Pareto Distribution: Theory and Applications

Hüseyin Adaş

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Applied Mathematics and Computer Science

Eastern Mediterranean University
July 2021
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Applied Mathematics and Computer Science.

Prof. Dr. Nazım Mahmudov
Chair, Department of Mathematics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Applied Mathematics and Computer Science.

Asst. Prof. Dr. Nidai Şemi
Supervisor

Examining Committee

1. Prof. Dr. Sonuç Zorlu Oğurlu
2. Asst. Prof. Dr. Halil Gezer
3. Asst. Prof. Dr. Nidai Şemi

ABSTRACT

The focus point of this thesis is the Pareto Distribution and Pareto Analysis. In this study, the history of the Pareto Distribution is given and types of the Pareto Distribution are analysed in detail. The necessary definitions are given and propositions about the subject are proved. On the other hand, it is clearly stated in what areas the Pareto Principle (80-20 Rule) is used in daily life. Pareto graphs are drawn using Microsoft Excel in Pareto Analysis Chapter.

Keywords: Vilfredo Pareto, Pareto Distribution, Pareto Principle, Pareto Analysis, Pareto Chart.

ÖZ

Bu tezin odak noktası Pareto Dağılımı ve Pareto Analizidir. Bu çalışmada öncelikle Pareto Dağılımının tarihçesine yer verilmiş ve Pareto Dağılımının türleri detaylı bir şekilde incelenmiştir. Gerekli tanımlar verilerek konu hakkındaki önermeler ispatlanmıştır. Öte yandan, Pareto İlkesi (80-20 Kuralı)'nin güncel hayatta ne gibi alanlarda kullanıldığı net bir biçimde ifade edilmiştir. Pareto Analizi bölümünde Microsoft Excel kullanılarak Pareto Grafikleri çizilmiştir.

Anahtar Kelimeler: Vilfredo Pareto, Pareto Dağılımı, Pareto İlkesi, Pareto Analizi, Pareto Grafiği.

TO MY LOVE

ACKNOWLEDGMENT

First of all, I would like to thank my supervisor Asst. Prof. Dr. Nidai Şemi who always showed invaluable contributions and guidance to me during the preparation process of my thesis. I am feeling blessed that I had the chance of studying with him. Then, I would like to thank all my instructors at the department of mathematics.

On the other hand, I would like to thank my friends for their help and support to me during the period of my studies and this thesis.

Special thanks go to my family for their priceless support and encouragements. I am so grateful to have such a family.

Last of all, I would like to extend my very special thank to my girlfriend Ayten for her continued support, encouragement and understanding throughout my masters programme.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDF	Cumulative Distribution Function
CRV	Continuous Random Variable
IQR	Interquartile Range
PDF	Probability Density Function

Chapter 1

INTRODUCTION

The Pareto Distribution is a continuous probability distribution that has many practical applications in probability theory and statistics.

The Pareto Distribution was found by Italian civil engineer, economist and sociologist Vilfredo Pareto. He discovered that $80\,\%$ of the land in Italy was owned by $20\,\%$ of the people in the country in 1906. Furthermore, he found out that the Wealth distribution was also same in all Europe. At the beginning of $20th$ century, Pareto created a mathematical model that describes the inequalities in Wealth distribution that existed in Italy. [3]

The Pareto Distribution is sometimes known as the Pareto Principle or the $80-20$ Rule. Pareto Analysis (80 – 20 Rule) is a statistical method in decision making. By doing 20% of the work, it generates 80% of the benefit of doing entire work. This distribution is used in various areas. For instance, this distribution is used in describing social, scientific and geophysical phenomena in society. In addition, Pareto Analysis $(80-20\,\text{Rule})$ can be applied almost all areas such as Business Management, Company Revenues, Employee Evaluation etc. [3], [4]

The Pareto Distribution can also be used for many other purposes. For example, a company can use $80-20$ Rule to determine the most significant portion that it can concentrate on its efficiency and it can go up its efficiency. More specifically; a restaurant can use the Pareto Principle to evaluate the performance of its cooks. Thus, the restaurant may observe that $80\,\%$ of its overall output is the direct result of about $20\,\%$ of its cooks. [3]

There exist two types of the Pareto distribution: The Basic Pareto distribution and The General Pareto distribution. In this study we have analysed them in details. Moreover, in this study we have introduced the necessary definitions from the Basic Pareto distribution and the General Pareto distribution which are needed to build up the proofs. This thesis consists of three chapters.

Chapter 1 is the introduction which contains historical background of the Pareto distribution and some important area of use of the Pareto distribution and Pareto Analysis $(80-20\text{ Rule})$.

In Chapter 2, we gave necessary definitions of the Basic Pareto distribution and General Pareto distribution. Then, significant propositions and examples related to both types of the Pareto distribution are proved and solved using these definitions.

In Chapter 3, we dealt with Pareto Analysis and where to use the Pareto Principle $(80-20\text{ Rule})$ in real life. Then, we mentioned Pareto Chart and how to draw a Pareto Chart. At the end of this chapter, we discussed various examples related to Pareto Chart. Here, we used Microsoft Excel to draw Pareto graphs.

Chapter 2

PARETO DISTRIBUTION

2.1 Basic Pareto Distribution

Definition 2.1.1. [1], [2]

A probability density function; $f(x)$, of the form

$$f(x) = \frac{a}{x^{a+1}}, \quad x \geq 1, \quad a > 0 \text{ (shape parameter)}$$

is called Basic Pareto distribution.

Proof.

Let us prove that $f(x)$ is a probability density function.

We should show that $\int_1^\infty f(x)dx = 1$ and $f(x) \geq 0$ for $x \geq 1$.

It is obvious that $f(x) \geq 0$ for $x \geq 1$ when $a > 0$.

$$\int_1^\infty f(x)dx = \int_1^\infty \frac{a}{x^{a+1}} dx = \lim_{t\to+\infty} \int_1^t \frac{a}{x^{a+1}} dx$$

$$= \lim_{t\to+\infty} a\int_1^t x^{-a-1} dx = \lim_{t\to+\infty} a\left[\frac{x^{-a}}{-a}\right]_1^t$$

$$= \lim_{t\to+\infty}\left[\frac{-1}{x^a}\right]_1^t = \left[\left(\frac{-1}{\infty^a}\right) - (-1)\right] = 1.$$

Thus, $f(x)$ is a probability density function. ■

The graph of a Basic Pareto distribution is given by

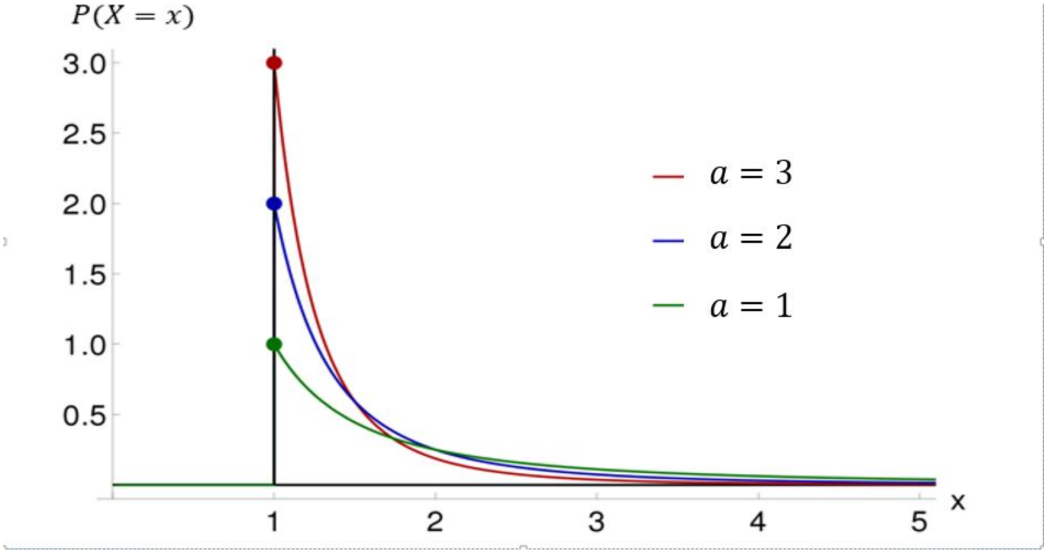

Figure 2.1: The Graph of a Basic Pareto Distribution

Example 2.1.1.

Let X be a continuous random variable having Basic Pareto Distribution with $a = 2$.

Then,

a) Compute $P(1 \le X \le 3)$.

Solution.

We know that;

$f(x)$ is defined by $f(x) = \frac{a}{x^{a+1}}$, then if $a = 2$, we obtain that;

$f(x) = \frac{2}{x^3}$. According to this, we have

$$P(1 \le X \le 3) = \int_1^3 \frac{2}{x^3} dx = 2\int_1^3 x^{-3} dx$$

$$= 2\left[\frac{x^{-2}}{-2}\right]_1^3 = \left[\frac{-1}{x^2}\right]_1^3$$

$$=\left(\frac{-1}{9}\right)-(-1)=\frac{8}{9}.$$

b) Compute $P(X<10)$.

Solution.

We know that: $f(x)$ was defined by $f(x)=\frac{a}{x^{a+1}}$ previously. In addition, $a=2$ was given in question. Thus, this means that, $f(x)=\frac{2}{x^3}$.

It follows that,

$$P(X<10)=\int_1^{10}\frac{2}{x^3}dx=2\int_1^{10}x^{-3}dx$$

$$=2\left[\frac{x^{-2}}{-2}\right]_1^{10}=\left[\frac{-1}{x^2}\right]_1^{10}$$

$$=\left[\left(\frac{-1}{10^2}\right)-(-1)\right]=\frac{99}{100}.$$

c) Show that $P(X\geq 5)=\frac{1}{25}$.

Solution.

$f(x)=\frac{a}{x^{a+1}}$, and $a=2$. Then, we get $f(x)=\frac{2}{x^3}$.

So,

$$P(X\geq 5)=\int_5^{\infty}\frac{2}{x^3}dx=\lim_{t\to+\infty}\int_5^{t}\frac{2}{x^3}dx$$

$$=\lim_{t\to+\infty}2\int_5^{t}x^{-3}dx=\lim_{t\to+\infty}2\left[\frac{x^{-2}}{-2}\right]_5^{t}$$

$$=\lim_{t\to+\infty}\left[\frac{-1}{x^2}\right]_5^{t}=\left[\left(\frac{-1}{\infty^2}\right)-\left(\frac{-1}{25}\right)\right]$$

$$= \frac{1}{25}.$$

Proposition 2.1.1. [1], [2]

The cumulative distribution function (cdf); $F(x)$, of a Basic Pareto Distribution is

$$F(x) = 1 - \frac{1}{x^a}, \quad x \geq 1, a > 0.$$

Proof.

We know that to obtain the Cumulative distribution function of Basic Pareto distribution, it is enough to show that $F(x) = \int_{-\infty}^{x} f(s)ds.$

Now, we will show that $F(x) = \int_{-\infty}^{x} f(s)ds$ where $f(x) = \frac{a}{x^{a+1}}, \quad x \geq 1, a > 0.$

$$F(x) = \int_{-\infty}^{x} f(s)ds = \int_{1}^{x} f(s)ds$$

$$= \int_{1}^{x} \frac{a}{s^{a+1}} ds = a\int_{1}^{x} s^{-a-1} ds$$

$$= a\left[\frac{s^{-a}}{-a}\right]_1^x = \left[\frac{-1}{s^a}\right]_1^x$$

$$= \left[\left(\frac{-1}{x^a}\right) - (-1)\right] = 1 - \frac{1}{x^a}.$$

Hence, function $F(x)$ is a cumulative distribution function. ■

Note that if $a = 1$ the Basic Pareto Distribution is called Standard Pareto Distribution. In this case;

a) The probability density function of the Standard Pareto Distribution is of the form

$$f(x) = \frac{1}{x^2}, \quad x \geq 1.$$

b) The cumulative distribution function of the Standard Pareto Distribution is of the form

$$F(x) = 1 - \frac{1}{x}, \quad x \geq 1.$$

Definition 2.1.2. [1], [2]

The inverse of a cumulative distribution function is called Quantile Function.

Quantile Function is used to determine quartiles, deciles and percentiles of a probability distribution.

Proposition 2.1.2. [1], [2]

The quantile function; $F^{-1}(p)$, of a Basic Pareto Distribution is

$$F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \leq p < 1, a > 0.$$

Proof.

Let $F(p) = 1 - \frac{1}{p^a}, \quad p \geq 1, a > 0.$ (We used cumulative distribution function to obtain this)

Now, we should show that $F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \leq p < 1, a > 0.$

Let $y = F(p) = 1 - \frac{1}{p^a}, \quad p \geq 1, a > 0.$

This means that,

$$\frac{1}{p^a} = 1 - y \Rightarrow 1 = (1-y)p^a$$

$$\Rightarrow p^a = \frac{1}{1-y}$$

$$\Rightarrow \sqrt[a]{p^a} = \frac{\sqrt[a]{1}}{\sqrt[a]{1-y}}$$

$$\Rightarrow p = \frac{1}{\sqrt[a]{1-y}}$$

$$\Rightarrow y = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0 \qquad \text{(Note that } y \to p\text{)}$$

Hence, $F^{-1}(p) = \dfrac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$ ■

Proposition 2.1.3. [1], [2]

The first quartile; q_1, of a Basic Pareto Distribution is

$$q_1 = \left(\frac{4}{3}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

First of all, we should use the quantile function for this proof.

Quantile function is $F^{-1}(p) = \dfrac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$

We take $p = \frac{1}{4}$ (because of first quartile) and substitute into the quantile function.

Then, we have

$$q_1 = \frac{1}{\left(1-\frac{1}{4}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{3}{4}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{4}{3}\right)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Proposition 2.1.4. [1], [2]

The second quartile; q_2, of a Basic Pareto Distribution is

$$q_2 = 2^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We should take $p = \frac{1}{2}$ and substitute into the quantile function

$$F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$$

Then, we have

$$q_2 = \frac{1}{\left(1-\frac{1}{2}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{1}{2}\right)^{\frac{1}{a}}}$$

$$= 2^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Note that the second quartile also equals to the median.

Proposition 2.1.5. [1], [2]

The third quartile; q_3, of a Basic Pareto Distribution is

$$q_3 = 4^{\frac{1}{a}}, \quad a > 0.$$

Proof.

Let $F^{-1}(p) = \dfrac{1}{(1-p)^{\frac{1}{a}}}, \quad p \in [0,1).$

Then, we can take $p = \dfrac{3}{4}$, because of third quartile.

It follows that

$$q_3 = \frac{1}{\left(1-\frac{3}{4}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{1}{4}\right)^{\frac{1}{a}}}$$

$$= 4^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Definition 2.1.3. [1], [2]

The interquartile range of a probability distribution is denoted by IQR, and it is defined by $IQR = q_3 - q_1$.

Proposition 2.1.6. [1], [2]

The interquartile range; IQR, of a Basic Pareto Distribution is

$$IQR = (4)^{\frac{1}{a}} - \left(\frac{4}{3}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We know that the interquartile range is $IQR = q_3 - q_1$.

We already obtained q_1 and q_3 as $q_1 = \left(\frac{4}{3}\right)^{\frac{1}{a}}$ and $q_3 = (4)^{\frac{1}{a}}$.

Thus, $IQR = q_3 - q_1 = (4)^{\frac{1}{a}} - \left(\frac{4}{3}\right)^{\frac{1}{a}}, \quad a > 0.$ $\blacksquare$

Proposition 2.1.7.

The first decile; d_1, of a Basic Pareto Distribution is

$$d_1 = \left(\frac{10}{9}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We take $p = \frac{1}{10}$ and substitute into the quantile function;

$$F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$$

Then, we have

$$d_1 = \frac{1}{\left(1-\frac{1}{10}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{9}{10}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{10}{9}\right)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Proposition 2.1.8.

The second decile; d_2, of a Basic Pareto Distribution is

$$d_2 = \left(\frac{10}{8}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

Now, we let $p = \frac{2}{10}$ and substitute it into the quantile function;

$$F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$$

So, it follows that

$$d_2 = \frac{1}{\left(1 - \frac{2}{10}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{8}{10}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{10}{8}\right)^{\frac{1}{a}} = \left(\frac{5}{4}\right)^{\frac{1}{a}}, \quad a > 0.$$ ■

Proposition 2.1.9.

The third decile; d_3, of a Basic Pareto Distribution is

$$d_3 = \left(\frac{10}{7}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

To start with,

Take $p = \frac{3}{10}$ in the quantile function.

Then, we have

$$d_3 = \frac{1}{\left(1 - \frac{3}{10}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{7}{10}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{10}{7}\right)^{\frac{1}{a}}, \quad a > 0.$$ ■

Proposition 2.1.10.

The fourth decile; d_4, of a Basic Pareto Distribution is

$$d_4 = \left(\frac{5}{3}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We know that;

We must take $p = \frac{4}{10}$ in the quantile function for this proof.

Thus, $d_4 = \frac{1}{\left(1 - \frac{4}{10}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{6}{10}\right)^{\frac{1}{a}}}$

$$= \left(\frac{10}{6}\right)^{\frac{1}{a}} = \left(\frac{5}{3}\right)^{\frac{1}{a}}, \quad a > 0.$$ ■

Proposition 2.1.11.

The fifth decile; d_5, of a Basic Pareto Distribution is

$$d_5 = 2^{\frac{1}{a}}, \quad a > 0.$$

Proof.

Take $p = \frac{5}{10}$ (because of fifth decile) in the quantile function.

So, we have

$$d_5 = \frac{1}{\left(1 - \frac{5}{10}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{5}{10}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{10}{5}\right)^{\frac{1}{a}} = 2^{\frac{1}{a}}, \quad a > 0.$$ ■

Property 2.1.1.

The fifth decile equals second quartile which is also median. That is

$$d_5 = q_2 = \text{median}.$$

In a similar way, we can easily find sixth; d_6, seventh; d_7, eighth; d_8, and ninth; d_9, deciles.

Proposition 2.1.12.

The ninth decile; d_9, of a Basic Pareto Distribution is

$$d_9 = (10)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We take $p = \frac{9}{10}$ and substitute it into the quantile function;

It follows that,

$$d_9 = \frac{1}{\left(1 - \frac{9}{10}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{1}{10}\right)^{\frac{1}{a}}} = (10)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Proposition 2.1.13.

The first percentile; p_1, of a Basic Pareto Distribution is

$$p_1 = \left(\frac{100}{99}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We know that,

If we let $p = \frac{1}{100}$ in the quantile function (because, p is first percentile)

Then, we have

$$p_1 = \frac{1}{\left(1 - \frac{1}{100}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{99}{100}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{100}{99}\right)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Note that, quantile function is $F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$

Proposition 2.1.14.

The second percentile; p_2, of a Basic Pareto Distribution is

$$p_2 = \left(\frac{100}{98}\right)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

First of all,

Let $p = \frac{2}{100}$ in the quantile function because of the second percentile.

This implies that,

$$p_2 = \frac{1}{\left(1 - \frac{2}{100}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{98}{100}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{100}{98}\right)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Proposition 2.1.15.

The fiftieth (median) percentile; p_{50}, of a Basic Pareto Distribution is

$$p_{50} = 2^{\frac{1}{a}}, \quad a > 0.$$

Proof.

To begin with,

We must take $p = \frac{50}{100}$ in the quantile function for proof.

On the other hand, quantile function is $F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$

Hence,

$$p_{50} = \frac{1}{\left(1-\frac{50}{100}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{50}{100}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{100}{50}\right)^{\frac{1}{a}} = 2^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Property 2.1.2.

Obviously, we see that the second quartile, fifth decile and fiftieth percentile are equal.

Moreover, $q_2 = d_5 = p_{50} = 2^{\frac{1}{a}}$ for Basic Pareto Distribution.

Proposition 2.1.16.

The ninetieth percentile; p_{90}, of a Basic Pareto Distribution is

$$p_{90} = (10)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

Take as $p = \frac{90}{100}$ in the quantile function.

It follows that,

$$p_{90} = \frac{1}{\left(1 - \frac{90}{100}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{10}{100}\right)^{\frac{1}{a}}}$$

$$= \left(\frac{100}{10}\right)^{\frac{1}{a}} = (10)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Proposition 2.1.17.

The last percentile; p_{99}, of a Basic Pareto Distribution is

$$p_{99} = (100)^{\frac{1}{a}}, \quad a > 0.$$

Proof.

We take $p = \frac{99}{100}$ and substitute it into the quantile function;

$$F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0.$$

Thus,

$$p_{99} = \frac{1}{\left(1 - \frac{99}{100}\right)^{\frac{1}{a}}} = \frac{1}{\left(\frac{1}{100}\right)^{\frac{1}{a}}}$$

$$= (100)^{\frac{1}{a}}, \quad a > 0. \qquad \blacksquare$$

Proposition 2.1.18. [1], [2]

The expectation; $E(X)$, of a Basic Pareto Distribution is

$$E(X) = \frac{a}{a-1}, \quad a > 1.$$

Proof.

The expectation is obtained by $E(X)=\int_a^b x\,f(x)dx$, where $f(x)=\frac{a}{x^{a+1}},\quad x\geq 1, a>1.$

This implies that,

$$E(X)=\int_1^{\infty} x\frac{a}{x^{a+1}}dx$$

$$=a\int_1^{\infty} x^{-a}dx$$

$$=\lim_{t\to+\infty} a\int_1^{t} x^{-a}dx$$

$$=\lim_{t\to+\infty} a\left[\frac{x^{-a+1}}{-a+1}\right]_1^t$$

$$=\lim_{t\to+\infty} \frac{a}{-a+1}\left[\frac{1}{x^{a-1}}\right]_1^t$$

$$=\lim_{t\to+\infty} \frac{a}{-a+1}\left[\left(\frac{1}{t^{a-1}}\right)-\left(\frac{1}{1^{a-1}}\right)\right]$$

$$=\frac{a}{-a+1}\left[\frac{1}{\infty^{a-1}}-\frac{1}{1^{a-1}}\right]$$

$$=\frac{-a}{-a+1}$$

$$=\frac{a}{a-1},\quad a>1.$$ ■

Proposition 2.1.19. [1], [2]

The variance; $Var(X)$, of a Basic Pareto Distribution is

$$Var(X)=\frac{a}{(a-1)^2(a-2)},\quad a>2.$$

Proof.

We know that: $Var(X)$ is obtained by

$$Var(X) = \int_a^b x^2 f(x)dx - \left[E(X)\right]^2.$$

This implies that,

$$Var(X) = \int_1^\infty x^2 \frac{a}{x^{a+1}} dx - \left(\frac{a}{a-1}\right)^2$$

$$= a\int_1^\infty x^2 x^{-a-1} dx - \left(\frac{a}{a-1}\right)^2$$

$$= a\int_1^\infty x^{-a+1} dx - \left(\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} a\int_1^t x^{-a+1} dx - \left(\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} a\left[\frac{x^{-a+2}}{-a+2}\right]_1^t - \left(\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} \frac{a}{-a+2}\left[\frac{1}{x^{a-2}}\right]_1^t - \left(\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} \frac{a}{-a+2}\left[\left(\frac{1}{t^{a-2}}\right) - \left(\frac{1}{1^{a-2}}\right)\right] - \left(\frac{a}{a-1}\right)^2$$

$$= \frac{-a}{-a+2} - \left(\frac{a}{a-1}\right)^2 = \frac{a}{a-2} - \left(\frac{a}{a-1}\right)^2$$

$$= \frac{a}{(a-1)^2(a-2)}, \quad a > 2.$$ ■

Proposition 2.1.20. [1], [2]

The generalized expectation; $E(X^n)$ of a Basic Pareto Distribution is

$$E(X^n) = \begin{cases} \dfrac{a}{a-n}, & 0 \le n < a, \quad a > 0. \\ \infty, & n \ge a, \quad a > 0. \end{cases}$$

Proof.

It is known that $E(X^n)=\int_a^b x^n f(x)dx.$

According to the above information, we have

$$E(X^n)=\int_1^\infty x^n \frac{a}{x^{a+1}}dx$$

$$=a\int_1^\infty x^{n-a-1}dx$$

$$=\lim_{t\to+\infty} a\int_1^t x^{n-a-1}dx$$

$$=\lim_{t\to+\infty} a\left[\frac{x^{n-a}}{n-a}\right]_1^t$$

$$=\lim_{t\to+\infty} \frac{a}{n-a}\left[\frac{1}{x^{-n+a}}\right]_1^t$$

$$=\lim_{t\to+\infty} \frac{a}{n-a}\left[\left(\frac{1}{t^{-n+a}}\right)-\left(\frac{1}{1^{-n+a}}\right)\right].$$

Here, there are two important states.

State 1: If $0<n<a \Rightarrow \lim_{t\to+\infty} \frac{a}{n-a}\left[\left(\frac{1}{t^{-n+a}}\right)-\left(\frac{1}{1^{-n+a}}\right)\right]=\frac{-a}{n-a}$

$$=\frac{a}{a-n}. \qquad \blacksquare$$

State 2: If $n\geq a \Rightarrow \lim_{t\to+\infty} \frac{a}{n-a}\left[\left(\frac{1}{t^{-n+a}}\right)-\left(\frac{1}{1^{-n+a}}\right)\right]=\infty.$ $\blacksquare$

Example 2.1.2. [1], [2]

Suppose that the income of a certain population has the Pareto Distribution with shape parameter $a=3.$ Find each of the following:

a) The proportion of the population with incomes between 2000 and $4000.$ [1], [2]

Solution.

We know that;

$f(x)$ is defined by $f(x)=\frac{a}{x^{a+1}}$, then if $a=3,$ we obtain that;

$f(x)=\frac{3}{x^4}$. According to this, we have

$$P(2000\le X\le 4000)=\int_{2000}^{4000}\frac{3}{x^4}dx$$

$$=3\int_{2000}^{4000}x^{-4}dx$$

$$=3\left[\frac{x^{-3}}{-3}\right]_{2000}^{4000}$$

$$=\left[-\frac{1}{x^3}\right]_{2000}^{4000}$$

$$=\left[\left(-\frac{1}{4000^3}\right)-\left(-\frac{1}{2000^3}\right)\right]$$

$$=1.09375\times 10^{-10}.$$

b) The median income. [1], [2]

Solution.

We can use second quartile or fifth decile or fiftieth percentile. Therefore, median is

$2^{\frac{1}{a}}$, where $a=3.$

This implies that,

Median income $=2^{\frac{1}{3}}\cong 1.2599.$

c) The interquartile range. [1], [2]

Solution.

To find interquartile range we need first and third quartiles.

First quartile is $q_1 = \left(\frac{4}{3}\right)^{\frac{1}{a}}$, where $a = 3$.

So, $q_1 = \left(\frac{4}{3}\right)^{\frac{1}{3}}$

$\cong 1.1006.$

Third quartile is $q_3 = (4)^{\frac{1}{a}}$.

Thus,

$$q_3 = (4)^{\frac{1}{3}}$$

$$\cong 1.5874.$$

Then, the interquartile range is obtained by

$$IQR = q_3 - q_1$$

$$= 1.5874 - 1.1006$$

$$= 0.4868.$$

d) The mean income. [1], [2]

Solution.

We should compute expectation to obtain mean income.

We know that: Expectation is $E(X) = \frac{a}{a-1}$, where $a = 3$.

Therefore,

$$E(X) = \frac{3}{2}$$

$$= 1.5.$$

e) The standard deviation of income. [1], [2]

Solution.

We know that;

Standard deviation is $\sigma = \sqrt{Var(X)}$, where $Var(X) = \dfrac{a}{(a-1)^2(a-2)}$

So, $Var(X) = \dfrac{3}{4}$

$= 0.75$

Hence, standard deviation is $\sigma = \sqrt{0.75} \cong 0.8660.$

f) The $90th$ percentile. [1], [2]

Solution.

We know that: $90th$ percentile formula of a Basic Pareto Distribution is

$p_{90} = (10)^{\frac{1}{a}}$, where $a = 3.$

Therefore,

$$p_{90} = (10)^{\frac{1}{3}}$$

$$\cong 2.1544.$$

Proposition 2.1.21. [2]

Suppose that X is a continuous random variable having a Standard Pareto Distribution. Then, the continuous random variable Y defined by $Y = X^{\frac{1}{a}}, \quad a > 0$ has a Basic Pareto Distribution with shape parameter $a.$

Proof.

Suppose X is a continuous random variable having a Standard Pareto Distribution.

Let Y be a continuous random variable defined by $Y = X^{\frac{1}{a}}, \quad a > 0.$

On the other hand,

$$Y = X^{\frac{1}{a}} \Rightarrow y = x^{\frac{1}{a}}$$

$$\Rightarrow y^a = \left(x^{\frac{1}{a}} \right)^a$$

$$\Rightarrow y^a = x$$

$$\Rightarrow x = y^a$$

This implies that,

$$P\left(X \le x \right) = P\left(X \le y^a \right)$$

$$= \int_1^{y^a} f(x) dx$$

$$= F(x) \Big|_1^{y^a}$$

$$= \left[1 - \frac{1}{x} \right]_1^{y^a}, \quad x \ge 1.$$

$$= 1 - \frac{1}{y^a}, \quad y \ge 1, a > 0.$$

Here, as a function of y, this is a cumulative distribution function for a Basic Pareto Distribution $Y = X^{\frac{1}{a}}$ with shape parameter a. ■

Proposition 2.1.22. [2]

Suppose that X is a continuous random variable having a Basic Pareto Distribution with shape parameter $a > 0$. Then, the continuous random variable Y defined by $Y = X^n, \quad n > 0$ has a Basic Pareto Distribution with shape parameter $\frac{a}{n}$.

Proof.

Suppose X is a continuous random variable having a Basic Pareto Distribution with shape parameter $a > 0$.

Let Y be a continuous random variable defined by $Y = X^n, \quad n > 0.$

On the other hand,

$$Y = X^n \Rightarrow y = x^n$$

$$\Rightarrow \sqrt[n]{y} = \sqrt[n]{x^n}$$

$$\Rightarrow x = y^{\frac{1}{n}}$$

This implies that,

$$P(X \le x) = P\left(X \le y^{\frac{1}{n}} \right)$$

$$= \int_1^{y^{\frac{1}{n}}} f(x)dx$$

$$= F(x) \Big|_1^{y^{\frac{1}{n}}}$$

$$= \left[1 - \frac{1}{x^a} \right]_1^{y^{\frac{1}{n}}}, \quad x \ge 1$$

$$= 1 - \frac{1}{\left(y^{\frac{1}{n}} \right)^a}, \quad y \ge 1.$$

$$= 1 - \frac{1}{y^{\frac{a}{n}}}, \quad y \ge 1, a > 0, n > 0.$$

Here, as a function of y, this is a cumulative distribution function for a Basic Pareto Distribution $Y = X^n$ with shape parameter $\frac{a}{n}$. ■

2.2 General Pareto Distribution

Definition 2.2.1. [1], [2]

A probability density function; $f(x)$, of the form

$$f(x) = \frac{ab^a}{x^{a+1}}, \quad a > 0, b > 0 \text{ and } x \geq b$$

where a is shape parameter and b is scale parameter; is called General Pareto distribution.

Proof.

Let us prove that $f(x)$ is a probability density function.

We should show that $\int_b^\infty f(x)dx = 1.$

$$\int_b^\infty \frac{ab^a}{x^{a+1}}dx = \lim_{t\to+\infty} \int_b^t \frac{ab^a}{x^{a+1}}dx$$

$$= \lim_{t\to+\infty} ab^a \int_b^t x^{-a-1}dx$$

$$= \lim_{t\to+\infty} ab^a \left[\frac{x^{-a}}{-a}\right]_b^t$$

$$= \lim_{t\to+\infty} b^a \left[\frac{-1}{x^a}\right]_b^t$$

$$= \lim_{t\to+\infty} b^a \left[\left(\frac{-1}{t^a}\right) - \left(\frac{-1}{b^a}\right)\right] = 1.$$

Thus, $f(x)$ is a probability density function. ■

The graph of a General Pareto Distribution is given by

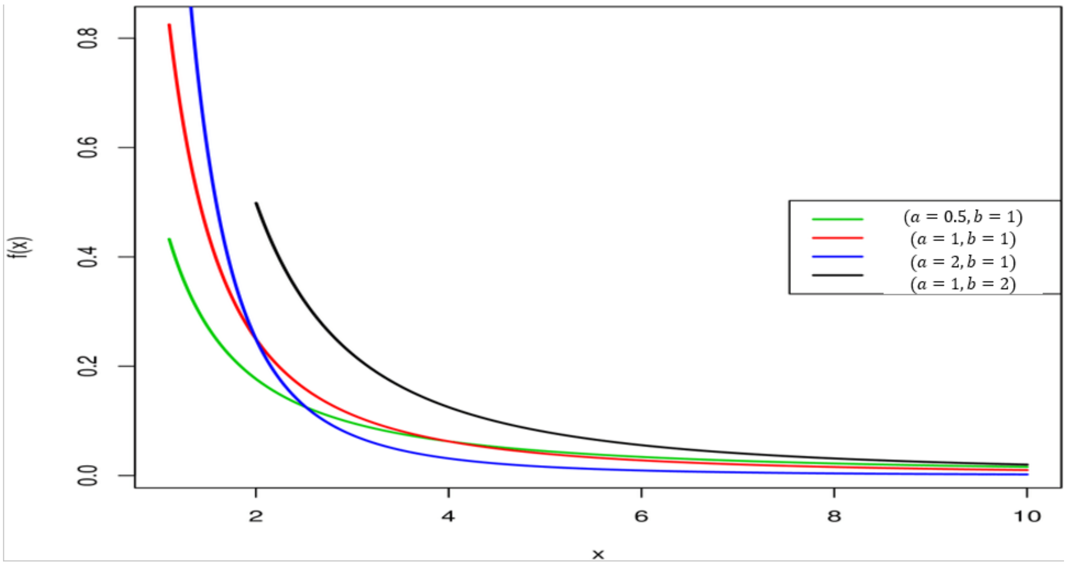

Figure 2.2: The Graph of a General Pareto Distribution

Example 2.2.1.

Let X be a continuous random variable having a General Pareto Distribution with $a=8$ and $b=2$. Then,

a) Compute $P(X \geq 5)$.

Solution.

We know that;

$f(x)$ is defined by $f(x)=\frac{ab^a}{x^{a+1}}$, then if $a=8$, and $b=2$, we obtain that;

$f(x)=\frac{8.2^8}{x^9}$. According to this, we have

$$P(X \geq 5)=\int_5^{\infty}\frac{2048}{x^9}\,dx$$

$$=2048\int_5^{\infty}x^{-9}dx$$

$$=\lim_{t\to+\infty}2048\left[\frac{x^{-8}}{-8}\right]_5^t$$

$$= \lim_{t\to+\infty} \frac{2048}{-8}\left[\frac{1}{x^8}\right]_5^t$$

$$= \lim_{t\to+\infty} \frac{2048}{-8}\left[\frac{1}{t^8}-\frac{1}{5^8}\right]$$

$$= \frac{2048}{-8}.\frac{-1}{5^8}$$

$$= 0.000655.$$

b) Compute $P(X<3)$.

Solution.

First of all,

We know that: $f(x)$ was defined by $f(x)=\frac{ab^a}{x^{a+1}}$ previously. In addition, $a=8$, and $b=2$ were given in question. Thus, this means that, $f(x)=\frac{8.2^8}{x^9}$.

It follows that,

$$P(X<3)=\int_2^3 \frac{8.2^8}{x^9}dx$$

$$= 2048\int_2^3 x^{-9}dx$$

$$= 2048\left[\frac{x^{-8}}{-8}\right]_2^3$$

$$= \frac{2048}{-8}\left[\frac{1}{x^8}\right]_2^3$$

$$= \frac{2048}{-8}\left[\frac{1}{3^8}-\frac{1}{2^8}\right]$$

$$= \frac{6305}{6561}$$

$$= 0.96.$$

c) Show that $P\left(X \geq 2\right) = 1$.

Solution.

$f(x) = \frac{ab^a}{x^{a+1}}$, and $a = 8, b = 2$. Then, we get;

$f(x) = \frac{8.2^8}{x^9}$. So,

$$P\left(X \geq 2\right) = \int_2^{\infty} \frac{8.2^8}{x^9} dx$$

$$= \lim_{t \to +\infty} \int_2^{t} \frac{8.2^8}{x^9} dx$$

$$= \lim_{t \to +\infty} 8.2^8 \int_2^{t} x^{-9} dx$$

$$= \lim_{t \to +\infty} 8.2^8 \left[\frac{x^{-8}}{-8} \right]_2^t$$

$$= \lim_{t \to +\infty} 2^8 \left[\frac{-1}{x^8} \right]_2^t$$

$$= \lim_{t \to +\infty} 2^8 \left[\left(\frac{-1}{t^8} \right) - \left(\frac{-1}{2^8} \right) \right]$$

$$= 2^8 . \frac{1}{2^8}$$

$$= 1.$$

Proposition 2.2.1. [1], [2]

The cumulative distribution function (cdf); $F(x)$, of a General Pareto Distribution is

$$F(x) = 1 - \left(\frac{b}{x} \right)^a, \quad a > 0, b > 0, x \geq b.$$

Proof.

We know that $F(x)=\int f(x)dx.$

Now, we will show that $F(x)=\int f(x)dx$, where $f(x)=\frac{ab^a}{x^{a+1}}, \quad a>0, b>0, x\geq b.$

$$F(x)=\int f(x)dx$$

$$=\int_b^x f(t)dt$$

$$=\int_b^x \frac{ab^a}{t^{a+1}}dt$$

$$=ab^a\int_b^x t^{-a-1}dt$$

$$=ab^a\left[\frac{t^{-a}}{-a}\right]_b^x$$

$$=-b^a\left[\frac{1}{t^a}\right]_b^x$$

$$=-b^a\left[\left(\frac{1}{x^a}\right)-\left(\frac{1}{b^a}\right)\right]$$

$$=\frac{-b^a}{x^a}+1$$

$$=1-\left(\frac{b}{x}\right)^a, \quad a>0, b>0, x\geq b.$$

Hence, function $F(x)$ is a cumulative distribution function. ■

Definition 2.2.2. [1], [2]

The inverse of a cumulative distribution function is called Quantile Function.

Quantile Function is used to determine quartiles, deciles and percentiles of a probability distribution.

Proposition 2.2.2. [1], [2]

The quantile function; $F^{-1}(p)$, of a General Pareto Distribution is

$$F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$$

Proof.

Let $F(p) = 1 - \left(\frac{b}{p}\right)^{a}, \quad a > 0, b > 0,\ p \ge b.$ (We used cumulative distribution function to obtain this)

Now, we should show that $F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$

Let $y = F(p) = 1 - \left(\frac{b}{p}\right)^{a}, \quad a > 0, b > 0,\ p \ge b.$

This means that,

$$\frac{b^a}{p^a} = 1 - y \Rightarrow b^a = (1-y)p^a$$

$$\Rightarrow p^a = \frac{b^a}{1-y}$$

$$\Rightarrow \sqrt[a]{p^a} = \frac{\sqrt[a]{b^a}}{\sqrt[a]{1-y}}$$

$$\Rightarrow p = \frac{b}{\sqrt[a]{1-y}}$$

$$\Rightarrow y = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1 \qquad \text{(Note that } y \to p\text{)}$$

Hence, $F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$ ■

Proposition 2.2.3. [1], [2]

The first quartile; q_1, of a General Pareto Distribution is

$$q_1 = b\left(\frac{4}{3}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

To begin with,

We should use the quantile function for this proof.

Quantile function is $F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$

We take $p = \frac{1}{4}$ (because of first quartile) and substitute into the quantile function.

Then, we have

$$q_1 = \frac{b}{\left(1-\frac{1}{4}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{3}{4}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{4}{3}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Proposition 2.2.4. [1], [2]

The second quartile; q_2, of a General Pareto Distribution is

$$q_2 = b(2)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We should take $p=\frac{1}{2}$ and substitute into the quantile function

$$F^{-1}(p)=\frac{b}{(1-p)^{\frac{1}{a}}},\quad 0\le p<1, a>0, b>0.$$

Then, we have

$$q_2=\frac{b}{\left(1-\frac{1}{2}\right)^{\frac{1}{a}}}$$

$$=\frac{b}{\left(\frac{1}{2}\right)^{\frac{1}{a}}}$$

$$=b(2)^{\frac{1}{a}},\quad a>0, b>0.$$ ■

Note that the second quartile also equals to the median.

Proposition 2.2.5. [1], [2]

The third quartile; q_3, of a General Pareto Distribution is

$$q_3=b(4)^{\frac{1}{a}},\quad a>0, b>0.$$

Proof.

Let $F^{-1}(p)=\frac{b}{(1-p)^{\frac{1}{a}}},\quad p\in[0,1), a>0, b>0.$

Then, we can take $p=\frac{3}{4}$, because of third quartile.

It follows that

$$q_3 = \frac{b}{\left(1 - \frac{3}{4}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{1}{4}\right)^{\frac{1}{a}}}$$

$$= b(4)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Definition 2.2.3. [1], [2]

The interquartile range of a probability distribution is denoted by IQR, and it is defined by $IQR = q_3 - q_1$.

Proposition 2.2.6. [1], [2]

The interquartile range; IQR, of a General Pareto Distribution is

$$IQR = b\left[(4)^{\frac{1}{a}} - \left(\frac{4}{3}\right)^{\frac{1}{a}}\right], \quad a > 0, b > 0.$$

Proof.

We know that the interquartile range is $IQR = q_3 - q_1$.

We already know q_1 and q_3 such that $q_1 = b\left(\frac{4}{3}\right)^{\frac{1}{a}}$ and $q_3 = b(4)^{\frac{1}{a}}$.

Thus, $IQR = q_3 - q_1 = b\left[(4)^{\frac{1}{a}} - \left(\frac{4}{3}\right)^{\frac{1}{a}}\right], \quad a > 0, b > 0.$ ■

Proposition 2.2.7.

The first decile; d_1, of a General Pareto Distribution is

$$d_1 = b\left(\frac{10}{9}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We take $p = \frac{1}{10}$ and substitute into the quantile function;

$$F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$$

Then, we have

$$d_1 = \frac{b}{\left(1-\frac{1}{10}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{9}{10}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{10}{9}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

■

Proposition 2.2.8.

The second decile; d_2, of a General Pareto Distribution is

$$d_2 = b\left(\frac{10}{8}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We know that;

We take $p = \frac{2}{10}$ and substitute into the quantile function;

$$F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$$

So, it follows that

$$d_2 = \frac{b}{\left(1-\frac{2}{10}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{8}{10}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{10}{8}\right)^{\frac{1}{a}}, \quad a > 0, b > 0. \qquad \blacksquare$$

Proposition 2.2.9.

The third decile; d_3, of a General Pareto Distribution is

$$d_3 = b\left(\frac{10}{7}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

Firstly,

We take $p = \frac{3}{10}$ in the quantile function.

Then, we have

$$d_3 = \frac{b}{\left(1-\frac{3}{10}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{7}{10}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{10}{7}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Proposition 2.2.10.

The fourth decile; d_4, of a General Pareto Distribution is

$$d_4 = b\left(\frac{10}{6}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We know that;

We must take $p = \frac{4}{10}$ in the quantile function for this proof.

Thus, $d_4 = \frac{b}{\left(1-\frac{4}{10}\right)^{\frac{1}{a}}}$

$$= \frac{b}{\left(\frac{6}{10}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{10}{6}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Proposition 2.2.11.

The fifth decile; d_5, of a General Pareto Distribution is

$$d_5 = b(2)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

Take as $p = \frac{5}{10}$ (because of fifth decile) in the quantile function.

So, we have

$$d_5 = \frac{b}{\left(1 - \frac{5}{10}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{5}{10}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{10}{5}\right)^{\frac{1}{a}}$$

$$= b(2)^{\frac{1}{a}}, \quad a > 0, b > 0. \qquad \blacksquare$$

Property 2.2.1.

The fifth decile equals second quartile which is also median. That is

$$d_5 = q_2 = \text{median}$$

In a similar way, we can easily find sixth; d_6, seventh; d_7, eighth; d_8, and ninth; d_9, deciles.

Proposition 2.2.12.

The ninth decile; d_9, of a General Pareto Distribution is

$$d_9 = b(10)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We take $p = \frac{9}{10}$ and substitute into the quantile function;

It follows that,

$$d_9 = \frac{b}{\left(1 - \frac{9}{10}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{1}{10}\right)^{\frac{1}{a}}}$$

$$= b(10)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Proposition 2.2.13.

The first percentile; p_1, of a General Pareto Distribution is

$$p_1 = b\left(\frac{100}{99}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

If we take $p = \frac{1}{100}$ in the quantile function (because, p is first percentile)

Then, we have

$$p_1 = \frac{b}{\left(1 - \frac{1}{100}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{99}{100}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{100}{99}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Note that, quantile function is $F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$

Proposition 2.2.14.

The second percentile; p_2, of a General Pareto Distribution is

$$p_2 = b\left(\frac{100}{98}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

Let $p = \frac{2}{100}$ in the quantile function because of the second percentile.

This implies that,

$$p_2 = \frac{b}{\left(1 - \frac{2}{100}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{98}{100}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{100}{98}\right)^{\frac{1}{a}}, \quad a > 0, b > 0.$$ ■

Proposition 2.2.15.

The fiftieth (median) percentile; p_{50} , of a General Pareto Distribution is

$$p_{50} = b(2)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We must take $p = \frac{50}{100}$ in the quantile function for proof.

On the other hand, quantile function is $F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$

Hence,

$$p_{50} = \frac{b}{\left(1 - \frac{50}{100}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{50}{100}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{100}{50}\right)^{\frac{1}{a}}$$

$$= b(2)^{\frac{1}{a}}, \quad a > 0, b > 0. \qquad \blacksquare$$

Property 2.2.2.

Obviously, we see that the second quartile, fifth decile and fiftieth percentile are equal.

Moreover, $q_2 = d_5 = p_{50} = b(2)^{\frac{1}{a}}$ for General Pareto Distribution.

Proposition 2.2.16.

The ninetieth percentile; p_{90}, of a General Pareto Distribution is

$$p_{90} = b(10)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We should take $p = \frac{90}{100}$ in the quantile function.

It follows that,

$$p_{90} = \frac{b}{\left(1 - \frac{90}{100}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{10}{100}\right)^{\frac{1}{a}}}$$

$$= b\left(\frac{100}{10}\right)^{\frac{1}{a}}$$

$$= b(10)^{\frac{1}{a}}, \quad a > 0, b > 0. \qquad \blacksquare$$

Proposition 2.2.17.

The last percentile; p_{99}, of a General Pareto Distribution is

$$p_{99} = b(100)^{\frac{1}{a}}, \quad a > 0, b > 0.$$

Proof.

We take $p = \frac{99}{100}$ and substitute it into the quantile function;

$$F^{-1}(p) = \frac{b}{(1-p)^{\frac{1}{a}}}, \quad 0 \le p < 1, a > 0, b > 0.$$

Thus,

$$p_{99} = \frac{b}{\left(1-\frac{99}{100}\right)^{\frac{1}{a}}}$$

$$= \frac{b}{\left(\frac{1}{100}\right)^{\frac{1}{a}}}$$

$$= b(100)^{\frac{1}{a}}, \quad a>0, b>0. \qquad \blacksquare$$

Proposition 2.2.18. [1], [2]

The expectation; $E(X)$, of a General Pareto Distribution is

$$E(X) = \frac{ab}{a-1}, \quad a>1, b>0.$$

Proof.

The expectation is obtained by $E(X) = \int_a^b x\, f(x)dx$,

where $f(x) = \frac{ab^a}{x^{a+1}}, \quad a>1, b>0, x \geq b.$

This implies that,

$$E(X) = \int_b^{\infty} x \frac{ab^a}{x^{a+1}} dx$$

$$= \lim_{t\to+\infty} \int_b^t x \frac{ab^a}{x^{a+1}} dx$$

$$= \lim_{t\to+\infty} ab^a \int_b^t x^{-a} dx$$

$$= \lim_{t\to+\infty} ab^a \left[\frac{x^{-a+1}}{-a+1} \right]_b^t$$

$$= \lim_{t\to+\infty} \frac{ab^a}{-a+1}\left[\frac{1}{x^{a-1}}\right]_b^t$$

$$= \lim_{t\to+\infty} \frac{ab^a}{-a+1}\left[\left(\frac{1}{t^{a-1}}\right)-\left(\frac{1}{b^{a-1}}\right)\right]$$

$$= \frac{ab^a}{-a+1}\left[\frac{1}{\infty^{a-1}}-\frac{1}{b^{a-1}}\right]$$

$$= \left(\frac{ab^a}{-a+1}\right).\left(-b^{-a+1}\right)$$

$$= \frac{-ab}{-a+1}$$

$$= \frac{ab}{a-1}$$

$$= b\frac{a}{a-1}, \quad a>1, b>0.$$ ■

Proposition 2.2.19. [1], [2]

The variance; $Var(X)$, of a General Pareto Distribution is

$$Var(X) = \frac{ab^2}{(a-1)^2(a-2)}, \quad a>2, b>0.$$

Proof.

We know that: $Var(X)$ is obtained by

$$Var(X) = \int_a^b x^2 f(x)dx - \left[E(X)\right]^2.$$

This implies that,

$$Var(X) = \int_b^\infty x^2 \frac{ab^a}{x^{a+1}}dx - \left(b\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty}\int_b^t x^2 \frac{ab^a}{x^{a+1}}dx - \left(b\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} ab^a \int_b^t x^{-a+1}dx - \left(b\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} ab^a \left[\frac{x^{-a+2}}{-a+2}\right]_b^t - \left(b\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} \frac{ab^a}{-a+2}\left[\frac{1}{x^{a-2}}\right]_b^t - \left(b\frac{a}{a-1}\right)^2$$

$$= \lim_{t\to+\infty} \frac{ab^a}{-a+2}\left[\left(\frac{1}{t^{a-2}}\right) - \left(\frac{1}{b^{a-2}}\right)\right] - \left(b\frac{a}{a-1}\right)^2$$

$$= \left(\frac{ab^a}{-a+2}\right).\left(-b^{-a+2}\right) - \left(b\frac{a}{a-1}\right)^2$$

$$= \left(\frac{-ab^2}{-a+2}\right) - \left(b\frac{a}{a-1}\right)^2$$

$$= \left(\frac{ab^2}{a-2}\right) - \left(b\frac{a}{a-1}\right)^2$$

$$= \frac{ab^2}{(a-1)^2(a-2)}, \quad a>2, b>0. \qquad \blacksquare$$

Proposition 2.2.20. [1], [2]

The generalized expectation; $E(X^n)$, of a General Pareto Distribution is

$$E(X^n) = \begin{cases} b^n \dfrac{a}{a-n}, & 0 \le n < a, a>0, b>0. \\ \infty, & n \ge a, a>0, b>0. \end{cases}$$

Proof.

It is known that $E(X^n) = \int_a^b x^n f(x)dx.$

According to the above information, we have

$$E(X^n) = \int_b^\infty x^n \frac{ab^a}{x^{a+1}} dx$$

$$= ab^a \int_b^{\infty} x^{n-a-1} dx$$

$$= \lim_{t \to +\infty} ab^a \int_b^t x^{n-a-1} dx$$

$$= \lim_{t \to +\infty} ab^a \left[\frac{x^{n-a}}{n-a} \right]_b^t$$

$$= \lim_{t \to +\infty} \frac{ab^a}{n-a} \left[\frac{1}{x^{-n+a}} \right]_b^t$$

$$= \lim_{t \to +\infty} \frac{ab^a}{n-a} \left[\left(\frac{1}{t^{-n+a}} \right) - \left(\frac{1}{b^{-n+a}} \right) \right].$$

Here, there are two important states.

State 1: If $0 < n < a$ and $b > 0 \Rightarrow \lim_{t \to +\infty} \frac{ab^a}{n-a} \left[\left(\frac{1}{t^{-n+a}} \right) - \left(\frac{1}{b^{-n+a}} \right) \right] = \left(\frac{ab^a}{n-a} \right) . \left(-b^{-a+n} \right)$

$$= \frac{-ab^n}{n-a}$$

$$= \frac{ab^n}{a-n}$$

$$= b^n \frac{a}{a-n}. \qquad \blacksquare$$

State 2: If $n \geq a$ and b>0 $\Rightarrow \lim_{t \to +\infty} \frac{ab^a}{n-a} \left[\left(\frac{1}{t^{-n+a}} \right) - \left(\frac{1}{b^{-n+a}} \right) \right] = \infty.$ $\blacksquare$

Example 2.2.2. [1], [2]

Suppose that the income of a certain population has the Pareto Distribution with shape parameter $a = 3$ and scale parameter $b = 1000$.

Find each of the following:

a) The proportion of the population with incomes between 2000 and 4000. [1], [2]

Solution.

We know that;

$f(x)$ is defined by $f(x)=\frac{ab^a}{x^{a+1}}$, then if $a=3, b=1000$, we obtain that;

$f(x)=\frac{3(1000)^3}{x^4}$. According to this, we have

$$P(2000 \le X \le 4000) = \int_{2000}^{4000} \frac{3(1000)^3}{x^4} dx$$

$$= 3(1000)^3 \int_{2000}^{4000} x^{-4} dx$$

$$= 3(1000)^3 \left[\frac{x^{-3}}{-3} \right]_{2000}^{4000}$$

$$= -(1000)^3 \left[\frac{1}{x^3} \right]_{2000}^{4000}$$

$$= -(1000)^3 \left[\frac{1}{4000^3} - \frac{1}{2000^3} \right]$$

$$= 0.109375.$$

b) The median income. [1], [2]

Solution.

We can use second quartile or fifth decile or fiftieth percentile, Therefore, median is

$b(2)^{\frac{1}{a}}$, where $a=3, b=1000.$

This implies that,

Median income $= b(2)^{\frac{1}{a}}$

$$= 1000(2)^{\frac{1}{3}}$$

$$\cong 1259.92.$$

c) The interquartile range. [1], [2]

Solution.

To find interquartile range we need first and third quartiles.

First quartile is $q_1 = b\left(\frac{4}{3}\right)^{\frac{1}{a}}$, where $a = 3, b = 1000.$

So, $q_1 = 1000\left(\frac{4}{3}\right)^{\frac{1}{3}}$

$\cong 1100.64.$

Third quartile is $q_3 = b\left(4\right)^{\frac{1}{a}}$.

Thus,

$$q_3 = 1000\left(4\right)^{\frac{1}{3}}$$

$$\cong 1587.40.$$

Then, the interquartile range is obtained by

$$IQR = q_3 - q_1$$

$$= 1587.40 - 1100.64$$

$$= 486.76.$$

d) The mean income. [1], [2]

Solution.

We should compute expectation to obtain mean income.

We know that: Expectation is $E(X) = b\frac{a}{a-1}$, where $a = 3, b = 1000.$

Therefore,

$$E\left(X\right) = 1000\frac{3}{2}$$

$$= 1500.$$

e) The standard deviation of income. [1], [2]

Solution.

We know that;

Standard deviation is $\sigma = \sqrt{Var(X)}$, where $Var(X) = b^2 \dfrac{a}{(a-1)^2(a-2)}$

So, $Var(X) = 1000^2\left(\dfrac{3}{4}\right)$

$$= 750000$$

Hence, standard deviation is $\sigma = \sqrt{750000} \cong 866.$

f) The $90th$ percentile. [1], [2]

Solution.

We know that: $90th$ percentile formula of a General Pareto Distribution is

$p_{90} = b(10)^{\frac{1}{a}}$, where $a = 3, b = 1000.$

Therefore,

$$p_{90} = 1000(10)^{\frac{1}{3}}$$

$$\cong 2154.43.$$

Proposition 2.2.21. [2]

Suppose that X is a continuous random variable having a Standard Pareto Distribution. Then, the continuous random variable Y defined by $Y = bX^{\frac{1}{a}}$, $a > 0$ and $b > 0$ has a General Pareto Distribution with shape parameter a and scale parameter b.

Proof.

Suppose X is a continuous random variable having a Standard Pareto Distribution.

Let Y be a continuous random variable defined by $Y = bX^{\frac{1}{a}}, \quad a > 0$ and $b > 0$.

On the other hand,

$$Y = bX^{\frac{1}{a}} \Rightarrow y = bx^{\frac{1}{a}}$$

$$\Rightarrow \frac{y}{b} = x^{\frac{1}{a}}$$

$$\Rightarrow \left(\frac{y}{b}\right)^a = \left(x^{\frac{1}{a}}\right)^a$$

$$\Rightarrow x = \left(\frac{y}{b}\right)^a$$

This implies that,

$$P(X \le x) = P\left(X \le \left(\frac{y}{b}\right)^a\right)$$

$$= \int_1^{\left(\frac{y}{b}\right)^a} f(x)dx$$

$$= F(x)\Big|_1^{\left(\frac{y}{b}\right)^a}$$

$$= \left[1 - \frac{1}{x}\right]_1^{\left(\frac{y}{b}\right)^a}, \quad x \ge 1$$

$$= 1 - \frac{1}{\left(\frac{y}{b}\right)^a}, \quad y \ge b.$$

$$= 1 - \left(\frac{b}{y}\right)^a, \quad a > 0, b > 0, y \ge b.$$

Here, as a function of y, this is a cumulative distribution function for a General Pareto Distribution $Y = bX^{\frac{1}{a}}$ with shape parameter $a > 0$ and scale parameter $b > 0$. ■

Proposition 2.2.22. [2]

Suppose that X is a continuous random variable having a Basic Pareto Distribution with shape parameter $a > 0$. Then, the continuous random variable Y defined by $Y = bX, \quad b > 0.$ has a General Pareto Distribution with shape parameter a and scale parameter b.

Proof.

Suppose X is a continuous random variable having a Basic Pareto Distribution with shape parameter $a > 0$.

Let Y be a continuous random variable defined by $Y = bX, \quad b > 0.$

On the other hand,

$$Y = bX \Rightarrow y = bx$$

$$\Rightarrow \frac{y}{b} = x$$

$$\Rightarrow x = \frac{y}{b}$$

This means that,

$$P(X \le x) = P\left(X \le \frac{y}{b}\right)$$

$$= \int_1^{\frac{y}{b}} f(x)dx$$

$$= F(x)\Big|_1^{\frac{y}{b}}$$

$$= \left[1 - \frac{1}{x^a}\right]_1^{\frac{y}{b}}, \quad x \geq 1$$

$$= 1 - \frac{1}{\left(\frac{y}{b}\right)^a}, \quad y \geq b$$

$$= 1 - \left(\frac{b}{y}\right)^a, \quad a > 0, b > 0, y \geq b.$$

Here, as a function of y, this is a cumulative distribution function for a General Pareto Distribution $Y = bX$ with shape parameter $a > 0$ and scale parameter $b > 0$. ■

Proposition 2.2.23. [2]

Suppose that X is a continuous random variable having a General Pareto Distribution with shape parameter $a > 0$ and scale parameter $b > 0$. Then, the continuous random variable Y defined by $Y = cX, \quad c > 0$ has a General Pareto Distribution with shape parameter a and scale parameter bc.

Proof.

Suppose X is a continuous random variable having a General Pareto Distribution with shape parameter $a > 0$ and scale parameter $b > 0$.

Let Y be a continuous random variable defined by $Y = cX, \quad c > 0$.

On the other hand,

$$Y = cX \Rightarrow y = cx$$

$$\Rightarrow x = \frac{y}{c}.$$

This means that,

$$P(X \leq x) = P\left(X \leq \frac{y}{c}\right)$$

$$= \int_{b}^{\frac{y}{c}} f(x)dx$$

$$= F(x) \Bigg|_{b}^{\frac{y}{c}}$$

$$= \left[1 - \left(\frac{b}{x} \right)^{a} \right]_{b}^{\frac{y}{c}}, \quad x \geq b$$

$$= 1 - \left(\frac{b}{\frac{y}{c}} \right)^{a}, \quad y \geq bc$$

$$= 1 - \left(\frac{bc}{y} \right)^{a}, \quad a > 0, b > 0, c > 0, y \geq bc.$$

Here, as a function of y, this is a cumulative distribution function for a General Pareto Distribution $Y = cX, c > 0,$ with shape parameter $a > 0$ and scale parameter $bc > 0.$ ■

Proposition 2.2.24. [2]

Suppose that X is a continuous random variable having a General Pareto Distribution with shape parameter $a > 0$ and scale parameter $b > 0.$ Then, the continuous random variable Y defined by $Y = X^n, \quad n > 0$ has a General Pareto Distribution with shape parameter $\frac{a}{n}$ and scale parameter $b^n.$

Proof.

Suppose X is a continuous random variable having a General Pareto Distribution with shape parameter $a > 0$ and scale parameter $b > 0.$

Let Y be a continuous random variable defined by $Y = X^n, \quad n > 0.$

On the other hand,

$$Y = X^n \Rightarrow y = x^n$$

$$\Rightarrow \sqrt[n]{y} = \sqrt[n]{x^n}$$

$$\Rightarrow y^{\frac{1}{n}} = x$$

$$\Rightarrow x = y^{\frac{1}{n}}.$$

This means that,

$$P\left(X \le x\right) = P\left(X \le y^{\frac{1}{n}} \right)$$

$$= \int_b^{y^{\frac{1}{n}}} f(x)dx$$

$$= F(x)\Big|_b^{y^{\frac{1}{n}}}$$

$$= \left[1 - \left(\frac{b}{x}\right)^a\right]_b^{y^{\frac{1}{n}}}, \quad x \ge b$$

$$= 1 - \left(\frac{b}{y^{\frac{1}{n}}}\right)^a, \quad y \ge b^n$$

$$= 1 - \frac{b^a}{y^{\frac{a}{n}}}, \quad y \ge b^n$$

$$= 1 - \left(\frac{b^n}{y}\right)^{\frac{a}{n}}, \quad a > 0, b > 0, n > 0, y \ge b^n.$$

Here, as a function of y, this is a cumulative distribution function for a General Pareto Distribution $Y = X^n$ with shape parameter $\frac{a}{n} > 0$ and scale parameter $b^n > 0$. ■

Proposition 2.2.25. [2]

Suppose that X has the General Pareto Distribution with shape parameter $a > 0$ and scale parameter $b > 0$. For $c \geq b$, the conditional distribution of X given $X \geq c$ is General Pareto Distribution with shape parameter $a > 0$ and scale parameter $c > 0$.

Proof.

We know that;

$F^c = 1 - F$ where F is the ordinary cumulative distribution function given below.

$$F(x) = 1 - \left(\frac{b}{x}\right)^a, \quad a > 0, b > 0, x \geq b.$$

If $x \geq c$, we have

$$P(X > x \mid X > c) = \frac{P(X > x)}{P(X > c)}$$

$$= \frac{1 - \int_1^x f(x)dx}{1 - \int_1^c f(c)dc}$$

$$= \frac{1 - F(x)}{1 - F(c)}$$

$$= \frac{1 - \left[1 - \left(\frac{b}{x}\right)^a\right]}{1 - \left[1 - \left(\frac{b}{c}\right)^a\right]}$$

$$= \frac{\left(\frac{b}{x}\right)^a}{\left(\frac{b}{c}\right)^a}$$

$$= \frac{b^a}{x^a} \cdot \frac{c^a}{b^a}$$

$$= \left(\frac{c}{x}\right)^a, \quad a > 0, c > 0.$$

Here, this is a cumulative distribution function for a General Pareto Distribution with $a > 0$ is shape parameter and $c > 0$ is scale parameter. ■

Chapter 3

PARETO ANALYSIS

3.1 History of Pareto Principle

In 1906, Italian civil engineer and economist Vilfredo Pareto found out that 20% of the people in Italy owned 80% of the land in the country. Moreover, the Wealth distribution was same in all Europe. This reality was explored by him. He also observed that 20% of the pea plants in his garden produce 80% of the peas. At the beginning of 20th century, Pareto created a mathematical model that describes the inequalities in wealth distribution that existed in Italy. This mathematical method is known as Pareto Analysis or 80 – 20 Rule. Pareto Analysis (80 – 20 Rule) is a statistical method in decision making. It states by doing 20% of the work, it generates 80% of the benefit of doing entire work. Pareto Analysis (80 – 20 Rule) can be applied almost all areas such as Business Management, Company Revenues, Employee Evaluation etc. [3], [4]

Example 3.1.1. [3], [4]

The following is related to Pareto Analysis (80 – 20 Rule), in different areas such as

o 20% of your time produces 80% of your work.

o 20% of sales produces 80% of firm’s revenue.

o 20% of salesperson make 80% of the sales.

o 20% of the customers represent 80% of the sales.

o 20% of a organization defects bring about 80% of its trouble.

o 20% of the hazards in a workplace cause 80% of the injuries.

o 80% of mobile phone user complaints occur 20% of company's product and services.

3.2 Pareto Chart [4], [5]

A Pareto chart is a type of bar graph that is used in Pareto Analysis. Frequency is represented by the lengths of bars. Furthermore, the lengths of bars are arranged with longest bars on the left and shortest to the right in descending order. A Pareto Chart is obtained in 8 steps as follows:

a. A vertical bar chart is formed with 'reasons' on the x – axis, on the other hand, with 'count' (number of events) on y – axis.

b. The bar chart is organized cause significance in descending order.

c. The cumulative count is computed for each reason in descending order.

d. The cumulative count percentage is computed for each reason in descending order.

e. A second y – axis is built with percentages ascending in raises of 10 from 0% to 100%.

f. The cumulative count percentage of each reason is sketched on the x – axis.

g. The points are united to obtain percentage ogive.

h. A line at 80% on the y – axis is plotted such that parallel to the x – axis.

i. The line is fallen at the point of intersection with the percentage ogive on the x – axis. The significant reasons on the left (Vital Few) are separated from the less significant reasons on the right (Trivial Many) by this point on the x – axis.

Example 3.2.1.

A government wants to determine and analyse traffic accidents in a country. Therefore, the government commissions a research company to conduct a survey about reasons of traffic accidents.

In this case, the research company considered 880 reasons, which are group into the following categories.

Table 3.1: Reasons by categories

Category	Number of reasons
Driving under the influence of alcohol	216
Sleeplessness	63
Using mobile phone while driving	100
Distractions to Driver	15
Over speeding	330
Disobeying traffic rules while driving	82
Avoiding Safety Gears like seat belts and helmets	17
Lack of road safety	52
Other	5

Table 3.2: Categories of data

No	Category	Absolute frequency	Unit relative frequency [%]	Cumulated absolute frequency	Cumulated relative frequency [%]
1	Over speeding	330	37.50	330	37.50
2	Driving under the influence of alcohol	216	24.55	546	62.05
3	Using mobile phone while driving	100	11.36	646	73.41
4	Disobeying traffic rules while driving	82	9.32	728	82.73
5	Sleeplessness	63	7.16	791	89.89
6	Lack of road safety	52	5.91	843	95.80
7	Avoiding Safety Gears like seat belts and helmets	17	1.93	860	97.73
8	Distractions to Driver	15	1.70	875	99.43
9	Others	5	0.57	880	100.00
	Total	880	100.00 %	–	–

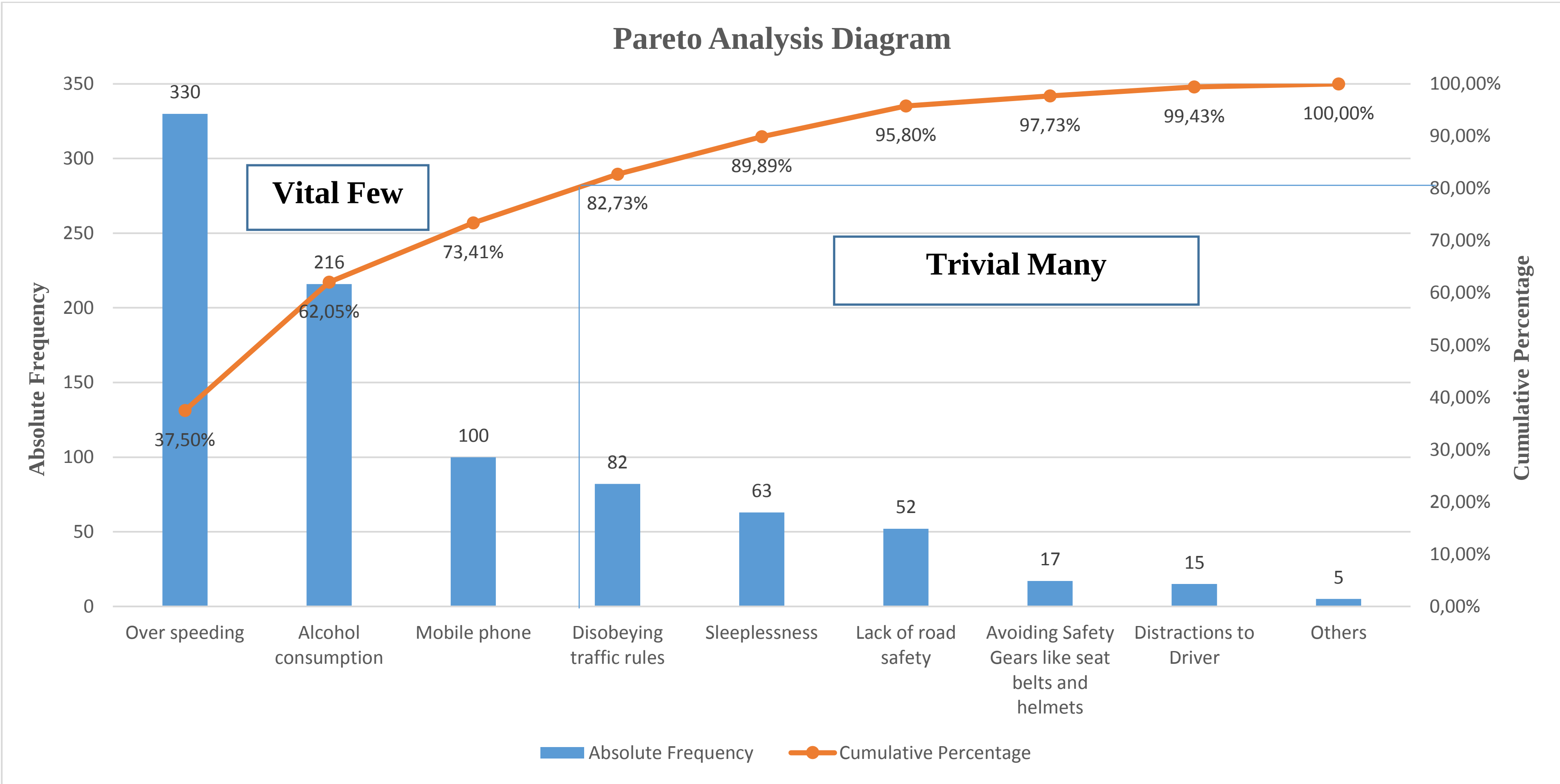

Figure 3.1: Pareto Analysis Diagram of Example 3.2.1

It can be easily seen that over speeding, driving under the influence of alcohol, using mobile phone while driving are the Vital Few and the others are Trivial Many factors of the reasons of traffic accidents in the country.

Example 3.2.2.

A research company gives a questionary in order to determine and analyse the problems in a country. The following example of a pollster who has to analyse various issues in a country. In this case, the pollster considered 950 issues, which are group into the following categories.

Table 3.3: Issues by categories

Category	**Number of issues**
Racism	120
International Relations	88
Poverty	432
Freedom of Press	41
Tax injustice	178
Unemployment	72
Freedom of Thought	16
Others	3

Table 3.4: Categories of data

No	Type of problem	Number of times	% of total	Number of problems (cumulative)	% (cumulative)
1	Poverty	432	45.47	432	45.47
2	Tax Injustice	178	18.74	610	64.21
3	Racism	120	12.63	730	76.84
4	International Relations	88	9.26	818	86.11
5	Unemployment	72	7.58	890	93.68
6	Freedom of Press	41	4.32	931	98.00
7	Freedom of Thought	16	1.68	947	99.68
8	Others	3	0.32	950	100.00
	Total	950	100.00 %	–	–

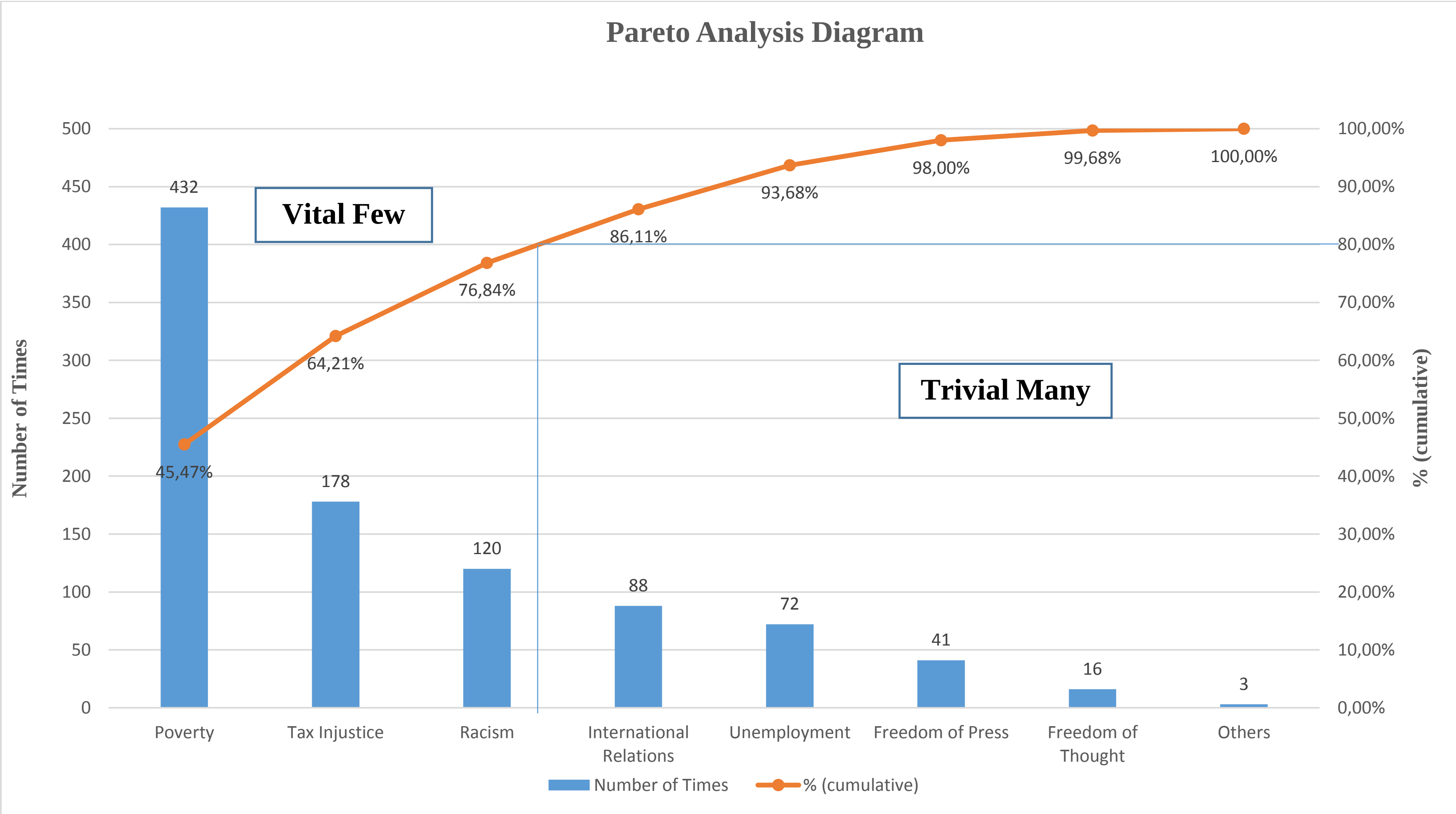

Figure 3.2: Pareto Analysis Diagram of Example 3.2.2

It can be easily seen that poverty, tax injustice, racism are the Vital Few and the others are Trivial Many factors of the problems of a country.

Example 3.2.3.

A company wants to determine and analyse the product defects of the cars they produces. The support officer observed 900 defects, which are group into the following categories.

Table 3.5: Defects by categories

Defect Category	**Number of Product Defects**
Hydraulics	20
Engine	75
Fuel System	360
Software	28
Tyres	79
Air System	65
Suspension	230
Driver Error	38
Others	5

Table 3.6: Categories of data

No	Type of Defect	Number of Times	% of total	Number of Defects (cumulative)	% (cumulative)
1	Fuel System	360	40.00	360	40.00
2	Suspension	230	25.56	590	65.56
3	Tyres	79	8.78	669	74.33
4	Engine	75	8.33	744	82.67
5	Air System	65	7.22	809	89.89
6	Driver Error	38	4.22	847	94.11
7	Software	28	3.11	875	97.22
8	Hydraulics	20	2.22	895	99.44
9	Others	5	0.56	900	100.00
	Total	900	100.00 %	–	–

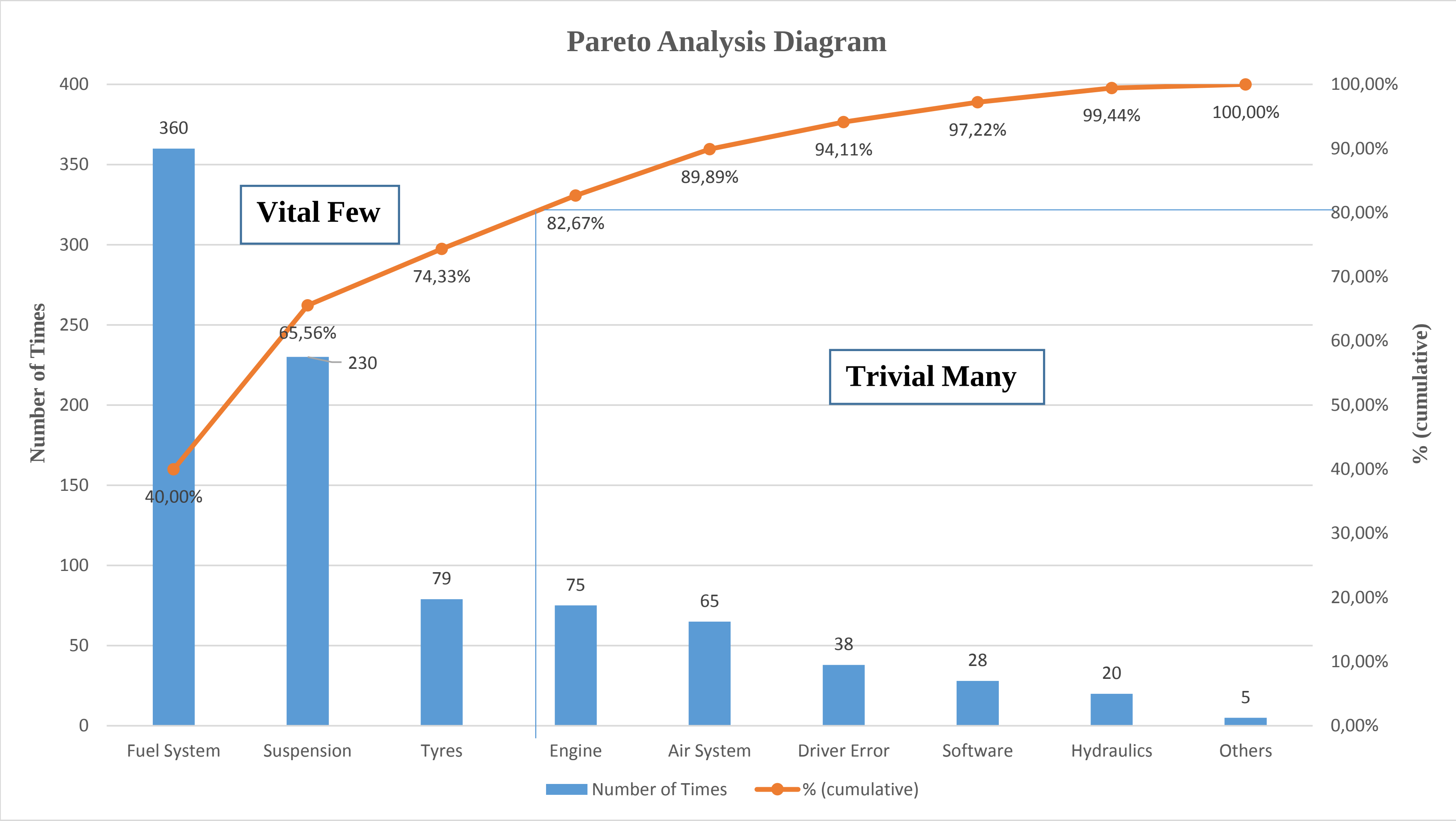

Figure 3.3: Pareto Analysis Diagram of Example 3.2.3

It can be easily seen that fuel system, suspension, tyres are the Vital Few and the others are Trivial Many factors of the product defects of the cars the company produces.

Example 3.2.4.

The following table shows the approximate daily sales of a cafe. Sketch the pareto chart and determine the vital few and trivial many products.

Table 3.7: Sales by categories

Sold Product	**Number of Sales**
Seafood	32
Cookies	5
Ice cream	6
Hot drinks	7
Daily deserts	44
Juice	5
Sandwiches	3
Various pizzas	38

Now, we should form table 3.8 to draw pareto chart.

Table 3.8: Categories of data

No	Sold Product	Number of Sales	% of total	Number of Sales (cumulative)	% (cumulative)
1	Daily deserts	44	31.43	44	31.43
2	Various pizzas	38	27.14	82	58.57
3	Seafood	32	22.86	114	81.43
4	Hot drinks	7	5.00	121	86.43
5	Ice cream	6	4.29	127	90.71
6	Cookies	5	3.57	132	94.29
7	Juice	5	3.57	137	97.86
8	Sandwiches	3	2.14	140	100.00
	Total	140	100.00 %	–	–

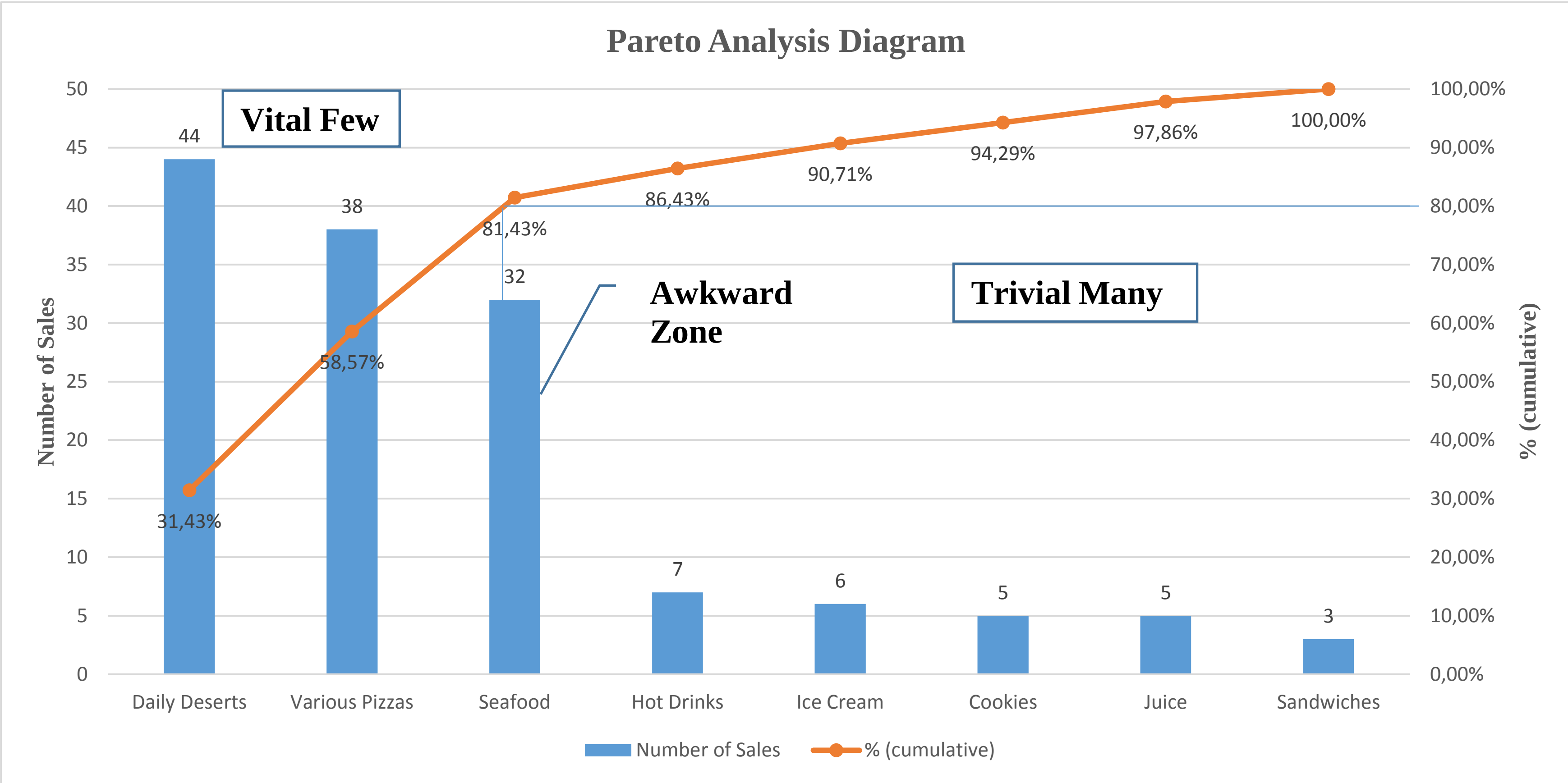

Figure 3.4: Pareto Analysis Diagram of Example 3.2.4

As it can be seen from the graph, Vital Few and Trivial Many can not be precisely distinguished from the 80 % line. Therefore, seafood is Awkward Zone. On the other hand, daily deserts, various pizzas are Vital Few and hot drinks, ice cream, cookies, juice, sandwiches are Trivial Many products.

REFERENCES

[1] The Pareto Distribution. (2019). Retrieved from http://math.bme.hu/~nandori/Virtual_lab/stat/special/Pareto.pdf

[2] The Pareto Distribution. (2019, April 12). Retrieved from https://www.randomservices.org/random/special/Pareto.html

[3] Pareto Distribution. (2019, November 26). Retrieved from https://corporatefinanceinstitute.com/resources/knowledge/economics/pareto-distribution/

[4] Pareto Analysis Step by Step. (2021). Retrieved from https://www.projectsmart.co.uk/pareto-analysis-step-by-step.php

[5] Pareto Principle (80/20 Rule) & Pareto Analysis Guide. (2019, March 19). Retrieved from https://www.juran.com/blog/a-guide-to-the-pareto-principle-80-20-rule-pareto-analysis/#:~:text=What%20is%20the%20Pareto%20Principle%20(80%2F20%20Rule)%3F&text=In%20the%20early%201950s%2C%20Juran,the%20bulk%20of%20the%20effect.

[6] Pareto Analysis. (2015, December). Retrieved from https://images.template.net/wp-content/uploads/2015/12/31065600/Paretho-Analysis-Chart-Template-PDF-File.pdf