FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

David Redek

Truncated data

Department of Probability and Mathematical Statistics

Supervisor of the bachelor thesis: doc. RNDr. Michal Pešta, Ph.D.
Study programme: General Mathematics
Study branch: Stochastics

Prague 2021

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

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

In .... date ....

Author's signature

I am grateful to my supervisor, doc. RNDr. Michal Pešta, Ph.D. for his time, effort, professional support, and swift responses whenever needed. I would like to thank my family for their endless support in my studies. Further, I would like to thank Sara for helping me find the motivation even in the darkest times of the coronavirus pandemic.

Title: Truncated data

Author: David Redek

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. RNDr. Michal Pešta, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This thesis deals with truncated distributions. Firstly, the case of the truncated normal distribution is analyzed in detail. Two estimation methods are described – method of moments and maximum likelihood – together with the discussion of their properties and confidence region construction. A more advanced method – Bayes estimator – is briefly presented. Secondly, the truncated gamma distribution is analyzed, however, in less detail than the normal case. The theoretical part is closed with a method for estimating truncation boundaries when not even those are known. Throughout the thesis, results from multiple articles by various authors are summarized and presented in a unified notation. The numerical part deals with the analysis of a real dataset, describing the height of soldiers in the U.S. Army. The built theory is transformed into R code and executed, indicating the correctness of our theoretical results.

Keywords: truncation, truncated normal distribution, truncated gamma distribution, maximum likelihood estimation, moment method estimation, moment generating function, confidence region.

Contents

Introduction

In statistics, we frequently work with real-world observations, forming conclusions about the process generating them. There are cases where the full range of certain processes cannot be completely observed. Instead, only a subset of the entire process is observed, while the rest remains unknown. Inability to observe the whole range may be caused, e.g., by technical limitations of measuring equipment or even natural reasons. The theory of truncated distributions then plays a crucial role.

In this thesis, we analyze the case of observations having upper and lower truncation boundary, meaning that the observations beyond those thresholds are not accounted for at all – unlike in the theory of censored samples, where information about the number of observations beyond the thresholds is present.

Even under truncation conditions, we still want to infer about the original data-generating process. In this thesis, we develop the needed theory. Chapter 1 states a motivation example with speed cameras and formally describes the general truncation problem.

In Chapters 2 and 3, we assume truncation boundaries to be known and analyze the case of truncated normal and gamma distribution, respectively, derive their expected value and higher moments together with their variance. For both distributions, we derive the method of moments and maximum likelihood estimators for their parameters. Properties of those estimators are then discussed, and asymptotic distribution with confidence region is derived for truncated normal distribution case.

Chapter 4 briefly describes techniques of truncation boundaries estimation when even those are unknown.

Chapter 5 applies the built theory to real data, presumably coming from the normal distribution. We analyze the height of the U.S. Army soldiers, estimate parameters in the model of truncated normal distribution, and construct confidence regions for the estimates.

1. The problem of truncation

1.1 Motivation example

Suppose we have a speed camera on a town road, with the speed limit being 50 km/h. Suppose the speed camera records the speed of the passing cars, whose speed follows a normal distribution with mean and standard deviation being 45 km/h and 10 km/h, respectively. Such an example is shown in the Figure 1.1.

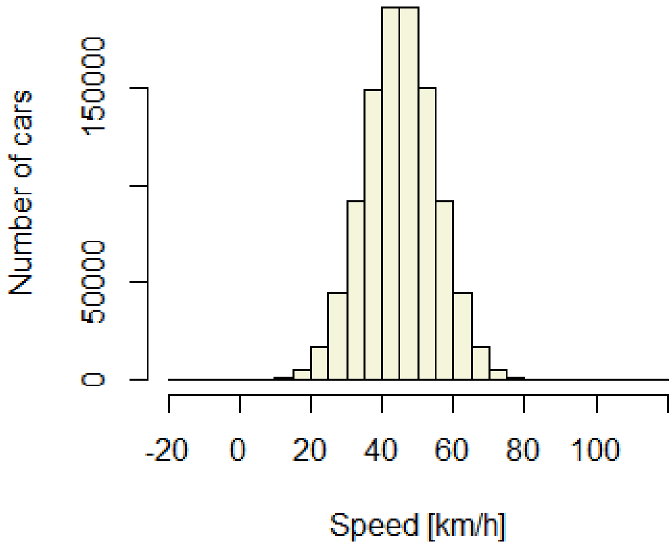

Figure 1.1: Histogram of the car speeds

Further, suppose the camera only records cars, violating the speed limit. Let us be generous to the not well-behaved drivers and give them a 5 km/h buffer zone. Our supposed camera is then only recording cars driving over 55 km/h, as can be seen in the Figure 1.2.

Suppose we want to estimate the average speed of all cars, based just on the known observations. Having all observations as in the Figure 1.1, the estimation would be straightforward, being just a sample mean. However, calculating the sample mean from the truncated observations would clearly be a biased estimate.

The goal of the thesis is to deal with the issues caused by truncation. We will develop the theory needed to estimate parameters from truncated samples. An example is to be seen in the Figure 1.3, where the truncated and original probability density function of a normal distribution with estimated parameters of mean and variance is plotted.

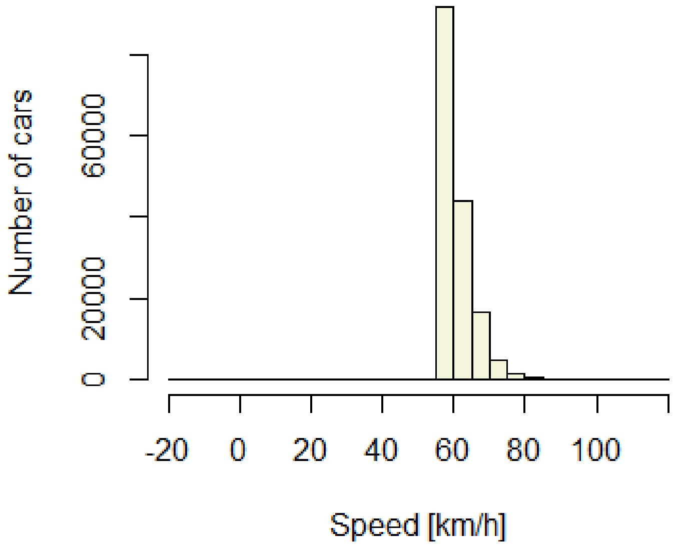

Figure 1.2: Histogram of the truncated car speeds

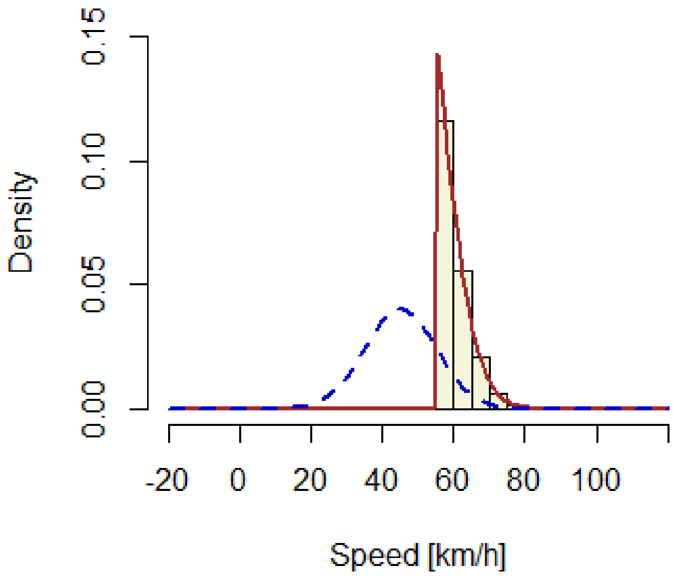

Figure 1.3: Histogram of the truncated car speeds with the truncated (red) and associated original (blue) normal probability density function.

1.2 General mathematical formulation

In this thesis, we focus on the truncation of continuous random variables. Truncation translates mathematically into either narrowing the support of the original probability density function or restricting the range of the original random variable. Simply put, the distribution remains the same on the not-truncated part of the density function support while being zero everywhere else. This leads to an issue – integrating over such a distribution does not result in the desired unit. Thus it is not a probability density function. We have to correct this. Firstly, let us define an indicator function

$$\mathbb{I}_{(a,b)}(x) = \begin{cases} 1 & \text{if } x \in (a, b) \\ 0 & \text{otherwise.} \end{cases}$$

Also, to recall an important property of cumulative distribution function:

$$\lim_{x\to\infty} F(x) = 1; \lim_{x\to-\infty} F(x) = 0$$

Now we can define the doubly truncated random variable.

Definition 1. *Let $a, b \in \mathbb{R}$ and $c, d \in \mathbb{R}^*$ such that $c < a < b < d$. Let X be a continuous random variable with a probability density function f_X on the support (c, d) and a cumulative distribution function F_X. Then, we say that a random variable Y with a probability density function*

$$f_Y(x) = f_X(x|a < X < b) = \frac{f_x(x)}{F(b) - F(a)}\mathbb{I}_{(a,b)}(y) \tag{1.1}$$

is a doubly truncated random variable with, truncation parameters a and b. We will denote such a random variable as $Y = X|a < X < b$.

Definition 2. *Let and $a, b, c, d \in \mathbb{R}^+$ such that $c \leq a < b \leq d$. Let X be a continuous random variable with a probability density function f_X on the support (c, d) and a cumulative distribution function F_X. We differentiate*

- *$a = c \wedge b < d$. Then, we say that a random variable Y with a probability density function*

  $$f_Y(x) = f_X(x|X < b) = \frac{f_x(x)}{F(b)}\mathbb{I}_{(c,b)}(y) \tag{1.2}$$

  is a singly upper (right) truncated random variable with, truncation parameter b. We will denote such a random variable as $Y = X|X < b$.

- *$a > c \wedge b = d$. Then, we say that a random variable Y with a probability density function*

  $$f_Y(x) = f_X(x|X > a) = \frac{f_x(x)}{1 - F(a)}\mathbb{I}_{(a,d)}(y) \tag{1.3}$$

  is a singly lower (left) truncated random variable with, truncation parameter a. We will denote such a random variable as $Y = X|X > a$.

- *$a = c \wedge b = d$. Then no truncation is applied.*

In practice we can meet samples from both singly and doubly truncated samples. This thesis focuses mainly on the doubly truncated case, since it is the most general one and the outcomes for singly truncated samples can be usually derived from them.

2. Truncated normal distribution

2.1 Introductory definitions

In this chapter, we will develop the theory for the truncated normal distribution. We chose normal distribution, as it is considered among the most important ones and offers wide variety of applications. We will investigate its moments, construct estimators for its parameters and discuss their properties.

We will denote the normal distribution with parameters of mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$ as $\mathsf{N}(\mu, \sigma^2)$. As a reminder, let us mention its probability density function

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, \quad x \in \mathbb{R}.$$

We will often use a special version, the standard normal distribution $\mathsf{N}(0, 1)$, denoting its probability density function and cumulative distribution function as ϕ and Φ, respectively. Let us recall some important properties we will use later:

$$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}; \\ f(x; \mu, \sigma^2) &= \frac{1}{\sigma}\phi\Big(\frac{x-\mu}{\sigma}\Big); \\ F(x; \mu, \sigma^2) &= \Phi\Big(\frac{x-\mu}{\sigma}\Big), \end{aligned}$$

where $F(x; \mu, \sigma^2)$ denotes the cumulative distribution function of $\mathsf{N}(\mu, \sigma^2)$.

Now, we can proceed to the truncated version of the normal distribution. We will assume doubly truncated normal distribution. Double truncation means that our original random variable is limited both for the upper (right) and lower (left) values. With the knowledge from Section 1.2, Definition 1, we can define doubly truncated normal distribution.

Definition 3. *Let $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$ and $a, b \in \mathbb{R}$ such, that $-\infty < a < b < \infty$. Let X be a random variable with normal distribution $\mathbb{N}(\mu, \sigma^2)$. Then we say, that a random variable $Y = X | a < X < b$ with a probability density function*

$$f(y; \mu, \sigma^2 | a < y < b) = \frac{1}{\sigma} \frac{\phi(\frac{y-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \mathbb{I}_{(a,b)}(y) \tag{2.1}$$

has a doubly truncated normal distribution $\mathsf{TN}(\mu, \sigma^2 | a, b)$.

Clearly, such a distribution does not have to possess the same properties as the original one, e.g., the expected value may change. As an example, assume $X \sim \mathsf{N}(0, 1)$ and a singly truncated random variable $Y = X | X > 0$, the expected value of Y will not be zero, but rather will be biased to the right, being greater than zero. Let us look at this difference in the Figure 2.1.

In the next section, we will further describe the properties of the truncated normal distribution.

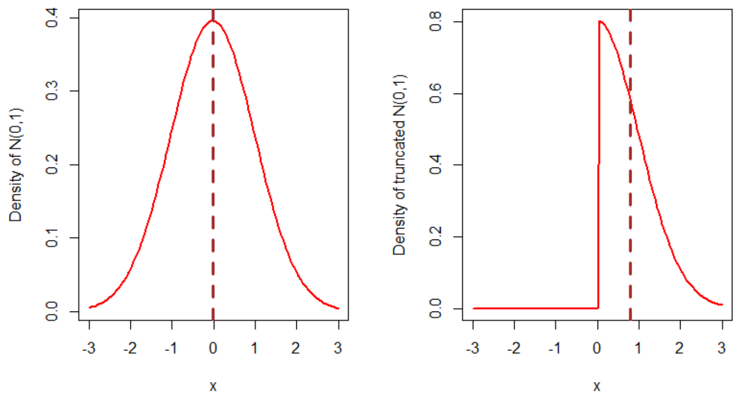

Figure 2.1: Densities of the original random variable $\mathsf{N}(0,1)$ and the truncated variant.

2.2 Calculation of moments

Let $X \sim \mathsf{TN}(\mu, \sigma^2|a, b)$. We will use a moment generating function $\varphi_X(t) = \mathsf{E}\, e^{tX}$ to find the expected value and variance.

$$\begin{aligned}\varphi_X(t) = \mathsf{E}[e^{tX}] &= \int_a^b e^{tx} f(x; \mu, \sigma^2 | a < x < b)\, dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \int_a^b e^{tx} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx.\end{aligned}$$

Since we are dealing with a long expression, we will now narrow our focus to the part with the integral. Using substitution $z = \frac{x-(t\sigma^2+\mu)}{\sigma}$, we arrive at

$$\begin{aligned}\int_a^b e^{tx} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\, dx &= \int_a^b e^{-\frac{(x-(t\sigma^2+\mu))^2+\mu^2-(t\sigma^2+\mu)^2}{2\sigma^2}}\, dx \\ &= e^{\frac{-\mu^2+(\sigma^2 t+\mu)^2}{2\sigma^2}} \int_a^b e^{-\frac{-(x-(t\sigma^2+\mu))^2}{2\sigma^2}}\, dx \\ &= \sigma e^{\frac{-\mu^2+(\sigma^2 t+\mu)^2}{2\sigma^2}} \int_{\frac{a-(t\sigma^2+\mu)}{\sigma}}^{\frac{b-(t\sigma^2+\mu)}{\sigma}} e^{-\frac{1}{2}z^2}\, dz \\ &= \sigma e^{\frac{-\mu^2+(\sigma^2 t+\mu)^2}{2\sigma^2}} \left(\Phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right) - \Phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right)\right).\end{aligned}$$

All combined together results in

$$\varphi_X(t) = e^{\mu t + \frac{t^2\sigma^2}{2}} \frac{\Phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right) - \Phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}. \tag{2.2}$$

Now we can calculate the derivative

$$\begin{aligned}\varphi'_X(t) = {} & e^{\mu t + \frac{t^2\sigma^2}{2}} (\mu + t\sigma^2) \frac{\Phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right) - \Phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \\ & + \sigma e^{\mu t + \frac{t^2\sigma^2}{2}} \frac{\phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right) - \phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\end{aligned}$$

and get the expected value

$$\mathsf{E}\,X = \varphi_X'(0) = \mu + \sigma\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}. \tag{2.3}$$

Since we already have the moment generating function, we can easily calculate the second moment $\mathsf{E}\,X^2$ and variance $\mathsf{var}\,X$. For this purpose, we will calculate the second derivative of the moment generating function

$$\begin{aligned}\varphi_X''(t) = (e^{\mu t+\frac{t^2\sigma^2}{2}}(\mu+t\sigma^2)^2 + e^{\mu t+\frac{t^2\sigma^2}{2}}\sigma^2)&\frac{\Phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right) - \Phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\\
&+ 2\sigma e^{\mu t+\frac{t^2\sigma^2}{2}}(\mu+t\sigma^2)\frac{\phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right) - \phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\\
&+ \sigma e^{\mu t+\frac{t^2\sigma^2}{2}}\frac{\phi\left(\frac{a-(t\sigma^2+\mu)}{\sigma}\right)(a-t\sigma^2-\mu) - \phi\left(\frac{b-(t\sigma^2+\mu)}{\sigma}\right)(b-t\sigma^2-\mu)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}.\end{aligned}$$

Similarly to the expected value, we use the formula $\mathsf{E}\,X^2 = \varphi_X''(0)$ and by further derivations we arrive at

$$\mathsf{E}\,X^2 = \mu^2 + \sigma^2 + 2\mu\sigma\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} + \sigma\frac{\phi\left(\frac{a-\mu}{\sigma}\right)(a-\mu) - \phi\left(\frac{b-\mu}{\sigma}\right)(b-\mu)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}. \tag{2.4}$$

Finally, we can calculate the variance

$$\begin{aligned}\mathsf{var}\,X &= \mathsf{E}\,X^2 - (\mathsf{E}\,X)^2 = \varphi_X''(0) - (\varphi_X'(0))^2\\
&= \sigma^2\left[1 + \frac{\phi\left(\frac{a-\mu}{\sigma}\right)\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2\right].\end{aligned} \tag{2.5}$$

2.3 Moment method estimation

Knowledge gained in the last section will enable us to build a moment method estimator for the parameters μ and σ^2.

Suppose we have a random sample $X_1, \ldots, X_n$ of size n from a doubly truncated normal distribution $\mathsf{TN}(\mu, \sigma^2|a, b)$, where $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$ are unknown parameters and truncation parameters $-\infty < a < b < \infty$ are assumed to be known.

Let's denote the real parameter $\boldsymbol{\theta}_X = (\mu, \sigma^2)$ and the parameter space $\boldsymbol{\Theta} \subset \mathbb{R} \times \mathbb{R}^+$. We will be working with the model

$$\mathcal{F} = \{\text{distributions with truncated normal density } f(x; \boldsymbol{\theta}_X|a, b);\ \boldsymbol{\theta}_X \in \boldsymbol{\Theta}\}.$$

Let's assume that the expected value and variance are a functions of parameters, i.e., $(\mathsf{E}\,X_i, \mathsf{var}\,X_i) = \boldsymbol{\tau}(\boldsymbol{\theta}_X)$, where $\boldsymbol{\tau} = (\tau_1, \tau_2)$ such, that $\tau_i : \Theta \to \mathbb{R}$. This is a reasonable assumption, since in Section 2.2 we found out, that even in the truncated normal distribution, the expected value and the variance are indeed functions of unknown parameters μ, σ^2.

Let us also denote a sample mean $\overline{X_n}$ and a sample variance s_n^2

$$\begin{aligned} \overline{X_n} &= \frac{1}{n}\sum_{i=1}^{n} X_i; \\ s_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^2. \end{aligned} \tag{2.6}$$

Both are consistent estimators of the expected value and $\overline{X_n}$ is also unbiased.

Estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$ of parameter $\boldsymbol{\theta}_X$ is then a solution of the system of equations

$$\begin{aligned} \overline{X_n} &= \tau_1(\hat{\boldsymbol{\theta}}_n); \\ s_n^2 &= \tau_2(\hat{\boldsymbol{\theta}}_n). \end{aligned}$$

In our case, we get the system

$$\begin{aligned} \overline{X_n} &= \hat{\mu}_n + \hat{\sigma}_n \frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)}; \\ s_n^2 &= \hat{\sigma}_n^2 \left[1 + \frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)} - \left(\frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)}\right)^2\right]. \end{aligned} \tag{2.7}$$

Obviously, τ_1, τ_2 are not injective functions and cannot be inverted. Inability to obtain analytical solution leads us to numerical methods with a possible solution being discussed later in the text. Properties of this estimator will be covered in the section dealing with the maximum likelihood estimator.

Notice that parameters μ and σ^2 in this context do not stand for the mean and the variance of the truncated distribution of our random sample. Instead, we are trying to estimate the mean and the variance of the original data-generating normal distribution we, for some reason, cannot fully observe – hence we have to use the theory of truncated distributions.

2.4 Maximum likelihood estimation

2.4.1 Score function

Now we will construct a maximum likelihood estimator (*MLE*) for parameters μ and σ^2. Again, suppose we have a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ of size n from a doubly truncated normal distribution $\mathsf{TN}(\mu, \sigma^2 | a, b)$, where $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$ are unknown parameters and the truncation parameters $-\infty < a < b < \infty$ are assumed to be known.

To briefly recall, the theory of maximum likelihood is bound to give us "the best" possible estimator of a given parameter. By "the best" we mean a consistent estimator with the lowest variance possible among consistent estimators. In other words – it is the statistically efficient estimator. We define a likelihood function

$\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{X})$ for the parameter $\boldsymbol{\theta} = (\mu, \sigma^2)$ and then find the estimator as $\hat{\boldsymbol{\theta}}_n = \operatorname{argmax}_{\theta \in \Theta} \mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{X})$, where $\Theta = \mathbb{R} \times \mathbb{R}^+$ is the parameter space. The model $\mathcal{F}$ is the same as in the Section 2.3.

$$\begin{aligned}\mathcal{L}_n(\mu, \sigma^2; \boldsymbol{X}) = \prod_{i=1}^{n} f(X_i) &= \prod_{i=1}^{n} \frac{1}{\sigma} \frac{\phi(\frac{X_i - \mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \\ &= (\sigma\sqrt{2\pi})^{-n} \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right)^{-n} e^{-\frac{1}{2}\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2}.\end{aligned} \tag{2.8}$$

Note that we do not have to write the indicator function since we assume a random sample from the truncated distribution.

Now we proceed to the log-likelihood

$$\begin{aligned}\ell_n(\mu, \sigma^2) &= \log \mathcal{L}_n(\mu, \sigma^2) \\ &= -n \log \sigma\sqrt{2\pi} - n \log \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2.\end{aligned} \tag{2.9}$$

Let us calculate the derivatives

$$\begin{aligned}\frac{\partial \ell_n}{\partial \mu} &= \frac{n}{\sigma} \frac{\phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + \frac{1}{\sigma}\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}; \\ \frac{\partial \ell_n}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} - \frac{n}{2\sigma^3} \frac{\phi(\frac{a-\mu}{\sigma})(a-\mu) - \phi(\frac{b-\mu}{\sigma})(b-\mu)}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2.\end{aligned}$$

By further adjusting those expressions and setting them to zero, we arrive at a system of two non-linear equations,

$$\begin{aligned}\overline{X_n} &= \hat{\mu}_n + \hat{\sigma}_n \frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)}; \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2 &= \hat{\sigma}_n^2 + \hat{\sigma}_n^2 \frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)}.\end{aligned}$$

Interestingly, our final maximum likelihood estimator is the same as the one obtained by the method of moments. Note that this is the case also with the original, full normal distribution. To prove the equivalency, we need to make an observation that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X_n} + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^2 + \sigma^2 \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2.\end{aligned}$$

With this observation, arriving at the same system of equations, as in (2.7), is only a matter of rearranging terms.

Maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$ of parameter $\boldsymbol{\theta}_X$ is then again the solution of the system of equations

$$\overline{X_n} = \hat{\mu}_n + \hat{\sigma}_n \frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)};$$

$$s_n^2 = \hat{\sigma}_n^2 \left[1 + \frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)} - \left(\frac{\phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right)}{\Phi\left(\frac{b-\hat{\mu}_n}{\hat{\sigma}_n}\right) - \Phi\left(\frac{a-\hat{\mu}_n}{\hat{\sigma}_n}\right)}\right)^2\right].$$

The system can be solved numerically, e.g., by the Newton method.

2.4.2 Properties of the estimators

We have found out that the method of moments estimator is equivalent to the maximum likelihood estimator. Therefore, it is enough to describe the properties of the MLE since those will apply to the method of moment estimators as well. We are now assuming that both a, b are known and both μ, σ^2 are unknown.

In this section we will follow a dissertation by M. M. Mittal. We chose this work, since it describes the problematic thoroughly, while being easy to understand. Mittal [1984] provides thorough discussion on the issue and proves that there are cases, where maximum likelihood estimator cannot be found with the previously described procedure. Let us go back to the system (2.7)

$$\overline{X_n} = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)};$$

$$s_n^2 = \sigma^2 \left[1 + \frac{\phi\left(\frac{a-\mu}{\sigma}\right)\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2\right],$$

and denote the right sides of these equations as

$$\overline{X_n} = \mu_T(\mu, \sigma); \tag{2.10}$$
$$s_n^2 = \sigma_T^2(\mu, \sigma). \tag{2.11}$$

Notice that the right sides are the expected value and the variance of the truncated normal distribution.

Mittal [1984] provides a discussion on the asymptotic behaviour of the equations (2.10), (2.11). He shows, that for a fixed parameter σ, μ_T is an increasing function of μ. Also, $\lim_{\mu \to -\infty} \mu_T = a$ and $\lim_{\mu \to +\infty} \mu_T = b$, for a fixed σ. Further, for a fixed μ, the $\lim_{\sigma \to 0} \mu_T$ is either a, b or μ, depending on if μ lies to the left of the interval (a, b), to the right of this interval or within, respectively. Thus (2.10) does not pose any nonexistence issues, since both sides of the equation acquire the same range of values, $a < \mu_T, \overline{X_n} < b$.

Similarly, when parameter μ is fixed, σ_T^2 is an increasing function of σ and $\lim_{\sigma \to 0} \sigma_T^2 = 0$, since the truncated distribution (2.1) degenerates at a, b or μ, depending on if μ lies to the left, to the right or within the interval respectively. Function σ_T^2 is then a variance of degenerate distribution, therefore zero. Further, $\lim_{\sigma \to \infty} \sigma_T^2 = (b-a)^2/12$. This value seems natural, since $(b-a)^2/12$ is the variance of a uniform distribution over interval (a, b). For a fixed σ and μ approaching either ∞ or $-\infty$, the expression σ_T^2 approaches zero, since the probability density function (2.1) then degenerates at a or b respectively.

Mittal [1984] then discusses cases, when for both (2.10) and (2.11), parameters μ and σ vary simultaneously. However, this part does not bring any new restrictions on either μ_T or σ_T^2, thus is not described further in detail. In any case, it holds that both μ_T and $\overline{X_n}$ are in the interval (a, b). Unfortunately this is not the case for σ_T^2 and s_n^2. We have found out, that $0 < \sigma_T^2 < (b-a)^2/12$, while s_n^2 can acquire values up to $(b-a)^2/4$. This occurs when one half of the observations lie on the lower truncation boundary a and the other half on the upper truncation boundary b. Thus, when the value of s_n^2 exceeds the threshold $(b-a)^{/}12$, the solution to the (2.11) does not exist.

However, this does not prove, that otherwise the solution exists. We have only derived conditions for the non-existence. In order to be able to estimate the parameters the way we derived, i.e., by the derivative of the log-likelihood, we need to fulfill some regularity conditions. We will state them for a vector parameter as they are in Anděl [2007].

(R1) Let $\Omega \subset \mathbb{R}^m$ be a parameter space, containing non-empty, open interval ω, such that the real parameter $\theta \in \omega$.

(R2) Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample, where X_i has the probability density function $f(x, \theta)$, given some σ-finite measure μ.

(R3) Let $M = \{x : f(x, \theta) > 0\}$ be independent of θ.

(R4) Let θ_1, $\theta_2 \in \Omega$. Then $f(x, \theta_1) = f(x, \theta_2)$, if and only if $\theta_1 = \theta_2$.

Having states these, we could still break, e.g., the (R1) regularity condition, since we cannot verify that the actual parameter can be covered by the open interval from Ω. Mittal [1984] further references here to work by Nielsen [2014], where sufficient conditions for the existence can be found. Mittal [1984] comments on them, stating that they are too complex, being hard to verify.

However, Mittal [1984] also states that the probability of nonexistence tends to zero as n approaches ∞. In the following, we will assume that the regularity conditions are fulfilled, and the MLE exists with probability one, at least asymptotically.

2.4.3 Fisher information matrix

As stated in the Section 2.4.2, we assume that the regularity conditions are fulfilled, and the solution can be found by the procedure described in the Section 2.4.1. Assuming this, it can be proved [Anděl, 2007] that MLE is asymptotically efficient. Estimators $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ have then asymptotically bivariate normal distribution. To find this distribution, we will now calculate the Fisher information matrix.

Let us first calculate the Hessian of the log-likelihood equations, denoted as

$$\mathbf{H}_n = \begin{pmatrix} \frac{\partial^2 \ell_n}{\partial \mu \partial \mu} & \frac{\partial^2 \ell_n}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 \ell_n}{\partial \mu \partial \sigma^2} & \frac{\partial^2 \ell_n}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix}.$$

For simplicity, let us denote

$$\alpha = \frac{a-\mu}{\sigma}; \quad \beta = \frac{b-\mu}{\sigma}.$$

With this notation, we will calculate the second partial derivatives

$$\begin{aligned}
\frac{\partial^2 \ell_n}{\partial^2 \mu} &= -\frac{n}{\sigma^2}\left[1 + \frac{\phi(\beta)\beta - \phi(\alpha)\alpha}{\Phi(\beta)-\Phi(\alpha)} - \left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right]; \\
\frac{\partial^2 \ell_n}{\partial \mu \partial \sigma^2} &= \frac{n}{2\sigma^3}\frac{(\phi(\alpha)-\phi(\beta))(\phi(\alpha)\alpha - \phi(\beta)\beta)}{(\Phi(\beta)-\Phi(\alpha))^2} - \frac{n}{2\sigma^3}\frac{\phi(\beta)(\beta^2+1) - \phi(\alpha)(\alpha^2+1)}{\Phi(\beta)-\Phi(\alpha)} \\
&\quad - \frac{1}{\sigma^4}\sum_{i=1}^n (X_i - \mu); \\
\frac{\partial^2 \ell_n}{\partial^2 \sigma^2} &= \frac{n}{2\sigma^4} + \frac{n}{2\sigma^4}\left[\frac{\phi(\alpha)\alpha - \phi(\beta)\beta}{\Phi(\beta)-\Phi(\alpha)}\right]^2 \\
&\quad + \frac{n}{4\sigma^4}\frac{3(\phi(\beta)\beta - \phi(\alpha)\alpha) + \phi(\beta)\beta^3 - \phi(\alpha)\alpha^3}{\Phi(\beta)-\Phi(\alpha)} - \frac{1}{\sigma^6}\sum_{i=1}^n (X_i-\mu)^2.
\end{aligned} \tag{2.12}$$

We can calculate the Fisher information matrix in the whole sample [Anděl, 2007] as

$$\mathbf{J}_n(\boldsymbol{\theta}) = -\mathsf{E}[\mathbf{H}_n] = \begin{pmatrix} J_n^{1,1}(\boldsymbol{\theta}) & J_n^{1,2}(\boldsymbol{\theta}) \\ J_n^{2,1}(\boldsymbol{\theta}) & J_n^{2,2}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} -\mathsf{E}\,\frac{\partial^2 \ell_n}{\partial^2 \mu} & -\mathsf{E}\,\frac{\partial^2 \ell_n}{\partial \mu \partial \sigma^2} \\ -\mathsf{E}\,\frac{\partial^2 \ell_n}{\partial \sigma^2 \partial \mu} & -\mathsf{E}\,\frac{\partial^2 \ell_n}{\partial^2 \sigma^2} \end{pmatrix}.$$

Notice, that most of the terms in (2.12) do not depend on any random variable. This results in rather easy computation of expectations. Moments needed are even pre-calculated in the Section 2.2. Taking those results into account, we arrive at

$$\begin{aligned}
\mathsf{E}\sum_{i=1}^n (X_i - \mu) &= \sum_{i=1}^n \mathsf{E}\,X_i - n\mu = n(\mathsf{E}\,X_i - \mu) \\
&= n\sigma\frac{\phi(\alpha)-\phi(\beta)}{\Phi(\beta)-\Phi(\alpha)}; \\
\mathsf{E}\sum_{i=1}^n (X_i-\mu)^2 &= \sum_{i=1}^n \mathsf{E}\,X_i^2 - 2\mu\sum_{i=1}^n \mathsf{E}\,X_i + n\mu^2 \\
&= n\sigma^2\left[1 + \frac{\phi(\alpha)\alpha - \phi(\beta)\beta}{\Phi(\beta)-\Phi(\alpha)}\right].
\end{aligned} \tag{2.13}$$

Here we used the knowledge of $\mathsf{E}\,X_i$ and $\mathsf{E}\,X_i^2$ from (2.3) and (2.4) respectively.

With these observations, we can calculate the mentioned expectation over (2.12)

$$
\begin{aligned}
J_n^{1,1} &= \frac{n}{\sigma^2}\left[1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)}\right)^2\right];\\
J_n^{1,2} &= \frac{n}{\sigma^3}\left[\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} + \frac{1}{2}\frac{\phi(\beta)(1+\beta^2) - \phi(\alpha)(1+\alpha^2)}{\Phi(\beta) - \Phi(\alpha)}\right.\\
&\qquad \left. - \frac{1}{2}\frac{(\phi(\alpha) - \phi(\beta))(\alpha\phi(\alpha) - \beta\phi(\beta))}{[\Phi(\beta) - \Phi(\alpha)]^2}\right];\\
J_n^{2,2} &= \frac{n}{2\sigma^4}\left\{1 + \frac{\phi(\alpha)\alpha - \phi(\beta)\beta}{\Phi(\beta) - \Phi(\alpha)} - \frac{1}{2}\left[\frac{\phi(\alpha)\alpha - \phi(\beta)\beta}{\Phi(\beta) - \Phi(\alpha)}\right]^2\right.\\
&\qquad \left. - \frac{1}{2}\frac{\phi(\beta)(3\beta+\beta^3) - \phi(\alpha)(3\alpha+\alpha^3)}{\Phi(\beta) - \Phi(\alpha)}\right\}.
\end{aligned}
$$

From Anděl [2007] we know that for a random sample, it holds that $\mathbf{J}_n(\boldsymbol{\theta}) = n\mathbf{J}(\boldsymbol{\theta})$. This allows us to easily compute the Fisher information matrix for one random variable $\mathbf{J}(\boldsymbol{\theta})$, needed for the asymptotic distribution of the estimators $\hat{\boldsymbol{\theta}}_n$.

$$
\begin{aligned}
J^{1,1} &= \frac{1}{\sigma^2}\left[1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)}\right)^2\right];\\
J^{1,2} &= \frac{1}{\sigma^3}\left[\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} + \frac{1}{2}\frac{\phi(\beta)(1+\beta^2) - \phi(\alpha)(1+\alpha^2)}{\Phi(\beta) - \Phi(\alpha)}\right.\\
&\qquad \left. - \frac{1}{2}\frac{(\phi(\alpha) - \phi(\beta))(\alpha\phi(\alpha) - \beta\phi(\beta))}{[\Phi(\beta) - \Phi(\alpha)]^2}\right];\\
J^{2,2} &= \frac{1}{2\sigma^4}\left\{1 + \frac{\phi(\alpha)\alpha - \phi(\beta)\beta}{\Phi(\beta) - \Phi(\alpha)} - \frac{1}{2}\left[\frac{\phi(\alpha)\alpha - \phi(\beta)\beta}{\Phi(\beta) - \Phi(\alpha)}\right]^2\right.\\
&\qquad \left. - \frac{1}{2}\frac{\phi(\beta)(3\beta+\beta^3) - \phi(\alpha)(3\alpha+\alpha^3)}{\Phi(\beta) - \Phi(\alpha)}\right\}.
\end{aligned}
\tag{2.14}
$$

2.4.4 Asymptotic distribution and confidence interval

From the theory of the maximum likelihood estimators, e.g., Anděl [2007], we can finally describe the asymptotic distribution of our estimator $\hat{\boldsymbol{\theta}}_n$ (under the previously stated assumptions of existence) as

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_X) \xrightarrow[n\to\infty]{d} \mathsf{N}_2(0, [\mathbf{J}(\boldsymbol{\theta})]^{-1}).$$

Inverse of $\mathbf{J}(\boldsymbol{\theta})$ can be evaluated easily with the well-known formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We will now use the result proved in Anděl [2007], describing test statistics based on the MLE. Let us assume a scenario of hypothesis testing – the null hypothesis $H_0 : \boldsymbol{\theta}_X = \boldsymbol{\theta}_0$ against the alternative $H_1 : \boldsymbol{\theta}_X \neq \boldsymbol{\theta}_0$ with a significance level α. Let us define the Wald test statistic

$$W_n = n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \boldsymbol{J}(\hat{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \tag{2.15}$$

where the $\boldsymbol{J}(\hat{\boldsymbol{\theta}}_n)$ is a consistent estimator of the Fisher information matrix $\boldsymbol{J}(\boldsymbol{\theta}_X)$ (if $\hat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}$, which we assume). Under the previously stated assumptions of regularity, it can be shown that under the null hypothesis $W_n \overset{\text{a.s.}}{\sim} \chi^2_p$, where p is the number of estimated parameters – in our case $p = 2$. We reject the null hypothesis H_0 in favor of the alternative H_1 for large values of W_n, i.e., if the test statistic W_n exceeds $(1-\alpha)$-quantile of of χ^2_p. We can translate this into the critical region $C(\alpha) = (q_{1-\alpha}, \infty)$, where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of of χ^2_p distribution.

Duality of the confidence regions and the hypothesis testing allows us to construct a confidence region based on the Wald statistic (2.15). A theorem describes this duality in the lecture notes by Omelka [2021], but only a version for the one-dimensional parameter can be found here. However, Roussas [1997] provides generalization to vector parameter. With this knowledge, we can finally construct the confidence region for the parameter $\boldsymbol{\theta}$ as

$$\mathbf{CI}(1-\alpha) = \left\{ \tilde{\boldsymbol{\theta}} \in \Theta : n \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}} \right)^\top \boldsymbol{J}(\hat{\boldsymbol{\theta}}_n) \left(\widehat{\boldsymbol{\theta}}_n - \widetilde{\boldsymbol{\theta}} \right) \leq q_{1-\alpha} \right\}. \tag{2.16}$$

2.5 Bayes estimator

In this section, we will shortly cover an advanced method for parameter estimation – Bayes model estimator – proposed by Mittal [1984]. Bayesian methods are not covered in the curriculum of the bachelor studies, and developing them formally would be beyond the scope of this thesis. Therefore, we will only outline the results derived in Mittal [1984].

To swiftly introduce the Bayes estimator, let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample from continuous distribution with probability density function $f(x; \theta)$. In Bayesian approach the likelihood of the truncated sample is multiplied by a “prior” density $p(\theta)$. This “prior” represents our beliefs about the distribution of unknown parameter θ before seeing the actual data. Hence

$$\mathcal{L}_n(\theta; \boldsymbol{X}) = \left[\prod_{i=1}^{n} f(X_i; \theta) \right] p(\theta).$$

The likelihood is then maximized as in the classical MLE.

Focusing on the $\mathsf{TN}(\mu, \sigma^2 | a, b)$ model again, Mittal [1984] claims inverted χ^2_m distribution with m degrees of freedom to be the appropriate prior distribution for σ^2. Similarly, the “non-informative” prior for μ is proposed, where the symbol “$\propto$” means equality except for the integration constant. Hence

$$\begin{aligned} p(\mu) &\propto c; \\ p(\sigma^2) &= \frac{2^{-\frac{m}{2}}}{\Gamma(\frac{m}{2})} \sigma^{-\frac{1}{2}(m-2)} e^{\frac{-1}{2\sigma}}. \end{aligned}$$

The modified likelihood proposed by Mittal [1984] is then

$$\mathcal{L}_n(\mu, \sigma^2; \boldsymbol{X}) = \frac{\psi(m)\sigma^{-(m-2)} \exp\{-\frac{\sum_{i=1}^{n}(X_i-\mu)^2+1}{2\sigma^2}\}}{(\sigma\sqrt{2\pi})^n \left(\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right)^n},$$

where ψ is only a function of m and does not play role in the maximization. Finally, the modified likelihood equations are

$$\overline{X_n} = \mu_T(\hat{\mu}_n, \hat{\sigma}_n^2);$$
$$s_n^2 + \frac{1}{n} = (m-2)\frac{\hat{\sigma}_n^2}{n} + \sigma_T^2(\hat{\mu}_n, \hat{\sigma}_n^2).$$

The solution has to be found numerically as in the Section 2.4. To minimize bias, Mittal [1984] proposes setting $m = 4$. Unlike the classical maximum likelihood estimator, the Bayes estimator exists with probability one [Mittal, 1984], making it an interesting topic for further studies.

We provided a brief summary, indicating other ways of coping with the possible nonexistence of the classical MLE. Further theory would have to be developed to cover this topic formally.

3. Truncated gamma distribution

3.1 Introductory definitions

In this chapter, we look into another essential distribution – gamma distribution. Even if we were not deeply interested in this particular distribution, we should still be motivated to study it, at least for its applications. Among very well-known special cases of the gamma distribution, famous exponential or χ^2 distributions can be found. We will investigate such a distribution under truncation.

Let us consider random variable with gamma distribution $X \sim \Gamma(\alpha, \beta)$, having a probability density function

$$f_\Gamma(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^\alpha \Gamma(\alpha)} \mathbb{I}_{(0,\infty)}(x), \quad \alpha, \beta > 0. \tag{3.1}$$

Cumulative distribution is defined the usual way and denoted as $F_\Gamma(x; \alpha, \beta)$ or just $F_\Gamma(x)$, when parameters a, b are obvious from the context.

Let us now define a truncated gamma distribution.

Definition 4. *Let $\alpha, \beta, a, b \in \mathbb{R}^+$ such that $0 < a < b < \infty$. Let X be a random variable with a gamma distribution $\mathsf{\Gamma}(\alpha, \beta)$. Then we say that a random variable $Y = X|a < X < b$ with a probability density function*

$$f_T(y; \alpha, \beta|a, b) = \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha)(F_\Gamma(b, \alpha, \beta) - F_\Gamma(a, \alpha, \beta))} \mathbb{I}_{(a,b)}(y). \tag{3.2}$$

has a doubly truncated gamma distribution $\mathsf{T\Gamma}(\alpha, \beta|a, b)$.

Definition 5. *et $\alpha, \beta, b \in \mathbb{R}^+$ such that $0 < b < \infty$. Let X be a random variable with a gamma distribution $\mathsf{\Gamma}(\alpha, \beta)$. Then we say that a random variable $Y = X|X < b$ with a probability density function*

$$f_T(y; \alpha, \beta|0, b) = \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha \Gamma(\alpha) F_\Gamma(b, \alpha, \beta)} \mathbb{I}_{(0,b)}(y). \tag{3.3}$$

has a singly (upper) truncated gamma distribution $\mathsf{T\Gamma}(\alpha, \beta|0, b)$.

Remark. We will also use rearranged version of the (3.2) and (3.3):

$$f_T(y; \alpha, \beta|a, b) = \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\int_a^b t^{\alpha-1} e^{-\frac{t}{\beta}}\, dt} \mathbb{I}_{(a,b)}(y). \tag{3.4}$$

3.2 Calculation of moments

In this section, we will find the moments of the gamma distribution. We could again calculate the moments, using the moment generating function as in the Section 2.2, but an article by Coffey and Muller [2000] suggests an easier way to get the results. Let $X \sim \mathsf{T\Gamma}(\alpha, \beta|a, b)$. Coffey and Muller [2000] provide the following lemma. We will also postulate a simple proof.

Lemma 1. *For any real number $m > -\alpha$ it holds*

$$x^m f_\Gamma(x;\alpha,\beta) = \frac{\beta^m \Gamma(\alpha+m)}{\Gamma(\alpha)} f_\Gamma(x;\alpha+m,\beta).$$

Proof. The proof is a direct corollary of the Definition 3.1.

$$\begin{aligned}\frac{\beta^m \Gamma(\alpha+m)}{\Gamma(\alpha)} f_\Gamma(x;\alpha+m,\beta) &= \frac{\beta^m \Gamma(\alpha+m)}{\Gamma(\alpha)} \frac{x^{\alpha+m-1}e^{-x/\beta}}{\beta^{\alpha+m}\Gamma(\alpha+m)} \mathbb{I}_{(0,\infty)}(x) \\ &= x^m \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \mathbb{I}_{(0,\infty)}(x) = x^m f_\Gamma(x;\alpha,\beta).\end{aligned}$$

□

With this lemma, we calculate the expected value

$$\begin{aligned}\mathsf{E}\,X &= \frac{\int_a^b x f_\Gamma(x;\alpha,\beta)\,dx}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)} = \frac{\int_a^b \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} f_\Gamma(x;\alpha+1,\beta)\,dx}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)} \\ &= \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{F_\Gamma(b,\alpha+1,\beta) - F_\Gamma(a,\alpha+1,\beta)}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)} \\ &= \beta\alpha \frac{F_\Gamma(b,\alpha+1,\beta) - F_\Gamma(a,\alpha+1,\beta)}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)}. \end{aligned} \tag{3.5}$$

Similarly we get the second moment

$$\begin{aligned}\mathsf{E}\,X^2 &= \frac{\beta^2\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{F_\Gamma(b,\alpha+2,\beta) - F_\Gamma(a,\alpha+2,\beta)}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)} \\ &= \beta^2\alpha(\alpha+1) \frac{F_\Gamma(b,\alpha+2,\beta) - F_\Gamma(a,\alpha+2,\beta)}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)}. \end{aligned} \tag{3.6}$$

Let us now, for simplicity, denote

$$\Lambda(m;\alpha,\beta) = \frac{F_\Gamma(b,\alpha+m,\beta) - F_\Gamma(a,\alpha+m,\beta)}{F_\Gamma(b,\alpha,\beta) - F_\Gamma(a,\alpha,\beta)}.$$

Using this notation, we can rewrite

$$\begin{aligned}\mathsf{E}\,X &= \alpha\beta\Lambda(1;\alpha,\beta); \\ \mathsf{E}\,X^2 &= \alpha(\alpha+1)\beta^2\Lambda(2;\alpha,\beta).\end{aligned}$$

Finally we arrive at the variance

$$\mathsf{var}\,X = \alpha^2\beta^2[\Lambda(2;\alpha,\beta) - (\Lambda(1;\alpha,\beta))^2] + \alpha\beta^2\Lambda(2;\alpha,\beta).$$

We can compare those values to the expected value and the variance of the original gamma distribution. Let $Y \sim \Gamma(\alpha,\beta)$, then it is well known, that $\mathsf{E}\,Y = \alpha\beta$ and $\mathsf{var}\,Y = \alpha\beta^2$. We can clearly see similarities in those values. When we compare $\mathsf{E}\,X$ and $\mathsf{E}\,Y$, we can see that the truncated gamma distribution expectation is the expectation of the original gamma distribution but weighted by some truncation factor. Similarly, when we compare $\mathsf{var}\,X$ and $\mathsf{var}\,Y$, we see that the original variance plays a role in the truncated one, again weighted by some truncation factor.

3.3 Moment method estimation

In this section, similarly to the Section 2.3, we will construct a method of moments estimator for parameters α, β. Suppose we have a random sample $X_1, \ldots, X_n$ of size n from doubly truncated gamma distribution $\mathsf{T\Gamma}(\alpha, \beta|a, b)$, where $\alpha, \beta \in \mathbb{R}^+$ are unknown parameters and truncation parameters $0 < a < b < \infty$ are assumed to be known.

Let us denote the real parameter $\boldsymbol{\theta}_X = (\alpha, \beta)$ and the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^+ \times \mathbb{R}^+$. We will be working with the model

$$\mathcal{F} = \{\text{distributions with density } f_\Gamma(x; \boldsymbol{\theta}_X|a, b), \boldsymbol{\theta}_X \in \boldsymbol{\Theta}\}.$$

Let us assume, that the expected value and variance are a function of the parameters, i.e., $(\mathsf{E}[X_i], \mathsf{var}\, X_i) = \boldsymbol{\tau}(\theta_X)$, where $\boldsymbol{\tau} = (\tau_1, \tau_2)$ such, that $\tau_i : \Theta \to \mathbb{R}$. This is a reasonable assumption, since in Section 3.2 we found out, that even in the truncated gamma distribution, the expected value and variance are indeed functions of unknown parameters α, β. The notation is the same as in (2.6).

Estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ of parameter $\boldsymbol{\theta}_X$ is then the solution of the system of equations

$$\begin{aligned} \overline{X_n} &= \tau_1(\hat{\boldsymbol{\theta}}_n); \\ s_n^2 &= \tau_2(\hat{\boldsymbol{\theta}}_n). \end{aligned}$$

In our case we get the system

$$\begin{aligned} \overline{X_n} &= \hat{\alpha}_n \hat{\beta}_n \Lambda(1; \hat{\alpha}_n, \hat{\beta}_n); \\ s_n^2 &= \hat{\alpha}_n^2 \hat{\beta}_n^2 [\Lambda(2; \hat{\alpha}_n, \hat{\beta}_n) - (\Lambda(1; \hat{\alpha}_n, \hat{\beta}_n))^2] + \hat{\alpha}_n \hat{\beta}_n^2 \Lambda(2; \hat{\alpha}_n, \hat{\beta}_n). \end{aligned} \tag{3.7}$$

As in the Section 2.3 functions τ_1 and τ_2 are not injective, thus cannot be inverted and in consequence, the system does not have an analytical solution and has to be solved numerically, e.g., by the Newton method. Also, not much can be said about consistency. We have to assume, that this estimator may not be consistent.

3.4 Maximum likelihood estimation

3.4.1 Score function

In this section we will construct a maximum likelihood estimator for the parameters α and β. Suppose we have a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ of size n from a doubly truncated gamma distribution $\mathsf{T\Gamma}(\alpha, \beta|a, b)$, where $\alpha, \beta \in \mathbb{R}^+$ are unknown parameters and truncation parameters $0 < a < b < \infty$ are assumed to be known.

The model $\mathcal{F}$ is the same as in the Section 3.3. We define the likelihood function $\mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{X})$ for the parameter $\boldsymbol{\theta} = (\alpha, \beta)$ and then find the estimator as $\hat{\boldsymbol{\theta}}_n = \operatorname{argmax}_{\theta \in \Theta} \mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{X})$, where $\Theta = \mathbb{R}^+ \times \mathbb{R}^+$ is the parameter space. Hence

$$\begin{aligned} \mathcal{L}_n(\boldsymbol{\theta}; \boldsymbol{X}) = \mathcal{L}_n(\alpha, \beta; \boldsymbol{X}) &= \prod_{i=1}^n f_\Gamma(x; \boldsymbol{\theta}_X|a, b) = \prod_{i=1}^n \frac{X_i^{\alpha-1} e^{-X_i/\beta}}{\int_a^b t^{\alpha-1} e^{-t/\beta}\, dt} \\ &= \left(\int_a^b t^{\alpha-1} e^{-t/\beta}\, dt \right)^{-n} e^{-\frac{1}{\beta} \sum_{i=1}^n X_i} \prod_{i=1}^n X_i^{\alpha-1}. \end{aligned} \tag{3.8}$$

Note that, again, we do not have to write the indicator function, since we assume random sample coming from a truncated distribution.

Now we proceed to the log-likelihood

$$\begin{aligned}\ell_n(\alpha,\beta) &= \log \mathcal{L}_n(\alpha,\beta) \\ &= -n\log\int_a^b t^{\alpha-1}e^{-t/\beta}\,dt + (\alpha-1)\sum_{i=1}^n \log X_i - \frac{1}{\beta}\sum_{i=1}^n X_i.\end{aligned} \tag{3.9}$$

Let us calculate the derivatives

$$\begin{aligned}\frac{\partial \ell_n}{\partial \alpha} &= -n\frac{\int_a^b t^{\alpha-1}e^{-t/\beta}\log t\,dt}{\int_a^b t^{\alpha-1}e^{-t/\beta}\,dt} + \sum_{i=1}^n \log X_i; \\ \frac{\partial \ell_n}{\partial \beta} &= -n\frac{\int_a^b t^{\alpha-1}e^{-t/\beta}t\beta^{-2}\,dt}{\int_a^b t^{\alpha-1}e^{-t/\beta}\,dt} + \beta^{-2}\sum_{i=1}^n X_i.\end{aligned}$$

By further adjusting those expressions and setting them to zero, we arrive at a system of two non-linear equations

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \log X_i &= \frac{\int_a^b t^{\hat\alpha_n-1}e^{-t/\hat\beta_n}\log t\,dt}{\int_a^b t^{\hat\alpha_n-1}e^{-t/\hat\beta_n}\,dt}; \\ \overline{X_n} &= \frac{\int_a^b t^{\hat\alpha_n-1}e^{-t/\hat\beta_n}t\,dt}{\int_a^b t^{\hat\alpha_n-1}e^{-t/\hat\beta_n}\,dt}.\end{aligned} \tag{3.10}$$

Solution of this system is the estimator $\hat{\boldsymbol{\theta}}_n = (\hat\alpha_n, \hat\beta_n)$. Equations (3.10) can be further adjusted. Notice, that by expanding the fraction on the right side by $\beta^\alpha\Gamma(\alpha)$, we get the truncated gamma densities as in (3.2), since the denominator can be written as a difference of two cumulative distribution functions. Hence

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \log X_i &= \int_a^b \log t\, f_T(t;\hat\alpha_n,\hat\beta_n|a,b)\,dt = \widehat{\mathsf{E}}\log X; \\ \overline{X_n} &= \int_a^b t f_T(t;\hat\alpha_n,\hat\beta_n|a,b)\,dt = \widehat{\mathsf{E}}X,\end{aligned}$$

where $\widehat{\mathsf{E}}X$ and $\widehat{\mathsf{E}}\log X$ are expectations with regard to the density with estimates $\hat a_n, \hat b_n$. Interestingly, in the maximum likelihood estimation, we are trying to match the sample mean with the theoretical, as it is in the case of the moment method estimator. But unlike the moment method estimator, here we are not matching higher moments, but rather sample and theoretical "log-moment".

Again, there is no analytical solution and the system has to be solved numerically.

3.4.2 Properties of the estimators

As with the normal distribution, we will follow the dissertation by Mittal [1984]. Mittal [1984] provides a discussion for a special case – singly upper truncated gamma distribution (3.3). Since the truncation point b is known, Mittal [1984] employs a transformation to the interval $(0,1)$, by dividing the random sample by the upper truncation boundary b. This allows him to simplify the calculation, only considering the case of $\mathsf{T\Gamma}(\alpha, \beta|0, 1)$. Estimating equations are then

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}\log X_i &= \int_0^1 \log t f_T(t; \hat{\alpha}_n, \hat{\beta}_n|0, 1)\, dt; \\ \overline{X_n} &= \int_0^1 t f_T(t; \hat{\alpha}_n, \hat{\beta}_n|0, 1)\, dt. \end{aligned} \tag{3.11}$$

Mittal [1984] provides a discussion similar to the discussion in the Section 2.4.2. However, he concludes that this discussion is not sufficient to prove, whether solution to (3.11) exists or not.

Hegde and Dahiya [1989] provide further discussion on the case of $\mathsf{T\Gamma}(\alpha, \beta|0, 1)$ and proves, that the MLE derived in the Section 3.4.1 does not have to always exist. Hegde and Dahiya [1989] even provide sufficient conditions for the existence. However, this requires further transformation. In our case, we have to assume that estimators derived in (3.10) may not be consistent.

Hegde and Dahiya [1989] conclude that different estimators, such as the modified maximum likelihood estimator (Bayes estimator) or the mixed estimator, would be more suitable. Bayes estimator would be analogical to the one in the Section 2.5. However, as Bayesian methods are not a subject of bachelor studies, we will not discuss them further.

4. Estimation of boundaries

So far, we have assumed the truncation points to be known. However, there are cases where their value is unknown and has to be estimated as well. We present a warning, what not to do, and a possible solution.

One could come up with an idea of following the same procedure as in the Section 2.4, i.e., trying to maximize the log-likelihood with regard to the truncation parameters a, b. However, such a procedure would violate the regularity conditions stated in the Section 2.4.2, needed for the maximum likelihood estimation to be valid. One of the regularity conditions states that the support of the probability density function must not depend on an unknown parameter. Thus, estimating the truncation parameter this way would be a severe violation of the regularity assumptions.

However, we could try to estimate those parameters in a different way. Let $X_1, \dots, X_n$ be a random sample from distribution with the probability density function on the support (a, b). Then, we could construct consistent estimators based on order statistics as $\hat{a}_n = \min_{i \in \{1,\dots,n\}} X_i$ and $\hat{b}_n = \max_{i \in \{1,\dots,n\}} X_i$.

To recall, we state the definitions of the consistency of an estimator and the convergence in probability.

- The estimator $\hat{\theta}_n$ of parameter θ_X is consistent, if $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathsf{P}} \theta_X$.
- $X_n \xrightarrow[n\to\infty]{\mathsf{P}} X \iff \forall \epsilon > 0 : \lim_{n\to\infty} \mathsf{P}(\omega : |X_n(\omega) - X(\omega)| > \epsilon) = 0.$

We will prove consistency of $\hat{a}_n$ from the definition of convergence in probability. Let $\epsilon > 0$, then

$$\begin{aligned}
\mathsf{P}(|\hat{a}_n - a| > \epsilon) &= \mathsf{P}(|\min_{i \in \{1,\dots,n\}} X_i - a| > \epsilon) = 1 - \mathsf{P}(-\epsilon + a < \min_{i \in \{1,\dots,n\}} X_i < \epsilon + a) \\
&= 1 + \mathsf{P}(\min_{i \in \{1,\dots,n\}} X_i > \epsilon + a) - \mathsf{P}(\min_{i \in \{1,\dots,n\}} > a - \epsilon) \\
&= 1 + \underbrace{\mathsf{P}(X_1 > \epsilon + a)}_{<1}{}^n - \underbrace{\mathsf{P}(X_1 > a - \epsilon)}_{=1}{}^n \xrightarrow{n\to\infty} 0,
\end{aligned}$$

since when the $\min_{i \in \{1,\dots,n\}} X_i > \epsilon + a$, then all $X_i > \epsilon + a$. Since they are independent and identically distributed, we can multiply them and get the n-th power. Also, the support of the random variable X_1 is the interval (a, b), therefore probability of $X_1 > a - \epsilon$ is one, while probability of $X_1 > a + \epsilon$ has to be lower then one since we are not integrating over the whole support of the random variable. This holds true for every $\epsilon > 0$.

We have proved the consistency of the $\hat{a}_n$. Proof for the $\hat{b}_n$ is analogous. However, both of them would be obviously biased – for $\hat{a}_n$ we always overestimate, while for $\hat{b}_n$ we always underestimate.

Robson and Whitlock [1964] offer further methods for constructing unbiased estimators. Robson and Whitlock [1964] propose a construction of unbiased estimator on the basis of order statistics. Let us rearrange the supposed random sample $X_1, \dots, X_n$ in such a way, that $X_{(1)} < \dots < X_{(n)}$. Estimator is then $\tilde{b}_n = X_{(n)} + (X_{(n)} - X_{(n-1)})$. Prove of consistency and unbiasedness can be found in Robson and Whitlock [1964].

5. Numerical study

5.1 Motivation and data exploration

In this section, we apply the previously built theory to a real dataset. Working with truncated distributions, the goal was to find a dataset with truncated observations. Since armies often employ physical limitations as an enrollment condition, the height of soldiers was a natural idea for a suitable dataset.

We chose the male Anthropometric Database provided by the Army Public Health Center, a U.S. Army Medical Department organization. Center [2020] database contains 93 anthropometric measurements of 4082 U.S. Army soldiers. Since normality of the population height is generally assumed, we chose soldiers' height to be further analyzed. Descriptive statistics of soldiers' height are to be found in the Table 5.1.

	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**	**Var.**
Height [cm]	149.1	171.0	175.5	175.6	180.2	199.3	47.0

Table 5.1: Descriptive statistics for the height of the males in the U.S. Army.

It seems that nowadays, there is no official cut-off height for entering the U.S. Army. However, according to unofficial sources, such as career-advising websites, the minimal and maximal height is often considered to be 58 inches (around 147 cm) and 80 inches (around 203 cm), respectively. That corresponds with characteristics in the Table 5.1

Since it can be assumed that most of the adult males in the U.S are in this range anyway, the height limits do not seem to be too restricting. By looking at the histogram in the Figure 5.1, we see that truncation is only mild indeed.

5.2 Additional truncation

For this thesis to be more interesting, we introduce further artificial truncation. Suppose we are in a scenario in which we recruit soldiers for submarines. Submarines are small. Thus, soldiers have to be shorter. Suppose the lower truncation point stays the same, and the upper one will be lowered. Let us set the truncation points as $a = 147$ cm, $b = 180$ cm for the lower and upper truncation point respectively. With this additional truncation, we arrive at a cropped dataset of 3016 soldiers. Descriptive statistics for this truncated dataset can be seen in the Table 5.2. When we compare the two tables, the mean and variance are clearly differing. The full effect of the truncation can be seen in the Figure 5.2.

	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**	**Var.**
Height [cm]	149.1	169.4	173.3	172.6	176.5	179.9	23.5

Table 5.2: Descriptive statistics for the height of the males in the U.S. Army with the additional truncation.

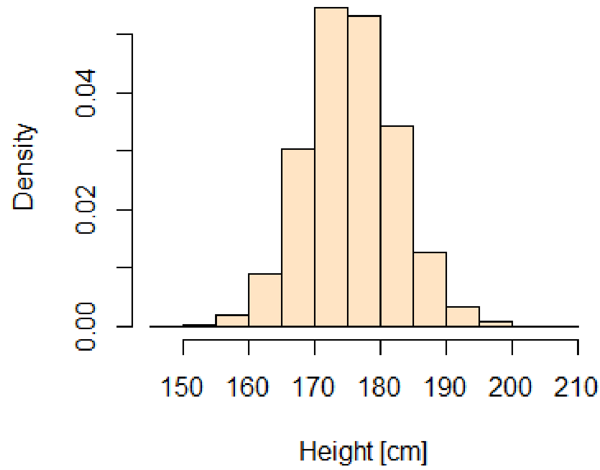

Figure 5.1: Histogram of the U.S. Army soldiers height.

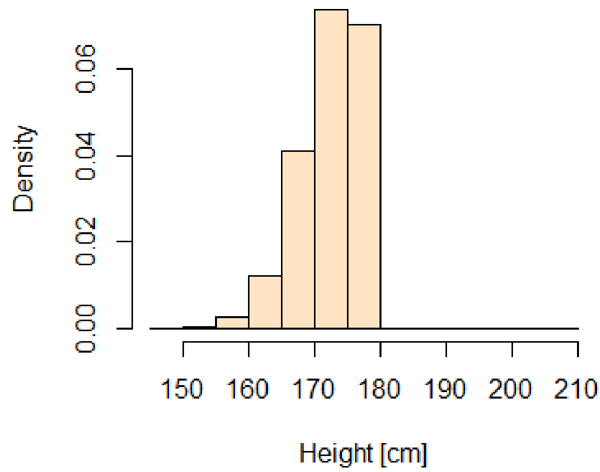

Figure 5.2: Histogram of the additionally truncated U.S. Army soldiers height.

5.3 Estimation of parameters

In the Section 5.1, we mentioned the normality of height assumption to be widely accepted. We assume, that the original generating distribution of our data is the normal distribution. Therefore, our truncated sample should follow the truncated normal distribution. The theory for such a distribution is covered in the Section 2.

We now formalize our task. Let $\boldsymbol{X} = (X_1, \ldots, X_{3016})$ be a random sample from $\mathsf{TN}(\mu, \sigma^2|a, b)$, where $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+$ are unknown parameters and the truncation parameters $a = 147$, $b = 180$ are assumed to be known. We will estimate the parameters μ and σ^2 with the method of maximum likelihood and construct a confidence region for those parameters.

For the numerical calculations we use statistical computing environment R. We employ function `optim` with `BFGS` method from the package `stats` to maximize the log-likelihood function (2.9). The R core package named `stats` is well documented; thus, it will not be discussed in more detail. `BFGS` algorithm implements the quasi-Newton method. Describing the algorithm and numerical method would be beyond the scope of this thesis. More on them can be found in Kelley [1999].

We provide a snippet of the executed code used to estimate μ and σ^2. The output gives us not only the values of the estimates $\hat{\mu}_n = 175,1$ and $\hat{\sigma}_n = 42.0$ but also the value of the log-likelihood in its maximum. Further, it provides technical details such as 0 output code for the convergence – meaning successful convergence. For us, only. Notice that as initialization, we use the mean and variance of the truncated dataset.

```
> (manual_estimator = optim(c(mean(truncated_height),var(truncated_height)),
+                           log_likelihood,y=truncated_height, method="BFGS",
+                           control = list(fnscale=-1)))
$par
[1] 175.07895  41.99259

$value
[1] -8710.083

$counts
function gradient
      20        9

$convergence
[1] 0

$message
NULL
```

With the estimated parameters $\hat{\mu}_n, \hat{\sigma}_n$ we are able to plot the probability density function of $\mathsf{TN}(175.1, 42|147, 180)$ over the truncated histogram – see the Figure 5.3.

5.4 Confidence region

In this section we utilize the confidence region formula (2.16) for the parameters μ and σ^2 derived in the Section 2.4.4.

$$\mathbf{CI}(1-\alpha) = \left\{ \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \in \Theta : n \begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\sigma}_n^2 - \sigma^2 \end{pmatrix}^\top \boldsymbol{J}(\hat{\mu}_n, \hat{\sigma}_n^2) \begin{pmatrix} \hat{\mu}_n - \mu \\ \hat{\sigma}_n^2 - \sigma^2 \end{pmatrix} \leq \chi_p^2(1-\alpha) \right\}.$$

Notice that after plugging in the actual values, the set is defined by a bilinear form describing an ellipse. This ellipse can be seen in the Figure 5.4.

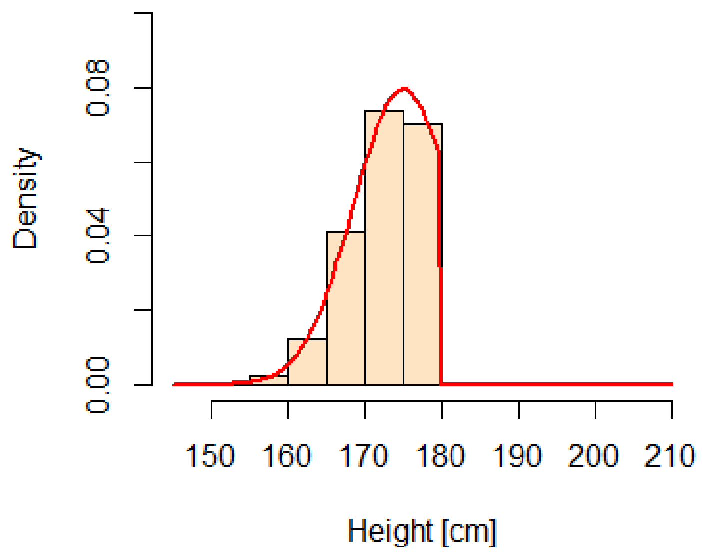

Figure 5.3: Histogram of the additionally truncated U.S. Army soldiers height with associated probability density function.

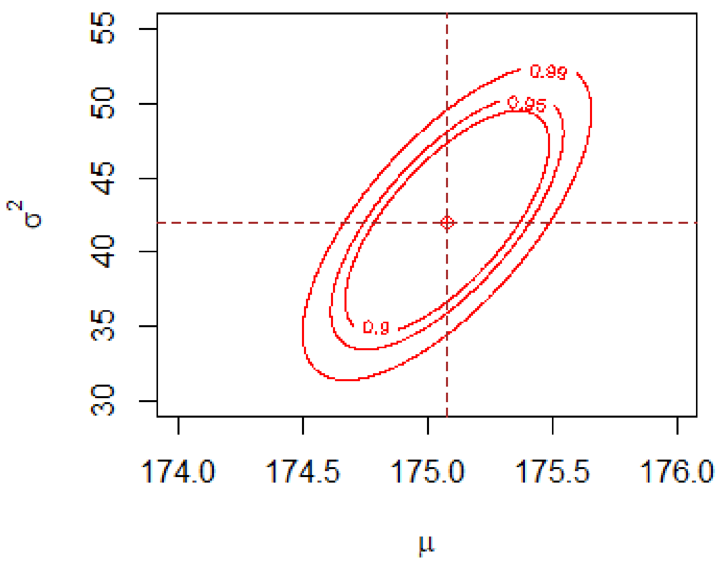

Figure 5.4: Confidence regions for the parameters μ and σ^2 with associated confidence levels 0.9, 0.95. and 0.9.

Conclusion

The goal of this thesis was to study stochastic models for truncated data and apply them to either real data or simulation study.

The first Chapter provided a comprehensible motivation example and definitions of both singly and doubly truncated distributions.

The second Chapter covered the model of truncated normal distribution. We covered the calculation of moments, derivation of two estimators for parameters μ and σ^2 and their asymptotic distribution. We would mention the detailed calculation of the moment generating function, score function, and Fisher information matrix as our contribution. Calculations were checked both by comparing to Mittal [1984] and by symbolic derivations in Mathematica software. Notably, we provided an alternative computation of the Fisher information matrix – Mittal [1984] obtained his results as a variance of the score function, whereas we calculated expectation of its derivatives. Fisher information matrix was then used to determine the asymptotic distribution of the estimators and construct confidence regions. Calculations were accompanied by a comprehensive discussion of existence, summarizing Mittals work. Bayes estimator was also presented as a possible variant for further research.

The third Chapter followed the structure of the second one, but in less detail, since most of the derivations and assertions would be analogous, probably with fewer closed-form solutions. Here we combined works of Mittal [1984] and Coffey and Muller [2000] to summarize both methods of moments and maximum likelihood estimators.

In the fourth Chapter, we presented a different perspective and examined the estimation of the truncation parameters, which were previously assumed to be known. A simple proof of consistency was constructed for a basic, biased estimator, complemented with a presentation of unbiased estimators derived by Robson and Whitlock [1964].

Finally, in the fifth Chapter, we covered a numerical study of a real-world dataset – an anthropometric database of U.S. Army soldiers. We explored the data regarding soldiers' height and estimated parameters of mean and variance in the truncated normal distribution model. Confidence regions were then constructed for the mentioned parameters. For numerical computations, statistical computing environment R together with RStudio integrated development environment was used.

Bibliography

J. Anděl. *Základy matematické statistiky.* Druhé opravené vydání. Matfyzpress, Praha, 2007. ISBN 80-7378-001-1.

Army Public Health Center. ANSURII Database Male, 2020. URL `https://phc.amedd.army.mil/topics/workplacehealth/ergo/Pages/Anthropometric-Database.aspx`. Last updated 7th Feb 2020, Accessed 14th Apr 2021.

C. S. Coffey and K. E. Muller. Properties of doubly-truncated gamma variables. *Communications in Statistics – Theory and Methods*, 29(4):851–857, 2000.

L. M. Hegde and R. C. Dahiya. Estimation of the parameters of a truncated gamma distribution. *Communications in Statistics – Theory and Methods*, 18 (2):561–577, 1989.

C. T. Kelley. *Iterative methods for optimization.* SIAM, Philadelphia, PA, 1999. ISBN 978-0-89871-433-3.

M. M. Mittal. *Estimating the Parameters of Truncated Distributions.* PhD thesis, Old Dominion University, VA, 1984.

O. E. Nielsen. *Information and exponential families: In statistical theory.* John Wiley & Sons, Chichester, UK, 2014. ISBN 978-1-118-85750-2.

M. Omelka. Matematická statistika 1 – Poznámky k přednášce, 2021. URL `https://www2.karlin.mff.cuni.cz/~omelka/Soubory/nmsa331/ms1.pdf`. Last updated 25th Jan 2021, Accessed 15th May 2021.

D. S. Robson and J. H. Whitlock. Estimation of a truncation point. *Biometrika*, 51(1-2):33–39, 1964.

G. Roussas. *A course in mathematical statistics.* Academic Press, San Diego, CA, 1997. ISBN 0-12-599315-3.

List of Figures

List of Tables

A. Attachments

A.1 Electronic Attachments

An electronic attachment consisting of two files is provided. R source code for real-data study in the Chapter 5 is provided in the file `soldiers_height.R`. Dataset for this study is attached under the name `ANSURIIMALEPublic.csv`.