Fat-Tailed Distributions and their Statistical Consequences

by

Anjali Susan Oommen

A dissertation submitted in partial fulfilment of the requirements for the degree of

B.Sc Honours in Mathematics

2020-2023

Supervised by: Shailaja Sharma
Evaluated by: Ajaykumar K

Fat-Tailed Distributions and their Statistical Consequences

Anjali Susan Oommen

Abstract

Fat Tailed Distributions rose to prominence in 2007 after the publication of the book titled The Black Swan by Nassim N Taleb, although their prevalence had already been pointed out much earlier by B. Mandelbrot. Despite the rising use of fat tailed distributions in the modern world, there is not much accessible literature on such distributions. The Law Of Large Numbers and the Central Limit Theorem, which are central requirements for statistical analysis in thin-tailed distributions, fail to hold in the case of fat tailed distributions. This leads to many consequences that require empirical methods of analysis.Through my thesis, I aim to understand what characterizes a fat tailed distribution and how they manifest in empirical data. I study the differences between such distributions and thin tailed distributions from their distributional properties and also based on their pre-asymptotic statistical properties using simulations. I discuss a metric based on the empirical rate of convergence of partial sums as a measure of the "fat tailedness" of a distribution. Finally, I critically examine the application of fat tailed distributions in network models and climate science models in recently published academic papers.

The case studies showed that phenomena may exhibit both fat-tailed and thin-tailed behaviour and sometimes maybe best modelled by mixtures of classical distributions. It was also concluded that the fat-tailedness behaviour aggregate distributions of business network structures is dependent not just on the proportion of fat-tailed distributions present but also on the difference between the tail exponents of the thin-tailed and fat-tailed distributions.

Dedication

I dedicate this to my late grandfather who always believed in me. Through his unwavering love and support, he did his best to make me believe that I could do anything I put my mind to.

Declaration

I hereby declare that the work in this thesis has been carried out by me, in the B. Sc (Honours) Program, under the supervision of Shailaja Sharma, and in the partial fulfillment of the requirements for the award of the degree of B. Sc (Honours) at the Azim Premji University, Bangalore. I further declare that this work has not been the basis for the award of any degree, diploma or any other title elsewhere.

Acknowledgements

I would like to express my gratitude to my advisor, Shailaja Sharma for the guidance she provided me without which this endeavour would not have been possible. Her expertise and encouragement helped me complete this research and write this thesis. Additionally I would like to express my appreciation for the math faculty of the university for their support and feedback through this project.

I would also like to acknowledge my family and friends, especially my mother for the moral support they provided through the course of the project. Without their encouragement and motivation, I would have not been able to complete this journey.

Contents

Chapter 1

Introduction

The Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) are two of the most important theorems in probability theory and statistics as they provide the basis for statistical inference. Empirical studies in various fields rely upon the convergence of observed sample means to the hypothetical true mean (LLN) and on the normality of accrued measurement errors (CLT). Both LLN and CLT require that the mean and variance exist for the population from which samples are taken. What happens when the mean or variance don't exist or are infinite? Ever since Mandelbrot pointed out the presence of fat-tailed distributions in his work on stock prices (c.1960), it has been clear that other methods are required [1]. More work on these ideas has been subsequently done to extend the theoretical as well as practical understanding of such phenomena.

In classical statistics, the distributions generally considered for modelling purposes are the normal distribution, binomial distribution, exponential/ gamma distributions – all of which are "thin-tailed distributions". The class of distributions called fat-tailed distributions are required to model or understand such phenomena. These distributions often do not have a defined mean or variance and even when they do, the sample size required to estimate these measures are very large. In more recent times, these distributions have received greater attention after the publica-

[1] Mandelbrot saw financial markets as an example of "wild randomness". In his early work, he found that the price changes in financial markets did not follow a Gaussian distribution, but rather Lévy stable distributions having infinite variance.

tion of, The Black Swan: The Impact of the Highly Improbable, (Taleb 2008). The distribution of wealth was what first aroused interest in such distributions in the 1890s with the work of the Italian economist Vilfredo Pareto. A naive classification of the different scenarios which map to the use of thin and fat-tailed distributions may be as respectively "natural" and "anthropogenic" or human-constructed - city traffic, wealth distribution and climate change being in the latter class. NN Taleb presents a more detailed view of two universes: "Mediocristan" where thin-tailed distributions preside and "Extremistan" where fat-tailed distributions are dominant (Taleb 2008).

In order to be able to identify whether a distribution is fat-tailed, we must first know what characterises a fat-tailed distribution. We do this by analysing the properties of known theoretical distributions, both fat-tailed and thin-tailed. A combination of heuristics, mathematical models and statistical computations are required for the purpose of identifying and working with data coming from fat-tailed phenomena. These methods help us recognise fat-tailedness in empirical data. The statistical behaviours we are concerned with are pre-asymptotic in nature, in order to be able to use them on empirical data and make practical decisions concerning such data.

We mainly refer to Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications (Technical Incerto), NN Taleb in our discussion of fat-tailed distributions. However, we note at the outset that the materials presented in the above book(Taleb 2022)are dense, whereas his popular writing is heavily metaphorical(Taleb 2004,Taleb 2008). In the present thesis, we try to provide the bridge between the two.

In what follows, we describe the important statistical distributions, along with the theoretical properties that establish a hierarchy among them in increasing degree of mathematical fat-tailedness. We state a set of heuristics for recognizing fat-tailedness and analyse an empirical metric proposed by Taleb. We test these methods with reference to selected economic models in climate data and business network analysis.

Chapter 2

Convergence of the Mean of *n* Summands

Consider a stream of data, i.e. a sequence of random numbers generated by a process. The behaviour of the mean of n summands is described by the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT). The LLN and the CLT together form the basis of inferential statistics and underlies scientific research methods across disciplinary boundaries.

2.1 Law Of Large Numbers

The Law of Large Numbers states that if an experiment is performed a large number of times, with the trials being independent, then the average value of the outcome will be close to the true (or expected) value. The Law has two forms - the strong and the weak form.

2.1.1 Weak Law Of Large Numbers

Let X_1,$X_2 \dots X_n$ be independent and identically distributed random variables with a finite expected value i.e $\mathbb{E}(X_i) = \mu < \infty$. Then $\forall \epsilon > 0$

$$\lim_{n\to\infty} P(|\bar{X_n} - \mu| \geq \epsilon) = 0$$

2.1.2 Strong Law Of Large Numbers

Let X_1,$X_2 \dots X_n$ be independent and identically distributed random variables with a finite expected value i.e $\mathbb{E}(X_i) = \mu < \infty$. Then with probability 1,

$$\bar{X_n} \to \mu \text{ as } n \to \infty$$

As seen above, the weak law discusses the convergence of the probability to 0 that the sample mean and population mean have a difference close to 0 whereas the strong law refers to almost sure convergence of the sample mean to the population mean. While the almost sure convergence implies that probability of the difference between the sample mean and the true population being 0 is 0, the converse is not true.

If we consider a large number of experiments, the weak law states that the error between the sample mean and the population may always exist but these errors occur less as the sample size increases. However the strong law tells us that given a large enough sample size, the sample size will give us the true population mean (no error).

In this thesis, we are concerned with the distributions that do not follow the weak law of large numbers or take much longer to converge under the weak law. The sample means of such distributions also definitely do not converge under the strong law of large numbers. As the decrease in size and frequency of errors either does not happen or takes much longer than it would for a thin tailed distribution, the properties of these distributions are different and need to be to studied in order to understand how to work with such distributions when encountered in the real world.

2.2 Central Limit Theorem

The central limit theorem states that if you have a population with finite mean μ and finite standard deviation σ and take sufficiently large random samples from the population with replacement , then the distribution of the sample means will be approximately normally distributed i.e.

$$\lim_{n\to\infty} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$$

The proof of the Central Limit Theorem uses the first and second derivative of the moment generating function i.e. the first and second raw moments. If the variance is finite, the second moment is also finite. Thus the finiteness of the mean and variance are prerequisites for the Central Limit Theorem. (Proof is given in Appendix)

The above 2 theorems imply that if the experiment is repeated a large number of times i.e. large enough sample is taken, we will be able to find the mean of the population. It can be seen that both these theorems hold only if the population has a finite mean and for the Central Limit Theorem to hold, it must also have a finite variance(Prerequisites of both theorems). So if the population has no finite mean, we can make no assumptions that, in finite time, a large enough sample size can give the true population mean.

2.3 Mediocristan vs Extremistan

Taleb 2008 describes two empirical scenarios using the metaphorical expressions: "Mediocristan" and "Extremistan"[1] He uses these metaphors to describe two different types of dominant statistical behaviours, which we explain below.

2.3.1 Mediocristan

Mediocristan according to Taleb 2008 is a world ruled by the following law: "When your sample is large, no single instance will significantly change the aggregate or total". Physical quantities (such as the heights and weights of people) are generally modelled by the distributions that behave as per this law. When the Law of Large Numbers and the Central Limit Theorem hold, the population can be modelled by such ("thin-tailed") distributions (thin and fat-tailed distributions will be properly described in a following chapter on the hierarchy of distributions). As a result, it is possible to make predictions from what is observed using the classical methods

[1] The choice of terminology points to his own ethnic origins in the Middle East (Lebanon).

of statistical inference, which use the Normal distribution (for example, regression analysis). All moments are defined for these distributions. In Mediocristan, there is mild randomness i.e. not much deviation from the average. Due to the low deviation from the average, the standard deviation can be used to measure variability in the distribution.

2.3.2 Extremistan

In Extremistan, one single extreme observation can have a disproportionate impact on the distributional parameters (Taleb 2008). The Law of Large Numbers and Central Limit Theorem do not hold in the corresponding distributional models. In such distributions, a few extreme events determine the statistical properties of the entire distribution. Anthropogenic events such as wealth, income and book sales are often modelled by distributions of this type. Not all of the moments of such distributions are defined. In Extremistan, there is said to be high randomness i.e. some events have a high deviation from the average. As the deviation from the mean is squared for standard deviation, the tail events in such distributions are overweighed and the standard deviation can be exceedingly large. For some distributions of this kind, the second moment is undefined or infinite which implies that the standard deviation cannot be found.

In his popular book "The Black Swan", Taleb describes a 'black swan' event as an event that comes as a surprise and has a huge (usually negative) impact (Taleb 2008). Such events while seemingly unexpected can be rationalised in hindsight (by considering the statistical properties of the distribution). When inferences are made based on samples taken from a fat-tailed distribution, black swan events may occur.

Chapter 3

Analysis Of Distributions

3.1 Hierarchy Of Distributions

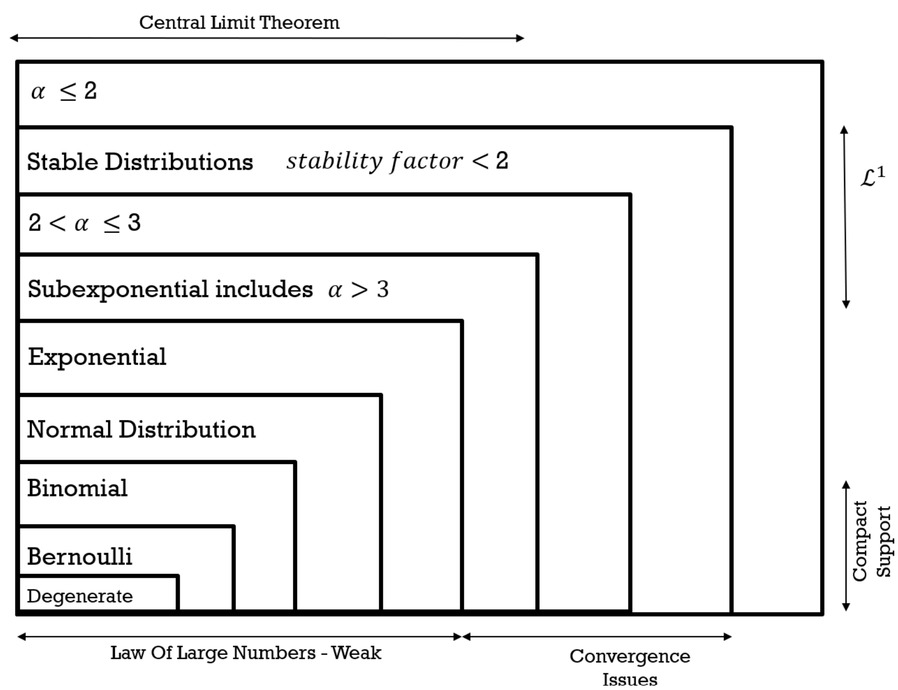


Figure 3.1: Hierarchy Of Distributions

The table in Fig 3.1 is based on the hierarchy of distributions as presented on pg 85, Fooled by Randomness (Taleb 2004). [1] [2] Recall the fat-tail heuristics described in Chapter 2. This table shows the hierarchy based on convergence of the sample mean to the true mean under LLN and CLT. The distributions found at the bottom of the table (degenerate distribution, Bernoulli Distribution, Binomial distribution, Normal Distribution) are all thin-tailed distributions. Degenerate distribution corresponds to the absence of randomness. The distributions just above it are discrete distributions which converge in distribution to the Normal when the sample size is large enough.

Above these lie an intermediate category of distributions, the semi-thick tails (such as the family of gamma distributions) and finally at the top are the fat-tailed distributions (such as the power law class of distributions). The Weak Law holds for the thin-tailed distributions and for some semi-thick tailed distributions(like the exponential distribution) i.e. all distributions where there exists a well defined mean. Beyond these distributions, the sample mean does not converge to the true population mean or the sample size must be much bigger for the sample mean to converge. The Central Limit Theorem holds for all distributions with defined variance and mean. $\mathcal{L}\infty$ refers to the norm-1 or linear measure of deviation, i.e. the class of distributions where the mean absolute deviation (a linear measure) is a better measure of variability than the variance (a quadratic measure) (this is further discussed subsection 3.4.1).

3.1.1 Categories of distribution on the basis of tails

Another representation of the above hierarchy is presented in the following table, Fig 3.2 by Taleb in the video lecture, "MINI-LESSON 8: Power Laws (maximally simplified)" *N N taleb's probability moocs* n.d. Here, we see the three distinct categories clearly spelt out.

A sub-exponential distribution is one whose tail decreases slower than any exponential tail (Goldie and Klüppelberg 1998). An example of such a distribution is $f(x) = e^{-x^{1/2}}$. It is shown in Fig 3.3. All subexponential distributions can be found to lie above the exponential distribution, $f(x) = e^{-x}$ on the graph pre-asymptotically. Semi-thick tails are those sub-exponential distributions which are not power laws.

[1] $\alpha =$ tail exponent of the power law class of the distributions

[2] When stability factor of stable distributions is less than 2, they have undefined variance

<table>
<tr><td rowspan="6">Subexponential</td><td rowspan="4">Power Laws</td><td>Pareto Distribution</td><td rowspan="4">Fat Tailed Distributions</td></tr>
<tr><td>Student-T Distribution</td></tr>
<tr><td>Levy Distribution</td></tr>
<tr><td>Cauchy Distribution</td></tr>
<tr><td rowspan="2"></td><td>Logarithmic Distribution</td><td rowspan="2">Semi Thick Tailed Distributions</td></tr>
<tr><td>Gamma Family–
eg Exponential Distribution</td></tr>
<tr><td rowspan="3">Gaussian</td><td rowspan="3"></td><td>Gaussian Distribution</td><td rowspan="3">Thin Tailed Distributions</td></tr>
<tr><td>Binomial Distribution</td></tr>
<tr><td>Bernoulli Distribution</td></tr>
</table>

Figure 3.2: Caption

The thin-tailed distributions are the Normal Distribution, Bernoulli Distribution and Binomial Distribution. They all have almost 0 probability in their tails. More specifically, they belong to the class of distributions that converge under the Central Limit Theorem to the Gaussian basin.

3.2 Understanding Classical Distributions

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 have derived in a few cases). As the focus of the project is fat-tailed distributions, a more detailed study has been done on the Pareto Distribution which is one the most widely used fat-tailed distributions.

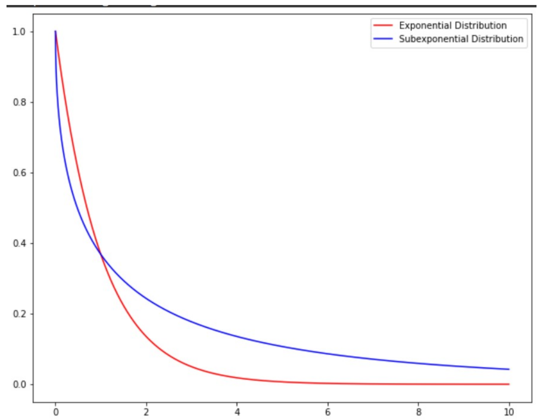


Figure 3.3: Understanding the exponential vs sub-exponential distributions

3.2.1 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	$\frac{1}{2}\left[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})\right]$
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, the 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. (*Probability distribution* 2023)

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.

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

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

3.2.2 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) (*Probability distribution* 2023) 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

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

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

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

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

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

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

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

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

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

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

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

3.2.3 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	$\sum_{k=0}^{m} \left\{ {m \atop k} \right\} n^{\underline{k}} p^k$
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) (*Probability distribution* 2023).

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.

3.2.4 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 (*Probability distribution* 2023). The exponential distribution is a special case of the gamma distribution where $k = 1$ and $\theta = \frac{1}{\lambda}$

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 (*Probability distribution* 2023). 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

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

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

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

$$\mu_4 = 9$$

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

3.2.5 Logarithmic Distributions

Parameters	$0 < p < 1$
Support	$k \in \{1, 2, 3 \ldots\}$
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 $t < -\ln p$

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 (*Logarithmic distribution* 2020).

3.2.6 Levy Distribution

Parameters	μ location $c > 0$ scale
Support	$x \in [\mu, \infty)$
Probability Density Function	$\sqrt{\frac{C}{2\pi}} \frac{e^{-\frac{e}{2(x-\mu)}}}{(x-\mu)^{3/2}}$
Cumulative Distribution Function	$\text{erfc}\left(\sqrt{\frac{c}{2(x-\mu)}}\right)$
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.(*Probability distribution* 2023)] It belongs to the stable family of distributions.

3.2.7 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	$\frac{1}{\pi\gamma[1+(\frac{x-x_0}{\gamma})^2]}$
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. (*Probability distribution* 2023)

3.2.8 Zipfs Distribution

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	$\frac{H_{N,s-2}}{H_{N,s}} - \frac{H^2_{N,s-1}}{H^2_{N,s}}$
Mode	1
Moment Generating Function	$\frac{1}{H_{N,s}} \sum_{n=1}^{N} \frac{e^{nm}}{n^s}$

The Zipf's distribution comes from the Zipf's Law which was originally formulated in terms of quantitative linguistics(*Probability distribution* 2023). 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.

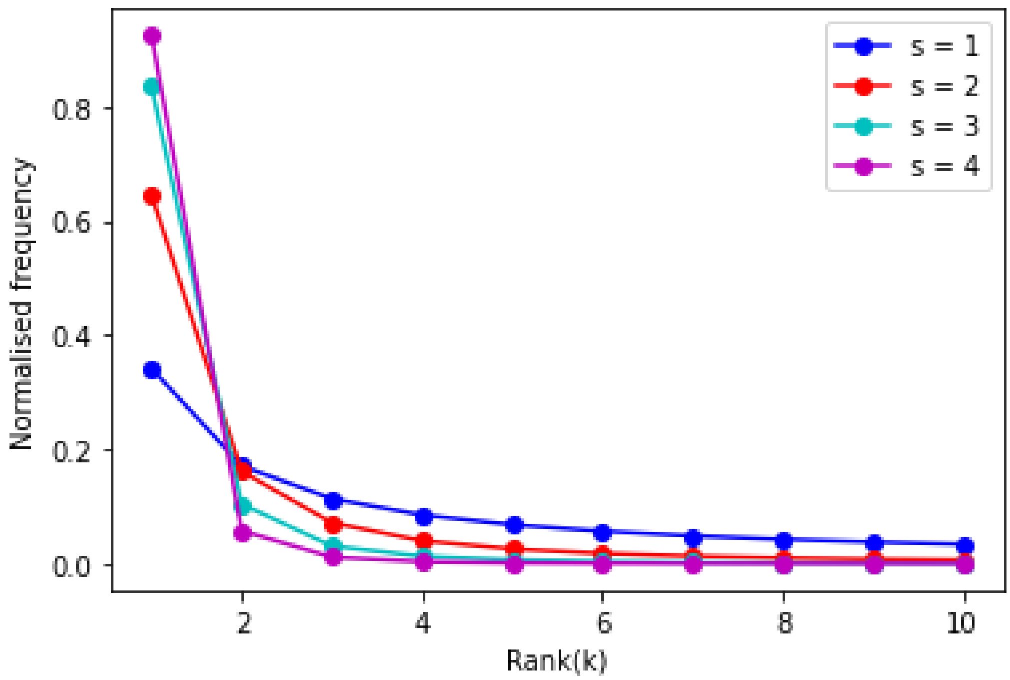
The limit as $N \to \infty$ is finite if $m > 1$, with the generalized harmonic number bounded by and converging to the Riemann zeta function.
(The Reimann zeta function is generally denoted by $\zeta(s)$. It is a function of a complex variable defined as

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

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.

3.2.9 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 (*Probability distribution* 2023) used to analyse life data, model failure times and access product reliability. It is a heavy tailed distribution

3.2.10 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	undefined if $\nu \leq 2$ ∞ for $2 < \nu \leq 4$ $\frac{6}{\nu-4}$ for $\nu > 4$

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 analysis, 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.(*Probability* *distribution* 2023)

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 $\nu < 4$

3.2.11 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	$1 - (\frac{x_m}{x})^\alpha$
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 (*Probability distribution* 2023). 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

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

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

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

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

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

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

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

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

Variance

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Higher Moments

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

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

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

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

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

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

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

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

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

.

To find Skewness

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

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

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

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

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

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

To find Kurtosis

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

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

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

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

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

Excess Kurtosis

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

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

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

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

3.3 Understanding the Power Law class of Distributions

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

<u>Finding the value of k</u>

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

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

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

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

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

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

<u>Mean</u>

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

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

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

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

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

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

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

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>

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

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

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

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

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

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

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

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

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

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

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

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

Higher Moments

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

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

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

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

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

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

3.4 Insights

3.4.1 Mean Absolute Deviation vs Standard Deviation

Above we see that for fat tailed distributions, higher moments need not always be defined. This proves to be an issue as we often cannot measure the variability of a fat-tailed distribution. This issue was identified by Taleb in his study of fat tailed distributions. In fact he has vehemently campaigned against the use of standard deviation in statistics. He claims that it should be replaced by mean absolute deviation. (*N N taleb's probability moocs* n.d.) The reasons for this are listed below

- Mean absolute deviation corresponds to real life much better.
- In standard deviation as we square the terms we give more weight to terms with large deviation and less weight to terms with small deviation, so we are overweighing the tail events.

In a Gaussian model, the ratio of standard deviation to mean absolute deviation is around 1.25. However if we look at a power law, the above ratio tends to infinity.

- As the mean is defined for more fat-tailed distributions than the variance has been defined, using the mean absolute deviation would be a more standard measure.
- Standard deviation is not additive while mean absolute deviation is

3.4.2 Understanding tail exponents

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 starts becoming thinner and at $\alpha = 4$ it becomes thinner than 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 $kx^{-\alpha}$.

Chapter 4

Recognizing Fat Tailed Distributions in Empirical Data

There are several definitions that pertain to the amount of probability in the tail region of the distribution. We come across terms such as "fat-tailed", "thick-tailed", "long-tailed", "semi-thick-tailed" and so on, in the literature on distributional tails. Here, we restrict our attention to only the fat-tailed distributions that belong to the Power Law class of distributions. These distributions have been defined and detailed in the previous chapter(Analysis of Distributions).

There are other ways to study fat-tailed distributions. In extreme value theory, one uses extremal values i.e. taking the probability of exceeding a maximum value, adjusted by the scale to make comparisons. One uses this value to classify a distribution as fat-tailed or thin-tailed. This measure is concerned with the behaviour of maxima or extreme values and thus is concerned with asymptotic measures. however in this project, following Taleb (Taleb 2019) , we focus on pre-asymptotic behaviour and pre-asymptotic measurement of fat-tailedness. We are here more concerned with the behaviour of averages rather than extreme values.

The following pre-asymptotic heuristics characterise fat-tailed distributions. Several of these heuristics are reported by Taleb in his video lecture, "MINI-LESSON 2: Fat Tails, a Very, Very Introductory Presentation." (*N N taleb's probability moocs* n.d.). We interpret and illustrate them with examples in this thesis.

1. A fat-tailed distribution exhibits a large skewness or kurtosis, relative to that of either a normal distribution or an exponential distribution.(**wikipedia˙2023˙ftd**) In extreme cases, skewness, kurtosis or other higher moments may not even be defined for such a distribution. The most commonly seen fat-tailed distribution, the Pareto distribution does not have skewness and kurtosis defined for tail exponent, $\alpha < 3$

2. In a fat-tailed distribution, a small number of observations will represent the bulk of the statistical properties. Remote events are not frequent, but they "command respect"(*N N taleb's probability moocs* n.d.) in the sense that they have a large impact. Here, impacts refer to the consequences upon the physical/ empirical system which shows fat-tailed behaviour. We will not concern ourselves with impacts, as much as try to understand how the remote events control the distribution. An example in this regard is the Pareto Distribution.

3. A distribution is said to be fat-tailed if the ratio of standard deviation to mean absolute deviation is greater than 1.5. (*N N taleb's probability moocs* n.d.) This is a purely empirical thumb-rule. The ratio of standard deviation to mean absolute deviation for a thin-tailed distribution is around 1,25

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 2021)

5. Fat-tailed distributions either don't follow the Law of Large Numbers or follow it much slower than the Gaussian distribution does. The rate of convergence of the mean of the Gaussian (i.e.$\sqrt{n}$)is taken as the standard. The sample means taken from such distributions do not form a normal distribution or a very large sample size is needed to form the normal distributions i.e. do follow the Central Limit Theorem or need a much larger sample size as compared to a normal distribution to follow the Central Limit Theorem. (*N N taleb's probability moocs* n.d.)

Here we note that the convergence of the mean of a sequence of i.i.d. r.v. is the same as the reduction of variance of the mean. The rate of reduction of the variance of the mean is $\sqrt{n}$, where n is the length of the sequence (This will be discussed in more detail in Chapter 5).

4.1 Simulated Demonstration Of the Heuristics

The following simulations show the difference between normal distributions and fat-tailed distributions using one of the most common cases of fat-tailed distributions - the Pareto Distribution. The code for these simulations can be found in the repository -Fat-Tailed-Distributions on GitHub [1]

1. The normal distribution has a skewness of 0 and an excess kurtosis of 0. Using simulation it was found that the skewness and excess kurtosis of a sample of size 1000 taken from a Pareto distribution with tail exponent 4 is widely varying and can even take values such as 8 and 115 respectively.

2. The following graph (Figure 4.1) shows the ratio of std deviation to mean absolute deviation for samples of size 200 taken from normal distributions (red) and Pareto distributions with tail exponent - 1.14 (blue). The x-axis are samples labelled 1-1000 and y-axis shows the ratio of standard deviation to mean absolute deviation for the samples. It is clear to see that the ratio is a little greater than 1 for samples taken from Normal distributions and can range from around 1.5 to around 7 for samples taken from the Pareto distribution

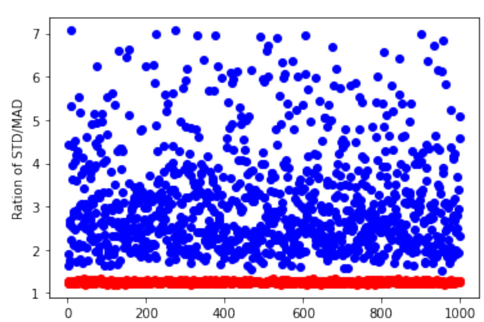


Figure 4.1: Ratio of Std Deviation to Mean Absolute Deviation

[1] The page also contains code to understand the properties of different fat tailed distributions

3. The following figure (Figure 4.2a) shows the variance of 1000 samples of size 200 taken from a Pareto distribution with tail exponent 1.14. It variance varies from 2.509 to 94018439.75. Figure 4.2b shows the variance of 1000 samples of size 200 taken from a Pareto distribution with tail exponent 1.14 (in blue) (same samples as those taken in figure 4.2a) and from the Standard Normal distribution (in red) but the y axis limit is set as -1 to 200.

(a) Pareto Distribution

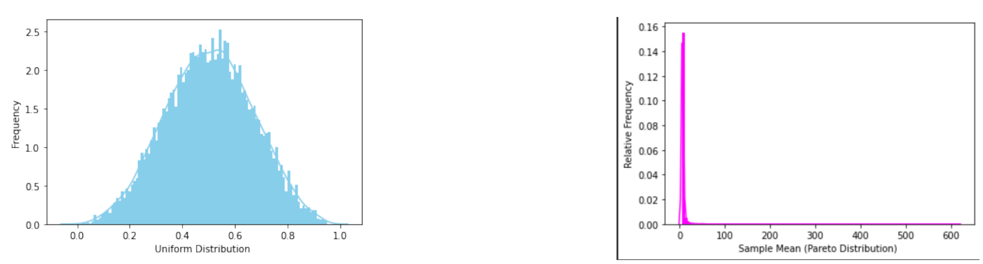
(b) Pareto Distribution(blue) and Standard Normal Distribution(red)

Figure 4.2: Variance of 1000 samples of size 200

4. **Central Limit Theorem**

   Fig 4.3 (a) shows the distribution of sample means for the uniform distribution. It can be seen that this looks like a normal distribution when the sample size is 3.

   Fig 4.3 (b) shows the distribution of sample means for the Pareto distribution with tail exponent 1.14. It can be seen that this does not look like a normal distribution even when the sample size is large (10,000).



(a) Uniform Distribution

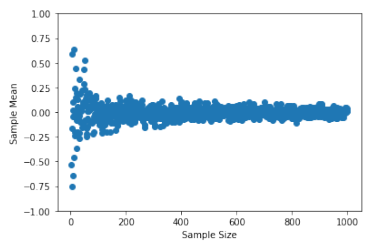
(b) Pareto Distribution

Figure 4.3: Distribution of Sample Means for fixed sample size

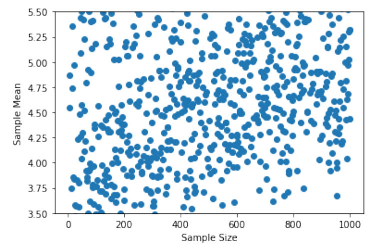
Law Of Large Numbers

Fig 4.4 (a) has the sample means on the y axis and the sample sizes on the x axis for the standard normal distribution. It can be seen that as the sample size increases, the variation of the sample mean decreases..

Fig 4.4 (b) has the sample means on the y axis and the sample sizes on the x axis for the Pareto distribution with tail exponent 1.14. It can be seen that as the sample size increases, there is a still high variation of the mean .



(a) Normal Distribution



(b) Pareto Distribution

Figure 4.4: Distribution of Sample Means as sample size increases

Chapter 5

The Kappa Metric

[1] Given two fat-tailed distributions, how does one tell which is more fat-tailed than the other? For example, how do we compare a infinite variance Pareto Distribution and a Student T Distribution? Conventional measures of fat-tailedness are the tail exponent (used when looking at power law distributions) and kurtosis (for distributions with a defined fourth moment). However the tail exponent can only be used to make comparisons within the power law class of distributions and most fat-tailed distributions do not have moments defined higher than the first moment which means kurtosis cannot be found for them which makes comparisons using this impossible. As explained earlier, extremal value would not be a good metric to measure fat-tailedness and thus to make comparisons.

Taleb proposed an operational metric for univariate, unimodal probability distributions with finite first moment in [0, 1] where 0 is maximally thin-tailed (Gaussian - used as the benchmark) and 1 is maximally fat-tailed. This metric, simply called the Kappa Metric has the following applications:

1. to make comparisons across different classes of univariate distributions when both have a well defined mean
2. to assess the sample size n needed for statistical significance outside the Gaussian,

[1] The contents of this chapter are based on - How much data do you need? An operational, pre-asymptotic metric for fat-tailedness, Taleb 2019

3. to measure the speed of convergence to the Gaussian (or stable basin)

4. Allows comparison of n-summed variables of different distributions for a given number of summands , or same distribution for different n, and assess the pre-asymptotic properties of a given distributions.

5. Provides a measure of the distance from the limiting distribution, namely the Lévy α-Stable basin (of which the Gaussian is a special case).

The metric has been constructed using the following characterising property of fat-tailed distributions: Fat-tailed Distributions follow the Law of Large Numbers at a much slower rate than thin-tailed distributions. It aims to show "how standard is standard, and measures the exact departure from the standard from the standpoint of statistical significance".

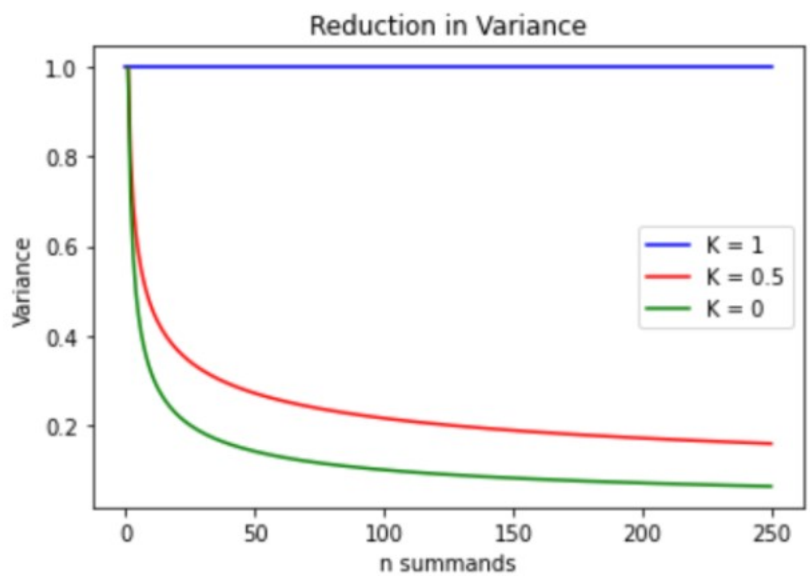


Figure 5.1: Reduction in Variance for n summands for different values of κ

Definition Let $X_1, \ldots, X_n$ be i.i.d random variables with finite mean, that is $\mathbb{E}(X) < +\infty$. Let $S_n = X_1 + X_2 + \cdots + X_n$ be a partial sum. Let $\mathbb{M}(n) = E(|Sn - E(Sn)|)$ be the expected mean absolute deviation from the mean for n summands. Define the rate of convergence for n additional summands starting with n_0:

$$\log \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \log \left(\frac{n}{n_0}\right)^{\frac{1}{2-\kappa_{n_0,n}}}$$

From the above equation, we get the formula for the kappa metric

$$\kappa_{n_0,n} = 2 - \frac{\log n - \log n_0}{\log \mathbb{M}(n) - \log \mathbb{M}(n_0)}$$

where $n > n_0 \geq 1$ ** For baseline values $n = n_0 + 1$, the shorthand κ_{n_0} is used

As a simple heuristic, the higher the κ, the more fat-tailed a distribution is. For any value of κ above 0.15, the "normally approximation" is highly unreliable

5.1 Understanding the Formula

5.1.1 Terms

$\mathbb{M}(n)$: Mean absolute Deviation of S_n i.e. $\mathbb{M}(n) = \mathbb{E}(|S_n - \mathbb{E}(S_n)|)$
$\frac{\mathbb{M}(n)}{n}$: Mean absolute Deviation of $\frac{S_n}{n}$ i.e. $\frac{\mathbb{M}(n)}{n} = \mathbb{E}(|\frac{S_n}{n} - \mathbb{E}(\frac{S_n}{n})|)$
$\frac{\mathbb{M}(n)}{n} = \mathbb{E}(|\frac{1}{n}S_n - \frac{1}{n}\mathbb{E}(s_n)|)$
$\frac{\mathbb{M}(n)}{n} = \frac{1}{n}\mathbb{E}(|S_n - \mathbb{E}(s_n)|)$
σ^2 : Variance of $S_N = X_1 + X_2 + \cdots + X_N$ where N is the size of the population

5.1.2 For a Normal Distribution

The Normal Distribution is a thin-tailed distribution so the Law of Large Numbers holds i.e. the sample mean converges to the true population mean as the sample size n increases. This means that the variance of the sample mean decreases with increase in n. We shall consider the mean absolute deviation analogous to the variance. As discussed earlier the mean absolute deviation is more robust so we shall proceed using that.

Consider mean absolute deviation of partial sum, S_n. We know the variance of S_n is proportional to $\sqrt{n}$. Thus

$$\mathbb{M}(n) \propto (\sqrt{n})$$

$$\frac{\mathbb{M}(n)}{n} \propto \frac{\sqrt{n}}{n}$$

$$\frac{\mathbb{M}(n)}{n} \div \frac{\mathbb{M}(n_0)}{n_0} = \frac{\sqrt{n}}{n} \div \frac{\sqrt{n_0}}{n_0}$$

$$\frac{\mathbb{M}(n)/n}{\mathbb{M}(n_0)/n_0} = \frac{\sqrt{n}/n}{\sqrt{n_0}/n_0}$$

$$\frac{\mathbb{M}(n)/n}{\mathbb{M}(n_0)/n_0} = \frac{n^{-1/2}}{n_0^{-1/2}}$$

$$\log \frac{\mathbb{M}(n)/n}{\mathbb{M}(n_0)/n_0} = \log \frac{n^{-1/2}}{n_0^{-1/2}}$$

$$\log \mathbb{M}(n)/n - \log \mathbb{M}(n_0)/n_0 = \log n^{-1/2} - \log n_0^{-1/2}$$

$$\log \mathbb{M}(n) - \log n - \log \mathbb{M}(n_0) + \log n_0 = -\frac{1}{2} \log n + \frac{1}{2} \log n_0$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \log n - \frac{1}{2} \log n + \frac{1}{2} \log n_0 - \log n_0$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \frac{1}{2} \log n - \frac{1}{2} \log n_0$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \log n^{1/2} - \log n_0^{1/2}$$

$$\log \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \log \frac{n^{1/2}}{n_0^{1/2}}$$

$$\frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left(\frac{n}{n_0}\right)^{1/2}$$

5.1.3 Other Distributions

For fat-tailed distributions, the Law of Large numbers take much longer to hold. This means that the variance of the sample mean for such distributions does not decrease at the same rate as it does for thin-tailed distributions. It decreases at a lower rate. As we have set Gaussian to be the benchmark, we shall look at the deviation of rate of decrease in variance from that of normal distributions.

$$\mathbb{M}(n) \propto n^{\frac{1}{2-\kappa}}$$

where κ is called the kappa metric and it shows the deviation from the normal distribution.

$$\frac{\mathbb{M}(n)}{n} \propto \frac{n^{\frac{1}{2-\kappa}}}{n}$$

$$\frac{\mathbb{M}(n)}{n} \propto n^{\frac{1}{2-\kappa}-1}$$

$$\frac{\mathbb{M}(n)}{n} \propto n^{\frac{1-2+\kappa}{2-\kappa}}$$

$$\frac{\mathbb{M}(n)}{n} \propto n^{\frac{\kappa-1}{2-\kappa}}$$

$$\frac{\mathbb{M}(n)/n}{\mathbb{M}(n_0)/n_0} = \frac{n^{\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}}}}{n_0^{\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}}}}$$

$$\log \frac{\mathbb{M}(n)/n}{\mathbb{M}(n_0)/n_0} = \log \frac{n^{\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}}}}{n_0^{\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}}}}$$

$$\log\left(\mathbb{M}(n)/n\right) - \log\left(\mathbb{M}(n_0)/n_0\right) = \log n^{\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}}} - \log n_0^{\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}}}$$

$$\log \mathbb{M}(n) - \log n - \log \mathbb{M}(n_0) + \log n_0 = \frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}} \log n - \frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}} \log n_0$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}} \log n + \log n - \frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}} \log n_0 - \log n_0$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \log n \Big(\frac{\kappa_{n_0,n}-1}{2-\kappa_{n_0,n}} + 1\Big) - \log n_0 \Big(\frac{\_n_0, nK - 1}{2-\kappa_{n_0,n}} + 1\Big)$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \log n \Big(\frac{\kappa_{n_0,n}-1+2-\kappa_{n_0,n}}{2-\kappa_{n_0,n}}\Big) - \log n_0 \Big(\frac{\kappa_{n_0,n}-1+2-\kappa_{n_0,n}}{2-\kappa_{n_0,n}}\Big)$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \log n \Big(\frac{1}{2-\kappa_{n_0,n}}\Big) - \log n_0 \Big(\frac{1}{2-\kappa_{n_0,n}}\Big)$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \frac{1}{2-\kappa_{n_0,n}} (\log n - \log n_0)$$

$$\log \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \frac{1}{2-\kappa_{n_0,n}}(\log \frac{n}{n_0})$$

$$\log \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \log \left(\frac{n}{n_0}\right)^{\frac{1}{2-\kappa_{n_0,n}}}$$

$$\frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left(\frac{n}{n_0}\right)^{\frac{1}{2-\kappa_{n_0,n}}}$$

$$\log \mathbb{M}(n) - \log \mathbb{M}(n_0) = \frac{1}{2-\kappa_{n_0,n}}(\log n - \log n_0)$$

$$2-\kappa_{n_0,n} = \frac{\log n - \log n_0}{\log \mathbb{M}(n) - \log \mathbb{M}(n_0)}$$

$$\implies \kappa_{n_0,n} = 2 - \frac{\log n - \log n_0}{\log \mathbb{M}(n) - \log \mathbb{M}(n_0)}$$

When the distribution is normal or thin-tailed, $\kappa = 0$

5.2 Kappa – for some Known Distributions

In order to compute the Kappa metric, one must be aware of how to compute the mean absolute deviation of the given distribution. Diaconis et al used a paper by De Moivre and Suzuki to prove that a closed form of mean absolute deviation does exist (Diaconis and Zabell 1991). The proof is shown below

Closed form of mean absolute deviation

$$\mathbb{E}(X-\mu|) = 2\int_m^\infty |x-\mu| f(x)dx$$

Proof

$$\mathbb{E}(X-\mu|) = \int_{-\infty}^\infty |x-\mu| f(x)dx$$

$$\mathbb{E}(X-\mu|) = \int_{-\infty}^m |x-\mu| f(x)dx + \int_m^\infty |x-\mu| f(x)dx$$

$$\mathbb{E}(X-\mu|) = \int_{-\infty}^m (\mu-x) f(x)dx + \int_m^\infty (x-\mu) f(x)dx$$

$$= \int_{-\infty}^{m} \mu f(x)dx - \int_{-\infty}^{m} xf(x)dx + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$= \mu \int_{-\infty}^{m} f(x)dx - \int_{-\infty}^{m} xf(x)dx + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$= \mu\Big(1 - \int_{m}^{\infty} f(x)dx\Big) - \int_{-\infty}^{m} xf(x)dx + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$= \mu\Big(1 - \int_{m}^{\infty} f(x)dx\Big) - \Big(\mu - \int_{m}^{\infty} xf(x)dx\Big) + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$= \mu - \mu \int_{m}^{\infty} f(x)dx - \mu + \int_{m}^{\infty} xf(x)dx + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$= -\int_{m}^{\infty} \mu f(x)dx + \int_{m}^{\infty} xf(x)dx + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$= \int_{m}^{\infty} (x-\mu)f(x)dx + \int_{m}^{\infty} (x-\mu)f(x)dx$$

$$\mathbb{E}(X-\mu|) = 2\int_{m}^{\infty} (x-\mu)f(x)dx$$

Hence proved

5.2.1 Gamma Distribution

Now computing the kappa value of the gamma distribution

In order to do this we need the probability density function of the gamma distribution(which we shall derive using convolution of the probability density function of the exponential distribution).

Exponential distribution probability density function

$$f_{exp}(x) = \lambda e^{-\lambda x}$$

Convolution Theorem

$$f(y) = \int_{0}^{y} f(x)f((y-x)dx$$

Let $X_i \sim Exp(\lambda)$ Let $Y = X_1 + X_2$

$$f_Y(y) = \int_0^y f_{X_1}(x) f_{X_2}((y-x)dx$$

$$f_Y(y) = \int_0^y \lambda e^{-\lambda x} \lambda e^{-\lambda(y-x)} dx$$

$$f_Y(y) = \lambda^2 \int_0^y e^{-\lambda x - \lambda y + \lambda x} dx$$

$$f_Y(y) = \lambda^2 \int_0^y e^{-\lambda y} dx$$

$$f_Y(y) = \lambda^2 e^{-\lambda y} \int_0^y dx$$

$$f_Y(y) = \lambda^2 e^{-\lambda y}(y-0)$$

$$f_Y(y) = \lambda^2 e^{-\lambda y} y$$

Now if $W \sim \sum_{i=1}^n X_i$ then $W \sim Gamma(n, \lambda)$

$$f_W(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$$

Finding constants

We know that the Gamma distribution is a type of Pearson distribution thus it is of the following form

$$\frac{f'(x)}{f(x)} = -\frac{(a_0 + a_1 x)}{b_0 + b_1 x + b_2 x^2}$$

$$\frac{f'(x)}{f(x)} = \frac{(-a_0 - a_1 x)}{b_0 + b_1 x + b_2 x^2}$$

$$f'(x) = \frac{\lambda^n}{(n-1)!}((n-1)x^{n-2}e^{-\lambda x} - \lambda e^{-\lambda x} x^{n-1})$$

$$f'(x) = \frac{\lambda^n}{(n-1)!} e^{-\lambda x} x^{n-1} \Big(\frac{(n-1)}{x} - \lambda \Big)$$

$$\frac{f'(x)}{f(x)} = \frac{\frac{\lambda^n}{(n-1)!} e^{-\lambda x} x^{n-1} \Big(\frac{(n-1)}{x} - \lambda \Big)}{\frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}}$$

$$\frac{f'(x)}{f(x)} = \Big(\frac{(n-1)}{x} - \lambda\Big)$$

From the above we see that

$$-a_0 = n-1 \implies a_0 = 1-n$$

$$-a_1 = -\lambda \implies a_1 = \lambda$$

$$b_0 = 0$$

$$b_1 = 1$$

$$b_2 = 0$$

Thus we know that if $W \sim Gamma(n, \lambda)$

$$f_W'(x) = \frac{1-n+\lambda x}{x} f_W(x)$$

$$f_W(x) = \frac{x}{1-n+\lambda x} f_W'(x)$$

Mean Absolute Deviation for a Pearson type Distribution is

$$\mathbb{E}(X-\mu|) = 2\int_m^\infty (x-\mu)(-f'(x))\frac{b_0+b_1x+b_2x^2}{a_0+a_1x}dx$$

On integrating by parts,

$$\mathbb{E}(|X-\mu|) = \frac{2(b_0+b_1\mu+b_2\mu^2)}{a_1+2b_2}f(\mu)$$

(Reference Kappa Paper here)

Mean Deviation of Gamma Distribution

$$\mathbb{E}(|X-\mu|) = \frac{2(0+1(\frac{n}{\lambda})+0(\frac{n}{\lambda})^2)}{\lambda+2\times 0}\frac{\lambda^n \frac{n}{\lambda}^{n-1} e^{-\lambda\frac{n}{\lambda}}}{(n-1)!}$$

$$\mathbb{E}(|X-\mu|) = \frac{2n}{\lambda^2}\frac{\lambda^n n^{n-1} e^{-n}}{\lambda^{n-1}(n-1)!}$$

$$\mathbb{E}(|X-\mu|) = \frac{2\lambda^n n^n e^{-n}}{\lambda^{n+1}(n-1)!}$$

$$\mathbb{E}(|X-\mu|) = \frac{2n^n e^{-n}}{\lambda \Gamma n}$$

Computing the Kappa Metric

$$\kappa_{(n_0,n)} = 2 - \frac{\log(n) - \log(n_0)}{\log\left(\frac{\mathbb{M}(n)}{\mathbb{M}(n_0)}\right)}$$

$$\kappa_{(n_0,n)} = 2 - \frac{\log(n) - \log(n_0)}{\log\left(\frac{n^n e^{-n}\Gamma(n_0)}{\Gamma(n) n_0^{n_0} e^{-n_0}}\right)}$$

5.2.2 For other distributions

The below table has been taken from Taleb 2019.

Distribution	κ_1
Student T (α)	$2 - \frac{2\log(2)}{2\log\left(\frac{2^{2-\alpha}\Gamma\left(\alpha-\frac{1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)^2}\right)+\log(\pi)}$
Exponential/Gamma	$2 - \frac{\log(2)}{2\log(2)-1} \approx .21$
Pareto (α)	$2 - \frac{\log(2)}{\log\left((\alpha-1)^{2-\alpha}\alpha^{\alpha-1}\int_0^{\frac{2}{\alpha-1}} -2\alpha^2(y+2)^{-2\alpha-1}\left(\frac{2}{\alpha-1}-y\right)\left(B_{\frac{1}{y+2}}(-\alpha,1-\alpha)-B_{\frac{y+1}{y+2}}(-\alpha,1-\alpha)\right)dy\right)}$[a]
Normal (μ,σ) with switching variance $\sigma^2 a$ w.p p^b.	$2 - \frac{\log(2)}{\log\left(\frac{\sqrt{2}\left(\sqrt{\frac{ap}{p-1}+\sigma^2}+p\left(-2\sqrt{\frac{ap}{p-1}+\sigma^2}+p\left(\sqrt{\frac{ap}{p-1}+\sigma^2}-\sqrt{2a\left(\frac{1}{p-1}+2\right)+4\sigma^2}+\sqrt{a+\sigma^2}\right)+\sqrt{2a\left(\frac{1}{p-1}+2\right)+4\sigma^2}\right)\right)}{p\sqrt{a+\sigma^2}-(p-1)\sqrt{\frac{ap}{p-1}+\sigma^2}}\right)}$
Lognormal (μ,σ)	$\approx 2 - \frac{\log(2)}{\log\left(\frac{2\,\mathrm{erf}\left(\frac{\sqrt{\log\left(\frac{1}{2}\left(e^{\sigma^2}+1\right)\right)}}{2\sqrt{2}}\right)}{\mathrm{erf}\left(\frac{\sigma}{2\sqrt{2}}\right)}\right)}.$

[a] $B_.(.,.)$ is the incomplete Beta function: $B_z(a,b) = \int_0^z t^{a-1}(1-t)^{b-1}dt$; erf(.) is the error function $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$.

Figure 5.2: $\kappa_{1,2}$ for a few distribution

From the above table, it is clear to us that the kappa for the Pareto Distribution and the Student T distribution (known fat-tailed distributions) are dependent on α. The below table was also taken from Taleb 2019. It shows this dependence

It is clear that as the tail exponent, α increases, the kappa value decreases. This means that

COMPARING PARETO TO STUDENT T (SAME TAIL EXPONENT α)

α	Pareto κ_1	Pareto $\kappa_{1,30}$	Pareto $\kappa_{1,100}$	Student κ_1	Student $\kappa_{1,30}$	Student $\kappa_{1,100}$
1.25	0.829	0.787	0.771	0.792	0.765	0.756
1.5	0.724	0.65	0.631	0.647	0.609	0.587
1.75	0.65	0.556	0.53	0.543	0.483	0.451
2.	0.594	0.484	0.449	0.465	0.387	0.352
2.25	0.551	0.431	0.388	0.406	0.316	0.282
2.5	0.517	0.386	0.341	0.359	0.256	0.227
2.75	0.488	0.356	0.307	0.321	0.224	0.189
3.	0.465	0.3246	0.281	0.29	0.191	0.159
3.25	0.445	0.305	0.258	0.265	0.167	0.138
3.5	0.428	0.284	0.235	0.243	0.149	0.121
3.75	0.413	0.263	0.222	0.225	0.13	0.10
4.	0.4	0.2532	0.211	0.209	0.126	0.093

Figure 5.3: $\kappa_{1,2}$ for Pareto and Student t distributions

the higher tail exponent, the more fat-tailed the distribution is.

5.3 Sample Size Metric

The formula can be also used to come up with a rigorous sample size metric to assess a sample size of statistical significance. The metric is given below

$$n_v = n_g^{-\frac{1}{\kappa_1 - 1}}$$

where n_v is the number of variables from a distribution other than the Gaussian with same variance as n_g(number of variables in the Gaussian distribution).

For example, the metric above tells us that a Student T distribution with 3 degrees of freedom (i.e $\alpha = 3$) requires 120 observations to get the same drop in variance as the Gaussian distribution has with 3 observations. The derivation of this metric will not be discussed as its beyond the scope of this project.

Chapter 6

Fat Tailedness in Real Life

The two case studies chosen have been chosen from the field of climate change and finance as these are the areas with respect to which fat-tailed distributions have been studied in great detail.

6.1 Case Study 1 : Testing Data for Paretianity

6.1.1 Data chosen

The data used for the following case study has been sourced from K Kulkarni (used in his paper, Quantifying Vulnerability of Crop Yields in India to Weather Extremes Kulkarni 2021 [1]). The data has been collected daily for 482 districts across India from January 1st, 1951, to December 31st, 2020.

As data was given district wise, we first had to decide which district's data should be concentrated on and checked for fat-tailedness. As discussed earlier there are a few heuristics which can be used to check if data given comes from a fat-tailed distribution. In this chapter, the heuristics will be checked using simulations. The code used to understand this data can be found in the repository -Fat-Tailed-Distributions on GitHub

[1]This data is not available on the public domain. If required, contact author

6.1.2 Ratio of Standard Deviation to Mean Absolute Deviation

For a normal/thin tailed distribution, the ratio of standard deviation to mean absolute deviation is around 1.25. However if the distribution is fat-tailed, the standard deviation will be blown up as when squaring the deviations from the mean, extreme values will be overweighed. So this ratio will get blown up. Thus the higher the ratio, the more fat-tailed a distribution is. The ratio of standard deviation to mean absolute deviation was examined for all the districts.On examination, it was found that the district with the highest ratio was Jaisalmer in Rajasthan with a ratio of 3.43. This is the district for which all the heuristics will be tested.

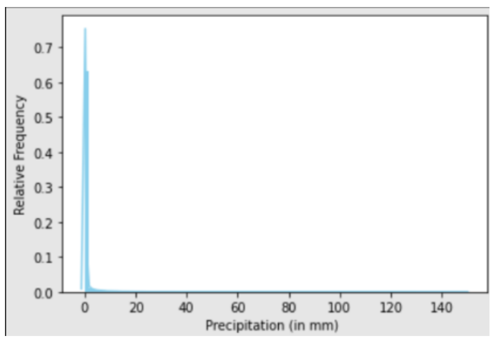


Figure 6.1: Jaisalmer, Rajasthan

The summary statistics for the precipitation data for this district is given below.

count	25568.000000
mean	0.539786
std	3.332083
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	148.903313

Figure 6.2: Summary statistics of Jaisalmer, Rajasthan

6.1.3 Moments

Though around 85 percent of the observations of this district are 0, the mean is not 0. This is due to the presence of a few extreme values which pull the mean away from the mode of the distribution.

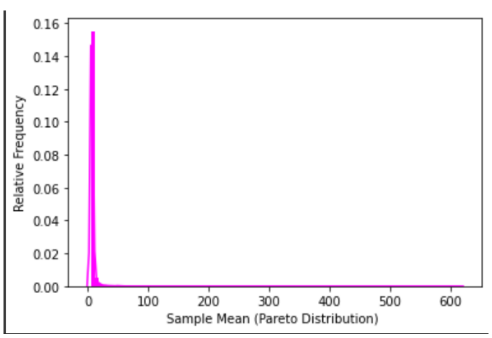
The skewness of this data set is 16.05 and the kurtosis is 441.72 which is much higher than that of the normal distribution which is 0 and 3 respectively.

6.1.4 Pillars Of Inferential Statistics

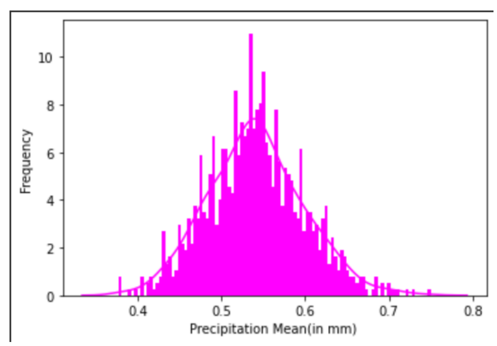
Central Limit Theorem

Fig 2 (a) shows the distribution of sample means for a known fat-tailed distribution (the Pareto distribution). It can be seen that this does not look like a normal distribution even when the sample size is large (10,000).

Fig 2 (b) shows the distribution of sample means for the precipitation data of Jaisalmer. It can be seen that this looks like a normal distribution when the sample size is 3000.



(a) Pareto Distribution



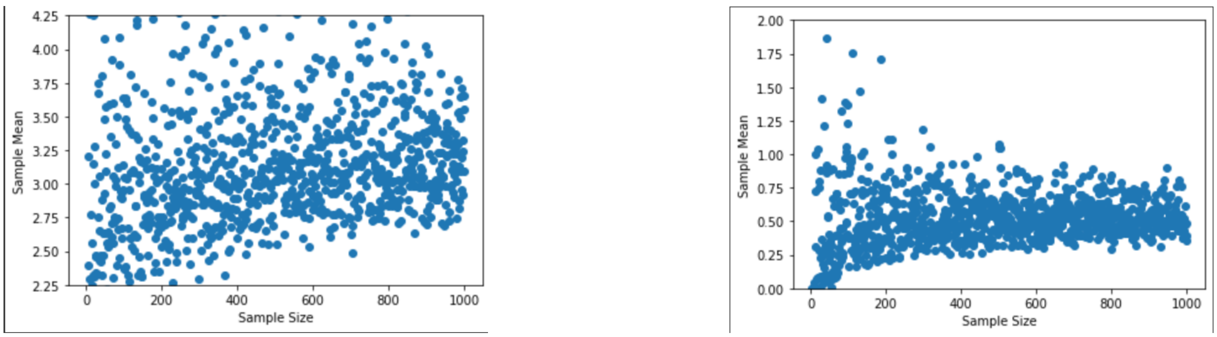
(b) Precipitation Data

Figure 6.3: Distribution of Sample Means

Law Of Large Numbers

Fig 3 (a) has the sample means on the y axis and the sample sizes on the x axis for a known fat-tailed distribution(the Pareto distribution). It can be seen that as the sample size increases, there is a still high variation of the mean (even at sample mean of 10,000).

Fig 3 (b) has the sample means on the y axis and the sample sizes on the x axis for the precipitation data of Jaisalmer. It can be seen that as the sample size increases, the variation of the sample mean decreases.



(a) Pareto Distribution

(b) Precipitation Data

Figure 6.4: Distribution of Sample Means

6.1.5 Kappa Metric

The Kappa Metric is based on the Law of Large Numbers. The Kappa metric for this data was calculated using the following equation.

$$n_v = n_g^{\frac{-1}{\kappa_1 - 1}} \tag{6.1}$$

where n_v is the size of a sample from an unknown distribution which has the same drop in variance as a given n_g (sample size) taken from a normal distribution. The standard deviation of a sample of size n is equal to standard deviation of the population by square root of n for the normal distribution. We then used trial and error to find the size of the sample that needed to be taken from precipitation data to get the same drop in variance as that of sample of size 30 taken from the normal distribution (that is scaled by a factor of $\frac{1}{\sqrt{30}}$) The Kappa Metric was found to be 0 for the observed data.

6.1.6 Fitting Known Fat-Tailed Distributions

As mentioned earlier, the above are just heuristics and cannot be used to definitively say whether the above data comes from a fat-tailed distribution. In order to know whether this comes from a fat-tailed distribution, we tried to fit known fat-tailed distributions to this data set.

Initially we tried to fit the Pareto distribution to this data with different tail exponents. Using the scipy.stats package, we fit a Pareto distribution (using method of moments estimation). The parameters for the fitted distribution found were $1.787, 0.5684, 1.092 \times 10^{-5}$. To check if the fitted distribution was actually a good fit for the data, a qqplot was constructed. If the fitted distribution was a good fit for the data, then the data points would lie around the $y = x$ line. However we see in the constructed qqplot, that the fitted distribution is definitely not a good fit for the data.

Also according to the constraints placed on the parameters for any Pareto distribution, the minimum value of x must be greater than 0 and for the Jaisalmer data, the minimum value of x is 0. So we decided to look at another form of the power law, the discretized version of the power law) - the Zipf distribution. In the Zipf Distribution, the data is ranked based on the frequency of the data.

To check if this distribution follows a Zipf distribution, we separated the data into 100 bins and checked the frequency of values in each bin and put it in a list. We then created a permutation of this list of frequencies such that the frequencies were now ordered from highest to lowest. This has now become the new observed distribution. A line plot of this distribution was drawn. Knowing the formula for the Zipf distribution, using trial and error we found a tail exponent such that a line plot of the Zipf distribution with this particular tail exponent visually matched the observed distribution. The tail exponent found was 4.3. (Fig 4). The blue line shows the simulated zipf distribution with tail exponent,4.3 and red line shows the frequency distribution of the precipitation data.

To confirm the hypothesis that the observed distribution was the Zipf distribution(tail exponent = 4.3), the chi square test was used. However using the chi square test turned out to be problematic due to the sparseness of observations in the tails. To use the chi square test as

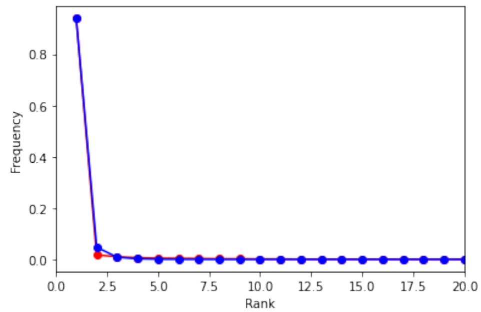


Figure 6.5: Zipf Distribution

a goodness of fit test, the expected /observed frequency must be at the least 5. However due to the low probability of any particular value in the tail, the expected frequency and observed frequency will be less than 5 thus the chi square test cannot be run. We tried decreasing the number of bins but unless the number of bins was low, the chisquare test could not be run. If the number of bins was low, then the number of observations would also be low. Hence no test run on it would have ay statistical significance.

6.1.7 Using metrics from "Are your data really Pareto distributed?" by P Cirillo

According to "Are your data really Pareto distributed?" by P Cirillo looking at the Zipf plot will tell us whether the data is fat-tailed or not. If the data is paretian in nature, the Zipf plot , i.e. a plot in which the logarithm of the empirical survival function is plotted against the logs of the ordered values of x will be linear. The following graph (Figure 5) (taken from the paper mentioned above) shows how the Zipf plot will look for some classical distributions. The Zipf plot for a power law distribution shows a negative linear relationship. The proof of this can be found below,

For a Pareto distribution,

$F(x) = 1 - (\frac{x}{x_m})^{-\alpha}$ where $0 < x_m \leq x$

$\bar{F}(x) = 1 - F(x) = (\frac{x}{x_m})^{-\alpha}$ Taking Logs on both sides of the equation

$\log \bar{F}(x) = \alpha \log x_m - \alpha \log x$

As α and x_m are parameters of the distribution, the term $\alpha \log x_m$ is a constant. Let us denote this by C.

$\log \bar{F}(x) = C - \alpha \log x$

It can be seen that there exists a negative linear relationship between the logarithm of the survival function and the logarithm of x. The slope of the line is equal to $-\alpha$.

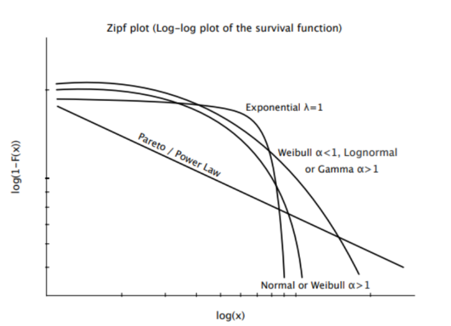


Figure 6.6: Zipf Plot for Some Classical Distributions

A zipf plot was created for the observed data.

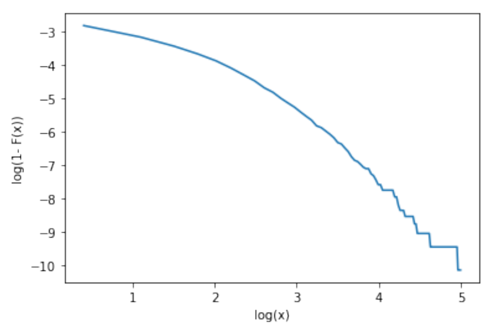


Figure 6.7: Zipf Plot for Precipitation data

On creating the Zipf plot for the observed data, it was clear to see it was not linear in nature. Comparing the Zipf plot got from the observed data (Figure 6) to the graph taken from the paper (Figure 5), it can be seen that the Zipf plot got is similar the Zipf plot of the Weibull distribution. However on fitting a Weibull distribution to the observed data, it could be seen that using trial and error no tail exponent could be found such that the Weibull distribution would visually match the observed distribution. We then tried using the scipy.stats module to fit the distribution.This module uses the Maximum Likelihood method of Estimation to fit distributions. However the exponent values returned by that also did not give a distribution that looked like the observed distribution.

Cirillo recommends using the Mean Excess Function to check whether the given data is paretian. The Pareto distribution (and its generalisations) are the only distributions characterised by a linear mean excess function. This is because they follow the van der Wijk's law which asserts that the average income of all the people above a given level u is proportional to u itself. The following graph (Figure 7) (taken from the paper) shows how the meplot looks for some classical distributions.

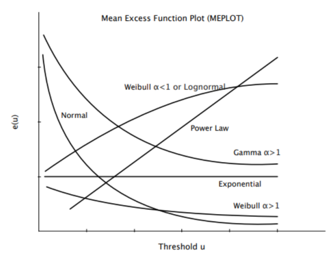


Figure 6.8: Meplot for some Classical Distributions

From an empirical point of view, the mean excess of a sample is

$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)}{\sum_{i=1}^{n} 1} \forall X_i > u \tag{6.2}$$

Using the above formula, a meplot was drawn from the observed data (Figure 8)

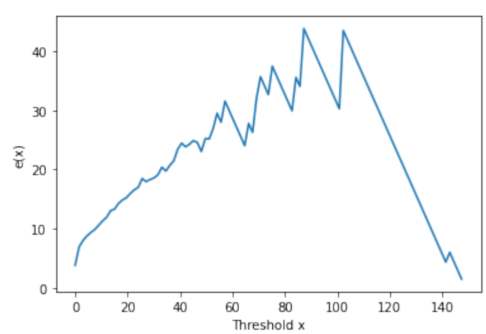


Figure 6.9: Meplot for Precipitation Data

The shape of mean excess function plot of the observed data looked nothing like the shapes of the meplots for any of the classical distributions as seen in Cirillo's paper. In fact it looked like the observed distribution could be a mixture of multiple classical distributions. It is hard to determine what the classical distributions in this mixture could be.

Due to the sparseness of values in the tail and the fact that some of the heuristics told us that the observed distribution of data could be fat-tailed while others said it could not, it can be concluded that though this data is not fat-tailed the tail does determine the properties.

6.1.8 Conclusions

- Observed precipitation data are neither strictly fat nor thin-tailed and exhibit mixed behaviour in the head and tail portions.
- Chisquare Goodness of fit test cannot be used to test empirical distributions or any other distribution with sparse data in the tail.
- *If there is sparseness in the tail and high probability in the lower values of the distribution, we could make log transformations in order to get a better picture of the distribution as a whole.*
- In order to understand a distribution where the tail determines the properties, the survival function and the mean excess function have proved to be extremely useful. The MEPLOT and the zipf plot provide alternative heuristic for such distributions.
- Distributions may exhibit both fat-tailed and thin tailed behaviour

6.2 Case Study 2: Aggregate Distributions in Business Network Structures

In 1977, RE Lucas put forward the famous diversification argument that stated microecomic shocks average out and have a negligible impact on the economy as a whole. However recent studies have shows that this is not always true. In fact, this argument only holds if the aggregate distribution of the economy (aggregated using firm level out degree distributions across sectors) is thin tailed. This case study aims to answer the question at what threshold(s) does the aggregate distribution become fat-tailed in nature. To understand this, the paper, "When Links Matter - Propagation of Firm-Level Shocks through Production Networks" by A Shrivastava and G Bahal [2] was used.

To answer the question above, one must first understand how these distributions are aggregated from sector level distributions and what those distributions are.

Firm level out degree distributions is a probability distribution specific to a sector where the random variable x denotes the number of customers a firm can supply to. These distributions can be generated by 2 types of linkages: thin-tailed linkages and fat-tailed linkages

Thin tailed linkage: The process by which customers choose their suppliers at random which leads to a thin tailed distribution. The most extreme case of this would be where each firm has only 1 customer.

Fat-tailed linkages: The process by which suppliers with more preexisting links attract more users. This leads to a fat-tailed distribution. The most extreme case would by when 1 firm has all the customers of the sector.

Given that there are sectors in the economy with either fat-tailed or thin-tailed linkages, what behaviour would the aggregate distribution display. The authors of this paper claim that aggregate distributions exhibit fat-tailed behaviour when the percentage of sectors with fat-tailed firm level out degree distributions is of the same order of magnitude as sectors with thin tailed firm level out degree distributions. In order to understand this claim, we must first understand what it means to aggregate sector distributions to an economy level.

Understanding firm level out degree distributions sector wise:

For any sector i, proportion of firms with degree k is denoted by $p_i(k)$

$$p_i(k) = \frac{k^{-\alpha_i}}{\sum_{j=1}^{k_i^{max}} k^{-\alpha_j}}$$

where k_{max} is the maximum firm level out degree and α_i is the tail exponent characterising the firm level out degree distribution for the particular sector.

Let $\sum_{j=1}^{k_i^{max}} k^{-\alpha_j} = S_i$

[2]This paper is yet to be published. The unpublished paper was obtained from A Shrivastava

S_i is constant for the i'th sector as it a function of k_{max}.

Let $\ln k = x \implies k = e^x$

$$p_i(x) = \frac{e^{-\alpha_i x}}{S_i}$$

Let the economy have 2 types of degree distributions: fat-tailed firm level out degree distribution and thin tailed firm level out degree distributions . For a fat-tailed distribution, the tail exponent, $\alpha_f = 2$ and for a thin tailed distribution $\alpha_t \geq 3$. In order to aggregate this over the economy, we shall make an assumption that all thin tailed sectors have the same tail exponent. Since the distribution of all fat-tailed sectors are the same (have the same tail exponent), we can collapse them into one sector

$$p_f(x) = \frac{e^{-\alpha_f x}}{S_f}$$

Similarly for thin tailed distributions,

$$p_t(x) = \frac{e^{-\alpha_t x}}{S_t}$$

Let proportion of fat-tailed sectors in economy $= n_f$

Thus proportion of thin tailed sectors in economy $= 1 - n_f$

Therefore when we aggregate to get the distribution of the economy, we get

$$p(x) = n_f p_f(x) + (1 - n_f) p_t(x)$$

$$p(x) = n_f \frac{e^{-\alpha_f x}}{S_f} + (1 - n_f) \frac{e^{-\alpha_t x}}{S_t}$$

$$p(x) = n_f \frac{e^{-\alpha_f x}}{S_f} \left[1 + \frac{1 - n_f}{n_f} \frac{S_f}{S_t} e^{-(\alpha_t - \alpha_f)x}\right]$$

Taking logs

$$\ln p(x) = \ln \frac{n_f}{S_f} - \alpha_f x + \ln \left[1 + \frac{1 - n_f}{n_f} \frac{S_f}{S_t} e^{-(\alpha_t - \alpha_f)x}\right]$$

Differentiating to get slope

$$\frac{d\ln p(x)}{dx} = -\alpha_f - \frac{1-n_f}{n_f}\frac{S_f}{S_t}(\alpha_t - \alpha_f)e^{-(\alpha_t-\alpha_f)x}$$

The slope will give us the tail exponent of the aggregate distribution at high values of k. In the paper, we are considering values around $k = 100$ i.e. $x = \ln 100 \approx 4.6$

$$\alpha_{agg} = (2 + \frac{1-n_f}{n_f}\frac{S_f}{S_t}(\alpha_t - \alpha_f)e^{-(\alpha_t-\alpha_f)x})$$

Bounds of the terms of the expression above:

1. $\frac{1-n_f}{n_f} \leq 9$ (given percentages are of the same order of magnitude)
2. $\frac{S_f}{S_t} = \frac{\sum_{j=1}^{kmax(f)} k^{-\alpha_f}}{\sum_{j=1}^{kmax(t)} k^{-\alpha_t}}$

   S_f is bounded above by the Reimann Zeta function. For $\alpha_f = 2$, it appears to be bounded by 1.65.

   S_t is bounded below by 1. (as $k \geq 1$ and $1^{-s} = 1 \forall s \in \mathbb{R}^+$)

   $\implies \frac{S_f}{S_t} \leq 1.65$
3. $(\alpha_t - \alpha_f) \geq 1$
4. $e^{-(\alpha_t-\alpha_f)4.6} \leq 0.01$

As $(\alpha_t - \alpha_f)$ increases, $e^{-(\alpha_t-\alpha_f)x}$ decreases. However as the first term increases linearly and the second term decreases exponentially, the second term dominates the first.

$$\alpha_{agg} \leq -2 - 0.1485(\alpha_t - \alpha_f)$$

With this understanding, 2 observations were made.

1. When $(\alpha_t - \alpha_f)$ is small, the aggregate distribution will be fat-tailed in nature only if the percentages of fat and thin-tailed distributions are of the same orders of magnitude.

   Example:

Let $n_f = 0.1$,

$k_{max}(t) = 50$

$k_{max}(f) = 100$

$\alpha_t = 3$

$\alpha_f = 2$

$\alpha_{agg} = 2 + \frac{1-0.1}{0.1}\frac{1.6349}{1}e^{-4.5}1 = 2.16$

Consider same parameters as above but $n_f = 0.01$,

$\alpha_{agg} = 2 + \frac{1-0.01}{0.01}\frac{1.6349}{1}e^{-4.5}1 = 3.79$

2. When $(\alpha_t - \alpha_f)$ is large, the aggregate distribution will be fat-tailed in nature even if the percentages of fat and thin tailed distributions are not of the same orders of magnitude.

   Example

   Let $n_f = 0.01$,

   $k_{max}(t) = 50$

   $k_{max}(f) = 100$

   $\alpha_t = 12$

   $\alpha_f = 2$

   $\alpha_{agg} = 2 + \frac{1-0.01}{0.01}\frac{1.6349}{1}e^{-45}10 = 2 + (4.6331323 \times 10^{-17}) = 2$

6.2.1 Conclusions

1. Percentage of fat-tailed and thin tailed distributions need not be of "same orders of magnitude" for aggregate distribution to be fat-tailed.

2. 2 parameters appear to determine whether the aggregate distribution is fat-tailed or not.

   Percentage of fat-tailed distributions (n_f)

   Difference in tail exponent $(\alpha_t - \alpha_f)$

Chapter 7

Conclusion

This thesis aimed to understand what fat-tailed distributions are and to understand how to classify a distribution as fat-tailed. The power law class of distributions exemplify fat-tailed behaviour and can be considered to be the analytical description of fat-tailed distributions. Based on the analysis done, it was found that a set of heuristics can be employed for detecting fat-tail behaviour in a pre-asymptotic manner (i.e. empirically from a stream of data). The empirical metric, κ (empirical in the sense of being indexed to observations) that measures the rate of convergence of the sample mean under the Law of Large Numbers can be used to assess the degree of fat-tailedness (for both theoretical and empirical distributions). This measure's importance lies in its use to make comparisons of fat-tailedness across distributions and assess sample size of statistical significance.

Through the case studies done in the fields of business network structure and climate change, it was concluded that phenomena in the real world may simultaneously exhibit both fat-tailed and thin-tailed behaviour and it is hard to classify them as fat or thin-tailed (such as precipitation data for Rajasthan). The case study done on business network structure led to the conclusion that even a small proportion of fat-tailed distributions in a business network may be sufficient for the aggregate distribution to be fat-tailed, apparently depending upon the differences in fat and thin-tail exponents (more thin vs less thin). Future research is needed to determine the relation between the two aforementioned factors that leads to fat-tailedness in aggregate distributions

Important Definitions

7.1 Moments

Moments of a function are quantitative measures related to the shape of a function's graph. If the function, f considered is a probability distribution, there are 3 types of moments to be considered - raw moments, central moments and standardised moments.

7.1.1 Raw moments

They are also known as crude moments. The n'th raw moment is defined as the expected value of X^n where X are the random variables possible in the distribution being studied.

$$E(X)^n = \int_{x \in Support(X)} x^n \dot{f}(x)dx$$

The first raw moment gives us the theoretical mean, μ.

7.1.2 Central Moments

These moments are defined around the mean of the distribution being considered. The n'th central moment is defined as expected value of $(X - \mu)^n$.

$$E(X - \mu)^n = \int_{x \in Support(X)} (x - \mu)^n \dot{f}(x)dx$$

The first central moment is 0 (as it is defined around the mean).

The second central moment is variance, σ^2. The square root of variance is standard deviation, σ

Variance

Variance is the square of standard deviation which is used as a measure of spread of a distribution.

7.1.3 Standardized Moments

A standardized moment is a central moment that has been normalised. This normalisation is done by division by standard deviation to a power. A standardised moment of degree n is the ratio of expected value of $(X - \mu)^n$ to the nth power of the standard deviation.

$$\tilde{\mu_n} = \frac{E(X - \mu)^n}{\sigma^n}$$

The first standardized moment is 0.

The second standardized moment is 1.

The third standardized moment is known as the skewness.

The fourth standardized moment is known as the kurtosis.

Skewness

Skewness essentially measures the relative size of the two tails. It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. It refers to a distortion of the given curve as it deviates from the symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed.

Skewness gives the direction of outliers, not the number of outliers. Skewness is used along with kurtosis to better judge the likelihood of events falling in the tails of a probability distribution. The skewness value can be positive, zero, negative, or undefined.

For a unimodal distribution, negative skew commonly indicates that the tail is on the left side of the distribution, and positive skew indicates that the tail is on the right. In cases where one tail is long but the other tail is fat, skewness does not obey a simple rule. For example, a zero value means that the tails on both sides of the mean balance out overall; this is the case for a symmetric distribution, but can also be true for an asymmetric distribution where one tail is long and thin, and the other is short but fat. (Wikipedia)

In the idea of nonparametric skew, the skewness is defined in terms of this relationship: positive/right nonparametric skew means the mean is greater than (to the right of) the median, while negative/left nonparametric skew means the mean is less than (to the left of) the median. However, the modern definition of skewness and the traditional nonparametric definition do not always have the same sign

Kurtosis

Kurtosis measures the combined sizes of the two tails. It measures the amount of probability in the tails. The normal distribution has a kurtosis of 3. Generally when we look at the kurtosis of a distribution, we give the excess kurtosis (which is kurtosis - 3) so it can be seen in comparison to the normal distribution. Higher kurtosis corresponds to greater extremity of deviations (or outliers), and not the configuration of data near the mean. There are 3 classes we can classify distributions into on the basis of kurtosis:

1. Mesokurtic : Excess kurtosis $= 0$ eg normal
2. Platykurtic : Excess kurtosis < 0 eg uniform
3. Leptokurtic : Excess kurtosis > 0 eg exponential / Poissons distribution

7.1.4 Mean Absolute Deviation

The mean absolute deviation is defined to be the average/mean of the absolute value of the deviation from the mean

$$MAD = \frac{1}{n}\sum_{i}^{n}|X_i - \mu|$$

7.2 Stable Family of Distributions

A distribution is said to be stable if a linear combination of two independent random variables with this distribution has the same distribution(including the same location and scale parameters). This is also called the Lévy alpha-stable distribution, after Paul Lévy, the first mathematician to have studied it. Examples of the stable distribution include Levy distribution, Cauchy Distribution and the Normal Distribution

7.3 Survival Function

The survival function of x is a function that gives the probability of an event being greater than x (the probability that the object of interest will survive past a certain time)

7.4 Mean Excess Function

Let X be a random variable with distribution F and right endpoint x_F (i.e. $x_F = sup\{x \in \mathbb{R} : F(x) < 1\}$). The mean excess function is defined as the following

$$e(u) = E[X - u|X > u] = \frac{\int_u^\infty (t-u)dF(t)}{\int_u^\infty dF(t)}, 0 < u < x_F$$

The linearity of the mean excess function characterizes the generalised pareto distribution class

7.5 Hazard Function

The hazard function models which periods have the highest or lowest chances of an event. The function is defined as the instantaneous risk that the event of interest happens, within a very narrow time frame. The hazard function is a conditional failure rate, in that it is conditional a person has actually survived until time t.

7.6 Characterisation theorem

A characterization theorem says that a particular object – a function, a space, etc. – is the only one that possesses properties specified in the theorem. A characterization of a probability distribution accordingly states that it is the only probability distribution that satisfies specified conditions.

7.7 Characteristic Functions

If a random variable admits a probability density function, then the characteristic function is the Fourier transform of the probability density function. Thus it provides an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions.

7.8 Memorylessness

In probability and statistics, memorylessness is a property of certain probability distributions. It usually refers to the cases when the distribution of a "waiting time" until a certain event does not depend on how much time has elapsed already. Only two kinds of distributions are memoryless: geometric distributions of non-negative integers and the exponential distributions of non-negative real numbers.

Bibliography

Taleb, Nassim Nicholas (2008). *The Black Swan: The Impact of the Highly Improbable*. 1st ed. London: Random House. ISBN: 1400063515 (cit. on pp. 2, 5, 6).

— (2022). *Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications*. arXiv: 2001.10488 [stat.OT] (cit. on p. 2).

— (2004). *Fooled by Randomness. The Hidden Role of Chance in Life and in the Markets*. 2nd ed. New York: Texere. 368 pp. ISBN: 978-1-58799-190-5 (cit. on pp. 2, 8).

N N taleb's probability moocs (n.d.). URL: https://www.youtube.com/channel/UC8uY6yLP9BS4BUc9BSc0Jww (cit. on pp. 8, 33, 37, 38).

Goldie, Charles M. and Claudia Klüppelberg (1998). "Subexponential distributions". In: (cit. on p. 8).

Probability distribution (Apr. 2023). URL: https://en.wikipedia.org/wiki/Probability_distribution (cit. on pp. 11, 12, 15–17, 19–21, 23, 24, 26).

Logarithmic distribution (Dec. 2020). URL: https://en.wikipedia.org/wiki/Logarithmic_distribution (cit. on p. 18).

Taleb, Nassim Nicholas (Apr. 2019). "How much data do you need? An operational, pre-asymptotic metric for fat-tailedness". In: *International Journal of Forecasting* 35.2, pp. 677–686. DOI: 10.1016/j.ijforecast.2018.10.003. URL: https://doi.org/10.1016%5C%2Fj.ijforecast.2018.10.003 (cit. on pp. 37, 42, 51).

Glen, Stephanie (June 2021). *Fat tail distribution: Definition, examples*. URL: https://www.statisticshowto.com/fat-tail-distribution/ (cit. on p. 38).

Diaconis, Persi and Sandy Zabell (1991). "Closed Form Summation for Classical Distributions: Variations on a Theme of De Moivre". In: *Statistical Science* 6.3, pp. 284–302. ISSN: 08834237. URL: http://www.jstor.org/stable/2245429 (visited on 04/29/2023) (cit. on p. 47).

Kulkarni, Kedar (Aug. 2021). *Quantifying Vulnerability of Crop Yields in India to Weather Extremes*. 2021 Annual Meeting, August 1-3, Austin, Texas 313879. Agricultural and Applied Economics Association. DOI: 10.22004/ag.econ.313879. URL: https://ideas.repec.org/p/ags/aaea21/313879.html (cit. on p. 54).

Appendix A

Proof of Central Limit Theorem

Statement of the Central Limit Theorem

$$\lim_{n\to\infty} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1) \tag{A.1}$$

PROOF

Consider the equation

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \sim N(0,1) \tag{A.2}$$

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{1}{\sigma\sqrt{n}} (n\bar{X}_n - n\mu)$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{n}{\sigma\sqrt{n}} (\bar{X}_n - \mu)$$

$$\Leftrightarrow \lim_{n\to\infty} \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) \sim N(0,1)$$

Hence (A.1) and (A.2) are equivalent.

Let $Z_i = \frac{X_i - \mu}{\sigma}$

$$T_n = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma}$$

Statement of CLT :

$$\lim_{n\to\infty} \frac{T_n}{\sqrt{n}} \sim N(0,1) \tag{A.3}$$

Moments of Z_i

$$E(Z_i) = E\Big(\frac{X_i - \mu}{\sigma}\Big)$$

$$E(Z_i) = \frac{1}{\sigma}\Big(E(X_i) - E(\mu)\Big)$$

$$E(Z_i) = \frac{1}{\sigma}\Big(\mu - \mu\Big)$$

$$E(Z_i) = 0$$

Variance of Z_i

$$Var(Z_i) = Var\Big(\frac{X_i - \mu}{\sigma}\Big)$$

$$Var(Z_i) = \frac{1}{\sigma^2}Var(X_i - \mu)$$

$$Var(Z_i) = \frac{1}{\sigma^2}Var(X_i)$$

$$Var(Z_i) = \frac{1}{\sigma^2}\sigma^2$$

$$Var(Z_i) = 1$$

Finding the MGF Proved above

$$E(Z_i) = 0$$

To find $E(Z_i^2)$

$$Var(Z_i) = E(Z_i^2) + E(Z_i)^2$$

$$1 = E(Z_i^2) + 0$$

$$E(Z_i^2) = 1$$

$$M_{Z_i}(t) = \mathbb{E}(e^{tZ_i})$$

$$M_{Z_i}(0) = \mathbb{E}(e^0) = 1$$

Therefore:
$M_{Z_i}(0) = 1$
$M'_{Z_i}(0) = 0$
$M''_{Z_i}(0) = 1$

To remember: If 2 random variables have the same MGF, then they are from the same distribution. Same can be said for the limit.

Proving equation A.3

$$\lim_{n\to\infty} \frac{T_n}{\sqrt{n}} \sim N(0,1)$$

Same as proving

$$\lim_{n\to\infty} M_{\frac{T_n}{\sqrt{n}}}(t) = M_Y(t) \text{ st } Y \sim N(0,1) \tag{A.4}$$

RHS

$$M_Y(t) = e^{\frac{t^2}{2}}$$

LHS

$$M_{\frac{T_n}{\sqrt{n}}}(t) = E(e^{t\frac{T_n}{\sqrt{n}}})$$

$$M_{\frac{T_n}{\sqrt{n}}}(t) = E(e^{\frac{t}{\sqrt{n}}T_n})$$

$$M_{\frac{T_n}{\sqrt{n}}}(t) = E(e^{\frac{t}{\sqrt{n}}\sum_{i=1}^{n} Z_i})$$

$$M_{\frac{T_n}{\sqrt{n}}}(t) = E(e^{\frac{t}{\sqrt{n}}Z_1})E(e^{\frac{t}{\sqrt{n}}Z_2})\dots E(e^{\frac{t}{\sqrt{n}}Z_n})$$

$$M_{\frac{T_n}{\sqrt{n}}}(t) = \left[M_{Z_i}(\frac{t}{\sqrt{n}})\right]^n$$

$$\lim_{n\to\infty} M_{\frac{T_n}{\sqrt{n}}}(t) = \lim_{n\to\infty}\left[M_{Z_i}(\frac{t}{\sqrt{n}})\right]^n$$

$$\lim_{n\to\infty} M_{\frac{T_n}{\sqrt{n}}}(t) = 1^\infty$$

The above is in indeterminate form so we can evaluate this using L'Hopital's rule after we do some manipulation

Taking natural logarithms

$$\ln M_{\frac{T_n}{\sqrt{n}}}(t) = n \ln M_{Z_i}(\frac{t}{\sqrt{n}}) \tag{A.5}$$

This still cannot be evaluated using L'Hopital's rule

$$\ln M_{\frac{T_n}{\sqrt{n}}}(t) = \frac{\ln M_{Z_i}(\frac{t}{\sqrt{n}})}{n^{-1}} \tag{A.6}$$

As $n \to \infty$, the RHS of the above evaluates to 0/0. Hence L'Hopitals rule can be used.

However n is an integer so we cannot take derivatives. Let $y \in \mathbb{R}$

$y = \frac{1}{\sqrt{n}}$

As $n \to \infty$, $y \to 0$

So RHS of equation A.6, can be written as

$$= \lim_{y\to 0} \frac{\ln(M_{Z_i}(yt))}{y^2}$$

As it's still of the form 0/0, we can use L'Hopital's rule

$$= \lim_{y\to 0} \frac{(M'_{Z_i}(yt))t}{(M_{Z_i}(yt))2y}$$

$$= \lim_{y\to 0} \frac{1}{(M_{Z_i}(yt))} \lim_{y\to 0} \frac{(M'_{Z_i}(yt))t}{2y}$$

We know that $M_{Z_i}(0) = 1$ and $M'_{Z_i}(0) = 0$

$$= \frac{1}{1} \lim_{y\to 0} \frac{0t}{0}$$

Using L'Hopitals Rule again

$$= 1 \lim_{y\to 0} \frac{t(M''_{Z_i}(yt))t}{2}$$

$$= \frac{t^2}{2}$$

From equation A.5

$$\ln M_{\frac{T_n}{\sqrt{n}}}(t) = \frac{t^2}{2}$$

To get the above back to our original form, we take antilogs

$$\lim_{n\to\infty} M_{\frac{T_n}{\sqrt{n}}}(t) = e^{\frac{t^2}{2}}$$

LHS = RHS

$\implies \lim_{n\to\infty} \frac{T_n}{\sqrt{n}} \sim N(0,1)$

$\implies \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i-\mu}{\sigma} \sim N(0,1)$

$\implies \lim_{n\to\infty} \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \sim N(0,1)$