A NONPARAMETRIC BOOTSTRAP LIKELIHOOD RATIO TEST FOR QUANTILE REGRESSION

by

Ziwei Jin

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2022

Contents

List of Tables . . . . **iv**

List of Figures . . . . **ix**

Abstract . . . . **x**

List of Abbreviations and Symbols Used . . . . **xi**

Chapter 1 Introduction **1**

Chapter 2 Standard Likelihood Inference **4**

2.1 Maximum Likelihood Estimation 4
2.2 The Wald Test 5
2.3 Likelihood Ratio Tests Using Chi-Square Distributions 7

Chapter 3 The Bootstrap **9**

3.1 Bootstrapping and the Bootstrap Principle 9
3.2 The Percentile Method 11
3.3 The Percentile-t Method 12

Chapter 4 Nonparametric Bootstrapping and the Likelihood Ratio Test **14**

4.1 The Testing Procedure 14
4.2 Nonparametric Advantages Over the Chi-Square Threshold 16

Chapter 5 The Asymmetric Laplace Distribution and Likelihood Inference **18**

5.1 The Symmetric and Asymmetric Laplace Distribution 18
5.2 Maximum Likelihood Estimation of Asymmetric Laplace Distribution 21
5.3 Quantile Estimation and Kullback-Leibler's Divergence 24

Chapter 6 Confidence Intervals for Asymmetric Laplace Distribution Inference **28**

6.1 Standard Error Calculation 28
6.2 The Wald Test 30
6.3 The Percentile Method: A Normal Approximation 30
6.4 The Nonparametric Bootstrap Likelihood Ratio Test 32

Chapter 7 Simulation Study of Asymmetric Laplace Distribution **34**

7.1 Simulation Settings 34
7.2 Simulation Results 38

Chapter 8 Quantile Regression **55**

8.1 Nonparametric Bootstrapping on Quantile Regression 55
8.2 Real Data Example 57
8.3 Simulation of Quantile Regression 60

Chapter 9 Conclusion **69**

Bibliography **71**

Appendix **74**

List of Tables

7.1 The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard Laplace distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR). 40

7.2 The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Laplace distribution. 44

7.3 The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard normal distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR). 45

7.4 The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard normal distribution. 46

7.5 The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard logistic distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR). 47

7.6 The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard logistic distribution. 48

7.7 The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard Cauchy distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR). 49

7.8 The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Cauchy distribution. 50

7.9 The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard folded distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR). 51

7.10 The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard folded normal distribution. 52

7.11 The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard log normal distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR). 53

7.12 The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard log normal distribution. 54

8.1 For the medfly data, the average lower bound/upper bounds and width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) of model A by the nonparametric bootstrap LRT and percentile method. 63

8.2 For the medfly data, the average lower bound/upper bounds and width of the 95% confidence intervals for the median ($\tau = 0.5$) of model B by the nonparametric bootstrap LRT and percentile method. 64

8.3 For the medfly data, the estimated coverages and width of confidence intervals of regression coefficents by chi-square and nonparametric bootstrap LRT threshold in model B of medfly data when $\tau = 0.9$. 65

8.4 For the medfly data, the estimated coverages and widths of confidence intervals of regression coefficents by chi-square and nonparametric bootstrap LRT threshold in model B of medfly data when $\tau = 0.99$. 66

8.5 The average lower bounds, upper bounds, widths, and coverages of the 90%, 95%, and 99% confidence intervals for the median ($\tau = 0.5$) of simulated quantile regression model. 67

8.6 The average lower bounds, upper bounds, widths, and coverages of the 90%, 95%, and 99% confidence intervals for the 90^{th} percentile ($\tau = 0.9$) of simulated quantile regression model. . . 67

8.7 The average lower bounds, upper bounds, widths, and coverages of the 90%, 95%, and 99% confidence intervals for the 99^{th} percentile ($\tau = 0.99$) of simulated quantile regression model. . . 68

1 The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard asymmetric Laplace distribution. . 74

2 The average lower bound/upper bound/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard asymmetric Laplace distribution. . 75

3 The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard normal distribution. 76

4 The average lower bound/upper bound/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard normal distribution. 77

5 The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the logistic distribution. 78

6 The average lower bound/upper bound/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the logistic distribution. 79

7 The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Cauchy distribution. 80

8 The average lower bound/upper bounds/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Cauchy distribution. 81

9 The average lower bound/upper bounds/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard folded normal distribution. 82

10 The average lower bound/upper bounds/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard folded normal distribution. 83

11 The average lower bound/upper bounds/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard log normal distribution. 84

12 The average lower bound/upper bounds/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard log normal distribution. 85

List of Figures

4.1 Given a random data set generated from standard normal distribution with size 5000 and 1000 bootstrap repetition, plot A shows the LR confidence interval of μ under chi-square (black) and nonparametric bootstrap thresholds (red) when the fitted model is misspecified as the standard ALD with $\tau = 0.5$. Plot B shows the comparision of LR confidence interval of μ under correct estimation. 17

5.1 Densities of ALD given $\mu = 0$ and $\sigma = 1$, with τ equals 0.5, 0.9, and 0.99 respectively. 20

5.2 The plot shows each value of $\tau \in [0.25, 0.75]$ verses its KL divergence of standard normal distribution on standard ALD with τ. 27

6.1 The blue curve is the smoothed rates curve between fitted standard error of standard ALD and true standard error of standard normal distribution verses $\tau \in [0.1, 0.9]$. The red horizontal line indicates the threshold 1. 29

7.1 Comparison of the simulating normal, logistic and Cauchy distributions (in blue curves) to approximating ALDS with scale parameters obtained by equating variances (in green curves) or as average MLEs (in red curves) with $\tau = 0.5$, 0.9, and 0.99 respetively. 36

7.2 Comparation of the simulating folded normal distribution and log-normal distribution (in blue curves) to approximating ALDS with scale parameters obtained by equating variances (in green curves) or as average MLEs (in red curves) with $\tau = 0.5$, 0.9, and 0.99 respetively. 37

Abstract

Likelihood ratio based confidence intervals often better accommodate asymmetric uncertainty about parameter estimates compared to other likelihood-based approaches. In practice, likelihood ratio tests (LRT) are mostly determined by chi-square thresholds or parametric bootstrapping thresholds, which give valid coverage probabilities under the assumption that parametric models are correct. But, in settings where likelihood estimation is robust to model misspecification, it is often the case that the likelihood theory leading to hypothesis tests and confidence intervals breaks down.

In this work, a nonparametric bootstrapping approach to LRT was developed to determine the critical value for misspecified data in the context of quantile regression models. The performance of the nonparametric LRT is compared with commonly used tests via simulated and real data. Examples of asymmetric Laplace distribution and quantile regression will be focused upon in the comparison. In addition a fast normal approximation of percentile method is derived in this thesis.

This thesis will show that compared to the Wald test, chi-square LRT, percentile method, and percentile-t method, the nonparametric bootstrapping likelihood ratio test often provides better confidence intervals. Finally, methods are illustrated through a real data example.

List of Abbreviations and Symbols Used

AL asymmetric Laplace. 2

ALD asymmetric Laplace distribution. 2

CDF cumulative density function. 20

CI confidence interval. 3

KL Kullback-Leibler's. 1

LLN law of large number. 10

LR likelihood ratio. 2

LRT likelihood ratio test. x

ML maximum likelihood. 1

MLE maximum likelihood estimator. 1

PDF probability density function. 19

QR quantile regression. 18

Chapter 1

Introduction

Maximum Likelihood (ML) estimation is one of the most widely used statistical estimation methods, as introduced by R. A. Fisher (1922) [11]. A proper ML estimation process requires that probability models are correctly assumed. However, it is often challenging to develop precise probability models for data. In reality, ML estimation may be applied by a wrong parametric family. P. J. Huber (1967) modified Wald's proof of consistency of ML estimator (MLE) [28] to deal with misspecified models [15]. When the model is misspecified, the MLE converges on the parameter minimizing Kullback-Leibler's (KL) divergence [22] between the true generating distribution for the data and model under ML estimation. The parameter that minimizes KL divergence need not be the true quantity of interest. Thus ML estimation is not generally consistent under model misspecification.

Examples of cases where ML estimation is consistent include ML estimation of means and covariances under a normal model; ML estimation of regression parameters are usually consistent when estimated with normal errors having constant variance even if the errors are non-normal and variances are heterogeneous in some way [13]; ML estimation of some parameters in regression models can be correct even if errors are correlated[2]. Examples of cases where ML estimation is not generally consistent include ML estimation of the 90th pecentile fitted to a normal model when the data are non-normal.

Even in cases where ML estimation is consistent, standard tests like the Wald test and likelihood ratio test are not robust under misspecified models [29]. The coverage probability of a 95% confidence interval need not be 0.95, for instance. In this thesis we investigate how to make confidence intervals robust to estimation under model misspecified conditions.

Nonparametric bootstrapping approaches can be used to improve confidence interval constructions when the model is misspecified. These methods include the bootstrap percentile and the bootstrap percentile-t methods. For bootstrap methods, the robustness of estimated confidence intervals increases as sample size increaes. The percentile-t confidence interval are expected to have coverages that converge to nominal level even faster than the percentile confidence interval [9].

It has not been appreciated that nonparametric bootstrapping can be used with likelihood ratio (LR) tests. LR tests may have advantages because they allow asymmetric intervals which are more natural for confidence intervals of quantities like quantiles that tend to have skewed sampling distributions. A major focus of this thesis will be the properties and application of LR tests with nonparametric bootstrap thresholds instead of chi-square thresholds in this thesis.

Quantile estimation and regression will be a focus throughout the thesis. Quantile estimation is a data-dependent estimation method that estimates conditional quantiles of the data[27]. It is valuable to obtain information about extreme quantiles conditional on data, for instance, in medical studies involving children and seniors, where questions about unusual growth patterns or longevity are questions best thought of as dependence of extremes on covariates. The goal of quantile regression in this thesis is to understand how extremes depend on covariates as opposed to averages (usual reqression). In quantile regression model the regression term is intended to be the τth quantile given that an individual with covariate x. The asymmetric Laplace (AL) density are considered in quantile regression by Koenker and et. (1999) [20] an allow consistent estimation of quantile regression.

The rest of the thesis are organized as follow. Chapter 2 reviews the likelihood inference used in AL distribution (ALD) and quantile estimation, and its comparable methods Wald test and chi-square LRT. Chapter 3 reviews the basic resampling method, bootstrapping, and discusses some alternative CI estimation methods like percentile method and percentile-t method. Chapter 4 presents the nonparametric bootstrap likelihood ratio test procedure and discusses the advantage of nonparametric bootstrap threshold over chi-square threshold in estimation. Chapter 5 introduces ALD and the likelihood inference of ALD. Chapter 6 discusses properties of confidence intervals. Methods for using nonparametric bootstrapping with LR tests are developed and a fast normal approximation to the percentile method is given. Chapter 7 gives the results of the simulation study for nonparametric bootstrap likelihood ratio test set side by side with the Wald test, the chi-square t test, the percentile method and the percentile-t method. The performance of these methods are compared by comparing properties such as the width of their confidence intervals (CI) and their coverage probabilties. Chapter 8 presents the performance of nonparametric bootstrap likelihood ratio CIs for quantile regression in both simulation study and practice, comparing to the percentile method under normal approximation. Finally, conclusions and future work introductions are made in Chhapter 9.

Chapter 2

Standard Likelihood Inference

In this chapter, we discuss the properties of the ML estimation and two classic methods of hypothesis testing. ML estimation often gives unbiased estimators (Section 5.4.3 from [24]), and the widely used Wald test and LRT are ML methods. The LRT is often the locally most powerful test (Section 5.4.4 from [24]). Thus, CIs corresponding to LRT are often uniformly most accurate regions.

2.1 Maximum Likelihood Estimation

ML estimation originally introduced by R.A Fisher in the 1920s [1], is a broadly applicable parameter estimation procedure. ML estimation has a number of optimal properties including sufficiency (complete information of data), consistency, efficiency (lowest asymptotic variance of estimated parameter) and parameterization invariance. The principle of ML estimation is to make the observed data most likely for a given probability distribution with estimated values of parameters.

Assume a parametric distribution $f(y|\theta)$ with parameter vector $\theta = (\phi, \lambda)$. Suppose we are only interest in parameter set ϕ and let λ denote the rest of the parameters. Note that a special case corresponds to $\phi = \theta$, in which case we are doing full ML estimation.

The likelihood measures the goodness of fit of a statistical model to a given sample. The likelihood function is the joint probability density of $y_1, ..., y_n$, where we treat the sample space as fixed observations but parameter set ϕ as unknowns over parameter space. The likelihood function is represented as

$$L(\phi, \lambda|y_1, ..., y_n) = \prod_{i=1}^{n} f(y_i|\phi, \lambda).$$

The log likelihood function is the natural logarithm transformation of likelihood function, which often simplifies the format of density function because converts a product to a sum and eliminates the exponential part. The log likelihood function is represented as

$$l(\phi, \lambda | y_1, ..., y_n) = \sum_{i=1}^{n} \log(f(y_i|\phi, \lambda)).$$

ML estimation is a procedure to estimate parameters from an assumed distribution $f(y_i|\phi, \lambda)$. The MLE of parameters of interest, $\hat{\phi}$, is a set of parameters that maximizes the likelihood function or log likelihood function in parameter space.

If the density function is differentiable, then the ϕ that maximizes the likelihood function can often be determined by first derivative test. Thus MLE of ϕ can be evaluated from

$$\frac{dL(\phi|y_1, ..., y_n, \lambda)}{d\phi} = 0.$$

Due to the monotonic property of the natural logarithm, $\hat{\phi}(\lambda)$ can usually also be evaluated by setting the first derivative of the log likelihood function $l(\phi|y_1, ..., y_n, \lambda)$ to zero, assuming the log likelihood function is differentiable.

MLEs are usually consistent ($\hat{\phi}$ converges almost surely to true parameter set ϕ) and efficient ($\hat{\phi}$ converges in distribution to a Normal distribution with minimum variance among all estimators that are approximately normal). Accordingly, the MLE of λ can be evaluated in general given ϕ,

$$\frac{dL(\lambda|y_1, ..., y_n, \phi)}{d\lambda} = 0 \quad or \quad \frac{dl(\lambda|y_1, ..., y_n, \phi)}{d\lambda} = 0,$$

which can be indicated as $\hat{\lambda}(\hat{\phi})$.

2.2 The Wald Test

The Wald test is a widely used sample-based statistical test. It can be used to test if the true parameter set of interest equals to some particular values. The results of Wald test indicate the difference between ML estimated parameters and the parameters under null hypothesis. Since we are only interested in ϕ, a scalar parameter in the parameter vector $\theta = (\phi, \lambda)$, we assume a null hypothesis for ϕ, $H_0 : \phi = \phi_0$, in which case ϕ_0 is known.

The standard error of $\hat{\phi}$, $se(\hat{\phi})$, can be calculated as the square root of the inverse Fisher information. The Fisher information contains all the information of ϕ

that can be explained by observations. The Fisher information is an expectation of the observed Fisher information. The observed Fisher information is evaluated as the negative second derivative of log likelihood function. The expression of Fisher information can be written as

$$I_n(\theta) = E_\phi\left[\frac{-d^2 l(\theta|y_1,...,y_n)}{d\theta^2}\right].$$

Note that $I_n(\theta)$ is a symmetric $(p+q)\times(p+q\ )$ dimensional matrix where p is the dimension of ϕ and q is the dimension of λ. Likelihood theory (Section 5.3.3 from [24]) gives that $\hat{\theta}$ is approximately $N(\theta_0, I_n(\hat{\theta})^{-1})$. Thus a standard error for $\hat{\theta}_j$ can be obtained through

$$se(\hat{\theta}_j) = \sqrt{[I_n(\hat{\theta})^{-1}]_{jj}}. \tag{2.1}$$

As a special case when $p = 1$, the standard error calculated from the appropriate entry of (2.1) is denoted as $se(\hat{\phi})$. Thus, the $se(\hat{\phi})$ will be calculated by replacing the Fisher information matrix of θ into $I(\hat{\phi}, \hat{\lambda}(\hat{\phi}))$. The Wald test uses the result that $\hat{\phi}$ is asymptotically normal distributed with the mean equal to ϕ_0 and the standard error $se(\hat{\phi})$. Thus, $\hat{\phi} - \phi_0$ follows a Normal distribution with standard deviation $se(\hat{\phi})$, which is equivalent to

$$\frac{\hat{\phi} - \phi_0}{se(\hat{\phi})} \sim N(0,1).$$

Let $\Phi^{-1}(\alpha/2)$ denote the $100\alpha/2^{th}$ percentile of standard normal distribution, then

$$P\Big(\Phi^{-1}(\alpha/2) \le \frac{\hat{\phi} - \phi}{se(\hat{\phi})} \le \Phi^{-1}(1-\alpha/2)\Big) = 1-\alpha,$$

which equivalent to

$$P\Big(\hat{\phi} - \Phi^{-1}(1-\alpha/2)se(\hat{\phi}) \le \phi \le \hat{\phi} - \Phi^{-1}(\alpha/2)se(\hat{\phi})\Big) = 1-\alpha.$$

Thus, a CI corresponding to the Wald test with coverage probability equals $1-\alpha$, is given by

$$\Big[\hat{\phi} - \Phi^{-1}(1-\alpha/2)se(\hat{\phi}), \hat{\phi} - \Phi^{-1}(\alpha/2)se(\hat{\phi})\Big],$$

where $\hat{\phi} - \Phi^{-1}(1-\alpha/2)se(\hat{\phi})$ gives us the lower bound of Wald test and $\hat{\phi} - \Phi^{-1}(\alpha/2)se(\hat{\phi})$ gives the upper bound. The version of the Wald test that I have described is the dual test to the usual Wald CIs. Some versions of the Wald test take into account the information that H_0 provides about the variance of $\hat{\phi}$ and replaces $I_n(\hat{\theta})$ with $I_n(\phi_0, \hat{\lambda}(\phi_0))$ in calculating the standard error.

2.3 Likelihood Ratio Tests Using Chi-Square Distributions

There are three classic methods to approach hypothesis tests, including the Wald test (as discussed above), the score test, and the likelihood ratio test. The LRT assesses the goodness of fit between competing statistical models, based on the ratio of the likelihoods. In the LRT estimation of variance is not necessary. The LRT is not only transformation invariant, but also range preserving. Moreover, the LRT is often locally most powerful.

The LRT tests whether the null hypothesis that the simpler of two nested models is correct for a given sample set. The first model is given by the alternative hypothesis, and the likelihood for it is obtained by maximizing over the entire parameter space. The second model is given by the null hypothesis under some constraints, which places some constraints on the values that the parameter can take on.

Thus, the null hypothesis can be written as $H_0 : \theta \in \Theta_0$, which imposes that the true parameter θ is in a possible parameter space Θ_0. In this model,

$$l(\hat{\theta}_0) = \sup_{\theta \in \Theta_0} l(\theta)$$

The alternative hypothesis, can be written as $H_\alpha : \ \theta \in \Theta_\alpha$. θ is in another group of parameter space Θ_α, which is in the complement of Θ_0. The log likelihood under the alternative hypothesis is the maximized log likelihood.

$$l(\hat{\theta}) = \sup_{\theta \in \Theta_\alpha} l(\theta)$$

The reason for this result is that usually Θ_0 is a set of measure 0 in the larger space Θ. So the supremum over Θ or the complement of Θ_0 give the same value. The likelihood ratio test statistic can be defined by

$$W = -2\left(l(\hat{\theta}_0) - l(\hat{\theta})\right),$$

where multiplication by two corrects the ratio so that it has a known large sample distribution. According to Wilk's theorem, W is asymptotically Chi-square distribution with degree of freedom equals to the the difference between the number of parameters estimated under the alternative hypothesis and the number estimated under the null hypothesis. Usually the number of constrains in null hypothesis equals the number of parameters mentioned in null hypothesis. The rejection region (rejecting H_0) is determined by $W > \chi^2_{\alpha,k}$, where k equals the degrees of freedom. The critical value is chosen to achieve a desired significant level α. Thus, the CIs are not always symmetric.

Chapter 3

The Bootstrap

This chapter introduces the bootstrap principle in the measurement of the CI and two methods of bootstrap intervals. The bootstrap is a computer-based statistical algorithm to resample data. The bootstrap can be either parametric or nonparametric. This thesis mainly talks about the nonparametric bootstrap. The percentile method uses the percentiles of bootstrap estimates, while the percentile-t method takes into account standard errors during bootstrapping.

3.1 Bootstrapping and the Bootstrap Principle

The basic idea of non parametric bootstrapping is re-sampling the observed data with the same sample size to infer and estimate an unknown population quantity. Each bootstrap sample can be considered as a random sample from the empirical distribution for the original data.

Suppose a random sample set $Y = (y_1, y_2, \ldots, y_n) \sim F$ with an unknown parameter vector θ that we are interested in. In nonparametric bootstrapping, a bootstrap sample is randomly drawn from Y of size n with a replacement . In equivalent, Y is drawn from the empirical distribution. Note that the ML estimate of distribution F is the empirical distribution of $y_1, y_2, \ldots, y_n$ [6]. Suppose $\hat{F}_n$ is defined as an empirical discrete distribution with a probability $\frac{1}{n}$ assigned to each y_i. Thus, the empirical probability of a subset A of Y under $\hat{F}_n$ is

$$\hat{P}(A) = \#\{y_i \in A\}/n.$$

Suppose a parameter set can expressed as a function of the true distribution, $\theta = t(F)$. Then, the plug-in estimator of $\theta = t(F)$ will be obtained corresponding to $\hat{\theta} = t(\hat{F}_n)$.

Bootstrap data and parameters can be denoted by superscript $*$. Let $Y^{*(b)} =$

$y_1^{*(b)}, y_2^{*(b)}, \ldots, y_n^{*(b)}$ denote the random bootstrap sample that are resampled independently with replacement from $\hat{F}_n$ and let $E_*[h(Y^{*(b)})]$ denote the expectation of $h(y_1^{*(b)}, y_2^{*(b)}, \ldots, y_n^{*(b)})$, where $b = 1, 2, \ldots, B$. Note that this expectation is calculated conditional upon the original data, $y_1, y_2, \ldots, y_n$ that determine the distribution $\hat{F}_n$.

In order to estimate parameters from bootstrap samples, the plug-in principle is applied. If we knew the true parameter θ, then in order to determine the distributional properties of some function of an estimator and the true parameter, $g(\hat{\theta}, \theta)$, we could generate B samples from F and calculate $\hat{\theta}_1, \ldots, \hat{\theta}_B$. Then, use the observed distribution of the $g(\hat{\theta}_1, \theta), ..., g(\hat{\theta}_B, \theta)$ as an approximation to the sampling distribution of $g(\hat{\theta}, \theta)$. With large B, the approximation would be almost exact. For instance, $E_\theta[g(\hat{\theta}, \theta)]$ can be approximated by the average of $g(\hat{\theta}_b, \theta)$, where $b \in \{1, \ldots, B\}$.

But we don't know the true parameter θ. Bootstrapping gets around this difficulty by generating data from the $\hat{F}_n$ (the bootstrap samples). Then, plug in $\hat{\theta}_b^*$ for $\hat{\theta}_b$ and $\hat{\theta}$ for θ in $g(\hat{\theta}_b, \theta)$. It then uses the observed distribution of the $g(\hat{\theta}_1^*, \hat{\theta}), \ldots, g(\hat{\theta}_B^*, \hat{\theta})$ as if it were the observed distribution of the $g(\hat{\theta}_1, \theta), \ldots, g(\hat{\theta}_B, \theta)$ in the process above.

We denote $E_*[g(\hat{\theta}_b^*, \hat{\theta})]$ as the expectation of $g(\hat{\theta}_b^*, \hat{\theta})$ when data are generated from the empirical distribution. The law of large number (LLN) gives that as $B \to \infty$,

$$\sum_{b=1}^{B} g(\hat{\theta}_b^*, \hat{\theta})/B \longrightarrow E_*[g(\hat{\theta}_b^*, \hat{\theta})].$$

Thus, $\sum_b g(\hat{\theta}_b^*, \hat{\theta})/B$ is an unbiased estimator of $E_*[g(\hat{\theta}_b^*, \hat{\theta})]$. Bootstrap theory implies that $E_*[g(\hat{\theta}_b^*, \hat{\theta})]$ provides a reasonable asymptotic approximation to $E_\theta[g(\hat{\theta}, \theta)]$. So as $n \to \infty$,

$$\sum_{b=1}^{B} g(\hat{\theta}_b^*, \hat{\theta})/B \longrightarrow E_\theta[g(\hat{\theta}, \theta)].$$

One example application of the nonparametric bootstrap is when $g(\hat{\theta}, \theta) = I\{\hat{\theta} - \theta \leq a\}$ and a is fixed. The average of bootstrap estimators equals to $\sum_{b=1}^{B} I\{\hat{\theta}_b^* - \hat{\theta} \leq a\}/B$, which approximates $E_*[I\{\hat{\theta}_b^* - \hat{\theta} \leq a\}]$ by the LLN. The bootstrap theory implies that $E_*[I\{\hat{\theta}_b^* - \hat{\theta} \leq a\}] = P_*[\hat{\theta}_b^* - \hat{\theta} \leq a]$ provides an estimation of $E_\theta[I\{\hat{\theta} - \theta \leq a\}] = P_\theta[\hat{\theta} - \theta \leq a]$, and thus that $P_*[\hat{\theta}_b^* - \hat{\theta} \leq a]$ estimates $P_\theta[\hat{\theta} - \theta \leq a]$.

Another example application is when $g(\hat{\theta}, \theta)$ is the variance of $\hat{\theta}$, where $Var_\theta(\hat{\theta}) = E_\theta((\hat{\theta} - \theta)^2)$. According to the LLN, the variance of $\hat{\theta}_b^*$, which can be written as

$V\hat{a}r_*(\hat{\theta}) = \sum_{b=1}^{B}(\hat{\theta}_b^* - \sum_b \hat{\theta}_b^*/B)^2/(B-1) \to Var_*(\hat{\theta}_b^*)$. Bootstrap theory gives that $Var_*(\hat{\theta}_b^*)$ converges to $Var_\theta(\hat{\theta})$.

Moreover, nonparametric bootstrap can also be used to estimate the bias of $\hat{\theta}$, where $g(\hat{\theta}, \theta) = bias(\hat{\theta}) = E_\theta(\hat{\theta}) - \theta$. Thus, the bootstrap estimate of $bias(\hat{\theta})$ is $\sum_{b=1}^{B} g(\hat{\theta}_b^*, \hat{\theta})/B = \sum_{b=1}^{B}(E^*(\hat{\theta}_b^*) - \hat{\theta})/B$, which equals to $\sum_{b=1}^{B}\hat{\theta}^*/B - \hat{\theta}$. Since bias of bootstrap estimators is $bias(\hat{\theta}^*) = E^*(\hat{\theta}_b^*) - \hat{\theta}$, by the LLN

$$\sum_{b=1}^{B}\hat{\theta}^*/B - \hat{\theta} \longrightarrow E^*(\hat{\theta}_b^*) - \hat{\theta}.$$

The bootstrap theory implies $E_*(\hat{\theta}_b^*) - \hat{\theta}$ approximate to $E_\theta(\hat{\theta}) - \theta$, then the bootstrap estimate of $bias(\hat{\theta})$ converges to $bias(\hat{\theta})$.

3.2 The Percentile Method

The percentile interval is a particular CI that does not require MLEs, but uses approximate normality of the estimators for its justification [7]. However, this procedure uses its own critical value from the distribution of data instead of using the critical value of the standard normal distribution.

Suppose a bootstrap sample that is generated from the original sample set $Y = (y_1, y_2, \ldots, y_n)$ with an unknown distribution $F(\theta)$, can be denoted as $Y^* = (y_1^*, y_2^*, \ldots, y_n^*)$. The MLE of θ in the b^{th} bootstrap sample Y_b^* is denoted as $\hat{\theta}_b^*$, where $b = 1,\ 2,\ ...,\ B$.

Let $G(\cdot)$ be the approximate empirical distribution of $\hat{\theta}^*$ with data generated from $\hat{F}_n$, which estimate the sample distribution of $\hat{\theta}$. Then, the probability of $\hat{\theta}_b^* \leq a$ can be represent by $G(a)$. Let $b_{(\alpha)}$ represents the order index of α^{th} percentile in Θ^*, then $\hat{\theta}_{b_{(\alpha)}}^*$ is the α^{th} percentile of $G(\cdot)$ and

$$P^*(\hat{\theta} \leq \hat{\theta}_{b_{(\alpha)}}^*) \approx G(\hat{\theta}_{b_{(\alpha)}}^*) = \alpha.$$

Thus, $\hat{\theta}_{b_{(\alpha)}}^*$ can also be represented as

$$\hat{\theta}_{b_{(\alpha)}}^* = G^{-1}(\alpha).$$

The $100 \times (1-\alpha)^{th}$ CI is generated from the $(1-\alpha/2)^{th}$ and the $\alpha/2^{th}$ percentile of $G(\cdot)$, which can be written as

$$P^*\Big(\hat{\theta} \leq \hat{\theta}_{b_{(\alpha/2)}}^*\Big) = \alpha/2$$
$$P^*\Big(\hat{\theta} \leq \hat{\theta}_{b_{(1-\alpha/2)}}^*\Big) = 1-\alpha/2$$

which are equivalent to

$$P^*\left(\hat{\theta}^*_{b_{(\alpha/2)}} \leq \hat{\theta} \leq \hat{\theta}^*_{b_{(1-\alpha/2)}}\right) = 1-\alpha$$

In other words, the $(1-\alpha)\times 100\%$ percentile CI can be written as

$$\left[\hat{\theta}^*_{b_{(\alpha/2)}}, \hat{\theta}^*_{b_{(1-\alpha/2)}}\right] = [\hat{\theta}_{lo}, \hat{\theta}_{up}].$$

The percentile CI is transformation-respecting, which means for increasing parameter transformation $\phi = t(\theta)$, the percentile interval has this equation satisfied:

$$[\hat{\phi}_{lo}, \hat{\phi}_{up}] = [t(\hat{\theta}_{lo}), t(\hat{\theta}_{up})].$$

3.3 The Percentile-t Method

The percentile-t approach estimates the distribution directly from the data, by generating B bootstrap samples and computing the bootstrap version of the standard normal test statistic Z. Then, the estimation of an empirical distribution Z, can be given by the distribution of

$$Z^*_b = \frac{\hat{\theta}^*_b - \hat{\theta}}{se(\hat{\theta}^*_b)}$$

where $\hat{\theta}^*_b$ is the estimated parameter set of the bootstrap sample Y^*_b and $se(\hat{\theta}^*_b)$ is the estimated standard error of $\hat{\theta}^*$ for the bootstrap sample $\hat{\theta}^*_b$, which for ML estimation of a scalar parameter can be evaluated by

$$se(\hat{\theta}^*_b) = \sqrt{\frac{1}{I_n(\hat{\theta}^*_b)}}$$

Let ξ_α denote the $\alpha \times 100^{th}$ sample percentile of the Z^*_b, with the same approach as the Wald test instead using different critical values

$$P\left(\xi_{\alpha/2} \leq \frac{\hat{\theta}^*_b - \hat{\theta}}{se(\hat{\theta}^*_b)} \leq \xi_{1-\alpha/2}\right) \approx 1-\alpha$$

$$P\left(\hat{\theta}^*_b - \xi_{1-\alpha/2}se(\hat{\theta}^*_b) \leq \hat{\theta} \leq \hat{\theta}^*_b - \xi_{\alpha/2}se(\hat{\theta}^*_b)\right) \approx 1-\alpha$$

By the LLN and bootstrap theory, then the average of bootstrap percentile-t CI of $\hat{\theta}$ converges to the CI of θ from original data

$$1-\alpha \approx P^*\{\hat{\theta} \in [\frac{1}{B}\sum_b \hat{\theta}^*_b - \xi_{1-\alpha/2}se(\hat{\theta}^*_b), \frac{1}{B}\sum_b \hat{\theta}^*_b - \xi_{\alpha/2}se(\hat{\theta}^*_b)]\}$$

$$\approx P\{\theta \in [\hat{\theta} - \xi_{1-\alpha/2}se(\hat{\theta}), \hat{\theta} - \xi_{\alpha/2}se(\hat{\theta})]\}$$

Thus, the general form of the Percentile-t CI with a coverage probability equivalent to $1-\alpha$ can be written as

$$\left[\hat{\theta}_b^* - \xi_{1-\alpha/2} se(\hat{\theta}_b^*), \hat{\theta}_b^* - \xi_{\alpha/2} se(\hat{\theta}_b^*)\right]$$

The percentile-t confidence procedure has a calculation problem, where a complicated parameter estimation increases the difficulty in the calculation of the estimated standard error. Moreover, for small sample and nonparametric condition, it is challenging to interpret.

Chapter 4

Nonparametric Bootstrapping and the Likelihood Ratio Test

In the LRT, the test statistics under the null hypothesis are mostly accurate when the assumed distribution for the data is the correct distribution of population.[26]. In the estimation of the theoretical sampling distribution, the result may be biased when the population distribution is incorrect. The use of the nonparametric bootstrap does not require a specified population distribution [3]. Instead, this method relies on the empirical distribution of samples. Therefore, it is promising to combine the nonparametric bootstrap and the LRT to achieve a more accurate result when the population distribution is unknown. It has not been widely appreciated that the nonparametric bootstrap can be used with likelihood ratio tests. Standard text book treatments of boostrapping like (Efron and Tibshirani [8]) and Hall [14] do not consider likelihood ratio tests. Davison and Hinkley [5] consider bootstrapping empirical likelihood methods or using kernel density estimation. In this chapter, details about the nonparametric bootstrap likelihood ratio test will be introduced.

4.1 The Testing Procedure

Suppose that a random sample $y_1, \ldots, y_n$ is generated from a true distribution G but fitted to a parametric model with density $f(y; \phi, \lambda)$ and cumulative distribution function $F(y; \phi, \lambda)$. We assume that ϕ is still a meaningful parameter under the true distribution G. For instance, $f(y; \phi, \lambda)$ might correspond to a normal model with mean ϕ. The mean ϕ would still be meaningful when the true distribution G was non-normal.

Suppose ϕ is the parameter of interest, let the null hypothesis H_0 be $\phi = \phi_0$. The MLE of λ under the null hypothesis can be denoted as $\hat{\lambda}(\phi_0)$. Then, the log likelihood function under the null hypothesis can be evaluated from $l_G(\phi_0, \hat{\lambda}(\phi_0))$.

Given the alternative hypothesis H_a that $\phi \neq \phi_0$ is under distribution G, the MLEs can be denoted as $\hat{\phi}$ and $\hat{\lambda}$. The log likelihood function under the alternative hypothesis is $l_G(\hat{\phi}, \hat{\lambda})$. Then, the likelihood ratio statistic is defined as

$$W_0 = -2\left(l_G\left(\hat{\phi}, \hat{\lambda}\right) - l_G\left(\phi_0, \hat{\lambda}(\phi_0)\right)\right).$$

Let $\hat{F}_n$ be the empirical discrete distribution of original data with a probability $\frac{1}{n}$ assigned to each observation. Let $Y^{*(1)}, \ldots, Y^{*(B)}$ represent B sets of bootstrap samples generated from $\hat{F}_n$, and $y_1^{*(b)}, \ldots, y_n^{*(b)}$ represent the $Y^{*(b)}$ bootstrap data, where $b \in \{1, \ldots, B\}$. The ML parameter estimate based on $Y^{*(b)}$ is $(\hat{\phi}^{*(b)}, \hat{\lambda}^{*(b)})$

The null hypothesis of a bootstrap sample $Y^{*(b)}$ is

$$H_0 : \phi = \hat{\phi}$$

where $\hat{\phi}$ is the MLE of ϕ under the original data. The bootstrap log likelihood function under H_0 can be evaluated from $l_G^*(\hat{\phi}, \hat{\lambda}^{*(b)}(\hat{\phi}))$, where the ML estimation of λ under H_0 is $\hat{\lambda}^{*(b)}(\hat{\phi})$.

While, the alternative hypothesis of bootstrap sample is

$$H_a : \phi \neq \hat{\phi}$$

In addition, the log likelihood function under the hypothesis is $l^{*(b)}(\hat{\phi}^{*(b)}, \hat{\lambda}^{*(b)})$, which only rely on bootstrap data.

Then, the likelihood ratio statistic of a bootstrap data $Y^{*(b)}$ is

$$W_b^* = -2\left(l_G^{*(b)}\left(\hat{\phi}^{*(b)}, \hat{\lambda}^{*(b)}\right) - l_G^{*(b)}\left(\hat{\phi}, \hat{\lambda}^{*(b)}(\hat{\phi})\right)\right).$$

One way to obtain the test result of the nonparametric bootstrap LRT procedure is through p-value, which is approximately the proportion of $W_b^* > W_0$, which can be derive from

$$p = \#\{W_b^* > W_0\}/B.$$

The α bootstrap threshold of the nonparametric bootstrap LRT is equivalent to the $100(1-\alpha)^{th}$ quantile of W_b^*. In this thesis, we are interested in the 0.1, 0.05, and 0.01 bootstrap threshold. These bootstrap thresholds are denoted as $Q_{0.1}(W_b^*)$, $Q_{0.05}(W_b^*)$, and $Q_{0.01}(W_b^*)$.

Another way to obtain the test result is through the $100(1-\alpha)\%$ CI of ϕ in nonparametric bootstrap LRT procedure. The lower bound is given by the minimum of ML estimations of ϕ among bootstrap samples when W^* is less or equal to the α thresholds, while the upper bound equals maximum of ML estimations of ϕ among bootstrap samples when W^* is less or equal to the α thresholds.

Assuming the profile likelihood is a unimodal function of ϕ, the $100(1-\alpha)\%$ nonparametric bootstrap LRT confidence interval can be express as

$$[\min\{\phi|W(\phi) \leq Q_\alpha(W_b^*)\}, \max\{\phi|W(\phi) \leq Q_\alpha(W_b^*)\}],$$

where $W(\phi)$ is the LR statistic for the original data.

4.2 Nonparametric Advantages Over the Chi-Square Threshold

Standard likelihood theory suggests that if the model is not misspecified then an appropriate LRT threshold is determined from the chi-square distribution. The α-level chi-square threshold equals the α^{th} quantile of chi-square distribution with degree of freedom k, $\chi^2(\alpha, k)$. The degree of freedom k is given by number of parameters in ϕ.

The $100(1-\alpha)\%$ chi-square CI is expressed as

$$\left[\min\left(\phi|W(\phi) \leq \chi^2_{(\alpha,k)}\right), \max\left(\phi|W(\phi) \leq \chi^2_{(\alpha,k)}\right)\right],$$

where $W(\phi)$ is the LR statistic for the original data.

Figure 4.1 indicates the performance of bootstrap thresholds and chi-square thresholds in estimating misspecified model and correct model. Figure 4.1 Plot A is an example of generating the original data from standard normal population while estimated by ALD with $\mu = 0$, $\sigma = 1$ and $\tau = 0$ ($ALD(0, 1, 0.5)$). Figure 4.1 Plot B is an example when data is correctly estimated $ALD(0, 1, 0.5)$. When the model is misspecified, the bootstrap threshold when $p = 0.05$ equals 6.1, which gives a 95% bootstrap CI of ϕ, $[-0.1017, 0.0474]$. While the 95% chi-square CI of ϕ is $[-0.0969, 0.0363]$, which is given by the chi-square threshold $\chi^2_{0.95,1} = 3.84$. The thresholds here are very different. In this case, however, the resulting confidence intervals are very similar because the LRT increases very quickly as μ moves away from the MLE.

In figure B, since the original data is also under ALD, the bootstrap CI and chi-square CI are all well estimated. In this case, the bootstrap threshold is 3.63 with is close to $\chi^2_{0.95,1}$. The bootstrap and chi-square condfidence intervals are $[-0.073, 0.160]$

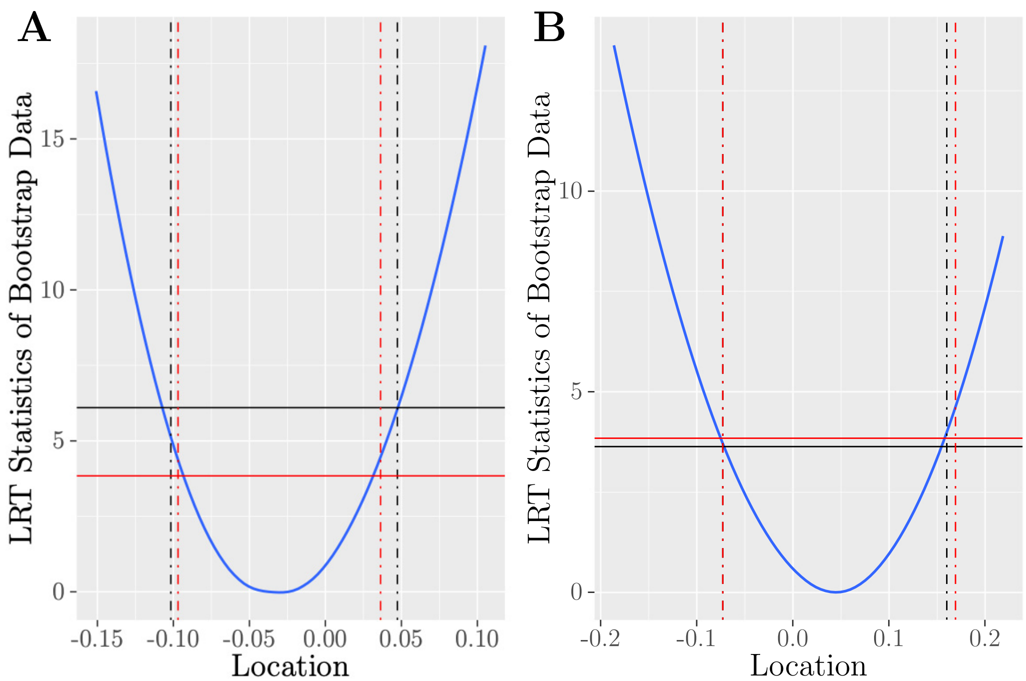


Figure 4.1: Given a random data set generated from standard normal distribution with size 5000 and 1000 bootstrap repetition, plot A shows the LR confidence interval of μ under chi-square (black) and nonparametric bootstrap thresholds (red) when the fitted model is misspecified as the standard ALD with $\tau = 0.5$. Plot B shows the comparision of LR confidence interval of μ under correct estimation.

and $[-0.073, 0.169]$, which are similar. Thus, under the correct model, the nonparametric bootstrap LRT gives similar results to the to chi-square LRT.

The justification for a chi-square threshold requires a correct model specification, but under real data situation its difficult to find the correct distribution. Thus, estimation with nonparametric bootstrap LRT is more robust than chi-square LRT.

Chapter 5

The Asymmetric Laplace Distribution and Likelihood Inference

In this chapter we discuss the properties of the ALD and develop maximum likelihood inference for this model. The location parameter of the ALD can be interpreted as a quantile of the distribution. We will show that the MLE of this parameter is statistically consistent even when the AL model is incorrect. Consequently, likelihood inference under the AL can be made robust to model misspecification via nonparametric bootstrap methods, which we explore in later chapters.

5.1 The Symmetric and Asymmetric Laplace Distribution

The symmetric Laplace distribution, also known as the double exponential distribution, can be used as an alternative distribution to the Gaussian distribution.

A random symmetric Laplace variable y has a density function

$$f(y|\mu, \sigma) = \frac{1}{2\sigma} \exp(-\frac{|y-\mu|}{\sigma}),$$

where location parameter μ has an interpretation as both the mean and the median of the distribution, and scale parameter σ, which has a positive value, and is related to the variance of the distribution. When $\mu = 0$, the symmetric Laplace variables is the distribution of a difference between exponential variables having the same mean, sigma. The symmetric Laplace distribution is a special case of the ALD without skewness.

The ALD is a continuous distribution that is widely used in quantile regression (QR) and beyond. The ALD pecifically in estimating the skewed error terms in QR. There are several different ways of expressing the ALD. In this thesis, we are using Koenker and Machado's formation[19] of ALD.

The probability density function of an AL random variable y with location μ, scale parameter σ, and skewness parameter τ is given by

$$f(y; \mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\left(\tau - I(y-\mu \leq 0)\right) \frac{y-\mu}{\sigma} \right),$$

where $I(y - \mu \leq 0)$ is an indicator of the event that $y - \mu \leq 0$. The probability density function (PDF) can be also written as:

$$f(y; \mu, \sigma, \tau) = \begin{cases} \frac{\tau(1-\tau)}{\sigma} \exp\left(-(\tau-1))\frac{y-\mu}{\sigma}\right), & y \leq \mu \\ \frac{\tau(1-\tau)}{\sigma} \exp\left(-\tau\frac{y-\mu}{\sigma}\right), & y > \mu \end{cases}$$

Here y denote the variable name of ALD, and the location parameter μ also represent the τ^{th} quantile of ALD.

An exponential variable x comes with a density function $f(x) = \lambda \exp(-\lambda x)$, where $x > 0$ and scale parameter $\lambda > 0$. The graph of density function will looks like a declining convex curve above x-axis, which starts from the y-axis and converges to 0 along the x-axis. As λ increases the curve flattens. Thus, the AL variables can be considered as combinations of exponential variables located at either side of μ with different scale parameters $\frac{-(\tau-1)}{\sigma}$ and $\frac{-\tau}{\sigma}$. Figure 5.1 shows $\tau = 0.5$ implies an equal scale on either side of μ and the distribution is symmetrically distributed. More extreme τ lead to a more skewed and flat distribution as Figure 5.1 shows.

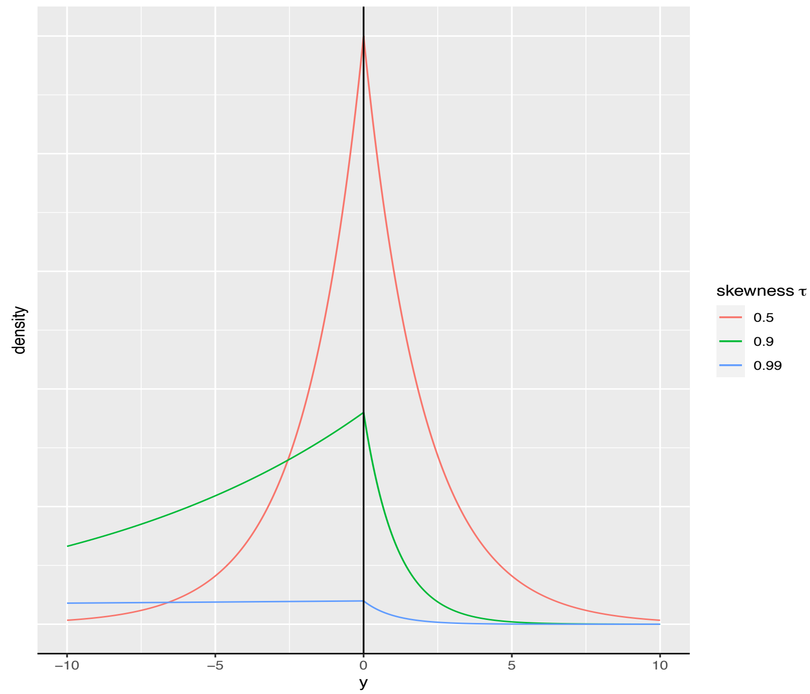


Figure 5.1: Densities of ALD given $\mu = 0$ and $\sigma = 1$, with τ equals 0.5, 0.9, and 0.99 respectively.

In addition, the cumulative density function (CDF) of an AL variable, y, can be written as

$$F(y;\mu,\sigma,\tau) = \begin{cases} \tau \exp\left((1-\tau)\frac{y-\mu}{\sigma}\right), & y \leq \mu \\ 1-(1-\tau)\exp\left(-\tau\frac{y-\mu}{\sigma}\right), & y > \mu \end{cases}$$

Note that $F(\mu;\mu,\sigma,\tau) = \tau$, establishing that one interpretation of μ is that it is the τ^{th} quantile of the ALD. Because the ALD has an explicit CDF, inversion transform sampling is an ideal method to generate random samples with their CDF, as we now show.

Suppose that p has a uniform distribution in the range $[0,1]$ and let F be the CDF of a specific continuous distribution. We seek to generate Y from F, let $Y = F^{-1}(p)$. Since p is uniformly distributed between $[0,1]$, $P[p \leq p_0] = p_0$. The event $Y \leq y$ is equivalent to the event that $p = F(Y) \leq F(y)$ which happens with probability $F(y)$. Thus the CDF of Y is $F(y)$. Therefore a random variable Y having CDF F can be generated as $F^{-1}(p)$.

Since the $F(y)$ is monotonically increasing and $F(\mu) = \tau$, then $\tau \geq F(y) = p$

if and only if $y \leq \mu$, and $\tau < F(y) = p$ if and only if $y > \mu$. Therefore, when $p \leq \tau$, $F^{-1}(p)$ is determined by solving $p = \tau \exp[(1-\tau)(y-\mu)/\sigma]$, which has solution $\sigma/(1-\tau)\log(p\sigma/\tau) + \sigma$. When $p > \tau$, $F^{-1}(p)$ is determined by solving $p = 1-(1-\tau)\exp[-\tau(y-\mu)/\sigma]$, which has solution $-\sigma/\tau \log[(1-p\sigma)p/(1-\tau)] + \sigma$.

Given all the parameters of ALD and p, we can get the expression of y by inversing the CDF of y. Then, a random AL sample can be generated by the equations below.

$$F^{-1}(p;\mu,\sigma,\tau) = \begin{cases} \frac{\sigma}{1-\tau}\log(\frac{p}{\tau}) + \mu, & p \leq \tau \\ -\frac{\sigma}{\tau}\log(\frac{1-p}{1-\tau}p) + \mu, & p > \tau \end{cases}$$

5.2 Maximum Likelihood Estimation of Asymmetric Laplace Distribution

We now derive the MLE of μ and σ for the ALD. It turns out that a non-standard argument is needed. Suppose we generated a random AL sample set with size n, $y_1, y_2, ..., y_n$. The likelihood function of the sample set can be written as

$$L_A(\mu,\sigma,\tau) = (\frac{\tau(1-\tau)}{\sigma})^n \times \prod_{(y_i\leq\mu)} \exp\left((1-\tau)\frac{y_i-\mu}{\sigma}\right) \times \prod_{(y_i>\mu)} \exp\left(-\tau\frac{y_i-\mu}{\sigma}\right)$$

It is valuable to reexpress the likelihood function in terms of order statistics, where the sample can written to $y_{(1)}, y_{(2)}, ..., y_{(n)}$. Then $y_{(1)}$ is the minimum value of all observations and $y_{(n)}$ gives the maximum value. It isn't difficult to see that the log-likelihood is decreasing for values of μ larger than all of the y_i and increasing for value of μ that are smaller than all of the y_i. So in determining the MLE we can assume that $y_{(i)} \leq \mu < y_{(i+1)}$ for some i. Thus, the natural logarithm of likelihood function can be written as

$$\begin{aligned} l(\mu,\sigma,\tau) =& n\log(\tau) + n\log(1-\tau) - n\log(\sigma) \\ &+ \frac{1-\tau}{\sigma}\sum_{j=1}^{i}(y_{(j)}-\mu) - \frac{\tau}{\sigma}\sum_{j=i+1}^{n}(y_{(j)}-\mu), \qquad (5.1) \\ &\textit{when } y_{(i)} \leq \mu < y_{(i+1)}. \end{aligned}$$

Note that here i is a function of μ, defined as the largest integer such that $y_{(i)} \leq \mu$.

In order to find the MLE of μ, we simplify the natural logarithm of likelihood function $L_A(\mu,\sigma,\tau)$ to $l_A(\mu)$ by treating σ and τ as constant numbers, which means

μ is the only unknown parameter. The score function (first order derivative) of $l_A(\mu)$ can be expressed as

$$S_A(\mu) = \frac{(n-i)\tau - (1-\tau)i}{\sigma} = \frac{n\tau - i}{\sigma}, \quad when\ y_{(i)} \leq \mu < y_{(i+1)}. \tag{5.2}$$

Usually, the log likelihood function is continuously differentiable and uni-modal. The MLE was solved by the first order derivative test. Because (5.2) gives a poly-line equation, we cannot directly solve the maximum solution by simply using first order derivative test.

In addition, we have

$$S_A(\mu) > 0 \longleftrightarrow n\tau - i > 0 \longleftrightarrow i < n\tau, \tag{5.3}$$

$$S_A(\mu) = 0 \longleftrightarrow n\tau - i = 0 \longleftrightarrow i = n\tau, \tag{5.4}$$

$$S_A(\mu) < 0 \longleftrightarrow n\tau - i < 0 \longleftrightarrow i > n\tau, \tag{5.5}$$

where i is the order of observation that may leads to MLE of μ.Because $\tau \in (0, 1)$, then $0 < n\tau < n$. We are looking for the estimator of μ that maximizes the likelihood function.

Case I $\quad n\tau$ **is not a integer and** $n\tau > 1$ In this case, $n\tau$ is located between two nearby integers, where

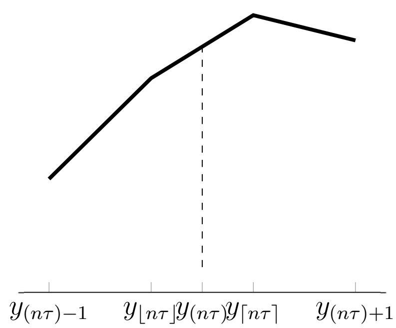
$$i < n\tau < i + 1 \quad for \quad 1 < i < n.$$

We can write $i = \lfloor n\tau \rfloor$ and $i + 1 = \lceil n\tau \rceil$,

$$y_{(\lfloor n\tau \rfloor)} < y_{(n\tau)} < y_{(\lceil n\tau \rceil)} \quad for \quad 1 < n\tau < n,$$

where $y_{(n\tau)}$ is an undefined order statistic of y.

Thus, $S_A(\mu) = 0$ will not be satisfied. Instead, when $i \leq \lfloor n\tau \rfloor$ then $S_A(\mu) > 0$ for $y_{(i)} \leq \mu < y_{(i+1)}$, by equation (5.3); and when $i \geq \lceil n\tau \rceil$ we can find $S_A(\mu) < 0$ for $y_{(i)} \leq \mu < y_{(i+1)}$ by equation (5.5).

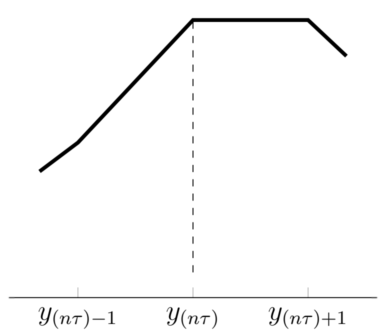


Graph 5.1

As the Graph 5.1 shows, the global maximum of the log likelihood function is

$$\hat{\mu} = y_{(\lceil n\tau \rceil)}.$$

Case II $n\tau$ **is a integer** In this case, $S(\mu) = 0$ when $y_{(i)} \leq \mu \leq y_{(i+1)}$ where $i = n\tau$. Moreover, as Graph 5.2 indicated, the order $n\tau$ and $n\tau + 1$ both stay at the maximum. Thus, the MLE $\hat{\mu}$ is any value between $y_{(n\tau)}$ and $y_{(n\tau+1)}$. Here $y_{(n\tau)}$ can be considered equals $y_{(\lceil n\tau \rceil)}$.

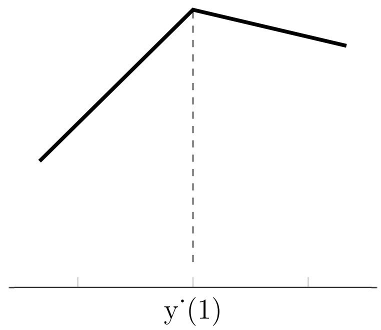


Graph 5.2

Case III $n\tau < 1$ In this case (5.5) gives that $S_A(\mu) < 0$ for all $y_{(i)} \leq \mu < y_{(i+1)}$ and all i. From Graph 5.3 we can see the first order statistic gives the global maximum of the log likelihood function,

$$\hat{\mu} = y_{(1)}.$$

As to $n\tau < 1$, 1 should be the ceiling of $n\tau$, then the result satisfied $\hat{\mu} = y_{(\lceil n\tau \rceil)}$.



Graph 5.3

In conclusion, the nonparametric MLE of μ can always be taken as

$$\hat{\mu} = y_{(\lceil n\tau \rceil)}.$$

A property of order statistics is that, for continuous data, regardless of what the distribution of the data is, $y_{(\lceil n\tau \rceil)}$ converges almost surely to the τ^{th} quantile of the distribution [25]. Thus we see that although the MLE was developed assuming an ALD for the data, the resulting estimator is reasonable even if that part of the model was misspecified.

5.3 Quantile Estimation and Kullback-Leibler's Divergence

Since $S_A(\mu)$ is not differentiable, the usual theory that gives the standard error of $\hat{\mu}$ ($se(\hat{\mu})$) from the information does not apply. One way to obtain the $se(\hat{\mu})$ is by quantile estimation. By the property of quantile estimation, the estimator location parameter has an asymptotically normal distribution[12], where

$$\hat{\mu} \sim AN\Big(\mu, \frac{\tau(1-\tau))}{n\ p(\mu)^2}\Big).$$

Here $p(\mu)$ is the density of an observation evaluated at μ and $\tau = P(X \leq \mu)$.

Moreover, the density of the ALD $f(y;\mu,\sigma,\tau)$ when $y = \mu$ is

$$f(\mu;\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma}.$$

Thus, we can find the $\hat{\mu}$ is asymptotically normal distribution with standard error

$$se(\hat{\mu}) = \sqrt{\frac{\tau(1-\tau)}{n\ p(\mu)^2}} = \sqrt{\frac{\sigma^2}{n\tau(1-\tau)}}.$$

For any $y_{(i)} < \mu \leq y_{(i+1)}$, the first derivative of of log likelihood function of ALD on rate parameter σ can be written as

$$\begin{aligned} S(\sigma) &= -\frac{n}{\sigma} - \frac{1-\tau}{\sigma^2}\sum_{j=1}^{i}(y_{(j)}-\mu) + \frac{\tau}{\sigma^2}\sum_{j=i+1}^{n}(y_{(j)}-\mu) \\ &= -\frac{n}{\sigma} - \frac{\sum_{j=1}^{i}(y_{(j)}-\mu)}{\sigma^2} + \frac{\tau\sum_{j=1}^{n}(y_{(j)}-\mu)}{\sigma^2}, \end{aligned}$$

which can be directly solved by set $S(\sigma) = 0$. The MLE of σ is determined by plugging in $\hat{\mu}$. Thus,

$$\hat{\sigma}(\hat{\mu}) = -\frac{\sum_{j=1}^{i}(y_{(j)}-\hat{\mu})}{n} + \frac{\tau\sum_{j=1}^{n}(y_{(j)}-\hat{\mu})}{n} \tag{5.6}$$

Thus, the estimated standard error of $\hat{\mu}$ can be also written as:

$$\widehat{se}(\hat{\mu}) = \Big(-\frac{\sum_{j=1}^{i}(y_{(j)}-\hat{\mu})}{n} + \frac{\tau\sum_{j=1}^{n}(y_{(j)}-\hat{\mu})}{n}\Big)\sqrt{\frac{1}{n\tau(1-\tau)}}. \tag{5.7}$$

In this thesis, we are interested in estimation when the model is misspecified. Then, the accuracy of estimation with ALD may not precise due to unknown distribution of original data. KL divergence describe how one probability distribution is different from another [23]. As an example, suppose a random sample set $y_1, \ldots, y_n$ follows the standard normal distribution and fitted with ALD, we are interested in the relative entropy from ALD to normal distribution. Denote standard normal distribution as P with PDF $p(y)$ and ALD as Q with PDF $q(y)$. Then the KL divergence from P to Q is

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(y)\log(p(y))dy - \int_{-\infty}^{\infty} p(y)\log(q(y))dy,$$

where $p(y) = \frac{1}{\sqrt{2\pi}}\exp(-y^2/2)$ and

$$q(y) = \begin{cases} \frac{\tau(1-\tau)}{\sigma}\exp\left((1-\tau)\frac{y-\mu}{\sigma}\right), & y \leq \mu \\ \frac{\tau(1-\tau)}{\sigma}\exp\left(-\tau\frac{y-\mu}{\sigma}\right), & y > \mu \end{cases}$$

Moreover, $\log(p(y)) = -\log(2\pi)/2 - y^2/2$ and

$$\log(q(x)) = \begin{cases} \log(\frac{\tau(1-\tau)}{\sigma}) + (1-\tau)\frac{y-\mu}{\sigma}, & y \leq \mu \\ \log(\frac{\tau(1-\tau)}{\sigma}) - \tau\frac{y-\mu}{\sigma}, & y > \mu \end{cases}$$

Then,

$$\begin{aligned} D_{KL}(P||Q) = & \underbrace{-\frac{1}{2}\log(2\pi) - \frac{1}{2} - \log(\tau(1-\tau))}_{\text{Regard as } C_1} \\ & + \log(\sigma) - \frac{1-\tau}{\sigma}\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\mu}(y-\mu)\exp\left(-\frac{y^2}{2}\right)dy}_{\text{Regard as } C_2} \\ & + \frac{\tau}{\sigma}\underbrace{\frac{1}{\sqrt{2\pi}}\int_{\mu}^{\infty}(y-\mu)\exp\left(-\frac{y^2}{2}\right)dy}_{\text{Regard as } C_3} \end{aligned}$$

Plugging $\hat{\mu}$ and $\hat{\sigma}$ into the equation we can get the estimated KL divergence equals

$$\hat{D}_{KL}(P||Q) = C_1 - C_2\frac{1-\tau}{\hat{\sigma}} + C_3\frac{\tau}{\hat{\sigma}} + \log(\hat{\sigma}),$$

where $C_2 = \int_{-\infty}^{\mu}(y-\hat{\mu})\exp(-y^2/2)dy/\sqrt{2\pi}$ and $C_3 = \int_{\mu}^{\infty}(y-\mu)\exp(-y^2/2)dy/\sqrt{2\pi}$.

Figure 5.2 indicates the KL divergence of standard normal distribution to ALD increases as τ increases from 0.25 to 0.75, indicating that the ALD distribution becomes an increasingly poor approximation to the normal distribution as τ increases.

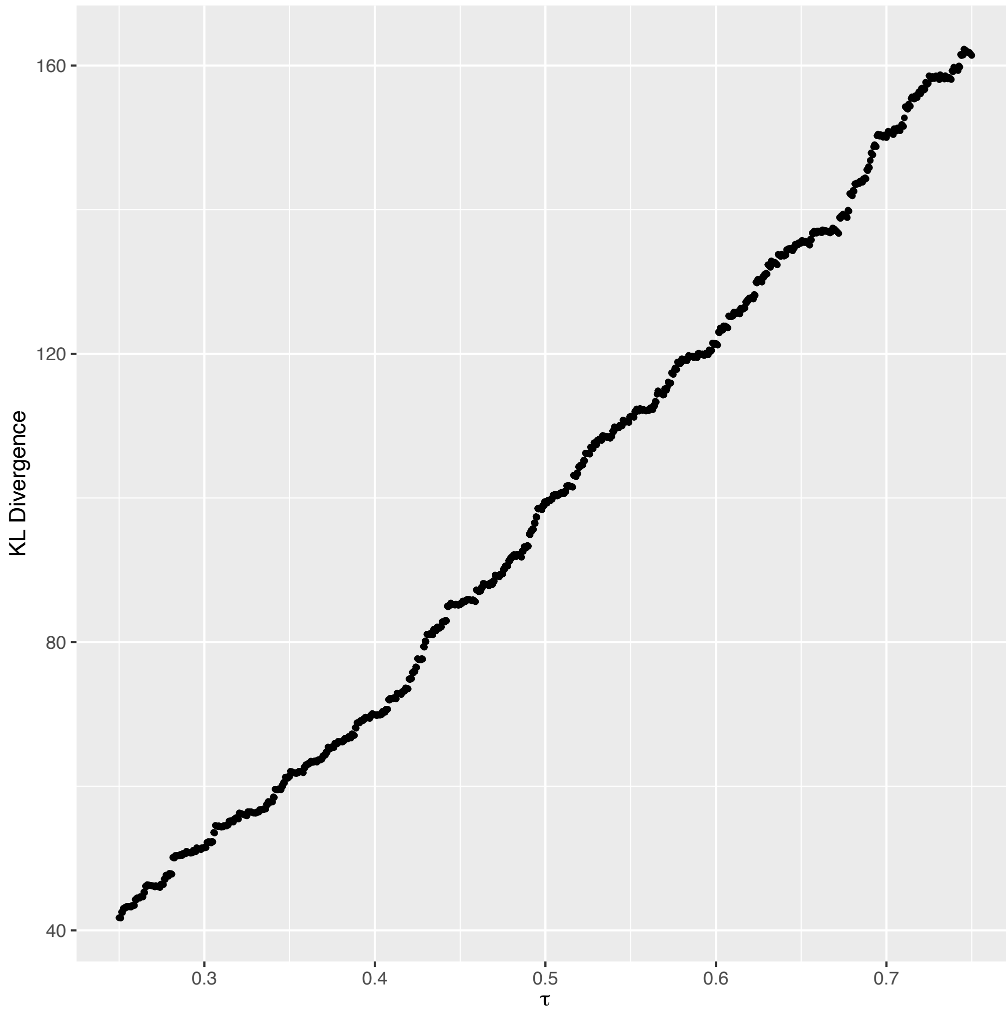


Figure 5.2: The plot shows each value of $\tau \in [0.25, 0.75]$ verses its KL divergence of standard normal distribution on standard ALD with τ.

Chapter 6

Confidence Intervals for Asymmetric Laplace Distribution Inference

Methods for CI construction using an ALD fitted model may not have correct coverage probabilities when the model is misspecified. Standard methods include the Wald test which may not be robust to model misspecification. Thus, a different method to estimate CI with nonparametric bootstrap LR test is introduced in this chapter, which is more robust to model misspecification than the Wald test.

6.1 Standard Error Calculation

Nonparametric quantile estimation is a technique to estimate the quantile under nonparametric condition[27]. By the theory of order statistics, the estimator of the τth quantile has a asymptotically normal distribution, where

$$\hat{\mu} \sim AN\Big(\mu, \frac{\tau(1-\tau)}{n\ p(\mu)^2}\Big).$$

Here $p(\mu)$ is the density function of variable Y.

As an example of model misspecification, suppose a sample was generated from standard normal distribution and fitted with ALD with fixed σ.

The density for the fitted model at μ is

$$p_L(\mu) = \frac{\tau(1-\tau)}{\sigma} \exp\left(\frac{\mu-\mu}{\sigma}(1-\tau)\right) = \frac{\tau(1-\tau)}{\sigma}.$$

Thus, if the model were correctly specified, the estimated parameter θ in the fitted distribution follows an asymptotically normal distribution with standard error

$$SE_{fitted} = \sqrt{\frac{q(1-q)}{n\ p_L(\mu)^2}} = \sqrt{\frac{\tau(1-\tau)\sigma^2}{n\ [\tau(1-\tau)]^2}} = \sqrt{\frac{\sigma^2}{n\tau(1-\tau)}}.$$

For the true distribution, standard normal distribution, we have $q = P(y \leq \mu) = \Phi(\mu) = \tau$. Then, $p_N(\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(\Phi^{-1}(\mu)^2/2\right)$, where $\mu = \Phi^{-1}(\tau)$. Thus, we have the asymptotically standard error of $\hat{\theta}$ for true model,

$$SE_{true} = \sqrt{\frac{\tau(1-\tau)}{n\ p_N(\mu)^2}} = \sqrt{\frac{2\pi\tau(1-\tau)}{n\ \exp(-\mu^2)}}.$$

The ratio of fitted standard error and true standard error can be written as

$$\frac{SE_{fitted}}{SE_{true}} = \frac{\sigma\ \exp\left(-\mu^2/2\right)}{\sqrt{2\pi}\tau(1-\tau)}.$$

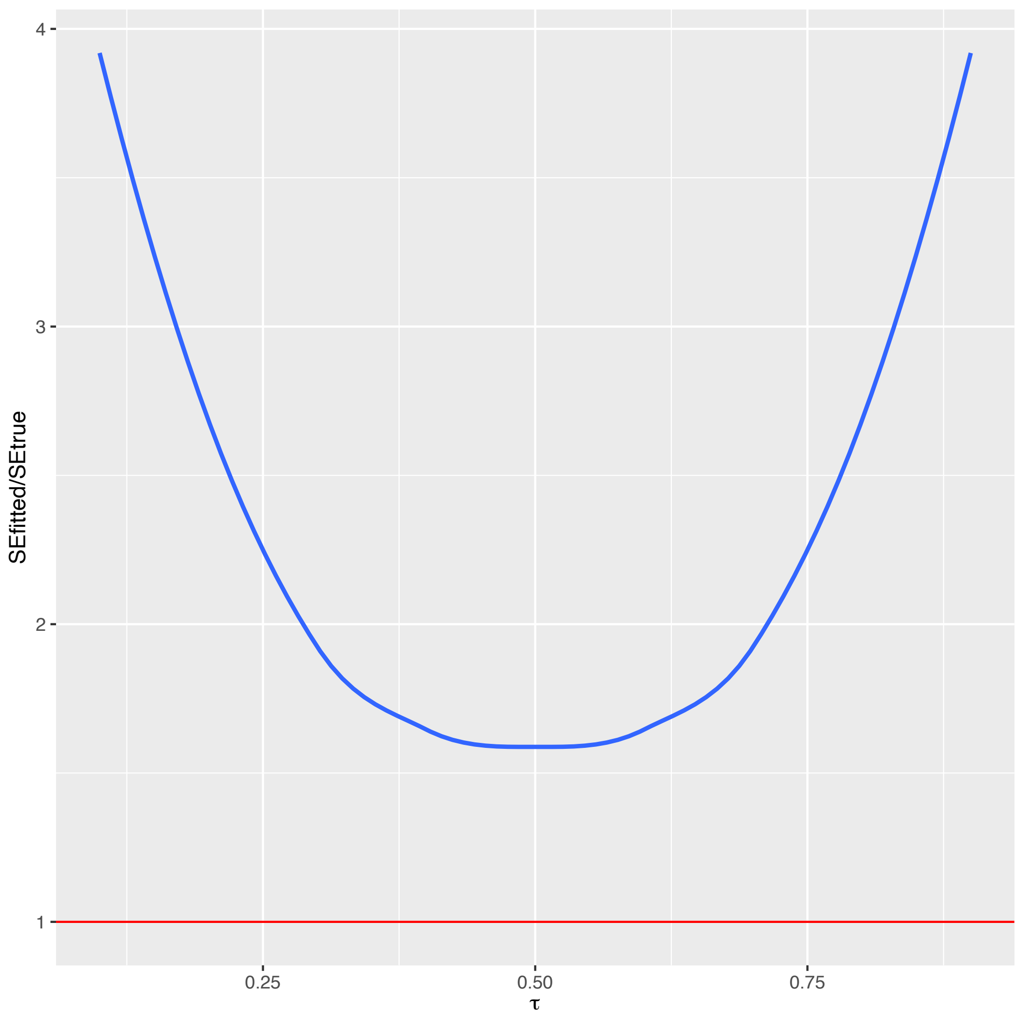


Figure 6.1: The blue curve is the smoothed rates curve between fitted standard error of standard ALD and true standard error of standard normal distribution verses $\tau \in [0.1, 0.9]$. The red horizontal line indicates the threshold 1.

Nonparametric quantile estimation of the standard error still requires approximation of the density at μ. As result of estimation with the incorrect density, Figure 6.1 that shows the rates of fitted standard error over true standard error are above 1.5, and the rates are getting larger (more than 4) as the value of τ away from 0.5. Rates are 4.43, 2.13, and 1.60 when τ equals 0.1, 0.25, and 0.5, and the rates are symmetrically distributed around $\tau = 0.5$. In brief, the fitted standard error does not approximate true standard error, which indicates that estimation of standard error under misspecified model is biased. Thus, the test procedures that require estimated standard error, such as the Wald test, may be questionable.

6.2 The Wald Test

The Wald test is based on an asymptotic distribution that holds in the absence of model misspecification,

$$t_w = \frac{\hat{\mu} - \mu_0}{se(\hat{\mu})} \sim N(0, 1).$$

This gives that $P(\Phi^{-1}(\alpha/2) \leq t_w \leq \Phi^{-1}(1 - \alpha/2)) = 1 - \alpha$. From the previous chapter we know $\hat{\mu} = y_{(\lceil n\tau \rceil)}$ and $se(\hat{\mu}) = \sqrt{\hat{\sigma}^2/(n\tau(1-\tau))}$. Thus, for each sample the $100(1 - \alpha)\%$ CI of μ corresponding to the Wald test is

$$\left[y_{(\lceil n\tau \rceil)} - \Phi^{-1}(1 - \alpha/2)\sqrt{\hat{\sigma}^2/(n\tau(1-\tau))}, y_{(\lceil n\tau \rceil)} - \Phi^{-1}(\alpha/2)\sqrt{\hat{\sigma}^2/(n\tau(1-\tau))} \right],$$

To evaluate the statistical properties of the Wald CIs or any other CIs we can generate a large number, m, of data sets from a population distribution of interest. For each of these samples we can calculate the CI for a confidence level of interest. By the LLN, the estimated coverage of probability p_w equals the proportion of times the true μ was in its CI,

$$p_w = \sum_m I\{y_{(\lceil n\tau \rceil)} - \Phi^{-1}(1 - \alpha/2)\sqrt{\hat{\sigma}^2/(n\tau(1-\tau))} \leq \hat{\mu} \leq y_{(\lceil n\tau \rceil)} - \Phi^{-1}(\alpha/2)\sqrt{\hat{\sigma}^2/(n\tau(1-\tau))}\}/m$$

6.3 The Percentile Method: A Normal Approximation

In the section we show how the CI for the percentile bootstrap method can be approximated without actually bootstrapping. Let variable x_k denote the number of time that $y_{(k)}$ was selected from the original sample to the bootstrap sample. Each time when we select one observation to the bootstrap sample, all the observation in the original sample are equally likely to be selected.

Thus, in the process of building up a bootstrap sample with n observation, each observation in the original sample has a probability $1/n$ to be selected and the number of times each observation was chosen is in the range of $[0, n]$. Then X follows a binomial distribution with n trails and success rate $1/n$.

Let z_k denote the summation of x from the first ordered observation to the k^{th} observation, $z_k = \sum_{i=1}^{k} x_i$. If $\hat{\mu}^*$ equals $y_{(k)}$, then $z_{k-1} < \lceil n\tau \rceil$ and $z_k \geq \lceil n\tau \rceil$, and vice versa. Moreover, If $z_{k-1} \geq \lceil n\tau \rceil$, then $\hat{\mu}^*$ is less than $y_{(k)}$, and vice versa. Thus

$$z_k \geq \lceil n\tau \rceil \cap z_{k-1} < \lceil n\tau \rceil \iff \hat{\mu}^* = y_{(k)}$$
$$z_k \geq \lceil n\tau \rceil \cap z_{k-1} \geq \lceil n\tau \rceil \iff \hat{\mu}^* < y_{(k)}.$$

In conclusion, we can say $z_k \geq \lceil n\tau \rceil \Leftrightarrow \hat{\mu}^* \leq y_{(k)}$, which implies that

$$P^*(z_k \geq \lceil n\tau \rceil) = P^*(\hat{\mu}^* \leq y_{(k)}).$$

As indicated above, $z_k = \sum_{i=1}^{n} I\{y_i^* \leq y_{(k)}\}$, where n trials of $I\{y_i^* \leq y_{(k)}\}$' s are independently distributed and each of the statement $y_i^* = y_{(k)}$ has a probability $1/n$. Then, each draw of y^* from bootstrap each bootstrap sample will be among $y_{(1)}, \ldots, y_{(k)}$, and hence $\hat{P}^*[y_i^* \leq y_{(k)}] = k/n$. Thus, Z_k follows a binomial distribution with size equals n and probability parameter equal to k/n. By the Central Limit Theorem, Z_k is approximately $N(k, k(n-k)/n)$, in which $\hat{\mu}^*$can be asymptotically distributed under

$$P^*(\hat{\mu}^* \leq y_{(k)}) \approx 1 - \Phi(\frac{\lceil n\tau \rceil - 1 - k}{\sqrt{k(n-k)/n}}).$$

By definition the $100(1-2\alpha)\%$ level of percentile CI is given by the lower bound (α^{th} quantile of $\hat{\theta}^*$) and the upper bound ($(1-\alpha)^{th}$ quantile of $\hat{\theta}^*$), which gives us the equations

$$\alpha = p^*(\hat{\theta}^* \leq L) \approx 1 - \Phi(\frac{\lceil n\tau \rceil - 1 - k_l}{\sqrt{k_l(n-k_l)/n}}), \tag{6.1}$$

$$1 - \alpha = p^*(\hat{\theta}^* \leq U) \approx 1 - \Phi(\frac{\lceil n\tau \rceil - 1 - k_u}{\sqrt{k_u(n-k_u)/n}}), \tag{6.2}$$

where L denote the lower bound of percentile CI and U denote the upper bound. k_l and k_u denote the position of lower bound and upper bound in the sequence of ascending ordered $\hat{\theta}^*$. To solve equation (6.1), we have

$$\alpha = 1 - \Phi\left(\frac{\lceil n\tau \rceil - 1 - k_l}{\sqrt{k_l(n-k_l)/n}}\right)$$

$$\Phi^{-1}(1-\alpha) = \frac{\lceil n\tau \rceil - 1 - k_l}{\sqrt{k_l(n-k_l)/n}}$$

Let $q_{1-\alpha}$ denotes $\Phi^{-1}(1-\alpha)$ and $n\tau \approx \lceil n\tau \rceil$, then

$$q_{1-\alpha}^2 k_l(n-k_l)/n = (n\tau - 1 - k_l)^2 (1 + q_{1-\alpha}^2/n)k_l^2 + (2 - q_{1-\alpha}^2 - 2n\tau)k_l + (n\tau - 1)^2 = 0$$

The solutions of this quadratic equation in k_l (k_l should be a positive value) are

$$k1 = \frac{C + q_{1-\alpha}^2 + 2\lceil n\tau \rceil - 2}{2(1 + q_{1-\alpha}^2/n)},$$

$$k2 = \frac{-C + q_{1-\alpha}^2 + 2\lceil n\tau \rceil - 2}{2(1 + q_{1-\alpha}^2/n)},$$

where $C = \sqrt{(2 - q_{1-\alpha}^2 - 2\lceil n\tau \rceil)^2 - 4(\lceil n\tau \rceil - 1)^2(1 + q_{1-\alpha}^2/n)}$.

To solve equation (6.2), we have

$$1 - \alpha = 1 - \Phi(\frac{\lceil n\tau \rceil - 1 - k_u}{\sqrt{k_u(n-k_u)/n}})$$

$$q_\alpha = \frac{\lceil n\tau \rceil - 1 - k_u}{\sqrt{k_u(n-k_u)}}$$

Since standard normal distribution is symmetric we have $q_\alpha = -q_{1-\alpha}$. This implies that solving the above equation gives the same solution as 6.2. Since C is a positive value, then $k_2 < k_1$. Thus, the percentile CI is $[y_{(k_2)}, y_{(k_1)}]$.

6.4 The Nonparametric Bootstrap Likelihood Ratio Test

The null hypothesis under the bootstrap is $\mu = \hat{\mu}$, where $\hat{\mu} = y_{(\lceil n\tau \rceil)}$. The original data gives $\hat{\sigma} = \sum_j (|y_j - \hat{\mu}|))/n$, while the bootstrap data gives $\hat{\mu}^{*(b)} = y^{*(b)}_{(\lceil n\tau \rceil)}$ and $\hat{\sigma}^{*(b)} = -\sum_j^{\lceil n\tau \rceil}(y_j^{*(b)} - \hat{\mu}^{*(b)}))/n + \tau \sum_j^n (y_j^{*(b)} - \hat{\mu}^{*(b)})/n$. Under the null hypothesis, the estimation of σ in the b^{th} bootstrap data given $\hat{\mu}$ is $\hat{\sigma}^{*(b)}(\hat{\mu})$ and the boostrap LR statistic is

$$W_b^* = -2\left(l_A^{*(b)}\left(\hat{\mu}^{*(b)}, \hat{\sigma}^{*(b)}\right) - l_A^{*(b)}\left(\hat{\mu}, \hat{\sigma}^{*(b)}(\hat{\mu})\right)\right),$$

where

$$\begin{aligned} l_A^{*(b)}(\hat{\mu}^{*(b)}, \hat{\sigma}^{*(b)}) =& n\log(\tau) + n\log(1-\tau) - n\log(\hat{\sigma}^{*(b)}) \\ &+ \frac{1-\tau}{\hat{\sigma}^{*(b)}} \sum_{j=1}^{i} (y_{(j)}^{*(b)} - \hat{\mu}^{*(b)}) - \frac{\tau}{\hat{\sigma}^{*(b)}} \sum_{j=i+1}^{n} (y_{(j)}^{*(b)} - \hat{\mu}^{*(b)}), \quad y_{(i)}^{*(b)} \leq \hat{\mu}^{*(b)} < y_{(i+1)}^{*(b)} \end{aligned}$$

and

$$\begin{aligned} l_A^{*(b)}(\hat{\mu}, \hat{\sigma}^{*(b)}(\hat{\mu})) =& n\log(\tau) + n\log(1-\tau) - n\log(\hat{\sigma}^{*(b)}(\hat{\mu})) \\ &+ \frac{1-\tau}{\hat{\sigma}^{*(b)}(\hat{\mu})} \sum_{j=1}^{i} (y_{(j)}^{*(b)} - \hat{\mu}) - \frac{\tau}{\hat{\sigma}^{*(b)}(\hat{\mu})} \sum_{j=i+1}^{n} (y_{(j)}^{*(b)} - \hat{\mu}), \quad y_{(i)}^{*(b)} \leq \hat{\mu} < y_{(i+1)}^{*(b)}. \end{aligned}$$

The nonparametric bootstrap threshold for significance level α is the $100(1-\alpha)^{th}$ quantile of W_b^*, $Q_\alpha(W_b^*)$. Then, the $100(1-\alpha)\%$ nonparametric bootstrap LR test CI can be expressed as

$$\left[\left[\min\{\mu | W(\mu) \leq Q_\alpha(W_b^*)\},\ \max\{\mu | W(\mu) \leq Q_\alpha(W_b^*)\}\right]\right].$$

As with the Wald CI, to approximate the coverage probabilities of the percentile CI or nonparametric bootstrap CI we repeatedly generate data sets from a distribution of interest, repeatedly calculate CIs for each of these data sets and then use the proportion of CIs containing the true parameter as an approximation for the coverage probability

Chapter 7

Simulation Study of Asymmetric Laplace Distribution

To compare the performance of nonparametric bootstrapping for parametric LRT and other commonly used tests (Wald test, percentile, percentile-t, and chi-square LRT) with misspecified models, data sets from different distributions were generated. In this thesis, we used data from six different original distributions and estimated with the same one. In each case of original distribution, the sample sizes of generated data are 100, 500, 1000, and 5000 respextively, while in each case 1000 sets of original data sets were generated. For each set of originial data, 1000 sets of bootstrap sample are generated corresponding to each original data set.

7.1 Simulation Settings

In this section, we compared performance of all methods by six different distributions.

Normal Distribution

Data sets were generated from standard normal distribution, with mean 0 and variance 1. The median of standard normal distribution equals its mean.

Logistic Distribution

Data sets were generated from logistic distribution mean and the location parameter equals 0 and the scale parameter equals 1 (standard logistic distribution), with density

$$f_L(y; 0, 1) = \exp(-y)/(1 + \exp(-y))^2$$

$f_L(y; 0, 1)$ is resemble to standard normal distribution except its more kurtosis and variance of $f_L(y; 0, 1)$ is $\pi^2/3$. The median of logistic distribution is equivlent to its location parameter, which is 0 in $f_L(y; 0, 1)$.

Cauchy distribution

Data sets that were generated from standard Cauchy distribution (location parameter 0 and scale parameter 1) have the density function

$$f_C(y; 0, 1) = \frac{1}{\pi(y^2 + 1)},$$

where the standard Cauchy variable Y is also defined by the ratio of two independent standard normal variables. All three distributions above are symmetrically distributed around zero and continuously through $[-\infty, \infty]$, besides the mean and variance of Cauchy distribution is undefined. However, the median of Cauchy distribution equals its location parameter.

Folded Normal Distribution

Data sets were generated from folded normal distribution with location parameter 0 and scale parameter 1 (standard folded normal distribution), with density

$$f_F(y; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-(x-1)^2/2\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-(x+1)^2/2\right).$$

When $y < 0$ the densities are folded by taking absolute value of y, *i.e.* $\{f_F(y; 0, 1) = 0 | y < 0\}$. The random variable Y in $f_F(y; 0, 1)$ is equivalent to $Y = |X|$, where X is the random variable in standard normal distribution. The mean of $f_F(y; 0, 1)$ is $\sqrt{2/\pi}$ and variance equals $1 - 2/\pi$. The median of this distribution does not exist.

Log-normal Distribution

Data sets are from standard log-normal distribution with location parameter 0 and scale parameter 1, with density

$$f_{LN}(y; 0, 1) = \frac{1}{x\sqrt{2\pi}} \exp\left(-(\log x - \mu)^2/2\right),$$

which gives mean $\exp(1/2)$ and variance $\exp(1)(1-\exp(1))$. The log-normal variable Y is maximum entropy probability distribution of standard normal distribution, *i.e.* $Y = \log(X)$. The median of this distribution is 1. Moreover, $f_{LN}(y; 0, 1)$ is only non-negatively defined.

Laplace Distribution

This is a control case, where data sets can be correctly estimated, *i.e.* data were generated from standard Laplace distribution (ALD with $\mu = 0$, $\sigma = 1$, and $\tau = 0.5$) with $\mu = 0$ and $\sigma = 1$. While other distribution above represent model misspecified, which are estimated by the ALD with different values of τ.

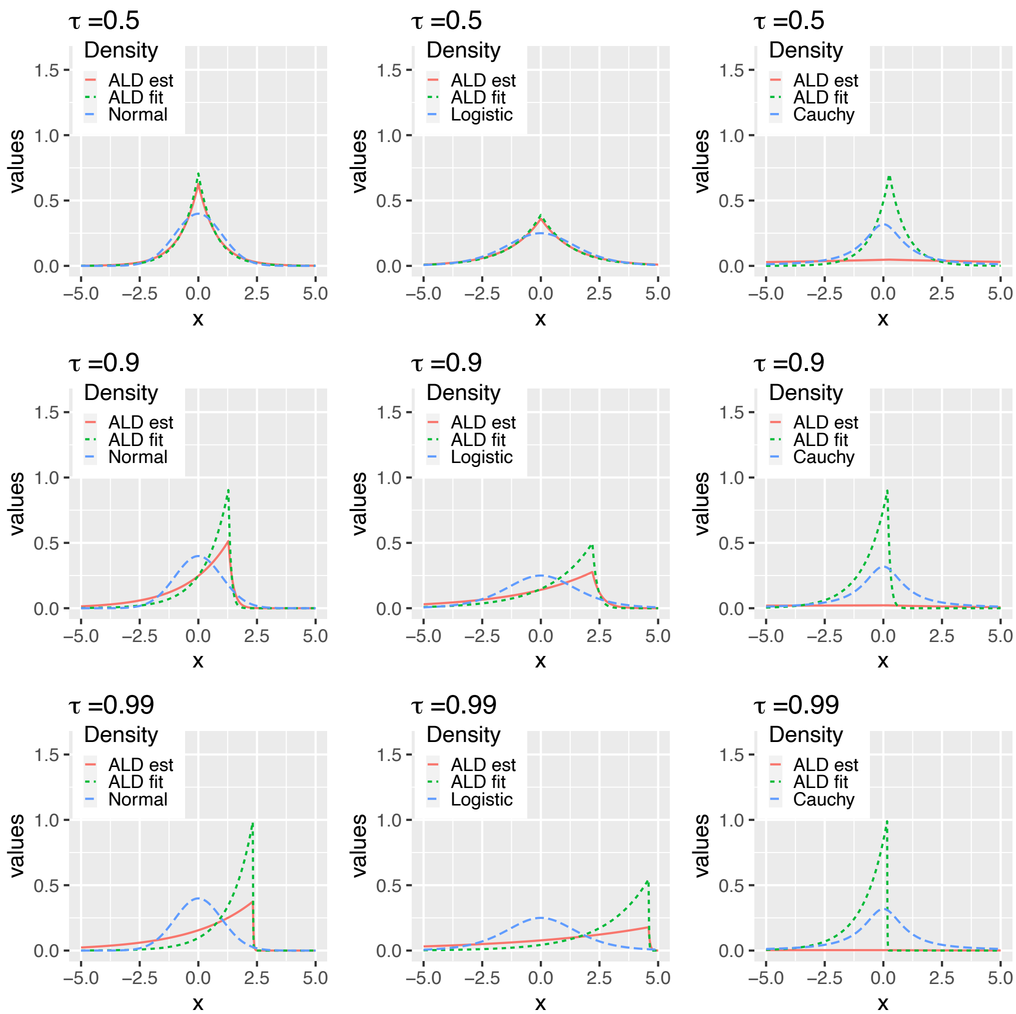


Figure 7.1: Comparison of the simulating normal, logistic and Cauchy distributions (in blue curves) to approximating ALDS with scale parameters obtained by equating variances (in green curves) or as average MLEs (in red curves) with $\tau = 0.5$, 0.9, and 0.99 respetively.

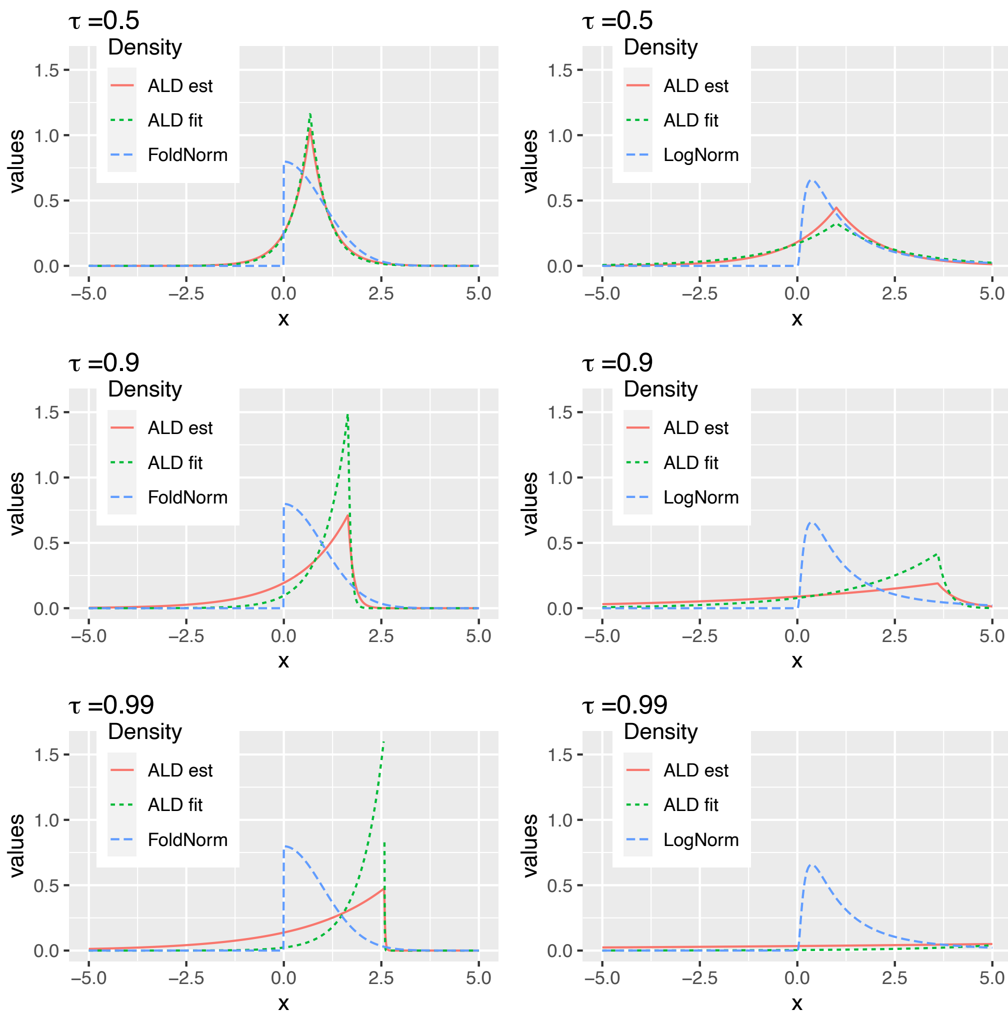


Figure 7.2: Comparation of the simulating folded normal distribution and log-normal distribution (in blue curves) to approximating ALDS with scale parameters obtained by equating variances (in green curves) or as average MLEs (in red curves) with $\tau = 0.5$, 0.9, and 0.99 respetively.

To give a sense for how well the simulating distributions are approximated by an ALD, in Figures 7.1 and 7.2 we plot the true distributions with the 'best' ALD approximations. In each case the ALD approximations had location parameter equal to the τth quantile of the simulating distribution. We used two approaches to get the best scale. The first approach was to determine a scale parameter for the ALD so that its variance matched the variance of the simulating distribution. The second approach took the scale parameter as the average MLE over 1000 samples from the simulating distribution, each of size $n = 5000$.

We illustrate with some examples. The variance of an AL variable is $\sigma^2((1-\tau)^2+\tau^2)/(\tau^2(1-\tau)^2)$, which equals $8\sigma^2$ when $\tau = 0.5$. The variance of the standard normal variable is 1,so setting the scale of the ALD to $\sigma = 0.3981$. The variance of the standard logistic variable equals $3/\pi^2$, which means $\sigma = 0.6413$. Since the variance of Cauchy variable is undefined, then we set $\sigma = 0.3536$, a rationale being that a Cauchy variable is equivalent to the ratio of two independent standard normal variables. When $\tau = 0.5$, the ALD fits reasonably well although the approximation to the Cauchy is not as good, particularly using the average MLEs as scale parameters. For $\tau = 0.9$ and 0.99, the ALDs as substantially different from the simulating distributions. Part of the reason for this is that the ALD always has its mode at the τth quantile, which doesn't coincide with the mode of the simulating distribution. The other is that the scale parameters tended to be large. The ALD is a light tailed distribution and so it appears that approximating heavier tailed distributions requires larger scale parameters.

In the case $\tau = 0.5$, setting the variance of the ALD to the variance of the folded normal, $1 - 2/\pi$, gives $\sigma = 0.2131$. Equating the variance of the ALD to that of the log-normal gives ALD $\sigma = 0.764$ because the variance of the log normal is $(e^1 - 1)e^1$. As it shown in Figure 7.2, when $\tau = 0.5$ the densities fitting ALDs are closed to the orignial densities. However, the differences become substantial when $\tau = 0.9$ and 0.99. In brief, the difficulties in approximating these distributions using an ALD tend to increase as τ increases.

Note that, in Figure 7.1 the estimated ALDs for Cauchy distribution are almost flat lines are because the comparatively large variance of Cauchy random samples from R, which are 5.25, 4.05, and 3.13 with respect to $\tau = 0.5$, 0.9, and 0.99.

7.2 Simulation Results

Given all the setting in last section, the results are shown as following. For each distribution, one table indicates the coverages of five estimating methods with different values of τ and n with respect to 0.1, 0.05, 0.01 significant levels. The other table shows the average of simulated 95% CIs and its widths corresponding to the first table. The tables of details of 90% and 99% CIs are in Appendix.

Asymmetric Laplace Distribution

For the case that there is no model misspecification, when data were generated from standard Laplace distribution and estimated by ALD with $\mu = 0$ and $\sigma = 1$ ($ALD(0, 1)$) when $\tau = 0.5$, 0.9, and 0.99 and significant level is 0.05, the estimation results are shown in Table 7.1 (coverages) and 7.2 (95% confidence intervals). In the case that $\tau = 0.5$, the coverages of all five methods were all close to the nominal when the sample size was larger than 100. When $n = 100$, the percentile-t and Wald intervals tend to undercover but the other intervals have coverage close to their stated levels.

However, when we generated data from $ALD(0, 1)$ with $\tau = 0.9$, the coverages of the Wald test and percentile-t method are below their stated levels when $n = 100$ and 500, whereas the other methods give accurate coverages. When $n = 1000$ all five methods showed accurate coverages.

When we raised τ to 0.99, the nonparametric bootstrap LRT performed well as it did in all settings considered for the ALD generating scenario. The Wald test, percentile method and percentile-t method gave coverages that were too small, especially when n is small ($n < 1000$). The chi-square LRT are are very accurate for $n \geq 500$. Overall, the 95% percentile-t CIs are slightly wider than the 95% percentile CIs than the 95% Wald CIs. The 95% chi-square and nonparametric bootstrap CIs are wider than the other CIs. In summary, for this setting with no misspecification, the LRT test, whether using bootstrap or chi-square thresholds is recommended. It performed well across all scenarios. The percentile bootstrap approach performed well but struggled with large quantiles. The Wald and percentile-t approaches were the worst performers particularly with smaller sample sizes.

Normal Distribution

Table 7.3 and 7.4 gives the coverages and CIs of data generated from standard normal distribution and estimated by $ALD(0, 1)$ when $\tau = 0.5$, 0.9, and 0.99 respectively. When $\tau = 0.5$, the coverages of nonparametric bootstrap LRT and percentile method are nearly $(1 - \alpha)100$, while the Wald and percentile-t methods gave coverages that were too small. The coverages of Wald test are very small relative to their stated values. When $\tau = 0.5$ the 95% nonparametric bootstrap CIs are substantially

Table 7.1: The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard Laplace distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR).

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.5$					
100	88/93/98	90/96/99	85/91/96	90/96/99	90/95/99
500	90/95/90	90/95/99	88/93/98	91/95/99	89/94/99
1000	90/95/89	90/94/99	90/94/99	90/95/99	90/95/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.9$					
100	84/90/87	87/94/98	81/89/96	89/94/99	90/94/99
500	87/93/89	89/95/99	85/91/98	89/95/99	90/95/99
1000	90/94/89	89/95/99	88/93/97	89/94/99	89/95/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.99$					
100	71/77/87	63/63/63	74/81/89	86/93/98	87/91/97
500	82/88/85	85/85/97	84/87/96	89/95/99	89/95/99
1000	85/91/84	84/92/97	82/89/96	89/94/99	90/95/99

wider than other CIS, and the 95% Wald CIs are tightest among them. Although having tighter confidence intervals is desirable, when coverages are close to nominal, the Wald tests tended to undercover dramatically.

When we estimated the original data with $\tau = 0.9$ in $ALD(0, 1)$, the Wald test gave about 40% coverages when $\alpha = 0.1$, 50% coverages when $\alpha = 0.05$, and about 60% coverages when $\alpha = 0.01$. The chi-squared LRT has slight better results than the Wald test, where chi-square has coverages 69%, 78%, and 90% when $\alpha = 0.1$, 0.05, and 0.01 respectively and $n = 1000$. Although the coverages of the percentile-t method are too small when $n = 100$ and 500, they are almost correct (87/93/99) when $n = 1000$. The coverages of the percentile method and the nonparametric bootstrap LRT are correct even $\tau = 0.9$. The 95% CIs of the nonparametric bootstrap when $\tau = 0.9$ are substantially wider than others, the CIs of the percentile-t method and the chi-square LRT are wider than percentile CI when n is small, and shrink to be tighter than percentile CIs when n approached 1000. However, as mentioned, this comes at the cost of coverage that is far from the nominal levels.

When $\tau = 0.99$, only a small number of Wald test results can cover the μ, which equals 2.33. The coverages of the chi-square LRT are better than the Wald test, but still below their stated levels. The percentile and percentile-t method are performed well when $n = 1000$, with coverages of 84/92/96 and 77/85/91 respectively for 90/95/99% CIs. The coverages of the nonparametric bootstrap LRT are close to and slight better than the percentile coverages when $n = 1000$. However, when n is smaller, the nonparametric bootstrap LRT performs better with coverages of 81/86/94 when $n = 100$ and 85/90/97 when $n = 500$. The widths of the 95% nonparametric bootstrap CIs are 1.15, 0.82 and 0.61 with respect to $n = 100$, 500, and 1000 when $\tau = 0.99$. Under this extreme situation, the widths of the 95% percentile and percentile-t CIs are all 0.46 when $n = 1000$. Although their widths are smaller than the bootstrap LRT, their coverages are smaller too. In summary, with model misspecification, the chi-square test no longer performs well. The Wald and percentile-t appraches continue to fare poorly. The percentile and bootstrap LRT are recommended. They both perform well although both approaches tend to undercover a little when $\tau = 0.99$. The percentile approach is slightly preferred in a number of settings because it tends to produce smaller confidence sets with roughly the same coverage as the bootstrap LRT. However, it undercovered to a much greater degree than the LRT when $\tau = 0.99$ and $n = 100$.

Logistic Distribution

When data are generated from a standard logistic distribution, the coverages in table 7.5 shows similar patterns to the coverages in table 7.3, where the data were generated from a $N(0, 1)$ distribution. This may be because the distributions are somewhat similar (Figure 7.1). From table 7.5, we can conclude the coverages of nonparametric bootstrap and percentile CIs are both robust to these forms of model misspecification. However, table 7.6 indicates although their coverages are similar, the nonparametric bootstrap CIs are wider than percentile CIs regardless the value of τ. One should note, however, that the percentile method uncovered substantiially when $n = 100$ and $\tau = 0.99$.

Cauchy Distribution

Because the Cauchy distribution does not have a mean, it may be expected that the properties of the methods will be unusual for this setting. As indicated by Figure 7.1, the ALD approximations for this distribution are not close. From Table 7.7 we see that the Wald test and chi-square LRT have 100% coverages when $\tau = 0.5$. However when $\tau = 0.9$, the coverages of the Wald test are 61/67/78 ($n = 100$) and 77/83/90 ($n = 1000$), and the coverages of the chi-square LRT are 77/85/93 ($n = 100$) and 83/89/96 ($n = 1000$). When $\tau = 0.99$, the coverages of these two methods are both very small with the Wald test giving particularly small coverages. The percentile and percentile-t method have similar coverages, which fairly accurate when $\tau = 0.5$ and 0.9. However, the coverages are noticeably smaller when $\tau = 0.99$, particularly when $n = 100$. This setting is one where the bootstrap LRT is clearly preferred. It has coverage that is close to nominal in almost all cases. The sole exception is when $\tau = 0.99$ and $n = 100$. In this it undercovers although it comes closer to having the right coverage than any of the other methods. The price paid is an extreme amount of uncertainty about what the true quantile is (Table 7.8).

The explanation for why the Cauchy is particularly problematic for most of the methods has to do with its large range, i.e. $[-4217.948, 2200.813]$, when $\tau = 0.99$ while other distributions are all between $[-8, 8]$, not only the estimated standard error but also the LRT statistics are large. Due to this reason, the methods that use constant thresholds or critical values like the Wald test and chi-square give thresholds that are too small. The percentile-t CIs have extreme large ranges also due to the large amount of variability in the estimated standard errors, due to highly variable and large $\hat{\sigma}$. The percentile and nonparametric bootstrap CIs are more comparable.

In conclusion, the nonparametric bootstrap LRT gives good coverage even with extreme value of τ (as long as n is 500 or more). Although the percentile and percentile-t methods had difficulties with $\tau = 0.99$, when $\tau = 0.05$ or 0.90 they provide reasonable estimations.

Folded normal Distribution

Because the folded normal distribution is only defined on the positive axis it makes sense that tests that make Gaussian assumptions about parameter estimates the Wald test, might not perform well. Table 7.9 shows that the Wald test has coverages that are very low even when $\tau = 0.5$. Although not as dramatic as for the

Wald test, the coverages of the chi-square LRT are also far too low. The coverages of the percentile-t are reasonable when $\tau = 0.5$ and $n \geq 500$ but are too low otherwise, often substantially so. When $\tau = 0.99$, the Wald, percentile, percentile-t, and chi-square coverages are all substaintly less than 90/95/99. The percentile and bootstrap LRT perform well for $\tau = 0.90$ and 0.5 and $n \geq 500$. They tend to undercover otherwise although the bootstrap LR test comes closer to the correct coverage.

From Table 7.10 we can conclude, the nonparametric bootstrap CIs are the widest, and wider than other methods as n increase. The percentile and percentile-t share the similar CIs, which are slight tighter than the nonparametric bootstrap CIs, and the chi-square CIs are even tighter. Although it may seem desirable to have tighter confidence intervals the coverages for these approaches is very low. In summary, the percentile and bootstrap LRT are the best performers with the bootstrap LRT giving better coverage when τ is large.

Log normal Distribution

For the percentile, percentile-t and bootstrap LRT, Table 7.11, which gives coverages, shows very similar patterns, over choices of n and τ, to the corresponding folded normal settings. The Wald test gave better coverages when $\tau = 0.5$ than the Wald coverages in folded normal distribution, but similarly had coverages that were substantially too small when $\tau = 0.9$ or 0.99. The chi-square LRT performed reasonably when $\tau = 0.5$ but had coverages that were far too small when $\tau = 0.9$ or 0.99. In terms of coverage, once again, the bootstrap LR test and percentile method are comparable with the bootstrap LR test doing better when τ is large. It is worth noting, however, that in cases where coverage was good for both methods, the percentile method tended to give smaller intervals (Table 7.12).

Table 7.2: The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Laplace distribution.

$\tau = 0.5$

n	μ	Wald	Perc	PercT
100	0	-0.41/0.37/0.78	-0.43/0.44/0.87	-0.48/0.41/0.89
500	0	-0.18/0.17/0.35	-0.19/0.17/0.36	-0.18/0.18/0.36
1000	0	-0.13/0.12/0.25	-0.13/0.12/0.25	-0.13/0.13/0.26
n	μ	χ^2	BootLR	
100	0	-0.56/0.55/1.11	-0.58/0.53/1.11	
500	0	-0.20/0.22/0.42	-0.21/0.21/0.42	
1000	0	-0.14/0.13/0.27	-0.14/0.14/0.28	

$\tau = 0.9$

n	μ	Wald	Perc	PercT
100	0	-0.73/0.57/1.30	-0.91/0.68/1.59	-0.88/0.76/1.64
500	0	-0.31/0.28/0.59	-0.34/0.29/0.63	-0.32/0.31/0.63
1000	0	-0.22/0.20/0.42	-0.24/0.20/0.44	-0.22/0.22/0.44
n	μ	χ^2	BootLR	
100	0	-1.73/0.78/2.51	-1.72/0.90/2.62	
500	0	-0.37/0.36/0.73	-0.42/0.36/0.78	
1000	0	-0.24/0.23/0.47	-0.26/0.24/0.50	

$\tau = 0.99$

n	μ	Wald	Perc	PercT
100	0	-2.91/1.00/3.91	-5.06/0.40/5.46	-2.52/3.22/5.74
500	0	-1.06/0.70/1.76	-1.43/0.68/2.11	-1.07/1.04/2.11
1000	0	-0.71/0.54/1.25	-0.81/0.60/1.41	-0.77/0.67/1.44
n	μ	χ^2	BootLR	
100	0	-6.91/1.91/8.82	-7.45/0.40/7.85	
500	0	-1.81/1.22/3.03	-1.93/1.06/2.99	
1000	0	-1.25/0.76/2.01	-1.35/0.76/2.11	

Table 7.3: The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard normal distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR).

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.5$					
100	71/79/90	91/95/98	83/90/96	83/90/97	90/94/98
500	72/82/92	91/95/99	89/93/98	83/90/97	90/95/99
1000	69/80/90	91/95/99	87/92/97	82/90/96	90/95/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.9$					
100	40/49/62	87/93/97	78/83/92	64/75/86	87/92/98
500	43/50/62	91/95/98	85/91/97	66/74/89	90/95/98
1000	44/52/64	91/95/99	87/93/98	69/78/90	91/95/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.99$					
100	08/09/12	65/65/65	61/68/76	45/48/54	81/86/94
500	08/09/13	87/88/96	76/80/90	31/37/48	85/90/97
1000	09/10/14	84/92/96	77/85/91	32/39/49	84/92/97

Table 7.4: The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.17/0.14/0.31	-0.26/0.23/0.49	-0.26/0.23/0.49
500	0	-0.07/0.07/0.14	-0.11/0.11/0.22	-0.11/0.11/0.22
1000	0	-0.05/0.05/0.10	-0.08/0.07/0.15	-0.08/0.08/0.16
n	$\hat{\mu}$	χ^2	BootLR	
100	0	-0.32/0.22/0.54	-0.34/0.29/0.63	
500	0	-0.10/0.10/0.20	-0.14/0.12/0.26	
1000	0	-0.08/0.07/0.15	-0.09/0.08/0.17	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1.28	1.13/1.36/0.23	0.93/1.59/0.66	0.92/1.60/0.68
500	1.28	1.22/1.32/0.10	1.13/1.42/0.29	1.13/1.43/0.30
1000	1.28	1.24/1.31/0.07	1.17/1.38/0.21	1.17/1.38/0.21
n	$\hat{\mu}$	χ^2	BootLR	
100	1.28	0.91/1.62/0.71	0.80/1.69/0.89	
500	1.28	1.15/1.38/0.23	1.09/1.46/0.37	
1000	1.28	1.20/1.36/0.16	1.15/1.40/0.25	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.33	2.10/2.20/0.10	1.59/2.50/0.91	1.77/2.88/1.11
500	2.33	2.26/2.31/0.05	1.99/2.56/0.57	2.01/2.61/0.60
1000	2.33	2.29/2.32/0.03	2.08/2.54/0.46	2.08/2.54/0.46
n	$\hat{\mu}$	χ^2	BootLR	
100	2.33	1.38/2.39/1.01	1.35/2.50/1.15	
500	2.33	2.00/2.38/0.38	1.87/2.69/0.82	
1000	2.33	2.15/2.38/0.23	2.00/2.61/0.61	

Table 7.5: The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard logistic distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR).

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.5$					
100	75/84/92	90/96/99	83/88/96	84/92/97	89/95/98
500	77/84/93	91/95/99	87/92/97	85/91/97	90/95/99
1000	76/84/93	90/94/99	89/92/97	84/90/96	90/94/98

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.9$					
100	39/46/58	87/93/98	77/85/92	63/72/86	87/93/98
500	41/49/62	89/95/99	83/89/95	67/75/86	88/95/99
1000	41/48/60	89/94/99	86/91/96	65/74/86	88/94/98

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.99$					
100	06/08/10	63/63/63	62/67/76	42/44/49	80/86/94
500	07/09/11	85/85/97	75/79/91	26/31/41	86/93/98
1000	08/09/12	85/92/97	77/85/93	27/31/41	87/98/94

Table 7.6: The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard logistic distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.29/0.25/0.54	-0.40/0.38/0.78	-0.42/0.37/0.79
500	0	-0.13/0.12/0.25	-0.18/0.17/0.35	-0.18/0.17/0.35
1000	0	-0.09/0.08/0.17	-0.13/0.12/0.25	-0.13/0.12/0.25

n	$\hat{\mu}$	χ^2	BootLR
100	0	-0.50/0.44/0.94	-0.55/0.48/1.03
500	0	-0.18/0.17/0.35	-0.22/0.20/0.42
1000	0	-0.12/0.11/0.23	-0.15/0.14/0.29

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.20	1.93/2.35/0.42	1.55/2.85/1.30	1.50/2.83/1.33
500	2.20	2.09/2.28/0.19	1.90/2.48/0.58	1.91/2.49/0.58
1000	2.20	2.12/2.26/0.14	1.99/2.40/0.41	1.99/2.40/0.41

n	$\hat{\mu}$	χ^2	BootLR
100	2.20	1.44/2.77/1.33	1.33/3.07/1.74
500	2.20	1.90/2.44/0.54	1.83/2.56/0.73
1000	2.20	2.02/2.32/0.30	1.95/2.44/0.49

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	4.60	4.06/4.26/0.20	2.86/5.14/2.28	3.12/6.07/2.95
500	4.60	4.44/4.54/0.10	3.74/5.28/1.54	3.74/5.39/1.65
1000	4.60	4.51/4.58/0.07	3.98/5.19/1.21	3.94/5.17/1.23

n	$\hat{\mu}$	χ^2	BootLR
100	4.60	2.50/4.75/2.25	2.35/5.14/2.79
500	4.60	3.89/4.70/0.81	3.46/5.66/2.20
1000	4.60	4.19/4.71/0.52	3.78/5.40/1.62

Table 7.7: The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard Cauchy distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR).

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.5$					
100	99/100/100	90/95/99	88/95/98	99/100/100	92/96/99
500	100/100/100	92/95/99	92/95/98	99/100/100	92/96/99
1000	100/100/100	91/94/99	92/96/99	100/100/100	92/96/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.9$					
100	61/67/78	88/94/99	86/92/97	77/85/93	91/95/99
500	72/78/87	88/93/99	87/93/98	81/88/95	90/95/99
1000	77/83/90	88/93/98	89/93/98	83/89/96	90/94/98

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.99$					
100	06/07/09	63/63/63	62/68/74	39/39/42	78/83/89
500	09/10/13	86/87/96	83/88/95	29/34/42	90/94/97
1000	11/13/16	87/93/97	87/91/96	29/34/43	91/95/98

Table 7.8: The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Cauchy distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-1.55/1.45/3.00	-0.23/0.17/0.40	-0.47/0.14/0.61
500	0	-0.81/0.80/1.61	-0.14/0.14/0.28	-0.27/0.27/0.54
1000	0	-0.65/0.65/1.30	-0.10/0.10/0.20	-0.18/0.17/0.35
n	$\hat{\mu}$	χ^2	BootLR	
100	0	-0.15/0.17/0.32	-0.24/0.19/0.43	
500	0	-0.21/0.18/0.39	-0.18/0.17/0.35	
1000	0	-0.14/0.14/0.28	-0.12/0.12/0.24	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	3.08	0.85/5.33/4.48	1.68/6.35/4.67	-1.86/9.86/11.72
500	3.08	1.94/4.24/2.30	2.35/4.11/1.76	1.48/5.08/3.60
1000	3.08	2.24/3.92/1.68	2.54/3.76/1.22	2.02/4.24/2.22
n	$\hat{\mu}$	χ^2	BootLR	
100	3.08	1.38/7.78/6.40	1.28/8.68/7.40	
500	3.08	2.18/4.24/2.06	2.14/4.58/2.44	
1000	3.08	2.30/4.05/1.75	2.39/4.03/1.64	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	31.82	27.22/36.87/9.65	6.38/244.15/237.77	-185.54/872.40/1057.94
500	31.82	29.19/34.66/5.47	14.56/78.39/63.83	-40.98/168.46/209.44
1000	31.82	29.98/33.88/3.90	17.93/65.00/47.07	-8.65/96.62/105.27
n	$\hat{\mu}$	χ^2	BootLR	
100	31.82	5.27/66.17/60.90	3.92/244.15/240.23	
500	31.82	15.15/53.20/38.05	10.67/144.71/134.04	
1000	31.82	16.70/53.98/37.28	14.44/96.16/81.72	

Table 7.9: The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard folded distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR).

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.5$					
100	68/76/87	89/94/98	83/88/94	79/87/95	87/93/98
500	71/79/91	93/96/99	89/94/98	84/91/97	93/96/99
1000	68/76/89	91/96/99	87/94/98	82/88/96	91/96/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.9$					
100	36/42/53	84/92/98	78/85/93	61/68/81	86/93/98
500	35/42/53	89/94/98	82/88/95	61/70/83	87/94/98
1000	34/40/51	89/94/99	85/91/96	59/68/84	90/94/98

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.99$					
100	08/09/12	63/63/63	64/70/78	43/45/50	81/86/94
500	07/08/12	86/87/96	75/81/90	29/33/42	85/92/98
1000	07/09/13	84/92/98	77/84/92	31/35/44	86/92/97

Table 7.10: The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard folded normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0.67	0.52/0.71/0.19	0.52/0.75/0.23	0.49/0.71/0.22
500	0.67	0.63/0.71/0.08	0.60/0.74/0.14	0.60/0.74/0.14
1000	0.67	0.64/0.70/0.06	0.62/0.72/0.10	0.62/0.72/0.10

n	$\hat{\mu}$	χ^2	BootLR
100	0.67	0.55/0.71/0.16	0.51/0.75/0.24
500	0.67	0.61/0.74/0.13	0.59/0.76/0.17
1000	0.67	0.63/0.72/0.09	0.62/0.73/0.11

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1.64	1.53/1.70/0.17	1.35/1.91/0.56	1.35/1.93/0.58
500	1.64	1.60/1.68/0.08	1.51/1.77/0.26	1.52/1.77/0.25
1000	1.64	1.61/1.67/0.06	1.55/1.73/0.18	1.55/1.73/0.18

n	$\hat{\mu}$	χ^2	BootLR
100	1.64	1.27/1.84/0.57	1.25/2.00/0.75
500	1.64	1.54/1.74/0.20	1.48/1.80/0.32
1000	1.64	1.58/1.70/0.12	1.53/1.75/0.22

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.58	2.37/2.45/0.08	1.91/2.74/0.83	2.07/3.12/1.05
500	2.58	2.52/2.56/0.04	2.27/2.80/0.53	2.29/2.85/0.56
1000	2.58	2.55/2.57/0.02	2.36/2.77/0.41	2.36/2.78/0.42

n	$\hat{\mu}$	χ^2	BootLR
100	2.58	1.85/2.62/0.77	1.70/2.74/1.04
500	2.58	2.25/2.62/0.37	2.16/2.92/0.76
1000	2.58	2.36/2.62/0.26	2.29/2.84/0.55

Table 7.11: The percentages of times (coverage) that the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) was in 90/95/99% confidence intervals when data were generated from the standard log normal distribution. Method abbreviations correspond to the Wald test (Wald), the percentile bootstrap confidence interval (Perc), the percentile-t bootstrap confidence interval (PercT), the LR test using chi-squared thresholds (χ^2) and the LR test using bootstrap thresholds (BootLR).

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.5$					
100	86/91/97	90/95/98	83/89/96	89/94/98	89/95/98
500	88/94/98	91/95/99	89/94/97	89/95/99	90/95/99
1000	87/92/98	91/95/99	87/93/97	89/94/98	90/95/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.9$					
100	30/36/48	87/93/97	79/86/94	56/65/80	87/92/97
500	33/38/49	91/95/98	85/92/97	60/67/82	89/95/99
1000	35/41/52	91/95/99	88/93/98	62/71/84	91/95/99

n	Wald	Perc	PercT	χ^2	BootLR
$\tau = 0.99$					
100	04/04/06	65/65/65	63/69/78	42/43/45	80/85/92
500	04/05/06	87/88/96	78/82/91	24/27/33	86/91/97
1000	04/05/07	84/92/96	79/85/93	23/27/35	85/92/97

Table 7.12: The average lower bound/upper bound/width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard log normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1	0.82/1.24/0.42	0.89/1.16/0.27	0.91/1.19/0.28
500	1	0.90/1.10/0.20	0.90/1.11/0.21	0.89/1.11/0.22
1000	1	0.93/1.07/0.14	0.92/1.08/0.16	0.92/1.08/0.16

n	$\hat{\mu}$	χ^2	BootLR
100	1	0.88/1.17/0.29	0.88/1.17/0.29
500	1	0.88/1.14/0.26	0.87/1.14/0.27
1000	1	0.91/1.09/0.18	0.91/1.09/0.18

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	3.60	3.22/3.83/0.61	2.56/5.00/2.44	2.29/4.93/2.64
500	3.60	3.45/3.72/0.27	3.10/4.16/1.06	3.06/4.15/1.09
1000	3.60	3.48/3.68/0.20	3.23/3.98/0.75	3.21/3.97/0.76

n	$\hat{\mu}$	χ^2	BootLR
100	3.60	2.47/4.66/2.19	2.26/5.65/3.39
500	3.60	3.16/3.96/0.80	2.98/4.34/1.36
1000	3.60	3.34/3.85/0.51	3.16/4.07/0.91

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	10.24	8.77/9.24/0.47	5.01/13.51/8.50	4.35/19.58/15.23
500	10.24	9.82/10.05/0.23	7.37/13.22/5.85	6.91/13.72/6.81
1000	10.24	9.98/10.15/0.17	8.08/12.82/4.74	7.68/12.65/4.97

n	$\hat{\mu}$	χ^2	BootLR
100	10.24	3.97/10.97/7.00	3.95/13.51/9.56
500	10.24	7.42/10.54/3.12	6.53/15.39/8.86
1000	10.24	8.75/10.52/1.77	7.45/13.94/6.49

Chapter 8

Quantile Regression

Quantile regression is a regression estimation method of how quantiles of a population vary as a function of covariates [16]. In the case that $\tau = 0.5$, quantile regression minimizes the weighted absolute residuals instead of the sum of squared residuals. This gives it some robustness to outlying observations [21]. Importantly, however, it is estimating something different in this case, the median responses as a function of the covariates rather than the mean responses. More generally, when τ is large or small, quantile regression allows one to model how extremes tend to vary with covariates. As such it is a good tool for modeling how, for instance, extreme weather patterns or longevity depend on covariates.

8.1 Nonparametric Bootstrapping on Quantile Regression

The quantile regression model can be expressed as

$$y_i = \mathbf{x_i}'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, ..., n$$

where y_i represents the i^{th} obsevation, $\boldsymbol{x}_i$ represents the covariate vector for the i^{th} individual, and ϵ_i is the ith independent error variable having an ALD with 0 location parameter. Thus $\boldsymbol{x}_i'\boldsymbol{\beta}$ is interpretable as the τth quantile of the conditional distribution of y_i, conditional upon the ith individual having covariate value $\boldsymbol{x}_i$, Thus β_j can be interpreted as the change in the τth quantile due to a unit change in x_j, holding all other values in $\boldsymbol{x}$ fixed.

We can write the likelihood function of $\boldsymbol{\beta}$ as

$$\begin{aligned} L(\boldsymbol{\beta};, \sigma, \tau) &= \prod_{i=1}^{n} f(\epsilon_i; \sigma, \tau) \\ &= \left(\frac{\tau(1-\tau)}{\sigma}\right)^n \exp\left(-\sum_{i=1}^{n} \rho_\tau(\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma})\right) \end{aligned}$$

where the quantile loss function is

$$\rho_\tau(u) = \begin{cases} (\tau - 1)u & u < 0 \\ \tau u & u \geq 0 \end{cases}$$

Then the log likelihood function can be written as

$$l(\boldsymbol{\beta}; \sigma, \tau) = n \log \tau(1-\tau) - n \log \sigma - \sum_{i=1}^{n} \rho_\tau \left(\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma} \right)$$

Thus, under quantile regression model, minimizing weighted absolute residuals is equivalent to solving the maximum likelihood function of $\boldsymbol{\beta}$. Thus,

$$\begin{aligned} \arg\max_{\boldsymbol{\beta}} l(\boldsymbol{\beta}; \sigma, \tau) &\Leftrightarrow \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau \left(\frac{y_i - \mathbf{x_i}'\boldsymbol{\beta}}{\sigma} \right) \\ &\Leftrightarrow \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau (y_i - \mathbf{x_i}'\boldsymbol{\beta}) \end{aligned}$$

Numerical algorithms are required to obtain the minimizer, $\hat{\text{fi}}$. Similarly as in Equation 5.6, one obtains that

$$\hat{\sigma}(\hat{\boldsymbol{\beta}}) = -\frac{\sum_{y_i - \mathbf{x_i}'\hat{\boldsymbol{\beta}} < \mu} (y_i - \mathbf{x_i}'\hat{\boldsymbol{\beta}} - \hat{\mu})}{n} + \tau \frac{\sum (y_i - \mathbf{x_i}'\hat{\boldsymbol{\beta}} - \hat{\mu})}{n}$$

We now describe how to obtain the confidence interval for the jth coefficient, β_j, using a LRT, in a quantile regression. First $\mathbf{y} - \beta_j \mathbf{x_j}$ is regressed on the rest regressors. The log likelihood coming from this quantile regression is $l(\hat{\boldsymbol{\beta}}(\beta_j),\ \beta_j)$, where $\hat{\boldsymbol{\beta}}(\beta_j)$ is the estimated coefficients of the regression model holding β_j fixed. The log likelihood obtained by regressing $\boldsymbol{y}$ on all of the regressors is denoted by $l(\hat{\boldsymbol{\beta}})$. Then the LRT statistic is

$$W_j = 2(l(\hat{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}}(\beta_j),\ \beta_j)).$$

In bootstrapping simulation, each bootstrap data set is randomly selected from the original data by entry with replications, each entry represent one independent variable y and the dependent variables $\boldsymbol{x}$. The size of bootstrap data is same to the size of original data, n. Let $\mathbf{y}^*$ denote the independent variable of bootstrap data and $\boldsymbol{x}^*$ represent the dependent variables of bootstrap data respectively. Let $\hat{\boldsymbol{\beta}}^*$ denote the bootstrap coefficients obtained by regressing $\mathbf{y}^*$ on $\boldsymbol{x}^*$.

Suppose the bootstrap coefficients are estimated with the restriction that $\hat{\beta}_j$ is fixed. Then we can obtain $\mathbf{y}^*(\hat{\beta}_j)$ by subtracting the $\boldsymbol{x}_j^{*'}\hat{\beta}_j$ from $\mathbf{y}^*$. The bootstrap coefficients estimated by regressing $\mathbf{y}^*(\hat{\beta}_j)$ on other independent variables except $\mathbf{x}_j$ is denote as $\hat{\boldsymbol{\beta}}^*(\hat{\beta}_j)$, and the associated log likelihood function is $l(\hat{\boldsymbol{\beta}}^*(\hat{\beta}_j),\ \hat{\beta}_j)$. Let the associated log likelihood function when there are no restrictions be $l($

$\beta^*)$. Then the LRT statistics of the bootstrap data is

$$\beta^*) - l(\hat{\beta}^*(\hat{\beta}_j),\ \hat{\beta}_j)).W_j^* = 2(l($$

The α^{th} threshold of nonparametric bootstrap LRT under null hypothesis of β_j^* is denoted as $W_{j,\alpha}^*$. The $100(1-\alpha)\%$ bootstrap CI of $\hat{\beta}_j$ is

$$\left[\min\{\beta_j^* | W_j \le W_{j,\alpha}^*)\}, \max\{\beta_j^* | W_j^* \le W_{j,\alpha}^*\}\right].$$

The percentile CI of $\hat{\beta}_j$ corresponding to original data is

$$\left[\hat{\beta}_{j,\alpha/2}^*, \hat{\beta}_{j,1-\alpha/2}^*\right],$$

where $\hat{\beta}_{j,\zeta}$ denotes the ζth quantile of the estimated $\hat{\beta}_j$ over bootstrap samples.

8.2 Real Data Example

The most common form of regression is mean regression, which only fits the mean of relationships between dependent variable and independent variables. By contrast, the quantile regression is used to assess the relationships between any particular quantile of the dependent variable and independent variables. This feature of quantile regression gives advantages in estimating data where it is the more extreme responses that are of interest, like salaries of same position [30], mortality rates [17], and etc.

One example of mortality rate data is the medfly data used by Koenker and Geling (2001)[17]. The medfly data from the Carey et al's experiment [4], which is known as "flies" in R package "REBayes" [18]. They monitored mortalities of Mediterranean fruit flies (*Ceratitis capitata*), where a cohort of more than 1 million medflies are used. One interesting feature of this data was that only 1% of the population survived past 50 days but mortality (the rate of death, given survival to the time point of interest) decreased dramatically from 15% to less than 5% by 100 days. Interest therefore is how the upper quantiles of age were dependent on explanatory variables.

This data frame includes 19072 observations and 17 variables. Koenker and Geling's weighted quantile regression model set the logarithm of ages in days of medflies ("age" in "flies" data) as the response. One of the covariates of substantial interest in their model is the sex the medflies ("sex" in "flies" data), since Carey et al. (1992) found out that male medflies have higher mortality rate during age 20 to 60 days. The variable "sex" has value female= 1 and male= 0. Other covariates included initial density of flies across cages ("begin"), initial proportion of male in each cage ("prebegin"), and five classes of pupae size in each cage ("size"), i.e. sorted by pupal sorter from 4mm to 8mm, are also included in their model. The last regressor been considered is batches ("batch"), i.e. pupaes are rised in eight distinct batches.

Two different quantile regression models, Model A and Model B, are established in Koenker's paper [17]. In this thesis, we will follow the same models, where model A regresses logarithm of age on female (indicator variable), pupal size, initial cage density and initial proportion. The model can be expressed as:

$$log(age) = \beta_{0,A} + \beta_{1,A} female + \beta_{2,A} size + \beta_{3,A} begin + \beta_{4,A} prebegin$$

Model B considered the effect of 8 different batches on logrithm of age, where we added seven indicator variables to represent batch 1 to batch 7 respectively.

$$\begin{aligned} log(age) = \beta_{0,B} + &\beta_{1,B} female + \beta_{2,B} size + \beta_{3,B} begin + \beta_{4,B} prebegin \\ &+ \beta_{5,B} batch1 + \beta_{6,B} batch2 + \beta_{7,B} batch3 + \beta_{8,B} batch4 \\ &+ \beta_{9,B} batch5 + \beta_{10,B} batch6 + \beta_{11,B} batch7 \end{aligned}$$

A total of 5000 sets of bootstrap samples are generated from the original data to model A and model B repectively. Compared to Koenker and Geling's results, when τ equals 0.5, 0.9, and 0.99 our results are close with wider CI.

Table 8.1 shows that when $\tau = 0.5$, while the quantile regression estimates the mean of response variable, the nonparametric boostrap and percentile CIs are close to each other. The coefficients of model A with the original medfly data are -0.1964 ($\hat{\beta}_{1,A}$), -0.0018 ($\hat{\beta}_{2,A}$), 0.4915 ($\hat{\beta}_{3,A}$), and 1.0937 ($\hat{\beta}_{4,A}$). The nonparametric CI of $\beta_{1,A}$ is $[-0.1993, -0.1938]$ with width 0.0055, and the percentile CI is slight tighter, which is $[-0.1982, -0.1943]$. The nonparametric and percentile CIs of $\beta_{2,A}$ are $[-0.0029, -0.0006]$ and $[-0.0023, -0.0011]$, where the width of the percentile CI smaller than the nonparametric bootstrap CI by 0.0011. The nonparametric bootstrap CIs of $\beta_{3,A}$ and $\beta_{4,A}$ are also wider than the percentile CIs. When $\tau = 0.9$ and $\tau = 0.9$, the quantile regression estimates the 90th and 99th percentile of response variable instead of mean.

Table 8.1 also indicates that all the nonparametric bootstrap CIs are wider than the percentile CIs, where the coefficients of model A with the original medfly data with $\tau = 0.9$ and 0.99 are all located in their corresponding CIs. The coeffcients of model A with $\tau = 0.9$ are -0.0345 ($\hat{\beta}_{1,A}$), -0.009 ($\hat{\beta}_{2,A}$), 0.4590 ($\hat{\beta}_{3,A}$), and 1.2624 ($\hat{\beta}_{4,A}$), and the coefficients of $\tau = 0.99$ are 0.0614, -0.0151, 0.3777, and 0.8848. Of particular interest is the coefficient for sex. This was negative for $\tau = 0.5$ and 0.90, suggesting that the median or even 90th percentile of lifetime was lower for females than males. For very long-lived flies, however, the results with $\tau = 0.99$ suggest that long lived flies are more likely to be female than male. It may be that other reasons for mortality make females more susceptible in the earlier stages of their life stages. The 95% CIs do not include 0, so the results are significant. Coefficients for the other variables remained relatively constant over different choices of τ.

Table 8.2, 8.3, and 8.4 indicated the coefficients CIs of model B with $\tau = 0.5$, 0.9, and0.99. The estimated coefficients of model B when $\tau = 0.5$ are -0.2048 (female), -0.0045 (size), 0.0371 (begin), 0.0421 (prebegin), and -0.1220, 0.1637, -0.2281, -0.1565, 0.0265, -0.0625, and 0.1335 (batch 1 to 7). Table 8.2 indicates nonparametric bootstrap CIs are all wider than percentile CIs, where in some special cases that differences between them are smaller than 0.01. The widths of nonparametric bootstrap and percentile CIs for β_1 are 0.0094 and 0.0042, for β_2 are 0.0041 and 0.0025, for β_7 are 0.0147 and 0.009, for β_9 are 0.00138 and 0.0076, for β_10 are 0.0116 and 0.0067, and for β_{11} are 0.0138 and 0.0065. Although the difference of widths of CIs of above coefficients are too small to tell in 8.2, the results still indicated the nonparametric bootstrap has wider CIs than the percentile method. The results including batches have not changed the coefficients much.

When $\tau = 0.9$, the first four estimated coefficients of model B with original data are $--0.0103$, -0.0151, 0.0323, and 0.4342. The estimated coefficients for 6 batches are -0.1127, 0.0795, -0.3715, -0.2276, -0.0862, -0.1725, and 0.0676. In Table 8.3, special cases that differences of nonparametric nootstrap CIs and percentile CIs are smaller than 0.01 are β_1 (0.0118 and 0.0080), β_2 (0.0053 and 0.0030), and β_8 (0.0165 and 0.0086). When $\tau = 0.99$, estimated coefficients of model B with original data are -0.0927 (female), -0.0142 (size), 0.0344 (begin), 0.5280 (prebegin), and -0.1835, 0.0052, -0.4249, -0.2842, -0.1368, -0.1769, and 0.0771 (batch 1 to 7). In Table 8.4,the only case that differences of nonparametric bootstrap CIs and percentile CIs are smaller than 0.01 is β_2 (0.0076 and 0.0041).

8.3 Simulation of Quantile Regression

In this section we consider the results of a small simulation study. The model is the one that Feng et al. (2011) used in their simulation study [10] is a multiple regression model with two independent variables. In their model, the dependent variable y_i with $i = 1, \ldots, n$ is generated from

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + 3^{-\frac{1}{2}}[2 + \frac{1}{10}\{1 + (x_{1,i} - 8)^2 + x_{2,i}\}]\epsilon_i, \tag{8.1}$$

where $\beta_0 = \beta_1 = \beta_2 = 1$ and $n = 1000$ in simulations. The residual variable $\boldsymbol{\epsilon}$ is generated from a t distribution with degree of freedom 3 (t_3). The variable $\mathbf{x_1}$ is generated from standard log-normal distribution, and 80% of the indicator indicator variables, $\boldsymbol{x}_2$, were independently random selected to be 1. Let ζ denote the τ^{th} quantile of $\mathbf{y}$, then

$$\begin{aligned}\tau &= P(y_i \leq \zeta | x_{1,i},\ x_{2,i}) \\ &= P\left(\epsilon_i \leq \frac{\zeta - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i}}{3^{-\frac{1}{2}}\left(2 + \frac{1}{10}\left(1 + (x_{1,i} - 8)^2 + x_{2,i}\right)\right)}\right), i \in 1, \ldots, n\end{aligned}$$

Given ϵ is distributed under t_3, then the τ^{th} quantile of ϵ is

$$t_{\tau,3} = \frac{\zeta - \beta_0 - \beta_1 x_{1,i} - \beta_2 x_{2,i}}{3^{-\frac{1}{2}}\left(2 + \frac{1}{10}\left(1 + (x_{1,i} - 8)^2 + x_{2,i}\right)\right)},$$

which gives

$$\zeta = (\beta_0 + 3^{-\frac{1}{2}} * 8.5 * t_{\tau,3}) + (\beta_1 - 3^{-\frac{1}{2}} * 1.6 * t_{\tau,3})x_{1,i}$$
$$+ (\beta_2 + 3^{-\frac{1}{2}} * 0.1 * t_{\tau,3})x_{2,i} + 3^{-\frac{1}{2}} * 0.1 * t_{\tau,3}x_{1,i}^2.$$

In our study, we also add a new variable $x_{3,i} = x_{1,i}^2$, where $i \in 1, \ldots, n$. Then the regression model can be expressed as

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + 3^{-\frac{1}{2}} * 8.5 * t_{\tau,3}\epsilon_i$$
$$- 3^{-\frac{1}{2}} * 1.6 x_{1,i}\epsilon + 3^{-\frac{1}{2}} * \frac{1}{10} x_{2,i}\epsilon_i + 3^{-\frac{1}{2}} * \frac{1}{10} x_{3,i}\epsilon_i,$$

where $\beta_3 = 0$. Because we simulate from $\beta_0 = \beta_1 = \beta_2 = 1$ in (8.1), the using (

$$)thetruecoefficientsforthequantileregressionforquantile$$

τ are quantile of coefficients can be expressed as

$$\beta_{\tau,1} = 1 - 3^{-\frac{1}{2}} * 1.6 * t_{\tau,3}$$
$$\beta_{\tau,2} = 1 + 3^{-\frac{1}{2}} * 0.1 * t_{\tau,3}$$
$$\beta_{\tau,3} = 3^{-\frac{1}{2}} * 0.1 * t_{\tau,3}$$

An example of comparison of percentile CIs and nonparametric CIs of coefficients in our model was taken below.

A total of 1000 data sets were simulated for each setting. To summarize results we consider the average lower and upper bounds of the intervals over the simulations and their width. We also calculate coverage as the proportions of the 1000 data sets where the true values of coefficients were in a confidence interval. For bootstrapping we used 1000 bootstrap samples for each data set.

The simulation results of our model are shown in Table 8.5, 8.6, and 8.7. In Table 8.5, given the true values of β_1, β_2, and β_3 are 1, 1, and 0 when $\tau = 0.5$, good coverage properties are shown by the nonparametric bootstrap LRT. The nonparametric bootstrap coverages of β_1 are 91%, 96% and 99% with respect to 90%, 95%, and 99% CIs. The percentile coverages are 84%, 90%, and 96%, which is a bit off. The widths of nonparametric bootstrap CIs are larger than the percentile CIs, being 0.93/1.02/1.15 compared to 0.66/0.77/0.99. Somewhat larger CIs were expected given the larger coverages. The results for β_2 are similar. The nonparametric bootstrap coverages are 92%, 96% and 99%. The percentile coverages of β_2 are slight better than the percentile coverages of β_1 and as good as the nonparametric bootstrap coverages of β_2. For the percentile method they are 91%, 95%, and 99%. However, the percentile CIs have tighter ranges than nonparametric bootstrap CIs (1.74/1.94/2.28), which are 1.29/1.53/2.00. The last coefficient β_3 has small CIs, where the nonparametric bootstrap CIs are wider (0.11/0.12/0.13) than the percentile CIs (0.07/0.08/0.11). The corresponding nonparametric bootstrap coverages are 88%, 94% and 99%, and the percentile method has worse coverages, which are 79%, 85% and 93%. Overall, the nonparametric bootstrap LRT had better coverage properties than the percentile method.

When $\tau = 0.9$, the true values of β_1, β_2, and β_3 are -0.51, 1.09, and 0.10. The nonparametric bootstrap coverages are 0.87/0.93/0.99 for β_1, 0.93/0.96/0.99 for β_2, and 0.84/0.90/0.96 for β_3. While the percentile coverages are 0.83/0.90/0.97, 0.93/0.95/0.99, and 0.77/0.84/0.93, where the perforamce of β_1 and β_2 are fine, but β_3 has smaller coverages than β_3 in nonoparametric bootstrap. The widths of nonparametric bootstrap CIs are all larger than percentile CIs. Similarly as when $\tau = 0.5$ and 0.90, the nonparametric bootstrap CIs are wider than percentile CIs. When $\tau = 0.99$ as Table 8.7 indicates, the advantages of nonparametic bootstrap CIs over percentile CIs are even larger than in Table 8.6. In summary, although the percentile intervals often performed well and tended to give slightly shorter confidence intervals, it tended to undercover, sometimes substantially so. By comparison, the LRT intervals tended to be closer to the nominal levels.

Table 8.1: For the medfly data, the average lower bound/upper bounds and width of the 95% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) of model A by the nonparametric bootstrap LRT and percentile method.

Method		β_1			β_2	
$\tau = 0.5$	$\hat{\beta}_1$	lb/ub	width	$\hat{\beta}_2$	lb/ub	width
Bootstrap	-0.20	-0.20/-0.19	0.01	0.00	0.00/0.00	0.00
Percentile	-0.20	-0.20/-0.19	0.01	0.00	0.00/0.00	0.00

Method		β_3			β_4	
$\tau = 0.5$	$\hat{\beta}_3$	lb/ub	width	$\hat{\beta}_4$	lb/ub	width
Bootstrap	0.49	0.47/0.50	0.03	1.09	1.09/1.17	0.08
Percentile	0.49	0.48/0.49	0.01	1.09	1.09/1.15	0.06

Method		β_1			β_2	
$\tau = 0.9$	$\hat{\beta}_1$	lb/ub	width	$\hat{\beta}_2$	lb/ub	width
Bootstrap	-0.04	-0.04/-0.03	0.01	-0.01	-0.01/-0.01	0.00
Percentile	-0.04	-0.04/-0.03	0.01	-0.01	-0.01/-0.01	0.00

Method		β_3			β_4	
$\tau = 0.9$	$\hat{\beta}_3$	lb/ub	width	$\hat{\beta}_4$	lb/ub	width
Bootstrap	0.46	0.43/0.47	0.04	1.26	1.20/1.32	0.12
Percentile	0.46	0.43/0.46	0.03	1.26	1.25/1.29	0.04

Method		β_1			β_2	
$\tau = 0.99$	$\hat{\beta}_1$	lb/ub	width	$\hat{\beta}_2$	lb/ub	width
Bootstrap	0.06	0.05/0.07	0.02	-0.02	-0.02/-0.01	0.01
Percentile	0.06	0.06/0.07	0.01	-0.02	-0.02/-0.01	0.01

Method		β_3			β_4	
$\tau = 0.99$	$\hat{\beta}_3$	lb/ub	width	$\hat{\beta}_4$	lb/ub	width
Bootstrap	0.38	0.35/0.40	0.05	0.89	0.83/0.96	0.13
Percentile	0.38	0.36/0.39	0.03	0.89	0.84/0.92	0.08

Table 8.2: For the medfly data, the average lower bound/upper bounds and width of the 95% confidence intervals for the median ($\tau = 0.5$) of model B by the nonparametric bootstrap LRT and percentile method.

Method	β_1			β_2		
$\tau = 0.5$	$\hat{\beta}_1$	lb/ub	width	$\hat{\beta}_2$	lb/ub	width
Bootstrap	-0.21	-0.21/-0.20	0.01	0.00	-0.01/0.00	0.01
Percentile	-0.21	-0.21/-0.20	0.01	0.00	-0.01/0.00	0.01

Method	β_3			β_4		
$\tau = 0.5$	$\hat{\beta}_3$	lb/ub	width	$\hat{\beta}_4$	lb/ub	width
Bootstrap	0.04	0.02/0.06	0.04	0.32	0.25/0.36	0.11
Percentile	0.04	0.02/0.04	0.02	0.32	0.30/0.35	0.05

Method	β_5			β_6		
$\tau = 0.5$	$\hat{\beta}_5$	lb/ub	width	$\hat{\beta}_6$	lb/ub	width
Bootstrap	-0.12	-0.14/-0.11	0.03	0.16	0.16/0.18	0.02
Percentile	-0.12	-0.13/-0.12	0.01	0.16	0.16/0.17	0.01

Method	β_7			β_8		
$\tau = 0.5$	$\hat{\beta}_7$	lb/ub	width	$\hat{\beta}_8$	lb/ub	width
Bootstrap	-0.23	-0.23/-0.22	0.01	-0.16	-0.16/-0.14	0.02
Percentile	-0.23	-0.23/0.22	0.01	-0.16	-0.16/-0.15	0.01

Method	β_9			β_{10}		
$\tau = 0.5$	$\hat{\beta}_9$	lb/ub	width	$\hat{\beta}_{10}$	lb/ub	width
Bootstrap	0.03	0.02/0.03	0.01	-0.06	-0.07/-0.06	0.01
Percentile	0.03	0.02/0.03	0.01	-0.06	-0.07/-0.06	0.01

Method	β_{11}		
$\tau = 0.5$	$\hat{\beta}_{11}$	lb/ub	width
Bootstrap	0.13	0.13/0.14	0.01
Percentile	0.13	0.13/0.14	0.01

Table 8.3: For the medfly data, the estimated coverages and width of confidence intervals of regression coefficents by chi-square and nonparametric bootstrap LRT threshold in model B of medfly data when $\tau = 0.9$.

Method	β_1			β_2		
$\tau = 0.9$	$\hat{\beta}_1$	lb/ub	width	$\hat{\beta}_2$	lb/ub	width
Bootstrap	-0.01	-0.02/-0.01	0.01	-0.02	-0.02/-0.01	0.01
Percentile	-0.01	-0.02/-0.01	0.01	-0.02	-0.02/-0.01	0.01

Method	β_3			β_4		
$\tau = 0.9$	$\hat{\beta}_3$	lb/ub	width	$\hat{\beta}_4$	lb/ub	width
Bootstrap	0.03	0.02/0.07	0.05	0.43	0.37/0.49	0.12
Percentile	0.03	0.02/0.06	0.04	0.43	0.39/0.48	0.09

Method	β_5			β_6		
$\tau = 0.9$	$\hat{\beta}_5$	lb/ub	width	$\hat{\beta}_6$	lb/ub	width
Bootstrap	-0.11	-0.13/-0.10	0.03	0.08	0.07/0.09	0.02
Percentile	-0.11	-0.12/-0.11	0.01	0.08	0.07/0.08	0.01

Method	β_7			β_8		
$\tau = 0.9$	$\hat{\beta}_7$	lb/ub	width	$\hat{\beta}_8$	lb/ub	width
Bootstrap	-0.37	-0.39/-0.36	0.03	-0.23	-0.23/-0.22	0.01
Percentile	-0.37	-0.38/-0.36	0.02	-0.23	-0.23/-0.22	0.01

Method	β_9			β_{10}		
$\tau = 0.9$	$\hat{\beta}_9$	lb/ub	width	$\hat{\beta}_{10}$	lb/ub	width
Bootstrap	-0.09	-0.09/-0.07	0.02	-0.17	-0.18/-0.16	0.02
Percentile	-0.09	-0.09/-0.08	0.01	-0.17	-0.18/-0.17	0.01

Method	β_{11}		
$\tau = 0.9$	$\hat{\beta}_{11}$	lb/ub	width
Bootstrap	0.07	0.06/0.08	0.02
Percentile	0.07	0.06/0.07	0.01

Table 8.4: For the medfly data, the estimated coverages and widths of confidence intervals of regression coefficents by chi-square and nonparametric bootstrap LRT threshold in model B of medfly data when $\tau = 0.99$.

Method	β_1			β_2		
$\tau = 0.99$	$\hat{\beta}_1$	lb/ub	width	$\hat{\beta}_2$	lb/ub	width
Bootstrap	0.09	0.08/0.10	0.02	-0.01	-0.02/-0.01	0.01
Percentile	0.09	0.09/0.10	0.01	-0.01	-0.02/-0.01	0.01

Method	β_3			β_4		
$\tau = 0.99$	$\hat{\beta}_3$	lb/ub	width	$\hat{\beta}_4$	lb/ub	width
Bootstrap	0.03	0.00/0.07	0.07	0.53	0.43/0.63	0.20
Percentile	0.03	0.01/0.06	0.05	0.53	0.46/0.58	0.12

Method	β_5			β_6		
$\tau = 0.99$	$\hat{\beta}_5$	lb/ub	width	$\hat{\beta}_6$	lb/ub	width
Bootstrap	-0.18	-0.21/-0.15	0.06	0.01	-0.01/0.02	0.03
Percentile	-0.18	-0.20/-0.17	0.03	0.01	-0.01/0.01	0.02

Method	β_7			β_8		
$\tau = 0.99$	$\hat{\beta}_7$	lb/ub	width	$\hat{\beta}_8$	lb/ub	width
Bootstrap	-0.42	-0.44/-0.39	0.05	-0.28	-0.30/-0.26	0.04
Percentile	-0.42	-0.44/-0.40	0.04	-0.28	-0.30/-0.27	0.03

Method	β_9			β_{10}		
$\tau = 0.99$	$\hat{\beta}_9$	lb/ub	width	$\hat{\beta}_{10}$	lb/ub	width
Bootstrap	-0.14	-0.15/-0.12	0.03	-0.18	-0.20/-0.17	0.03
Percentile	-0.14	-0.15/-0.13	0.02	-0.18	-0.19/-0.17	0.02

Method	β_{11}		
$\tau = 0.99$	$\hat{\beta}_{11}$	lb/ub	width
Bootstrap	0.08	0.05/0.10	0.05
Percentile	0.08	0.07/0.09	0.02

Table 8.5: The average lower bounds, upper bounds, widths, and coverages of the 90%, 95%, and 99% confidence intervals for the median ($\tau = 0.5$) of simulated quantile regression model.

β_1		Bootstrap Interval				Percentile Interval			
α	β_1	lb	ub	width	cov	lb	ub	width	cov
0.10	1	0.54	1.47	0.93	0.91	0.68	1.33	0.66	0.84
0.05	1	0.49	1.52	1.02	0.96	0.62	1.39	0.77	0.90
0.01	1	0.43	1.58	1.15	0.99	0.51	1.50	0.99	0.96

β_2		Bootstrap Interval				Percentile Interval			
α	β_2	lb	ub	width	cov	lb	ub	width	cov
0.10	1	0.13	1.87	1.74	0.92	0.35	1.65	1.29	0.91
0.05	1	0.03	1.97	1.94	0.96	0.23	1.77	1.53	0.95
0.01	1	-0.13	2.14	2.28	0.99	-0.00	2.00	2.00	0.99

β_3		Bootstrap Interval				Percentile Interval			
α	β_3	lb	ub	width	cov	lb	ub	width	cov
0.10	0	-0.06	0.06	0.11	0.88	-0.04	0.04	0.07	0.79
0.05	0	-0.06	0.06	0.12	0.94	-0.04	0.04	0.08	0.85
0.01	0	-0.07	0.07	0.13	0.99	-0.05	0.05	0.11	0.93

Table 8.6: The average lower bounds, upper bounds, widths, and coverages of the 90%, 95%, and 99% confidence intervals for the 90^{th} percentile ($\tau = 0.9$) of simulated quantile regression model.

β_1		Bootstrap Interval				Percentile Interval			
α	β_1	lb	ub	width	cov	lb	ub	width	cov
0.10	-0.51	-1.31	0.49	1.81	0.87	-0.96	0.20	1.16	0.83
0.05	-0.51	-1.40	0.58	1.98	0.93	-1.07	0.31	1.39	0.90
0.01	-0.51	-1.51	0.69	2.20	0.99	-1.29	0.53	1.82	0.97

β_2		Bootstrap Interval				Percentile Interval			
α	β_2	lb	ub	width	cov	lb	ub	width	cov
0.10	1.09	-1.29	2.92	4.21	0.93	-0.46	2.37	2.83	0.93
0.05	1.09	-1.52	3.10	4.62	0.96	-0.79	2.60	3.39	0.95
0.01	1.09	-1.92	3.35	5.27	0.99	-1.45	3.03	4.47	0.99

β_3		Bootstrap Interval				Percentile Interval			
α	β_3	lb	ub	width	cov	lb	ub	width	cov
0.10	0.10	-0.00	0.19	0.19	0.84	0.03	0.14	0.11	0.77
0.05	0.10	-0.01	0.20	0.21	0.90	0.02	0.15	0.13	0.84
0.01	0.10	-0.01	0.21	0.23	0.96	0.00	0.18	0.17	0.93

Table 8.7: The average lower bounds, upper bounds, widths, and coverages of the 90%, 95%, and 99% confidence intervals for the 99^{th} percentile ($\tau = 0.99$) of simulated quantile regression model.

β_1		Bootstrap Interval				Percentile Interval			
α	β_1	lb	ub	width	cov	lb	ub	width	cov
0.10	-3.19	-8.96	2.89	11.86	0.89	-5.54	0.30	5.84	0.87
0.05	-3.19	-9.54	3.31	12.86	0.94	-6.33	0.93	7.26	0.92
0.01	-3.19	-10.29	3.99	14.29	0.99	-8.16	2.47	10.62	0.97

β_2		Bootstrap Interval				Percentile Interval			
α	β_2	lb	ub	width	cov	lb	ub	width	cov
0.10	1.26	-16.66	14.14	30.80	0.91	-13.39	7.99	21.37	0.92
0.05	1.26	-16.94	14.89	31.83	0.94	-14.40	9.30	23.70	0.96
0.01	1.26	-17.25	15.59	32.85	0.98	-16.02	12.01	28.03	0.99

β_3		Bootstrap Interval				Percentile Interval			
α	β_3	lb	ub	width	cov	lb	ub	width	cov
0.10	0.26	-0.18	0.99	1.17	0.86	0.00	0.51	0.51	0.86
0.05	0.26	-0.22	1.05	1.28	0.92	-0.04	0.61	0.65	0.92
0.01	0.26	-0.27	1.17	1.43	0.98	-0.14	0.86	1.00	0.97

Chapter 9

Conclusion

Generally, the nonparametric bootstrap LRT confidence intervals worked better than other methods that we discussed in this thesis. However, the coverages of the percentile approach we often comparable. The coverages of nonparametric bootstrap CI almost always did well at estimating their stated significance level when $\tau = 0.5$. When τ has extreme values like 0.9 and 0.99, the coverages of the nonparametric bootstrap CI tended to be closer to the significance level than the other methods considered.The widths of the nonparametric bootstrap CI are often among the widest, but this was usually because other methods had a tendency to undercover.

Even though the percentile-t method is commonly recognized as a more accurate method than the percentile method, in our thesis the percentile CI worked better than the percentile-t CI, which is the second best method among those considered. One of our contributions to the percentile CI methods was to improve calculation in estimating percentile CI with ALD by normal approximation, as introduced in Chapter 5. One possible reason that the percentile-t CI didn't work as expected may be the the calculation of standard error estimates. Since the distributions are misspecified with ALD in many of our examples the benefits of estimating standard errors may not be as apparent as when there is no misspecification. When we estimating the percentile-t CI for bootstrap data, the bootstrap standard errors of parameter of interest are estimated on ALD during this process even though that need not be the true generating process. Consequently, the arguments that suggest the percentile-t should perform best, which are based on the standard errors being good approximations to the actual standard errors, no longer hold.

The results of chi-square LR CI and Wald CI are the worst two methods from the results we got. The reason that the chi-square LR didn't perform well is the limitation of constant chi-square thresholds. Since the value of chi-square thresholds are below 6.63 with significant level less than 0.01 and constant, their coverage probabilities will tend to be low when the true generating distribution tends to give very large LR statistics. The Wald test is based on a standard normal approximation, which assume the estimated parameters are symmetrically distributed. Such approximations do not work well, particularly with extreme value of τ (i.e. extremely skewness). The results for the LR CI with chi-square thresholds and Wald tests indicate the importance of making robustness adjustments.

Finally, the nonparametric bootstrap LRT also the best CI of coefficients in quantile regression. Since the quantile regression contain only fixed effects, it may be valuable to consider adding random effects into the regression model as a follow-up to the work in this thesis. It may also be valuable to consider the use of bootstrap LR methods in other contexts where likelihood estimation is robust to model misspecification.

Bibliography

[1] John Aldrich. Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science*, 12(3):162–176, 1997.

[2] Charles M Beach and James G MacKinnon. A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica: journal of the Econometric Society*, pages 51–58, 1978.

[3] Daniel Berrar. Introduction to the non-parametric bootstrap., 2019.

[4] James R Carey, Pablo Liedo, Dina Orozco, and James W Vaupel. Slowing of mortality rates at older ages in large medfly cohorts. *Science*, 258(5081):457–461, 1992.

[5] AC Davison and DV Hinkley. Bootstrap methods and their application, cambridge univ. *Press, Cambridge*, 1997.

[6] Thomas DiCiccio and Robert Tibshirani. Bootstrap confidence intervals and bootstrap approximations. *Journal of the American Statistical Association*, 82(397):163–170, 1987.

[7] Bradley Efron and Trevor Hastie. Computer age statistical inference. 2022.

[8] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[9] Christopher J Elias. Percentile and percentile-t bootstrap confidence intervals: a practical comparison. *Journal of Econometric Methods*, 4(1):153–161, 2015.

[10] Xingdong Feng, Xuming He, and Jianhua Hu. Wild bootstrap for quantile regression. *Biometrika*, 98(4):995–999, 2011.

[11] Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character*, 222(594-604):309–368, 1922.

[12] Carol A Francisco and Wayne A Fuller. Quantile estimation with a complex survey design. *The Annals of statistics*, pages 454–469, 1991.

[13] Shelby J Haberman. Maximum likelihood estimates in exponential response models. *The annals of statistics*, 5(5):815–841, 1977.

[14] Peter Hall. Principles of edgeworth expansion. In *The bootstrap and edgeworth expansion*, pages 39–81. Springer, 1992.

[15] Peter J Huber. Under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather Modification; University of California Press: Berkeley, CA, USA*, page 221, 1967.

[16] Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[17] Roger Koenker and Olga Geling. Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*, 96(454):458–468, 2001.

[18] Roger Koenker and Jiaying Gu. REBayes: An R package for empirical bayes mixture methods. *Journal of Statistical Software*, 82(8):1–26, 2017.

[19] Roger Koenker and J. A. F. Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448):1296–1310, 1999.

[20] Roger Koenker and Jose AF Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310, 1999.

[21] Roger Koenker and Zhijie Xiao. Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612, 2002.

[22] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[23] Fernando Perez-Cruz. Kullback-leibler divergence estimation of continuous distributions. In *2008 IEEE International Symposium on Information Theory*, pages 1666–1670, 2008.

[24] Kjell A. Doksum Peter .J. Bickel. *Mathematical statistics: Basic ideas and selected topics*. San Francisco: Holden-Day, 1977.

[25] James Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1):119–131, 1975.

[26] Rostislav Protassov, David A Van Dyk, Alanna Connors, Vinay L Kashyap, and Aneta Siemiginowska. Statistics, handle with care: detecting multiple model components with the likelihood ratio test. *The Astrophysical Journal*, 571(1):545, 2002.

[27] Ichiro Takeuchi, Quoc Le, Timothy Sears, Alexander Smola, et al. Nonparametric quantile estimation. 2006.

[28] Abraham Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949.

[29] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25, 1982.

[30] Keming Yu, Zudi Lu, and Julian Stander. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):331–350, 2003.

Appendix: Table

Table 1: The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard asymmetric Laplace distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.35/0.31/0.66	-0.38/0.34/0.72	-0.39/0.33/0.72
500	0	-0.15/0.14/0.29	-0.16/0.14/0.30	-0.15/0.15/0.30
1000	0	-0.11/0.10/0.21	-0.11/0.11/0.22	-0.10/0.11/0.21

n	$\hat{\mu}$	χ^2	BootLR
100	0	-0.56/0.49/1.05	-0.56/0.44/1.00
500	0	-0.19/0.18/0.37	-0.19/0.17/0.36
1000	0	-0.14/0.14/0.28	-0.13/0.12/0.25

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.63/0.46/1.09	-0.79/0.48/1.27	-0.70/0.60/1.30
500	0	-0.26/0.23/0.49	-0.29/0.23/0.52	-0.27/0.26/0.53
1000	0	-0.18/0.16/0.34	-0.20/0.16/0.36	-0.18/0.18/0.36

n	$\hat{\mu}$	χ^2	BootLR
100	0	-1.17/0.78/1.95	-1.54/0.85/2.39
500	0	-0.34/0.30/0.64	-0.36/0.30/0.66
1000	0	-0.23/0.23/0.46	-0.24/0.20/0.44

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-2.60/0.69/3.29	-4.00/0.40/4.40	-2.46/2.20/4.66
500	0	-0.92/0.56/1.48	-1.23/0.68/1.91	-1.04/0.87/1.91
1000	0	-0.61/0.43/1.04	-0.72/0.42/1.14	-0.61/0.55/1.16

n	$\hat{\mu}$	χ^2	BootLR
100	0	-6.91/1.49/8.40	-7.29/0.40/7.69
500	0	-1.81/1.02/2.83	-1.87/0.82/2.69
1000	0	-1.25/0.66/1.91	-1.30/0.63/1.93

Table 2: The average lower bound/upper bound/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard asymmetric Laplace distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.53/0.49/1.02	-0.58/0.54/1.12	-0.62/0.51/1.13
500	0	-0.24/0.23/0.47	-0.25/0.24/0.49	-0.24/0.25/0.49
1000	0	-0.17/0.16/0.33	-0.19/0.16/0.35	-0.17/0.18/0.35

n	$\hat{\mu}$	χ^2	BootLR
100	0	-0.59/0.55/1.14	-0.61/0.58/1.19
500	0	-0.27/0.24/0.51	-0.27/0.26/0.53
1000	0	-0.20/0.19/0.39	-0.20/0.19/0.39

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.94/0.77/1.71	-1.15/0.91/2.06	-1.11/1.03/2.14
500	0	-0.40/0.37/0.77	-0.45/0.41/0.86	-0.44/0.44/0.88
1000	0	-0.28/0.26/0.54	-0.29/0.30/0.59	-0.32/0.27/0.59

n	$\hat{\mu}$	χ^2	BootLR
100	0	-1.73/0.95/2.68	-1.79/0.91/2.70
500	0	-0.54/0.45/0.99	-0.54/0.44/0.98
1000	0	-0.31/0.35/0.66	-0.3/0.33/0.65

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-3.53/1.62/5.15	-7.25/0.40/7.65	-2.65/5.08/7.73
500	0	-1.33/0.98/2.31	-1.83/1.19/3.02	-1.56/1.39/2.95
1000	0	-0.91/0.73/1.64	-1.23/0.79/2.02	-1.00/1.06/2.06

n	$\hat{\mu}$	χ^2	BootLR
100	0	-6.91/2.33/9.24	-7.66/0.40/8.06
500	0	-1.94/1.40/3.34	-2.21/1.20/3.41
1000	0	-1.25/1.02/2.27	-1.41/0.99/2.40

Table 3: The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.14/0.12/0.26	-0.22/0.19/0.41	-0.22/0.19/0.41
500	0	-0.06/0.06/0.12	-0.09/0.09/0.18	-0.10/0.09/0.19
1000	0	-0.04/0.04/0.08	-0.07/0.06/0.13	-0.07/0.06/0.13

n	$\hat{\mu}$	χ^2	BootLR
100	0	-0.27/0.21/0.48	-0.32/0.26/0.58
500	0	-0.09/0.10/0.19	-0.12/0.11/0.23
1000	0	-0.07/0.06/0.13	-0.08/0.07/0.15

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1.28	1.15/1.34/0.19	0.97/1.51/0.54	0.98/1.55/0.57
500	1.28	1.23/1.32/0.09	1.15/1.40/0.25	1.15/1.40/0.25
1000	1.28	1.24/1.31/0.07	1.19/1.36/0.17	1.19/1.36/0.17

n	$\hat{\mu}$	χ^2	BootLR
100	1.28	0.97/1.51/0.54	0.98/1.55/0.57
500	1.28	1.15/1.40/0.25	1.15/1.40/0.25
1000	1.28	1.21/1.35/0.14	1.16/1.38/0.22

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.33	2.11/2.19/0.08	1.68/2.50/0.82	1.77/2.75/0.98
500	2.33	2.26/2.30/0.04	2.03/2.56/0.53	2.03/2.56/0.53
1000	2.33	2.29/2.32/0.03	2.12/2.49/0.37	2.12/2.50/0.38

n	$\hat{\mu}$	χ^2	BootLR
100	2.33	1.38/2.37/0.99	1.36/2.50/1.14
500	2.33	2.00/2.36/0.36	1.89/2.62/0.73
1000	2.33	2.17/2.36/0.19	2.02/2.57/0.55

Table 4: The average lower bound/upper bound/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.22/0.19/0.41	-0.33/0.30/0.63	-0.34/0.31/0.65
500	0	-0.09/0.09/0.18	-0.15/0.14/0.29	-0.15/0.14/0.29
1000	0	-0.07/0.06/0.13	-0.10/0.10/0.20	-0.10/0.10/0.20
n	$\hat{\mu}$	χ^2	BootLR	
100	0	-0.37/0.28/0.65	-0.38/0.34/0.72	
500	0	-0.15/0.14/0.29	-0.16/0.15/0.31	
1000	0	-0.10/0.09/0.19	-0.11/0.10/0.21	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1.28	1.10/1.39/0.29	0.84/1.69/0.85	0.83/1.72/0.89
500	1.28	1.21/1.34/0.13	1.09/1.47/0.38	1.08/1.47/0.39
1000	1.28	1.23/1.32/0.09	1.14/1.41/0.27	1.14/1.42/0.28
n	$\hat{\mu}$	χ^2	BootLR	
100	1.28	0.90/1.62/0.72	0.77/1.76/0.99	
500	1.28	1.12/1.44/0.32	1.06/1.49/0.43	
1000	1.28	1.17/1.36/0.19	1.13/1.43/0.30	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.33	2.08/2.21/0.13	1.47/2.50/1.03	1.76/3.12/1.36
500	2.33	2.25/2.31/0.06	1.91/2.72/0.81	1.88/2.72/0.84
1000	2.33	2.28/2.32/0.04	2.03/2.61/0.58	2.01/2.61/0.60
n	$\hat{\mu}$	χ^2	BootLR	
100	2.33	1.38/2.42/1.04	1.33/2.50/1.17	
500	2.33	2.00/2.42/0.42	1.84/2.77/0.93	
1000	2.33	2.10/2.42/0.32	1.98/2.67/0.69	

Table 5: The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the logistic distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.25/0.21/0.46	-0.34/0.31/0.65	-0.36/0.31/0.67
500	0	-0.19/0.17/0.36	-0.15/0.14/0.29	-0.15/0.14/0.29
1000	0	-0.07/0.07/0.14	-0.11/0.10/0.21	-0.11/0.10/0.21
n	$\hat{\mu}$	χ^2	BootLR	
100	0	-0.47/0.35/0.82	-0.51/0.42/0.93	
500	0	-0.18/0.15/0.33	-0.19/0.17/0.36	
1000	0	-0.12/0.10/0.22	-0.13/0.12/0.25	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.20	1.97/2.32/0.35	1.64/2.67/1.03	1.63/2.71/1.08
500	2.20	2.11/2.27/0.16	1.95/2.43/0.48	1.95/2.44/0.49
1000	2.20	2.13/2.25/0.12	2.02/2.36/0.34	2.02/2.37/0.35
n	$\hat{\mu}$	χ^2	BootLR	
100	2.20	1.53/2.69/1.16	1.38/2.93/1.55	
500	2.20	1.96/2.41/0.45	1.87/2.51/0.64	
1000	2.20	2.07/2.32/0.25	1.97/2.41/0.44	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	4.60	4.07/4.24/0.17	3.07/5.14/2.07	3.13/5.72/2.59
500	4.60	4.45/4.53/0.08	3.83/5.28/1.45	3.78/5.25/1.47
1000	4.60	4.51/4.57/0.06	4.06/5.04/0.98	4.07/5.07/1.00
n	$\hat{\mu}$	χ^2	BootLR	
100	4.60	2.50/4.67/2.17	2.38/5.14/2.76	
500	4.60	3.89/4.68/0.79	3.51/5.47/1.96	
1000	4.60	4.19/4.67/0.48	3.83/5.28/1.45	

Table 6: The average lower bound/upper bound/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the logistic distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-0.37/0.34/0.71	-0.52/0.49/1.01	-0.55/0.50/1.05
500	0	-0.16/0.15/0.31	-0.23/0.22/0.45	-0.23/0.22/0.45
1000	0	-0.12/0.11/0.23	-0.16/0.16/0.32	-0.16/0.16/0.32

n	$\hat{\mu}$	χ^2	BootLR
100	0	-0.50/0.49/0.99	-0.62/0.56/1.18
500	0	-0.20/0.21/0.41	-0.26/0.24/0.50
1000	0	-0.16/0.15/0.31	-0.18/0.17/0.35

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.20	1.87/2.42/0.55	1.40/3.07/1.67	1.31/3.05/1.74
500	2.20	2.06/2.31/0.25	1.82/2.58/0.76	1.82/2.58/0.76
1000	2.20	2.10/2.28/0.18	1.93/2.46/0.53	1.93/2.46/0.53

n	$\hat{\mu}$	χ^2	BootLR
100	2.20	1.39/2.95/1.56	1.27/3.25/1.98
500	2.20	1.90/2.49/0.59	1.78/2.63/0.85
1000	2.20	2.00/2.38/0.38	1.91/2.49/0.58

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	4.60	4.02/4.29/0.27	2.58/5.14/2.56	3.09/6.76/3.67
500	4.60	4.43/4.56/0.13	3.56/5.78/2.22	3.36/5.66/2.30
1000	4.60	4.50/4.59/0.09	3.83/5.39/1.56	3.77/5.36/1.59

n	$\hat{\mu}$	χ^2	BootLR
100	4.60	2.50/4.83/2.33	2.32/5.14/2.82
500	4.60	3.84/4.78/0.94	3.40/5.92/2.52
1000	4.60	4.11/4.78/0.67	3.72/5.56/1.84

Table 7: The average lower bound/upper bound/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Cauchy distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-1.31/1.21/2.52	-0.21/0.16/0.37	-0.43/0.12/0.55
500	0	-0.68/0.67/1.35	-0.12/0.11/0.23	-0.21/0.20/0.41
1000	0	-0.55/0.54/1.09	-0.08/0.08/0.16	-0.14/0.13/0.27
n	$\hat{\mu}$	χ^2	BootLR	
100	0	-0.15/0.17/0.32	-0.19/0.15/0.34	
500	0	-0.20/0.18/0.38	-0.17/0.15/0.32	
1000	0	-0.14/0.14/0.28	-0.11/0.11/0.22	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	3.08	1.21/4.97/3.76	1.83/5.30/3.47	-0.17/8.22/8.39
500	3.08	2.12/4.05/1.93	2.45/3.92/1.47	1.88/4.57/2.69
1000	3.08	2.38/3.78/1.40	2.62/3.64/1.02	2.29/3.97/1.68
n	$\hat{\mu}$	χ^2	BootLR	
100	3.08	1.38/7.78/6.40	1.35/7.48/6.13	
500	3.08	2.18/4.24/2.06	2.21/4.40/2.19	
1000	3.08	2.44/4.05/1.61	2.45/3.91/1.46	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	31.82	27.99/36.10/8.11	7.95/244.15/236.20	-185.13/595.68/780.81
500	31.82	29.63/34.22/4.59	16.09/77.08/60.99	-14.17/126.83/141.00
1000	31.82	30.30/33.56/3.26	19.52/54.95/35.43	2.56/77.76/75.20
n	$\hat{\mu}$	χ^2	BootLR	
100	31.82	5.27/66.17/60.90	4.00/244.15/240.15	
500	31.82	15.15/53.20/38.05	11.15/105.34/94.19	
1000	31.82	16.70/53.98/37.28	15.00/83.81/68.81	

Table 8: The average lower bound/upper bounds/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard Cauchy distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0	-2.02/1.93/3.95	-0.24/0.19/0.43	-0.50/0.16/0.66
500	0	-1.06/1.06/2.12	-0.18/0.18/0.36	-0.39/0.40/0.79
1000	0	-0.86/0.85/1.71	-0.13/0.13/0.26	-0.27/0.26/0.53

n	$\hat{\mu}$	χ^2	BootLR
100	0	-0.41/0.17/0.58	-0.24/0.19/0.43
500	0	-0.21/0.18/0.39	-0.21/0.20/0.41
1000	0	-0.14/0.14/0.28	-0.14/0.14/0.28

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	3.08	0.15/6.03/5.88	1.43/7.95/6.52	-5.37/13.12/18.49
500	3.08	1.58/4.60/3.02	2.17/4.51/2.34	0.59/6.11/5.52
1000	3.08	1.98/4.18/2.20	2.40/4.01/1.61	1.48/4.86/3.38

n	$\hat{\mu}$	χ^2	BootLR
100	3.08	1.38/7.78/6.40	1.21/10.20/8.99
500	3.08	2.18/5.19/3.01	2.05/4.86/2.81
1000	3.08	2.30/4.05/1.75	2.31/4.20/1.89

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	31.82	25.70/38.39/12.69	4.84/244.15/239.31	-188.05/1390.52/1578.57
500	31.82	28.33/35.52/7.19	11.99/149.91/137.92	-92.83/277.95/370.78
1000	31.82	29.37/34.49/5.12	15.50/82.94/67.44	-35.30/139.90/175.20

n	$\hat{\mu}$	χ^2	BootLR
100	31.82	5.27/86.07/80.80	3.77/244.15/240.38
500	31.82	15.10/73.20/58.10	10.18/321.51/311.33
1000	31.82	14.63/62.12/47.49	13.78/115.83/102.05

Table 9: The average lower bound/upper bounds/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard folded normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0.67	0.54/0.69/0.15	0.53/0.74/0.21	0.49/0.70/0.21
500	0.67	0.64/0.71/0.07	0.62/0.73/0.11	0.61/0.73/0.12
1000	0.67	0.65/0.70/0.05	0.63/0.71/0.08	0.63/0.71/0.08

n	$\hat{\mu}$	χ^2	BootLR	
100	0.67	0.55/0.71/0.16	0.52/0.72/0.20	
500	0.67	0.61/0.73/0.12	0.60/0.74/0.14	
1000	0.67	0.64/0.72/0.08	0.62/0.72/0.10	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1.64	1.55/1.68/0.13	1.39/1.84/0.45	1.40/1.87/0.47
500	1.64	1.61/1.67/0.06	1.53/1.74/0.21	1.54/1.75/0.21
1000	1.64	1.62/1.66/0.04	1.57/1.72/0.15	1.57/1.72/0.15

n	$\hat{\mu}$	χ^2	BootLR	
100	1.64	1.27/1.82/0.55	1.27/1.95/0.68	
500	1.64	1.55/1.72/0.17	1.50/1.78/0.28	
1000	1.64	1.58/1.69/0.11	1.55/1.73/0.18	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.58	2.38/2.44/0.06	1.99/2.74/0.75	2.07/2.99/0.92
500	2.58	2.52/2.55/0.03	2.31/2.79/0.48	2.30/2.80/0.50
1000	2.58	2.55/2.57/0.02	2.39/2.72/0.33	2.40/2.75/0.35

n	$\hat{\mu}$	χ^2	BootLR	
100	2.58	1.85/2.61/0.76	1.72/2.74/1.02	
500	2.58	2.25/2.61/0.36	2.18/2.85/0.67	
1000	2.58	2.36/2.61/0.25	2.30/2.81/0.51	

Table 10: The average lower bound/upper bounds/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard folded normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	0.67	0.49/0.73/0.24	0.51/0.75/0.24	0.48/0.72/0.24
500	0.67	0.62/0.73/0.11	0.59/0.76/0.17	0.58/0.76/0.18
1000	0.67	0.63/0.71/0.08	0.61/0.74/0.13	0.61/0.73/0.12
n	$\hat{\mu}$	χ^2	BootLR	
100	0.67	0.53/0.78/0.25	0.51/0.75/0.24	
500	0.67	0.60/0.76/0.16	0.58/0.77/0.19	
1000	0.67	0.62/0.73/0.11	0.61/0.74/0.13	

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1.64	1.51/1.72/0.21	1.28/1.99/0.71	1.28/2.03/0.75
500	1.64	1.59/1.69/0.10	1.48/1.81/0.33	1.48/1.81/0.33
1000	1.64	1.61/1.67/0.06	1.53/1.76/0.23	1.53/1.76/0.23
n	$\hat{\mu}$	χ^2	BootLR	
100	1.64	1.27/1.95/0.68	1.22/2.07/0.85	
500	1.64	1.52/1.78/0.26	1.46/1.83/0.37	
1000	1.64	1.55/1.72/0.17	1.52/1.77/0.25	

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	2.58	2.36/2.46/0.10	1.80/2.74/0.94	2.06/3.38/1.32
500	2.58	2.51/2.56/0.05	2.20/2.95/0.75	2.17/2.95/0.78
1000	2.58	2.54/2.58/0.04	2.30/2.84/0.54	2.30/2.85/0.55
n	$\hat{\mu}$	χ^2	BootLR	
100	2.58	1.84/2.66/0.82	1.69/2.74/1.05	
500	2.58	2.25/2.66/0.41	2.14/2.99/0.85	
1000	2.58	2.36/2.65/0.29	2.26/2.89/0.63	

Table 11: The average lower bound/upper bounds/width of the 90% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard log normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1	0.85/1.20/0.35	0.90/1.15/0.25	0.92/1.18/0.26
500	1	0.92/1.08/0.16	0.91/1.10/0.19	0.91/1.09/0.18
1000	1	0.94/1.06/0.12	0.94/1.07/0.13	0.93/1.06/0.13

n	$\hat{\mu}$	χ^2	BootLR
100	1	0.88/1.17/0.29	0.90/1.14/0.24
500	1	0.91/1.11/0.20	0.89/1.12/0.23
1000	1	0.93/1.08/0.15	0.92/1.08/0.16

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	3.6	3.27/3.78/0.51	2.67/4.61/1.94	2.52/4.67/2.15
500	3.6	3.47/3.70/0.23	3.17/4.06/0.89	3.14/4.05/0.91
1000	3.6	3.50/3.66/0.16	3.28/3.92/0.64	3.27/3.91/0.64

n	$\hat{\mu}$	χ^2	BootLR
100	3.6	2.67/4.31/1.64	2.31/5.30/2.99
500	3.6	3.24/3.90/0.66	3.03/4.23/1.20
1000	3.6	3.36/3.85/0.49	3.20/4.01/0.81

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	10.24	8.81/9.20/0.39	5.52/13.51/7.99	4.38/17.25/12.87
500	10.24	9.84/10.03/0.19	7.68/13.19/5.51	7.20/13.04/5.84
1000	10.24	10.00/10.14/0.14	8.38/12.15/3.77	8.16/12.19/4.03

n	$\hat{\mu}$	χ^2	BootLR
100	10.24	3.97/10.97/7.00	3.99/13.51/9.52
500	10.24	7.42/10.43/3.01	6.64/14.18/7.54
1000	10.24	8.75/10.45/1.70	7.59/13.39/5.80

Table 12: The average lower bound/upper bounds/width of the 99% confidence intervals for the median ($\tau = 0.5$), the 90^{th} percentile ($\tau = 0.9$), and the 99^{th} percentile ($\tau = 0.99$) when data were generated from the standard log normal distribution.

$\tau = 0.5$

n	$\hat{\mu}$	Wald	Perc	PercT
100	1	0.75/1.30/0.55	0.88/1.17/0.29	0.90/1.20/0.30
500	1	0.87/1.13/0.26	0.87/1.15/0.28	0.85/1.14/0.29
1000	1	0.91/1.09/0.18	0.90/1.10/0.20	0.90/1.10/0.20

n	$\hat{\mu}$	χ^2	BootLR
100	1	0.88/1.17/0.29	0.88/1.17/0.29
500	1	0.87/1.15/0.28	0.85/1.17/0.32
1000	1	0.90/1.10/0.20	0.90/1.11/0.21

$\tau = 0.9$

n	$\hat{\mu}$	Wald	Perc	PercT
100	3.6	3.12/3.92/0.80	2.33/5.52/3.19	1.88/5.50/3.62
500	3.6	3.40/3.77/0.37	2.97/4.37/1.40	2.89/4.33/1.44
1000	3.6	3.45/3.71/0.26	3.13/4.12/0.99	3.10/4.10/1.00

n	$\hat{\mu}$	χ^2	BootLR
100	3.6	2.46/5.13/2.67	2.18/6.07/3.89
500	3.6	3.07/4.14/1.07	2.90/4.49/1.59
1000	3.6	3.27/3.91/0.64	3.09/4.18/1.09

$\tau = 0.99$

n	$\hat{\mu}$	Wald	Perc	PercT
100	10.24	8.70/9.31/0.61	4.42/13.51/9.09	4.30/24.45/20.15
500	10.24	9.79/10.09/0.30	6.80/15.72/8.92	5.41/15.16/9.75
1000	10.24	9.96/10.18/0.22	7.61/13.81/6.20	6.93/13.55/6.62

n	$\hat{\mu}$	χ^2	BootLR
100	10.24	3.97/11.14/7.17	3.87/13.51/9.64
500	10.24	7.42/10.89/3.47	6.37/16.92/10.55
1000	10.24	8.75/10.81/2.06	7.26/14.78/7.52