Parameter Estimation for the Heavy Tailed Distributions with the Empirical Distribution

H. Jabbari Khamnei, H. Bevrani and A.A. Haydari
Department of Statistics, Faculty of Mathematical Science,
University of Tabriz and Research Institute for Fundamental Science, Tabriz, Iran

Abstract: One of the most challenging characteristics of heavy tailed distribution is their high or even infinite variance or burstiness. Heavy-tailed distributions are the distributions, the tails of which cannot be cut off. So, we cannot neglect the large-scale but rare events. The most evident and natural problem of statistical estimation connected with heavy tailed distributions is the problem of their parameter estimation. For estimating their parameter, known statistical procedures can be used, for example, estimation methods of Hill, de Haan and others. This research addresses a new method for estimating the parameters of heavy tailed distributions based on the empirical distribution and, using the list squares method. To demonstrate the ability of the proposed approach, it is applied to famous Pareto and Student distributions as two different examples. The related numerical results are reported.

Key words: Empirical distribution, Glivenko theorem, least squares method

INTRODUCTION

Let \( F(x) \) be the distribution function of random variable \( X \). Then \( F(x) \) is said to be heavy tailed distributions class, if there exists positive constants \( C \) and \( 0 < \gamma < 2 \) such that:

\[
\lim_{x \to 0^+} x^\gamma F(x) = C
\]

Where:
\[
F(x) = 1 - F(x).
\]

One of the most challenging characteristics of heavy tailed distribution is its high or even infinite variance or burstiness. This guarantees the condition that \( 0 < \gamma < 2 \). Heavy tailed distributions have been observed in many natural phenomena including hydrology, geology, electrotechnology, informatics, physics, insurance, they describe the behavior of financial markets (Mandelbrot, 1982; Hosking and Wallis, 1987; Brockwell and Davis, 1991; Janicki and Weron, 1994; Adler et al., 1998; Nikias and Shao, 1995). When analyzing statistical dependences we often neglect the probabilities of large-scale events on the tail of distribution.

Heavy tailed distributions are the distributions, the tail of which cannot be cut off. So, we cannot neglect the large-scale but rare events.

Further we will investigate Pareto and Student distributions, as heavy tailed distributions class representatives.

Pareto distribution: We know that Pareto density function is given by:

\[
p_{\alpha,\beta}(x) = \frac{\beta x^\beta}{x^{\alpha+1}}, \quad x > \beta, \quad \alpha > 0.
\]

Its distribution function has the form:

\[
F(x) = 1 - \left( \frac{\beta}{x} \right)^\alpha
\]

In case \( \beta = 1 \) this distribution is considered to be the distribution with one parameter. Mean and variance of this distribution are:

\[
E(X) = \frac{\alpha \beta}{\alpha - 1}, \quad \alpha > 1;
\]

\[
\text{Var}(X) = \alpha \beta^2 \frac{(\alpha - 2)}{(\alpha - 1)^2}, \quad \alpha > 2.
\]

The distribution has infinite mean and variance if \( \alpha < 1 \) and \( \alpha < 2 \), respectively. Especially in case \( \alpha < 2 \), when variance is very high, Pareto distribution is a heavy tailed distribution. Comparing Eq. 1 and 3 we realize that...
heavy tailed distributions and Pareto distribution are similar. That is why we often use the first definition for heavy tailed distribution.

**Student distribution**: A random variable X has a Student distribution with probability mass function:

\[
p_X(x) = \frac{\Gamma((\gamma + 1)/2)}{\sqrt{\pi\Gamma(\gamma/2)}} (1 + x^2/\gamma)^{-\gamma/2}, \quad -\infty < x < \infty.
\]  

It's not complicated to realize that for the Student distribution with parameter \( \gamma \) moments of degrees \( \delta \geq \gamma \) do not exist.

Moreover if \( Y_\gamma \) is random value with Student's distribution and \( x \to \infty \) then:

\[
P(Y_\gamma > x) = \frac{C_\gamma}{x^\gamma},
\]

Where:

\[
C_\gamma = \frac{2\Gamma((\gamma + 1)/2)\gamma^{1/2}}{\sqrt{\pi\Gamma(\gamma/2)}}
\]

Equation 5 satisfies the heavy tailed distributions definition. That is why Student distribution belongs to heavy tailed distributions class. The parameter \( \gamma \) is the index of tail gravity.

We pay attention that \( x \to \infty \) only when \( \gamma \leq 2 \) exist. This condition exactly is equivalent to infinite variance.

The most evident and natural problem of statistical estimation connected with heavy tailed distributions is the problem of parameter \( \gamma \) estimation. For estimating the parameter known procedures can be used, for example, estimations of Hill (1974), de Haan (1970) and others (Hosking and Wallis, 1987). The mentioned estimation methods, as usual, are calculated easier than maximum likelihood or moments estimation methods. In this study we shall consider the estimation of using the least squares method and empirical distribution. Furthermore, an unbiased estimation is introduced for \( \frac{1}{\gamma_k} \).

**ESTIMATION OF \( \gamma \) USING EMPIRICAL DISTRIBUTION**

Let \( X_1, X_2, \ldots, X_n \) be a random sample of size \( n \) from a heavy tailed distribution. Let \( X_{(1)}, X_{(2)}, \ldots, X_{(n)} \) be the corresponding order statistics and \( F(x) \) be the appropriate empirical distribution function. Then

\[ E^*_\gamma(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i < x), \quad -\infty < x < \infty. \]

According to Glivenko theorem for large \( n \)

\[ p(X_i > x) = 1 - F(x) \]  

It's obvious that:

\[ F_n(X_{(i)}) = \frac{i-1}{n}, \quad i = 1, \ldots, n. \]  

Comparing 4, 6 and 7, we note that for sufficiently large \( i \), the equations

\[ C_\gamma = n - i + 1 \]

or

\[ \log(C_\gamma) - \gamma \log(X_{(i)}) = \log\left(\frac{n-i+1}{n}\right) \]  

are fulfilled.

It's natural to consider that Eq. 8a and b are fulfilled for all \( i \), that are greater or equal than some \( k \), so, for \( i \geq k \). Using Eq. 8a and b for \( i = k, \ldots, n \), we'll calculate estimation \( \hat{\gamma}_k \) of \( \gamma \) using condition

\[ \hat{\gamma}_k = \arg \min\left\{ \sum_{i=k}^{n} \left[ \log(C_{(i)}) - \gamma \log(X_{(i)}) - \log\left(\frac{n-i+1}{n}\right) \right]^2 \right\} \]  

or

\[ \hat{\gamma}_k = \arg \min\left\{ \sum_{i=k}^{n} \left[ \log(X_{(i)}) - \frac{1}{\gamma} \log\left(\frac{n-i+1}{n}\right) \right]^2 \right\} \]

Conditions Eq. 9a and b represent the two-parameter model of the least squares. The solution of Eq. 9a is the estimation:

\[ \hat{\gamma}_k = \frac{(n - k + 1) \sum_{i=k}^{n} \log(X_{(i)}) \log\left(\frac{n-i+1}{n}\right) - \sum_{i=k}^{n} \log(X_{(i)}) \sum_{j=k}^{n} \log\left(\frac{n-j+1}{n}\right)}{(n - k + 1) \sum_{i=k}^{n} \log^2(X_{(i)}) - \left( \sum_{i=k}^{n} \log(X_{(i)}) \right)^2} \]

and for (9b) the estimation is
\[ \hat{c}_k = \exp \left( \frac{1}{n-k+1} \sum_{i=k}^{n} \log \left( n-i+1 \right) \right) \] (11a)

\[ \text{RESULTS ANALYSIS} \]

In this research, we have estimated the parameter of heavy tailed distribution using empirical distribution. Results of these estimation are given in the Fig. 1a-c for Student distribution and the Fig. 2a-c for Pareto distributions.

We have made the most accurate estimations of parameter when \( \gamma \leq 2 \). For clearness there are graphics of \( \bar{\gamma} \) and \( \bar{\gamma} \pm S_\gamma \).

Where:

- \( S_\gamma \) - Deviation.

For Student distribution when \( \gamma > 2 \) we get more accurate results.

Fig. 1: \( \bar{\gamma} \) and \( \bar{\gamma} \pm S_\gamma \) for student distribution, where (a) \( \gamma = 0.5 \), (b) \( \gamma = 1 \) and (c) \( \gamma = 2 \)
We can also use our method not only for estimating parameter of Student and Pareto distributions but also for other distributions, which belong to the heavy tailed distributions class.

ACKNOWLEDGMENT

This research has been supported by the Research Institute for Fundamental Science, Tabriz, Iran. The authors would like to thank this support.

REFERENCES


