University of Nevada, Reno

A Discrete Truncated Power-law Distribution

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics

by

Kwame Boamah-Addo

Tomasz J. Kozubowski, Ph.D., Thesis Adviser

August, 2018

University of Nevada, Reno

THE GRADUATE SCHOOL

We recommend that the thesis
prepared under our supervision by

KWAME BOAMAH-ADDO

entitled

A Discrete Truncated Power-law Distribution

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Tomasz J. Kozubowski, Ph.D., Adviser

Anna Panorska, Ph.D., Committee Member

Forister Mathew, Ph.D., Graduate School Representative

David Zeh, Ph.D., Dean, Graduate School

August, 2018

Abstract

This thesis concerns discrete probability distributions with power-law tail behavior. Discrete power-law distributions are reviewed together with their truncated versions. The Zipf distribution (Zipf, 1932) is then focused on and a truncated version of the Zipf distribution is further studied. This distribution is known as the discrete truncated power-law distribution ($DTPL$). The basic properties of the $DTPL$ are derived and procedure for estimating the parameters of the distribution are developed. The usefulness of the distribution is shown through several data examples.

Acknowledgements

I would first like to thank my thesis advisor, Professor Tomasz J. Kozubowski of the Department of Mathematics and Statistics at University of Nevada, Reno. Prof. Kozubowski always opened his door to me whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it.

I would also like to thank the members of my research committee who provided their invaluable input for this research project: Professor Anna K. Panorska of the Mathematics Department and Prof. Matthew Forister of the Department of Biology. Without their passionate participation and input, this work would not have been completed.

Finally, I must express my very profound gratitude to my parents and to my friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Thank you.

Kwame Boamah-Addo

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Many empirical quantities are approximately normally distributed, that is, they are normally clustered around the means of these empirical quantities. Examples of these quantities include: the speeds of cars on a highway, the weights of apples in a store, air pressure, sea level, the temperature in New York at noon on midsummer's day.

However, not all empirical quantities have distributions that have this centering-around-the-mean property. In particular, some phenomena are known to have power-law distributions. Mathematically, a distribution is said to follow power-law when it is drawn from the probability distribution where the probability density function (PDF) scales like a power law,

$$p(x) \propto x^{-\alpha}, \tag{1.1}$$

where α is a constant parameter of the distribution known as the exponent or scaling parameter (see Clauset et al., 2009).

Studies of power-law distributions include power law in cities population, financial markets, and internet sites (see Blank and Solomon, 2000) and power-law distributions in financial market fluctuations studied by Gabaix et al. (2003).

In this work, we look at a subclass discrete power-law distributions, where the relevant observations are non-negative integers.

A standard discrete power-law distribution which has been used widely in literature, is the Zipf distribution, first introduced by George Kingsley Zipf in his 1932 paper titled "Selective Studies and the Principle of Relative Frequency in Language". Zipf applied this distribution to word frequency. The probability mass function (PMF) of this discrete distribution is given by

$$f_X(x) = \frac{1}{Tx^\alpha}, \;\; x \in N = \{1, 2, ..\}, \tag{1.2}$$

where T is the norming constant,

$$T = \sum_{x=1}^{\infty} \frac{1}{x^\alpha}. \tag{1.3}$$

The above Zipf distribution has the support of all natural numbers. In contrast, in this work we shall look at the truncated version of this distribution, with our support being bounded by two natural numbers,γ and ν. We term it *discrete truncated power law distribution (DTPL) distribution.*

Here is the organization of this work. First, we review a number of discrete power-law distributions in Chapter 2, highlighting some works done in the past as well as their applications. These distributions include the Zipf distribution, Lokta distribution, Zipf-Mandelbrot distribution, Estoup distribution, discrete Pareto distribution, generalized Sibuya distribution and Yule-Simon distribution. In Chapter 3, the truncated versions of these distributions are discussed, which include some new distributions. Chapter 4 is devoted to the new discrete truncated power law distribution, where we develop its basic properties. Then, in Chapter 5, we discuss procedures for parameter estimation. In Chapter 6, we provide several data examples, illustrating the modeling potential of our distribution. Finally, programming routines in R are collected in the Appendix.

Chapter 2

Here we offer a brief review of important power-law distributions supported on the set of positive integers.

2.1 Zipf Distribution

Zipf law is attributed to George Kingsley Zipf, a natural scientist and a linguist who was a Professor at Havard. Zipf first introduced the law in 1932 in his paper titled "Studies and the Principle of Relative Frequency in Language". Zipf is also credited with having discovered that word frequencies and the sizes of cities follow a power law distribution (Zipf, 1941). Zipf further used this law in his subsequent publications Zipf (1935) and Zipf (1941).

While, for historical reasons, Zipf distribution is often applied in the area of linguistics, it also appears in connection with modeling income distribution, city sizes and run length distribution of active typing (see Meier-Heilstern et al., 1991).

The probability mass function (PMF) of the Zipf distribution is of the form:

$$f(x) = P(X = x) = \frac{1}{Tx^{\alpha}}, \quad x \in N = \{1, 2, ..\}, \tag{2.1}$$

where $\alpha > 1$ and T is a norming constant, given by

$$T = \sum_{x=1}^{\infty} \frac{1}{x^{\alpha}}. \tag{2.2}$$

Note that $T = \zeta(\alpha)$, where

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s > 1, \tag{2.3}$$

is the Riemann-Zeta function. It is worth noting that the survival function of a Zipf-distributed random variable X behaves like a power-law as well. More precisely, the quantity

$$P(X > x)$$

is asymptotically equivalent to

$$C \cdot X^{-\beta}$$

as $x \longrightarrow \infty$, where $\beta = \alpha - 1 > 0$. We express this by writing

$$P(X > x) \sim Cx^{-\beta}, x \longrightarrow \infty. \tag{2.4}$$

The meaning of the relation (2.4) is that the ratio of the two functions converges to 1 as $x \longrightarrow \infty$. The constant C in (2.4) is given in the following result.

Proposition 2.1 *Let X have a Zipf distribution with parameter $\alpha > 1$, given by the PMF (2.1). Then we have (2.4), where $\beta = \alpha - 1$ and $C = [(\alpha - 1)\zeta(\alpha)]^{-1}$.*

Proof Note that for any integer $x \in N$, we have

$$P(X > x) = \frac{D(x)}{T}$$

where

$$D(x) = \left(\frac{1}{x+1}\right)^{\alpha} + \left(\frac{1}{x+2}\right)^{\alpha} + \cdots .$$

Referring to the graph of the function $h(t) = t^{-\alpha}$, it is not hard to see that

$$\int_{x+1}^{\infty} \left(\frac{1}{t^{\alpha}}\right) dt < D(x) < \int_{x+1}^{\infty} \left(\frac{1}{t-1}\right)^{\alpha} dt.$$

Next, since

$$\int_{x+1}^{\infty} \left(\frac{1}{t^{\alpha}}\right) dt = \frac{1}{\alpha - 1} \left(\frac{1}{x+1}\right)^{\alpha-1}$$

and

$$\int_{x+1}^{\infty} \left(\frac{1}{t-1}\right)^{\alpha} dt = \frac{1}{\alpha - 1} \left(\frac{1}{x+1}\right)^{\alpha-1}$$

we conclude that

$$\frac{1}{\alpha - 1} \left(\frac{1}{x+1}\right)^{\alpha-1} < D(x) < \frac{1}{\alpha - 1} \left(\frac{1}{x}\right)^{\alpha-1}. \tag{2.5}$$

Thus, since

$$\frac{P(X > x)}{Cx^{-(\alpha-1)}} = \frac{D(x)[\zeta(\alpha)]^{-1}}{[(\alpha-1)\zeta(\alpha)]^{-1} x^{-(\alpha-1)}},$$

in the view of (2.5) we conclude that

$$\left(\frac{x}{x+1}\right)^{\alpha-1} < \frac{P(X > x)}{Cx^{-(\alpha-1)}} < 1.$$

Finally, when we take into account that

$$\left(\frac{x}{x+1}\right)^{\alpha-1} \longrightarrow 1$$

as $x \longrightarrow \infty$, we conclude that

$$\frac{P(X > x)}{x^{-(\alpha-1)}} \longrightarrow 1 \text{ as } x \longrightarrow \infty,$$

as desired. This concludes the proof.

Remark 2.2 *To accomodate the case where the support of the distribution is the set of non-negative integers $N_0 = \{0, 1, 2, \cdots\}$, one can define a shifted version of the Zipf distribution. We transform the Zipf variable X viz*

$$Y = X - 1.$$

All shifted versions of distributions seen in this chapter follow the same scheme.The probability mass function of the shifted Zipf distribution is given by:

$$f_Y(y) = \frac{1}{T(y+1)^{\alpha}}, \quad y \in N_0 = \{0, 1, 2, ...\}, \tag{2.6}$$

where T is the same as before.

2.2 Lotka Distribution

A special case of the Zipf distribution is known as Lotka distribution, and is given by the PMF

$$f_X(x) = \frac{1}{Tx^2}, \quad x \in N, \tag{2.7}$$

where

$$T = \frac{6}{\pi^2}. \tag{2.8}$$

This distribution was introduced by Alfred Lokta, a statistician who worked with the Metropolitan Life Insurance company in New York in 1926. Lokta's main goal was to examine the number of articles produced by chemist Michael Mitzenmacher (2004). In the preamble to his publication, Lokta stated that it "would be of interest to determine, if possible, the part which men of different caliber contribute to the progress"(Lokta, 1926). According to Rousseau (2002), it was through Lokta's work that Zipf got his idea to use the inverse square law to model word frequency in Chinese. Lotka distribution is often used in the area of documentation and scientometrics as well as in the study of publication activity, citation frequency and related fields. For more information, see Zörnig and Altmann (1995).

Remark 2.3 *The probability mass function of a shifted Lotka distribution is given by*

$$f_Y(y) = \frac{6}{\pi^2(y+1)^2}, \quad y \in N_0. \tag{2.9}$$

2.3 Zipf-Mandelbrot Distribution

A more general version of Zipf distribution is obtained by replacing x by $x+a$ in the PMF, leading to

$$f_X(x) = \frac{1}{T(a+x)^\alpha} \quad x \in N, \tag{2.10}$$

where,

$$T = \sum_{x=1}^{\infty} \frac{1}{(a+x)^\alpha}. \tag{2.11}$$

The formula for T above can be written in terms of the special Lerch Zeta function defined by

$$\Phi(p, a, \alpha) = \sum_{n=1}^{\infty} \frac{p^n}{(a+n)^\alpha}, \tag{2.12}$$

so that $T = \Phi(1, a, \alpha)$.

This distribution is known as Zipf-Mandelbrot distribution as it was introduced Benoit Mandelbrot, a Polish born mathematician (Mandelbrot, 1953). According to Montemurro (2001), the standard Zipf's law can account for the statistical behaviour of words frequencies in a rather limited zone in the middle-low to low range of the rank variable. Mandelbrot introduced this modification by using arguments on the fractal structure of lexical trees. Due to its history, the Zipf-Mandelbrot law is generally applied in the area of linguistics and well as other areas.

Remark 2.4 *The probability mass function of a shifted Zipf-Mandlebrot distribution is given by*

$$f_Y(y) = \frac{1}{T(a+(y+1))^{\alpha}}, \quad y \in N_0, \tag{2.13}$$

where T is as before.

2.4 Estoup Distribution

The so-called Estoup distribution is given by the PMF

$$f_X(x) = \frac{1}{Tx}, \quad x \in N, \tag{2.14}$$

where

$$T = \sum_{x=1}^{n} \frac{1}{x}. \tag{2.15}$$

The distribution is named after Jean-Baptiste Estoup, who was a French stenographer and scientist. According to Mitzenmacher (2004), "the first known attribution of the power law distribution of word frequencies appears to be due to Estoup in 1916". This is noted by Rousseau (2002), where he indicated that Zipf made reference to Estoup in his very first publication on the relative frequency of words. The Estoup distribution, unlike the other discrete power law distributions seen thus far, exists only in truncated form since the harmonic series is not convergent (Zörnig and Altmann, 1995). The Estoup distribution is used in linguistics for modeling word frequencies.

Remark 2.5 *The probability mass function of the shifted Estoup Distribution is given by*

$$f_Y(y) = \frac{1}{T(y+1)}, \quad y \in N_0, \tag{2.16}$$

where

$$T = \sum_{y=0}^{n} \frac{1}{y}.$$

2.5 Discrete Pareto Distribution

The Pareto distribution is a power law probability distribution introduced by Italian civil engineer, economist, and sociologist Vilfredo Pareto. Pareto introduced this distribution to describe income distribution in 1896 (see Mitzenmacher, 2004). Since then, Pareto distribution has been studied extensively and provided an inspiration for introducing many related power-law distributions. In particular, Buddana and Kozubowski (2014) studied discrete Pareto distribution, while Zaninetti and Ferraro (2008) also studied truncated Pareto distribution with applications. Kozubowski et al. (2015) also studied the discrete truncated Pareto distribution. Pareto distribution has a wide reach of application in economics, actuarial science and other fields.

The discrete Pareto distribution studied by Buddana and Kozubowski (2014) is described by the PMF

$$f_X(x) = \left(\frac{1}{1+\frac{x-m}{\sigma}}\right)^{\alpha} - \left(\frac{1}{1+\frac{x-1-m}{\sigma}}\right)^{\alpha}, \quad x = m, m+1, ..., \tag{2.17}$$

where $m \in Z$ is a shift parameter, $\sigma > 0$ is a "size" parameter, and $\alpha > 0$ is a tail parameter. An alternative parameterization with $m = 1$ allows for writing the PMF as

$$f_X(x) = \left(\frac{1}{1-\alpha(x-1)\log(1-p)}\right)^{\frac{1}{\alpha}} - \left(\frac{1}{1-\alpha x\log(1-p)}\right)^{\frac{1}{\alpha}}, \quad x \in N. \qquad (2.18)$$

Remark 2.6 *A shifted version of the distribution is obtained viz. the transformation* $Y = X - 1$*, leading to the PMF:*

$$f_Y(y) = \left(\frac{1}{1-\alpha(y)\log(1-p)}\right)^{\frac{1}{\alpha}} - \left(\frac{1}{1-\alpha(y+1)\log(1-p)}\right)^{\frac{1}{\alpha}}, \quad y \in N_0. \qquad (2.19)$$

2.6 Generalized Sibuya Distribution

The PMF of a generalized Sibuya distribution as studied by Kozubowski and Podgorski (2016), is given by

$$f_X(x) = \left(1 - \frac{\alpha}{k+1}\right)\cdots\left(1 - \frac{\alpha}{k+x-1}\right)\frac{\alpha}{k+x}, \; x \in N, \qquad (2.20)$$

where the two parameters are restricted by the relation $0 < \alpha < k + 1$.
As discussed in Kozubowski and Podgorski (2016), this distribution arises as the distribution of the waiting time for the first success in Bernoulli trials, where the probabilities of success are inversely proportional to the number of a trial. The generalized Sibuya distribution is linked to the Sibuya distribution ($k = 0$), which was first introduced in Sibuya (1979). Sibuya introduced this distribution "as substitutes of the logarithmic series when the observed frequency data have such a long tail that cannot be fitted by the latter distributions "(Sibuya, 1979).

Remark 2.7 *A shifted version of the distribution is obtained viz* $Y = X - 1$*, leading* *to the PMF*

$$f_Y(y) = \left(1 - \frac{\alpha}{k+1}\right) \cdots \left(1 - \frac{\alpha}{k+y-1}\right) \frac{\alpha}{k+y}, \quad y \in N_0. \tag{2.21}$$

2.7 Yule-Simon Distribution

The Yule-Simon distribution is a discrete power-law distribution that was first proposed by Yule in 1925 to explain experimental data on the abundances of biological genera (see Cattuto et al., 2006). The model was subsequently revised by Simon (1955) to explain the observed power-law distribution of word frequencies in texts (Cattuto et al., 2006). According to Mitzenmacher (2004), Simon (1955) lists five applications of this type of model in his introduction, which include distributions of word frequencies in documents, distributions of numbers of papers published by scientists, distribution of cities by population, distribution of incomes, and distribution of species among genera.

The probability mass function of the Yule-Simon distribution is given by:

$$f(x;\rho) = \rho \frac{\Gamma(x)\Gamma(\rho+1)}{\Gamma(x+\rho+1)}, \tag{2.22}$$

where $x \in N$, $\rho > 0$ and Γ is the Gamma function. If ρ is an integer and PMF is written in terms of the falling factorial, then the PMF is

$$f(x;\rho) = \frac{\rho\rho!(x-1)!}{(x+\rho)!}. \tag{2.23}$$

Remark 2.8 *Note that for $x \in N$ we have*

$$\Gamma(x + p + 1) = (x + p)(x - 1 + p) \cdots (1 + p) \cdot \Gamma(p + 1), \tag{2.24}$$

so that the PMF (2.22) of Yule distribution can be written as

$$f(x; p) = \frac{\rho\Gamma(x)}{(x + p)(x - 1 + p) \cdots (1 + p)}, \quad x \in N. \tag{2.25}$$

A comparison of this with (2.20) shows that Yule distribution with parameter $\rho > 0$ is a special case of generalized Sibuya distribution with parameters $k = \alpha = \rho$.

Chapter 3

Truncated Discrete Power-Law Distributions

In this chapter we shall consider the truncated versions of the discrete power-law distributions discussed in the previous chapter. As shown below, the process of truncation leads to distributions with finite support.

3.1 Truncated Discrete Power law Distributions

Given a random variable X supported on the set of positive integers N, one can obtain a new random variable Y, which describes X conditionally on the event that X is bounded by two integers. That is, we define

$$Y \stackrel{d}{=} X | \gamma \leq X \leq \nu, \tag{3.1}$$

where γ , $\nu \in N$. Then, the PMF of Y becomes

$$P(Y = y) = \begin{cases} \dfrac{P(X = y)}{P(X \leq \nu) - P(X \leq \gamma - 1)}, & y = \gamma, \gamma + 1, \cdots, \nu, \\ 0, \quad otherwise. \end{cases} \tag{3.2}$$

The equation above shows that the normalizing constants with infinite sums in some of the discrete power law distributions seen in Chapter 2 will disappear, leaving constants that have finite sum and are easier to work with. It is with this framework that we shall develop truncated versions of the distributions discussed in Chapter 2.

3.2 Truncated Lotka Distribution

Generating the probability mass function of the truncated Lokta distribution from (3.2), we have the following

$$f_Y(y) = \frac{y^{-2}}{\sum_{x=\gamma}^{\nu} x^{-2}}, \quad y = \gamma, \gamma + 1, \cdots, \nu. \tag{3.3}$$

3.3 Truncated Zipf-Mandlebrot Distribution

The probability mass function of the truncated Zipf-Mandlebrot distribution generated by (3.2) is as shown below:

$$f_Y(y) = \frac{(y + a)^{-\alpha}}{\sum_{x=\gamma}^{\nu} (x + a)^{-\alpha}}, \quad \gamma, \gamma + 1, \cdots, \nu. \tag{3.4}$$

Note that the parameter α is now no longer restricted by the condition $\alpha > 1$, and can assume any real number, so that $\alpha \in R$ for this distribution.

3.4 Truncated Estoup Distribution

Generating the probability mass function of the truncated Estoup distribution from (3.2) we have the following

$$f_Y(y) = \frac{y^{-1}}{\sum_{x=\gamma}^{\nu} x^{-1}}, \quad y = \gamma, \gamma+1, \cdots, \nu. \tag{3.5}$$

Note that this is special case $\alpha = 1, a = 0$ fo the truncated Zipf-Mandelbrot distribution.

3.5 Discrete Truncated Pareto Distribution

The discrete truncated Pareto distribution has been studied by Kozubowski et.al., (2015). The probability mass function of this distribution is given by

$$f_Y(y) = \frac{\frac{1}{y^\alpha} - \frac{1}{(y+1)^\alpha}}{\frac{1}{\gamma^\alpha} - \frac{1}{(\nu+1)^\alpha}}, \ \ y = \gamma, \gamma+1, ..., \nu, \tag{3.6}$$

where γ and $\nu \in N$. One advantage of the discrete truncated Pareto distribution is that α is no longer restricted to $\alpha > 0$ and that α can assume negative values. For $\alpha < 0$, writing $\beta = -\alpha > 0$ the PMF becomes

$$f_Y(y) = \frac{(y+1)^\beta - y^\beta}{(\nu+1)^\beta - \gamma^\beta}, \ \ x = \gamma, \gamma+1, ..., \nu. \tag{3.7}$$

3.6 Truncated Generalized Sibuya Distribution

Kozubowski and Podgorski (2016) worked extensively on the generalized Sibuya distribution where the survival function is of the form

$$P(N > n) = \left(1 - \frac{\alpha}{k+1}\right) \cdots \left(1 - \frac{\alpha}{k+n}\right). \tag{3.8}$$

We can now generate a truncated generalized Sibuya distribution from (3.8) and (**??**) using the definition given in (3.2). Thus, PMF of a truncated generalized Sibuya distribution is given by

$$f_Y(y) = \frac{\left(1 - \frac{\alpha}{k+1}\right) \cdots \left(1 - \frac{\alpha}{k+y-1}\right) \frac{\alpha}{k+y}}{\left[\left(1 - \frac{\alpha}{k+1}\right) \cdots \left(1 - \frac{\alpha}{k+\gamma-1}\right)\right] - \left[\left(1 - \frac{\alpha}{k+1}\right) \cdots \left(1 - \frac{\alpha}{k+\nu}\right)\right]}, \tag{3.9}$$

where $y = \gamma, \gamma + 1, \cdots, \nu$.

3.7 Truncated Yule-Simon Distribution

Generating the probability mass function of the truncated Yule-Simon distribution from (3.2) is straightforward, if we take into account the fact that this distribution is a special case $\alpha = k = \rho$ of generalized Sibuya distribution. Upon substitution of the values into (3.9), after some algebra we obtain

$$f_Y(y) = \frac{\rho\Gamma(y)}{\Gamma(y + \rho + 1)} \left(\frac{\Gamma(\gamma)}{\Gamma(\gamma + \rho)} - \frac{\Gamma(\nu + 1)}{\Gamma(\rho + \nu + 1)}\right)^{-1}, \tag{3.10}$$

for $y = \gamma, \gamma + 1, \cdots, \nu$.

Chapter 4

Truncated Zipf Distribution and its Properties

Here we develop basic properties of truncated Zipf distribution. Let X be a random variable that follows the Zipf distribution as described in Chapter 2, which is given by the PMF

$$f_X(x) = \frac{1}{Tx^\alpha}, \quad \text{x} \in N, \tag{4.1}$$

Where T is as in (2.2),

$$T = \sum_{j=1}^{\infty} \frac{1}{j^\alpha}. \tag{4.2}$$

Now, for any $1 \leq \gamma < \nu < \infty$, where $\gamma, \nu \in N$, define a new random variable Y by restricting the distribution of X to be the interval $[\gamma, \nu]$. More precisely, we have

$$Y \stackrel{d}{=} X | \gamma \leq X \leq \nu. \tag{4.3}$$

Thus, the PMF of Y is

$$f_Y(y) = \begin{cases} \dfrac{P(X = y)}{P(X \le \nu) - P(X \le \gamma - 1)}, & y = \gamma, \gamma + 1, \cdots, \nu \\ 0, \quad otherwise. \end{cases}$$

Consequently, the normalizing constant T cancels out, yielding the probability mass function of the form

$$f_Y(y) = \frac{\frac{1}{y^\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}, \quad y = \gamma, \gamma + 1, \cdots, \nu. \tag{4.4}$$

We shall refer to the distribution given by the PMF (4.4) as the $DTPL$ distribution, which stands for the discrete truncated power law, and write $Y \sim DTPL(\alpha, \gamma, \nu)$. The parameters γ and ν are integers, restricted by the relation $1 \le \gamma < \nu < \infty$. Note that after the truncation, the parameter α is no longer restricted to be greater than 1, and if fact (4.4) is a genuine PMF for any $\alpha \in R$. Thus, α can take on any positive or negative value. In the special case $\alpha = 0$ we obtain a uniform distribution on the set $\{\gamma, \gamma + 1, \cdots, \nu\}$, with the PMF

$$f_Y(y) = \frac{1}{\nu - \gamma + 1}, \ y = \gamma, \gamma + 1, \cdots, \nu. \tag{4.5}$$

4.1 Cumulative Distribution Function

The cumulative distribution function (CDF) of the truncated Zipf distribution can be derived from its PMF, leading to is given as

$$F(z) = P(Y \le z) = \sum_{y=1}^{z} P(Y = y), \tag{4.6}$$

where $z \in N$. Thus the CDF of the DTPL(α, γ, ν) distribution is of the form

$$F(z) = \frac{\sum_{y=\gamma}^{z} \frac{1}{y^\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}, \tag{4.7}$$

where $\gamma \leq z \leq \nu$ and $z \in N$. Clearly, $F(z)$ is 0 for $z < \gamma$ and $F(z) = 1$ for $z > \nu$. Similarly, the survival function of DTPL(α, γ, ν) distribution is given by

$$S(z) = 1 - F(z) = \frac{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha} - \sum_{y=\gamma}^{z} \frac{1}{y^\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}, \tag{4.8}$$

where $\gamma \leq z \leq \nu$ and $z \in N$. The above formula can be extended to account for a non-integer arguments by means of the greatest integer function $\lfloor z \rfloor$. Recall that $\lfloor z \rfloor$ is the greatest integer that is less or equal to z, so that $\lfloor z \rfloor = n$ whenever $z \in [n, n+1)$. With this notation we have the following result.

Proposition 4.1 *The cumulative distribution function and the survival function of a $DTPL(\alpha, \gamma, \nu)$ random variable Y are given by*

$$F(z) = P(Y \leq z) = \frac{\sum_{y=\gamma}^{\lfloor z \rfloor} \frac{1}{y^\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}, \ \gamma \leq z < \nu. \tag{4.9}$$

And the survival function can be defined thus:

$$S(z) = P(Y > z) = \frac{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha} - \sum_{y=\gamma}^{\lfloor z \rfloor} \frac{1}{y^\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}, \ \gamma \leq z < \nu. \tag{4.10}$$

For $z < \gamma$ the CDF is zero while $S(z) = 1$, and for $z \geq \nu$, $F(z) = 1$ while $S(z) = 0$.

As we see from Figure 4.1, as α is increasing from $-\infty$ to ∞, the probability mass appears to shift from the right to left, which suggests that for larger α it is more likely to observe a smaller value of Y. Mathematically we can express thus by saying that the distribution of $Y_1 \sim DTPL(\alpha_1, \gamma, \nu)$ is *stochastically dominated* by distribution of $Y_2 \sim DTPL(\alpha_2, \gamma, \nu)$ when $\alpha_1 > \alpha_2$ or that Y_2 is stochastically larger than Y_1. This occurs, when for each fixed y the CDF of Y_1, $F_{Y_1}(y)$, is greater or equal to the CDF of Y_2, $F_{Y_2}(y)$:

$$F_{Y_1}(y) \geq F_{Y_2}(y), \ y \in R.$$

In the current situation, this will occur if the CDF of $Y \sim DTPL(\alpha, \gamma, \nu)$ is an increasing function of α when y is held fixed. We have the following result in connection with this, showing that this is actually the case.

Proposition 4.2 *The CDF of $Y \sim DTPL(\alpha, \gamma, \nu)$, $F_Y(y)$, is a non-decreasing function of α on $(-\infty, \infty)$ for each fixed $y \in R$.*

4.2 Basic Properties

In this section we shall develop basic properties of the $DTPL$ distribution. We start with the behavior of the PMF with respect to changes of the argument y or the parameters. Our first result shows that the paramter α controls the monotonicity of the PMF.

Proposition 4.3 *The PMF (4.4) of a DTPL(α, γ, ν) random variable Y is monotonically increasing in y when $\alpha < 0$, the PMF is constant when $\alpha = 0$, and is monotonically decreasing in y when $\alpha > 0$.*

Proof The result follows from the properties of the power function $y^{-\alpha}$, which is increasing in y on $[1, \infty)$ for $\alpha < 0$, constant for $\alpha = 0$, and decreasing for $\alpha > 0$.

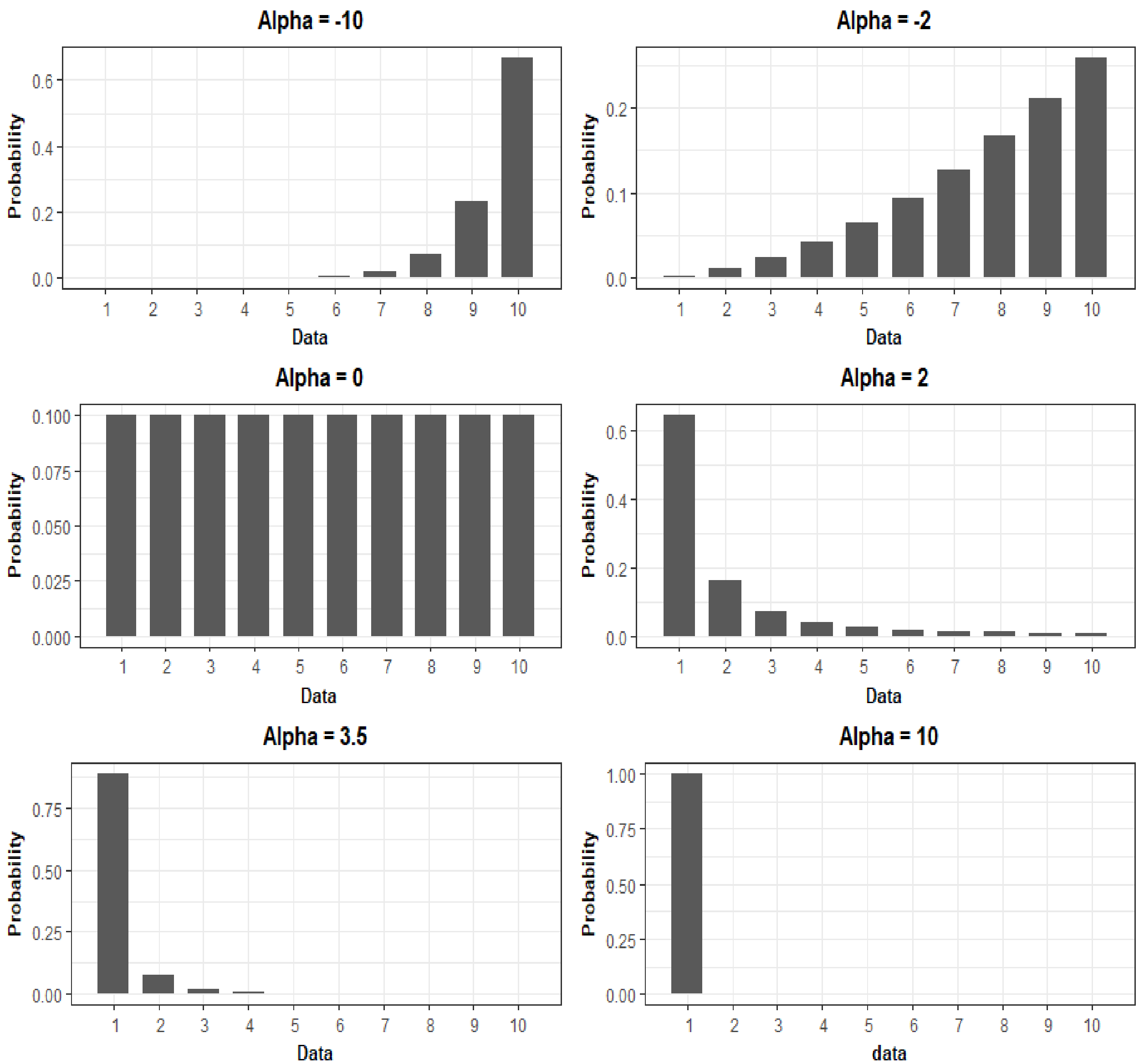

Figure 4.1: The PMF of DTPL distribution with, $\gamma = 1$, $\nu = 10$, and selected values of α

This monotonic behavior is illustrated in Figure 4.1, which shows $DTPL$ probabilities for $\gamma = 1$, $\nu = 10$ and selected values of α.

Next, we consider the behavior of the PMF when y is fixed while the parameters vary. Note that here the parameter γ is restricted to be in the interval $[1, y]$ when y is held fixed, while the parameter ν is restricted to the interval $[y, \infty)$.

We have the following result concerning the behavior of the PMF with respect to changes in γ or ν.

Proposition 4.4 *For any fixed $y > 0$ and $\alpha \in (-\infty, \infty)$, the PMF of DTPL($\alpha, \gamma, \nu$) is monotonically increases in γ on the interval (0, y) and monotonically decreases in ν on the interval (y, ∞).*

Proof For y fixed and α fixed, consider (4.4) and let γ_1 and γ_2 such that $1 \leq \gamma_1 < \gamma_2 \leq y$. Let f_1 be the PMF of Y $\sim$ DTPL(α, γ_1, ν). Then, we have

$$f_1(y) = \frac{(\frac{1}{y})^\alpha}{\sum_{j=\gamma_1}^{\gamma_2-1}(\frac{1}{j})^\alpha + \sum_{j=\gamma_2}^{\nu}(\frac{1}{j})^\alpha}. \tag{4.11}$$

Similarly, if f_2 is the PMF of Y $\sim$ DTPL(α, γ_2, ν), then

$$f_2(y) = \frac{(\frac{1}{y})^\alpha}{\sum_{j=\gamma_2}^{\nu}(\frac{1}{j})^\alpha}. \tag{4.12}$$

Since $\sum_{j=\gamma_2}^{\nu}(\frac{1}{j})^\alpha > 0$, we conclude that $f_1(y) < f_2(y)$.Thus, the PMF a of DTPL(α, γ, ν) random variable monotonically increases in γ.

Next, for y and α fixed, consider again (4.4) and let ν_1 and ν_2 be such that $y \leq \nu_1 < \nu_2$. Then, with f_1 denoting the PMF of Y $\sim$ DTPL(α, γ, ν_1), we have

$$f_1(y) = \frac{(\frac{1}{y})^\alpha}{\sum_{j=\gamma}^{\nu_1}(\frac{1}{j})^\alpha}. \tag{4.13}$$

Similarly, with f_2 denoting the PMF of Y $\sim$ DTPL(α, γ, ν_2), we have

$$f_2(y) = \frac{(\frac{1}{y})^\alpha}{\sum_{j=\gamma}^{\nu_1}(\frac{1}{j})^\alpha + \sum_{j=\nu_1+1}^{\nu_2}(\frac{1}{j})^\alpha}. \tag{4.14}$$

Since $\sum_{j=\nu_1+1}^{\nu_2}(\frac{1}{j})^\alpha > 0$, we conclude that $f_1(y) > f_2(y)$. Hence, the PMF of a DTPL(α, γ, ν) random variable monotonically decreases in ν. This concludes the proof.

Our next result shows that truncation of a $DTPL$ distribution leads to another $DTPL$ distribution.

Proposition 4.5 *Suppose that Y has DTPL(α, γ, ν) distribution, and let $Z \stackrel{d}{=} Y|\tilde{\gamma} \leq Y \leq \tilde{\nu}$, where $\gamma \leq \tilde{\gamma} \leq \tilde{\nu} \leq \nu$. Then $Z \sim$ DTPL($\alpha, \tilde{\gamma}, \tilde{\nu}$) with the PMF*

$$f_Z(z) = \frac{z^{-\alpha}}{\sum_{i=\tilde{\gamma}}^{\tilde{\nu}} i^{-\alpha}}, \quad z = \tilde{\gamma}, \tilde{\gamma}+1, ..., \tilde{\nu}. \tag{4.15}$$

Proof The PMF of the new distribution is

$$f_Z(z) = \frac{\frac{z^{-\alpha}}{\sum_{i=\gamma}^{\nu} i^{-\alpha}}}{\sum_{j=\tilde{\gamma}}^{\tilde{\nu}} \frac{j^{-\alpha}}{\sum_{i=\gamma}^{\nu} i^{-\alpha}}}. \tag{4.16}$$

Upon simplification of the above, we obtain

$$f_Z(z) = \frac{z^{-\alpha}}{\sum_{i=\tilde{\gamma}}^{\tilde{\nu}} j^{-\alpha}}, \quad z = \tilde{\gamma}, \tilde{\gamma}+1, ..., \tilde{\nu}, \tag{4.17}$$

as desired.

Proposition 4.6 *The PMF f(y) of a DTPL(α, γ, ν) random variable Y is a continuous and differentiable function of α on the interval ($-\infty, \infty$).*

Proof Clearly, f(y) is a ratio of two differentiable functions of α, and thus it is differentiable (and continuous) as well.

Proposition 4.7 *If $y = \gamma$, then the PMF of a DTPL(α, γ, ν) random variable Y is monotonically increasing in α on (-∞,∞) with the limits of zero and one at -infinity and infinity respectively. If $y = \nu$, then the PMF of a DTPL(α, γ, ν) is monotonically decreasing in α on (-∞,∞) with the limits of one and zero at -infinity and infinity, respectively.*

Proof First, consider the case $y = \gamma$. By dividing the numerator and the denominator of the PMF (4.4) by $\left(\frac{1}{\gamma}\right)^{\alpha}$, we obtain

$$f_Y(\gamma) = \frac{1}{1 + \left(\frac{\gamma}{\gamma+1}\right)^{\alpha} + \cdots + \left(\frac{\gamma}{\nu}\right)^{\alpha}}. \tag{4.18}$$

Since for each $j = \gamma + 1, \cdots, \nu$ we have

$$0 < \frac{\gamma}{j} < 1, \tag{4.19}$$

the quantity

$$\left(\frac{\gamma}{\gamma+1}\right)^{\alpha} + \cdots + \left(\frac{\gamma}{\nu}\right)^{\alpha} \tag{4.20}$$

in (4.18) is monotonically decreasing in α, with the limits of ∞ and 0 at $-\infty$ and $+\infty$, respectively. In view of this, the function $f_Y(\gamma)$ in (4.18) is monotonically increasing in α on $(-\infty, \infty)$, with the limits of 0 and 1 at $-\infty$ and ∞, respectively. Now, consider the case $y = \nu$. By proceeding analogously as above, we write

$$f_Y(\nu) = \frac{1}{\left(\frac{\nu}{\gamma}\right)^{\alpha} + \left(\frac{\nu}{\gamma+1}\right)^{\alpha} + \cdots + \left(\frac{\nu}{\nu-1}\right)^{\alpha} + 1}. \tag{4.21}$$

Since for each $j = \gamma, \gamma + 1, \cdots, \nu - 1$ we have

$$\frac{\nu}{j} > 1, \tag{4.22}$$

the quantity

$$\left(\frac{\nu}{\gamma}\right)^{\alpha} + \cdots + \left(\frac{\nu}{\nu-1}\right)^{\alpha} \tag{4.23}$$

in (4.21) is increasing in α on $(-\infty, \infty)$, with the limits of 0 and ∞ at $-\infty$ and ∞, respectively. Thus, the function $f_Y(\nu)$ is decreasing in α on $(-\infty, \infty)$, with the limits of 1 and 0 at $-\infty$ and ∞, respectively. This concludes the proof.

4.3 Moments

Since DTPL(α, γ, ν) distribution has finite and positive support, its moment of any order $\eta \in (-\infty, \infty)$ exist, and

$$E(Y^\eta) = \sum_{y=\gamma}^{\nu} y^\eta f_Y(y). \tag{4.24}$$

Thus, the ηth moment of the DTPL(α, γ, ν) distribution is given by

$$E(Y^\eta) = \frac{\sum_{y=\gamma}^{\nu} y^\eta \frac{1}{y^\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}, \tag{4.25}$$

which can be simplified as

$$E(Y^\eta) = \frac{\sum_{y=\gamma}^{\nu} y^{\eta-\alpha}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}. \tag{4.26}$$

Proposition 4.8 *The mean of a DTPL(α, γ, ν) random variable Y is given by*

$$E(Y) = \frac{\sum_{y=\gamma}^{\nu} y^{-(\alpha-1)}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^\alpha}}. \tag{4.27}$$

Remark 4.9 *When $\alpha = 0$, the mean of the DTPL(α, γ, ν) random variable Y becomes*

$$E(Y) = \frac{\nu(\nu+1) - \gamma(\gamma-1)}{2(\nu+\gamma-1)}. \tag{4.28}$$

When $\alpha = 1$, the mean of the DTPL(α, γ, ν) random variable Y is

$$E(Y) = \frac{\nu+\gamma-1}{\sum_{j=\gamma}^{\nu} j^{-1}}. \tag{4.29}$$

Close examination of the formula (4.27) shows that the mean converges to γ as α approaches infinity and converges to ν as α converges to negative infinity. Actually, as shown below, the mean is actually a monotonically decreasing function as α increases from negative to positive infinity.

Proposition 4.10 *When $\gamma < \nu$ then the mean of a $DTPL$ random variable is monotonically decreasing function of α on $(-\infty, \infty)$. Moreover, the limits are γ and ν at infinity and minus infinity, respectively.*

Proposition 4.11 *The second moment of a DTPL(α, γ, ν) random variable Y is given by*

$$E(Y^2) = \frac{\sum_{y=\gamma}^{\nu} y^{-(\alpha-2)}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^{\alpha}}}. \tag{4.30}$$

Remark 4.12 *When $\alpha = 2$, then we have*

$$E(Y^2) = \frac{\nu + \gamma - 1}{\sum_{j=\gamma}^{\nu} j^{-2}}. \tag{4.31}$$

Remark 4.13 *The variance of a DTPL(α, γ, ν) random variable Y is*

$$Var(Y) = \frac{\sum_{y=\gamma}^{\nu} y^{-(\alpha-2)}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^{\alpha}}} - \left(\frac{\sum_{y=\gamma}^{\nu} y^{-(\alpha-1)}}{\sum_{j=\gamma}^{\nu} \frac{1}{j^{\alpha}}} \right)^2. \tag{4.32}$$

Chapter 5

Estimation

In this chapter, we consider the problem of estimating the parameters of the discrete truncated power law distribution. The estimates of the parameters shall be based on the maximum likelihood method. We shall first look at estimating γ and ν before estimating α.

5.1 Likelihood Function

Consider a vector $(Y_1, Y_2, \cdots, Y_n)$ of independently and identically distributed variables that follow the $DTPL(\alpha, \gamma, \nu)$ distribution. Then the likelihood function can be generated by finding product of the density functions, leading to

$$L(\alpha, \gamma, \nu) = \prod_{i=1}^{n} f_Y(y_i). \tag{5.1}$$

By taking into account the formula for the PMF $f_Y(y)$, given in (4.4), we can write the likelihood function as

$$L(\alpha, \gamma, \nu) = \frac{\prod_{i=1}^{n} y_i^{-\alpha}}{\left(\sum_{j=\gamma}^{\nu} j^{-\alpha}\right)^n}. \tag{5.2}$$

5.2 Estimating γ and ν

In order to estimate the parameters γ and ν we shall first order the random sample $Y_1, ..., Y_n$ to obtain the order statistics,

$$Y_{(1)} \leq \cdots \leq Y_{(n)}, \tag{5.3}$$

with $y_{(1)} \leq \cdots \leq y_{(n)}$ denoting their particular values. The likelihood function then becomes

$$L(\alpha, \gamma, \nu) = \frac{\prod_{i=1}^{n} y_{(i)}^{-\alpha}}{\left(\sum_{j=\gamma}^{\nu} j^{-\alpha}\right)^{n}} \;\; for \;\; \gamma \leq y_{(1)}, \;\; \nu \geq y_{(n)}, \;\; -\infty < \alpha < \infty. \tag{5.4}$$

Now, by Proposition 4.4, the likelihood function is monotonically increasing in γ and decreasing in ν (over their respective domains) for any fixed value of $\alpha \in (-\infty, \infty)$ and for the above sample. From this, we can draw the conclusion that the MLEs of these two parameters are the extreme order statistics,

$$\hat{\gamma} = Y_{(1)} \;\text{ and }\; \hat{\nu} = Y_{(n)}. \tag{5.5}$$

Subsequently, we need to maximize the function

$$L(\alpha, \hat{\gamma}, \hat{\nu}) = \frac{\prod_{i=1}^{n} y_{(i)}^{-\alpha}}{\left(\sum_{j=\hat{\gamma}}^{\hat{\nu}} j^{-\alpha}\right)^{n}}, \;\; \alpha \in R, \tag{5.6}$$

with respect to the parameter α.

5.3 Estimating α

We begin by finding the log-likelihood function, which according to (5.6), is of the form

$$l(\alpha) = -\alpha \sum_{i=1}^{n} \log x_i - n \log \sum_{j=\hat{\gamma}}^{\hat{\nu}} j^{-\alpha}, \tag{5.7}$$

with x_i denoting $y_{(i)}$ for convenience.

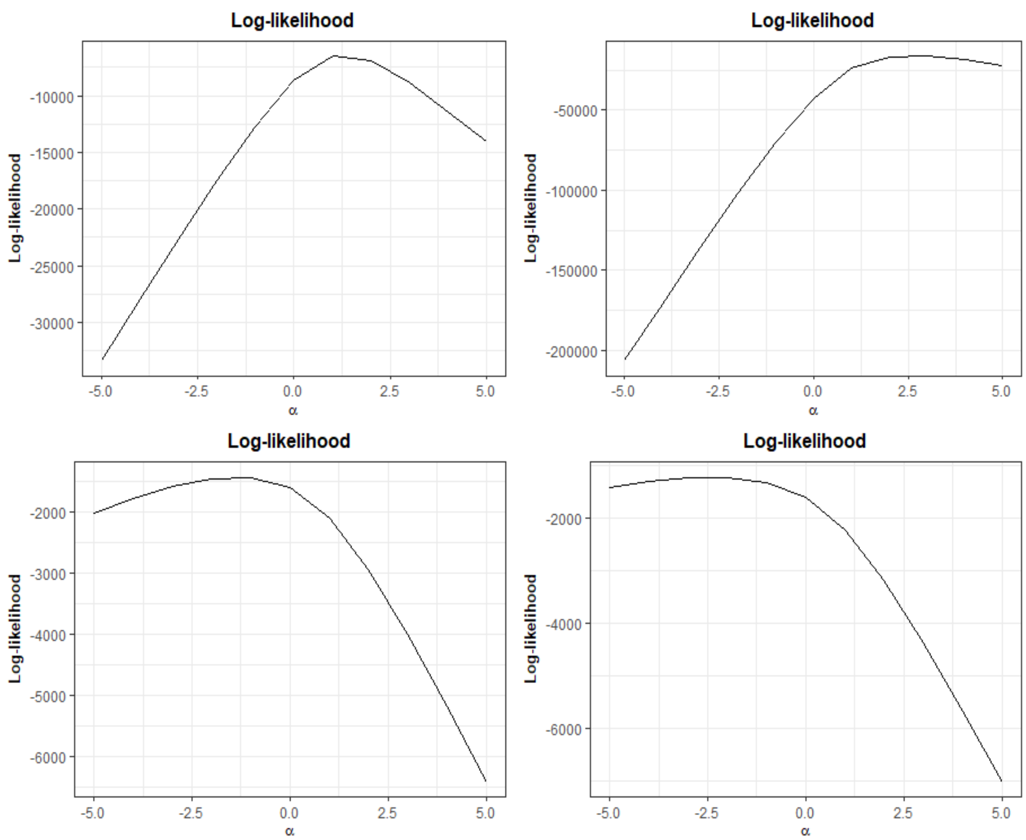

Figure 5.1: Log-likelihood function of α for simulated data

As shown in Figure 5.1, the plot of the log-likelihood function from the simulated data gradually increases till it gets to a maximum value and then gradually decreases. This suggests that there actually is a unique value of α that maximizes the log-likelihood.

To find the maximum value of $l(\alpha)$ with respect to α, we shall find the derivative of $l(\alpha)$ and set it to 0. A straightforward calculation shows that

$$l'(\alpha) = -\sum_{i=1}^{n} \log x_i + n\frac{\sum_{\hat{\gamma}}^{\hat{\nu}} j^{-\alpha} \log j}{\sum_{j=\hat{\gamma}}^{\hat{\nu}} j^{-\alpha}}. \tag{5.8}$$

As shown below, the existence and uniqueness of the MLE of α are closely related to the behavior of the function

$$w(\alpha) = \frac{\sum_{x_1}^{x_n} j^{-\alpha} \log j}{\sum_{j=x_1}^{x_n} j^{-\alpha}}, \tag{5.9}$$

where $x_1 = y_{(1)}$ and $x_n = y_{(n)}$.

Let us first consider the limiting behavior of $w(\alpha)$.

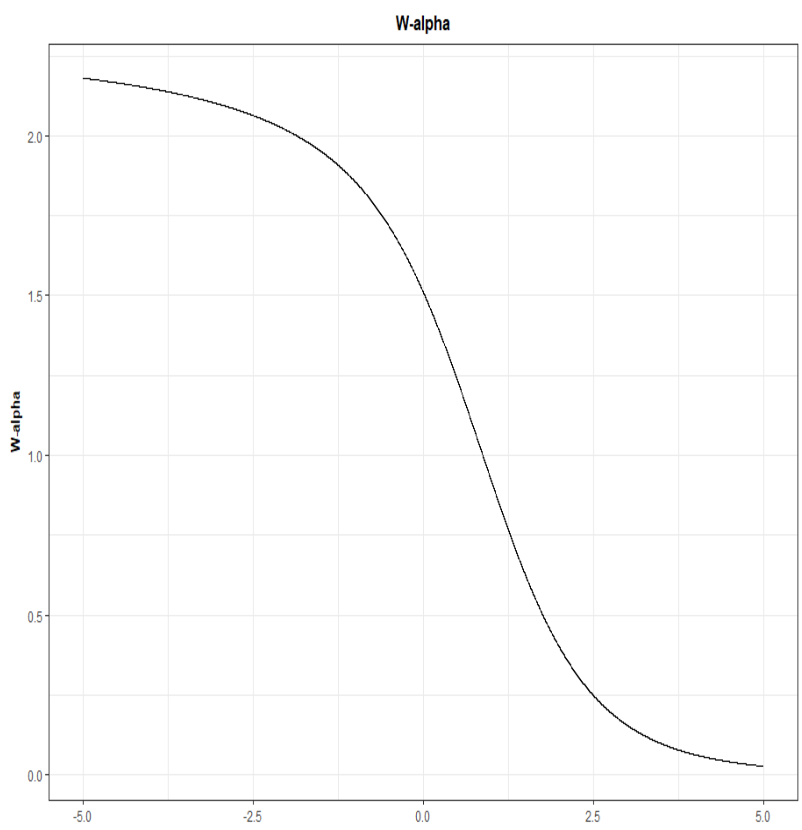

Figure 5.2: Typical behavior of w(α) suggesting that the function is decreasing

As we show below, this function has limits at $\pm\infty$ that are related to x_1 and x_n, the two extreme order statistics.

Proposition 5.1 *As* $\alpha \longrightarrow \infty$, $w(\alpha) \longrightarrow \log x_1$.

Proof From (5.9), we have

$$w(\alpha) = \frac{\left(\frac{1}{x_1}\right)^{\alpha}\log(x_1) + \left(\frac{1}{x_1+1}\right)^{\alpha}\log(x_1+1) + \cdots + \left(\frac{1}{x_n}\right)^{\alpha}\log(x_n)}{\left(\frac{1}{x_1}\right)^{\alpha} + \left(\frac{1}{x_1+1}\right)^{\alpha} + \cdots + \left(\frac{1}{x_n}\right)^{\alpha}}. \tag{5.10}$$

By multiplying the numerator and the denominator in (5.10) by x_1^{α}, we obtain

$$w(\alpha) = \frac{\log(x_1) + \left(\frac{x_1}{x_1+1}\right)^{\alpha}\log(x_1+1) + \cdots + \left(\frac{x_1}{x_n}\right)^{\alpha}\log(x_n)}{1 + \left(\frac{x_1}{x_1+1}\right)^{\alpha} + \cdots + \left(\frac{x_1}{x_n}\right)^{\alpha}}. \tag{5.11}$$

Since $x_1 < x_1 + 1 < \cdots < x_n$, we have that, as $\alpha \longrightarrow \infty$,

$$\left(\frac{x_1}{x_1+i}\right)^{\alpha} \longrightarrow 0$$

for each $i = 1, 2, \cdots, x_n - x_1$. Consequently, in the view of (5.11), we conclude that

$$w(\alpha) \longrightarrow \log(x_1)$$

as $\alpha \longrightarrow \infty$. Thus concludes the proof.

Proposition 5.2 *As* $\alpha \longrightarrow -\infty$, $w(\alpha) \longrightarrow \log x_n$.

Proof For convenience, we let $\beta = -\alpha > 0$, so that

$$w(\alpha) = \frac{\left(\frac{1}{x_1}\right)^{-\beta}\log(x_1) + \left(\frac{1}{x_1+1}\right)^{-\beta}\log(x_1+1) + \cdots + \left(\frac{1}{x_n}\right)^{-\beta}\log(x_n)}{\left(\frac{1}{x_1}\right)^{-\beta} + \left(\frac{1}{x_1+1}\right)^{-\beta} + \cdots + \left(\frac{1}{x_n}\right)^{-\beta}}. \tag{5.12}$$

$$= \frac{x_1^{\beta}\log x_1 + (x_1+1)^{\beta}\log(x_1+1) + \cdots + x_n^{\beta}\log x_n}{x_1^{\beta} + (x_1+1)^{\beta} + \cdots + x_n^{\beta}}. \tag{5.13}$$

By multiplying the numerator and denominator in the above expression by $\frac{1}{x_n^{\beta}}$, we obtain

$$w(\beta) = \frac{\left(\frac{x_1}{x_n}\right)^{\beta}\log(x_1) + \left(\frac{x_1+1}{x_n}\right)^{\beta}\log(x_1+1) + \cdots + \log(x_n)}{\left(\frac{x_1}{x_n}\right)^{-\beta} + \left(\frac{x_1+1}{x_n}\right)^{\beta} + \cdots + 1}. \tag{5.14}$$

Since $x_1 < x_1 + 1 < \cdots < x_n$, as $\beta \longrightarrow \infty$ we have

$$\left(\frac{x_1+i}{x_n}\right)^{\beta} \longrightarrow 0$$

as $\beta \longrightarrow \infty$ for each $i = 0, 1, 2, \cdots, x_n - x_1 - 1$. Thus, we have

$$w(\alpha) \longrightarrow \log(x_n)$$

as $\alpha \longrightarrow -\infty$, as desired. This concludes the proof.

Figure 5.2 shows a typical behavior of the function $w(\alpha)$ obtained from simulated data, where smallest value was $x_1 = 1$ and the largest value was $x_n = 10$. The limiting behavior of $\pm\infty$ is clearly seen here.

In addition, the figure shows the function is decreasing over its domain. As shown by the analysis below, this is actually the case. Indeed, when we calculate its derivative, we obtain

$$w^{'}(\alpha) = \frac{-\sum_{j=x_1}^{x_n}(\log j)^2 j^{-\alpha}\sum_{j=x_1}^{x_n} j^{-\alpha} + \left(\sum_{j=x_1}^{x_n} j^{-\alpha}\log j\right)^2}{\left(\sum_{j=x_1}^{x_n} j^{-\alpha}\right)^2}. \tag{5.15}$$

To show that $w(\alpha)$ is decreasing, we need to show that the numerator of $w^{'}(\alpha)$ is negative, which is equivalent to

$$\left(\sum_{j=x_1}^{x_n} j^{-\alpha}\log j\right)^2 \leq \sum_{j=x_1}^{x_n} (\log j)^2\, j^{-\alpha} \sum_{j=x_1}^{x_n} j^{-\alpha}. \tag{5.16}$$

Since for any $x_1 \leq j \leq x_n$ we have $j^{-\alpha} \geq 0$ and $\log j \geq 0$, the above is true if

$$\left(\sum_{j=1}^{k} a_j b_j\right)^2 \leq \sum_{j=1}^{k} b_j^2 a_j \sum_{j=1}^{k} a_j \tag{5.17}$$

holds for any $k \geq 1$ and non-negative a_j,b_j. The result below shows that this is indeed true.

Lemma 5.3 *For any $k \leq 1$ and non-negative $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$, we have*

$$\left(\sum_{j=1}^{k} a_j b_j\right)^2 \leq \sum_{j=1}^{k} b_j^2 a_j \sum_{j=1}^{k} a_j. \tag{5.18}$$

Proof We prove this by induction.

Let us first assume that $k = 1$. In this case, the equation (5.18) becomes

$$a_1^2 b_1^2 \leq a_1^2 b_1^2,$$

which is obviously true. For $k = 2$, the equation becomes

$$[a_1 b_1 + a_2 b_2]^2 \leq \left[a_1 b_1^2 + a_2 b_2^2\right][a_1 + b_1],$$

which, after some algebra, reduces to

$$0 \leq a_1 a_2 b_1^2 + a_1 a_2 b_2^2 - 2a_1 a_2 b_1 b_2.$$

Since the right-hand-side above is

$$(b_1 - b_2)^2 a_1 a_2,$$

this quantity is non-negative, since by assumption $a_i \geq 0$. We shall now assume the statement is true for some $k \geq 1$, and then show that it is also true for $k+1$. Now, with k replaced by $k+1$, the statement (5.18) becomes

$$\left(\sum_{j=1}^{k+1} a_j b_j\right)^2 \leq \sum_{j=1}^{k+1} b_j^2 a_j \sum_{j=1}^{k+1} a_j.$$

By grouping the first k terms in the sums above, we obtain

$$\left(\sum_{j=1}^{k} a_j b_j + a_{k+1} b_{k+1}\right)^2 \leq \left(\sum_{j=1}^{k} b_j^2 a_j + b_{k+1}^2 a_{k+1}\right) \left(\sum_{j=1}^{k} a_j + a_{k+1}\right),$$

which, after some algebra, becomes

$$\left(\sum_{j=1}^{k+1} a_j b_j\right)^2 + (a_{k+1} b_{k+1})^2 + 2 \sum_{j=1}^{k} a_j b_j (a_{k+1} b_{k+1})$$

$$\leq \sum_{j=1}^{k} b_j^2 a_j \sum_{j=1}^{k} a_j + (a_{k+1} b_{k+1})^2 + \sum_{j=1}^{k} a_j b_j^2 a_{k+1} + \sum_{j=1}^{k} a_j a_{k+1} b_{k+1}^2.$$

Now, by the induction step, equation (5.18) holds for k, so the above will be true if

$$2 \sum_{j=1}^{k} a_j b_j (a_{k+1} b_{k+1}) \leq \sum_{j=1}^{k} a_j b_j^2 a_{k+1} + \sum_{j=1}^{k} a_j a_{k+1} b_{k+1}^2.$$

When we move all the terms on one side and factor out a_{k+1}, this becomes

$$0 \leq a_{k+1} \left(\sum_{j=1}^{k} a_j b_j^2 + \sum_{j=1}^{k} a_j b_{k+1}^2 - 2 \sum_{j=1}^{k} a_j b_j b_{k+1}\right).$$

After combing the three sums, we obtain an equivalent inequality to be

$$0 \le \sum_{j=1}^{k} a_j \left(b_j^2 + b_{k+1}^2 - 2b_j b_{k+1}\right),$$

or

$$0 \le \sum_{j=1}^{k} a_j (b_j - b_{k+1})^2$$

upon completing the squares.

The above inequality is satisfied since, by assumption, $a_j > 0$. This completes the proof.

Thus, by Lemma 5.3, we can conclude that (5.16) is true and thus $w(\alpha)$ is a monotonically decreasing function.

We now connect this property with the log-likelihood function. According to (5.8), we have

$$l^{'}(\alpha) = n\left(w(\alpha) - \frac{1}{n}\sum_{i=1}^{n}\log x_i\right). \tag{5.19}$$

Now according to the above results, the function $w(\alpha)$ is decreasing in α with limits of $\log x_n$ and $\log x_1$ at $-\infty$ and ∞, respectively. On the other hand, as long as all the data points are not equal $(x_1 < x_n)$, we must have

$$\log x_1 < \frac{1}{n}\sum_{i=1}^{n}\log x_i < \log x_n.$$

Thus, in the view of the above facts, we can conclude that there exists a unique $\hat{\alpha}$ such that $l^{'}(\alpha) > 0$ for $\alpha < \hat{\alpha}$, $l^{'}(\alpha) = 0$ for $\alpha = \hat{\alpha}$, and $l^{'}(\alpha) < 0$ for $\alpha > \hat{\alpha}$. Consequently, the log-likelihood function $l(\alpha)$ is monotonically increasing on $(0, \hat{\alpha})$ and its is decreasing on $(\hat{\alpha}, \infty)$, with the largest value occurring at $\hat{\alpha}$. This value $\hat{\alpha}$ is the maximum likelihood estimator of α, which according to the above analysis exists and is unique. This is summarized in the result below.

Proposition 5.4 *Let $Y_1, \ldots, Y_n$ be a random sample from a DTPL distribution, where not all sample values are equal. Then, there are unique MLEs of γ, ν and α, where $\hat{\gamma} = Y_{(1)}$ (the smallest data value), $\hat{\nu} = Y_{(n)}$ (the largest data value) and $\hat{\alpha}$ is the unique α that solves the equation*

$$\frac{\sum_{j=\hat{\gamma}}^{\hat{\nu}} j^{-\alpha} \log j}{\sum_{j=\hat{\gamma}}^{\hat{\nu}} j^{-\alpha}} = \frac{1}{n} \sum_{i=1}^{n} \log Y_i. \tag{5.20}$$

Note that the values of $\hat{\gamma}$ and $\hat{\nu}$ are given explicitly, while $\hat{\alpha}$ has to be found by a numerical search by standard methods such as Newton-Raphson or bisection as discussed below.

Remark 5.5 *Our analysis above shows that if $w(0) > \frac{1}{n} \sum_{i=1}^{n} \log Y_i$, we will have $\hat{\alpha} > 0$, while for $w(0) < \frac{1}{n} \sum_{i=1}^{n} \log Y_i$, we will have $\hat{\alpha} < 0$. Since*

$$w(0) = \frac{1}{Y_{(n)} - Y_{(1)} + 1} \sum_{j=Y_{(1)}}^{Y_{(n)}} \log j,$$

the sign of α is determined by the sign of the statistic

$$S_n = \frac{1}{Y_{(n)} - Y_{(1)} + 1} \sum_{j=Y_{(1)}}^{Y_{(n)}} \log j - \frac{1}{n} \sum_{i=1}^{n} \log Y_{(i)}, \tag{5.21}$$

which seems to play an important role in this problem.

5.3.1 Bisection Method

As discussed above, one of the standard procedures for estimating α is the bisection method. To understand the reason why bisection method works in this case, let us first consider the plots of $l'(\alpha)$ from simulated data.

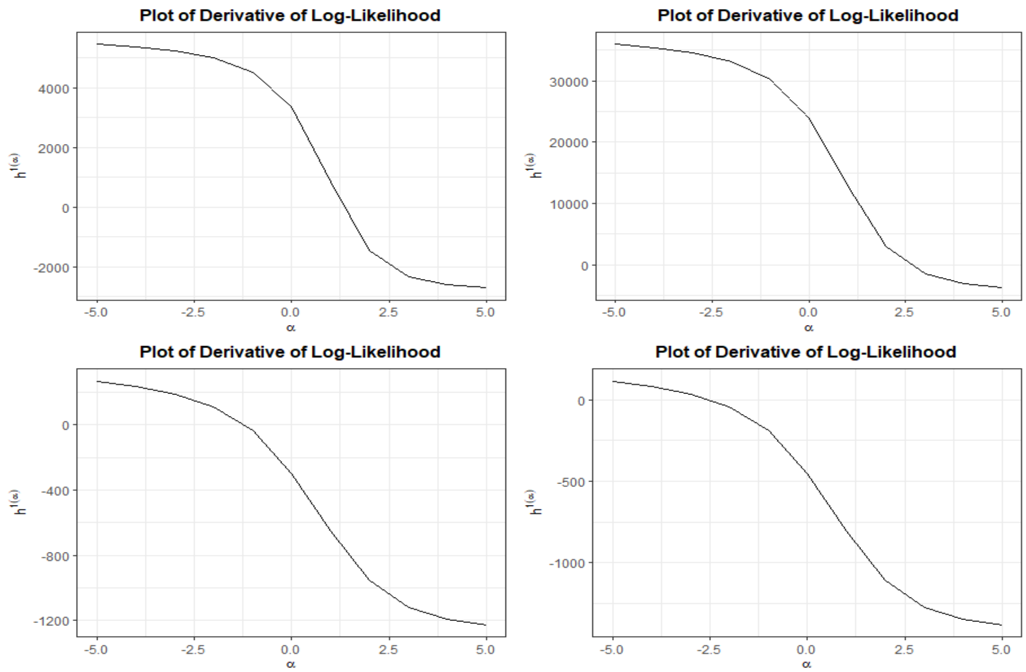

Figure 5.3: Plots of $l^{'}(\alpha)$ from simulated data.

The plots shows the behavior of $l^{'}(\alpha)$ which shows a monotonically decreasing function with 0 in its range. Due to this behavior the bisection method can be used as there exists only one unique root as discussed above.

The algorithm for finding this root using the bisection method is as follows:

- Select an interval you suspect contains the root, $[a_1, b_1]$, such that $f(a_1) > 0$ and $f(b_1) < 0$, where $f = l^{'}$.
- Select a tolerance level.
- Find the midpoint of this interval, $c_1 = \frac{a_1+b_1}{2}$
- Evaluate $f(a_1), f(b_1), f(c_1)$.
- If $f(c_1) > 0$, select the new interval to be $[a_1, c_1]$, else select the interval to be $[c_1, b_1]$.
- Continue this iteration until $f(c_n) - f(c_{n-1}) <$ tolerance level.

- Return the value $\hat{\alpha} = c_n$.

5.4 Simulation

Using the estimation methods developed in this chapter for the parameters α,γ and ν in (5.5) and (5.20), we now simulate data from the DTPL(α,γ,ν) distribution, calculate the MSE for each parameter and use this result to test how good the estimation is. The following table presents the estimated values of DTPL(α,γ,ν)

Table 5.1: Estimated parameters of DTPL(α,γ,ν) distribution

n	k	α	$\hat{\alpha}$	MSE($\hat{\alpha}$)	γ	$\hat{\gamma}$	MSE($\hat{\gamma}$)	ν	$\hat{\nu}$	MSE($\hat{\nu}$)
10	10000	-1.2	0.64815	4.144047	1	5.2	17.8	10	10	0
10	10000	1.8	2.2349200	0.6488452	1	1	0	10	4.8	29.2
100	10000	0	-0.002965	0.02177839	1	1	0	10	10	0
500	10000	0.4	0.357665200	0.005499714	1	1	0	10	10	0
1000	10000	1.2	1.25366	0.004780834	1	1	0	10	10	0

The estimated values of $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\nu}$ are averaged values based on $k = 10000$ estimated values. The mean squared errors are calculated based on k iterations. From Table 5.1 above, it is clear that the $\hat{\alpha}$, $\hat{\gamma}$ and $\hat{\nu}$ converge to α, γ and ν as sample size (n) gets larger. It is also interesting to note that for small sample sizes, $\hat{\gamma}$ and $\hat{\nu}$ depend on the value of α for accuracy and this as a result of Propositon 4.5. Also, $\hat{\gamma}$ and $\hat{\nu}$ converge faster to γ and ν respectively than $\hat{\alpha}$ converges to α.

Chapter 6

Examples

In this section, in order to demonstrate the usefulness of this new distribution and its properties as described in this thesis paper, we apply them to real life data. These datasets, from their univariate plots, are suspected to follow a DTPL(γ, ν, α) distribution. The datasets are drawn from a broad variety of different human endeavors, including insurance, linguistics and social science. These datasets are:

- The severity of terrorist attacks worldwide from February 1968 to June 2006, measured as the number of deaths directly resulting from the attacks.
- The frequency of occurrence of unique words in the novel Moby Dick by Herman Melville.
- The total area of external buildings other than the main house from contracts concerning home insurance policies from this decade in France.
- Surface area of terrain surface from contracts concerning home insurance policies from this decade in France.

We first look at the Total Surface Area of External Building and the Terrain Surface datasets. In order to check if this dataset follows a DTPL(γ, ν, α) distribution, the relative frequencies of the observations are compared with the estimated PMFs at each observation in the support, and compared using bar-charts and scatterplots as seen below. The parameters are estimated by the method of maximum likelihood, as described in Chapter 5. Table 6.1 shows the estimated values across all datasets.

Table 6.1: The results of estimation of α, γ and ν when DTPL(α, γ, ν) distribution is fitted to the four datasets.

Datasets \ Parameters	$\hat{\alpha}$	$\hat{\gamma}$	$\hat{\nu}$
Terrorist attacks	1.89344	1	2749
Words Frequency	1.77206	1	14086
Contract count of total area of external buildings	2.59386	1	14
Contract count of total area of terrain surface area	1.74562	1	11

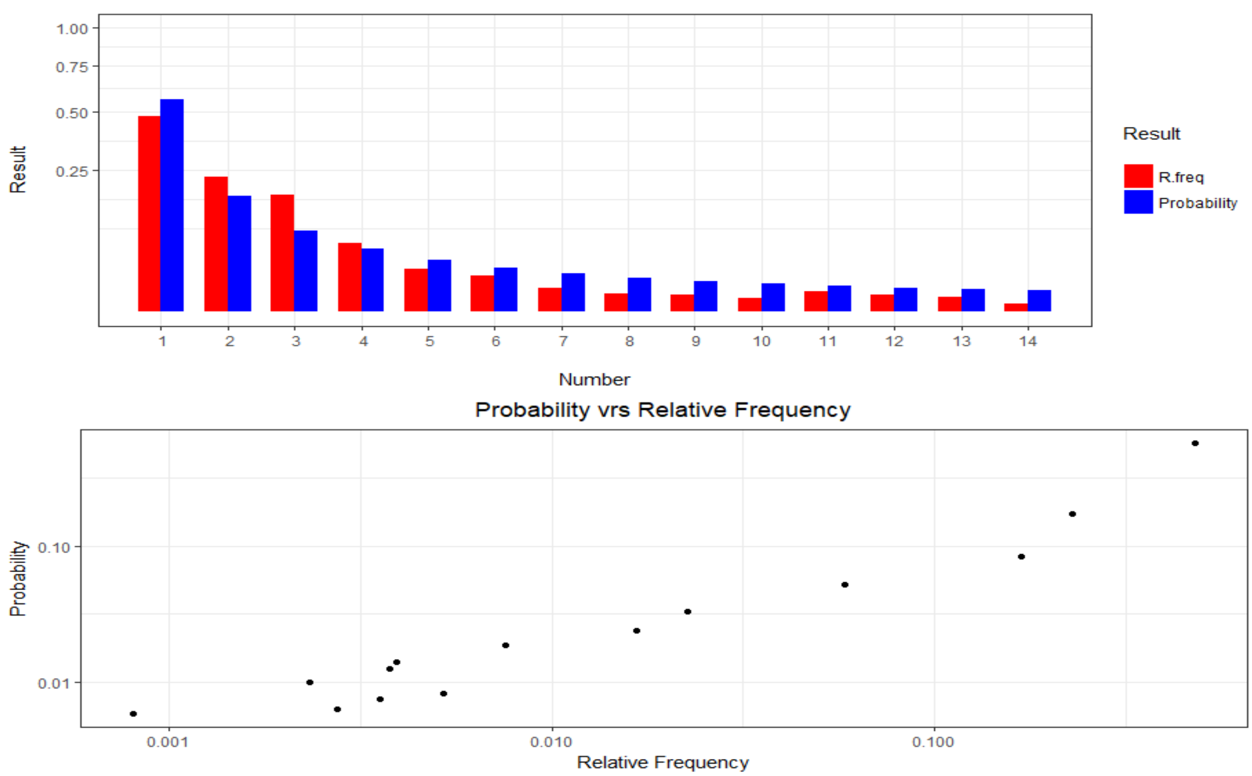

Figure 6.1: Total Surface Area of External Building

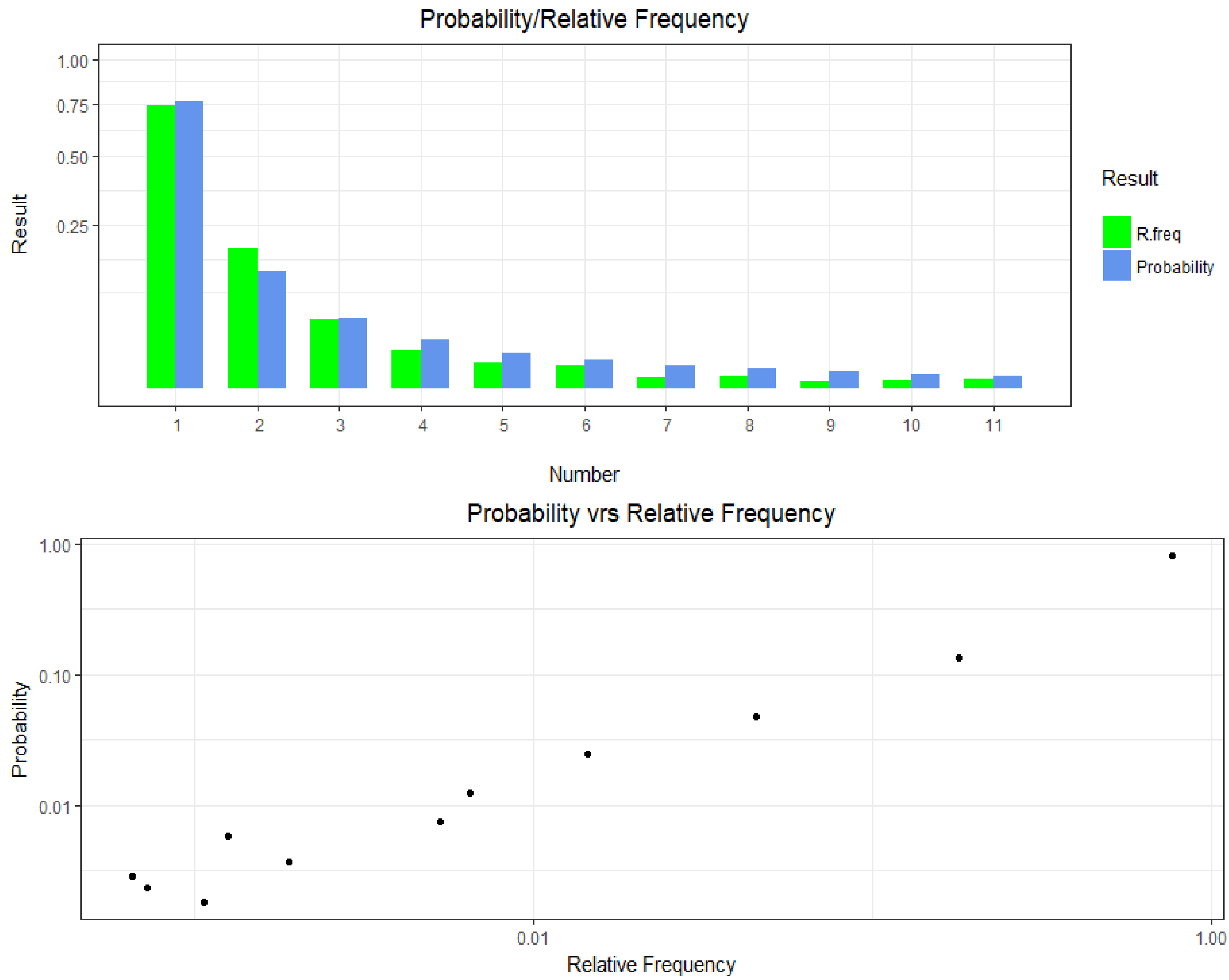

Figure 6.2: Terrain Surface

From the plots above, we can see that the both datasets appear to follow the DTPL(γ, ν, α) distribution as there is an almost linear relationship with the relative frequency of the data and probabilities.

The same process as above is applied to the word frequency and number of death datasets. But in this case only scatter-plots are used because the maximum of the ordered dataset was too high thus a bar-plot would not be appropriate. Again, we see an almost linear relation between relative frequency and probability indicating that the datasets have a DTPL(γ, ν, α) distribution. These are shown below:

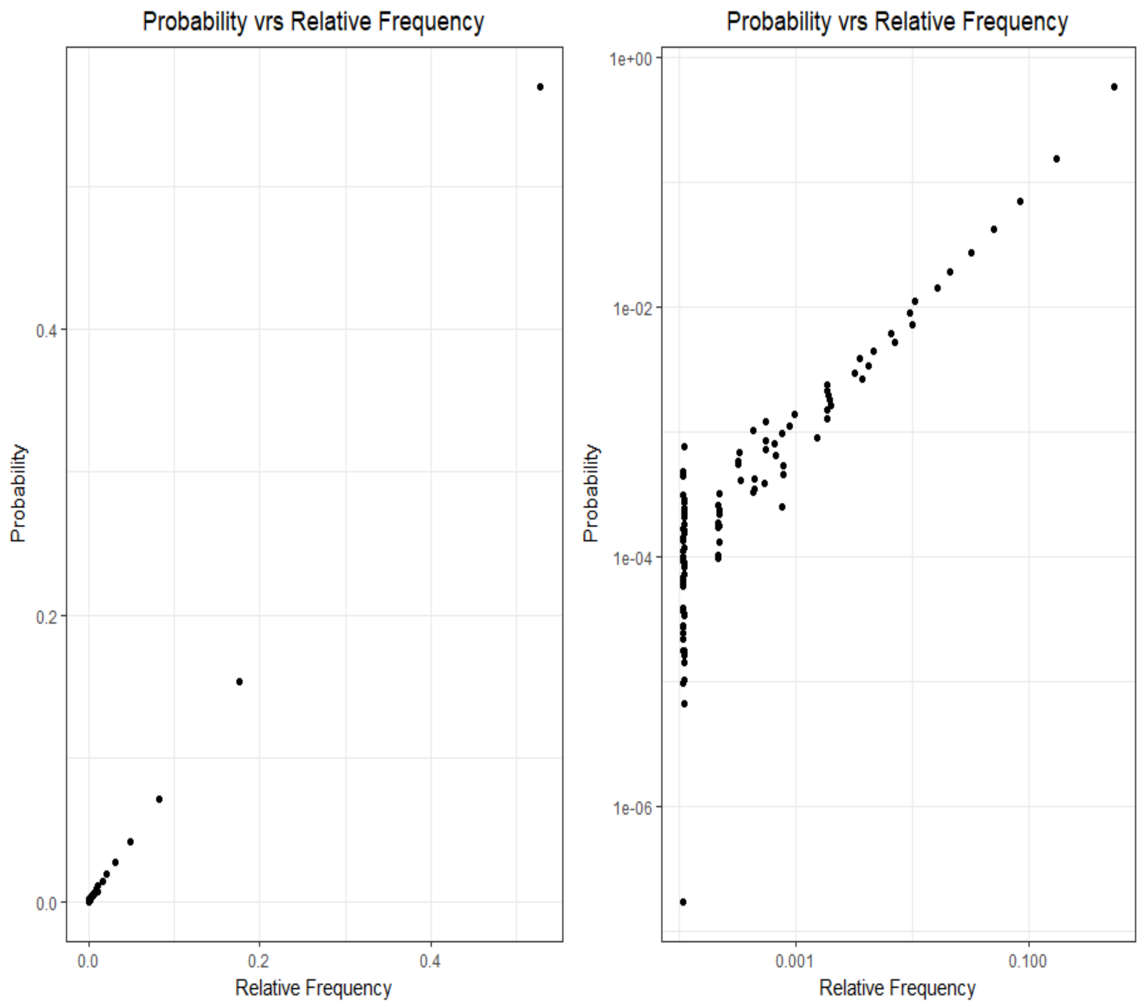

Figure 6.3: Number of Deaths

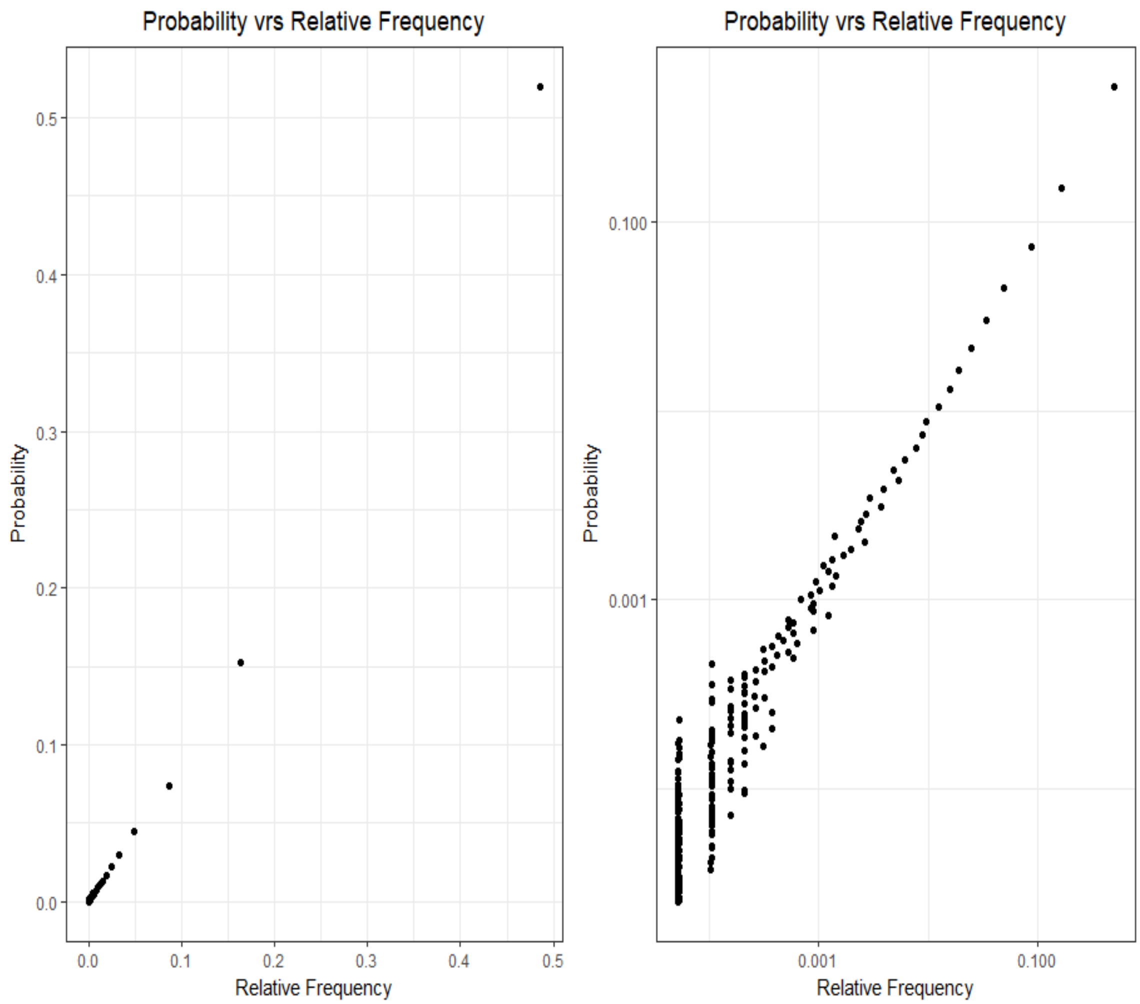

Figure 6.4: Word Frequency

Chapter 7

Appendix

Here we include several functions written in the R language, which were used in this work.

7.1 Function for Computing Loglikelihood

```
loglik_dptl <- function(a,b,X,plot=TRUE){
alpha <- seq(a,b,1)
alpha <- as.matrix(alpha,nrow=length(alpha),
ncol=1,byrow=FALSE)

X <- sort(X)
X <- as.matrix(X,nrow=length(X),ncol=1,
byrow=FALSE)
n <- length(X)
logy <- apply(X,2,log)
logy <- as.matrix(logy,nrow=length(X),
ncol=1,byrow=FALSE)
```

```
sumlog <- apply(logy, 2, sum)
logfirst <- -sumlog*alpha
logfirst <- as.matrix(logfirst,
nrow=length(logfirst),ncol=1, byrow=FALSE)
Y <- seq(X[1],X[n],1)
Y <- as.matrix(Y,nrow=length(X),
ncol=1,byrow=FALSE)
f <- function(x,alpha){
func <- x^(-alpha)
func
}
logsecond <- outer(Y,alpha,f)
logsecond <- as.data.frame(logsecond,
 nrow=length(Y),ncol=length(alpha), byrow=FALSE)
logsecond1 <- apply(logsecond,2,sum)
logsecond1 <- as.matrix(logsecond1,
nrow=length(logsecond1), ncol = 2, byrow = FALSE)
logsecond2 <- apply(logsecond1, 2, log)
logsecond2 <- n*logsecond2
logsecond2 <- as.matrix(logsecond2,
nrow=length(logsecond1), ncol = 2, byrow = FALSE)
loglik1 <- logfirst[,1] - logsecond2[,1]
loglik1 <- as.matrix(loglik1, nrow=length(loglik1),
ncol = 2, byrow = FALSE)
loglik1 <- cbind(alpha,loglik1)
colnames(loglik1) <- list(c("alpha"),c("Loglik"))
```

```
if (plot==TRUE){
alpha <- seq(a,b,0.001)
alpha <- as.matrix(alpha,nrow=length(alpha),
ncol=1,byrow=FALSE)

X <- sort(X)
X <- as.matrix(X,nrow=length(X),ncol=1,byrow=FALSE)
n <- length(X)
logy <- apply(X,2,log)
logy <- as.matrix(logy,nrow=length(X),ncol=1,byrow=FALSE)
sumlog <- apply(logy, 2, sum)
logfirst <- -sumlog*alpha
logfirst <- as.matrix(logfirst,
nrow=length(logfirst),ncol=1, byrow=FALSE)
Y <- seq(X[1],X[n],1)
Y <- as.matrix(Y,nrow=length(X),ncol=1,byrow=FALSE)
f <- function(x,alpha){
func <- x^(-alpha)
func
}
logsecond <- outer(Y,alpha,f)
logsecond <- as.data.frame(logsecond, nrow=length(Y),
ncol=length(alpha), byrow=FALSE)
logsecond1 <- apply(logsecond,2,sum)
logsecond1 <- as.matrix(logsecond1,
=length(logsecond1), ncol = 2, byrow = FALSE)
```

```
logsecond2 <- apply(logsecond1 , 2, log)
logsecond2 <- n*logsecond2
logsecond2 <- as.matrix(logsecond2 ,
 nrow=length(logsecond1), ncol = 2, byrow = FALSE)
loglik1 <- logfirst[,1] - logsecond2[,1]
loglik1 <- as.matrix(loglik1 ,
nrow=length(loglik1), ncol = 2, byrow = FALSE)
loglik1 <- cbind(alpha,loglik1)
colnames(loglik1) <- list(c("alpha"),c("Loglik"))
g      <- as.data.frame(loglik1)
a      <- g[which.max(g$Loglik),][,1]
textlab <- paste("alpha␣=",a, sep = "")
library(ggplot2)
p <- ggplot(g,aes(alpha,y=Loglik))
p <- p+geom_line()
p <- p+geom_vline(xintercept =
  g[which.max(g$Loglik),][,1], color="red")
p <- p+ylab("Log-likelihood")
p <- p+xlab(expression(alpha))
p <- p+ggtitle("Log-likelihood")
p <- p+
theme(plot.title = element_text(hjust = 0.5))+
annotate("text", x = a,
y =quantile(g[,2],0.25), label =textlab, colour="blue", angle=90, vjust =
t2<-theme(
axis.title.x =
```

```
element_text(face="bold", color="black", size=10),
axis.title.y =
element_text(face="bold", color="black", size=10),
plot.title = element_text(face="bold",
color = "black", size=12,hjust = 0.5)
)
p <- p+theme_bw() + t2
p
}
else{
loglik1
}

}
```

7.2 Function for Probability Mass Function of DTPL distribution

```
p.dtpl <- function(x,gamma,nu,alpha){
X <- seq(gamma,nu,1)
X <- as.matrix(X)
f <- function(x,alpha){
func <- x^(-alpha)
func
}
```

7.3 Function for Estimating α

```
bisect.y <- function(a,b,X, tol=0.00001){
fun <- function(x) f.d.1(x,X)
m = (a+b)/2
y1 = fun(a)
y2 = fun(b)
if (y1*y2 > 0){
print("wrong input")
}
i = 0
options(digits = 6)
while(abs((a-b)/2)> tol){
i=i+1
y1 = fun(a)
y2 = fun(b)
y3 = fun(m)
print(i)
print (m)
if((y1*y3)<0){
b=m
}
else{
a = m
}
m =(a+b)/2
}
```

```
}
```

7.4 Function for Simulating from DTPL distribution

```
rdtpl <- function(gamma,nu,alpha,n, plot=TRUE){
w <- runif(n,0,1)
Y <- seq(gamma,nu,1)
Y <- as.matrix(Y,nrow=length(Y),ncol=1,byrow=FALSE)
k <- length(Y)
f <- function(x,alpha){
func <- x^(-alpha)
func
}
Y1 <- sum(outer(Y,alpha,f))
u <- NULL
i <- gamma
for(i in gamma:nu ){
u[i] <- (outer(i,alpha,f))/Y1
}

a <- as.matrix(u)
if (gamma==1){
a <- a


}
```

```
else{
a <- as.matrix(a[-c(1:gamma-1),])
}

b <- as.matrix(cumsum(a))
c <- cbind(Y,b)
f7 <- function(x){
for (i in 1:length(c)){
if (x <= c[i,2]){
return(c[i,1])
}
}
}

d <- sapply(w,f7)
library(ggplot2)
if (plot==TRUE){
d <- as.data.frame(d)
colnames(d) <- list(c("Sample"))
p <- ggplot(d,aes(Sample))
p <- p+geom_bar(fill = "lightblue")
p <- p+ggtitle("Sample From DTPL")+
theme(plot.title = element_text(hjust = 0.5))
p <- p+scale_x_discrete(limits=gamma:nu)
t2<-theme(
axis.title.x = element_text
```

```
(face="bold", color="black", size=10),
axis.title.y = element_text(face=
"bold", color="black", size=10),
plot.title = element_text(face="bold",
color = "black", size=12,hjust = 0.5)
)
p <- p+theme_bw() + t2


p
}
else{
as.data.frame(d)
}
}
```

7.5 Function for Derivative of Loglikehood

```
f.d <- function(a,b,X,plot=TRUE){
alpha <- seq(a,b,1)
G <- sum(log(X))
n <- length(X)
X <- sort.int(X)
Y <- seq(X[1],X[n],1)
Y <- as.matrix(Y,nrow=length(X),ncol=1,byrow=FALSE)
f1 <- function(x,alpha){
func1 <- x^(-alpha) * log(x)
func1
```

```
}

S <- outer(Y, alpha , f1 )
S <- as.data.frame(S, nrow=length(Y),
ncol=length(alpha), byrow=FALSE)
S <- colSums(S)
f <- function(x, alpha){
func <- x^(-alpha)
func
}
R <- outer(Y, alpha , f)
R <- as.data.frame(R, nrow=length(Y),
ncol=length(alpha), byrow=FALSE)
R <- colSums(R)
C <- S/R
C <- n*C
W <- -G + C
W <- data.frame(alpha ,W)
if(plot==TRUE){
alpha <- seq(a,b,0.001)
G <- sum(log(X))
n <- length(X)
X <- sort.int(X)
Y <- seq(X[1] ,X[n] ,1)
Y <- as.matrix(Y,nrow=length(X) ,ncol=1,byrow=FALSE)
f1 <- function(x, alpha){
```

```
func1 <- x^(-alpha) * log(x)
func1
}

S <- outer(Y, alpha , f1)
S <- as.data.frame(S, nrow=length(Y),
ncol=length(alpha), byrow=FALSE)
S <- colSums(S)
f <- function(x, alpha){
func <- x^(-alpha)
func
}
R <- outer(Y, alpha , f)
R <- as.data.frame(R, nrow=length(Y),
ncol=length(alpha), byrow=FALSE)
R <- colSums(R)
C <- S/R
C <- n*C
W <- -G + C
W <- data.frame(alpha ,W)
c    <- est.alpha(a,b,X)
textlab <- paste("alpha =",c, sep = "")
library(ggplot2)
p <- ggplot(W, aes(alpha ,y=W))
p <- p+geom_line()+geom_vline(xintercept =c , color="red")
p <- p+xlab(expression(alpha))+ylab(expression(h^1 (alpha)))
```

```
p <- p+ggtitle("Plot of Derivative of Log-Likelihood")+
theme(plot.title = element_text(hjust = 0.5))+
annotate("text", x = c, y =quantile(W[,2],0.50),
label =textlab,
colour="blue", angle=90, vjust = -0.5,
 text=element_text(size=50))
t2<-theme(
axis.title.x = element_text(face="bold", color="black", size=10),
axis.title.y = element_text(face="bold", color="black", size=10),
plot.title = element_text(face="bold",
 color = "black", size=12,hjust = 0.5)
)
p <- p+theme_bw() + t2
p
}
else{
W
}
}
```

References

[1] Blank, A. and Solomon, S. (2000). Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A: Statistical Mechanics and its Applications*, 287(1-2):279–288.

[2] Buddana, A. and Kozubowski, T. J. (2014). Discrete Pareto distributions. *Economic Quality Control*, 29(2):143–156.

[3] Cattuto, C., Loreto, V., and Servedio, V. D. (2006). A Yule-Simon process with memory. *EPL (Europhysics Letters)*, 76(2):208.

[4] Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703.

[5] Clauset, A., Young, M., and Gleditsch, K. S. (2007). On the frequency of severe terrorist events. *Journal of Conflict Resolution*, 51(1):58–87.

[6] Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267.

[7] Kozubowski, T. J., Panorska, A. K., and Forister, M. L. (2015). A discrete truncated Pareto distribution. *Statistical Methodology*, 26:135–150.

[8] Kozubowski, T. J. and Podgórski, K. (2016). A generalized sibuya distribution. *Annals of the Institute of Statistical Mathematics (in press).*

[9] Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12):317–323.

[10] Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251.

[11] Montemurro, M. A. (2001). Beyond the Zipf–Mandelbrot law in quantitative linguistics. *Physica A: Statistical Mechanics and its Applications*, 300(3-4):567–578.

[12] Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5):323–351.

[13] Sibuya, M. (1979). Generalized hypergeometric, digamma and trigamma distributions. *Annals of the Institute of Statistical Mathematics*, 31(1):373–390.

[14] Zaninetti, L. and Ferraro, M. (2008). On the truncated Pareto distribution with applications. *Open Physics*, 6(1):1–6.