# Continuous Statistical Models: With or Without Truncation Parameters?

Vancak Valentin

Thesis Submitted in Partial Fulfilment of the Requirements
for the Master's Degree

University of Haifa
Faculty of Social Sciences
Department of Statistics

October, 2013

# Continuous Statistical Models: With or Without Truncation Parameters?

By: **Valentin Vancak**

Supervised by: **Dr. Yair Goldberg**

Thesis Submitted in Partial Fulfilment of the Requirements
for the Master's Degree

University of Haifa
Faculty of Social Sciences
Department of Statistics

October, 2013

Approved by:____________________ Date:____________
(Supervisor)

Approved by:____________________ Date:____________
(Chairperson of Master's studies Committee)

# Acknowledgments

I would like to thank my advisor, Dr. Yair Goldberg, for his dedication, endless knowledge and rare mathematical erudition for guiding and advising me on every step of this work. This work owns a lot to valuable and helpful comments from Prof. Shaul Bar-Lev and Prof. Benzion Boukai. I would like to express my appreciation to the administrative staff of the Department of Statistics as well, especially to Hana Avraham for providing excess to departments facilities and valuable administrative support.

# Contents

Continuous Statistical Models: With or Without Truncation Parameters?

Valentin Vancak

## ABSTRACT

Life time data are usually assumed to stem from a continuous distribution supported on $(0, b)$ , $b \leq \infty$. The continuity assumption implies that the support of the distribution does not have atom points, particularly not at 0. Accordingly, it seems reasonable that with an accurate measurement tool all data observations will be positive. This may lead to conjecturing that the true support is truncated from left and is in fact has the form $(\gamma, b), \gamma \geq 0$ , with $\gamma$ unknown. Consequently, under such a situation, we will face the two typical errors (of False Models). To describe these we let Model I denotes the model which is linked with the assumption that the true support is $(0, b)$, while Model II denotes the truncated model for which the true support is $(\gamma, b), \gamma \geq 0$. We shall therefore say that a False Model I error has occurred if Model I has been incorrectly used for inference while the correct model is Model II. False Model II error is defined similarly. The question that naturally arises then is which of the two types of errors is more severe?

Two main settings are discussed in this work. The first setting concerns general statistical continuous models. For this setting we consider the following two scenarios. First we assume that there is a left truncation when in fact there is none, i.e., the actual model support is $S = (0, \theta)$ where $\theta$ is a parameter of the model. We show that this type of error do not cause any severe deficit in the estimation process in terms of asymptotic accuracy and efficiency. This result stems from the fast rate of convergence of Bar-Lev and Boukai's estimator (Bar-lev and Boukai, 1985) to the function of interest. However, we demonstrated that in the converse case, when we assume that there is no left truncation, but in fact the model support is $S = (\gamma, \theta)$, Tate's estimator (Tate, 1959) does not converge to the function of interest. In fact, in this case Tate's estimator has a constant bias term, and therefore, for large left truncation, Tate's estimator will result in inaccurate and inefficient estimation.

The second setting concerns the exponential families of distributions and the MLE for the natural parameter $\theta$. We present two examples when assuming that there is left truncation when actually there is none does not results in any substantial inefficiency. This

result stems from the fast rate of convergence of $X_{(1)}$ to the left boundary 0. We show that the MLE for the natural parameter $\theta$ that was derived Model's II support converges to the function of interest; it is asymptotically unbiased and efficient (the MSE meets the Cramer-Rao Lower Bound). Moreover, we show that the converse situation is inefficient as is exampled in Setting I; assuming that there is no left truncation where actually there is, results in an inaccurate and inconsistent estimator that converges almost surely to the wrong function. We pose as an open question for further research the generalization of the results to the whole exponential families.

# List of Figures

# 1 Introduction

Life time data are usually assumed to stem from a continuous distribution supported on $(0, b), b \leq \infty$. The continuity assumption implies that the support of the distribution does not have atom points, particularly not at 0. Accordingly, it seems reasonable that with an accurate measurement tool all data observations will be positive. This may lead to conjecturing that the true support is truncated from left and in fact has the form $(\gamma, b), \gamma \geq 0$ , with $\gamma$ unknown. Consequently, we can discuss two models. The first model, which we denote as Model I is the one for which the true support is $(0, b)$. The second model, Model II, denotes the truncated model for which the true support is $(\gamma, b), \gamma \geq 0$. Note that Model I is included in Model II. We shall therefore say that a False Model I error has occurred if Model I has been incorrectly used for inference while the correct model is Model II. False Model II error is defined similarly. The question that naturally arises then is which of the two types of errors is more severe?

It seems reasonable that even if $(0, b)$ is the correct support then the use of $(\gamma, b)$ will not result in much loss of information. While if $(\gamma, b)$ for some $\gamma > 0$ is the correct support, there will be substantial loss of information. This claim can also be justified in term of sufficiency. To realize this, let $\mathbf{X} = (X_1, ..., X_n)$ be a size $n$ random sample taken from the population being studied. Assume that Model I depends on an unknown parameter $\theta$ (possibly a vector), and is associated with a minimal sufficient statistic $S_n = s(\mathbf{X})$. Model II, which obtained by a truncation of Model I, is therefore parameterized by $(\theta, \gamma)$ and is associated with the minimal sufficient statistic $(S_n, X_{(1)})$. Note that $(S_n, X_{(1)})$ while being minimal sufficient for Model II, is still sufficient for Model I; whereas $S_n$ while being minimal sufficient for Model I, is not even sufficient for Model II. Hence False Model I error is more critical since inference is based on a statistic which is not sufficient, while for False Model II error inference is based on a sufficient statistic, not minimal though.

Two main settings will be investigated in this work. In the first one, we assume that the density function known up to a right truncation parameters. In this setting, under Model I, there is only right side truncation parameter $\theta$. Under Model II, we assume also left truncation. For this setting, two candidate estimators (Tate, 1959; Bar-lev and Boukai, 1985) will be compared by their finite-sample bias and MSE, as well as their asymptotic efficiency. We also compare these estimator for the case that Model II (two side truncation) holds.

The second setting deals with two representative distribution from the exponential family of distributions with possible left truncation. Here, Model I holds when no left truncation is introduced, and Model II holds when left truncation exists. In this setting the investigation will focus on the effects of left truncation on the MLE of the natural parameter $\theta$. Both asymptotic and finite-sample behaviour of the estimators will be investigated.

As stated above, our main concern is the cross-model behaviour of the estimators. In other words, we are interested in finite-sample and the asymptotic properties (distribution function, expectation and MSE) of each of the estimators under the “wrong” model. More specifically, for the right truncation with possible left truncation setting, we are interested in the behaviour of Bar-lev and Boukai (1985) (hereafter abbreviated BB) estimator when there is no left truncation, and the behaviour of Tate estimator (Tate, 1959) when left truncation is introduced.

In fact, when Tate’s estimator is considered, but there is left truncation, we will show that the estimator will have an asymptotic constant bias term, i.e., the estimator does not converge to the function of interest. Therefore, Tate’s estimator is inconsistent w.r.t. Model II support. In this case, it is obvious that the asymptotic relative efficiency of the estimators will tend to infinity due to the constant bias term. As such, the main interest and concern will be the behaviour of BB’s estimator w.r.t. Model I support. Note that in BB’s estimator, left truncation parameter is estimated with the minimal order statistic $X_{(1)}$, which converges in probability to 0, therefore, the bias of the BB’s estimator, when such bias exists, will asymptotically vanish, and the estimator will converge in probability to the function of interest. Therefore, we should take into account (i) the rate of convergence and (ii) the relative asymptotic efficiency of the estimators. Hereafter, in the case of model misspecification, while considering BB’s estimator when in fact Model I holds, our main interest is the ”price” that we pay (in terms of accuracy and efficiency of the estimation) if we are estimating the left truncation parameter whilst the left support bound is 0.

In the exponential families distributions setting, the same scenarios will be considered for the maximum likelihood estimators. As in continuous statistical models setting, our main concerns will be with the MLE derived under Model II support, while in fact Model I holds. The converse case should not yield any surprises due to the expected constant bias term that would cause asymptotic inefficiency and convergence of the estimator to the wrong function. Therefore, in this setting, as in the previous one, the main effort will be directed to the investigation of the properties of MLE derived under Model II support, while Model I holds. Nevertheless, the converse situation in which the MLE derived under Model I support while Model II support holds, will also be considered.

The work is organized as follows. The analysis of continuous statistical right truncated models with possible left truncation is presented in Chapter 2. The main focus of this chapter is to estimate the tail probability and the distribution’s expectation w.r.t. the two different supports. In Chapter 3 we discuss the exponential families with possible left truncation. The chapter focuses on both finite-sample and asymptotic behaviour of maximum likelihood estimators for the natural parameter of the exponential and Erlang-2 distributions. Concluding remarks appear in Chapter 4. All proofs appear in the Appendix.

# 2 Setting I: Continuous Statistical Right Truncated Models with Possible Left Truncation

This chapter is organized as follows. In Section 2.1 we present the notations for the whole chapter. Afterwards, in Section 2.2, we present the general form of the estimators and its density function w.r.t. two possible supports. Next, in Section 2.3, we perform the analysis of the tail probability function for the uniform case. In Section 2.4, the same analysis is done for a Beta distribution example. Finally, in Section 2.5 we analyse the estimators for the expectation of uniform distribution. Conclusions can be found in Section 2.6. All computations of this Chapter appear in Appendix A.

## 2.1 Construction and Notations

The purpose of this section is to present the notations and preliminary calculations. Note that the density of any continuous distribution with possible left truncation can be decomposed into two basic functions: The normalizing constant $g_0(\gamma, \theta)$, and the invariant term $h(x)$.

Let $h(\cdot)$ be a positive integrable function over $[0,\ \infty)$. For any $0 \leq \gamma < \theta$ define:

$$g_k(\gamma, \theta) = \int_{\gamma}^{\theta} x^k\, h(x) dx, \quad k = 0, 1, 2, \ldots \tag{1}$$

Using (1), we construct the p.d.f. of the continues type random variable $X$ as:

$$f^{II}(x; \gamma, \theta) = \frac{h(x)}{g_0(\gamma, \theta)} I[\gamma < x < \theta]. \tag{2}$$

Note that with the notation is (1), the moments of $X$ are easily defined by

$$E(X^k) = \frac{g_k(\gamma, \theta)}{g_0(\gamma, \theta)}, \quad k = 0, 1, 2, \ldots$$

In particular the mean of $X$ is

$$E(X) = \frac{g_1(\gamma, \theta)}{g_0(\gamma, \theta)}.$$

The c.d.f. of $X$ is given, for any $\tau \in \mathbb{R}$, by

$$F_\eta(\tau) \equiv P_\eta(X \leq \tau) = \frac{g_0(\gamma, \tau)}{g_0(\gamma, \theta)} I[\gamma < \tau < \theta] + I[\theta \leq \tau].$$

The tail probability is given by

$$P_\eta(X > \tau) = 1 - F_\eta(\tau) = I[\tau \leq \gamma] + \frac{g_0(\tau, \theta)}{g_0(\gamma, \theta)} I[\gamma < \tau < \theta].$$

Here, $\gamma$ and $\theta$ are possible unknown parameters of $f(x;\gamma,\theta)$. Accordingly we consider two possible models:

- Model I: $\gamma \equiv \gamma_0 = 0$ known, while $\theta > \gamma$ is an unknown parameter, so that $\eta_0 \equiv (\gamma_0, \theta)$ designates the model's only unknown parameter, $\theta$.
- Model II: Both $\gamma$ and $\theta$ are unknown parameters, $0 < \gamma < \theta$, so that $\eta \equiv (\gamma\,, \theta)$ designates the model's two unknown parameters.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a sample of $n$ i.i.d. observations from $f(x;\gamma,\theta)$ in (2), and let $X_{(1)} \leq X_{(2)} \leq, \cdots \leq X_{(n)}$ denote the corresponding ordered statistics. It is a standard exercise to show that under Model I, the minimal sufficient statistics (MSS) $t(\mathbf{X})$ for $\eta_0 = \theta$ is $t(\mathbf{X}) = X_{(n)}$. Similarly, it can be shown that under Model II (with $\eta \equiv (\gamma\,, \theta)$) , the MSS is $t(\mathbf{X}) = (X_{(1)},\, X_{(n)})$.

Under Model I, the p.d.f of the MSS statistic $t(\mathbf{X}) = X_{(n)}$ is readily available

$$f^{I}_{X_{(n)}}(t) = \frac{n\, h(t)\, \big(g_0(0,\, t)\big)^{n-1}}{\big(g_0(0,\, \theta)\big)^{n}}\; I[0 \leq t \leq \theta],$$

whereas, under Model II, the p.d.f of $t(\mathbf{X}) = (X_{(1)},\, X_{(n)})$ can be shown to be

$$f^{II}_{X_{(1)},X_{(n)}}(y,\, t) = \frac{n\,(n-1)\, h(y)\, h(t)\, \big(g_0(y,\, t)\big)^{n-2}}{\big(g_0(\gamma,\, \theta)\big)^{n}}\; I[\gamma \leq y \leq t \leq \theta].$$

Finally under Model I, it can shown that the p.d.f of $t(\mathbf{X}) = (X_{(1)},\, X_{(n)})$ is

$$f^{I}_{X_{(1)},X_{(n)}}(y,\, t) = \frac{n\,(n-1)\, h(y)\, h(t)\, \big(g_0(y,\, t)\big)^{n-2}}{\big(g_0(0,\, \theta)\big)^{n}}\; I[0 \leq y \leq t \leq \theta].$$

## 2.2 Estimators

Let $\xi(\eta)$ be any estimable function of the model's unknown parameter $\eta$. For instance, $\xi(\eta) = E_\eta(X)$, or $\xi(\eta) = F_\eta(a)$, for some fixed $a \in \mathbb{R}$ (see Tate, 1959, and Bar-lev and Boukai, 1985, for specific expressions for both models).

Based on the sample data $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we are interested in constructing a UMVUE $q(t(\mathbf{X}))$ for $\xi(\eta)$. Clearly, this estimator satisfies

$$E_\eta(q(t(\mathbf{X}))) = \xi(\eta).$$

Tate (1959) considered this problem under Model I (i.e., $\eta_0 \equiv (\gamma_0, \theta)$ and $t(\mathbf{X}) = X_{(n)}$) and obtained that the general form of the UMVUE for $\xi(\eta_0) = \xi(\theta)$, is

$$q_I(X_{(n)}) = \xi(X_{(n)}) + \frac{\xi'(X_{(n)}) g_0(0, X_{(n)})}{n h(X_{(n)})},$$

whenever the derivative $\xi'(\theta) = \partial\xi(\theta)/\partial\theta$ exists and is continuous almost everywhere on the support $\Theta = \{(0, \theta) : a < \theta < b\}$.

Example for Tate's estimators include

**Example 1.** *Under Model I, if* $\xi(\eta_0) = E_{\eta_0}(X) = g_1(\gamma_0,\, \theta)/g_0(\gamma_0,\, \theta)$*, with known* $\gamma_0 = 0$*, it can be shown that*

$$q_I(X_{(n)}) = \frac{g_1(0, X_{(n)})}{g_0(0, X_{(n)})}\left(1 - \frac{1}{n}\right) + \frac{X_{(n)}}{n}.$$

*The computations for this example can be found in Appendix A.1.*

**Example 2.** *Under Model I, if* $\xi(\eta_0) = 1 - F_{\eta_0}(\tau) = g_0(\tau,\, \theta)/g_0(\gamma_0,\, \theta)$*, with known* $\gamma_0 = 0$*, and* $\tau \geq \gamma_0$*, one can show, using the Leibniz integral rule, that the general form of Tate's estimator is:*

$$q_I(X_{(n)}) = 1 - \left(1 - \frac{1}{n}\right)\frac{g_0(0, \tau)}{g_0(0, X_{(n)})}. \tag{3}$$

Similarly, Bar-lev and Boukai (1985) consider the same estimation problem, but under Model II (i.e., $\eta \equiv (\gamma,\, \theta)$ and $t(\mathbf{X}) = (X_{(1)},\, X_{(n)})$). They showed that the general form of the UMVUE for any estimable function $\xi(\eta) \equiv \xi(\gamma,\, \theta)$ is

$$\begin{aligned} q_{II}(X_{(1)}, X_{(n)}) =& \xi(X_{(1)}, X_{(n)}) - \frac{g_0(X_{(1)}, X_{(n)})\xi_1(X_{(1)}, X_{(n)})}{(n-1)h(X_{(n)})} \\ &+ \frac{g_0(X_{(1)}, X_{(n)})\xi_2(X_{(1)}, X_{(n)})}{(n-1)h(X_{(n)})} \\ &- \frac{g_0^2(X_{(1)}, X_{(n)})\xi_{12}(X_{(1)}, X_{(n)})}{n(n-1)h(X_{(1)})h(X_{(n)})}, \end{aligned} \tag{4}$$

whenever the partial derivations $\xi_1(\gamma, \theta) = \partial\xi(\gamma, \theta)/\partial\gamma$, $\xi_2(\gamma, \theta) = \partial\xi(\gamma, \theta)/\partial\theta$, and $\xi_{12}(\gamma, \theta) = \partial^2\xi(\gamma, \theta)/\partial\gamma\partial\theta$ exist and are continuous almost everywhere on $\Theta = \{(\gamma, \theta) : a < \gamma < \theta < b\}$.

Examples for BB's estimators include

**Example 3.** *Under Model II, if* $\xi(\eta) = E_\eta(X) = \xi(\gamma,\, \theta) = g_1(\gamma,\, \theta)/g_0(\gamma,\, \theta)$*, using the Leibniz integral rule, one can derive the general form of BB's estimator:*

$$q_{II}(X_{(1)},\, X_{(n)}) = \frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})}.$$

*The computations for this example can be found in Appendix A.2.*

**Example 4.** *Under Model II, if* $\xi(\eta) = 1 - F_\eta(\tau) = \xi(\gamma,\, \theta) = g_0(\tau,\, \theta)/g_0(\gamma,\, \theta)$*,* $\tau \geq 0$*, one obtains that*

$$q_{II}(X_{(1)},\, X_{(n)}) = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{g_0(X_{(1)}, \tau)}{g_0(X_{(1)}, X_{(n)})}.$$

## 2.3 Analysis of the Tail Probability Function $P_\eta(X > \tau)$ for Uniform Distribution

The analysis in this Section focuses on model misspecification, where the properties of interest are (i) the estimator's expectations and (ii) the estimator's MSE w.r.t. the incorrect support. In other words, what is the deficiency (if there is one) when we derive the estimators w.r.t. Model I support while actually Model II support holds and vice versa.

Let $X_1, ..., X_n$ be a sample size $n \geq 4$ from $\mathrm{U}(0, \theta)$, where $\theta$ is an unknown right truncation parameter. Before we introduce the appropriate estimators for this problem, we construct the relevant density functions.

Note that the density function for a uniform random variable $\mathrm{U}(0, \theta)$ is given by

$$f^I(x; \gamma_0, \theta) = \frac{h(x)}{g_0(0, \theta)}\mathbf{1}\{0 \leq x \leq \theta\} = \frac{1}{\theta}\mathbf{1}\{0 \leq x \leq \theta\}\,.$$

The function of interest w.r.t. Model I support, for any $0 \leq \tau \leq \theta$, is given by

$$P_{\eta_0}(X > \tau) = \frac{g_0(\tau, \theta)}{g_0(0, \theta)}\mathbf{1}\{0 < \tau < \theta\} = 1 - \frac{\tau}{\theta}\mathbf{1}\{0 < \tau < \theta\}\,.$$

The function of interest w.r.t. Model II support, for any $\gamma \leq \tau \leq \theta$, is

$$P_\eta(X > \tau) = \frac{g_0(\tau, \theta)}{g_0(\gamma, \theta)}\mathbf{1}\{\gamma < \tau < \theta\} = 1 - \frac{\tau - \gamma}{\theta - \gamma}\mathbf{1}\{\gamma < \tau < \theta\}\,.$$

Finally, the density function of the order statistics can be written as follows. With respect to Model I support, i.e., $\gamma = \gamma_0 = 0$ such that $S = (0, \theta)$, the joint density of the vector of order statistics $(X_{(1)}, X_{(n)})$ is given by

$$\begin{aligned} f^I_{X_{(1)},X_{(n)}}(y, t) &= \frac{n(n-1)h(y)h(t)(g_0(y, t))^{n-2}}{(g_0(y, t))^n}\mathbf{1}\{0 \leq y \leq t \leq \theta\} \\ &= \frac{n(n-1)(t-y)^{n-2}}{\theta^n}\mathbf{1}\{0 \leq y \leq t \leq \theta\}. \end{aligned} \tag{5}$$

In order to compute cross-model properties of the estimators, we have to present some functions of interest w.r.t. Model II support, i.e., when $\gamma > 0$ and $S = (\gamma, \theta)$. The density function of the random variable w.r.t. Model II support is modified and given by

$$f^{II}(x; \gamma, \theta) = \frac{h(x)}{g_0(\gamma, \theta)}\mathbf{1}\{\gamma \leq x \leq \theta\} = \frac{1}{\theta - \gamma}\mathbf{1}\{\gamma \leq x \leq \theta\}.$$

Due to the density function modification, the joint density of the following vector $(X_{(1)}, X_{(n)})$

is given by

$$\begin{aligned} f^{II}_{X_{(1)},X_{(n)}}(y,t) &= \frac{n(n-1)h(y)h(t)(g_0(y,t))^{n-2}}{(g_0(y,t))^n}\mathbf{1}\{\gamma \le y \le t \le \theta\} \\ &= \frac{n(n-1)(t-y)^{n-2}}{(\theta-\gamma)^n}\mathbf{1}\{\gamma \le y \le t \le \theta\}. \end{aligned} \tag{6}$$

The density of the maximal order statistic $X_{(n)}$ w.r.t. Model I support is

$$f^{I}_{X_{(n)}}(t) = \frac{nt^{n-1}}{\theta^n}\mathbf{1}\{0 \le t \le \theta\}.$$

Using these densities, we are now ready to present the estimators. Tate's UMVUE is given by

$$q_I(X_{(n)}) = 1 - \left(1 - \frac{1}{n}\right)\frac{\tau}{X_{(n)}}, \quad 0 < \tau < X_{(n)}. \tag{7}$$

Its variance w.r.t. Model I support is given by

$$MSE_I\left(q_I(X_{(n)})\right) = Var_I\left(q_I(X_{(n)})\right) = \left(\frac{\tau}{\theta}\right)^2\frac{1}{n(n-2)}.$$

Tate's estimator variance tends to zero in a polynomial rate of order -2, i.e., $Var_I\left(q_I(X_{(n)})\right) = \mathcal{O}(n^{-2})$. In order to compute the cross model asymptotic efficiency, we have to calculate Tate's estimator MSE w.r.t. Model II support as well. The calculation of Tate's estimator expectation w.r.t. Model II support yields the following complicated result:

$$\frac{(n-1)\tau(\theta-\gamma)^{-n}\left(\theta^n\left(\frac{\gamma}{\theta}\right)^n B_{\frac{\gamma}{\theta}}(1-n,n) - \pi\gamma^n\csc(\pi n)\right)}{\gamma} + 1 \tag{8}$$

where *csc* of some variable $x$ is the *cosec* function which is defined to be $sin^{-1}(x)$, and $B_{\frac{\gamma}{\theta}}(1-n,n)$ is incomplete Beta function such that $\frac{\gamma}{\theta}$ is the upper bound of the variable's support, i.e.,

$$B_{\frac{\gamma}{\theta}}(1-n,n) = \int_0^{\gamma/\theta} x^{-n}(1-x)^{n-1}dx.$$

The second moment of Tate's estimator is even more complicated and given by

$$\begin{aligned} &\frac{\gamma^{n-2}\left(\gamma\left(\frac{\theta}{\gamma}-1\right)^n(\gamma\theta n + (n-1)^2\tau^2)\right)}{\theta n(\theta-\gamma)^n} \\ &+ \frac{\gamma^{n-2}\left(\theta(n-1)\tau\left(2\gamma n + (n-1)^2\tau\right)\left(\left(\frac{\gamma}{\theta}\right)^n\left(\frac{\theta}{\gamma}\right)^n B_{\frac{\gamma}{\theta}}(1-n,n) - \pi\csc(\pi n)\right)\right)}{\theta n(\theta-\gamma)^n}. \end{aligned}$$

The MSE of Tate's estimator w.r.t. Model II support can be obtained from the two results described above. Note however that the bias term does not tend to zero and hence neither the MSE.

We are now ready to present BB's estimator, which is given by

$$q_{II}(X_{(1)}, X_{(n)}) = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right) \frac{\tau - X_{(1)}}{X_{(n)} - X_{(1)}}, \quad X_{(1)} < \tau < X_{(n)} \tag{9}$$

We computed the expected value of BB's estimator w.r.t. Model I support, and surprisingly it can be shown to be unbiased estimator. Therefore, the mean-squared-error of BB's estimator equals its variance which is given by

$$Var_I(q_{II}(X_{(1)}, X_{(n)})) = \frac{\theta^2(n-1) - 2\theta n\tau + 2n\tau^2}{\theta^2(n-3)n^2}.$$

We can see that BB's estimator variance tends to zero in a polynomial rate of order -2, i.e., $Var_I(q_I(X_{(1)}, X_{(n)})) = \mathcal{O}(n^{-2})$, therefore BB's estimator is consistent as well and converges to the function on interest in the same rate as Tate's estimator.

Now we can compute the asymptotic relative efficiency of the estimators w.r.t. Model I support:

$$\lim_{n\to\infty} \left( \frac{MSE_I(q_I(X_{(n)}))}{MSE_I(q_{II}(X_{(1)}, X_{(n)}))} \right) = \frac{\tau^2}{(\theta - \tau)^2 + \tau^2}.$$

We can see that asymptotically, for any $\tau < \theta$, Tate's estimator has smaller risk.

Figure 1 illustrates the convergence rate of both estimators under Model I support as a function of the sample size $n$, for $\theta = 1$, when $\tau = 0.25$ and when $\tau = 0.75$. We can see that under Model I support, Tate's estimator has smaller risk for any finite sample size. However asymptotic relative efficiency depends on $\tau$. Figure 2 illustrates the asymptotic MSE ratio of the estimators under Model I support for $0 \leq \tau \leq 1$ while $\theta = 1$. We can see that for small values of $\tau$, Tate's estimator posses much smaller risk. Tate's estimators preserves that superiority up to very large values of $\tau$, such that if $\theta = \tau + \delta$ for arbitrary small $\delta$ there is no significant difference between the two estimator's asymptotic risks.

Now consider the converse situation, i.e., Model II support, $\gamma > 0$ when $S = (\gamma, \theta)$, holds. In this case Tate's estimator is clearly biased and does not converge to the function of interest. In other words, in the case of relative large left truncation, the estimator will converge in probability to the wrong function and will be inadequate estimator to the tail probability. More specifically, the bias term will go asymptotically to some positive proportional function of the unknown truncation parameters $\eta = (\gamma, \theta)$.

To conclude this example, we can say that mistakenly assuming left truncation will result in significant deficiency in the efficiency and accuracy of the estimation for small values of $\tau$ and small sample sizes as well; Tate's estimator has uniformly (w.r.t. $\tau$ and $n$, ceteris paribus) smaller risk for any reasonable $\tau$ and $n$, therefore Tate's estimator should be preferred when our goal is to estimate $P_\eta(X > \tau)$ and there is no reason to assume left truncation.

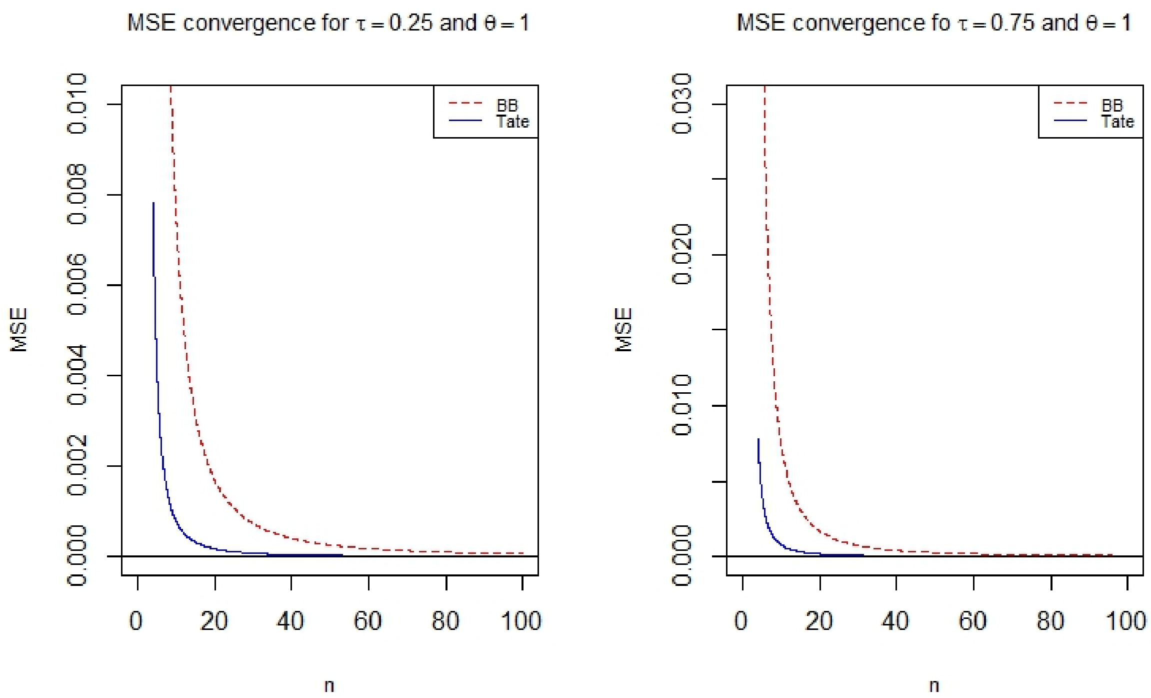


Figure 1: The BB's estimator MSE (dashed red) and Tate's estimator MSE (solid blue) convergence rate w.r.t Model I support as a function of $n$ for $\theta = 1$, when $\tau = 0.25$ (left panel) and when $\tau = 0.75$ (right panel).

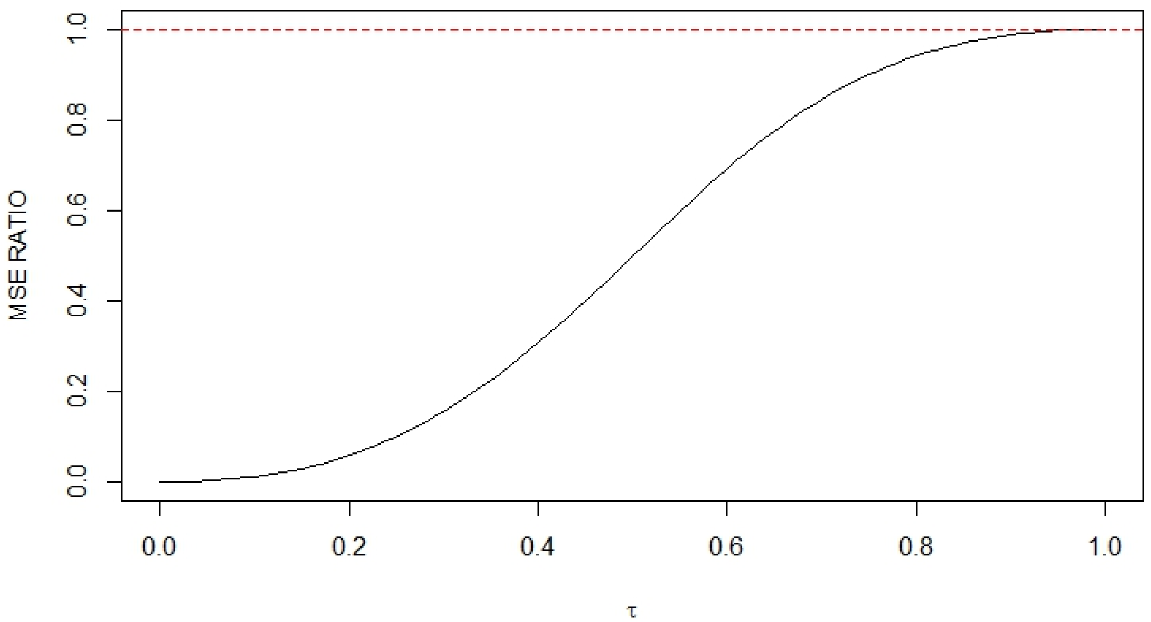


Figure 2: Asymptotic MSE ratio w.r.t. Model I support as a function of $\tau$ for $\theta = 1$. The red line in height 1 was added as a guideline.

.

## 2.4 Analysis of the Tail Probability Function for $B(\alpha+1,1)$ Distribution

In this example we consider the Beta distribution with parameters $\alpha+1$ and 1. As in the uniform distribution example of Section 2.3, we start with construction of the density function, by applying the principles described in Section 2.1. Afterwards, we will use the distribution's density function in order to derive the function of interest and compute the densities of the relevant order statistics. Finally, we will present the estimators under both models' support and compute their cross-model properties, such as expectation and mean-squared error. To conclude this example, we will compute the estimator's asymptotic efficiency w.r.t. both models support and draw conclusions regarding their appropriate use under different sets of model assumptions.

We begin with computing the density function under Model I support. In other words, suppose that $S=(0,\theta)$ holds, therefore the density of the $B(\alpha+1,1)$ random variable is given by

$$f^I(x;\gamma_0,\theta)=\frac{h(x)}{g_0(\gamma,\theta)}\mathbf{1}\{0\leq x\leq\theta\}=\frac{(\alpha+1)x^\alpha}{\theta^{\alpha+1}}\mathbf{1}\{0\leq x\leq\theta\}. \tag{10}$$

From this density function one can derive the function of interest of this example w.r.t. Model I support (when $\eta=\eta_0=(0,\theta)$):

$$P_{\eta_0}(X>\tau)=\frac{g_0(\tau,\theta)}{g_0(0,\theta)}\mathbf{1}\{0<\tau<\theta\}=1-\frac{\tau^{\alpha+1}}{\theta^{\alpha+1}}\mathbf{1}\{0<\tau<\theta\}.$$

The function of interest w.r.t. Model II support (when $\eta=(\gamma,\theta)$, $\gamma>0$) is given by

$$P_\eta(X>\tau)=\frac{g_0(\tau,\theta)}{g_0(\gamma,\theta)}\mathbf{1}\{\gamma<\tau<\theta\}=1-\frac{\tau^{\alpha+1}-\gamma^{\alpha+1}}{\theta^{\alpha+1}-\gamma^{\alpha+1}}\mathbf{1}\{\gamma<\tau<\theta\}.$$

Now, using (10) we can compute the joint density of $(X_{(1)},X_{(n)})$ under Model I support, which is a vector of the minimal and maximal order statistic. Its density given by

$$\begin{aligned} f^I_{X_{(1)},X_{(n)}}(y,t)&=\frac{n(n-1)h(y)h(t)(g_0(y,t))^{n-2}}{(g_0(y,t))^n}\mathbf{1}\{0\leq y\leq t\leq\theta\}\\ &=\frac{n(n-1)(yt)^\alpha(t^{\alpha+1}-y^{\alpha+1})^{n-2}(\alpha+1)^2}{\theta^{n\alpha+n}}\mathbf{1}\{0\leq y\leq t\leq\theta\}. \end{aligned} \tag{11}$$

The density of the maximal order statistic $X_{(n)}$ w.r.t. Model I support is given by

$$f^I_{X_{(n)}}(t)=\frac{(n\alpha+n)t^{n\alpha+n-1}}{\theta^{n\alpha+n}}\mathbf{1}\{0\leq t\leq\theta\}.$$

In order to compute BB estimator's variance we have to construct the joint density function of $(X_{(1)}, X_{(n)})$ under Model II support as well (when $\gamma > 0$). It can be shown that:

$$\begin{aligned} f^{II}_{X_{(1)},X_{(n)}}(y,t) &= \frac{n(n-1)h(y)h(t)(g_0(y,t))^{n-2}}{(g_0(y,t))^n}\mathbf{1}\{\gamma \le y \le t \le \theta\} \\ &= \frac{n(n-1)(yt)^{\alpha}(t^{\alpha+1}-y^{\alpha+1})^{n-2}(\alpha+1)^2}{(\theta^{\alpha+1}-\gamma^{\alpha+1})^n}\mathbf{1}\{\gamma \le y \le t \le \theta\}. \end{aligned} \tag{12}$$

Using (3), we can show that Tate's UMVUE w.r.t Model I support is given by

$$q_I(X_{(n)}) = 1 - \left(1-\frac{1}{n}\right)\left(\frac{\tau}{X_{(n)}}\right)^{\alpha+1}, \quad 0 < \tau < X_{(n)}. \tag{13}$$

Its variance, w.r.t. Model I support can be computed explicitly and is given by

$$MSE_I(q_I(X_{(n)})) = Var_I(q_I(X_{(n)})) = \left(\frac{\tau}{\theta}\right)^{2(\alpha+1)} \frac{1}{n(n-2)}. \tag{14}$$

Using (4), we can show that BB's UMVUE is given by

$$q_{II}(X_{(1)}, X_{(n)}) = \left(1-\frac{1}{n}\right) - \left(1-\frac{2}{n}\right)\frac{\tau^{\alpha+1}-X_{(1)}^{\alpha+1}}{X_{(n)}^{\alpha+1}-X_{(1)}^{\alpha+1}}, \quad X_{(1)} < \tau < X_{(n)}.$$

We are now move to discuss Model misspecification. Assume that Model I holds, i.e., $S = (0,\theta)$, then BB's estimator is finite-sample biased estimator such that

$$E_I\left(q_{II}(X_{(1)}, X_{(n)})\right) = \left(1+\frac{1}{n}\right) - \left(\frac{\tau}{\theta}\right)^{a+1}. \tag{15}$$

Recall that the MSE can be written as

$$MSE_I(q_{II}(X_{(1)}), X_{(n)})) = E_I(q_{II}(X_{(1)}, X_{(n)}))^2 - E_I^2(q_{II}(X_{(1)}, X_{(n)})) + b_I^2(q_{II}(X_{(1)}, X_{(n)}))$$

where $E_I^2(q_{II}(X_{(1)}, X_{(n)}))$ and $b_I^2(q_{II}(X_{(1)}, X_{(n)}))$ can be computed from the cross-model expectation presented above. Therefore, if we compute $E_I^2(q_{II}(X_{(1)}, X_{(n)}))$ (for the computation see Appendix A.5 ) we can find BB estimator's cross-model MSE. Hence, by change of variables and iterating use of integration by parts, one can obtain the following result:

$$\begin{aligned} MSE_I(q_{II}(X_{(1)}, X_{(n)})) &= \left(1-\frac{1}{n}\right)^2 - 2\left(\left(\frac{\tau}{\theta}\right)^{\alpha+1} - \frac{2}{n}\right)^2 \\ &+ \frac{(n-2)}{n^2(n-3)}\left(\left(\frac{\tau}{\theta}\right)^{2(\alpha+1)}\left(\frac{\tau}{\theta}\right)^{\alpha+1} 2n+2\right) \\ &- \left(1+\frac{1}{n}-\left(\frac{\tau}{\theta}\right)^{\alpha+1}\right)^2 + \left(\frac{1}{n}\right)^2. \end{aligned}$$

Figure 3 illustrates the MSE ratio Tate/BB w.r.t. Model I support as a functions of $\tau$ and $\alpha$, whereas Figure 4 illustrates both estimators' MSE convergence. All the computations

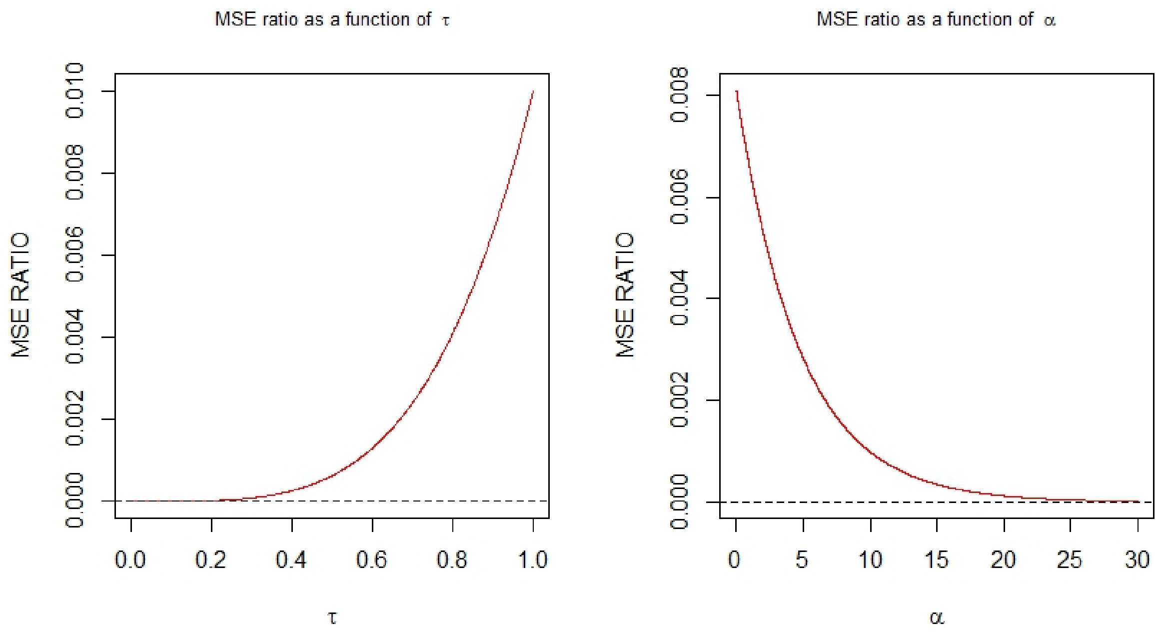


Figure 3: In the left panel we can see the approximately asymptotic MSE ratio Tate/BB w.r.t. Model I support as a function of $\tau$ for $\theta = 1$, $\alpha = 1$. The approximation is done by substituting $n = 10^4$ in expression for the MSE ratio. In the right panel we can see the MSE ratio Tate/BB w.r.t. Model I support as a function of $\alpha$ for the same set of parameters. Dashed black lines were added as a guidelines.

for this example can be found in Appendices A.3, A.4 and A.5.

Now consider the converse False Model error when Model II support holds, it can be shown that Tate's estimator is biased by a constant term and its MSE diverges (to infinity).

## 2.5 Analysis of the Expectation Function $E_\eta(X)$ for Uniform Distribution

Let $X_1, ..., X_n$ be a sample size of $n \geq 4$ from $\mathrm{U}(0, \theta)$ where $\theta$ is unknown truncation parameter. As in the previous examples, before we introduce the estimators, we start with the construction of the density function, proceed with the construction of the function of interest and density functions of the relevant order statistics.

We start with a construction of the density function w.r.t. Model I support, i.e., $\gamma \equiv \gamma_0 = 0$, and $S = (0, \theta)$.

$$f^I(x; \gamma_0, \theta) = \frac{h(x)}{g_0(0, \theta)} \mathbf{1}\{0 \leq x \leq \theta\} = \frac{1}{\theta} \mathbf{1}\{0 \leq x \leq \theta\}.$$

Now, using the density function, we can derive the function of interest w.r.t. Model I support:

$$E_{\eta_0}(X) = \frac{g_1(0, \theta)}{g_0(0, \theta)} = \frac{\theta}{2}.$$

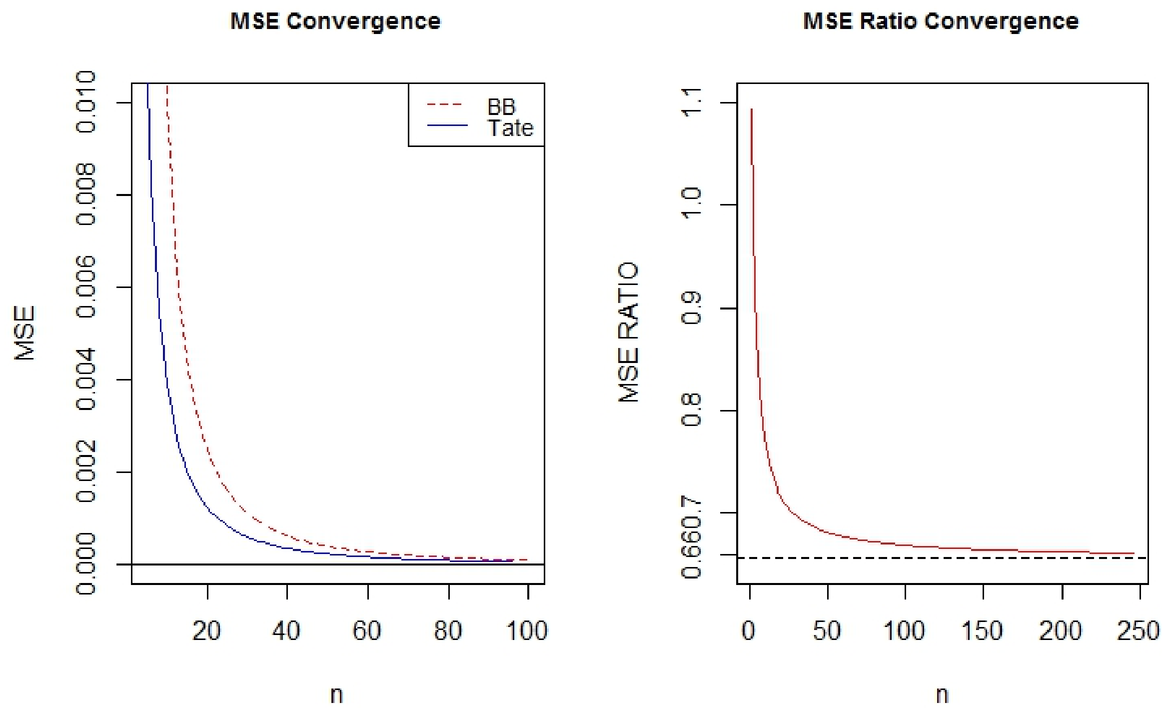


Figure 4: In the left panel we can see the convergence rate of the estimator's MSE w.r.t. Model I support as a function of $n$ for $\theta = 1$, $\alpha = 1$ and $\tau = 0.9$. The red dashed line designates BB estimator's MSE, where the solid blue line is for Tate's estimator MSE. In the right panel, we can see the convergence rate of the Tate/BB MSE ratio w.r.t. Model I support as a function of $n$ for the same set of parameters. Dashed black line was added as a guideline.

Under Model II support, using the density function given above, we can present the new function of interest which has undergone some changes as well:

$$E_\eta(X) = \frac{g_1(\gamma, \theta)}{g_0(\gamma, \theta)} = \frac{\gamma + \theta}{2}.$$

Finally, we are ready to present the estimators. Starting with Tate's estimator (see Eq. 1), that is given by the following expression:

$$q_I(X_{(n)}) = \frac{1}{2} X_{(n)} \left(1 + \frac{1}{n}\right).$$

Computation of the MSE w.r.t. Model I support yields

$$MSE_I(q_I(X_{(n)})) = Var_I(q_I(X_{(n)})) = \left(\frac{\theta}{2}\right)^2 \frac{1}{n(n+2)}.$$

Proceeding with BB's estimator (see Eq. 4), derived under Model II support, is given by

$$q_{II}(X_{(1)}, X_{(n)}) = \frac{1}{2} \left(X_{(1)} + X_{(n)}\right).$$

Computation of the variance w.r.t Model II support yields

$$MSE_{II}(q_{II}(X_{(1)}, X_{(n)})) = Var_{II}(q_{II}(X_{(1)}, X_{(n)})) = \frac{(\gamma-\theta)^2}{2(n+1)(n+2)}.$$

We are now ready to present the cross-model expectations. Suppose that Model II holds, i.e., $S = (\gamma, \theta)$. Hence the expectation of Tate's estimator w.r.t. Model II support is given by

$$E_{II}\left(q_I(X_{(n)})\right) = \frac{\gamma + n\theta}{2n}.$$

From the result above, Tate's estimator bias is obtained to be

$$b_{II}\left(q_{II}(X_{(1)}, X_{(n)})\right) = \frac{\gamma + n\theta}{2n} - \frac{\theta+\gamma}{2} = \frac{\gamma + n\theta - n\theta - n\gamma}{2n} = \frac{\gamma(1-n)}{2n}$$

Therefore, asymptotically Tate's estimator converges to the following function:

$$\lim_{n\to\infty}\left(E_{II}\left(q_I(X_{(n)})\right)\right) = \frac{1}{2}\theta$$

which is the expected value of uniform r.v. with only right truncation parameter $\theta$, and left known bound that is given by 0. Under Model II support Tate's estimator bias term asymptotically converges to $-\frac{\gamma}{2}$, hence it is inconsistent estimator.

Suppose now that Model I support holds, i.e., $\eta \equiv \eta_0 = (\gamma_0, \theta)$ when $\gamma_0 = 0$, and $S = (0, \theta)$ :

$$E_I\left(q_{II}(X_{(1)}, X_{(n)})\right) = \frac{1}{2}\theta.$$

We can see that although BB's estimator was derived w.r.t the improper support, it is still unbiased.

We are now ready to compute the cross-models mean squared errors, this is in order to (i) determine whether or not the estimators converge in distribution, and if so (ii) we would like to compare their MSE's in order to discuss their asymptotic efficiency. If Model I holds, i.e., $S = (\gamma = 0, \theta)$, we obtain that the MSE of BB's estimator is given by

$$\begin{aligned} MSE_I\left(q_{II}(X_{(1)}, X_{(n)})\right) &= Var_I\left(q_{II}(X_{(1)}, X_{(n)})\right) \\ &= \frac{\theta^2}{(n+1)(n+2)}. \end{aligned} \tag{16}$$

Next we compute Tate's estimator mean squared error w.r.t. Model II support, i.e., $S = (\gamma, \theta)$, this MSE is given by

$$\begin{aligned} &MSE_{II}\left(q_I(X_{(n)})\right) \\ &= \frac{\gamma^2\left(n^2-5\right) - \theta\left(\theta + (\theta-2)n^2 + 2(\theta-2)n\right) - 2\gamma(\theta + (\theta-1)n - 2)}{4n(n+2)}. \end{aligned} \tag{17}$$

Figure 5 illustrates the convergence rate of the estimators' MSE as a function of the

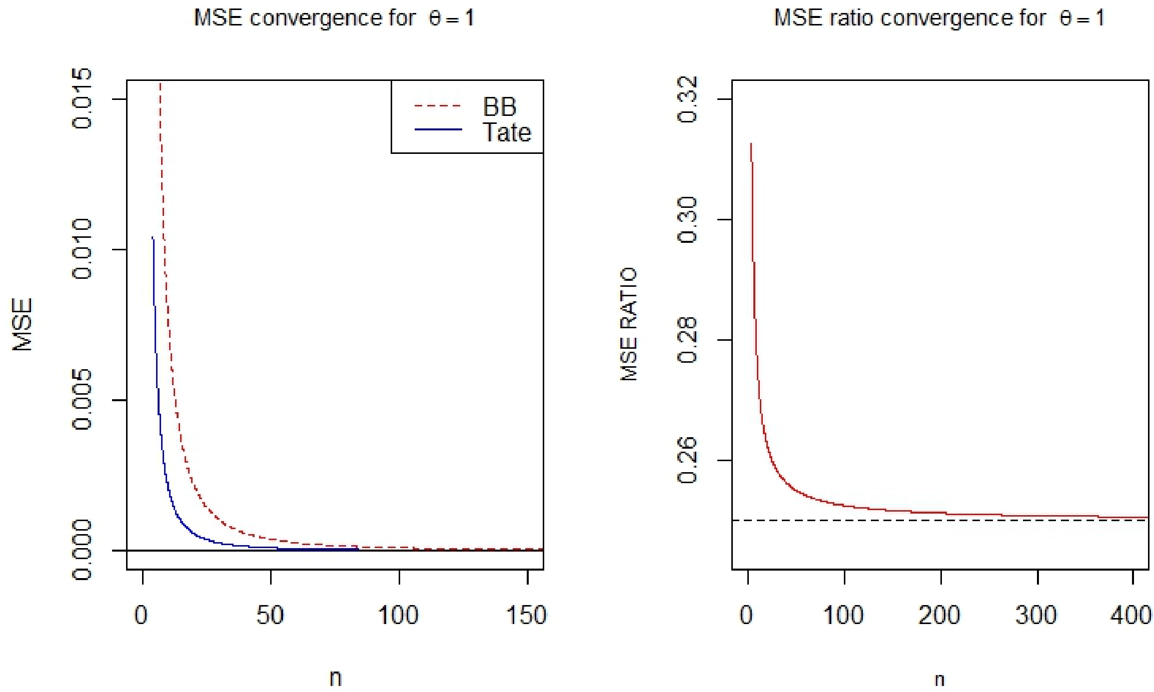


Figure 5: In the left panel we can see the convergence rate of the estimator's MSE w.r.t. Model I support as a function of $n$ for $\theta = 1$, the red dashed line designates BB's estimator MSE when solid blue line is for Tate's estimator MSE. In the left panel we can see the convergence rate of the MSE ratio Tate/BB w.r.t. Model I support as a function of $n$ for $\theta = 1$. Dashed black line added as a guideline.

sample size w.r.t. Model I support, and the convergence rate of the MSE ratio Tate/BB w.r.t. Model I support as a function of the sample size, when $\theta = 1$. From the illustration we can learn that Tate's estimator is uniformly better w.r.t. Model I support (i.e., has smaller risk) for every reasonable sample size.

We now compare the estimator's relative asymptotic efficiency. Assume that Model I holds. The following interesting result is follows from 16

$$\frac{MSE_I\left(q_I(X_{(n)})\right)}{MSE_I\left(q_{II}(X_{(1)}, X_{(n)})\right)} = \frac{n+1}{n}\frac{1}{4} \underset{n\to\infty}{\longrightarrow} \frac{1}{4}.$$

We can see that w.r.t. Model I support, Tate's estimator, which is UMVUE in that case, is uniformly better (w.r.t. any sample size $n$) than the BB alternative, and asymptotically has four times smaller risk (w.r.t. to any convex loss-function) than BB's estimator.

Needles to say that due to Tate's estimator constant bias term w.r.t. Model II support, the estimator does not converge in distribution to the function of interest. Therefore it inconsistent estimator and clearly asymptotically inefficient, since

$$\lim_{n\to\infty}\left(\frac{MSE_{II}\left(q_I(X_{(n)})\right)}{MSE_{II}\left(q_{II}(X_{(1)}, X_{(n)})\right)}\right) = \infty.$$

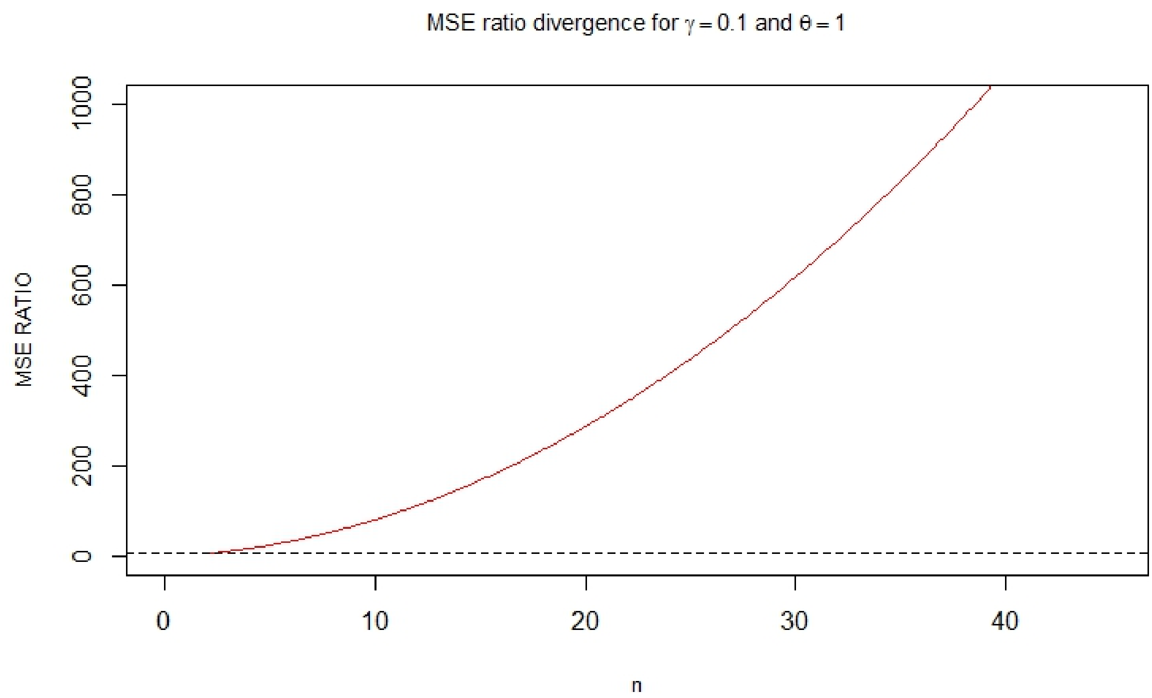


Figure 6: Divergence rate of the MSE ratio w.r.t. Model II support as a function of $n$ for $\theta = 1$ and $\gamma = 0.1$.

Figure 6 illustrates this behaviour. The figure shows that already for small sample size the MSE ratio is very large in favour of BB's estimator, and the divergence rate is exponential such that already for $n = 40$ the ratio exceed any reasonable value. Therefore, it is clear that w.r.t. Model II support Tate's estimators will result in inefficient and inaccurate estimation.

In conclusion of this example we can say that if we mistakenly assume that there is left truncation where in fact there is none, we can still perform quite good estimation with BB's estimator, although in this case the MSE will be at least four time larger then the optimal scenario. Nevertheless, False Model II error, i.e., assuming that there is no left truncation and using Tate's estimator for the mean value, will result in a severe deficiency since the estimator does not converge to the function of interest.

## 2.6 Conclusions

We have started our analysis with the estimation of the tail probability $P_\eta(X > \tau)$ in the uniform distribution. We discovered that assuming left truncation when the actual support is $(0, \theta)$ has its price. BB's estimator, although converges to the function of interest in polynomial rate of order -2, has uniformly inferior MSE comparing to Tate's UMVUE. Moreover the inefficiency grows as $\tau$ decreases. However, the converse mistake will lead to totally inappropriate estimation. The second example that deals with $Beta(\alpha + 1, 1)$ random variable yields the same result. Finally we examined the expectation in the uniform distribution. First we discovered that due to the symmetry property of the density function, BB's estimator is unbiased w.r.t. Model I support as well as w.r.t. Model II support. This result assured once again that BB's estimator will converge to the function of interest, therefore we focused the investigation on its MSE. Similar to the tail probability example, Tate's estimator is uniformly better then BB's estimator such that asymptotically Tate's estimator has exactly four times smaller MSE then BB's estimator. In this scenario, as well as in the others examples, making the converse False Model error results in great deficiency in estimation.

# 3 Setting II: Exponential Family with Possible Left Truncation

This chapter is organized as follows. An introduction to natural exponential families is given in Section 3.1. In Section 3.2, we present an exponential distribution example. In Section 3.3 an Erlang distribution example is presented. Finally, in Section 3.4, we discuss the results of this chapter. All computations of this Chapter appear in Appendix B.

## 3.1 Construction and Notations

Consider a one-dimensional natural exponential family (hereafter abbreviated NEF) supported on $S = (0, \infty)$ generated by a function $h : S \to [0, \infty)$, as follows. Define the Laplace transform

$$\mathcal{L}(\theta, 0) = \int_0^\infty e^{\theta x} h(x) dx$$

from the random variable's support to the natural parameter effective domain $D_0 \equiv \{\theta \in \mathcal{R}; \mathcal{L}(\theta, 0) < \infty\}$. We assume that $D_0$ has a non-empty interior $\Theta_0$. Define

$$k(\theta, 0) = \ln \mathcal{L}(\theta, 0).$$

The NEF corresponding to the natural parameter domain $\Theta_0$ is given by:

$$\mathcal{F}_0 = \{P_{\theta,0}(dx) = h(x) e^{\theta x - k(\theta,0)} \mathbf{1}\{0 < x < \infty\} dx, \theta \in \Theta\}.$$

Thus, due to the fact that

$$\int_0^\infty h(x) e^{\theta x - k(\theta,0)} dx = 1,$$

one can easily show that

$$\mathcal{L}(\theta, 0) = \int_0^\infty h(x) e^{\theta x} dx = e^{k(\theta,0)}.$$

It can be shown that $k(\theta, 0)$ is a strictly convex real analytic function on $\Theta_0$. From the analytical properties of $k(\theta, 0)$, we conclude that any finite derivative $\partial^j k_j(\theta, 0)/\partial\theta^j$ exists. Define $k_j(\theta, 0), j \geq 1$, to be the $j$-th cumulant of $\mathcal{F}_0$, where in particular, $\mu_0 = k_1(\theta, 0)$ is the mean function of $\mathcal{F}_0$. The mean function of $\mathcal{F}_0$ defines a bijective function from the natural parameter domain $\Theta_0$ to the mean domain $\Omega_0 \equiv k_1(\Theta, 0)$.

Set $\gamma > 0$, and define the Laplace transform from the support $S = (\gamma, \infty)$ to the natural parameters domain $D_\gamma = \{\theta \in \mathcal{R}; \mathcal{L}(\theta, \gamma) < \infty\}$ which is $\mathcal{L}(\theta, \gamma)$'s effective domain with a non-empty interior $\Theta_\gamma$. The family of probability densities generated by the support $S = (\gamma, \infty)$ is of the following form:

$$\mathcal{F}_\gamma = \{P_{\theta,\gamma}(dx) = h(x)e^{\theta x - k(\theta,\gamma)}\mathbf{1}\{\gamma < x < \infty\}dx, \theta \in \Theta_\gamma, \gamma > 0\}.$$

Similarly to the role of $\mu_0$ as the mean function of $\mathcal{F}_0$, $\mu_\gamma = k_1(\theta, \gamma)$ is the mean function of $\mathcal{F}_\gamma$; where $k_1(\theta, \gamma)$ is a bijective function from the natural parameters domain $\Theta_0$ to $\Omega_\gamma$. Clearly $\Theta_0 \subseteq \Theta_\gamma$ because $\mathcal{L}(\theta, \gamma) \le \mathcal{L}(\theta, 0)$, however for the sake of simplicity we shall assume that $\Theta_0 = \Theta_\gamma$, for all $\gamma > 0$. Due to the fact that $k_1(\theta, 0)$ is a bijective function between the $\Theta_0$ set and the corresponding mean domain $\Omega_0$, we will perform the analysis w.r.t. the mean domain and then apply the inverse function on $k_1(\theta, \gamma)$ to translate the results back to the parameters set $\Theta_0$.

Our interest will focus on the first order and second order derivatives of $k(\theta, 0)$ and $k(\theta, \gamma)$, which are the first and second cumulants, i.e., $k_1(\theta, \gamma) = \mu_\gamma$ and $k_2(\theta, \gamma) = \sigma^2_\gamma$. The first cumulant will serve to estimate functionals of $\mu_0$ and the second cumulant to derive the asymptotic variance of the estimators.

## 3.2 Exponential Distribution Example

Consider an exponential distribution under Model I support $S = (0, \infty)$, i.e., $X_1, ..., X_n \sim \exp(\lambda, 0)$, such that $\lambda = -\theta$, where $\theta$ is the natural parameter of the distribution, $h(x) = \mathbf{1}\{0 < x < \infty\}$, the Laplace transform

$$\mathcal{L}(\theta, 0) = \int_0^\infty e^{\theta x} h(x) dx = -\frac{1}{\theta},$$

the natural parameter domain is given by $\Theta_0 = (-\infty, 0)$. Note that $k(\theta, 0) = \ln(-\theta)$ and hence the mean domain is given by $\Omega_0 = (0, \infty)$. The maximum likelihood equation is given by

$$k_1(\theta, 0) = -\frac{1}{\theta} = \bar{X}_n = \hat{\mu}_0^I.$$

Now consider an exponential distribution under Model II support, $S = (\gamma, \infty)$, i.e., $X_1, ..., X_n \sim \exp(\lambda, \gamma)$, where $\lambda = -\theta, h(x) = \mathbf{1}\{\gamma < x < \infty\}$, the Laplace transform

$$\mathcal{L}(\theta, \gamma) = \int_\gamma^\infty e^{\theta x} h(x) dx = -\frac{1}{\theta} e^{\theta\gamma},$$

and the natural parameter domain is $\Theta_0 = (-\infty, 0)$. Note that $k(\theta, \gamma) = \theta\gamma - \ln(-\theta)$, and hence the mean domain is given by $\Omega_\gamma = (\gamma, \infty)$. In this case the maximum likelihood equation is given by

$$k_1(\theta, \gamma) = \gamma - \frac{1}{\theta} = \bar{X}_n = \hat{\mu}_\gamma. \quad (18)$$

Due to the fact that $\mu_0 = -\frac{1}{\theta}$ and by pluging $X_{(1)}$ instead of $\gamma$ in 18, we can derive the following MLE for $\mu_0$ w.r.t $\Omega_\gamma$:

$$\hat{\mu}_0^{II} = \bar{X}_n - X_{(1)},$$

and hence, the MLE for the natural parameter is

$$\hat{\theta}_{II} = \frac{1}{X_{(1)} - \bar{X}_n}.$$

We now move to discuss model misspecification. Consider the scenario when the MLE for $\theta$ was derived under Model I, however Model II actually holds. We should notice that $\mu_\gamma = -\frac{1}{\theta} + \gamma = \mu_0 + \gamma$. Therefore,

$$-\frac{1}{\bar{X}_n} - \theta \xrightarrow[II]{a.s} \frac{\theta^2\gamma}{1-\theta\gamma}, \tag{19}$$

which means that the sequence $\sqrt{n}(-\frac{1}{\bar{X}_n} - \theta)$ goes to infinity. For computations of this example see Appendix B.1. By simple calculations, we obtain that $k_2(\theta,\gamma) = k_2(\theta,0) = \frac{1}{\theta^2}$ which is quite unique observation which means that in the exponential case, truncation of the r.v. is equivalent to shifting of the r.v. by a factor of $\gamma$. Indeed, Bar-Lev and Boukai (2009) showed that this is the only case to which this property holds. These two results are illustrated in Figure 7 .

We now move to consider the case when the MLE was derived under Model II, but Model I holds. We first deal with a finite-simple behaviour of the MLE for the natural parameter $\theta$, and we then explore the asymptotic behaviour of the ML estimator.

We start with a finite sample behaviour of the MLE for the natural parameter $\theta$. It was shown by (Bar-lev and Boukai, 1985) that $\sum_{i=1}^{n}(X_i - X_{(1)})$ is distributed Erlang with $n-1$ and $\theta$. Therefore, $(\sum_{i=0}^{n}(X_i - X_{(1)}))^{-1}$ has an Inverse-Gamma distribution. We computed the MLE's density function, which can be found to be

$$f^I_{-\hat{\theta}_{II}}(y) = \frac{\theta^{n-1}}{\Gamma(n-2)} n^{n-1} \frac{1}{y^n} e^{-\theta n/y}. \tag{20}$$

For computations see Appendix B.1. For illustration see Figure 8. The MLE's expected value is $\frac{n\theta}{n-2}$, and therefore its bias is given by $\frac{2}{n-2}$. We computed its MSE, which is given by $\frac{(n(n+4)-12)\theta^2}{(n-2)^2(n-3)}$. Clearly it is asymptotically unbiased, because the bias is $\mathcal{O}(n^{-1})$. Figure 9 compares this MSE to the MSE of the MLE derived under the correct model, i.e., Model I. Since the MSE converges to 0, we conclude that $\hat{\theta}_{II}$ converges in probability to the natural parameter $\theta$.

Now we are ready to discuss the asymptotic behaviour of the MLE for the natural parameter $\theta$. The rate of convergence of $X_{(1)}$ to $\gamma$ is $\mathcal{O}(n^{-1})$, however the rate of convergence of $\bar{X}_n$ to $\mu_0$ is $\mathcal{O}(n^{-1/2})$. Hence, $\hat{\mu}_0^{II}$ goes to $\bar{X}_n$, and therefore asymptotically it remains unbiased and the effect on the MLE's variance vanishes asymptotically as well.

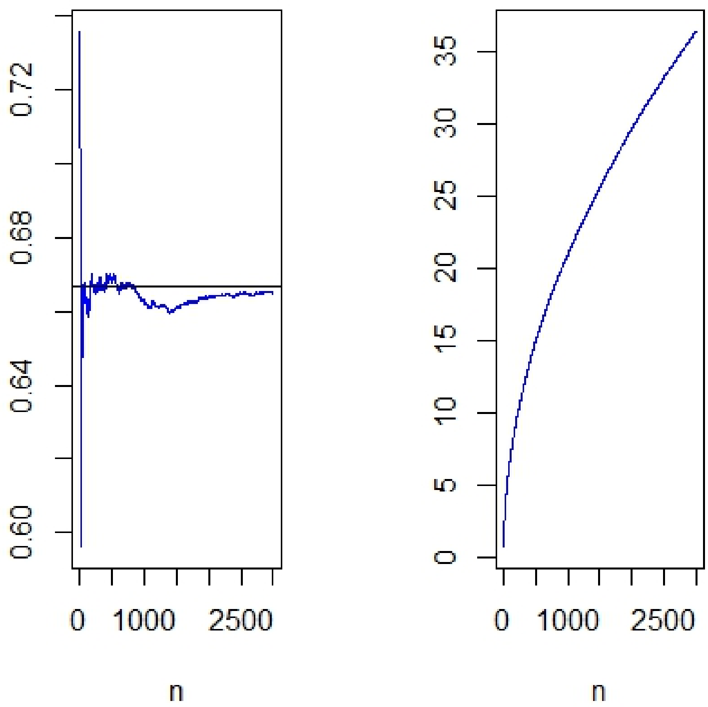


Figure 7: The left panel demonstrates the almost sure convergence of the sequence $-\frac{1}{\bar{X}_n} - \theta$ to $\frac{\theta^2 \gamma}{1-\theta\gamma} = 2/3$, for $\theta = -1$ and $\gamma = 2$. The right panel demonstrates the divergence of the sequence $\sqrt{n}(-\frac{1}{\bar{X}_n} - \theta)$ (to infinity). The simulation based on 3,000 random values from exponential truncated distribution with parameters stated above.

Consequently

$$\sqrt{n}\left(\hat{\mu}_0^{II} - \mu_0\right) \equiv \sqrt{n}(\bar{X}_n - X_{(1)} - \mu_0) = \sqrt{n}(\bar{X}_n - \mu_0) - o_p(1) \xrightarrow[I]{D} \mathcal{N}\left(0, k_2(\theta, 0)\right). \quad (21)$$

Hence, using delta method (Ferguson, 1996), we can state that:

$$\sqrt{n}\left(\frac{1}{X_{(1)} - \bar{X}_n} + \frac{1}{\mu_0}\right) \equiv \sqrt{n}\left(\hat{\theta}_{II} - \theta\right) \xrightarrow[I]{D} \mathcal{N}\left(0, \frac{1}{k_2(\theta, 0)}\right) \quad (22)$$

where $\frac{1}{X_{(1)} - \bar{X}_n} = -\frac{1}{\hat{\mu}_0^{II}} = \hat{\theta}_{II}$ is the MLE for the natural parameter $\theta$ under Model II support. For illustration of the convergence to the normal distribution see Figure 10.

### 3.3 Erlang Distribution Example

Consider an Erlang-2 distribution under Model I support, $S = (0, \infty)$, i.e., $X_1, ..., X_n \sim$ Erlang$(2, \lambda, 0)$, where $\lambda = -\theta$, and where $\theta$ is the distribution's natural parameter. We have $h(x) = x\mathbf{1}\{0 < x < \infty\}$, the Laplace transform $\mathcal{L}(\theta, 0) = \frac{1}{\theta^2}$, the natural parameter domain is $\Theta_0 = (-\infty, 0)$, and similarly to the exponential distribution example, $k_1(\theta, 0) =$

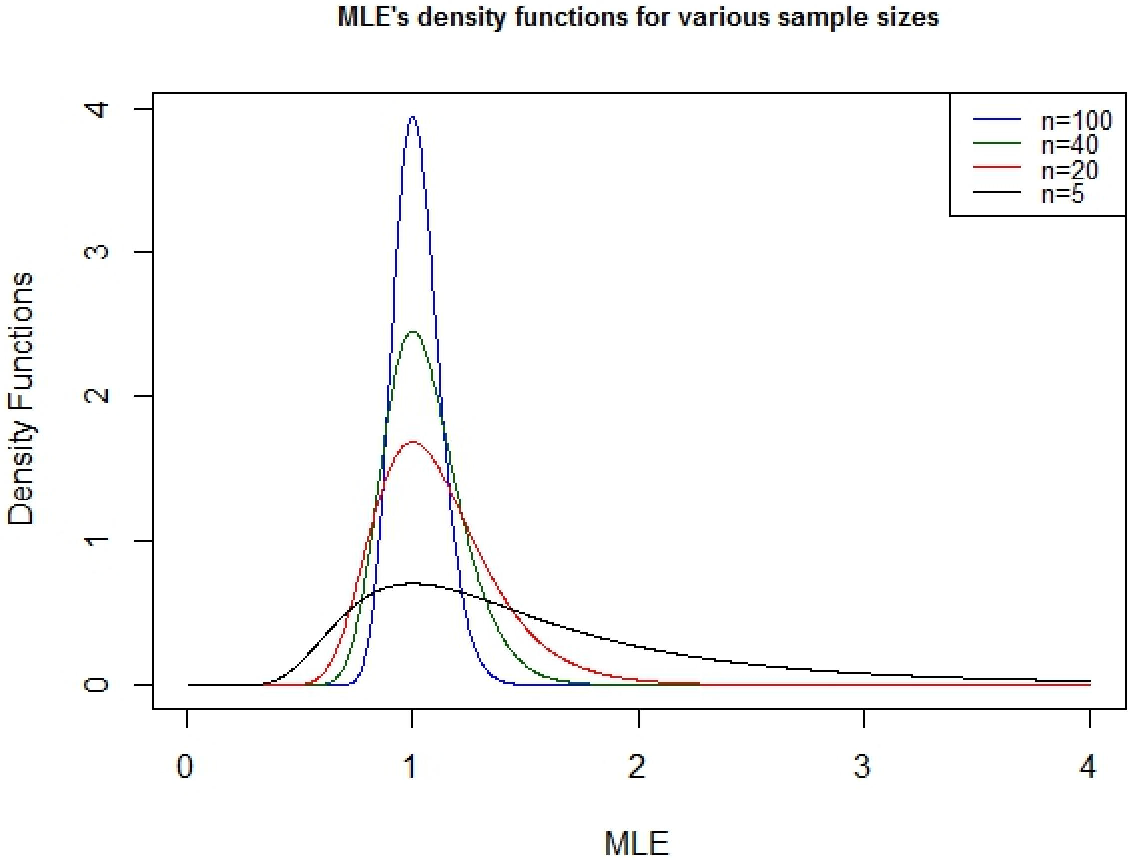


Figure 8: The figure shows the density functions of the natural parameter's MLE for various sample sizes. The exponential distribution rate parameter is taken to be 1. We can see that as $n$ increases the distribution function becomes more symmetric and resembles the normal asymptotic distribution which is its limit.

$-\frac{2}{\theta} = \mu_0$, and hence the mean domain is $\Omega_0 = (0, \infty)$. The ML equation is given by

$$-\frac{2}{\theta} = \bar{X}_n = \hat{\mu}_0^I.$$

Therefore, the MLE for $\theta$ under Model I support is $\hat{\theta}_I = -\frac{2}{\hat{\mu}_0^I} = -\frac{2}{\bar{X}_n}$.

Now consider the Erlang-2 distribution under Model II support, $S = (\gamma, \infty)$, i.e., $X_1, ..., X_n \sim \text{Erlang}(2, \lambda, \gamma)$, where $\lambda = -\theta$, and $h(x) = x\mathbf{1}\{\gamma < x < \infty\}$, the Laplace transform $\mathcal{L}(\theta, \gamma) = \frac{1}{\theta^2} e^{\theta\gamma} (1 - \gamma\theta)$ (for computations see Appendix B.4), and hence the mean domain is given by $\Omega_\gamma = (\gamma, \infty)$, and the natural parameter domain is given by $\Theta_0 = (-\infty, 0)$. It is easy to show that $k_1 (\theta, \gamma) = -\frac{2}{\theta} + \gamma - \frac{\gamma}{1-\theta\gamma} = \mu_\gamma$. But since $\mu_0 = -\frac{2}{\theta}$, simple computations shows

$$\mu_\gamma = \mu_0 + \gamma - \frac{\mu_0\gamma}{\mu_0 + 2\gamma}.$$

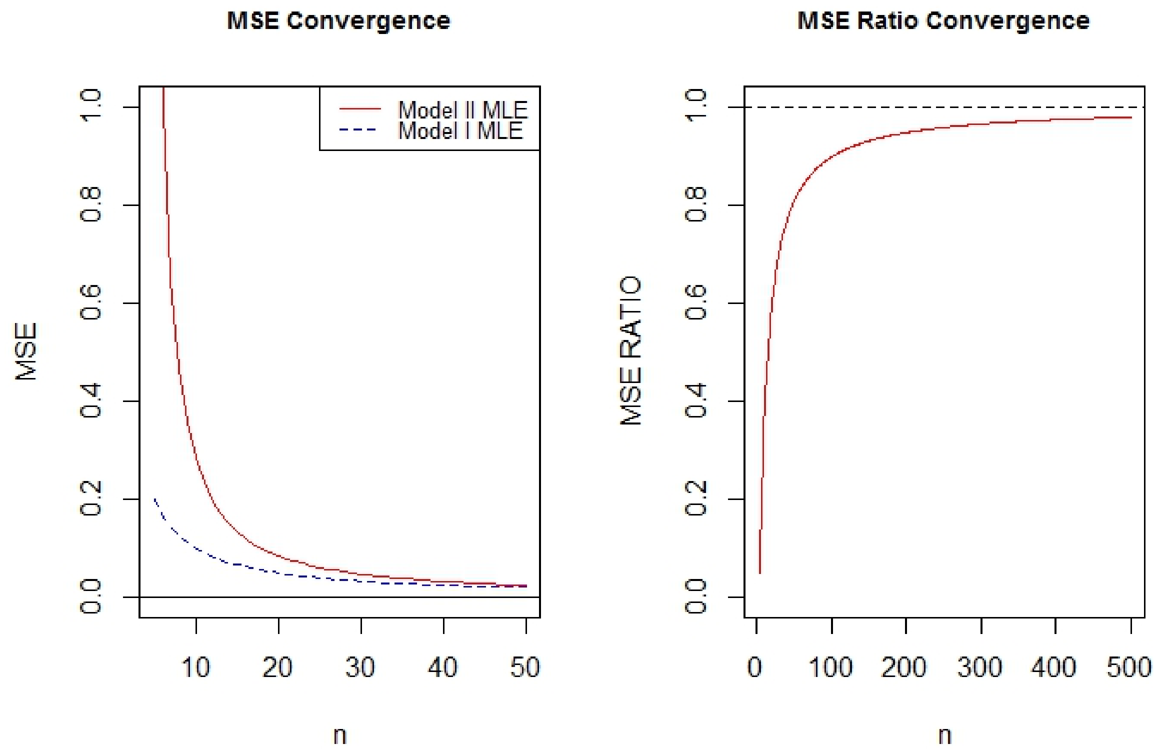


Figure 9: Illustration of the MSE convergence of MLEs for the rate parameter $\lambda$ in exponential distribution with $\lambda = 1$, where $\lambda = -\theta$, . In the left panel we can see the MSE convergence of both MLEs w.r.t. Model I support. The solid red line represents the MSE of the MLE derived under Model II support. Dashed blue line represents the MSE of the MLE derived w.r.t. Model I support. The solid black line was added as a guideline. In the right panel we can see the Model I/Model II MSE ratio convergence. The dashed black line was added as a guideline.

Therefore, we can derive the maximum likelihood equation for $\mu_\gamma$ w.r.t. $\Omega_\gamma$ mean domain:

$$\bar{X}_n = \hat{\mu}_0^{II} + X_{(1)} - \frac{\hat{\mu}_0^{II} X_{(1)}}{\hat{\mu}_0^{II} + 2X_{(1)}}.$$

Some simple algebra gives us the following maximum likelihood quadratic equation w.r.t $\hat{\mu}_0^{II}$ in $\Omega_\gamma$ mean domain:

$$0 = (\hat{\mu}_0^{II})^2 + \hat{\mu}_0^{II}(2X_{(1)} - \bar{X}_n) + 2X_{(1)}(X_{(1)} - \bar{X}_n).$$

This equation has a.s. two real roots, of which only the root with the plus sign is positive and therefore consistent with the $\Omega_0$ mean domain. Therefore, under Model II support, the MLE for $\mu_0$ is:

$$\hat{\mu}_0^{II} = \frac{1}{2}\left(\bar{X}_n - 2X_{(1)} + \sqrt{4X_{(1)}(\bar{X}_n - X_{(1)}) + \bar{X}_n^2}\right).$$

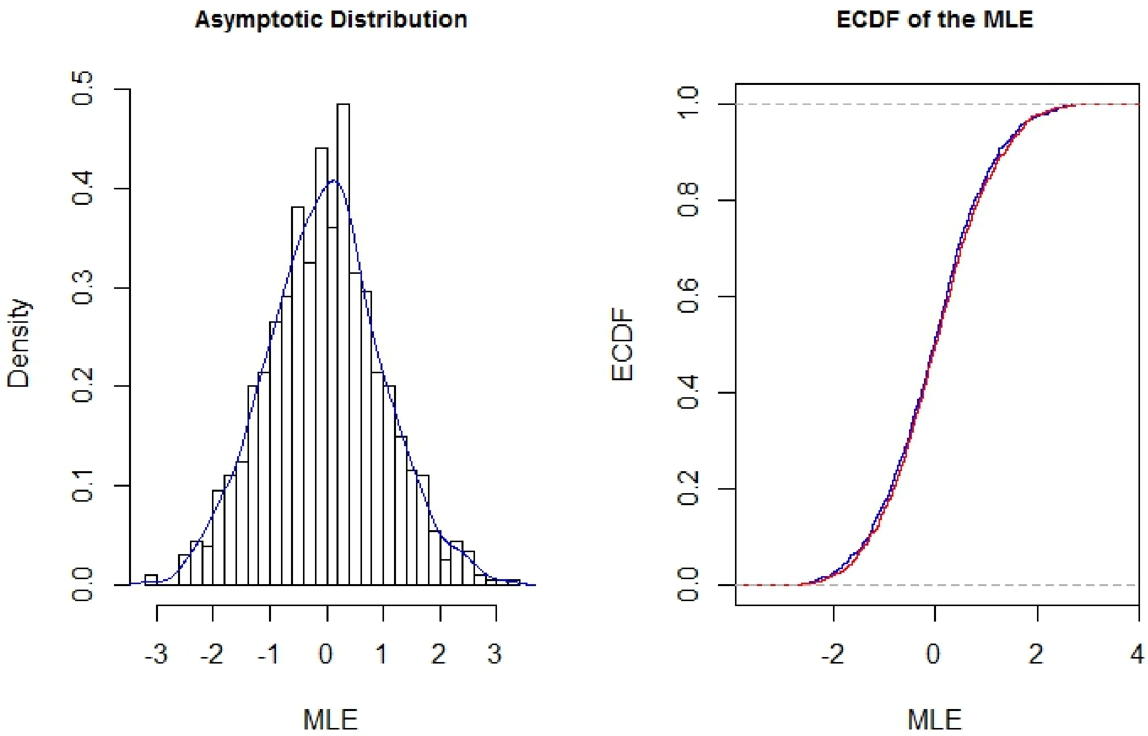


Figure 10: In the left panel we can see the asymptotic cross-model distribution of the MLE for the rate parameter $\lambda$, such that $\lambda = -\theta$. In the right panel we can see the empirical cumulative distribution of the MLE (solid blue) vs. ECDF of Normal $(0, 1)$ (solid red). The simulation based on 1,000 ML estimators, where each of the estimators constructed from 10,000 random values from exponential distribution with rate parameter equals 1.

Consequently we obtained the MLE for $\theta$ under Model II support

$$\hat{\theta}_{II} = -\frac{2}{\hat{\mu}_0^{II}} = -4\left(\bar{X}_n - 2X_{(1)} + \sqrt{4X_{(1)}(\bar{X}_n - X_{(1)}) + \bar{X}_n^2}\right)^{-1}.$$

We now move to discuss model misspecification. Consider the scenario when the MLE was derived under Model I support, however Model II holds. One can prove that

$$-\frac{2}{\bar{X}_n} - \theta \xrightarrow[II]{a.s} \frac{-\theta^3\gamma^2}{2 + \theta\gamma(\theta\gamma - 2)}. \tag{23}$$

Hence, the sequence $\sqrt{n}(-\frac{2}{\bar{X}_n} - \theta)$ goes to infinity. For computations of this example see Appendix B.2.

One can see that assuming $S = (0, \infty)$ while actually $S = (\gamma, \infty)$, for significantly large truncation $\gamma$ will result in a meaningful estimation deficiency.

We move to consider the scenario that the MLE was derived under Model II support, but Model I holds. In this case we will derive the asymptotic behaviour of the MLE for the natural parameter explicitly. The finite sample behaviour is analytically complicated and is demonstrated using simulations.

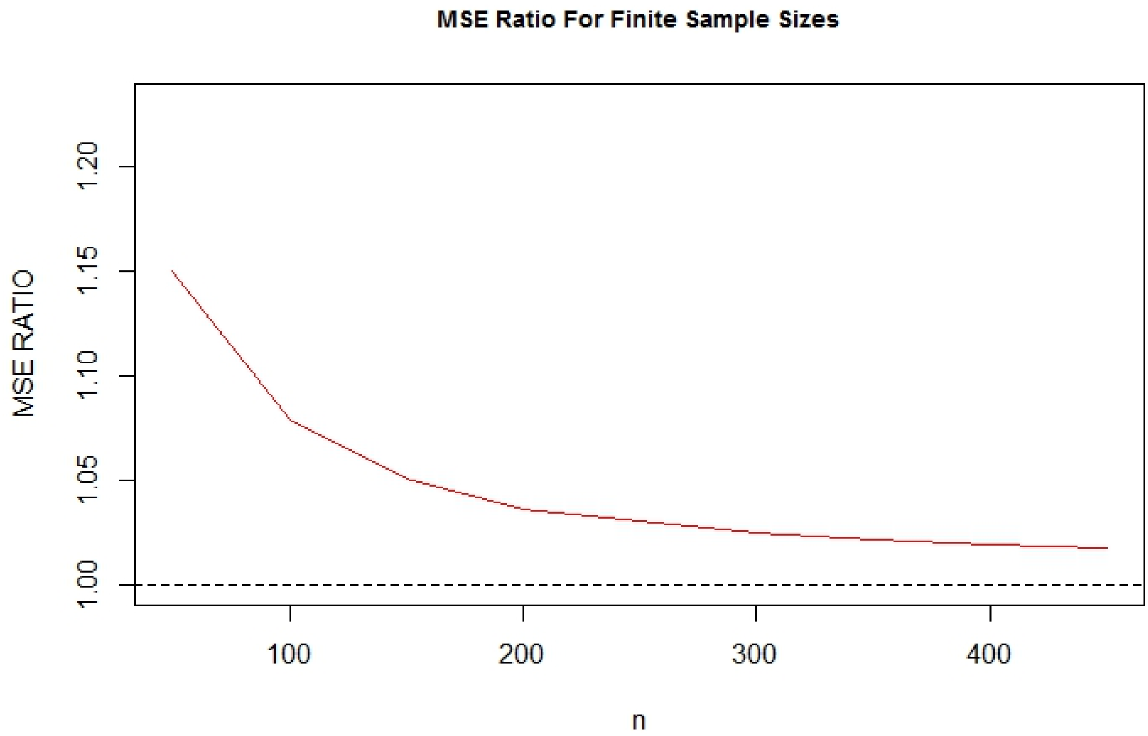


Figure 11: MLEs MSE ratio (Model I/Model II) for various sample sizes. The simulation based on 100,000 ML estimators each of one constructed from various number of random values (sample sizes) drawn from Gamma distribution with $\lambda = 1$ and $k = 2$.

.

In Figure 11 we can observe the rate of convergence of the MLEs MSE ratio for finite sample sizes. As we can see, already for $n = 200$ the MSE ratio just slightly differs from 1. Therefore there is no need in asymptotic sample sizes in order to perform efficient estimation with Model II MLE when Model I support holds.

Now we are ready to discuss the asymptotic behaviour of the MLE for the natural parameter. Similar to the exponential case, the rate of convergence of $X_{(1)}$ to $\gamma$ is $\mathcal{O}(n^{-1})$, but the rate of convergence of $\bar{X}_n$ to $\mu_0$ is $\mathcal{O}(n^{-1/2})$, therefore $\hat{\mu}_0^{II}$ goes to $\bar{X}_n$. Hence by iterating use of Slutsky lemma and Continuous Mapping theorem (see Ferguson, 1996, Theorem 6) we can state that asymptotically $\hat{\mu}_0^{II}$ goes to $\mu_0$ with probability 1, and therefore, $\hat{\mu}_0^{II}$ remains unbiased and efficient, i.e., the MLE's variance meets the Cramer-Rao lower bound for unbiased estimators of the natural parameter.

$$\sqrt{n}(\hat{\mu}_0^{II} - \mu_0) \xrightarrow[I]{D} \mathcal{N}\left(0, k_2(\theta, 0)\right).$$

Hence, using the delta method, and the fact that $k_2(\theta,0)/(k_1(\theta,0))^4 = \frac{2/\theta^2}{(-2/\theta)^4} = \frac{\theta^2}{2^3} = \frac{1}{2^2}\frac{1}{k_2(\theta,0)}$ we can deduce the following asymptotic behaviour:

$$\sqrt{n}\left(-\frac{2}{\hat{\mu}_0^{II}} - -\frac{2}{\mu_0}\right) \equiv \sqrt{n}\left(\theta - \hat{\theta}_{II}\right) \xrightarrow[I]{D} \mathcal{N}\left(0, \frac{1}{2^2 k_2(\theta, 0)}\right).$$

For conclusion of the two examples presented above, we saw that assuming mistakenly that Model II support holds and deriving the MLE for the natural parameter $\theta$ w.r.t. $S = (\gamma, \infty)$ does not result in severe inefficiency. In this situation, due to the fast rate (by a quadratic order faster then $\bar{X}_n$ rate of convergence to the distribution mean) of convergence of $X_{(1)}$ (which is the estimator of the left side truncation parameter $\gamma$) to $\gamma$, the estimator remains asymptotically (i) unbiased and (ii) efficient. However, if we derive the MLE for the natural parameter $\theta$ under Model I support, while actually Model II holds, there will be a bias that will be proportional to the truncation parameter $\gamma$, i.e., the estimator will converge almost surely to some function of $\theta$ and $\gamma$. Hence, using Model's I MLE for $\theta$ when actually Model II holds result in biased and inaccurate estimator.

## 3.4 Conclusions

In exponential and Erlang-2 distributions we saw that deriving the MLE for the natural parameter w.r.t. Model II support will not cause severe deficiency to the estimations process, that is due to the fact that the MLE asymptotically remains (i) unbiased and (ii) efficient. However for finite samples the MLE for the natural parameter is a biased and slightly inefficient estimator with an Inverse-Gamma like distribution up to multiplication by a sample size. Further investigation is required in order to generalize the results to the whole exponential families.

# 4 Conclusions and Discussion

In this work we analysed the effect of left truncation parameter on estimation in continuous distribution functions. We discussed two main settings; general continuous right truncated models with possible left truncation and exponential families with possible left truncation. In Setting I we illustrated the effect of introducing left truncation on estimation of the tail probabilities. These analysis performed using uniform and Beta distributions. For the uniform distribution the effect on the estimation of the mean function considered as well. In Settings II we illustrated the effect of left truncation on maximum likelihood estimators of the natural parameter in two representative examples of the NEF; The exponential and Erlang-2 distributions used to derive a finite-sample and the asymptotic effect of left truncation on the estimation process.

In conclusion of the investigation of continuous right truncated models with possible left truncation, we learned that if we mistakenly assume that there is left truncation where in fact there is none, we can still perform quite good estimation with BB's estimator. Although Tate's estimator is uniformly better (w.r.t. to any convex risk function), BB's estimator converge to the function of interest in the same rate as the estimator that was derived w.r.t. Model I support. Nonetheless, it is important to note that there is a significant price for that mistake such that the MSE of BB's estimator can be much larger than Tate's. For example, see the uniform distribution expectation estimators where using BB's estimator while Model I holds, will result in asymptotic four time larger MSE. However, the converse error, i.e., assuming mistakenly that there is no left truncation and using Tate's estimator for the mean value, may result in severe deficit; the estimator will not converge to the function of interest and will be generally inappropriate.

In conclusion of the investigation of exponential families with possible left truncation, we can state that similar results yield the analysis of maximum likelihood estimators in the exponential and Erlang-2 distributions as in Setting I. In the case where the right model is Model I, however the MLE was derived under Model II support, this error will not cause any deficiency in terms of asymptotic results. More specifically, the MLE for the natural parameter will be asymptotically (i) unbiased and (ii) efficient when derived under Model II support (where the right model is Model I). Furthermore, finite sample analysis shows that the MLE distribution is Inverse-Gamma related. In the exponential example we computed the explicit bias and MSE for finite sample sizes. These computations showed that there is finite-sample deficiency (bias and large MSE) that vanishes asymptotically in this type of error. However, the converse error, i.e., assuming mistakenly that there is no left truncation and using the MLE w.r.t. Model I support may cause severe deficiency because in this case the estimator will be biased by a constant term, therefore it will diverge (to infinity) and be generally inappropriate. Further research is required in order to generalize (if it is possible) the results and draw conclusions to the whole exponential families.

Finally, we can state that based on the two examples described above, if there is a good reason to suspect that the model involves left truncation, we need not hesitate to use the more complex estimator. This is due to the fact that a more complicated model does not involve substantial loss in the estimation efficiency. Nevertheless, there is still a price for using Model II estimators where Model I support holds, therefore if there is no reason to assume left truncation - we should avoid doing so.

# A Computations for Chapter 2

## A.1 Section 2.2, Example 1

**Lemma 5.** *Let* $\xi(\eta_0) = E_{\eta_0}(X) = \xi(\theta) = g_1(\gamma_0, \theta)/g_0(\gamma_0, \theta)$ *with known* $\gamma_0 = 0$. *Then Tate's estimator is given by*

$$q_I(X_{(n)}) = \frac{g_1(0, X_{(n)})}{g_0(0, X_{(n)})}\left(1 - \frac{1}{n}\right) + \frac{X_{(n)}}{n},$$

*whenever the derivative* $\xi'(\theta) = \partial\xi(\theta)/\partial\theta$ *exists and is continuous almost everywhere on* $\Theta = \{(0, \theta) : a < \theta < b\}$.

*Proof.* Tate's estimator is given by

$$q_I(X_{(n)}) = \xi(X_{(n)}) + \frac{\xi'(X_{(n)})g_0(0, X_{(n)})}{nh(X_{(n)})},$$

for any estimable function $\xi(\eta_0) \equiv \xi(\theta)$ of the model's only unknown parameter $\theta$. Therefore,

$$q_I(\theta) = \frac{g_1(0, \theta)}{g_0(0, \theta)} + \frac{\partial}{\partial\theta}\left(\frac{g_1(0, \theta)}{g_0(0, \theta)}\right)\frac{g_0(0, \theta)}{nh(\theta)},$$

recall that,

$$g_k(\gamma, \theta) = \int_{\gamma}^{\theta} x^k h(x)dx, k = 0, 1, 2, ...$$

Hence, using Liebniz rule w.r.t. Model I support, i.e., while $\gamma \equiv \gamma_0 = 0$ yields

$$\frac{\partial}{\partial\theta}g_k(\gamma_0, \theta) = \frac{\partial}{\partial\theta}\int_{0}^{\theta} x^k h(x)dx = \theta^k h(\theta).$$

Applying the rule we obtain that

$$\begin{aligned} q_I(\theta) &= \frac{g_1(0, \theta)}{g_0(0, \theta)} + \frac{h(\theta)\left(\theta g_0(0, \theta) - g_1(0, \theta)\right)}{g_0(0, \theta)^2}\frac{g_0(0, \theta)}{nh(\theta)} \\ &= \frac{g_1(0, \theta)}{g_0(0, \theta)} + \frac{g_1(0, \theta)}{ng_0(0, \theta)} + \frac{\theta}{n}. \end{aligned}$$

By rearranging the equation and plugging $X_{(n)}$ instead of $\theta$, one can show that

$$q_I(X_{(n)}) = \frac{g_1(0, X_{(n)})}{g_0(0, X_{(n)})}\left(1 - \frac{1}{n}\right) + \frac{X_{(n)}}{n}. \quad \square$$

## A.2 Section 2.2, Example 3

**Lemma 6.** *Under Model II, if* $\xi(\eta) = E_\eta(X) = \xi(\gamma, \theta) = g_1(\gamma, \theta)/g_0(\gamma, \theta)$ *the general form of BB's estimator is:*

$$q_{II}(X_{(1)}, X_{(n)}) = \frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})}.$$

*Proof.* BB's estimator general form is given by (4),whenever the partial derivations $\xi_1(\gamma, \theta) = \partial\xi(\gamma, \theta)/\partial\gamma, \xi_2(\gamma, \theta) = \partial\xi(\gamma, \theta)/\partial\theta$, and $\xi_{12}(\gamma, \theta) = \partial^2\xi(\gamma, \theta)/\partial\gamma\partial\theta$ exist and are continuous almost everywhere on $\Theta = \{(\gamma, \theta) : a < \gamma < \theta < b\}$. In order to simplify the calculations, we first calculate all the required derivation of the function of interest. Using Leibniz rule w.r.t. Model II support yields

$$\frac{\partial}{\partial\gamma} g_k(\gamma, \theta) = \frac{\partial}{\partial\gamma} \int_\gamma^\theta x^k h(x)dx = -\gamma^k h(\gamma),$$

hence,

$$\xi_1(\gamma, \theta) = \frac{h(\gamma)\left(g_1(\gamma, \theta) - \gamma g_0(\gamma, \theta)\right)}{\left(g_0(\gamma, \theta)\right)^2},$$

similarly,

$$\xi_2(\gamma, \theta) = \frac{h(\theta)\left(\theta g_0(\gamma, \theta) - g_1(\gamma, \theta)\right)}{\left(g_0(\gamma, \theta)\right)^2},$$

finally,

$$\xi_{12}(\gamma, \theta) = \frac{h(\gamma)h(\theta)\left(\theta g_0(\gamma, \theta) - \gamma g_0(\gamma, \theta) - 2g_1(\gamma, \theta) + 2\gamma g_0(\gamma, \theta)\right)}{\left(g_0(\gamma, \theta)\right)^3}.$$

Plugging the results to the BB's estimator and substituting $\theta$ with $X_{(n)}$ and $\gamma$ with $X_{(1)}$, yields

$$\begin{aligned} q_{II}(X_{(1)}, X_{(n)}) &= \frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})} - \frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})n(n-1)} + \frac{X_{(1)}}{n-1} \\ &+ \frac{X_{(n)}}{n-1} - \frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})n(n-1)} \\ &- \frac{X_{(1)}}{n-1} - \frac{X_{(n)}}{n-1} + 2\frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})n(n-1)} \\ &= \frac{g_1(X_{(1)}, X_{(n)})}{g_0(X_{(1)}, X_{(n)})}. \quad \square \end{aligned}$$

## A.3 Section 2.4, Beta Distribution Example, Eq. 14

**Lemma 7.** *Let* $X_1, ..., X_n \sim B(\alpha + 1, 1)$*. Suppose that Model I holds. Tate's estimator for*

$$P_{\eta_0}(X > \tau) = 1 - \frac{\tau^{\alpha+1}}{\theta^{\alpha+1}} \mathbf{1}\{0 < \tau < \theta\}.$$

*is given by*

$$q_I(X_{(n)}) = 1 - \left(1 - \frac{1}{n}\right)\left(\frac{\tau}{X_{(n)}}\right)^{\alpha+1}, \quad X_{(1)} < \tau < X_{(n)}$$

*Moreover,*

$$E_I\left(q_I(X_{(n)})\right)^2 = 1 - 2\left(\frac{\tau}{\theta}\right)^{a+1} + \frac{(n-1)^2}{n(n-2)}\left(\frac{\tau}{\theta}\right)^{2(a+1)}.$$

*Proof.* The first assertion follows from application of Example 2. Note that

$$E_I\left(q_I(X_{(n)})\right)^2 = \int_0^\theta q_I^2(t) f_{X_{(n)}}(t)dt.$$

Recall that the density of the maximal order statistic $X_{(n)}$ w.r.t. Model I support is given by

$$f^I_{X_{(n)}}(t) = \frac{(n\alpha+n)t^{n\alpha+n-1}}{\theta^{n\alpha+n}}\mathbf{1}\{0 \le x \le \theta\},$$

hence, one can simplify the equation by expending $\left(q_I(X_{(n)})\right)^2$,

$$\begin{aligned} E_I\left(q_I(X_{(n)})\right)^2 &= 1 - \left(1 - \frac{1}{n}\right)\frac{2(na+n)\tau^{a+1}}{\theta^{na+n}}\int_0^\theta t^{a(n-1)+n-2}dt \\ &+ \left(1 - \frac{1}{n}\right)^2 \frac{(na+n)\tau^{2(a+1)}}{\theta^{na+n}}\int_0^\theta t^{a(n-2)+n-3}dt \\ &= 1 - 2\left(\frac{\tau}{\theta}\right)^{a+1} + \frac{(n-1)^2}{n(n-2)}\left(\frac{\tau}{\theta}\right)^{2(a+1)}. \quad \square \end{aligned}$$

## A.4 Section 2.4, Beta Distribution Example, Eq. 15

**Lemma 8.** *Let $X_1, ..., X_n \sim B(\alpha+1, 1)$. Suppose that Model I holds. BB's estimator for*

$$P_{\eta_0}(X > \tau) = 1 - \frac{\tau^{\alpha+1}}{\theta^{\alpha+1}}\mathbf{1}\{0 < \tau < \theta\},$$

*is given by*

$$q_{II}(X_{(1)}, X_{(n)}) = \left(1 - \frac{1}{n}\right) - \left(1 - \frac{2}{n}\right)\frac{\tau^{\alpha+1} - X_{(1)}^{\alpha+1}}{X_{(n)}^{\alpha+1} - X_{(1)}^{\alpha+1}}, \quad X_{(1)} < \tau < X_{(n)}$$

*Moreover,*

$$E_I\left(q_{II}(X_{(1)}, X_{(n)})\right) = 1 + \frac{1}{n} - \left(\frac{\tau}{\theta}\right)^{a+1}.$$

*Proof.* The first assertion follows from Example 4. Note that

$$E_I\left(q_{II}(X_{(1)}, X_{(n)})\right) = \int_0^\theta \int_0^t q_{II}(y,t) f^I_{X_{(1)},X_{(n)}}(y,t)dydt,$$

while,

$$q_{II}(X_{(1)},X_{(n)}) = \left(1-\frac{1}{n}\right) - \left(1-\frac{2}{n}\right)\frac{\tau^{\alpha+1}-X_{(1)}^{\alpha+1}}{X_{(n)}^{\alpha+1}-X_{(1)}^{\alpha+1}}\,, \quad X_{(1)}<\tau<X_{(n)}$$

and

$$f^{I}_{X_{(1)},X_{(n)}}(y,t) = \frac{n(n-1)(yt)^{\alpha}(t^{\alpha+1}-y^{\alpha+1})^{n-2}(\alpha+1)^2}{\theta^{n\alpha+n}}\mathbf{1}\{0\le y\le t\le\theta\}.$$

Hence, one can express the expectation of BB's estimator w.r.t. Model I support as

$$\begin{aligned} E_I\left(q_{II}(X_{(1)},X_{(n)})\right) &= \left(1-\frac{1}{n}\right)-\left(1-\frac{2}{n}\right)\Bigg(\int_0^{\theta}\frac{n(n-1)(a+1)\tau^{a+1}t^{a}}{\theta^{na+n}}dt \\ & \int_0^{t} y^{a}\left(t^{a+1}-y^{a+1}\right)^{n-3}dy \\ &- \int_0^{\theta}\frac{n(n-1)t^{a}(a+1)^2}{\theta^{na+n}}dt\int_0^{t} y^{2a+1}\left(t^{a+1}-y^{a+1}\right)^{n-3}dy\Bigg). \end{aligned}$$

By changing the integration variable to $x$, such that $x=y^{a+1}$ then $dy=\frac{dx}{y^{a}(a+1)}$ we obtain

$$\begin{aligned} E_I\left(q_{II}(X_{(1)},X_{(n)})\right) &= \left(1-\frac{1}{n}\right)-\left(1-\frac{2}{n}\right)\Bigg(\int_0^{\theta}\frac{n(n-1)(a+1)\tau^{a+1}t^{a}}{\theta^{na+n}}dt \\ &= \int_0^{t^{a+1}}\left(t^{a+1}-x\right)^{n-3}dx \\ &- \int_0^{\theta}\frac{n(n-1)t^{a}(a+1)}{\theta^{na+n}}dt\int_0^{t^{a+1}} x\left(t^{a+1}-x\right)^{n-3}dx\Bigg). \end{aligned}$$

Note that the first integral is a simple integral, but the second involves integration by parts when $v=x$ and $u'=(t^{a+1}-x)$. Solving these integrals yields,

$$E_I\left(q_{II}(X_{(1)},X_{(n)})\right) = 1+\frac{1}{n}-\left(\frac{\tau}{\theta}\right)^{a+1}. \quad \square$$

## A.5 Section 2.4, Beta Distribution Example, Computation of $E_I\left(q_{II}(X_{(1)},X_{(n)})\right)^2$

**Lemma 9.** *Let $X_1,...,X_n \sim B(\alpha+1,1)$. Suppose that Model I holds. Hence the Second moment of BB's estimator is*

$$\begin{aligned} E_I\left(q_{II}^2(X_{(1)},X_{(n)})\right) &= \left(1-\frac{1}{n}\right)^2 - 2\left(\left(\frac{\tau}{\theta}\right)^{a+1}-\frac{2}{n}\right) \\ &+ \left(1-\frac{2}{n}\right)^2 \frac{1}{(n-3)(n-2)}\left(\left(\frac{\tau}{\theta}\right)^{2(a+1)} n(n-1)+\left(\frac{\tau}{\theta}\right)^{a+1} 2n+2\right). \end{aligned}$$

*Proof.* By definition, the second moment of BB's estimator w.r.t. Model I support can be obtained by

$$E_I\left(q_{II}^2(X_{(1)},X_{(n)})\right) = \int_0^\theta \int_0^t q_{II}^2(y,t) f^I_{X_{(1)},X_{(n)}}(y,t)\,dy\,dt,$$

using the functions from Lemma 3, and the same change of variables, i.e., $x = y^{a+1}$ where $dy = \frac{dx}{y^a(a+1)}$, one can write the following equation as a sum of three integrals

$$\begin{aligned} E_I\left(q_{II}^2(X_{(1)},X_{(n)})\right) &= \int_0^\theta \frac{n(n-1)t^a(a+1)\tau^{2(a+1)}}{\theta^{na+n}}dt \int_0^{t^{a+1}} \left(t^{a+1}-x\right)^{n-4} dx \\ &- \int_0^\theta \frac{n(n-1)t^a(a+1)2\tau^{a+1}}{\theta^{na+n}}dt \int_0^{t^{a+1}} x\left(t^{a+1}-x\right)^{n-4} dx \\ &+ \int_0^\theta \frac{n(n-1)t^a(a+1)}{\theta^{na+n}}dt \int_0^{t^{a+1}} x^2\left(t^{a+1}-x\right)^{n-4} dx, \end{aligned}$$

note that the first integral is a simple one, while the second and the third involve, as before, iterative use of integration by parts where $v = x$ and $v = x^2$ respectively and $u' = (t^{a+1}-x)$. Solving these integrals yields,

$$\begin{aligned} E_I\left(q_{II}^2(X_{(1)},X_{(n)})\right) &= \left(1-\frac{1}{n}\right)^2 - 2\left(\left(\frac{\tau}{\theta}\right)^{a+1}-\frac{2}{n}\right) \\ &+ \left(1-\frac{2}{n}\right)^2 \frac{1}{(n-3)(n-2)}\left(\left(\frac{\tau}{\theta}\right)^{2(a+1)} n(n-1)+\left(\frac{\tau}{\theta}\right)^{a+1} 2n+2\right). \quad \square \end{aligned}$$

# B Computations for Chapter 3

## B.1 Section 3.2, Exponential Distribution Example, Eq. 20

**Lemma 10.** *Let $X_1, ..., X_n \sim \exp(\lambda)$, where $\lambda = -\theta$. Then the MLE for $\lambda$ w.r.t. Model II support is given by*

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n}(X_i - X_{(1)})}.$$

*Moreover,*

$$f^I_{\hat{\lambda}}(w) = \frac{\theta^{n-1}}{(n-2)!}\frac{n^{n-1}}{w^n}\exp\{\theta n/w\}.$$

*Proof.* Let $\sum_{i=1}^{n}(X_i - X_{(1)})$ be $Y$, where $Y \sim \text{Erlang}(n-1, \theta)$ (Bar-Lev and Boukai, 2009). Define $W = n/Y$, hence

$$\begin{aligned} F_W(w) &= P(W \leq w) \\ &= P\left(\frac{n}{w} \leq Y\right) \\ &= 1 - F_Y\left(\frac{n}{w}\right), \end{aligned}$$

therefore,

$$\begin{aligned} f_W(w) &= \frac{\partial}{\partial w}\left(1 - F_Y\left(\frac{n}{w}\right)\right) \\ &= \frac{n}{w^2} f_Y\left(\frac{n}{w}\right) \\ &= \frac{\theta^{n-1}}{(n-2)!}\frac{n^{n-1}}{w^n}\exp\{\theta n/w\}. \quad \square \end{aligned}$$

## B.2 Section 3.2, Exponential Distribution Example, Eq. 19

**Lemma 11.** *Let $X_1, ..., X_n \sim \exp(\gamma, -\theta)$. Assume that the MLE for $\theta$ was derived under Model I, however Model II actually holds. Then*

$$-\frac{1}{\bar{X}_n} - \theta \xrightarrow[II]{a.s} \frac{\theta^2\gamma}{1-\theta\gamma}.$$

*Proof.* The maximum likelihood equation for the mean function $\mu_\gamma$ w.r.t. Model II support can be written as

$$\bar{X}_n = -\frac{1}{\theta} + \gamma,$$

recall that the MLE for the natural parameter $\theta$ w.r.t. Model I support is $-1/\bar{X}_n$, hence by rearranging the equation one can show that the maximum likelihood equation can be written as

$$\frac{\theta}{1-\theta\gamma} = -\frac{1}{\bar{X}_n}.$$

It follows form the SLLN (Ferguson, 1996, Theorem 4) that the sample mean converges a.s. to $\mu_\gamma = -\frac{1}{\theta} + \gamma$ . Therefore, using the continuous mapping theorem (van der Vaart, 2000, Theroem 2.3), the sequence $-\frac{1}{\bar{X}_n}$ converge a.s. to $\frac{\theta}{1-\theta\gamma}$, hence one can show that the sequence $-\frac{1}{\bar{X}_n} - \theta$ converge a.s. to

$$\frac{\theta}{1-\theta\gamma} - \theta = \frac{\theta^2\gamma}{1-\theta\gamma}. \quad \square$$

## B.3 Section 3.3, Erlang Distribution Example. Eq. 23

**Lemma 12.** *Let* $X_1, ..., X_n \sim \text{Erlang}(\gamma, 2, -\theta)$. *If the MLE for the natural parameter* $\theta$ *was derived under Model I support, however Model II holds, therefore*

$$-\frac{2}{\bar{X}_n} - \theta \xrightarrow[II]{a.s} \frac{-\theta^3\gamma^2}{2+\theta\gamma(\theta\gamma-2)}.$$

*Proof.* The maximum likelihood equation for the mean function $\mu_\gamma$ w.r.t. Model II support is given by

$$-\frac{2}{\theta} + \gamma - \frac{\gamma}{1-\theta\gamma} = \bar{X}_n.$$

Recall that $-2/\bar{X}_n$ is the MLE for the natural parameter $\theta$ w.r.t. Model I support, hence by rearranging the equation one can show that

$$-\frac{2}{\bar{X}_n} = \frac{2\theta(1-\theta\gamma)}{2+\theta\gamma(\theta\gamma-2)},$$

Clearly, by the SLLN (Ferguson, 1996, Theorem 4), the sample mean converge a.s. to the mean function $\mu_\gamma = -\frac{2}{\theta} + \gamma - \frac{\gamma}{1-\theta\gamma}$. Hence, using the continuous mapping theorem (van der Vaart, 2000, Theroem 2.3), one can show that the sequence $-\frac{2}{\bar{X}_n}$ converge a.s. to $\frac{2\theta(1-\theta\gamma)}{2+\theta\gamma(\theta\gamma-2)}$, therefore the sequence $-\frac{2}{\bar{X}_n} - \theta$ converge a.s. to

$$\frac{2\theta(1-\theta\gamma)}{2+\theta\gamma(\theta\gamma-2)} - \theta = \frac{-\theta^3\gamma^2}{2+\theta\gamma(\theta\gamma-2)}. \quad \square$$

## B.4 Section 3.3, Laplace Transform - Erlang Distribution Example

**Lemma 13.** *For the Erlang-2 distribution under Model II support,* $S = (\gamma, \infty)$, *i.e.,* $X_1, ..., X_n \sim \text{Erlang}(2, \theta, \gamma)$ *where the natural parameter domain is* $\Theta = (-\infty, 0)$, *the Laplace transform is*

$$\mathcal{L}(\theta, \gamma) = \frac{1}{\theta^2} e^{\theta\gamma} (1 - \gamma\theta).$$

*Proof.* By definition,

$$\mathcal{L}(\theta, \gamma) = \int_{\gamma}^{\infty} e^{\theta x} h(x) dx.$$

Note that $h(x) = x\mathbf{1}\{\gamma < x < \infty\}$, hence

$$\mathcal{L}(\theta,\gamma) = \int_{\gamma}^{\theta} xe^{\theta x}dx,$$

define $v = x$ and $u' = e^{\theta x}$ and using integration by parts, one can show that

$$\begin{aligned}\mathcal{L}(\theta,\gamma) &= \left(\frac{x}{\theta}e^{\theta x}\right|_{\gamma}^{\infty} - \frac{1}{\theta}\int_{\gamma}^{\infty} e^{\theta x}dx \\ &= -\frac{\gamma}{\theta}e^{\theta\gamma} + \frac{1}{\theta^2}e^{\theta\gamma} \\ &= \frac{1}{\theta^2}e^{\theta\gamma}\left(1-\gamma\theta\right). \quad \square\end{aligned}$$

# References

M Faris Muslim Al-Athari. Estimation of the mean of truncated exponential distribution. *Journal of Mathematics and Statistics*, 4(4):284, 2008.

S. K. Bar-Lev. Large sample properties of the MLE and MCLE for the natural parameter of a truncated exponential family. *Annals of the Institute of Statistical Mathematics*, 36(1):217–222, 1984.

S. K. Bar-lev and B. Boukai. Minimum variance unbiased estimation for families of distributions involving two truncation parameters. *Journal of Statistical Planning and Inference*, 12:379–384, 1985.

Shaul K Bar-Lev and Benzion Boukai. A characterization of the exponential distribution by means of coincidence of location and truncated densities. *Statistical Papers*, 50(2): 403–405, 2009.

T. M. Dubinin and S. B. Vardeman. Likelihood-Based Inference in Some Continuous Exponential Families with Unknown Threshold Parameters. *Journal of the American Statistical Association*, 98(463), 2003.

Thomas Ferguson. *A course in large sample theory*. Chapman & Hall, London New York, 1996.

R. F. Tate. Unbiased Estimation: Functions of Location and Scale Parameters. *The Annals of Mathematical Statistics*, 30(2):341–366, 1959.

Aad W. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 2000.

להנחה (השגויה) שישנה קטימה שמאלית כאשר בפועל היא איננה, אין השלכות הרות גורל על יעילות האמידה. תוצאה זו נובעת מקצב ההתכנסות המהיר של $X_{(1)}$ לגבול השמאלי – 0.

הראנו כי אמד הנראות המרבית תחת הנחת *מודל II* מתכנס ל-$\theta$. בנוסף, אמד זה הוא אסימפטוטית חסר הטיה ויעיל (השגיאה הריבועית משיגה את החסם התחתון של קרמר-ראו). יתרה מזאת, הראנו כי המקרה ההפוך בדוגמאות אלו, בדומה לפרק הקודם, הינו בעל השלכות חמורות. רוצה לומר, להניח בשוגג כי אין קטימה משמאל גורם לאמידה לא מדויקת ולא עקיבה וזאת משום שהאומד יתכנס almost surely לפונקציה שגויה.

# מודלים סטטיסטיים רציפים: עם או בלי פרמטר קטימה?

ולנטין ונצק

## תקציר

בנתוני אורך חיים שכיחה ההנחה כי מקור הנתונים הוא מהתפלגות רציפה הנתמכת על הקטע $\left(0,b\right)$, כך ש-$b\geq 0$. מהנחת הרציפות נובע שתומך ההתפלגות אינו מכיל atom points, במיוחד לא ב-0. לפיכך, אך טבעי שעם כלי מדידה מדויק נקבל שכל התצפיות תהינה חיוביות. הבחנה זו יכולה להניב השערה שהתומך האמיתי קטום משמאל והוא למעשה בעל הצורה $\left(\gamma,b\right)$, כך ש-$\gamma\geq 0$ כאשר $\gamma$ הוא פרמטר לא-ידוע. אי לכך, במקרה זה, אנו ניצבים בפני שני סוגי שגיאות אפשריות (של מודל שגוי). כדי לתאר זאת, נסמן ב*מודל I* את המודל שבו התומך הנכון הוא $\left(0,b\right)$, ואילו *מודל II* יציין את המקרה שבו התומך הנכון הוא $\left(\gamma,b\right)$. לפיכך, נציין כי שגיאת False Model I תארע כל אימת שבוצעה הסקה ע"ס *מודל I* כאשר המודל הנכון הוא *מודל II*. באופן סימטרי תוגדר שגיאת False Model II. השאלה המתעוררת במקרה דנן היא איזו משני סוגי הטעויות חמור יותר?

עבודה זו מחולקת לשני פרקים עיקריים. הפרק הראשון ידון במודלים סטטיסטיים רציפים כלליים. עבור פרק זה נבחן ונחקר התסריט הבא: הנחנו תחילה כי ישנה קטימה משמאל כאשר בפועל היא איננה, דהיינו, התומך האמיתי של המודל הוא $S=\left(0,\theta\right)$ כאשר $\theta$ הוא פרמטר המודל. הראנו כי שגיאה מסוג זה אינה גורמת לגרעון גדול בתהליך האמידה במונחים של דיוק ויעילות אסימפטוטיים. תוצאה זו נובעת מקצב ההתכנסות המהיר של האמד של בר-לב ובוקעי (Bar-lev & Boukai, 1985) לפונקציה הנאמדת. בנוסף לכך, הראנו כי במקרה ההפוך, כאשר אנו מניחים (בשוגג) כי אין קטימה כאשר בפועל התומך של המודל הוא $S=\left(\gamma,\theta\right)$, האמד של Tate (Tate, 1959) אינו מתכנס לפונקציה הנאמדת. לאמיתו של דבר, במקרה זה לאמד של Tate ישנה הטיה בגודל קבוע, ולכן עבור קטימה שמאלית גדולה, האמד של Tate יגרום לאמידה לא מדויקת ובלתי יעילה.

הפרק השני ידון באמדי נראות-מרבית עבור הפרמטר הטבעי $\theta$ בשני מקרים פרטיים של המשפחה האקספוננציאלית. הראנו כי בהתפלגות אקספוננציאלית ובהתפלגות ארלנג-2

# מודלים סטטיסטיים רציפים: עם או בלי פרמטר קטימה?

מאת: ולנטין ונצק
בהנחיית: ד"ר יאיר גולדברג

עבודת גמר מחקרית (תזה) המוגשת כמילוי חלק מהדרישות
לקבלת התואר "מוסמך האוניברסיטה"

אוניברסיטת חיפה
הפקולטה למדעי החברה
החוג לסטטיסטיקה

אוקטובר, 2013

# מודלים סטטיסטיים רציפים: עם או בלי פרמטר קטימה?

**ולנטין ונצק**

עבודת גמר מחקרית (תזה) המוגשת כמילוי חלק מהדרישות
לקבלת התואר "מוסמך האוניברסיטה"

אוניברסיטת חיפה
הפקולטה למדעי החברה
החוג לסטטיסטיקה

אוקטובר, 2013