2015-06-19

Maximum Lq-Likelihood Estimation for Gamma Distributions

Xing, Nana

UNIVERSITY OF CALGARY

Maximum Lq-Likelihood Estimation for Gamma Distributions

by

Nana Xing

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

JUNE, 2015

Abstract

MLqE is an extension of MLE which introduces a distortion parameter q to MLE to make the estimation more adaptive. The purpose of this thesis is to examine MLqE for specific distribution models. Particularly, for exponential and standard gamma distributions, we look at their asymptotics, finite sample performance in terms of efficiency and robustness, and the choice of the distortion parameter q. We investigate these aspects of MLqE, compared with MLE, in parameter estimation and tail probability estimation through both Monte Carlo simulation and a real data analysis. Our results show that, when exponential or standard gamma models are concerned, MLqE and MLE perform competitively for large sample sizes while MLqE outperforms MLE for small or moderate sample size in terms of reducing MSE. In addition, MLqE generally has better robustness properties than MLE with respect to outlying observations.

Acknowledgements

Many valuable assistance and encouragement helped me finalize this thesis in the past four years. The precious time spent at the university makes my life and career rich and colourful.

First and foremost I want to thank my supervisor Dr. Jingjing Wu. It has been an honour for me to be her graduate student. She teaches me, both consciously and un-consciously, how good statistical research could be done. I appreciate all her contributions of time, ideas and financial support to make my graduate study experience stimulating and productive. She is informative, considerate and always encourages me to attend and present talks at seminars and meetings such as the PIMS Young Researcher Conference and the Statistical Society of Canada Annual Meeting. These conference experiences helped me to improve presentation skills, enhance my understanding of the research, learn the frontier of the field and accumulate my confidence on what I have been doing. The joy and enthusiasm she has for her research was contagious and motivational for me. I am also thankful for the excellent example she has proved to be as a successful female statistician and professor. She not only has strict research attitude but also conveys a spirit of adventure and excitement in regard to research, presentation and teaching. Dr. Wu made a lot of valuable suggestions and inspired many ideas that led to improvement of this thesis. Without her guidance and persistent help this dissertation would not have been possible.

I would also like to give my great appreciation to other members of my committee, Dr. Alexander de Leon and Dr. Robert Lund, for serving as members of my committee, reading my thesis and providing valuable advice for my thesis. I thank them for their time, interest, and helpful comments.

Many thanks go to my graduate course instructors including Dr. Xuewen Lu and Dr. Burke Murray. I cannot say thank you enough for their tremendous advice and help in my studies. I learned much important and useful knowledge, and feel motivated and encouraged while attending their classes. Dr. Lu asked us to be more professional on statistical softwares which helped me to figure out simulation problems. Dr. Burke provided students a great deal of advice on giving presentations and defences. I thank all the professors who instructed, illuminated or cultivated me. The members of the Department of Mathematics and Statistics have contributed immensely to my personal and professional time at University of Calgary. The department has been a source of friendships as well as good advice and collaboration. I am really thankful to the staff and colleagues in the department, who are warmhearted and friendly, for their time and insightful advice.

Finally, I am grateful for the time I spent with friends in Canada, for our memorable trips and happy leisure activities here. And I also want to thank my family for all their love and encouragement.

Table of Contents

List of Tables

List of Figures and Illustrations

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
MLE	maximum likelihood estimation/estimator/estimate
MLqE	maximum Lq-likelihood estimation/estimator/estimate
CLT	Central Limit Theorem
p.d.f.	probability distribution function
c.d.f.	cumulative distribution function
i.i.d.	independent and identically distributed
MSE	mean squared error
IF	influence function
$Exp(\lambda)$	exponential distribution with parameter λ
$Gamma(\theta, 1)$	standard gamma distribution with shape parameter θ and scale parameter 1
se	standard error

Chapter 1

INTRODUCTION

In this chapter, we give some background to the estimation methods under consideration. In Section 1.1, some fundamental theory and results of maximum likelihood estimation (MLE) are reviewed. In Section 1.2, we introduce the nonextensive entropy theory, which is the basis of maximum Lq-likelihood estimation (MLqE), and its scope of application. Following that, MLqE is introduced in Section 1.3. Finally Section 1.4 aims to briefly illustrate the organization of the thesis.

1.1 Maximum Likelihood Estimation (MLE)

MLE was recommended, analyzed and vastly popularized by R. A. Fisher between 1912 and 1922, although it had been used earlier by Gauss, Laplace, Thiele and Edgeworth.

Let Y be a random variable with distribution of known type up to some unknown finite-dimensional parameter θ. Denote the probability distribution function (p.d.f.) of Y by $f(\cdot;\theta)$. Suppose we have independent and identically distributed (i.i.d.) sample values $y_1, y_2, \cdots, y_n$ from this distribution. Define the likelihood of θ, given data $y_1, \cdots, y_n$, as

$$L(\theta; y_1, \cdots, y_n) = f_{Y_1,\cdots,Y_n}(y_1, \cdots, y_n; \theta) = \prod_{i=1}^{n} f(y_i; \theta)$$

The value $\hat{\theta}$ that maximizes the likelihood $L(\theta; y_1, \cdots, y_n)$ is defined as the MLE. The MLE is used to estimate θ based on data values. Often, it is found that $\frac{\partial L}{\partial \hat{\theta}} = 0$ and $\frac{\partial^2 L}{\partial \hat{\theta}^2} < 0$. These relationship may help find maxima, but one also needs to check boundary values of θ. In practice it is often more convenient to work with the logarithm of the likelihood function, called the log-likelihood function: $\ln L(\theta; y_1, \cdots, y_n) = \sum_{i=1}^{n} \ln f(y_i; \theta)$. One can also use optimization methods to find $\hat{\theta} = \underset{\hat{\theta}\in\Theta}{\arg\max} L(\theta; y_1, \cdots, y_n)$, with Θ denoting the parameter space of θ. The MLE will be the same regardless of whether we maximize the likelihood function or the log-likelihood function, since log is a monotonically increasing transformation.

For many models, the MLE can be found as an explicit function of the observed data $y_1, \cdots, y_n$. For many other models, however, no closed-form solution to the maximization problem is known or available, and the MLE has to be found numerically using optimization methods. For some problems, there may be multiple estimates that maximize the likelihood. For other problems, no MLE exists (meaning that the log-likelihood function increases without attaining its supremum).

MLE has some desirable properties including that the distribution of $\hat{\theta}$ is known and is narrowly distributed around the true value of θ. Increasing the sample size n will improve the estimate and guarantee its efficiency. MLE has no optimum properties for finite samples in the sense that (when evaluated on finite samples) other estimators could have greater concentration around the true parameter value, however, MLE possesses a number of attractive limiting properties. As the sample size increases to infinity, sequences of MLEs have following properties:

(1) Consistency: sequence of MLEs(in univariate) converges in probability to the value being estimated.

(2) Asymptotic normality: the distribution of MLE(in multivariate) tends to the Gaussian distribution with mean θ and covariance matrix the inverse of the Fisher information matrix.

(3) Efficiency: MLE achieves the Cramér-Rao lower bound when the sample size tends to infinity. This means that no consistent estimator has smaller asymptotic variance than the MLE (or other estimators attaining this bound).

(4) Second-order efficiency after correction for bias: with θ of dimension p, a second-order "bias-corrected" MLE of θ can then be obtained as $\tilde{\theta} = \hat{\theta} - \hat{K}^{-1} * \hat{A} * vec(\hat{K}^{-1})$, where $\hat{K} = K|_{\theta=\hat{\theta}}$, $K = [-k_{ij}]_{i,j=1,\cdots,p}$ with k_{ij} the element of the inverse of the (expected)

information matrix and $\hat{A} = A|_{\theta=\hat{\theta}}$ with A defined as follows. With $l(\theta)$ the (total) log-likelihood based on a sample of n observations, $A = [A^{(1)} \mid A^{(2)} \mid \cdots \mid A^{(p)}]$, where $A^{(m)} = [k_{ij}^{(m)} - (k_{ijm}/2)]_{i,j=1,\cdots,p}$ with $k_{ij}^{(m)} = \partial k_{ij}/\partial\theta_m$ and $k_{ijm} = E(\partial^3 l/\partial\theta_i\partial\theta_j\partial\theta_m)$. Using the bias-corrected MLE $\tilde{\theta}$ of the parameter is extremely effective.

MLE is commonly used in statistics. Large sample theory guarantees that MLE is asymptotically efficient, which means that when the sample size is large, MLE is at least as accurate as any other estimator. MLE is used for a wide range of statistical models, including linear models and generalized linear models, exploratory and confirmatory factor analysis, structural equation modelling, and so on.

1.2 Nonextensive Entropy Theory

In the late 1940s, Claude Shannon established the information theory which became one of the major scientific advances in the last century. The Shannon's information theory has been successfully applied in a variety of scientific areas including statistics. The key point of Shannon's information theory is the so called Shannon's entropy defined as $H(X) = -E[\log p(X)]$. Here $p(x)$ represents the p.d.f. of random variable X.

After the Shannon's entropy was introduced, the relationship between $\log p(X)$ and $H(X)$ was widely studied. A statistical model that was expected to minimize the Shannon's entropy was brought by Akaike (1973) in which it was stated that the minimization of $-\sum_{i=1}^{n} \log p(X_i)$ (empirical version of Shannon's entropy) is equivalent to the maximization of the log-likelihood function. Then model comparison based on the minimum description length criterion was established by Barron, Rissanen and Yu (1998). Later, Shannon's entropy became widely used, which brought newly proposed measures of information such as Rényi entropies. Rényi entropies use a more general definition of mean and keep additivity of independent information; see Aczél and Daróczy (1975) and Rényi (1961).

Havrda and Charvát (1967) proposed nonextensive entropies, sometimes referred to as q-

order entropy. The q-order entropy is an important extension of Shannon's entropy where the logarithm is replaced by the more general function $L_q(u) = (u^{1-q} - 1)/(1-q)$ for $q > 0$. Note that $L_q(u) \to \log(u)$ when $q \to 1$, recovering the usual Shannon's entropy. Recently, q-order entropies have been applied in different scientific areas. In thermodynamics, the q-entropy functional is usually minimized subject to some properly chosen constraints, according to the formalism proposed by Jaynes (1957a, 1957b). Tsallis (1988) and Tsallis, Mendes and Plastino (1998) successfully exploited q-order entropies in physics. In statistics, Altun and Smola (2006) concluded that the classical maximum entropy estimation and MLE are convex duals of each other.

1.3 Maximum Lq-likelihood Estimation (MLqE)

In this section, we review an alternative parametric estimation to MLE, i.e. the MLqE. MLqE was first introduced by Ferrari and Yang (2010) and it is based on the nonextensive q-order entropy function introduced in the last section.

Let $X_1, \cdots, X_n$ be an i.i.d. sample from p.d.f. $f(\cdot;\theta)$ with some $\theta \in \Theta$. The MLqE of θ is defined as

$$\tilde{\theta}_n = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} L_q[f(X_i;\theta)], \quad q > 0,$$

where

$$L_q(u) = \begin{cases} \log u, & \text{if } q = 1, \\ (u^{1-q} - 1)/(1-q), & \text{otherwise.} \end{cases}$$

From the definition of L_q and L'Hôspital's rule we can see that if $q \to 1$, then $L_q(u) \to \log u$, i.e. $L_q(u)$ is a continuous function of q for any fixed $u > 0$. Therefore, when q is close to 1, the value of $\tilde{\theta}_n$ will be close to the MLE of θ. In this sense, MLqE extends the classic MLE method, resulting in a general inferential procedure that inherits most of the desirable features of traditional MLE and at the same time can improve MLE via variance reduction.

For the multivariate parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)$, under some regularity conditions, its

MLqE is the solution to the following p equations system

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} L_q[f(X_i;\boldsymbol{\theta})] = 0, \quad j = 1, 2, \cdots, p.$$

In this thesis, we only consider the univariate case for a natural parameter or tail probability.

Ferrari and Yang (2010) provided theoretical insights concerning the statistical usage of the generalized entropy function. In particular, they highlighted the role of the distortion parameter q. When the sample size is large and q tends to 1, Ferrari and Yang (2010) established a necessary and sufficient condition to ensure asymptotic normality and efficiency of MLqE. MLE is asymptotically efficient, however, for a small or moderate sample size, when q is properly chosen MLqE can offer a dramatically reduced mean squared error (MSE) at the expense of a slightly increased bias when compared to MLE. In the framework of Ferrari and Yang (2010), MLqE has been shown through simulation to be very useful when estimating the exponential distribution parameter and its small tail probability.

For finite sample performance of MLqE, not only the size of $q-1$ but also its sign (i.e., the direction of distortion) is important. It turns out that for different families or different parametric functions of the same family, the beneficial direction of distortion can be different. In addition, for some parameters, MLqE does not produce any improvement. Ferrari and Yang (2010) found that an asymptotic variance expression of the MLqE is very helpful to decide the direction of distortion for applications.

To our knowledge, there are only a few papers on MLqE so far. Qin and Priebe (2013) introduced a MLqE for mixture models using their proposed expectation-maximization (EM) algorithm, namely the EM algorithm with Lq-likelihood (EM-Lq). Ferrari and Paterlini (2009) applied MLqE to estimate quantiles of the Generalized Extreme Value (GEV) and the Generalized Pareto (GP) distributions in finance. In Huang, Lin and Ren (2013), the hypothesis testing problem for the shape parameter of the GEV distribution is investigated using the Lq-likelihood ratio statistic, a generalized form of the classical likelihood ratio statistic.

1.4 Organization of Thesis

In this thesis, we attempt to examine the MLqEs of parameter and tail probability for exponential distribution $Exp(\lambda) = Gamma(1, 1/\lambda)$ and standard gamma distribution $Gamma(\theta, 1)$, both of which are important special cases of the gamma distribution $Gamma(\theta, 1/\lambda)$. The standard gamma distribution is a gamma distribution with scale parameter being 1. We examine their asymptotics, finite sample performance in terms of efficiency and robustness, and the choice of distortion parameter q in terms of direction and value.

This thesis is organized as follows. In Chapter 2, we review the asymptotic normality of MLqE for exponential families. For both exponential and standard gamma distributions, we present the MLqEs of parameter and tail probabilities and derive their asymptotic variances. Then we discuss a method of choosing distortion parameter q based on the MSE. In Chapter 3, we implement Monte Carlo simulation studies to examine, for both exponential and standard gamma distributions, the MLqE's finite sample performance in terms of efficiency and robustness, the accuracy of confidence intervals based on MLqE, and the choice of the distortion parameter q. In Chapter 4, we demonstrate how to implement the MLqE through a real data analysis. Final concluding remarks are presented in Chapter 5.

Chapter 2

MLqE FOR GAMMA DISTRIBUTIONS

In this chapter, we present some theoretical results on MLqE for both exponential and standard gamma distributions. Particularly, in Section 2.1 we derive the MLqEs of model parameters. Section 2.2 reviews some asymptotic results of Ferrari and Yang (2010) on the MLqE for exponential families. In Section 2.3, we derive the explicit expression of the asymptotic variances of the MLqEs. Section 2.4 is devoted to the plug-in MLqEs of tail probabilities. Finally in Section 2.5, we discuss a method of choosing the distortion parameter q based on MSE.

2.1 MLqEs of the Parameters in Two Gamma Distributions

2.1.1 MLqE of the parameter in exponential distribution

Let $X_1, \cdots, X_n$ be an i.i.d. sample from the p.d.f. $f(\cdot;\boldsymbol{\theta})$ with some $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. The ML$q$E of $\boldsymbol{\theta}$ is defined as

$$\tilde{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\Theta}{\arg\max} \sum_{i=1}^{n} L_q[f(X_i;\boldsymbol{\theta})], \quad q > 0, \tag{2.1}$$

where

$$L_q(u) = \begin{cases} \log u, & \text{if } q = 1, \\ (u^{1-q} - 1)/(1-q), & \text{otherwise.} \end{cases} \tag{2.2}$$

Define

$$\begin{aligned} U(x;\boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}} \log\{f(x;\boldsymbol{\theta})\}, \\ U^*(x;\boldsymbol{\theta},q) &= U(x;\boldsymbol{\theta}) f^{1-q}(x;\boldsymbol{\theta}). \end{aligned} \tag{2.3}$$

Then in general, the estimating equation for the MLqE $\tilde{\boldsymbol{\theta}}_n$ solves

$$\sum_{i=1}^{n} U^*(X_i;\boldsymbol{\theta},q) = 0. \tag{2.4}$$

Equation (2.4) offers a natural interpretation of the MLqE as a solution to a weighted likelihood. When $q \neq 1$, (2.4) provides a relative-to-the-model reweighing. Observations that disagree with the model receive low or high weight depending on $q < 1$ or $q > 1$. In the case of $q = 1$, all the observations receive the same weight.

Ferrari and Yang (2010) has given the MLqE of the parameter in the exponential distribution and investigated its asymptotic properties. Consider an i.i.d. sample of size n from exponential distribution with p.d.f. $f(x; \lambda) = \lambda \exp(-\lambda x)$ for $x > 0$ and some fixed $\lambda > 0$. For this model, the Lq-likelihood equation (2.4) is

$$\sum_{i=1}^{n} e^{-(\lambda X_i - \log \lambda)(1-q)}(-X_i + \frac{1}{\lambda}) = 0. \tag{2.5}$$

With $q = 1$, the usual MLE of λ is $\hat{\lambda} = (\sum_{i=1}^{n} X_i/n)^{-1} = \bar{X}^{-1}$. Generally, for any $q > 0$, the MLqE $\tilde{\lambda}$ is the solution to (2.5) or equivalently the solution to

$$\tilde{\lambda} = \left(\frac{\sum_{i=1}^{n} X_i w_i(X_i, \tilde{\lambda}, q)}{\sum_{i=1}^{n} w_i(X_i, \tilde{\lambda}, q)} \right)^{-1}, \tag{2.6}$$

where $w_i(X_i, \tilde{\lambda}, q) = e^{-(\tilde{\lambda} X_i - \log \tilde{\lambda})(1-q)}$. As a result, the MLE and plug-in MLqE of the upper tail probability $\alpha(x; \lambda) = e^{-\lambda x}$ are respectively $\alpha(x; \hat{\lambda})$ and $\alpha(x; \tilde{\lambda})$.

2.1.2 MLqE of the parameter in standard gamma distribution

Here we consider the problem of estimating the shape parameter of a gamma distribution when the scale parameter is known. Since any gamma distribution can be expressed in terms of the standard gamma distribution after transformation, without generality, we assume the scale parameter is 1; i.e. we consider estimating the shape parameter of standard gamma distribution.

Consider an i.i.d. sample of size n from the standard gamma distribution $Gamma(\theta, 1)$ with density $f(x; \theta) = (x^{\theta-1} e^{-x})/\Gamma(\theta)$ for some $\theta > 0$. The Lq-likelihood equation is then

$$\sum_{i=1}^{n} e^{[(\theta-1)\log X_i - X_i](1-q)} \left[\Gamma(\theta) \log X_i - \Gamma'(\theta)\right] = 0. \tag{2.7}$$

The MLqE $\tilde{\theta}$ of θ is the solution to (2.7). With $q = 1$, the MLE $\hat{\theta}$ is the root of equation

$$\psi(\theta) = \left(\sum_{i=1}^{n} \log X_i\right) / n, \tag{2.8}$$

where

$$\psi(\theta) = \Gamma'(\theta)/\Gamma(\theta). \tag{2.9}$$

As a result, the MLE and plug-in MLqE of upper tail probability $\alpha(x;\theta) = 1 - \int_0^x \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}dy$ with $x > 0$ are $\alpha(x;\hat{\theta})$ and $\alpha(x;\tilde{\theta})$ respectively.

2.2 Asymptotics of the MLqE for Exponential Families

Consider density functions of the exponential family

$$f(x;\boldsymbol{\theta}) = \exp[\boldsymbol{\theta}^T \boldsymbol{b}(x) - A(\boldsymbol{\theta})], \tag{2.10}$$

where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is a real valued natural parameter vector, $\boldsymbol{b}(x)$ is the vector of functions with elements $b_j(x)$, $j = 1, \cdots, p$, and $A(\boldsymbol{\theta}) = \log \int e^{\boldsymbol{\theta}^T \boldsymbol{b}(x)} dx$ is the cumulant generating function (or log normalizer). The true parameter will be denoted by $\boldsymbol{\theta}_0$.

This exponential family include the exponential distribution and standard gamma distribution as special cases. For $Exp(\lambda)$, $\boldsymbol{\theta} = \lambda$, $\boldsymbol{b}(x) = -x$ and $A(\boldsymbol{\theta}) = -\log\lambda$. For $Gamma(\theta, 1)$, $\boldsymbol{\theta} = (1, \theta)^T$, $\boldsymbol{b}(x) = (-x - \log x, \log x)^T$ and $A(\boldsymbol{\theta}) = \log \Gamma(\theta)$. Ferrari and Yang (2010) have studied this exponential family and established the consistency and asymptotic normality of the MLqE. Their results are given as follows.

Consider $\boldsymbol{\theta}_n^*$, the value such that

$$E_{\boldsymbol{\theta}_0} U^*(X; \boldsymbol{\theta}_n^*, q_n) = 0. \tag{2.11}$$

Here we use q_n instead of q, since we want $q_n \to 1$ as $n \to \infty$, so that the MlqE is asymptotically equivalent to MLE. It can be easily shown that $\boldsymbol{\theta}_n^* = \boldsymbol{\theta}_0 / q_n$. Since the actual target of $\tilde{\boldsymbol{\theta}}_n$ is $\boldsymbol{\theta}_n^*$, to retrieve asymptotic unbiasedness of $\tilde{\boldsymbol{\theta}}_n$, q_n must converge to 1. Ferrari and Yang (2010) called $\boldsymbol{\theta}_n^*$ the surrogate parameter of $\boldsymbol{\theta}_0$. The following conditions are imposed:

A.1 $q_n > 0$ is a sequence such that $q_n \to 1$ as $n \to \infty$.

A.2 The parameter space Θ is compact and $\boldsymbol{\theta}_0$ is an interior point of Θ.

Theorem 2.1. *Under assumptions A.1 and A.2, with probability going to 1, the Lq -likelihood equation yields a unique solution $\tilde{\boldsymbol{\theta}}_n$ that is the maximizer of the Lq-likelihood function in Θ. Furthermore, we have $\tilde{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$.*

When Θ is compact, the MLqE always exists under their conditions, although it is not necessarily unique with probability one.

Theorem 2.2. *If assumptions A.1 and A.2 hold, then we have*

$$\sqrt{n} V_n^{-1/2} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^*) \xrightarrow{\mathcal{D}} N_p(0, \boldsymbol{I}_p) \quad \text{as } n \to \infty, \tag{2.12}$$

where $\boldsymbol{I}_p$ is the $(p \times p)$ identity matrix, and

$$V_n = J_n^{-1} K_n J_n^{-1} \tag{2.13}$$

$$K_n = E_{\boldsymbol{\theta}_0} [U^*(X; \boldsymbol{\theta}_n^*, q_n)]^T [U^*(X; \boldsymbol{\theta}_n^*, q_n)], \tag{2.14}$$

$$J_n = E_{\boldsymbol{\theta}_0} [\nabla_{\boldsymbol{\theta}_n^*} U^*(X; \boldsymbol{\theta}_n^*, q_n)]. \tag{2.15}$$

Here J_n is symmetric. A necessary and sufficient condition for asymptotic normality of MLqE around $\boldsymbol{\theta}_0$ is $\sqrt{n}(q_n - 1) \to 0$.

Let $m(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} A(\boldsymbol{\theta})$ and $D(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}^2 A(\boldsymbol{\theta})$. Note that K_n and J_n can be expressed as

$$K_n = c_{2,n} \left(D(\boldsymbol{\theta}_{2,n}) + [m(\boldsymbol{\theta}_{2,n}) - m(\boldsymbol{\theta}_n^*)][m(\boldsymbol{\theta}_{2,n}) - m(\boldsymbol{\theta}_n^*)]^T \right) \tag{2.16}$$

and

$$J_n = c_{1,n}(1-q_n) D(\boldsymbol{\theta}_{1,n}) - c_{1,n} D(\boldsymbol{\theta}_n^*) + c_{1,n}(1-q_n)[m(\boldsymbol{\theta}_{1,n}) - m(\boldsymbol{\theta}_n^*)][m(\boldsymbol{\theta}_{1,n}) - m(\boldsymbol{\theta}_n^*)]^T, \tag{2.17}$$

where $c_{k,n} = \exp\{A(\boldsymbol{\theta}_{k,n}) - A(\boldsymbol{\theta}_0)\}$ and $\boldsymbol{\theta}_{k,n} = k\boldsymbol{\theta}_0(1/q_n - 1) + \boldsymbol{\theta}_0$. When $q_n \to 1$, it is seen that $V_n \to -D(\boldsymbol{\theta}_0)$, the asymptotic variance of the MLE. When $\Theta \subseteq \mathbb{R}^1$ we use the notation σ_n^2 for the asymptotic variance in place of V_n. Note that the existence of moments are ensured by the functional form of the exponential families (Lehmann and Casella, 1998).

When q is fixed, the MLqE is a regular M-estimator (Huber, 1981), which converges in probability to $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0/q$. With the explicit expression of $\boldsymbol{\theta}^*$, one may consider correcting the bias of MLqE by using the estimator $q_n\tilde{\boldsymbol{\theta}}_n$.

2.3 Asymptotic Variances of the MLqEs for Two Gamma Distributions

2.3.1 Asymptotic variance of the MLqE for the exponential distribution

Ferrari and Yang (2010) discussed the asymptotic distribution of the MLqE for exponential distribution. For $Exp(\lambda_0)$, the surrogate parameter is $\theta_n^* = \lambda_0/q_n$ and a lengthy but straightforward calculation shows that the variance of the MLqE of λ_0 is

$$\sigma_n^2 = \left(\frac{\lambda_0}{q_n}\right)^2 \left[\frac{q_n^2 - 2q_n + 2}{q_n^3(2-q_n)^3}\right] \longrightarrow \lambda_0^2 \tag{2.18}$$

as $n \to \infty$. By Theorem 2.2, $n^{1/2}\sigma_n^{-1}(\tilde{\lambda}_n - \lambda_0/q_n)$ converges weakly to a standard normal distribution as $n \to \infty$. Note that λ_0^2 is the asymptotic variance of the MLE, i.e. the inverse of the Fisher information. Clearly, the asymptotic calculation does not produce any advantage of MLqE in terms of reducing the limiting variance. However, Ferrari and Yang (2010) found that $\sigma_n^2 < \lambda_0^2$ for some choices of q_n for some finite sample sizes.

2.3.2 Asymptotic variance of the MLqE for the standard gamma distribution

Here we will derive the explicit form of the asymptotic variance of the MLqE of the shape parameter in the standard gamma distribution given in Section 2.1.2.

Consider the standard gamma distribution $Gamma(\theta_0, 1)$. Again the surrogate parameter for this model is $\theta_n^* = \theta_0/q_n$. Written in the form of (2.10), $Gamma(\theta, 1)$ has $\boldsymbol{\theta} = (1, \theta)^T$, $\boldsymbol{b}(x) = (-x - \log x, \log x)^T$ and $A(\boldsymbol{\theta}) = \log \Gamma(\theta)$. While $\boldsymbol{\theta}$ is two dimensional, it is essentially the one dimensional parameter θ. With ψ given in (2.9), we have

$$m(\theta) = \nabla_\theta A(\theta) = \psi(\theta),$$
$$D(\theta) = \nabla^2_\theta A(\theta) = \psi'(\theta).$$

Since $\theta_{k,n} = [k(1/q_n - 1) + 1]\theta_0$, we have

$$c_{1,n} = e^{A(\theta_{1,n})-A(\theta_0)} = \tfrac{\Gamma(\theta_0/q_n)}{\Gamma(\theta_0)},$$
$$c_{2,n} = e^{A(\theta_{2,n})-A(\theta_0)} = \tfrac{\Gamma((2/q_n-1)\theta_0)}{\Gamma(\theta_0)},$$

and then by (2.16) and (2.17)

$$\begin{aligned} K_n &= c_{2,n}\left(D(\theta_{2,n}) + [m(\theta_{2,n}) - m(\theta_n^*)]^2\right) \\ &= \frac{\Gamma((2/q_n-1)\theta_0)}{\Gamma(\theta_0)}\left\{\psi'((2/q_n-1)\theta_0) + [\psi((2/q_n-1)\theta_0) - \psi(\theta_0/q_n)]^2\right\}, \end{aligned}$$

$$\begin{aligned} J_n &= c_{1,n}(1-q_n)D(\theta_{1,n}) - c_{1,n}D(\theta_n^*) + c_{1,n}(1-q_n)[m(\theta_{1,n}) - m(\theta_n^*)]^2 \\ &= -q_n\frac{\Gamma(\theta_0/q_n)}{\Gamma(\theta_0)}\psi'(\theta_0/q_n). \end{aligned}$$

Finally by (2.13), the variance of the MLqE of θ_0 is

$$\begin{aligned} \sigma_n^2 &= K_n/J_n^2 \\ &= \frac{\frac{\Gamma((2/q_n-1)\theta_0)}{\Gamma(\theta_0)}\left\{\psi'((2/q_n-1)\theta_0) + [\psi((2/q_n-1)\theta_0) - \psi(\theta_0/q_n)]^2\right\}}{\left[\frac{q_n\Gamma(\theta_0/q_n)}{\Gamma(\theta_0)}\psi'(\theta_0/q_n)\right]^2} \qquad (2.19) \\ &\longrightarrow [\psi'(\theta_0)]^{-1} = \{[\log\Gamma(\theta_0)]''\}^{-1} \end{aligned}$$

as $n \to \infty$. By Theorem 2.2, we can conclude that $n^{1/2}\sigma_n^{-1}(\tilde{\theta}_n - \theta_0/q_n)$ converges weakly to a standard normal distribution as $n \to \infty$. Note that $\{[\log\Gamma(\theta_0)]''\}^{-1}$ is the asymptotic variance of the MLE.

2.4 Estimation of tail probabilities

Let $\alpha(x;\theta) = P_\theta(X \le x)$ or $\alpha(x;\theta) = 1 - P_\theta(X \le x)$, depending on whether we are considering the lower tail or the upper tail of the distribution. Without loss of generality, we only consider the upper tail of the distribution. Assume $\alpha(x;\theta) > 0$ for all x. Of course

$\alpha(x;\theta) \to 0$ as $x \to \infty$. When x is fixed, under some conditions, the delta method shows that an asymptotically normally distributed and efficient estimator of θ makes the plug-in estimator of $\alpha(x;\theta)$ also asymptotically normal and efficient. However, in most applications a large sample size is demanded in order for this asymptotic behavior to be accurate for a small tail probability. As a consequence, the setup with x fixed but $n \to \infty$ presents an overly optimistic view, as it ignores the possible difficulty due to the smallness of the tail probability in relation to the sample size n. Instead, allowing x to increase in n (so that the tail probability to be estimated becomes smaller as the sample size increases) more realistically addresses the problem.

2.4.1 Asymptotic normality of the MLqE of tail probabilities

We are interested in estimating $\alpha(x_n;\theta_0)$, where $x_n \to \infty$ as $n \to \infty$. For $\theta^* \in \Theta$ and $\delta > 0$, Ferrari and Yang (2010) defined

$$\begin{aligned} \beta(x;\theta^*;\delta) &= \sup_{\theta \in \Theta \cap [\theta^* - \delta/\sqrt{n}, \theta^* + \delta/\sqrt{n}]} \left| \tfrac{\alpha''(x;\theta)}{\alpha''(x;\theta^*)} \right|, \\ \gamma(x;\theta) &= \alpha''(x;\theta)/\alpha'(x;\theta), \end{aligned} \tag{2.20}$$

where the derivatives are with respect to θ, and gave the following results.

Theorem 2.3. *Let θ_n^* be the solution to (2.11) such that $\theta_n^* \to \theta_0$ as $n \to \infty$. Under assumptions A.1 and A.2, if*

$$n^{-1/2} \left| \gamma(x_n;\theta_n^*) \right| \beta(x_n;\theta_n^*;\delta) \to 0 \quad for\ each \quad \delta > 0, \tag{2.21}$$

then

$$\sqrt{n} \frac{\alpha(x_n;\tilde{\theta}_n) - \alpha(x_n;\theta_n^*)}{\sigma_n \alpha'(x_n;\theta_n^*)} \xrightarrow{\mathcal{D}} N(0,1),$$

where

$$\sigma_n = \frac{-\{E_{\theta_0}[U^*(X;\theta_n^*,q_n)]^2\}^{1/2}}{E_{\theta_0}[\nabla_{\theta_n^*} U^*(X;\theta_n^*,q_n)]}. \tag{2.22}$$

For the main requirement (2.21) of the theorem on the order of the sequence x_n, it is easiest to verify this on a case by case basis. For the exponential distribution, it can be easily shown that if $x_n n^{-1/2} \to 0$ as $n \to \infty$ then (2.21) is satisfied. However, this condition is not easy to check for the standard gamma distribution. Note that for $Gamma(\theta, 1)$, the upper tail probability is $\alpha(x;\theta) = 1 - \int_0^x \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}dy$, and then by the definitions in (2.20)

$$\beta(x_n;\theta_n^*;\delta) = \sup_{\theta \in [\theta_n^* - \frac{\delta}{\sqrt{n}}, \theta_n^* + \frac{\delta}{\sqrt{n}}]} \left| \frac{\int_0^{x_n} \frac{\partial^2}{\partial \theta^2} \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)} dy}{\int_0^{x_n} \frac{\partial^2}{\partial {\theta_n^*}^2} \frac{y^{\theta_n^*-1}e^{-y}}{\Gamma(\theta_n^*)} dy} \right|,$$

$$\gamma(x_n;\theta_n^*) = \sup_{\theta \in [\theta_n^* - \frac{\delta}{\sqrt{n}}, \theta_n^* + \frac{\delta}{\sqrt{n}}]} \frac{\int_0^{x_n} \frac{\partial^2}{\partial \theta^2} \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)} dy}{\int_0^{x_n} \frac{\partial}{\partial \theta} \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)} dy}.$$

It is difficult to give the explicit convergence order of x_n such that (2.21) holds, for which we will rely on our simulation studies in Chapter 3. When $x_n \to \infty$ too fast so as to violate the condition, the asymptotic normality is not guaranteed, which indicates the extreme difficulty in estimating a tiny tail probability.

2.4.2 Relative efficiency between MLE and MLqE

Theorem 2.2 shows that when $(q_n - 1)\sqrt{n} \to 0$, the MLqE is asymptotically as efficient as the MLE. This subsection looks at the efficiency of the MLqE of tail probability when $x_n \to \infty$.

Consider w_n and v_n, two estimators of a parametric function $g_n(\theta)$ such that both $\sqrt{n}(w_n - a_n)/\sigma_n$ and $\sqrt{n}(v_n - b_n)/\tau_n$ converges weakly to a standard normal distribution as $n \to \infty$ for some deterministic sequences $a_n, b_n, \sigma_n > 0$ and $\tau_n > 0$. Let

$$\Lambda(w_n, v_n) = \frac{(b_n - g_n(\theta))^2 + \tau_n^2/n}{(a_n - g_n(\theta))^2 + \sigma_n^2/n}. \tag{2.23}$$

Ferrari and Yang (2010) defined $\lim_{n\to\infty} \Lambda(w_n, v_n)$ as the bias adjusted asymptotic relative efficiency of w_n with respect to v_n, provided that the limit exists. It can be easily verified that the definition does not depend on the specific choice of a_n, b_n, σ_n and τ_n among equivalent expressions. Following directly from Theorem 2.3, Ferrari and Yang (2010) gave the following result on the relative efficiency of the MLqE with respect to the MLE of the tail probability;

i.e., when q_n is chosen sufficiently close to 1, asymptotically speaking, the MLqE is as efficient as the MLE.

Corollary 2.1. *Under the conditions of Theorem 2.3, when* q_n *is chosen such that*

$$n^{1/2}\alpha(x_n;\theta_n^*)\alpha(x_n;\theta_0)^{-1} \longrightarrow 1 \quad \text{and} \quad \alpha'(x_n;\theta_n^*)\alpha'(x_n;\theta_0)^{-1} \longrightarrow 1, \tag{2.24}$$

then $\Lambda(\alpha(x_n;\hat{\theta}_n),\alpha(x_n;\tilde{\theta}_n)) = 1$.

For the standard gamma distribution, $\alpha(x_n;\theta) = 1 - \int_0^{x_n} \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}dy$ and $\alpha'(x_n;\theta) = -\int_0^{x_n} \frac{\partial}{\partial\theta}\frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}dy$. For sequences $x_n \to \infty$ and $q_n \to 1$ such that (2.21) holds, we have $\sqrt{n}\frac{\alpha(x_n;\tilde{\theta}_n)-\alpha(x_n;\theta_0/q_n)}{\sigma_n\alpha'(x_n;\theta_0/q_n)} \xrightarrow{\mathcal{D}} N(0,1)$ with σ_n given in (2.22). For a one dimensional parameter, $\sigma_n = \sqrt{V_n}$ with V_n defined in (2.13). Particularly for $Gamma(\theta,1)$, σ_n^2 is calculated and given in (2.19). When $q_n = 1$ for all n, the usual plug-in estimator based on the MLE is recovered. With the asymptotic expressions given above,

$$\Lambda(\alpha(x;\hat{\theta}_n),\alpha(x;\tilde{\theta}_n)) = n[\log\Gamma(\theta_0)]''\left[\frac{\alpha(x_n;\theta_0/q_n)-\alpha(x_n;\theta_0)}{\alpha'(x_n;\theta_0)}\right]^2 + \left[\frac{\alpha'(x_n;\theta_0/q_n)}{\alpha'(x_n;\theta_0)}\right]^2.$$

Again, not as for the exponential distribution, it is not clear for the standard gamma distribution which q_n values will give this bias adjusted asymptotic relative efficiency Λ less than 1. Based on our simulation results in Chapter 3, we can conclude that MLqE is not more efficient than MLE in limits, but MLqE can be much better than MLE for small sample size due to variance reduction.

2.5 Choice of Distortion Parameter q

When estimating the parameter in either the exponential distribution or standard gamma distribution, with $q_n \to 1$, the asymptotic variance of the MLqE is equivalent to that of the MLE in limit, but can be smaller for small sample sizes. In this section we discuss the choice of q such that the MLqE has reduced variance.

Ferrari and Yang (2010) discussed the choice of q for the exponential distribution. When estimating λ_0, they observed from the expression of σ_n in (2.18) that $(q^2-2q+2)/[q^5(2-q)^3] < 1$ for $1 < q < 1.40$; thus, choosing the distortion parameter in such a range gives $\sigma_n^2 < \lambda_0^2$. When estimating the tail probability $\alpha(x_n, \lambda_0)$, the MLqE is asymptotically as efficient as the MLE when $(q_n - 1)x_n \to 0$ and one needs $0 < q_n < 1$ to minimize the MSE. The method above is the main method we use to get q values in this thesis.

As Ferrari and Yang (2010) suggested, one can choose the q which minimizes an estimated asymptotic MSE of the estimator when it is mathematically tractable. In the case of the exponential distribution, by Theorem 2.2 and (2.18),

$$\mathrm{MSE}_{\tilde{\lambda}}(q, \lambda_0) = \left(\frac{\lambda_0}{q} - \lambda_0\right)^2 + n^{-1}\left(\frac{\lambda_0}{q}\right)^2 \left[\frac{q^2 - 2q + 2}{q^3(2-q)^3}\right]. \tag{2.25}$$

However, since λ_0 is unknown, we consider

$$q^* = \underset{q \in (0,2)}{\arg\min}\{\mathrm{MSE}_{\tilde{\lambda}}(q, \hat{\lambda})\}, \tag{2.26}$$

where $\hat{\lambda}$ is the MLE. The reason why we choose interval (0,2) is due to the positiveness of σ_n^2 in (2.18). This will be used in some of our simulation studies in Chapter 3, similar to what Ferrari and Yang (2010) did for estimating the tail probability for the exponential distribution.

Now we look at how to choose q for the standard gamma distribution. By Theorem 2.2, the MLqE $\tilde{\theta}$ of θ_0 has asymptotic MSE

$$\mathrm{MSE}_{\tilde{\theta}}(q, \theta_0) = \left(\frac{\theta_0}{q} - \theta_0\right)^2 + \frac{\sigma_n^2}{n}, \tag{2.27}$$

where σ_n is given in (2.19) with q_n replaced by q. As a result, when estimating θ_0, we choose q^* such that

$$q^* = \underset{q \in (0,2)}{\arg\min}\{\mathrm{MSE}_{\tilde{\theta}}(q, \hat{\theta})\}, \tag{2.28}$$

where $\hat{\theta}$ is the MLE. By Theorem 2.3, the MLqE $\alpha(x_n; \tilde{\theta}_n)$ of the upper tail probability $\alpha(x_n; \theta_0)$ has asymptotic MSE

$$\mathrm{MSE}_{\tilde{\alpha}}(q, \theta_0) = [\alpha(x_n; \theta_0/q) - \alpha(x_n; \theta_0)]^2 + [\alpha'(x_n; \theta_0/q)]^2 \frac{\sigma_n^2}{n}, \tag{2.29}$$

where σ_n is given in (2.19) with q_n replace by q, $\alpha(x;\theta) = 1 - \int_0^x \frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}dy$ and $\alpha'(x;\theta) = -\int_0^x \frac{\partial}{\partial\theta}\frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}dy$. Note that $\alpha'(x;\theta)$ is the derivative with respect to θ instead of x. Now (2.29) can be written more explicitly as

$$\mathrm{MSE}_{\tilde{\alpha}}(q,\theta_0) \;=\; \left[\int_0^{x_n} \frac{y^{\theta_0-1}e^{-y}}{\Gamma(\theta_0)}dy - \int_0^{x_n} \frac{y^{\theta_0/q-1}e^{-y}}{\Gamma(\theta_0/q)}dy\right]^2 + \left[\int_0^{x_n} \left(\frac{\partial}{\partial\theta}\frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}\right)\bigg|_{\theta=\theta_0/q} dy\right]^2 \frac{\sigma_n^2}{n} \tag{2.30}$$

As a result, when estimating $\alpha(x_n;\theta_0)$, we choose q^* such that

$$q^* = \underset{q\in(0,2)}{\arg\min}\{\mathrm{MSE}_{\tilde{\alpha}}(q,\hat{\theta})\}, \tag{2.31}$$

where $\hat{\theta}$ is the MLE. The choices of q in (2.28) and (2.31) will be also used in some of our Chapter 3 simulation studies.

Chapter 3

MONTE CARLO SIMULATION

In this chapter, we implement Monte Carlo simulation studies to examine, for both exponential and standard gamma distributions, the MLqE's finite sample performance compared with the traditional MLE. In Section 3.1, we assess the accuracy of MLqEs and MLEs for both parameter and tail probabilities by looking at their MSE ratio with a varying but deterministic distortion parameter q. In Section 3.2, we assess the reliability of confidence intervals produced by MLEs and MLqEs with data-driven optimal distortion parameter. Section 3.3 is devoted to a robustness study of the MLqEs and MLEs.

3.1 MSE: Role of Distortion Parameter q

In the first group of simulations, we compare the two estimators of the natural parameter in either the exponential distribution or standard gamma distribution, obtained via the MLq method and the traditional ML approach respectively. Particularly, we are interested in assessing the relative performance of the two estimators for different choices of sample size by taking the ratio of the their MSEs, i.e. MSE$(\hat{\lambda}_n)$/MSE$(\tilde{\lambda}_n)$ for λ_0 in $Exp(\lambda_0)$, or MSE$(\hat{\theta}_n)$/MSE$(\tilde{\theta}_n)$ for θ_0 in $Gamma(\theta_0, 1)$, or MSE$(\hat{\alpha}_n)$/MSE$(\tilde{\alpha}_n)$ for upper tail probabilities α_0 of either $Exp(\lambda_0)$ or $Gamma(\theta_0, 1)$.

The simulations are structured as follows:

(i) For any given sample size $n \geq 2$, $B = 10,000$ of Monte Carlo samples $X_1, \cdots, X_n$ are generated from an exponential distribution with parameter λ_0 (i.e. $Exp(\lambda_0)$) or a standard gamma distribution with parameter θ_0 (i.e. $Gamma(\theta_0, 1)$).

(ii) For each sample, the MLqEs and MLEs of λ_0 or θ_0 and the corresponding α_0 are calculated.

(iii) For each sample size n, the relative performance between the two estimators is evaluated by the ratio $\hat{R}_n = \mathrm{MSE}_{\mathrm{MC}}(\hat{\beta}_n)/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\beta}_n)$, where β denotes λ (the exponential parameter) or θ (the standard gamma parameter) or α (the corresponding upper tail probability) and $\mathrm{MSE}_{\mathrm{MC}}$ denotes the Monte Carlo estimate of the MSE.

To find $\hat{R}_n$ in (iii), let $\bar{y}_1 = B^{-1}\sum_{k=1}^{B}(\hat{\beta}_{n,k}-\beta_0)^2$ and $\bar{y}_2 = B^{-1}\sum_{k=1}^{B}(\tilde{\beta}_{n,k}-\beta_0)^2$, where $\hat{\beta}_{n,k}$ and $\tilde{\beta}_{n,k}$ denote the estimates based on the k-th sample with size n, $k = 1, 2, \cdots, B$, and β_0 is either λ_0 or θ_0 or α_0. Then we use $\hat{R}_n = \bar{y}_1/\bar{y}_2$ and the rationale is as follows. By the Central Limit Theorem (CLT), for large values of B, $\bar{\boldsymbol{y}} = (\bar{y}_1, \bar{y}_2)^T$ has approximately a bivariate normal distribution with mean $\boldsymbol{\mu} = (\mathrm{MSE}(\hat{\beta}_n), \mathrm{MSE}(\tilde{\beta}_n))^T$ and a certain covariance matrix $\boldsymbol{\Gamma}$. Thus we could use $\bar{y}_1/\bar{y}_2$ to estimate $\hat{R}_n$. The standard error of $\hat{R}_n$ can be computed by the delta method (Ferguson, 1996) as

$$se(\hat{R}_n) = B^{-1/2}\left(\frac{\hat{\gamma}_{11}}{\bar{y}_2^2} - 2\hat{\gamma}_{12}\frac{\bar{y}_1}{\bar{y}_2^3} + \hat{\gamma}_{22}\frac{\bar{y}_1^2}{\bar{y}_2^4}\right)^{1/2}, \tag{3.1}$$

where $\hat{\gamma}_{11}$, $\hat{\gamma}_{22}$ and $\hat{\gamma}_{12}$ denote, respectively, the Monte Carlo estimates of the components of the covariance matrix $\boldsymbol{\Gamma}$. The standard error (3.1) is derived as followes. By the CLT, as $B \to \infty$ we have

$$\sqrt{B}\,(\bar{\boldsymbol{y}} - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{pmatrix}\right).$$

Let $g(\bar{\boldsymbol{y}}) = \bar{y}_1/\bar{y}_2$, then by the delta method we have

$$\sqrt{B}\,[g(\bar{\boldsymbol{y}}) - g(\boldsymbol{\mu})] \xrightarrow{\mathcal{D}} N\left(0, \dot{g}^T(\boldsymbol{\mu})\boldsymbol{\Gamma}\dot{g}(\boldsymbol{\mu})\right),$$

where $\dot{g}(\boldsymbol{\mu})$ denotes the gradient of $g(\boldsymbol{\mu})$, i.e.

$$\dot{g}^T(\boldsymbol{\mu}) = \left(\frac{\partial}{\partial\mu_1}g(\boldsymbol{\mu}), \frac{\partial}{\partial\mu_2}g(\boldsymbol{\mu})\right) = \left(\frac{1}{\mu_2}, -\frac{\mu_1}{\mu_2^2}\right).$$

Thus

$$\dot{g}^T(\boldsymbol{\mu})\boldsymbol{\Gamma}\dot{g}(\boldsymbol{\mu}) = \left(\frac{1}{\mu_2}, -\frac{\mu_1}{\mu_2^2}\right)\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{pmatrix}\begin{pmatrix} 1/\mu_2 \\ -\mu_1/\mu_2^2 \end{pmatrix} = \frac{\gamma_{11}}{\mu_2^2} - 2\gamma_{12}\frac{\mu_1}{\mu_2^3} + \gamma_{22}\frac{\mu_1^2}{\mu_2^4},$$

and as a result we can use (3.1) to estimate the standard error of $\hat{R}_n$.

3.1.1 Simulation results for exponential distribution

For the exponential distribution $Exp(\lambda_0)$, we can estimate the true parameter λ_0 and the upper tail probability $\alpha(x;\lambda_0)$. Since the latter has been presented in Ferrari and Yang (2010), we only give the simulation results for the parameter estimation in this section.

Study I: fixed λ_0 and q for parameter estimation.

Figure 3.1 displays results for $\lambda_0 = 0.5, 1, 1.5$ and fixed distortion parameter $q = 1.2$. The plot illustrates the behaviour of $\hat{R}_n$ for different choices of sample size. From Figure 3.1 we find that $\hat{R}_n > 1$ for most sample sizes considered, i.e. $n < 80$. This indicates that MSE($\hat{\lambda}$) is greater than MSE($\tilde{\lambda}$) and thus MLqE clearly outperforms the traditional MLE when sample size $n < 80$. When the sample size increases, the bias component becomes more relevant and we observe that $\hat{R}_n$ increases first from 1, reaches the peak around sample size $n = 20$, and then decreases slowly down to 1 when $n > 80$. From Figure 3.1 we also observe that the advantage of using MLqE for small sample sizes is more accentuated for larger values of λ_0.

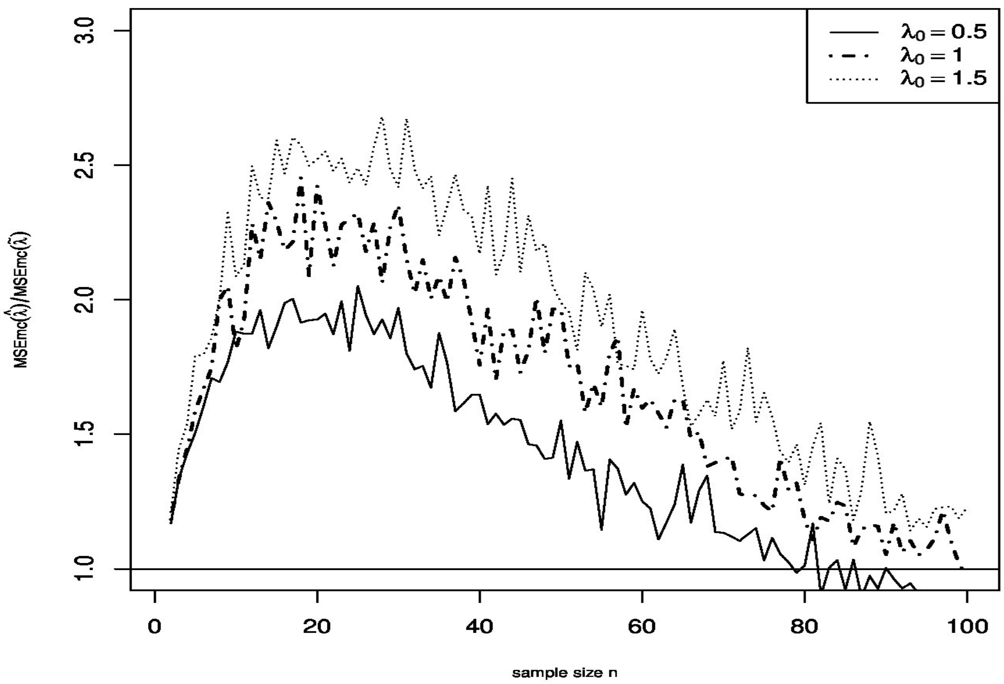

Figure 3.1: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\lambda}_n)/\text{MSE}_{\text{MC}}(\tilde{\lambda}_n)$ as a function of sample size n for $Exp(\lambda_0)$ with $\lambda_0 = 0.5, 1, 1.5$ and $q = 1.2$ (Study I).

Figure 3.2 shows results for $\lambda_0 = 1$ and the three different distortion parameter values $q = 1.35, 1.15, 1.05$. From this plot we observe that $\hat{R}_n > 1$ for most sample sizes considered, i.e. when $n < 90$ MLqE performs better than MLE in terms of MSE. When the sample size n increases, the $\hat{R}_n$ value increases first from 1, reaches the peak around sample size $n = 20$, and then decreases down to 1 when $n > 90$. Moreover, larger values of the distortion parameter q accentuate the benefits of MLqE for relatively small sample size.

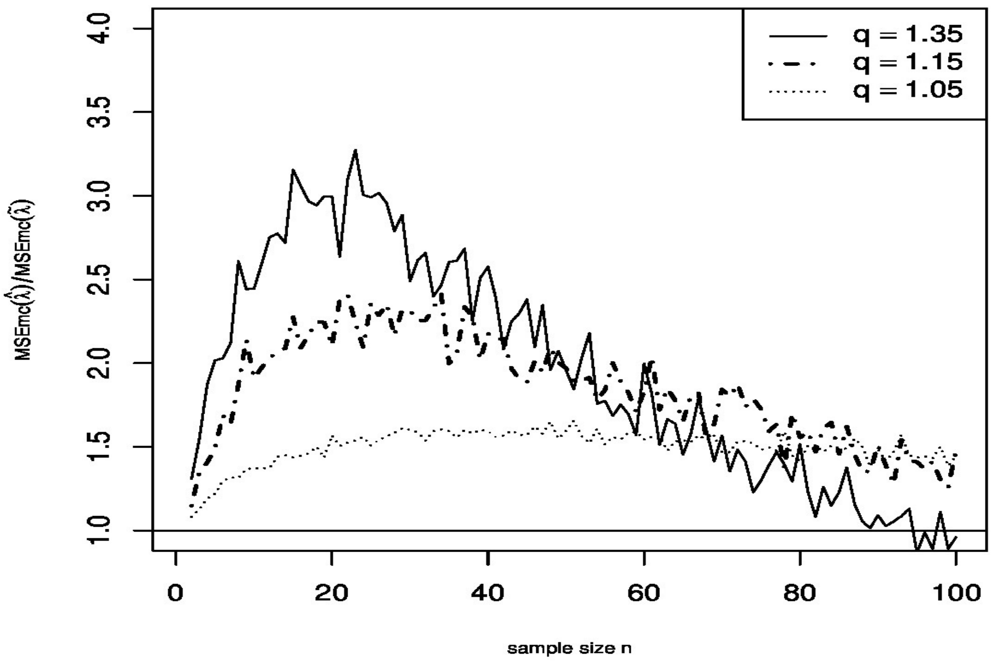

Figure 3.2: Plot of MSE ratio $\hat{R}_n = \mathrm{MSE}_{\mathrm{MC}}(\hat{\lambda}_n)/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\lambda}_n)$ as a function of sample size n for $Exp(\lambda_0)$ with $\lambda_0 = 1$ and $q = 1.35, 1.15, 1.05$ (Study I).

Study II: fixed λ_0 and $q_n \searrow 1$ for parameter estimation.

Figure 3.3 shows results for $\lambda_0 = 0.5, 1, 1.5$ and varying $q_n = [1 + e^{0.3(n-20)}]/[0.5 + e^{0.3(n-20)}]$, a decreasing function of sample size n. Note that q_n is a sequence such that $1 < q_n < 2$ and $q_n \searrow 1$ as $n \to \infty$. We choose this sequence for illustrative purposes and study $\hat{R}_n$ for different choices of λ_0. For small values of the sample size, the chosen sequence q_n is way bigger than 1 and thus produces benefits in terms of smaller variance. As a consequence, for small sample sizes, $\hat{R}_n > 1$ and the MLqE outperforms the traditional MLE in terms of MSE. In contrast, when the sample size becomes larger, q_n adjusts quickly to 1. As a consequence, for large sample sizes, $\hat{R}_n$ is close to 1 and the MLqE exhibits the same behaviour as the MLE. We also observe that the advantage of using the MLqE for small sample sizes is much more accentuated for larger values of λ_0.

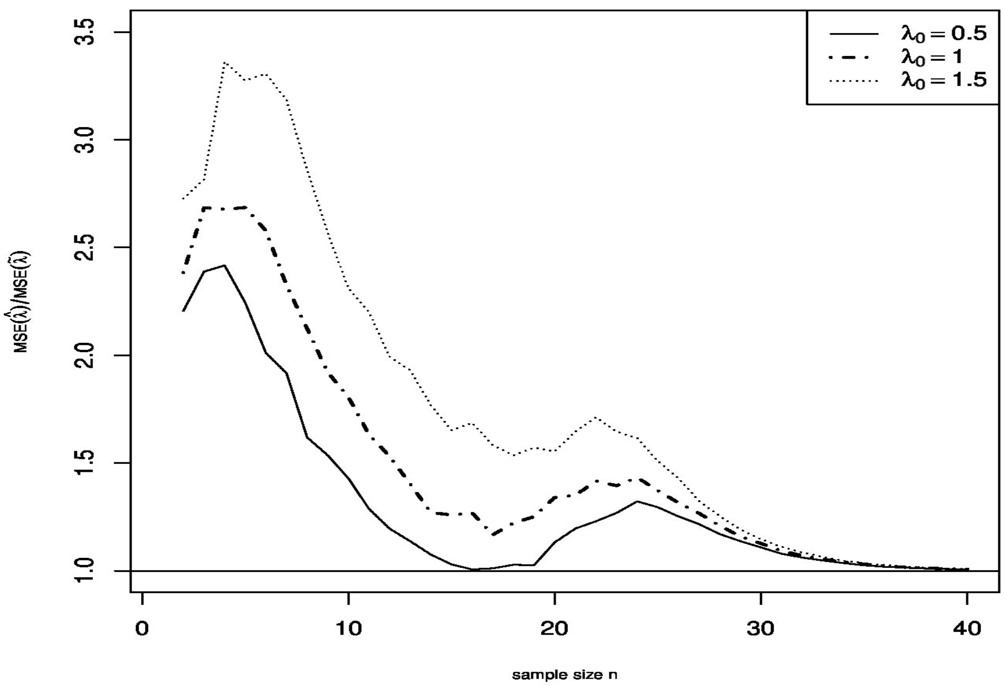

Figure 3.3: Plot of MSE ratio $\hat{R}_n = \mathrm{MSE}_{\mathrm{MC}}(\hat{\lambda}_n)/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\lambda}_n)$ as a function of sample size n for $Exp(\lambda_0)$ with $\lambda_0 = 0.5, 1, 1.5$ and $q_n = [1 + e^{0.3(n-20)}]/[0.5 + e^{0.3(n-20)}]$ (Study II).

3.1.2 Simulation results for the standard gamma distribution

In this section we investigate, for the standard gamma distribution Gamma $(\theta_0, 1)$, the MLqEs of the true parameter θ_0 and the upper tail probability $\alpha(x; \theta_0)$.

Study I: fixed θ_0 and q for parameter estimation.

In Figure 3.4, we consider $\theta_0 = 1, 5, 10$ and fixed distortion parameter $q = 1.5$. Figure 3.4 shows that $\hat{R}_n > 1$ for very small sample sizes, i.e., $n < 10$, and $\theta_0 = 5, 10$. When $\theta_0 = 1$, $\hat{R}_n < 1$ for any sample size. When the sample size increases, $\hat{R}_n$ is decreasing, $\hat{R}_n < 1$ and converges to 0. This indicates that the MLE is much better than MLqE for large sample sizes, e.g. $n > 20$, with fixed $q = 1.5$.

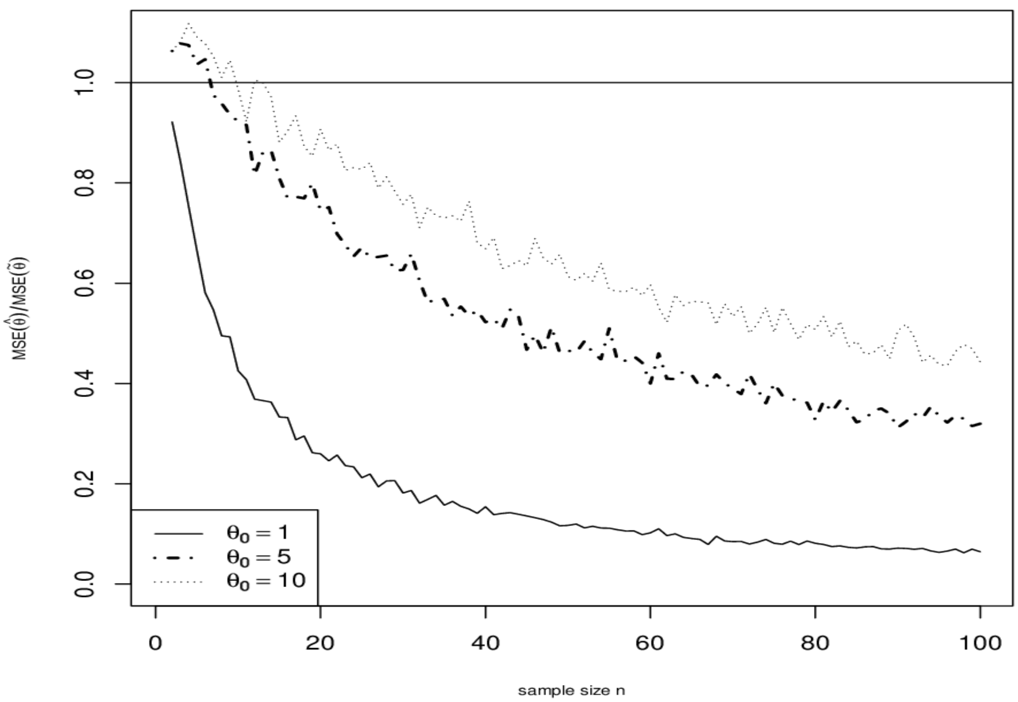

Figure 3.4: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\theta}_n)/\text{MSE}_{\text{MC}}(\tilde{\theta}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $\theta_0 = 1, 5, 10$ and $q = 1.5$ (Study I).

Figure 3.5 considers fixed $\theta_0 = 10$ and the three different distortion parameter values $q = 1.35, 1.15, 1.05$. From this plot we observe that $\hat{R}_n > 1$ for very small sample sizes, i.e. $n < 10$. When sample size n increases, the $\hat{R}_n$ value decreases slowly especially when $q = 1.05, 1.15$. Moreover, larger values of the distortion parameter q accentuate the benefits of MLqE for very small sample sizes.

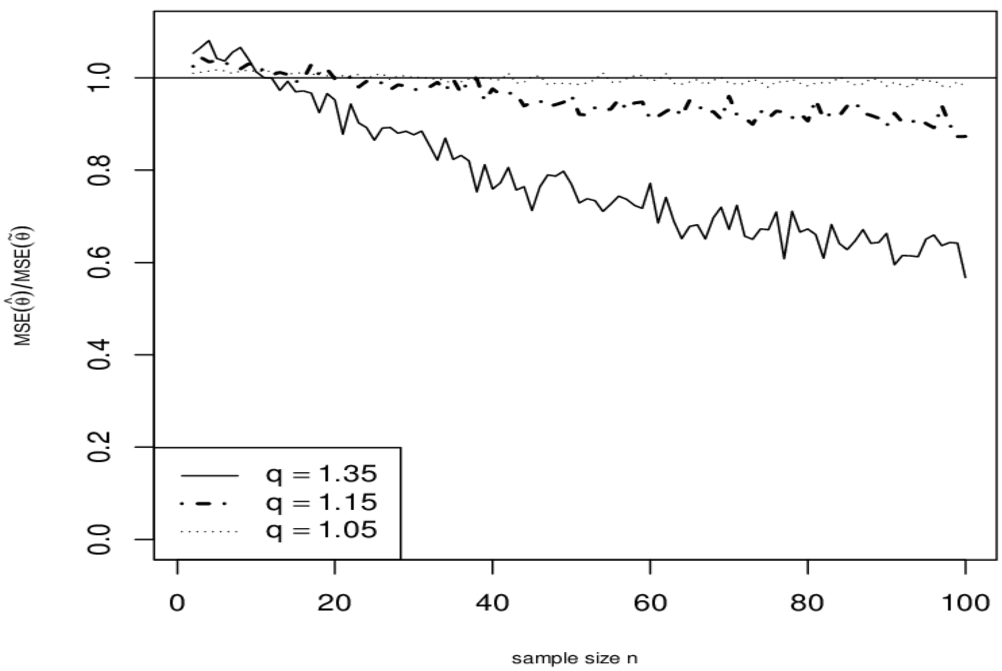

Figure 3.5: Plot of MSE ratio $\hat{R}_n = \mathrm{MSE_{MC}}(\hat{\theta}_n)/\mathrm{MSE_{MC}}(\tilde{\theta}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $\theta_0 = 10$ and $q = 1.35, 1.15, 1.05$ (Study I).

Study II: fixed θ_0 and $q_n \searrow 1$ for parameter estimation.

Figure 3.6 considers $\theta_0 = 1, 5, 10$ and varying $q_n = [1 + e^{0.3(n-20)}]/[0.5 + e^{0.3(n-20)}]$. When $\theta_0 = 5, 10$, $\hat{R}_n > 1$ for very small sample sizes $n < 7$ and $\hat{R}_n < 1$ for sample sizes $n > 10$. When $\theta_0 = 1$, $\hat{R}_n < 1$ for all sample sizes. For large sample sizes, e.g. $n > 30$, q_n converges quickly to 1 and thus $\hat{R}_n$ converges to 1 and the MLqEs and MLEs perform equivalently.

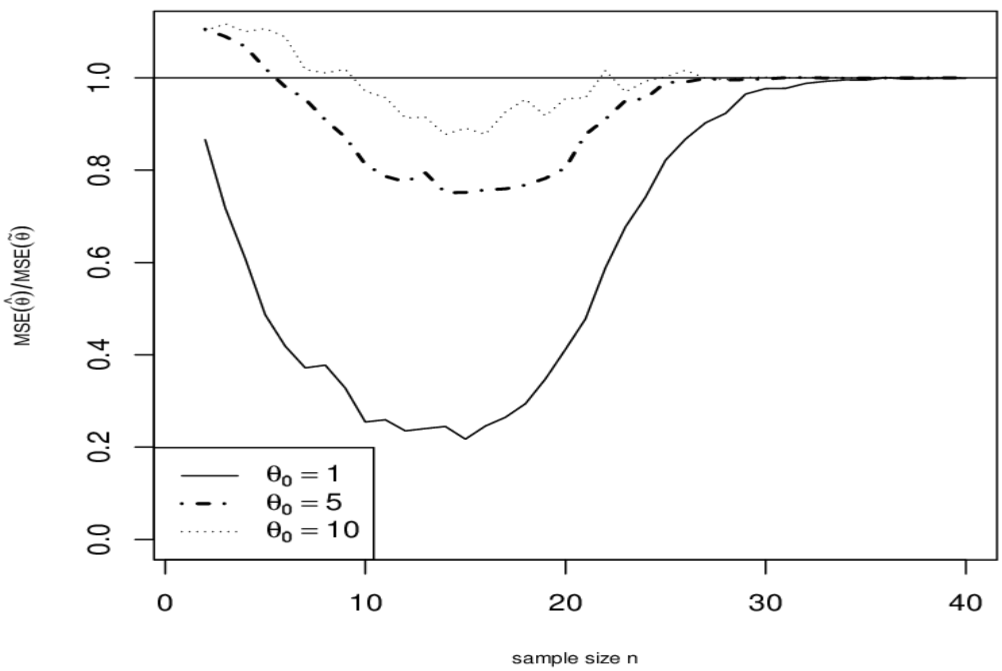

Figure 3.6: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\theta}_n)/\text{MSE}_{\text{MC}}(\tilde{\theta}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $\theta_0 = 1, 5, 10$ and $q_n = [1 + e^{0.3(n-20)}]/[0.5 + e^{0.3(n-20)}]$ (Study II).

Study III: fixed α_0 and q for tail probability estimation.

Figures 3.7 and 3.8 illustrate the behaviour of $\hat{R}_n$ for different choices of tail probability α_0 and q. For relatively small sample sizes, i.e. $n < 10$ for $q = 0.5$ in Figure 3.7 and $n < 20$ for $q = 0.65, 0.85, 0.95$ in Figure 3.8, we observe $\hat{R}_n > 1$ which means that the MLqE performs better than the MLE. Such behaviour is more accentuated for smaller values of the α_0 and smaller values of the distortion parameter q. In contrast, when the sample size is larger, the bias plays an increasingly relevant role and we observe that $\hat{R}_n < 1$.

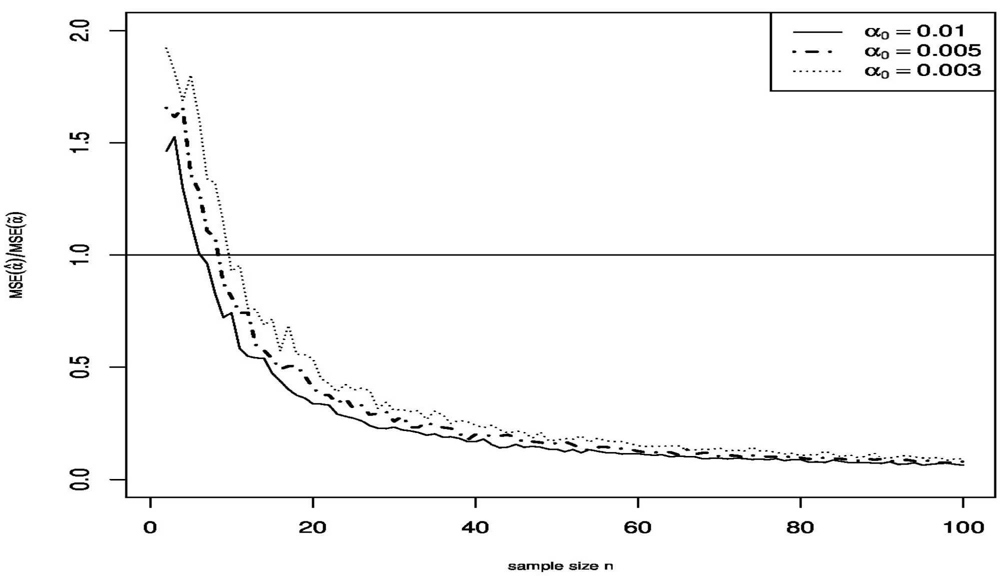

Figure 3.7: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\alpha}_n)/\text{MSE}_{\text{MC}}(\tilde{\alpha}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $\alpha_0 = 0.01, 0.005, 0.003$ and $q = 0.5$ (Study III).

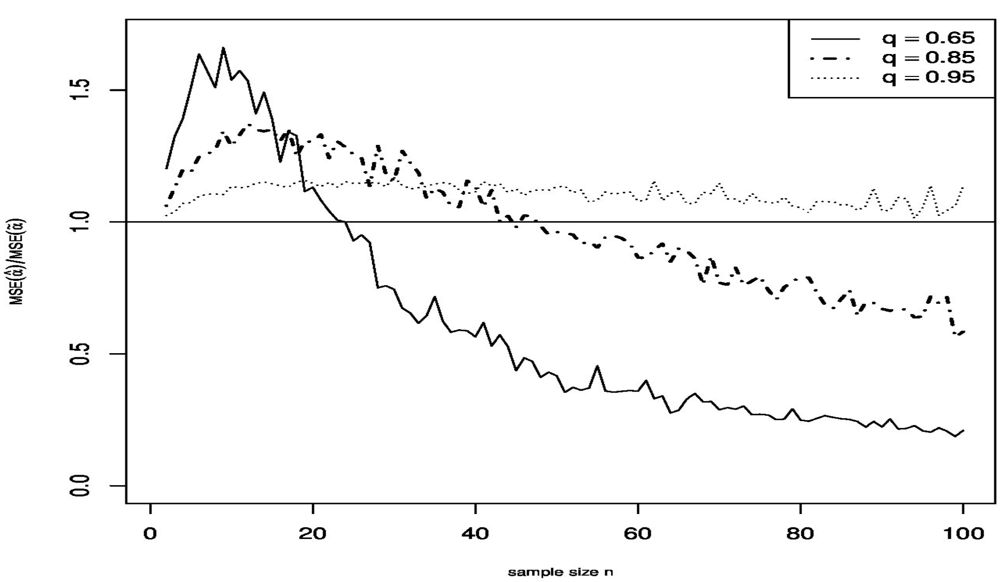

Figure 3.8: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\alpha}_n)/\text{MSE}_{\text{MC}}(\tilde{\alpha}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $\alpha_0 = 0.003$ and $q = 0.65, 0.85, 0.95$ (Study III).

Study IV: fixed α_0 and $q_n \nearrow 1$ for tail probability estimation.

Figure 3.9 considers fixed upper tail probabilities for $\alpha_0 = 0.01, 0.005, 0.003$ but varying $q_n = [0.5+e^{0.3(n-20)}]/[1+e^{0.3(n-20)}]$ so that $q_n \nearrow 1$ and $0 < q_n < 1$. For small values of sample size, the chosen sequence q_n converges relatively slowly to 1 and the distortion parameter produces benefits in terms of a smaller variance. As a consequence, for small sample sizes, $\hat{R}_n > 1$ and the MLqE outperforms the MLE in terms of MSE. In contrast, when the sample size becomes larger, q_n adjusts quickly to one. As a consequences, for large sample sizes, the MLqE exhibits the same behaviour as the MLE. We also observe that the advantage of using the MLqE for small sample sizes is much more accentuated for smaller values of α_0.

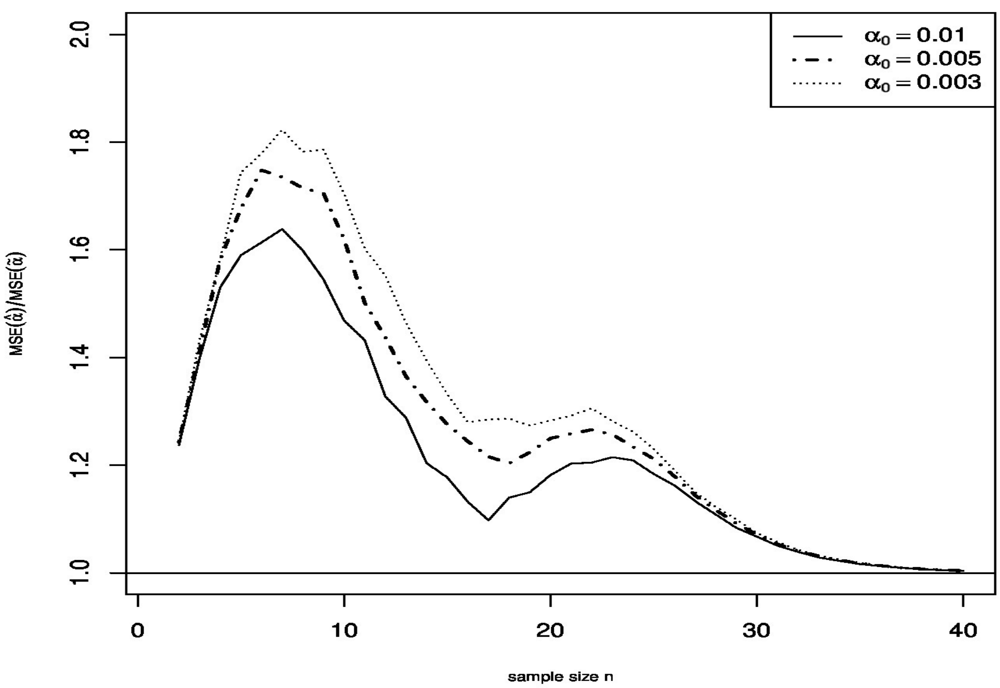

Figure 3.9: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\alpha}_n)/\text{MSE}_{\text{MC}}(\tilde{\alpha}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $\alpha_0 = 0.01, 0.005, 0.003$ and $q_n = [0.5 + e^{0.3(n-20)}]/[1 + e^{0.3(n-20)}]$ (Study IV).

Study V: $\alpha_n \searrow 0$ **and** $q_n \nearrow 1$ **for tail probability estimation.**

Figure 3.10 considers the case where both the true tail probability and the distortion parameter change with sample size. From Study III and Study IV we observe that, in order to produce benefit of smaller variance, one may choose relatively larger q values for smaller α_0 values and smaller q values for larger α_0 values, i.e. α_n decreases and q_n increases as sample size n increases. We consider sequences of distortion parameters converging slowly relative to the sequence of quantiles x_n. In particular we set $q_n = 1 - [10 \log(n + 10)]^{-1}$ and $x_n = n^{1/(2+\delta)}$ as in Ferrari and Yang (2010). In Figure 3.10, we illustrate the behaviour of the estimator for $\delta = 0.5, 1, 1.5$. Smaller δ means smaller α_n and thus, similar to our observation in Figures 3.7 and 3.9, better performance of the MLqE.

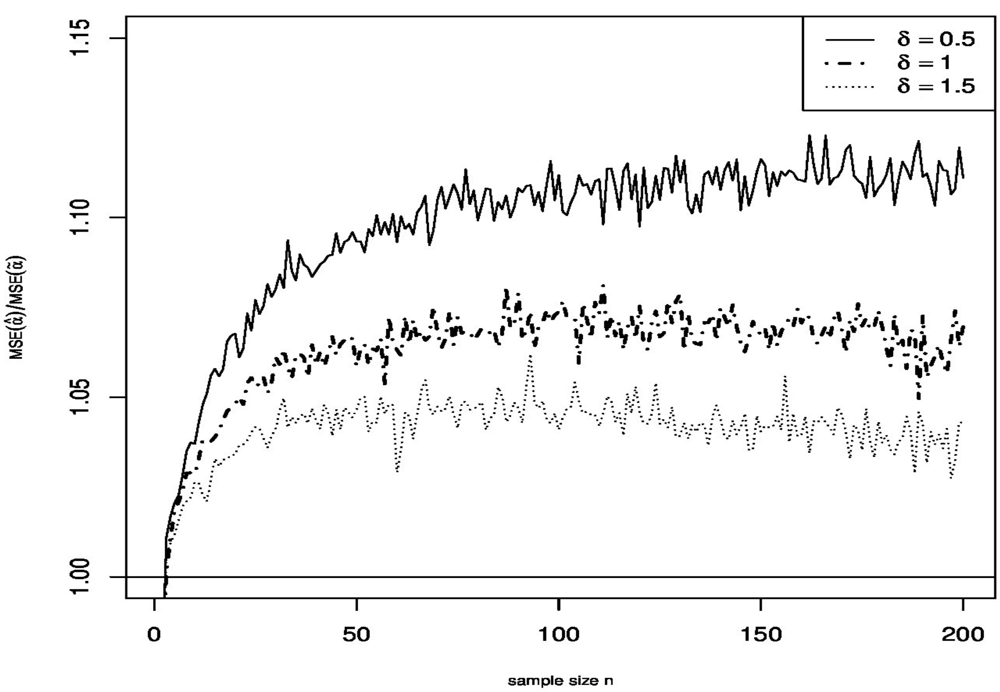

Figure 3.10: Plot of MSE ratio $\hat{R}_n = \text{MSE}_{\text{MC}}(\hat{\alpha}_n)/\text{MSE}_{\text{MC}}(\tilde{\alpha}_n)$ as a function of sample size n for $Gamma(\theta_0, 1)$ with $x_n = n^{1/(2+\delta)}$, $\delta = 0.5, 1.0, 1.5$, and $q_n = 1 - [10 \log(n + 10)]^{-1}$ (Study V).

Studies IV and V indicate that the choice of q_n depends on the size of the probability to be estimated. If q_n approaches 1 too quickly from below, the gain obtained in terms of variance vanishes rapidly as n becomes larger. On the other hand, if q_n converges to 1 too slowly, the bias dominates the variance and the MLE outperforms the MLqE.

3.2 Asymptotic and Bootstrap Confidence Intervals

In this section, we study the reliability of the MLqE based confidence intervals using three commonly used methods: (a) asymptotic normality; (b) parametric bootstraps; (c) nonparametric bootstraps. We compare the results with those obtained using the MLE. Again we consider estimating λ_0 for the exponential distribution and θ_0 and tail probability α_0 for the standard gamma distribution, as the case of tail probability for exponential distribution has been discussed in Ferrari and Yang (2010).

The structure of the simulations in this section is similar to that of Section 3.1, but a data-driven choice of q_n is used:

(i) For each sample, first we compute $\hat{\beta}$, the MLE of β_0, where β is either λ for the exponential distribution or θ for standard gamma distribution. We substitute $\hat{\beta}$ into either (2.26) if $\beta = \lambda$ or (2.28) or (2.31) if $\beta = \theta$, and solve it numerically in order to obtain q^* as described in (2.26) or (2.28) or (2.31).

(ii) With q^* calculated in (i), the MLqE $\tilde{\beta}$ of β_0 is obtained. The standard errors of the estimates are computed using three different methods: the asymptotic formula, parametric bootstrap and nonparametric bootstrap. The number of replicates employed in bootstrap re-sampling is 500. We construct 95% bootstrap confidence intervals based on the bootstrap quantiles and check the coverage of the true value β_0.

We take $B = 1,000$ repetitions for each simulation in this Section. The standard error based on asymptotic formula is derived in either (2.18) $(\sigma_n/\sqrt{n})$ for estimating λ_0 or (2.19) $(\sigma_n/\sqrt{n})$ for estimating θ_0 or (2.30) (square-root of the second term) for estimating α_0. More explicitly, for the exponential distribution, the standard error of the MLqE for λ_0 and upper tail probability $\alpha(x_n;\lambda_0)$ are respectively

$$\mathrm{SE}_{\mathrm{asy}}(\tilde{\lambda}) = \frac{\sigma_n}{\sqrt{n}} = n^{-1/2}\frac{\lambda_0}{q_n}\sqrt{\frac{q_n^2-2q_n+2}{q_n^3(2-q_n)^3}},$$

$$\mathrm{SE}_{\mathrm{asy}}(\tilde{\alpha}) = \frac{\sigma_n\alpha'(x_n;\lambda_n^*)}{\sqrt{n}} = n^{-1/2}\frac{\lambda_0 x_n e^{-\lambda_0 x_n/q_n}}{q_n}\sqrt{\frac{q_n^2-2q_n+2}{q_n^3(2-q_n)^3}}$$

For the standard gamma distribution, the standard error of the MLqE for θ_0 and upper tail probability $\alpha(x_n;\theta_0)$ are respectively

$$\mathrm{SE}_{\mathrm{asy}}(\tilde{\theta}) = \frac{\sigma_n}{\sqrt{n}} = n^{-1/2}\frac{\frac{\Gamma^{1/2}((2/q_n-1)\theta_0)}{\Gamma^{1/2}(\theta_0)}\left\{\psi'((2/q_n-1)\theta_0)+[\psi((2/q_n-1)\theta_0)-\psi(\theta_0/q_n)]^2\right\}^{1/2}}{\frac{q_n\Gamma(\theta_0/q_n)}{\Gamma(\theta_0)}\psi'(\theta_0/q_n)}$$

$$\begin{aligned}\mathrm{SE}_{\mathrm{asy}}(\tilde{\alpha}) &= \frac{\sigma_n\alpha'(x_n;\theta_n^*)}{\sqrt{n}}\\ &= n^{-1/2}\int_0^{x_n}\left(\frac{\partial}{\partial\theta}\frac{y^{\theta-1}e^{-y}}{\Gamma(\theta)}\right)\Bigg|_{\theta=\theta_0/q} dy\\ &\quad\cdot\frac{\frac{\Gamma^{1/2}((2/q_n-1)\theta_0)}{\Gamma^{1/2}(\theta_0)}\left\{\psi'((2/q_n-1)\theta_0)+[\psi((2/q_n-1)\theta_0)-\psi(\theta_0/q_n)]^2\right\}^{1/2}}{\frac{q_n\Gamma(\theta_0/q_n)}{\Gamma(\theta_0)}\psi'(\theta_0/q_n)}\end{aligned}$$

In these formulas, $q=1$ corresponds to the MLE.

3.2.1 Simulation results for the exponential distribution

For the exponential distribution $Exp(\lambda_0)$, we only present the simulation results for estimating the true parameter λ_0 as that for tail probability has been presented in Ferrari and Yang (2010). Without loss of generality, we take $\lambda_0=1$.

In Table 3.1, we present the Monte Carlo means and standard deviations of the MLE $\hat{\lambda}$ and the MLqE $\tilde{\lambda}$ over $B=1,000$ repetitions, with standard errors computed using the three methods described above. In addition, we report the Monte Carlo average of the optimal distortion parameter q^*. Again $q^*=1$ refers to the MLE. We examine different sample sizes $n=15,25,50,100,500$. From Table 3.1 we can see that, not surprisingly, q^* approaches 1 as the sample size increases. In addition, the optimal q^* is always higher than 1 regardless of sample size, which is consistent with our choice in Section 3.1. For all sample size considered, the MLqE always has smaller standard deviation and thus better performance than the MLE, though this advantage of MLqE diminishes with increasing sample size. The MLqE also has smaller bias than the MLE for all sample sizes considered. When comparing the standard errors calculated using the three methods, the parametric bootstrap provides values closest to the Monte Carlo standard deviation. The asymptotic formula gives better estimation of standard error than nonparametric bootstrap when sample size is small (15, 25 or 50), while the nonparametric bootstrap gives better estimation when sample size is large (100 or 500). No matter which of the three methods is used, the MLqE always demonstrates smaller standard error than the MLE for all sample sizes considered.

Table 3.1: Mean, standard deviation and standard error of the MLE $\hat{\lambda}$ and MLqE $\tilde{\lambda}$.

n	q^*	Estimate	St. dev.	se_{asy}	se_{boot}	se_{pboot}
15	1.087	0.985843	0.233474	0.241559	0.287935	0.232463
	1.000	1.047873	0.244609	0.258076	0.304693	0.248623
25	1.034	1.010206	0.200986	0.223410	0.236957	0.216375
	1.000	1.041656	0.206214	0.231035	0.243282	0.224222
50	1.018	1.003199	0.128298	0.138897	0.141865	0.129142
	1.000	1.020331	0.130608	0.141443	0.143734	0.131644
100	1.009	1.017528	0.104521	0.107096	0.106713	0.104592
	1.000	1.026400	0.105389	0.108068	0.107465	0.105755
500	1.002	1.000736	0.045950	0.046142	0.046083	0.045965
	1.000	1.002474	0.046054	0.046468	0.046314	0.046336

In Table 3.2, we compare the accuracy of 95% confidence intervals for the MLE and MLqE and report both the coverage probability/rate (Coverage) and the relative length of intervals (RL) for MLqE over those for MLE. Here RL is the averaged ratio, over $B = 1,000$ repetitions, of the interval length for MLqE over that for MLE. From Table 3.2 we observe that when either nonparametric bootstrap or parametric bootstrap is used to calculate standard error, the coverage probability for MLqE is always larger than that of MLE. Although the coverage probability for MLqE is slightly smaller than that of the MLE (within 1%) when asymptotic formula is used, the interval length is reduced for all considered cases and all the three standard error calculation methods. The reduction in interval length is more evident when the sample size is small. For all the sample size considered, the parametric bootstrap provides the most accurate (closest to 95%) confidence interval followed by nonparametric bootstrap.

Table 3.2: Coverage rate and relative length of interval (RL) for the MLqE $\tilde{\lambda}$ over MLE $\hat{\lambda}$.

		Asympt.		Boot.		Par. boot.	
n	q^*	Coverage	RL	Coverage	RL	Coverage	RL
15	1.087	82.8	0.936	85.8	0.945	87.6	0.935
	1.000	83.4		84.5		87.2	
25	1.034	85.7	0.967	88.1	0.974	90.3	0.965
	1.000	86.5		87.7		90.2	
50	1.018	89.3	0.982	90.6	0.987	92.1	0.981
	1.000	89.9		90.2		91.7	
100	1.009	91.0	0.991	92.9	0.993	93.9	0.989
	1.000	91.2		92.5		93.6	
500	1.002	94.1	0.993	94.6	0.995	95.1	0.992
	1.000	94.3		94.5		94.8	

3.2.2 Simulation results for the standard gamma distribution

For the standard gamma distribution $Gamma(\theta_0, 1)$, we present the simulation results for estimating both the true parameter θ_0 (Tables 3.3 and 3.4) and the upper tail probability α_0 (Tables 3.5 and 3.6). Without loss of generality, we take $\theta_0 = 1$ and $\alpha_0 = 0.01$.

In Table 3.3, we present the means, standard deviations and standard errors of the MLE $\hat{\theta}$ and the MLqE $\tilde{\theta}$ of gamma parameter θ_0. From Table 3.3 we observe a similar phenomena as in Table 3.1. The optimal q^* approaches 1 as the sample size increases. It is always higher than 1 regardless of sample size, which is consistent with our choice in Section 3.1. For all sample sizes considered the MLqE has smaller standard deviation than the MLE, though the advantage of MLqE diminishes with increasing sample size. When comparing the standard errors calculated using the three methods, the parametric bootstrap provides values closest to the Monte Carlo standard deviation.

Table 3.3: Mean, standard deviation and standard error of the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$.

n	q^*	Estimate	St. dev.	se_{asy}	se_{boot}	se_{pboot}
15	1.033	1.062884	0.220070	0.249909	0.266232	0.235482
	1.000	1.044836	0.221192	0.269299	0.285963	0.265781
25	1.022	1.012154	0.165725	0.185742	0.190547	0.179941
	1.000	0.999072	0.167732	0.195312	0.199735	0.196228
50	1.011	1.019137	0.100756	0.109429	0.108276	0.108026
	1.000	1.012545	0.100918	0.112350	0.110712	0.114313
100	1.006	1.009953	0.075361	0.081522	0.078320	0.076018
	1.000	1.006411	0.075527	0.082596	0.079111	0.078127
500	1.001	0.999521	0.037271	0.037668	0.037853	0.037382
	1.000	0.998792	0.037289	0.038010	0.038082	0.037798

In Table 3.4, we compare the accuracy of 95% confidence intervals for MLE and MLqE of gamma parameter θ_0 and report both the Coverage and the RL. From Table 3.4 we observe similar phenomena as those demonstrated in Table 3.2. The coverage probability for MLqEs is always larger than that of MLEs when either nonparametric bootstrap or parametric bootstrap is used to calculate standard error, while it is smaller than that of MLEs (within 1%) when asymptotic formula is used. Regardless of sample size and method of calculating the standard error, the interval length of MLqEs is always reduced especially when the sample size is small. For all the sample size considered, the parametric bootstrap provides most accurate confidence interval followed by nonparametric bootstrap.

Table 3.4: Coverage rate and relative length of interval (RL) for the MLqE $\tilde{\theta}$ over MLE $\hat{\theta}$.

		Asympt.		Boot.		Par. boot.	
n	q^*	Coverage	RL	Coverage	RL	Coverage	RL
15	1.033	87.3	0.928	90.5	0.931	91.7	0.886
	1.000	87.8		89.4		90.9	
25	1.022	89.5	0.951	91.7	0.954	92.4	0.917
	1.000	89.9		91.3		92.0	
50	1.011	91.1	0.974	93.5	0.978	93.8	0.945
	1.000	91.3		93.2		93.6	
100	1.006	92.0	0.987	94.6	0.990	94.9	0.973
	1.000	92.5		94.1		94.8	
500	1.001	93.2	0.991	94.8	0.994	95.1	0.989
	1.000	93.6		94.7		94.9	

In Table 3.5, we present the means, standard deviations and standard errors of the MLE $\hat{\alpha}$ and the MLqE $\tilde{\alpha}$ of gamma upper tail probability α_0. From Table 3.5 we observe a similar phenomena to that in Tables 3.1 and 3.3. The optimal q^* approaches 1 as the sample size increases. It is always smaller than 1 regardless of sample size, which is consistent with our choice in Section 3.1. The MLqE has smaller bias and standard deviation than the MLE, though this advantage diminishes with increasing sample size. When comparing the standard errors calculated using the three methods, the parametric bootstrap generally provides values closest to the Monte Carlo standard deviation, followed by the asymptotic formula.

Table 3.5: Mean, standard deviation and standard error of the MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$.

n	q^*	Estimate	St. dev.	se_{asy}	se_{boot}	se_{pboot}
15	0.975	0.011633	0.005910	0.005725	0.005656	0.005891
	1.000	0.011997	0.006102	0.006312	0.006043	0.006740
25	0.985	0.010302	0.004419	0.004663	0.004950	0.004236
	1.000	0.010503	0.004475	0.005036	0.005167	0.004686
50	0.993	0.010376	0.002470	0.002836	0.002946	0.002799
	1.000	0.010477	0.002498	0.002989	0.003022	0.002984
100	0.996	0.010110	0.001776	0.001792	0.001848	0.001784
	1.000	0.010251	0.001784	0.001844	0.001884	0.001874
500	0.999	0.009987	0.000863	0.000866	0.000867	0.000865
	1.000	0.009997	0.000864	0.000878	0.000873	0.000885

In Table 3.6, we compare the accuracy of 95% confidence intervals for the MLE and the MLqE of gamma upper tail probability α_0 and report both the Coverage and the RL. From Table 3.6 we observe a similar phenomena to that in Tables 3.2 and 3.4. The coverage probability for MLqE is always larger than that of MLE when either nonparametric bootstrap or parametric bootstrap is used to calculate standard error, while it is smaller than that of the MLE (within 1%) when asymptotic formula is used. Regardless of the sample size and method of calculating the standard error, the interval length of MLqE is always reduced especially when the sample size is small. For all the sample size considered, the parametric bootstrap provides most accurate confidence interval followed by nonparametric bootstrap.

Table 3.6: Coverage rate and relative length of interval (RL) for the MLqE $\tilde{\alpha}$ over MLE $\hat{\alpha}$.

		Asympt.		Boot.		Par. boot.	
n	q^*	Coverage	RL	Coverage	RL	Coverage	RL
15	0.975	79.9	0.907	84.5	0.936	89.2	0.874
	1.000	80.3		83.8		88.7	
25	0.985	84.1	0.926	87.7	0.958	91.3	0.904
	1.000	84.5		87.5		89.6	
50	0.993	88.9	0.949	90.3	0.975	92.5	0.938
	1.000	89.3		89.9		92.2	
100	0.996	92.1	0.972	92.8	0.981	94.2	0.952
	1.000	92.6		92.7		94.0	
500	0.999	94.5	0.986	94.9	0.994	95.2	0.977
	1.000	94.6		94.7		95.0	

3.3 Robustness Study

The MLqE may have some robustness properties when compared with the traditional MLE. In this section, we briefly investigate this aspect for the exponential and standard gamma distributions.

We look at whether the MLqE is resistant to outlying observations. Particularly, for simplicity we examine how MLqE behaves when a single outlying observation is present. For this purpose, the α-influence function (IF) given in Beran (1977) is a suitable measurement. The α-IF measures the change in the estimate when a component with probability α is added to the original model. Here we use an adapted version of the α-IF proposed by Lu, Hui and Lee (2003).

In this section, we always use the sample size $n = 20$. Without loss of generality, we take the true parameters $\lambda_0 = 1$ for the exponential distribution, $\theta_0 = 1$ for standard gamma distribution and $\alpha_0 = 0.01$ for upper tail probability of both distributions. We randomly select a sample of $n = 20$ observations from either $Exp(1)$ or $Gamma(1, 1)$. Based on this sample, we can calculate the MLE and the MLqE with a fixed q value. To calculate the α-IF, we replace the last observation (data are not sorted) with an outlying observation x, where x is an integer varying from 1 to 20 (large enough for both $Exp(1)$ and $Gamma(1, 1)$ as an outlier). Thus the contamination rate is $\alpha = 1/n = 1/20$. Now the α-IF is defined as

$$\text{IF}(x) = \frac{W[(X_i)_{i=1}^{n-1}, x] - W[(X_i)_{i=1}^{n}]}{1/n},$$

where W represents a functional (estimator of β) based on the data. In our simulation, W is either the MLE $\hat{\beta}$ or MLqE $\tilde{\beta}$ with β being λ, θ or α.

The α-IFs are averaged over $B = 100$ repetitions and results are presented in Figures 3.11-3.14. Figures 3.11 and 3.12 give the α-IFs of MLE and MLqE of parameter λ_0 and upper tail probability α_0 in $Exp(\lambda_0)$, while Figures 3.13 and 3.14 give those of parameter θ_0 and upper tail probability α_0 in $Gamma(\theta_0, 1)$. The fixed distortion parameters are chosen in the same direction as discussed and observed in Sections 3.1 and 3.2. In another words, we choose $q > 1$ for parameter estimation but $q < 1$ for tail probability estimation. For the parameter estimations in Figures 3.11 and 3.13, we observe that the α-IFs of MLE and MLqE increases in their absolute values at about the same rate when the outlying observation increases from 0 to 20, though the MLqE performs a bit worse than the MLE. However when looking at the tail probabilities estimators in Figures 3.12 and 3.14, MLqE performs much better than MLE in the sense that the α-IF of MLqE increases mildly, keeps the same level or even decreases as outlying observation increases, while that of MLE increases dramatically.

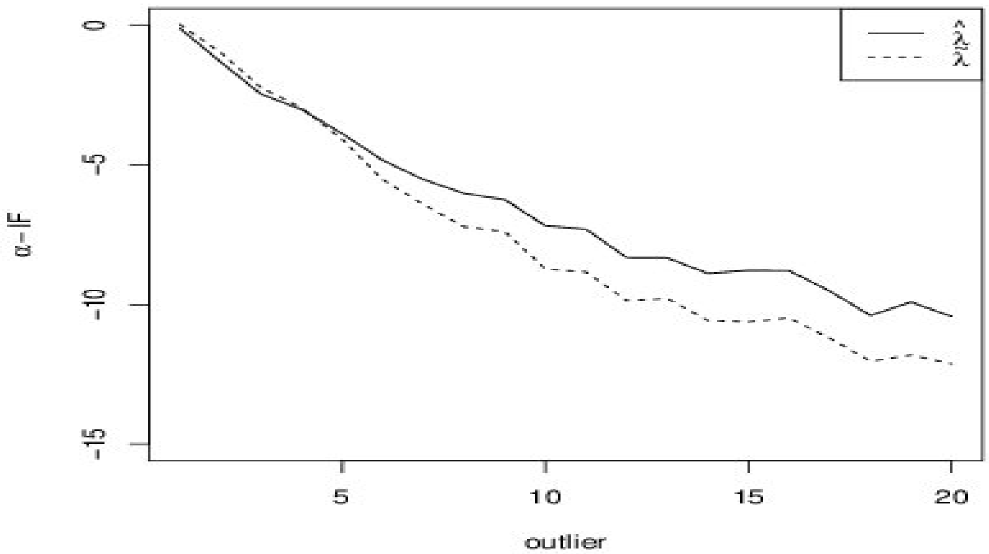

Figure 3.11: The α-IF of MLE $\hat{\lambda}$ and MLqE $\tilde{\lambda}$ with $q = 1.5$ for $Exp(\lambda_0)$.

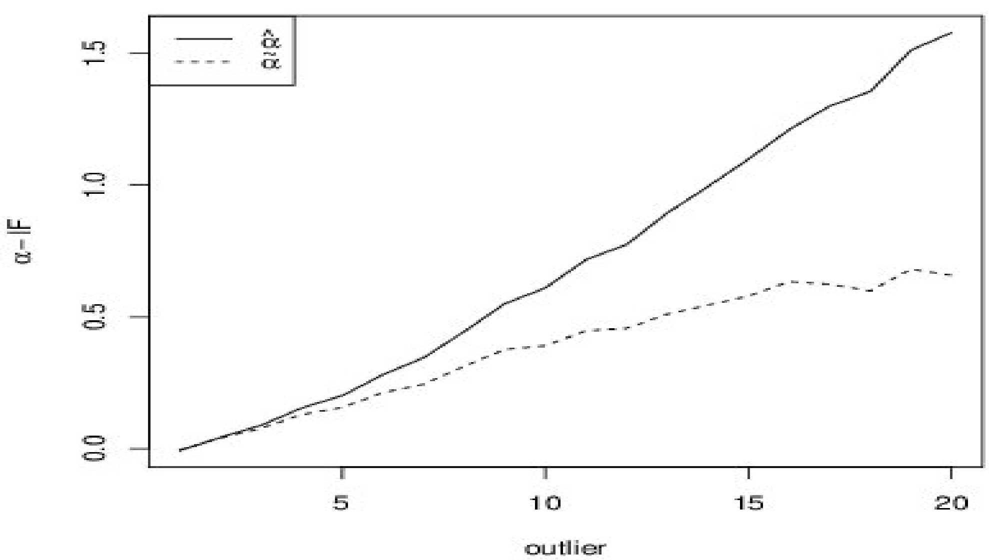

Figure 3.12: The α-IF of MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$ with $q = 0.95$ for $Exp(\lambda_0)$.

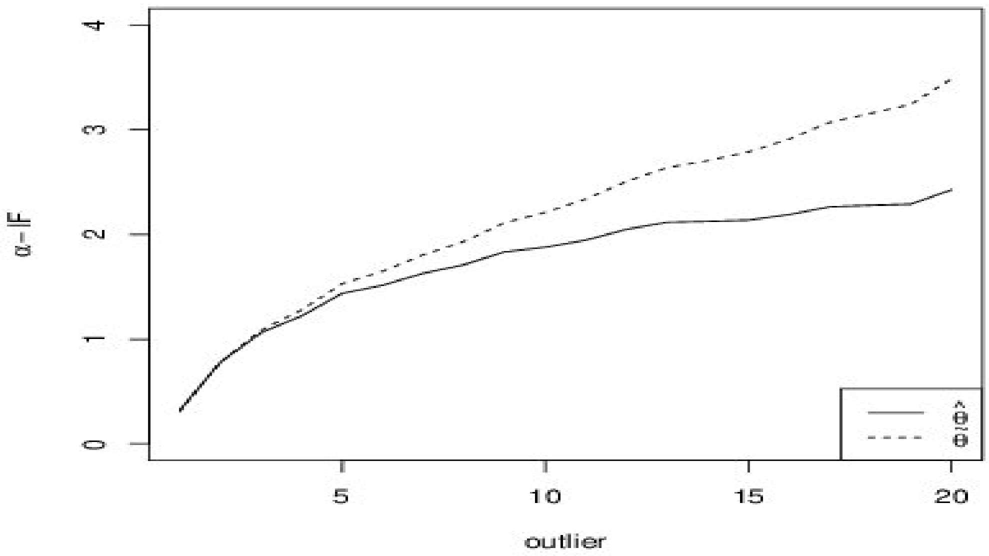

Figure 3.13: The α-IF of MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$ with $q = 1.5$ for $Gamma(\theta_0, 1)$.

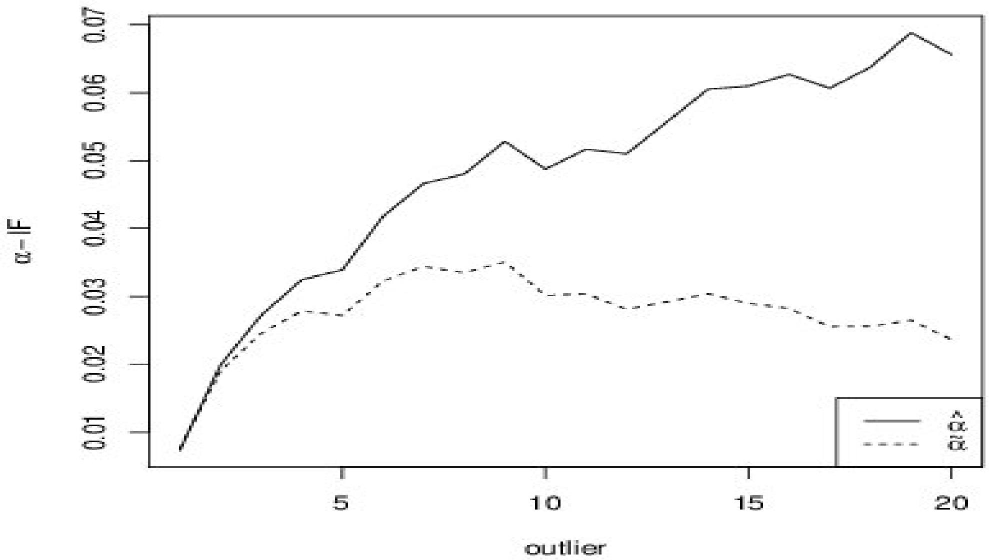

Figure 3.14: The α-IF of MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$ with $q = 0.95$ for $Gamma(\theta_0, 1)$.

To give a more complete discussion, we also present the results with q in the opposite direction. The results with $q = 0.8$ ($q < 1$) are presented in Figures 3.15 and 3.17, while those for tail probabilities estimators with $q = 1.1$ ($q > 1$) are presented in Figures 3.16 and 3.17. Note that the curves for MLE do not change with different q values. We observe a similar phenomena but in opposite manners for parameter estimation and tail probability estimation. In another words, in terms of α-IF, the MLE and MLqE perform competitively for tail probability estimation although the MLqE is a little bit worse, while the MLqE perform dramatically better than the MLE for parameter estimation.

The study presented in this section illustrates that the MLqE is generally more robust with respect to outlying observations than the MLE. Also the choice of distortion parameter q provides enough flexibility according to the purpose of an analysis. When exponential and standard gamma models are concerned, if the concentration is on the accuracy of estimation, then one should choose the distortion parameter q in the same direction as the optimal value q^*; if the resistance to outliers is more important, then one should choose q in the same direction as q^* for tail probability estimation and the opposite direction for parameter estimation.

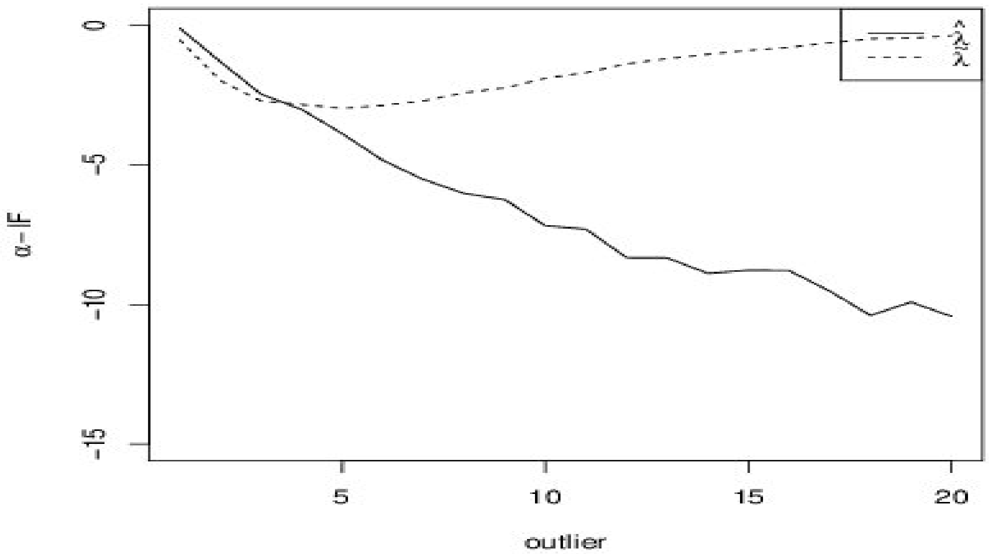

Figure 3.15: The α-IF of MLE $\hat{\lambda}$ and MLqE $\tilde{\lambda}$ with $q = 0.8$ for $Exp(\lambda_0)$.

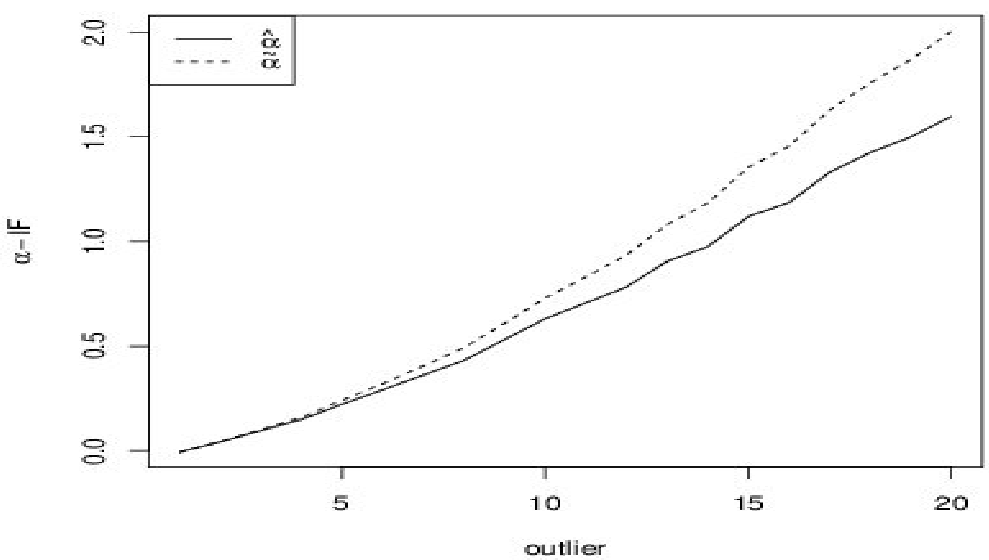

Figure 3.16: The α-IF of MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$ with $q = 1.1$ for $Exp(\lambda_0)$.

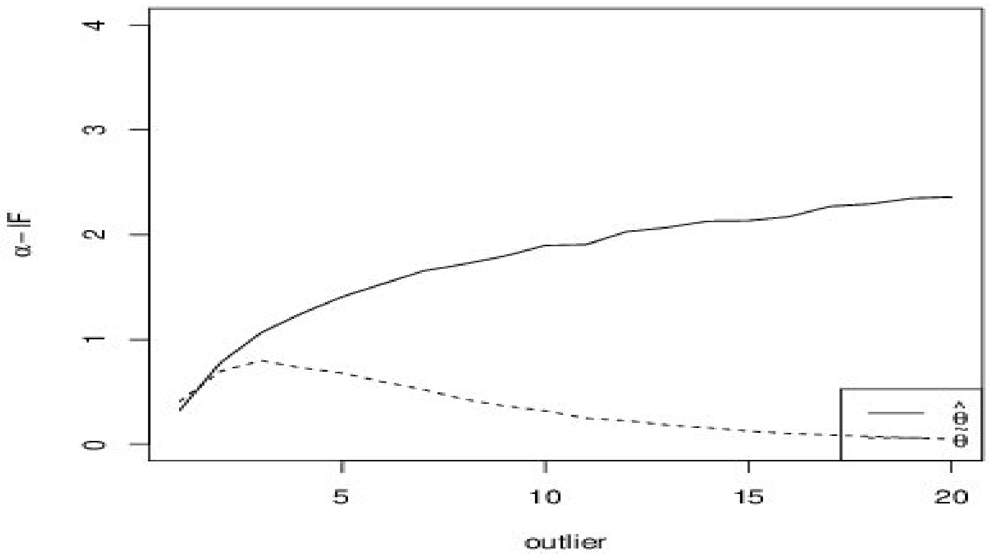

Figure 3.17: The α-IF of MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$ with $q = 0.8$ for $Gamma(\theta_0, 1)$.

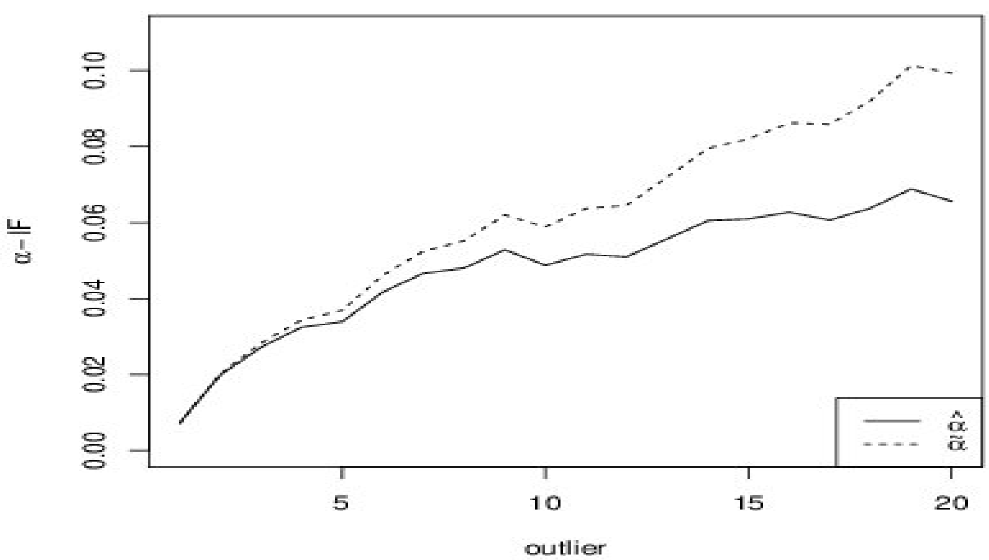

Figure 3.18: The α-IF of MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$ with $q = 1.1$ for $Gamma(\theta_0, 1)$.

Chapter 4

REAL DATA ANALYSIS

In this chapter we analyze a real data set, the Guinea Pigs data. We apply the MLqE to this data and, for the purpose of comparison, provide the results from the MLE approach as well. In Section 4.1 we fit exponential models to the data. For the fitted model, we calculate the quadratic error and the Gain(%) and compare them between the MLqE and the MLE for different subsample sizes. In Section 4.2 we fit the more general gamma model to the data and compare the fitted distributions based on the respective MLqE and MLE.

The Guinea Pigs data was presented in Bjerkedal (1960) and comprises survival times, in days, of 72 Guinea pigs injected with different amount of tubercle. This species of Guinea pigs are known to have high susceptibility of human tuberculosis, which is one of the reasons for selecting them. We consider only the study in which animals in a single cage are under the same regimen. The data (in days) are given below:

12 15 22 24 24 32 32 33 34 38 38 43 44 48 52 53 54 54 55 56 57 58 58 59 60 60 60 60 61 62 63 65 65 67 68 70 70 72 73 75 76 76 81 83 84 85 87 91 95 96 98 99 109 110 121 127 129 131 143 146 146 175 175 211 233 258 258 263 297 341 341 376.

4.1 Fitted Exponential Models

Preliminary analysis of the data shows that an exponential density is appropriate to model its distribution. The distortion parameter q^* is selected by using (2.26). The MLE and MLqE (with optimal distortion parameter $q^* = 1.01$) along with their standard errors are respectively

$$\hat{\lambda} = 0.010018 \ (se = 0.00117) \quad \text{and} \quad \tilde{\lambda} = 0.009937 \ (se = 0.00116).$$

Here the standard errors are calculated by using the asymptotic formula $\sigma_n/\sqrt{n}$ with σ_n derived in (2.18), $q_n = q^*$ for the MLqE and $q_n = 1$ for the MLE. The two fitted exponential models $Exp(\hat{\lambda})$ and $Exp(\tilde{\lambda})$ based on respective MLEs and MLqEs give the following Kolmogorov-Smirnov(ks) test statistics and p-values:

$$\hat{\lambda} : ks = 0.0931,\ p\text{−value} = 0.2868 \quad \text{and} \quad \tilde{\lambda} : ks = 0.0961,\ p\text{−value} = 0.2645.$$

The two high p-values (> 0.1) mean that both of the two fitted exponential models are reasonably appropriate. In addition, the two p-values are very close which indicates that the two gamma models fit the data equivalently well.

We examine in Figure 4.1 the fitted exponential models based on the MLE $\hat{\lambda}$ and MLqE $\tilde{\lambda}$, along with the histogram of the survival times. As expected from the estimates, the fitted exponential model based on the MLqE fits the data the same as well as that based on the MLE.

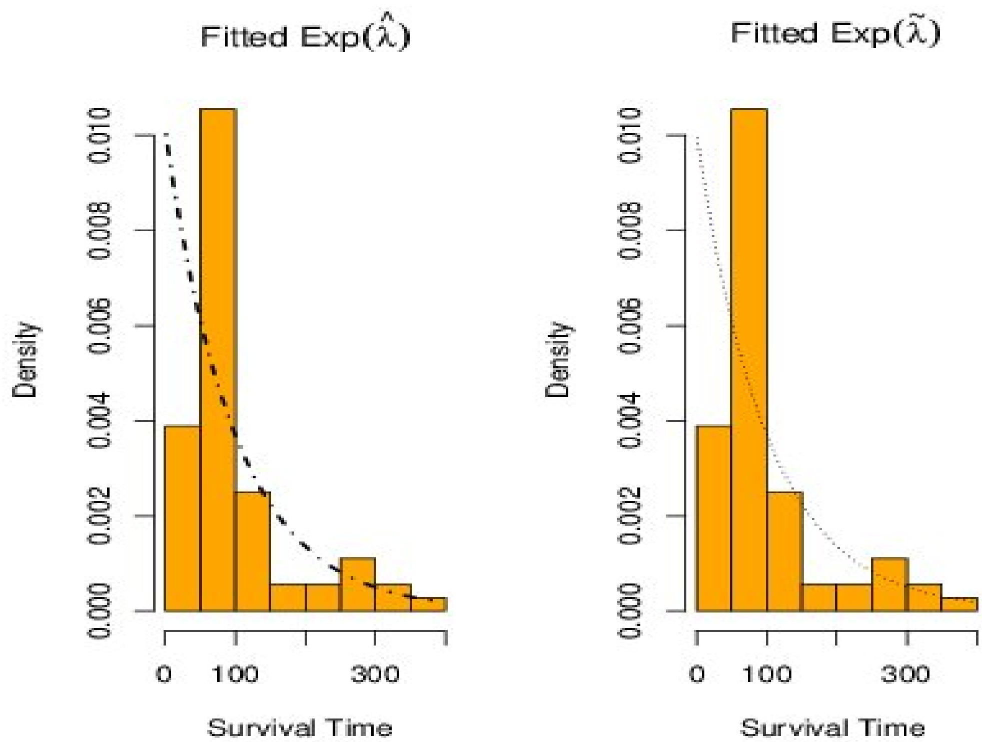

Figure 4.1: Fitted exponential distributions based on the MLE $\hat{\lambda}$ and MLqE $\tilde{\lambda}$.

We also examine, in Figure 4.2, the fitted models based on the MLE and MLqE by looking at the empirical cumulative distribution functions (c.d.f.) and QQ-plots. In the two c.d.f. plots, the dots show the empirical c.d.f. and the solid curves are the c.d.f.s of the two fitted exponential models. In the two QQ-plots, the x-axis represents the quantiles of the fitted exponential model $Exp(\hat{\lambda})$ or $Exp(\tilde{\lambda})$ and y-axis represents the empirical quantiles based on the data. From these plots we can see that the exponential model based on both MLE and MLqE fits the data equivalently well.

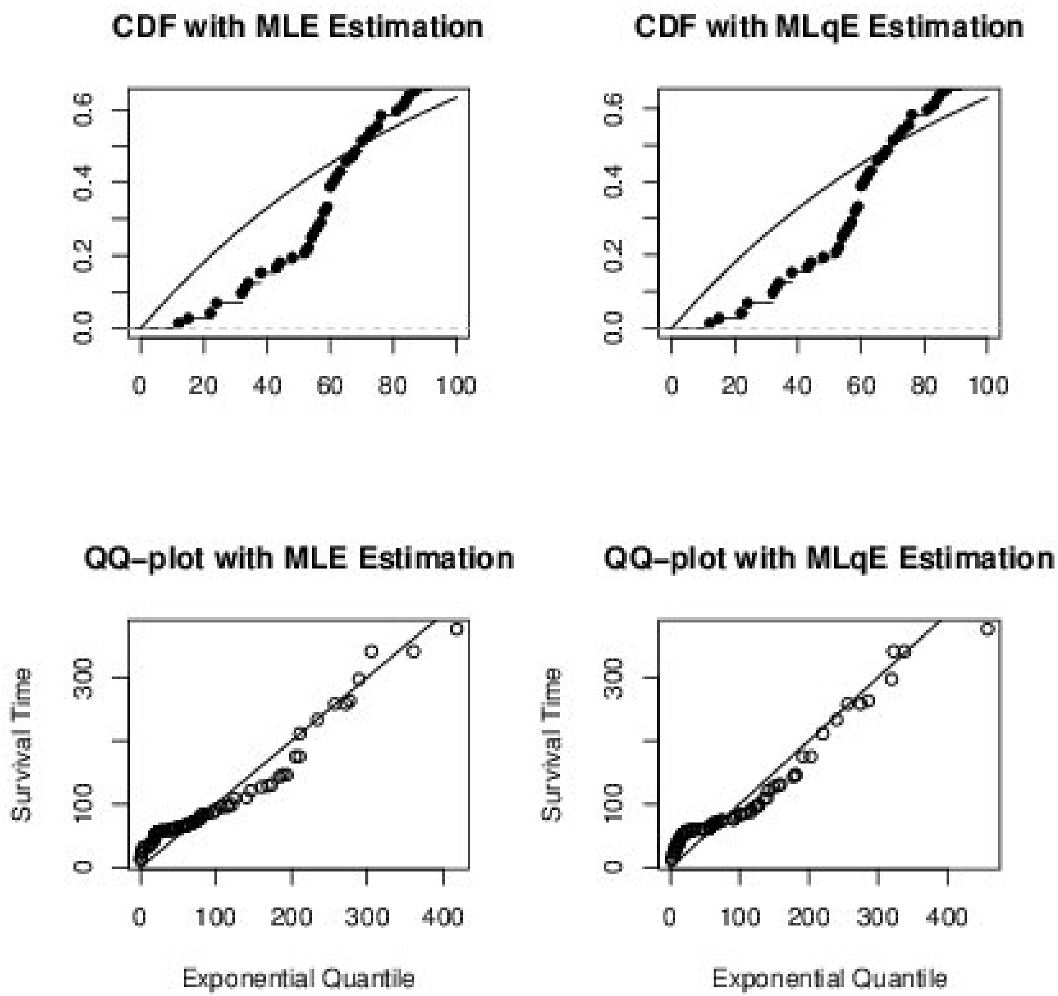

Figure 4.2: The c.d.f. and QQ-plot of the fitted exponential distributions based on the MLE $\hat{\lambda}$ and MLqE $\tilde{\lambda}$.

We also employ a simple hold-out procedure to evaluate the performance of the MLE and MLqE. We draw $B = 200$ independent subsamples of size $n^* < n$ from the original data and for each subsample compute the MLE $\hat{\lambda}_{n^*,b}$, optimal q^* and MLqE $\tilde{\lambda}_{n^*,b}$, $b = 1, \cdots, B$. Then the quadratic error of MLqE and MLE are defined respectively as

$$\varepsilon(q^*, n^*) = B^{-1} \sum_{b=1}^{B} (\tilde{\lambda}_{n^*,b} - 0.0100181)^2,$$

$$\varepsilon(1, n^*) = B^{-1} \sum_{b=1}^{B} (\hat{\lambda}_{n^*,b} - 0.0099375)^2.$$

Now the Gain(%) is defined as

$$\text{Gain}(\%) = \left(\frac{\varepsilon(1, n^*)}{\varepsilon(q^*, n^*)} - 1 \right) \times 100.$$

The Gain(%) quantifies how much more variability is in MLE than that in MLqE. The results are presented in Table 4.1. In the table, for different n^* values, we give the quadratic errors of the MLqE and MLE, the averaged optimal distortion parameter q^* over $B = 200$ subsamples, and the Gain(%).

Table 4.1: MLqE and MLE of exponential parameter for Guinea Pigs data.

n^*	10	20	30	40	50	60	70
$\varepsilon(1, n^*)$	$7.55 \cdot 10^{-6}$	$2.78 \cdot 10^{-6}$	$1.58 \cdot 10^{-6}$	$8.23 \cdot 10^{-7}$	$4.21 \cdot 10^{-7}$	$3.55 \cdot 10^{-7}$	$7.52 \cdot 10^{-8}$
$\varepsilon(q^*, n^*)$	$6.88 \cdot 10^{-6}$	$2.63 \cdot 10^{-6}$	$1.50 \cdot 10^{-6}$	$7.82 \cdot 10^{-7}$	$4.08 \cdot 10^{-7}$	$3.48 \cdot 10^{-7}$	$8.31 \cdot 10^{-8}$
q^*	1.09	1.05	1.04	1.03	1.02	1.02	1.01
Gain(%)	9.65	5.47	5.10	4.94	3.95	1.87	-9.52

From Table 4.1 we can see that the Gain(%) decreases dramatically when subsample size n^* increases. It is always positive when sample size $n^* \leq 60$ but negative when $n^* = 70$. Note that a positive Gain(%) value means the variability in MLqE is smaller than that in MLE. So when the subsample size n^* increases, the benefit of using MLqE over MLE in terms of reducing variability diminishes, which is consistent with our observation in the simulation studies in Chapter 3. The fact here that the optimal q^* is higher than 1 also agrees with our observations in Chapter 3. The analysis of this data shows that, for small to moderate sample sizes, the MLqE is superior to the MLE in terms of Gain(%) for estimating the exponential parameter λ.

4.2 Fitted Gamma Models

To demonstrate the implementation of MLqE for the standard gamma distribution, in this section we assume that the data follows a more general gamma distribution and will test this model assumption once parameters are estimated. Since the methodology discussed in this thesis assumes the scale parameter λ is known, we try different values $\lambda = 0.01, 0.02, 0.03, 0.04, \cdots, 0.10$ and pick up the one gives the best fit to the data with Kolmogorov-Smirnov test. For each fixed λ value, we

Step 1. Divide the data by λ to make it standard gamma distributed after re-scaling.

Step 2. Calculate the optimal q^* (defined in (2.28)) based on the re-scaled data.

Step 3. Calculate the MLE and MLqE (with the q^*) based on the re-scaled data.

Step 4. Use Kolmogorov-Smirnov test to test the fitted gamma model with MLE $\hat{\theta}$ (MLqE $\tilde{\theta}$) as the shape parameter and the fixed λ as the scale parameter.

Repeat Steps 1-4 for each λ value and choose the one that gives the best fit, i.e. the one that gives the largest p-value. It turns out that $\lambda^* = 0.05$ gives the best fit. The corresponding MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$ (with optimal $q^* = 1.00134$) along with their standard errors are respectively

$$\hat{\theta} = 4.341229 \ (se = 0.51155) \quad \text{and} \quad \tilde{\theta} = 4.341704 \ (se = 0.511606).$$

The standard errors are calculated by using the asymptotic formula $\sigma_n/\sqrt{n}$ with σ_n derived in (2.20), $q_n = q^* = 1.00134$ for MLqE and $q_n = 1$ for MLE. Furthermore, the 99% asymptotical confidence intervals of θ produced by MLqEs and MLEs are $(3.023807, 5.659601)$ and $(3.023476, 5.658982)$ respectively.

These two best fitted gamma models $Gamma(\hat{\theta}, \lambda^*)$ and $Gamma(\tilde{\theta}, \lambda^*)$ based on respective MLE and MLqE give the following Kolmogorov-Smirnov test statistics and p-values:

$$\hat{\theta} : ks = 0.1267,\ p\text{−value} = 0.1979 \quad \text{and} \quad \tilde{\theta} : ks = 0.1268,\ p\text{−value} = 0.1972.$$

The two high p-values (> 0.1) mean that both of the two fitted gamma models are reasonably appropriate. In addition, the two p-values are very close, which indicates that the two gamma models fit the data equivalently well.

In Figure 4.3, we examine the fitted gamma models based on the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$ along with a histogram of the survival times. As expected from the estimates, the two gamma models fit the data equivalently well. In Figure 4.4, we look at the c.d.f. of the fitted gamma model compared with the empirical c.d.f., as well as the QQ-plot based on the fitted gamma model. From Figure 4.4 we observe again that the gamma model based on both MLE and MLqE fits the data equivalently well. When compared with the fitted exponential models, the fitted gamma models seem more appropriate.

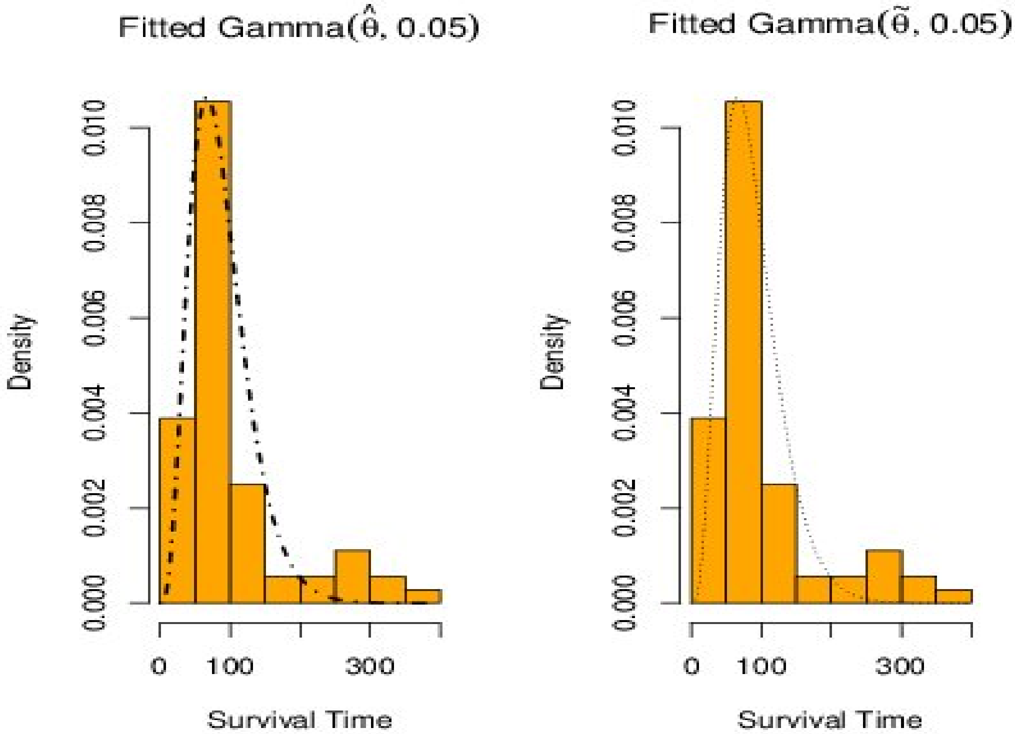

Figure 4.3: Fitted gamma distributions based on the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$.

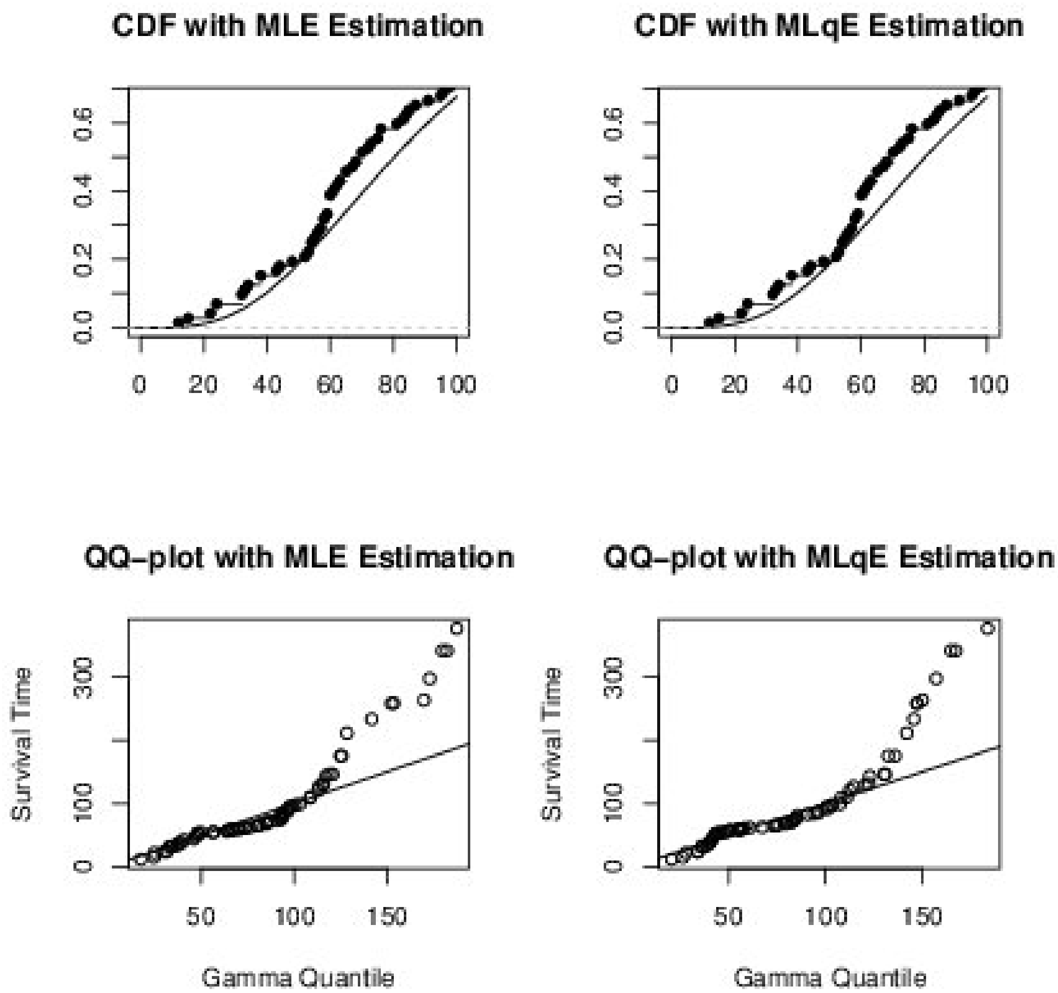

Figure 4.4: CDF and QQ-plot of the fitted exponential distributions based on the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$.

Chapter 5

CONCLUDING REMARKS

The MLqE was first introduced in Ferrari and Yang (2010). It is a novel estimation procedure based on the empirical version of Havrda-Charvát-Tsallis entropy. Different from the traditional MLE, MLqE uses a more general function $L_q(u) = (u^{1-q} - 1)/(1 - q)$ to take place of the log function in MLE. Note that when $q \to 1$, $L_q(u) \to \log(u)$ and thus one can recover the traditional MLE. As an extension of the traditional MLE, MLqE has been shown in Ferrari and Yang (2010) to be very useful and efficient for small and moderate samples. This thesis aims to examine this newly proposed MLqE in two important types of gamma distributions: the exponential distribution $Exp(\lambda)$ and the standard gamma distribution $Gamma(\theta, 1)$. Our results confirm the findings in Ferrari and Yang (2010) but with a more detailed analysis and additional robustness studies.

As special cases of the exponential family described in Ferrari and Yang (2010), the MLqEs of parameter and tail probabilities for both exponential and standard gamma distributions obeys consistency and asymptotic normality. By straightforward calculation, the asymptotic variances of the MLqEs are derived. As suggested by Ferrari and Yang (2010), we examine, via simulation studies and a real data analysis, a method of choosing the distortion parameter q by minimizing the estimated MSE that depends on the derived asymptotic variance. Our results indicate that the optimal distortion parameter q^* is always bigger than 1 for parameter estimation and smaller than 1 for tail probability estimation. Thus, the optimal choice of q^* is not a characteristic of the family alone but also depends on the parametric function to be estimated. The optimal q^* converges to 1 as the sample size increases, which indicates that the MLqE with optimal q^* is asymptotically equivalent to MLE.

With a properly chosen distortion parameter q, we examine the finite sample performance of the MLqE compared with the MLE for varying parameters and tail probabilities in the exponential and standard gamma distributions. Our simulation results demonstrate that when the sample size is small and moderate, the MLqE reduces estimator variance. When the sample size is large, the bias component becomes more relevant and the advantage of using MLqE diminishes. Moreover, the benefit of using MLqE techniques for small sample sizes is accentuated by taking a distortion parameter further away from 1 in the direction of optimal value q^*.

The robustness, particularly the resistance to outliers, is examined for MLqE. Our results show that the MLqE is generally more robust with respect to outlying observations than the MLE. Also the choice of the distortion parameter q provides enough flexibility according to the purpose of an analysis. When exponential and standard gamma models are concerned, if the concentration is on the accuracy of estimation, then one should choose the distortion parameter q in the same direction as the optimal value q^*; if the resistance to outliers is more important, then one should choose q in the same direction as q^* for tail probability estimation, which is the opposite direction for parameter estimation.

Despite of the benefits of MLqE, it has its own limitations. The computation of MLqE is a bit more complicated than the traditional MLE, especially with the calculation of the optimal distortion parameter q^*. Also one should not expect the same benefit when using MLqE for different models. As shown in our simulation results, the MLqE of the parameter for the standard gamma model is not as beneficial as that for the exponential model (Figures 3.4-3.6 compared with Figures 3.1-3.3).

As a generalization of the study in this thesis, we may consider the general gamma model. In the general gamma model, there are two unknown parameters and this two dimensional parameter may make issues more complex both analytically and numerically. We may also consider gamma regression model. A second direction of future study is to generalize the MLqE method to nonparametric or even semiparametric models. As a third direction of future research, we would like to find the optimal q^* value by maximizing Lq-likelihood function over both parameter space and q simultaneously, i.e. the maximization in (2.1) with respect to both θ and q instead of θ only. This maximizing procedure could be implemented by using profile likelihood technique.

Bibliography

[1] Aczél, J.D. and Daróczy, Z. (1975). On measures of information and their characterizations. *Mathematics in Science and Engineering* **115**, 26-49 & 71-82.

[2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium of Information Theory* **2**, 267-281.

[3] Altun, Y. and Smola, A. (2006). Unifying divergence minimization and statistical inference via convex duality. *In Learning Theory. Lecture Notes in Computer Science* **4005**, 139-153.

[4] Barron, A., Rissanen, J. and Yu, B. (1998). The minimum description length principle in coding and modelling. *IEEE Transactions on Information Theory*, **44**, 2743-2760.

[5] Beran, R. (1977). Minimum Hellinger distance estimators for parametric models. *The Annals of Statistics* **5**, 445-463.

[6] Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Public Hygiene* **72**, 130-148.

[7] Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Chapman & Hall, London.

[8] Ferrari, D. and Paterlini, S. (2009). The Maximum Lq-Likelihood Method: An Application to Extreme Quantile Estimation in Finance. *Methodology and Computing in Applied Probability* **11**, 3-19.

[9] Ferrari, D. and Yang, Y. (2010). Maximum Lq-Likelihood Estimation. *The Annals of Statistics* **38**, 753-783.

[10] Havrda, J. and Charvát, F. (1967). Quantification method of classification processes: Concept of structural entropy. *Kibernetika* **3**, 30-35.

[11] Huang, C., Lin, J. and Ren, Y.Y. (2013). Testing for the shape parameter of generalized extreme value distribution based on the Lq-likelihood ratio statistic. *Metrika* **76**, 641-671.

[12] Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

[13] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York.

[14] Lu, Z., Hui, Y.V. and Lee, A.H. (2003). Minimum Hellinger distance estimation for finite mixtures of Poisson regression models and its applications. *Biometrics* **59**, 1016-1026.

[15] Jaynes, E.T. (1957a). Information theory and statistical mechanics. *Physical Review* **106**, 620-630.

[16] Jaynes, E.T. (1957b). Information theory and statistical mechanics II. *Physical Review* **108**, 171-190.

[17] Qin, Y. and Priebe, E.C. (2013). Maximum Lq-Likelihood Estimation via the Expectation Maximization Algorithm: A Robust Estimation of Mixture Models. *Journal of the American Statistical Association* **108**, 914-928.

[18] Rényi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**, 547-461.

[19] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics* **52**, 479-487.

[20] Tsallis, C., Mendes, R.S. and Plastino, A.R. (1998). The role of constraints within generalized nonextensive statistics. *Physica A: Statistical Mechanics and its Applications* **261**, 534-554.