VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

ÚSTAV MATEMATIKY

FACULTY OF MECHANICAL ENGINEERING

INSTITUTE OF MATHEMATICS

STATISTICAL ANALYSIS OF SAMPLES FROM THE GENERALIZED EXPONENTIAL DISTRIBUTION

STATISTICKÁ ANALÝZA VÝBĚRŮ ZE ZOBECNĚNÉHO EXPONENCIÁLNÍHO ROZDĚLENÍ

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE Bc. HELENA PASCHKEOVÁ

AUTHOR

VEDOUCÍ PRÁCE doc. RNDr. JAROSLAV MICHÁLEK, CSc.

SUPERVISOR

BRNO 2013

Vysoké učení technické v Brně, Fakulta strojního inženýrství

Ústav matematiky
Akademický rok: 2013/2014

ZADÁNÍ DIPLOMOVÉ PRÁCE

student(ka): Bc. Helena Paschkeová

který/která studuje v **magisterském navazujícím studijním programu**

obor: **Matematické inženýrství (3901T021)**

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

Statistická analýza výběrů ze zobecněného exponenciálního rozdělení

v anglickém jazyce:

Statistical analysis of samples from the generalized exponential distribution

Stručná charakteristika problematiky úkolu:

Při statistické analýze spolehlivostních dat nebo dat popisujících životnosti se často pracuje s Weibullovým rozdělením. Je to proto, že toto rozdělení je velmi flexibilní a umožňuje kvalitně popsat rozdělení dob přežití v nejrůznějších technických i společenskovědních aplikacích. V roce 1999 navrhli Gupta a Kundu v článku [3] použít zobecněné exponenciální rozdělení jako alternativu k Weibullovu rozdělení. V článku [4] je pak uvedeno srovnání vlastností odhadů parametrů Weibullova rozdělení a zobecněného exponenciálního rozdělení, které byly získány metodou maximální věrohodnosti. Ukazuje se, že asymptotické vlastnosti obou odhadů jsou odlišné a je vhodné detailněji studovat odhady parametrů zobecněného exponenciálního rozdělení (viz [3]). Dále vzhledem k tomu, že pro některé vybrané typy cenzorování tyto odhady ještě nebyly studovány je vhodné zaměřit se na maximálně věrohodné odhady a jejich vlastnosti pro vybrané typy cenzorování.

Cíle diplomové práce:

V práci popište zobecněné exponenciální rozdělení, jeho charakteristiky a zaměřte se na maximálně věrohodné odhady jeho parametrů, které můžete popsat pomocí teorie uvedené v [1, 2]. Odvoďte asymptotické vlastnosti těchto odhadů pro necenzorovaná i cenzorovaná data. Výpočet odhadů algoritmizujte a proveďte počítačovou implementaci navržených algoritmů. Vlastnosti získaných odhadů ověřte pomocí simulací. Získané odhady použijte k analýze reálných environmentálních dat.

ABSTRACT

Thesis deals with generalized exponential distribution as an alternative distribution to Weibull and log-normal distributions. At first, properties of the generalized exponential distribution are presented, followed by the methods of parameter estimation. Separate chapter describes goodness of fit tests. Second part of the thesis deals with censored samples. Demonstrative examples of censoring on exponential distribution are presented. Moreover the type I left censored case on generalized exponential distribution, which has not been studied before, is elaborated at the end of the chapter. Simulations for this particular case of censoring are presented and studied in detail. EM algorithm is developed and its efficiency is compared to the maximum likelihood method. The derived theory is then applied on set of environmental data.

KEYWORDS

Generalized exponential distribution, maximum likelihood method, Fisher information matrix, censored samples, EM algorithm

ABSTRAKT

Diplomová práce se zabývá zobecněným exponenciálním rozdělením jako alternativou k Weibullovu a log-normálnímu rozdělení. Jsou popsány základní charakteristiky tohoto rozdělení a metody odhadu parametrů. Samostatná kapitola je věnována testům dobré shody. Druhá část práce se zabývá cenzorovanými výběry. Jsou uvedeny ukázkové příklady pro exponenciální rozdělení. Dále je studován případ cenzorování typu I zleva, který dosud nebyl publikován. Pro tento speciální případ jsou provedeny simulace s podrobným popisem vlastností a chování. Dále je pro toto rozdělení odvozen EM algoritmus a jeho efektivita je porovnána s metodou maximální věrohodnosti. Vypracovaná teorie je aplikována pro analýzu environmentálních dat.

KLÍČOVÁ SLOVA

Zobecněné exponenciální rozdělení, metoda maximální věrohodnosti, Fisherova informační matice, cenzorované výběry, EM algoritmus

PASCHKEOVÁ, Helena *Statistical analysis of samples from the generalized exponential distribution*: master's thesis. Brno: Brno University of Technology, Faculty of Mechanical Engineering, Department of Mathematics, 2014. 72 p. Supervised by Doc. RNDr. Jaroslav Michálek, CSc.

DECLARATION

I declare that I have written my master's thesis on the theme of "Statistical analysis of samples from the generalized exponential distribution" independently, under the guidance of the master's thesis supervisor and using the technical literature and other sources of information which are all quoted in the thesis and detailed in the list of literature at the end of the thesis.

As the author of the master's thesis I furthermore declare that, as regards the creation of this master's thesis, I have not infringed any copyright. In particular, I have not unlawfully encroached on anyone's personal and/or ownership rights and I am fully aware of the consequences in the case of breaking Regulation § 11 and the following of the Copyright Act No 121/2000 Sb., and of the rights related to intellectual property right and changes in some Acts (Intellectual Property Act) and formulated in later regulations, inclusive of the possible consequences resulting from the provisions of Criminal Act No 40/2009 Sb., Section 2, Head VI, Part 4.

Brno ....

(author's signature)

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my advisor Doc. RNDr. Jaroslav Michálek, CSc. for the continuous support of my master study and research, for his patience, motivation and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor. Moreover I would like to express my gratitude to RNDr. Marie Forbelská, Ph. D. for the support and valuable advice.

Brno . . . .

(author's signature)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Reliability theory is important for any design of product in any industry where we want to assure the quality of the product. So far the behavior of reliability was modeled by Weibull, Gamma or log-normal distributions. Professors Gupta and Kundu suggested in 1999 an alternative distribution that is appropriate for evaluation of reliability data. This distribution is the generalized exponential distribution (abbreviated as GE distribution).

This thesis gives the general overview of the properties of the GE distribution in chapter 2. The theory continues with basic methods of parameters estimation in chapter 3 and goodness of fit tests in chapter 4. Short chapter 5 gives the comparison among GE, Weibull and log-normal distributions and shows the pitfalls of discriminating among these distributions.

Chapter 6 sets the key theory for the censored data that are closely related to reliability theory. Fundamental definitions of censoring are introduced as well as examples of selected types of censoring. Chapter 6 moreover contains the analysis of type I left censoring, that has not been studied before, as an original work by the author. The type I left censoring is completely described in details as well as its asymptotic behavior. Simulations of this type of censoring are done at the end of the chapter.

Chapter 7 shows an alternative way of parameter estimation by EM algorithm in case of censored data. EM algorithm is adjusted for the GE distribution. The comparison of this approach and maximum likelihood estimation is done at the end of this chapter.

Chapter 8 shows the application of the theory and simulations on type I left censored data set. Data processed in this chapter was gathered as a part of project CzechGlobe (CZ.1.05/1.1.00/02.0073, CzechGlobe – Global Change Research Centre).

1 BASIC DEFINITIONS

At first some basic concepts of probability theory, that will be used in the following chapters, will be listed. The chapter emphasizes the important definitions of the topic, not the whole theory of statistics. Text will hold the notation and definitions used in [1], [14] and [10].

1.1 Probability space and random variable

Probability theory models random experiments so that we can draw inferences about them. The triple (Ω;$\mathcal{A}$;P) is called the probability space and it is the fundamental object of probability theory. A probability space is needed for each experiment that we wish to describe mathematically. The probability space is defined by its sample space Ω, a collection $\mathcal{A}$ of events, which is an σ-algebra, and a probability measure P. The properties of σ-algebra $\mathcal{A}$ and measure P are summarized in the following two definitions.

Definition 1.2 (σ-algebra).
Set $\mathcal{A}$ is a σ-algebra on Ω iff

- $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ (i.e. $\mathcal{A}$ is subset of the powerset of Ω)
- $\mathcal{A} \neq \emptyset$
- $M \in \mathcal{A} \Rightarrow \Omega \backslash M \in \mathcal{A}$
- $M_i \in \mathcal{A}, \forall i \in \mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty} M_i \in \mathcal{A}$

Definition 1.3 (Probability function). The probability function $P : \mathcal{A} \to \langle 0, 1\rangle$ is defined for all $M \in \mathcal{A}$ such that

1. $\forall M_i \in \mathcal{A}$ such that $M_i \cap M_j = \emptyset$ for $i \neq j$ and $i, j = 1, 2, \ldots$
   holds that
   $$P\left(\cup_{i=1}^{\infty} M_i\right) = \sum_{i=1}^{\infty} P(M_i)$$
2. $P(\Omega) = 1$,

For each subset $M \subseteq \mathcal{A}$, the number $P(M)$ is called the probability of the event M. P is also a measure on $\mathcal{A}$.

The system of Borel sets in $\mathbb{R}$ will be denoted as $\mathcal{B}$.

Definition 1.4 (Random variable). Given a random experiment with a sample space (Ω;$\mathcal{A}$;P). A measurable function $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ is called a random variable. Every Borel set $B \in \mathcal{B}$ can be assigned their preimage $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ and probability measure $\nu(B) = P\{X^{-1}(B)\}$. Measure ν is called the induced measure.

Definition 1.5 (Vector of random variables). Given a random experiment with the sample space $(\Omega; \mathcal{A}; \mathrm{P})$. Measurable function $\mathbb{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}^n)$, where $\mathcal{B}^n$ denotes the system of Borel sets in $\mathbb{R}^n$, is called a random vector $\mathbb{X} = (X_1, \ldots, X_n)$.

Definition 1.6 (Cumulative distribution function and probability density function). Let X be a random variable. Function $F(x) = P(\{\omega : X(\omega) \leq x\})$ is called cumulative distribution function (abbreviated c.d.f.) of the random variable X. If there exists function $f(w) \geq 0, w \in \mathbb{R}$ such that

$$F(x) = \int_{-\infty}^{x} f(w)\, dw, \quad \forall x \in \mathbb{R}$$

then X is said to have continuous distribution. Function $f(x)$ is called the probability density function (abbreviated as p.d.f.) of X.

Equivalently function $F(x_1, \ldots, x_n) = P(\cap_{i=1}^{n} \{\omega : X_i(\omega) \leq x_i\})$ is called c.d.f. of random vector. Moreover if there exists function $f(w_1, \ldots, w_n) \geq 0, (w_1, \ldots, w_n) \in \mathbb{R}^n$ such that

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(w_1, \ldots, w_n)\, dw_1 \ldots dw_n, \quad \forall x_1, \ldots, x_n \in \mathbb{R}$$

then random vector $\mathbb{X}$ is said to have continuous distribution. Function $f(x_1, \ldots, x_n)$ is called the probability density function of $\mathbb{X}$.

Definition 1.7 (Survival function). Let the random variable X have the c.d.f. $F(x)$. Function $S(x) = 1 - F(x)$ is called survival function (in some literature reliability function) of the random variable X.

Definition 1.8 (Hazard function). Let the random variable X have the p.d.f. $f(x)$ and survival function $S(x)$. Function $h(x) = \frac{f(x)}{S(x)}$ for $S(x) > 0$ is called hazard function (in some literature hazard rate) of the random variable X.

Survival and hazard functions are very important for the reliability theory. The survival function gives us information of the percentual survival. The hazard function is often interpreted as the probability of fail at time x under the condition that the observed sample survives till time x.

Definition 1.9 (Mathematical expectation). Let the random variable X have the p.d.f. $f(x)$ and let $Y = u(X)$ be a transformed random variable of X. Value $E(Y)$ such that

$$E(Y) = E(u(X)) = \int_{-\infty}^{\infty} u(x) f(x)\, dx$$

is called mathematical expectation (or expected value) of transformed random variable Y when the integral above is well defined.

There are several special transformed random variables and their mathematical expectations that will be used in the following chapters.

Definition 1.10 (Mathematical expectation for special transformed random variables). Let the random variable X have the p.d.f. $f(x)$. Let us define special mathematical expectations under the condition that the following integrals exist:

- n-th moment of random variable X: $E(X^n) = \int_{-\infty}^{\infty} x^n f(x)\,dx$, $n \in \mathbb{N}$
  In the rest of the thesis will be the n-th moment denoted as μ'_n.
- Expected value of random variable X: $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$
- n-th central moment of random variable X:
  $E((X - E(X))^n) = \int_{-\infty}^{\infty} (x - E(X))^n f(x)\,dx$, $n \in \mathbb{N}$
  In the rest of the thesis will be the n-th central moment denoted as μ_n.
- Variance of random variable X: $E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\,dx$
  In the rest of the thesis will be variance denoted as $Var(X)$.

Definition 1.11 (Coefficient of variation). Let the random variable X have finite variance and $E(X) \neq 0$. The coefficient of variation is defined as

$$CV = \frac{\sqrt{Var(X)}}{E(X)}$$

Definition 1.12 (Skewness and curtosis). Let the random variable X have finite fourth central moment μ_4 and $Var(X) \neq 0$. The skewness of the distribution γ_1 is defined as

$$\gamma_1 = \frac{\mu_3}{Var(X)^{\frac{3}{2}}}$$

The curtosis of the distribution γ_2 is defined as

$$\gamma_2 = \frac{\mu_4}{Var(X)^2}$$

Definition 1.13 (Mode). The mode of a continuous probability density is the value x at which its p.d.f. $f(x)$ has its maximum value.

Definition 1.14 (Quantile function and median). Let the random variable X have the c.d.f. $F(x)$ then function $Q(p)$ defined as

$$Q(p) = \inf\{x \in \mathbb{R}, p < F(x)\}, \quad p \in \langle 0, 1\rangle$$

is called the quantile function of the random variable X. Value $Q(0.5)$ is called median of the random variable X.

Definition 1.15 (Characteristic function). Let X be a random variable. Function $\phi_X(t)$ given by

$$\phi_X(t) = E(e^{itX}) = E(\cos(tX)) + iE(\sin(tX)), \quad t \in \mathbb{R}$$

is called characteristic function of X.

Theorem 1.15.1. [1]
Let $\phi_X(t)$ be a characteristic function of X and let the k-th derivation of $\phi_X(t)$ be well defined. Then the following equality holds

$$\phi_X(0)^{(k)} = i^k E(X^k), \quad \forall k \in \mathbb{N}$$

Definition 1.16 (Independence of two random variables). Let X and Y be the random variables with corresponding c.d.f. $F_X(x)$ and $F_Y(y)$. Random variables X and Y are said to be independent if and only if the c.d.f. of joint distribution of random vector (X, Y) has the property that

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y), \quad \forall x, y \in \mathbb{R}$$

Equivalently for their corresponding p.d.f. $f_X(x)$ and $f_Y(y)$ and joint p.d.f. $f_{(X,Y)}(x, y)$ holds that

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y), \quad \forall x, y \in \mathbb{R}$$

Definition 1.17 (Independence of n random variables). Let $\mathbb{X}$ be a vector $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ of random variables. Random variables $X_1, X_2, \ldots, X_n$ are independent if and only if joint distribution of this vector $F_{\mathbb{X}}$ holds

$$F_{\mathbb{X}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall x_1, x_2, \ldots, x_n \in \mathbb{R}^n$$

Independence is a very important property and assumption in many theorems in statistics. Random vector $\mathbb{X} = (X_1, \ldots, X_n)$ of independent and identically distributed variables X_i with the c.d.f. F_X (abbreviated as i.i.d.) will be called the random sample from distribution with c.d.f. F_X.

Definition 1.18 (Covariance and Correlation). Let X and Y be the random variables then

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

is called covariance $Cov(X, Y)$ of the random variables X and Y if all the expectations are well defined. Moreover if $Var(X) \neq 0$ and $Var(Y) \neq 0$ then

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

is called the correlation $Cor(X, Y)$ of the random variables X and Y if all the expectations are well defined.

Theorem 1.18.1. [1]
Let X and Y be the random variables. If X and Y are independent then

$$Cov(X, Y) = 0.$$

Definition 1.19 (Ordered random sample). Let $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ be random sample from distribution with c.d.f. F_X. Let $X_{(1)}(\omega) = \min(X_1(\omega), \ldots, X_n(\omega))$, $X_{(2)}(\omega)$ be the second smallest value from $(X_1(\omega), \ldots, X_n(\omega))$ and so on till $X_{(n)}(\omega) = \max(X_1(\omega), \ldots, X_n(\omega))$.

$$X_{(1)}(\omega) \le X_{(2)}(\omega) \le \ldots \le X_{(n)}(\omega), \quad \forall \omega \in \Omega$$

Then random vector $\mathbb{X}_{()} = (X_{(1)}(\omega), \ldots, X_{(n)}(\omega))$ for $\omega \in \Omega$ is called ordered random sample. Random variable $X_{(i)}$ is called the i-th ordered random variable.

Theorem 1.19.1. [21]
Let $\mathbb{X}$ be i.i.d. vector of random variables with c.d.f. $F_{\mathbb{X}}$ of the size n and let each X_i have p.d.f. f_X and c.d.f. F_X. The p.d.f. of the ordered random variable $X_{(i)}$ is:

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!\,(n-i)!} \left(F_X(x)\right)^{i-1} \left(1 - F_X(x)\right)^{n-i} f(x), \quad i = 1, \ldots, n$$

Definition 1.20 (Empirical distribution function). Given an random sample $\mathbb{X}$ of size n. The empirical distribution function $\widehat{F}(x)$ is defined as

$$\widehat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[X_i \le x]}$$

where I_A is so called indicator random event which is defined to be equal to 1 if property A holds and 0 otherwise.

1.2 Basic statistic methods

This section will focus primary on method of maximum likelihood estimation. This method is used in the Chapter 3 as well as the Chapter 6, where is effectively used for advanced problems of estimation theory. Notation and definitions are taken from [1], theorems are taken from [15]. Other methods are unfortunately not listed because of the limited range of pages of this thesis. For this reason has every method a short introduction in the corresponding part of this thesis.

Definition 1.21 (System of regular p.d.f.). Let $\Theta \subset \mathbb{R}^m$ be an open Borel set. System $\mathcal{F}_{reg} = \{f(\mathbb{x}, \theta) : \theta \in \Theta, \mathbb{x} \in \mathbb{R}^n\}$ of p.d.f. depending on parameter θ is said to be regular if:

- set $M = \{\mathbb{x} : \mathbb{x} \in \mathbb{R}^n, f(\mathbb{x}, \theta) > 0\}$ is independent of θ
- there exists vector of finite partial derivations

$$\frac{\partial f(\mathbb{x}, \theta)}{\partial \theta_i}, \quad \theta \in \Theta, \quad \forall x \in M, \quad \forall i = 1, \dots, m$$

- for all $\theta = (\theta_1, \theta_2, \dots, \theta_m) \in \Theta$ holds

$$\int_M \frac{\frac{\partial f(\mathbb{x},\theta)}{\partial \theta_i}}{f(\mathbb{x}, \theta)}\, dF(\mathbb{x}, \theta) = \int_M \frac{\partial \ln f(\mathbb{x}, \theta)}{\partial \theta_i}\, dF(\mathbb{x}, \theta) = 0 \qquad i = 1, \dots, m$$

  where $F(\mathbb{x}, \theta)$ is the corresponding c.d.f.
  (i.e. derivation can be interchanged with integration)
- for all $\theta = (\theta_1, \theta_2, \dots, \theta_m) \in \Theta$ is

$$\mathcal{I}_{ij}(\theta) = \int_M \frac{\partial \ln f(\mathbb{x}, \theta)}{\partial \theta_i} \frac{\partial \ln f(\mathbb{x}, \theta)}{\partial \theta_j}\, dF(\mathbb{x}, \theta) \qquad i, j = 1, \dots, m$$

  is finite and matrix $\mathcal{I} = \mathcal{I}(\theta) = (\mathcal{I}_{ij}(\theta))_{i,j=1}^m$ is positive definite matrix. Matrix $\mathcal{I}$ is called Fisher information matrix.

Definition 1.22 (Score vector). Let $f(x, \theta) \in \mathcal{F}_{reg}$. The random vector

$$U(\theta) = (U_1(\theta), \dots, U_m(\theta))$$

where $U_i(\theta) = \frac{\partial \ln f(X,\theta)}{\partial \theta_i}$ is called the score vector (or the score) with respect to the p.d.f. $f(x, \theta)$.

Theorem 1.22.1. [15]
Let $f(x, \theta) \in \mathcal{F}_{reg}$. If $f''_{ij}(x, \theta) = \frac{\partial^2 f(X,\theta)}{\partial \theta_i \partial \theta_j}$ then

$$E(U_i(\theta)) = 0, \quad Var(U(\theta)) = \mathcal{I}(\theta), \quad \forall i = 1, \dots, m$$

Moreover if $E\left(\frac{f''_{ij}(X,\theta)}{f(X,\theta)}\right) = 0$ then

$$\mathcal{I}(\theta) = -E(U'(\theta)) = -E \begin{pmatrix} \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_1^2} & \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_1 \partial \theta_m} \\ \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_2^2} & \cdots & \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_2 \partial \theta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_m \partial \theta_1} & \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_m \partial \theta_2} & \cdots & \frac{\partial^2 \ln f(X,\theta)}{\partial \theta_m^2} \end{pmatrix}$$

THEOREM 1.22.2. Let $\mathbb{X} = (X_1, \dots, X_n)$ be an i.i.d. random vector with p.d.f. $f(x, \theta) \in \mathcal{F}_{reg}$. Then the score vector of $\mathbb{X}$ is

$$U_n^{\mathbb{X}}(\theta) = \sum_{j=1}^{n} U(X_j, \theta)$$

Fisher information matrix of $\mathbb{X}$ is:

$$\mathcal{I}_n(\theta) = n\mathcal{I}(\theta)$$

where $\mathcal{I}$ is the Fisher information matrix defined above.

In the following text let us denote $l(\mathbb{x}, \theta) = \prod_{i=1}^{n} f_i(x_i, \theta) = f_{\mathbb{X}}(\mathbb{x}, \theta)$ the probability density of i.i.d. vector $\mathbb{X}$ considered as a function of parameter θ for given x. Function $l(\mathbb{x}, \theta)$ is called the likelihood function. Moreover $L(\mathbb{x}, \theta)$ is called the log-likelihood function which is the logarithm of likelihood function

$$L(\mathbb{x}, \theta) = \ln l(\mathbb{x}, \theta)$$

where $f_i(x, \theta) \in \mathcal{F}_{reg}, \forall i = 1, \dots, n$ and $x_i \in M, \forall i = 1, \dots, n$.

THEOREM 1.22.3. Let $\mathbb{X} = (X_1, \dots, X_n)$ be an i.i.d. random vector with p.d.f. $f(x, \theta) \in \mathcal{F}_{reg}$. If $f''_{ij}(x, \theta) = \frac{\partial^2 f(x,\theta)}{\partial\theta_i \partial\theta_j}$ then

$$E(U_n^{\mathbb{X}}(\theta)) = 0, \quad Var(U_n^{\mathbb{X}}(\theta)) = n\mathcal{I}(\theta)$$

Moreover if $E\left(\frac{f''_{ij}(X,\theta)}{f(X,\theta)}\right) = 0$ then

$$\mathcal{I}(\theta) = -\frac{1}{n}E(U_n^{\mathbb{X}\prime}(\theta)) = -\frac{1}{n}E\begin{pmatrix} \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_1^2} & \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_1\partial\theta_m} \\ \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_2^2} & \cdots & \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_2\partial\theta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_m\partial\theta_1} & \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_m\partial\theta_2} & \cdots & \frac{\partial^2 L(\mathbb{X},\theta)}{\partial\theta_m^2} \end{pmatrix}$$

THEOREM 1.22.4. Let $\mathbb{X} = (X_1, \dots, X_n)$ be a random sample with p.d.f. $f(x, \theta) \in \mathcal{F}_{reg}$. If for all $x \in M, \theta \in \Theta$ and $i, j = 1, \dots, m$ exists second partial derivations of $f(x, \theta)$ then

$$\frac{1}{\sqrt{n}} U_n^{\mathbb{X}}(\theta) \overset{A}{\sim} N_m(0, \mathcal{I}(\theta))$$

where $\overset{A}{\sim}$ means that as sample size $n \to \infty$ the distribution of the statistic approaches a normal distribution.

THEOREM 1.22.5. Let $\mathbb{X} = (X_1, \ldots, X_n)$ be an i.i.d. random vector with p.d.f. $f(x, \theta_0) \in \mathcal{F}_{reg}$. Let $M = \{x \in \mathbb{R}, f(x, \theta_0) > 0\}$. Let $\forall x \in M, \theta \in \Theta$ and $i, j = 1, \ldots, m$ exist second partial derivations of $f(x, \theta)$ and let $E\left(\frac{f''_{ij}(Y,\theta)}{f(X,\theta)}\right) = 0$ be true. Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{A}{\sim} N_m(0, \mathcal{I}(\theta_0)^{-1})$$

or equivalently

$$\hat{\theta}_n \overset{A}{\sim} N_m\left(\theta_0, \frac{\mathcal{I}(\theta_0)^{-1}}{n}\right)$$

where $\hat{\theta}_n$ is the maximum likelihood estimator.

1.23 Special functions

This short section will introduce two important functions that will be used in the following text.

Gamma function: Gamma function Γ is defined as:

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x}\, dx$$

for all t for which integral converges.

Digamma function: Digamma function ψ is defined as:

$$\psi(t) = \frac{\Gamma'(t)}{\Gamma(t)} = \int_0^\infty \left(\frac{e^x}{x} - \frac{e^{-xt}}{1 - e^{-x}}\right) dx$$

for all t for which integral converges.

2 GENERALIZED EXPONENTIAL DISTRIBUTION

The two-parametrized general exponential distribution (GE distribution) was introduced by authors Gupta and Kundu in 1999 in article [7] as a generalization of exponential distribution. This distribution is from exponential family of distributions [1] and can be used in the description of natural phenomena. Since then many papers were published by these authors describing the properties of this distribution a comparisons with other similar distributions (mainly log-normal and Weibull distributions).[8],[9] This chapter will summarize all the important properties of this distribution.

2.1 Distribution and density function of the GE distribution

The GE distribution is characterized by shape parameter α and by the scale parameter λ. The GE distribution has the c.d.f.:

$$F_{GE}(x; \alpha, \lambda) = (1 - e^{-\lambda x})^{\alpha}, \qquad \alpha, \lambda > 0, \quad x \geq 0$$

and p.d.f.:

$$f_{GE}(x; \alpha, \lambda) = \alpha\lambda(1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}, \qquad \alpha, \lambda > 0, \quad x \geq 0$$

The p.d.f. for different values α and λ can be seen on the figures 2.1 and 2.2. Figure 2.1 demonstrates the change of shape of the p.d.f. when the shape parameter α changes. Figure 2.2 demonstrates the change of scale of the p.d.f. when the scale parameter λ changes. For parameter $\alpha = 1$ the distribution passes to exponential distribution. We can see that for $\alpha \leq 1$, it is decreasing function and for $\alpha > 1$, it is unimodal, skewed, right tailed density function. It is observed that even for very large shape parameter it is not symmetric.[10]

The potential application of this distribution comes right from its p.d.f. because density is different from zero only for $x \geq 0$. In most cases the variable x represents time and the value $\int_x^{x+\epsilon} f_{GE}(t; \alpha, \lambda)\, dt$ probability that some natural phenomena occurs in time interval $(x, x + \epsilon)$.

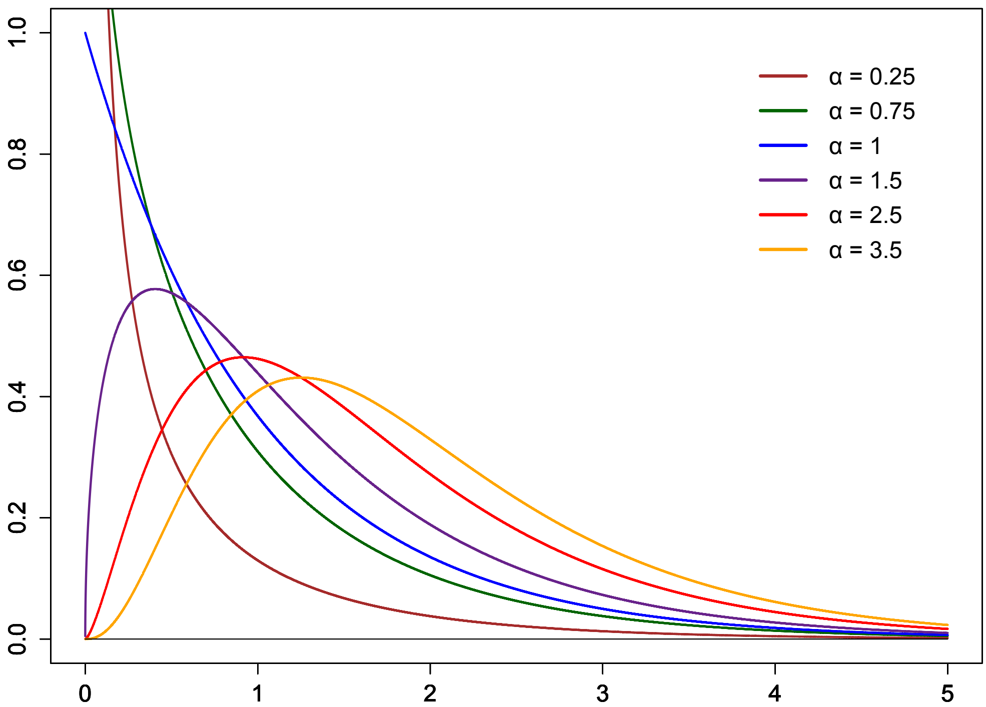

Fig. 2.1: Density functions of the GE distribution for different values of α when $\lambda = 1$

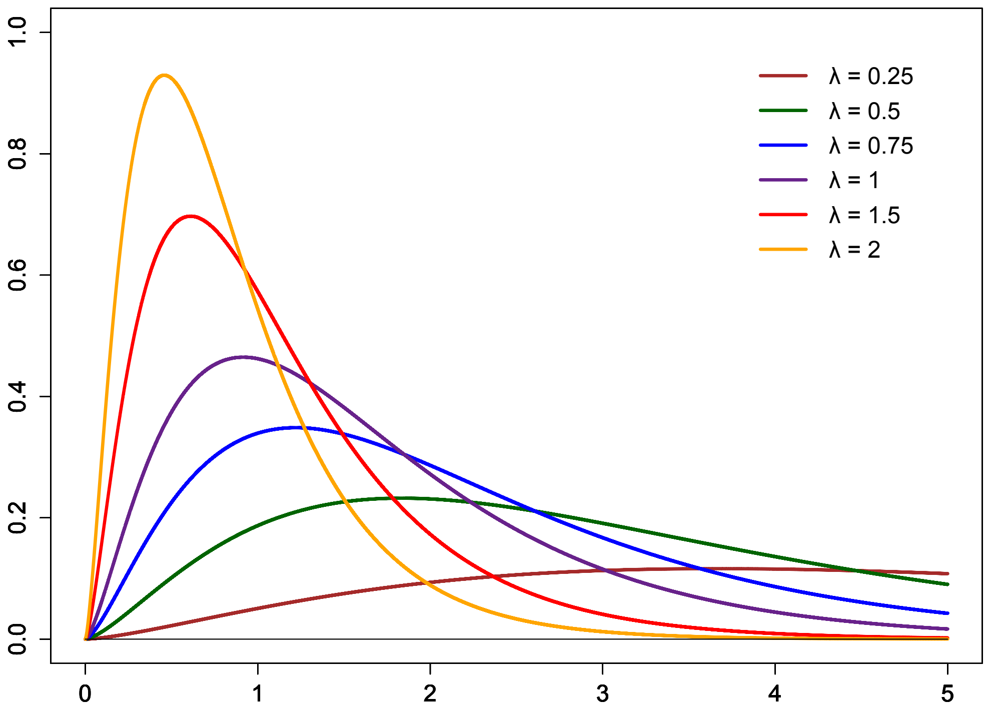

Fig. 2.2: Density functions of the GE distribution for different values of λ when $\alpha = 2.5$

2.2 Survival and hazard function of the GE distribution

Basic functions used in reliability theory are the survival and the hazard function. The survival function of GE distribution is:

$$S_{GE}(x;\alpha,\lambda) = 1 - (1 - e^{-\lambda x})^{\alpha}, \qquad \alpha, \lambda > 0, \quad x \geq 0$$

and the hazard function is:

$$h_{GE}(x;\alpha,\lambda) = \frac{\alpha\lambda(1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}}{1 - (1 - e^{-\lambda x})^{\alpha}}, \qquad \alpha, \lambda > 0, \quad x \geq 0$$

For $\alpha = 1$ (i.e. exponential distribution) the hazard function simplifies to constant function equal to λ. For $\alpha \neq 1$ is the hazard function a monotone function which is approaching the λ value in ∞ as it is shown in theorem 2.2.1. The behavior of the hazard function can be observed in figure 2.3.

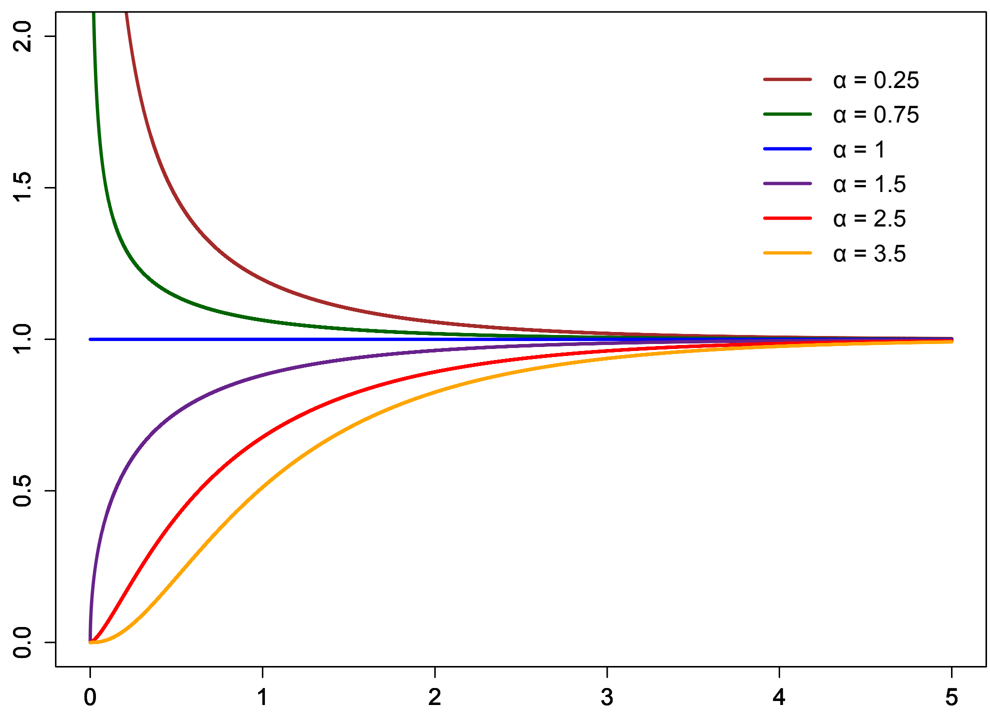

Fig. 2.3: Hazard function of the GE distribution for different values of α when $\lambda = 1$

THEOREM 2.2.1. Let $h_{GE}(x;\alpha,\lambda)$ be hazard function of GE distribution. Then

$$\lim_{x\to\infty} h_{GE}(x;\alpha,\lambda) = \lambda$$

Proof [Convergence of the hazard function]
We use L'Hospital rule repeatedly and simplify the expressions.

$$\begin{aligned}
\lim_{x\to\infty} h_{GE}(x;\alpha,\lambda) &= \lim_{x\to\infty} \frac{\alpha\lambda(1-e^{-\lambda x})^{\alpha-1}e^{-\lambda x}}{1-(1-e^{-\lambda x})^{\alpha}} \\
&\overset{L'H.}{=} \lim_{x\to\infty} \frac{\frac{\alpha\lambda^2\left(1-e^{-\lambda x}\right)^{\alpha-1}(\alpha-1)\left(e^{-\lambda x}\right)^2}{1-e^{-\lambda x}} - \alpha\lambda^2\left(1-e^{-\lambda x}\right)^{\alpha-1}e^{-\lambda x}}{-\frac{\left(1-e^{-\lambda x}\right)^{\alpha}\alpha\lambda e^{-\lambda x}}{1-e^{-\lambda x}}} \\
&= \lim_{x\to\infty} \frac{e^{\lambda x}\left(-1+e^{-\lambda x}\right)\lambda\left(\alpha-e^{\lambda x}\right)}{\left(e^{\lambda x}-1\right)^2} \\
&\overset{L'H.}{=} \lim_{x\to\infty} \frac{\lambda^2 e^{\lambda x}\left(-1+e^{-\lambda x}\right)\left(\alpha-e^{\lambda x}\right)-\lambda^2\left(a-e^{\lambda x}\right)-\left(e^{\lambda x}\right)^2\left(-1+e^{-\lambda x}\right)\lambda^2}{2\left(e^{\lambda x}-1\right)\lambda e^{\lambda x}} \\
&= -\frac{1}{2}\lim_{x\to\infty} \frac{\lambda\left(a-2\,e^{\lambda x}+1\right)}{e^{\lambda x}-1} \\
&\overset{L'H.}{=} -\frac{1}{2}\lim_{x\to\infty} \frac{-2\,\lambda^2 e^{\lambda x}}{\lambda e^{\lambda x}} \\
&= -\frac{1}{2}(-2\lambda) \\
&= \lambda \qquad \square
\end{aligned}$$

The λ parameter thus denotes the failure rate when time is approaching infinity. This theorem has the key role in understanding the significance of both parameters α and λ. This can be seen and compared on the figure 2.1 of the p.d.f. with fixed λ parameter and the figure 2.3 of the hazard function.

2.3 Mode and median of the GE distribution

Mode and median are the robust characteristics of the GE distribution. For $\alpha \leq 1$ is mode at zero value (see figure 2.1). For $\alpha > 1$ is mode equal to the $\frac{\ln(\alpha)}{\lambda}$ (see the peak in the figure 2.1).[10]

$$\text{Mode}_{GE}(\alpha,\lambda) = \begin{cases} 0, & 0 < \alpha \leq 1 \\ \frac{\ln(\alpha)}{\lambda}, & \alpha > 1 \end{cases}$$

Quantile function for the GE distribution is:

$$Q_{GE}(p;\alpha,\lambda) = -\frac{\ln\left(1-\sqrt[\alpha]{p}\right)}{\lambda}, \quad 0 \leq p \leq 1$$

Median is value that separates data in half.

$$\text{Median}_{GE}(\alpha,\lambda) = -\frac{\ln\left(1-\sqrt[\alpha]{0.5}\right)}{\lambda}$$

2.4 Moments of the GE distribution

GE distribution can be equivalently fully described by its characteristic function. All moments of the distribution can be derived from the characteristic function. The most important moments of every distribution are its expected value and variance. The characteristic function of $GE(\alpha, \lambda)$ is:

$$\phi_{GE}(t;\alpha,\lambda) = Ee^{itX} = \frac{\Gamma(\alpha+1)\,\Gamma\left(1-\frac{it}{\lambda}\right)}{\Gamma\left(\alpha-\frac{it}{\lambda}+1\right)}[10]$$

From the property of this function derived in the section 1.1, expectation value and variance can be obtained.

$$E(X) = \frac{1}{\lambda}[\psi(\alpha+1) - \psi(1)]$$

$$Var(X) = \frac{1}{\lambda^2}[\psi'(1) - \psi'(\alpha+1)]$$

Where ψ denotes the digamma function and Γ denotes the gamma function. Both functions are described in section 1.23. From the properties of digamma function we can derive the properties of $E(X)$ for given parameters. For fixed λ and $\alpha \to \infty$ the expected value tends to ∞, the variance also increases up to value $\frac{\pi^2}{6\lambda}$. For fixed α and $\lambda \to \infty$ the expected value and variance go to 0 but with different rate. [1] [10]

2.5 Skewness and curtosis of the GE distribution

Skewness γ_1 and curtosis γ_2 are important parameters of the GE distribution. As it was referred in Chapter 1.1 they are given by the second (μ_2'), third (μ_3') and fourth (μ_4') moment that can be computed by the characteristic function $\phi_{GE(\alpha,\lambda)}(x)$.

$$\gamma_1 = \frac{\mu_3' - 3\mu_2' E(X) + 2E(X)^3}{Var(X)^{\frac{3}{2}}}, \quad \gamma_2 = \frac{\mu_4' - 4\mu_3' E(X) + 6\mu_2' E(X)^2 - 3E(X)^4}{Var(X)^2}$$

where

$$\mu_2' = \frac{1}{\lambda^2}\left[\psi'(1) - \psi'(\alpha+1) + [\psi(\alpha+1) - \psi(1)]^2\right]$$

$$\mu_3' = \frac{1}{\lambda^3}\left[\psi''(\alpha+1) - \psi''(1) + 3[\psi(\alpha+1) - \psi(1)][\psi'(1) - \psi'(\alpha+1)] + [\psi(\alpha+1) - \psi(1)]^3\right]$$

$$\mu_4' = \tfrac{1}{\lambda^4}\,[\psi'''(1) - \psi'''(\alpha+1) + 3[\psi'(1) - \psi(\alpha+1)]^2 + 4[\psi(\alpha+1) - \psi(1)][\psi''(\alpha+1) - \psi''(1)] + 6[\psi(\alpha+1) - \psi(1)]^2[\psi'(1) - \psi(\alpha+1)] + [\psi'''(1) - \psi'''(\alpha+1)]^4)]$$

Skewness and curtosis are independent of the scale parameter λ. It is numerically observed that kurtosis and skewness are both decreasing functions with respect to α. Moreover the limiting value of skewness is approximately 1.139547. [10]

[1] This feature is different from gamma or Weibull distribution. In case of gamma distribution, the variance tends to infinity as the shape parameter increases. In case of Weibull distribution the variance is approximately $\frac{\pi^2}{6\theta\tau^2}$ for large values of the shape parameter τ.[10]

2.6 Ordered sample properties of the GE distribution

Ordered sample from GE distribution is important for several methods of parameters estimation as well as for problems where censoring on the data occurs (see chapter 6). The key is to understand the behavior of the i-th ordered random variable $X_{(i)}$ because generally $X_{(i)}$ does not have the GE distrubution. Using the theorem 1.19.1 in Chapter 1.1, the i-th ordered random variable $X_{(i)}$ has the p.d.f. function:

$$f_{X_{(i)}}(x) = \frac{n!\,\alpha\lambda}{(i-1)!\,(n-i)!}\left(1-e^{-\lambda x}\right)^{i\alpha-1}\left[1-\left(1-e^{-\lambda x}\right)^{\alpha}\right]^{n-i} e^{-\lambda x}, \text{ [11]}$$

The illustration of the p.d.f. can be seen on the Figure 2.4 for the case when $n = 5$. The two special cases that should be emphasized are for $i = 1$ and $i = n$.

$$X_{(1)} = \min\left\{X_1, X_2, \ldots, X_n\right\}$$

$$X_{(n)} = \max\left\{X_1, X_2, \ldots, X_n\right\}$$

The corresponding p.d.f. are:

$$f_{X_{(1)}}(x) = n\alpha\lambda\left[1-\left(1-e^{-\lambda x}\right)^{\alpha}\right]^{n-1}\left(1-e^{-\lambda x}\right)^{\alpha-1} e^{-\lambda x}$$

$$f_{X_{(n)}}(x) = n\alpha\lambda(1-e^{-\lambda x})^{\alpha n-1}e^{-\lambda x}$$

The transformed random variable $X_{(n)}$ has also a GE distribution with shape parameter $n\alpha$ and scale parameter λ.

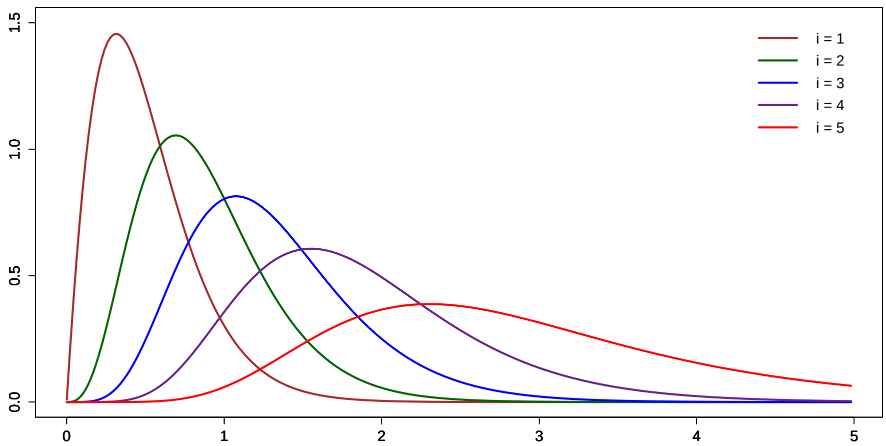

Fig. 2.4: P.d.f. of $X_{(i)}$ under assumption that X_i are i.i.d. taken from GE(2,1), $n = 5$

3 ESTIMATION OF GE DISTRIBUTION PARAMETERS

3.1 Estimation by maximum likelihood estimators

Method of maximum likelihood is a very popular method of parameter estimation of parameters of the distributions for a given random sample $\mathbb{x}$. Theory was introduced in section 1.2. Let the random variables X_i be from GE distribution with p.d.f.

$$f_{GE}(x;\alpha,\lambda) = \alpha\lambda(1-e^{-\lambda x})^{\alpha-1}e^{-\lambda x}, \qquad \alpha,\lambda > 0.$$

The likelihood function is

$$l(\mathbb{x};\alpha,\lambda) = \alpha^n\lambda^n\prod_{i=1}^{n}(1-e^{-\lambda x_i})^{\alpha-1}e^{-\lambda x_i}$$

In order to find the maximum likelihood estimators, function $l(\mathbb{x};\alpha,\lambda)$ for given $\mathbb{x}$ must be maximized. This maximizing values of $\hat{\alpha}$ and $\hat{\lambda}$ are maximum likelihood estimates of α and λ. Since logarithm is an injective function, the l function is maximized for the same values as $L = \ln l$ (i.e. the log-likelihood function). Then:

$$L(\mathbb{x};\alpha,\lambda) = n\ln\alpha + n\ln\lambda + (\alpha-1)\sum_{i=1}^{n}\ln(1-e^{-\lambda x_i}) - \lambda\sum_{i=1}^{n}x_i$$

So we have:

$$\frac{\partial L(\mathbb{x};\alpha,\lambda)}{\partial\alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n}\ln(1-e^{-\lambda x_i})$$

$$\frac{\partial L(\mathbb{x};\alpha,\lambda)}{\partial\lambda} = \frac{n}{\lambda} + (1-\alpha)\sum_{i=1}^{n}\frac{x_i e^{-\lambda x_i}}{1-e^{-\lambda x_i}} - \sum_{i=1}^{n}x_i$$

The maximum likelihood estimator of α is obtained from the first equation as the function $\hat{\alpha}(\lambda)$:

$$\hat{\alpha}(\lambda) = \frac{-n}{\sum_{i=0}^{n}\ln(1-e^{-\lambda x_i})}$$

This function is dependent on λ parameter only thus we need to estimate the λ parameter first before estimating the α parameter.

The maximum likelihood estimator (MLE) of λ is obtained from the second equation by solving the equation $g(\lambda) = \lambda$ where:

$$g(\lambda) = n\left[(\alpha-1)\sum_{i=1}^{n}\frac{x_i e^{-\lambda x_i}}{1-e^{-\lambda x_i}} + \sum_{i=1}^{n}x_i\right]^{-1}$$

In this equation is α estimated by $\hat{\alpha}(\lambda)$ and $g(\lambda)$ is then dependent only on λ. It is observed in article [11] that $g(\lambda)$ is an unimodal function. Its maximum can be obtained by an iterative procedure.

The maximum likelihood estimators $\hat{\alpha}$ and $\hat{\lambda}$ are consequently found first by solving $g(\lambda) = \lambda$ equation iteratively and obtaining $\hat{\lambda}$ and then by substituting $\hat{\lambda}$ to $\hat{\alpha}(\lambda)$ and obtaining the $\hat{\alpha}$ estimator.

In case that parameter α is known from the beginning, the equation $g(\lambda) = \lambda$ simplifies in the α term. In case that parameter λ is known, the estimate of the shape parameter α is obtained by $\hat{\alpha} = \hat{\alpha}(\lambda)$.

Fisher Information Matrix

More information on estimates of α and λ can be gained from the Fisher information matrix by exploring the second partial derivations of the likelihood function. Fisher information matrix of the GE distribution can be written as:

$$\mathcal{I}(\alpha, \lambda) = -\frac{1}{n} \begin{pmatrix} A_{\alpha\alpha} & A_{\alpha\lambda} \\ A_{\lambda\alpha} & A_{\lambda\lambda} \end{pmatrix}, [18]$$

where the elements of the matrix are the partial derivations of the log-likelihood function (as it was presented in chapter 1.2):

$$A_{\alpha\alpha} = E\left(\frac{\partial^2 L(\mathbb{X}; \alpha, \lambda)}{\partial \alpha^2}\right) = \frac{-n}{\alpha^2}$$

$$A_{\alpha\lambda} = A_{\lambda\alpha} = E\left(\frac{\partial^2 L(\mathbb{X}; \alpha, \lambda)}{\partial \alpha \partial \lambda}\right) = E\left(\frac{\partial^2 L(\mathbb{X}; \alpha, \lambda)}{\partial \lambda \partial \alpha}\right) = E\left(\sum_{i=1}^{n} \frac{X_i e^{-\lambda X_i}}{1 - e^{-\lambda X_i}}\right)$$

$$A_{\lambda\lambda} = E\left(\frac{\partial^2 L(\mathbb{X}; \alpha, \lambda)}{\partial \lambda^2}\right) = -E\left(\frac{n}{\lambda^2} + (\alpha - 1)\sum_{i=1}^{n} \frac{X_i^2 e^{-\lambda X_i}}{(1 - e^{-\lambda X_i})^2}\right)$$

The expectations are computed in [12]. For clarity, only the outcomes from the computations are presented.

$$A_{\alpha\lambda} = A_{\lambda\alpha} = \begin{cases} -\frac{n}{\lambda}\left[\frac{\alpha}{\alpha-1}\left[\psi(\alpha) - \psi(1)\right] - \psi(\alpha+1) + \psi(1)\right], & \alpha \neq 1 \\ \frac{n}{\lambda}\sum_{i=0}^{\infty} \frac{1}{(2+i)^2} \approx \frac{0.645n}{\lambda}, & \alpha = 1 \end{cases}$$

$$A_{\lambda\lambda} = \begin{cases} -\frac{n}{\lambda^2}\left[1 + \frac{\alpha(\alpha-1)}{\alpha-2}\left[(\psi'(1) - \psi'(\alpha-1) + \left[\psi(\alpha-1) - \psi(1)\right]^2\right]\right] + \\ \frac{n\alpha}{\lambda}\left[\psi'(1) - \psi(\alpha) + \left[\psi(\alpha) - \psi(1)\right]^2\right], & \alpha \neq 2 \\ \\ -\frac{n}{\lambda^2} - \frac{4n}{\lambda^2}\sum_{i=0}^{\infty} \frac{1}{(2+i)^3} \approx -\frac{1.308n}{\lambda^2}, & \alpha = 2 \end{cases}$$

In case that parameters α and λ are not known but estimated from the given sample and expectations are not taken, we are speaking about *observed Fischer information matrix* $\widetilde{\mathcal{I}}(\hat{\alpha}, \hat{\lambda})$.

$$\widetilde{\mathcal{I}}(\hat{\alpha}, \hat{\lambda}) = \begin{pmatrix} \frac{1}{\hat{\alpha}^2} & A_1 \\ A_1 & \frac{1}{\hat{\lambda}^2} + A_2 \end{pmatrix}$$

where

$$A_1 = -\frac{1}{n} \sum_{i=1}^{n} \frac{x_i e^{-\hat{\lambda} x_i}}{1 - e^{-\hat{\lambda} x_i}}$$

$$A_2 = \frac{\hat{\alpha} - 1}{n} \sum_{i=1}^{n} \frac{x_i^2 e^{-\hat{\lambda} x_i}}{\left(1 - e^{-\hat{\lambda} x_i}\right)^2}$$

Fisher information matrix is then the expected value of $\widetilde{\mathcal{I}}(\hat{\alpha}, \hat{\lambda})$.

$$\mathcal{I}(\alpha, \lambda) = E(\widetilde{\mathcal{I}}(\hat{\alpha}, \hat{\lambda}))$$

3.2 Estimation by method of moment estimators

Estimation by moment method uses the expectation value and variance of the distribution computed in section 2.4. This method involves equating sample moments with theoretical moments. The coefficient of variation (CV) mentioned in chapter 1.1 is used. For GE distribution the CV excludes the λ parameter and depends only on α.

$$CV = \frac{\sqrt{Var(\mathbb{X})}}{E(\mathbb{X})} = \frac{\sqrt{\psi'(1) - \psi'(\alpha + 1)}}{\psi(\alpha + 1) - \psi(1)}$$

Corresponding sample moments S^2 and $\bar{X}$ are computed in order to make an estimation of CV.

$$\frac{\sqrt{S^2}}{\bar{X}} = \frac{\sqrt{\psi'(1) - \psi'(\hat{\alpha} + 1)}}{\psi(\hat{\alpha} + 1) - \psi(1)}$$

Equation can be solved iteratively in order to obtain the moment estimator $\hat{\alpha}$. For an efficient guess of the starting value of $\hat{\alpha}^{(1)}$, tables in the article [11] can be used. The moment estimator $\hat{\lambda}$ is then directly obtained from the following equation

$$\hat{\lambda} = \bar{X}\left[\psi(\hat{\alpha} + 1) - \psi(1)\right]$$

which is derived from the expression of the expectation value of the GE distribution that was presented in Section 2.4.

3.3 Least squares estimators and weighted least squares estimators

The method of least squares is about estimating parameters by minimizing the squared discrepancies between observed data and their expected values. Suppose that $\mathbb{X}$ is i.i.d. vector of size n taken from the GE distribution. Then an ordered sample $\mathbb{X}_{(i)}$ can be defined. The density of $X_{(i)}$ is given by formula in section 2.6. The least squares method uses the distribution function $F(X_{(i)})$. For a sample size n, we have

$$E(F_{GE}(X_{(i)})) = \frac{i}{n+1}, \qquad Var(F_{GE}(X_{(i)})) = \frac{j(n-i+1)}{(n+1)^2(n+2)},$$

$$Cov(F_{GE}(X_{(i)}), F_{GE}(X_{(j)})) = \frac{i(n-j+1)}{(n+1)^2(n+2)} \quad \text{for } i < j,[11]$$

Then we can obtain estimators by minimizing the difference between $F_{GE}(x_{(i)}; \alpha, \lambda)$ and the term $\frac{i}{n+1}$:

$$\sum_{i=1}^{n} \left(F_{GE}(x_{(i)}; \alpha, \lambda) - \frac{i}{n+1} \right)^2, [11]$$

Which means the $\hat{\alpha}$ and $\hat{\lambda}$ estimators are given by minimizing:

$$\sum_{i=1}^{n} \left((1 - e^{-\lambda x_{(i)}})^\alpha - \frac{i}{n+1} \right)^2$$

If we want to include the effect of changing variance for each $X_{(i)}$, we can suggest so called Weighted least squares method.

$$\sum_{i=1}^{n} w_i \left(F_{GE}(x_{(i)}; \alpha, \lambda) - \frac{i}{n+1} \right)^2$$

We introduce the weight w_i

$$w_i = \frac{1}{Var(F_{GE}(X_{(i)}))} = \frac{(n+1)^2(n+2)}{j(n-i+1)}.[11]$$

such that weight w_i gives more importance to the $X_{(i)}$ with smaller variance. Then the $\hat{\alpha}$ and $\hat{\lambda}$ estimators are given by minimizing

$$\sum_{i=1}^{n} \frac{(n+1)^2(n+2)}{i(n-i+1)} \left((1 - e^{-\lambda x_{(i)}})^\alpha - \frac{i}{n+1} \right)^2.$$

4 GOODNESS OF FIT TESTS ON GE DISTRIBUTION

This chapter introduces a group of tests that will test whether given sample comes from GE distribution of not. These tests typically sumarize the discrepancy between observed values and the values expected under the predicted model.

4.1 Pearson χ^2 test

Pearson χ^2 test (commonly referred as χ^2 test) is nonparametric test. This test is based on fact that random variable with multinomial distribution can be transformed to a random vector with χ^2 distribution. The null hypothesis is that the frequency distribution of considered sample is consistent with the theoretical distribution (see figure 4.1). χ^2 test is an asymptotic test so it can be recommended only for large samples.[1]

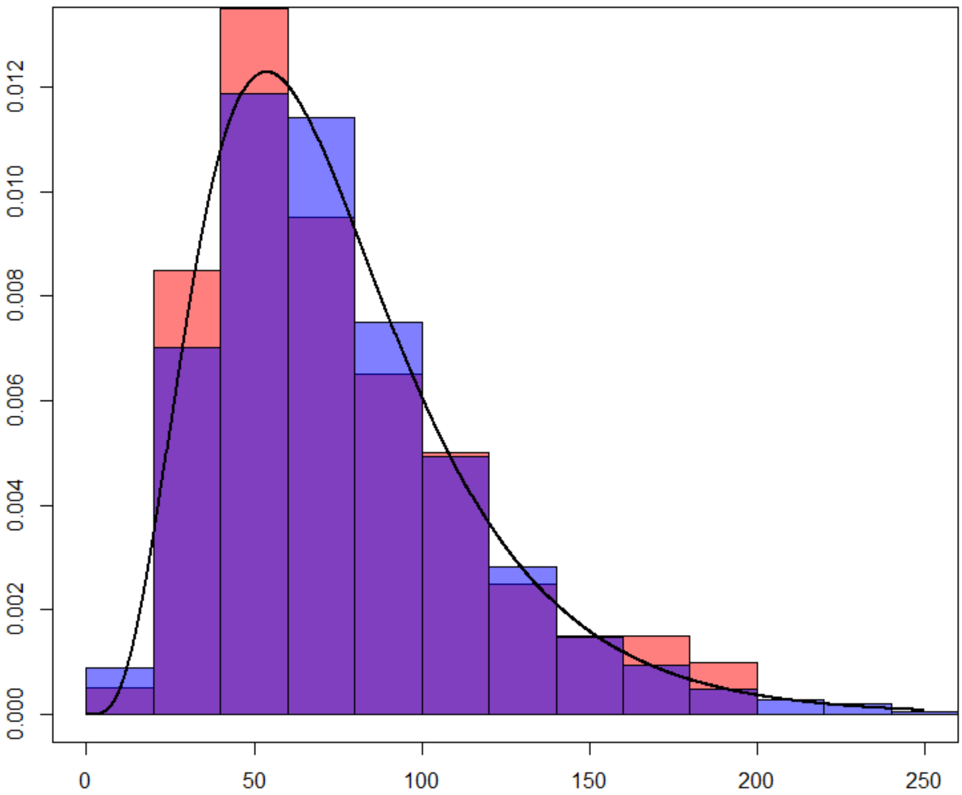

Fig. 4.1: Histogram of measured sample a fitted density

Let the set M defined in section 1.2 be divided into k intervals. In case of GE distribution $M = \mathbb{R}^+$. Let a_i and b_i be the lower and upper bound of each interval $I_i = (a_i, b_i\rangle$. Let $Z_1, Z_2, \ldots, Z_{k-1}, Z_k$ be the particular frequencies of each interval I_i. $Z_1, Z_2, \ldots, Z_{k-1}, Z_k$ have a multinomial distribution with parameters p_1, $p_2, \cdots, p_k$ (see the histogram 4.1 with blue columns representing the p_i value). Then $p_1 + \ldots + p_k = 1$ thus p_k is set by the others p_i such that $p_k = 1 - (p_1 + \ldots + p_{k-1})$. Now we define Q_{k-1}:

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(Z_i - np_i)^2}{np_i}$$

It is proved that for $n \to \infty$ Q_n has an limiting distribution $\chi^2(k-1)$. We can say that Q_{k-1} is an approximate χ^2 distribution with $k-1$ degrees of freedom. [14]

For known parameters α and λ of the GE distribution we can easily compute the $p_1, \ldots, p_k$.

$$p_i = F_{GE(\alpha,\lambda)}(b_i) - F_{GE(\alpha,\lambda)}(a_i)$$

where F is the cumulative distribution function of the GE distribution with known parameters α and λ.

If the parameters α and λ are not known then we estimate them by using the maximum likelihood method described in the chapter 3. Then the statistic Q_{k-1} has asymptotically χ^2 distribution with $(k-1-m)$ degrees of freedom where m is the number of parameters of the distribution that has been estimated (i.e. $m = 0$ for known α and λ parameters; $m = 1$ if one of the parameter is estimated; $m = 2$ if both α and λ are estimated by one of the methods of 3).

Literature commonly recommends to have n such that $np_i \geq 5$ for every $i = 1, \ldots, k$.

χ^2 **test for GE distribution** is defined as follows:

- H_0: The random sample is taken from the GE distribution
- H_a: The random sample is not taken from the GE distribution

Statistic:

$$\chi^2_{GE} = \sum_{i=1}^{k} \frac{(Z_i - np_i)^2}{np_i}$$

Where p_i is the frequency of observations in the interval I_i that was defined in the formula above, n is the number of measured samples and Z_i is the number of measured samples in interval i. Hypothesis H_0 is not rejected if

$$\chi^2_{GE} \leq \chi^2_{(1-\alpha)}(k-1-r).$$

Quantiles $\chi^2_{(1-\alpha)}(k-1-r)$ can be found in statistical tables or computed in language `R` by command `qchisq`.[19]

4.2 One-sample Kolmorov–Smirnov test

One-sample Kolmogorov–Smirnov test is a nonparametric test. Test statistic is the distance between the theoretical c.d.f. and the empirical c.d.f. The statistic finds the biggest difference between empirical and theoretical c.d.f. and compare it to the critical value. For a better clarification observe figure 4.2.

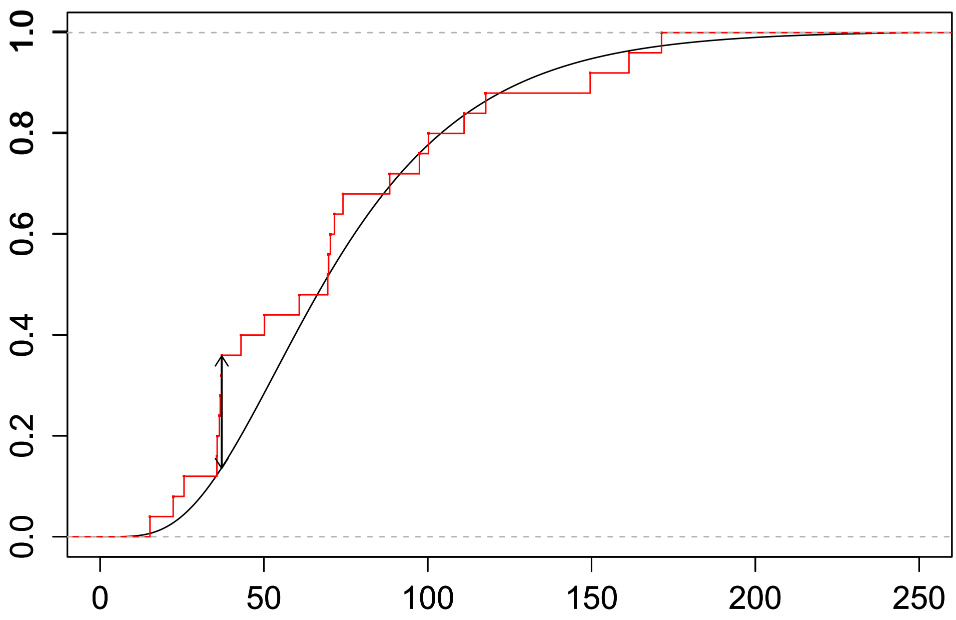

Fig. 4.2: Empirical cumulative distribution function and the hypothetical distribution function

---

The Kolmogorov–Smirnov test for GE distribution is defined as follows:

- H_0: The random sample is taken from the GE distribution
- H_a: The random sample is not taken from the GE distribution

Statistic:

$$D_n = \sup_x \left| \widehat{F}(x) - F_{GE(\hat{\alpha},\hat{\lambda})}(x) \right|$$

Where $\widehat{F}(x)$ is the empirical c.d.f. Values of D_n are compared to the critical value $D_\alpha(n)$ that can be found in tables in [13]. For large sizes of n can be used the aproximated critical value

$$D_\alpha(n) \approx \sqrt{\frac{1}{2n} \ln \frac{2}{\alpha}}.[17]$$

Hypothesis H_0 is not rejected if $D_n \leq D_\alpha(n)$. This test for GE distribution is included for language `R` in package `reliaR` as a function `ks.gen.exp`.

4.3 Anderson–Darling test

Anderson–Darling test compares the empirical cumulative distribution function with the hypothetical distribution function (see Fig.4.3). Anderson–Darling test is studied in detail in [13].

We treat the following measure:

$$W = n \int_0^{\infty} \psi\left(F_{GE(\hat{\alpha},\hat{\lambda})}(x)\right)\left[\widehat{F}(x) - F_{GE(\hat{\alpha},\hat{\lambda})}(x)\right]^2 dF, [2]$$

where $\psi\left(F_{GE(\hat{\alpha},\hat{\lambda})}(x)\right)$ is some preassigned weight function.

When $\psi\left(F_{GE(\hat{\alpha},\hat{\lambda})}(x)\right) = 1$, the statistic is the Cramér von Mies statistic. [13] Cramér von Mies statistic gives the same weight along the domain. Anderson–Darling test is realized when $\psi\left(F_{GE(\hat{\alpha},\hat{\lambda})}(x)\right) = \left[F_{GE(\hat{\alpha},\hat{\lambda})}(x)(1 - F_{GE(\hat{\alpha},\hat{\lambda})}(x))\right]^{-1}$, thus gives more weight on observations in the tails of the distribution.

$$W_{AD} = n \int_0^{\infty} \frac{\left[\widehat{F}(x) - F_{GE(\hat{\alpha},\hat{\lambda})}(x)\right]^2}{F_{GE(\hat{\alpha},\hat{\lambda})}(x)(1 - F_{GE(\hat{\alpha},\hat{\lambda})}(x))} dF$$

.

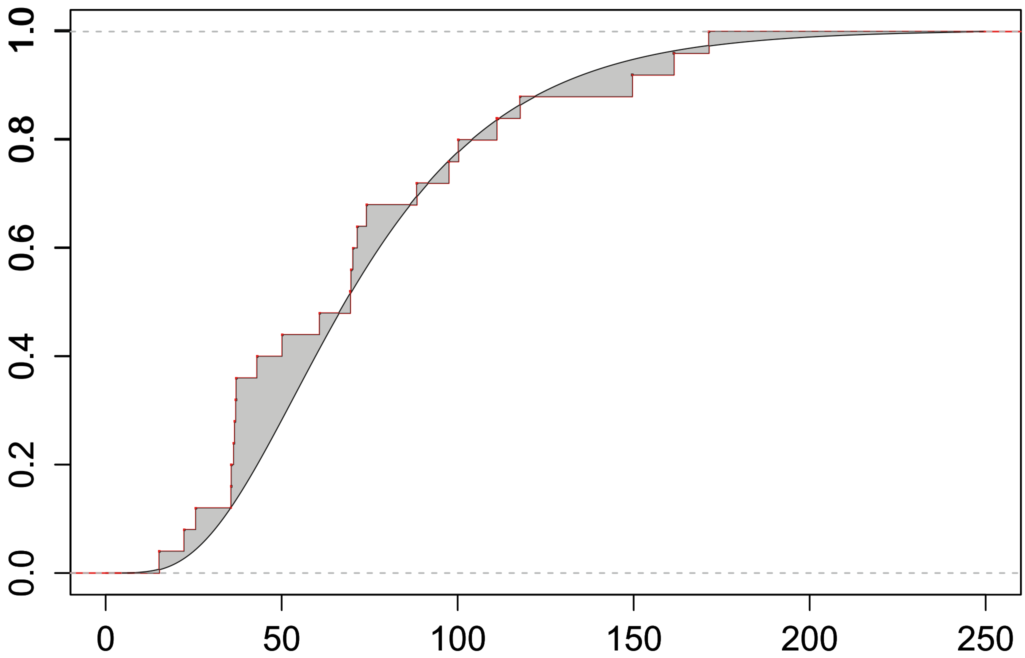

Fig. 4.3: Empirical c.d.f. and the hypothetical c.d.f.

Anderson–Darling test for GE distribution is defined as follows:

- H_0: The random sample is taken from the GE distribution
- H_a: The random sample is not taken from the GE distribution

Statistic:

$$A^2 = -n - \sum_{i=1}^{n} \frac{(2i-1)}{n} \left[\ln F(x_{(i)}, \hat{\alpha}, \hat{\lambda}) + \ln(1 - F(x_{(n+1-i)}, \hat{\alpha}, \hat{\lambda})) \right]$$

where F is the c.d.f. of the hypothetical distribution with known [estimated] parameters $\alpha[\hat{\alpha}]$ and $\lambda[\hat{\lambda}]$. $x_{(i)}$ are the ordered sample data. This A^2 is then compared to the critical value A^2_α that can be found in the tables [13]. Critical values are determinated by Monte Carlo methods for cases when both parameters are known, one parameter is known and one estimated or both parameters are estimated by maximum likelihood method derived in section 3. Hypothesis H_0 is not rejected if $A^2_\alpha \leq A^2$.

5 COMPARISONS OF WEIBULL, LOG-NORMAL AND GE DISTRIBUTION

Weibull, log-normal and GE distribution play an important role in reliability analysis. All of them are positively skewed distributions thus appropriate to analyze positively skewed data. We often assume that the data are coming from a specific parametric family and then build a model based on this assumption. But this assumption of a particular model is quite difficult. Basic properties of Weibull and log-normal distributions are summed up in the Appendix A (A.1, A.2).

It is observed that for certain range of parameters of Weibull, log-normal and GE distribution the probability density functions and cumulative distribution functions are very close to each other. On the other hand, hazard functions can differ very significantly with the same parameters.[6] One of the cases is shown on the Figure 5.1. In cases when empirical c.d.f. is close to these distribution functions it is very hard to distinguish between them and choose the correct distribution and thus the correct model.

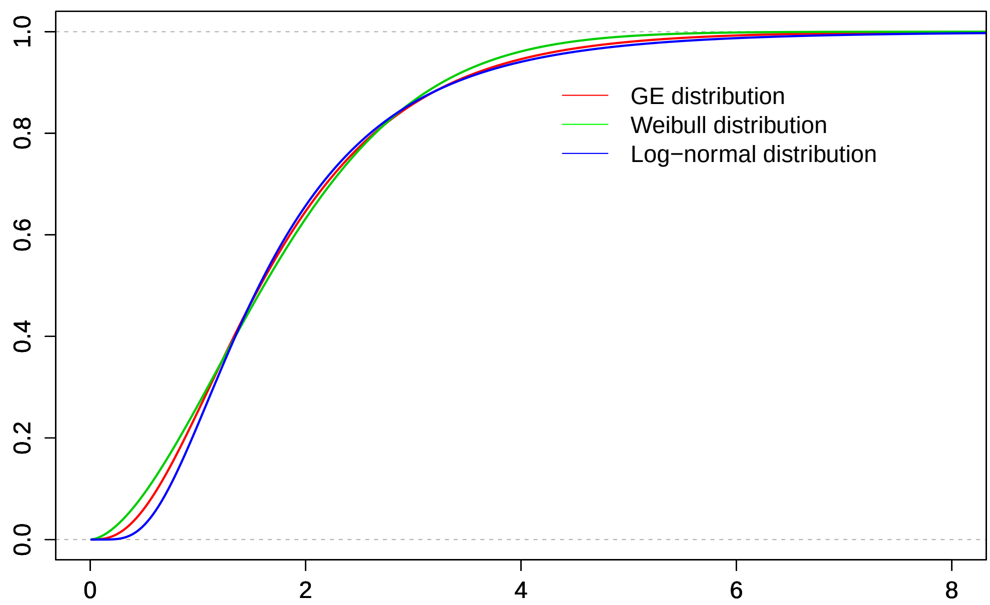

Fig. 5.1: C.d.f of Weibull ($\tau = 1.7, \theta = 2.1$), log-normal ($\mu = 0.45, \sigma = 0.6$) and GE ($\alpha = 3$ and $\lambda = 1$) distributions

The choice of the right model is crucial because as you can see on the Figure 5.2 the hazard functions of all the distributions have completely different behavior and this can lead to misinterpretation of the data. Distributions are close only for certain parameters such as the presented one. The key is to set a proper number of observations n such that likelihood ratio test is able to distinguish among the distributions. The method, which determines the minimum sample size n needed to discriminate among these three distributions for a given user specified probability of correct selection, is described in article [6]. Sadly it is observed that higher n is needed to distinguish between GE distributions and Weibull or log-normal distribution rather than distinguish between log-normal and Weibull distributions.

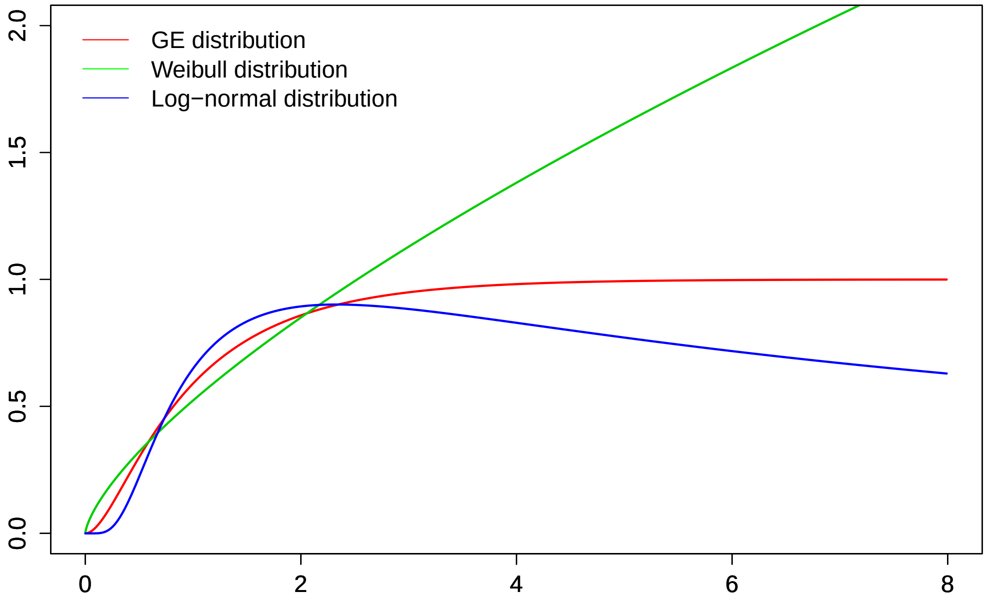

Fig. 5.2: Hazard functions of Weibull ($\tau = 1.7, \theta = 2.1$), log-normal ($\mu = 0.45, \sigma = 0.6$) and GE ($\alpha = 3$ and $\lambda = 1$) distributions

6 CENSORING

6.1 Introduction

In many practical experiments we can measure only some part of the data set due to limited measurement conditions or costs of the experiment.

In mechanical engineering we typically test mechanical parts for durability over time or lifetime (especially when the metal fatigue occurs). All the parts are tested during the experiment and we make some conclusions on durability of the parts based on how many parts did not last the experiment. The parts that survived the experiment are then censored because we did not get the information about the part's lifetime.

Most typically censoring occurs in medical research. When we want to estimate the survival of patients after organ transplant (or any other serious operation) and after some time patients stop attending their regular checkups or move to another region. For those patients we do not know any health information after this time.

Another case when censoring can play an important role is in experiments with measuring devices that are sensitive only in some range. For example if a scale can measure samples from 1 kg up to 100 kg we obviously cannot measure samples lighter than 1 kg. But we know that unmeasured lighter samples have 0–1 kg. Particularly this case is solved in the last chapter on the environmental data of organic and elemental carbon.

As you can see, censoring is very important in real experiments and it is crucial to recognize it and include it in the model for a better understanding of what is happening. It can be observed in the previous example with the mechanical parts. If we throw away all the data of parts that did not break during the experiment, we would lost a huge part of the information. Often only 10% of parts breaks during the experiment so we could lost a 90% of the information. In some cases, like this one, the effect is obvious, in some cases it isn't and by ignoring it we can get very different conclusions on what is happening in reality.

The following subsections deal with the basic types of censoring and give practical examples of the application on exponential distribution with censoring. Exponential distribution is a special case of GE distribution for $\alpha = 1$. This simplification is done mainly because we want to focus on the approach and not glut this text with long equations.

The last section of this chapter deals with type I left censored case on GE distribution and its asymptotic properties. The other types of the censoring were elaborated in several articles by the authors Gupta, Kundu and Mitra.

6.1.1 Type I censoring (Time censored samples)

Type I censoring is used when the experiment ends in a given fixed time T. Thus the number of censored variables is random and T is constant. The situation is illustrated in the figures 6.1,6.2,6.3.

Definition 6.2 (Type I left censoring). Let $\mathbb{X}$ be i.i.d. sample of size n and constant $T \in \mathbb{R}$ is given. Let $X_{(N)}$ be such that $X_{(N)} < T \leq X_{(N+1)}$. The tuple $(N, X_{(N+1)}, \dots, X_{(n)})$ is then called the type I left censored sample.

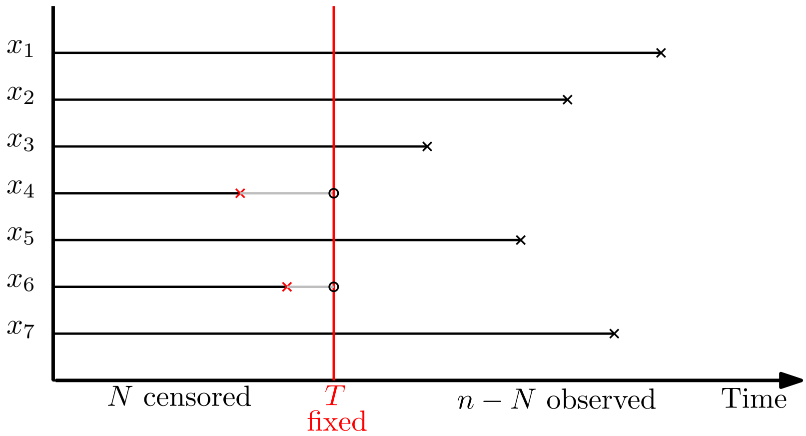

Fig. 6.1: Illustration of type I left censoring

Definition 6.3 (Type I right censoring). Let $\mathbb{X}$ be i.i.d. sample of size n and constant $T \in \mathbb{R}$ is given. Let $X_{(N)}$ be such that $X_{(N)} < T \leq X_{(N+1)}$. The tuple $(N, X_{(1)}, \dots, X_{(N)})$ is then called the type I right censored sample.

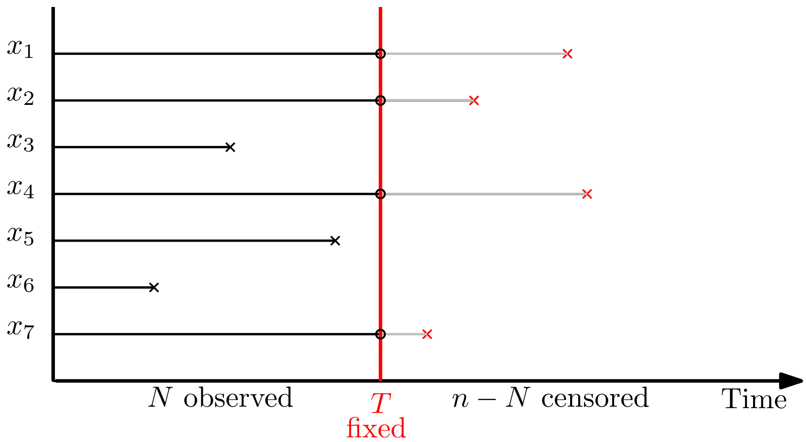

Fig. 6.2: Illustration of type I right censoring

Definition 6.4 (Type I interval censoring). Let $\mathbb{X}$ be i.i.d. sample of size n and constants $T_1, T_2 \in \mathbb{R}$ are given. Let $X_{(N_1)}$ be such that $X_{(N_1)} < T_1 \leq X_{(N_1+1)}$ and $X_{(N_2)}$ be such that $X_{(N_2)} < T_2 \leq X_{(N_2+1)}$.
The tuple $(N_1, N_2, X_{(1)}, \dots, X_{(N_1)}, X_{(N_2+1)}, \dots, X_{(n)})$ is then called the type I interval censored sample.

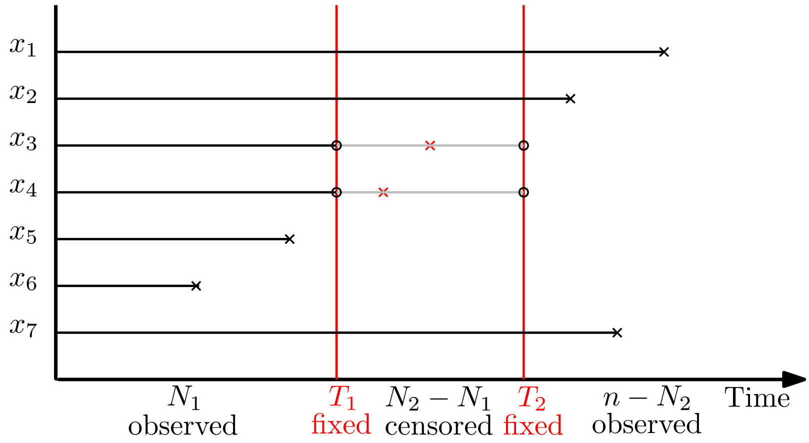

Fig. 6.3: Illustration of type I right censoring

THEOREM 6.4.1. [3]
Let $\mathbb{X}$ be the i.i.d. and the tuple $(N, X_{(N+1)}, \dots, X_{(n)})$ be the type I left censored sample. Then the likelihood function of this sample is:

$$l(\theta, T, x_{(N+1)}, \dots, x_{(n)}) = \frac{n!}{N!} \left(F(T,\theta)\right)^N \prod_{i=N+1}^{n} f\left(x_{(i)}, \theta\right).$$

EXAMPLE 1. [1] Find the maximum likelihood estimation of parameter λ under assumption that $\mathbb{X}$ of size n is i.i.d. with distribution $X_i \sim Ex(\lambda)$. Some observations are left censored by the time censor T.

SOLUTION: Because exponential distribution is a special type of GE distribution with parameter $\alpha = 1$, the p.d.f. of exponential distribution is:

$$f_{Ex(\lambda)}(x) = \lambda e^{-\lambda x}, \quad \text{for } x \geq 0$$

and the c.d.f. is given by:

$$F_{Ex(\lambda)}(x) = 1 - e^{-\lambda x}, \quad \text{for } x \geq 0$$

Because $\mathbb{X} = (X_1, ..., X_n)$ are independent variables then we can build a likelihood function. Assume that from n values, N are censored and $n - N$ are observed. Then the random variable N has binomial distribution thus

$$N \sim Bi(n, \theta)$$

where $\theta = F_{Ex(\lambda)}(T)$. Then the p.d.f. of this distribution is:

$$\begin{aligned} f_{Bi}(n, N, \theta) &= \tbinom{n}{N} (\theta)^N (1-\theta)^{n-N} \\ &= \tfrac{n!}{N!(n-N)!} \left(F_{Ex(\lambda)}(T)\right)^N \left(1 - F_{Ex(\lambda)}(T)\right)^{n-N} \end{aligned}$$

[1]Similar example on right censoring can be found in [17]

Then we can use the theorem 6.4.1 and write the likelihood function such as:

$$\begin{aligned} l(\lambda, T, x_{(N+1)}, \ldots, x_{(n)}) &= \tfrac{n!}{N!} \left(F_{Ex(\lambda)}(T)\right)^N \prod_{i=N+1}^n f_{Ex(\lambda)}\left(x_{(i)}\right) \\ &= \tfrac{n!}{N!} \left(1 - e^{-\lambda T}\right)^N \prod_{i=N+1}^n \lambda e^{-\lambda x_{(i)}} \end{aligned}$$

The log-likelihood function is then

$$\begin{aligned} L(\lambda, T, x_{(N+1)}, \ldots, x_{(n)}) &= \ln\left(\tfrac{n!}{N!}\right) + N \ln\left(F_{Ex(\lambda)}(T)\right) + \textstyle\sum_{i=N+1}^n \ln\left(f_{Ex(\lambda)}\left(x_{(i)}\right)\right) \\ &= \ln\left(\tfrac{n!}{N!}\right) + N \ln\left(1 - e^{-\lambda T}\right) + \textstyle\sum_{i=N+1}^n \ln\left(\lambda e^{-\lambda x_{(i)}}\right) \\ &= \ln\left(\tfrac{n!}{N!}\right) + N \ln\left(1 - e^{-\lambda T}\right) + (n-N)\ln(\lambda) \\ &\quad -\lambda \textstyle\sum_{i=N+1}^n x_{(i)} \end{aligned}$$

We want to find the maximum of this function. Then $\frac{\partial L}{\partial \lambda} = 0$.

$$\frac{\partial L}{\partial \lambda} = N \frac{T e^{-\lambda T}}{1 - e^{-\lambda T}} + \frac{n-N}{\lambda} - \sum_{i=N+1}^n x_{(i)} = 0$$

Then we get the implicit equation which must be solved by iterations such that $g(\lambda) = \lambda$.

$$\lambda = g(\lambda) = \left[\frac{1}{n-N} \sum_{i=N+1}^n x_{(i)} - \frac{N}{n-N} \frac{T e^{-\lambda T}}{1 - e^{-\lambda T}}\right]^{-1}$$

6.4.1 Type II censoring (Failure-censored samples)

This type of censoring is used when the experiment ends after given number of observations $N \in \mathbb{N}$. Thus the number of censored variables (left/interval/right) is fixed and the time T when we start the censoring is a random variable. The situation is illustrated by the figures 6.4 and 6.5.

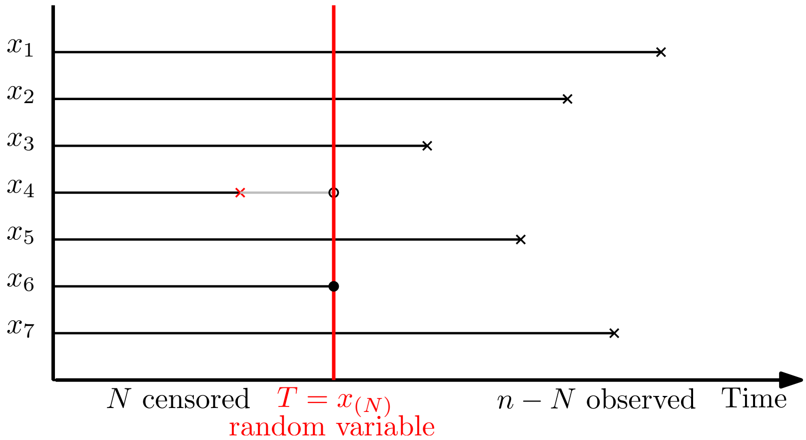

Fig. 6.4: Illustration of Type II left censoring

Definition 6.5. Let $\mathbb{X}$ be i.i.d. random sample of size n and constant $N_1, N_2 \in \mathbb{N}$ such that $1 \leq N_1 < N_2 \leq n$. Let $X_{(i)}$ be the ordered sample. The tuple $(X_{(1)}, \ldots, X_{(N_1)})$ is then called the type II right censored sample. The tuple $(X_{(N_1+1)}, \ldots, X_{(n)})$ is then called the type II left censored sample. The tuple

$(X_{(1)},\ldots,X_{(N_1)},X_{(N_2+1)},\ldots,X_{(n)})$ is then called the type II interval censored sample.

THEOREM 6.5.1. [3]
Let $\mathbb{X}$ be the i.i.d. random sample of size n with p.d.f. $f(x,\theta)$ and survival function $S(x,\theta)$. The tuple $(X_{(1)},\ldots,X_{(N)})$ be the type II right censored sample. Then the likelihood function of this sample is:

$$l(\theta,n,x_{(1)},...x_{(N)})=\frac{n!}{(n-N)!}\left(S(x_{(N)},\theta)\right)^{n-N}\prod_{i=1}^{N}f\left(x_{(i)},\theta\right).$$

EXAMPLE 2. Find the maximum likelihood estimation of parameter λ under assumption that $\mathbb{X}$ is a i.i.d. random vector of size n with distribution $X_i\sim Ex(\lambda)$. n units are being observed during the experiment and $n-N$ of them are right censored (see figure 6.5).[17]

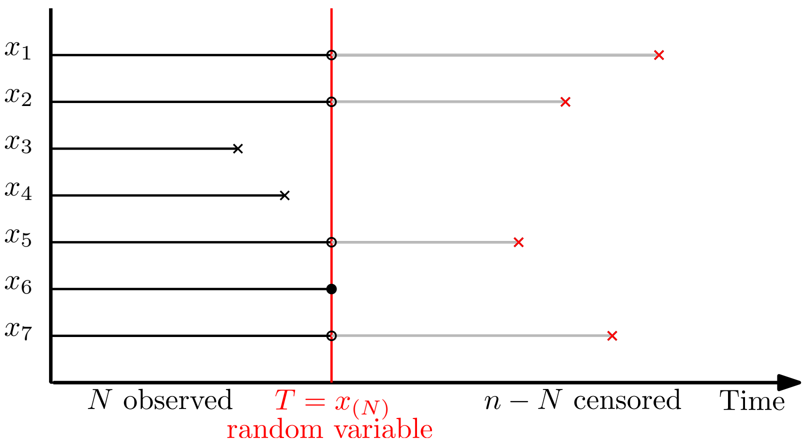

Fig. 6.5: Illustration of Type II right censoring

SOLUTION: Because exponential distribution is a special type of GE distribution with parameter $\alpha=1$ the survival function is given by:

$$S_{Ex(\lambda)}(x)=e^{-\lambda x},\quad \text{for } x\geq 0$$

Survival function is used because we have right censoring thus we need the $(1-F_{Ex(\lambda)}(x))$ probability which is $S_{Ex(\lambda)}(x)$. Because $\mathbb{X}=(X_1,\ldots,X_n)$ are all independent variables then we can use the theorem 6.5.1 and build a likelihood function l.

$$l(\lambda,n,x_{(1)},...,x_{(N)})=\frac{n!}{(n-N)!}\left(S_{Ex(\lambda)}(x_{(N)})\right)^{n-N}\prod_{i=1}^{N}f_{Ex(\lambda)}\left(x_{(i)}\right)$$

After substitution of $S_{Ex(\lambda)}(x)$ and $f_{Ex(\lambda)}(x)$ we obtain

$$l(\lambda,n,x_{(1)},...,x_{(N)})=\frac{n!}{(n-N)!}\left(e^{-\lambda x_{(N)}}\right)^{n-N}\prod_{i=1}^{N}\lambda e^{-\lambda x_{(i)}}$$

The log-likelihood function is then

$$L(\lambda, n, x_{(1)}, ..., x_{(N)}) = \ln\left(\frac{n!}{(n-N)!}\right) - (n-N)\lambda x_{(N)} + N\ln(\lambda) - \lambda\sum_{i=1}^{N} x_{(i)}$$

We want to find the maximum of this function. Then $\frac{\partial L}{\partial \lambda} = 0$.

$$\frac{\partial L}{\partial \lambda} = -(n-N)x_{(N)} + \frac{N}{\lambda} - \sum_{i=1}^{N} x_{(i)} = 0$$

By solving the equation we obtain the solution of the estimated λ parameter.

$$\hat{\lambda} = \left(\frac{1}{N}\sum_{i=1}^{N} x_{(i)} + \frac{n-N}{N}x_{(N)}\right)^{-1}$$

Note that we arrived to the explicit formula of $\hat{\lambda}$. This is caused by the right censoring and the survival function in the step of taking logarithm of the maximum likelihood function L.

6.5.1 Random censoring

Random censoring often occurs in case of complex real systems. We cannot design the experiment such that we stop the experiment in a given time. The most common examples are the studies in medicine, where patients can stop attending their regular checkups or move to another city. Some of the patients are then censored but not by the same time T. The time censor T_i is then a random variable for each observation X_i. [17]

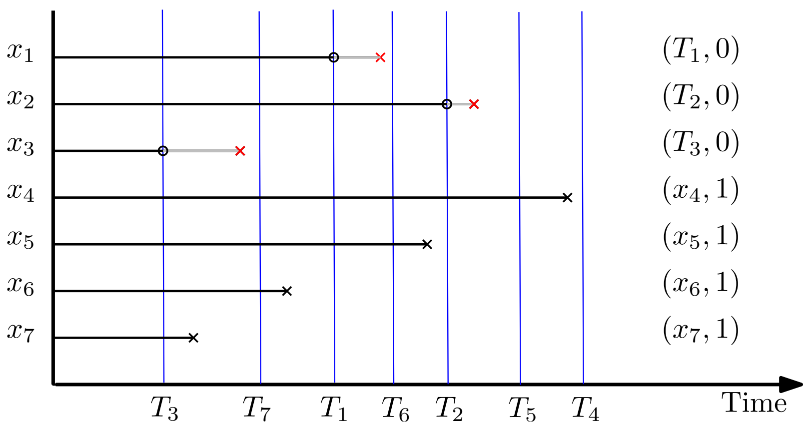

Fig. 6.6: Illustration of random right censoring

Definition 6.6 (Random right censored observations). Let $\mathbb{X}$ be i.i.d. sample and $\mathbb{T}$ be i.i.d. sample. $\mathbb{X}$ and $\mathbb{T}$ are not necessarily taken from the same distribution. Let $\mathbb{W}$ be a random variable such that $W_i = \min(X_i, T_i)$ and $I_i \in \{0, 1\}$. Then each observation can be defined as the couple (W_i, I_i) where W_i represents the time information about the observation and I_i represents if the observation was censored $(I = 0)$ or not $(I = 1)$. Observation is called randomly right censored if $X_i > T_i$ thus $I_i = 0$. Couple (W_i, I_i) is called the random censored sample. (see figure 6.6)

The left random censoring and interval random censoring are defined in the same way as the previous right random censoring. This type of censoring is the most complicated one. No example is given in this section in order to prevent the incomprehensibility of the text due to limited range of this thesis. Examples on random censoring can be found in the [17].

6.7 Type I left censored data from GE distribution

After introduction to censoring, this chapter will discuss type I left censored data of the GE distribution. Other types of censoring for GE distribution were mainly described by autors Gupta, Kundu and Mitra in [18]. The maxumim likelihood function l is again introduced. The size of the sample is n and the time censor will be denoted as T. N represents how many samples were left censored. Thus $0 \leq N \leq n$. Likelihood function is then

$$l(\alpha, \lambda, N, x_{(N+1)}, .., x_{(n)}) = \frac{n!}{N!} F_{GE}(T, \alpha, \lambda)^N \prod_{i=N+1}^{n} f_{GE}(x_{(i)}).$$

Log-likelihood function L is

$$L(\alpha, \lambda, N, x_{(N+1)}, .., x_{(n)}) = \ln\left(\frac{n!}{N!}\right) + N \ln(F_{GE}(T, \alpha, \lambda)) + \sum_{i=N+1}^{n} \ln[f_{GE}(x_{(i)})].$$

By substituting the F_{GE} and f_{GE} functions the following expression is obtained.

$$\begin{aligned} L(\alpha, \lambda, N, x_{(N+1)}, .., x_{(n)}) &= \ln\left(\tfrac{n!}{N!}\right) + N\alpha \ln(1 - e^{-\lambda T}) \\ &+ \textstyle\sum_{i=N+1}^{n} \ln[\alpha\lambda(1 - e^{-\lambda x_{(i)}})^{\alpha-1} e^{-\lambda x_{(i)}}] \end{aligned}$$

Next step is just formal breakdown of the sum.

$$\begin{aligned} L(\alpha, \lambda, N, x_{(N+1)}, .., x_{(n)}) = \ln\left(\tfrac{n!}{N!}\right) + N\alpha \ln\left(1 - e^{-\lambda T}\right) + (n-N)\ln(\lambda) + (n-N)\ln(\alpha) \\ + (\alpha - 1) \textstyle\sum_{i=N+1}^{n} \ln\left(1 - e^{-\lambda x_{(i)}}\right) - \lambda \textstyle\sum_{i=N+1}^{n} x_{(i)} \end{aligned}$$

This expression is fundamental for the following section. Finding the maximum likelihood estimators means finding the maximum $[\hat{\alpha}, \hat{\lambda}]$ of this function. Necessary conditions for the maximum of this function are $\frac{\partial L}{\partial \alpha} = 0$ and $\frac{\partial L}{\partial \lambda} = 0$. Both partial derivations of the L are computed

$$\frac{\partial L}{\partial \alpha} = N \ln\left(1 - e^{-\lambda T}\right) + \frac{n - N}{\alpha} + \sum_{i=N+1}^{n} \ln\left(1 - e^{-\lambda x_{(i)}}\right) = 0$$

$$\frac{\partial L}{\partial \lambda} = \frac{n-N}{\lambda} + \frac{N\alpha}{1-e^{-\lambda T}} T e^{-\lambda T} + (\alpha - 1) \sum_{i=N+1}^{n} \frac{x_{(i)} e^{-\lambda x_{(i)}}}{1-e^{-\lambda x_{(i)}}} - \sum_{i=N+1}^{n} x_{(i)} = 0$$

These two equations must be solved in order to find the maximum of this function. Similarly to chapter 3, the first equation gives an estimation of the α parameter which is denoted as $\hat{\alpha}$. Function $\hat{\alpha}(\lambda)$ depends only on λ parameter.

$$\hat{\alpha}(\lambda) = -\frac{n-N}{N \ln\left(1-e^{-\lambda T}\right) + \sum_{i=N+1}^{n} \ln\left(1-e^{-\lambda x_{(i)}}\right)}$$

The second equation can be modified into $g(\lambda) = \lambda$, form where the $g(\lambda)$ is the following expression:

$$g(\lambda) = \left[\frac{1}{n-N} \sum_{i=N+1}^{n} \frac{x_{(i)}}{1-e^{-\lambda x_{(i)}}} + \frac{\frac{NTe^{-\lambda T}}{1-e^{-\lambda T}} + \sum_{i=N+1}^{n} \frac{x_{(i)} e^{-\lambda x_{(i)}}}{1-e^{-\lambda x_{(i)}}}}{N \ln\left(1-e^{-\lambda T}\right) + \sum_{i=N+1}^{n} \ln\left(1-e^{-\lambda x_{(i)}}\right)}\right]^{-1}$$

The equation $g(\lambda) = \lambda$ can be solved iteratively. Algorithm, which is presented in the appendix, uses function `uniroot` in `R`, which is based on Fortran subroutine `zeroin`. The solution of this equation is denoted as $\hat{\lambda}$. The parameter estimator $\hat{\alpha}$ is obtained by substituting $\hat{\lambda}$ into $\hat{\alpha}(\lambda)$.

6.7.1 Fisher information matrix of type I left censored data

Fisher information matrix can give us the variance of the estimations $\hat{\alpha}$ and $\hat{\lambda}$. Theory was introduced in section 1.2. Fisher information matrix can be written as:

$$\mathcal{I}(\alpha, \lambda) = -\frac{1}{n} \begin{pmatrix} E\left(\frac{\partial^2 L(\alpha,\lambda)}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 L(\alpha,\lambda)}{\partial \alpha \lambda}\right) \\ E\left(\frac{\partial^2 L(\alpha,\lambda)}{\partial \lambda \alpha}\right) & E\left(\frac{\partial^2 L(\alpha,\lambda)}{\partial \lambda^2}\right) \end{pmatrix}$$

where the partial derivations are:

$$E\left(\frac{\partial^2 L(\alpha, \lambda)}{\partial \alpha^2}\right) = \frac{N-n}{\alpha^2}$$

$$E\left(\frac{\partial^2 L(\alpha, \lambda)}{\partial \alpha \partial \lambda}\right) = E\left(\frac{\partial^2 L(\alpha, \lambda)}{\partial \lambda \partial \alpha}\right) = E\left(\frac{NTe^{-\lambda T}}{1-e^{-\lambda T}} + \sum_{i=N+1}^{n} \frac{X_{(i)} e^{-\lambda X_{(i)}}}{1-e^{-\lambda X_{(i)}}}\right)$$

$$E\left(\frac{\partial^2 L(\alpha, \lambda)}{\partial \lambda^2}\right) = -E\left(\frac{n-N}{\lambda^2} + \frac{N\alpha T^2 e^{-\lambda T}}{\left(1-e^{-\lambda T}\right)^2} + (\alpha - 1) \sum_{i=N+1}^{n} \frac{X_{(i)}^2 e^{-\lambda X_{(i)}}}{\left(1-e^{-\lambda X_{(i)}}\right)^2}\right)$$

where $X_{(i)}$ and N are the random variables and $E(N)$ is computed as: $E(N) = nF_{GE}(T; \alpha, \lambda)$, where $F_{GE}(T; \alpha, \lambda)$ gives the percentual censoring and n is the sample range. $E(N)$ is the mean of the number of left censored variables. The p.d.f. of the i-th ordered variable is $f_{X_{(i)}}(x)$ and was introduced in section 2.6. Fisher information matrix in fact depends only on ratio $\frac{N}{n}$ and not the values N and n itselves. Thus every empirical Fisher information matrix can be computed numerically such that $F_{GE}(T;\alpha,\lambda) = \frac{N}{n}$ where $n, N \in \mathbb{N}$.

Values of the empirical Fisher information matrix were computed numerically in `R` and corresponding algorithm can be found on the attached CD. .

6.7.2 Simulation results

The maximum likelihood method for type I left censored data was programmed in `R` language in the R studio.[19] The algorithm can be found in appendix and on CD. At first the alpha and lambda values were set in combinations that can be seen in the table 6.1.

$\alpha\backslash\lambda$	0.5	1	1.5	2.5
0.5	(0.5,0.5)	(0.5,1)	(0.5,1.5)	(0.5,2.5)
1	(1,0.5)	(1,1)	(1,1.5)	(1,2.5)
1.5	(1.5,0.5)	(1.5,1)	(1.5,1.5)	(1.5,2.5)
2.5	(2.5,0.5)	(2.5,1)	(2.5,1.5)	(2.5,2.5)

Tab. 6.1: Table of all test choices of α and λ parameters for simulations

Random samples of sample size n were generated with these pairs (α, λ). As sample size n were chosen the representative samples of size 30, 50, 100 and 500.

n	30	50	100	500

Tab. 6.2: Table of all test choices for n

The censoring level T was set such that T will censor some % of observed values. The censoring level T is thus given in % out of observed values. T is computed by the quantile function Q_{GE}

$$T = Q_{GE}(p) = -\frac{\ln\left(1 - \sqrt[\alpha]{p}\right)}{\lambda}, \quad 0 \le p \le 1$$

where p is the given percentage that has been censored. At first the censoring levels were set at 0%, 5%, 10%, 15%, 20%, 25%, 50% and 75%.

p	0	0.05	0.1	0.15	0.2	0.25	0.5	0.75

Tab. 6.3: Table of all test choices for p

This gives four parameters (α, λ, n, T) on which the estimations $\hat{\alpha}$ and $\hat{\lambda}$ can depend. A 5-dimensional graph would be needed to explore all the aspects at once. For this reason we fix some of the parameters and explore the behavior of the estimations only partially.

For each combination (α, λ, n, T) was generated 1000 samples and the on each of them were computed the estimations $\hat{\alpha}$ and $\hat{\lambda}$. An average value of all estimations was taken and this value is presented as the result $\hat{\alpha}$ and $\hat{\lambda}$. Also variance and covariance of all estimations is presented. All the results can be seen in files on the attached CD.

In the following subsections will present results of the simulations when certain parameters will be fixed.

Fixed α and λ parameters and sample size n

By fixing both α and λ parameters and the sample size we are able to see how censoring affects the accuracy of the estimation of both parameters. As the representative example was chosen $\alpha = 1.5$ and $\lambda = 1$ and $n = 100$. The graph of dependence of estimation on censoring level can be seen on figure 6.7.

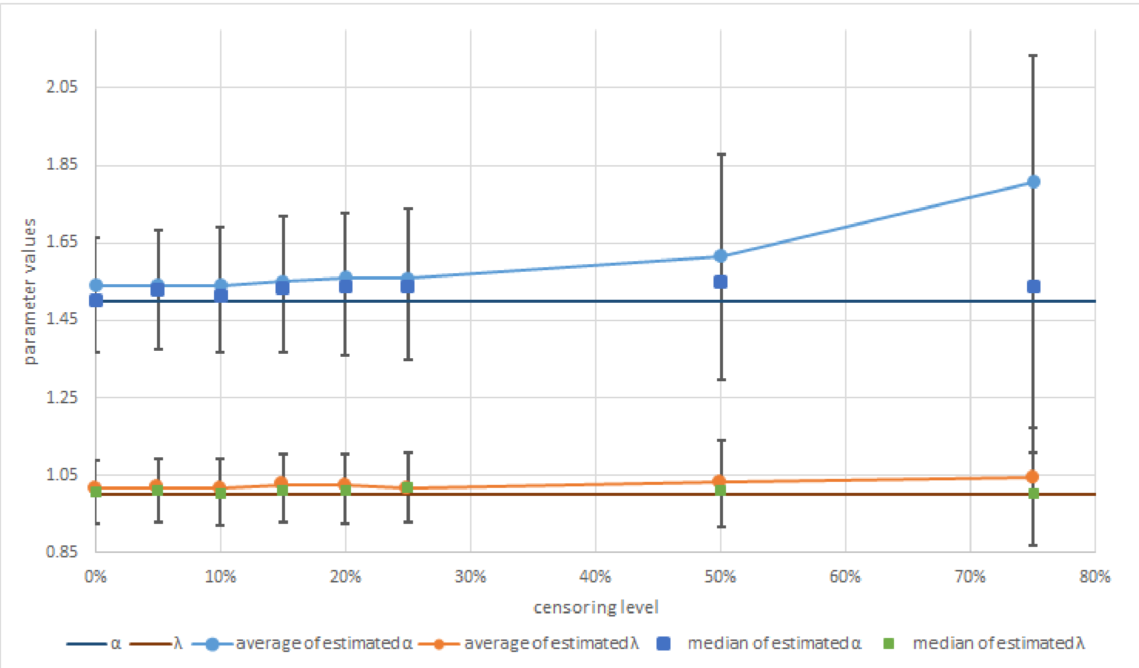

Fig. 6.7: Dependence of estimation on censoring level

The graph shows the average value of estimations α and λ. Moreover it shows the medians of the estimations and the first and third quartile. Generally for all combinations of parameters α, λ and n is observed that both parameters λ and α are biased. Also the λ parameter is estimated better than the α parameter.

Fixed α parameter and sample size n

If we fix only α parameter and n then it is reasonable to ask if the precision of estimator $\hat{\lambda}$ is affected by the value of λ. For comparing the quality of estimatior $\hat{\lambda}$ an ratio $\frac{\hat{\lambda}}{\lambda}$ is used. In fact the simulations did not showed any evidence that the quality of the estimator $\hat{\lambda}$ is affected by the value of λ. For fixed $\alpha = 2.5$ and $n = 100$ was produced a Fig. 6.8.

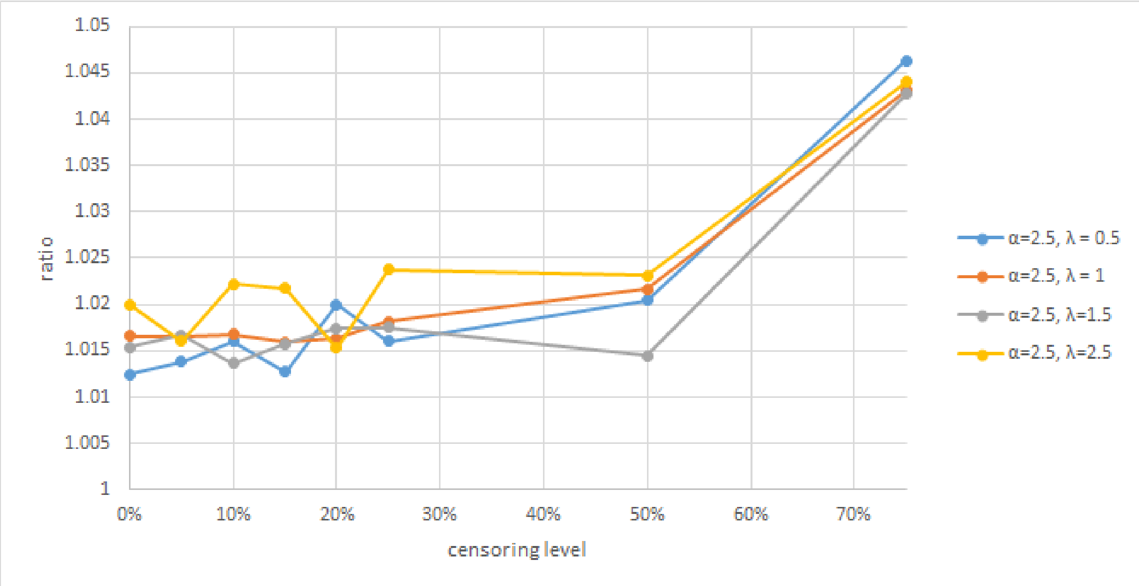

Fig. 6.8: Dependence of $\hat{\lambda}$ on censoring level and λ

Either the estimator $\hat{\alpha}$ did not showed any evidence that value of λ affects the estimation.

Fixed λ parameter and sample size n

When the parameter λ is fixed it is reasonable to ask if the precision of estimator $\hat{\alpha}$ is affected by the value of α. For most of the cases it was observed that the greater parameter α gets the worse is the precision of the estimatior $\hat{\alpha}$. Again a ratio $\frac{\hat{\alpha}}{\alpha}$ is introduced in order to compare the results. For fixed $\lambda = 1$ and $n = 100$ it can be seen on Fig. 6.9 that the greater the α parameter gets the less accurate is the estimation $\hat{\alpha}$. Specially for higher levels of censoring T.

On the other hand simulations showed that by increasing α parameter the more accurate is the estimation $\hat{\lambda}$. For fixed $\lambda = 1$ and $n = 100$ it can be seen on Fig. 6.10.

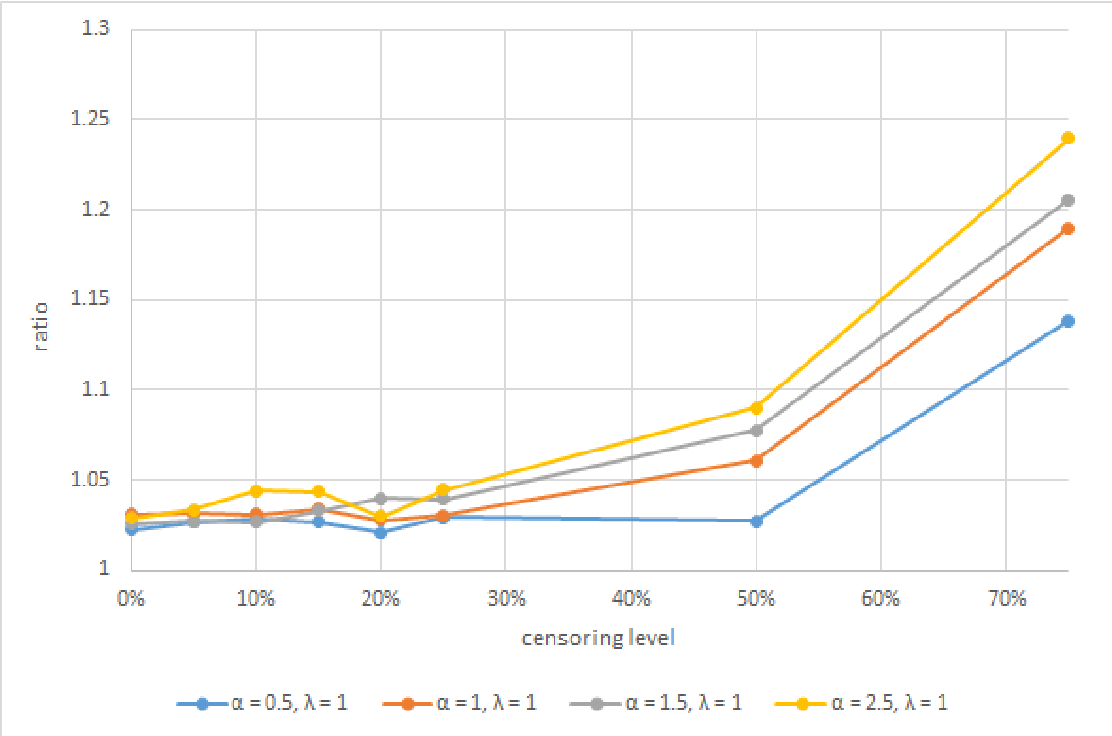

Fig. 6.9: Dependence of $\hat{\alpha}$ on censoring level and α

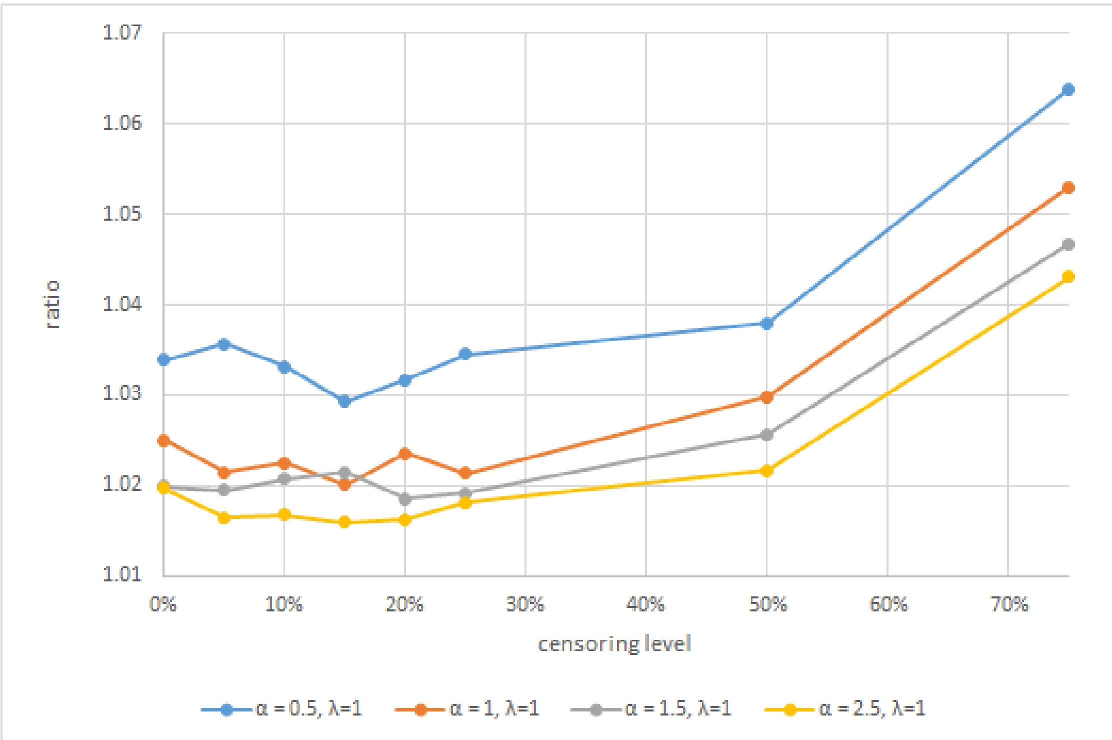

Fig. 6.10: Dependence of $\hat{\lambda}$ on censoring level and α

This sensitivity of estimators to change of α parameter can be explained by looking at the figures 2.1 and 2.2 of p.d.f. in Chapter 2. The shape parameter α affects the shape p.d.f. more than parameter λ.

Fixed α and λ parameters

For fixed parameters α and λ can be observed how accurate are the estimations $\hat{\alpha}$ and $\hat{\lambda}$ when we change the size of the sample n and the censoring level T. At first the previous simulations were analyzed but the situation showed that more deep insight will be needed. The T was chosen such that interval $\langle 0\%, 100\%)$ was divided to 20 representing points by 5%. Sample sizes n were chosen to 15, 30, 50, 100, 500 and 5000. As the fixed values were chosen $\alpha = 2.5$ and $\lambda = 2$. The results of the simulations can be seen on graph 6.11 and 6.11. Some results for cases when n is small and T is large are not dispayed in order to make the graph clear and easily understandable.

This graph again shows that maximum likelihood method estimates the λ parameter better than the α parameter. Both estimators are biased and maximum likelihood method tends to overestimate them both. Moreover we can see that for fixed n the estimators from some censoring level T grow exponentially. This for example shows that for $n = 100$ and $T > 70\%$ the maximum likelihood method won't give us any reliable information about the α parameter because the cesoring level will be simply too high. This problem can be fought by increasing the sample size n. If sample size n is raised from 100 to 500 the estimator $\hat{\alpha}$ improves rapidly. This shows that if the sample size n is high enough the censoring level T is not a problem for the calculations.

The natural question that would follow is about the variances of the estimators. So far only behavior of mean of the outcomes of the simulations was studied. The variance of the estimators is strongly related to the Fisher information matrix $\mathcal{I}$ which was introduced in chapter 1.1.

$$(\hat{\alpha}_n, \hat{\lambda}_n) \overset{A}{\sim} N_2\left((\alpha, \lambda), \frac{\mathcal{I}(\alpha, \lambda)^{-1}}{n}\right)$$

Fisher information matrix for type I left censored data was derived at the end of the previous section. The computation of Fisher information matrix is done in file `Fisher.r` on attached CD and the outcomes of this computation are presented in Excel file `FI.xls`. Parameters are again set to $\alpha = 2.5$ and $\lambda = 2$. Only limiting Fisher information matrix where the censoring level is set to $\frac{N}{n}\%$ is studied. The inverse of the Fisher information matrix is taken and presented in the table 6.4. Element A_{12} is equal to the element A_{21}. Correlation of parameters is computed

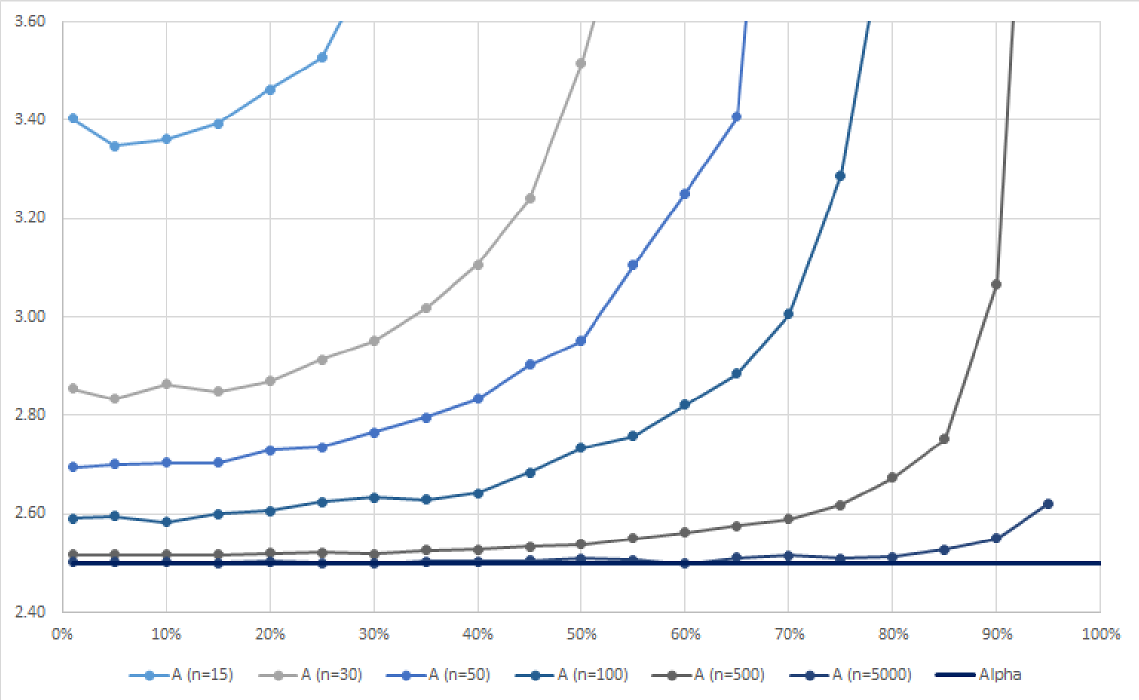

Fig. 6.11: Graph of average values of $\hat{\alpha}$ with respect to censoring level in % and number of observations n (maximum likelihood estimation)

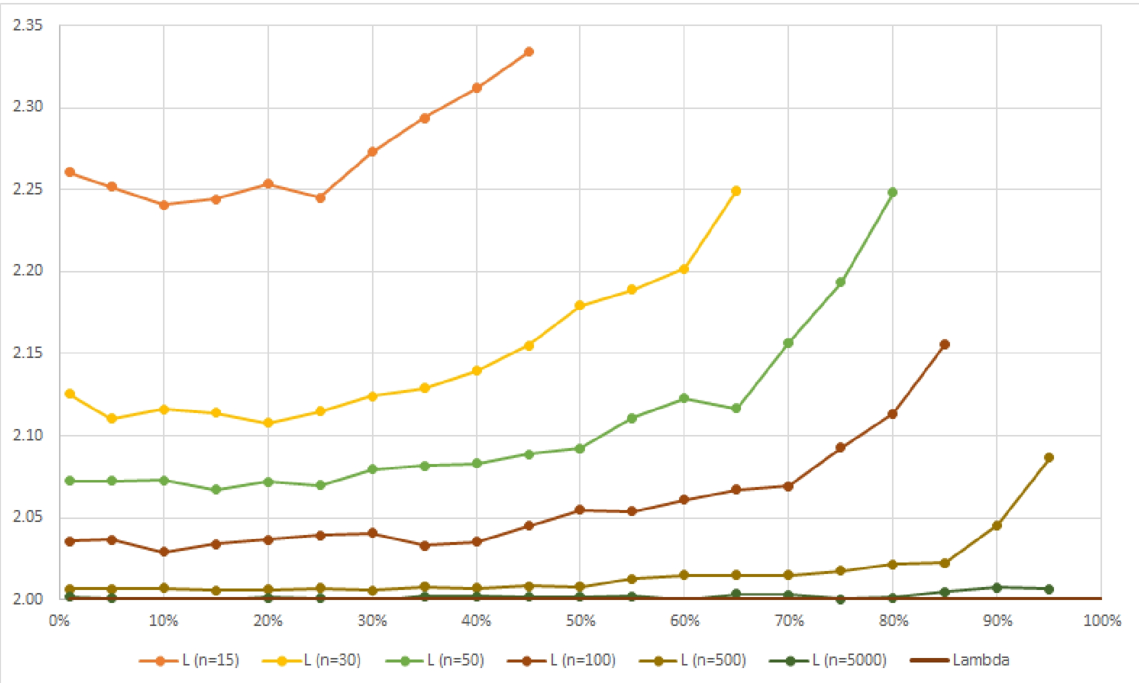

Fig. 6.12: Graph of average values of $\hat{\lambda}$ with respect to censoring level in % and number of observations n (maximum likelihood estimation)

and presented in the last column of the table.

$$\mathcal{I}(\alpha,\lambda)^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

Simulations supported this theory. The variances of $\hat{\alpha}$ and $\hat{\lambda}$ that was observed

T [%]	A_{11}	A_{22}	A_{12}	Correlation
0%	15.378	4.547	6.443	0.770
10%	19.352	5.155	7.997	0.801
20%	24.521	5.839	9.877	0.825
30%	31.720	6.683	12.341	0.848
40%	42.280	7.780	15.744	0.868
50%	58.790	9.289	20.737	0.887
60%	87.095	11.535	28.708	0.906
70%	141.910	15.233	42.945	0.924
80%	276.536	22.633	74.509	0.942
90%	826.124	45.086	185.550	0.961

Tab. 6.4: Values of inverse Fischer information matrix for $\alpha = 2.5$ and $\lambda = 2$

in simulations correspond to the theoretically computed. The variance of the $\hat{\alpha}$ can be seen on figure 6.13. The variance of the $\hat{\lambda}$ can be seen on figure 6.14. Pay attention to the logarithmic scale on the vertical axis. The correlation observed in simulations corresponds to the theoretically computed correlation. The outcomes of the simulation and the comparison can be seen on Fig. 6.15. Correlation has the advantage that it is dimensionless thus outcomes for different n can be visually compared. Fig. 6.15 shows that for different n there exist a certain level of censoring $T'(n)$ for which the simulations works quite well. But when the censoring level $T'(n)$ is exceeded the individual values start to fall down. This means that one of the variances grows rapidly and the estimatiors $\hat{\alpha}$ or $\hat{\lambda}$ are not useful anymore. For example for $n = 30$ is not recommended to have T more than 40%.

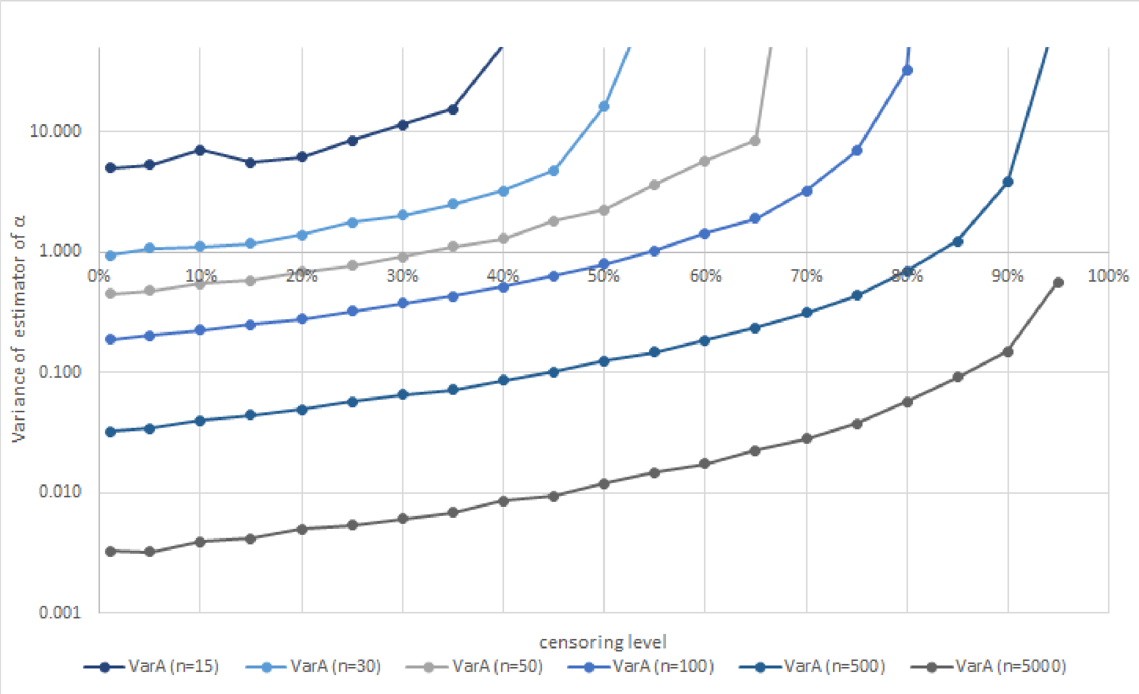

Fig. 6.13: Variance of $\hat{\alpha}$ from simulations

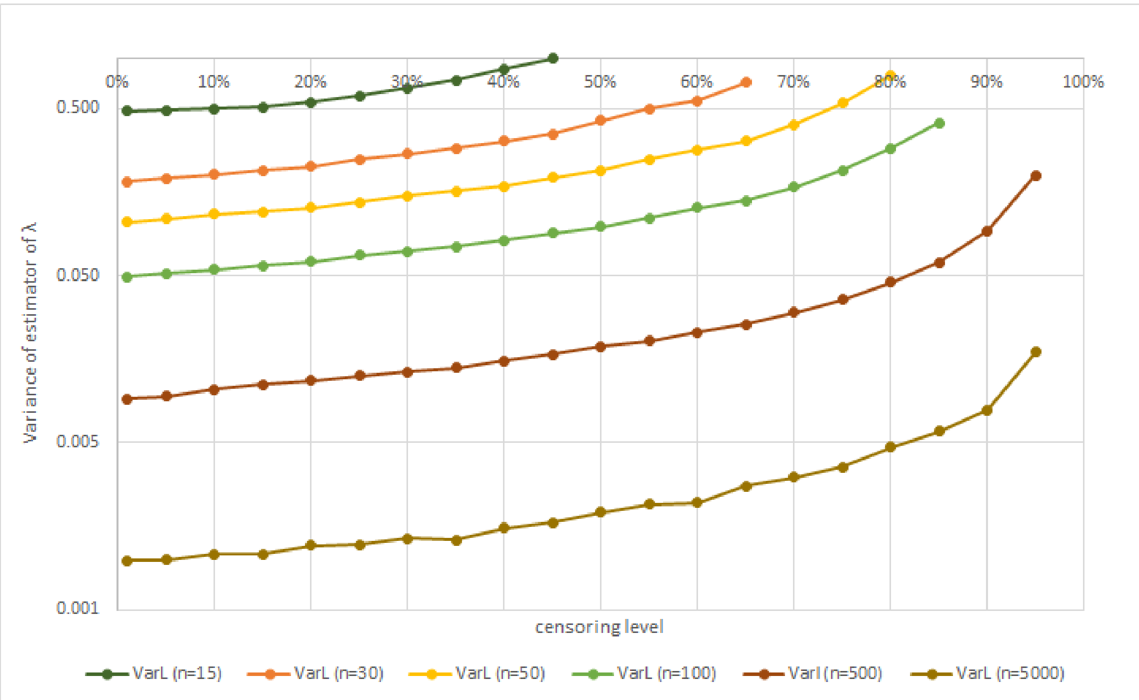

Fig. 6.14: Variance of $\hat{\lambda}$ from simulations

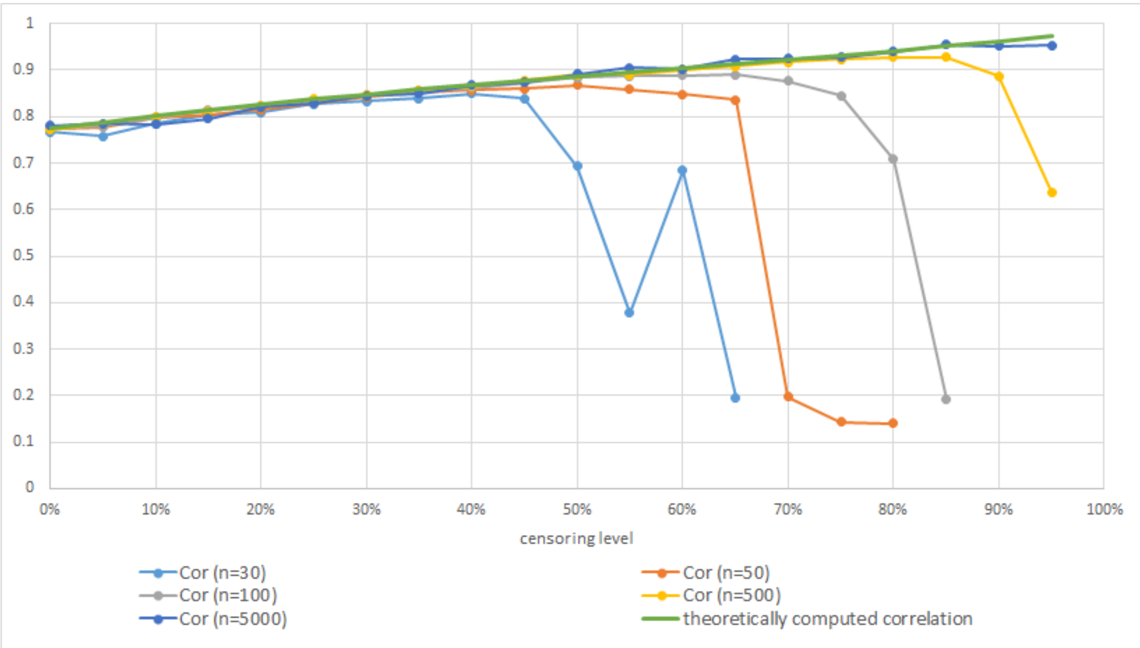

Fig. 6.15: Correlation of $\hat{\alpha}$ and $\hat{\lambda}$

7 EM ALGORITHM

Introduction

Expectation-maximization algorithm (EM algorithm) is an iterative method for finding the maximum likelihood estimates of parameters in statistical models with latent variables. It is broadly applicable to the interactive computation and even for very complex incomplete data problems. EM algorithm was firstly suggested in article [5]. Book [16] describes many applications of this algorithm.

The theory developed in this section will demonstrate the alternative algorithm for maximum likelihood estimations and will be compared with the algorithm developed in the previous chapter. EM algorithm developed here is purposed for more general cases than just type I left censored samples. Censoring is generalized to type I interval censoring and can be adjusted to left or right censoring eventually.

The general EM algorithm starts with some initial parameters $\theta^{(0)}$ which can be chosen from some given domain Θ of the estimated parameters. A new value of the θ is computed in each iteration as the algorithm proceeds. The value in the k-th iteration will be denoted as $\theta^{(k)}$. The data can be split into complete data, which will be marked as X, and incomplete data, which will be marked as Y. Each iteration consists of two steps, the Expectation step (E step) and the Maximization step (M step).[16]

E step: In this step a function $Q(\theta, \theta^{(k)})$ is constructed, which is an function of expectation of the log-likelihood function which depends on the known parameter $\theta^{(k)}$ form the previous step and parameter θ that should be estimated.

$$Q(\theta, \theta^{(k)}) = E_{\theta^{(k)}} \{L(\theta)|Y\}$$

M step: This step is used for finding $\theta^{(k+1)} \in \Theta$ such that function $Q(\theta, \theta^{(k)})$ is maximized.

$$Q(\theta^{(k+1)}, \theta^{(k)}) = \max_{\theta} Q(\theta, \theta^{(k)})$$

Weaker version of this step just requires to find value $\psi^{(k+1)} \in \Theta$ such that

$$Q(\psi^{(k+1)}, \psi^{(k)}) \geq Q(\psi, \psi^{(k)})$$

E steps and M steps are repeated until $\left|\psi^{(k)} - \psi^{(k+1)}\right| < \epsilon$ where ϵ is the chosen precision value.

Convergence analysis of this algorithm was sketched in the original paper [5] however it was fully developed in 1983 by C. F. Jeff Wu in article [20]. The algorithm converges for exponential family (GE distribution is in exponential family) of distributions as well as for the problems outside the exponential family (this was the major proof by Wu). EM algorithm is due to this property used in many applications.

Unfortunately the convergence of the algorithm to some value θ does not necessary mean that θ in the global maximum of the function. In general, if the log-likelihood function L has several maximums and stationary points, convergence of the EM sequence to either type of point depends on the choice of starting point. This situation can be prevented by using several starting points and choosing the θ which returns maximum value of the $Q(\theta)$ function. [16]

EM algorithm for the GE distribution

The EM algorithm in this particular problem uses an maximum likelihood function L_c, where censored data can be thought of as the missing data of the EM algorithm. Complete data set $\mathbb{W}$ can be formed by combining the observed data X_i with interval censored data Y_i. The interval censored data Y_i are censored on the interval (a_i, b_i). Y_i forms generalized case for type I censored data. When the data Y_i are left censored, then $a_i = 0$. If Y_i are right censored, then $b_i = \infty$. Number of censored variables is N and number of observed variables X_i is $n - N$. Together they form data set of size n. We assume that $\mathbb{W}$ is i.i.d. and taken from GE distribution and we treat Y_i as an observed value. The likelihood function l_c is:

$$l_c(\alpha, \lambda, N, \mathbb{W}) = \prod_{i=1}^{n-N} f_{GE}(x_i, \alpha, \lambda) \prod_{j=1}^{N} f_{GE}(y_j, \alpha, \lambda)[4]$$

This function in fact coincides with the likelihood function introduced in section 3.1. Logarithm of the likelihood function is:

$$L_c(\alpha, \lambda, N, \mathbb{W}) = \sum_{i=1}^{n-N} \ln f_{GE}(x_i, \alpha, \lambda) + \sum_{j=1}^{N} \ln f_{GE}(y_j, \alpha, \lambda)$$

After substituting the f_{GE}:

$$\begin{aligned} L_c(\alpha, \lambda, N, \mathbb{W}) = &\sum_{i=1}^{n-N} \ln\left(\alpha\lambda(1 - e^{-\lambda x_i})^{\alpha-1} e^{-\lambda x_i}\right) + \\ &\sum_{j=1}^{N} \ln\left(\alpha\lambda(1 - e^{-\lambda y_j})^{\alpha-1} e^{-\lambda y_j}\right) \end{aligned}$$

Then we can expand all the terms. But values y_i are not known since Y_i are censored on the intervals. This problem is solved by taking the expected value of the outcome as you can see in the following expression.

$$\begin{aligned} L_c(\alpha, \lambda, N, \mathbb{W}) = &n \ln(\alpha) + n \ln(\lambda) + (\alpha - 1) \sum_{i=1}^{n-N} \ln\left(1 - e^{-\lambda x_i}\right) - \lambda \sum_{i=1}^{n-N} x_i + \\ &(\alpha - 1) \sum_{j=1}^{N} E\left(\ln\left(1 - e^{-\lambda Y_j}\right)\right) - \lambda \sum_{j=1}^{N} E\left(Y_j\right) \end{aligned}$$

Now the main task is to deal with the expectations $E\left(\ln\left(1-e^{-\lambda Y_j}\right)\right)$ and $E\left(Y_j\right)$. The idea is that estimations between steps k and $k+1$ won't differ that much, so for step $k+1$ the estimations from the previous step will be partially used in such way that:

$$E\left(\ln\left(1-e^{-\lambda Y_j}\right)\right) = \int_{a_i}^{b_i} \ln\left(1-e^{-\lambda y}\right) f_{Y_j}(y, \alpha^{(k)}, \lambda^{(k)})\, dy$$

$$E\left(Y_j\right) = \int_{a_i}^{b_i} y\, f_{Y_j}(y, \alpha^{(k)}, \lambda^{(k)})\, dy$$

where f_{Y_j} is the p.d.f. of Y_j and is defined as:

- for interval censored data:

$$f_{Y_j}(y_j, \alpha, \lambda) = \frac{f_{GE(\alpha,\lambda)}(y_i)}{F_{GE(\alpha,\lambda)}(b_j) - F_{GE(\alpha,\lambda)}(a_j)}, \quad a_i \le y_i \le b_i$$

- for right censored data

$$f_{Y_j}(y_j, \alpha, \lambda) = \frac{f_{GE(\alpha,\lambda)}(y_i)}{1 - F_{GE(\alpha,\lambda)}(a_i)}, \quad a_i \le y_i \le \infty$$

- for left censored data

$$f_{Y_j}(y_j, \alpha, \lambda) = \frac{f_{GE(\alpha,\lambda)}(y_i)}{F_{GE(\alpha,\lambda)}(b_i)}, \quad 0 \le y_i \le b_i$$

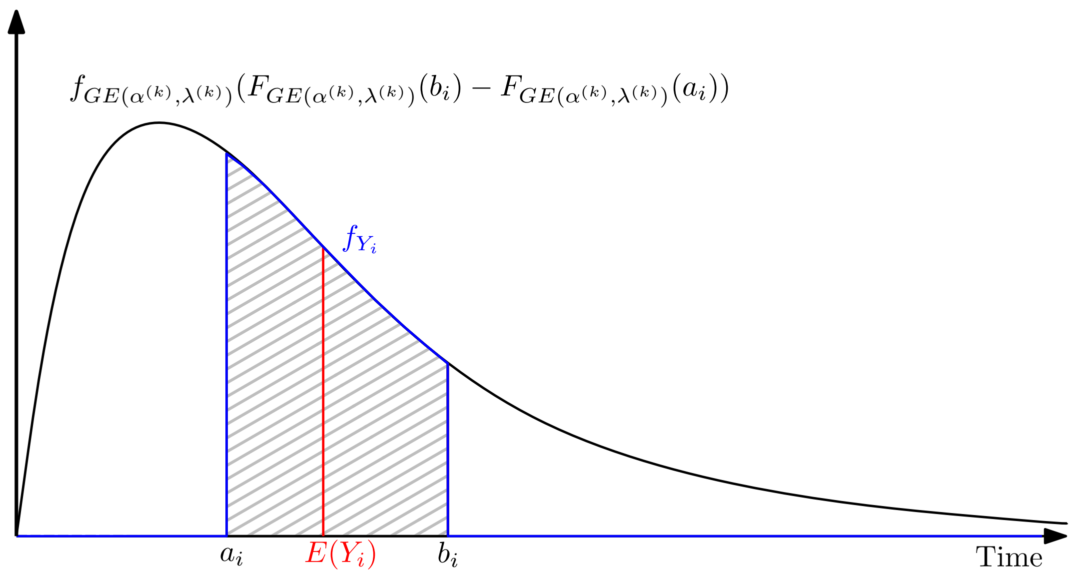

Fig. 7.1: Illustration of estimation of $E(Y_i)$ from interval censored variables Y_i in EM algorithm

This is the key step that simplifies the most of the calculations. First expectation depends now only on λ and the second is independent on both parameters. Illustrative examples of both expectations in case of interval censoring can be seen on

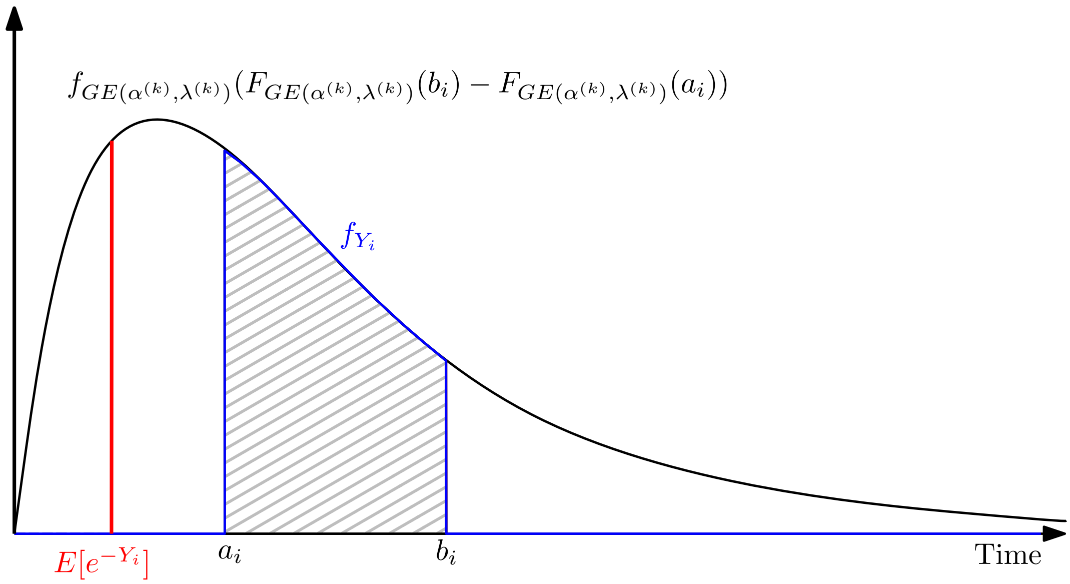

Fig. 7.2: Illustration of estimation of $E[e^{-Y_i}]$ from interval censored variables Y_i in EM algorithm

figures 7.1, 7.2. First figure shows the estimated expected value of Y_i. Second figure shows the expected value of e^{-Y_i}.
By substituting these terms, the following equation is obtained.

$$
\begin{aligned}
L_c(\alpha, \lambda, N, \mathbb{W}) =& n\ln(\alpha) + n\ln(\lambda) + (\alpha-1)\sum_{i=1}^{n-N} \ln\left(1-e^{-\lambda x_i}\right) - \lambda\sum_{i=1}^{n-N} x_i + \\
& (\alpha-1)\sum_{j=1}^{N} E\left(\ln\left(1-e^{-\lambda Y_j}\right)\right) - \lambda\sum_{j=1}^{N} E\left(Y_j\right)
\end{aligned}
$$

The parameters that maximizes this function must be found. In order to maximize the L_c, derivations with respect to parameters α and λ must be taken. Both partial derivations must be equal to zero as a necessary condition of the maximum.

$$
\frac{\partial L_c(\alpha, \lambda, N, \mathbb{W})}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^{n-N} \ln\left(1-e^{-\lambda x_i}\right) + \sum_{j=1}^{N} E\left(\ln\left(1-e^{-\lambda Y_j}\right), \alpha^{(k)}, \lambda^{(k)}\right) = 0
$$

$$
\begin{aligned}
\frac{\partial L_c(\alpha, \lambda, N, \mathbb{W})}{\partial \lambda} =& \frac{n}{\lambda} - \sum_{i=1}^{n-N} x_i + (\alpha-1)\sum_{i=1}^{n-N} \frac{x_i e^{-\lambda x_i}}{1-e^{-\lambda x_i}} - \\
& \sum_{j=1}^{N} E\left(Y_j, \alpha^{(k)}, \lambda^{(k)}\right) + (\alpha-1)\sum_{j=1}^{N} E\left(\frac{Y_j e^{-\lambda Y_j}}{1-e^{-\lambda Y_j}}, \alpha^{(k)}, \lambda^{(k)}\right) = 0
\end{aligned}
$$

From the first equation arise the estimation of $\alpha^{(k+1)}$.

$$
\alpha^{(k+1)} = \hat{\alpha}(\lambda) = -n\left[\sum_{i=1}^{n-N} \ln\left(1-e^{-\lambda x_i}\right) + \sum_{j=1}^{N} E\left(\ln\left(1-e^{-\lambda Y_j}\right), \alpha^{(k)}, \lambda^{(k)}\right)\right]^{-1}
$$

Unfortunately $\alpha^{(k+1)}$ is dependent on λ. Approximation of λ is needed and is obtained from the second equation. Second equation can be again expressed as $g(\lambda) = \lambda$ where $g(\lambda)$ is

$$\begin{aligned} g(\lambda) =& n\Bigg[\sum_{i=1}^{n-N} x_i - (\alpha - 1)\sum_{i=1}^{n-N} \frac{x_i e^{-\lambda x_i}}{1 - e^{-\lambda x_i}} + \\ & \sum_{j=1}^{N} E\left(Y_j, \alpha^{(k)}, \lambda^{(k)}\right) - (\alpha - 1)\sum_{j=1}^{N} E\left(\frac{Y_j e^{-\lambda Y_j}}{1 - e^{-\lambda Y_j}}, \alpha^{(k)}, \lambda^{(k)}\right)\Bigg]^{-1} \end{aligned}$$

Function $g(\lambda)$ is also dependent on the α parameter, but this can be estimated as $\hat{\alpha}(\lambda)$ and thus $g(\lambda)$ will depend only on parameter λ.

The steps are now similar to estimation in Chapter 6 except that the calculations are done in an iterative way. The algorithm is summarized here and the implementation can be found on the attached CD.

EM algorithm for GE distribution

Data: Vector $\mathbb{x}$, Vectors $\mathbb{a}$ and $\mathbb{b}$ for censored variables.
Result: Estimated parameters $\hat{\alpha}$ and $\hat{\lambda}$
Start with arbitrary positive values $\alpha^{(1)}$ and $\lambda^{(1)}$;
$\alpha^{(k+1)} := \alpha^{(1)}$, $\lambda^{(k+1)} := \lambda^{(1)}$;
repeat
  $\alpha^{(k)} := \alpha^{(k+1)}$, $\lambda^{(k)} := \lambda^{(k+1)}$;
  Define $g(\lambda) = g(\lambda, \mathbb{x}, \mathbb{a}, \mathbb{b}, \alpha^{(k)}, \lambda^{(k)})$;
  $\lambda^{(k+1)}$:=Solve($g(\lambda) - \lambda = 0$);
  $\alpha^{(k+1)} = \hat{\alpha}(\lambda^{(k+1)})$;
until $\left|\alpha^{(k)} - \alpha^{(k+1)}\right| < \epsilon$ *and* $\left|\lambda^{(k)} - \lambda^{(k+1)}\right| < \epsilon$;

7.1 Simulation results

As in the simulation results of the maximum likelihood method the same simulations were done for type I left censored data by EM algorithm. The results of the simulations are virtually the same and both methods are consistent with each other. On the figures 7.4 and 7.5 can be seen a similar outcome of the simulations by using EM algorithm as it was on figures 6.11 and 6.12, where maximum likelihood method was used.

The biggest difference between this two approaches is the computing time. As we could see, the EM algorithm is more general and is able to handle even very diverse types of censoring. This is reflected in the computing time. On Fig. 7.3 can be seen, that for higher censoring levels, the EM algorithm must do more iterations and the demands on time grows. This particular simulation was set such that both algorithms had to compute 1000 problems on the set censoring level T with $n = 100$ and for $\alpha = 2.5, \lambda = 2$. The time displayed is the overall time for each censoring level T.

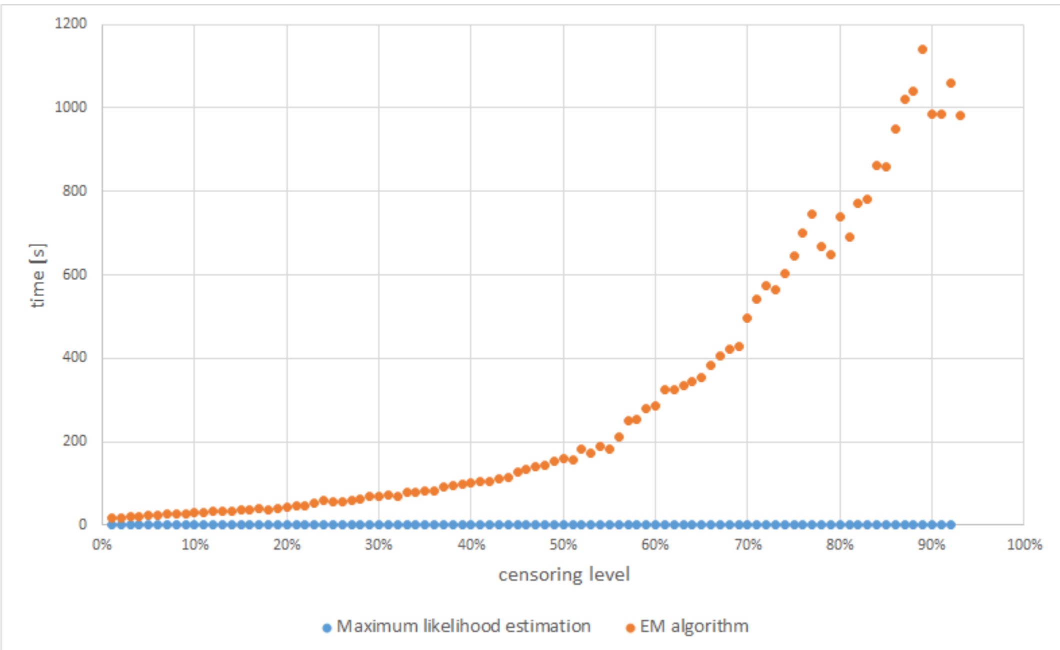

Fig. 7.3: Time demands on both algorithms

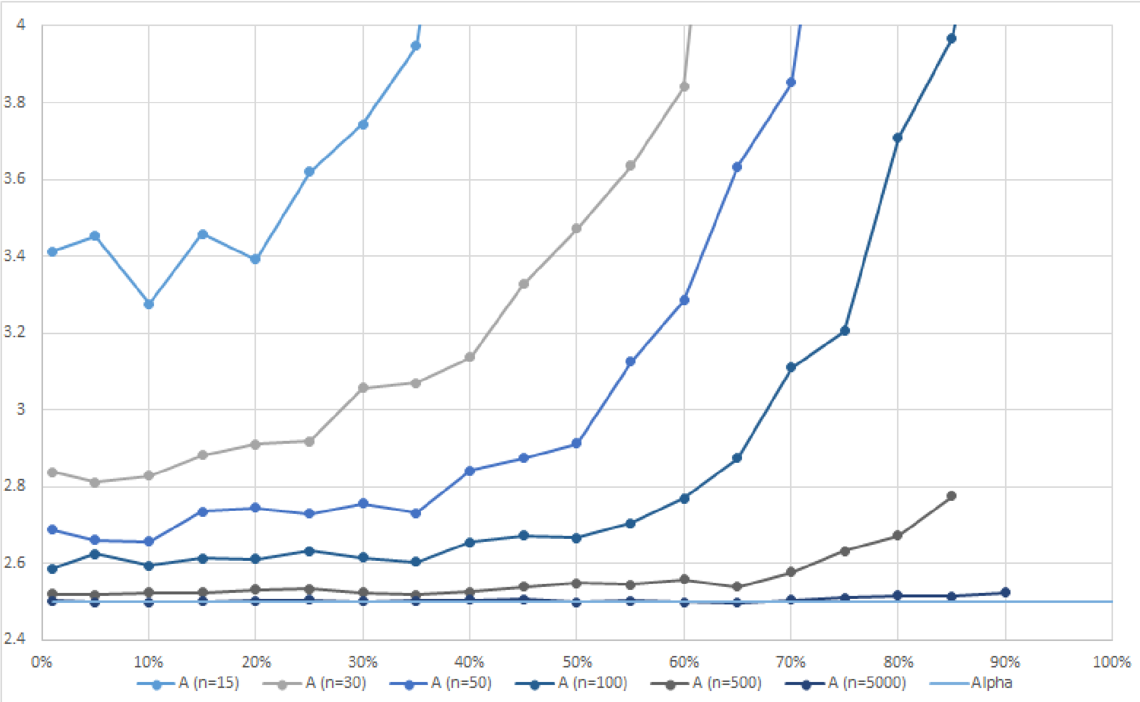

Fig. 7.4: Graph of average values of $\hat{\alpha}$ with respect to censoring level in % and number of observations n (EM algorithm)

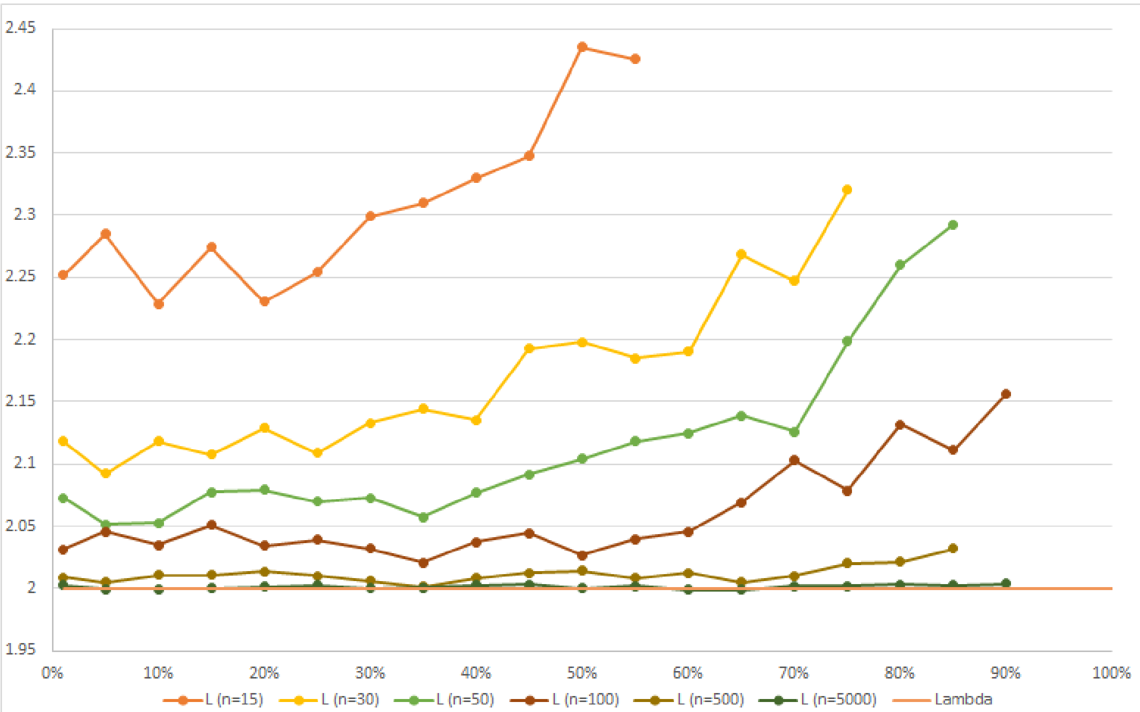

Fig. 7.5: Graph of average values of $\hat{\lambda}$ with respect to censoring level in % and number of observations n (EM algorithm)

8 APPLICATION

This chapter shows the application of derived theory on given measured sets of data. First set of data deals with elemental carbon. Second set deals with organic carbon. These two sets were measured at atmospheric station at Křešín u Pacova from August 2013 till December 2013. Measurements were taken every day in 4 hour intervals.

Elemental carbon (also called black carbon, abbreviated as EC) is emitted during the combustion of fossil fuels as small aerosol particles that are smaller than 2,5 µm. EC particles have often other chemicals attached to their surface. EC forms include charcoal, soot, graphite, and coal. EC causes higher human morbidity and premature mortality.

Naturally-occurring organic carbon (abbreviated as OC) forms are derived from the decomposition of plants and animals. Sources of organic carbon include industrial combustion and the degradation of carbon-containing materials.

Both samples are type I left censored with censoring level $T = 0.5$ µgC/m^3. This censoring is caused by the lower resolution of measurements for samples that have values less than 0.5 µgC/m^3. Both data sets are time series. The corresponding autocorrelation functions drop quickly to zero thus both time series can be considered as stationary.

8.1 Elemental carbon

First set of data consists of 740 measurements of elemental carbon. Unfortunately almost 58% of the data is left censored. Estimated parameters of GE distribution by maximum likelihood method for this set of data are $\hat{\alpha} = 2.53$ and $\hat{\lambda} = 2.58$. The histogram of the data with the corresponding p.d.f. of this distribution can be seen of figure 8.1.

Figure 8.2 shows the Q-Q plot (quantile-quantile plot) of the data and the fitted distribution. The censored part of the data is not shown.

The censoring of this sample was quite high thus the Pearson χ^2 test was the most appropriate one to test the goodness of fit. Figure 8.3 shows the comparison between theoretical distribution and the data set. The outcome of the Pearson χ^2 test was the following:

```
        Pearson's Chi-squared test
data: dataEC
X-squared = 5.8157, df = 13, p-value = 0.9526
```

The p-value $= 0.9526 > 0.05$ and that means that the hypothesis that the data are coming from GE distribution is not rejected.

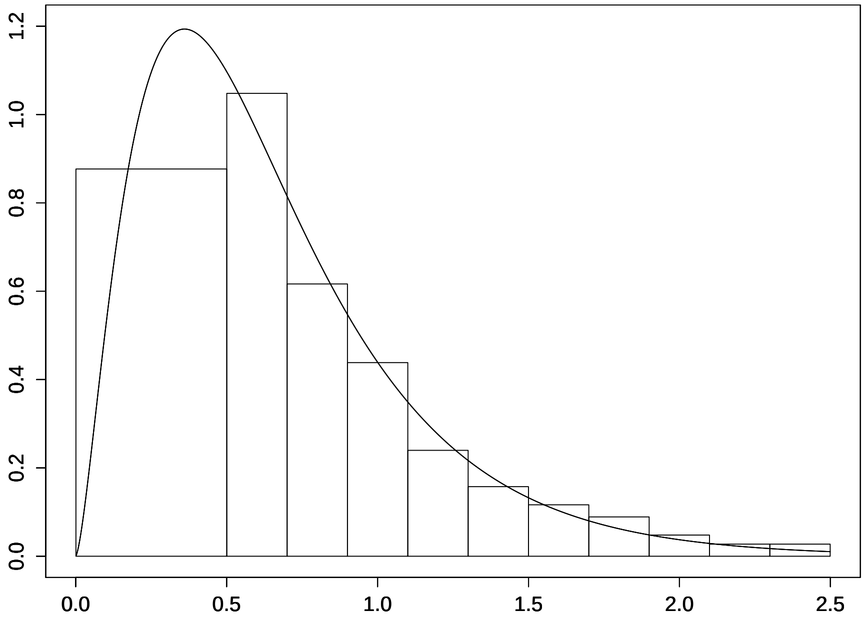

Fig. 8.1: Histogram of EC data and p.d.f. of fitted GE distribution

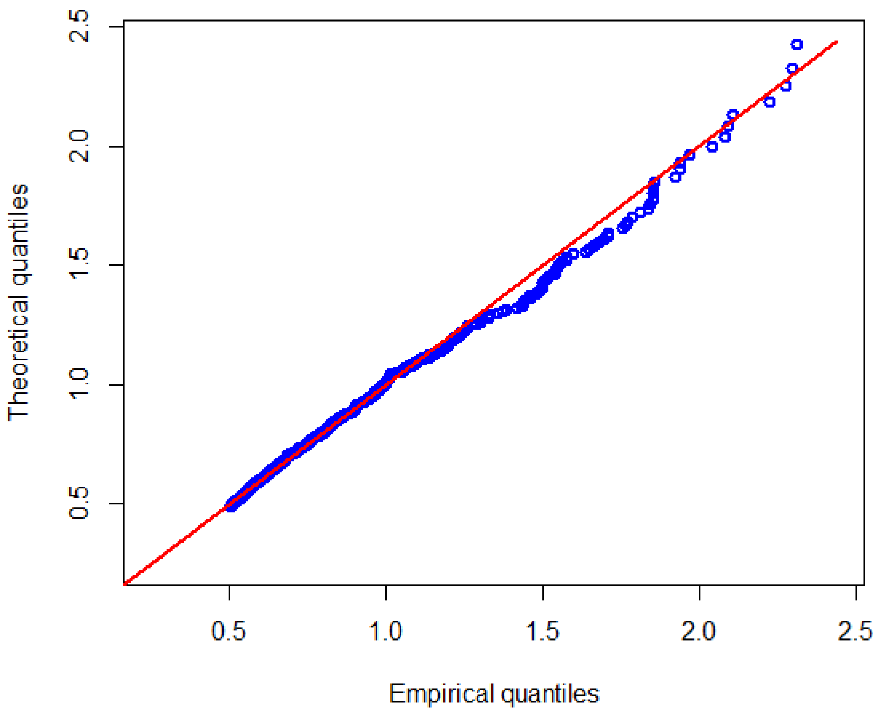

Fig. 8.2: Q-Q plot of EC

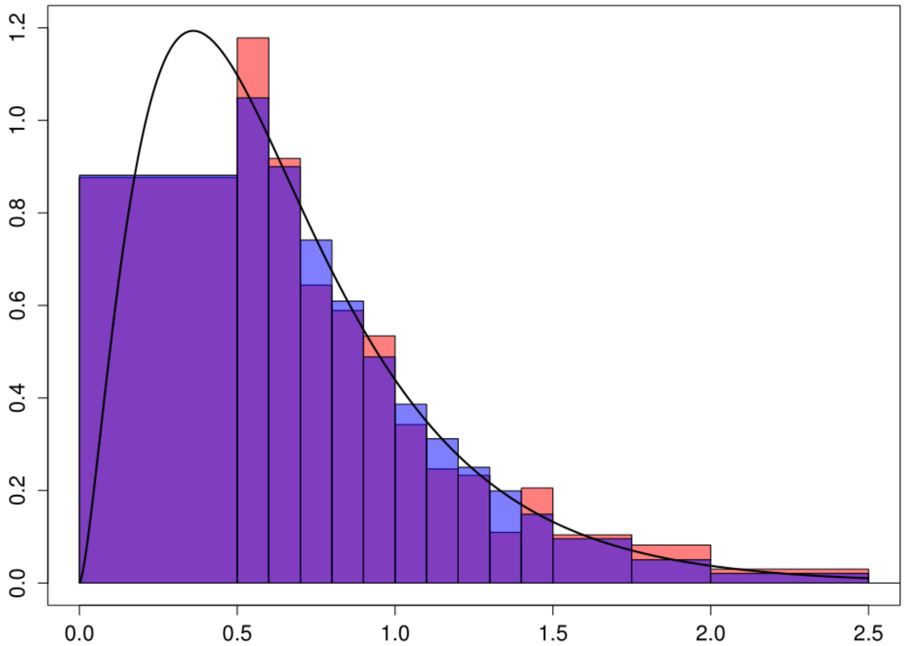

Fig. 8.3: Illustration of the Pearson χ^2 test on EC data

8.2 Organic carbon

Second set of data consists of 742 measurements of organic carbon. OC has higher concentrations in the air thus only two observations were censored during the experiment. Estimated parameters of GE distribution for this set of data are $\alpha = 4.61$ and $\lambda = 0.79$. The histogram of the data with the corresponding p.d.f. of this distribution can be seen of figure 8.4.

Figure 8.5 shows the Q-Q plot of the data and the fitted distribution. The censored part of the data is not shown.

There are only 2 censored values of 742. For this case is used Kolmogorov-Smirnov test as the goodness of fit test. The outcome of the Kolmogorov-Smirnov test was the following:

```
One-sample Kolmogorov-Smirnov test
data:  dataOC
D = 0.0296, p-value = 0.5502
alternative hypothesis: two-sided
```

The p-value $= 0.5502 > 0.05$ and that means that the hypothesis that the data are coming from GE distribution is not rejected.

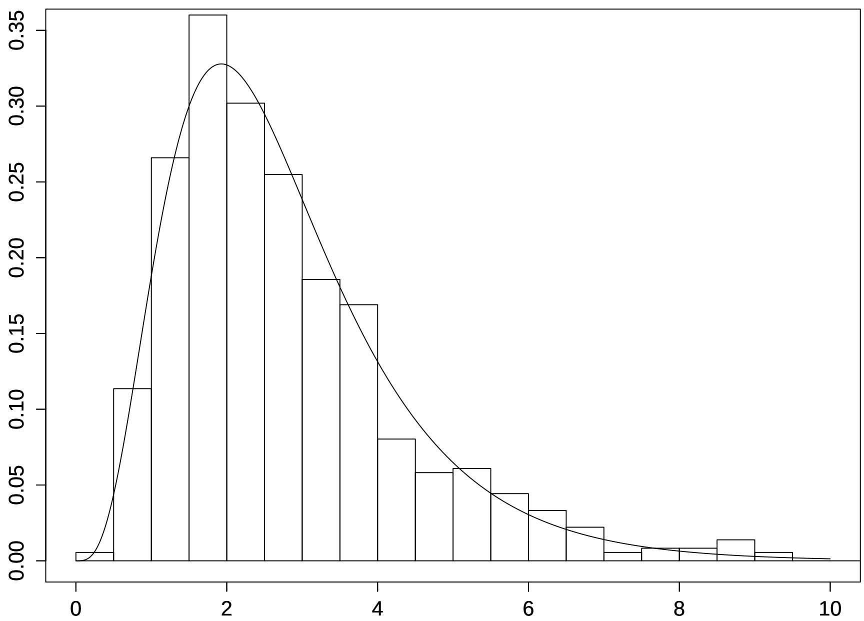

Fig. 8.4: Histogram of OC data and p.d.f. of fitted GE distribution

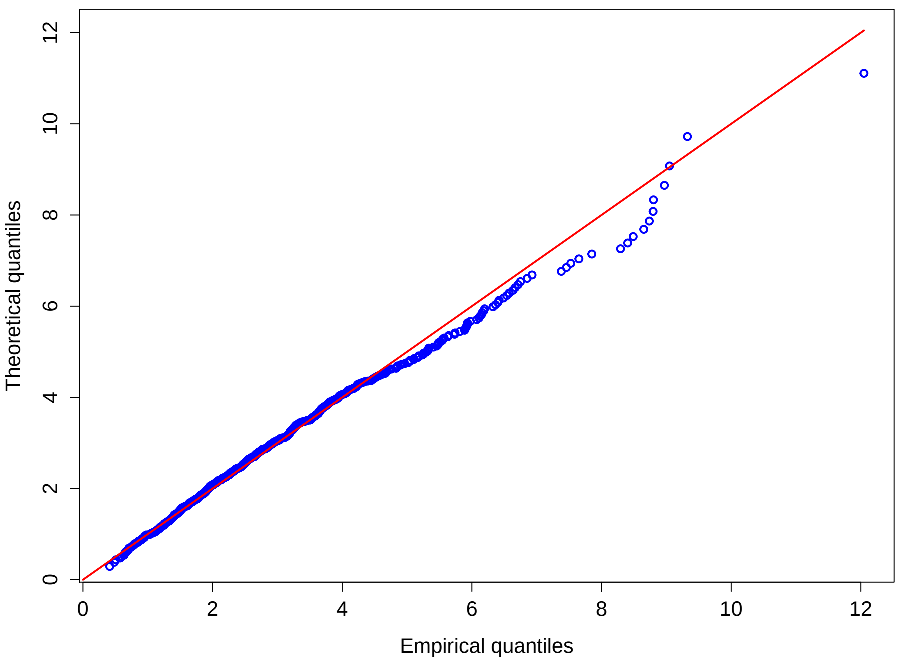

Fig. 8.5: Q-Q plot of OC

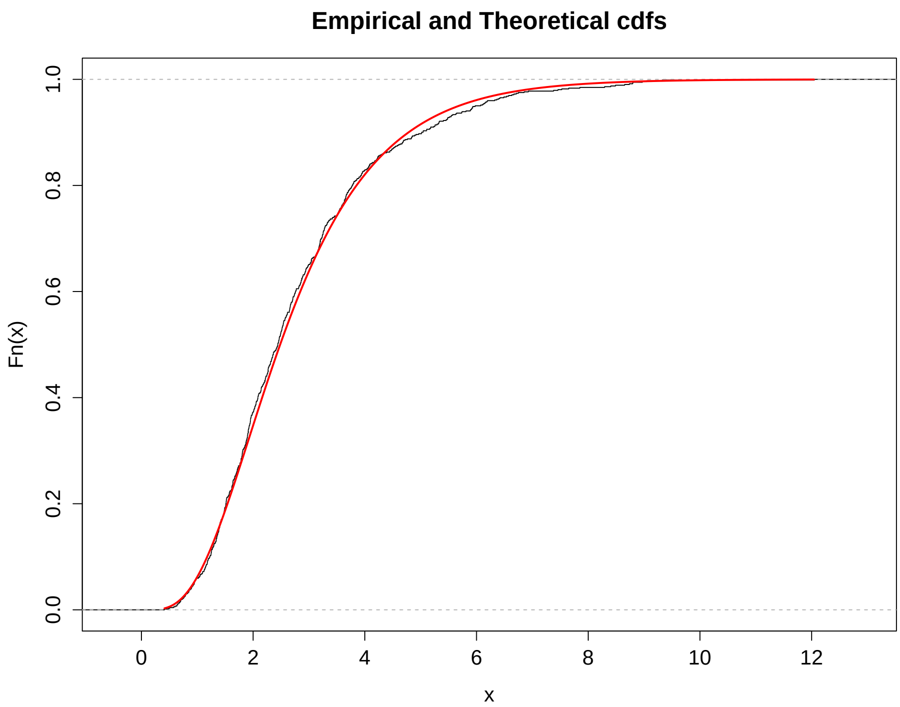

Fig. 8.6: Illustration of Kolmogorov–Smirnov test on OC data

Analysis of both samples showed that GE distribution can be used as an alternative distribution for analyzing unimodal, positively skewed data.

9 CONCLUSION

The thesis describes the GE distribution and its characteristics. Fundamental characteristics are described in chapter 2. Chapter 3 summarizes the methods of parameter estimations with special attention to the maximum likelihood estimations. Chapter 4 describes goodness of fit tests for this distributions and chapter 5 shows the comparisons among the GE distribution, Weibull distribution and log-normal distribution. Chapter 6 introduces the fundamentals of censoring and the maximum likelihood estimations with censored variables.

Chapter 6 contains section where I derived the maximum likelihood method for type I left censored data of the GE distribution. Type I left censored case has not been studied before. The maximum likelihood estimations are derived as well as their asymptotic behavior through the Fisher information matrix. The theoretically derived properties are verified by using simulations. Simulation results are presented in detail and the practical recommendations on sample censoring are done. Moreover observations of parameter sensitivity are done.

Chapter 7 introduces the EM algorithm that I developed for the GE distribution. EM algorithm can be used for simulations of various types of censoring. The performance of this algorithm is compared to the maximum likelihood method on the same type I left censored data. Although the numerical results on both algorithms are almost the same, it is shown that the EM algorithm is much more time-consuming than the derived maximum likelihood method. This shows the efficiency of the maximum likelihood approach.

As an example, environmental data of elemental carbon and organic carbon are presented. Data were collected at atmospheric station at Křešín u Pacova from August 2013 till December 2013. Analysis shows that GE distribution can be used as an alternative distribution for analyzing the environmental data.

BIBLIOGRAPHY

[1] ANDĚL, Jiří. *Základy matematické statistiky.* 3 edition Praha: Matfyzpress, 2011, 358 pg. ISBN 978-80-7378-162-0.

[2] ANDERSON, T. W., DARLING,D. A. *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes* Ann. Math. Statist. Volume 23, Number 2 (1952), 193–212.

[3] COHEN, A. *Truncated and censored samples: theory and applications.* New York: M. Dekker, c1991, xiv, 312 p. ISBN 08-247-8447-2.

[4] CHENG Cong-hua, CHEN Jin-yuan. *The MLE of GE distribution with interval censored data*, Journal of Lanzhou University, Vol. 48 No. 3, Jun. 2012

[5] DEMPSTER, A.P.; LAIRD, N.M.; RUBIN, D.B. (1977). *Maximum Likelihood from Incomplete Data via the EM Algorithm.* Journal of the Royal Statistical Society, Series B 39 : 1–38.

[6] DEY, Arabin Kumar, KUNDU, Debasis. *Discriminating Among the Log-Normal,Weibull and Generalized Exponential Distributions* Reliability, IEEE Transactions on, On page(s): 416–424, Vol. 58, Issue: 3, Sept. 2009

[7] GUPTA, Rameshwar D., KUNDU, Debasis. *Generalized Exponential Distributions*, Australian and New Zealand Journal of Statistics , Vol. 41, No. 2, 173–188, 1999.

[8] GUPTA, Rameshwar D., KUNDU, Debasis. *On comparison of the Fisher information of the Weibull and GE distributions.* Journal of the Statistical Planning and Inference, vol. 136, no. 9, 3130–3144, 2006.

[9] GUPTA, Rameshwar D., KUNDU, Debasis. *Discriminating between the Weibull and the GE distributions*, Computational Statistics and Data Analysis, vol. 43, 179–196. 2003.

[10] GUPTA, Rameshwar D., KUNDU, Debasis. *Generalized Exponential Distribution: Existing Results and Some Recent Developments.*, Journal of Statistical Planning and Inference, vol. 137, pages 3537–3547. 2007.

[11] GUPTA, R. D. and KUNDU, D. (2002), *Generalized exponential distributions: statistical inferences*, Journal of Statistical Theory and Applications, vol. 1, pages 101–118.

[12] GUPTA, R. D. and KUNDU, D. (2006), *On the comparison of Fisher information of the Weibull and GE distributions*, Journal of Statistical Planning and Inference, vol. 136, Issue 9, pages 3130—3144

[13] HASSAN, Amal S. *Goodness-Of-Fit For The Generalized Exponential Distribution.* 2005, Institute of Statistical Studies and Research, Cairo University

[14] HOGG, Robert V., CRAIG, Allen T. *Introduction to Mathematical Statistics.* New York: Macmillan, 1978. Print.

[15] LEHMANN, E. and CASELLA, G. *Theory of point estimation.* 2nd edition. New York: Springer, c1998, xxvi, 589 p. ISBN 03-879-8502-6.

[16] MCLACHLAN, Geoffrey J. *The EM algorithm and extensions.* 2nd ed. Hoboken: Wiley-Interscience, c2008. Series in probability and statistics (Wiley). ISBN 978-0-471-20170-0.

[17] MICHÁLEK, Jaroslav, Michal ŠMEREK a Jarmila ŠOTOVÁ. *Matematické modelování rizik.* Vyd. 1. Brno: Univerzita obrany, 2007. ISBN 978-80-7231-251-1.

[18] MITRA, Sharmishtha, KUNDU, Debasis. *Analysis of left censored data from the generalized exponential distribution.* Journal of Statistical Computation and Simulation. 2008, vol. 78, issue 7, s. 669–679. DOI: 10.1080/00949650701344158. Available on: http://www.tandfonline.com/doi/abs/10.1080/00949650701344158

[19] *The R Project for Statistical Computing.* [online]. [cit. 2013-11-23]. Available on: http://www.r-project.org/

[20] WU, C. F. Jeff. *On the Convergence Properties of the EM Algorithm.* Annals of Statistics 11 (1): 95-–103. doi:10.1214/aos/1176346060

[21] Weisstein, Eric W. *Order Statistic.* From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/OrderStatistic.html

LIST OF ABBREVIATIONS

i.i.d.	independent and identically distributed
p.d.f.	probability density function
c.d.f.	cumulative distribution function
GE distribution	Generalized exponential distribution

LIST OF APPENDICES

A PROPERTIES OF WEIBULL AND LOG-NORMAL DISTRIBUTION

A.1 Weibull distribution

The Weibull distribution is characterized by shape parameter τ and the scale parameter θ. The Weibull distribution has the cumulative distribution function:

$$F_W(x;\tau,\theta) = 1 - e^{-(\frac{x}{\theta})^\tau}, \qquad \tau,\theta > 0, \quad x \geq 0$$

and probability density function:

$$f_W(x;\tau,\theta) = \tau\theta^{-\tau}x^{\tau-1}e^{-(\frac{x}{\theta})^\tau}, \qquad \tau,\theta > 0, \quad x \geq 0$$

For a better picture of how the density function looks like for different τ and θ observe the figures A.2 and A.3.

The hazard function is given by:

$$h_W(x;\tau,\theta) = \frac{\tau}{\theta}\left(\frac{x}{\theta}\right)^{\tau-1}, \qquad \tau,\theta > 0, \quad x \geq 0$$

The different behavior of the hazard function is on figure A.1. Observe that for $\tau = 1$ the function is a constant function equal to $\frac{1}{\theta}$. For $\tau < 1$ the hazard function rises to infinity. For $\tau > 1$ the hazard function approaches zero.

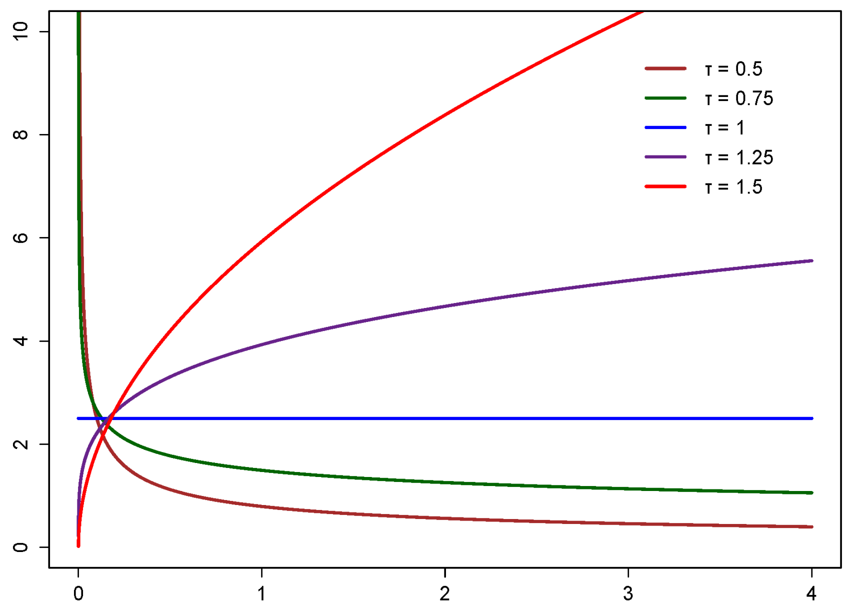

Fig. A.1: Hazard function of Weibull distriution with $\theta = 0.4$ fixed

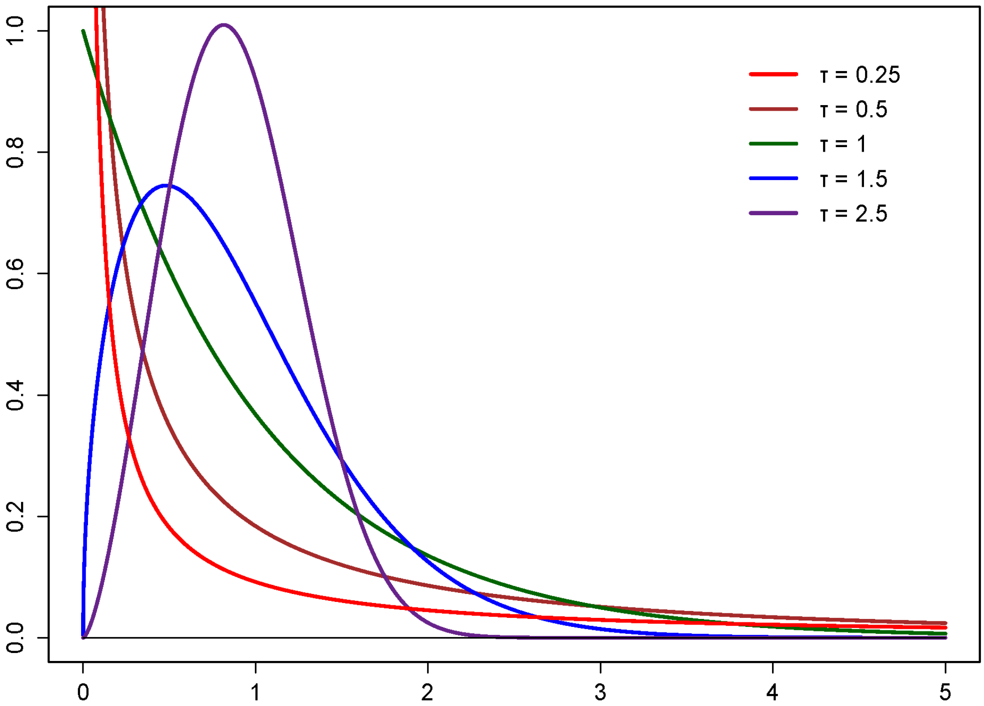

Fig. A.2: P.d.f. of Weibull distribution with $\theta = 1$ fixed

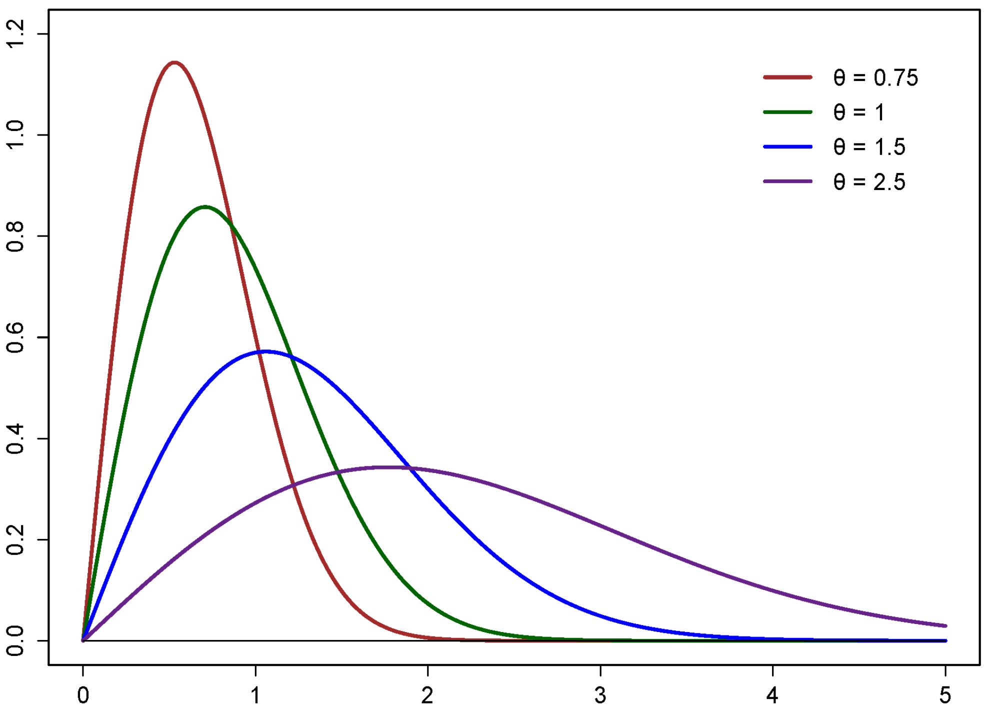

Fig. A.3: P.d.f. of Weibull distribution with $\tau = 2$ fixed

A.2 Log-normal distribution

The log-normal distribution is characterized by parameters μ and σ. The log-normal distribution has the cumulative distribution function:

$$F_{LN}(x;\alpha,\beta) = \phi\left(\frac{\ln x - \mu}{\sigma}\right), \qquad \mu \in \mathbb{R},\ \sigma > 0,\ x \geq 0$$

and probability density function:

$$f_{LN}(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \qquad \mu \in \mathbb{R},\ \sigma > 0,\ x \geq 0$$

For a better picture of how the density function looks like for different μ and σ observe the figures A.5 and A.6.

The hazard function is given by:

$$h_{LN}(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \left[\phi\left(\frac{\ln x - \mu}{\sigma}\right)\right]^{-1}, \qquad \mu \in \mathbb{R},\ \sigma > 0,\ x \geq 0$$

The behavior of the hazard function is on Figure A.4. The log-normal hazard function approaches zero for any parameters μ and σ.

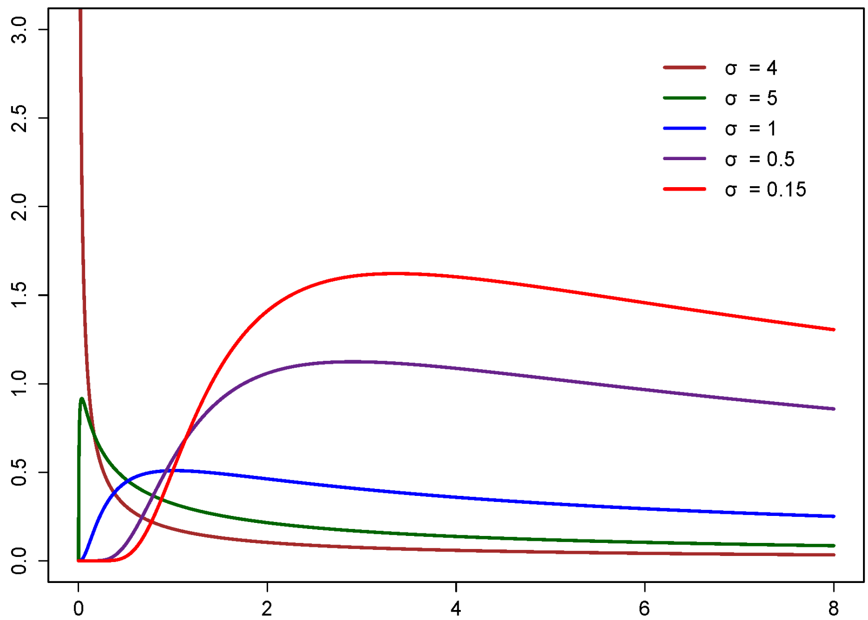

Fig. A.4: Hazard function of Log-normal distriution with $\mu = 0.5$ fixed

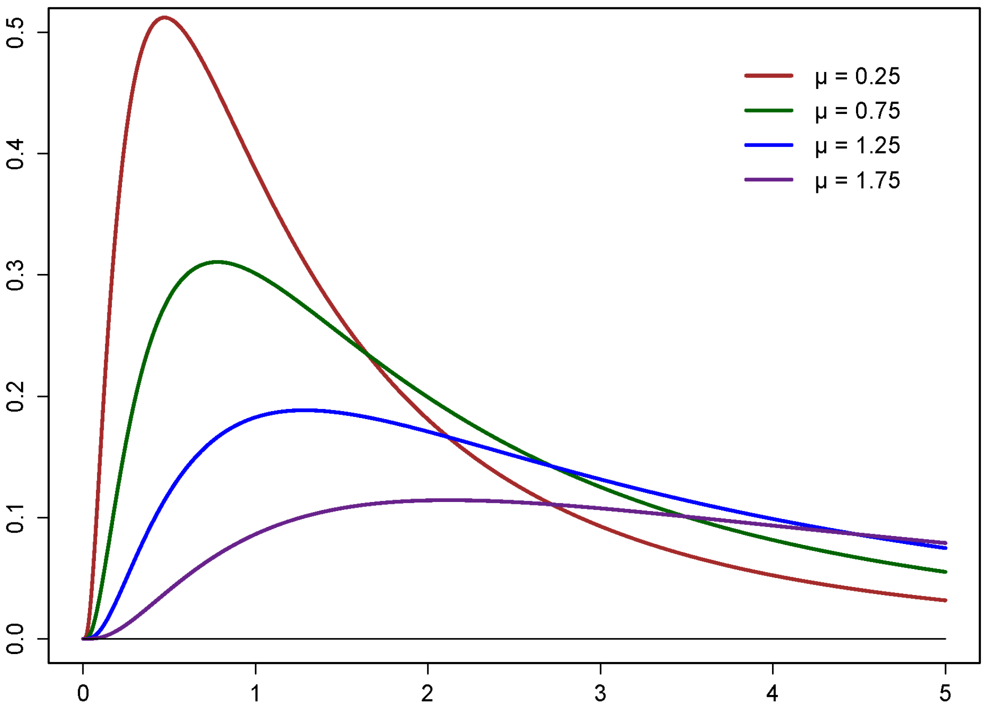

Fig. A.5: P.d.f. of Log-normal distriution with $\sigma = 1$ fixed

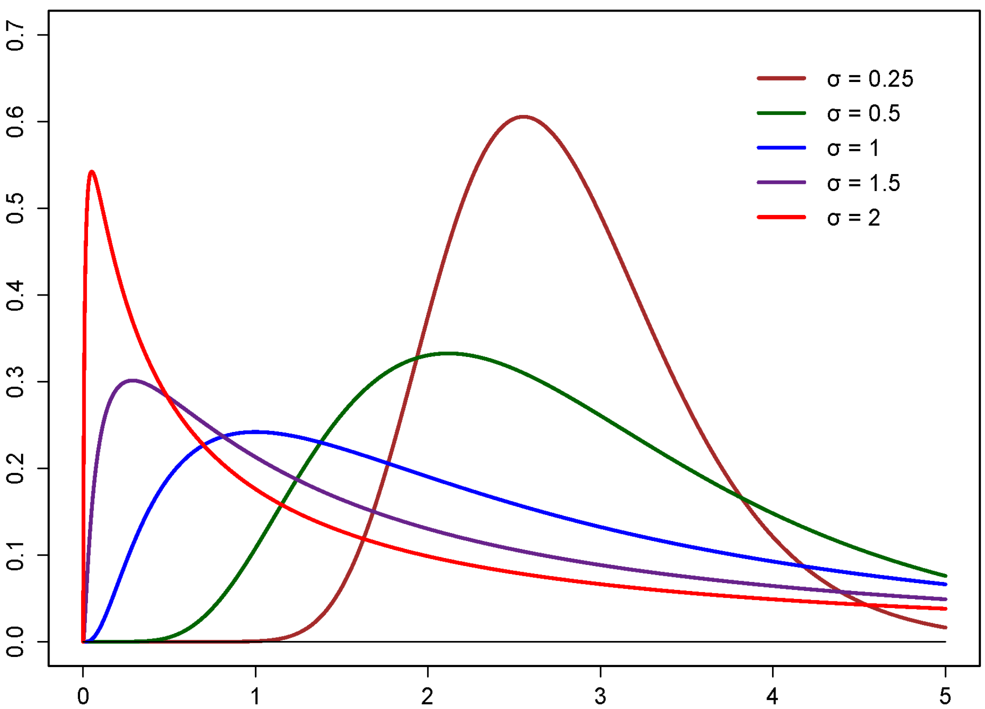

Fig. A.6: P.d.f. of Log-normal distriution with $\mu = 1$ fixed