fat tailed distributions in today's world. Taleb's heirarchy of the distributions can be found in Table 1.

Subexponential	Power Laws	Pareto Distribution	**Fat Tailed Distributions**
		Student-T Distribution	
		Levy Distribution	
		Cauchy Distribution	
		Logarithmic Distribution	**Semi Thick Tailed Distributions**
		Gamma Family– eg Exponential Distribution	
Gaussian		Gaussian Distribution	**Thin Tailed Distributions**
		Binomial Distribution	
		Bernoulli Distribution	

Table 1

In sections 1 - 11 , the properties of these distributions as well as a few other notable ones have been looked at. Section 12 of this paper includes the derivation of the moments of the Power Law Family of Distributions as these are a class of fat tailed distributions. Section 13 has a few insights of mine that I had when I did the analysis.

1   Normal Distribution

Parameters	$\mu \in \mathbb{R}$ (mean) (location parameter) $\sigma^2 \in \mathbb{R}_{>0}$ (variance)
Support	$x \in \mathbb{R}$
Probability Density Function	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Cumulative Distribution Function	$\frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Mean	μ
Variance	σ^2
Median	μ
Mode	μ
Moment Generating Function	$\exp(\mu t + \frac{\mu^2 t^2}{2})$
Skewness	0
Excess Kurtosis	$3(\sigma^4 - 1)$

Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve. While normal

distributions are symmetric, but not all symmetric distributions are normal. In a normal distribution, mean = median = mode. It is also unimodal in nature. It belongs to the stable family of distributions. (Wikipedia)

It is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ is the mean of the distribution and σ is the standard deviation.

For a normal distribution, 68 percent of the observations are within $+/-$ one standard deviation of the mean, 95 percent are within $+/-$ two standard deviations, and 99.7 percent are within $+/-$ three standard deviations.

1.1 Standard Normal Distribution

In a standard normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3. Every normal distribution can be standardised. A normal dist is a standard normal distribution stretched by a the standard deviation of the normal distribution and translated by the mean of the normal distribution. The general probability density function of a standard normal distribution is

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

A normal distribution can be standardised by giving the observations in the dataset of the distribution a z-score

$$z = \frac{x-\mu}{\sigma}$$

2 Bernoulli Distribution

Parameters	$p \in [0,1]$
	$q = 1 - p$
Support	$k = \{0, 1\}$
Probability Mass Function	p if $k = 1$
	$q = 1 - p$ if $k = 0$
Cumulative Distribution Function	0 if $k < 0$
	1-p if $0 \leq k < 1$
	1 if $k \geq 1$
Mean	p
Variance	pq
Median	0 if $p < 1/2$
	[0,1] if $p = 1/2$
	1 if $p > 1/2$
Mode	0 if $p < 1/2$
	[0,1] if $p = 1/2$
	1 if $p > 1/2$
Moment Generating Function	$q + pe^m$
Skewness	$\frac{q-p}{\sqrt{pq}}$
Excess Kurtosis	$\frac{1-6pq}{pq}$

The Bernoulli distribution (which was named after Swiss mathematician Jacob Bernoulli) is the discrete probability distribution of a random variable which takes only 2 values: 1 and 0.
$P(1) = p$
$P(0) = 1 - p = q$

To find Skewness

$$\mu_3 = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

$$\mu_3 = \frac{p - 3p \cdot pq - p^3}{pq\sqrt{pq}}$$

$$\mu_3 = \frac{p - 3p^2(1-p) - p^3}{pq\sqrt{pq}}$$

$$\mu_3 = \frac{p(1-p^2) - 3p^2(1-p)}{pq\sqrt{pq}}$$

$$\mu_3 = \frac{p(1-p)(1+p) - 3p \cdot p(1-p)}{pq\sqrt{pq}}$$

$$\mu_3 = \frac{p(1-p)(1+p-3p)}{pq\sqrt{pq}}$$

$$\mu_3 = \frac{pq(1-2p)}{pq\sqrt{pq}}$$

$$\mu_3 = \frac{(q-p)}{\sqrt{pq}}$$

To find Kurtosis

$$\mu_4 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}$$

$$\mu_4 = \frac{p - 4p \cdot p + 6p^2 \cdot p - 3p^4}{(pq)^2}$$

$$\mu_4 = \frac{p(1 - 4p + 6p^2 - 3p^3)}{p^2 q^2}$$

$$\mu_4 = \frac{(1 - 4p + 6p^2 - 3p^3)}{pq^2}$$

$$\mu_4 = \frac{(1-p)(3p^2 - 3p + 1)}{pq^2}$$

$$\mu_4 = \frac{(3p^2 - 3p + 1)}{pq}$$

$$\text{Excess Kurtosis} = \frac{(3p^2 - 3p + 1)}{pq} - 3$$

$$\text{Excess Kurtosis} = \frac{3p^2 - 3p + 1 - 3pq}{pq}$$

$$\text{Excess Kurtosis} = \frac{3p(1-q) - 3p + 1 - 3pq}{pq}$$

$$\text{Excess Kurtosis} = \frac{3p - 3pq - 3p + 1 - 3pq}{pq}$$

$$\text{Excess Kurtosis} = \frac{1 - 6pq}{pq}$$

3 Binomial Distribution

Parameters	$n \in \{0, 1, 2, \dots\}$
	$p \in [0, 1]$
	$q = 1 - p$
Support	$k \in \{0, 1 \dots n\}$
Probability Mass Function	$\binom{n}{x} p^x (1-p)^{n-x}$
Cumulative Distribution Function	$\sum_{i=1}^{k} \binom{n}{i} p^i (1-p)^{n-i}$
Mean	np
Variance	$np(1-p)$
Median	$\lfloor np \rfloor$
Mode	$\lfloor (n+1)p \rfloor$
Moment Generating Function	$\sum_{k=0}^{m} \{{m \atop k}\} n^k p^k$
Skewness	$\frac{q-p}{\sqrt{npq}}$
Excess Kurtosis	$\frac{1-6pq}{npq}$

The binomial distribution is a discrete probability distribution. Such a distribution consists of n Bernoulli trials(independent trials which have only 2 possible outputs and every trial has the same probability of success).

If X is a random variable which follows the Binomial distribution, it can be expressed as the following $X \sim Bin(n, p)$ where n : no of trials and p : probability of success.

$$Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

The shape of the graph depends on the values of n and p.

4 Gamma Distribution

Parameters	$k > 0$ (shape)
	$\theta > 0$ (scale)
Support	$x \in (0, \infty)$
Probability Density Function	$\frac{x^{k-1} e^{-x/\theta}}{\theta^k \Gamma(k)}$
Cumulative Distribution Function	$\frac{1}{\Gamma(k)} \gamma(k, \frac{x}{\theta})$
Mean	$k\theta$
Variance	$k\theta^2$
Median	No simple closed form
Mode	$(k-1)\theta$ for $k \geq 1$
	0 for $k < 1$
Moment Generating Function	$(1 - \theta t)^{-k}$ for $t < \frac{1}{\theta}$
Skewness	$\frac{2}{\sqrt{k}}$
Excess Kurtosis	$\frac{6}{k}$

The gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-square distribution are special cases of the gamma distribution. The exponential distribution is a special case of the gamma distribution where $k = 1$ and $\theta = \frac{1}{\lambda}$

4.1 Exponential Distribution

Parameters	$\lambda > 0$
Support	$x \in [0, \infty)$
Probability Density Function	$\lambda e^{-\lambda x}$
Cumulative Distribution Function	$1 - e^{-\lambda x}$
Mean	$1/\lambda$
Variance	$1/\lambda^2$
Median	$\frac{ln2}{\lambda}$
Mode	0
Moment Generating Function	$\frac{\lambda}{\lambda - t}$ for $t < \lambda$
Skewness	2
Excess Kurtosis	6

The exponential distribution is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant rate. It is the continuous analogue of the geometric distribution. The distribution has the memorylessness property.

To find Skewness

$$\mu_3 = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

$$\mu_3 = \frac{\frac{3!}{\lambda^3} - 3\frac{1}{\lambda}\frac{1}{\lambda^2} - \frac{1}{\lambda^3}}{\frac{1}{\lambda^3}}$$

$$\mu_3 = 3! - 3 - 1$$

$$\mu_3 = 2$$

To find Kurtosis

$$\mu_4 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}$$

$$\mu_4 = \frac{\frac{4!}{\lambda^4} - 4\frac{1}{\lambda}\frac{3!}{\lambda^3} + 6\frac{1}{\lambda^2}\frac{2}{\lambda^2} - 3\frac{1}{\lambda^4}}{\frac{1}{\lambda^4}}$$

$$\mu_4 = 24 - 24 + 12 - 3$$

$$\mu_4 = 9$$

$$\text{Excess Kurtosis} = 6$$

5 Logarithmic Distributions

Parameters	$0 < p < 1$
Support	$k \in \{1, 2, 3 \dots\}$
Probability Density Function	$\frac{-1}{\ln(1-p)}\frac{p^k}{k}$
Cumulative Distribution Function	$1 + \frac{B(p;k+1,0)}{ln(1-p)}$
Mean	$\frac{-1}{\ln(1-p)}\frac{p}{1-p}$
Variance	$-\frac{p^2 + p\ln(1-p)}{(1-p)^2(\ln(1-p))^2}$
Median	
Mode	1
Moment Generating Function	$\frac{ln(1-pe^t)}{ln(1-p)}$ for $t < -\ln p$
Skewness	
Excess Kurtosis	

The logarithmic distribution (sometimes known as the Logarithmic Series distribution) is a discrete, positive distribution, peaking at x = 1, with one parameter and a long right tail.

6   Levy Distribution

Parameters	μ location
	$c > 0$ scale
Support	$x \in [\mu, \infty)$
Probability Density Function	$\sqrt{\frac{C}{2\pi}}\, \dfrac{e^{-\frac{e}{2(x-\mu)}}}{(x-\mu)^{3/2}}$
Cumulative Distribution Function	$\mathrm{erfc}\left(\sqrt{\frac{c}{2(x-\mu)}}\right)$
Mean	∞
Variance	∞
Median	$\mu + c/2(erfc^{-1}(1/2))^2$
Mode	$\mu + \frac{c}{3}$
Moment Generating Function	undefined
Skewness	undefined
Excess Kurtosis	undefined

The Lévy distribution,(which was named after Paul Lévy) is a continuous probability distribution for a non-negative random variable. [In spectroscopy, this distribution, with frequency as the dependent variable, is known as a van der Waals profile.] It belongs to the stable family of distributions.

7   Cauchy Distribution

Parameters	$x_0 \in \mathbb{R}$ (location)
	$\gamma > 0$ and $\gamma \in \mathbb{R}$ (scale)
Support	$x \in (-\infty, +\infty)$
Probability Density Function	$\dfrac{1}{\pi\gamma[1+(\frac{x-x_0}{\gamma})^2]}$
Cumulative Distribution Function	$\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma} + \frac{1}{2}\right)$
Mean	Undefined
Variance	Undefined
Median	x_0
Mode	x_0
Moment Generating Function	does not exist
Skewness	undefined
Excess Kurtosis	undefined

The Cauchy distribution (which is named after Augustin Cauchy) is a continuous probability distribution. It is also known, especially among physicists, as the Lorentz distribution. It belongs to the stable family of distributions.

The first known occurrence of the Cauchy distribution was in 1659 when a function with the form of the density function of the Cauchy distribution was studied geometrically by Fermat.It later was known as the witch of Agnesi, after Agnesi included it as an example in her calculus textbook published in 1784. Contrary to common belief, the first explicit analysis of the properties of the Cauchy distribution was published by the French mathematician Poisson in 1824, with Cauchy only becoming associated with it during an academic controversy regarding the importance of it's undefined mean and variance in 1853. (Cauchy Distribution, Wikipedia)

8 Zipfs Law

Parameters	$s \geq 0$ (Real)
	$N \in \{1, 2, 3, 4 \ldots\}$ (Integers)
Support	$k \in \{1, 2 \ldots N\}$
Probability Density Function	$\frac{1}{k^s H_{N,s}}$
Cumulative Distribution Function	$\frac{H_{k,s}}{H_{N,s}}$
Mean	$\frac{H_{N,s-1}}{H_{N,s}}$
Variance	$\frac{H_{N,s-2}}{H_{N,s}} - \frac{H_{N,s-1}^2}{H_{N,s}^2}$
Mode	1
Moment Generating Function	$\frac{1}{H_{N,s}} \sum_{n=1}^{N} \frac{e^{nm}}{n^s}$

Zipf's law is an empirical law formulated using mathematical statistics. It states that for many types of data studied in the physical and social sciences, the rank-frequency distribution is an inverse relation.

Zipf's law then predicts that out of a population of N elements, the normalized frequency of the element of rank k, f(k;s,N), where s is the value of the exponent characterizing the distribution

$$f(k : s, N) = \frac{1/k^s}{\sum_{n=1}^{N}(1/n^s)} = \frac{1}{k^s H_{N,s}}$$

where $H_{N,s}$ is the Nth generalised Harmonic number.

The limit as $N \to \infty$ is finite if $m > 1$, with the generalized harmonic number bounded by and converging to the Riemann zeta function.
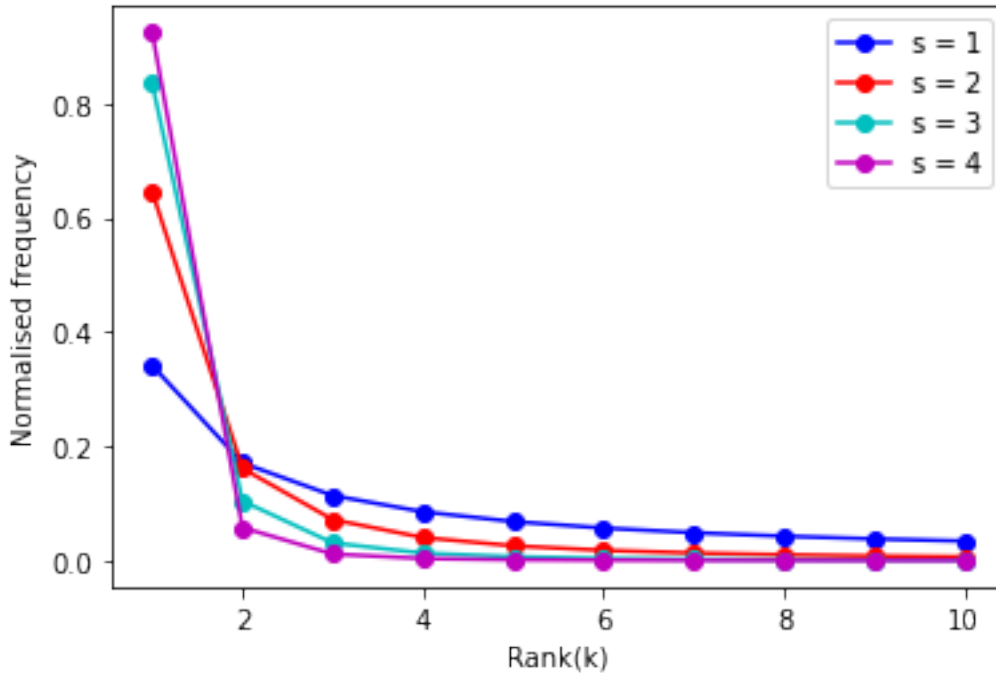
(The Reimann zeta function is generally denoted by $\zeta(s)$. It is a function of a complex variable defined as

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s})$$

As long as the exponent s exceeds 1, it is possible for such a law to hold with infinitely many words, since if $s > 1$,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} < \infty$$

As the exponent, s increases, given a rank, it can be seen below that the normalised frequency of the element decreases.

*** Zipf's law holds if the number of elements with a given frequency is a random variable with power

law distribution, $p(x) = \alpha x^{-1-1/s}$

$***$ The skewness and kurtosis are rarely talked about for this distribution as it comes from an empirical law.

9 Weibull Distribution

Parameters	$\lambda \in (0, +\infty)$ (Scale)
	$k \in +\infty)$ (Shape)
Support	$x \in [0, +\infty)$
Probability Density Function	$\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}e^{-(x/\lambda)^k}$
Cumulative Distribution Function	$1 - e^{-(x/\lambda)^k}$
Mean	$\lambda\Gamma(1 + 1/k)$
Median	$\lambda(ln2)^{1/k}$
Variance	$\lambda^2\left[\Gamma\left(1+\frac{2}{k}\right) - \left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right]$
Mode	$\lambda(\frac{k-1}{k})^{1/k}$
Moment Generating Function	$\sum_n^\infty = 0\frac{t^n\lambda^n}{n!}\Gamma(1+n/k), k \geq 1$
Skewness	$\frac{\Gamma(1+3/k)\lambda^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$
Excess Kurtosis	$\frac{-6\Gamma_1^4 + 12\Gamma_1^2\Gamma_2 - 3\Gamma_2^2 - 4\Gamma_1\Gamma_3 + \Gamma_4}{[\Gamma_2 - \Gamma_1^2]^2}$

The Weibull Distribution is a continuous probability distribution (which was named after Swedish mathematician Waloddi Weibull) used to analyse life data, model failure times and access product reliability. It is a heavy tailed distribution

10 Student T-Distribution

Parameters	$\nu > 0$ (degrees of freedom)
Support	$x \in (-\infty, \infty)$
Probability Density Function	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
Cumulative Distribution Function	$\frac{1}{2} + x\Gamma(\frac{\nu+1}{2})\frac{{}_2F_1(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu})}{\sqrt{pi\nu}\Gamma(\frac{\nu}{2})}$
Mean	0 if $\nu > 1$
	Undefined otherwise
Variance	$\frac{\nu}{\nu-2}$ if $\nu > 2$
	∞ if $1 < \nu \leq 2$
	Undefined otherwise
Median	0
Mode	0
Moment Generating Function	does not exist
Skewness	0 for $\nu > 3$
	undefined otherwise
Excess Kurtosis	**TO BE UNDERSTOOD**

The Student's t-distribution was developed by English statistician William Sealy Gosset under the pseudonym "Student". The t-distribution plays a role in a number of widely used statistical analyses, including Student's t-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis. Student's t-distribution also arises in the Bayesian analysis of data from a normal family. If we take a sample of n observations from a normal distribution, then the t-distribution with $nu = n - 1$ degrees of freedom can be defined as the distribution of the location of the sample mean relative to the true mean, divided by the sample standard deviation, after multiplying by the standardizing term $\sqrt{n}$. In this way, the t-distribution can be used to construct a confidence interval for the true mean.(Student's t-distribution, Wikipedia)

The t-distribution is shaped like the normal distribution (symmetric and bell shaped). However it has heavier tails. The tails become heavier as ν gets less. At $\nu = \infty$, the student t distribution resembles the normal distribution

To find Skewness

$$\mu_3 = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

$E(X^3) = \mu = 0$ when $\nu > 3$
$E(X^3)$ does not exist when $\nu < 3$

$$\mu_3 = 0 \text{ for } \nu > 3 \text{ ,undefined o/w}$$

To find Kurtosis

$$\mu_4 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}$$

$E(X^4)$ does not exist when $\nu < 4$

11 Pareto Distribution

Parameters	$x_m > 0$ (scale)
	$\alpha > 0$ (shape)
Support	$x \in [x_m, \infty)$
Probability Density Function	$\frac{\alpha x_m^\alpha}{x^{\alpha+1}}$ if $x \geq x_m$
	0 if $x < x_m$
Cumulative Distribution Function	$1 - \left(\frac{x_m}{x}\right)^\alpha$
Mean	∞ if $\alpha \leq 1$
	$\frac{\alpha x_m}{\alpha - 1}$ if $\alpha > 1$
Variance	∞ if $\alpha \in (1, 2]$
	$\left(\frac{x_m}{\alpha-1}\right)^2 \frac{\alpha}{\alpha-2}$ if $\alpha > 2$
Median	$x_m \sqrt[\alpha]{2}$
Mode	x_m
Moment Generating Function	does not exist
Skewness	$\frac{2(1+\alpha)}{(\alpha-3)} \sqrt{\frac{\alpha-2}{\alpha}}$
Excess Kurtosis	$\frac{6(\alpha^3 + \alpha^2 - 6\alpha - 2)}{\alpha(\alpha-3)(\alpha-4)}$

The Pareto distribution is one of the most commonly found distributions in the real world. (It was first used to describe the distribution of wealth i.e. how a small percentage of the population hold a large percentage of the wealth. Now it is used in description of social, quality control, scientific, geophysical, actuarial, and many other types of observable phenomena) It is a type of power law which is named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto. It is a fat tailed distribution like all other power laws. The most widely known form of the Pareto distribution is that which follows the Pareto principle or "80-20 rule" which states that 80 percent of outcomes are due to 20 percent of causes. Only Pareto distributions with shape value (α) of $log_4 5 \approx 1.16$ precisely reflect it. Empirical observation has shown that this 80-20 distribution fits a lot of phenomena, both natural ones and human activities.

Mean

$$E(X) = \int_{x_{min}}^{\infty} x f(x) dx$$

$$E(X) = \int_{x_{min}}^{\infty} x \frac{\alpha x_{min}^\alpha}{x^{\alpha+1}} dx$$

$$E(X) = \alpha x_{min}^\alpha \int_{x_{min}}^{\infty} x^{1-\alpha-1} dx$$

$$E(X) = \alpha x_{min}^\alpha \left[\frac{x^{-\alpha+1}}{-\alpha + 1}\right]_{x_{min}}^{\infty} dx$$

$$E(X) = \frac{\alpha x_{min}^\alpha x_{min}^{-\alpha+1}}{\alpha - 1}$$

$$E(X) = \frac{\alpha x_{min}}{\alpha - 1}$$

If $\alpha > 1$, $E(X) = \frac{\alpha x_{min}}{\alpha - 1}$

If $\alpha \leq 1$, $E(X) = \infty$

<u>Variance</u>

$Var(x) = E(X^2) - (E(X))^2$

$$E(X^2) = \int_{x_{min}}^{\infty} x^2 f(x) dx$$

$$= \int_{x_{min}}^{\infty} x^2 \frac{\alpha x_{min}^{\alpha}}{x^{\alpha+1}} dx$$

$$= \alpha x_{min}^{\alpha} \int_{x_{min}}^{\infty} x^{2-\alpha-1}$$

$$= \alpha x_{min}^{\alpha} \int_{x_{min}}^{\infty} x^{1-\alpha}$$

$$= \alpha x_{min}^{\alpha} \left[\frac{x^{1-\alpha+1}}{1-\alpha+1} \right]_{x_{min}}^{\infty}$$

$$= \alpha x_{min}^{\alpha} \frac{x_{min}^{2-\alpha}}{\alpha - 2}$$

$$= x_{min}^2 \frac{\alpha}{\alpha - 2} \text{ when } \alpha > 2$$

$$Var(x) = x_{min}^2 \frac{\alpha}{\alpha - 2} - \left(\frac{\alpha x_{min}}{\alpha - 1} \right)^2$$

$$= x_{min}^2 \left(\frac{\alpha}{\alpha - 2} - \frac{\alpha^2}{(\alpha - 1)^2} \right)$$

$$= x_{min}^2 \left(\frac{\alpha(\alpha - 1)^2 - \alpha^2(\alpha - 2)}{(\alpha - 2)(\alpha - 1)^2} \right)$$

$$= x_{min}^2 \alpha \left(\frac{(\alpha - 1)^2 - \alpha(\alpha - 2)}{(\alpha - 2)(\alpha - 1)^2} \right)$$

$$= x_{min}^2 \alpha \left(\frac{(\alpha^2 + 1 - 2\alpha - \alpha^2 + 2\alpha)}{(\alpha - 2)(\alpha - 1)^2} \right)$$

$$= x_{min}^2 \alpha \left(\frac{1}{(\alpha - 2)(\alpha - 1)^2} \right)$$

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

$$Var(x) = \infty \text{ when } \alpha \in (1, 2]$$

<u>Higher Moments</u>

$$E(X^m) = \int_{x_{min}}^{\infty} x^m f(x) dx$$

$$E(X^m) = \int_{x_{min}}^{\infty} x^m \frac{\alpha x_{min}^{\alpha}}{x^{\alpha+1}} dx$$

$$E(X^m) = \alpha x_{min}^{\alpha} \int_{x_{min}}^{\infty} x^{m-\alpha-1} dx$$

$$E(X^m) = \alpha x_{min}^{\alpha} \left[\frac{x^{m-\alpha-1+1}}{m-\alpha-1+1} \right]_{x_{min}}^{\infty}$$

$$E(X^m) = \alpha x_{min}^{\alpha} \left[\frac{x^{m-\alpha}}{m-\alpha} \right]_{x_{min}}^{\infty}$$

$$E(X^m) = \alpha x_{min}^{\alpha} \left[\frac{x_{min}^{m-\alpha}}{\alpha - m} \right]$$

$$E(X^m) = x_{min}^m \frac{\alpha}{\alpha - m}$$

If $n < \alpha$, $E(X^m) = x_{min}^m \frac{\alpha}{\alpha-m}$.

If $n \geq \alpha$, $E(X^m) = \infty$

.

To find Skewness

$$\mu_3 = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}$$

$$\mu_3 = \frac{\frac{\alpha x_m^3}{\alpha-3} - 3\left(\frac{\alpha x_m}{\alpha-1}\right)\frac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)} - \left(\frac{\alpha x_m}{\alpha-1}\right)^3}{\frac{x_m^2 \alpha}{(\alpha-1)^2(\alpha-2)}\frac{x_m}{(\alpha-1)}\sqrt{\frac{\alpha}{\alpha-2}}}$$

$$\mu_3 = \frac{\frac{\alpha x_m^3}{\alpha-3} - 3\frac{\alpha^2 x_m^3}{(\alpha-1)^3(\alpha-2)} - \frac{\alpha^3 x_m^3}{(\alpha-1)^3}}{\frac{x_m^3 \alpha}{(\alpha-1)^3(\alpha-2)}\sqrt{\frac{\alpha}{\alpha-2}}}$$

$$\mu_3 = \frac{(\alpha-1)^3(\alpha-2) - 3\alpha(\alpha-3) - \alpha^2(\alpha-2)(\alpha-3)}{(\alpha-3)\sqrt{\frac{\alpha}{\alpha-2}}}$$

$$\mu_3 = \frac{\alpha^4 - 3\alpha^3 + 3\alpha^2 - \alpha - 2\alpha^3 + 6\alpha^2 - 6\alpha + 2 - 3\alpha^2 + 9\alpha - \alpha^4 + 2\alpha^3 + 3\alpha^3 - 6\alpha^2}{(\alpha-3)\sqrt{\frac{\alpha}{\alpha-2}}}$$

$$\mu_3 = \frac{2(\alpha+1)}{\alpha-3}\sqrt{\frac{\alpha-2}{\alpha}}$$

To find Kurtosis

$$\mu_4 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}$$

$$\mu_4 = \frac{\frac{\alpha x_m^4}{\alpha-4} - \frac{4\alpha x_m}{\alpha-1}\frac{\alpha x_m^3}{(\alpha-3)} + 6\left(\frac{\alpha x_m}{\alpha-1}\right)^2\frac{\alpha x_m^2}{(\alpha-2)} - 3\left(\frac{\alpha x_m}{\alpha-1}\right)^4}{\frac{x_m^4 \alpha^2}{(\alpha-1)^4(\alpha-2)^2}}$$

$$\mu_4 = \frac{\frac{\alpha x_m^4}{\alpha-4} - \frac{4\alpha^2 x_m^4}{(\alpha-1)(\alpha-3)} + 6\frac{\alpha^3 x_m^4}{(\alpha-1)^2(\alpha-2)} - 3\frac{\alpha^4 x_m^4}{(\alpha-1)^4}}{\frac{x_m^4 \alpha^2}{(\alpha-1)^4(\alpha-2)^2}}$$

$$\mu_4 = \frac{(\alpha-1)^4(\alpha-2)(\alpha-3) - 4\alpha(\alpha-1)^3(\alpha-2)(\alpha-4) + 6\alpha^2(\alpha-1)^2(\alpha-3)(\alpha-4) - 3\alpha^3(\alpha-2)(\alpha-3)(\alpha-4)}{\frac{\alpha(\alpha-3)(\alpha-4)}{(\alpha-2)}}$$

$$\mu_4 = \frac{(9\alpha^2 + 3\alpha + 6)(\alpha-2)}{\alpha(\alpha-3)(\alpha-4)} \text{when } \alpha > 4$$

Excess Kurtosis

$$\text{Excess kurtosis} = \frac{(9\alpha^2 + 3\alpha + 6)(\alpha-2)}{\alpha(\alpha-3)(\alpha-4)} - 3$$

$$\text{Excess kurtosis} = \frac{9\alpha^3 - 15\alpha^2 - 12 - 3\alpha(\alpha-3)(\alpha-4)}{\alpha(\alpha-3)(\alpha-4)}$$

$$\text{Excess kurtosis} = \frac{9\alpha^3 - 15\alpha^2 - 12 - 3\alpha^3 + 21\alpha^2 - 36\alpha}{\alpha(\alpha-3)(\alpha-4)}$$

$$\text{Excess kurtosis} = \frac{6(\alpha^3 + \alpha^2 - 6\alpha - 2)}{\alpha(\alpha-3)(\alpha-4)} \text{when } \alpha > 4$$

12   Power Law

$f(x) = kx^{-\alpha}$

Finding the value of k

$$\int_{x_{min}}^{\infty} f(x)dx = 1$$

$$\int_{x_{min}}^{\infty} kx^{-\alpha}dx = 1$$

$$k \int_{x_{min}}^{\infty} x^{-\alpha}dx = 1$$

$$k \left[\frac{x^{-\alpha+1}}{-\alpha+1} \right]_{x_{min}}^{\infty} = 1$$

$$k \frac{x_{min}^{-\alpha+1}}{\alpha-1} = 1$$

$$k = \frac{\alpha-1}{x_{min}^{1-\alpha}}$$

Mean
A well defined mean exists $\in [x_{min}, \infty)$ iff $\alpha \geq 2$

$$E(X) = \int_{x_{min}}^{\infty} x f(x)dx$$

$$= k \int_{x_{min}}^{\infty} x \frac{1}{x^{\alpha}}dx$$

$$= k \int_{x_{min}}^{\infty} x^{1-\alpha}dx$$

$$= \frac{\alpha-1}{x_{min}^{1-\alpha}} \left[\frac{x^{2-\alpha}}{2-\alpha} \right]_{x_{min}}^{\infty}$$

$$= \frac{\alpha-1}{x_{min}^{1-\alpha}} \left[\frac{x_{min}^{2-\alpha}}{\alpha-2} \right]$$

$$= \frac{\alpha-1}{\alpha-2} x_{min}$$

If $\alpha = 2$

$$E(X) = k \int_{x_{min}}^{\infty} \frac{1}{x}dx$$

$log(\infty)$ does not exist, $\therefore \alpha \neq 2$

If $\alpha < 2$, the integral does not converge and hence does not exist.

<u>Variance</u>

$$Var(x) = E(X^2) - (E(X))^2$$

$$E(X)^2 = \int_{x_{min}}^{\infty} x^2 f(x) dx$$

$$= \int_{x_{min}}^{\infty} x^2 \frac{\alpha - 1}{x_{min}^{1-\alpha}} x^{-\alpha} dx$$

$$= \frac{\alpha - 1}{x_{min}^{1-\alpha}} \int_{x_{min}}^{\infty} x^{2-\alpha} dx$$

$$= \frac{\alpha - 1}{x_{min}^{1-\alpha}} \left[\frac{x^{2-\alpha+1}}{2 - \alpha + 1} \right]_{x_{min}}^{\infty}$$

$$= \frac{\alpha - 1}{x_{min}^{1-\alpha}} \left[\frac{x^{3-\alpha}}{3 - \alpha} \right]_{x_{min}}^{\infty}$$

$$= \frac{\alpha - 1}{x_{min}^{1-\alpha}} \left[\frac{x_{min}^{3-\alpha}}{\alpha - 3} \right]$$

$$= \frac{\alpha - 1}{\alpha - 3} x_{min}^{3-\alpha-1+\alpha}$$

$$= \frac{\alpha - 1}{\alpha - 3} x_{min}^2$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$Var(X) = \frac{\alpha - 1}{\alpha - 3} x_{min}^2 - (\frac{\alpha - 1}{\alpha - 2} x_{min})^2$$

$$Var(X) = x_{min}^2 \left(\frac{\alpha - 1}{\alpha - 3} - (\frac{\alpha - 1}{\alpha - 2})^2 \right)$$

<u>Higher Moments</u>

$$E(X^m) = \int_{x_{min}}^{\infty} x^m f(x) dx$$

$$E(X^m) = k \int_{x_{min}}^{\infty} x^m x^{-\alpha} dx$$

$$E(X^m) = \frac{\alpha - 1}{x_{min}^{1-\alpha}} \int_{x_{min}}^{\infty} x^{m-\alpha} dx$$

$$E(X^m) = \frac{\alpha - 1}{x_{min}^{1-\alpha}} \left[\frac{x^{m-\alpha+1}}{m - \alpha + 1} \right]_{x_{min}}^{\infty}$$

$$E(X^m) = \frac{\alpha - 1}{x_{min}^{1-\alpha}} \left[\frac{x_{min}^{-\alpha+1} x_{min}^m}{\alpha - m - 1} \right]$$

$$E(X^m) = \frac{\alpha - 1}{\alpha - m - 1} x_{min}^m$$

13 Insights

When we talk about fat tailed distributions we are generally most concerned with those distributions (mostly power laws) which have tail exponents from 1 to 3. As the tail exponent gets higher, the distribution becomes less and less fat tailed.Below is a tabulation of power law distributions which have pdf $= kx^{-\alpha}$ with different tail exponents.It can be seen that as the exponent gets greater, the tails become thinner.

Tail exponent,α	(0,1]	(1,2]	(2, 3)	[3, ∞)
	Extreme case of fat tailed distribution, (rarely considered)	Commonly used Fat tailed distributions	Commonly used Fat tailed distributions	Tail is thinner than tail of exponential distribution
	Infinite mean, skewness, kurtosis, variance	Defined Mean and Infinite Variance, skewness and kurtosis	Infinite skewness, Infinite kurtosis but defined mean and variance	The m'th moment is defined when $m > \alpha$

Below is a tabulation of distributions where $f(x) = e^{-x^{\beta}}$ where β is the exponent of x. It can be seen that as the exponent gets greater, the tails become thinner.

Exponent of x,β	(0,1]	1	(1,2)	2	(2, ∞)
	Sub-exponential Distributions	Exponential Distribution	Fatter tail than normal distribution but thinner than tail of exponential distribution	Normal Distribution	Extremely thin tailed / No tail Distribution
	lies above the exponential distribution preasymptotically on the graph		Lies between the exponential distribution and the normal distribution preasymptotically on the graph		

It can be seen on comparing the two tabulations that tails of distributions of the form $e^{-x^{\beta}}$ become thinner faster than tails of distributions of the form $kx^{-\alpha}$.

Bibliography

1. Wikipedia (2022) Fat-tailed Distribution. Available at: https://en.wikipedia.org/wiki/Fat-tailed_distribution

2. Taleb, N.N. (2022) 'MINI-LESSON 1: Breaking down intuitively the concept of standard deviation. Why pple don't get it.'. N N Taleb's Probability Moocs, . Available at: https://www.youtube.com/watch?v=oMl-SbuQUYc&list=PLMV8UXQuOWKPAIjvnyMN2317LHF3ydvnG&index=8 (Accessed: date).

3. Taleb, N.N. (2022) 'MINI-LESSON 8: Power Laws(maximally simplified'. N N Taleb's Probability Moocs, . Available at: https://www.youtube.com/watch?v=oMl-SbuQUYc&list=PLMV8UXQuOWKPAIjvnyMN2 index=8 (Accessed: date).

4. Taleb, N.N. (2022) 'MINI-LESSON 2: Fat Tails, a Very, Very Introductory Presentation.'. N N Taleb's Probability Moocs, . Available at: https://www.youtube.com/watch?v=t7Fr6iGhmBM&list=PLMV8UXQuOWKPAIjvnyMN2317LHF3ydvnG&index=2 (Accessed: date).

5. Taleb, N.N. (2022) 'MINI-LESSON 3: The Law of Large Numbers. A very intuitive introduction.'. N N Taleb's Probability Moocs, . Available at: https://www.youtube.com/watch?v=zyBXNIskQK0&list=PLMV8UXQuOWKPAIjvnyMN2317LHF3ydvnG&index=3 (Accessed: date).

6. Glen S. "Fat Tail Distribution: Definition, Examples" StatisticsHowTo.com: Elementary Statistics for the rest of us! Available at: https://www.statisticshowto.com/fat-tail-distribution/

7. Wikipedia (2022) Normal Distribution. Available at: https://en.wikipedia.org/wiki/Normal_distribution (Accessed: date)

8. Wikipedia (2022) Binomial Distribution. Available at: https://en.wikipedia.org/wiki/Binomial_distribution (Accessed: date)

9. Wikipedia (2022) Bernoulli Distribution. Available at: https://en.wikipedia.org/wiki/Bernoulli_distribution (Accessed: date)

10. Wikipedia (2022) Gamma Distribution. Available at: https://en.wikipedia.org/wiki/Gamma_distribution (Accessed: date)

11. Wikipedia (2022) Exponential Distribution. Available at: https://en.wikipedia.org/wiki/Exponential_distribution (Accessed: date)

12. Wikipedia (2022) Logarithmic Distribution. Available at: https://en.wikipedia.org/wiki/Logarithmic_distribution (Accessed: date)

13. Wikipedia (2022) Cauchy Distribution. Available at: https://en.wikipedia.org/wiki/Cauchy_distribution (Accessed: date)

14. Wikipedia (2022) Levy Distribution. Available at: https://en.wikipedia.org/wiki/L%C3%A9vy_distribution (Accessed: date)

15. Wikipedia (2022) Student's t-distribution. Available at: https://en.wikipedia.org/wiki/Student%27s_t-distribution (Accessed: date)

16. Wikipedia (2022) Pareto Distribution. Available at: https://en.wikipedia.org/wiki/Pareto_distribution (Accessed: date)

17. Wikipedia (2022) Power Law. Available at: https://en.wikipedia.org/wiki/Power_law (Accessed: date)

18. Evans, M Hastings, Nicholas Peacock, Brian. (2000). Statistical Distributions, Third Edition. Measurement Science and Technology. Available at: http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/201_06_Forbes_-Statistical-Distributions-Fourth-Edition-Wiley-2011.pdf