Penalized Maximum Likelihood Estimation of Two-Parameter Exponential Distributions

A PROJECT
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mengjie Zheng

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Advisor: Yongcheng Qi

May, 2013

Acknowledgements

I would like to express my sincere appreciation to my advisor, Professor Yongcheng Qi, for his guidance, support, encouragement, and great patience to me all the way through my graduation process.

I am very grateful to the members of my supervisory committee: Professor Kang L. James and Professor Richard Green for serving on my examination committee.

I would like to thank Professor Kang James for giving me precious practice and advice, and help me out all along the seminar.

Abstract

The two-parameter exponential distribution has many applications in real life. In this project we consider estimation problem of the two unknown parameters. The most widely used method Maximum Likelihood Estimation(MLE) always uses the minimum of the sample to estimate the location parameter, which is too conservative. Our idea is to add a penalty multiplier to the regular likelihood function so that the estimate of the location parameter is not too close to the sample minimum. The new estimates for both parameters are unbiased and also Uniformly Minimum Variance Unbiased Estimators(UMVUE). The penalized MLE for incomplete data is also discussed.

Contents

List of Tables

Chapter 1

Introduction

Consider a random variable X having two-parameter exponential distribution $\text{EXP}(\theta, \eta)$, with probability density function(pdf) given by

$$f(x;\theta,\eta) = \begin{cases} \dfrac{1}{\theta}e^{-(x-\eta)/\theta} & x \geqslant \eta, \\ \\ 0 & \text{otherwise}, \end{cases} \tag{1.1}$$

where $\theta > 0$ is a scale parameter and $\eta \in R$ is a location parameter. The cumulative distribution function (CDF) is

$$F(x) = P[X \leqslant x] = \int_{\eta}^{x} \frac{1}{\theta}e^{-(x-\eta)/\theta}dx = 1 - e^{-(x-\eta)/\theta}, x \geqslant \eta. \tag{1.2}$$

The two-parameter exponential distribution has many real world applications. It can be used to model the data such as the service times of agents in a system (Queuing Theory), the time it takes before your next telephone call, the time until a radioactive particle decays, the distance between mutations on a DNA strand, and the extreme values of annual snowfall or rainfall.

Given a sample of size n from a two-parameter exponential distribution, we are interested in estimating both θ and η. The most widely used method to do estimation is Maximum Likelihood Estimation(MLE). Under some regularity conditions, the MLE method has nice properties such as consistency and efficiency. The regular MLE is too conservative because it always chooses the minimum of the sample to estimate the location parameter. Our idea is to add a penalty multiplier to the regular MLE. By introducing a proper penalty, the penalized maximum likelihood estimators for both parameters are Uniformly Minimum Variance Unbiased Estimators (UMVUE).

The rest of the project is organized as follows. In chapter 2, for complete data set, we first introduce the conventional MLE and penalized MLE. In chapter 3, we consider incomplete data including Type-II censoring and Type-II hybrid censoring, and extend the penalized MLE to do estimation for incomplete data. In chapter 4, we present the simulation results for regular and penalized MLE for both complete and incomplete data sets, and give analysis of these results.

Chapter 2

Estimation for Complete Data

In this chapter, we introduce the likelihood function and penalized likelihood function. Then we discuss the properties of both regular and penalized likelihood estimators from the two-parameter exponential distributions.

2.1 MLE for complete data

Maximum likelihood estimation (MLE) is a method to provide estimates for the parameters of a statistical model by maximizing likelihood functions. For an independent and identically distributed(i.i.d) sample $x_1, x_2, \cdots, x_n$ with pdf as (1.1), the joint density function is

$$f(x_1, x_2, \cdots, x_n \mid \theta, \eta) = f(x_1 \mid \theta, \eta) f(x_2 \mid \theta, \eta) \cdots f(x_n \mid \theta, \eta).$$

A likelihood function provides a look at the joint density function from a different perspective by considering the observed values $x_1, x_2, \cdots, x_n$ to be fixed, while θ and η are the variables of the function. The likelihood function is

$$L(\theta, \eta \mid x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta, \eta) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i - \eta)}, x_{1:n} \geqslant \eta,$$

where $x_{1:n} \leqslant x_{2:n} \leqslant \cdots \leqslant x_{n:n}$ are order statistics based on $x_1, x_2, \cdots, x_n$, and $x_{1:n}$ is the minimum of the sample. Note that $x_{1:n} \geqslant \eta$ is equivalent to $x_i \geqslant \eta$ for all i.

It is more convenient to work with the logarithm of the likelihood function.

$$\begin{aligned}\ln L(\theta, \eta \mid x_1, x_2, \cdots, x_n) &= \prod_{i=1}^{n} \ln f(x_i \mid \theta, \eta) \\ &= -\frac{1}{\theta}\sum_{i=1}^{n} x_i + \frac{n\eta}{\theta} - n\ln\theta, \quad x_{1:n} \geqslant \eta.\end{aligned} \tag{2.1}$$

The likelihood function is maximized with respect to η by taking $\hat{\eta} = x_{1:n}$. To maximize relative to θ, differentiate (2.1) with respect to θ and solve the equation

$$\frac{d\ln L(\theta, \eta)}{d\theta} = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n}(x_i - \eta)}{\theta^2} = 0.$$

The MLE for θ is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{n}(x_i - \hat{\eta})}{n} = \bar{x} - \hat{\eta} = \bar{x} - x_{1:n}. \tag{2.2}$$

The CDF of $x_{1:n}$ is

$$\begin{aligned}F_{(1)}(x) &= P(x_{1:n} \leqslant x) = 1 - P(x_{1:n} > x) \\ &= 1 - P(\text{all } x_i > x) = 1 - (1 - F(x))^n \\ &= 1 - e^{-n(x-\eta)/\theta},\end{aligned}$$

with the pdf given by

$$f_{(1)}(x) = F'_{(1)}(x) = \frac{n}{\theta}e^{-n(x-\eta)/\theta}. \tag{2.3}$$

From (2.3), $x_{1:n} \sim \text{EXP}\left(\frac{\theta}{n}, \eta\right)$.

An estimator $\hat{\theta}$ is said to be an **unbiased estimator** of θ if $E(\hat{\theta}) = \theta$ for all θ. Otherwise, $\hat{\theta}$ is said to be a **biased estimator** of θ, and the **bias** is $\mathrm{b}(\hat{\theta}) = E(\hat{\theta}) - \theta$. The **mean square error(MSE)** of θ is $\mathrm{MSE}(\hat{\theta}) = E[\hat{\theta} - \theta]^2 = \mathrm{Var}(\hat{\theta}) + [\mathrm{b}(\hat{\theta})]^2$.

The expectations of $\hat{\eta}$ and $\hat{\theta}$ are

$$E(\hat{\eta}) = E(x_{1:n}) = \frac{\theta}{n} + \eta \neq \eta, \tag{2.4}$$

and

$$E(\hat{\theta}) = E(\bar{x} - x_{1:n}) = \theta + \eta - \left(\frac{\theta}{n} + \eta\right) = \theta - \frac{\theta}{n} \neq \theta. \tag{2.5}$$

By (2.4) and (2.5), $\hat{\eta}$ and $\hat{\theta}$ are not unbiased estimators. And the biases are

$$\mathrm{b}\left(\hat{\eta}\right) = E\left(\hat{\eta}\right) - \eta = \frac{\theta}{n} + \eta - \eta = \frac{\theta}{n}, \tag{2.6}$$

and

$$\mathrm{b}(\hat{\theta}) = E(\hat{\theta}) - \theta = \theta - \frac{\theta}{n} - \theta = -\frac{\theta}{n}. \tag{2.7}$$

Note that the traditional MLE of η picks the smallest value of the sample to estimate the location parameter. It always overestimates the location parameter since $P(x_{1:n} > \eta) = 1$.

The variance of $\hat{\eta}$ is

$$\mathrm{Var}\left(\hat{\eta}\right) = \mathrm{Var}\left(x_{1:n}\right) = \frac{\theta^2}{n^2}, \tag{2.8}$$

which follows from the fact that $x_{1:n} \sim \mathrm{EXP}\left(\frac{\theta}{n}, \eta\right)$.

The variance of $\hat{\theta}$ is

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\bar{x} - x_{1:n}) = \frac{n-1}{n^2}\theta^2. \tag{2.9}$$

This result is obtained as follows:

$$\begin{aligned}
\mathrm{Var}(\hat{\theta}) &= \mathrm{Var}\left(\bar{x} - x_{1:n}\right) = \mathrm{Var}\left(\bar{x}\right) + \mathrm{Var}\left(x_{1:n}\right) - 2\mathrm{Cov}\left(\bar{x}, x_{1:n}\right), \\
\mathrm{Cov}\left(\bar{x}, x_{1:n}\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} x_i/n, x_{1:n}\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}\left(x_i, x_{1:n}\right) \\
&= \frac{1}{n}\binom{n}{1}\mathrm{Cov}\left(x_{1:n}, x_{1:n}\right) = \mathrm{Var}\left(x_{1:n}\right), \qquad (2.10) \\
\mathrm{Var}(\hat{\theta}) &= \mathrm{Var}\left(\bar{x}\right) + \mathrm{Var}\left(x_{1:n}\right) - 2\mathrm{Var}\left(x_{1:n}\right) = \frac{\theta^2}{n} - \frac{\theta^2}{n^2} = \frac{n-1}{n^2}\theta^2.
\end{aligned}$$

From (2.6), (2.7), (2.8), (2.9), the MSEs of $\hat{\eta}$ and $\hat{\theta}$ are

$$\mathrm{MSE}\left(\hat{\eta}\right) = \mathrm{Var}\left(\hat{\eta}\right) + \left[\mathrm{b}\left(\hat{\eta}\right)\right]^2 = \frac{\theta^2}{n^2} + \left(\frac{\theta}{n}\right)^2 = 2\frac{\theta^2}{n^2}, \tag{2.11}$$

and

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{b}(\hat{\theta})]^2 = \frac{n-1}{n^2}\theta^2 + \left(-\frac{\theta}{n}\right)^2 = \frac{\theta^2}{n}. \tag{2.12}$$

2.2 Penalized MLE for complete data

The regular MLE of the two-parameter exponential distribution does not give unbiased estimators due to the fact that the likelihood function is monotone increasing as a function of location parameter. Our approach is to add a penalty to the likelihood function such that the new function is no longer monotone as a function of the location parameter. The penalty term we used is $x_{1:n} - \eta$, and the regular likelihood function is multiplied by this penalty. For i.i.d sample $x_1, x_2, \cdots, x_n$ with pdf as (1.1), the penalized likelihood function is

$$L^*(\theta, \eta) = (x_{1:n} - \eta) \prod_{i=1}^{n} f(x_i \mid \theta, \eta) = (x_{1:n} - \eta) \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i - \eta)}, \quad x_{1:n} \geqslant \eta. \tag{2.13}$$

After doing this, $L^*(\theta, \eta)$ is not monotone respect to η. And $L^*(\theta, x_{1:n}) = 0$, which forces η to depart from $x_{1:n}$.

By taking the logarithm of likelihood function, we have,

$$\ln L^*(\theta, \eta) = \ln(x_{1:n} - \eta) - n \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n}(x_i - \eta), \quad x_{1:n} \geqslant \eta. \tag{2.14}$$

In order to get MLEs, we differentiate the logarithm of the likelihood function with respect to θ, η respectively, and set the derivatives equal to 0.

$$\begin{cases} \dfrac{d \ln L^*(\theta, \eta)}{d\eta} = -\dfrac{1}{x_{1:n} - \eta} + \dfrac{n}{\theta} = 0, \\[2em] \dfrac{d \ln L^*(\theta, \eta)}{d\theta} = -\dfrac{n}{\theta} + \dfrac{\sum_{i=1}^{n}(x_i - \eta)}{\theta^2} = 0. \end{cases}$$

The solutions are the penalized MLEs for θ and η.

$$\theta^* = \frac{\sum x_i - n x_{1:n}}{n-1} = \frac{n(\bar{x} - x_{1:n})}{n-1}, \tag{2.15}$$

$$\eta^* = \frac{-\sum x_i + n^2 x_{1:n}}{n(n-1)} = \frac{n x_{1:n} - \bar{x}}{n-1}. \tag{2.16}$$

These estimators were previously constructed by Sarhan [1], and Cohen and Helm[2]. Sarhan employed the least-square technique. Cohen and Helm applied a method of modified moments.

Theorem 1. *(Sarhan[1]) θ^* and η^* are unbiased and uniformly minimum variance estimators for θ and η.*

The expectations for the penalized MLEs are

$$E(\theta^*) = E\left[\frac{n(\bar{x} - x_{1:n})}{n-1}\right] = \frac{n}{n-1}\left(\theta + \eta - \left(\frac{\theta}{n} + \eta\right)\right) = \theta, \tag{2.17}$$

and

$$\begin{aligned} E(\eta^*) &= E\left[\frac{nx_{1:n} - \bar{x}}{n-1}\right] = \frac{1}{n-1}(nE(x_{1:n}) - \bar{x}) \\ &= \frac{1}{n-1}\left(n\left(\frac{\theta}{n} + \eta\right) - (\theta + \eta)\right) = \frac{1}{n-1}(n\eta - \eta) = \eta. \end{aligned} \tag{2.18}$$

From (2.17), (2.18), θ^* and η^* are unbiased estimators for θ and η. And thus the biases are

$$\mathrm{b}(\theta^*) = E(\theta^*) - \eta = 0, \tag{2.19}$$

and

$$\mathrm{b}(\eta^*) = E(\eta^*) - \eta = 0. \tag{2.20}$$

The variance of θ^* is

$$\begin{aligned} \mathrm{Var}(\theta^*) &= \mathrm{Var}\left(\frac{n(\bar{x} - x_{1:n})}{n-1}\right) = \frac{n^2}{(n-1)^2}\left(\mathrm{Var}(\bar{x}) + \mathrm{Var}(x_{1:n}) - 2\mathrm{Cov}(\bar{x}, x_{1:n})\right) \\ &= \frac{n^2}{(n-1)^2}\left(\mathrm{Var}(\bar{x}) + \mathrm{Var}(x_{1:n}) - 2\mathrm{Var}(x_{1:n})\right) \\ &= \frac{n^2}{(n-1)^2}\left(\frac{\theta^2}{n} + \frac{\theta^2}{n^2} - 2\frac{\theta^2}{n^2}\right) = \frac{\theta^2}{n-1}. \end{aligned} \tag{2.21}$$

The variance of η^* is

$$\begin{aligned} \mathrm{Var}(\eta^*) &= \mathrm{Var}\left(\frac{nx_{1:n} - \bar{x}}{n-1}\right) = \frac{1}{(n-1)^2}\mathrm{Var}(nx_{1:n} - \bar{x}) \\ &= \frac{1}{(n-1)^2}\left(n^2\mathrm{Var}(x_{1:n}) + \mathrm{Var}(\bar{x}) - 2\mathrm{Cov}(nx_{1:n}, \bar{x})\right) \\ &= \frac{1}{(n-1)^2}\left(n^2\mathrm{Var}(x_{1:n}) + \mathrm{Var}(\bar{x}) - 2n\mathrm{Var}(x_{1:n})\right) \\ &= \frac{1}{(n-1)^2}\left(n^2\frac{\theta^2}{n^2} + \frac{\theta^2}{n} - 2n\frac{\theta^2}{n^2}\right) = \frac{\theta^2}{n(n-1)}. \end{aligned} \tag{2.22}$$

From (2.19), (2.20), (2.21), (2.22) the MSEs of θ^* and η^* are

$$\mathrm{MSE}\left(\theta^*\right) = \mathrm{Var}\left(\theta^*\right) + \left[\mathrm{b}\left(\theta^*\right)\right]^2 = \frac{\theta^2}{n-1}, \tag{2.23}$$

and

$$\mathrm{MSE}\left(\eta^*\right) = \mathrm{Var}\left(\eta^*\right) + \left[\mathrm{b}\left(\eta^*\right)\right]^2 = \frac{\theta^2}{n\left(n-1\right)}. \tag{2.24}$$

From (2.12) and (2.23), $\mathrm{MSE}(\hat{\theta}) = \frac{\theta^2}{n}$ is a little bit smaller than $\mathrm{MSE}(\theta^*) = \frac{\theta^2}{n-1}$, but when the sample size is large, $\mathrm{MSE}(\hat{\theta})$ and $\mathrm{MSE}(\theta^*)$ don't have much difference; from (2.11) and (2.24), $\mathrm{MSE}(\hat{\eta}) = \frac{2\theta^2}{n^2}$ is almost two times of $\mathrm{MSE}(\eta^*) = \frac{\theta^2}{n(n-1)}$, which means $\mathrm{MSE}(\eta^*)$ is much less than $\mathrm{MSE}(\hat{\eta})$. Therefore, respect to MSE, the penalized MLE is much better than regular MLE for estimation of the location parameter, but the penalized MLE is not much better for the scale parameter. In terms of biases, penalized MLE does a good job by removing the biases.

Proof of Theorem 1

A **uniformly minimum variance unbiased estimator(UMVUE)** is an estimator which achieves the smallest variance among all unbiased estimators. **Lehmann - Scheff Theorem**[3] states that if $\boldsymbol{S}$ is a vector of jointly **complete sufficient** statistics for θ and η, and if $\boldsymbol{S}$ are **unbiased** statistics for θ and η, then they are the UMVUEs for θ and η. **Sufficient** means the statistic tells everything about the parameter, no other statistic could give any additional information. A family of density function is **complete** if $E[u(\boldsymbol{S})] = 0$ for all $\boldsymbol{\theta} \in \Omega$ implies $u(\boldsymbol{S}) = 0$ with probability 1 for all $\boldsymbol{\theta} \in \Omega$.

From (2.17), (2.18), we already have unbiased statistics $\theta^* = \frac{n\left(\bar{x} - x_{1:n}\right)}{n-1}$ and $\eta^* = \frac{n x_{1:n} - \bar{x}}{n-1}$ for θ and η. So we only need to know whether they are jointly sufficient and complete.

Fisher's **Factorization Theorem**[3] provides a criterion for sufficient statistics. $\boldsymbol{S}$ is sufficient for $\boldsymbol{\theta}$ if and only if $f(x_1, \cdots, x_n; \boldsymbol{\theta}) = g(\boldsymbol{S}; \boldsymbol{\theta}) h(x_1, \cdots, x_n)$, where h function does not depend on $\boldsymbol{\theta}$ and g function depends on $x_1, \cdots, x_n$ only through $\boldsymbol{S}$.

$$
\begin{aligned}
f\left(x_1, x_2, \cdots, x_n \mid \theta, \eta\right) &= \frac{1}{\theta^n} e^{-\frac{1}{\theta}\left(\sum_{i=1}^{n} x_i - n\eta\right)} \qquad\qquad x_{1:n} \geqslant \eta \\
&= \frac{1}{\theta^n} \exp\left[\frac{n\eta}{\theta}\right] \exp\left[-\sum_{i=1}^{n} x_i/\theta\right] I_{(\eta,\infty)}\left(x_{1:n}\right)
\end{aligned}
\tag{2.25}
$$

This satisfies the Factorization Criterion equation with $h(x_1, \cdots, x_n) = 1$ and $g(s;\theta,\eta)$ depends on $x_1, \cdots, x_n$ only through $x_{1:n}$ and $\sum_{i=1}^{n} x_i$. So $x_{1:n}$ and $\sum_{i=1}^{n} x_i$ are joint sufficient statistics for θ and η. Notice that θ^* and η^* correspond to a one-to-one transformation of $x_{1:n}$ and $\sum_{i=1}^{n} x_i$, so θ^* and η^* are jointly sufficient for θ and η.

It follows from [4] that $x_{1:n}$ and $\sum_{i=1}^{n}(x_i - x_{1:n})$ jointly complete of θ and η. Note that θ^* and η^* are one-to-one function of $x_{1:n}$ and $\sum_{i=1}^{n}(x_i - x_{1:n})$. So θ^* and η^* are jointly complete for θ and η. Thus we conclude that θ^* and η^* are UMVUE for θ and η.

Chapter 3

Estimation for Incomplete Data

In real life, sometimes it is hard to get a complete data set; often the data are censored. Scientific experiments might have to stop before all items fail because of the limit of time or out of money. Type-I and Type-II censoring are the most basic among the different censoring schemes. **Type-I censoring** happens when the experimental time T is fixed, but the number of failures is random. **Type-II censoring** occurs when the number of failures r is fixed, the experimental time is random.

Hybrid censoring is a mixture of Type-I and Type-II censoring scheme. **Type-I hybrid censoring(Type-I HCS)**[5] considers the experiment being terminated at a random time point $T^* = \min\{T_{r:n}, T\}$, where $T \in (0, \infty)$ is a pre-determined time and r is a predetermined number of failures out of total n items, where $1 \leq r \leq n$. Under this method, the experiment time will be no more than T, which leads to a problem that there might be very few failures that occur before time T, which may result in extremely low efficiency of estimation. Because of that, in this paper, we choose **Type-II hybrid censoring (Type-II HCS)**[6], the experiment ends at a random time $T^* = \max\{T_{r:n}, T\}$, where again T and r are predetermined. This scheme guarantees that at least r failures are observed. And it may be applied in the situation that at least r failures must be observed. If r failures happen before time T, the experiment can continue up to time T to make full use of the facility; if the rth failure does not occur before time T, then the experiment has to continue until the rth failure.

In this chapter, the maximum likelihood method and penalized method are investigated for data from two-parameter exponential distributions under Type-II hybrid censoring. A special case of Type-II hybrid censoring is also discussed.

3.1 MLE for incomplete data

Let T be a pre-chosen experimental time, r be a pre-determined number of items out of total n items. If, at the end of experiment, there is only one observation, then it will be impossible to do estimation since we need to estimate two parameters. Therefore, it is reasonable to assume that at least two failures must be observed, that is, $r \geqslant 2$. The experiment ends at time $T^* = \max\{x_{r:n}, T\}$. Let N be the number of failures that happen before time T, that is, $N = \sum_{i=1}^{n} I\{x_i < T\}$. Let r^* be the number of observations when the experiment stopped, $r^* = \max\{r, N\}$. The Type-II hybrid censoring likelihood function of the observed data is given by

$$L(\theta,\eta) = \frac{n!}{(n-r^*)!\theta^{r^*}} e^{-\frac{1}{\theta}\sum_{i=1}^{r^*}(x_{i:n}-\eta)-\frac{1}{\theta}(n-r^*)(T^*-\eta)}, \quad x_{1:n} \geqslant \eta. \tag{3.1}$$

By taking the logarithm of the likelihood function, we have

$$\ln L(\theta,\eta) = \ln\frac{n!}{(n-r^*)!} - r^*\ln\theta - \frac{1}{\theta}\sum_{i=1}^{r^*}(x_{i:n}-\eta) - \frac{1}{\theta}(n-r^*)(T^*-\eta), \quad x_{1:n} \geqslant \eta. \tag{3.2}$$

Again, the log likelihood function is maximized with respect to η by taking $\eta = x_{1:n}$. To get the MLE for θ, we solve the equation

$$\frac{d\ln L(\theta,\eta)}{d\theta} = -\frac{r^*}{\theta} + \frac{\sum_{i=1}^{r^*}(x_{i:n}-\eta)}{\theta^2} + \frac{(n-r^*)(T^*-\eta)}{\theta^2} = 0.$$

So for $r \geqslant 2$, MLEs of the unknown parameters exist for all values of N and they are given by

$$\hat{\eta} = x_{1:n}, \tag{3.3}$$

and

$$\hat{\theta} = \frac{1}{r^*}\left[\sum_{i=1}^{r^*} x_{i:n} + (n-r^*)T^* - nx_{1:n}\right]. \tag{3.4}$$

When $T < \eta$, this reduce to regular **Type-II censoring**, that is, the experiment is terminated when the first r failures become available. In this case, $r^* = r$ and $T^* = x_{r:n}$. The MLEs for Type-II censoring are

$$\begin{aligned} \hat{\eta} &= x_{1:n}, \\ \hat{\theta} &= \frac{1}{r}\left[\sum_{i=1}^{r} x_{i:n} + (n-r)x_{r:n} - nx_{1:n}\right]. \end{aligned} \tag{3.5}$$

Since $\hat{\eta} = x_{1:n}$ is exponentially distributed as discussed before, the expectation of $\hat{\eta}$ is

$$E[\hat{\eta}] = \frac{\theta}{n} + \eta, \tag{3.6}$$

and the variance of $\hat{\eta}$ is

$$\mathrm{Var}(\hat{\eta}) = \frac{\theta^2}{n^2}. \tag{3.7}$$

We need the following theorem:

Exponential Spacing Theorem[7]: Let $x_{1:n}, x_{2:n}, \cdots, x_{n:n}$ be order statistics form EXP (θ, η). Then $W_0 = n(x_{1:n} - \eta), W_i = (n-i)(x_{(i+1):n} - x_{i:n}), 1 \leqslant i \leqslant n-1$ are independent and identically distributed with common distribution EXP(θ).

By **exponential spacing**,

$$\begin{aligned} \sum_{i=1}^{n-r+1} (n-i)(x_{i+1:n} - x_{i:n}) &= x_{2:n} + \cdots + x_{r-1:n} + (n-r+1)x_{r:n} - (n-1)x_{1:n} \\ &= \sum_{i=1}^{r} x_{i:n} + (n-r)x_{r:n} - nx_{1:n} \sim \mathrm{GAM}(\theta, r-1). \end{aligned} \tag{3.8}$$

So the expectation of $\hat{\theta}$ is

$$E[\hat{\theta}] = \frac{1}{r}(r-1)\theta = (1 - \frac{1}{r})\theta, \tag{3.9}$$

the bias of $\hat{\theta}$ is

$$\mathrm{b}(\hat{\theta}) = E[\hat{\theta}] - \hat{\theta} = -\frac{1}{r}\theta, \tag{3.10}$$

and the variance of $\hat{\theta}$ is

$$\mathrm{Var}(\hat{\theta}) = \frac{r-1}{r^2}\theta^2. \tag{3.11}$$

The MSE of $\hat{\eta}$ is the same as calculated before,

$$\mathrm{MSE}(\hat{\eta}) = \frac{2\theta^2}{n^2}. \tag{3.12}$$

From (3.10), (3.11), the MSE of $\hat{\theta}$ is

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [b(\hat{\theta})]^2 = \frac{r-1}{r^2}\theta^2 + (\theta - \frac{\theta}{r} - \theta)^2 = \frac{\theta^2}{r}. \tag{3.13}$$

When $T > \eta$,

$$\hat{\theta} = \begin{cases} \dfrac{1}{N}\left[\sum\limits_{i=1}^{N} x_{i:n} + (n-N)T - nx_{1:n}\right] & \text{if } x_{r:n} < T, \\ \\ \dfrac{1}{r}\left[\sum\limits_{i=1}^{r} x_{i:n} + (n-r)x_{r:n} - nx_{1:n}\right] & \text{if } x_{r:n} > T. \end{cases} \tag{3.14}$$

This result could be found in [8].

3.2 Penalized MLE for Incomplete Data

The regular Type-II hybrid censoring MLE again uses the smallest observation $x_{1:n}$ to estimate the location parameter η, since the likelihood function is monotone increasing as a function of η. Again, T is a fixed experimental time, r is a fixed number of failures, and N is the number of failures that happened before time T. The experiment ends at time $T^* = \max\{x_{r:n}, T\}$. When the experiment ends, the number of observations $r^* = \max\{r, N\}$. Again, assume $r \geq 2$. Our approach is to add the penalty multiplier $(x_{1:n} - \eta)$ in the usual likelihood function so that the new likelihood function is no longer a monotone function of the location parameter. The penalized likelihood function of the observed data is given by

$$L^*(\theta,\eta) = (x_{1:n} - \eta)\frac{n!}{(n-r^*)!\theta^{r^*}} e^{-\frac{1}{\theta}\sum\limits_{i=1}^{r^*}(x_{i:n}-\eta) - \frac{1}{\theta}(n-r)(T^*-\eta)}, \quad x_{1:n} \geqslant \eta. \tag{3.15}$$

Logarithm of the likelihood function is

$$\ln L^*(\theta,\eta) = \ln(x_{1:n}-\eta) + \ln\frac{n!}{(n-r^*)!} - r^*\ln\theta \\ - \frac{1}{\theta}\sum_{i=1}^{r^*}(x_{i:n}-\eta) - \frac{1}{\theta}(n-r^*)(T^*-\eta), \quad x_{1:n} \geqslant \eta.$$

By solving the following equations

$$\begin{cases} \dfrac{d(\ln L^*(\theta,\eta))}{d\eta} = -\frac{1}{x_{1:n}-\eta} + \dfrac{r^*}{\theta} + \dfrac{n-r^*}{\theta} = 0, \\ \\ \dfrac{d(\ln L^*(\theta,\eta))}{d\theta} = -\dfrac{r^*}{\theta} + \dfrac{\sum\limits_{i=1}^{r^*}(x_{i:n}-\eta)}{\theta^2} + \dfrac{(n-r^*)(T^*-\eta)}{\theta^2} = 0, \end{cases}$$

we have the penalized MLEs

$$\eta^* = \frac{nr^*x_{1:n} - \sum\limits_{i=1}^{r^*} x_{i:n} - (n-r^*)T^*}{n(r^*-1)}, \tag{3.16}$$

and

$$\theta^* = n(x_{1:n}-\eta^*) = \frac{-nx_{1:n} + \sum\limits_{i=1}^{r^*} x_{i:n} + (n-r^*)T^*}{r^*-1}. \tag{3.17}$$

When $T < \eta$, this yields to **Type-II censoring** with $r^* = r$ and $T^* = x_{r:n}$. The Type-II censoring MLEs are

$$\eta^* = \frac{nrx_{1:n} - \sum\limits_{i=1}^{r} x_{i:n} - (n-r)x_{r:n}}{n(r-1)}, \tag{3.18}$$

and

$$\theta^* = \frac{-nx_{1:n} + \sum\limits_{i=1}^{r} x_{i:n} + (n-r)x_{r:n}}{r-1}. \tag{3.19}$$

These estimators are the same results as constructed by Epstein and Sobel[7] before. And they had proved that these estimators are also UMVUE for θ and η.

By using (3.8), the expectation of θ^* is

$$E[\theta^*] = \frac{1}{(r-1)}(r-1)\theta = \theta. \tag{3.20}$$

From (3.20), θ^* is unbiased estimator of θ, and the bias is

$$\mathrm{b}(\theta^*) = E[\theta^*] - \theta = 0. \tag{3.21}$$

The variance of θ^* is

$$\mathrm{Var}(\theta^*) = \frac{1}{(r-1)^2}(r-1)\theta^2 = \frac{\theta^2}{r-1}. \tag{3.22}$$

Note that $\eta^* = x_{1:n} - \dfrac{\theta^*}{n}$, by exponential spacing theorem [7], $x_{1:n}$ and θ^* are independent, and the expectation of η^* is given by

$$E[\eta^*] = E[x_{1:n} - \frac{\theta^*}{n}] = E[x_{1:n}] - \frac{1}{n}E[\theta^*] = \frac{\theta}{n} + \eta - \frac{\theta}{n} = \eta. \tag{3.23}$$

From (3.23), η^* is an unbiased estimator of η, and the bias is

$$\mathrm{b}(\eta^*) = E[\eta^*] - \eta = 0. \tag{3.24}$$

The variance of η^* is

$$\mathrm{Var}(\eta^*) = \mathrm{Var}(x_{1:n} - \frac{\theta^*}{n}) = \frac{\theta^2}{n^2} + \frac{\theta^2}{n^2(r-1)} = \frac{r\theta^2}{n^2(r-1)}. \tag{3.25}$$

From (3.21), (3.22), the MSE of θ^* is

$$\mathrm{MSE}(\theta^*) = \mathrm{Var}(\theta^*) + [b(\theta^*)]^2 = \frac{\theta^2}{r-1}. \tag{3.26}$$

From (3.24), (3.25), MSE of η^* is

$$\mathrm{MSE}(\eta^*) = \mathrm{Var}(\eta^*) + [b(\eta^*)]^2 = \frac{r\theta^2}{n^2(r-1)}. \tag{3.27}$$

Notice that when $r = n$, the penalized Type-II censoring MLE has the same results as penalized MLE for complete data. From (3.12), (3.13), (3.26), (3.27), the penalized Type-II censoring MLE is consistently better than the regular Type-II censoring MLE. Again, the penalized method removes the biases completely.

When $T > \eta$,

$$\eta^* = \begin{cases} \dfrac{nNx_{1:n} - \sum\limits_{i=1}^{N} x_{i:n} - (n-N)T}{n(N-1)}, & \text{if } x_{r:n} < T, \\ \\ \dfrac{nrx_{1:n} - \sum\limits_{i=1}^{r} x_{i:n} - (n-r)x_{r:n}}{n(r-1)}, & \text{if } x_{r:n} > T. \end{cases} \tag{3.28}$$

$$\theta^* = \begin{cases} \dfrac{-nx_{1:n} + \sum\limits_{i=1}^{N} x_{i:n} + (n-N)T}{r-1}, & \text{if } x_{r:n} < T, \\ \\ \dfrac{-nx_{1:n} + \sum\limits_{i=1}^{r} x_{i:n} + (n-r)x_{r:n}}{r-1}, & \text{if } x_{r:n} > T. \end{cases} \tag{3.29}$$

For Type-II hybrid censoring results, since a theoretical justification is difficult, the comparison for different approaches will be made through Monte Carlo simulation in the next chapter.

Chapter 4

Simulation Results

In this chapter, we demonstrate simulation results for both complete data and incomplete data with different approaches.

4.1 Simulation for complete data

For a complete data set, we choose the value of the location parameter η to be 0 and scale parameter θ to be 1. We take different values for sample size n: 10, 20, 50, 100. For each sample size, the simulation is repeated 1000 times. The results of biases and MSEs of regular MLE and penalized MLE are reported in Table 4.1. In the table, the numbers with parenthesis are MSEs, and the numbers without parenthesis are biases.

Table 4.1: The biases and MSEs of estimators for complete data

$\theta = 1, \eta = 0$	Regular MLE		Penalized MLE	
n	$\hat{\eta}$	$\hat{\theta}$	$\hat{\eta}$	$\hat{\theta}$
10	0.1013 (0.0203)	-0.1072 (0.0888)	0.0021 (0.0107)	-0.0081 (0.0955)
20	0.0525 (0.0056)	-0.0417 (0.0482)	0.0021 (0.0030)	0.0088 (0.0515)
50	0.0199 (0.0008)	-0.0184 (0.0207)	-0.0001 (0.0004)	0.0016 (0.0212)
100	0.0106 (0.0002)	-0.0075 (0.0099)	0.0006 (0.0001)	0.0026 (0.0101)

From Table 4.1, it is clear that for each sample size n, the penalized MLEs are better than normal MLEs in terms of bias; the penalized estimators have strong superiority especially for small n; and MSE(θ^*) is grater than MSE($\hat{\theta}$), and the differences between them become smaller as n increasing ; MSE(η^*) is much smaller than MSE($\hat{\eta}$), when n is larger, MSE(η^*) is only half of MSE($\hat{\eta}$).

4.2 Simulation for incomplete data

For incomplete data sets, we choose the values of location η to be 0 and scale parameter θ to be 1. We take $n = 10, r = 2, 5, 8$, $n = 20, r = 4, 10, 18$, $n = 50, r = 10, 25, 48$, and take different values for T: 1.5, 2.5, 5. Each simulation is repeated 1000 times.

Table 4.2 reports the biases and MSEs of estimators for regular MLE and penalized MLE, when the total sample size is 10, the predetermined number of failures are 2, 5, 8, and the predetermined times are 1.5, 2.5, 5. In the table, the numbers inside parenthesis are MSEs, and the numbers without parenthesis are biases.

Table 4.2: The biases and MSEs of estimators for Type-II HCS, with $n = 10$

$\theta = 1, \eta = 0$		Regular MLE		Penalized MLE	
T	r	$\hat{\eta}$	$\hat{\theta}$	$\hat{\eta}$	$\hat{\theta}$
1.5	2	0.1013	-0.0865	-0.0050	0.0637
		(0.0203)	(0.1230)	(0.0115)	(0.1945)
	5	0.0966	-0.0739	-0.0113	0.0785
		(0.0188)	(0.1344)	(0.0115)	(0.2118)
	8	0.1020	-0.1201	0.0025	-0.0048
		(0.0203)	(0.1230)	(0.0115)	(0.1945)
2.5	2	0.1008	-0.0773	-0.0033	0.0408
		(0.0192)	(0.1090)	(0.0103)	(0.1436)
	5	0.0983	-0.0649	-0.0072	0.05487
		(0.0193)	(0.1054)	(0.0113)	(0.1431)
	8	0.1005	-0.0911	-0.0017	0.0222
		(0.0208)	(0.1071)	(0.0119)	(0.1308)
5	2	0.1069	-0.0905	0.0057	0.0119
		(0.0232)	(0.0982)	(0.0119)	(0.0130)
	5	0.1052	-0.0945	0.0045	0.0076
		(0.0215)	(0.1047)	(0.0118)	(0.1202)
	8	0.1020	-0.1007	0.0019	0.0006
		(0.0203)	(0.1037)	(0.0108)	(0.1177)

Table 4.3 reports the biases and MSEs of estimators for regular MLE and penalized MLE, when the total sample size is 20, the fixed number of failures are 4, 10, 18, and the fixed times are 1.5, 2.5, 5. In the table, the numbers inside parenthesis are MSEs, and the numbers without parenthesis are biases.

Table 4.3: The biases and MSEs of estimators for Type-II HCS, with $n = 20$

$\theta = 1, \eta = 0$		Normal MLE		Penalized MLE	
T	r	$\hat{\eta}$	$\hat{\theta}$	$\hat{\eta}$	$\hat{\theta}$
1.5	4	0.0489	-0.0353	-0.0028	0.0347
		(0.0049)	(0.0682)	(0.0026)	(0.0843)
	10	0.0513	-0.0439	0.0001	0.0256
		(0.0054)	(0.0704)	(0.0029)	(0.0856)
	18	0.0507	-0.0521	0.0006	0.0009
		(0.0050)	(0.0587)	(0.0026)	(0.0627)
2.5	4	0.0479	-0.04010	-0.0029	0.0153
		(0.0045)	(0.0541)	(0.0023)	(0.0609)
	10	0.0498	-0.0463	-0.0007	0.0096
		(0.0050)	(0.0541)	(0.0027)	(0.0601)
	18	0.0497	-0.0488	-0.0006	0.0048
		(0.0047)	(0.0533)	(0.0023)	(0.0572)
5	4	0.0526	-0.0406	0.0021	0.0104
		(0.0055)	(0.0496)	(0.0028)	(0.0535)
	10	0.0503	-0.0418	-0.0002	0.0092
		(0.0049)	(0.0496)	(0.0024)	(0.0535)
	18	0.0517	-0.0525	0.0019	-0.0022
		(0.0055)	(0.0480)	(0.0029)	(0.0504)

Table 4.4 reports the biases and MSEs of estimators for regular MLE and penalized MLE, when the total sample size is 50, the fixed number of failures are 10, 25, 48, and the fixed times are 1.5, 2.5, 5. In the table, the numbers inside parenthesis are MSEs, and the numbers without parenthesis are biases.

Table 4.4: The biases and MSEs of estimators for Type-II HCS, with $n = 50$

$\theta = 1, \eta = 0$		Normal MLE		Penalized MLE	
T	r	$\hat{\eta}$	$\hat{\theta}$	$\hat{\eta}$	$\hat{\theta}$
1.5	10	0.0199	-0.0069	-0.0005	0.0198
		(0.0008)	(0.0275)	(0.0004)	(0.0300)
	25	0.0208	-0.0159	0.0005	0.0106
		(0.0009)	(0.0259)	(0.0005)	(0.0279)
	48	0.0203	-0.0243	0.0004	-0.0039
		(0.0008)	(0.0209)	(0.0004)	(0.0212)
2.5	10	0.0200	-0.0164	-0.0002	0.0056
		(0.0008)	(0.0223)	(0.0004)	(0.0234)
	25	0.0205	-0.0147	0.0003	0.0074
		(0.0008)	(0.0234)	(0.0004)	(0.0245)
	48	0.0202	-0.0218	0.0002	-0.0013
		(0.0008)	(0.0203)	(0.0004)	(0.0206)
5	10	0.0188	-0.0242	-0.0011	-0.0041
		(0.0007)	(0.0202)	(0.0004)	(0.0205)
	25	0.0194	-0.0163	-0.0007	0.0039
		(0.0008)	(0.0210)	(0.0004)	(0.0217)
	48	0.0201	-0.0237	0.0001	-0.0037
		(0.0008)	(0.0203)	(0.0004)	(0.0206)

From Table 4.2, Table 4.3 and Table 4.4, for each sample size n, we see that the penalized MLEs are consistently better than normal MLEs in terms of biases; the penalized estimators have strong superiority, especially for small n; and MSE(θ^*) is slightly greater than MSE($\hat{\theta}$), the differences between them become smaller with n increasing; MSE(η^*) is much smaller than MSE($\hat{\eta}$), about half of MSE($\hat{\eta}$).

References

[1] A. E. Sarhan (1954). Estimation of the Mean and Standard Deviation by Order Statistics. The Annals of Mathematical Statistics, Vol. 25, 317-328.

[2] A. Clifford Cohen, Frederick Russell Helm (1973). Estimation in the Exponential Distribution. Technometrics, Vol. 15, 415-418.

[3] Lee J. Bain, Max Engelhardt (1991). Introduction to probability and mathematical statistics. Second edition. Duxbury Classic Series.

[4] E. L. Lehmann, George Casella (1998). Theory of Point Estimation. Second edition. Springer-Verlag, New York.

[5] B. Epstein (1954), Truncated Life Tests in the Exponential Case. The Annals of Mathematical Statistics, Vol. 25, 555-564.

[6] A. Childs, B. Chandrasekhar, N. Balakrishnan, and D. Kundu (2003). Exact likelihood inference based on type-I and type-II hybrid censored samples from the exponential distribution. Annals of the Institute of Statistical Mathematics, Vol. 55, 319-330.

[7] B. Epstein, M. Sobel (1954). Some theorems relevant to life testing from an exponential distribution. The Annals of Mathematical Statistics, Vol. 25, 373-381.

[8] A. Ganguly, S. Mitra, D. Samanta, D. Kundu (2012). Exact inference for the two-parameter exponential distribution under Type-II hybrid censoring. Journal of Statistical Planning and Inference, Vol 142, 613-625.

Appendix A

R code

```
#Mengjie Zheng Project Simulation
# For complete data

# fix the location to be 0
locate = 0
# fix the scale to be 1
scale = 1

# number of times for repeat
B=1000

#n is total number of observations

set.seed(1234)
for(n in c(10, 20, 30, 50, 100, 1000,10000)){
nscale = rep(0, B)
nlocate = rep(0, B)
pscale = rep(0, B)
plocate = rep(0, B)
```

```
for(redu in c(1:B)){
#First generate random uniform numbers,
#then transfer these number into two-parameter exponential
u = runif(n)
data = rep(0,n)
data = - scale*log(1-u)+locate

#sort data set to get the order statistics
data = sort(data)

#The normal MLE to estimate parameters
nlocate[redu] = data[1]
nscale[redu] = mean(data) - data[1]

#The penalized MLE to estimate parameters
plocate[redu] = (n*data[1] - mean(data))/(n-1)
pscale[redu] = (sum(data)-n*data[1])/(n-1)
}
cat("Sample size:", n, "\n")

nolocate = mean(nlocate)
noscale = mean(nscale)
cat("Normal MLE parameter estimation:\n")
cat("    location:",nolocate, "\n")
cat("    scale:",noscale, "\n")

bnolocate = nolocate - locate
bnoscale = noscale - scale
cat("    location bias:",bnolocate, "\n")
cat("    scale bias:",bnoscale, "\n")
```

```
nmselocate = var(nlocate)+(mean(nlocate)-locate)*(mean(nlocate)-locate)
nmsescale = var(nscale)+(mean(nscale)-scale)*(mean(nscale)-scale)
cat("    MSE for location:",nmselocate, "\n")
cat("    MSE for scale:",nmsescale, "\n")

pelocate = mean(plocate)
pescale = mean(pscale)
cat("Penalized MLE estimation:\n")
cat("    location:",pelocate, "\n")
cat("    scale:",pescale, "\n")

pbnolocate = pelocate - locate
pbnoscale = pescale - scale
cat("    location bias:",pbnolocate , "\n")
cat("    scale bias:",pbnoscale, "\n")

pmselocate = var(plocate)+(mean(plocate)-locate)*(mean(plocate)-locate)
pmsescale = var(pscale)+(mean(pscale)-scale)*(mean(pscale)-scale)
cat("    MSE for location:",pmselocate, "\n")
cat("    MSE for scale:",pmsescale, "\n")
cat("\n")
}

#Mengjie Zheng Project Simulation
# For incomplete data

#n is total number of observations
#r is the fixed number of observations
#T is the fixed time
n = 10
```

```
# fix the location to be 0
locate = 0
# fix the scale to be 1
scale = 1

# number of times for repeat
B=1000

set.seed(1234)
for(n in c(10, 20, 50)){
for(r in c(n/5, n/2, n-2)){
for(T in c(1.5, 2.5, 5, 20, 100)){

nscale = rep(0, B)
nlocate = rep(0, B)
pscale = rep(0, B)
plocate = rep(0, B)

for(redu in c(1:B)){
#First generate random uniform numbers,
#then transfer these number into two-parameter exponential
u = runif(n)
data = rep(0,n)
data = - scale*log(1-u)+locate

#sort data set to get the order statistics
data = sort(data)

#Time for the rth observation
timeR = data[r]
```

```
#Get censored data set
j=0
for(i in c(1:n)){
if(data[i] < T || i < r){
j = j+1
}
}

#The normal MLE to estimate parameters

nlocate[redu] = data[1]

if( timeR < T){
scale1 = (sum(data[2:j])+(n - j)*T - (n-1)*data[1])/j
}
if( timeR > T){
scale1 = (sum(data[2:r-1])+(n - r + 1)*data[r] - (n - 1)*data[1])/r
}
nscale[redu] = scale1

#The penalized MLE to estimate parameters
if( timeR < T){
locate2 = (n*j*data[1] - sum(data[1:j])-(n - j)*T)/(n*(j-1))
scale2 = (-n*data[1]+sum(data[1:j])+(n-j)*T)/(j-1)
}
if( timeR > T){
locate2 = (n*r*data[1] - sum(data[1:r])-(n - r)*data[r])/(n*(r-1))
scale2 = (-n*data[1]+sum(data[1:r])+(n-r)*data[r])/(r - 1)
}

plocate[redu] = locate2
```

```
pscale[redu] = scale2

}

cat("Scale:",scale, "\n")
cat("Fixed number:", r, "\n")
cat("Total number:", n, "\n")
cat("Fixed time:", T, "\n")

nolocate = mean(nlocate)
noscale = mean(nscale)
cat("Normal MLE parameter estimation:\n")
cat("    location:",nolocate, "\n")
cat("    scale:",noscale, "\n")

blocation = nolocate - locate
bscale = noscale - scale
cat("    location bias:",blocation, "\n")
cat("    scale bias:",bscale, "\n")

nmselocate = var(nlocate)+(mean(nlocate)-locate)*(mean(nlocate)-locate)
nmsescale = var(nscale)+(mean(nscale)-scale)*(mean(nscale)-scale)
cat("    MSE for location:",nmselocate, "\n")
cat("    MSE for scale:",nmsescale, "\n")

pelocate = mean(plocate)
pescale = mean(pscale)
cat("Penalized MLE estimation:\n")
cat("    location:",pelocate, "\n")
cat("    scale:",pescale, "\n")
```

```
pblocation = pelocate - locate
pbscale = pescale - scale
cat("    location bias:",pblocation, "\n")
cat("    scale bias:",pbscale, "\n")

pmselocate = var(plocate)+(mean(plocate)-locate)*(mean(plocate)-locate)
pmsescale = var(pscale)+(mean(pscale)-scale)*(mean(pscale)-scale)
cat("    MSE for location:",pmselocate, "\n")
cat("    MSE for scale:",pmsescale, "\n")
cat("\n")
}
}
}
```