Comparison of Semi-parametric and Parametric Maximum Likelihood Estimators under Random Censoring

Mavis Amoa-Dadzeasah

A Thesis

in

The Department

of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Mathematics and Statistics)

at

Concordia University

Montreal, Quebec, Canada

August 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mavis Amoa-Dadzeasah**

Entitled: Comparison of Semi-parametric and Parametric Maximum Likelihood Estimators under Random Censoring

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics and Statistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality. Signed by the final Examining Committee:

Signed by the final Examining Committee:

_____ Thesis Supervisor

Dr. A. Sen

_____ Thesis Supervisor

Dr. Y. Chaubey

_____ Examiner

Dr. J. Zhang

Approved by _____

Dr. C. Hyndman (Graduate Program Director)

_____

Dr. P. Sicotte (Dean of Faculty)

_____

Date

Abstract

Comparison of Semi-parametric and Parametric Maximum Likelihood Estimators under Random Censoring

Mavis Amoa-Dadzeasah

In the presence of censored data, selecting an appropriate estimation method is critical for obtaining reliable parameter estimates. This thesis compares the performance of the parametric maximum likelihood estimator (MLE) and a modified, semi-parametric MLE under random censoring. The comparison focuses on estimator of variance, mean squared error, and confidence regions, using theoretical derivations and simulation studies. We examine various combinations of continuous distributions for the event and censoring variables, including Exponential, Weibull, Gamma, Beta, and Pareto models. Our findings show that the modified estimator performs comparably to or better than the parametric estimator, particularly at low censoring rates and for specific distributional configurations. In cases with high censoring, the parametric estimator largely yielded lower variances. These results provide practical guidance for applied statisticians working with randomly censored data, especially in fields such as medical research, reliability engineering, and finance.

Acknowledgments

I acknowledge the assistance I received in completing this work. My deepest gratitude goes first to God Almighty for His grace and sustenance throughout this period. I would also like to express my appreciation to my supervisors Dr. Arusharka Sen, and Dr. Yogendra Chaubey for all the support and guidance in making this work a success. I also appreciate my family and friends for the continued backing and encouragement. My last thanks goes to the authors of the various articles and books that provided relevant information.

This work was supported by the Faculty of Arts and Science Graduate Fellowship, and the Department of Mathematics and Statistics at Concordia University.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The maximum likelihood estimator has been long recognized as a useful and reliable tool in statistical inference and analysis. It has been widely applied across diverse areas, including regression analysis, model selection in machine learning, Bayesian decision theory, and parameter estimation in time series models. Beyond traditional statistical domains, its applications extend to fields such as biotechnology and health research. As a result of its desired properties like efficiency, consistency and asymptotic normality, many researchers and statisticians use this method as a benchmark to make comparisons. In this thesis, we employ maximum likelihood estimation to compare two likelihood approaches for randomly censored data in survival analysis. This area of statistics and its methodologies are particularly important due to their relevance in medical research, engineering, finance, and the social sciences. See the references Cramer and Cramer (1989); Smith, Phillips, Luque-Fernandez, and Maringe (2023).

1.2 Preliminaries

In this section, we present definitions, notation, and distributional forms that will be used throughout the thesis. The focus is on the probability density functions and survival functions of the event and censoring variables considered in the subsequent chapters. These include the Exponential, Weibull, Gamma, Beta, and Pareto distributions. Additionally, we present select theorems and

results from statistical theory that are instrumental in the derivations and theoretical analyses.

1.2.1 Exponential Distribution

The probability density function (pdf) of an exponential distribution is

$$f(x \mid \theta) = \begin{cases} \theta e^{-\theta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where the parameter, $\theta > 0$. The survival function of this distribution, which is the complement of the cumulative distribution function is also given by

$$\bar{F}(x \mid \theta) = e^{-\theta x}, \quad x \geq 0$$

We write $X \sim Exponential(1/\theta)$ or $X \sim Exp(1/\theta)$ to represent a random variable X that has this distribution, where $1/\theta$ is the mean of X.

1.2.2 Weibull Distribution

The pdf of the Weibull distribution with shape parameter p, and scale parameter θ, is given as

$$f(x \mid \theta, p) = \begin{cases} p\theta x^{p-1} e^{-\theta x^p}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

where $\theta > 0,\ p > 0$. The survival function of this distribution, is also given by

$$\bar{F}(x \mid \theta, p) = e^{-\theta x^p}, \quad x \geq 0$$

The notation $X \sim Weibull(\theta, p)$ is used to represent a random variable X which follows Weibull distribution.

1.2.3 Beta Distribution

A beta random variable X with shape parameters α and β is denoted as $X \sim Beta(\alpha, \beta)$, and has pdf

$$f(x \mid \alpha, \beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha > 0,\ \beta > 0$, and $\Gamma(s) = \int_0^\infty t^{s-1}e^{-s}dt$. The survival function is

$$\bar{F}(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_x^1 t^{\alpha-1}(1-t)^{\beta-1}dt, \quad 0 \le x \le 1$$

1.2.4 Gamma Distribution

A gamma random variable X with shape parameter α, and scale parameter θ is denoted as $X \sim Gamma(\alpha, \theta)$, and has the probability density function

$$f(x \mid \alpha, \theta) = \begin{cases} \dfrac{x^{\alpha-1}\ e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\ \theta^{\alpha}}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

where $\alpha > 0,\ \theta > 0$. The corresponding survival function is given as

$$\bar{F}(x \mid \alpha, \theta) = \frac{1}{\Gamma(\alpha)}\, \Gamma\left(\alpha, \frac{x}{\theta}\right), \quad x \ge 0$$

where

$$\Gamma\left(\alpha, \frac{x}{\theta}\right) = \int_{\frac{x}{\theta}}^\infty t^{\alpha-1}\ e^{-t}dt$$

1.2.5 Pareto Distribution

If X is a random variable with Pareto distribution having shape parameter λ and scale parameter θ, the pdf of X is

$$f_X(x) = \begin{cases} \dfrac{\lambda\, \theta^{\lambda}}{(\theta+x)^{\lambda+1}}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

where $\lambda > 0,\ \theta > 0$, and the survival function is

$$\bar{F}_X(x) = \left(\frac{\theta}{\theta+y}\right)^{\lambda}, \quad x \ge 0$$

Denoted as $X \sim Pareto(\lambda, \theta)$, this distribution is mostly used as a censoring distribution in this thesis.

See Walck et al. (2007), Forbes, Evans, Hastings, and Peacock (2011), and Ross (2014) as references for these distributions.

1.2.6 Supporting Theorems

Theorem 1.1. *(The Central Limit Theorem) If $X_1, \dots X_n$ are independent and identically distributed (iid) with finite mean $\mathbb{E}_\theta(X_i) = \mu(\theta)$, and finite variance $Var_\theta(X_i) = \sigma^2(\theta) > 0$, then as $n \to \infty$,*

$$\frac{\bar{X}_n - \mu(\theta)}{\sigma(\theta)/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu(\theta))}{\sigma(\theta)} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\xrightarrow{d}$ denotes *convergence in distribution.* See Section 4.2 of Hogg, McKean, and Craig (2018).

Theorem 1.2. *(Slutsky's Theorem) If X_n, A_n, B_n are random variables, a and b are constants, where $A_n \xrightarrow{d} a$, $B_n \xrightarrow{d} b$, and $X_n \xrightarrow{d} X$, then it follows that $A_n + B_n X_n \xrightarrow{d} a + bX$.*

Note that for a constant a, $A_n \xrightarrow{d} a$ is equivalent to $A_n \xrightarrow{P} a$, where $\xrightarrow{P}$ denotes *convergence in probability.* See Section 5.2 of Hogg et al. (2018).

Theorem 1.3. *Suppose X has a $\mathcal{N}_k(\mu, \Sigma)$ distribution, where Σ is positive definite. Then the random variable $Y = (X - \mu)^\top \Sigma^{-1} (X - \mu)$ has $\chi^2(k)$ distribution.*

See Section 3.5 of Hogg et al. (2018).

1.3 Maximum Likelihood Estimator without Random Censoring

We provide a brief overview of the maximum likelihood estimator (MLE) without random censoring, which would later be useful in establishing the limiting variance of the semi-parametric estimator we propose. The maximum likelihood estimator of an unknown parameter θ of a probability distribution is that value $\hat{\theta}$ of θ which maximizes the probability or probability density of an observed sample from that distribution, Lehmann and Casella (2006). Let $X_1, \dots, X_n$ be iid random variables which follow a *regular model* (see Appendix A for the definition and some details of the *regular model*) with a common probability density function (pdf) or mass function (pmf) $f(x \mid \theta)$,

where $\theta \in \Theta$, the parameter space. It follows that the likelihood function, or the joint probability distribution and its log are given by

$$L(\theta \mid X) = \prod_{i=1}^{n} f(X_i \mid \theta) \tag{1.1}$$

$$\ln L(\theta \mid X) = \sum_{i=1}^{n} \ln f(X_i \mid \theta) \tag{1.2}$$

The MLE maximizes Equation 1.2. In *regular models* (see Appendix A) where the $\ln L(\theta \mid X)$ is differentiable and the maxima occurs at an interior point of Θ, the derivative of the log-likelihood function, also known as the score function is equal to zero at $\hat{\theta}$, i.e. $\hat{\theta}$ is obtained as the solution of θ such that

$$\frac{d}{d\theta} \ln L(\theta \mid X) = 0.$$

Suppose $\theta \in \Theta$ is the true parameter value. Then, the score function $S(\theta \mid X)$, defined as

$$S(\theta \mid X) = \sum_{i=1}^{n} \frac{d}{d\theta} \ln f(X_i \mid \theta) \ , \tag{1.3}$$

has the following properties in a regular model, as seen in Hogg et al. (2018):

1. $S(\hat{\theta} \mid X) = 0$, since $\hat{\theta}$ maximizes the log-likelihood.
2. Its expectation is zero, i.e. $\mathbb{E}_\theta[S(\theta \mid X)] = 0$.
3. The negative expectation of the first derivative of $\dfrac{d}{d\theta} \ln f(X \mid \theta)$ gives the Fisher Information $I(\theta)$.

$$I(\theta) = -\mathbb{E}_\theta \left[\frac{d^2}{d\theta^2} \ln f(X \mid \theta) \right] \tag{1.4}$$

The asymptotic distribution of $\hat{\theta}$ is derived using a Taylor series expansion of the score function around the true parameter θ:

$$\begin{aligned} S(\hat{\theta} \mid X) &= S(\theta \mid X) + (\hat{\theta} - \theta) \left. \frac{dS(\theta \mid X)}{d\theta} \right|_{\theta=\hat{\theta}} \\ S(\hat{\theta} \mid X) &\approx S(\theta \mid X) + (\hat{\theta} - \theta) \frac{dS(\theta \mid X)}{d\theta} \end{aligned}$$

$$\begin{aligned}
\frac{dS(\theta \mid X)}{d\theta} &\approx -nI(\theta) \text{ since } I(\theta) = \lim_{n\to\infty} -\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\ln f(X_i \mid \theta) \\
0 &\approx S(\theta \mid X) - (\hat{\theta} - \theta)\, nI(\theta) \\
(\hat{\theta} - \theta) &\approx \frac{S(\theta \mid X)}{nI(\theta)}
\end{aligned}$$

By the Central Limit Theorem, $S(\theta \mid X)$ is asymptotically normally distributed, where θ is the true parameter value.

$$\frac{S(\theta \mid X)}{\sqrt{n}} \xrightarrow{d} \mathcal{N}\left(0, I(\theta)\right).$$

Therefore, the limiting distribution of the maximum likelihood estimator $\hat{\theta}$ under regularity conditions is

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, I^{-1}(\theta)\right). \tag{1.5}$$

The regularity conditions are outlined in Appendix A. The likelihood function used to obtain the MLE above is the specific case of an uncensored data. In survival analysis, we are interested in studying the time until some population survives or experiences an event of interest. However, many factors could prevent the entire population from experiencing the event. This results in censored data. Therefore, censoring is said to have occurred when information on the start or end of some event of interest is missing, Turkson, Ayiah-Mensah, and Nimoh (2021) and Kalbfleisch and Prentice (2002). Generally, observed data comprises both censored and uncensored data, thus, the probability distribution of this kind of data has a different form than the case without random censoring.

1.4 Maximum Likelihood Estimator under Random Censoring

The statistical inference methods employed in this research are based on likelihood functions derived from the observed data. Let X_i be the lifetime of observed data, which characterizes the time between entry into the study and the event, with density function $f(x \mid \theta)$, and cumulative density function $F(x \mid \theta)$. Let Y_i be the random censoring variable independent of X_i, Stute (1995), which characterizes the time from entry into the study and the end of the study, with density function $g(y)$, and cumulative density function $G(y)$. We define the lifetime random variable as $Z = \min(X, Y), z_i, i = 1, \ldots, n$ is similarly defined. We also define a censoring indicator δ_i as

$\delta_i = \mathbb{I}(X_i \leq Y_i)$ such that

$$\delta = \begin{cases} 1, & \text{if } X \leq Y \quad \text{i.e. uncensored} \\ 0, & \text{if } X > Y \quad \text{i.e. censored} \end{cases}$$

The joint distribution $h(\delta, z \mid \theta)$ is based on the probability distribution of (δ_i, z_i):

$$\begin{aligned} h(\delta, z \mid \theta) &= \begin{cases} \mathbb{P}[\delta = 1,\, X] = \bar{G}(x)\, f(x \mid \theta) \\ \mathbb{P}[\delta = 0,\, Y] = \bar{F}(y \mid \theta)\, g(y) \end{cases} \\ &= \Big(\bar{G}(x)\, f(x \mid \theta)\Big)^{\delta_i} \Big(\bar{F}(y \mid \theta)\, g(y)\Big)^{1-\delta_i}, \quad \text{see Lawless (2011).} \end{aligned}$$

where $z = \min(x, y)$, and, $\bar{G}(.)$ and $\bar{F}(.)$ are the respective survival functions corresponding to $G(.)$ and $F(.)$, respectively. Therefore, the likelihood function is given as

$$\begin{aligned} L(\theta \mid \delta, Z) &= \prod_{i=1}^{n} h(\delta_i, Z_i \mid \theta) \\ L(\theta \mid \delta, Z) &= \prod_{i=1}^{n} f(Z_i \mid \theta)^{\delta_i}\, \bar{G}(Z_i)^{\delta_i}\; g(Z_i)^{1-\delta_i}\, \bar{F}(Z_i \mid \theta)^{1-\delta_i} \qquad (1.6) \\ \ln L(\theta \mid \delta, Z) &= \sum_{i=1}^{n} \Big[\delta_i \ln \bar{G}(Z_i) + \delta_i \ln f(Z_i \mid \theta) + (1-\delta_i) \ln \bar{F}(Z_i \mid \theta) + (1-\delta_i) \ln g(Z_i)\Big] \qquad (1.7) \end{aligned}$$

See Kalbfleisch and Prentice (2002) and Klein and Moeschberger (2003). If the data is entirely uncensored, we note that $\delta_i = 1\ \forall i$ and although $g(y)$ does not appear in the likelihood, $\bar{G}(x) = \mathbb{P}(Y > x)$ still appears since $Y > X$ always. Thus, even in uncensored cases, the censoring distribution matters in the likelihood through $\bar{G}(x)$. The MLE under this likelihood solves $\dfrac{d}{d\theta} \ln L(\theta \mid \delta, Z) = 0$. That is,

$$\frac{1}{n} \sum_{i=1}^{n} \delta_i \frac{d}{d\theta} \ln f(Z_i \mid \theta) + \frac{1}{n} \sum_{i=1}^{n} (1-\delta_i) \frac{d}{d\theta} \ln \bar{F}(Z_i \mid \theta) \Big|_{\theta = \hat{\theta}} = 0$$

Given that $S(\theta \mid \delta, Z) = \dfrac{d}{d\theta} \ln L(\theta \mid \delta, Z)$, using a Taylor series expansion around θ gives

$$(\hat{\theta} - \theta) \approx - \frac{S(\theta \mid \delta, Z)}{\frac{1}{n} \sum_{i=1}^{n} \left(\delta_i \dfrac{d^2}{d\theta^2} \ln f(Z_i \mid \theta) + (1-\delta_i) \dfrac{d^2}{d\theta^2} \ln \bar{F}(Z_i \mid \theta) \right)} = \frac{A_n}{B_n}$$

$$\sqrt{n}\,(A_n) \xrightarrow{d} \mathcal{N}\,(0, I_{RC}(\theta)) \quad \text{and} \quad B_n \xrightarrow{d} I_{RC}(\theta)$$

Thus, by Slutsky's Theorem, the limiting distribution of $\hat{\theta}$ under random censoring is

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0 \ , \ I_{RC}^{-1}(\theta)\right) \tag{1.8}$$

where $I_{RC}(\theta)$ is the Fisher information under random censoring:

$$I_{RC}(\theta) = \begin{cases} \mathbb{E}_\theta \left(\delta \dfrac{d}{d\theta} \ln f(Z \mid \theta) + (1-\delta) \dfrac{d}{d\theta} \ln \bar{F}(Z \mid \theta) \right)^2 \\ - \ \mathbb{E}_\theta \left(\delta \dfrac{d^2}{d\theta^2} \ln f(Z \mid \theta) + (1-\delta) \dfrac{d^2}{d\theta^2} \ln \bar{F}(Z \mid \theta) \right) \end{cases} \tag{1.9}$$

Properties of $S(\theta \mid \delta, Z)$ include:

1. The expectation of the score function is equal to zero, i.e. $\mathbb{E}_\theta[S(\theta \mid \delta, Z)] = 0$.
2. The negative expectation of its first derivative gives the Fisher Information, i.e.

$$\mathbb{E}\left(\frac{d^2}{d\theta^2} \ln h(\delta, Z \mid \theta) \right) = -\mathbb{E}\left(\frac{d}{d\theta} \ln h(\delta, Z \mid \theta) \right)^2$$

Proofs of these properties are shown in Section 1.4.1 below.

1.4.1 Further proofs

First, we show that the expectation of the score function in the uncensored data case equals 0.

Proof of $\mathbb{E}_\theta[S(\theta \mid X)] = 0$.

Recall that for a pdf $f(x \mid \theta)$,

$$\begin{aligned} \int f(x \mid \theta)\, dx &= 1 \\ \frac{d}{d\theta} \int f(x \mid \theta)\, dx &= 0 \\ \int \frac{\frac{d}{d\theta} f(x \mid \theta)}{f(x \mid \theta)} \cdot f(x \mid \theta)\, dx &= 0 \\ \int \left(\frac{d}{d\theta} \ln f(x \mid \theta) \right) f(x \mid \theta)\, dx &= 0 \\ \therefore \ \ \mathbb{E}_\theta \left[\frac{d}{d\theta} \ln f(X \mid \theta) \right] &= 0 \end{aligned}$$

Similarly, we can show that the expectation of the score function in the censored case is zero.

Proof of $\mathbb{E}_\theta[S(\theta \mid \delta, Z)] = 0$.

$$\mathbb{E}_\theta \left[\delta \, \frac{\dfrac{d}{d\theta} f(Z \mid \theta)}{f(Z \mid \theta)} + (1 - \delta) \, \frac{\dfrac{d}{d\theta} \bar{F}(Z \mid \theta)}{\bar{F}(Z \mid \theta)} \right] = 0$$

That is,

$$\int \mathbb{I}(X \leq Y) \cdot \frac{\dfrac{d}{d\theta} f(x \mid \theta)}{f(x \mid \theta)} \cdot f(x \mid \theta) \, g(y) \, dx \, dy + \int \mathbb{I}(X > Y) \cdot \frac{\dfrac{d}{d\theta} \bar{F}(y \mid \theta)}{\bar{F}(y \mid \theta)} \cdot f(x \mid \theta) \, g(y) \, dx \, dy = 0$$

Since the joint distribution $h(\delta, z \mid \theta)$ is a pdf,

$$\begin{aligned} \sum_{\delta=0}^{1} \int h(\delta, z \mid \theta) \, dz &= 1 \\ \text{i.e.,} \quad \int \bar{G}(x) \, f(x \mid \theta) \, dx &+ \int \bar{F}(y \mid \theta) \, g(y) \, dy = 1, \quad \text{for all } \theta \qquad (1.10) \\ \int \bar{G}(x) \, \frac{d}{d\theta} f(x \mid \theta) \, dx &+ \int \frac{d}{d\theta} \bar{F}(y \mid \theta) \, g(y) \, dy = 0, \quad \text{for all } \theta \qquad (1.11) \end{aligned}$$

Hence for all θ, and for all $r \geq 1$,

$$\int \bar{G}(x) \, \frac{d^r}{d\theta^r} f(x \mid \theta) \, dx + \int \frac{d^r}{d\theta^r} \bar{F}(y \mid \theta) \, g(y) \, dy = 0 \qquad (1.12)$$

$\mathbb{E}_\theta[S(\theta \mid \delta, Z)]$ can be further expressed as

$$\mathbb{E}_\theta[S(\theta \mid \delta, Z)] = \int \bar{G}(x) \cdot \frac{d}{d\theta} f(x \mid \theta) \, dx + \int \frac{d}{d\theta} \bar{F}(y \mid \theta) \cdot g(y) \, dy$$

Then, it follows from Equation 1.11 that $\mathbb{E}_\theta[S(\theta \mid \delta, Z)] = 0$.

Proof of $\mathbb{E}\left(\dfrac{d^2}{d\theta^2} \ln h(\delta, Z \mid \theta) \right) = -\mathbb{E}\left(\dfrac{d}{d\theta} \ln h(\delta, Z \mid \theta) \right)^2$.

We start by taking the derivative of the score function, and show that its expectation is equal to the right hand side.

$$
\begin{aligned}
& \frac{d}{d\theta}\left(\delta\,\frac{\frac{d}{d\theta}f(z\mid\theta)}{f(z\mid\theta)}+(1-\delta)\,\frac{\frac{d}{d\theta}\bar{F}(z\mid\theta)}{\bar{F}(z\mid\theta)}\right)= \\
= \quad & \delta\left(\frac{f(z\mid\theta)\,\frac{d^2}{d\theta^2}f(z\mid\theta)-\left(\frac{d}{d\theta}f(z\mid\theta)\right)^2}{f^2(z\mid\theta)}\right)+(1-\delta)\left(\frac{\bar{F}(z\mid\theta)\,\frac{d^2}{d\theta^2}\bar{F}(z\mid\theta)-\left(\frac{d}{d\theta}\bar{F}(z\mid\theta)\right)^2}{\bar{F}^2(z\mid\theta)}\right) \\
= \quad & \int \mathbb{I}(X\le Y)\cdot\frac{f(x\mid\theta)\,\frac{d^2}{d\theta^2}f(x\mid\theta)-\left(\frac{d}{d\theta}f(x\mid\theta)\right)^2}{f^2(x\mid\theta)}\cdot f(x\mid\theta)\,g(y)\,dx\,dy \quad + \\
& \int \mathbb{I}(X>Y)\cdot\frac{\bar{F}(y\mid\theta)\,\frac{d^2}{d\theta^2}\bar{F}(y\mid\theta)-\left(\frac{d}{d\theta}\bar{F}(y\mid\theta)\right)^2}{\bar{F}^2(y\mid\theta)}\cdot f(x\mid\theta)\,g(y)\,dx\,dy \\
= \quad & -\int\frac{\left(\frac{d}{d\theta}f(x\mid\theta)\right)^2}{f^2(x\mid\theta)}\cdot f(x\mid\theta)\,\bar{G}(x)\,dx-\int\frac{\left(\frac{d}{d\theta}\bar{F}(y\mid\theta)\right)^2}{\bar{F}^2(y\mid\theta)}\cdot\bar{F}(y\mid\theta)\,g(y)\,dy \\
= \quad & -\int\left(\frac{d}{d\theta}\ln f(x\mid\theta)\right)^2 f(x\mid\theta)\,\bar{G}(x)\,dx-\int\left(\frac{d}{d\theta}\ln\bar{F}(y\mid\theta)\right)^2\bar{F}(y\mid\theta)\,g(y)\,dy \\
= \quad & -\mathbb{E}\left[\delta\cdot\left(\frac{d}{d\theta}\ln f(Z\mid\theta)\right)+(1-\delta)\cdot\left(\frac{d}{d\theta}\ln\bar{F}(Z\mid\theta)\right)\right]^2 \\
= \quad & -\mathbb{E}\left(\frac{d}{d\theta}\ln h(\theta\mid\delta,Z)\right)^2
\end{aligned}
$$

hence the proof.

1.4.2 Censoring Rate

One can find the random censoring rate by finding the probability that $\delta = 0$, or alternatively, the failure rate as the probability $\delta = 1$. Therefore, the random censoring probability is $\mathbb{P}[\delta = 0] = \mathbb{P}(X > Y)$.

$$
\begin{aligned}
\mathbb{P}(X>Y) &= \int\mathbb{P}(X>y)\,g(y)\;dy \\
&= \int\bar{F}(y\mid\theta)\,g(y)\;dy
\end{aligned}
$$

The probability of failure, $\mathbb{P}[\delta = 1] = \mathbb{P}(X \le Y) = \mathbb{P}(X < Y)$ since X and Y are continuous.

$$\begin{aligned}\mathbb{P}(X < Y) &= \int \mathbb{P}(Y > x)\, f(x \mid \theta)\, dx \\ &= \int \bar{G}(x)\, f(x \mid \theta)\, dx\end{aligned}$$

1.5 Research Objectives and Methods

In this research, we consider two likelihoods: the full parametric model as outlined above, and a semi-parametric one based on the Non-Parametric Maximum Likelihood Estimator by Kaplan and Meier (1958). The latter will be referred to as "compact" likelihood or "modified" estimator (defined in Section 2.1), and the former as "parametric" estimator. The compact likelihood modifies the fully parametric model by incorporating Kaplan–Meier (KM) weights (defined in Section 2.1) in place of the censoring indicators. We evaluate the performance of the maximum likelihood estimators of these two likelihoods by examining their limiting variances and confidence regions where applicable. The limiting variances are obtained using the main result of Stute (1995). We study which estimator performs better under varying censoring probabilities and distributional setups. We would also gain more insights on conditions under which comparable results are obtained.

Other researchers have also explored alternative estimation procedures to the traditional maximum likelihood estimator in the presence of censoring. In particular, weighted likelihood methods and semi-parametric approaches have been developed (Murray (2001); Ren (2008); Zhou and Liang (2011); Ren and Lyu (2024)) which have shown promise in reducing bias and improving efficiency. This thesis builds upon such methodologies.

To carry out this analysis effectively, we consider a variety of distributions for the event variable X along with corresponding distributions for the censoring variable Y. For each distributional pair (X, Y), the limiting variances of the estimators are evaluated both theoretically and via simulation. The behavior of the variances are assessed under varying censoring rates, controlled by the parameter values from the censoring distribution. Both single-parameter distributions (e.g., $Exp(1/\theta)$) and two-parameter distributions (e.g., $Gamma(\alpha, \theta)$) for X are considered. All computational analyses were performed using R version 4.5.1.

1.6 Plan of the Thesis

The rest of the thesis is organized as follows. A comprehensive analysis of the single parameter case of X is outlined in Chapter 2. This includes a detailed derivation of the limiting variances of the MLEs for both the parametric and compact likelihood cases, based on general forms of X and Y. This chapter also explores specific sampled distributions for X and Y. Chapter 3 extends the discussion to the multi-parameter case, providing generalized forms of the vectorized likelihood functions and corresponding MLEs. As in Chapter 2, selected distributions are analyzed. Additionally, this chapter revisits key concepts from the multivariate normal distribution, which serves as the foundation for evaluating the limiting variance-covariance matrices. The thesis concludes with a summary of key findings in Chapter 4. Regularity conditions based on which the theory of the maximum likelihood estimators are established, and complementary results are outlined in Appendix A and Appendix B respectively.

Chapter 2

Single Parameter Case

This chapter begins with an introduction to the compact likelihood. Unlike the fully parametric likelihood discussed in Section 1.4, which incorporates the censoring indicator δ, the compact likelihood, $L^M(\theta \mid Z)$ replaces these indicators with Kaplan–Meier weights W_{in}, as seen in Stute (1995).

$$\begin{aligned} \ln L^M(\theta \mid Z) &= \sum_{i=1}^{n} W_{in} \ln f(Z_i \mid \theta) \\ \frac{d}{d\theta} \ln L^M(\theta \mid Z) &= \sum_{i=1}^{n} W_{in} \frac{d}{d\theta} \ln f(Z_i \mid \theta) \end{aligned}$$

where,

$$W_{in} = \frac{\delta_i}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{\delta_j}{n-j+1}\right) \tag{2.1}$$

The modified likelihood above is motivated by the fact that a sample average of the form $n^{-1} \sum_{i=1}^{n} \varphi(X_i)$ based on an uncensored sample, is to be replaced by $\sum_{i=1}^{n} W_{in} \varphi(Z_{in})$ in a censored sample, where $Z_{in},\ 1 \le i \le n$, are the order-statistics of the censored sample. The MLE of θ for this likelihood, $\hat{\theta}^M$ is obtained at $\frac{d}{d\theta} \ln L^M(\theta \mid Z)\Big|_{\theta=\hat{\theta}^M} = 0$.

Proceeding as in the case of uncensored data, we have

$$\sqrt{n}\left(\hat{\theta}^M - \theta\right) \approx \sqrt{n}\, \frac{\sum_{i=1}^{n} W_{in} \frac{d}{d\theta} \ln f(Z_i \mid \theta)}{-\sum_{i=1}^{n} W_{in} \frac{d^2}{d\theta^2} \ln f(Z_i \mid \theta)} = \frac{A_n^M}{B_n^M}$$

Further, using Theorem 1 of Stute (1995),

$$A_n^M \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(\theta)\right) \quad \text{and} \quad B_n^M \xrightarrow{d} I(\theta)$$

where

$$\sigma^2(\theta) = \mathbb{E}_\theta \left[\delta \, \frac{\varphi_\theta(Z)}{\bar{G}(Z)} + (1-\delta)\gamma_{\varphi_\theta}(Z) - \Gamma_{\varphi_\theta}(Z) \right]^2 \tag{2.2}$$

and,

$$\begin{aligned} \varphi_\theta(z) &= \frac{d}{d\theta} \ln f(z \mid \theta) \\ \gamma_{\varphi_\theta}(z) &= \frac{1}{\bar{F}(z \mid \theta)\, \bar{G}(z)} \cdot \int \mathbb{I}(t > z)\, \varphi(t)\, f(t \mid \theta)\, dt \\ \Gamma_{\varphi_\theta}(z) &= \int \frac{\mathbb{I}(z > s)\, \dfrac{d}{d\theta} \bar{F}(s \mid \theta)}{\bar{F}(s \mid \theta)\, \bar{G}^2(s)} \cdot g(s)\, ds \end{aligned}$$

Section 2.2.2 describes in detail the terms in Equation 2.2 above. By Slutsky's Theorem,

$$\sqrt{n}\left(\hat{\theta}^M - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \, \frac{\sigma^2(\theta)}{I^2(\theta)}\right) \tag{2.3}$$

2.1 Non-Parametric Maximum Likelihood Estimator

The Kaplan-Meier estimator, as proposed by Kaplan and Meier (1958) is also interpreted as a non-parametric maximum likelihood estimator (NPMLE), Rodrıguez (2005). This estimator assumes that the censored and uncensored subjects have the same chances of survival at any time, censoring probabilities are the same for all subjects regardless of their entry time into the study, and events occur exactly at the recorded times, with no lag, see Goel, Khanna, and Kishore (2010). The Kaplan-Meier estimate of the ith order statistic, $\bar{F}(Z_{i:n})$ is defined as

$$\bar{F}(Z_{i:n}) = \prod_{j=1}^{i} \left(1 - \frac{\delta_j}{n-j+1}\right), \quad 1 \le i \le n$$

where $Z_{i:n}$ are the order-statistics.

Hence, if $\delta_1 = \cdots = \delta_n = 1$, then $\bar{F}(Z_{i:n}) = \left(1 - \frac{1}{n}\right)\left(1 - \frac{1}{n-1}\right) \ldots \left(1 - \frac{1}{n-i+1}\right)$

$\implies \bar{F}(Z_{i:n}) = \dfrac{n-i}{n}$ when $\delta_1 = \cdots = \delta_n = 1$.

Note that,

$$\bar{F}(Z_{i:n}) = \prod_{j=1}^{i}\left(1-\frac{\delta_j}{n-j+1}\right) \implies \bar{F}(Z_{i:n}) = \left(1-\frac{\delta_i}{n-i+1}\right)\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{n-j+1}\right)$$

Thus,

$$\begin{aligned}
\bar{F}(Z_{i-1:n}) - \bar{F}(Z_{i:n}) &= \prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{n-j+1}\right) - \left(1-\frac{\delta_i}{n-i+1}\right)\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{n-j+1}\right) \\
&= \prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{n-j+1}\right)\left[1-\left(1-\frac{\delta_i}{n-i+1}\right)\right] \\
&= \prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{n-j+1}\right)\left[\frac{\delta_i}{n-i+1}\right]
\end{aligned}$$

Therefore as seen in Stute (1995), $W_{in} = \bar{F}(Z_{i-1:n}) - \bar{F}(Z_{i:n}) = \dfrac{\delta_i}{n-i+1}\displaystyle\prod_{j=1}^{i-1}\left(1-\frac{\delta_j}{n-j+1}\right)$.

2.2 Limiting Distribution of Estimators

One way to assess the efficiency of an estimator is by examining its variance, Saleem, Sanaullah, Al-Essa, Bashir, and Al Mutairi (2023). Given that maximum likelihood estimators (MLEs) are asymptotically normal, we compute the limiting variances of the two estimators as a basis for comparison. The variance provides insight into an estimator's performance, with lower variance generally indicating greater efficiency, thus being more desirable, Lehmann and Casella (2006). These variances can be evaluated analytically by computing the integrals presented in the subsequent sections for specific choices of the event and censoring distributions. Generally, desirable results are said to be obtained when the limiting variance of the modified estimator, $Var_M(\theta)$ is less than, or approximately equal to that of the parametric estimator, $Var_P(\theta)$, i.e. where $Var_M(\theta) \leq Var_P(\theta)$, and where $Var_M(\theta) \approx Var_P(\theta)$. Nonetheless, we will also present cases where this is not achieved.

To support the theoretical results, we carry out a computational analysis. This involves selecting a family of distributions for the event variable X by varying the parameter θ across a specified range, while holding the parameter λ of the censoring variable Y fixed at several points within that range. We then compute the variance for each (θ, λ) pair across the entire range of θ, capturing the behavior of the estimators under varying censoring rates and parameter values. Additionally, we simulate 1000 datasets each of size 100 from the specified distributions and evaluate the estimators

by comparing the mean squared errors (MSEs) of the maximum likelihood estimators, DeGroot and Schervish (2012), defined as

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2$$

where $\hat{\theta}_i$ is the maximum likelihood estimate for the ith simulation, and n is the number of datasets simulated.

2.2.1 Parametric Estimator

From expression 1.8, the limiting variance of $\hat{\theta}$ under random censoring is:

$$Var_P(\theta) = \frac{1}{I_{RC}(\theta)} \tag{2.4}$$

where,

$$I_{RC}(\theta) = \quad \mathbb{E}_\theta\left(\delta\frac{d}{d\theta}\ln f(Z \mid \theta) + (1-\delta)\frac{d}{d\theta}\ln \bar{F}(Z \mid \theta)\right)^2$$

We showed in Section 1.4.1 that

$$-I_{RC}(\theta) = -\int \frac{\left(\frac{d}{d\theta}f(x \mid \theta)\right)^2}{f^2(x \mid \theta)} \cdot f(x \mid \theta)\,\bar{G}(x)\,dx - \int \frac{\left(\frac{d}{d\theta}\bar{F}(y \mid \theta)\right)^2}{\bar{F}^2(y \mid \theta)} \cdot \bar{F}(y \mid \theta)\,g(y)\,dy$$

$$\implies I_{RC}(\theta) = \int \left(\frac{d}{d\theta}\ln f(x \mid \theta)\right)^2 f(x \mid \theta)\,\bar{G}(x)\,dx + \int \frac{\left(\frac{d}{d\theta}\bar{F}(y \mid \theta)\right)^2}{\bar{F}(y \mid \theta)} \cdot g(y)\,dy$$

Thus, the limiting variance in the parametric likelihood case is the multiplicative inverse of the integral above.

2.2.2 Modified Estimator

From expression 2.3, the limiting variance of the modified estimator is given as

$$Var_M(\theta) = \frac{\sigma^2(\theta)}{I^2(\theta)} \tag{2.5}$$

where $I(\theta)$ is the inverse of the limiting variance in the non-censored case as outlined in Equation 1.5.

From Stute (1995),

$$\sigma^2(\theta) = \mathbb{E}_\theta \left[\delta \, \frac{\varphi_\theta(Z)}{\bar{G}(Z)} + (1-\delta)\gamma_{\varphi_\theta}(Z) - \Gamma_{\varphi_\theta}(Z) \right]^2, \text{ where } \varphi_\theta(z) = \frac{d}{d\theta} \ln f(z \mid \theta)$$

$\varphi_\theta(.)$ and $\varphi(.)$ are used interchangeably to denote the same function, i.e., $\varphi_\theta(.) = \varphi(.)$.

$$\sigma^2(\theta) = \mathbb{E}_\theta \left[\delta \, \frac{\varphi^2(Z)}{\bar{G}^2(Z)} + (1-\delta)\, \gamma_\varphi^2(Z) + \Gamma_\varphi^2(Z) - 2\,\delta \, \frac{\varphi(Z)}{\bar{G}(Z)} \cdot \Gamma_\varphi(Z) - 2\,(1-\delta)\, \gamma_\varphi(Z)\, \Gamma_\varphi(Z) \right]$$

$$= \underbrace{\mathbb{E}_\theta \left[\frac{\varphi^2(X)}{\bar{G}(X)} \right]}_{\mathbf{A}} + \underbrace{\int \bar{F}(y)\, \gamma_\varphi^2(y)\, g(y)\, dy}_{\mathbf{B}} + \underbrace{\mathbb{E}_\theta \left[\Gamma_\varphi^2(Z) \right]}_{\mathbf{C}} - \underbrace{2\, \mathbb{E}_\theta \left[\varphi(X)\, \Gamma_\varphi(X) \right]}_{\mathbf{D}} - \underbrace{2\, \mathbb{E}_\theta \left[\bar{F}(Y)\, \gamma_\varphi(Y)\, \Gamma_\varphi(Y) \right]}_{\mathbf{E}}$$

where,

$$\gamma_\varphi(y) = \frac{S_\varphi(y)}{\bar{H}(y)} = \frac{1}{\bar{H}(y)} \cdot \int \mathbb{I}(t > y)\, \varphi(t)\, f(t)\, dt, \;\; \bar{H}(y) = \bar{F}(y)\, \bar{G}(y),$$

$$S_\varphi(y) = \int \mathbb{I}(t > y)\, \varphi(t)\, dF(t) = \int \mathbb{I}(t > y)\, \varphi(t)\, f(t) dt$$

$$\Gamma_\varphi(Z) = \iint \mathbb{I}(z > s) \cdot \frac{\mathbb{I}(t > s)\, \varphi(t)\, f(t)}{\bar{H}(s)\, \bar{G}(s)} \cdot g(s)\, ds\, dt \;=\; \int \frac{\mathbb{I}(z > s)\, S_\varphi(s)}{\bar{H}(s)\, \bar{G}(s)} \cdot g(s)\, ds, \;\; \bar{H}(s) = \bar{F}(s)\, \bar{G}(s)$$

Term-wise, we have that:

$$\mathbf{A}\text{:} \quad \mathbb{E} \left[\frac{\varphi^2(X)}{\bar{G}(X)} \right] \;=\; \int \varphi^2(x) \cdot \frac{f(x)}{\bar{G}(x)}\, dx$$

$$\mathbf{B}\text{:} \quad \int \bar{F}(y)\, \gamma_\varphi^2(y)\, g(y)\, dy \;=\; \int \bar{F}(y) \cdot \frac{S_\varphi^2(y)}{\bar{H}^2(y)} \cdot g(y)\, dy \;=\; \int \frac{S_\varphi^2(y)}{\bar{F}(y)\, \bar{G}^2(y)} \cdot g(y)\, dy$$

$$\begin{aligned}
\mathbf{C}\text{:} \quad \mathbb{E} \left[\Gamma_\varphi^2(Z) \right] &= \mathbb{E} \left[\iint \mathbb{I}(Z > s) \cdot \frac{S_\varphi(s)}{\bar{H}(s)\, \bar{G}(s)} \cdot \mathbb{I}(Z > w) \cdot \frac{S_\varphi(w)}{\bar{H}(w)\, \bar{G}(w)} \cdot g(s)\, g(w)\, ds\, dw \right] \\
&= \iint \frac{\bar{H}(s \vee w)}{\bar{H}(s)\, \bar{H}(w)} \cdot \frac{S_\varphi(s)}{\bar{G}(s)} \cdot \frac{S_\varphi(w)}{\bar{G}(w)} \cdot g(s)\, g(w)\, ds\, dw \\
&= \iint \frac{1}{\bar{H}(s)} \cdot \frac{S_\varphi(s)}{\bar{G}(s)} \cdot \frac{S_\varphi(w)}{\bar{G}(w)} \cdot g(s)\, g(w)\, ds\, dw \;+ \\
&\quad \iint \frac{1}{\bar{H}(w)} \cdot \frac{S_\varphi(s)}{\bar{G}(s)} \cdot \frac{S_\varphi(w)}{\bar{G}(w)} \cdot g(s)\, g(w)\, ds\, dw
\end{aligned}$$

$$\begin{aligned}
\mathbf{D}\text{:}\quad 2\,\mathbb{E}\left[\varphi(X)\,\Gamma_{\varphi}(X)\right] &= 2\iint \varphi(x)\,\mathbb{I}(x>s)\cdot\frac{S_{\varphi}(s)}{\bar{H}(s)\,\bar{G}(s)}\cdot g(s)\,f(x)\,ds\,dx \\
&= 2\int \frac{S_{\varphi}^{2}(s)}{\bar{H}(s)\,\bar{G}(s)}\cdot g(s)\,ds \\
&= 2\int \frac{S_{\varphi}^{2}(s)}{\bar{F}(s)\,\bar{G}^{2}(s)}\cdot g(s)\,ds
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\text{:}\quad 2\,\mathbb{E}\left[\bar{F}(Y)\,\gamma_{\varphi}(Y)\,\Gamma_{\varphi}(Y)\right] &= 2\iint \bar{F}(y)\cdot\frac{S_{\varphi}(y)}{\bar{H}(y)}\,\mathbb{I}(y>s)\cdot\frac{S_{\varphi}(s)}{\bar{H}(s)\,\bar{G}(s)}\cdot g(y)\,g(s)\,dy\,ds \\
&= 2\iint \frac{S_{\varphi}(y)}{\bar{G}(y)}\cdot\mathbb{I}(y>s)\cdot\frac{S_{\varphi}(s)}{\bar{G}(s)}\cdot g(y)\,g(s)\,dy\,ds
\end{aligned}$$

$$\begin{aligned}
\mathbf{A}+\mathbf{B}+\mathbf{C}-\mathbf{D}-\mathbf{E} &= \mathbb{E}\left(\frac{\varphi^{2}(X)}{\bar{G}(X)}\right)-\mathbb{E}\left(\bar{F}(Y)\gamma_{\varphi}^{2}(Y)\right) \\
&= \int \varphi^{2}(x)\cdot\frac{f(x)}{\bar{G}(x)}\,dx-\int\frac{S_{\varphi}^{2}(y)}{\bar{F}(y)\,\bar{G}^{2}(y)}\cdot g(y)\,dy
\end{aligned}$$

$$S_{\varphi}(y)=\int \mathbb{I}(t>y)\,\frac{d}{d\theta}\ln f(t\mid\theta)\cdot f(t\mid\theta)\,dt \;=\; \int \mathbb{I}(t>y)\,\frac{\frac{d\,f(t|\theta)}{d\theta}}{f(t\mid\theta)}\,f(t\mid\theta)\,dt$$

$$\implies\quad S_{\varphi}(y)=\frac{d}{d\theta}\,\bar{F}(y\mid\theta)$$

$$\therefore\quad \sigma^{2}(\theta)=\int\left(\frac{d}{d\theta}\ln f(x\mid\theta)\right)^{2}\frac{f(x\mid\theta)}{\bar{G}(x)}\;dx-\int\frac{\left(\frac{d}{d\theta}\bar{F}(y\mid\theta)\right)^{2}}{\bar{F}(y\mid\theta)\bar{G}^{2}(y)}\;g(y)dy \tag{2.6}$$

As mentioned in Section 2.2 above, desirable results are obtained if $Var_M(\theta)\leq Var_P(\theta)$. Therefore, we can find conditions on which this is attained by finding parameter values at which

$$\begin{aligned}
&\frac{\sigma^{2}(\theta)}{I^{2}(\theta)} \;\leq\; \frac{1}{I_{RC}(\theta)} \\
\iff\quad &\frac{\sigma^{2}(\theta)}{I(\theta)} \;\leq\; \frac{I(\theta)}{I_{RC}(\theta)} \\
\iff\quad &\frac{\sigma^{2}(\theta)}{I(\theta)} \;\leq\; \frac{1}{I_{RC}(\theta)/I(\theta)}
\end{aligned} \tag{2.7}$$

Note that for uncensored data $\bar{G}(x)=1$ and $g(x)=0$ for all $x\geq 0$, hence $\sigma^{2}(\theta)=I(\theta)=I_{RC}(\theta)$.

$Var_M(\theta) \leq Var_P(\theta)$ resolves to:

$$\frac{1}{I(\theta)} \int \left(\frac{d}{d\theta} \ln f(x \mid \theta) \right)^2 \frac{1}{\bar{G}(x)} f(x \mid \theta)\ dx - \frac{1}{I(\theta)} \int \frac{\left(\frac{d}{d\theta} \bar{F}(y \mid \theta) \right)^2}{\bar{F}(y \mid \theta)} \frac{1}{\bar{G}^2(y)}\ g(y) dy$$

$$\leq \quad \frac{1}{\frac{1}{I(\theta)} \int \left(\frac{d}{d\theta} \ln f(x \mid \theta) \right)^2 \bar{G}(x) f(x \mid \theta)\ dx + \frac{1}{I(\theta)} \int \frac{\left(\frac{d}{d\theta} \bar{F}(y \mid \theta) \right)^2}{\bar{F}(y \mid \theta)}\ g(y) dy}$$

2.3 Exponentially Distributed Event

Generally, we select continuous distributions for both the event variable X and the censoring variable Y such that integrability conditions are satisifed, particularly for the variance of the modified estimator. Therefore, in the case where $X \sim Exp(1/\theta)$, as in other cases we consider, we ensure that $\sigma^2(\theta)$ remains finite by choosing Y such that the survival function $\bar{G}(x)$ has a heavier tail than $f(x \mid \theta)$, and such that the ratio $\frac{f(x \mid \theta)}{\bar{G}(x)}$ remains finite.

We begin by computing the theoretical variances, followed by the estimation of the maximum likelihood estimators (MLEs).

$$\begin{aligned}
f(x \mid \theta) &= \theta e^{-\theta x},\ x > 0,\ \theta > 0 \\
\ln f(x \mid \theta) &= \ln \theta - \theta x \\
\frac{d}{d\theta} \ln f(x \mid \theta) &= \frac{1}{\theta} - x \\
\bar{F}(x \mid \theta) &= e^{-\theta x},\ \theta > 0 \\
\ln \bar{F}(x \mid \theta) &= -\theta x \\
\frac{d}{d\theta} \ln \bar{F}(x \mid \theta) &= -x
\end{aligned}$$

2.3.1 Exponentially Distributed Censoring

Under $f(x \mid \theta) = \theta e^{-\theta x}$, $\theta > 0$, we take $g(y) = \lambda e^{-\lambda y}$, $\lambda > 0$ and $\bar{G}(y) = e^{-\lambda y}$.

Computing all relevant integrals including $\sigma^2(\theta)$, $I_{RC}(\theta)$, and $I^2(\theta)$ from the simplified expressions above, we have

$$
\begin{aligned}
I(\theta) &= -\mathbb{E}\left[\frac{d^2}{d\theta^2}\ln f(X \mid \theta)\right] \\
&= -\mathbb{E}\left[-\frac{1}{\theta^2}\right] = \frac{1}{\theta^2} \\
\implies I^2(\theta) &= \frac{1}{\theta^4}
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2(\theta) &= \theta\int_0^\infty \left(\frac{1}{\theta} - x\right)^2 e^{(\lambda-\theta)x}\, dx - \lambda\int_0^\infty y^2 e^{(\lambda-\theta)y}\, dy \\
&= \frac{1}{\theta}\int_0^\infty e^{-(\theta-\lambda)x}dx - 2\int_0^\infty xe^{-(\theta-\lambda)x}dx + \theta\int_0^\infty x^2e^{-(\theta-\lambda)x}dx - \lambda\int_0^\infty y^2e^{-(\theta-\lambda)y}dy \\
&= \frac{1}{\theta(\theta-\lambda)} - \frac{2}{(\theta-\lambda)^2} + \frac{2\theta}{(\theta-\lambda)^3} - \frac{2\lambda}{(\theta-\lambda)^3} \\
\sigma^2(\theta) &= \frac{1}{\theta(\theta-\lambda)},\ \theta > \lambda
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{\sigma^2(\theta)}{I^2(\theta)} &= \frac{\dfrac{1}{\theta(\theta-\lambda)}}{\dfrac{1}{\theta^4}} \\
\frac{\sigma^2(\theta)}{I^2(\theta)} &= \frac{\theta^3}{(\theta-\lambda)}
\end{aligned}
$$

$$
\begin{aligned}
I_{RC}(\theta) &= \theta\int_0^\infty \left(\frac{1}{\theta} - x\right)^2 e^{-(\theta+\lambda)x}\, dx + \lambda\int_0^\infty y^2 e^{-(\theta+\lambda)y}\, dy \\
&= \frac{1}{\theta}\int_0^\infty e^{-(\theta+\lambda)x}dx - 2\int_0^\infty xe^{-(\theta+\lambda)x}dx + \theta\int_0^\infty x^2e^{-(\theta+\lambda)x}dx - \lambda\int_0^\infty y^2e^{-(\theta+\lambda)y}dy \\
&= \frac{1}{\theta(\theta+\lambda)} - \frac{2}{(\theta+\lambda)^2} + \frac{2\theta}{(\theta+\lambda)^3} - \frac{2\lambda}{(\theta+\lambda)^3} \\
I_{RC}(\theta) &= \frac{1}{\theta(\theta+\lambda)}
\end{aligned}
$$

Notice that as $\lambda \to 0$, $Var_M(\theta) \approx Var_P(\theta)$, as $\sigma^2(\theta) \to I(\theta)$, and $I_{RC}(\theta) \to I(\theta)$. Additionally, we find that Equation (2.7) holds when

$$
\begin{aligned}
\frac{\sigma^2(\theta)}{I^2(\theta)} &\leq \frac{1}{I_{RC}(\theta)} \\
\frac{\theta^3}{(\theta-\lambda)} &\leq \theta(\theta+\lambda)
\end{aligned}
$$

$$
\begin{aligned}
\theta^3 &\leq \theta(\theta+\lambda)(\theta-\lambda) \\
\theta^3 &\leq \theta(\theta^2-\lambda^2) \\
\theta^3 &\leq \theta^3-\theta\lambda^2 \\
\theta\lambda^2 &\leq 0 \\
\lambda &\leq 0
\end{aligned} \tag{2.8}
$$

Since $\lambda > 0$, we conclude that $Var_M(\theta) \not\leq Var_P(\theta)$, and that the modified estimator attains minimal variance as the censoring rate approaches 0. The censoring rate is:

$$
\mathbb{P}(X > Y) = \lambda \int_0^\infty e^{-(\theta+\lambda)y}\, dy \;=\; \lambda \left. \frac{e^{-(\theta+\lambda)y}}{-(\theta+\lambda)} \right|_0^\infty \;=\; \frac{\lambda}{\theta+\lambda}
$$

Computationally, we examine $0 < \theta < 20$, with the condition $\theta > \lambda$ to identify regions where comparable results are observed for $\lambda \in \{0.5, 1, 2, 3\}$. The corresponding results are presented below.

While these graphs support the theoretical results, they also illustrate how close the variances of the estimators are, particularly as $\lambda \to 0$. Similarly, as $\lambda \to \infty$, the difference in variance increases, with a more pronounced gap observed at $\theta \to 0$. The graphs of the MSEs obtained from the simulated data further confirm these results (see Figure 2.3). These observations were made at a decreasing censoring rate.

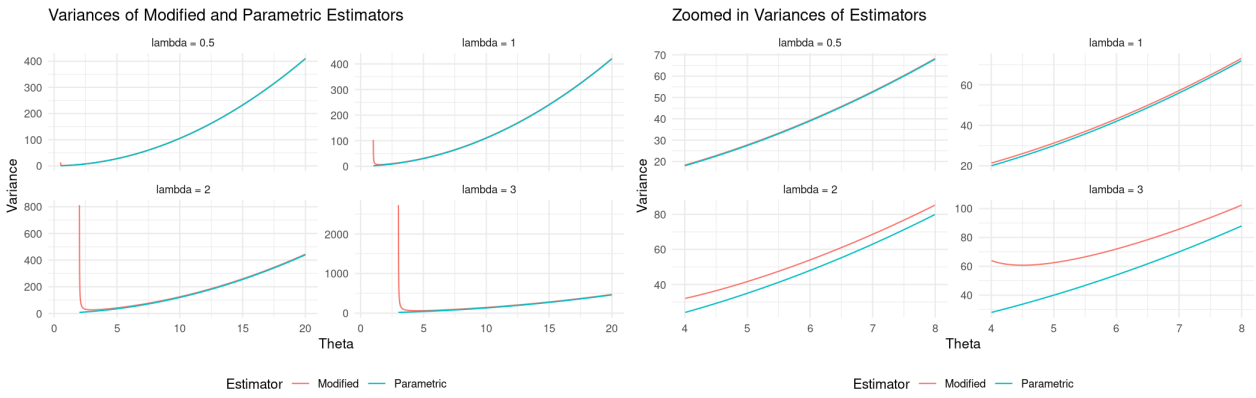

Figure 2.1: Variance of Estimators for $X \sim Exp(1/\theta),\ Y \sim Exp(1/\lambda)$

For all values of λ, the random censoring rate steadily decreases as θ increases, from approximately 0.5 to 0. As $\mathbb{P}(\delta = 0) \to \infty$, the variances of both estimators also approach zero. Throughout this range, the variance of the modified estimator is slightly higher than that of the parametric estimator, as shown in Figure 2.2. Notably, the modified estimator exhibits a sharp increase in variance toward high censoring rates. This behavior corresponds to the wider variance gap observed as $\theta \to 0$ in Figure 2.1.

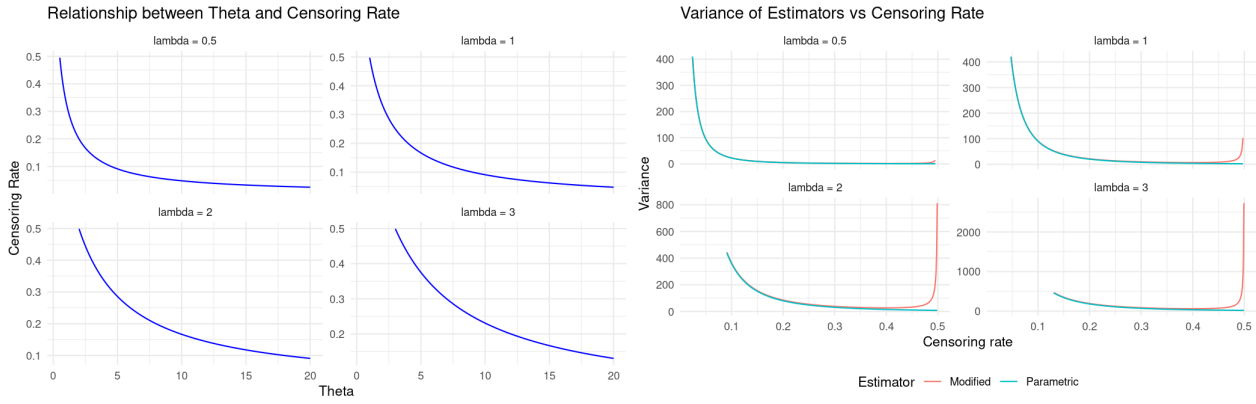

Figure 2.2: Comparison by Censoring Rates for $X \sim Exp(1/\theta),\ Y \sim Exp(1/\lambda)$

Additionally, we simulated data from the respective distributions and computed the MLEs of the estimators as follows:

Parametric MLE

$$\sum_{i=1}^{n}\left[\delta_i\left(\frac{1}{\theta}-z_i\right)+(1-\delta_i)(-z_i)\right]=0$$

$$\sum_{i=1}^{n}\left[\frac{\delta_i}{\theta}-\delta_i z_i-z_i(1-\delta_i)\right]=0$$

$$\frac{1}{\theta}\sum_{i=1}^{n}\delta_i-\sum_{i=1}^{n}z_i=0$$

$$\therefore\quad \hat{\theta}^P=\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}z_i}$$

Semi-parametric MLE

$$\sum_{i=1}^{n}W_{in}\left(\frac{1}{\theta}-z_i\right)=0$$

$$\sum_{i=1}^{n}W_{in}\frac{1}{\theta}-\sum_{i=1}^{n}W_{in}z_i=0$$

$$\frac{1}{\theta}\sum_{i=1}^{n}W_{in}=\sum_{i=1}^{n}W_{in}z_i$$

$$\therefore\quad \hat{\theta}^M=\frac{\sum_{i=1}^{n}W_{in}}{\sum_{i=1}^{n}W_{in}z_i}$$

The mean squared errors of the MLEs were thus computed and compared in Figure 2.3.

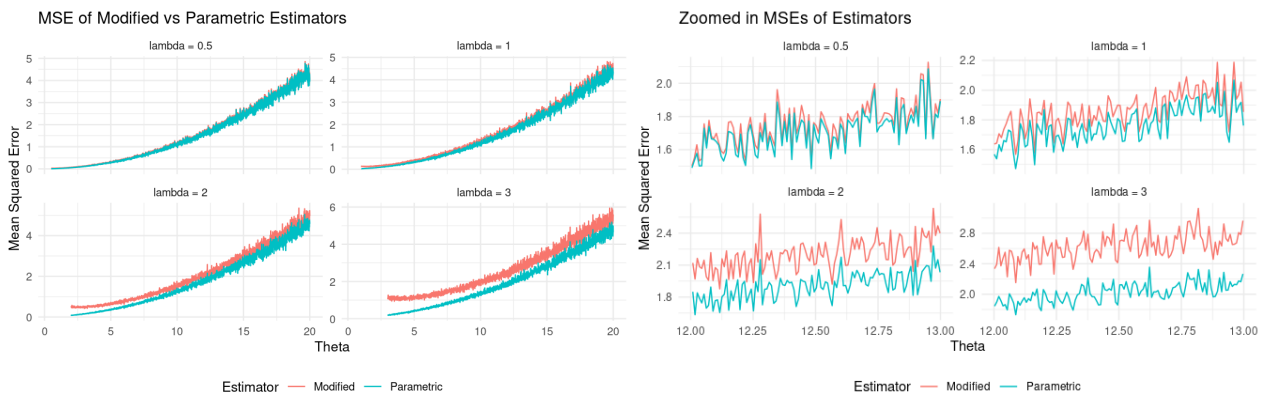

Figure 2.3: Mean Squared Error Comparison when $X \sim Exp(1/\theta),\ Y \sim Exp(1/\lambda)$

Similar to the graphs from the theoretical computations, the two estimators are comparable as $\lambda \to 0$.

2.3.2 Pareto Distributed Censoring

We also consider Y to have a Pareto distribution with shape parameter $= \lambda$ and scale $= 1$, i.e. $g(y) = \dfrac{\lambda}{(1+y)^{\lambda+1}}$, $\lambda > 0$ and $\bar{G}(y) = \dfrac{1}{(1+y)^{\lambda}}$. The theoretical variances are computed as

$$
\begin{aligned}
\sigma^2(\theta) &= \int_0^\infty \left(\frac{1}{\theta} - x\right)^2 \theta e^{-\theta x}(1+x)^{\lambda}\, dx - \int_0^\infty y^2 e^{-\theta y}\ \lambda(1+y)^{\lambda-1}\, dy \\
&= \theta\left(\frac{1}{\theta^3} + \frac{\lambda}{\theta^3}\int_0^\infty e^{-\theta x}(1+x)^{\lambda-1}dx + \frac{\lambda}{\theta}\int_0^\infty x^2 e^{-\theta x}(1+x)^{\lambda-1}dx\right) - \\
&\quad \lambda\int_0^\infty y^2 e^{-\theta y}(1+y)^{\lambda-1}dy \\
&= \frac{1}{\theta^2}\left(1 + \lambda\int_0^\infty e^{-\theta x}(1+x)^{\lambda-1}dx\right), \quad \text{let } u = 1+x,\ x = u-1,\ du = dx \\
&= \frac{1}{\theta^2}\left(1 + \lambda\int_1^\infty e^{-\theta(u-1)}u^{\lambda-1}du\right) = \frac{1}{\theta^2}\left(1 + \lambda e^{\theta}\int_1^\infty e^{-\theta u}u^{\lambda-1}du\right) \quad \text{let } t = \theta u,\ dt = \theta du \\
&= \frac{1}{\theta^2}\left(1 + \lambda e^{\theta}\int_\theta^\infty e^{-t}\cdot\frac{t}{\theta}^{\lambda-1}\frac{dt}{\theta}\right) = \frac{1}{\theta^2}\left(1 + \frac{\lambda e^{\theta}}{\theta^{\lambda}}\int_\theta^\infty e^{-t}\cdot t^{\lambda-1}dt\right) \\
\sigma^2(\theta) &= \frac{1}{\theta^2}\left(1 + \frac{\lambda e^{\theta}}{\theta^{\lambda}}\,\Gamma(\lambda,\theta)\right), \text{ where } \Gamma(\lambda,\theta) \text{ is the upper incomplete gamma function.}
\end{aligned}
$$

i.e.

$$\Gamma(\lambda,\theta) = \int_\theta^\infty t^{\lambda-1}e^{-t}dt\,.$$

For the modified estimator,

$$\frac{\sigma^2(\theta)}{I^2(\theta)} = \frac{\frac{1}{\theta^2}\left(1+\frac{\lambda e^{\theta}}{\theta^{\lambda}}\,\Gamma(\lambda,\theta)\right)}{\frac{1}{\theta^4}} = \theta^2\left(1+\frac{\lambda e^{\theta}}{\theta^{\lambda}}\,\Gamma(\lambda,\theta)\right)$$

For the parametric estimator, the theoretical variance is the reciprocal of $I_{RC}(\theta)$:

$$\begin{aligned} I_{RC}(\theta) &= \int_0^{\infty}\left(\frac{1}{\theta}-x\right)^2\frac{\theta e^{-\theta x}}{(1+x)^{\lambda}}\,dx + \int_0^{\infty} y^2 e^{-\theta y}\frac{\lambda}{(1+y)^{\lambda+1}}\,dy \\ &= \frac{1}{\theta^2} - \frac{\lambda}{\theta^2}\int_0^{\infty} e^{-\theta x}(1+x)^{-(\lambda+1)}dx \end{aligned}$$

with random censoring rate,

$$\mathbb{P}(X>Y) = \int_0^{\infty} e^{-\theta y}\,\frac{\lambda}{(1+y)^{\lambda+1}}\,dy = \lambda\int_0^{\infty} e^{-\theta y}(1+y)^{-(\lambda+1)}dy$$

These integrals are solved computationally using R. The corresponding results from this case are included below.

We find that the theoretical variances of the two estimators are fairly similar, especially as $\lambda \to 0$. However, we do not find intervals of θ where $Var_M(\theta) \leq Var_P(\theta)$. Additionally, we noticed that $\mathbb{P}(\delta = 0) \to 0$ as $\theta \to \infty$ and as $\mathbb{P}(\delta = 0) \to \infty$, variances also approach 0; Figure 2.5 shows that the variance of the modified estimator is only marginally higher than that of the parametric estimator.

Figure 2.4: Variance of Estimators for $X \sim Exp(1/\theta),\ Y \sim Pareto(\lambda, 1)$

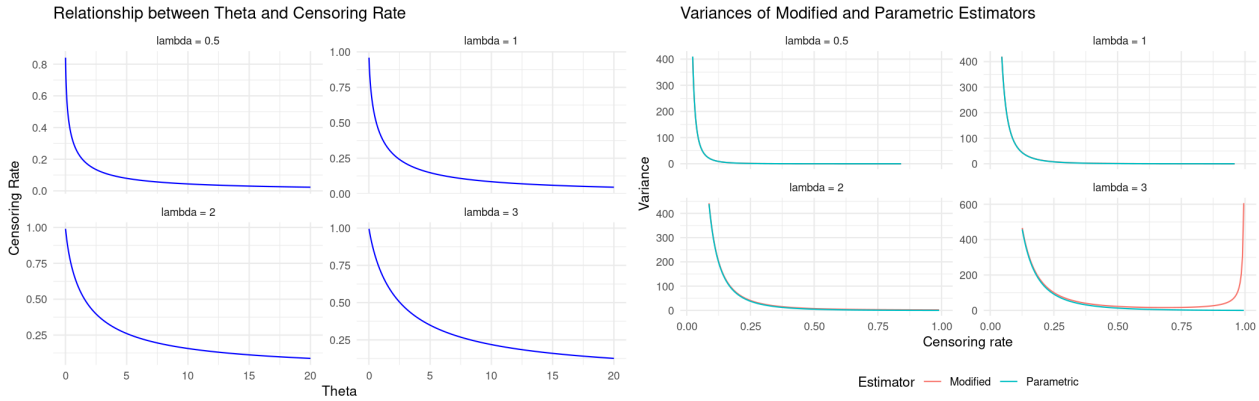

Figure 2.5: Comparison by Censoring Rates for $X \sim Exp(1/\theta),\ Y \sim Pareto(\lambda, 1)$

The comparability of the estimators as $\lambda \to 0$, as shown in Figure 2.4, is confirmed by the simulated data results with the MLEs (Figure 2.6).

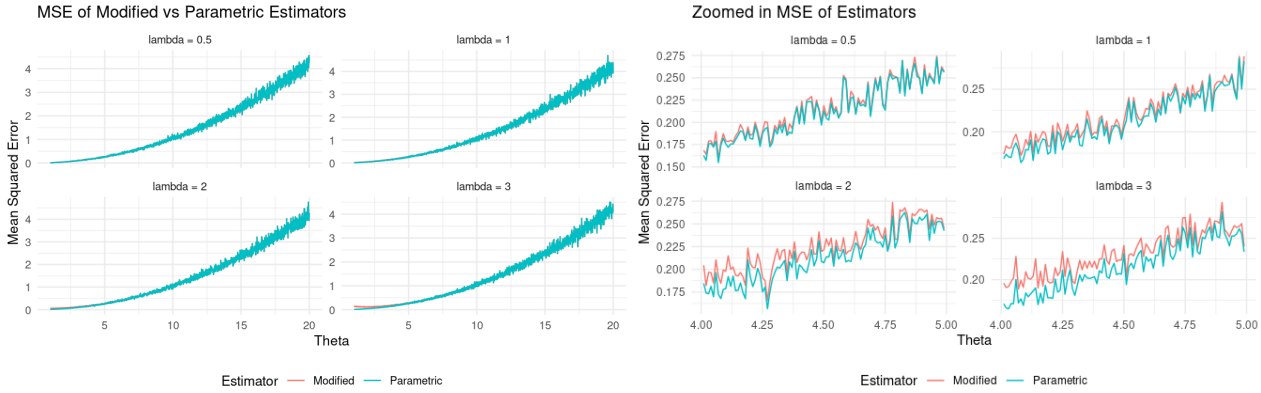

Figure 2.6: Mean Squared Error Comparison when $X \sim Exp(1/\theta),\ Y \sim Pareto(\lambda, 1)$

Generally for $X \sim Exponential(1/\theta)$, we saw that $Var_M(\theta) \approx Var_P(\theta)$ for small λ, and despite the gap in variance at $\theta \to 0$, the estimators behave more closely as $\theta \to \infty$.

2.4 Weibull Distributed Event

We also consider the event variable to have a Weibull distribution with scale $= \theta$, and shape $= 2$. Some distributions of Y include $Weibull(\theta, 2)$, and $Pareto(\lambda, 1)$. Let

$$f(x \mid \theta) \quad = \quad 2\theta x e^{-\theta x^2},\ x > 0,\ \theta > 0$$

$$
\begin{aligned}
\ln f(x \mid \theta) &= \ln 2 + \ln \theta + \ln x - \theta x^2 \\
\frac{d}{d\theta} \ln f(x \mid \theta) &= \frac{1}{\theta} - x^2 \\
\bar{F}(x \mid \theta) &= e^{-\theta x^2},\ x > 0 \\
\ln \bar{F}(x \mid \theta) &= -\theta x^2 \\
\frac{d}{d\theta} \ln \bar{F}(x \mid \theta) &= -x^2
\end{aligned}
$$

2.4.1 Weibull Distributed Censoring

Under $f(x \mid \theta) = 2\theta x e^{-\theta x^2}$, $\theta > 0$, we consider $g(y) = 2\lambda y e^{-\lambda y^2}$, $\lambda > 0$ and $\bar{G}(y) = e^{-\lambda y^2}$, which is also the specific case of the shape parameter, $p = 2$. However, to show that $\forall p$ the variances do not change, we will solve the theoretical variances for a general p. These results support the findings for $X \sim Exp(1/\theta),\ Y \sim Exp(1/\theta)$, which is the case when $p = 1$. Let $f(x \mid \theta) = p\theta x^{p-1} e^{-\theta x^p}$ and $g(y) = p\lambda y^{p-1} e^{-\lambda y^p}$.

$$
\begin{aligned}
\sigma^2(\theta) &= p\theta \int_0^\infty \left(\frac{1}{\theta} - x^p\right)^2 x^{p-1} e^{(\lambda-\theta)x^p}\ dx - p\lambda \int_0^\infty y^{3p-1} e^{(\lambda-\theta)y^p}\ dy \\
&= p\theta \left(\frac{1}{\theta^2} \int_0^\infty x^{p-1} e^{-(\theta-\lambda)x^p}\ dx - \frac{2}{\theta} \int_0^\infty x^{2p-1} e^{-(\theta-\lambda)x^p} + \int_0^\infty x^{3p-1} e^{-(\theta-\lambda)x^p}\right) - \\
&\quad p\lambda \int_0^\infty y^{3p-1} e^{-(\theta-\lambda)y^p}\ dy \\
&= p\theta \left(\frac{\theta-\lambda}{\theta^2} \int_0^\infty x^{2p-1} e^{-(\theta-\lambda)x^p} dx - \frac{2}{\theta} \int_0^\infty x^{2p-1} e^{-(\theta-\lambda)x^p} dx + \int_0^\infty x^{3p-1} e^{-(\theta-\lambda)x^p}\right) - \\
&\quad p\lambda \int_0^\infty y^{3p-1} e^{-(\theta-\lambda)y^p}\ dy \\
&= p\theta \left(-\frac{(\theta+\lambda)(\theta-\lambda)}{2\theta^2} \int_0^\infty x^{3p-1} e^{-(\theta-\lambda)x^p}\ dx + \int_0^\infty x^{3p-1} e^{-(\theta-\lambda)x^p}\right) - \\
&\quad p\lambda \int_0^\infty y^{3p-1} e^{-(\theta-\lambda)y^p}\ dy \\
&= \frac{p\theta\left(\theta^2+\lambda^2\right)}{2\theta^2} \int_0^\infty x^{3p-1} e^{-(\theta-\lambda)x^p} - p\lambda \int_0^\infty y^{3p-1} e^{-(\theta-\lambda)y^p}\ dy \\
&= \frac{p(\theta-\lambda)^2}{2\theta} \int_0^\infty x^{3p-1} e^{-(\theta-\lambda)x^p}\ dx \\
&\quad \text{Let } u = (\theta-\lambda)x^p,\ x = \left(\frac{u}{\theta-\lambda}\right)^{\frac{1}{p}},\ du = p(\theta-\lambda)x^{p-1}dx \\
&= \frac{p(\theta-\lambda)^2}{2\theta} \int_0^\infty \left(\frac{u}{\theta-\lambda}\right)^{\frac{3p-1}{p}} \frac{e^{-u}\ du}{p(\theta-\lambda)^{\frac{1}{p}}\ u^{\frac{p-1}{p}}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{p(\theta-\lambda)^2}{2\theta}\cdot\frac{1}{(\theta-\lambda)^{\frac{3p-1}{p}}}\cdot\frac{1}{p(\theta-\lambda)}\int_0^\infty u^{\frac{3p-1}{p}}\cdot\frac{1}{u^{\frac{p-1}{p}}}\ e^{-u}\ du \\
&= \frac{p(\theta-\lambda)^2}{2\theta}\cdot\frac{1}{p(\theta-\lambda)^{\frac{1}{p}+\frac{3p-1}{p}}}\int_0^\infty u^{(\frac{3p-1}{p}-\frac{p-1}{p})}\ e^{-u}\ du = \frac{p(\theta-\lambda)^2}{2\theta\cdot p(\theta-\lambda)^3}\int_0^\infty u^2\ e^{-u}\ du \\
&= \frac{\Gamma(3)}{2\theta\ (\theta-\lambda)} = \frac{1}{\theta(\theta-\lambda)} \\
\sigma^2(\theta) &= \frac{1}{\theta(\theta-\lambda)},\ \theta>\lambda
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{\sigma^2(\theta)}{I^2(\theta)} &= \frac{\frac{1}{\theta(\theta-\lambda)}}{\frac{1}{\theta^4}} \\
\frac{\sigma^2(\theta)}{I^2(\theta)} &= \frac{\theta^3}{(\theta-\lambda)}
\end{aligned}
$$

$$
\begin{aligned}
I_{RC}(\theta) &= p\theta\int_0^\infty\left(\frac{1}{\theta}-x^p\right)^2 x^{p-1}e^{-(\theta+\lambda)x^p}\ dx + p\lambda\int_0^\infty y^{3p-1}e^{-(\theta+\lambda)y^p}\ dy \\
&= p\theta\cdot\frac{\theta^2+\lambda^2}{2\theta^2}\int_0^\infty x^{3p-1}e^{-(\theta+\lambda)x^p}dx + p\lambda\int_0^\infty y^{3p-1}e^{-(\theta+\lambda)y^p}dy \\
&= \left(\frac{p\ (\theta^2+\lambda^2)}{2\theta}+p\lambda\right)\int_0^\infty x^{3p-1}e^{-(\theta+\lambda)x^p}dx \\
&= \frac{p(\theta+\lambda)^2}{2\theta}\int_0^\infty x^{3p-1}e^{-(\theta+\lambda)x^p}dx \\
&= \frac{p(\theta+\lambda)^2}{2\theta}\cdot\frac{1}{p(\theta+\lambda)^{\frac{1}{p}}(\theta+\lambda)^{\frac{3p-1}{p}}}\int_0^\infty u^2e^{-u}\ du \\
&= \frac{p(\theta+\lambda)^2}{2\theta}\cdot\frac{\Gamma(3)}{p(\theta+\lambda)^3} \\
&= \frac{p(\theta+\lambda)^2}{2\theta}\cdot\frac{2}{p(\theta+\lambda)^3} \\
I_{RC}(\theta) &= \frac{1}{\theta(\theta+\lambda)} \\
\implies I_{RC}^{-1}(\theta) &= \theta(\theta+\lambda)
\end{aligned}
$$

Since $\lambda > 0$, we conclude that $Var_M(\theta) \not\leq Var_P(\theta)$, see breakdown in Equation 2.8. However, we remark that comparable results are obtained from the computational analysis. We examine $0 < \theta < 20$, $\theta > \lambda$ for $\lambda \in \{0.5, 1, 2, 3\}$.

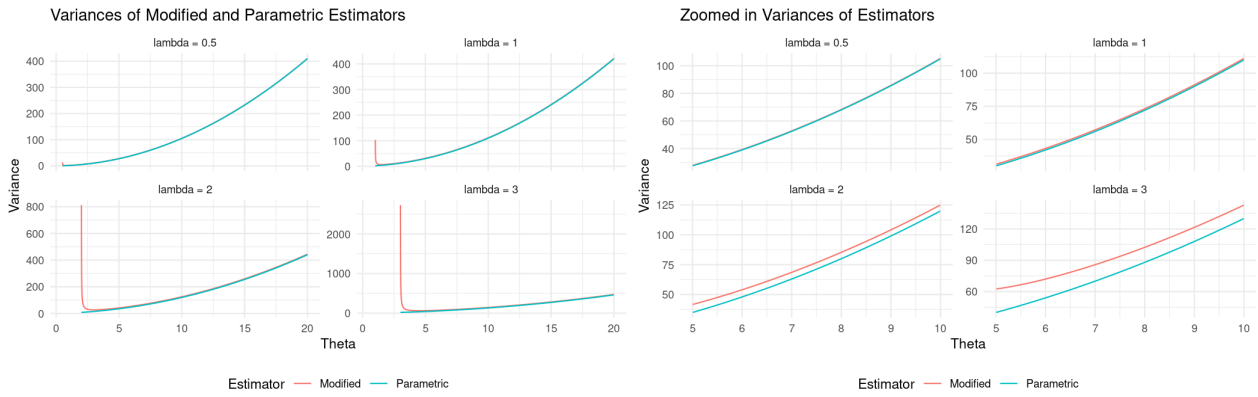

Figure 2.7: Variance of Estimators for $X \sim Weibull(\theta, 2),\ Y \sim Weibull(\lambda, 2)$

The same observations made for $X \sim Exp(1/\theta)$, $Y \sim Exp(1/\theta)$ are replicated here with random censoring rate

$$\mathbb{P}(X > Y) = p\lambda \int_0^\infty y^{p-1} e^{-(\theta+\lambda)y^p} dy \ = \ p\lambda \ \frac{1}{p(\theta+\lambda)^{\frac{p-1}{p}+\frac{1}{p}}} \int_0^\infty e^{-u}\, du \ = \ \frac{\lambda}{\theta+\lambda}$$

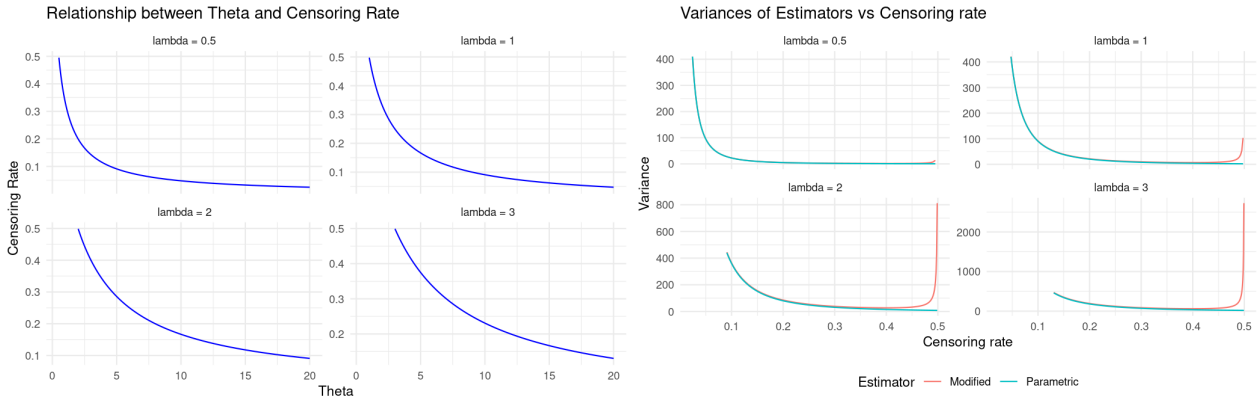

Figure 2.8: Comparison with Censoring Rates for $X \sim Weibull(\theta, 2),\ Y \sim Weibull(\lambda, 2)$

Censoring rates increase with decreasing variance, with variance of the modified estimator being slighly greater than the parametric estimator. The MLEs used to compute the mean squared errors from simulated data are outlined as follows:

Parametric MLE	**Semi-parametric MLE**
$\sum_{i=1}^{n}\left[\delta_i\left(\frac{1}{\theta}-z_i^2\right)+(1-\delta_i)(-z_i^2)\right]=0$	$\sum_{i=1}^{n}W_{in}\left(\frac{1}{\theta}-z_i^2\right)=0$
$\sum_{i=1}^{n}\left[\frac{\delta_i}{\theta}-\delta_i z_i^2-z_i^2(1-\delta_i)\right]=0$	$\sum_{i=1}^{n}W_{in}\frac{1}{\theta}-\sum_{i=1}^{n}W_{in}z_i^2=0$
$\frac{1}{\theta}\sum_{i=1}^{n}\delta_i-\sum_{i=1}^{n}z_i^2=0$	$\frac{1}{\theta}\sum_{i=1}^{n}W_{in}=\sum_{i=1}^{n}W_{in}z_i^2$
$\therefore\quad \hat{\theta}^P=\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}z_i^2}$	$\therefore\quad \hat{\theta}^M=\frac{\sum_{i=1}^{n}W_{in}}{\sum_{i=1}^{n}W_{in}z_i^2}$

See Balakrishnan and Kateri (2008). The mean squared errors of the MLEs were thus computed and compared in Figure 2.9.

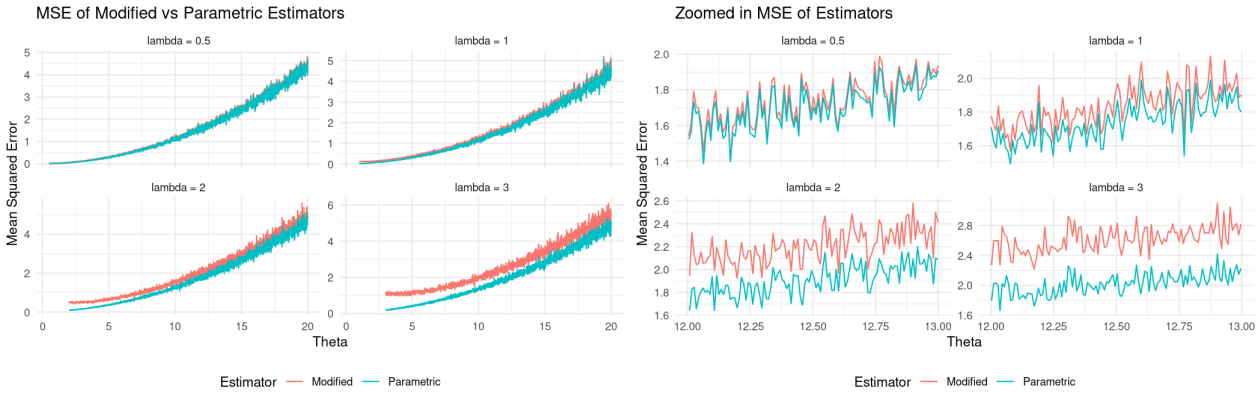

Figure 2.9: Mean Squared Error Comparison when $X \sim Weibull(\theta, 2),\ Y \sim Weibull(\lambda, 2)$

The estimators are comparable as $\lambda \to 0$. These results confirm the theoretical computations and Figure 2.7.

2.4.2 Pareto Distributed Censoring

Considering $Y \sim Pareto(\lambda, 1),\ g(y) = \frac{\lambda}{(1+y)^{\lambda+1}}$ and $\bar{G}(y) = \frac{1}{(1+y)^{\lambda}}$.

The relevant integrals are evaluated computationally. The outcomes obtained are included below.

$$
\begin{aligned}
\sigma^2(\theta) &= 2\theta \int_0^\infty \left(\frac{1}{\theta} - x^2\right)^2 x e^{-\theta x^2} (1+x)^\lambda \, dx - \lambda \int_0^\infty y^4 e^{-\theta y^2} \, (1+y)^{\lambda-1} \, dy \\
I_{RC}(\theta) &= 2\theta \int_0^\infty \left(\frac{1}{\theta} - x^2\right)^2 \frac{x e^{-\theta x^2}}{(1+x)^\lambda} \, dx + \lambda \int_0^\infty \frac{y^4 e^{-\theta y^2}}{(1+y)^{\lambda+1}} \, dy \\
\mathbb{P}[X > Y] &= \int_0^\infty e^{-\theta y^2} \, \frac{\lambda}{(1+y)^{\lambda+1}} \, dy
\end{aligned}
$$

Whilst $Var_M(\theta) \nleq Var_P(\theta)$, we observe that $Var_M(\theta) \approx Var_P(\theta)$ particularly as $\lambda \to 0$, as shown in Figure 2.10. As $\lambda \to \infty$, the variance gap between the two estimators widen, with the modified estimator's variance being marginally higher. With respect to the random censoring rate, little to no difference in variance is detected for small λ. Furthermore, $\mathbb{P}(\delta = 0) \to 0$ as $\theta \to \infty$, and as $\mathbb{P}(\delta = 0) \to \infty$, the variances of both estimators approach zero as well (see Figure 2.11).

Figure 2.10: Variance of Estimators for $X \sim Weibull(\theta, 2),\ Y \sim Pareto(\lambda, 1)$

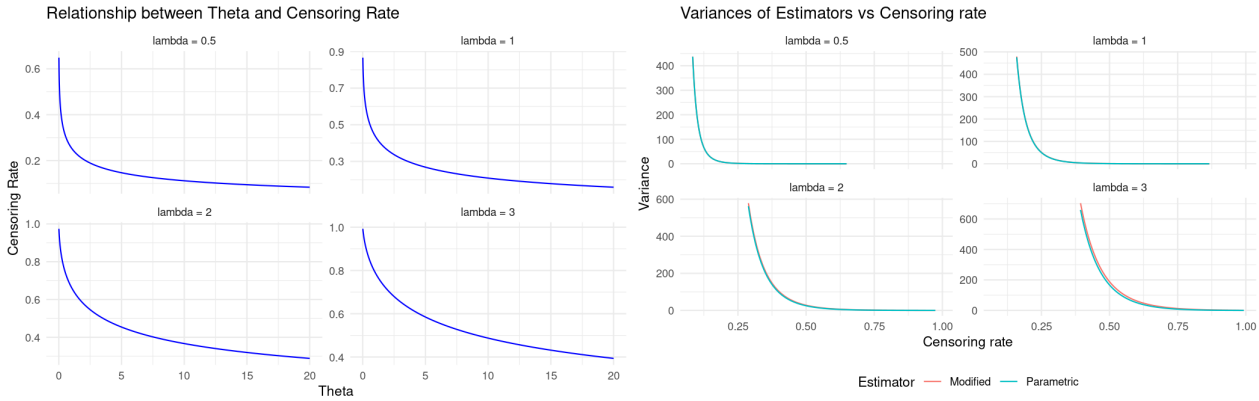

Figure 2.11: Comparison by Censoring Rates for $X \sim Weibull(\theta, 2),\ Y \sim Pareto(\lambda, 1)$

The difference in the MSEs of the estimators is negligible for small values of λ. Contrary to the case where $X \sim Exponential(1/\theta)$ and $Y \sim Pareto(\lambda, 1)$, as $\theta \to \infty$, very little difference is observed

in the MSEs. These results (Figure 2.12) further confirm the comparability of the estimators as $\lambda \to 0$, consistent with the observations in Figure 2.10.

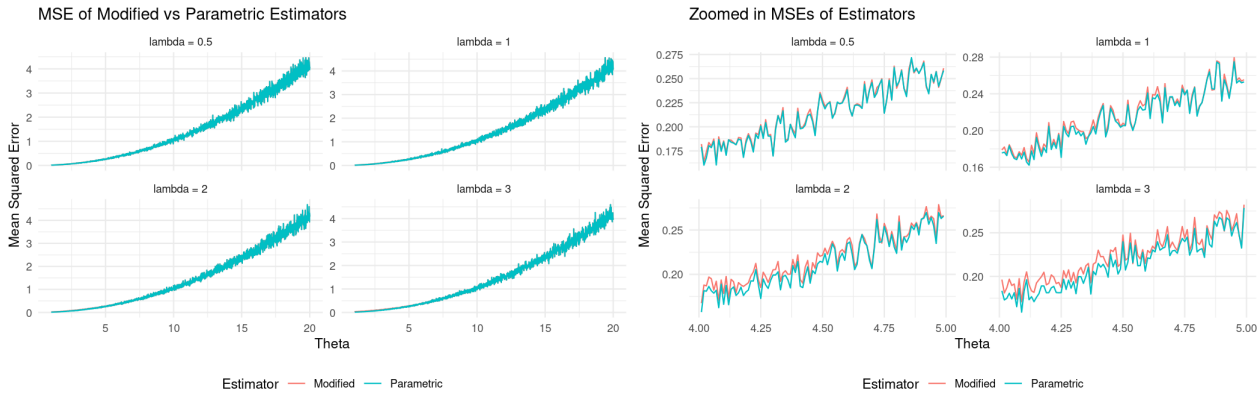

Figure 2.12: Mean Squared Error Comparison when $X \sim Weibull(\theta, 2),\ Y \sim Pareto(\lambda, 1)$

For $X \sim Weibull(\theta, p = 2)$ and $Y \sim Weibull(\theta, p = 2)$, we observed a strong similarity with the case $X \sim Exponential(1/\theta)$ and $Y \sim Exponential(1/\theta)$, as the latter corresponds to the special case where $p = 1$. Consequently, similar patterns and observations were found in both scenarios. More desirable results were obtained when $X \sim Weibull(\theta, p = 2)$ and $Y \sim Pareto(\lambda, 1)$, compared to the combination $X \sim Exponential(1/\theta)$ and $Y \sim Pareto(\lambda, 1)$. In this case, comparability between the estimators was evident even at $\lambda = 1$.

2.5 Beta Distributed Event

Fixing one shape parameter, we consider $X \sim Beta(\alpha = \theta, \beta = 1)$ which allows us to explore several choices for the distribution of Y specifically at $0 \leq y \leq 1$. Let

$$\begin{aligned}
f(x \mid \theta) &= \theta x^{\theta-1},\ 0 \leq x \leq 1,\ \theta > 0 \\
\ln f(x \mid \theta) &= \ln \theta + (\theta - 1) \ln x \\
\frac{d}{d\theta} \ln f(x \mid \theta) &= \frac{1}{\theta} + \ln x \\
\bar{F}(x \mid \theta) &= 1 - x^{\theta},\ 0 \leq x \leq 1 \\
\ln \bar{F}(x \mid \theta) &= \ln(1 - x^{\theta}) \\
\frac{d}{d\theta} \ln \bar{F}(x \mid \theta) &= \frac{-x^{\theta} \ln x}{1 - x^{\theta}}
\end{aligned}$$

2.5.1 Exponentially Distributed Censoring

Under $f(x \mid \theta) = \theta x^{\theta-1}$, $0 \leq x \leq 1$, we take $g(y) = \lambda e^{-\lambda y}, \lambda > 0$ and $\bar{G}(y) = e^{-\lambda y}$.

$$\begin{aligned}
\sigma^2(\theta) &= \theta \int_0^1 \left(\frac{1}{\theta} + \ln x\right)^2 x^{\theta-1} e^{\lambda x}\, dx - \lambda \int_0^1 \frac{y^{2\theta} \ln^2 y \cdot e^{\lambda y}}{1 - y^\theta}\, dy \\
I_{RC}(\theta) &= \theta \int_0^1 \left(\frac{1}{\theta} + \ln x\right)^2 x^{\theta-1} e^{-\lambda x}\, dx + \lambda \int_0^1 \frac{y^{2\theta} \ln^2 y \cdot e^{-\lambda y}}{1 - y^\theta}\, dy \\
I(\theta) &= \frac{1}{\theta^2}, \implies I^2(\theta) = \frac{1}{\theta^4} \\
\mathbb{P}[X > Y] &= \lambda \int_0^1 (1 - y^\theta) e^{-\lambda y}\, dy
\end{aligned}$$

These integrals are evaluated computationally.

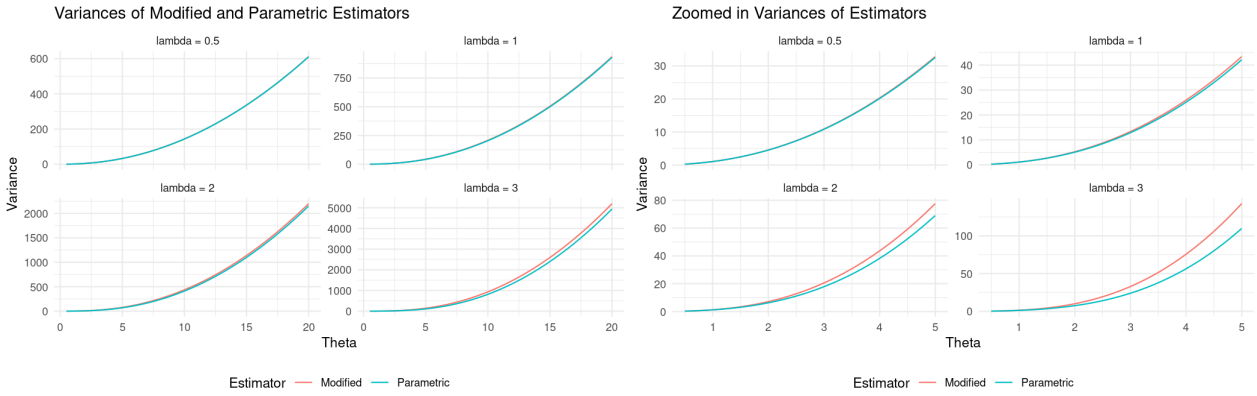

Figure 2.13: Variance of Estimators for $X \sim Beta(\theta, 1),\ Y \sim Exp(1/\lambda)$

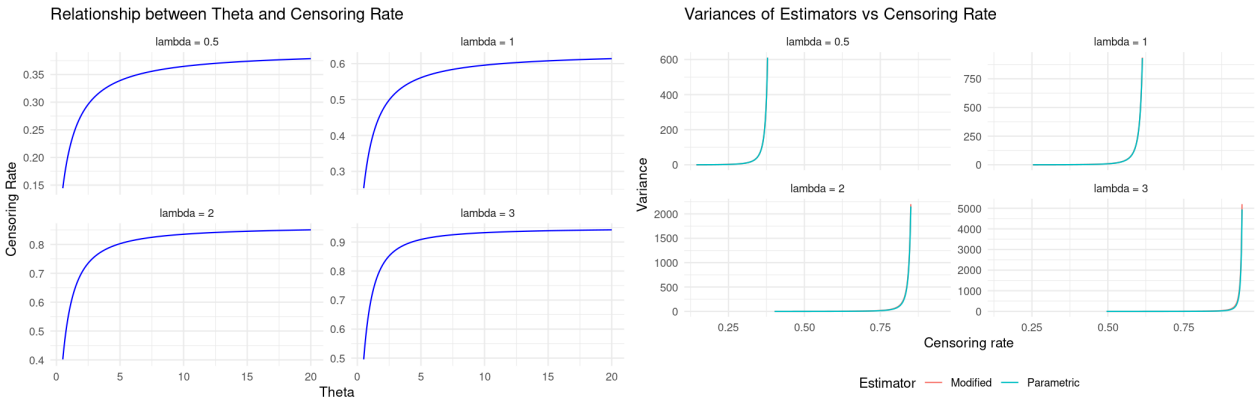

Figure 2.14: Comparison by Censoring Rates for $X \sim Beta(\theta, 1),\ Y \sim Exp(1/\lambda)$

The modified and parametric estimators have similar variances especially when $\lambda \to 0$. In contrast to the cases where X follows an exponential or Weibull distribution, $\mathbb{P}(\delta = 0)$ exhibits a positive

relationship with θ. Likewise, the variances increase with increasing censoring rates. When $\lambda \leq 1$, there is no significant difference between the variances as $\mathbb{P}(\delta = 0) \to \infty$. However, for larger values of λ, the modified estimator exhibits moderately higher variances.

Computing the MLE of $X \sim Beta(\theta, 1)$ requires numerical methods, especially for the parametric estimator owing to the form of $\dfrac{d}{d\theta} \ln \bar{F}(x \mid \theta)$.

Parametric MLE	**Semi-parametric MLE**
$\sum_{i=1}^{n} \left(\delta_i \frac{d}{d\theta} \ln f(z_i \mid \theta) + (1 - \delta_i) \frac{d}{d\theta} \ln \bar{F}(z_i \mid \theta) \right) = 0$	$\sum_{i=1}^{n} W_{in} \left(\frac{1}{\theta} + \ln z_i \right) = 0$
$\sum_{i=1}^{n} \left[\delta_i \left(\frac{1}{\theta} + \ln z_i \right) + (1 - \delta_i) \left(\frac{-z_i^\theta \ln z_i}{1 - z_i^\theta} \right) \right] = 0$	$\sum_{i=1}^{n} W_{in} \frac{1}{\theta} + \sum_{i=1}^{n} W_{in} \ln z_i = 0$
$\sum_{i=1}^{n} \left(\frac{\delta_i}{\theta} + \delta_i \ln z_i - \frac{z_i^\theta \ln z_i}{1 - z_i^\theta} + \frac{\delta_i z_i^\theta \ln z_i}{1 - z_i^\theta} \right) = 0$	$\frac{1}{\theta} \sum_{i=1}^{n} W_{in} = -\sum_{i=1}^{n} W_{in} \ln z_i$
No explicit solution exists for $\hat{\theta}^P$.	$\therefore \ \hat{\theta}^M = -\frac{\sum_{i=1}^{n} W_{in}}{\sum_{i=1}^{n} W_{in} \ln z_i}$

The inbuilt R `uniroot` function was used to find the roots of the score function where explicit expressions were not found. On examining the mean squared errors, the estimators were highly comparable as $\lambda \to 0$. See Figure 2.15.

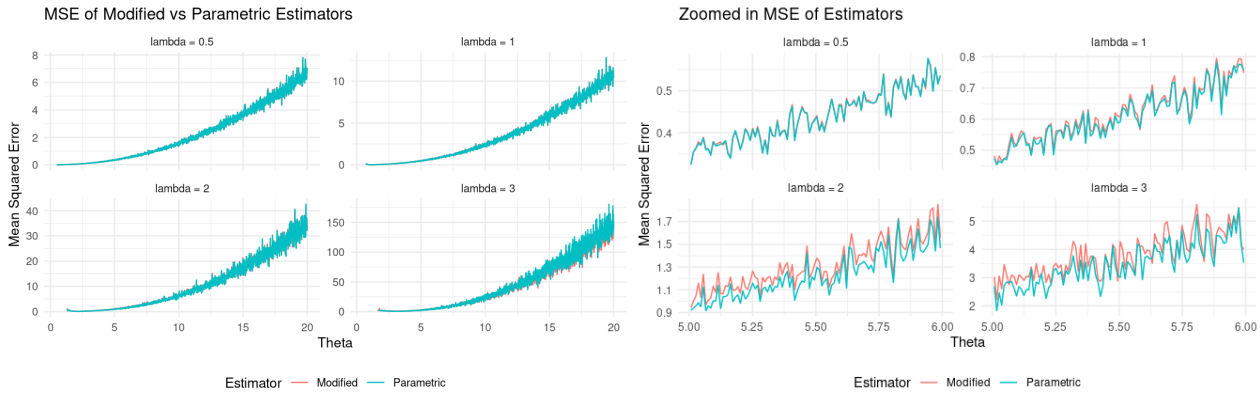

Figure 2.15: Mean Squared Error Comparison when for $X \sim Beta(\theta, 1),\ Y \sim Exp(1/\lambda)$

2.5.2 Weibull Distributed Censoring

Taking $g(y) = 2\lambda y e^{-\lambda y^2}$ and $\bar{G}(y) = e^{-\lambda y^2}$:

$$\begin{aligned}
\sigma^2(\theta) &= \theta \int_0^1 \left(\frac{1}{\theta} + \ln x\right)^2 x^{\theta-1} e^{\lambda x^2}\, dx - 2\lambda \int_0^1 \frac{y^{2\theta+1} \ln^2 y \cdot e^{\lambda y^2}}{1 - y^\theta}\, dy \\
I_{RC}(\theta) &= \theta \int_0^1 \left(\frac{1}{\theta} + \ln x\right)^2 x^{\theta-1} e^{-\lambda x^2}\, dx + 2\lambda \int_0^1 \frac{y^{2\theta+1} \ln^2 y \cdot e^{-\lambda y^2}}{1 - y^\theta}\, dy \\
\mathbb{P}[X > Y] &= 2\lambda \int_0^1 (1 - y^\theta) y\, e^{-\lambda y^2}\, dy
\end{aligned}$$

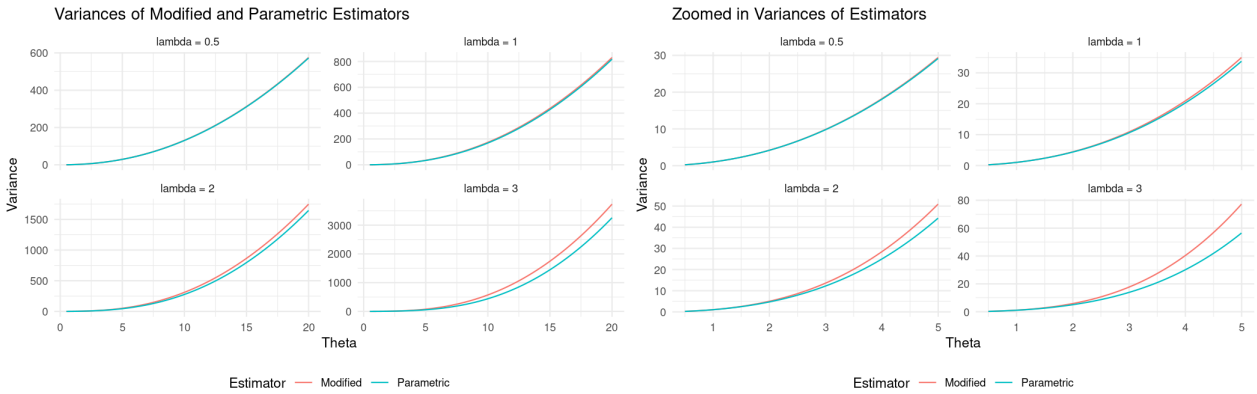

Figure 2.16: Variance of Estimators for $X \sim Beta(\theta, 1),\ Y \sim Weibull(\lambda, 2)$

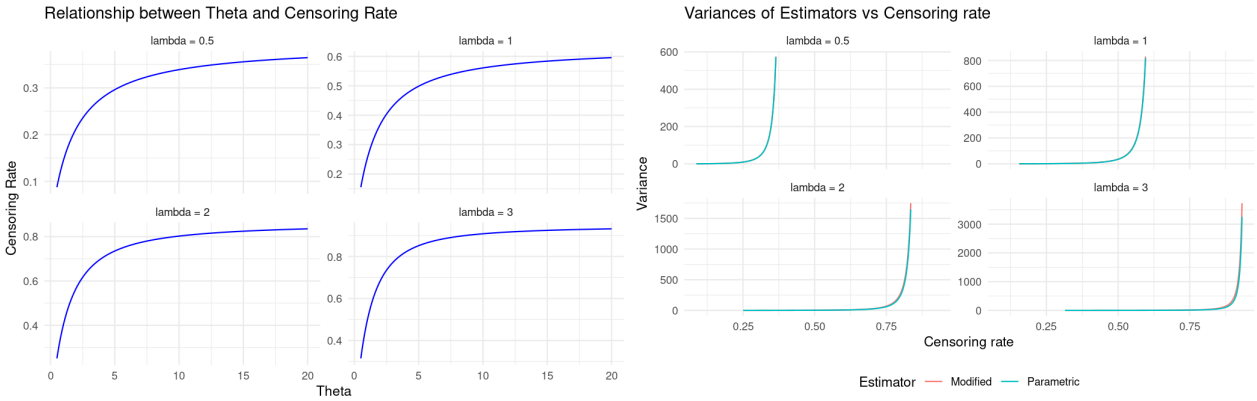

Figure 2.17: Comparison by Censoring Rates for $X \sim Beta(\theta, 1),\ Y \sim Weibull(\lambda, 2)$

Figures 2.16 – 2.18 display similar results to the previous scenarios, where the estimators are comparable at small values of λ, and random censoring rate increases with increasing θ. The MSEs further corroborate these findings.

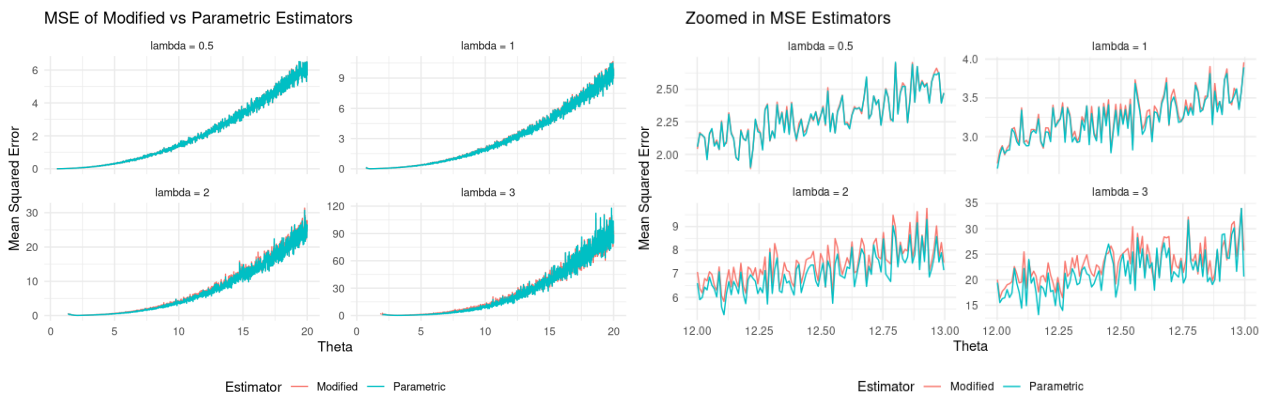

Figure 2.18: Mean Squared Error Comparison when $X \sim Beta(\theta, 1),\ Y \sim Weibull(\lambda, 2)$

2.5.3 Pareto Distributed Censoring

Finally for $X \sim Beta(\theta, 1)$, we let $g(y) = \dfrac{\lambda}{(1+y)^{\lambda+1}}$ and $\bar{G}(y) = \dfrac{1}{(1+y)^{\lambda}}$.

$$\begin{aligned}
\sigma^2(\theta) &= \theta \int_0^1 \left(\frac{1}{\theta} + \ln x\right)^2 x^{\theta-1}(1+x)^{\lambda}\, dx - \lambda \int_0^1 \frac{y^{2\theta} \ln^2 y}{1-y^{\theta}}\, (1+y)^{\lambda-1}\, dy \\
I_{RC}(\theta) &= \theta \int_0^1 \left(\frac{1}{\theta} + \ln x\right)^2 \frac{x^{\theta-1}}{(1+x)^{\lambda}}\, dx + \lambda \int_0^1 \frac{y^{2\theta} \ln^2 y}{(1-y^{\theta})(1+y)^{\lambda+1}}\, dy \\
\mathbb{P}[X > Y] &= \lambda \int_0^1 \frac{(1-y^{\theta})}{(1+y)^{\lambda+1}}\, dy
\end{aligned}$$

See the results from this distribution appended below.

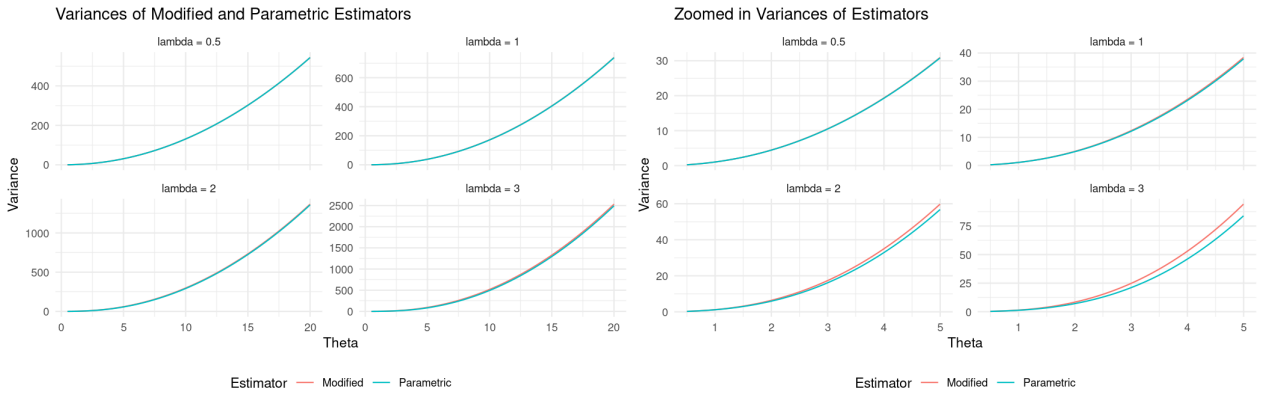

Figure 2.19: Variance of Estimators for $X \sim Beta(\theta, 1),\ Y \sim Pareto(\lambda, 1)$

This case yields the most desirable results for the modified estimator so far. No significant differences are noticed between the estimators, except at large values of θ, where the modified estimator

exhibits a moderately higher variance. As shown in Figure 2.21, the MSE comparison highlights the strong comparability of the estimators. Moreover, the curves in Figure 2.19, representing the theoretical variances, are nearly indistinguishable.

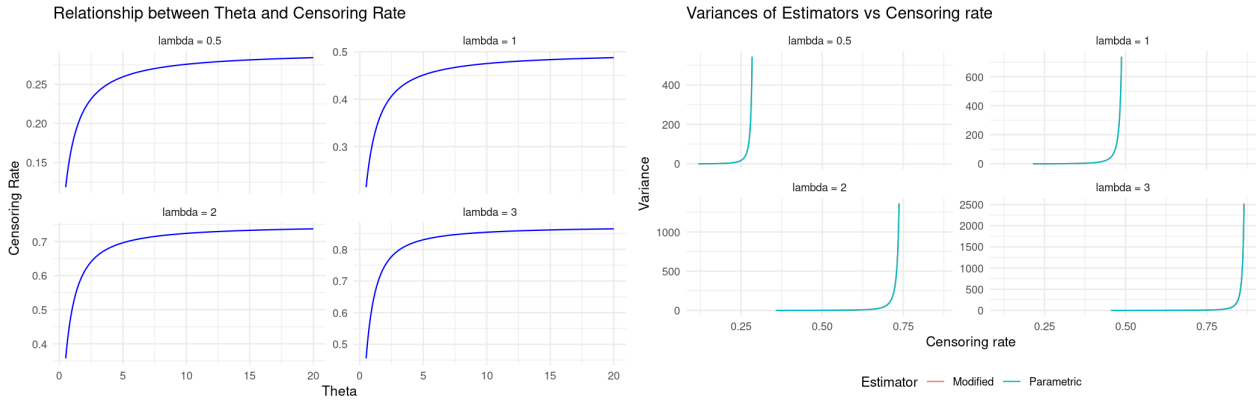

Figure 2.20: Comparison by Censoring Rates for $X \sim Beta(\theta, 1),\ Y \sim Pareto(\lambda, 1)$

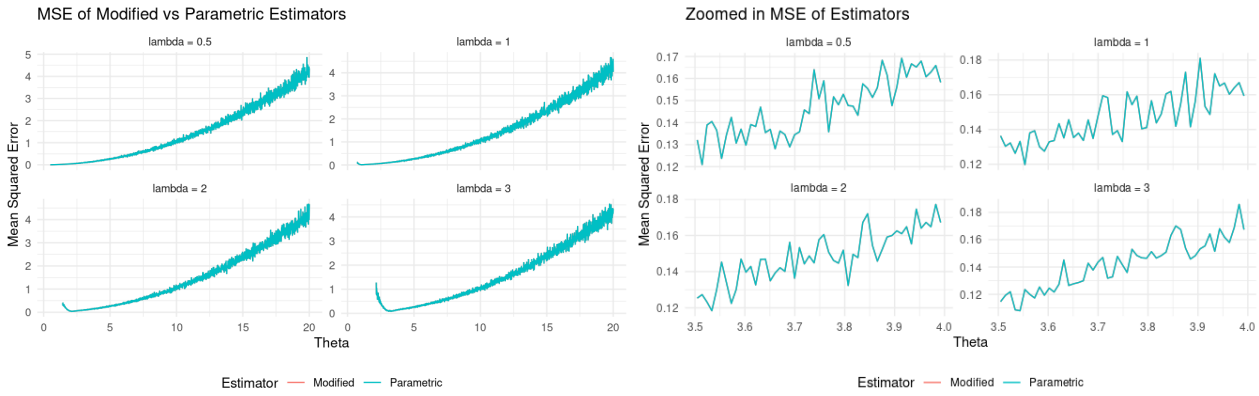

Figure 2.21: Mean Squared Error Comparison when $X \sim Beta(\theta, 1),\ Y \sim Pareto(\lambda, 1)$

The two MLEs yielded identical results from the simulated data. Consequently, there was little to no distinction in the curves of the estimators, which explains the absence of the red curve corresponding to the modified estimator in the graph. Given the outcome of the theoretical computations, these results are entirely consistent and expected.

2.6 Summary

In this chapter, we examined three continuous distributions for the event variable, and some distributions for the censoring variable. Integrability and finiteness were key in making suitable choices

for Y. Therefore, combinations such as $(X \sim Exp(1/\theta), Y \sim Weibull(\theta, 2))$, $(X \sim Beta(\theta, 1), Y \sim Uniform(0, 1))$ and $(X \sim Weibull(\theta, p), Y \sim Weibull(\theta, q), p \neq q)$ were excluded. In cases where both X and Y followed the same distribution, as in Sections 2.3.1 and 2.4.1, finiteness was ensured by choosing the same shape parameters where applicable, and by fixing the scale parameters such that $\theta > \lambda$. Although in these cases $Var_M(\theta) \not\leq Var_P(\theta)$, the estimators exhibited similar behavior across the range of θ for small values of λ. Overall, the estimates derived from the compact and full parametric likelihoods were found to be comparable for small λ, typically $\lambda \leq 1$. A higher comparability was also observed at small θ values.

For $X \sim Exponential$ and $X \sim Weibull$, the random censoring rates were negatively correlated with the parameter being estimated. In contrast, for $X \sim Beta(\theta, 1)$, the random censoring rate increased as $\theta \to \infty$. Moreover, as $\mathbb{P}(\delta = 0) \to \infty$, the difference in variances between the two estimators became negligible. In a few instances, the modified estimator exhibited a slightly higher variance with increasing censoring rate. The most notable results in the single-parameter case were observed in the case of $X \sim Beta(\theta, 1)$ and $Y \sim Pareto(\lambda, 1)$, discussed in Section 2.5.3, where no significant distinction was seen between the estimated variances across censoring rates, and the mean squared errors of the two maximum likelihood estimators were identical.

Chapter 3

Multiparameter Case

In the previous chapter, we discussed and illustrated the performance of the full parametric and modified estimators for the special case of the observed variable X which depends only on a single parameter. This chapter extends the discussion to consider observed distributions with mutliple parameters to be estimated. We begin the chapter with an introduction to maximum likelihood estimation in the absence of random censoring.

Let $X_1, \ldots, X_n$ be a random sample from uncensored data with common pdf or pmf $f(x \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top$, $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k$, the parameter space. The likelihood function and its logarithm are defined by

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta} \mid X_1, \ldots, X_n) &= \prod_{i=1}^{n} f(X_i \mid \boldsymbol{\theta}) \\ \ln \mathcal{L}(\boldsymbol{\theta} \mid X_1, \ldots, X_n) &= \sum_{i=1}^{n} \ln f(X_i \mid \boldsymbol{\theta}) \end{aligned}$$

The regularity conditions assumed for this theory are detailed in Appendix (A). The maximum likelihood estimator of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ that miximizes the likelihood or log-likelihood function. Potential candidates for the MLE are values of $(\theta_1, \ldots, \theta_n)$ obtained by solving

$$\frac{\partial}{\partial \theta_j} \ln \mathcal{L}(\boldsymbol{\theta} \mid X_1, \ldots, X_n) = 0, \quad 1 \leq j \leq k \tag{3.1}$$

The solutions to Equation 3.1 are potential maxima because one needs to verify if they are indeed the maxima. Another approach to determining the MLE is the direct maximization method. This

method involves finding a global upper bound on the likelihood function and then proving that there is a unique point for which the upper bound is attained, see Casella and Berger (2024).

In the single parameter case, we established that the expectation of the score function is zero, and that the negative expectation of its derivative gives the Fisher Information. In the multiparameter case, the Fisher Information becomes a matrix with element $\mathcal{I}_{ij}(\boldsymbol{\theta})$ given below, see Lehmann and Casella (2006).

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) \;=\; \mathbb{E}\left(\frac{\partial}{\partial\theta_i}\ln f(X\mid\boldsymbol{\theta})\cdot\frac{\partial}{\partial\theta_j}\ln f(X\mid\boldsymbol{\theta})\right),\quad 1\le i,\,j\le k \tag{3.2}$$

An alternative and simpler expression for the Information may be obtained using the expression:

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = -\mathbb{E}\left(\frac{\partial^2\ln f(X\mid\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right)$$

We consider the specific case where $k=2$, so that $\boldsymbol{\theta}=(\theta_1,\theta_2)^\top$. The Fisher Information matrix is

$$\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} -\mathbb{E}\left(\frac{\partial^2\ln f(X|\boldsymbol{\theta})}{\partial\theta_1^2}\right) & -\mathbb{E}\left(\frac{\partial^2\ln f(X|\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2}\right) \\ -\mathbb{E}\left(\frac{\partial^2\ln f(X|\boldsymbol{\theta})}{\partial\theta_2\partial\theta_1}\right) & -\mathbb{E}\left(\frac{\partial^2\ln f(X|\boldsymbol{\theta})}{\partial\theta_2^2}\right) \end{bmatrix}$$

Performing a multivariate Taylor series expansion of the score function $S(\boldsymbol{\theta}\mid X)$, around the true parameter θ_l, we have

$$0 \;=\; \frac{\partial}{\partial\theta_j}\frac{1}{n}\sum_{i=1}^{n}\ln f(X_i\mid\boldsymbol{\theta}) + \sum_{l=1}^{k}(\hat{\theta}_l-\theta_l)\;\underbrace{\frac{\partial^2}{\partial\theta_l\partial\theta_j}\;\frac{1}{n}\;\left(\sum_{i=1}^{n}\ln f(X_i\mid\boldsymbol{\theta})\right)}_{\hat{B}(\boldsymbol{\theta})},\quad 1\le j\le k$$

Let $\hat{B}(\boldsymbol{\theta})$ be the matrix identified in the above equation. Then,

$$\hat{B}(\boldsymbol{\theta}) \;\approx\; -\mathcal{I}(\boldsymbol{\theta}),\;\; \text{since } \mathcal{I}(\boldsymbol{\theta}) = \lim_{n\to\infty} -\frac{\partial^2}{\partial\theta_l\partial\theta_j}\;\frac{1}{n}\;\left(\sum_{i=1}^{n}\ln f(X_i\mid\boldsymbol{\theta})\right)\;\text{ by law of large numbers.}$$

$$0 \;\approx\; S(\boldsymbol{\theta}\mid X) + \hat{B}(\boldsymbol{\theta})\,(\widehat{\boldsymbol{\theta_l}}-\boldsymbol{\theta_l}),\;\; \text{where } S(\boldsymbol{\theta}\mid X) = \left[\frac{\partial}{\partial\theta_j}\;\sum_{i=1}^{n}\ln f(X_i\mid\boldsymbol{\theta}),\quad 1\le j\le k\right]$$

is the score vector, as in Equation (3.1). This implies,

$$\begin{aligned} -\hat{B}(\boldsymbol{\theta})\ (\widehat{\boldsymbol{\theta}}_{\boldsymbol{l}} - \boldsymbol{\theta}_{\boldsymbol{l}}) &\approx S(\boldsymbol{\theta} \mid X) \\ \mathcal{I}(\boldsymbol{\theta})\ (\widehat{\boldsymbol{\theta}}_{\boldsymbol{l}} - \boldsymbol{\theta}_{\boldsymbol{l}}) &\approx S(\boldsymbol{\theta} \mid X) \end{aligned}$$

Therefore, the limiting distribution of the $\widehat{\boldsymbol{\theta}}$ when $k = 2$ is:

$$\begin{aligned} \sqrt{n}\ \ \mathcal{I}(\boldsymbol{\theta}) \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} &\longrightarrow \mathcal{N}_2\left(\mathbf{0},\ \mathcal{I}(\boldsymbol{\theta})\right) \\ \sqrt{n}\ \ \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} &\longrightarrow \mathcal{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \mathcal{I}^{-1}(\boldsymbol{\theta})\ \mathcal{I}(\boldsymbol{\theta})\ \mathcal{I}^{-1}(\boldsymbol{\theta})\right) \\ \therefore\ \ \sqrt{n}\ \ \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} &\longrightarrow \mathcal{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \mathcal{I}^{-1}(\boldsymbol{\theta})\right) \end{aligned}$$

See Pawitan (2001) and Bickel and Doksum (2015) for a comprehensive overview of the limiting distribution.

3.1 Multivariate Normal Distribution and Confidence Ellipsoid

To assess the performance of the above estimators in terms of their limiting Normal distribution, we make use of so-called *confidence ellipsoids*. Let $X = (X_1, \ldots, X_k)$ have a k-variate Normal distribution, $\mathcal{N}_k(\mu, \Sigma)$. Now, define Z by

$$Z \ = \ \Sigma^{-1/2}(X - \mu)$$

where $\Sigma^{-1/2}$ is the square-root matrix of Σ, i.e., $\Sigma^{-1/2}\Sigma^{-1/2} = \Sigma$. Then Z has $\mathcal{N}_k(0, I_k)$ distribution, where I_k is the $k \times k$ identity matrix. Let $Y = (X - \mu)^\top \Sigma^{-1} (X - \mu)$, then from the above

$$Y = Z^\top \Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2} Z = Z^\top Z = \sum_{i=1}^{k} Z_i^2$$

Since $Z_1, \ldots, Z_k$ are iid $\mathcal{N}(0, 1)$, Y has χ^2 distribution with k degrees of freedom, see Section B.4 – B.7 of Bickel and Doksum (2015). Hence we have

Theorem 3.1. *Suppose X has a $\mathcal{N}_k(\mu, \Sigma)$ distribution, where Σ is positive definite. Then the random variable $Y = (X-\mu)^\top \Sigma^{-1} (X-\mu)$ has $\chi^2(k)$ distribution.*

See Section 3.5 of Hogg et al. (2018). Since the limiting ditribution of $\widehat{\boldsymbol{\theta}}$ (the MLE of $\boldsymbol{\theta}$) is N_2, we have that $X^\top \Sigma^{-1} X \sim \chi^2_2$ where $X = \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. A $100(1-\alpha)\%$ confidence region for $\boldsymbol{\theta}$ is given by

$$X^\top \Sigma^{-1} X \leq \chi^2_{2,\alpha}$$

where $\chi^2_{2,\alpha}$ is the upper α quantile of the χ^2_2 distribution, that is $\mathbb{P}\left(\chi^2_2 > \chi^2_{2,\alpha}\right) = \alpha$. This confidence region will be the interior of the ellipse:

$$x_1^2\, \Sigma_{11}^{-1} + x_2^2\, \Sigma_{22}^{-1} + 2\, \Sigma_{12}^{-1}\, x_1 x_2 \;\leq \chi^2_{2,\alpha}$$

where we denote the entries of Σ^{-1} by $\Sigma^{-1}_{ij}, 1 \leq i, j \leq 2$. Here, we consider the significance level $\alpha = 0.01$, and 0.05. Note that the boundary of such an ellipse is also a constant-density contour of the corresponding bi-variate Normal density:

$$f(\boldsymbol{x}) = \frac{1}{2\pi|\Sigma|} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right), \quad \text{for } \boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^2$$

where $|\Sigma| = \det \Sigma = \sigma_1 \sigma_2 \sqrt{1-\rho^2}$. In this chapter, we propose to use this *confidence ellipsoid* as a graphical illustration of the performance of $\widehat{\boldsymbol{\theta}}$.

From Bickel and Doksum (2015), when $k = 2$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]$

Thus, the inverse of the variance-covariance matrix

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$$

ρ is the correlation between x_1 and x_2. If x_1 and x_2 are independent then $\rho = 0$. If $\rho = 0$, and $\sigma_1^2 = \sigma_2^2$, the eigenvalues would also be equal, and a circle will be obtained instead of an ellipse. If $\rho > 0$, then the major axis of the ellipse will have a positive slope. Likewise, if $\rho < 0$, then the major axis will have a negative slope. Further, the lengths of the two axes are proportional to $\sqrt{\lambda_j}$,

where $\lambda_j, j = 1, 2$, are the eigenvalues of Σ. Note that,

$$\lambda_1 + \lambda_2 = \sigma_1^2 + \sigma_2^2.$$

3.1.1 Parametric Estimator

In the multiparameter case under random censoring, the score function is given as a vector. Specifically for $k = 2$,

$$S(\boldsymbol{\theta} \mid \delta, Z) \quad = \quad \begin{bmatrix} \frac{\partial}{\partial \theta_1} \left[\delta \ln f(Z \mid \boldsymbol{\theta}) + (1 - \delta) \ln \bar{F}(Z \mid \boldsymbol{\theta}) \right] \\ \frac{\partial}{\partial \theta_2} \left[\delta \ln f(Z \mid \boldsymbol{\theta}) + (1 - \delta) \ln \bar{F}(Z \mid \boldsymbol{\theta}) \right] \end{bmatrix}$$

This yields a 2×2 Fisher information matrix, where each entry $\mathcal{I}_{RC}(\boldsymbol{\theta})_{ij}$ is given by:

$$\mathcal{I}_{RC}(\boldsymbol{\theta})_{ij} = \int \frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_j} \cdot f(x \mid \boldsymbol{\theta}) \bar{G}(x) dx \; + \; \int \frac{\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_j}}{\bar{F}(y \mid \boldsymbol{\theta})} \cdot g(y) dy$$

The limiting variance-covariance matrix would be symmetric, which implies that $\mathcal{I}_{RC}(\boldsymbol{\theta})_{12} = \mathcal{I}_{RC}(\boldsymbol{\theta})_{21}$, with elements given to be,

$$\mathcal{I}_{RC}(\boldsymbol{\theta})_{11} = \int \left(\frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_1} \right)^2 \cdot f(x \mid \boldsymbol{\theta}) \, \bar{G}(x) dx \; + \; \int \frac{\left(\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_1} \right)^2}{\bar{F}(y \mid \boldsymbol{\theta})} \cdot g(y) dy$$

$$\mathcal{I}_{RC}(\boldsymbol{\theta})_{22} = \int \left(\frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_2} \right)^2 \cdot f(x \mid \boldsymbol{\theta}) \, \bar{G}(x) dx \; + \; \int \frac{\left(\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_2} \right)^2}{\bar{F}(y \mid \boldsymbol{\theta})} \cdot g(y) dy$$

$$\mathcal{I}_{RC}(\boldsymbol{\theta})_{12} = \int \frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_1} \cdot \frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_2} \cdot f(x \mid \boldsymbol{\theta}) \, \bar{G}(x) dx \; + \; \int \frac{\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_1} \cdot \frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_2}}{\bar{F}(y \mid \boldsymbol{\theta})} \cdot g(y) dy$$

The limiting distribution of $\widehat{\boldsymbol{\theta}}$ is

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \longrightarrow \mathcal{N}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathcal{I}_{RC}^{-1}(\boldsymbol{\theta}) \right), \text{ where}$$

$$\mathcal{I}_{RC}^{-1}(\boldsymbol{\theta}) = \frac{1}{\mathcal{I}_{RC}(\boldsymbol{\theta})_{11} \, \mathcal{I}_{RC}(\boldsymbol{\theta})_{22} - \mathcal{I}_{RC}(\boldsymbol{\theta})_{12}^2} \begin{pmatrix} \mathcal{I}_{RC}(\boldsymbol{\theta})_{22} & -\mathcal{I}_{RC}(\boldsymbol{\theta})_{12} \\ -\mathcal{I}_{RC}(\boldsymbol{\theta})_{12} & \mathcal{I}_{RC}(\boldsymbol{\theta})_{11} \end{pmatrix}$$

and,

$$\begin{aligned} \Sigma &= \mathcal{I}_{RC}^{-1}(\boldsymbol{\theta}) \\ \Longrightarrow \Sigma^{-1} &= \mathcal{I}_{RC}(\boldsymbol{\theta}) = \begin{pmatrix} \mathcal{I}_{RC}(\boldsymbol{\theta})_{11} & \mathcal{I}_{RC}(\boldsymbol{\theta})_{12} \\ \mathcal{I}_{RC}(\boldsymbol{\theta})_{12} & \mathcal{I}_{RC}(\boldsymbol{\theta})_{22} \end{pmatrix} \end{aligned}$$

3.1.2 Modified Estimator

The limiting distribution of $\widehat{\boldsymbol{\theta}}$ under the compact likelihood is derived as follows. Performing a multivariate Taylor series expansion around the true parameter, we have

$$\begin{aligned} 0 &= \sum_{i=1}^{n} W_{in} \frac{\partial}{\partial \theta_1} \ln f(Z_i|\boldsymbol{\theta}) + (\hat{\theta}_1 - \theta_1) \sum_{i=1}^{n} W_{in} \frac{\partial^2}{\partial \theta_1^2} \ln f(Z_i|\boldsymbol{\theta}) + (\hat{\theta}_2 - \theta_2) \sum_{i=1}^{n} W_{in} \frac{\partial^2}{\partial \theta_1 \partial \theta_2} \ln f(Z_i|\boldsymbol{\theta}) \\ 0 &= \sum_{i=1}^{n} W_{in} \frac{\partial}{\partial \theta_2} \ln f(Z_i|\boldsymbol{\theta}) + (\hat{\theta}_1 - \theta_1) \sum_{i=1}^{n} W_{in} \frac{\partial^2}{\partial \theta_2 \partial \theta_1} \ln f(Z_i|\boldsymbol{\theta}) + (\hat{\theta}_2 - \theta_2) \sum_{i=1}^{n} W_{in} \frac{\partial^2}{\partial \theta_2^2} \ln f(Z_i|\boldsymbol{\theta}) \end{aligned}$$

And the MLE of θ_j is obtained at

$$S_j(\boldsymbol{\theta} \mid Z) = \sum_{i=1}^{n} W_{in} \frac{\partial}{\partial \theta_j} \ln f\left(Z_i \mid \boldsymbol{\theta}\right) \Big|_{\theta_j = \widehat{\theta}_j} = 0 \,, \quad j = 1, 2.$$

For $\varphi(Z_i) = \begin{bmatrix} \dfrac{\partial}{\partial \theta_1} \ln f\left(Z_i \mid \boldsymbol{\theta}\right) \\ \dfrac{\partial}{\partial \theta_2} \ln f\left(Z_i \mid \boldsymbol{\theta}\right) \end{bmatrix}$, we have that

$$\sqrt{n} \sum_{i=1}^{n} W_{in}\, \varphi(Z_i) \longrightarrow \mathcal{N}_2\left(\mathbf{0},\ D(\boldsymbol{\theta})\right) \tag{3.3}$$

where,

$$D(\boldsymbol{\theta}) = \begin{bmatrix} D_{11}(\boldsymbol{\theta}) & D_{12}(\boldsymbol{\theta}) \\ D_{21}(\boldsymbol{\theta}) & D_{22}(\boldsymbol{\theta}) \end{bmatrix}$$

To determine the entries of $D(\boldsymbol{\theta})$ we use Cramèr-Wold device, i.e., consider an arbitrary linear combination $\sum_{i=1}^{n} W_{in}\, \varphi(Z_i)$, so that

$$\sum_{i=1}^{n} W_{in} \left[c_1 \frac{\partial}{\partial \theta_1} \ln f(Z_i \mid \boldsymbol{\theta}) + c_2 \frac{\partial}{\partial \theta_2} \ln f(Z_i \mid \boldsymbol{\theta}) \right] \longrightarrow \mathcal{N}\left(0,\ c^\top D\, c\right),$$

where,

$$c^\top D\, c = c_1^2 D_{11} + c_2^2 D_{22} + 2 c_1 c_2 D_{12} \,.$$

Expression for the terms is found by applying Equation (2.6) to the function

$$\varphi(Z) = c_1 \, \frac{\partial}{\partial \theta_1} \ln f(Z \mid \boldsymbol{\theta}) \, + \, c_2 \, \frac{\partial}{\partial \theta_2} \ln f(Z \mid \boldsymbol{\theta})$$

We thus get

$$D_{11}(\boldsymbol{\theta}) \, = \int \left(\frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_1} \right)^2 \cdot \frac{f(x \mid \boldsymbol{\theta})}{\bar{G}(x)} dx - \int \frac{\left(\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_1} \right)^2}{\bar{F}(y \mid \boldsymbol{\theta}) \cdot \bar{G}^2(y)} \cdot g(y) dy$$

$$D_{22}(\boldsymbol{\theta}) \, = \int \left(\frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_2} \right)^2 \cdot \frac{f(x \mid \boldsymbol{\theta})}{\bar{G}(x)} dx - \int \frac{\left(\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_2} \right)^2}{\bar{F}(y \mid \boldsymbol{\theta}) \cdot \bar{G}^2(y)} \cdot g(y) dy$$

$$D_{12}(\boldsymbol{\theta}) \, = \, D_{21}(\boldsymbol{\theta}) = \int \frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_1} \cdot \frac{\partial \ln f(x \mid \boldsymbol{\theta})}{\partial \theta_2} \cdot \frac{f(x \mid \boldsymbol{\theta})}{\bar{G}(x)} \, dx - \int \frac{\frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_1} \cdot \frac{\partial \bar{F}(y|\boldsymbol{\theta})}{\partial \theta_2}}{\bar{F}(y \mid \boldsymbol{\theta}) \cdot \bar{G}^2(y)} \cdot g(y) dy$$

We are able to find that,

$$\begin{aligned} \sqrt{n} \;\; \mathcal{I}(\boldsymbol{\theta}) \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \quad &\longrightarrow \quad \mathcal{N}_2 \left(\mathbf{0}, \; D(\boldsymbol{\theta}) \right) \\ \implies \quad \sqrt{n} \quad \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \quad &\longrightarrow \quad \mathcal{N}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; \mathcal{I}^{-1}(\boldsymbol{\theta}) \; D(\boldsymbol{\theta}) \; \mathcal{I}^{-1}(\boldsymbol{\theta}) \right) \end{aligned}$$

Here, $\Sigma = \mathcal{I}^{-1}(\boldsymbol{\theta}) \; D(\boldsymbol{\theta}) \; \mathcal{I}^{-1}(\boldsymbol{\theta})$, so that $\Sigma^{-1} \; = \; \mathcal{I}(\boldsymbol{\theta}) \, D^{-1}(\boldsymbol{\theta}) \; \mathcal{I}(\boldsymbol{\theta})$.

Let the elements of $\mathcal{I}(\boldsymbol{\theta}) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where:

$$\begin{aligned} a = \mathcal{I}_{11} &= -\mathbb{E} \left[\frac{\partial^2 \ln f(X \mid \boldsymbol{\theta})}{\partial \theta_1^2} \right] \\ d = \mathcal{I}_{22} &= -\mathbb{E} \left[\frac{\partial^2 \ln f(X \mid \boldsymbol{\theta})}{\partial \theta_2^2} \right] \\ b = c = \mathcal{I}_{12} = \mathcal{I}_{21} &= -\mathbb{E} \left[\frac{\partial^2 \ln f(X \mid \boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} \right] \end{aligned}$$

so that $\mathcal{I}^{-1}(\boldsymbol{\theta}) = \dfrac{1}{ad - b^2} \begin{bmatrix} d & -b \\ -b & a \end{bmatrix}$, since $b = c$.

Now, let the elements of matrix $D(\boldsymbol{\theta}) = \begin{bmatrix} e & f \\ f & g \end{bmatrix}$, so that the limiting variance-covariance matrix, $Cov_M(\boldsymbol{\theta}) = \mathcal{I}^{-1}(\boldsymbol{\theta})\ D(\boldsymbol{\theta})\ \mathcal{I}^{-1}(\boldsymbol{\theta})$:

$$
\begin{aligned}
Cov_M(\boldsymbol{\theta}) &= \frac{1}{(ad-b^2)^2} \begin{bmatrix} d & -b \\ -b & a \end{bmatrix} \times \begin{bmatrix} e & f \\ f & g \end{bmatrix} \times \begin{bmatrix} d & -b \\ -b & a \end{bmatrix} \\
&= \frac{1}{(ad-b^2)^2} \begin{bmatrix} de-bf & df-bg \\ -be+af & -bf+ag \end{bmatrix} \times \begin{bmatrix} d & -b \\ -b & a \end{bmatrix} \\
Cov_M(\boldsymbol{\theta}) &= \frac{1}{(ad-b^2)^2} \begin{bmatrix} d(de-bf)-b(df-bg) & -b(de-bf)+a(df-bg) \\ d(-be+af)-b(-bf+ag) & -b(-be+af)+a(-bf+ag) \end{bmatrix}
\end{aligned}
$$

where,

$$
\begin{aligned}
(ad-b^2)^2 &= (\mathcal{I}_{11}\mathcal{I}_{22}-\mathcal{I}_{12}^2)^2 \\
d(de-bf)-b(df-bg) &= D_{11}\mathcal{I}_{22}^2 - 2\mathcal{I}_{12}\mathcal{I}_{22}D_{12} + D_{22}\mathcal{I}_{12}^2 \\
-b(de-bf)+a(df-bg) &= -\mathcal{I}_{12}\mathcal{I}_{22}D_{11} + D_{12}\mathcal{I}_{12}^2 + \mathcal{I}_{11}\mathcal{I}_{22}D_{12} - \mathcal{I}_{11}\mathcal{I}_{12}D_{22} \\
d(-be+af)-b(-bf+ag) &= -\mathcal{I}_{12}\mathcal{I}_{22}D_{11} + \mathcal{I}_{11}\mathcal{I}_{22}D_{12} + D_{12}\mathcal{I}_{12}^2 - \mathcal{I}_{11}\mathcal{I}_{12}D_{22} \\
-b(-be+af)+a(-bf+ag) &= D_{11}\mathcal{I}_{12}^2 - 2\mathcal{I}_{11}\mathcal{I}_{22}D_{12} + D_{22}\mathcal{I}_{11}^2
\end{aligned}
$$

3.2 Weibull Distributed Event

Under this case, we consider $X \sim Weibull(\theta, p)$ and $Y \sim Pareto(\lambda, 1)$. We begin by computing the theoretical variances, followed by the maximum likelihood estimation of the scale parameter, θ and the shape parameter, p.

$$
\begin{aligned}
f(x \mid \theta, p) &= p\theta x^{p-1} e^{-\theta x^p}, x > 0,\ p > 0,\ \theta > 0 \\
\ln f(x \mid \theta, p) &= \ln p + \ln \theta + (p-1)\ln x - \theta x^p \\
\frac{\partial}{\partial \theta} \ln f(x \mid \theta, p) &= \frac{1}{\theta} - x^p \\
\frac{\partial^2}{\partial \theta^2} \ln f(x \mid \theta, p) &= -\frac{1}{\theta} \\
\frac{\partial^2}{\partial \theta \partial p} \ln f(x \mid \theta, p) &= -x^p \ln x
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial p}\ln f(x \mid \theta, p) &= \frac{1}{p} + \ln x - \theta x^p \ln x \\
\frac{\partial^2}{\partial p^2}\ln f(x \mid \theta, p) &= -\frac{1}{p^2} - \theta\, x^p \ln^2 x \\
\bar{F}(x \mid \theta, p) &= e^{-\theta x^p},\ x > 0,\ \theta > 0,\ p > 0 \\
\ln \bar{F}(x \mid \theta, p) &= -\theta x^p \\
\frac{\partial}{\partial \theta}\ln \bar{F}(x \mid \theta, p) &= -x^p \\
\frac{\partial}{\partial p}\ln \bar{F}(x \mid \theta, p) &= -\theta x^p \ln x
\end{aligned}
$$

Let $g(y) = \dfrac{\lambda}{(1+y)^{\lambda+1}}$ and $\bar{G}(y) = \dfrac{1}{(1+y)^{\lambda}}$. The components of the variance-covariance matrices are computed below.

For $\mathcal{I}(\boldsymbol{\theta})$:

$$
\mathcal{I}_{11} = -\mathbb{E}\left[\frac{\partial^2 \ln f(X \mid \theta, p)}{\partial \theta^2}\right] = \frac{1}{\theta^2}
$$

$$
\mathcal{I}_{12} = -\mathbb{E}\left[\frac{\partial^2 \ln f(X \mid \theta, p)}{\partial \theta \partial p}\right] = \mathbb{E}[X^p \ln X]
$$

$$
\begin{aligned}
\mathbb{E}[X^p \ln X] &= \int_0^\infty x^p \ln x \cdot p\theta x^{p-1} e^{-\theta x^p} dx,\ \text{let } z = x^p,\ dz = px^{p-1}dx \\
&= \theta \int_0^\infty z\, \ln z^{\frac{1}{p}} \cdot p \cdot x^{p-1} \cdot e^{-\theta z}\, \frac{dz}{px^{p-1}},\quad \text{since } x = z^{\frac{1}{p}},\ dx = \frac{dz}{px^{p-1}} \\
&= \frac{\theta}{p}\int_0^\infty z\, \ln z e^{-\theta z},\quad \text{let } u = \theta z,\ z = \frac{u}{\theta},\ dz = \frac{du}{\theta} \\
&= \frac{\theta}{p}\int_0^\infty \frac{u}{\theta} \cdot \ln \frac{u}{\theta} \cdot e^{-u}\, \frac{du}{\theta} = \frac{1}{p\theta}\int_0^\infty u(\ln u - \ln \theta)\, e^{-u} du \\
&= \frac{1}{p\theta}\left[\int_0^\infty u \ln u\, e^{-u} du - \ln\theta \int_0^\infty u e^{-u} du\right] \\
\implies \mathbb{E}[X^p \ln X] &= \frac{1}{p\theta}\left(1 - \gamma - \ln\theta\right),\quad \text{where } \gamma \text{ is the Euler-Mascheroni constant.} \\
\therefore\ \ \mathcal{I}_{12} &= \frac{1}{p\theta}\left(1 - \gamma - \ln\theta\right)
\end{aligned}
$$

$$
\mathcal{I}_{22} = -\mathbb{E}\left[\frac{\partial^2 \ln f(X \mid \theta, p)}{\partial p^2}\right] = \frac{1}{p^2} + \theta\, \mathbb{E}\left[X^p \ln^2 X\right]
$$

$$
\begin{aligned}
\mathbb{E}\left[X^p \ln^2 X\right] &= \int_0^\infty x^p \ln^2 x \cdot p\theta x^{p-1} e^{-\theta x^p} dx,\ \text{let } z = x^p,\ dz = px^{p-1}dx \\
&= \frac{\theta}{p}\int_0^\infty z \ln^2 z\ e^{-\theta z},\quad \text{let } u = \theta z,\ z = \frac{u}{\theta},\ dz = \frac{du}{\theta}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[X^p \ln^2 X\right] &= \frac{\theta}{p}\int_0^\infty \frac{u}{\theta}\cdot\left(\ln\frac{u}{\theta}\right)^2\cdot e^{-u}\,\frac{du}{\theta} \\
&= \frac{1}{p\theta}\left[\int_0^\infty u\ln^2 u\, e^{-u}du - 2\ln\theta\int_0^\infty u\ln u\, e^{-u}du + \ln^2\theta\int_0^\infty u\, e^{-u}du\right] \\
&= \frac{1}{p\theta}\left(\frac{\pi^2}{6} - 2\gamma + \gamma^2 - 2\ln\theta\cdot(1-\gamma) + \ln^2\theta\right)
\end{aligned}
$$

Hence,

$$
\mathcal{I}_{22} = \frac{1}{p^2} + \frac{1}{p}\left(\frac{\pi^2}{6} - 2\gamma + \gamma^2 - 2\ln\theta\cdot(1-\gamma) + \ln^2\theta\right)
$$

Recall that

$$
[\sigma^2(\boldsymbol{\theta})]_{ij} = \int \frac{\partial \ln f(x\mid\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial \ln f(x\mid\boldsymbol{\theta})}{\partial\theta_j}\cdot\frac{f(x\mid\boldsymbol{\theta})}{\bar{G}(x)}\,dx - \int \frac{\frac{\partial\bar{F}(y|\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial\bar{F}(y|\boldsymbol{\theta})}{\partial\theta_j}}{\bar{F}(y\mid\boldsymbol{\theta})\bar{G}^2(y)}\cdot g(y)\,dy
$$

$$
[\mathcal{I}_{RC}(\boldsymbol{\theta})]_{ij} = \int \frac{\partial \ln f(x\mid\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial \ln f(x\mid\boldsymbol{\theta})}{\partial\theta_j}\cdot f(x\mid\boldsymbol{\theta})\bar{G}(x)\,dx + \int \frac{\frac{\partial\bar{F}(y|\boldsymbol{\theta})}{\partial\theta_i}\cdot\frac{\partial\bar{F}(y|\boldsymbol{\theta})}{\partial\theta_j}}{\bar{F}(y\mid\boldsymbol{\theta})}\cdot g(y)\,dy
$$

Thus,

$$
\begin{aligned}
\sigma^2(\boldsymbol{\theta})_{11} &= p\theta\int_0^\infty\left(\frac{1}{\theta}-x^p\right)^2\cdot x^{p-1}e^{-\theta x^p}(1+x)^\lambda\,dx - \lambda\int_0^\infty x^{2p}e^{-\theta x^p}(1+x)^{\lambda-1}\,dx \\
\sigma^2(\boldsymbol{\theta})_{12} &= p\theta\int_0^\infty\left(\frac{1}{\theta}-x^p\right)\left(\frac{1}{p}+\ln x-\theta x^p\ln x\right)x^{p-1}e^{-\theta x^p}(1+x)^\lambda\,dx \;- \\
&\quad \lambda\theta\int_0^\infty x^{2p}\ln x\cdot e^{-\theta x^p}(1+x)^{\lambda-1}\,dx \\
\sigma^2(\boldsymbol{\theta})_{22} &= p\theta\int_0^\infty\left(\frac{1}{p}+\ln x-\theta x^p\ln x\right)^2 x^{p-1}e^{-\theta x^p}(1+x)^\lambda\,dx \;- \\
&\quad \lambda\theta^2\int_0^\infty x^{2p}\ln^2 x\cdot e^{-\theta x^p}(1+x)^{\lambda-1}\,dx \\
\mathcal{I}_{RC}(\boldsymbol{\theta})_{11} &= p\theta\int_0^\infty\left(\frac{1}{\theta}-x^p\right)^2 x^{p-1}e^{-\theta x^p}(1+x)^{-\lambda}\,dx + \lambda\int_0^\infty x^{2p}e^{-\theta x^p}(1+x)^{-\lambda-1}\,dx \\
\mathcal{I}_{RC}(\boldsymbol{\theta})_{12} &= p\theta\int_0^\infty\left(\frac{1}{\theta}-x^p\right)\left(\frac{1}{p}+\ln x-\theta x^p\ln x\right)x^{p-1}e^{-\theta x^p}(1+x)^{-\lambda}\,dx \;+ \\
&\quad \lambda\theta\int_0^\infty x^{2p}\ln x\cdot e^{-\theta x^p}(1+x)^{-\lambda-1}\,dx \\
\mathcal{I}_{RC}(\boldsymbol{\theta})_{22} &= p\theta\int_0^\infty\left(\frac{1}{p}+\ln x-\theta x^p\ln x\right)^2 x^{p-1}e^{-\theta x^p}(1+x)^{-\lambda}\,dx \;+ \\
&\quad \lambda\theta^2\int_0^\infty x^{2p}\ln^2 x\cdot e^{-\theta x^p}(1+x)^{-\lambda-1}\,dx
\end{aligned}
$$

with censoring rate,

$$
\mathbb{P}(X>Y) = \lambda\int_0^\infty \frac{e^{-\theta y^p}}{(1+y)^{\lambda+1}}\,dy
$$

The integrals are evaluated computationally using software. The variance-covariance matrices

obtained from this case produced the following results.

Table 3.1: Numerical Results from Weibull Distributed Event at significance level $\alpha = 0.01$

p	θ	λ	Parametric Ellipse Area		Modified Ellipse Area		Censoring Rate
			Theoretical	Estimated	Theoretical	Estimated	
1	0.1	0.5	4.939092	5.026896	0.127872	0.134701	0.5944349
1	0.1	1	9.393342	9.436232	0.387559	0.401798	0.7985357
0.75	0.25	0.5	7.925077	8.053711	1.153352	1.221890	0.5007178
0.75	0.25	1	12.85312	12.68529	3.120171	2.973647	0.6950513
1	0.5	0.5	16.48671	16.49241	8.133768	8.184578	0.3443205
1	0.5	1	22.87485	22.86229	13.45165	13.58318	0.5385447
0.5	0.1	0.5	3.986597	3.961047	0.172574	0.173800	0.7490133
0.5	0.1	1	8.370920	8.249199	2.511489	2.299068	0.8708996
0.5	1	0.5	15.16020	15.22512	17.07878	17.39303	0.2453899
0.5	1	1	18.75805	18.80608	30.21760	30.04615	0.3785504

Due to the complexity of graphing multiple ellipses across a range of parameter values, we selected a sample of values for θ and p at $\lambda = 0.5$ and $\lambda = 1$. For each combination (or triple), we computed the corresponding variance-covariance matrices and plotted the resulting confidence ellipses. The parameter values are listed in Table 3.1 above. Comparisons are made based on the area covered, calculated as

$$\text{Area} = \pi ab, \qquad \text{where } a, b \text{ are the half lengths of the ellipse.}$$

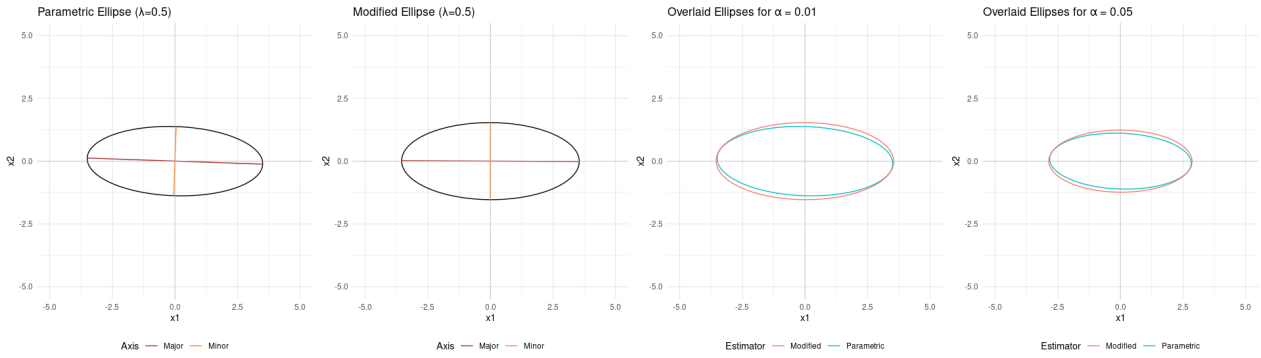

Figure 3.1: Theoretical Ellipse Comparison for $Weibull(p = 0.5,\ \theta = 1)$ at $\lambda = 0.5$

We observed that ellipses associated with the modified estimator generally covered smaller areas than those of the parametric estimator across different censoring rates. This suggests that, for the selected (p, θ, λ) triples, the modified estimator tends to yield lower variance in the maximum likelihood estimates. This observation was consistent across all examined combinations, with the exception of the case $p = 0.5$, $\theta = 1$. Graphical illustrations of some triples are included below.

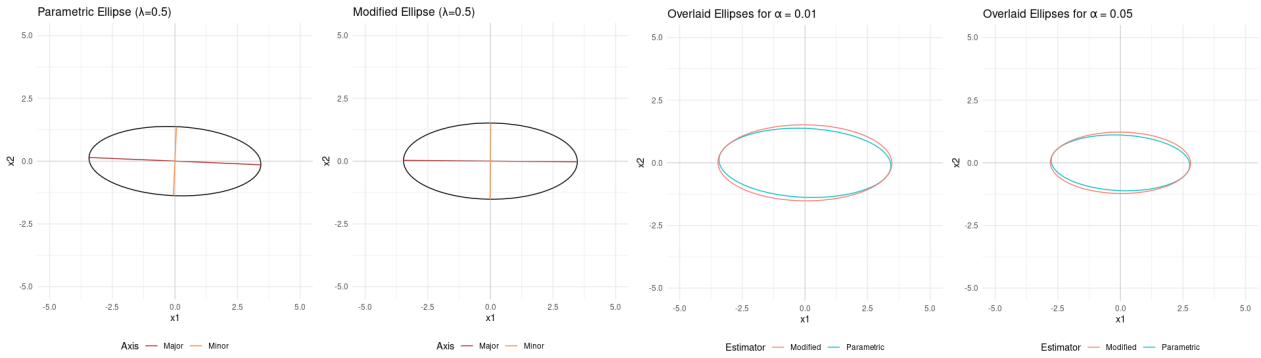

Figure 3.2: Estimated Ellipse Comparison for $Weibull(p = 0.5,\ \theta = 1)$ at $\lambda = 0.5$

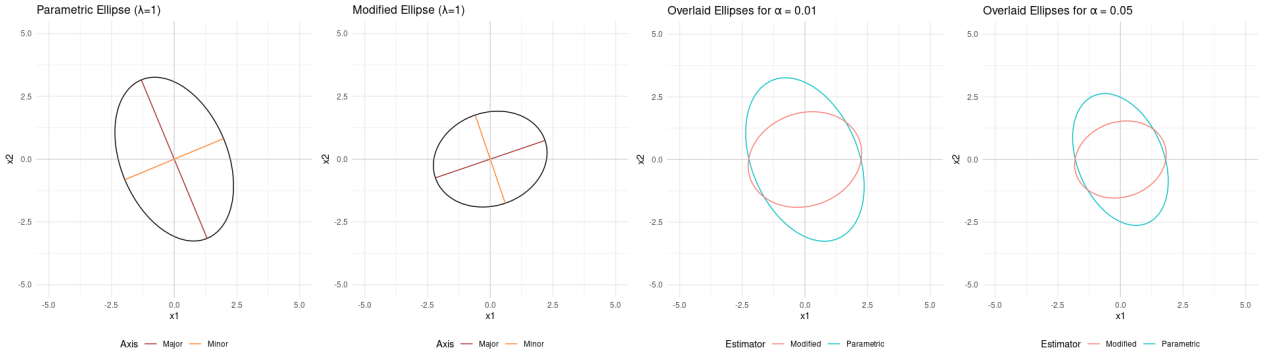

Figure 3.3: Theoretical Ellipse Comparison for $Weibull(p = 1,\ \theta = 0.5)$ at $\lambda = 1$

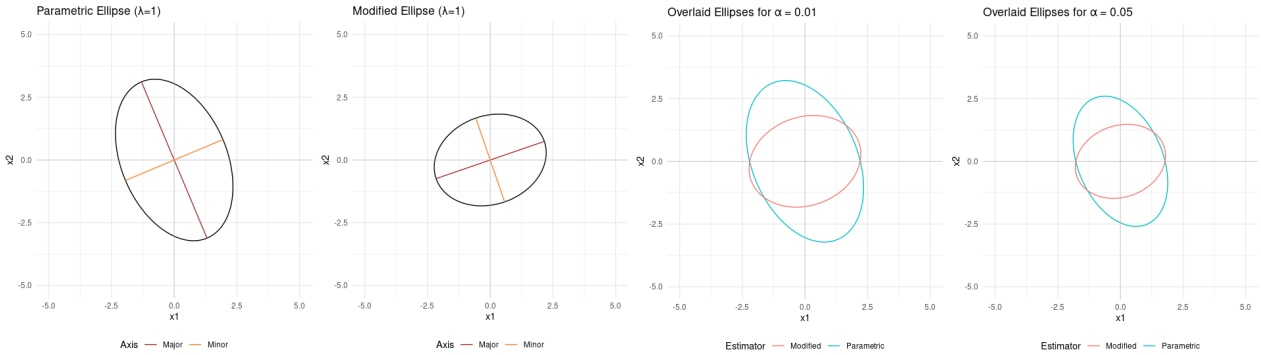

Figure 3.4: Estimated Ellipse Comparison for $Weibull(p = 1,\ \theta = 0.5)$ at $\lambda = 1$

The maximum likelihood estimators used to generate the ellipses in Figures 3.2 and 3.4 are shown below. In both likelihoods, no closed-form expressions was obtained for the shape parameter. Consequently, the `uniroot` function in R was employed to numerically solve for the roots of the respective score functions.

For $\hat{p}$

Parametric MLE

$$\sum_{i=1}^{n}\left[\delta_i\left(\frac{1}{p}+\ln z_i-\theta z_i^p\ln z_i\right)+(1-\delta_i)(-\theta z_i^p\ln z_i)\right]=0$$

$$\sum_{i=1}^{n}\left[\frac{\delta_i}{p}+\delta_i\ln z_i-\theta z_i^p\ln z_i\right]=0$$

Semi-parametric MLE

$$\sum_{i=1}^{n}W_{in}\frac{\partial}{\partial p}\ln f(z_i\mid\boldsymbol{\theta})=0$$

$$\sum_{i=1}^{n}W_{in}\left(\frac{1}{p}+\ln z_i-\theta z_i^p\ln z_i\right)=0$$

For $\hat{\theta}$

Parametric MLE

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\left[\delta_i\ln f(z_i\mid\boldsymbol{\theta})+(1-\delta_i)\ln\bar{F}(z_i\mid\boldsymbol{\theta})\right]=0$$

$$\sum_{i=1}^{n}\left[\delta_i\left(\frac{1}{\theta}-z_i^p\right)+(1-\delta_i)(-z_i^p)\right]=0$$

$$\frac{1}{\theta}\sum_{i=1}^{n}\delta_i-\sum_{i=1}^{n}z_i^p=0$$

$$\hat{\theta}^P=\frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}z_i^p}$$

Semi-parametric MLE

$$\sum_{i=1}^{n}W_{in}\frac{\partial}{\partial\theta}\ln f(z_i\mid\boldsymbol{\theta})=0$$

$$\sum_{i=1}^{n}W_{in}\left(\frac{1}{\theta}-z_i^p\right)=0$$

$$\frac{1}{\theta}\sum_{i=1}^{n}W_{in}=\sum_{i=1}^{n}W_{in}z_i^p$$

$$\hat{\theta}^M=\frac{\sum_{i=1}^{n}W_{in}}{\sum_{i=1}^{n}W_{in}z_i^p}$$

Table 3.2 below provides the maximum likelihood estimates and their corresponding mean squared errors for $\hat{p}$ and $\hat{\theta}$, based on 10,000 simulations drawn from the specified distributions.

Table 3.2: MLE Results for Weibull Distributed Event at significance level $\alpha = 0.01$

	$\hat{p}$					θ				
λ	p	$\hat{p}^P$	$\hat{p}^M$	MSE_P	MSE_M	θ	$\hat{\theta}^P$	$\hat{\theta}^M$	MSE_P	MSE_M
0.5	1	1.00630	1.00670	0.00004	0.00004	0.1	0.10208	0.10216	0.00000	0.00000
1	1	1.00999	1.01120	0.00010	0.00013	0.1	0.10043	0.10249	0.00000	0.00001
0.5	0.75	0.75858	0.75998	0.00007	0.00010	0.25	0.25329	0.25533	0.00001	0.00003
1	0.75	0.73879	0.73875	0.00013	0.00013	0.25	0.24709	0.24224	0.00001	0.00006
0.5	1	1.00752	1.00960	0.00006	0.00009	0.5	0.49614	0.49979	0.00001	0.00000
1	1	0.99868	1.00276	0.00000	0.00001	0.5	0.50043	0.50201	0.00000	0.00000
0.5	0.5	0.50107	0.50412	0.00000	0.00002	0.1	0.09855	0.10092	0.00000	0.00000
1	0.5	0.49053	0.51270	0.00009	0.00016	0.1	0.09427	0.10388	0.00003	0.00002
0.5	0.5	0.50089	0.50494	0.00000	0.00002	1	1.00366	1.01087	0.00001	0.00012
1	0.5	0.50262	0.49589	0.00001	0.00002	1	0.99590	0.98768	0.00002	0.00015

The observed similarity in the mean squared errors accounts for the consistency between the theoretical and estimated confidence ellipses.

3.3 Gamma Distributed Event

In the second scenario, we let $X \sim Gamma(\alpha, \theta)$ and $Y \sim Pareto(\lambda, 1)$.

$$\begin{aligned}
f(x \mid \alpha, \theta) &= \frac{x^{\alpha-1}\, e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\, \theta^{\alpha}},\ x > 0,\ \alpha > 0,\ \theta > 0 \\
\ln f(x \mid \alpha, \theta) &= (\alpha - 1)\ln x - \frac{x}{\theta} - \ln\Gamma(\alpha) - \alpha \ln\theta \\
\frac{\partial}{\partial\theta} \ln f(x \mid \alpha, \theta) &= \frac{x}{\theta^2} - \frac{\alpha}{\theta} \\
\frac{\partial^2}{\partial\theta^2} \ln f(x \mid \alpha, \theta) &= \frac{\alpha}{\theta^2} - \frac{x}{\theta^3} \\
\frac{\partial^2}{\partial\theta\partial\alpha} \ln f(x \mid \alpha, \theta) &= -\frac{1}{\theta} \\
\frac{\partial}{\partial\alpha} \ln f(x \mid \alpha, \theta) &= \ln x - \psi(\alpha) - \ln\theta \\
\frac{\partial^2}{\partial\alpha^2} \ln f(x \mid \alpha, \theta) &= -\psi'(\alpha) \\
\bar{F}(x \mid \alpha, \theta) &= \frac{1}{\Gamma(\alpha)}\, \Gamma\left(\alpha, \frac{x}{\theta}\right) \\
\ln \bar{F}(x \mid \alpha, \theta) &= \ln\Gamma\left(\alpha, \frac{x}{\theta}\right) - \ln\Gamma(\alpha) \\
\frac{\partial}{\partial\theta} \ln \bar{F}(x \mid \alpha, \theta) &= \frac{\partial}{\partial\theta} \ln\Gamma\left(\alpha, \frac{x}{\theta}\right) \\
\frac{\partial}{\partial\theta} \ln \bar{F}(x \mid \alpha, \theta) &= \frac{\frac{\partial}{\partial\theta}\Gamma\left(\alpha, \frac{x}{\theta}\right)}{\Gamma\left(\alpha, \frac{x}{\theta}\right)} \\
\frac{\partial}{\partial\theta} \ln \bar{F}(x \mid \alpha, \theta) &= \frac{1}{\Gamma\left(\alpha, \frac{x}{\theta}\right)} \cdot \frac{x^{\alpha}}{\theta^{\alpha+1}} \cdot e^{-\frac{x}{\theta}} \\
\frac{\partial}{\partial\alpha} \ln \bar{F}(x \mid \alpha, \theta) &= \frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{x}{\theta}\right)}{\Gamma\left(\alpha, \frac{x}{\theta}\right)} - \psi(\alpha)
\end{aligned}$$

where

$$\psi(\alpha) = \frac{d}{d\alpha}\ln\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)},\ \text{ and } \psi'(\alpha) = \frac{d^2}{d\alpha^2}\ln\Gamma(\alpha) = \frac{d}{d\alpha}\frac{\Gamma'(\alpha)}{\Gamma(\alpha)},\quad \text{Kalbfleisch and Prentice (2002)}$$

Simplification of $\frac{\partial}{\partial\theta}\ln\bar{F}(x|\alpha,\theta)$ and $\frac{\partial}{\partial\alpha}\ln\bar{F}(x|\alpha,\theta)$ are provided below:

$$\frac{\partial}{\partial\theta} \ln \bar{F}(x|\alpha, \theta) = \frac{1}{\Gamma\left(\alpha, \frac{x}{\theta}\right)} \cdot \frac{\partial}{\partial\theta}\, \Gamma\left(\alpha, \frac{x}{\theta}\right)$$

Recall that the upper incomplete Gamma function is defined as:

$$\Gamma\left(\alpha, \frac{x}{\theta}\right) = \int_{\frac{x}{\theta}}^{\infty} t^{\alpha-1}\ e^{-t} dt, \quad \text{let } u = \frac{x}{\theta},\ \frac{du}{d\theta} = -\frac{x}{\theta^2}$$

Applying the Leibniz integral rule which says:

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(x,t)dt = f(x,b(x)) \cdot \frac{d}{dx}b(x) - f(x,a(x)) \cdot \frac{d}{dx}a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)dt$$

since $u = u(\theta)$, we get

$$\frac{d}{d\theta}\int_{u(\theta)}^{\infty} h(t)dt = -h(u(\theta))\,\frac{d\,u(\theta)}{d\theta}.$$

In this case,

$$h(t) = t^{\alpha-1}e^{-t}, \text{ and } \frac{du}{d\theta} = -\frac{x}{\theta^2}$$

That implies,

$$\frac{\partial}{\partial\theta}\,\Gamma\left(\alpha, \frac{x}{\theta}\right) = -\left(-\frac{x}{\theta^2}\right) \cdot \left(\frac{x}{\theta}\right)^{\alpha-1}\ e^{-\frac{x}{\theta}} = \frac{x}{\theta^2} \cdot \left(\frac{x}{\theta}\right)^{\alpha-1}\ e^{-\frac{x}{\theta}} \ = \ \frac{x^\alpha}{\theta^{\alpha+1}} \cdot e^{-\frac{x}{\theta}}$$

Hence,

$$\frac{\partial}{\partial\theta}\ln \bar{F}(x|\alpha,\theta) = \frac{1}{\Gamma\left(\alpha, \frac{x}{\theta}\right)} \cdot \frac{x^\alpha}{\theta^{\alpha+1}} \cdot e^{-\frac{x}{\theta}}$$

$$\frac{\partial}{\partial\alpha}\ln \bar{F}(x|\alpha,\theta) = \frac{\partial}{\partial\alpha}\ln\Gamma\left(\alpha, \frac{x}{\theta}\right) - \frac{\partial}{\partial\alpha}\ln\Gamma(\alpha)$$

$$\frac{\partial}{\partial\alpha}\ln \bar{F}(x|\alpha,\theta) = \frac{1}{\Gamma\left(\alpha, \frac{x}{\theta}\right)} \cdot \frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{x}{\theta}\right) - \psi(\alpha)$$

$$\text{We see that, } \ \frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{x}{\theta}\right) = \int_{\frac{x}{\theta}}^{\infty} \frac{\partial}{\partial\alpha} t^{\alpha-1}\ e^{-t} dt = \int_{\frac{x}{\theta}}^{\infty} t^{\alpha-1} \ln t \cdot e^{-t} dt$$

For $f(x \mid \alpha,\theta) = \dfrac{x^{\alpha-1}\ e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\ \theta^\alpha}$, we only consider Y to be Pareto with shape $= \lambda$ and scale $= 1.$, i.e. $g(y) = \dfrac{\lambda}{(1+y)^{\lambda+1}}$ and $\bar{G}(y) = \dfrac{1}{(1+y)^\lambda}$. The information matrix for $X \sim \text{Gamma}(\alpha,\theta)$ in the non-censored case, $\mathcal{I}(\alpha,\theta)$ is known and given as

$$\mathcal{I}(\alpha,\theta) = \begin{bmatrix} \psi^{(1)}(\alpha) & \theta^{-1} \\ \theta^{-1} & \alpha\theta^{-2} \end{bmatrix},$$

where $\psi^{(1)}(\alpha)$ is the trigamma function, the first derivative of the digamma function $\psi(\alpha)$, see

Lehmann and Casella (2006). Some relevant integrals include:

$$
\begin{aligned}
[\sigma^2(\boldsymbol{\theta})]_{11} &= \int_0^\infty (\ln x - \psi(\alpha) - \ln\theta)^2 \cdot \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha}(1+x)^\lambda \, dx \\
&\quad - \int_0^\infty \left(\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{y}{\theta}\right) - \Gamma\left(\alpha, \frac{y}{\theta}\right)\psi(\alpha)}{\Gamma(\alpha)} \right)^2 \cdot \frac{\Gamma(\alpha)}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot \frac{\lambda\,(1+y)^{2\lambda}}{(1+y)^{\lambda+1}} \, dy \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty (\ln x - \psi(\alpha) - \ln\theta)^2 \cdot \; x^{\alpha-1}\, e^{-\frac{x}{\theta}}\, (1+x)^\lambda \, dx \\
&\quad - \frac{\lambda}{\Gamma(\alpha)} \int_0^\infty \frac{\left(\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{y}{\theta}\right) - \Gamma\left(\alpha, \frac{y}{\theta}\right)\psi(\alpha)\right)^2}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot (1+y)^{\lambda-1} \, dy
\end{aligned}
$$

$$
\begin{aligned}
[\sigma^2(\boldsymbol{\theta})]_{12} &= \int_0^\infty \left(\frac{x}{\theta^2} - \frac{\alpha}{\theta}\right)(\ln x - \psi(\alpha) - \ln\theta) \cdot \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha}(1+x)^\lambda \, dx \\
&\quad - \int_0^\infty \frac{ye^{-\frac{y}{\theta}}}{\Gamma(\alpha)\,\theta^2} \cdot \left(\frac{y}{\theta}\right)^{\alpha-1} \cdot \frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{y}{\theta}\right) - \Gamma\left(\alpha, \frac{y}{\theta}\right)\psi(\alpha)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot \frac{\lambda\,(1+y)^{2\lambda}}{(1+y)^{\lambda+1}} \, dy \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty \left(\frac{x}{\theta^2} - \frac{\alpha}{\theta}\right)(\ln x - \psi(\alpha) - \ln\theta) \cdot \; x^{\alpha-1}\, e^{-\frac{x}{\theta}}\, (1+x)^\lambda \, dx \\
&\quad - \frac{\lambda}{\Gamma(\alpha)\theta^{\alpha+1}} \int_0^\infty y^\alpha\, e^{-\frac{y}{\theta}} \cdot \frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{y}{\theta}\right) - \Gamma\left(\alpha, \frac{y}{\theta}\right)\psi(\alpha)}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot (1+y)^{\lambda-1} \, dy
\end{aligned}
$$

$$
\begin{aligned}
[\sigma^2(\boldsymbol{\theta})]_{22} &= \int_0^\infty \left(\frac{x}{\theta^2} - \frac{\alpha}{\theta}\right)^2 \cdot \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha}(1+x)^\lambda \, dx \\
&\quad - \int_0^\infty \left(\frac{1}{\Gamma(\alpha)} \cdot \frac{-y}{\theta^2} \left(\frac{y}{\theta}\right)^{\alpha-1} e^{-\frac{y}{\theta}} \right)^2 \cdot \frac{\Gamma(\alpha)}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot (1+y)^{2\lambda} \cdot \frac{\lambda}{(1+y)^{\lambda+1}} \, dy \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty \left(\frac{x}{\theta^2} - \frac{\alpha}{\theta}\right)^2 \cdot \; x^{\alpha-1}\, e^{-\frac{x}{\theta}}\, (1+x)^\lambda \, dx \\
&\quad - \frac{\lambda}{\Gamma(\alpha)\theta^{2\alpha+2}} \int_0^\infty y^{2\alpha}\, e^{-\frac{2y}{\theta}} \cdot \frac{(1+y)^{\lambda-1}}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \, dy
\end{aligned}
$$

For the parametric likelihood,

$$
\begin{aligned}
[\mathcal{I}_{RC}(\boldsymbol{\theta})]_{11} &= \int_0^\infty (\ln x - \psi(\alpha) - \ln\theta)^2 \cdot \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha\,(1+x)^\lambda} \, dx \\
&\quad + \int_0^\infty \left(\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{y}{\theta}\right) - \Gamma\left(\alpha, \frac{y}{\theta}\right)\psi(\alpha)}{\Gamma(\alpha)} \right)^2 \cdot \frac{\Gamma(\alpha)}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot \frac{\lambda}{(1+y)^{\lambda+1}} \, dy \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha} \int_0^\infty (\ln x - \psi(\alpha) - \ln\theta)^2 \cdot \; \frac{x^{\alpha-1}\, e^{-\frac{x}{\theta}}}{(1+x)^\lambda} \, dx \\
&\quad + \frac{\lambda}{\Gamma(\alpha)} \int_0^\infty \frac{\left(\frac{\partial}{\partial\alpha}\Gamma\left(\alpha, \frac{y}{\theta}\right) - \Gamma\left(\alpha, \frac{y}{\theta}\right)\psi(\alpha)\right)^2}{\Gamma\left(\alpha, \frac{y}{\theta}\right)} \cdot \frac{1}{(1+y)^{\lambda+1}} \, dy
\end{aligned}
$$

$$
\begin{aligned}
[\mathcal{I}_{RC}(\boldsymbol{\theta})]_{12} &= \int_0^\infty \left(\frac{x}{\theta^2}-\frac{\alpha}{\theta}\right)(\ln x-\psi(\alpha)-\ln\theta)\cdot\ \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha\,(1+x)^\lambda}\,dx \\
&+\int_0^\infty \frac{ye^{-\frac{y}{\theta}}}{\Gamma(\alpha)\,\theta^2}\cdot\left(\frac{y}{\theta}\right)^{\alpha-1}\cdot\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha,\frac{y}{\theta}\right)-\Gamma\left(\alpha,\frac{y}{\theta}\right)\psi(\alpha)}{\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha)}{\Gamma\left(\alpha,\frac{y}{\theta}\right)}\cdot\frac{\lambda}{(1+y)^{\lambda+1}}\,dy \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha}\int_0^\infty \left(\frac{x}{\theta^2}-\frac{\alpha}{\theta}\right)(\ln x-\psi(\alpha)-\ln\theta)\cdot\ \frac{x^{\alpha-1}\,e^{-\frac{x}{\theta}}}{(1+x)^\lambda}\,dx \\
&+\frac{\lambda}{\Gamma(\alpha)\theta^{\alpha+1}}\int_0^\infty y^\alpha\,e^{-\frac{y}{\theta}}\cdot\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha,\frac{y}{\theta}\right)-\Gamma\left(\alpha,\frac{y}{\theta}\right)\psi(\alpha)}{\Gamma\left(\alpha,\frac{y}{\theta}\right)}\cdot\frac{1}{(1+y)^{\lambda+1}}\,dy
\end{aligned}
$$

$$
\begin{aligned}
[\mathcal{I}_{RC}(\boldsymbol{\theta})]_{22} &= \int_0^\infty \left(\frac{x}{\theta^2}-\frac{\alpha}{\theta}\right)^2\cdot\ \frac{x^{\alpha-1}e^{-\frac{x}{\theta}}}{\Gamma(\alpha)\theta^\alpha\,(1+x)^\lambda}\,dx \\
&+\int_0^\infty \left(\frac{1}{\Gamma(\alpha)}\cdot\frac{y}{\theta^2}\,\left(\frac{y}{\theta}\right)^{\alpha-1}e^{-\frac{y}{\theta}}\right)^2\cdot\frac{\Gamma(\alpha)}{\Gamma\left(\alpha,\frac{y}{\theta}\right)}\cdot\frac{\lambda}{(1+y)^{\lambda+1}}\,dy \\
&= \frac{1}{\Gamma(\alpha)\theta^\alpha}\int_0^\infty \left(\frac{x}{\theta^2}-\frac{\alpha}{\theta}\right)^2\cdot\ \frac{x^{\alpha-1}\,e^{-\frac{x}{\theta}}}{(1+x)^\lambda}\,dx \\
&+\frac{\lambda}{\Gamma(\alpha)\theta^{2\alpha+2}}\int_0^\infty y^{2\alpha}\,e^{-\frac{2y}{\theta}}\cdot\frac{1}{\Gamma\left(\alpha,\frac{y}{\theta}\right)\,(1+y)^{\lambda+1}}\,dy
\end{aligned}
$$

and censoring rate is also given as

$$
\begin{aligned}
\mathbb{P}(X>Y) &= \int \bar{F}(y\mid\alpha,\theta)\,g(y)\,dy \\
&= \int_0^\infty \frac{\Gamma\left(\alpha,\frac{y}{\theta}\right)}{\Gamma(\alpha)}\cdot\frac{\lambda}{(1+y)^{\lambda+1}}\,dy \\
&= \frac{\lambda}{\Gamma(\alpha)}\int_0^\infty \frac{\Gamma\left(\alpha,\frac{y}{\theta}\right)}{(1+y)^{\lambda+1}}dy
\end{aligned}
$$

Table 3.3 below shows results of the ellipses obtained from the theoretical computation of the variance-covariance matrices of the likelihoods for selected values of $\alpha,\ \theta$, and λ.

Table 3.3: Numerical Results from Gamma Distributed Event at significance level = 0.01

α	θ	λ	Parametric Ellipse Area		Modified Ellipse Area		Censoring Rate
			Theoretical	Estimated	Theoretical	Estimated	
0.5	0.1	0.5	2.439091	2.448590	2.440983	2.450497	0.02264331
0.5	0.1	1	2.488794	2.536911	2.496540	2.545101	0.04391339
1	0.1	0.5	3.729703	3.714726	3.734178	3.719167	0.04391339
1	0.1	1	3.856470	3.865999	3.875031	3.884613	0.08436666
0.2	0.2	0.5	2.877299	2.860837	2.882318	2.864727	0.01728026
0.2	0.2	1	2.945821	3.042820	2.966536	3.065323	0.03308677
0.5	0.25	0.5	6.257164	6.210355	6.278819	6.229643	0.05039196
0.5	0.25	1	6.536152	6.660282	6.627420	6.749412	0.09464590
0.1	0.1	0.5	0.967690	1.035191	0.968106	1.035714	0.00464514
0.1	0.1	1	0.978154	0.933697	0.979930	0.935235	0.00907924

The confidence ellipses generated by the two estimators were highly comparable at low censoring rates. However, as the censoring rate increased, the ellipses corresponding to the modified estimator exhibited substantially larger areas than those of the parametric estimator. The comparability at low censoring rates was consistently observed across all (α, θ) pairs evaluated at $\lambda = 0.5$ and 1. Graphical illustrations of some sampled cases are presented below.

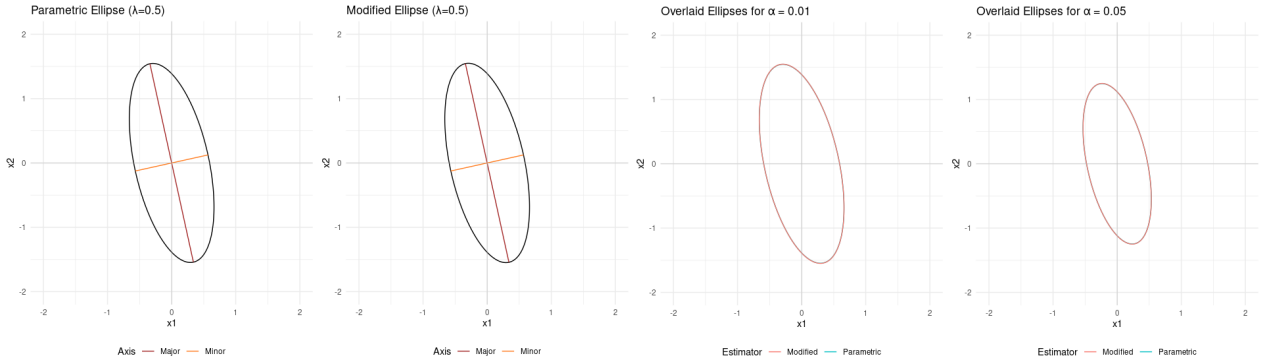

Figure 3.5: Theoretical Ellipse Comparison for $Gamma(\alpha = 0.2,\ \theta = 0.2)$ at $\lambda = 0.5$

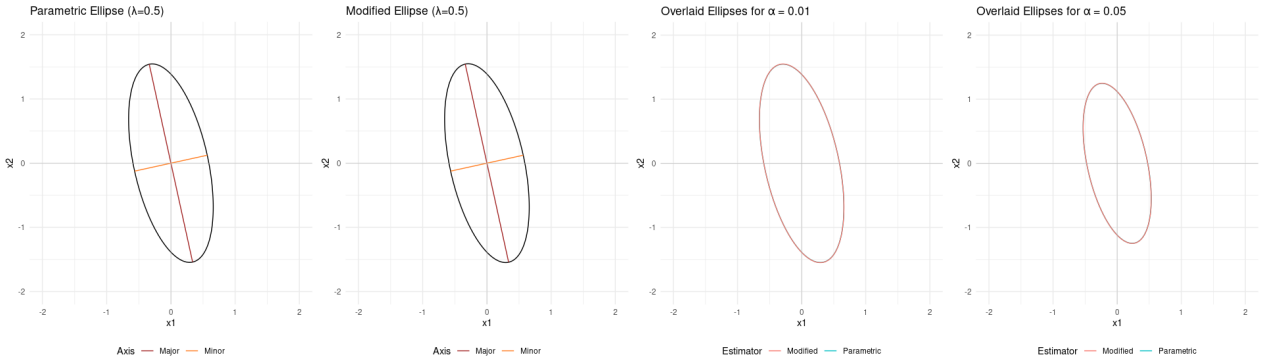

Figure 3.6: Estimated Ellipse Comparison for $Gamma(\alpha = 0.2,\ \theta = 0.2)$ at $\lambda = 0.5$

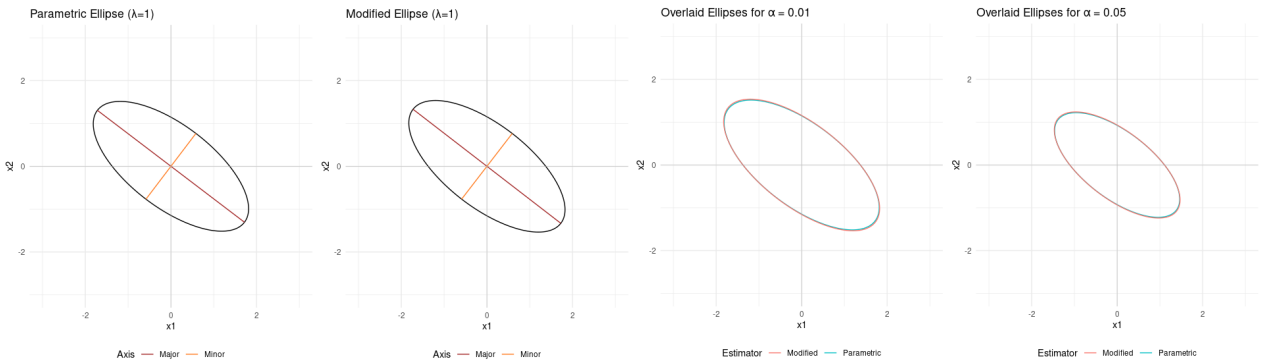

Figure 3.7: Theoretical Ellipse Comparison for $Gamma(\alpha = 0.5,\ \theta = 0.25)$ at $\lambda = 1$

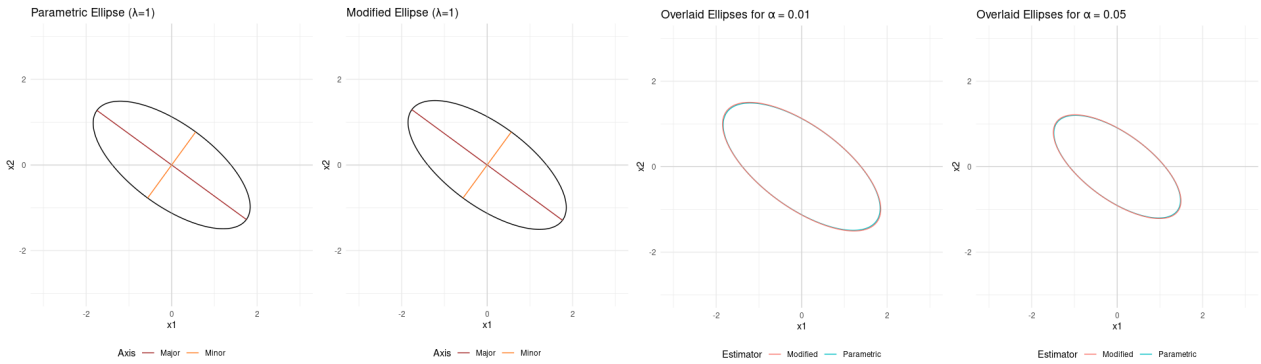

Figure 3.8: Estimated Ellipse Comparison for $Gamma(\alpha = 0.5,\ \theta = 0.25)$ at $\lambda = 1$

In Figures 3.5 - 3.8, the ellipses largely overlap such that there is minimal to no distinction between the interior regions. A negative correlation between the shape and scale parameters was also evident across all parameter combinations examined. In addition, the maximum likelihood estimates and their corresponding mean squared errors are presented in Table 3.4. The derivations for the MLEs are outlined below.

A closed-form expression was obtained only for $\hat{\theta}^M$, while the remaining estimators required numerical methods for their computation.

Semi-parametric MLE

For $\hat{\alpha}$

$$\sum_{i=1}^{n} W_{in} \frac{\partial}{\partial \alpha} \ln f(z_i \mid \boldsymbol{\theta}) = 0$$

$$\sum_{i=1}^{n} W_{in} \left(\ln z_i - \psi(\alpha) - \ln \theta\right) = 0$$

$$\psi(\alpha) \sum_{i=1}^{n} W_{in} = \sum_{i=1}^{n} W_{in} \ln z_i - \ln \theta \sum_{i=1}^{n} W_{in}$$

For $\hat{\theta}$

$$\sum_{i=1}^{n} W_{in} \left(\frac{z_i}{\theta^2} - \frac{\alpha}{\theta}\right) = 0$$

$$\frac{1}{\theta^2} \sum_{i=1}^{n} W_{in}\, z_i = \frac{\alpha}{\theta} \sum_{i=1}^{n} W_{in}$$

$$\hat{\theta}^M = \frac{\sum_{i=1}^{n} W_{in}\, z_i}{\alpha \sum_{i=1}^{n} W_{in}}$$

Parametric MLE

For $\hat{\alpha}$

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \left[\delta_i \ln f(z_i \mid \boldsymbol{\theta}) + (1 - \delta_i) \ln \bar{F}(z_i \mid \boldsymbol{\theta})\right] = 0$$

$$\sum_{i=1}^{n} \left[\delta_i \left(\ln z_i - \psi(\alpha) - \ln \theta\right) + (1 - \delta_i) \left(\frac{\partial}{\partial \alpha} \ln \Gamma\left(\alpha, \frac{z_i}{\theta}\right) - \psi(\alpha)\right)\right] = 0$$

$$\sum_{i=1}^{n}\delta_i \ln z_i - \psi(\alpha)\sum_{i=1}^{n}\delta_i - \ln\theta\sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha,\frac{z_i}{\theta}\right)}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} - \sum_{i=1}^{n}\psi(\alpha) - \sum_{i=1}^{n}\delta_i\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha,\frac{z_i}{\theta}\right)}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} + \psi(\alpha)\sum_{i=1}^{n}\delta_i = 0$$

$$\ln\theta\sum_{i=1}^{n}\delta_i - \sum_{i=1}^{n}\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha,\frac{z_i}{\theta}\right)}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} + \sum_{i=1}^{n}\delta_i\frac{\frac{\partial}{\partial\alpha}\Gamma\left(\alpha,\frac{z_i}{\theta}\right)}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} = \sum_{i=1}^{n}\delta_i \ln z_i - n\psi(\alpha)$$

For $\hat{\theta}$

$$\sum_{i=1}^{n}\left[\delta_i\left(\frac{z_i}{\theta^2}-\frac{\alpha}{\theta}\right) + (1-\delta_i)\left(\frac{z_i}{\theta^2}\right)\left(\frac{z_i}{\theta}\right)^{\alpha-1}\cdot\frac{e^{-\frac{z_i}{\theta}}}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)}\right] = 0$$

$$\frac{1}{\theta^2}\sum_{i=1}^{n}\delta_i z_i - \frac{\alpha}{\theta}\sum_{i=1}^{n}\delta_i + \sum_{i=1}^{n}\frac{z_i^{\alpha}}{\theta^{\alpha+1}}\cdot\frac{e^{-\frac{z_i}{\theta}}}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} - \sum_{i=1}^{n}\delta_i\frac{z_i^{\alpha}}{\theta^{\alpha+1}}\cdot\frac{e^{-\frac{z_i}{\theta}}}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} = 0$$

$$\frac{1}{\theta}\sum_{i=1}^{n}\delta_i z_i + \frac{1}{\theta^{\alpha}}\sum_{i=1}^{n}\frac{z_i^{\alpha}\,e^{-\frac{z_i}{\theta}}}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} - \frac{1}{\theta^{\alpha}}\sum_{i=1}^{n}\frac{\delta_i\,z_i^{\alpha}\,e^{-\frac{z_i}{\theta}}}{\Gamma\left(\alpha,\frac{z_i}{\theta}\right)} = \alpha\sum_{i=1}^{n}\delta_i$$

Table 3.4: MLE Results for Gamma Distributed Event at significance level = 0.01

	$\hat{\alpha}$					θ				
λ	α	$\hat{\alpha}^P$	$\hat{\alpha}^M$	MSE_P	MSE_M	θ	$\hat{\theta}^P$	$\hat{\theta}^M$	MSE_P	MSE_M
0.5	0.5	0.50287	0.50287	0.00001	0.00001	0.1	0.10004	0.10004	0.00000	0.00000
1	0.5	0.49554	0.49554	0.00002	0.00002	0.1	0.10241	0.10241	0.00001	0.00001
0.5	1	0.98954	0.98954	0.00011	0.00011	0.1	0.10024	0.10024	0.00000	0.00000
1	1	1.01155	1.01155	0.00013	0.00013	0.1	0.09954	0.09954	0.00000	0.00000
0.5	0.2	0.19947	0.19950	0.00000	0.00000	0.2	0.19917	0.19909	0.00000	0.00000
1	0.2	0.20018	0.20018	0.00000	0.00000	0.2	0.20621	0.20621	0.00000	0.00000
0.5	0.5	0.50107	0.50127	0.00000	0.00000	0.25	0.24788	0.24775	0.00000	0.00000
1	0.5	0.49110	0.49163	0.00008	0.00007	0.25	0.25719	0.25677	0.00005	0.00005
0.5	0.1	0.10190	0.10190	0.00000	0.00000	0.1	0.10583	0.10583	0.00003	0.00003
1	0.1	0.10179	0.10179	0.00000	0.00000	0.1	0.09464	0.09464	0.00003	0.00003

In Table 3.4, we observe that the estimates produced by both estimators are identical in most cases, with only minimal differences in the remaining instances. This explains the substantial overlap of the confidence ellipses observed in almost all scenarios in this case. Consequently, for small values of the parameters and λ, the variances of the estimators are almost indistinguishable.

3.4 Summary

This chapter began with an introduction to the maximum likelihood estimation of multiple parameters in the absence of random censoring, with a focus on the two-parameter case estimation. We explored the estimation of the variance-covariance matrix using confidence ellipses derived from the bivariate normal distribution. This concept was also applicable in the censored data case. Considering $X \sim Weibull(p, \theta)$ and $X \sim Gamma(\alpha, \theta)$, both paired with $Y \sim Pareto(\lambda, 1)$, we identified some triples of the scale parameter (θ), shape parameter (p or α), and the parameter of the censoring distribution (λ) where comparisons were particularly insightful.

Motivated by findings from the single-parameter case, we focused on relatively small parameter values. Thus, we examined the shape-scale pairs of the event variable at $\lambda = 0.5$ and $\lambda = 1$. For $X \sim Weibull(p, \theta)$, random parameter combinations were tested across a range of censoring rates. Among the sampled triples, the modified estimator generally demonstrated better efficiency, with the exception of $(p = 0.5, \theta = 1)$, where the estimators were comparable at $\lambda = 0.5$. Notably, the case $(p = 1, \theta = 0.1)$ yielded especially favorable results for the modified estimator (see Table 3.1 and B.3).

In the case of $X \sim Gamma(\alpha, \theta)$, high censoring rates were associated with large confidence ellipse areas for the modified estimator. As such, our comparisons focused on scenarios where $\mathbb{P}(\delta = 0) \to 0$. Across all triples examined, the estimators showed highly comparable performance, particularly at $\lambda = 0.5$. The maximum likelihood estimates obtained from simulated data showed strong similarity between the estimators, as evidenced by identical mean squared errors in most cases.

Chapter 4

Conclusion

The aim of this research was to compare two maximum likelihood estimators under random censoring. Specifically, we sought to examine the traditional maximum likelihood estimator of the fully parametric likelihood and a modified likelihood estimator constructed by replacing the censoring indicators in the parametric model with Kaplan-Meier-type weights. The study considered various estimation scenarios, including both single-parameter and multi-parameter cases. For each scenario, appropriate distributions were specified for the event and censoring variables. Generally, suitable choices were made to ensure finiteness and integrability in the variance components. The comparison was conducted from a theoretical perspective and further supported by simulated data. Therefore, variances and mean squared errors were the primary tools for evaluating and contrasting the performance of the estimators.

In the single parameter case, we observed that although the variance of the modified estimator was not consistently smaller than that of the parametric estimator, the two estimators exhibited strong comparability at lower parameter values. The most notable similarity was observed at $\lambda \leq 1$ in nearly all three cases of the event variable X. In particular, when $X \sim Beta(\theta, 1)$ and $Y \sim Pareto(\lambda, 1)$, the estimators produced identical mean squared errors for $\lambda \in \{0.5, 1, 2, 3\}$. Comparisons in the multi-parameter case were based on confidence ellipses, as the focus was on estimating $k = 2$ parameters. Remarkably, when $X \sim Weibull(\theta, p)$, the modified estimator consistently produced confidence ellipses with smaller areas than those of the parametric estimator across most of the randomly selected parameter triples at different censoring rates. This suggests

that the modified estimator produces more precise estimates than the parametric estimator at those triples. On top of that, comparable ellipses were produced under $X \sim Gamma(\alpha, \theta)$ with minimal differences in area. This similarity was further supported by the fact that both estimators produced identical maximum likelihood estimates for the shape and scale parameters.

This research posed a few challenges, particularly in selecting appropriate distributions for both the single and multi-parameter scenarios, due to the varying shapes and characteristics of the candidate distributions. Another challenge was estimating the shape parameters in the parametric models, particularly in the multi-parameter cases, where closed-form solutions were largely unattainable. Nonetheless, these challenges presented the opportunity to apply novel and numerical techniques in solving statistical problems. A limitation of this study is that, in instances where the relevant integrals were not explicitly solved, the comparisons were made at selected ranges of the parameters. Future research could extend the analysis to wider parameter ranges to provide a more comprehensive evaluation. Moreover, it would be valuable to validate these findings using real-world data. The observation of high comparability between the estimators in the case of $X \sim Gamma(\alpha, \theta)$ at very low censoring rates may also inspire future studies to explore similar comparisons in the absence of random censoring.

Appendix A

Regular Model

Let the density of X be $f(x \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n$, the parameter space. Here, X can be a vector in $\mathbb{R}^n$ or a scalar. Let $\boldsymbol{I}(\boldsymbol{\theta})$ be the $n \times n$ Fisher information matrix. From Hogg et al. (2018), a model is said to be *regular* if the following regularity conditions are satisfied:

1. The cumulative density functions are distinct, that is, for $\theta \neq \theta'$, it follows that $F(x_i \mid \theta) \neq F(x_i \mid \theta')$.

2. The probability density functions share a common support $\forall\ \theta$.

3. The point θ_0 lies in the interior of the parameter space Θ.

4. $f(x \mid \theta)$ is twice differentiable.

5. $\int f(x \mid \theta)\ dx$ is twice differentiable under the integral sign as a function of θ.

6. $\exists$ an open subset $\Theta_0 \subset \Theta$ such that $\boldsymbol{\theta}_0 \in \Theta_0$ and the third partial derivatives of $f(x \mid \boldsymbol{\theta})$ exists $\forall \boldsymbol{\theta} \in \Theta_0$.

7. The expectation of the score function is zero, and the information matrix is the negative expectation of the second derivative of the log-likelihood function.

8. $\forall \theta \in \Theta_0$, the information matrix $\boldsymbol{I}(\boldsymbol{\theta})$ is positive definite.

9. $\exists$ functions $M_{jkl}(x)$ such that $\left|\dfrac{\partial^3}{\partial\theta_j \partial\theta_k \partial\theta_l} \log f(x \mid \boldsymbol{\theta})\right| \leq M_{jkl}(x),\ \forall \theta \in \Theta_0$ and $E_{\boldsymbol{\theta}_0}(M_{jkl}) < \infty,\ \forall j, k, l \in 1, ..., n$.

Appendix B

Complementary Results

This section contains additional results and figures. Below are some graphs from the multi-parameter case.

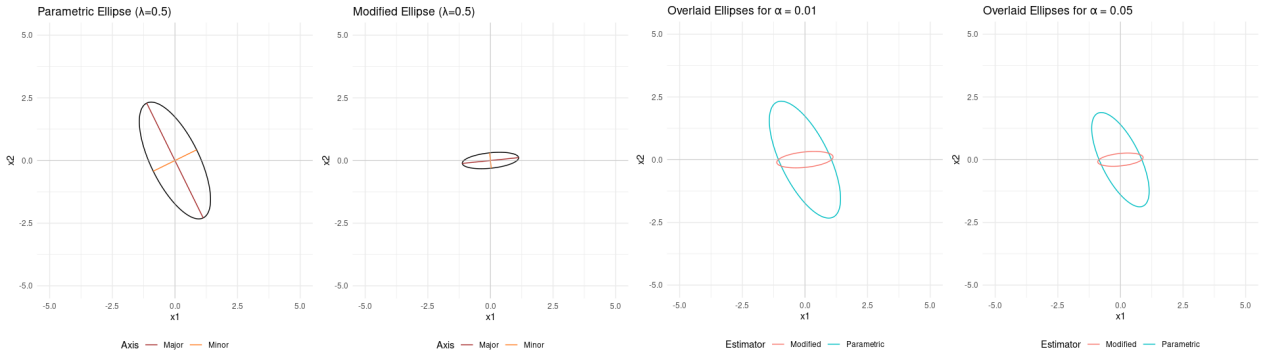

Figure B.1: Estimated Ellipse Comparison for $Weibull(\theta = 0.25, p = 0.75)$ at $\lambda = 0.5$

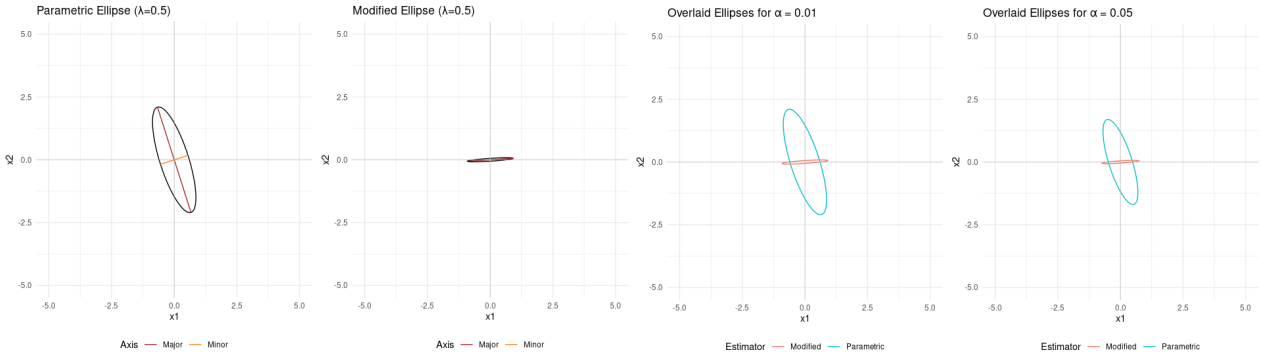

Figure B.2: Estimated Ellipse Comparison for $Weibull(\theta = 0.1, p = 0.5)$ at $\lambda = 0.5$

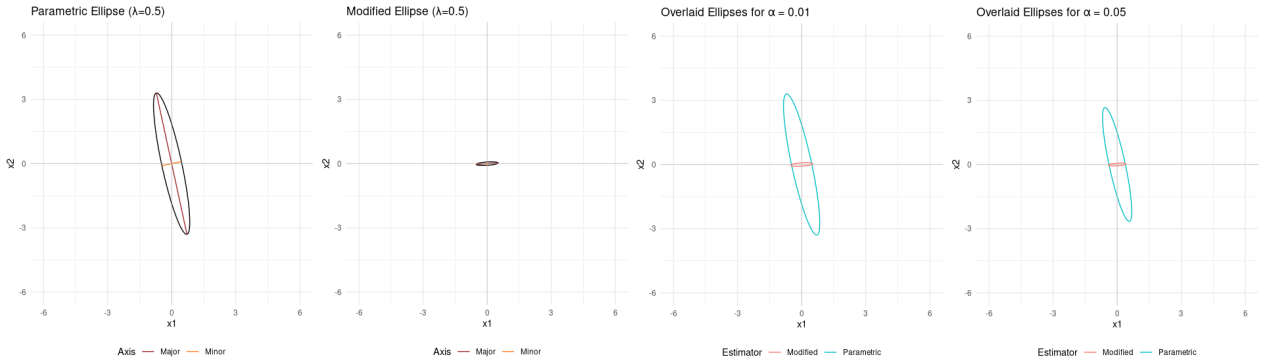

Figure B.3: Estimated Ellipse Comparison for $Weibull(\theta = 0.1, p = 1)$ at $\lambda = 0.5$

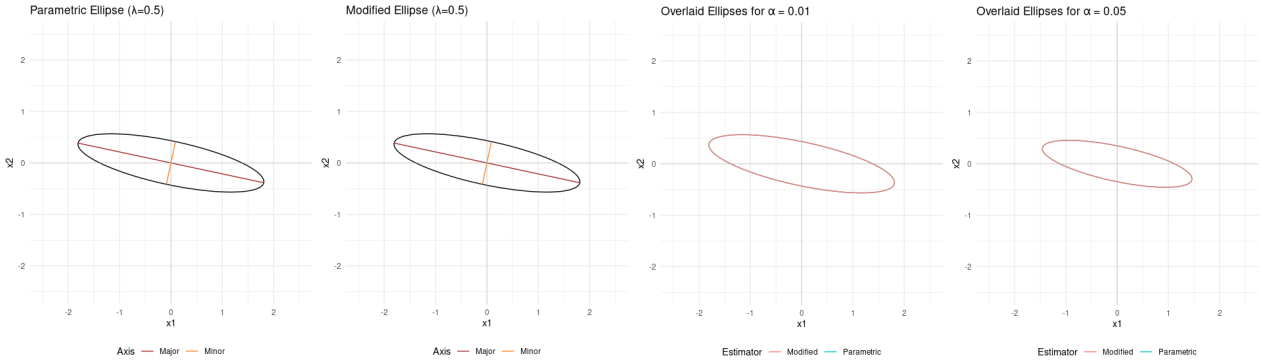

Figure B.4: Estimated Ellipse Comparison for $Gamma(\alpha = 0.5, \theta = 0.1)$ at $\lambda = 0.5$

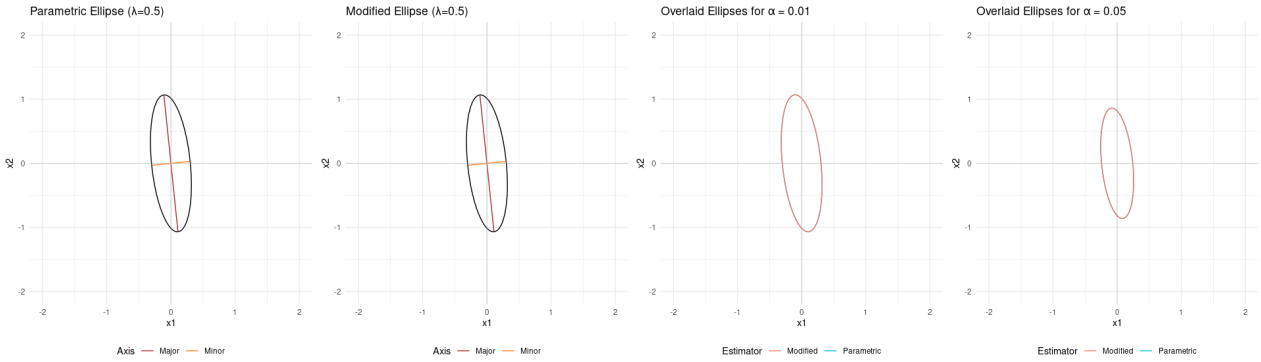

Figure B.5: Estimated Ellipse Comparison for $Gamma(\alpha = 0.1, \theta = 0.1)$ at $\lambda = 0.5$

References

Balakrishnan, N., & Kateri, M. (2008). On the maximum likelihood estimation of parameters of weibull distribution based on complete and censored data. *Statistics & Probability Letters*, *78*(17), 2971–2975.

Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes i-ii package*. Chapman and Hall/CRC.

Casella, G., & Berger, R. (2024). *Statistical inference*. CRC press.

Cramer, J. S., & Cramer, J. S. (1989). *Econometric applications of maximum likelihood methods*. CUP Archive.

DeGroot, M. H., & Schervish, M. J. (2012). *Probability and statistics* (4th ed.). Pearson.

Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2011). *Statistical distributions* (4th ed.). Wiley.

Goel, M. K., Khanna, P., & Kishore, J. (2010). Understanding survival analysis: Kaplan-meier estimate. *International journal of Ayurveda research*, *1*(4), 274.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2018). *Introduction to mathematical statistics* (8th ed.). Boston: Pearson Education.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), 457–481.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). Springer.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.

Murray, S. (2001). Using weighted kaplan–meier statistics in nonparametric comparisons of paired censored survival outcomes. *Biometrics*, *57*(2), 361–368.

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Ren, J.-J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics*.

Ren, J.-J., & Lyu, Y. (2024). Weighted empirical likelihood for accelerated life model with various types of censored data. *Stats*, *7*(3), 944–954.

Rodrıguez, G. (2005). Non-parametric estimation in survival models. *cited on*, 20.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Saleem, I., Sanaullah, A., Al-Essa, L. A., Bashir, S., & Al Mutairi, A. (2023). Efficient estimation of population variance of a sensitive variable using a new scrambling response model. *Scientific Reports*, *13*, 19913.

Smith, M. J., Phillips, R. V., Luque-Fernandez, M. A., & Maringe, C. (2023). Application of targeted maximum likelihood estimation in public health and epidemiological studies: a systematic review. *Annals of epidemiology*, *86*, 34–48.

Stute, W. (1995). The central limit theorem under random censorship. *The Annals of Statistics*, 422–439.

Turkson, A. J., Ayiah-Mensah, F., & Nimoh, V. (2021). Handling censoring and censored data in survival analysis: a standalone systematic literature review. *International journal of mathematics and mathematical sciences*, *2021*(1), 9307475.

Walck, C., et al. (2007). *Hand-book on statistical distributions for experimentalists*. Stockholms universitet.

Zhou, M., & Liang, K.-Y. (2011). Semi-parametric hybrid empirical likelihood inference for two-sample comparison with censored data. *Lifetime Data Analysis*, *17*, 533–551.