Maximum Lq-likelihood Estimation for Gamma Distributions

Jingjing Wu[1][*], Nana Xing[1] and Shawn Liu[2]

[1]Department of Mathematics and Statistics, University of Calgary, Calgary, Canada
[2]Department of Mathematics and Computing, Mount Royal University, Calgary, Canada
Email: `jinwu@ucalgary.ca`

Abstract In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. Standard large sample theory guarantees asymptotic efficiency of MLE. On the other hand, MLE does not perform as well as expected for moderate or small sample size. In 2010, a new parameter estimator based on nonextensive entropy ([1]), named Maximum Lq-likelihood Estimator (MLqE), was first introduced and studied by [2]. MLqE is an extension of MLE which introduces a distortion parameter q to make the estimation more adaptive. The purpose of this work is to examine this methodology for gamma distributions that are widely used in engineering, science and business to model continuous but skewed distributions. For specifically standard gamma models, we look at the MLqE's asymptotics, finite sample performance in terms of efficiency and robustness, and the choice of the distortion parameter q. We investigate these aspects of MLqE and compare it with MLE in parameter estimation and tail probability estimation, through both Monte Carlo simulation and a real data analysis. Our results show that, with appropriately chosen q, MLqE and MLE perform competitively for large sample sizes while MLqE outperforms MLE for small or moderate sample sizes in terms of reducing MSE. In addition, MLqE with $q < 1$ has much better robustness properties than MLE when outlying observations are present.

Keywords: Maximum Lq-likelihood estimation, gamma distribution, distortion parameter, efficiency, robustness.

1 Introduction

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 a 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 [3] 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 [4]. 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 [5] and [6].

Article [1] 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 [7] and [8]. Articles [9] and [10] successfully exploited q-order entropies in physics. In statistics, [11] concluded that the classical maximum entropy estimation and MLE are convex duals of each other.

As an alternative parametric estimation to MLE, the Maximum Lq-likelihood Estimator (MLqE) was first introduced by [2] and it is based on the nonextensive q-order entropy function. 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

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

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 $\tilde{\theta}$ 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.

Article [2] 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, [2] 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. [2] gave a good discussion on the efficiency of MLqE but left the robustness properties undiscovered. To fill the gap, in this paper we will discuss the robustness properties of MLqE.

To our knowledge, there are only a few papers on MLqE so far. Article [12] introduced a MLqE for mixture models using their proposed expectation-maximization (EM) algorithm, namely the EM algorithm with Lq-likelihood (EM-Lq). Article [13] applied MLqE to estimate quantiles of the Generalized Extreme Value (GEV) and the Generalized Pareto (GP) distributions in finance. In [14], 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.

In this paper, we will investigate the MLqE of parameters and tail probabilities for gamma distributions. Gamma distributions are widely used in engineering, science and business to model continuous but skewed distributions. For example, it is well known that gamma distribution is commonly used to model failure time. We will examine both the MLqE's asymptotic properties and its finite sample performance. We will not only look at its efficiency but more importantly its robustness properties. In addition, we will discuss the choice of distortion parameter q in terms of direction and value. We will investigate these aspects of MLqE and compare it with MLE in parameter estimation and tail probability estimation, through both Monte Carlo simulation and a real data analysis. Our results show that MLqE and MLE perform competitively for large sample sizes while MLqE outperforms MLE for small or moderate sample size in terms of reducing MSE. More promisingly, MLqE when $q < 1$ has much better robustness properties than MLE when outlying observations are present.

This paper is organized as follows. In Section 2, we will review the properties of MLqE for exponential families and further, for a special class of gamma distributions, derive the MLqE of the parameter and its asymptotic variance. We will also discuss a method of choosing the distortion parameter q based on the MSE. In Section 3, we will implement Monte Carlo simulation studies to examine the proposed MLqE's finite sample performance in terms of both efficiency and robustness, the accuracy of the MLqE-based confidence intervals, and the choice of the distortion parameter q. In Section 4, we demonstrate how to implement the MLqE through a real data analysis of the Guinea Pigs data. Final concluding remarks are presented in Section 5.

2 MLqE of Standard Gamma Distributions

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 in (1). Define

$$\begin{aligned} U(x; \boldsymbol{\theta}) &= \nabla_{\boldsymbol{\theta}} \log\{f(x; \boldsymbol{\theta})\}, \\ U^*(x; \boldsymbol{\theta}, q) &= U(x; \boldsymbol{\theta}) f^{1-q}(x; \boldsymbol{\theta}), \end{aligned} \tag{2}$$

where $\nabla_{\boldsymbol{\theta}}$ denotes the operator of first derivative with respect to $\boldsymbol{\theta}$. Then in general, the estimating equation for the MLqE $\tilde{\boldsymbol{\theta}}$ solves

$$\sum_{i=1}^{n} U^*(X_i; \boldsymbol{\theta}, q) = 0. \tag{3}$$

Equation (3) offers a natural interpretation of the MLqE as a solution to a weighted likelihood. When $q \neq 1$, (3) 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.

Consider density functions of the exponential family

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

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). Article [2] has studied this exponential family and established the consistency and asymptotic normality of the MLqE. In this paper we will examine closely gamma distributions which are special cases of this exponential family. More specifically, we are interested in 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 loss of generality, we assume the scale parameter is 1; i.e. we consider estimating the shape parameter of standard gamma distributions. We will derive the MLqE of the shape parameter and its asymptotic variance.

2.1 MLqE of the Parameter and Tail Probability

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 (3) is reduced to

$$\sum_{i=1}^{n} e^{[(\theta-1)\log X_i - X_i](1-q)} \left[\Gamma(\theta) \log X_i - \Gamma'(\theta) \right] = 0. \tag{4}$$

The MLqE $\tilde{\theta}$ of θ is the solution to (4). With $q = 1$, the MLE $\hat{\theta}$ is the solution to equation

$$\psi(\theta) = \left(\sum_{i=1}^{n} \log X_i \right) / n, \tag{5}$$

where $\psi(\theta) = \Gamma'(\theta)/\Gamma(\theta)$. With a lengthy but straightforward calculation, we obtain the variance of the MLqE $\tilde{\theta}$ as

$$\sigma_n^2 = \frac{\frac{\Gamma((2/q_n - 1)\theta)}{\Gamma(\theta)} \left\{ \psi'((2/q_n - 1)\theta) + [\psi((2/q_n - 1)\theta) - \psi(\theta/q_n)]^2 \right\}}{\left[\frac{q_n \Gamma(\theta/q_n)}{\Gamma(\theta)} \psi'(\theta/q_n) \right]^2} \tag{6}$$
$$\longrightarrow [\psi'(\theta)]^{-1} = \{[\log \Gamma(\theta)]''\}^{-1}, \quad \text{as } n \to \infty,$$

where q_n is chosen such that $q_n \to 1$ in order to achieve asymptotic efficiency. By Theorem 3.2 in [2], we can conclude that $n^{1/2} \sigma_n^{-1}(\tilde{\theta} - \theta/q_n)$ converges weakly to a standard normal distribution as $n \to \infty$. Note that $\{[\log \Gamma(\theta)]''\}^{-1}$ is the asymptotic variance of the MLE.

In fields such as finance, people are more interested in the tail probability estimation. Denote the upper tail probability of standard gamma distribution by

$$\alpha(x; \theta) = P_\theta(X \leq x) = 1 - \int_0^x \frac{y^{\theta-1} e^{-y}}{\Gamma(\theta)} dy, \quad x > 0. \tag{7}$$

Note that one can also consider the lower tail probability, but without loss of generality we only consider the upper tail probability. Based on the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$, the MLE and plug-in MLqE of the upper tail probability $\alpha(x; \theta)$ are $\hat{\alpha} = \alpha(x; \hat{\theta})$ and $\tilde{\alpha} = \alpha(x; \tilde{\theta})$ respectively.

Assume $\alpha(x; \theta) > 0$ for all x. Obviously $\alpha(x; \theta) \to 0$ 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 behaviour 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. Thus here we are interested in estimating $\alpha(x_n; \theta)$, where $x_n \to \infty$ as $n \to \infty$.

2.2 Choice of Distortion Parameter q

When estimating the parameter in standard gamma distributions, 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.

As [2] suggested, one can choose the q which minimizes an estimated asymptotic MSE of the estimator when it is mathematically tractable. Article [2] discussed the choice of q for exponential distributions. Now we look at how to choose q for standard gamma distributions. By Theorem 3.2 in [2], the MLqE $\tilde{\theta}$ of θ has asymptotic MSE

$$\text{MSE}_{\tilde{\theta}}(q, \theta) = \left(\frac{\theta}{q} - \theta \right)^2 + \frac{\sigma_n^2}{n}, \tag{8}$$

where σ_n is given in (6) with q_n replaced by q. As a result, when estimating θ, we choose q^* such that

$$q^* = \underset{q \in (0,2)}{\arg \min} \{ \text{MSE}_{\tilde{\theta}}(q, \hat{\theta}) \}, \tag{9}$$

where $\hat{\theta}$ is the MLE.

We use the same way to choose distortion parameter p when estimating tail probabilities. By Theorem 5.1 of [2], the plug-in MLqE $\alpha(x_n; \tilde{\theta})$ of the upper tail probability $\alpha(x_n; \theta)$, given in (7), has asymptotic MSE

$$\text{MSE}_{\tilde{\alpha}}(q, \theta) = [\alpha(x_n; \theta/q) - \alpha(x_n; \theta)]^2 + [\alpha'(x_n; \theta/q)]^2 \frac{\sigma_n^2}{n}, \tag{10}$$

where σ_n is given in (6) with q_n replaced by q, 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. The asymptotic MSE in (10) can be written more explicitly as

$$\text{MSE}_{\tilde{\alpha}}(q, \theta) = \left[\int_0^{x_n} \frac{y^{\theta-1} e^{-y}}{\Gamma(\theta)} dy - \int_0^{x_n} \frac{y^{\theta/q-1} e^{-y}}{\Gamma(\theta/q)} dy \right]^2 + \left[\int_0^{x_n} \left(\frac{\partial}{\partial \theta} \frac{y^{\theta-1} e^{-y}}{\Gamma(\theta)} \right) \bigg|_{\theta=\theta/q} dy \right]^2 \frac{\sigma_n^2}{n}. \tag{11}$$

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

$$q^* = \underset{q \in (0,2)}{\arg \min} \{ \text{MSE}_{\tilde{\alpha}}(q, \hat{\theta}) \}, \tag{12}$$

where $\hat{\theta}$ is the MLE. The choices of q in (9) and (12) will be also used in some of our simulation studies in Section 3.

3 Simulation Studies

In this section we will implement an extensive Monte Carlo simulation study to examine the MLqE for standard gamma distributions. We will compare their finite sample performance with that of the traditional MLE. In Section 3.1, we assess the accuracy of MLqE and MLE for both the 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 MLqE and MLE with data-driven optimal distortion parameter. Section 3.3 is devoted to a robustness study of the MLqE and MLE.

3.1 Relative Efficiency

In the first group of simulations, we compare the two estimators of the natural parameter in standard gamma distributions, obtained via the MLq method and the traditional ML approach respectively. Particularly, we are interested in assessing the relative performance of MLqE over MLE for different choices of sample size by taking the ratio of their MSEs, i.e. $\text{MSE}(\hat{\theta})/\text{MSE}(\tilde{\theta})$ for θ and $\text{MSE}(\hat{\alpha})/\text{MSE}(\tilde{\alpha})$ for upper tail probability α.

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 a standard gamma distribution with parameter θ (i.e. $Gamma(\theta, 1)$).

(ii) For each sample, the MLqEs and MLEs of θ and α are calculated.

(iii) For each sample size n, the relative performance of MLqE over MLE is evaluated by the ratio $\hat{R} = \text{MSE}_{\text{MC}}(\hat{\beta})/\text{MSE}_{\text{MC}}(\tilde{\beta})$, where β denotes θ or α and MSE_{MC} denotes the Monte Carlo estimate of the MSE.

To find $\hat{R}$ in (iii), let $\bar{y}_1 = B^{-1} \sum_{k=1}^{B} (\hat{\beta}_k - \beta)^2$ and $\bar{y}_2 = B^{-1} \sum_{k=1}^{B} (\tilde{\beta}_k - \beta)^2$, where $\hat{\beta}_k$ and $\tilde{\beta}_k$ denote the estimates based on the k-th sample, $k = 1, 2, \cdots, B$. Then we use $\hat{R} = \bar{y}_1/\bar{y}_2$ and the rationale is as follows. By the Central Limit Theorem (CLT), when $B \to \infty$, $\bar{\boldsymbol{y}} = (\bar{y}_1, \bar{y}_2)^T$ has an asymptotic bi-variate normal distribution with mean $\boldsymbol{\mu} = (\text{MSE}(\hat{\beta}), \text{MSE}(\tilde{\beta}))^T$ and a certain covariance matrix $\boldsymbol{\Gamma}$, i.e.

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

Thus we could use $\bar{y}_1/\bar{y}_2$ to estimate the relative performance of MLqE over MLE. To estimate the standard error of $\hat{R}$, 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 estimate the standard error of $\hat{R}$ by

$$se(\hat{R}) = B^{-1/2} \left(\frac{\hat{\gamma}_{11}}{\bar{y}_2^2} - 2\hat{\gamma}_{12}\frac{\bar{y}_1}{\bar{y}_2^3} + \hat{\gamma}_{22}\frac{\bar{y}_1^2}{\bar{y}_2^4} \right)^{1/2}, \tag{13}$$

where $\hat{\gamma}_{11}$, $\hat{\gamma}_{22}$ and $\hat{\gamma}_{12}$ denote the Monte Carlo estimates of the components of the covariance matrix $\boldsymbol{\Gamma}$.

Study I: fixed θ and q for parameter estimation

In Figure 1, we consider $\theta = 1, 5, 10$ and fixed distortion parameter $q = 1.5$. Figure 1 shows that $\hat{R} > 1$ for very small sample sizes, i.e., $n < 10$, and $\theta = 5, 10$. When $\theta = 1$, $\hat{R} < 1$ for any sample size. When the sample size increases, $\hat{R}$ is decreasing, $\hat{R} < 1$ 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 2 considers fixed $\theta = 10$ and three different distortion parameter values $q = 1.35, 1.15, 1.05$. From this plot we observe that $\hat{R} > 1$ for very small sample sizes, i.e. $n < 10$. When sample size n increases, the $\hat{R}$ 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.

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

Figure 3 considers $\theta = 1, 5, 10$ and varying $q_n = [1 + e^{0.3(n-20)}]/[0.5 + e^{0.3(n-20)}]$. When $\theta = 5, 10$, $\hat{R} > 1$ for very small sample sizes $n < 7$ and $\hat{R} < 1$ for sample sizes $n > 10$. When $\theta = 1$, $\hat{R} < 1$ for all
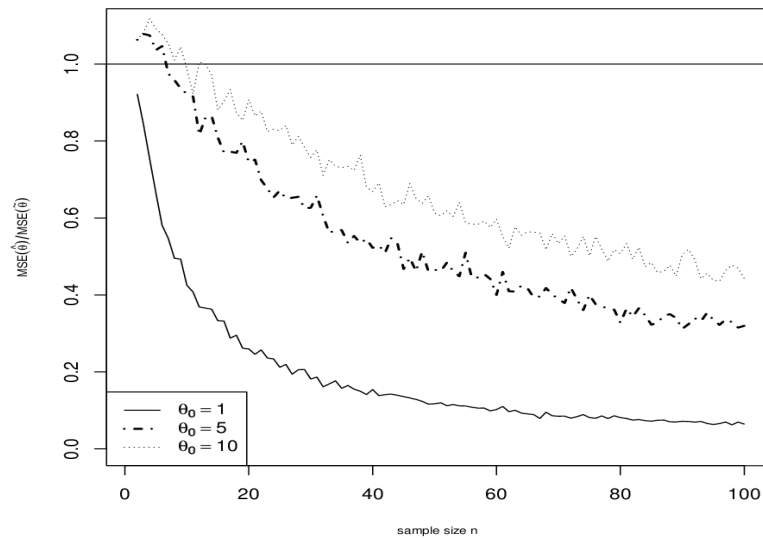
Figure 1. Plot of MSE ratio $\hat{R} = \mathrm{MSE}_{\mathrm{MC}}(\hat{\theta})/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\theta})$ as a function of sample size n with $\theta = 1, 5, 10$ and $q = 1.5$ (Study I).
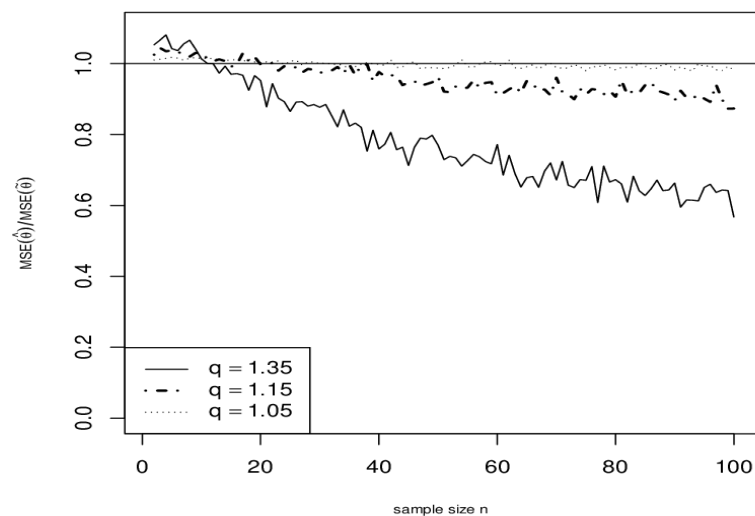

Figure 2. Plot of MSE ratio $\hat{R} = \mathrm{MSE}_{\mathrm{MC}}(\hat{\theta})/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\theta})$ as a function of sample size n with $\theta = 10$ and $q = 1.35, 1.15, 1.05$ (Study I).

sample sizes. For large sample sizes, e.g. $n > 30$, q_n converges quickly to 1 and thus $\hat{R}$ converges to 1 and the MLqEs and MLEs perform equivalently.

Study III: fixed α and q for tail probability estimation

Figures 4 and 5 illustrate the behaviour of $\hat{R}$ for different choices of q and the tail probability α. For relatively small sample sizes, i.e. $n < 10$ for $q = 0.5$ in Figure 4 and $n < 20$ for $q = 0.65, 0.85, 0.95$ in Figure 5, we observe $\hat{R} > 1$ which means that the MLqE performs better than the MLE. Such behaviour
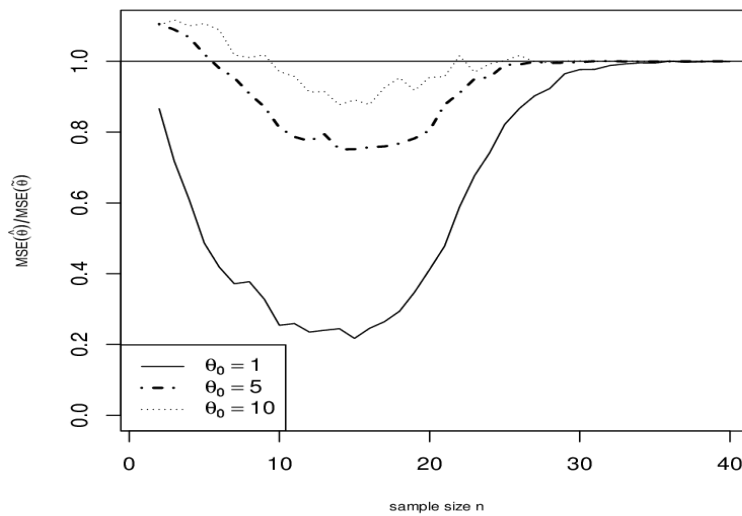
Figure 3. Plot of MSE ratio $\hat{R} = \mathrm{MSE}_{\mathrm{MC}}(\hat{\theta})/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\theta})$ as a function of sample size n with $\theta = 1, 5, 10$ and $q_n = [1 + e^{0.3(n-20)}]/[0.5 + e^{0.3(n-20)}]$ (Study II).

is more accentuated for smaller values of α and smaller values of distortion parameter q. In contrast, when the sample size is larger, the bias plays an increasingly relevant role and we observe that $\hat{R} < 1$.
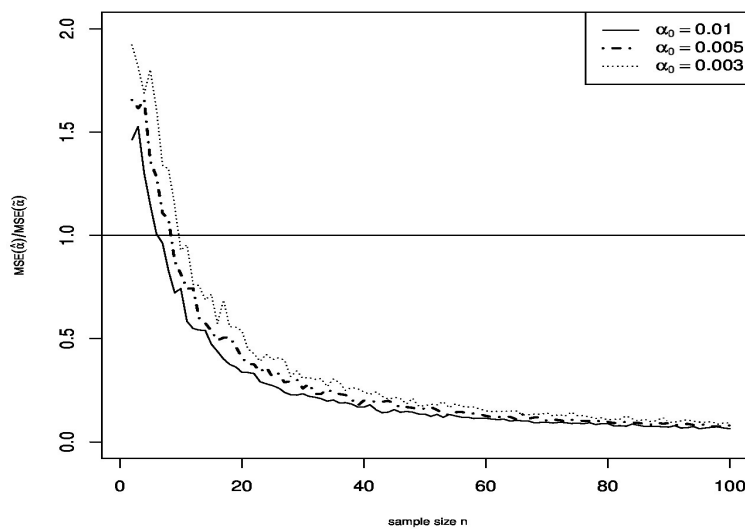

Figure 4. Plot of MSE ratio $\hat{R} = \mathrm{MSE}_{\mathrm{MC}}(\hat{\alpha})/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\alpha})$ as a function of sample size n with $\alpha = 0.01, 0.005, 0.003$ and $q = 0.5$ (Study III).

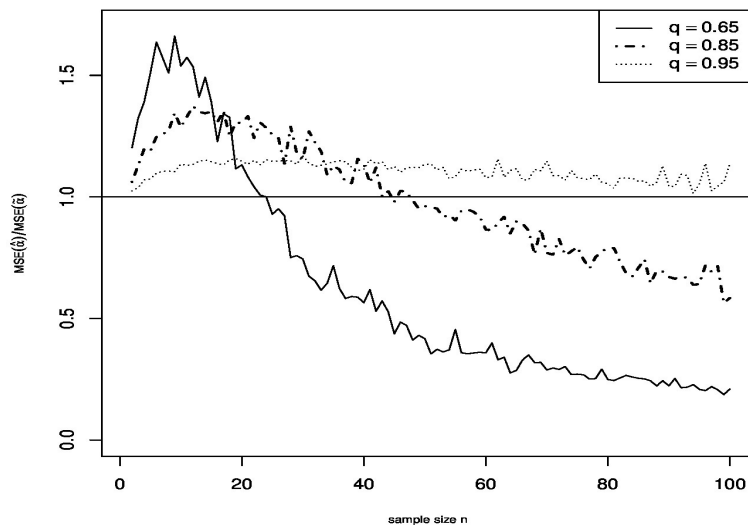**Study IV: fixed α and $q_n \nearrow 1$ for tail probability estimation**

Figure 5. Plot of MSE ratio $\hat{R} = \mathrm{MSE}_{\mathrm{MC}}(\hat{\alpha})/\mathrm{MSE}_{\mathrm{MC}}(\tilde{\alpha})$ as a function of sample size n with $\alpha = 0.003$ and $q = 0.65, 0.85, 0.95$ (Study III).

Figure 6 considers fixed upper tail probabilities $\alpha = 0.01, 0.005, 0.003$ 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 sample sizes, 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, $\hat{R} > 1$ 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.

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

Figure 7 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 values and smaller q values for larger α 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 [2]. In Figure 7, 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 4 and 6, better performance of the MLqE.

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 different methods: (a) asymptotic formula; (b) nonparametric bootstrap; (c) parametric bootstrap. We compare the results with those obtained using the MLE.

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. More specifically, for each sample we first compute the MLE $\hat{\theta}$. Then substitute it into (9) and (12) and solve them numerically in order to obtain the optimal q^* for MLqEs of θ and α
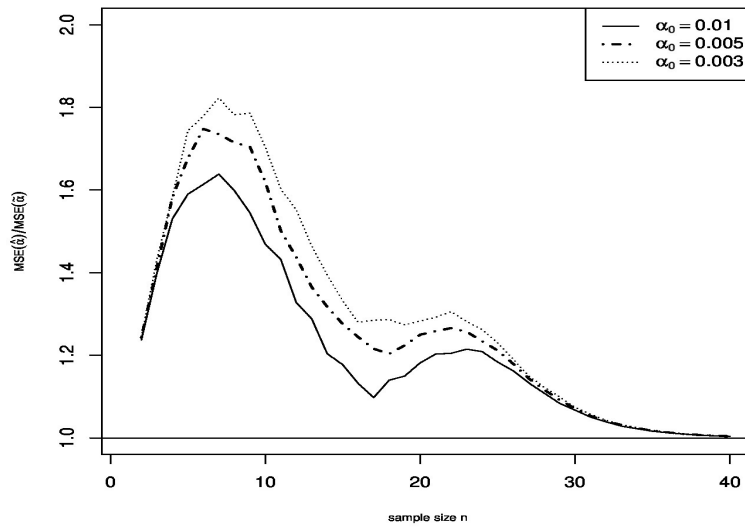
Figure 6. Plot of MSE ratio $\hat{R} = \text{MSE}_{\text{MC}}(\hat{\alpha})/\text{MSE}_{\text{MC}}(\tilde{\alpha})$ as a function of sample size n with $\alpha = 0.01, 0.005, 0.003$ and $q_n = [0.5 + e^{0.3(n-20)}]/[1 + e^{0.3(n-20)}]$ (Study IV).
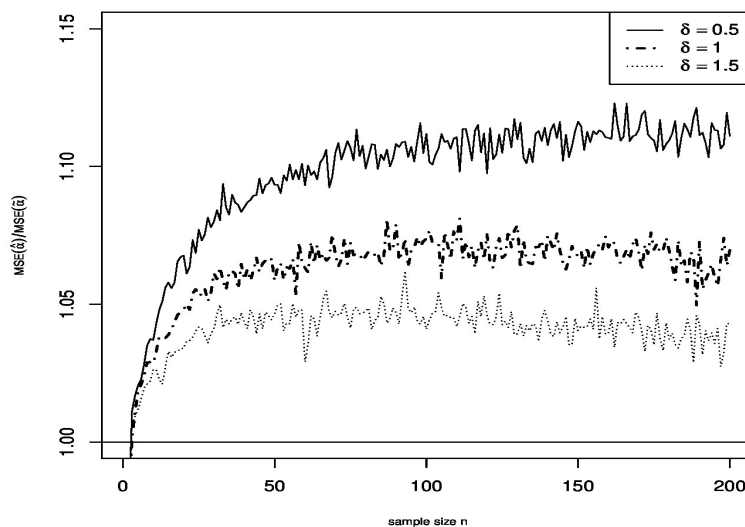

Figure 7. Plot of MSE ratio $\hat{R} = \text{MSE}_{\text{MC}}(\hat{\alpha})/\text{MSE}_{\text{MC}}(\tilde{\alpha})$ as a function of sample size n with $x_n = n^{1/(2+\delta)}$, $\delta = 0.5, 1.0, 1.5$, and $q_n = 1 - [10\log(n + 10)]^{-1}$ (Study V).

respectively. With the optimal q^*, the MLqE $\tilde{\theta}$ and $\tilde{\alpha}$ are obtained. The standard errors of the estimates are computed using the above three methods (a), (b) and (c). The standard error based on asymptotic formula is derived in either (6) ($\sigma_n/\sqrt{n}$) for estimating θ or (11) (square-root of the second term) for estimating α. In (6) and (11), $q = 1$ corresponds to the MLE. We take $B = 1,000$ repetitions for each simulation. The number of replicates employed in bootstrap re-sampling is 500. Without loss of generality, we take $\theta = 1$ and $\alpha = 0.01$. We use the level 95% for the confidence intervals and check the coverage of the true values θ and α.

In Table 1, we present the means, standard deviations and standard errors of the MLE $\hat{\theta}$ and the MLqE $\tilde{\theta}$. From Table 1 we observe that the optimal q^* approaches 1 as 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 followed by nonparametric bootstrap.

Table 1. 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 2, we compare the accuracy of 95% confidence intervals for MLE and MLqE of θ 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 2 we observe that the coverage probability of MLqE is always larger than that of MLE when either nonparametric bootstrap or parametric bootstrap is used, while it is smaller than that of MLE (within 1%) when the asymptotic formula is used. Regardless of the sample size and the 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 sizes considered, the parametric bootstrap provides most accurate confidence interval followed by nonparametric bootstrap.

In Tables 3 and 4 we present the results for tail probability estimation. From Table 3 we observe a similar phenomena to that in Table 1 except that the optimal q^* is always smaller than 1 here regardless of sample size, which is also consistent with our choice in Section 3.1. In Table 4, we compare the accuracy of 95% confidence intervals for the MLE and the MLqE of the upper tail probability α. From Table 4 we observe a similar phenomena to that in Table 2.

3.3   Robustness Study

In this section, we will examine the robustness properties of MLqE. Particularly, we will look at whether the MLqE is resistant to outlying observations. For simplicity we will examine how MLqE behaves when a single outlying observation is present, but the results could be generalized to multiple outlying observations. For this purpose, the a-influence function (IF) given in [15] is a suitable measurement. The a-IF measures the change in the estimate when a component with probability a is added to the original model. Here we use an adapted version of the a-IF proposed by [16].

In this section, we always use the sample size $n = 20$ and, without loss of generality, we take the true parameters $\theta = 1$ and $\alpha = 0.01$. We randomly select a sample of $n = 20$ observations from the population, based on which we calculate the MLE and MLqE with a fixed q value. To calculate the a-IF, we replace

Table 2. Coverage rate and relative length of interval (RL) for the MLqE $\tilde{\theta}$ over MLE $\hat{\theta}$.

n	q^*	Asympt.		Boot.		Par. boot.	
		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	

Table 3. 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

Table 4. Coverage rate and relative length of interval (RL) for the MLqE $\tilde{\alpha}$ over MLE $\hat{\alpha}$.

n	q^*	Asympt.		Boot.		Par. boot.	
		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	

the last observation (data are not sorted) with an outlying observation x, where x is an integer varying from 1 to 20. Thus the contamination rate is $a = 1/n = 1/20$. Now the a-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) based on the data. In our simulation, W is either the MLEs $\hat{\theta}$ and $\hat{\alpha}$ or MLqEs $\tilde{\theta}$ and $\tilde{\alpha}$.

For the fixed distortion parameter, we consider both cases of $q < 1$ and $q > 1$. For $q < 1$ we choose $q = 0.8$ for estimating θ and $q = 0.95$ for estimating α, while for $q > 1$ we choose $q = 1.5$ and $q = 1.1$ respectively. The a-IFs are averaged over $B = 100$ repetitions and results are presented in Figures 8 - 11. From Figures 8 and 9 ($q < 1$) we observe that MLqE performs much better than MLE in the sense that the a-IF of MLqE increases mildly, keeps the same level or even decreases as outlying observation increases, while that of MLE increases dramatically. However, in Figures 10 and 11 ($q > 1$), the a-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. Therefore, the MLqE is much more robust than MLE with respect to outlying observations when $q < 1$ and perform competitively when $q > 1$.

To conclude this section, we believe that the choice of distortion parameter q provides good flexibility according to the purpose of an analysis. When 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 $q < 1$.

4   Real Data Analysis

In this section we will use MLqE to analyze a real data set, the Guinea Pigs data. This data was presented in [17] 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.
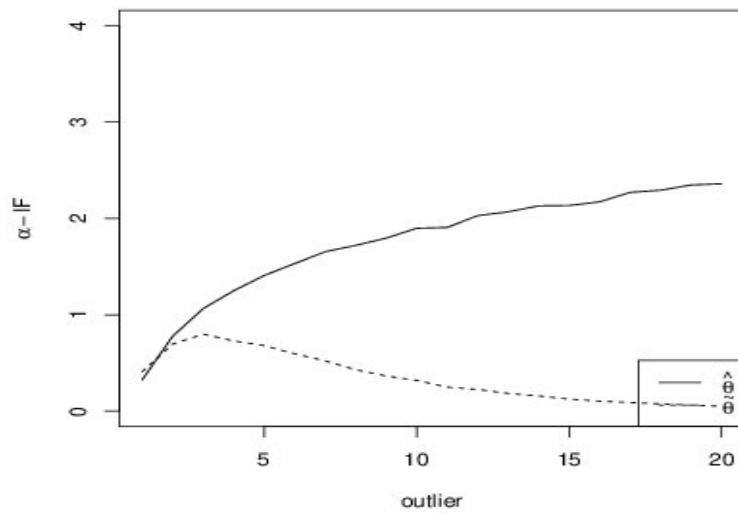
Figure 8. The α-IF of MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$ with $q = 0.8$.
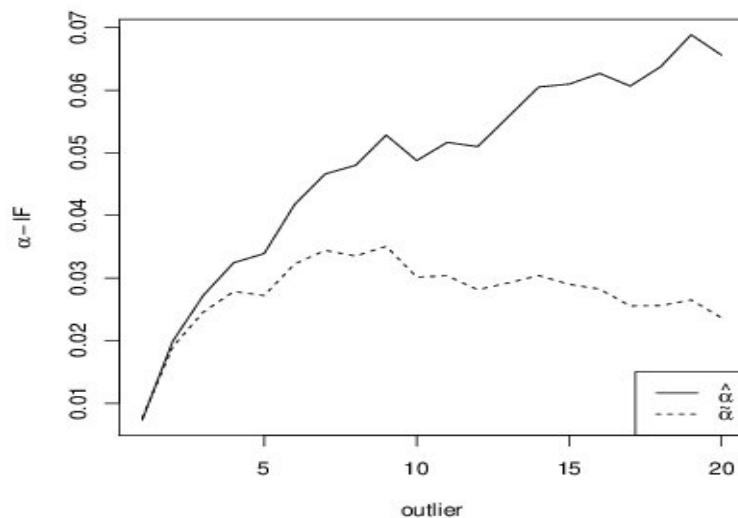

Figure 9. The α-IF of MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$ with $q = 0.95$.

Since the methodology discussed in this paper 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 (9)) 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 the Kolmogorov-Smirnov test to test the fitted gamma model with MLE $\hat{\theta}$ or MLqE $\tilde{\theta}$ as the shape parameter and the fixed λ as the scale parameter.
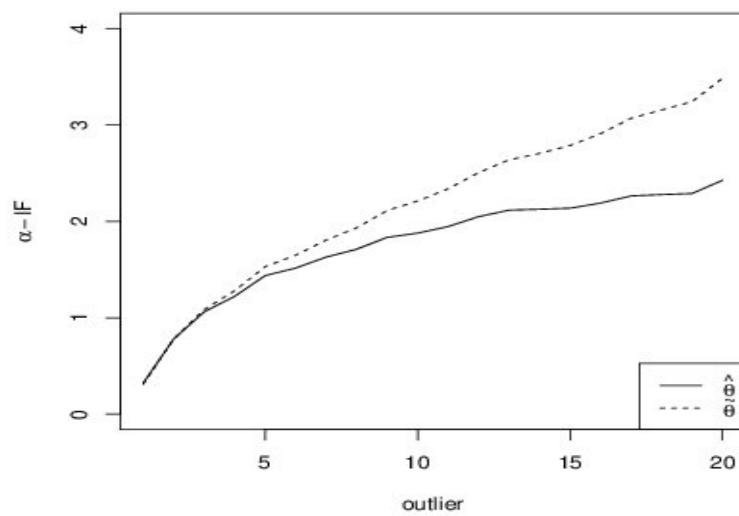
Figure 10. The α-IF of MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$ with $q = 1.5$.
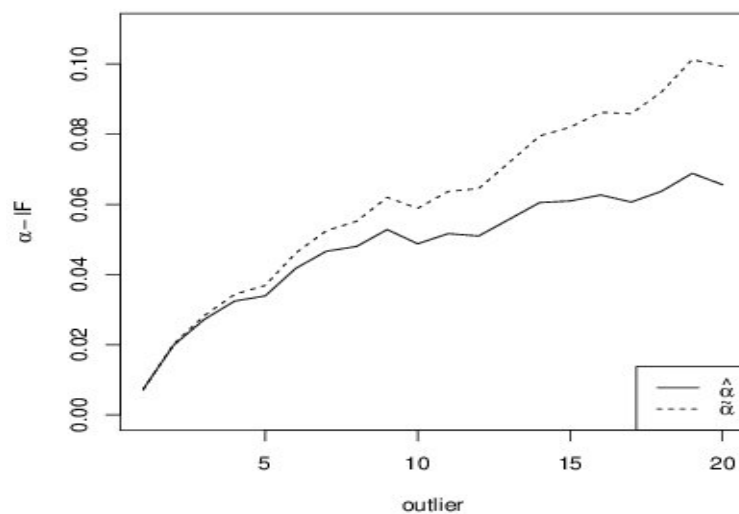

Figure 11. The α-IF of MLE $\hat{\alpha}$ and MLqE $\tilde{\alpha}$ with $q = 1.1$.

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 (6), $q_n = q^* = 1.00134$ for MLqE and $q_n = 1$ for MLE. Furthermore, the 99% asymptotical confidence intervals of θ produced by MLqE and MLE 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 12, 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 13, we look at the c.d.f.s of the fitted gamma models compared with the empirical c.d.f., as well as the QQ-plots based on the fitted gamma models. From Figure 13 we observe again that the gamma model based on both MLE and MLqE fits the data equivalently well.
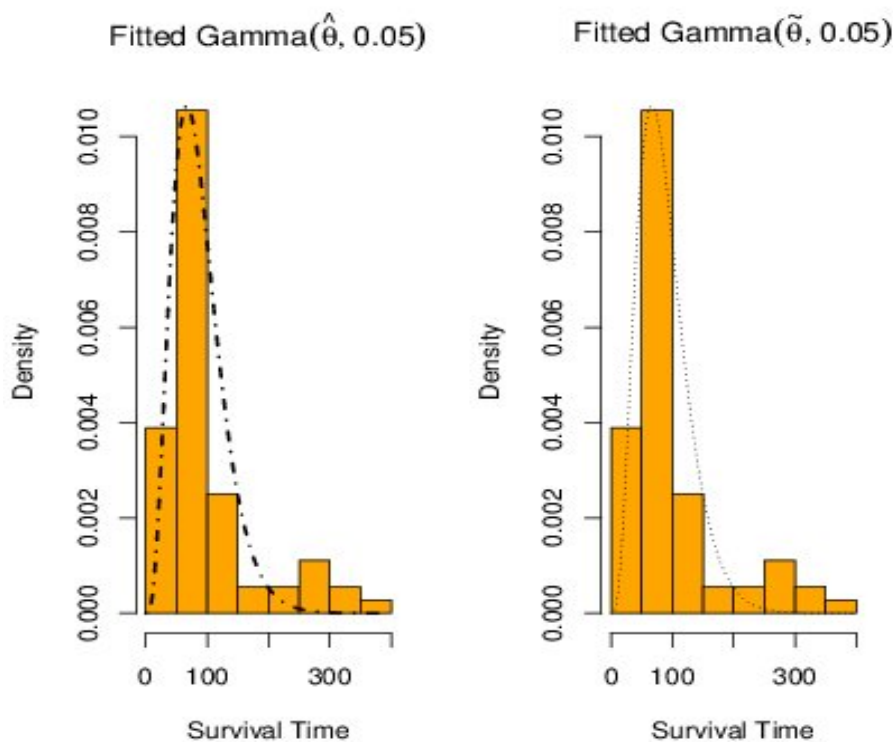

Figure 12. Fitted gamma distributions based on the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$.

5   Concluding Remarks

The MLqE was first introduced in [2]. As an extension of the traditional MLE, MLqE has been shown to be very useful and efficient for small and moderate samples. Our work aims to examine this newly proposed MLqE for gamma distributions. Our results are consistent with the findings in [2]. More significantly, we give the very first time the robustness study of MLqE and demonstrate that MLqE is much more robust than MLE when distortion parameter $q < 1$.

For standard gamma distributions, 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. The
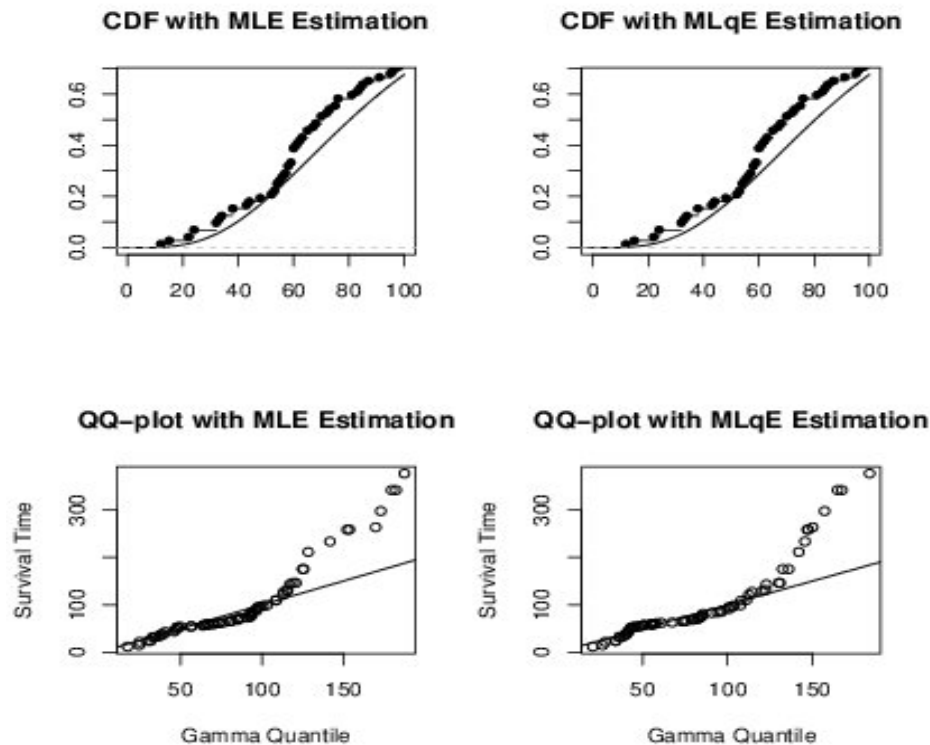
Figure 13. CDF and QQ-plot of the fitted gamma distributions based on the MLE $\hat{\theta}$ and MLqE $\tilde{\theta}$.

optimal q^* converges to 1 as the sample size increases, which indicates that the MLqE with optimal q^* is asymptotically equivalent to MLE. 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 the optimal value q^*.

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. Comparing our results on the standard gamma model with those on the exponential model in [2], the use of MLqE of the parameters in standard gamma model is not as beneficial as that in exponential model.

As demonstrated in previous and current work, the MLqE is very promising and ensures further investigation. As a generalization of the study in this paper, we may consider the general gamma model. In the general gamma model, there are two unknown parameters and this two dimensional parameter will increase the calculation burden 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 (1) with respect to both θ and q instead of θ only. This maximizing procedure could be implemented by using profile likelihood techniques.

References

1. J. Havrda and F. Charvát, "Quantification method of classification processes. concept of structural a-entropy," *Kybernetika*, vol. 3, no. 1, pp. 30–35, 1967.
2. D. Ferrari and Y. Yang, "Maximum lq-likelihood estimation," *The Annals of Statistics*, vol. 38, no. 2, pp. 753–783, 2010.
3. H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Selected Papers of Hirotugu Akaike*. Springer, 1998, pp. 199–213.
4. A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2743–2760, 1998.
5. J. Aczél and Z. Daróczy, *On measures of information and their characterizations*. AP, 1975.
6. A. Rényi *et al.*, "On measures of entropy and information," in *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 1, 1961, pp. 547–561.
7. E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, pp. 620–630, 1957.
8. ——, "Information theory and statistical mechanics. ii," *Physical review*, vol. 108, no. 2, pp. 171–190, 1957.
9. C. Tsallis, "Possible generalization of boltzmann-gibbs statistics," *Journal of statistical physics*, vol. 52, no. 1, pp. 479–487, 1988.
10. C. Tsallis, R. Mendes, and A. R. Plastino, "The role of constraints within generalized nonextensive statistics," *Physica A: Statistical Mechanics and its Applications*, vol. 261, no. 3, pp. 534–554, 1998.
11. Y. Altun and A. Smola, "Unifying divergence minimization and statistical inference via convex duality," in *International Conference on Computational Learning Theory*. Springer, 2006, pp. 139–153.
12. Y. Qin and C. E. Priebe, "Maximum lq-likelihood estimation via the expectation-maximization algorithm: A robust estimation of mixture models," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 914–928, 2013.
13. D. Ferrari and S. Paterlini, "The maximum lq-likelihood method: an application to extreme quantile estimation in finance," *Methodology and Computing in Applied Probability*, vol. 11, no. 1, pp. 3–19, 2009.
14. C. Huang, J.-G. Lin, and Y.-Y. Ren, "Testing for the shape parameter of generalized extreme value distribution based on the lq-likelihood ratio statistic," *Metrika*, vol. 76, no. 5, pp. 641–671, 2013.
15. R. Beran, "Minimum hellinger distance estimates for parametric models," *The Annals of Statistics*, pp. 445–463, 1977.
16. Z. Lu, Y. V. Hui, and A. H. Lee, "Minimum hellinger distance estimation for finite mixtures of poisson regression models and its applications," *Biometrics*, vol. 59, no. 4, pp. 1016–1026, 2003.
17. T. Bjerkedal, "Acquisition of resistance in guinea pies infected with different doses of virulent tubercle bacilli." *American Journal of Hygiene*, vol. 72, no. 1, pp. 130–148, 1960.