

Modeling Bursts and Heavy Tails in Inter-Arrival Claims in Non-Life Insurance

Mohamed Hanafy

Department Statistics, Zhejiang Gongshang University, Hangzhou, China

Email: mhanafy102@gmail.com

How to cite this paper: Hanafy, M. (2020). Modeling Bursts and Heavy Tails in Inter-Arrival Claims in Non-Life Insurance. *Journal of Financial Risk Management*, 9, 314-333.

<https://doi.org/10.4236/jfrm.2020.93017>

Received: August 24, 2020

Accepted: September 26, 2020

Published: September 29, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Current insurance models, assuming that inter-arrival time of claims, are distributed randomly and thus well approximated by Poisson processes. Here we provide clear proof that the timing of inter-claims fits by non-Poisson patterns, marked by rapid events, separated by long periods of inactivity. The time of inter-arrival claims will be heavy tailed, most claims will be executed quickly, while a few will have very long waiting times. We will model and analysis of insurance based on claim inter-arrival time, the time interval between two successive claims and the ability to carry out such modeling was limited by a lack of ecologically relevant data collected on claims inter-arrival. We propose a structured process behavior model based on data from Egyptian fire insurance company. Our analysis shows that claim activities can be represented by non-Poisson processes and that the subsequent distribution of inter-arrival activity times follows the Pareto distribution. These results will help researchers understand daily behavioral trends and create more sophisticated predictive models of claims.

Keywords

Inter-Arrival Time, Claims, Non-Life Insurance, Goodness of Fit, Pareto

1. Introduction

It is becoming increasingly important to understand the nature of the claims acts. Indeed, the quantitative discovery of the laws governing the probability of ruin is of major scientific importance, and requires us to tackle the factors that determine the timing of claims. Certainly, the interest in addressing the timing of statements in ruin probability is not new: it has a long history in mathematical literature, contributing to the emergence of some of the core principles in probability theory, Feller (1971). But most existing ruin probability models presume

that claim inter-arrival time is measured at a constant rate. Anderson (2003), meaning a claim has a fixed probability of engaging in a particular action within a given time interval. These models estimate the timing of claims by the Poisson method, in which there is a time interval between two consecutive claims. Called the time to wait or to intervene fits the exponential distribution, Haight (1967).

Poisson processes are at the root of the famous Erlang method, Erlang (1917). Nevertheless, a growing number of recent studies suggest that the timing of many acts is systematically deviating from the Poisson predictions. See Vazquez (2005), Grais et al. (2003), Pieropan et al. (2013) and Kwon et al. (2016).

We find that waiting or inter-event claims are best matched with heavy tails or distribution of Pareto, Bees et al. (2005). There's a striking contrast between a Poisson and heavy tailed activity: The exponential decay of the Poisson distribution forces successive events to match each other at relatively frequent time intervals and prevents very long waiting periods. Oliveira & Barabási (2005), on the other hand, the slowly decaying heavy tailed processes allow for very long periods of inactivity, which distinguish the bursts of intense activity.

In this paper, we propose formal models of (arrival process) inter-arrival claims. Specifically, we study and model the sequences and pacing of the Egyptian insurance company's claims time inter-arrival. The availability of these models offers a basis for allegations related to fire claims. Due to constraints on real world data collection techniques, previous ruin probability models did not provide adequate details on the complex properties of inter-arrival claims. Usually, they believed that inter-arrival claims could be based on Poisson processes and that inter-arrival time, or time interval between two successive claims, follows an exponential distribution. This statement indicates that claims take place at a constant rate. Nevertheless, this model does not capture the variations that occur in the arrival rate of the operation. Recently, researchers proposed using heavy tailed distributions to explain the many dynamics our approach in this paper is to create a general model of the arrival process involving the collection of real-time data in daily environments based on Egyptian fire insurance companies. To investigate that behavior, we use a case study with 10 years of data from one Egyptian insurance company. This behavior-driven by the company shows that Claims inter-arrival time routines can be modeled by non-Poisson processes. The time of inter-arrival operation is accompanied by a heavy tailed distribution, precisely the Pareto distribution. Our analysis offers evidence to support an inter-arrival claims hypothesis, the Pareto model and its properties, such as the 80/20 law, may be useful for the analysis of inter-arrival claims. The results of this study will provide the ability to simplify the treatment and design of claims behavioral interventions.

1.1. Problem Statement

After a detailed study of research papers, articles and books related to reliability and other statistical analysis it has been found that in maximum of researches show that Current ruin probability models, assuming that inter-arrival time of

claims, is distributed randomly and thus well approximated by Poisson processes. Here we provide clear proof that the timing of claims fits by non-Poisson patterns, our analysis shows that claims activities can be represented by non-Poisson processes and that the subsequent distribution of inter-arrival activity times follows the Pareto, distribution. These results will help researchers understand daily behavioral trends and create more sophisticated predictive models of claims and their timing.

1.2. Objectives of Study

The main objective is to model inter-arrival time of claims in Insurance Company.

1.3. Justification of the Study

These results will help researchers understand daily behavioral trends and create more sophisticated predictive models of claims.

By estimating commercial fire loss insurance risk on business-line and event-type levels, we are able to present the estimates in a more balanced fashion and the results may help non-life insurance companies to manage their risk.

1.4. Research Structure

- Test if inter-arrival time of claims be heavy tailed and follow pareto distribution.
- I used “easy-fit” software for:
 - 1) Exploratory data analysis;
 - 2) Goodness of fit tests included KS-test, AD test, chi-squared test.
- I find that show that the pareto distribution is the best one among 56 continuous distribution according to KS test, and also chi-squared test.

The remainder of the paper is organized as follows, in Section 2, the Poisson process for inter-arrival claims, which predicts an exponential distribution of interevent times. In Section 3, related works. In Section 4, pareto distribution, in Section 5, we present proof that the power law tail characterizes the interevent time probability density function of claims. In Section 6, the conclusion.

2. Poisson Process

Inter-arrival and waiting time distributions

Let $\{N(t) : t \geq 0\}$ be a Poisson process with arrival rate $\lambda > 0$. Set $T_0 \equiv 0$. For $n = 1, 2, \dots$ define $T_n = \inf\{t \geq 0 : N(t) = n\}$ = time of arrival of n -th claim (or waiting time until the n -th claim arrival). Put $A_n = T_n - T_{n-1}, n = 1, 2, \dots$ so that A_n time between $(n - 1)$ -th and n -th claim arrivals. Recall from our initial comments that we had in fact defined the process, see [Rolski et al. \(1999\)](#).

$\{N(t)\}$ starting from $\{T_i\}$. The random variables T_0, T_1, T_2, \dots are called claim arrival times (or waiting times); the sequence $\{A_n : n = 1, 2, \dots\}$ is called the sequence of inter-arrival times. See [Bingham et al. \(1987\)](#).

For any $s > 0$ note that $\{T_1 > s\} = \{N(s) = 0\}$; hence

$$P(A_1 > s) = P(T_1 > s) = P(N(s) = 0) = \exp(-\lambda s). \quad (1)$$

So $P(A_1 \leq s) = 1 - e^{-\lambda s}$, $s \geq 0$. Therefore the random variable A_1 has an EXP (λ) distribution (= exponential distribution with parameter $\lambda > 0$); that is,

$$P(A_1 \in (a, b)) = \int_a^b \lambda e^{-\lambda s} ds, \quad 0 \leq a < b < \infty \quad (2)$$

Next let us consider the joint distribution of (T_1, T_2) let $F_{(T_1, T_2)}$ denote the joint distribution function of (T_1, T_2) ; that is, $F_{(T_1, T_2)}(t_1, t_2) = P(T_1 \leq t_1, T_2 \leq t_2)$. as $0 \leq T_1 \leq T_2$ it is enough to look at $F_{(T_1, T_2)}(t_1, t_2)$ for $0 \leq t_1 \leq t_2$. It is clear that for $0 \leq t_1 \leq t_2$,

$$\begin{aligned} \{T_1 \leq t_1, T_2 \leq t_2\} &= \{N(t_1) \geq 1, N(t_2) \geq 2\} \\ &= \{N(t_1) = 1, N(t_2) - N(t_1) \geq 1\} \cup \{N(t_1) \geq 2\} \end{aligned} \quad (3)$$

where the r.h.s. is a disjoint union.

$$\begin{aligned} F_{(T_1, T_2)}(t_1, t_2) &= P(N(t_1) = 1, N(t_2) - N(t_1) \geq 1) + P(N(t_1) \geq 2) \\ &= \lambda t_1 e^{-\lambda t_1} (1 - e^{-\lambda(t_2 - t_1)}) + [1 - (e^{-\lambda t_1} + \lambda t_1 e^{-\lambda t_1})] \\ &= -\lambda t_1 e^{-\lambda t_2} + H(t_1) \end{aligned} \quad (4)$$

where H is a function depending only on t_1 : Consequently the joint probability density function $f_{(T_1, T_2)}$ of (T_1, T_2) is given by

$$\begin{aligned} f_{(T_1, T_2)}(t_1, t_2) &\triangleq \frac{\partial^2}{\partial t_2 \partial t_1} F_{(T_1, T_2)}(t_1, t_2) \\ &= \begin{cases} \lambda^2 e^{-\lambda t_2}, & \text{if } 0 < t_1 < t_2 < \infty \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

To find the joint distribution of (A_1, A_2) from the above, note that

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 - T_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} \quad (6)$$

The linear transformation given by the (2×2) matrix in (13) has determinant 1, and transforms the region $\{(t_1, t_2) : 0 < t_1 < t_2 < \infty\}$ in 1-1 fashion onto $\{(a_1, a_2) : a_1 > 0, a_2 > 0\}$. So the joint probability density function $f_{(A_1, A_2)}$ of (A_1, A_2) is given by

$$\begin{aligned} f_{(A_1, A_2)}(a_1, a_2) &= F_{(T_1, T_2)}(a_1, a_1 + a_2) \\ &= \begin{cases} (\lambda e^{-\lambda a_1})(\lambda e^{-\lambda a_2}), & \text{if } a_1 > 0, a_2 > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (7)$$

Thus A_1, A_2 are independent random variables each having an exponential distribution with parameter λ . See Billingsley (1968).

With more effort, one can prove

Theorem 1 Let $\{N(t) : t \geq 0\}$ be a time homogeneous Poisson process with arrival rate $\lambda > 0$. Let A_1, A_2, \dots denote the inter-arrival times. Then $\{A_n : n = 1, 2, \dots\}$ is a sequence of independent, identically distributed random

variables (or in other words an i.i.d. sequence) having Exp (λ) distribution. See [Feller \(1969\)](#).

In view of the argument above for the case $n=2$, the general idea of the proof is clear. One proves rst that the joint distribution function of T_1, T_2, \dots, T_n is given by

$$F_{(T_1, T_2, \dots, T_n)}(t_1, t_2, \dots, t_n) = -\lambda^{n-1} \left(\prod_{i=1}^{n-1} t_i \right) e^{-\lambda t_n} + H(t) \tag{8}$$

If $0 \leq t_1 < t_2 < \dots < t_n < \infty$, where $H(\cdot)$ is a function such that $\partial^n H / (\partial t_1 \partial t_2 \dots \partial t_n) = 0$. In fact $H(\cdot)$ is a sum of a finite number of terms; each term is a product of powers of t_i and $e^{-\lambda t_j}$ with at least one $t_k, k \geq 2$ missing! Establishing this is the tedious part of the proof. Once this is done the joint probability density function of T_1, T_2, \dots, T_n is given by

$$F_{(T_1, T_2, \dots, T_n)}(t_1, t_2, \dots, t_n) = \begin{cases} \lambda^n \exp(-\lambda t_n), & \text{if } 0 < t_1 < t_2 < \dots < t_n < \infty \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Note that the analogue is see [Delampady et al. \(2001\)](#)

$$\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 - T_1 \\ T_3 - T_2 \\ \vdots \\ T_n - T_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ \vdots \\ T_n \end{pmatrix} \tag{10}$$

One can now proceed exactly as in the earlier case to obtain the theorem. The reader is invited to work out the details at least when $n = 3, 4$.

Note: As A_1 has Exp (λ) distribution, its expectation is given by $E(A_1) = \frac{1}{\lambda}$; so $\frac{1}{\lambda}$ is the mean arrival time. Thus the arrival rate being λ is consistent with this conclusion. See [Bingham et al. \(1987\)](#).

Note: It is an easy corollary of the theorem that $T_n = A_1 + A_2 + \dots + A_n$ has the gamma distribution $\Gamma(n, \lambda)$.

Remark 4: One can also go in the other direction. That is, let $0 = T_0 \leq T_1 \leq T_2 \leq \dots$ be the claim arrival times; let $A_n = T_n - T_{n-1}, n \geq 1$. Suppose $\{A_n\}$ is an i.i.d. sequence having EXP (λ) distribution. Define $\{N(t)\}$ by (1). See [Ethier & Kurtz \(1986\)](#), Then the stochastic process $\{N(t) : t \geq 0\}$ can be shown to be time homogeneous Poisson process with rate λ . In the jargon of the theory of stochastic processes, Poisson process is the renewal process with i.i.d. exponential arrival rates.

3. Related Work

Maturing pervasive computing technologies have sparked a new wave of human behavior analysis and resulted in new theories regarding human behavior patterns. Barabasi’s study of the timing of consecutive electronic and physical mail messages sparked a model of human dynamics as a heavy-tailed distribution see

Oliveira & Barabási (2005) and Bees et al. (2005). A queuing model and heavy-tailed distribution were introduced in Barabasi's study to explain the large time gap between sent messages after a burst of responses.

After Barabasi's discovery, scientists use heavy tailed distributions to explain human behavior in diverse domains, ranging from social science to health care, see Andriani & McKelvey (2009). In the social network field, heavy-tailed distributions are used to characterize the dynamics of popularity based on diverse digital platforms, such as Wikipedia, blog posts, Android applications, Web pages, and Twitter see Leskovec et al. (2007) and Yu et al. (2017). As an example, Li et al. (2015) show that the behavior-based popularity of Android applications follows the Pareto principle. Tsompanidis et al. (2014) also discover that web traffic flow size can be explained by the Pareto distribution. Similarly, researchers presented a list of social and organizational power laws, one kind of heavy-tailed distribution, to describe human behavior see Scholz (2015) and Andriani & McKelvey (2009). Specifically, the power law distribution identifies the number of inter-firm relationships observed from linkages between firms: suppliers, customers, and owners see Dewes et al. (2003) and Saito et al. (2007).

Further, scientists use heavy-tailed distributions to model and predict human mobility see Mainardi et al. (2000) and Gallotti et al. (2016). For example, GPS-based human movement patterns can be captured by heavy-tailed flights for different transportation modes, including walking/running and car/taxi see Hong (2010) Regardless of transportation modes, the distribution of user's moving distances, from visited locations to the target location, can be modelled by the Pareto distribution see Zhu et al. (2015).

Evidence that non-Poisson activity patterns characterize human activity has first emerged in computer communications, where the timing of many human driven events is automatically recorded, see Gonzalez et al. (2008). For example, measurements identifying the distribution of the time differences between consecutive instant messages sent by individuals during online chats see Dewes et al. (2003) have found evidence of heavy tailed statistics. Professional tasks, such as the timing of job submissions on a supercomputer, directory listings and file transfers [FTP requests] initiated by individual users see Mainardi et al. (2000) were also reported to display non-Poisson features. Similar patterns emerge in economic transactions see Reberto et al., in the number of hourly trades in a given security see Plerou et al. (2000) or the time interval distribution between individual trades in currency futures see Masoliver et al. (2003). Finally, heavy tailed distributions characterize entertainment related events, such as the time intervals between consecutive online games played by users see Henderson & Henderson (2001). Note, however, that while these datasets provide clear evidence for non-Poisson human activity patterns, most of them do not resolve individual human behavior, but capture only the aggregated behavior of a large number of users. For example, the dataset recording the timing of the job sub-

missions looks at the timing of all jobs submitted to a computer, by any user. Thus for these measurements the interevent time does not characterize a single user but rather a population of users. Given the extensive evidence that the activity distribution of the individuals in a population is heavy tailed, these measurements have difficulty capturing the origin of the observed heavy tailed patterns. For example, while most people send only a few emails per day, a few send a very large number on a daily basis see [Eckmann et al. \(2004\)](#) and [Ebel et al. \(2002\)](#).

4. Pareto Distribution

The Pareto distribution is the classic heavy-tailed distribution. In comparison with the exponential, it has a much higher probability of generating extreme values. This means that jobs with very long service times account for a significant fraction of the queue's total work. The Pareto distribution is often associated with the famous 80 - 20 rule, which holds that 80% of outputs are attributable to only 20% of inputs in applications with heavy-tailed behavior. For example, it's been observed that 20% of a population tends to hold about 80% of total wealth, or that 80% of business sales revenue tends to come from only 20% of customers. An extension of this rule holds that the top 1% of inputs account for 50% of outputs. If a system's jobs are Pareto distributed, then half of the total system running time will be dedicated to serving only 1% of jobs! It's important to remember that the numbers 80 and 20 are not magical. The actual values will vary for different applications. They don't even need to sum to one, since they're measures of two different quantities. The significant part of the "law of the vital few," as it's sometimes called, is the relative importance of a surprisingly small portion of the population, see ([Amoroso, 1938](#) & [Pareto, 1898](#)).

A power-law probability distribution that is used in description of social, scientific, geophysical, actuarial, and many other types of observable phenomena. Originally applied to describing the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population see [Amoroso \(1938\)](#), the Pareto distribution has colloquially become known and referred to as the Pareto principle, or "80 - 20 rule", and is sometimes called the "Matthew principle". This rule states that, for example, 80% of the wealth of a society is held by 20% of its population. However, one should not conflate the Pareto distribution with the Pareto Principle as the former only produces this result for a particular power value, ($\alpha = \log 45 \approx 1.16$). While is variable, empirical observation has found the 80 - 20 distribution to fit a wide range of cases, including natural see [Van Montfort \(1986\)](#) phenomena and human activities. See [Oancea \(2017\)](#).

If X is a random variable with a Pareto (Type I) distribution see [Arnold \(1983\)](#), then the probability that X is greater than some number x , i.e. the survival function (also called tail function), is given by

$$\bar{F}(x) = pr(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m \\ 1 & x < x_m \end{cases}$$

where x_m is the (necessarily positive) minimum possible value of X , and α is a positive parameter. The Pareto Type I distribution is characterized by a scale parameter x_m and a shape parameter α , which is known as the *tail index*. When this distribution is used to model the distribution of wealth, then the parameter α is called the Pareto index.

Cumulative distribution function

From the definition, the cumulative distribution function of a Pareto random variable with parameters α and x_m is

$$FX(x) = \begin{cases} 1 - \left(\frac{x_m}{x}\right)^\alpha & x \geq x_m \\ 0 & x < x_m \end{cases}$$

Probability density function

It follows (by differentiation) that the probability density function is

$$FX(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \\ 0 & x < x_m \end{cases}$$

When plotted on linear axes, the distribution assumes the familiar J-shaped curve which approaches each of the orthogonal axes asymptotically. All segments of the curve are self-similar (subject to appropriate scaling factors). When plotted in a log-log plot, the distribution is represented by a straight line.

Properties

Moments and characteristic function

The expected value of a random variable following a Pareto distribution is

$$E(x) = \begin{cases} \infty & \alpha \leq 1 \\ \frac{\alpha x_m^\alpha}{\alpha - 1} & \alpha > 1 \end{cases}$$

The variance of a random variable following a Pareto distribution is

$$Var(x) = \begin{cases} \infty & \alpha \in (1, 2] \\ \left(\frac{\alpha x_m^\alpha}{\alpha - 1}\right)^2 \frac{\alpha}{\alpha - 2} & \alpha > 2 \end{cases}$$

(If $\alpha \leq 1$, the variance does not exist).

Figure 1, Figure 2 show pdf and cdf of pareto distribution for various α (1, 2, 3, ∞).

5. Methodology

This section presents the procedure which was used in the study. It explains in

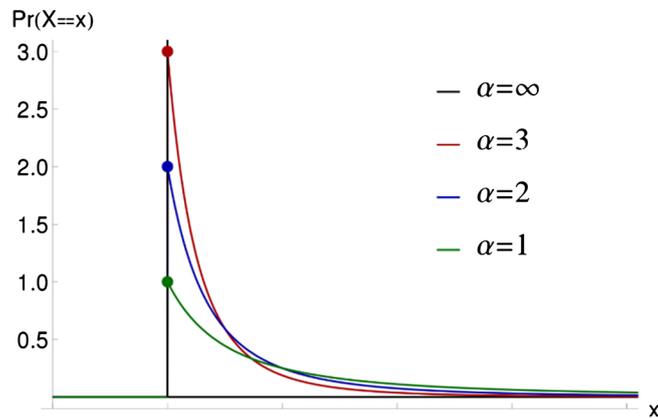


Figure 1. Pareto Type I probability density functions for various α with $x_m=1$. As $\alpha \rightarrow \infty$ the distribution approaches $\delta(x-x_m)$ where δ is the Dirac delta function.

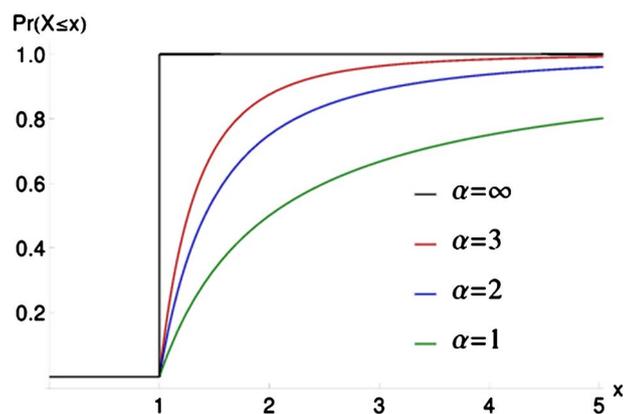


Figure 2. Pareto Type I cumulative distribution functions for various α with $x_m=1$.

detail the steps that were encountered in the modeling process which includes the data processing and analysis. There are 939 observations in the data set. All commercial fire insurance loss data sets used in this study were obtained from a non-life insurance company in Egypt.

5.1. Scope of the Data

Secondary data from E.G. insurance company regarding fire industrial claims for the period 2000-2011 was used in this study.

5.2. Actuarial Modeling Process

This section will describe the steps that were followed in fitting a statistical distribution to the extreme claim severity. These steps include

- 1) Exploratory data analysis.
- 2) Goodness of fit test.

5.3. Exploratory Data Analysis

It was necessary to do some descriptive analysis of the data to obtain the salient features. This involves the Mean, Median, Mode, Standard Deviation, Skewness

and Kurtosis. This was done using easy fit programming language and also manual calculation.

5.4. Specific Objectives

Testing for the appropriate statistical distribution for the claim inter-arrival time.

Test the goodness of fit of the chosen distribution.

5.5. Variable

The random variables used in the study were the fire claim inter-arrival time reported and claimed at EG Insurance.

5.6. Descriptive Data Analysis

Table 1 represents the summary statistics of fire insurance claims in years (2000-2011). The mean of claims is 9.539, the Variance 409.28 and the Range 266.63.

Data Fitting process involves the use of certain statistical techniques which enable us to estimate fitness parameters according to the data sample. One benefit of using software to fit the data and interpret probability data is that they can automatically fit data simultaneously with a number of known distribution patterns. Easy Fit is a data analyzer and simulation program that helps us to fit probabilistic distributions to define data samples, to simulate them, to pick the best fit sample and to apply the analytical results to make better decisions. Goodness of fit Test is a technique used to determine the appropriate distribution to be fitted for the given data. The theoretical history of this test is clarified initially and then the whole test is applied to live data collected from Egyptian insurance company. The traditional assessment of fitness test goodness in statistics is interested in testing precision for the sample produced from the supposed PDF. Moreover, it is also worth emphasizing the opportunity to reject the hypotheses when the supposed PDF is different from actual PDF. Furthermore, the opportunity to reject the hypotheses is also worth highlighting when the supposed

Table 1. Descriptive analysis of data set.

Statistic	Value
Sample Size	939
Range	266.63
Mean	9.539
Variance	409.28
Std. Deviation	20.231
Coef. of Variation	2.1208
Std. Error	0.6602
Skewness	7.7114
Excess Kurtosis	72.484

PDF is different from the actual PDF. This study applies three methods of fitness testing which include tests for Chi-Square (C-S), Kolmogorov-Smirnov (K-S), and Anderson-Darling (A-D).

The results indicate that the distribution of pareto is one of the best distributions for the inter-arrival time claims.

5.7. Goodness-of-Fit Tests

As their very name implies, can be used to assess whether or not a particular distribution is properly fit to the data. The measurement of goodness-of-fit statistics also helps to rank the fitted distributions over the raw data according to fit consistency. This particular function of the app is very useful when comparing fitted models. The most widely used tests for goodness-of-fit include Kolmogorov-Smirnov, Anderson-Darling and Chi-squared tests. For these tests, the principle of use is identical. They differ in functional method (and use type) however. Could be called the Kolmogorov-Smirnov test as the most used Goodness-of-Fit test.

5.8. Easy Fit Software

Easy Fit is a data analysis and simulation software which enables us to fit and simulate statistical distributions with sample data, choose the best model, and use the obtained result of analysis to take better decisions. This software can function as a stand-alone windows application or as an add-on for Excel spread sheet.

Prominent features of this program are:

- Supports more than 50 discrete and continuous distributions.
- Automatic and manual settings.
- Ability to test performed operations.

Easy Fit supports all the commonly used continuous distributions. Some of them have alternative names (indicated in parentheses): 1. Beta 2. Burr (Burr Type 12, or Singh-Maddala), 3. Burr [4P], 4. Cauchy (Lorentz), 5. Chi-Squared, 6. Chi-Squared [2P], 7. Dagum (Burr Type 3, or Inverse Burr), 8. Dagum [4P], 9. Erlang, 10. Error (Exponential Power, or Generalized Error), 11. Error Function, 12. Exponential, 13. Exponential [2P], 14. F Distribution, 15. Fatigue Life (Birnbau-Saunders), 16. Fatigue Life [3P], 17. Frechet (Maximum Extreme Value Type 2), 18. Frechet [3P], 19. Gamma, 20. Gamma [3P], 21. Generalized Gamma, 22. Generalized Gamma [4P], 23. Gen.extreme value, 24. Gumbel Max (Maximum Extreme Value Type 1), 25. Gumbel Min (Minimum Extreme Value Type 1), 26. genpareto, 27. Hyperbolic Secant, 28. Inverse Gaussian, 29. Inverse Gaussian [3P], 30. Johnson SB, 31. Johnson SU, 32. Kumaraswamy, 33. Levy, 34. Laplace (Double Exponential), 35. Logistic, 36. Log-Gamma, 37. Log-Logistic (Fisk), 38. Lognormal, 39. Nakagami (Nakagami-m), 40. Normal (Gaussian), 41. Pareto—first kind, 42. Pareto—second kind (Lomax), 43. Pearson Type 5 (Inverse Gamma), 44. Pearson Type 6 (Beta dist. of the second kind), 45. Pearson

Type 6 [4] 46. Pert, 47. Power Function, 48. Rayleigh, 49. Rayleigh [2P], 50. Reciprocal, 51. Rice (Ricean, or Nakagami-n), 52. Student's t, 53. Triangular, 54. Uniform, 55. Weibull, 56. Weibull [3P].

In EasyFit, you can use almost all the Goodness-Of-Fit tests including Kolmogorov-Smirnov, Anderson-Darling, and Chi-square tests. When the distributions are fitted, EasyFit will generate a report of goodness-of-fit values which includes calculated test statistics and critical values for various significance levels. We will compare the process of fitting for several kinds of distribution. Since Goodness-Of-Fit statistics are in form of distance between data and fitted distributions, clearly the distribution with minimum statistics value has been best fitted with data. Based on this fact, EasyFit will attribute a ranking number to each distribution (1—the best model, 2—best model after the first one ... etc.). This allows you to select the most reliable model easily.

5.9. Methods

The goodness of fit of the statistical model explains how well the set of observations fits. Fit goodness measures typically sum up the discrepancy between observed values and expected values under the model in question. These tests can be used in statistical hypothesis testing, to test whether two samples are taken from the same distributions, or whether the resulting frequencies fit a particular distribution. There are various approaches used for the fitness test. The most significant of them are: • Kolmogorov-Smirnov • Anderson-Darling • Chi-Squared (method 2) And The methods used in this paper include goodness of fit tests, C-S, K-S, A-D tests, probability distribution function (PDF) parameter estimation.

5.10. Problem Identification

After a detail study of research papers, articles and books related to reliability and other statistical analysis it has been found that in maximum of researches show that Current ruin probability models, assuming that inter-arrival time of claims, is distributed randomly and thus well approximated by Poisson processes. Here we provide clear proof that the timing of claims fits by non-Poisson patterns, our analysis shows that claims activities can be represented by non-Poisson processes and that the subsequent distribution of inter-arrival activity times follows the Pareto, distribution. These results will help researchers understand daily behavioral trends and create more sophisticated predictive models of claims and their timing.

5.11. Summary of Goodness-of-Fit

Table 2 shows that the pareto distribution is the best one among 56 continuous distribution according to KS test, and also chi-squared test show pareto distribution is the best, while the Anderson test indicate pareto ranked 7 between 56 continuous distribution.

Table 2. Goodness-of-fit tests.

Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
	Statistic	Rank	Statistic	Rank	Statistic	Rank
Pareto	0.01816	1	2.2587	7	3.2094	1
Pearson 6 (4P)	0.01877	2	4.3282	11	N/A	
Dagum (4P)	0.02032	3	0.35633	1	5.4068	4
Log-Logistic (3P)	0.02146	4	0.58452	2	6.9813	6
Burr (4P)	0.02223	5	4.2821	10	N/A	
Lognormal (3P)	0.02975	6	0.7423	3	3.7114	2
Gen. Pareto	0.03098	7	1.6016	4	4.5452	3
Inv. Gaussian (3P)	0.04038	8	2.1557	6	17.966	8
Log-Pearson 3	0.04065	9	2.1431	5	6.0442	5
Pearson 5 (3P)	0.04121	10	2.9102	9	26.673	10
Frechet (3P)	0.0415	11	2.781	8	21.23	9
Gen. Extreme Value	0.07233	12	6.0418	12	10.895	7
Pearson 5	0.10769	13	23.488	18	66.948	15
Pearson 6	0.10943	14	22.838	17	64.565	13
Fatigue Life (3P)	0.11065	15	15.325	13	62.524	12
Levy (2P)	0.11716	16	19.679	15	132.06	18
Weibull (3P)	0.12044	17	25.749	19	135.87	19
Frechet	0.13169	18	17.489	14	49.008	11
Gamma (3P)	0.14025	19	40.773	21	N/A	
Log-Gamma	0.14105	20	21.828	16	65.85	14
Lognormal	0.15359	21	41.181	22	115.63	17
Burr	0.15361	22	382.41	49	N/A	
Log-Logistic	0.15647	23	38.795	20	103.63	16
Chi-Squared (2P)	0.19292	24	108.79	29	198.12	20
Gen. Gamma (4P)	0.19467	25	58.949	23	347.85	26
Weibull	0.1948	26	99.293	27	230.06	22
Fatigue Life	0.2034	27	83.944	24	203.71	21
Cauchy	0.24559	28	100.08	28	261.54	23
Exponential	0.24673	29	90.209	26	286.56	25
Rayleigh (2P)	0.26225	30	204.45	40	N/A	
Exponential (2P)	0.26411	31	169.35	31	415.52	29
Kumaraswamy	0.27082	32	127.53	30	N/A	
Error	0.3109	33	184.03	34	394.5	27
Laplace	0.3109	34	184.03	33	394.5	28
Pareto 2	0.31804	35	88.52	25	276.02	24

Continued

Gen. Gamma	0.32178	36	180.03	32	546.01	35
Hypersecant	0.33846	37	190.47	35	422.62	31
Levy	0.3468	38	207.91	41	612.33	36
Logistic	0.35139	39	196.09	38	439.44	32
Power Function	0.35715	40	195.93	37	N/A	
Normal	0.36771	41	216.29	42	448.92	33
Rayleigh	0.37008	42	330.94	48	825.79	38
Chi-Squared	0.38529	43	393.68	50	1153.8	39
Gumbel Min	0.39505	44	287.7	44	N/A	
Uniform	0.40245	45	299.7	45	N/A	
Gumbel Max	0.42061	46	191.45	36	416.61	30
Inv. Gaussian	0.46214	47	197.65	39	687.46	37
Beta	0.48425	48	323.76	47	N/A	
Reciprocal	0.54998	49	709.4	51	1724.3	41
Error Function	0.55314	50	284.05	43	478.6	34
Gamma	0.58597	51	322.24	46	1206.6	40
Rice	0.64241	52	1188.2	52	3728.7	42
Pert	0.70911	53	1484.7	53	4591.7	43
Triangular	0.81258	54	2264.7	55	6786.5	44
Dagum	0.89083	55	2060.6	54	12042.0	45
Student's t	0.94302	56	2426.9	56	17027.0	46

I will provide four classical goodness-of-fit plots for pareto distribution presented on:

- p-p Graph;
- Q-Q Graph;
- Probability Difference (PD) graph;
- Cumulated Distribution Function (CDF) graph.

Figure 3, Figure 4 represent the PP and QQ diagnostic plots for the fitted pareto. Since out of the diagnostic plots the probability plot and quantile plot are approximately linear and the straight line fits most of the data points, it is safe to conclude that the pareto fits the inter-arrival time of insurance claims data points and the model we chose is valid.

Figure 5 shows Probability Difference Graph, The probability difference graph is a plot of the difference between the empirical CDF and the theoretical CDF, This graph can be used to determine how well the theoretical distribution fits to the observed data, and **Figure 5** shows a fairly linear relationship between the sample, and the theoretical quantiles.

Figure 6 shows Cumulative Distribution Function (CDF) graph for pareto distribution and is displayed as a stepped discontinuous line depending on the

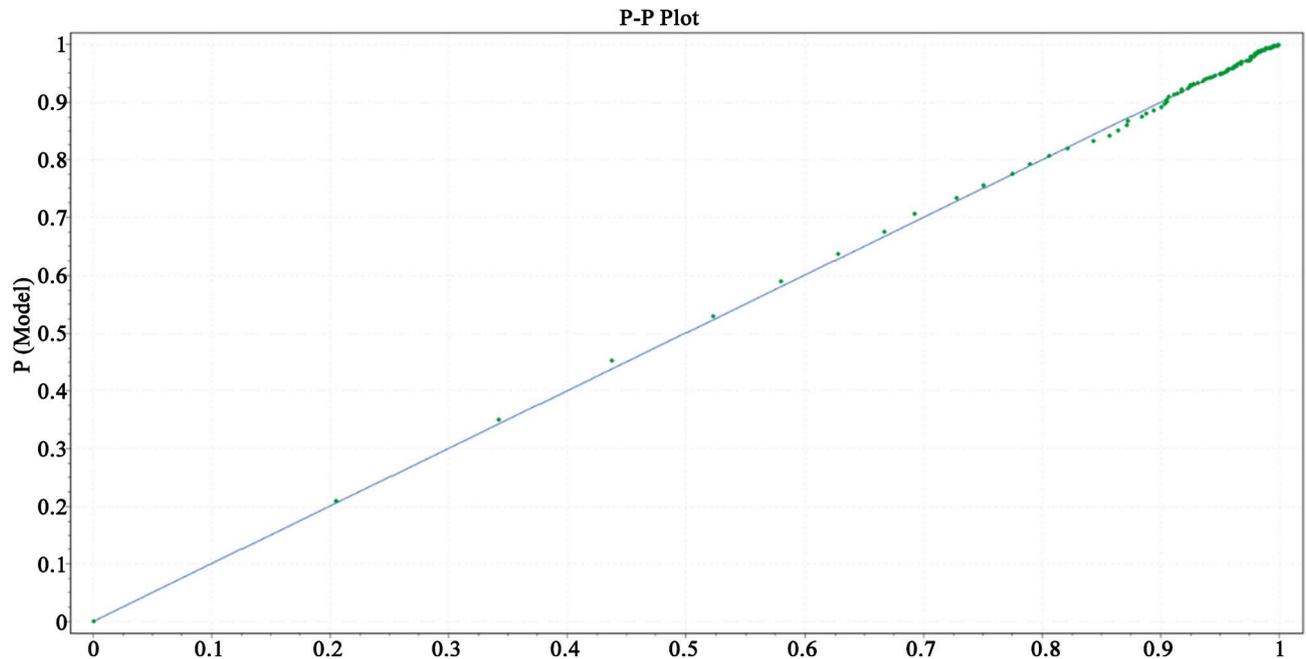


Figure 3. QQ plot of pareto distribution.

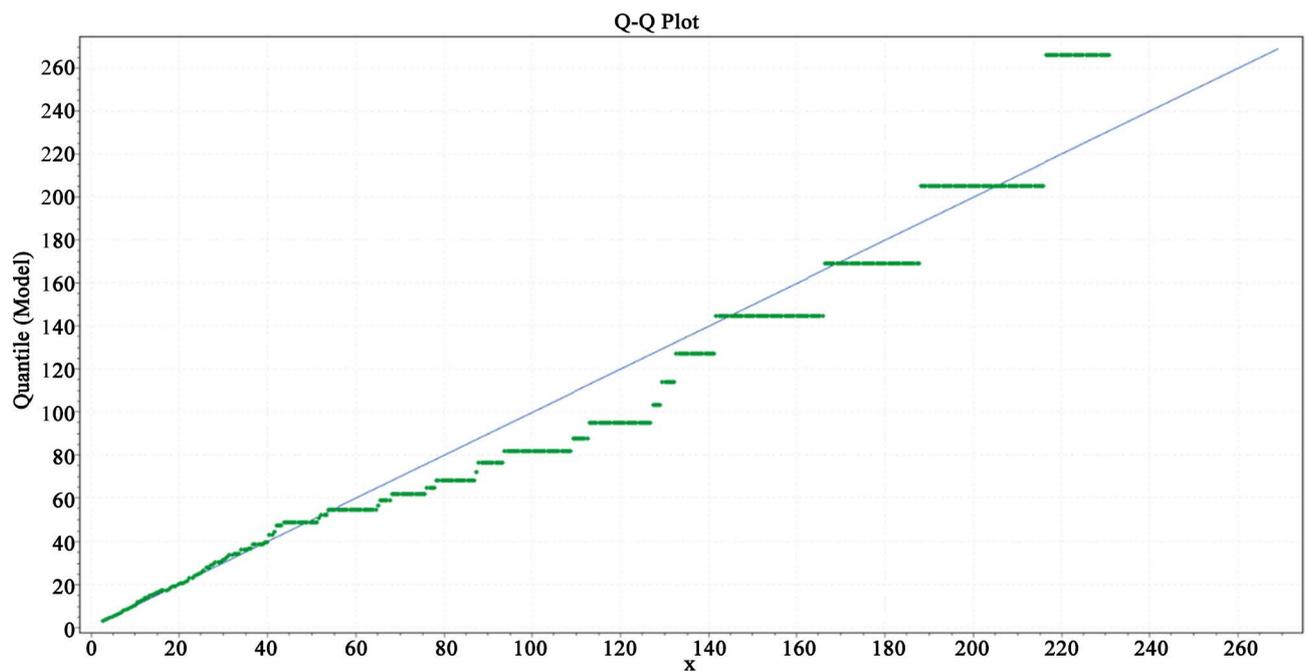


Figure 4. PP plot of pareto distribution.

number of bins.

5.12. Hypothesis Testing

5.12.1. KS Test

The null and the alternative hypotheses are:

- H0: the data follow the pareto distribution.
- HA: the data do not follow the pareto distribution.

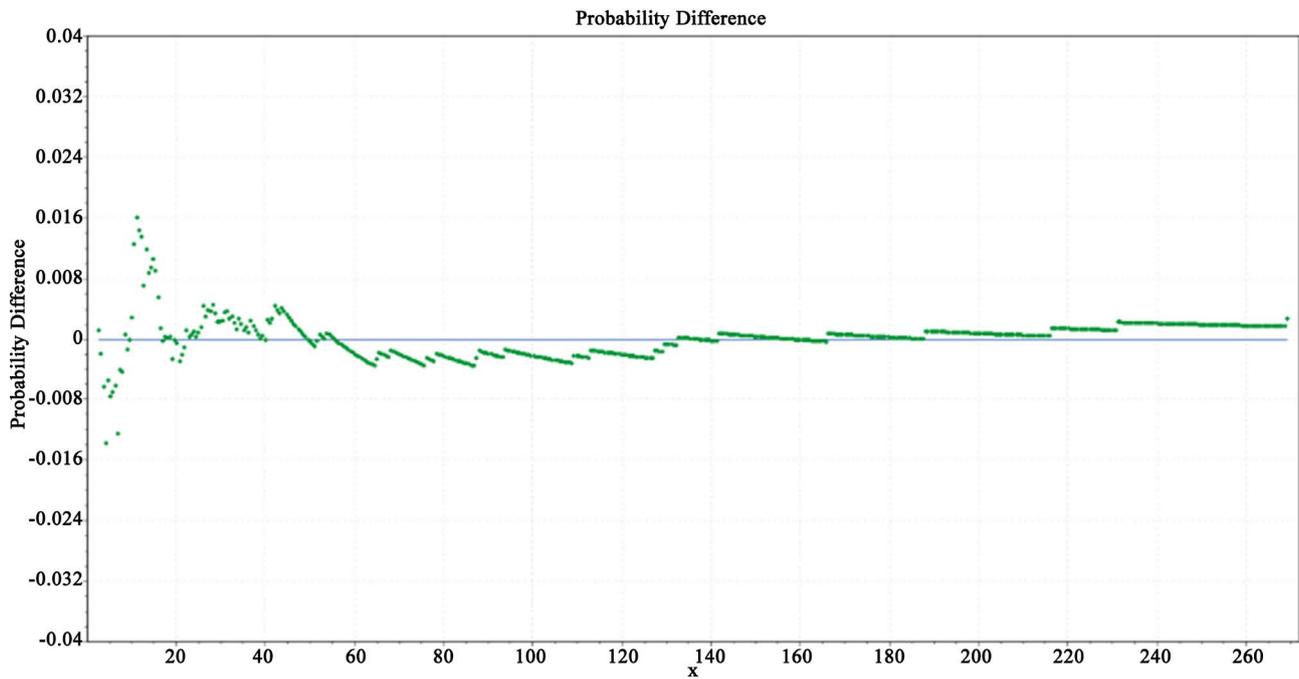


Figure 5. Probability Difference (PD) graph of pareto.

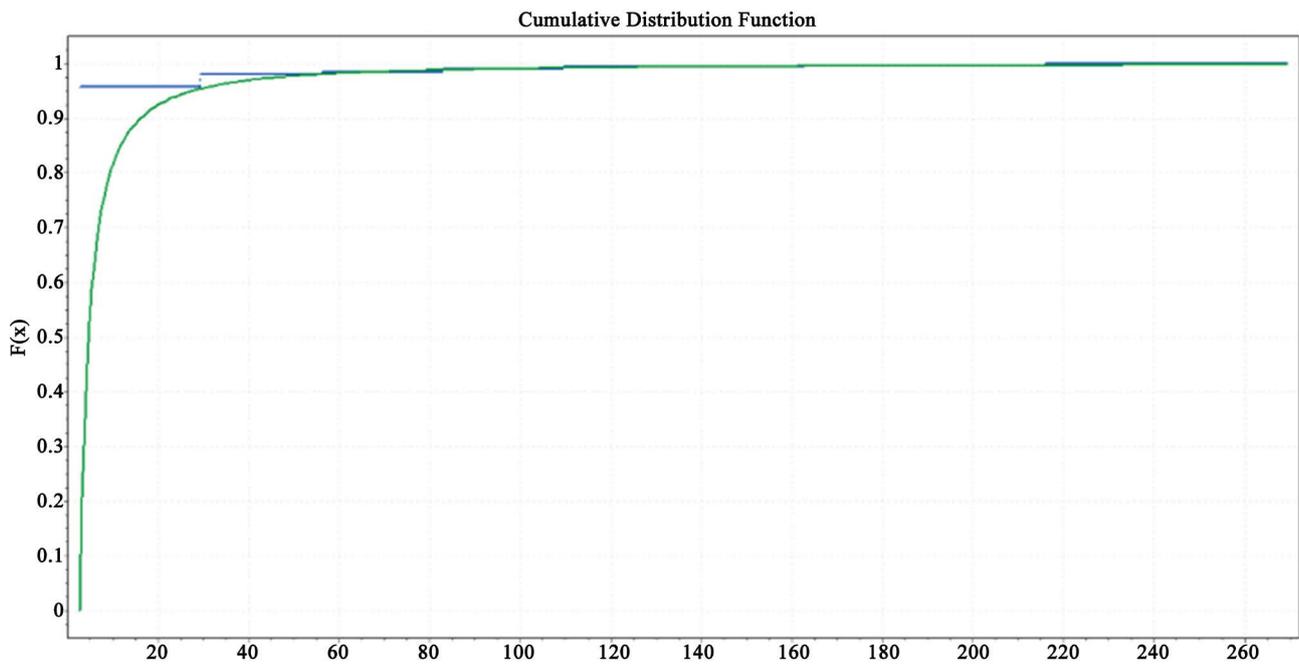


Figure 6. Cumulated Distribution Function (CDF) graph.

Table 3 shows that pareto distribution is accepted by KS Test having P value 0.91054 at all level of significance.

5.12.2. AD Test

The null and the alternative hypotheses are:

- H_0 : the data follow the pareto distribution.
- H_A : the data do not follow the pareto distribution.

Table 4 shows that pareto distribution is accepted by AD Test at 1% and 2% and 5% level of significance, But rejected at 10% and 20% level of significance.

5.12.3. Chi Test

The null and the alternative hypotheses are:

- H0: the data follow the pareto distribution.
- HA: the data do not follow the pareto distribution.

Table 5 shows that Pareto distribution is accepted by chi-squared Test having P -value 0.95541 at all level of significance.

6. Conclusion

In many applications of claim inter-arrival time data distributions, a key concern

Table 3. KS test for pareto distribution.

Kolmogorov-Smirnov					
Sample Size	939				
Statistic	0.01816				
P -Value	0.91054				
Rank	1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.03502	0.03991	0.04432	0.04954	0.05316
Reject?	No	No	No	No	No

Table 4. AD test for pareto distribution.

Anderson-Darling					
Sample Size	939				
Statistic	2.2587				
Rank	7				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	Yes	Yes	No	No	No

Table 5. Chi-squared Test for pareto distribution.

Chi-Squared					
Deg. of freedom	9				
Statistic	3.2094				
P -Value	0.95541				
Rank	1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	12.242	14.684	16.919	19.679	21.666
Reject?	No	No	No	No	No

is fitting the claim inter-arrival time in the tail. As mentioned above, good estimates of the tails of fire claim inter-arrival time distributions are essential for pricing and risk management of commercial fire insurance loss. We execute an exploratory claim inter-arrival time analysis using a goodness of fit. The goodness of fit revealed the some distributions to be poorly fitted, while pareto distributions can be seen to fit the claim inter-arrival time data much better.

The Q-Q plots indicate that most points of the Pareto distribution are lying along the reference line thus making it one of the best distributions for claim inter-arrival time. A histogram of claims and goodness of fit with Probability Density Function (PDF) graph, Cumulated Distribution Function (CDF) graph, p-p Graph, Probability Difference (PD) graph and also pointed that Pareto distribution was one of the best fitting distribution among the 56 distributions.

Preceding page and the goodness-of-fit shows that after analyzing the results of **Table 2** it has been found that pareto distribution is the best distributions among 56 probability distributions. This distribution is accepted by KS Test having P -value 0.91054 at all level of significance, is accepted by AD Test at 1% and 2% and 5% level of significance, but rejected at 10% and 20% level of significance and accepted by chi-squared Test having P -value 0.95541 at all level of significance.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Amoroso, L. (1938). Vilfredo Pareto. *Econometrica*, 6, 1-21. <https://doi.org/10.2307/1910081>
- Anderson, H. R. (2003). *Fixed Broadband Wireless System Design*. New York: Wiley. <https://doi.org/10.1002/0470861290>
- Andriani, P., & McKelvey, B. (2009). Perspective—From Gaussian to Paretian Thinking: Causes and Implications of Power Laws in Organizations. *Organization Science*, 20, 1053-1071. <https://doi.org/10.1287/orsc.1090.0481>
- Arnold, B. C. (1983). *Pareto Distributions*. Fairland, MD: International Cooperative Publishing House.
- Bees, A., York, N., & Barabasi, A. (2005). The Origin of Bursts and Heavy Tails in Human Dynamics. *Nature*, 435, 207-211. <https://doi.org/10.1038/nature03459>
- Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.
- Bingham, N. H., Goldie, C. M., & Teugels, J. L. (1987). *Regular Variation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511721434>
- Delampady, M., Krishnan, T., & Ramasubramanian, S. (2001). *Probability and Statistics*. A Volume in “Echoes from Resonance”, Hyderabad: Universities Press.
- Dewes, C., Wichmann, A., & Feldman, A. (2003). *Proceedings of the 2003 ACM SIGCOMM Conference on Internet Measurement (IMC-03)*. New York: ACM.
- Ebel, H., Mielsch, L.-I., & Bornholdt, S. (2002). Scale-Free Topology of E-Mail Networks.

- Physical Review E*, 66, R35103.
- Eckmann, J.-P., Moses, E., & Sergi, D. (2004). Entropy of Dialogues Creates Coherent Structures in E-Mail Traffic. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 14333-14337. <https://doi.org/10.1073/pnas.0405728101>
- Erlang, A. K. (1917) Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. *Post Office Electrical Engineer's Journal*, 10, 189-197.
- Ethier, S., & Kurtz, T. (1986). *Markov Processes: Characterization and Convergence*. New York: Wiley. <https://doi.org/10.1002/9780470316658>
- Feller, W. (1969). *An Introduction to Probability Theory and Its Applications* (Vol. II). New Delhi: Wiley-Eastern.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (Volume II, p. 704) (2 ed.). New York: John Wiley & Sons Inc.
- Gallotti, R., Bazzani, A., Rambaldi, S., & Barthelemy, M. (2016). A Stochastic Model of Randomly Accelerated Walkers for Human Mobility. *Nature Communications*, 7, 12600. <https://doi.org/10.1038/ncomms12600>
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding Individual Human Mobility Patterns. *Nature*, 453, 779-782. <https://doi.org/10.1038/nature06958>
- Grais, R. F., Ellis, J. H., & Glass, G. E. (2003). Assessing the Impact of Airline Travel on the Geographic Spread of Pandemic Influenza. *European Journal of Epidemiology*, 18, 1065-1072.
- Haight, F. A. (1967). *Handbook of the Poisson Distribution*. New York: Wiley.
- Henderson, S., & Henderson, E. (2001). A Note on the Public Interest and Ethical Behaviour. *Australian Accounting Review*, 11, 68-72. <https://doi.org/10.1111/j.1835-2561.2002.tb00391.x>
- Hong, S. (2010). *Human Movement Patterns, Mobility Models and Their Impacts on Wireless Applications*. Raleigh, NC: North Carolina State University.
- Kwon, O., Son, W.-S., & Jung, W.-S. (2016). The Double Power Law in Human Collaboration Behavior: The Case of Wikipedia. *Physica A: Statistical Mechanics and Its Applications*, 461, 85-91. <https://doi.org/10.1016/j.physa.2016.05.010>
- Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., & Hurst, M. (2007). Patterns of Cascading Behavior in Large Blog Graphs. In *Proceedings of the 2007 SIAM International Conference on Data Mining* (pp. 551-556). <https://doi.org/10.1137/1.9781611972771.60>
- Li, H. et al. (2015). Characterizing Smartphone Usage Patterns from Millions of Android Users. In *Proceedings of the 2015 Internet Measurement Conference* (pp. 459-472). <https://doi.org/10.1145/2815675.2815686>
- Mainardi, F., Raberto, M., Gorenflo, R., & Scalas, E. (2000). Fractional Calculus and Continuous-Time Finance II: The Waiting-Time Distribution. *Physica A: Statistical Mechanics and Its Applications*, 287, 468-481. [https://doi.org/10.1016/S0378-4371\(00\)00386-1](https://doi.org/10.1016/S0378-4371(00)00386-1)
- Masoliver, J., Montero, M., & Weiss, G. H. (2003). Continuous-Time Random-Walk Model for Financial Distributions. *Physical Review E*, 67, Article ID: 021112. <https://doi.org/10.1103/PhysRevE.67.021112>
- Oancea, B. (2017). Income Inequality in Romania: The Exponential-Pareto Distribution. *Physica A: Statistical Mechanics and Its Applications*, 469, 486-498. <https://doi.org/10.1016/j.physa.2016.11.094>

- Oliveira, J. G., & Barabási, A.-L. (2005). Human Dynamics: Darwin and Einstein Correspondence Patterns. *Nature*, *437*, 1251. <https://doi.org/10.1038/4371251a>
- Pareto, V. (1898). Cours d'économie politique. *Journal of Political Economy*, *6*, 549-552. <https://doi.org/10.1086/250536>
- Pieropan, A., Ek, C. H., & Kjellström, H. (2013). Functional Object Descriptors for Human Activity Modeling. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on* (pp. 1282-1289). <https://doi.org/10.1109/ICRA.2013.6630736>
- Plerou, V., Gopikrishnan, P., Amaral, A. N., Gabaix, X., & Stanley, H. E. (2000). Economic Fluctuations and Anomalous Diffusion. *Physical Review E*, *62*, R3023. <https://doi.org/10.1103/PhysRevE.62.R3023>
- Rolski, T., Schmidli, H., Schmidt, V., & Teugels, J. L. (1999). *Stochastic Processes for Insurance and Finance*. Chichester: Wiley. <https://doi.org/10.1002/9780470317044>
- Saito, Y. U., Watanabe, T., & Iwamura, M. (2007). Dolarger Firms Have More Interfirm Relationships? *Physica A: Statistical Mechanics and Its Applications*, *383*, 158-163. <https://doi.org/10.1016/j.physa.2007.04.097>
- Scholz, T. M. (2015). The Human Role within Organizational Change: A Complex System Perspective. In *Change Management and the Human Factor* (pp. 19-31). Berlin: Springer. https://doi.org/10.1007/978-3-319-07434-4_3
- Tsompanidis, I., Zahran, A. H., & Sreenan, C. J. (2014). Mobile Network Traffic: A User Behaviour Model. In *2014 7th IFIP Wireless and Mobile Networking Conference (WMNC)* (pp. 1-8). <https://doi.org/10.1109/WMNC.2014.6878862>
- Van Montfort, M. A. J. (1986). The Generalized Pareto Distribution Applied to Rainfall Depths. *Hydrological Sciences Journal*, *31*, 151-162. <https://doi.org/10.1080/02626668609491037>
- Vazquez, A. (2005). Exact Results for the Barabási Model of Human Dynamics. *Physical Review Letters*, *95*, Article ID: 248701. <https://doi.org/10.1103/PhysRevLett.95.248701>
- Yu, L., Cui, P., Song, C., Zhang, T., & Yang, S. (2017). A Temporally Heterogeneous Survival Framework with Application to Social Behavior Dynamics. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1295-1304). <https://doi.org/10.1145/3097983.3098189>
- Zhu, W.-Y., Peng, W.-C., Chen, L.-J., Zheng, K., & Zhou, X. (2015). Modeling User Mobility for Location Promotion in Location-Based Social Networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1573-1582). <https://doi.org/10.1145/2783258.2783331>