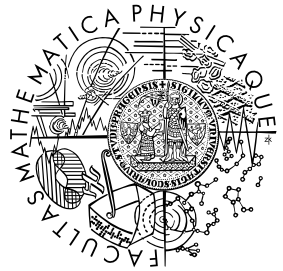

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Robert Oganesyan

Goodness-of-fit tests for Pareto distribution

Department of Probability and Mathematical Statistics

Supervisor of the bachelor thesis: doc. RNDr. Zdeněk Hlávka, Ph.D.
Study programme: Mathematics
Study branch: Financial mathematics

Prague 2023

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

Dedication. I would like to thank my thesis supervisor doc. RNDr. Zdeněk Hlávka, Ph.D.invaluable advice and dedication he has shown me. I am deeply thankful to my parents, who have provided me with the opportunity to study abroad and offered support throughout my academic journey. Last but not least, I thank my friends, without whom my student life would be less exciting.

Title: Goodness-of-fit tests for Pareto distribution

Author: Robert Oganesyan

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Zdeněk Hlávka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Pareto model is widely used in insurance and reinsurance, as well as in finance (capital management, stock returns), and in modelling of natural disasters. As a motivation, we introduce a case study drawn from the pricing of reinsurance excess of loss contracts. In the first chapter of the thesis, we define four types of Pareto distributions and explore such properties as moments, parameter estimation and characterizations. In the following chapter, we examine the theory of unbiased estimation in order to better understand the construction of goodness-of-fit tests based on characterizations. Subsequently, we introduce an overview of some goodness-of-fit tests. The final chapter focuses on simulating the type I error rates and power of these tests, as well as conducting a comparative analysis. Finally, we return to the motivational example and complete the calculation of the price for the reinsurance coverage.

Keywords: Pareto distribution, reinsurance pricing, goodness-of-fit test, distribution characterizations, U-statistics, Bahadur efficiency, kernel statistics, power study.

Contents

Introduction

We initiate our study by exploring one of the many applications of the Pareto model, specifically its implementation in pricing reinsurance excess of loss contracts.

Pricing reinsurance excess of loss (XL) contracts

Reinsurance is a transfer of risks taken by a *direct insurer (cedent)* to a second insurance carrier *(reinsurer)* with no direct connection to the policyholder [Schwepcke et al., 2004]. It enables cedents to limit and spread the direct insurer's risks in order to increase underwriting capacity, lower capital requirements, stabilize financial results, etc. In return for taking these risks, the reinsurer charges a *reinsurance premium*. The part that is passed on to the reinsurer is called *cession* and the part that direct insurer keeps for his own account is known as the *deductible*. Further, we introduce a simplified example of a real-life case where cedent is seeking to transfer some portion of fire risks on his property portfolio.

Example 1. The insurer, in order to protect himself from any fire losses occurring within one full calendar year that exceed a 0.4m EUR deductible, issues a non-proportional excess of loss (XL) contract. The reinsurer agrees to provide cover up to the limit of a 4m EUR loss burden, with unlimited and free restoration of coverage after a loss occurs (reinstatements). The insurer provides the reinsurer with claim experiences (see Table A.1) and additional information on his fire portfolio, such as a list of the largest insured risks and historical time series of collected premiums on the fire portfolio. Our goal as a reinsurer is to estimate the amount of premium that would be sufficient to cover the loss burden of a contract.

Based on the assumption that an adequately high number of observed claims would make it possible to gain a systematic picture of the expected number and size of claims, a mathematical model could be used by the reinsurer in order to estimate the expected amount of ceded losses. We start by constructing an empirical distribution of claim size based on the data provided. Then, we assume a Pareto distribution for modeling losses and find the maximum likelihood estimate (MLE) of its parameters. Formally, we define the Pareto Type (I) distribution.

Definition 1 (Pareto Type (I) distribution)**.** *Random variable X with the distribution function of the form:*

$$P(X \leq x) = 1 - \left[\frac{\sigma}{x}\right]^{\alpha}, \quad \forall x \geq \sigma,\ \sigma > 0,\ \alpha > 0 \tag{1}$$

has Pareto Type (I) probability distribution.

The Pareto model is frequently used for pricing excess of loss covers due to its simplicity and the interpretability of its parameters, which can be clearly understood in this context. The position of the distribution curve is determined by the chosen severity σ, which denotes the threshold at which claims are observed. The parameter α determines the curvature and steepness of the curve. The larger the value of α, the steeper the curve becomes, indicating a greater number of medium-sized and small claims, and fewer large claims, and vice versa.

Collective risk model

We start with the notation for key parameters:

- S is total cedent's loss in a given underwriting year;
- $S^{(C)}$ is total reinsurer's loss (*total ceded loss*);
- $p^{(C)} = \mathsf{E}\, S^{(C)}$ is *pure ceded premium*;
- $p^{(risk)} = p^{(C)} + \epsilon\sqrt{\mathsf{Var}\,(S^{(C)})}$ is *risk-adjusted premium* and some $\epsilon > 0$ (*trend*).

Then, under the assumption of the collective risk model [Kaas et al., 2008]:

- N is number of claims occurred in a given period, $N \sim Poisson(\lambda)$,
- $X_1, \ldots, X_N$ are i.i.d. r.v., where $X_i \sim F_X$ represents the i^{th} claim size,
- $S = \sum_{i=1}^{N} X_i$, where X_i is independent of $N\ \forall i = 1, \ldots, N$,

the random variable S follows a compound Poisson distribution. By using the Law of Total Probability and properties of the conditional distribution of independent random variables X_i and N, we derive the distribution for S:

$$\begin{aligned} F_S(x) = \mathsf{P}(S \le x) &= \sum_{n=0}^{\infty} \mathsf{P}(S \le x | N = n)\, \mathsf{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathsf{P}(\sum_{i=1}^{n} X_i \le x)\, \mathsf{P}(N = n) \\ &= \sum_{n=0}^{\infty} P^{*n}(x)\, \mathsf{P}(N = n), \end{aligned}$$

where P^{*n} is the distribution of sum of n independent random variable with CDF $F_X(x)$.
Subsequently, the total ceded loss is expresses $S^{(C)} = \sum_{i=1}^{N} X_i^{(C)}$, where

$$X_i^{(C)} = \min\{\max\{(X_i - d), 0\}, l\}, d\text{- deductible}, l\text{-limit of the cover},$$

has compound Poisson distribution as well. So then, according to [Kaas et al., 2008, p. 43], we are able to estimate the total risk premium $p^{(risk)}$ using the fact that

$$\begin{aligned} p^{(C)} = \mathsf{E}\, S^{(C)} &= \Big(\mathsf{E}\, N\Big)\Big(\mathsf{E}\, X^{(C)}\Big); \\ \sqrt{\mathsf{Var}\,(S^{(C)})} &= \sqrt{\Big(\mathsf{E}\, N\Big)\, \mathsf{E}\,(X^{(C)})^2}. \end{aligned} \tag{2}$$

In order to continue with the calculation of risk premium we have to make some assumption on the distribution F_X. Schwepcke et al. [2004] proposes to use Pareto distribution for modelling claim sizes with shape parameter α within the range $[1.5, 2, 5]$.

Further in the work we are going to introduce goodness-of-fit tests that would make possible to test whether the hypothesis borough by Schwepcke et al. [2004] is valid or not. We would also discover some of the properties of Pareto distribution, including moments, parameter estimation and distributional characterisations. In the last chapter, we would use all the theory presented further in order to finish the calculation of risk premium needed to cover all the risks coming from signing such reinsurance excess of loss contract.

1. Family of Pareto distributions

In this chapter, we define four modifications of Pareto distributions that are commonly used when studying the whole family of Generalized Pareto distributions. In addition, a more general Feller-Pareto family from [Arnold, 2008] will be introduced in order to derive some distributional results for the Pareto family. We will describe its properties, such as moments and parameter estimation. At the end of the chapter, we will present some characterizations for the classical Pareto distribution.

1.1 The Generalized Pareto Distributions

The hierarchy of the whole family of generalized Pareto models starts with the classical Pareto distribution, as described by Arnold [2008]. We say that a random variable X has a Pareto Type (I) distribution and denote it by $X \sim P(\mathrm{I})(\sigma, \alpha)$ if it has the following survival function:

$$\overline{F}_X(x) = \mathsf{P}\left(X > x\right) = \begin{cases} \left(\frac{x}{\sigma}\right)^{-\alpha}, & x \geq \sigma; \\ 1, & \text{otherwise.} \end{cases}$$

Here, both $\sigma > 0$ and $\alpha > 0$, where σ is the *scaling parameter* and represents the minimum possible value that X could reach, and α is the *shape parameter*. In practice, we often assume $\alpha > 1$ so that the distribution has a finite mean.

In some literature, a standard Pareto distribution may also have an additional real *location parameter* μ. In that case, we will denote it by $P(\mathrm{II})(\mu, \sigma, \alpha)$, where α and σ have the same properties as in the Pareto (I) distribution, with a survival function of the form:

$$\overline{F}_X(x) = \begin{cases} \left[1 + \left(\frac{x-\mu}{\sigma}\right)\right]^{-\alpha}, & x \geq \mu; \\ 1, & \text{otherwise.} \end{cases}$$

Pareto (III) distribution has similar tail behavior to Pareto (II) and its survival function is of the form:

$$\overline{F}_X(x) = \begin{cases} \left[1 + \left(\frac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{-1}, & x \geq \mu; \\ 1, & \text{otherwise,} \end{cases} \tag{1.1}$$

where γ is positive and is called the *inequality parameter*. We denote that random variable X has the survival function (1.1) by $X \sim P(\mathrm{III})(\mu, \sigma, \gamma)$.

Finally generalizing Pareto (III) distribution by adding both the shape α and the inequality γ parameter, we introduce the Pareto (IV) distribution defined by the following survival function:

$$\overline{F}_X(x) = \begin{cases} \left[1 + \left(\frac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{-\alpha}, & x \geq \mu; \\ 1, & \text{otherwise.} \end{cases}$$

We denote that random variable X follows the Pareto (IV) probability law by $X \sim P(\mathrm{IV})(\mu, \sigma, \gamma, \alpha)$.

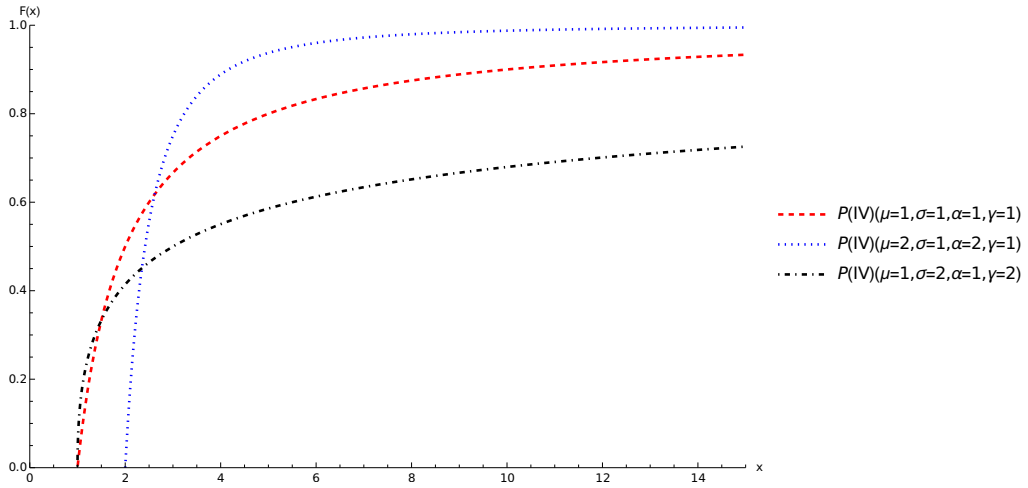

Figure 1.1: Pareto Type (IV) distribution functions with different values of $(\mu, \sigma, \alpha, \gamma)$.

Under the above assumptions, it is clear that Pareto Types $(\mathrm{I}-\mathrm{III})$ are special cases of the generalized Pareto (IV) distribution and they satisfy the following properties:

$$\begin{aligned} P(\mathrm{IV})(\sigma, \sigma, 1, \alpha) &= P(\mathrm{I})(\sigma, \alpha); \\ P(\mathrm{IV})(\mu, \sigma, 1, \alpha) &= P(\mathrm{II})(\mu, \sigma, \alpha); \\ P(\mathrm{IV})(\mu, \sigma, \gamma, 1) &= P(\mathrm{III})(\mu, \sigma, \gamma). \end{aligned} \tag{1.2}$$

1.1.1 Feller-Pareto distribution

In addition to Pareto Type $(\mathrm{I}-\mathrm{IV})$ distributions, it is useful to introduce the Feller-Pareto distribution from [Arnold, 2008] as it provides one more level of generalization of the Pareto Type (IV). It opens the possibility to derive moments for $P(\mathrm{I}-\mathrm{IV})$ using the connection between the Feller-Pareto and beta and beta prime distributions.

Assuming that $B(\gamma_1, \gamma_2)$ stands for *beta distribution*, we denote that a random variable U have *beta prime distribution* with parameters $\gamma_1, \gamma_2 > 0$ by $B'(\gamma_1, \gamma_2)$. Its probability density function is then of the form:

$$f_U(u) = \frac{u^{\gamma_2 - 1}(1+u)^{-\gamma_1 - \gamma_2}}{\mathsf{B}(\gamma_1, \gamma_2)}, \qquad u > 0.$$

Then a random variable $W = \mu + \sigma U^{\gamma}$ follows a *Feller-Pareto* distribution, and we denote it by $W \sim FP(\mu, \sigma, \gamma, \gamma_1, \gamma_2)$ for $\mu \in \mathbb{R}$ and $\sigma, \gamma, \gamma_1, \gamma_2 > 0$. We derive the distribution of W by using the standard change-of-variable technique:

$$\mathsf{P}(W > w) = \mathsf{P}(\mu + \sigma U^{\gamma} > w) = \mathsf{P}\left(U > \left(\frac{w-\mu}{\sigma}\right)^{1/\gamma}\right), \tag{1.3}$$

and finally, assuming $\gamma_2 = 1$, we achieve the simplest form of the survival function $\overline{F}_U(u) = (1+u)^{-\gamma_1},\ u > 0$ and combining that result with (1.3) we obtain:

$$\mathsf{P}(W > w) = \left[1 + \left(\frac{w-\mu}{\sigma}\right)^{1/\gamma}\right]^{-\gamma_1}.$$

That explicitly proves the fact that $FP(\mu, \sigma, \gamma, \alpha, 1) = P(\mathrm{IV})(\mu, \sigma, \gamma, \alpha)$.

1.2 Properties of Pareto (I-IV) distributions

1.2.1 Moments

Moments for the Pareto Type $(\mathrm{I}-\mathrm{IV})$ can be derived from moments for the Feller-Pareto distribution. Under the assumption that $W \sim FP(\mu, \sigma, \gamma, \gamma_1, \gamma_2)$, we define the random variable $W^* = U^\gamma$, then $W^* \sim FP(0, 1, \gamma, \gamma_1, \gamma_2)$. Using the property of the beta prime distribution, we can redefine $W^* = (Y^{-1} - 1)^\gamma$, where $Y \sim B(\gamma_1, \gamma_2)$. So then the k^{th} moment for W^* can then be calculated as follows:

$$\begin{aligned}\mathsf{E}(W^{*k}) = \mathsf{E}\left(Y^{-1}-1\right)^{\gamma k} &= \frac{1}{\mathsf{B}(\gamma_1,\gamma_2)}\int_0^1 \left(\frac{1-y}{y}\right)^{\gamma k} y^{\gamma_1-1}(1-y)^{\gamma_2-1}\mathrm{d}y \\ &= \frac{\mathsf{B}(\gamma_1-\gamma k, \gamma_2+\gamma k)}{\mathsf{B}(\gamma_1,\gamma_2)} = \frac{\Gamma(\gamma_1-\gamma k)\Gamma(\gamma_2+\gamma k)}{\Gamma(\gamma_1)\Gamma(\gamma_2)}.\end{aligned}$$

The k^{th} moment of W^* exists for $k \in (-\gamma_2/\gamma, \gamma_1/\gamma)$. By straightforward application of the Binomial theorem we derive the k^{th} moment of W for $k < \gamma_1/\gamma$:

$$\mathsf{E}\left(W^k\right) = \mathsf{E}\left(\mu + \sigma(Y^{-1}-1)^\gamma\right)^k = \sum_{j=0}^{k}\binom{k}{j}\mu^{k-j}\sigma^j\,\mathsf{E}(W^{*j}).$$

Implicitly, we can obtain the k^{th} moments for Pareto Type $(\mathrm{II}-\mathrm{IV})$ distributions with fixed $\gamma_2 = 1$ by substituting parameter values using (1.2) properties.

However, moments for Pareto (I) cannot be derived from moments of a general Feller-Pareto distribution. The k^{th} moment for a random variable X from $P(\mathrm{I})(\sigma, \alpha)$ with density defined as:

$$f_X(x) = \begin{cases}\frac{\alpha\sigma^\alpha}{x^{\alpha+1}}, \text{ for } x \geq \sigma, \\ 0, \qquad \text{otherwise,}\end{cases} \tag{1.4}$$

could be calculated for $\alpha > k$ via standard approach:

$$\begin{aligned}\mathsf{E}\left(X^k\right) &= \int_{-\infty}^{\infty} x^k f_X(x)\mathrm{d}x = \int_\sigma^\infty \frac{\alpha\sigma^\alpha}{x^{\alpha-k+1}}\mathrm{d}x = \frac{\alpha\sigma^\alpha}{k-\alpha}\left[\frac{1}{x^{\alpha-k}}\right]_\sigma^\infty \\ &= \frac{\alpha\sigma^\alpha}{\alpha-k}\sigma^{-(\alpha-k)} = \frac{\alpha}{\alpha-k}\sigma^k.\end{aligned}$$

1.2.2 MLE of parameters σ and α for P(I) distribution

Let's assume that we work with a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, $n \in \mathbb{N}$, with univariate continuous density $f(x|\boldsymbol{\theta}_X)$ with respect to given measure μ from *parametric model* $\mathcal{F}_{\boldsymbol{\theta}}$:

$$\mathcal{F}_{\boldsymbol{\theta}} = \{\text{distribution with density } f(x|\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d\}, \tag{1.5}$$

where Θ is a *parameter space*. *Maximum likelihood estimate* (*MLE*) $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_X$ is such a point from Θ that maximizes the joint density evaluated at observed values $X_1, \ldots, X_n$ for $\boldsymbol{\theta} \in \Theta$.

We calculate the MLE of $\boldsymbol{\theta}$ for $\boldsymbol{X}$, where $X_i \sim P(\mathrm{I})(\sigma, \alpha)$, $\forall i \in \{1, \ldots, n\}$, with density (1.4), $\boldsymbol{\theta} = (\sigma, \alpha)$ and $\Theta \subseteq \Big((0, \infty) \times (0, \infty)\Big)$. The *likelihood function* takes the following form:

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{\alpha \sigma^{\alpha}}{X_i^{\alpha+1}} = \alpha^n \sigma^{n\alpha} \prod_{i=1}^{n} \frac{1}{X_i^{\alpha+1}}.$$

Since the logarithm is a strictly monotonic function, $L_n(\boldsymbol{\theta})$ and the *log-likelihood function* $\ell_n(\boldsymbol{\theta}) = \ln(L_n(\boldsymbol{\theta}))$ have their maxima at the same point, we continue our calculation using $\ell_n(\boldsymbol{\theta})$:

$$\ell_n(\boldsymbol{\theta}) = n \ln(\alpha) + n\alpha \ln(\sigma) - (\alpha + 1) \sum_{i=1}^{n} \ln(X_i).$$

As Θ is an open set, then $\widehat{\boldsymbol{\theta}}_n$ is the root of a system of the *likelihood equations* defined as:

$$\left(\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) (\widehat{\boldsymbol{\theta}}_n) = \mathbf{0} \in \mathbb{R}^d.$$

Then we calculate the partial derivatives of $\ell_n(\boldsymbol{\theta})$ at point $\widehat{\boldsymbol{\theta}}_n$ and solve the system of the likelihood equations:

$$\left(\frac{\partial \ell_n(\sigma, \alpha)}{\partial \sigma} \right) \left(\widehat{\boldsymbol{\theta}}_n \right) = \left(\frac{n\alpha}{\sigma} \right) \left(\widehat{\boldsymbol{\theta}}_n \right) = 0; \tag{1.6}$$

$$\left(\frac{\partial \ell_n(\sigma, \alpha)}{\partial \alpha} \right) \left(\widehat{\boldsymbol{\theta}}_n \right) = \left(\frac{n}{\alpha} + n \ln(\sigma) - \sum_{i=1}^{n} \ln X_i \right) \left(\widehat{\boldsymbol{\theta}}_n \right) = 0. \tag{1.7}$$

From equation (1.6), we find that for MLE $\widehat{\sigma}_n$ of σ holds $\widehat{\sigma}_n \to \infty$. However, by definition of the Pareto (I) distribution, it holds that $X_i \geq \sigma$ for all $i \in \{1, \ldots, n\}$. Therefore, we maximize the likelihood function by setting $\widehat{\sigma}_n$ as high as possible:

$$\widehat{\sigma}_n = \min_{i \in \{1, \ldots, n\}} X_i. \tag{1.8}$$

Finally, we derive the MLE $\widehat{\alpha}_n$ of α by solving (1.7) and which is of the form:

$$\widehat{\alpha}_n = \frac{n}{\sum_{i=1}^{n} \ln X_i - n \ln(\widehat{\sigma}_n)}. \tag{1.9}$$

1.2.3 Relation to exponential distribution

Since most of goodness-of-fit tests based on characterization for Pareto distribution use its relation to exponential distribution, we introduce this property of Pareto Type (I) distribution in the following theorem.

Theorem 1. *If random variable $X \sim P(\mathrm{I})(\sigma, \alpha)$ then the random variable Y defined as*

$$Y = \ln\left(\frac{X}{\sigma} \right)$$

is exponentially distributed random variable with the rate α (i.e. $Y \sim Exp(\alpha)$).

Proof. By using standard change-of-variable technique we derive the distribution for random variable Y:

$$\begin{aligned}\mathsf{P}(Y \leq y) &= \mathsf{P}\left(\ln\left(\frac{X}{\sigma}\right) \leq y\right) = \mathsf{P}(X \leq \sigma e^{y}) \\ &= 1 - \left(\frac{\sigma}{\sigma e^{y}}\right)^{\alpha} = 1 - e^{-\alpha y}, \quad \forall y \in [\sigma, \infty).\end{aligned}$$

□

Theorem 2. *If random variable $Y \sim Exp(\alpha)$ then the random variable X defined as*

$$X = \sigma e^{Y}$$

follows Pareto Type (I) *distribution with shape $\alpha > 0$ and scale $\sigma > 0$ parameters (i.e. $X \sim P(\mathrm{I})(\sigma, \alpha)$).*

Remark 1. If we define a random variable $X' = \frac{X}{\sigma}$, where $X \sim P(\mathrm{I})(\sigma, \alpha)$,then it holds that $X' \sim P(\mathrm{I})(1, \alpha)$.

1.3 Characterizations of Pareto (I) distribution

In general, a *characterization theorem* in mathematics states that there exists one and only one object that satisfies the properties defined in the theorem. In statistics, a *characterization of a probability distribution* is a set of properties defined in the theorem that holds only for that particular distribution. For simplicity, we will assume that a random variable X follows the $P(\mathrm{I})(1, \alpha)$ distribution. Based on Remark 1, characterizations will hold for any $\sigma > 0$.

Theorem 3 (Characterization based on transformation, [Obradovic et al., 2014])**.** *Let X and Y i.i.d. be non-negative absolutely continuous random variables. Then, X and $\max\left\{\frac{X}{Y}, \frac{Y}{X}\right\}$ have the same distribution if and only if the distribution of X belongs to the family $P(\mathrm{I})(1, \alpha)$.*

Proof. Let's start with the proof of reverse implication. We suppose that random variables X and Y belong to the family $P(\mathrm{I})(1, \alpha)$. We then denote random variable Z as:

$$Z = \max\left\{\frac{X}{Y}, \frac{Y}{X}\right\}.$$

After that, we use Theorem 1 and apply it on both random variables X and Y. This provides a more convenient representation of Z.

$$\widetilde{X} = \ln X \text{ and } \widetilde{Y} = \ln Y \implies \widetilde{X}, \widetilde{Y} \sim Exp(\alpha);$$

$$Z = \max\left\{e^{\widetilde{X}-\widetilde{Y}}, e^{\widetilde{Y}-\widetilde{X}}\right\} = e^{|\widetilde{X}-\widetilde{Y}|}.$$

To derive the distribution of Z, we find the joint distribution of the random variable $W = \widetilde{X} - \widetilde{Y}$, which is derived using the independence of the random

variables $\widetilde{X}$ and $\widetilde{Y}$ and the conditional distribution of $\widetilde{X}$:

$$\begin{aligned}
F_W(w) &= \mathsf{P}\left(\widetilde{X}-\widetilde{Y}\le w\right) = \int_0^\infty \mathsf{P}\left(\widetilde{X}-\widetilde{y}\le w|\widetilde{Y}=\widetilde{y}\right)\mathsf{P}\left(\widetilde{Y}=\widetilde{y}\right)\mathrm{d}\widetilde{y} \\
&= \int_0^\infty \mathsf{P}\left(\widetilde{X}\le w+\widetilde{y}\right)\alpha e^{-\alpha\widetilde{y}}\mathrm{d}\widetilde{y} = \int_0^\infty\left(\int_0^{w+\widetilde{y}}\alpha e^{-\alpha\widetilde{x}}\mathrm{d}\widetilde{x}\right)\alpha e^{-\alpha\widetilde{y}}\mathrm{d}\widetilde{y} \\
&= \int_0^\infty\left(1-e^{-\alpha(w+\widetilde{y})}\right)\alpha e^{-\alpha\widetilde{y}}\mathrm{d}\widetilde{y} = 1-\int_0^\infty \alpha e^{-\alpha(2\widetilde{y}+w)}\mathrm{d}\widetilde{y} \\
&= 1-\frac{1}{2}e^{-\alpha w}.
\end{aligned}$$

After that we are able to derive the distribution function for Z and we see that finally Z follows $P(\mathrm{I})(1,\alpha)$ probability law:

$$\begin{aligned}
F_Z(z) = \mathsf{P}(Z\le z) &= \mathsf{P}\left(e^{|\widetilde{X}-\widetilde{Y}|}\le z\right) = \mathsf{P}\left(|\widetilde{X}-\widetilde{Y}|\le \ln z\right) \\
&= \mathsf{P}\left(-\ln z\le \widetilde{X}-\widetilde{Y}\le \ln z\right) \\
&= \mathsf{P}\left(\widetilde{X}-\widetilde{Y}\le \ln z\right) - \mathsf{P}\left(\widetilde{X}-\widetilde{Y}\le -\ln z\right) \\
&= \mathsf{P}\left(\widetilde{X}-\widetilde{Y}\le \ln z\right) - \mathsf{P}\left(\widetilde{Y}-\widetilde{X}> \ln z\right) \\
&= \mathsf{P}\left(\widetilde{X}-\widetilde{Y}\le \ln z\right) - 1 + \mathsf{P}\left(\widetilde{Y}-\widetilde{X}\le \ln z\right) \\
&= 1-\frac{1}{2}e^{-\alpha\ln z} - 1 + 1 - \frac{1}{2}e^{-\alpha\ln z} = 1-z^{-\alpha}.
\end{aligned}$$

We now move to the proof of the converse implication. Given that the logarithm is a monotonous transformation, X and Z are identically distributed if and only if the random variables $\widetilde{X}$ and $\widetilde{Z}=|\widetilde{X}-\widetilde{Y}|$ are identically distributed. Our goal is to demonstrate that only exponentially distributed random variables $\widetilde{X}$ and $\widetilde{Z}$ satisfy this characterization, as then, according to Theorem 2, X and Z will belong to $P(\mathrm{I})(1,\alpha)$.

To begin, we assume a more general case where $\widetilde{X}$ and $\widetilde{Y}$ are independent, non-negative random variables derived from absolutely continuous distributions. We denote the distribution functions of $\widetilde{X}$ and $\widetilde{Y}$ by F and G, respectively. Given that $\widetilde{X}$ and $\widetilde{Z}$ are identically distributed, we can infer about the distribution F the following property:

$$\begin{aligned}
F_{\widetilde{Z}}(\widetilde{z}) &= \mathsf{P}(\widetilde{X}\le\widetilde{z}+\widetilde{Y},\widetilde{X}\ge\widetilde{Y}-\widetilde{z}) = \int_{-\infty}^{\infty}\int_{\widetilde{y}-\widetilde{z}}^{\widetilde{y}+\widetilde{z}} f_{\widetilde{X},\widetilde{Y}}(\widetilde{x},\widetilde{y})\mathrm{d}\widetilde{x}\mathrm{d}\widetilde{y} \\
&= \int_0^\infty g(\widetilde{y})\int_{\widetilde{y}-\widetilde{z}}^{\widetilde{y}+\widetilde{z}} f(\widetilde{x})\mathrm{d}\widetilde{x}\mathrm{d}\widetilde{y} \\
&= \int_0^\infty g(\widetilde{y})\Big(F(\widetilde{y}+\widetilde{z})-F(\widetilde{y}-\widetilde{z})\Big)\mathrm{d}\widetilde{y} \\
&= \int_0^\infty g(\widetilde{y})F(\widetilde{y}+\widetilde{z})\mathrm{d}\widetilde{y} - \int_0^\infty g(\widetilde{y})F(\widetilde{y}-\widetilde{z})\mathrm{d}\widetilde{y} = F(\widetilde{z}).
\end{aligned} \tag{1.10}$$

After that we can use the *Leibniz's rule* for differentiating under the integral sign and find the density $f_{\widetilde{Z}}(\widetilde{z}) = f(\widetilde{z})$:

$$\begin{aligned}
f(\widetilde{z}) &= \int_0^\infty g(\widetilde{y})\frac{\mathrm{d}}{\mathrm{d}\widetilde{z}}\Big(F(\widetilde{y}+\widetilde{z})\Big)\mathrm{d}\widetilde{y} - \int_0^\infty g(\widetilde{y})\frac{\mathrm{d}}{\mathrm{d}\widetilde{z}}\Big(F(\widetilde{y}-\widetilde{z})\Big)\mathrm{d}\widetilde{y} \\
&= \int_0^\infty g(\widetilde{y})f(\widetilde{y}+\widetilde{z})\mathrm{d}\widetilde{y} + \int_0^\infty g(\widetilde{y})f(\widetilde{y}-\widetilde{z})\mathrm{d}\widetilde{y} \\
&= \int_0^\infty g(\widetilde{y})f(\widetilde{y}+\widetilde{z})\mathrm{d}\widetilde{y} + \int_0^\infty g(\widetilde{y}+\widetilde{z})f(\widetilde{y})\mathrm{d}\widetilde{y}, \quad \forall\widetilde{z}\in(0,\infty).
\end{aligned} \tag{1.11}$$

Here, we obtain a functional equation for $f(\widetilde{z})$. The solution in the general case is complex. According to Puri and Rubin [1970], under the assumptions of $\mathsf{E}\,X < \infty$ and G having an absolutely continuous component, $F(t)$ satisfies equation (1.10) if and only if F is absolutely continuous, with the probability density function $f(t)$ given by

$$f(t) = \frac{1 - G(t)}{\mathsf{E}\,\widetilde{X}}, \quad \forall t \geq 0. \tag{1.12}$$

We show it further by substituting (1.12) into (1.11):

$$\begin{aligned}
\frac{1 - G(\widetilde{z})}{\mathsf{E}\,\widetilde{X}} &= \int_0^\infty g(\widetilde{y}) \frac{1 - G(\widetilde{y} + \widetilde{z})}{\mathsf{E}\,\widetilde{X}}\,\mathrm{d}\widetilde{y} + \int_0^\infty g(\widetilde{y} + \widetilde{z}) \frac{1 - G(\widetilde{y})}{\mathsf{E}\,\widetilde{X}}\,\mathrm{d}\widetilde{y};\\
1 - G(\widetilde{z}) &= 1 - \int_0^\infty g(\widetilde{y}) G(\widetilde{y} + \widetilde{z})\,\mathrm{d}\widetilde{y} + 1 - G(\widetilde{z}) - \int_0^\infty g(\widetilde{y} + \widetilde{z}) G(\widetilde{y})\,\mathrm{d}\widetilde{y};\\
0 &= 1 - [G(\widetilde{y}) G(\widetilde{y} + \widetilde{z})]_0^\infty + \int_0^\infty g(\widetilde{y} + \widetilde{z}) G(\widetilde{y})\,\mathrm{d}\widetilde{y} - \int_0^\infty g(\widetilde{y} + \widetilde{z}) G(\widetilde{y})\,\mathrm{d}\widetilde{y};\\
0 &= 1 - 1.
\end{aligned}$$

Now, it remains to demonstrate that (1.12) is the only solution of (1.11). To show this, we utilize the property of the *stationary distribution* of a *Markov chain*. We consider a sequence of i.i.d. random variables $\{Y_i\}_{i=1}^\infty$ with the same distribution as $\widetilde{Y}$. We then define a second sequence of random variables $\{Z_i\}_{i=1}^\infty$, where $Z_1 = Y_1$ and $Z_{n+1} = |Z_n - Y_{n+1}|$. Since the transition from Z_n to Z_{n+1} depends only on the *current state* Z_n, the sequence Z_k forms a Markov chain with a continuous *state space* $[0, \infty)$ and a *transition distribution* H, given by $\mathrm{d}H(y|x) = \mathrm{d}G(x - y) + \mathrm{d}G(x + y)$.

We can observe that if F_n is the distribution of Z_n, then $F_1 = G$ and for $n = 1, 2, \ldots$, we obtain a relationship somewhat similar to (1.10) but in the recurrent form for F_n:

$$F_n(z) = \int_0^\infty F_{n-1}(y + \widetilde{z})\mathrm{d}G(y) - \int_0^\infty F_{n-1}(y - \widetilde{z})\mathrm{d}G(y). \tag{1.13}$$

It is evident that the solution of (1.10) is a stationary distribution of the Markov chain Z_n. The uniqueness of $f(t)$ given by (1.12) follows from the fact that Z_n is *irreducible* Markov chain, which is a consequence of G having an absolutely continuous part.

Thus, the expression (1.12) is unique and, given that we expect $F = G$, we obtain an *ordinary separable partially differential equation* by considering $y(t) = F(t)$ and $y'(t) = f(t)$ and solving it further:

$$\begin{aligned}
y'(t) &= \frac{1 - y(t)}{\mathsf{E}\,X};\\
\frac{\mathrm{d}y(t)}{1 - y(t)} &= \frac{\mathrm{d}t}{\mathsf{E}\,X};\\
\int \frac{\mathrm{d}y(t)}{1 - y(t)} &= \int \frac{\mathrm{d}t}{\mathsf{E}\,X};\\
-\ln|1 - y(t)| &= \frac{t}{\mathsf{E}\,X} + C;\\
y(t) &= 1 - e^{-C} e^{-(t/\mathsf{E}\,X)}.
\end{aligned}$$

In our case, the constant C equals zero due to the fact that $\int_0^\infty y(t)\mathrm{d}y = 1$. Finally, we see that the distribution function F characterises the alternative parametrization of exponential distribution with the scale parameter $\beta = \mathsf{E}\, X = \frac{1}{\alpha}$. Since we see that both random variables $\widetilde{X}$ and $\widetilde{Z}$ belong to the exponential distribution, than using Theorem 2 leads us to desired result.

□

Further in work, we use the following notation $X_{(1,n)} < X_{(2,n)} < \cdots < X_{(n,n)}$ of the order statistics of a random sample $X_1, X_2, \ldots, X_n$.

Theorem 4 (Characterization based on order statistics, [Volkova, 2014])**.** *Let $X_1, \ldots, X_n$ be random sample, with positive and bounded random variables, having an absolutely continuous distribution function. Then, X_1 and $X_{(n,n)}/X_{(n-1,n)}$ have the same distribution if and only if the distribution of X_1 belongs to $P(\mathrm{I})(1, \alpha)$ distribution.*

Proof. Using already mentioned property of log-transformed random variables from Pareto distribution, Ahsanullah [1977] proves that characterization by proving existing characterization of exponential distribution.

□

In the subsequent chapter, we introduce a detailed examination of the theory of unbiased estimates and derive the limiting distribution for U-statistics. These results would be further applied for a limiting distribution of the goodness-of-fit test statistics derived from characterizations of probability distributions.

2. Theory of unbiased estimates

Goodness-of-fit tests based on characterisations require a detailed understanding of *U-statistics*. First introduced by Hoeffding [1948], U-statistics are so named for their defining property of being "unbiased". They enable the derivation of *unbiased estimators* for a broad set of statistical functionals and extensive classes of probability distributions. It is worth noting that the sample mean and variance both belong to the class of U-statistics.

Consider $\boldsymbol{\theta} = \theta(G)$ as a real, numerically valued functional of probability distribution functions defined on $\mathbb{R}$. The functional $\boldsymbol{\theta}$ could represent the expected value or standard deviation of the distribution G. The challenge in unbiased estimation lies in identifying a function (*statistic*) g_n from a sample of size n from a population with distribution G, such that the expected value of this function equals $\theta(G)$. This concept is formally described in the following definition, with a further overview based on the works of Serfling [1980] and Halmos [1946].

Definition 2 (Unbiased estimate)**.** *Let $\mathcal{G}$ be a subset of all distribution functions defined on $\mathbb{R}$, $X_1, \ldots, X_n$ is a real random sample from G and $\theta(G)$, $\forall G \in \mathcal{G}$ is a real functional. Then we say that $\theta(G)$ grant unbiased estimate, if*

$$\exists n_0 : \forall n > n_0 \; \exists g_n(X_1, \ldots, X_n) : \mathsf{E}\left[g_n(X_1, \ldots, X_n)\right] = \theta(G). \tag{2.1}$$

Then we introduce a necessary and sufficient condition for $\theta(G)$ to have an unbiased estimate of order n over the domain $\mathcal{G}$. This involves the definition of a homogeneous functional.

Definition 3 (Homogeneous functional)**.** *A functional $\theta(G)$ is called homogeneous over $\mathcal{G}$ of degree $k = 1, \ldots,$ if there exists such a measurable mapping $\Psi : \mathbb{R}^k \to \mathbb{R}$ such that the Lebesgue-Stieltjes integral*

$$\theta(G) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \Psi(x_1, \ldots, x_k) \mathrm{d}G(x_1) \ldots \mathrm{d}G(x_k) \tag{2.2}$$

exists and if k is minimal with respect to the property of existence of such representation (see Definition 2). The function Ψ is called the kernel of functional $\theta(G)$.

Theorem 5. *A necessary and sufficient condition that $\boldsymbol{\theta}$ have an unbiased estimate of order n over the domain $\mathcal{G}$ is that it be homogeneous over $\mathcal{G}$ of degree k, where $k \leq n$.*

The previous theorem indicated which functionals $\theta(G)$ admit unbiased estimates. However, these estimates may not always be satisfactory. For instance, let's assume a random sample $X_1, \ldots, X_n$ is drawn from normal distribution $\mathcal{N}(\mu, \sigma^2)$. We seek an unbiased estimate of the expected value $\mu = \theta\Big(\Phi(\frac{x-\mu}{\sigma})\Big)$. There exists such statistic g_n that $\mathsf{E}\left[g_n(X_1, X_2, \ldots, X_n)\right] = \mathsf{E}\, X_1 = \mu$. Although the first element of a sample is an unbiased estimate of μ, it is not a good estimate as it ignores most of the available information. Since we are interested in defining "the best unbiased estimate", we should examine the symmetric ones.

Definition 4 (Symmetric function)**.** *We say that a function g_n is symmetric if and only if $\forall i, j = 1, \ldots, n$:*

$$g_n(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) = g_n(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n).$$

In case g_n would not be symmetric we will define symmetric function g_n^* as

$$g_n^* = \frac{1}{n!} \sum_{(n)} g_n(x_{i_1}, \ldots, x_{i_n}), \tag{2.3}$$

where $\sum_{(n)}$ denotes sum over all permutations of arguments g_n.

The next two theorems show the existence of the only symmetric unbiased estimate and declare that these estimates have the smallest variance.

Theorem 6. *Let $\mathcal{G}$ be a set of all probability distributions with finite support defined on a Borel set E. $\boldsymbol{\theta}$ is a homogeneous functional. Then there exist the only one symmetric unbiased estimate $\Psi^{[n]}$ of $\boldsymbol{\theta}$ on E. If $\Psi = \Psi(x_1, \ldots, x_k)$ is a functional of k real variables, we denote $\Psi^{[n]}$, where $n \geq k$ by*

$$\Psi^{[n]}(X_1, \ldots, X_n) = \frac{(n-k)!}{n!} \sum_{P(k,n)} \Psi(X_{i_1}, \ldots, X_{i_k}), \tag{2.4}$$

where we sum over all $\frac{n!}{(n-k)!}$ partial permutations without replacement of indices $(i_1, \ldots, i_k)$ from $\{1, \ldots, n\}$.

Proof. See [Halmos, 1946], page 39.

□

Theorem 7. *Let $X_1, \ldots, X_n$ be a random sample from $G \in \mathcal{G}$, where $\mathcal{G}$ is a space of all continues distribution functions on $\mathbb{R}$. And $\theta(G)$ is a homogeneous functional with kernel $\Psi(x_1, \ldots, x_k)$ of a degree k. Then it follows:*

- *$\Psi^{[n]}(X_1, \ldots, X_n)$ is the only unbiased symmetric estimate of $\boldsymbol{\theta}$;*
- *$\Psi^{[n]}(X_1, \ldots, X_n)$ has the least variance among all unbiased estimates of $\boldsymbol{\theta}$.*

Proof. The proof of equivalent theorem could be found in [Serfling, 1980].

□

Based on previous theorems we could take $\widehat{\boldsymbol{\theta}} = \Psi^{[n]}$ as an estimate of $\boldsymbol{\theta}$. But then in order to define U-statistics, using remark from Definition 4 we firstly symmetrize the kernel $\Psi_{[k]}$:

$$\Psi_{[k]} = \frac{1}{k!} \sum_{(k)} \Psi(X_{i_1}, \ldots, X_{i_k}),$$

so then we define $\widehat{\boldsymbol{\theta}}$ with symmetrized kernel by substituting the kernel Ψ from Theorem 6 by $\Psi_{[k]}$ and get:

$$\widehat{\boldsymbol{\theta}} = \Psi^{[n]} = \binom{n}{k}^{-1} \sum_{(n,k)} \Psi_{[k]}(X_{i_1}, \ldots, X_{i_k}).$$

In the following definition we formally introduce U-statistics as being "the best unbiased estimate" of functional $\boldsymbol{\theta}$.

Definition 5 (U-statistics)**.** *Let $X_1, \ldots, X_n$ be a random sample, we say that symmetric unbiased estimate $U_n = U_n(X_1, \ldots, X_n)$ of $\boldsymbol{\theta}$ is U-statistics if*

$$U_n(X_1, \ldots, X_n) = \binom{n}{k}^{-1} \sum_{(n,k)} \Psi(X_{i_1}, \ldots, X_{i_k}), \tag{2.5}$$

where we sum over all subsets of $\{i_1, \ldots, i_k\}$, $1 \leq i_1 \leq \cdots \leq i_k \leq n$ and $\Psi(x_{i_1}, \ldots, x_{i_k})$ is symmetric kernel of functional U_n.

To investigate the limiting distribution of U-statistics, we need introduce the following notations: we denote a *one-dimensional projection* of the kernel Ψ from Definition 2.5 by $\psi(x)$ and the *variance of projection* $\psi(X_1)$ by Δ^2. Under the null hypothesis that a random sample $X_1, \ldots, X_k$ follows distribution G, we obtain:

$$\begin{aligned} \psi(x) &= \mathsf{E}\Big[\Psi(X_1, \ldots, X_k)|X_1 = x\Big]; \\ \Delta^2 = \mathsf{Var}\Big[\psi(X_1)\Big] &= \mathsf{E}\Big[\psi^2(X_1)\Big] - \Big(\mathsf{E}[\psi(X_1)]\Big)^2 \\ &= \mathsf{E}\Big[\psi^2(X_1)\Big] - \Big(\theta(G)\Big)^2 \end{aligned} \tag{2.6}$$

We say that U-statistics (2.5) which has $\Delta^2 = 0$ is *degenerate* and *nondegenerate* otherwise. In [Hoeffding, 1948], is stated the for nondegenerate U-statistics the following central limit result is valid.

Theorem 8 (Limiting distribution of U-statistics)**.** *If under the valid null hypothesis $\mathsf{E}\Big[\Psi^2(X_1, \ldots, X_k)\Big] < \infty$ and $\Delta^2 > 0$ (i.e. U_n is nondegenerate), then*

$$\frac{\sqrt{n}}{k\sqrt{\Delta^2}}\Big(U_n - \theta(G)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, 1\Big). \tag{2.7}$$

3. Goodness-of-fit tests

Goodness-of-fit (GOF) tests play a pivotal role in data analysis, with applications spanning a wide range of fields, from economics and finance to computer science and hypothesis testing. We are going to define such statistical tests are used to test the *composite null hypothesis* — observed data sample follows a hypothesized probability distribution with unknown parameters. Tests based on the characterization of probability distributions are particularly useful, as they can reveal essential and latent properties of these distributions, and may be more efficient.

3.1 Overview of goodness-of-fit tests

Definition 6 (Goodness-of-fit test)**.** *Let $X_1, \dots, X_n$ be a random sample from some continuous distribution $F_X(x)$ from parametric model $\mathcal{F}_{\boldsymbol{\theta}}$. We say that statistical test that tests the composite null hypothesis:*

$$H_0 : F_X(x) \in \mathcal{F}_{\boldsymbol{\theta}}{}^*$$

against the alternative

$$H_1 : F_X(x) \notin \mathcal{F}_{\boldsymbol{\theta}},$$

under the assumption that the alternative distribution has the same support is called a goodness-of-fit test.

In Chu et al. [2019], a review of 23 distinct GOF tests is presented for Pareto Type (I) distribution. It includes Kolmogorov-Smirnov and Cramer-von Mises tests, alongside other GOF tests based on transforms (Maintanis and Bassiakos), kernel functions (Jackson, Lewis), spacings (Gulati and Shapiro), Euclidean distances (Rizzo), Kullback-Liebler information (Lequesne), and the characterizations already mentioned in Theorems 3, 4 and etc.

The simulation part in [Chu et al., 2019] provided a *power comparison* for 8 tests for the Pareto (I) distribution for two sample sizes $n \in \{100, 200\}$. The authors simulated the power of these tests for a hundred set of parameters' values each ($\sigma = 1, \forall \alpha \in \{0.01, 0.02, \dots, 1\}$), when the null distribution was $P(\mathrm{I})(1, 1)$. According to the simulated power results, the Kolmogorov-Smirnov test performs the best, with the test based on kernel statistics showing the second-best performance. The remaining 6 tests demonstrated considerably worse simulated power. Several questions raised by the authors in this area will be further explored in this work, specifically, the power of these tests against some distributions that are considered alternatives for Pareto (e.g. lognormal, Weibull, gamma, Burr, etc.).

Further, we focus on some of GOF tests based on probability characterizations, detailing their properties and the *asymptotic distributions* of their statistics.

---

* $\mathcal{F}_{\boldsymbol{\theta}}$ parametric model defined in (1.5)

3.2 GOF test based on characterization

Theorem 9 (Glivenko-Cantelli)**.** *Let $X_1, X_2 \dots$ be a sequence of i.i.d. real valued random variables with common distribution function F and let F_n denote an empirical distribution function. Then as $n \to \infty$*

$$\sup_{x \in \mathbb{R}} \Big| F_n(x) - F(x) \Big| \to 0 \text{ almost surely.}$$

Let $X_1, \dots, X_n$ be a random sample drawn from some continuous distribution $F_X(x)$. Suppose we have a characterization of some probability law given by the identical distribution of two statistics $\phi_1(X_1, \dots, X_r)$ and $\phi_2(X_1, \dots, X_s)$. Then we built two U-empirical CDFs such that:

$$\mathcal{L}_n^1 = \binom{n}{r}^{-1} \sum_{1 \le i_1 \cdots \le i_r \le r} \mathbb{1}\{\phi_1(X_{i_1}, \dots, X_{i_r}) \le t\}, \quad t \in \mathbb{R};$$
$$\mathcal{L}_n^2 = \binom{n}{s}^{-1} \sum_{1 \le i_1 \cdots \le i_s \le s} \mathbb{1}\{\phi_2(X_{i_1}, \dots, X_{i_s}) \le t\}, \quad t \in \mathbb{R},$$

where the sum is taken over all subsets of $\{i_1, \dots, i_r\}$, $1 \le i_1 \le \cdots \le i_r \le n$; $\{i_1, \dots, i_s\}$, $1 \le i_1 \le \cdots \le i_s \le n$

Since the theory of U-empirical CDFs is very similar to theory of usual CDFs we can apply the result of Glivenko-Cantelli theorem on U-empirical CDFs as well. Therefore we state that U-empirical CDFs *converge uniformly* to its' theoretical distributions. Then we are able to construct goodness-of-fit test based on closeness of U-empirical CDFs. We assume under the null hypothesis for large n that $\mathcal{L}_n^1 \approx \mathcal{L}_n^2$ on $\mathbb{R}$. In our study we work with the *integral type statistics*:

$$\int_1^\infty \Big(\mathcal{L}_n^1(t) - \mathcal{L}_n^2(t) \Big) \mathrm{d}F_n(t), \tag{3.1}$$

but *Kolmogorov type test statistics* is often used as well. It is of the following form:

$$\sup_{t \in \mathbb{R}} \Big| \mathcal{L}_n^1(t) - \mathcal{L}_n^2(t) \Big|. \tag{3.2}$$

In the following two sections we introduce two GOF tests for Pareto Type (I) distribution based on integral type statistics and characterizations from both Theorems 3 and 4.

3.3 GOF test based on Theorem 3

In this test we assume the following null hypothesis H_0 and alternative H_1:

$$\begin{aligned} &H_0 : \textit{random sample } X_1, \dots, X_n \textit{ follows } P(\mathrm{I})(1, \alpha) \textit{ distribution;} \\ &H_1 : \textit{random sample } X_1, \dots, X_n \textit{ doesn't follow } P(\mathrm{I})(1, \alpha) \textit{ distribution.} \end{aligned} \tag{3.3}$$

Obradovic et al. [2014] proposes the integral type statistics T_n which is of the form:

$$T_n = \int_1^\infty \Big(M_n(t) - F_n(t) \Big) \mathrm{d}F_n(t). \tag{3.4}$$

Here we assume $\mathcal{L}_n^1(t) = M_n(t)$ and $\mathcal{L}_n^2(t) = F_n(t)$ from (3.3). The equivalence in distribution of $F_n(t)$ and $M_n(t)$ holds as a consequence of characterization defined in the Theorem 3. Then the U-empirical distribution function $M_n(t)$ is defined as

$$M_n(t) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq t\right\}, \quad t \geq 1.$$

In the following two steps, for the purpose of simulation study we firstly simplify the expression of T_n and then show that its limiting distribution is equivalent to U-statistics of the 3$^{\text{rd}}$ degree with the symmetric kernel $\Psi(X, Y, Z)$.

$$\begin{aligned}
T_n =& \int_1^{\infty} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq t\right\} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{X_i \leq t\right\} \mathrm{d}F_n(t) \\
=& \frac{1}{n}\sum_{m=1}^{n} \int_1^{\infty} \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq t\right\} \\
& - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{X_i < t\right\} \mathrm{d}\mathbb{1}\left\{X_m \leq t\right\} \\
=& \frac{1}{n}\sum_{m=1}^{n} \left[\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq X_m\right\} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{X_i \leq X_m\right\}\right] \\
=& \frac{1}{n}\sum_{m=1}^{n} \left[\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq X_m\right\} - \frac{m}{n}\right];
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} T_n =& \left[\lim_{n\to\infty} \frac{2}{n^2(n-1)} \sum_{m=1}^{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq X_m\right\} - \frac{1}{n^2}\sum_{m=1}^{n} m\right] \\
=& \lim_{n\to\infty}\left[\frac{2}{n^2(n-1)} \sum_{m=1}^{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{3!} \sum_{(3)} \mathbb{1}\left\{\max\left(\frac{X_i}{X_j}, \frac{X_j}{X_i}\right) \leq X_m\right\}\right. \\
& \left. - \frac{(n-n^2)}{2n^2}\right] = \binom{n}{3}^{-1} \sum_{(n,3)} \Psi(X, Y, Z).
\end{aligned}$$

Kernel $\Psi(X, Y, Z)$ is symmetrized according to Definition 4 and is of the form:

$$\begin{aligned}
\Psi(X, Y, Z) =& \frac{1}{3}\mathbb{1}\left\{\max\left(\frac{X}{Y}, \frac{Y}{X}\right) \leq Z\right\} + \frac{1}{3}\mathbb{1}\left\{\max\left(\frac{X}{Z}, \frac{Z}{X}\right) \leq Y\right\} \\
& + \frac{1}{3}\mathbb{1}\left\{\max\left(\frac{Z}{Y}, \frac{Y}{Z}\right) \leq Y\right\} - \frac{1}{2}.
\end{aligned} \tag{3.5}$$

Consequently, using derived results for U-statistics mentioned in equasions (2.6) we evaluate under the null hypothesis the projection $\psi(x)$ of kernel $\Psi(X, Y, Z)$ and the variance Δ^2 of $\psi(X)$ as following:

$$\begin{aligned}
\psi(x) &= \mathsf{E}\left[\Psi(X, Y, Z)|X = x\right] = \frac{2}{3}\frac{\alpha \ln(x)}{x^{\alpha}} - \frac{1}{6}; \\
\Delta^2 = \mathsf{Var}\left[\psi(X)\right] &= \int_1^{\infty} \psi^2(x)\alpha x^{-\alpha-1}\mathrm{d}x - \left(\int_1^{\infty} \psi(x)\alpha x^{-\alpha-1}\mathrm{d}x\right)^2 \\
&= \frac{5}{972} - \left(\left[-\frac{2a\ln(x) - x^a + 1}{6x^{2a}}\right]_1^{\infty}\right)^2 = \frac{5}{972}.
\end{aligned}$$

From the calculation above we see that $\mathsf{E}\left[\psi(X)\right] = 0$. The test statistics T_n is asymptotically equivalent to nondegenerate ($\Delta^2 > 0$) U-statistics with centered kernel (3.5) of a degree 3. Using these facts according to properties of nondegenerate U-statistics (see Theorem 8), we obtain that T_n has asymptotically normal distribution:

$$\sqrt{n}T_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{5}{108}\right). \tag{3.6}$$

3.4 GOF test based on Theorem 4

In this test we assume the following null hypothesis H_0 and alternative H_1:

H_0 : *random sample* $X_1, \ldots, X_n$ *follows* $P(\mathrm{I})(1, \alpha)$ *distribution*;
H_1 : *random sample* $X_1, \ldots, X_n$ *doesn't follow* $P(\mathrm{I})(1, \alpha)$ *distribution*.

In [Volkova, 2014] there were proposed two test statistics: integral type statistics (3.1) and Kolmogorov type statistic (3.2). They both use the characterization from Theorem 4. We will focus just on integral statistics since it was discovered in that article that given the same alternatives integral type statistics has higher *asymptotic local Bahadur efficiencies* with lower degree of U-statistics needed than Kolmogorov one. Bahadur local efficiency is used to quantify the performance of a statistics by evaluating its *asymptotic relative efficiency* as the alternative hypothesis approaches the null hypothesis. The theory of asymptotic relative efficiency of a statistical test is out of scope in this work, the outline of Bahadur theory could be found in e.g. [Nikitin, 2017].

In the similar manner as in Section 3.3, we define the test statistics $I_n^{(k)}$:

$$I_n^{(k)} = \int_1^\infty (H_n^{(k)}(t) - F_n(t))\mathrm{d}F_n(t), \tag{3.7}$$

where the equivalence in distribution of $F_n(t)$ and $H_n^{(k)}(t)$ is granted by characterization from Theorem 4. $H_n^{(k)}(t)$ is a U-empirical CDF defined as:

$$H_n^{(k)}(t) = \binom{n}{k}^{-1} \sum_{1 \le i_1 < \cdots < i_k \le n} \mathbb{1}\left\{\frac{X_{(k,\{i_1,\ldots,i_k\})}}{X_{(k-1,\{i_1,\ldots,i_k\})}} \le t\right\}, \quad t \ge 1,$$

where $X_{(k,\{i_1,\ldots,i_k\})}$ denotes the k^{st} order statistics of a subsample $X_{i_1}, \ldots, X_{i_k}$.

Statistics $I_n^{(k)}$ is asymptotically equivalent to U-statistics of a degree $(k+1)$ with the symmetric kernel which is of the form:

$$\Psi_k(X_{i_1}, \ldots, X_{i_{k+1}}) = \frac{1}{k+1} \sum_{\pi(i_1,\ldots,i_{k+1})} \mathbb{1}\left\{\frac{X_{(k,\{i_1,\ldots,i_k\})}}{X_{(k-1,\{i_1,\ldots,i_k\})}} \le X_{i_{k+1}}\right\} - \frac{1}{2},$$

where the sum is over all permutations of indices from set $\{i_1, \ldots, i_{k+1}\}$.

For the purpose of simulation study we choose a special case of $I_n^{(k)}$, where $k = 4$,since among all k it shows maximum local Bahadur efficiency for the 1^{st} and 2^{st} Ley-Paindaveine distribution alternatives and good efficiency for log-Weibull alternative for $k = 4$. Using somewhat similar approach as was used for the test

statistics T_n (3.4) the expression of test statistics $I_n^{(4)}$ then could be simplified to the form:

$$I_n^{(4)} = \frac{1}{n}\sum_{m=1}^{n}\left[\binom{n}{4}^{-1}\sum_{1\le i_1<\cdots<i_k\le n}\mathbb{1}\left\{\frac{X_{(k,\{i_1,\ldots,i_k\})}}{X_{(k-1,\{i_1,\ldots,i_k\})}}\le X_m\right\}-\frac{m}{n}\right].$$

In [Volkova, 2014], it was stated that U-statistics with kernel Ψ_k is nondegenerate. Then according to Theorem 8, U-statistics have asymptotically normal distribution. Considering $k=4$ under the null hypothesis we evaluate $\psi_4(x)$ and $(\Delta_4)^2$. Consequently, obtaining that $\sqrt{n}I^{(4)}$ as $n\to\infty$ is asymptotically normal:

$$\sqrt{n}I^{(4)} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\frac{271}{2100}\right). \tag{3.8}$$

3.5 Kolmogorov-Smirnov GOF test

In order to implement Kolmogorov-Smirnov (K-S) goodness-of-fit test for testing the same null hypothesis against the alternative (3.3) as in the above mentioned tests, we need to estimate unknown parameters σ, α from data. In this context, have to reformulate the null hypothesis H_0 and the alternative H_1 to

$$\begin{aligned} &H_0 : \textit{random sample } X_1,\ldots,X_n \textit{ follows } P(\mathrm{I})(1,\widehat{\alpha}_n) \textit{ distribution};\\ &H_1 : \textit{random sample } X_1,\ldots,X_n \textit{ doesn't follow } P(\mathrm{I})(1,\widehat{\alpha}_n) \textit{ distribution}, \end{aligned} \tag{3.9}$$

where the MLE of α is denoted by $\widehat{\alpha}_n$ (see equation (1.9)). In accordance with reformulated hypothesis we define the test statistics K_n assuming $F_0 = P(\mathrm{I})(1,\widehat{\alpha}_n)$:

$$K_n = \sup_{x\in\mathbb{R}}\Big|F_n(x)-F_0(x)\Big|, \tag{3.10}$$

where F_n denotes empirical CDF of a random sample $X_1,\ldots,X_n$.

In this context, the default critical values for K-S GOF test are not valid as they are not dependant on specific value of $\widehat{\alpha}_n$. Therefore, we simulate new rejection criterion for K-S GOF test with unknown parameter α using the following parametric bootstrap methodology:

1. let $X_1,\ldots,X_n$ be a random sample drawn from population;
2. find MLE $\widehat{\alpha}_n$ based on $X_1,\ldots,X_n$;
3. generate a bootstrap random sample denoted by $X_1^{*i},\ldots,X_n^{*i}$ of size n from $P(\mathrm{I})(1,\widehat{\alpha}_n)$;
4. find MLE $\widehat{\alpha}_n^{*i}$ from data $X_1^{*i},\ldots,X_n^{*i}$;
5. calculate bootstrapped K-S statistics $K_n^{*i} = \sup_{x\in\mathbb{R}}\left|F_n^{*i}(x)-F_0(x)\right|$, where $F_n^*(x)$ stands for the ECDF of $X_1^{*i},\ldots,X_n^{*i}$ and $F_0 = P(\mathrm{I})(1,\widehat{\alpha}_n^{*i})$;
6. run for $i=1,\ldots,1000$ bootstrap replicates (repeat 1000 times steps 3-5), get the values $K_n^{*1},\ldots,K_n^{*1000}$;
7. critical value $\mathcal{C}$ at significance level α is equal to $(1-\alpha)$ quantile of the ECDF of $K_n^{*1},\ldots,K_n^{*1000}$;
8. reject the null hypothesis in prospect to alternative for $K_n > \mathcal{C}$.

4. Kernel GOF statistics

Kernel GOF statistics could be interesting from the perspective of pricing reinsurance. Since clients often do not report extreme losses (due to the high return period of such losses), it is challenging to estimate the tail of the fitted distribution and assess whether it is reasonably heavy. We need to consider the maximum loss that a client could experience from the list of the largest insured risks and adjust our modeling accordingly based on the client's exposure, as well as common market and peril practices (legal limits, extreme market losses). Underestimating extreme losses can lead to significant issues, particularly when reinsurers deal with unlimited coverage. This could result in an underestimated solvency capital requirement. This test could potentially uncover, based on the k largest losses, the heaviness already present in the data.

In [Chu et al., 2019], authors introduce GOF tests based on general kernel statistics as well as two special cases with Jackson and Lewis kernel functions. We test the following null hypothesis against the alternative:

$$H_0 : \textit{the upper tail behaves as Pareto } (\mathrm{I})(1, \alpha);$$
$$H_1 : \textit{the upper tail does not behave as Pareto } (\mathrm{I})(1, \alpha).$$

Originally, these tests were introduced to test for exponentiality. However, once again the relationship between Pareto and exponential distribution allows us to adapt test for testing the tail of Pareto distribution. In this context, we introduce some theory underlying the general kernel based GOF test. And further we modify Jackson kernel statistics in the way, that it measures the linearity of the k largest observations on the Pareto quantile plot.

In order define GOF test based on kernel statistics, we need some results that were presented in Goegebeur et al. [2008]. According to the extreme value theory for heavy tailed distributions, we define a *first order condition* for function $L_U(x)$ in terms of a tail quantile function $U(x) = \inf\{y : F(y) \geq 1 - 1/x\}$ $\forall x > 1$:

$$U(x) = x^{\lambda} L_U(x), \tag{4.1}$$

where $L_U(x)$ denotes a slowly varying function such that:

$$\frac{L_U(\delta x)}{L_U(x)} \to 1 \text{ as } x \to \infty \text{ and } \forall \delta > 0.$$

We say that, a slowly varying function $L(x)$ satisfies *the second order condition* $\mathcal{R}_L$ of tail behavior if there exist a real constant $\rho \leq 0$ and a rate function b satisfying $b(x) \to \infty$ as $x \to 0$ such that $\forall \delta > 1$ it holds:

$$\frac{L(\delta x)}{L(x)} - 1 \sim b(x) \frac{\delta^{\rho} - 1}{\rho} \text{ as } x \to \infty. \tag{4.2}$$

In that context, we reformulate the null hypothesis and alternative for random sample from some distribution F in the following manner:

$$H_0 : F \textit{ is Pareto type } (\mathrm{I})(1, \alpha) \textit{ with } L_U \textit{ satisfying } \mathcal{R}_L.$$

Exponential goodness-of-fit test statistics often involve a comparison between two estimators for the exponential scale parameter. Building upon this idea and using the properties of Pareto (I) order statistics mentioned the proof of Theorem 4 or in greater detail in [Ahsanullah, 1977], we adopt a similar approach by applying a ratio to the k largest order statistics. Let a random sample $X_1, \ldots, X_n$ be from Pareto (I) distribution with parameters $\sigma = 1$ and $\alpha = 1/\lambda$. Then we derive the following test statistic:

$$T_{n,k} = \frac{\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) Z_j}{H_{k,n}}, \tag{4.3}$$

where

- K denotes a kernel function satisfying $\int_0^1 K(u)du = 0$;
- $Z_j = j\left(\ln X_{(n-j+1,n)} - \ln X_{(n-j,n)}\right) \implies Z_j \sim \text{Exp}(1/\lambda),\ j = 1, \ldots, n$;
- $H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j$ is the Hill estimator of λ (which is in fact empirical mean excess function of the log-transformed data) [Hill, 1975].

According to Theorem 2.1 from [Goegebeur et al., 2008], under the null hypothesis $\sqrt{k}T_{n,k}$ have asymptotically normal distribution as $k, n \to \infty$, $k/n \to 0$ and $\sqrt{k}b(n/k) \to c$, where $b(t)$ satisfies the equation (4.2):

$$\sqrt{k}T_{n,k} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{c}{\lambda}\int_0^1 K(u)u^{-\rho}\mathrm{d}u, \int_0^1 K^2(u)\mathrm{d}u\right). \tag{4.4}$$

We see that the limiting distribution of $T_{n,k}$ is not centered at zero and centering is dependent on unknown constant c and unknown parameter λ. That could potentially lead to some bias in test statistics. Then the decision rule for testing the null hypothesis at the significance level α is to reject H_0 if

$$\begin{aligned}\sqrt{k}\left|\frac{1}{kH_{k,n}}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) Z_j - \frac{b(n/k)}{\lambda}\int_0^1 K(u)u^{-\rho}\mathrm{d}u\right| > \\ > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\int_0^1 K^2(u)\mathrm{d}u},\end{aligned} \tag{4.5}$$

where Φ^{-1} denotes the standard normal quantile function. However, the rule isn't particularly useful for practical applications as it relies on the unknown function b, in addition to the parameters λ and ρ. One solution is to select a relatively small k, i.e. sufficiently small to ensure $\sqrt{k}b(n/k) \approx 0$. This then leads to the directive to reject H_0 if

$$\frac{\sqrt{k}}{H_{k,n}}\left|\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) Z_j\right| > \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\int_0^1 K^2(u)\mathrm{d}u}$$

Then one may ask a question, what is the optimal threshold for approximate tail parameter λ (4.1). In [Goegebeur et al., 2008] it was proposed to choose the number of upper order statistics k based on the minimum asymptotic mean squared error of the Hill estimator $AMSE\left(H_{k,n}\right)$. The method uses its connection to the bias component (4.4) of kernel statistics $T_{n,k}$ expressed as following:

$$AMSE\left(H_{k,n}\right) = \lambda^2\left[\frac{1}{k} + \left(\frac{b(n/k)}{\lambda(1-\rho)}\right)^2\right]. \tag{4.6}$$

Using the property of the test statistics $T_{n,k}$ and its bias component from equation (4.5) we find its connection with (4.6) and then derive the estimate for $AMSE\left(H_{k,n}\right)$:

$$\widehat{AMSE}\left(H_{k,n}\right) = \gamma^2 \left\{ \frac{1}{k} + \left[\frac{\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right) Z_j}{(1-\rho) H_{k,n} \int_0^1 K(u) u^{-\rho} \mathrm{d}u} \right]^2 \right\}.$$

The optimal choice of k for kernel statistics is then approximated by

$$\widehat{k}_{\text{opt}} = \arg \min_{k \in \{1,\ldots,n\}} \left\{ \frac{1}{k} + \left[\frac{T_{n,k}}{(1-\widehat{\rho}_k) \int_0^1 K(u) u^{-\widehat{\rho}_k} \mathrm{d}u} \right]^2 \right\}, \tag{4.7}$$

where $\widehat{\rho}_k$ denotes a consistent estimate of ρ inspired by Genberg et al. [2008]. It is expressed as a log-ratio of Hill estimators for some $\delta \in (0,1)$:

$$\widehat{\rho}_k = -\left| \frac{1}{\ln \delta} \ln \left| \frac{H_{\lfloor \delta^2 k \rfloor, n} - H_{\lfloor \delta k \rfloor, n}}{H_{\lfloor \delta k \rfloor, n} - H_{k,n}} \right| \right|. \tag{4.8}$$

4.1 Jackson and Lewis kernel functions

The Jackson and Lewis kernel goodness-of-fit statistics, $J_{n,k}$ and $L_{n,k}$, are defined according to general kernel statistics $T_{n,k}$ (4.3), with the kernel functions respectively set to $K_J(u) = -1 - \ln(u)$ and $K_L(u) = u - 0.5$. Under the same set of assumptions as for general kernel statistics for testing the tail of Pareto distribution, the Jackson and Lewis statistics are asymptotically normally distributed:

$$\begin{aligned} \sqrt{k} J_{n,k} &\xrightarrow{\mathcal{D}} \mathcal{N}\left(-\frac{c\rho}{\lambda(1-\rho)^2}, 1 \right); \\ \sqrt{k} L_{n,k} &\xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{c\rho}{2\lambda(1-\rho)(2-\rho)}, \frac{1}{12} \right). \end{aligned} \tag{4.9}$$

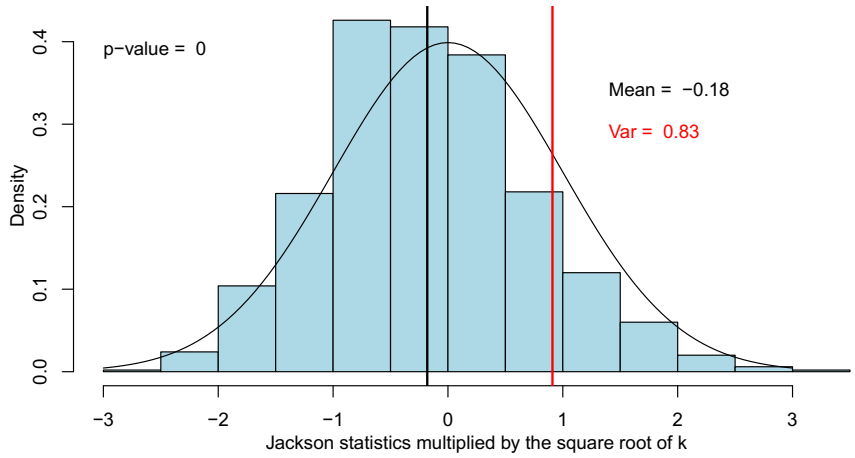

Figure 4.1: Simulated empirical distribution of $\sqrt{k} J_{n,k}$ for $n = 150$, $k = 120$; theoretical PDF of $\mathcal{N}(0,1)$, Shapiro-Wilk test rejects normality.

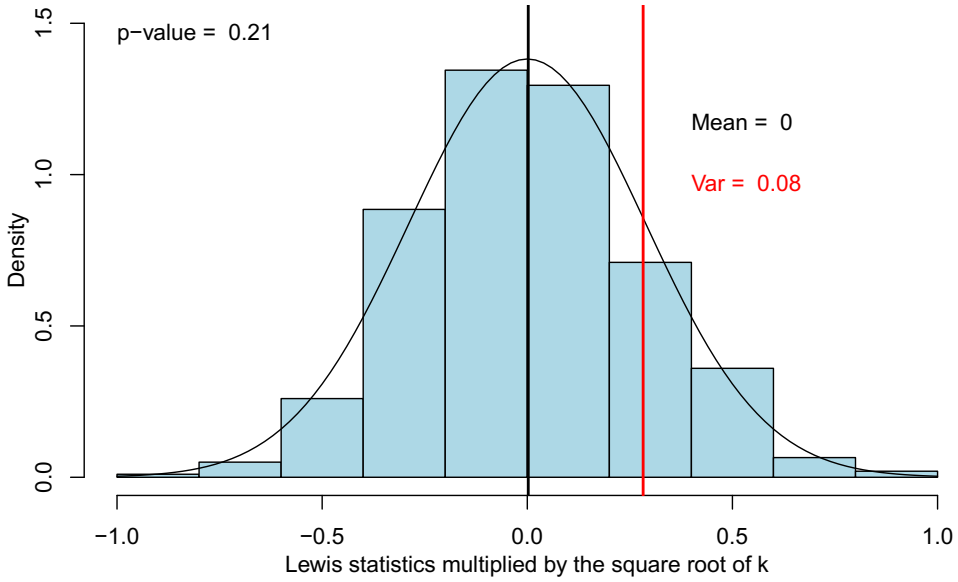

Figure 4.2: Simulated empirical distribution of $\sqrt{k}L_{n,k}$ for $n = 150$, $k = 120$; theoretical PDF of $\mathcal{N}(0, 1/12)$, Shapiro-Wilk test does not reject normality.

Note that for the same value of constant c, the absolute value of bias of Lewis statistics is considerably smaller than the bias of Jackson statistics. As it was already mentioned above, from a practical standpoint, in the simulation study we assume $c \approx 0$, so then both Jackson and Lewis kernel statistics are centred at zero. This assumption is inline with e.g. the empirical distribution of Jackson and Lewis statistics for sample size 150 and $k = 120$ displayed on histograms above. In the simulation part it would also be interesting to explore how the level of significance will change based on different values of k.

4.2 Bias correction

The bias of the kernel statistics could make it difficult determine tail behavior. Therefore, we introduce a bias corrected test statistics $K_{BC}(i/k + 1, \rho)$. Under valid null hypothesis with $\rho < 0$ it holds according to [Goegebeur et al., 2008] for log-spacings of the successive order statistics:

$$Z_j \sim \lambda + b(n/k)\left(\frac{j}{k+1}\right)^{-\rho} + \epsilon_j, \quad \forall j \in \{1, 2, \ldots, k\}, \tag{4.10}$$

where function $b(n/k) = b_{n,k}$ is from (4.2) and ϵ_j are centered error terms. Then bias corrected kernel statistics could be expressed as following:

$$\frac{\frac{1}{k}\sum_{j=1}^{k} K\left(\frac{j}{k+1}\right)\left(Z_j - \widehat{b}_{\mathrm{LS},k}(\rho)\left(\frac{j}{k+1}\right)^{-\rho}\right)}{\widehat{\lambda}_{\mathrm{LS},k}(\rho)}, \tag{4.11}$$

where $\widehat{\lambda}_{\mathrm{LS},k}$ and $\widehat{b}_{\mathrm{LS},k}$ denote the least squared estimates of λ and $b_{n,k}$ obtained from (4.10) respectively. For a given k they are both defined as functions of

unknown parameter ρ:

$$\widehat{b}_{\mathrm{LS},k}(\rho) = \frac{(1-\rho)^2(1-2\rho)}{\rho^2}\frac{1}{k}\sum_{j=1}^{k}\left(\left(\frac{j}{k+1}\right)^{-\rho} - \frac{1}{1-\rho}\right)Z_j; \tag{4.12}$$

$$\widehat{\lambda}_{\mathrm{LS},k}(\rho) = H_{k,n} - \frac{\widehat{b}_{\mathrm{LS},k}(\rho)}{1-\rho}. \tag{4.13}$$

In order to obtain the limiting distribution, we define the bias corrected kernel function from (4.11) substituting $\widehat{b}_{\mathrm{LS},k}$ with the expression above and get the following representation:

$$K_{BC}(u,\rho) = K(u) - \frac{(1-\rho)^2(1-2\rho)}{\rho^2}\left(u^{-\rho} - \frac{1}{1-\rho}\right)\int_0^1 K(t)t^{-\rho}\mathrm{d}t. \tag{4.14}$$

Then under the null hypothesis the limiting distribution of the bias corrected statistics is centered normal distribution:

$$\frac{\sqrt{k}}{\widehat{\lambda}_{\mathrm{LS},k}(\rho)}\frac{1}{k}\sum_{j=1}^{k}K_{BC}\left(\frac{i}{k+1},\rho\right)Z_j \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \int_0^1 K_{BC}^2(u,\rho)\mathrm{d}u\right). \tag{4.15}$$

For the purpose of the simulation study, the kernel function cannot be dependent on unknown parameter ρ, therefore we substitute it with its consistent estimator $\widehat{\rho}_k$ defined in equation (4.8).

4.3 Bias corrected Jackson and Lewis statistics

Using the limiting distribution obtained for general bias corrected statistics (4.15) and substituting Jackson and Lewis kernel function into (4.14), we obtain limiting distribution for bias corrected Lewis $\widetilde{L}_{n,k}$ and Jackson $\widetilde{J}_{n,k}$ statistics which are now centered at zero:

$$\sqrt{k}\widetilde{J}_{n,k} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \left(\frac{\rho}{1-\rho}\right)^2\right);$$

$$\sqrt{k}\widetilde{L}_{n,k} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{12}\left(\frac{1+\rho}{2-\rho}\right)^2\right).$$

Remark 2. However, in [Beirlant et al., 2006] it was discovered that the bias of kernel statistics prevented to distinguish between Pareto-type and non-Pareto-type tail behaviour it is still quite challenging to define and later use the bias corrected kernel statistics for testing the null hypothesis. On simulation stage both Jackson and Lewis bias corrected statistics were defined, but they had not delivered the expected result. First of all, Jackson BC statistic was not centered and secondly the hypothesis that the limiting distribution of the bias corrected test statistics according to Shapiro-Wilk test rejected normality for sample sizes e.g. $n = 150$ and $k = 120$. The main reason why these test statistics behave in contrary to theoretical results is a quite indistinct definition of the consistent estimator $\widehat{\rho}_k$ of ρ. In various sources it was defined in five different ways in

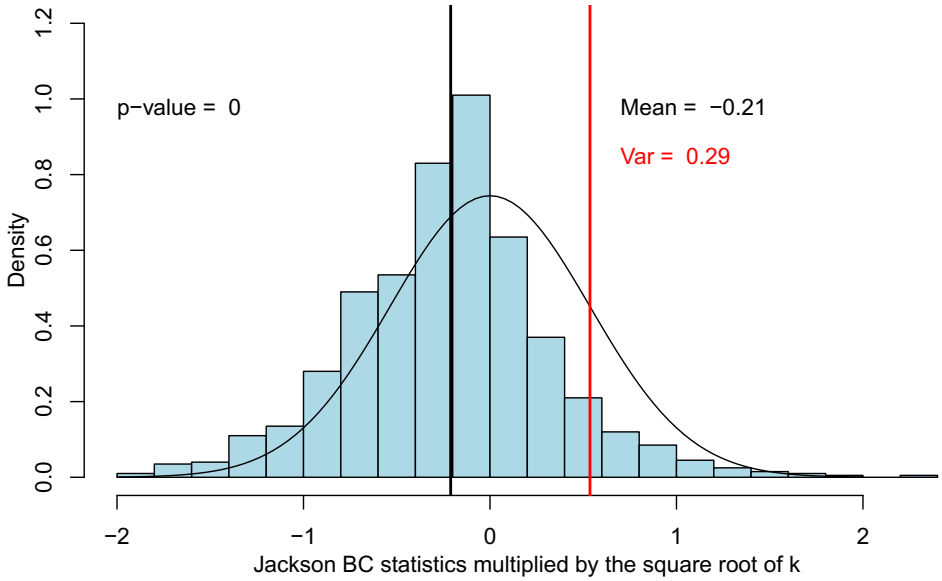

Figure 4.3: Simulated empirical distribution of $\sqrt{k}\widetilde{J}_{n,k}$ for $n = 150,\ k = 120$; theoretical PDF of $\mathcal{N}(0, \mathsf{Var}(\sqrt{k}\widetilde{J}_{n,k}))$, Shapiro-Wilk test rejects normality.

[Goegebeur et al., 2008], [Genberg et al., 2008], [Chu et al., 2019] and [Drees and Kaufmann, 1998]. In four out of five cases $\widehat{\rho}_k$ was also dependent on some other parameter, which value should be picked from some fixed range (e.g. in our case some $\delta \in (0, 1)$), what really complicated the process. The other suggestion using a fixed value for $\widehat{\rho}_k$ results in losing the bias correcting effect of the test statistics.

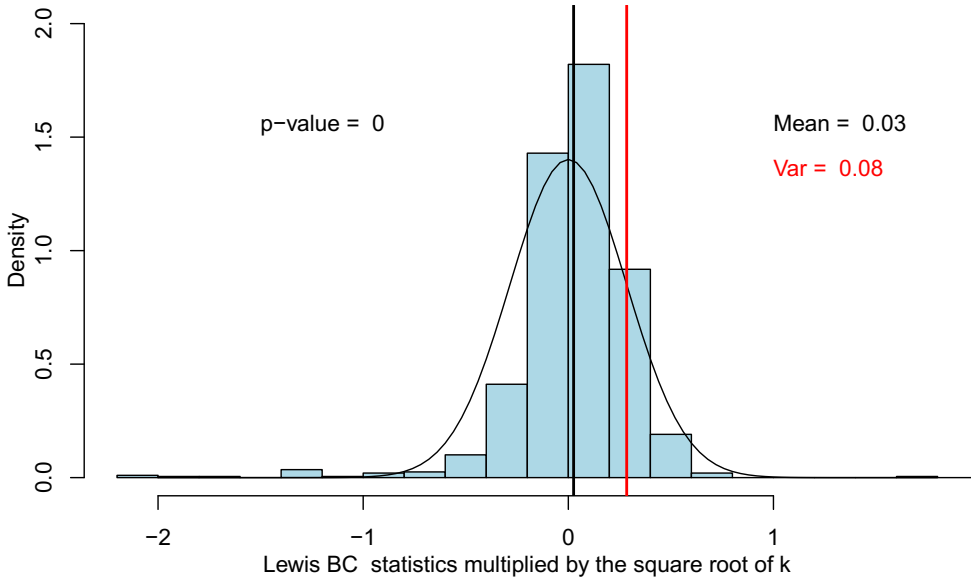

Figure 4.4: Simulated empirical distribution of $\sqrt{k}\widetilde{L}_{n,k}$ for $n = 150,\ k = 120$; theoretical PDF of $\mathcal{N}(0, \mathsf{Var}(\sqrt{k}\widetilde{L}_{n,k}))$, Shapiro-Wilk test rejects normality.

5. Simulation study

In this chapter, we want to study in more detail GOF tests that were mentioned in that work. The first part of the chapter focuses on examining whether the level $\alpha = 0.05$ is maintained for the test statistics. Then we compare the power of these GOF tests for some distributions that are commonly considered as alternatives for Pareto distribution. And finally we take a step back to the case study presented in the introduction and attempt to answer the question weather selecting of the simple Pareto model for fitting the claim data was a reasonable choice. And finally complete the calculating of risk adjusted premium for (XL) contract.

5.1 Type I error

We conducted simulations to determine the type I error rates of the test statistics T_n, $I_n^{(4)}$, and K_n for different sample sizes ($n = 20, 50, 150$), assuming a significance level of $\alpha = 0.05$. The results for the test statistics are presented in the following table:

n	T_n	$I_n^{(4)}$	K_n
20	0.050	0.058	0.055
50	0.048	0.048	0.048
150	0.067	0.060	0.061

Table 5.1: Simulated significance level of test statistic for different sample sizes.

We simulate the significance level of $T_n, I_n^{(4)}$ by calculating p-values based on properties of their asymptotic normality derived in (3.6) and (3.8). P-values for Kolmogorov-Smirnov statistics K_n were calculated from the critical values simulated via parametric bootstrap methodology described in Section 3.5. For all considered sample sizes and for all test statistics from the table above, the simulated levels conform to the theoretical $\alpha = 0.05$. This validates the definition of the test, at least in terms of the level. It is worth noting that the computational complexity of the test statistic $I_n^{(4)}$ is $O\left(\binom{n}{5}\right)$, which makes this test statistic computationally demanding for large sample sizes.

As already mentioned in the paragraph after the equation (4.5), we are interested in how Jackson and Lewis kernel statistics perform based on different sample sizes and the fraction of upper order statistics k that are considered. Table 5.2 provides an overview of the achieved simulated level of all kernel GOF statistics, depending on the number k of upper order statistics considered. In addition, we provide the value of the simulated level for the estimate $\widehat{k}_{\text{opt}}$ (4.7) of an optimal threshold k.

Reviewing the Table 5.2, several key points stand out. We see that the simulated level of the Lewis bias-corrected test statistics is quite high. During the simulation process, we found that the consistent estimate $\widehat{\rho}_k$, dependent on parameter $\delta \in (0, 1)$, was not suitable for Lewis kernel statistics with the same value $\delta = 0.5$ as the Jackson case. The simulated level for Lewis bias-corrected statistics, shown in the table above, was calculated for $\delta = 0.85$, as this was one

n	k	k/n in %	$J_{n,k}$	$\widetilde{J}_{n,k}$	$L_{n,k}$	$\widetilde{L}_{n,k}$	ave$(J_{n,k}, \widetilde{J}_{n,k}, L_{n,k})$
20	4	20	0.000	0.029	0.001	0.132	0.010
	10	50	0.004	0.046	0.019	0.207	**0.023**
	16	80	0.007	0.037	0.023	0.208	0.022
	$\widehat{k}_n$	-	0.000	0.002	0.001	0.091	0.001
50	10	20	0.004	0.027	0.025	0.188	0.019
	25	50	0.022	0.038	0.045	0.226	**0.035**
	40	80	0.022	0.034	0.035	0.237	0.030
	$\widehat{k}_n$	-	0.003	0.027	0.002	0.098	0.011
150	30	20	0.019	0.037	0.039	0.236	0.032
	75	50	0.029	0.035	0.045	0.215	0.036
	120	80	0.034	0.040	0.043	0.214	**0.039**
	$\widehat{k}_n$	-	0.003	0.048	0.003	0.090	0.018

Table 5.2: Simulated level for kernel test statistics for different sample sizes and different values of upper order statistics.

of the better options identified during adjustments of δ. The simulation with the threshold set to $\widehat{k}_n$ did not perform well. The only reasonable level was achieved for $\widetilde{J}_{n,k}$, and $n = 150$ seems to be a coincidence. However, for increasing sample sizes, the simulated level becomes closer to the theoretical. Contrary to the assumption derived from the rejection criterion 4.5 for general kernel GOF statistics — we cannot draw any conclusions based on the obtained simulated level for considering different fractions for selection of k.

For the power study, we need to consider a fixed k. We set the ratio k/n to 50 % for sample sizes $n = 20, 50$. and consider the ratio equal to 80 % for sample size $n = 50$. We decided on this particular selection based on the proximity of the average achieved simulated level for kernel statistics (not considering Lewis biased statistics) to the theoretical $\alpha = 0.05$, as seen in column ave$(J_{n,k}, \widetilde{J}_{n,k}, L_{n,k})$ in the Table 5.2 above.

5.2 Power comparison

In this section, we set the significance level $\alpha = 0.05$ and we compare the power of all GOF statistics $T_n,\ K_n,\ J_{n,k},\ L_{n,k}\ \widetilde{J}_{n,k},\ \widetilde{L}_{n,k}$ for sample sizes $n = 20, 50, 150$. Due computational complexity of $I_n^{(4)}$ we simulate the power just for sample sizes $n = 20, 50$. We propose six distributions that are often considered alternatives for Pareto distribution:

- log-normal with $\mu = 1,\ \sigma = 0$;
- gamma with $\alpha = 2,\ \beta = 1$;
- Weibul with $\alpha = 2$;
- the mixture of two Pareto distributions that has the following distribution function:

  $$P_{\text{mix}}(x, p) = (1-p)\Big(P(\mathrm{I})(\sigma, \alpha)\Big) + p\Big(P(\mathrm{I})(\sigma, 2\alpha)\Big),\ \alpha, \sigma > 0,\ x > \sigma, p \in [0, 1],$$

  with fixed parameter values $\sigma = 1$, $\alpha = 1,\ p \in \{0.2, 0.4\}$
- Burr with $\alpha = 1,\ \rho = -1, \eta = 1$.

n	alternative	T_n	$I_n^{(4)}$	K_n	$J_{n,k}$	$L_{n,k}$	$\widetilde{J}_{n,k}$	$\widetilde{L}_{n,k}$
20	log-normal	1.000	0.856	1.000	0.231	0.431	0.156	0.462
	gamma	0.763	0.179	0.945	0.518	0.689	0.226	0.628
	Weibull	1.000	0.979	1.000	0.588	0.737	0.266	0.648
	mixPareto $p = 0.2$	0.241	0.177	0.258	0.006	0.035	0.035	0.206
	mixPareto $p = 0.4$	0.328	0.221	0.357	0.009	0.032	0.036	0.189
	Burr	1.000	0.904	1.000	0.179	0.362	0.126	0.392
50	log-normal	1.000	0.990	1.000	0.156	0.256	0.099	0.350
	gamma	0.914	0.165	1.000	0.356	0.457	0.133	0.507
	Weibull	1.000	1.000	1.000	0.431	0.540	0.194	0.534
	mixPareto $p = 0.2$	0.555	0.332	0.528	0.014	0.035	0.036	0.210
	mixPareto $p = 0.4$	0.722	0.479	0.702	0.015	0.040	0.026	0.217
	Burr	1.000	0.996	1.000	0.076	0.135	0.067	0.259
150	log-normal	1.000	-	1.000	1.000	1.000	0.349	0.746
	gamma	0.999	-	1.000	1.000	1.000	0.507	0.793
	Weibull	1.000	-	1.000	1.000	1.000	0.571	0.804
	mixPareto $p = 0.2$	0.969	-	0.932	0.054	0.079	0.050	0.225
	mixPareto $p = 0.4$	0.994	-	0.986	0.074	0.111	0.064	0.233
	Burr	1.000	-	1.000	0.986	0.998	0.280	0.688

Table 5.3: Power of the test for different alternatives and sample sizes.

We have observed that the power of the GOF test, based on Jackson bias corrected statistics $\widetilde{J}_{n,k}$, is low for all alternatives and sample sizes. Lewis bias-corrected test statistics $\widetilde{L}_{n,k}$ displays a greater power compared $L_{n,k}$ across all alternatives and sample sizes $n = 20, 50.$, but shows considerably less power for the increased sample sample size $n = 150$. This could likely be due to the complications and uncertainties caused by estimating numerous unknown parameters. An interesting observation to note is the unexpected behavior of the power for all kernel tests. As the sample size increases from 20 to 50 observations, and the same ratio of 50 % for k is maintained, we observe a decrease in the power of the test. On the whole, the simulated power of kernel GOF statistics remains unsatisfactory until the sample size reaches $n = 150$. Even then, the power against the mixture Pareto distribution remains very low.

Regarding the simulated power for GOF statistics $T_n,\ I_n^{(4)},\ K_n$, we have identified the K-S GOF statistics as the best performing among all when the sample size is small ($n = 20$). For $n = 50$, the powers of T_n, and K_n are comparable. However, for $n = 150$, the power of T_n surpasses all other test statistics. Although the power of $I_n^{(4)}$ is comparable to the previous two, it is always slightly lower, especially it is always very low for the gamma alternative.

Despite their struggles, all test statistics had difficulty distinguishing the mixture Pareto distribution. Reasonable power was only achieved by $T_n,\ I_n^{(4)},\ K_n$ for the sample size $n = 150$. This level of performance is considered acceptable, given that small sample sizes may not efficiently provide sufficient information for rejecting false hypotheses in the case of the mixture of two Pareto distributions with close rate $\alpha = 1$ and scale parameters.

5.3 Application in fitting of real data

In this section, we try to assess whether the Pareto (I) model, proposed in introduction as an estimate of claim size distribution on fire portfolio was reasonable. Taking into account performance results, we rely on just two test statistics: Kolmogov-Smirnov K_n and the one T_n (3.4) based on characterization property of Pareto distribution.

We want to test if claims with captured time value of money (we use historical consumer price index in combination with an additional charge for more rapid cost growth of property repair) could be described with a Pareto model. Because of initial assumption for testing the null hypothesis with $\sigma = 1$, we need to normalize the data by MLE $\widehat{\sigma}_n$ of the scale parameter σ and before testing the hypothesis. The resulting p-values for K_n and the one T_n are presented in th following table:

statistics	p-value
T_n	0.53
K_n	0.87

Table 5.4: Calculated p-value for the test statistics on client data.

We see that two the most reliable tests do not reject the hypothesis, so the the assumption that Pareto probability law with some shape parameter α in the range proposed by is reasonable for that particular case. So then we can complete the calculation of risk premium $p^{(risk)}$ from the motivation Example 1.

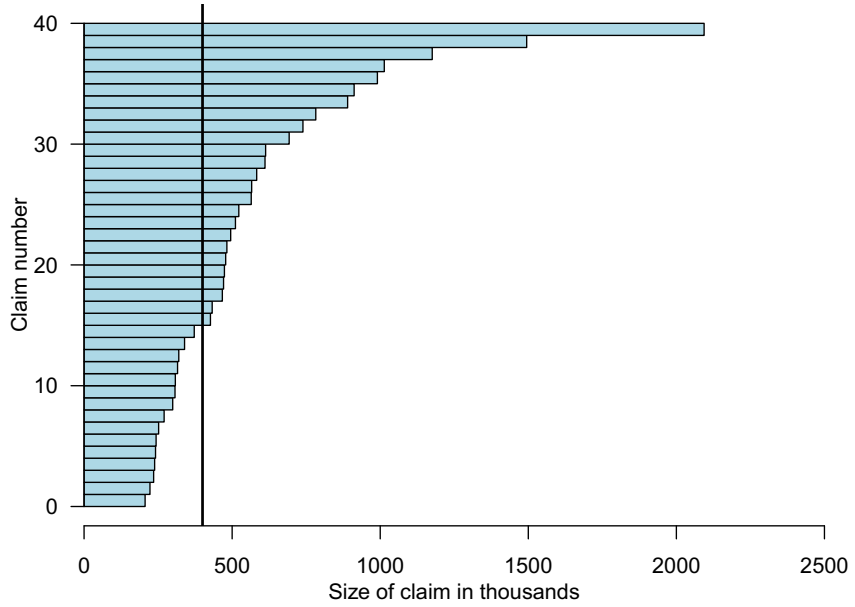

Figure 5.1: Inflation adjusted claim experience provided by the cedent. Forty claims arranged by size, the deductible $d = 400$ thousand EUR is indicated by a perpendicular line.

5.4 Calculation of risk adjusted premium

We begin with the definition of risk adjusted premium:

$$p^{(risk)} = p^{(C)} + \epsilon\sqrt{\mathsf{Var}\left(S^{(C)}\right)} \text{ and some } \epsilon > 0,$$

where ϵ is chosen positive or negative in presents of some trend development of claim sizes of increased frequency. Based on assumptions that are valid for collective risk model we obtained that the risk premium could be expressed as:

$$p^{(risk)} = \Big(\mathsf{E}\, N\Big)\Big(\mathsf{E}\, X^{(C)}\Big) + \epsilon\sqrt{\Big(\mathsf{E}\, N\Big)\,\mathsf{E}\,(X^{(C)})^2}$$

. We see that in order to estimate the premium we need to estimate:

- $\mathsf{E}\, N$ –expected number of claim per contract year;
- $\mathsf{E}\, X^{(C)}$ — expected ceded amount per contract year;
- $\mathsf{E}\,(X^{(C)})^2$ — the second moment of $\mathsf{E}\, X^{(C)}$.

So the under the assumption that $X_1, \ldots, X_N$ follow Pareto Type (I) probability law we calculate $\mathsf{E}\, X^{(C)}$ for $l < \infty$:

$$\begin{aligned}\mathsf{E}\, X^{(C)} &= \int_\sigma^\infty \max\left\{(x-d)_+, l\right\} f_X(x)\mathrm{d}x = \alpha\sigma^\alpha \int_d^{d+l} (x-d)x^{-\alpha-1}\mathrm{d}x \\ &= \alpha\sigma^\alpha\left[\left[\frac{x^{-\alpha+1}}{1-\alpha}\right]_d^{d+l} - d\left[\frac{x^{-\alpha}}{-\alpha}\right]_d^{d+l}\right] = \alpha\sigma^\alpha \cdot \left[\frac{d^{1-\alpha}}{\alpha^2-\alpha} - \frac{\alpha l + d}{(\alpha^2-\alpha)\,(l+d)^\alpha}\right].\end{aligned}$$

Solution of $\mathsf{E}\,(X_i^{(C)})^k$ for $k > 1$ has too complicated form and is omitted.

We estimate the unknown values of parameters σ and α by method of maximum likelihood presented in Section 1.2.2. We get $\widehat{\alpha}_{25} = 2.2$, $\widehat{b}_{25} = 426716.8$. We can see that Pareto distribution describes the empirical distribution of inflated claims well on the (see Figure 5.1). The deductible $d = 400$ thousand and the limit $l = 4$mil.

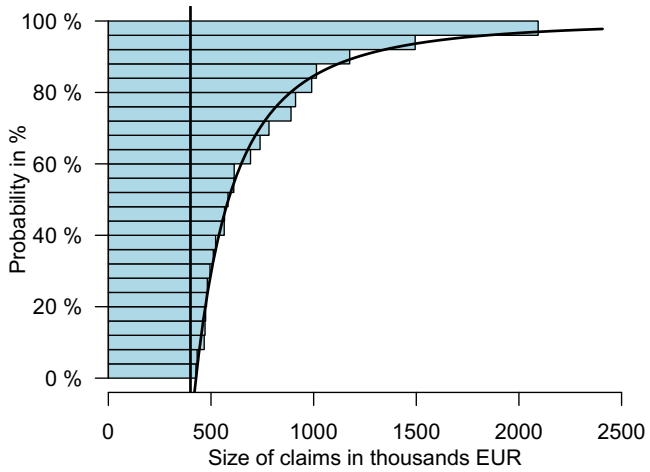

Figure 5.2: There are 25 claims excess d = 400 thousand EUR fitted with Pareto model with MLE $\widehat{\alpha}_n$ and $\widehat{\sigma}_n$ of parameters α and σ.

In the next step we need to estimate $\mathsf{E}\, N$ (see Figure 5.4). Under the assumption of collective risk model we assume that the random variable $N \sim Poisson(\lambda)$. According to the property of Poisson distribution the moment estimate of parameter λ is the sample mean. So we estimate based on client date is equal to $\widehat{\lambda}_{20} = \overline{N} = 3.13$

By using results from the above sections we can calculate the total premium for the excess of loss reinsurance contract from Example 1. By evaluating $\mathsf{E}\,(X_i^{(C)})^k$ for $k = 1, 2$; $d = 0.4$m and $l = 4$m, we calculate the risk-adjusted reinsurance premium, assuming $\epsilon = 0.23$ (see Figure 5.3):

$$\mathsf{E}\, S^{(C)} = \Big(\mathsf{E}\, N\Big)\Big(\mathsf{E}\, X^{(C)}\Big) = 339403.74 * 3.13 = 1062333,7062;$$
$$\epsilon\sqrt{\mathsf{Var}\,(S^{(C)})} = 229768.91; \quad \Longrightarrow \; p^{(Risk)} = 1292102,62.$$

Figure 5.3: Number of ceded loses occurred per contract year. There are 25 claims in excess $d = 400$ thousand EUR. The moment estimate of $\lambda = 3.13$ is the expected ceded claim number per contract year.

Figure 5.4: Ceded claims occurred per contract year. There are 25 claims in excess $d = 400$ thousand EUR. Red horizontal line expresses the average ceded claim per contract year.

Conclusion

We started the thesis with exploring one of many applications of Pareto model — pricing of reinsurance excess of loss contract. We used some theory defined by collective risk model in order to formally state the problem. We draw the hypothesis that claim sized could be estimated by Pareto model.

The main goal of this thesis was to define the family of Pareto distributions and introduce some goodness-of-fit tests for Pareto distribution. We accurately defined the four types of Pareto distributions and summarized their fundamental properties such as moments, parameter estimation, and characterizations. The second chapter focused on definition on unbiased estimates — U-statistics, which properties were used for constructing GOF test based on characterizations. Then a Kolmogorov-Smirnov test for Pareto distribution with unknown shape parameter was introduced. We used bootstrap mythology in order to simulate critical values for the updated rejection of the hypothesis. After that kernel goodness-of-fit tests were introduced, with Jackson and Lewis kernel statistics. We faced some issues as the tests statistics were very dependent on many estimators.

In the last chapter we simulated the level and powers for all tests mentioned above. We draw some conclusions and revealed, the two best performing tests were Kolmogorov-Smirnov and test based on characterization introduced in [Obradovic et al., 2014].

Finally we haven't rejected the initial hypothesis stated in introduction, that claim data came from Pareto distribution. And then completed the estimation of reinsurance premium.

Bibliography

M. Ahsanullah. A characteristic property of the exponential distribution. *The Annals of Statistics*, 5(3):580–582, 1977. ISSN 00905364. URL http://www.jstor.org/stable/2958914.

Barry C Arnold. Pareto and generalized Pareto distributions. In *Modeling income distributions and Lorenz curves*, pages 119–145. Springer, 2008.

Jan Beirlant, Tertius de Wet, and Yuri Goegebeur. A goodness-of-fit statistic for pareto-type behaviour. *Journal of Computational and Applied Mathematics*, 186(1):99–116, 2006. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2005.01.036. URL https://www.sciencedirect.com/science/article/pii/S0377042705001913. Special Issue: Jef Teugels.

J. Chu, O. Dickin, and S. Nadarajah. A review of goodness of fit tests for pareto distributions. *Journal of Computational and Applied Mathematics*, 361:13–41, 2019. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2019.04.018. URL https://www.sciencedirect.com/science/article/pii/S0377042719302067.

Holger Drees and Edgar Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications*, 75(2):149–172, 1998. ISSN 0304-4149. doi: https://doi.org/10.1016/S0304-4149(98)00017-9. URL https://www.sciencedirect.com/science/article/pii/S0304414998000179.

B. L. Genberg, M. Kulich, S. Kawichai, P. Modiba, A. Chingono, G. P. Kilonzo, L. Richter, A. Pettifor, M. Sweat, and D. D. Celentano. HIV risk behaviors in sub-Saharan Africa and Northern Thailand: Baseline behavioral data from project Accept. *Journal of Acquired Immune Deficiency Syndrome*, 49:309–319, 2008.

Yuri Goegebeur, Jan Beirlant, and Tertius de Wet. Linking pareto-tail kernel goodness-offit statistics with tail index at optimal threshold and second order estimation. *REVSTAT-Statistical Journal*, 6(1):51–69, Mar. 2008. doi: 10.57805/revstat.v6i1.57. URL https://revstat.ine.pt/index.php/REVSTAT/article/view/57.

Paul R. Halmos. The Theory of Unbiased Estimation. *The Annals of Mathematical Statistics*, 17(1):34 – 43, 1946. doi: 10.1214/aoms/1177731020. URL https://doi.org/10.1214/aoms/1177731020.

Bruce M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163 – 1174, 1975. doi: 10.1214/aos/1176343247. URL https://doi.org/10.1214/aos/1176343247.

Wassily Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325, 1948. doi: 10.1214/aoms/1177730196. URL https://doi.org/10.1214/aoms/1177730196.

Rob Kaas, Marc Goovaerts, Jan Dhaene, and Michel Denuit. *Modern Actuarial Risk Theory: Using R*. Springer, 01 2008. ISBN 978-3-540-70992-3. doi: 10.1007/978-3-540-70998-5.

Ya. Nikitin. Tests based on characterizations, and their efficiencies: a survey. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 21(1):4–24, jul 2017. doi: 10.12697/acutm.2017.21.01. URL https://doi.org/10.12697%2Facutm.2017.21.01.

Marko Obradovic, Milan Jovanović, and Bojana Milosevic. Goodness-of-fit tests for pareto distribution based on a characterization and their asymptotics. *Statistics*, 05 2014. doi: 10.1080/02331888.2014.919297.

Prem S. Puri and Herman Rubin. A Characterization Based on the Absolute Difference of Two I. I. D. Random Variables. *The Annals of Mathematical Statistics*, 41(6):2113 – 2122, 1970. doi: 10.1214/aoms/1177696709. URL https://doi.org/10.1214/aoms/1177696709.

A. Schwepcke, D. Arndt, and S.R.G. AG. *Reinsurance: Principles and State of the Art - A Guidebook for Home Learners*. Bildungsnetzwerk Versicherungswirtschaft. Verlag Versicherungswirtschaft, 2004. ISBN 9783899521597. URL https://books.google.cz/books?id=nr_5YrljLaUC.

Robert J. Serfling. *Approximation theorems of mathematical statistics*. Wiley series in probability and mathematical statistics : Probability and mathematical statistics. Wiley, New York, NY [u.a.], [nachdr.] edition, 1980. ISBN 0471024031. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+024353353&sourceid=fbw_bibsonomy.

K. Volkova. Goodness-of-fit tests for the pareto distribution based on its characterization. *Statistical Methods and Applications*, 25, 08 2014. doi: 10.1007/s10260-015-0330-y.

List of Figures

List of Tables

A. Attachments

A.1 First Attachment

Year	Reported	Inflated	Claim xs 0.4m = X	$X^{(C)}$	$X/\widehat{\sigma}_n$
2015	179 961	243 250	0	0	-
2017	186 275	251 785	0	0	-
2021	199 999	270 335	0	0	-
2019	221 468	299 355	0	0	-
2018	227 200	307 102	0	0	-
2022	228 000	308 184	0	0	-
2020	233 566	315 708	0	0	-
2020	236 620	319 835	0	0	-
2016	251 010	339 286	0	0	-
2017	275 257	372 061	0	0	-
2017	315 693	426 717	426 717	26 717	1,00
2018	320 034	432 584	432 584	32 584	1,01
2018	345 435	466 919	466 919	66 919	1,09
2020	348 434	470 972	470 972	70 972	1,10
2015	350 633	473 945	473 945	73 945	1,11
2017	353 783	478 203	478 203	78 203	1,12
2018	356 745	482 206	482 206	82 206	1,13
2018	366 270	495 080	495 080	95 080	1,16
2019	378 153	511 143	511 143	111 143	1,20
2016	386 573	522 524	522 524	122 524	1,22
2019	417 588	564 446	564 446	164 446	1,32
2019	418 682	565 925	565 925	165 925	1,33
2021	431 256	582 921	582 921	182 921	1,37
2018	452 000	610 961	610 961	210 961	1,43
2016	453 551	613 057	613 057	213 057	1,44
2021	512 367	692 558	692 558	292 558	1,62
2018	546 780	739 073	739 073	339 073	1,73
2015	578 866	782 444	782 444	382 444	1,83
2016	658 374	889 913	889 913	489 913	2,09
2019	674 631	911 888	911 888	511 888	2,14
2017	732 786	990 494	990 494	590 494	2,32
2021	749 999	1 013 760	1 013 760	613 760	2,38
2016	869 765	1 175 647	1 175 647	775 647	2,76
2022	1 105 700	1 494 556	1 494 556	1 094 556	3,50
2020	1 548 932	2 093 665	2 093 665	1 693 665	4,91

Table A.1: Thirty five largest claims provided by client on his property portfolio with fire risks. In addition, some calculations are made. Calculated ceded amount and normalization applied.