Risk Analysis of Life-time Acceptance Sampling Plans under Model Uncertainties

A thesis submitted for the degree of

Master of Science

by

Ruizhe Yang

School of Fundamental Sciences (SFS)

Massey University

New Zealand

April 2020

Contents

Acknowledgements

First of all, I would like to express my profound gratitude to Dr. Xun Xiao, my main supervisor, who gives me support. I am very grateful for his guidance and encouragement. His efforts to explain things clearly and simply, his constant optimism and immense patience are essential for my project. Furthermore, I sincerely appreciate his revision of this thesis.

I am deeply thankful to my co-supervisor, Dr. Kondaswamy Govindaraju, for providing all the necessary means for writing codes, scientific guidance, good advice, cheerful mood, and support whenever needed. I also deeply appreciate his revision of this thesis.

I would also like to thank my parents for their unconditional support during my graduate study. Thank you for your spiritual support and company. When I wanted to give up, it was you who gave me encouragement. You say that no matter what decision I make, you will always be on my side. Thank you very much, I love you forever.

I deeply appreciate and thank the poor performance of Arsenal in recent seasons. Let me abandon impetuousness and concentrate on learning and research.

I would like to thank my friends, Yue Wang, Yujie Liu, Qingyu Chen, Qianyu Zhang, Bingxue Zhao, Minjia Zhu, Jingyu Guo, Xintong Lu and Haoxiang Yin for their emotional support. Thanks to jet lag, we can keep in touch even in the middle of the night.

Last but not the least, I would like to thank my boyfriend, who has never appeared in my life for more than 25 years, allowing me to concentrate on academics and complete my graduation thesis.

Statement of Originality

This is to certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Ruizhe Yang

Abstract

Lifetime acceptance sampling is one of the important branches of quality engineering because lifetime is a critical characteristic of many industrial and agricultural products. Due to budget and time constraints, lifetime acceptance sampling plans usually suffer from the curse of small sample sizes. Given a sufficiently large sample size, the test or sample data can identify its parent distribution easily. However, it is a challenge to find out the parent distribution for small sample sizes, especially when the data comes from a lifetime distribution having a shape parameter. In this thesis, I propose a variables sampling plan, called the M-method plan, to resolve the distribution-data identification issue in lifetime acceptance sampling with small sample sizes. Extensive Monte Carlo simulation studies were carried out to compare the Operating Characteristic (OC) curves of the M-method plans, and two existing alternative plans. Furthermore, I show that the lognormal distribution, which is a shape free lifetime distribution, can be used as a surrogate for Weibull or gamma distributions when the sample size is small. In other words, model uncertainties can be ignored when designing a lifetime acceptance sampling plan under the M-method. The M-method based sampling plan, under the correctly-specified distribution, is compared with various M-method based sampling plans under the scenario of misspecified distributions. Even though the OC curves are distinct from each other significantly depending on the operating procedure, the OC curves can be matched under the proposed method when the parent distribution is fully misspecified as a lognormal distribution for small sample sizes.

Key words: lifetime acceptance sampling plan; model misspecification; lognormal distribution; Weibull distribution; gamma distribution; small sample size; censoring

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

The term *lifetime distributions* refers to the collection of the probability distributions of non-negative random variables that can be used in lifetime data analysis, biomedical sciences, and reliability engineering. According to Lawless (2003), lifetime distributions are employed to analyze lifetime data, survival time, and failure time in the study of drug durability and human diseases. Although lifetime data can be modeled by using many different lifetime distributions, we will focus on three common lifetime distributions in this thesis: the lognormal, Weibull, and gamma distributions.

Lognormal Distribution

The lognormal distribution is typically used to simulate the repair time of maintainable systems in reliability analysis (Lawless, 2003). Moreover, it can be used in biology and medicine, such as modeling the measurement of living tissue size in problems of relative growth (Julian, 1932) and modeling the number of hospitalized cases when an epidemic occurs (Wang et al., 2013).

The random variable X is said to be lognormally distributed if $\log(X)$ is normally distributed with the mean μ and variance σ^2. It can be denoted as $X \sim \mathrm{LN}(\mu, \sigma^2)$. The probability density function (PDF) is given by

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)}, \;\; x > 0,$$

where $\sigma > 0$. The cumulative distribution function (CDF) is

$$F_X(x) = \Phi\left((\ln x - \mu)/\sigma\right), \;\; x > 0,$$

where $\Phi()$ is the CDF of the standard normal distribution. The quantile and survival functions are given by

$$Q_p(p; \mu, \sigma^2) = \exp(\mu + \sigma\Phi^{-1}(p)), \;\; 0 \leq p < 1,$$

$$S_X(x) = 1 - \Phi\left((\ln x - \mu)/\sigma\right), \;\; x > 0,$$

where $\Phi^{-1}()$ is the quantile function of the standard normal distribution. The mean and variance of X are given by

$$\mathrm{E(X)} = \exp\left(\mu + \sigma^2/2\right),$$

$$\mathrm{Var(X)} = \left[\exp\left(\sigma^2\right) - 1\right] \exp\left(2\mu + \sigma^2\right).$$

The mode of X is $\mathrm{Mode(X)} = \exp(\mu - \sigma^2)$, and the coefficient of variation (CV) is equal to $\mathrm{CV(X)} = \sqrt{\exp(\sigma^2) - 1}$.

For a lognormally distributed random variable X, the parameters μ and σ are not location and scale parameters because they are the location and scale parameters of $\log(X)$, which follows a normal distribution. It is important to note that the lognormal distribution has no shape parameter.

Given the values of parameters $\mu = 0.25$ and $\sigma = 0.40, 1.00, 1.15$, the curves of the PDF, CDF, hazard function (HF), and cumulative hazard function (CHF) of the lognormal distribution are shown in Figure 1.1. It can be observed that the curves of $\sigma < 1$ are significantly different from the curves of $\sigma = 1$ and $\sigma > 1$ when the parameter μ is kept the same.

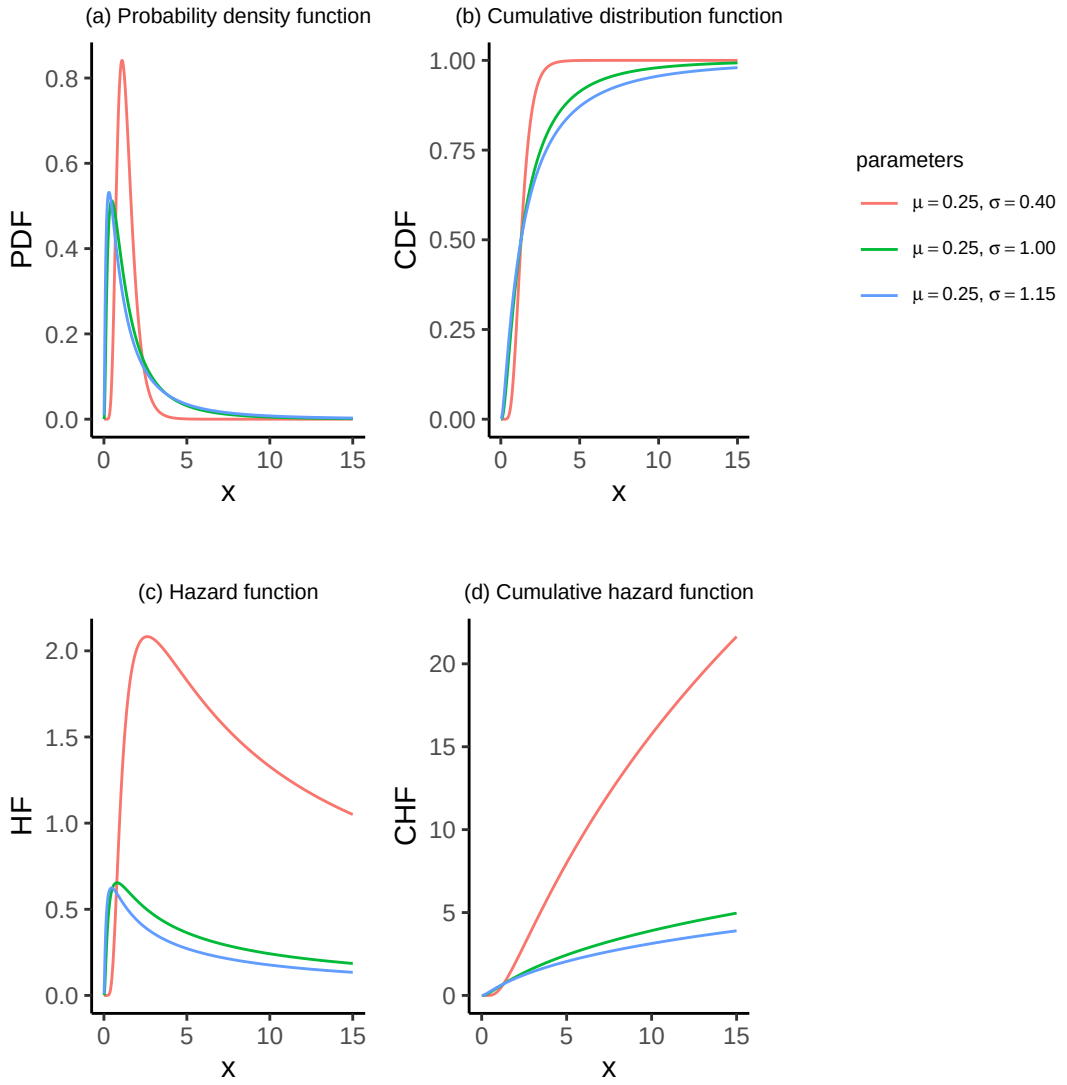

Figure 1.1: *Functions of the lognormal distribution*

Weibull Distribution

The Weibull distribution probably is the most-widely used distribution in reliability analysis. It is also used for survival analysis and communication systems (Lawless, 2003).

A random variable Y which follows a Weibull distribution with the shape parameter k and scale parameter λ can be denoted as $Y \sim \text{Wei}(k, \lambda)$. The PDF and CDF of Weibull distribution are

$$f_Y(y) = (k/\lambda)\,(y/\lambda)^{k-1}\, e^{\left(-\left(\frac{y}{\lambda}\right)^k\right)},\ \ y \geq 0,$$

$$F_Y(y) = 1 - \exp(-(y/\lambda)^k),\ \ y \geq 0.$$

It also includes the exponential distribution as a special case if $k = 1$. The mean and variance can be expressed as

$$\mathrm{E}(\mathrm{Y}) = \lambda\Gamma\left(1 + 1/k\right),$$

$$\mathrm{Var}(\mathrm{Y}) = \lambda^2\left[\Gamma\left(1 + 2/k\right) - \left(\Gamma\left(1 + 1/k\right)\right)^2\right],$$

where $\Gamma()$ is the gamma function. The mode is $\mathrm{Mode}(\mathrm{Y}) = \lambda\left[(k-1)/k\right]^{1/k}$, and the CV is equal to

$$\mathrm{CV}(\mathrm{Y}) = (2/k\Gamma(2/k) - \Gamma(1 + 1/k)^2)^{1/2}(1/k\Gamma(1/k))^{-1}.$$

Furthermore, the quantile function, survival function, and hazard function of the Weibull distribution are given as follows

$$Q_p(p; k, \lambda) = \lambda(-\ln(1-p))^{1/k}, \ \ 0 \leq p < 1,$$

$$S_Y(y) = \exp(-(y/\lambda)^k), \ \ y > 0,$$

$$h_Y(y) = (k/\lambda)\,(y/\lambda)^{k-1}, \ \ y > 0.$$

If the quantity Y is the "failure time", the Weibull distribution gives a distribution in which the failure rate is proportional to a power of life. The shape parameter k is equal to that power plus one. The failure rate is decreasing if $k < 1$, increasing if $k > 1$, and constant if $k = 1$.

In medical statistics, the parameterization is different from those mentioned above. $b = \lambda^{-k}$ is used to represent the scale parameter, and the shape parameter k remains the same. In addition, the parameter b is called the characteristic life in engineering.

Given the values of the shape parameters $k = 0.70, 1.00, 1.40$, and scale parameter $\lambda = 2.00$, the curves of the PDF, CDF, HF, and CHF of the Weibull distribution are shown in Figure 1.2. Under the same scale parameter λ, it can be seen that the hazard function is increasing if $k > 1$, decreasing if $k < 1$ and keeping a straight line if $k = 1$. In addition, the probability density function of $k < 1$ is apparently different from $k = 1$ and $k > 1$.

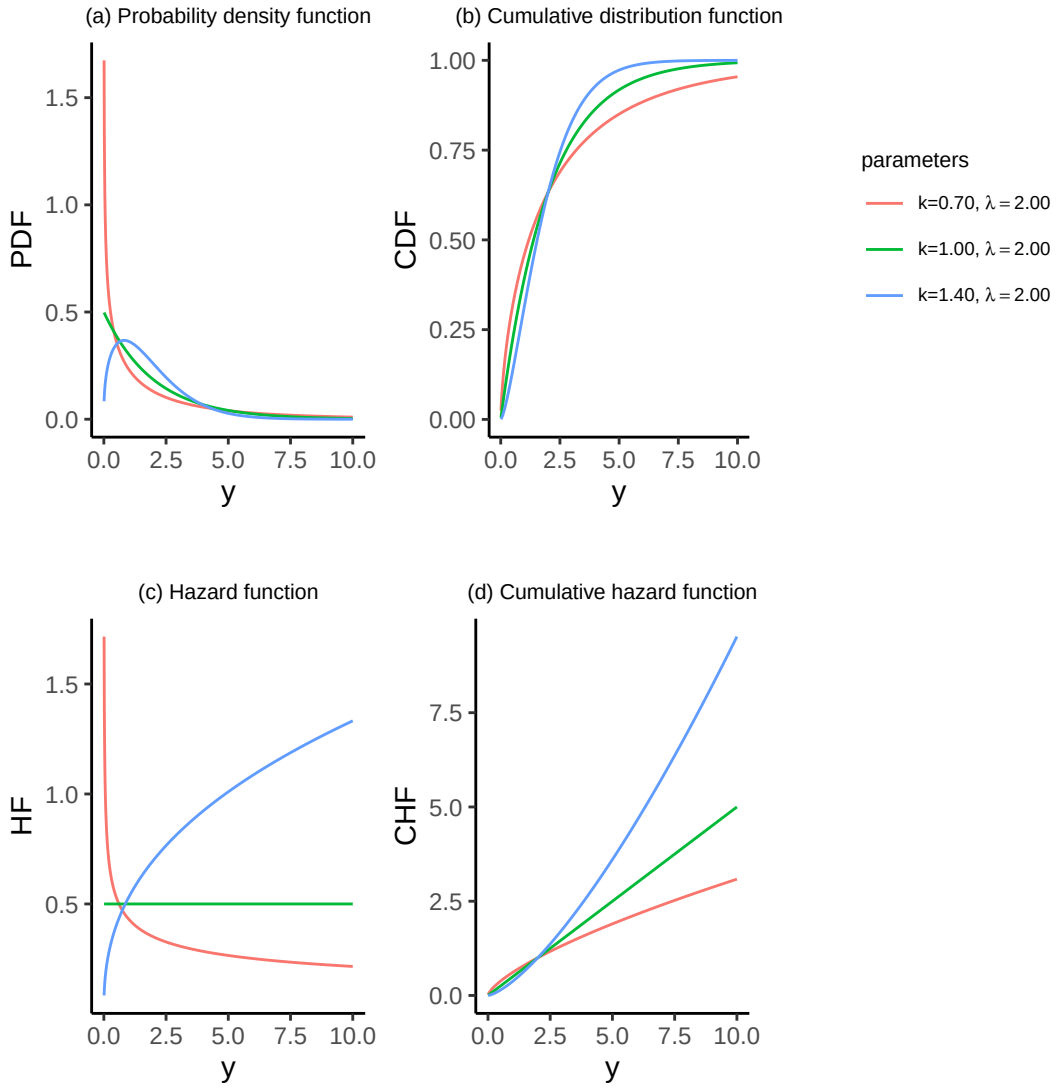

Figure 1.2: *Functions of the Weibull distribution*

Gamma Distribution

The gamma distribution is used to simulate insurance claims and rainfall (Aksoy, 2000; Boland, 2007). It is also used in medical applications, such as modeling the age distribution of cancer incidence in oncology with the gamma distribution. Although the gamma distribution is less popular as a lifetime model than the lognormal and Weibull distributions, it is also suitable for various lifetime data. Particularly, the gamma distribution is often used in cases involving the sum of exponential distributions because the sum of

independently and identically distributed (i.i.d.) exponential random variables follows a gamma distribution.

A random variable Z which follows a gamma distribution with shape parameter ω and scale parameter θ can be denoted as $Z \sim Gamma(\omega, \theta)$. However, the gamma distribution is usually parameterized by the rate parameter $\psi = 1/\theta$ and shape parameter ω in Bayesian statistics. It also has another type of parameterization with the same shape parameter ω and a mean parameter equal to $\omega\theta$. Like the Weibull distribution, the gamma distribution becomes an exponential distribution when the shape parameter $\omega = 1.0$.

The PDF and CDF of the gamma distribution are given by

$$f_Z(z) = z^{\omega-1}\theta^{-\omega}(\Gamma(\omega))^{-1}\exp(-z/\theta), \;\; z \geq 0,$$

$$F_Z(z) = \int_0^z f(u)\,du = \gamma\,(\omega, z/\theta)/\Gamma(\omega), \;\; z \geq 0,$$

where $\gamma\,(\omega, x/\theta)$ is the lower incomplete gamma function. The mean of Z is $\mathrm{E(Z)} = \omega\theta$, variance is $\mathrm{Var(Z)} = \omega\theta^2$, CV is equal to $\mathrm{CV(Z)} = \omega^{-1/2}$, and mode is $\mathrm{Mode(Z)} = (\omega - 1)\theta$ for $\omega \geq 1$. The quantile and survival functions are given by

$$Q_p(p; \omega, \theta) = F_Y^{-1}(p), \;\; 0 \leq p < 1,$$

$$S_Z(z) = 1 - \gamma\,(\omega, z/\theta)/\Gamma(\omega), \;\; z \geq 0.$$

In addition, the chi-squared distribution is another special case of the gamma distribution. If the shape parameter $\omega = n/2$ and scale parameter $\theta = 1/2$, the gamma distribution reduces to the chi-squared distribution with n degrees of freedom. Furthermore, the gamma distribution has the property of additivity. If $Z_i \sim$ Gamma (ω_i, θ) (i = 1,2,...,N) and all Z_i are independent, then $\sum_{i=1}^{N} Z_i \sim$ Gamma $\left(\sum_{i=1}^{N} \omega_i, \theta\right)$.

Given the values of the shape parameters $\omega = 0.80, 1.00, 2.00$ and scale parameter $\theta = 2.00$, the curves of the PDF, CDF, HF, and CHF of the gamma distribution are shown in Figure 1.3. It can be found that when the scale parameter θ remains unchanged, the hazard function will be increasing for $\omega > 1$, decreasing for $\omega < 1$ and becomes a straight line for

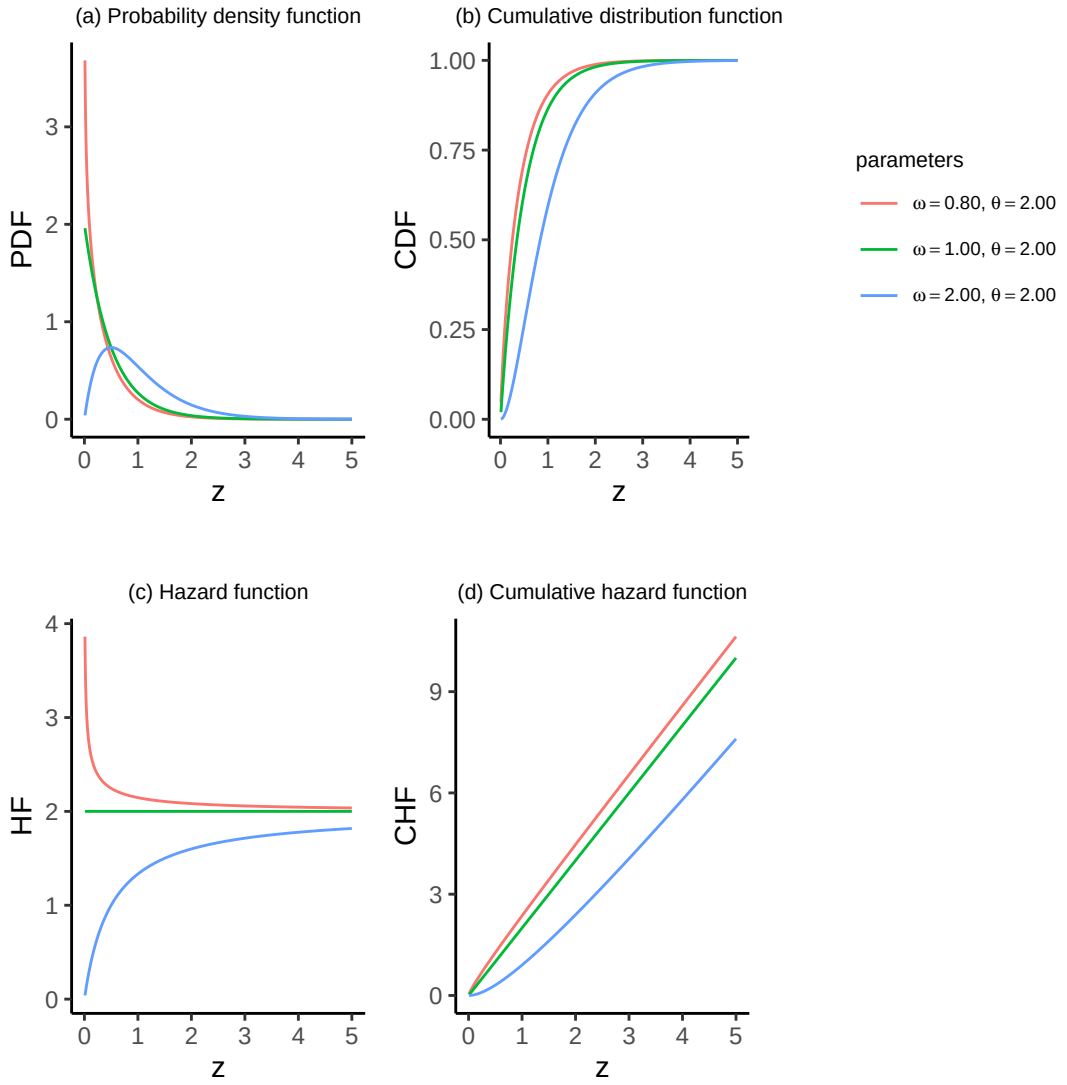

Figure 1.3: *Functions of the gamma distribution*

$\omega = 1$. In addition, the probability density function of $\omega > 1$ is apparently different from $\omega = 1$ and $\omega < 1$.

When using the lifetime distributions for modeling lifetime data, we face the distribution-data identification issue. According to Marshall, Meza, and Olkin (2001), several factors affect the ability of lifetime data to identify its parent distribution, such as the size of the sample, the number of alternative distributions, and the parameter estimation methods. Given a sufficiently large sample size, the lifetime data can identify its parent distribution easily due to the Glivenko–Cantelli theorem. However, it is difficult to find out the parent distribution with small sample sizes.

1.2 Matching Two Lifetime Distributions

The issue of the data-distribution identification is pertinent for small sample sizes. To resolve the identification issue, one can attempt to match two different lifetime distributions. Dubey (1967) found that the normal distribution can be used to approximate the Weibull distribution. This means that a simple (shape-free) lifetime distribution can be used as a surrogate for a complicated one. Among the three popular lifetime distributions, the lognormal distribution is a distribution without a shape parameter, and hence it was chosen to match the Weibull and gamma distributions in our study.

Many criteria can be used to match two different lifetime distributions, such as based on the mean and CV, based on the mean and standard variance, based on the mean and median (Dey and Kundu, 2009; Loistl, 2015). In this section, three following naive methods for matching are discussed:

- matching based on the mean and variance,
- matching based on the mode and coefficient of variation,
- matching based on the CDF.

Matching Based on the Mean and Variance

Matching two different distributions at the mean and variance means that we should establish the equivalence of the means and variances of the two distributions. The matching is carried out by the following two steps:

- Step 1: Find the mean and variance of the two lifetime distributions. For example, the lognormal distribution and any other distribution, such as Weibull or gamma. Force the means to be equal, and repeat the same to their variances.
- Step 2: Transform the equation obtained from Step 1 to represent the parameters of the lognormal distribution with the parameters of the other lifetime distribution we were wanting to match.

An important fact is that three lifetime distributions we study in this thesis just have two parameters. The lognormal distribution is different from the Weibull and gamma distributions because it has no shape parameter. Therefore, we have to tackle the shape issue in our matching. The shape parameter is hard to identify in practice, and this is the reason why we want to use the lognormal distribution to approximate the Weibull and gamma distributions. Case (a) given below describes matching $\mathrm{Wei}(k, \lambda)$ for $k > 1$ with $\mathrm{LN}(\mu, \sigma^2)$, while Case (b) deals with matching $Gamma(\omega, \theta)$ with $\mathrm{LN}(\mu, \sigma^2)$.

- Case (a): Employ the lognormal distribution $X \sim \mathrm{LN}(\mu, \sigma^2)$ to approximate the Weibull distribution $\mathrm{Y} \sim \mathrm{Wei}(k, \lambda)$ by performing the above steps. Consider the following equations

$$\mathrm{E(X)} = e^{(\mu+\sigma^2/2)} = \lambda\Gamma(1+1/k) = \mathrm{E(Y)} \tag{1.1}$$

and

$$\mathrm{Var(X)} = (e^{\sigma^2}-1)e^{(2\mu+\sigma^2)} = \lambda^2\left[\Gamma\left(1+2/k\right) - \left(\Gamma\left(1+1/k\right)\right)^2\right] = \mathrm{Var(Y)} \tag{1.2}$$

Solving equations (1.1) and (1.2) for the parameters $\mu_{(a)}$ and $\sigma_{(a)}$

$$\sigma_{(a)} = (\log(\Gamma(1+2/k)/(\Gamma(1+1/k))^2))^{1/2},$$

$$\mu_{(a)} = \log(\lambda\Gamma\left(1+1/k\right)) - \log(\lambda^2\left(\Gamma\left(1+2/k\right) - ((\Gamma\left(1+1/k\right))^2\right))/2.$$

- Case (b): Employ the lognormal distribution $X \sim \mathrm{LN}(\mu, \sigma^2)$ to approximate the gamma distribution $Z \sim Gamma(\omega, \theta)$ by performing the above steps, and the equations are given by

$$\mathrm{E(X)} = e^{(\mu+\sigma^2/2)} = \omega\theta = \mathrm{E(Z)} \tag{1.3}$$

and

$$\mathrm{Var(X)} = (e^{\sigma^2}-1)e^{(2\mu+\sigma^2)} = \omega\theta^2 = \mathrm{Var(Z)} \tag{1.4}$$

Solving equations (1.3) and (1.4) for the parameters $\mu_{(b)}$ and $\sigma_{(b)}$

$$\sigma_{(b)} = (\log(1/\omega + 1))^{1/2},$$

$$\mu_{(b)} = \log(\omega\theta) - (\log(\omega\theta^2))/2.$$

As an example, let the values of parameters be: $k = 3$, $\lambda = 2$, $\omega = 2$, and $\theta = 3$. Furthermore, we repeat the derivation from the given values, $\mu_{(a)} = 1.01$, $\sigma_{(a)} = 0.12$, $\mu_{(b)} = 0.35$, and $\sigma_{(b)} = 0.41$. The plots of these two matching lifetime distributions are shown in Figure 1.4. As illustrated, the distance between these two distinct distribution types is greater. Figure 1.4(a) shows the CDF curves of the lognormal and Weibull distributions. Figure 1.4(b) shows the CDF curves of lognormal and gamma distributions. Therefore, we can conclude that the method of matching at the mean and variance do not work well when the lognormal distribution is employed to match the Weibull and gamma distributions.

Matching Based on the Mode and Coefficient of Variation

To make sure the two different distributions approximately possess the same mode and coefficient of variation, we should establish equivalence relationships based on the mode and CV. The matching is carried out by the following two steps:

- Step 1: Find the mode and CV of two lifetime distributions: the lognormal distribution and the other distribution; Weibull or gamma distribution. Force their modes equal, and then repeat the same to their CV.
- Step 2: Transform the equation obtained from Step 1 to represent the parameters of the lognormal distribution by using the parameters of the other lifetime distribution.

The shape parameter of the Weibull and gamma distributions is assumed to be greater than one in the following two cases considered in this section:

- Case (a): Employ the lognormal distribution $X \sim \mathrm{LN}(\mu, \sigma^2)$ to approximate the Weibull distribution $Y \sim \mathrm{Wei}(k, \lambda)$ by following the above steps. The relevant

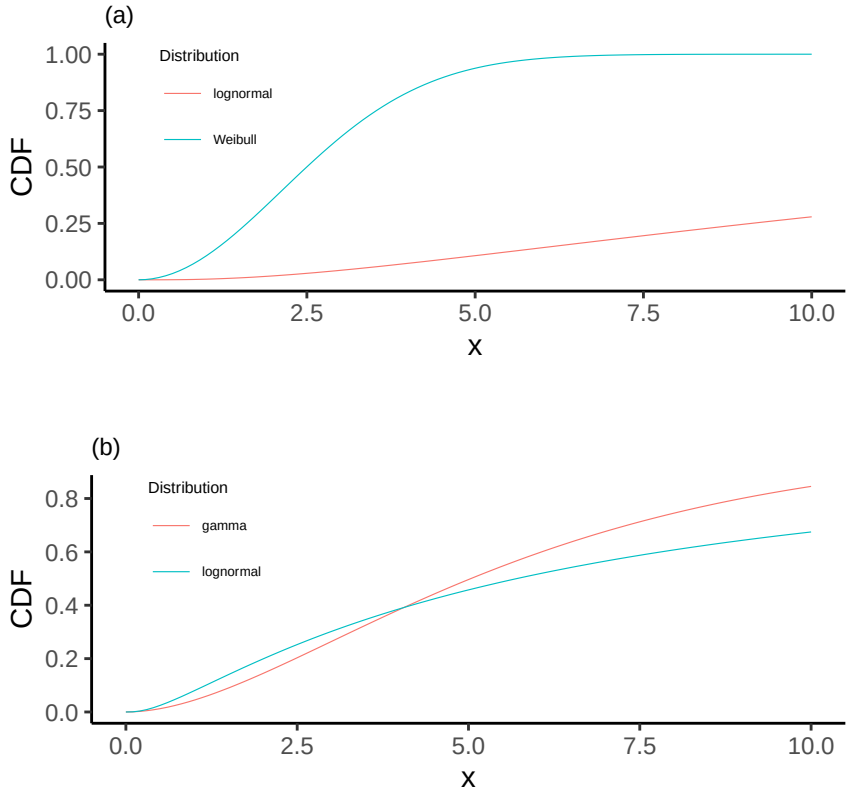

Figure 1.4: *Plots of two different matching distributions based on mean and variance: (a) LN(1.01,0.12) → Wei(3,2); (b) LN(0.35,0.41) → Gamma(2,3)*

equations are given by

$$\text{Mode}(\text{X}) = e^{\mu-\sigma^2} = \lambda\left[(k-1)/k\right]^{1/k} = \text{Mode}(\text{Y}) \tag{1.5}$$

and

$$\text{CV}(\text{X}) = (e^{\sigma^2}-1)^{1/2} = (2/k\Gamma(2/k) - \Gamma(1+1/k)^2)^{1/2}(1/k\Gamma(1/k))^{-1} = \text{CV}(\text{Y}) \tag{1.6}$$

Solving the equations (1.5) and (1.6) for the parameters $\mu_{(a)}$ and $\sigma_{(a)}$

$$\mu_{(a)} = \log(\lambda(1-1/k)^{1/k}(\Gamma(1+2/k)/(\Gamma(1+1/k)^2)),$$

$$\sigma_{(a)} = (\log(\Gamma(1+2/k)/(\Gamma(1+1/k))^2))^{1/2}.$$

- Case (b): Employ the lognormal distribution $X \sim \mathrm{LN}(\mu, \sigma^2)$ to approximate the gamma distribution $Z \sim Gamma(\omega, \theta)$ with the following equations

$$\mathrm{Mode(X)} = e^{\mu - \sigma^2} = (\omega - 1)\theta = \mathrm{Mode(Z)} \tag{1.7}$$

and

$$\mathrm{CV(X)} = (e^{\sigma^2} - 1)^{1/2} = \omega^{-1/2} = \mathrm{CV(Z)} \tag{1.8}$$

Solve the equations (1.7) and (1.8) for the parameters $\mu_{(b)}$ and $\sigma_{(b)}$

$$\mu_{(b)} = \log(\theta\omega(1 - 1/\omega^2)),$$

$$\sigma_{(b)} = (\log(1/\omega + 1))^{1/2}.$$

We present an example for this, and let the given values of parameters be: $k = 1.5$, $\lambda = 3$, $\omega = 2$, and $\theta = 1.75$. We then repeat the derivation from the given values, $\mu_{(a)} = 1.26$, $\sigma_{(a)} = 0.81$, $\mu_{(b)} = 0.86$, and $\sigma_{(b)} = 0.46$. The plots of these two matching different lifetime distributions are shown in Figure 1.5. As illustrated, the inter-distance between these two distributions is large. Figure 1.5(a) shows the CDF curves of the lognormal and Weibull distributions. Figure 1.5(b) shows the CDF curves of the lognormal and gamma distributions. Based on these examples, we conclude that the matching two distributions based on the two population numerical characteristics are not an appealing strategy.

Matching Based on the CDF

We can also match two different lifetime distributions by finding the minimum distance between the cumulative distribution functions of two lifetime distributions. For this type of matching, the `BB optim` function in `R` can be used. In this section, we also consider the following two cases:

- Case (a): Employ the lognormal distribution $X \sim \mathrm{LN}(\mu, \sigma^2)$ to approximate the Weibull distribution $\mathrm{Y} \sim \mathrm{Wei}(k, \lambda)$, and search for the optimal parameters of the

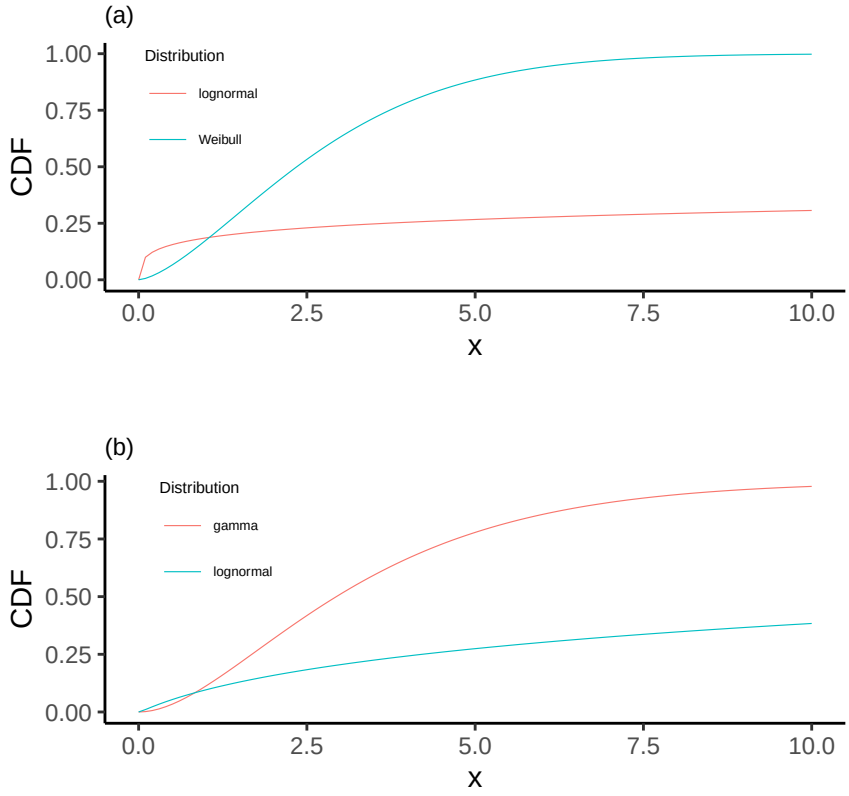

Figure 1.5: *Plots of two different matching distributions based on mode and CV: (a) LN(1.26,0.81) → Wei(1.5,3); (b) LN(0.86,0.46) → Gamma(2,1.75)*

lognormal distribution. The equations are given by

$$\|F_X - F_Y\| = \int_0^{+\infty} |F_X(t) - F_Y(t)|\, dt = h(\mu_{(a)}, \sigma^2_{(a)}) \tag{1.9}$$

Solving the equation (1.9) for the optimal parameters $\mu_{(a)}$ and $\sigma_{(a)}$

$$(\mu_{(a)}, \sigma^2_{(a)}) = \operatorname{argmin} \hat{h}(\mu_{(a)}, \sigma^2_{(a)}),$$

where $\hat{h}(\mu_{(a)}, \sigma^2_{(a)}) = \sum_{n=1}^{N} |F_X(t_n) - F_Y(t_n)|\Delta t$, $\Delta t = t_n - t_{n-1}$ and $t_0 = 0$.

- Case (b): Employ the lognormal distribution $X \sim \mathrm{LN}(\mu, \sigma^2)$ to approximate the gamma distribution $Z \sim Gamma(\omega, \theta)$, and search for the optimal parameters of the lognormal distribution. The equations are given by

$$\|F_X - F_Z\| = \int_0^{+\infty} |F_X(t) - F_Z(t)|\, dt = h(\mu_{(b)}, \sigma^2_{(b)}) \tag{1.10}$$

Solving the equation (1.10) for the optimal parameters $\mu_{(b)}$ and $\sigma_{(b)}$

$$(\mu_{(b)}, \sigma^2_{(b)}) = \operatorname{argmin} \hat{h}(\mu_{(b)}, \sigma^2_{(b)}),$$

where $\hat{h}(\mu_{(b)}, \sigma^2_{(b)}) = \sum_{n=1}^{N} |F_X(t_n) - F_Z(t_n)|\Delta t$, $\Delta t = t_n - t_{n-1}$ and $t_0 = 0$.

As an example, let the given the values of parameters be: $k = 6.0$, $\lambda = 5.0$, $\omega = 2.0$, and $\theta = 3.0$. This derivation is repeated for the given values, $\mu_{(a)} = 0.42$, $\sigma_{(a)} = 0.01$, $\mu_{(b)} = 0.36$, and $\sigma_{(b)} = 0.21$. The plots of these two matching lifetime distributions are shown in Figure 1.6. It can be seen that the curves of these two distributions look somewhat similar. Figure 1.6(a) shows the CDF curves of the lognormal and Weibull distributions. Figure 1.6(b) shows the CDF curves of the lognormal and gamma distributions. Therefore, we can conclude that the method of matching based on the CDF works better than the previous two naive methods.

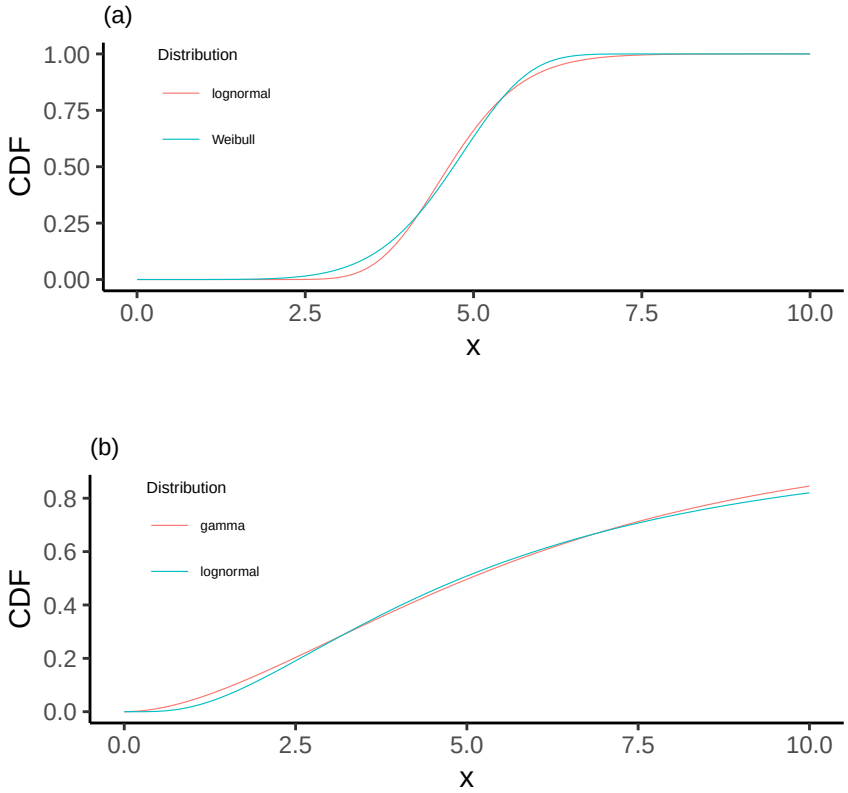

Figure 1.6: *Plots of two different matching distributions based on the difference between CDF: (a) LN(0.42,0.01) → Wei(6,5); (b) LN(0.36,0.21) → Gamma(2,3)*

Although many researchers studied statistical methods for model misspecification, there is only a very few published works that focused on the life test sampling under model uncertainty; see, for example, Zhao et al. (2019) analyzed the accelerated life test sampling plan under model uncertainty.

1.3 Synopsis of new contribution

The purpose of our work is to develop a method to resolve the distribution-data identification issue in lifetime acceptance sampling for small sample sizes. We would like to show that the lognormal distribution can be a good surrogate for Weibull and gamma distributions under certain circumstances.

The rest of this thesis is organized as follows. Chapter 2 discusses the parameter estimation methods for three lifetime distributions, and type I censored data. Chapter 3 designs a variables sampling plan, called the M-method sampling plan, which is particularly suitable for small sample sizes. Some existing alternative methods are also discussed in this Chapter. Chapter 4 shows that when the underlying distribution is misspecified, the proposed method of using a simple distribution instead of a complicated distribution works well for small sample sizes. The M-method developed in Chapter 3 is used in Chapter 4 to design life test sampling plans. Since the shape parameter is the difficult one to estimate precisely in practice, the lognormal distribution can be used as a surrogate for the Weibull and gamma distributions for small sample sizes. Finally, the general conclusions and future perspectives are given in the last Chapter.

Chapter 2

Parameter Estimation of Lifetime Distributions

2.1 Introduction

The estimation of parameters of distribution is intrinsically tied to the identification of the parent distribution. Various estimation procedures can be used to estimate the parameters of the lognormal, Weibull, and gamma distributions; the popular parameter estimation methods include maximum likelihood estimation (MLE), the method of moments (MOM), least-square estimation, and the method of logarithm moment (Teimouri, Hoseini, and Nadarajah, 2013). Among all these estimation methods, MLE is the most commonly used approach, and MOM is the oldest estimation method (Casella and Berger, 2002). Hence, the MLE and MOM of three lifetime distributions will be discussed in Section 2.2. In Section 2.3, a brief discussion on right-censored data is provided, and the parametric inference for the maximum likelihood estimation with type I censored data is presented. A numerical study of MLE with small sample sizes for three lifetime distributions is given in Section 2.4. The `mledist` function of the `fitdistrplus` R package is discussed.

2.2 Review of Estimation Methods for Lifetime Distributions

2.2.1 Lognormal Distribution

Maximum Likelihood Estimation For MLE, there are three steps to obtain the estimators: first, let $x_1, ..., x_n$ be a random sample from a lognormal distribution with two parameters μ and σ. The likelihood function is

$$L(\mu, \sigma \mid x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} (1/x_i)\sigma^{-n}(2\pi)^{-n/2} exp(-\sum_{i=1}^{n}(\log x_i - \mu)^2/2\sigma^2);$$

second, take the logarithm of the likelihood function. The log-likelihood function is

$$\ell(\mu, \sigma \mid x_1, x_2, \ldots, x_n) = -\sum_{i=1}^{n} \log x_i - n \log \sigma - (n/2) \log 2\pi - \sum_{i=1}^{n}(\log x_i - \mu)^2/2\sigma^2;$$

finally, take the partial derivatives for μ and σ^2, and then let these two equations equal to 0. Hence, the estimators are given by

$$\widehat{\mu} = (\sum_{i=1}^{n} \log x_i)/n,$$

$$\widehat{\sigma}^2 = (\sum_{i=1}^{n}(\log x_i - \widehat{\mu})^2)/n.$$

Method of Moments The regular MOM uses the first two sample moments equal to the corresponding two population moments, $\overline{X} = \mathrm{E(X)}$ and $S_X^2 = \mathrm{Var(X)}$. The sample mean is $\overline{X} = (\sum_{i=1}^{n} x_i)/n$ and sample variance is

$$S_X^2 = (\sum_{i=1}^{n}(x_i - \overline{X})^2)/(n-1).$$

After a little algebra, the moment estimators of μ and σ^2 can be found as

$$\widetilde{\mu} = (1/3) \log(\overline{X} S_X)^2,$$

$$\widetilde{\sigma}^2 = (2/3) \log(\overline{X}/S_X^2).$$

2.2.2 Weibull Distribution

Maximum Likelihood Estimation For MLE, three steps are involved to obtain the estimators: first, let $y_1, ..., y_n$ be a random sample from a Weibull distribution with shape parameter k and scale parameter λ. The likelihood function is

$$L(k, \lambda \mid y_1, y_2, \ldots, y_n) = k^n \lambda^{-kn} \prod_{i=1}^{n} y_i^{k-1} exp(-\sum_{i=1}^{n} y_i^k / \lambda^k);$$

second, take the logarithm of the likelihood function. The log-likelihood function is

$$\ell(k, \lambda \mid y_1, y_2, \ldots, y_n) = n \log k - kn \log \lambda + (k-1) \sum_{i=}^{n} \log y_i - \sum_{i=}^{n} y_i^k / \lambda^k;$$

finally, take the partial derivatives for k and λ, and then let these two equations equal to 0. According to Johnson, Kotz, and Balakrishman (1994), the maximum likelihood estimators of parameter k and λ can be found by solving

$$\widehat{\lambda} = (1/n)(\sum_{i=1}^{n} y_i^k)^k,$$

$$1/\widehat{k} = [(\sum_{i=1}^{n} y_i^k)(\log(y_i))] / \sum_{i=1}^{n} y_i^k - (1/n) \sum_{i=1}^{n} \log(y_i).$$

It can be found that $\widehat{\lambda}$ depends on the parameter k, and hence $\widehat{k}$ must be calculated numerically. In addition, some graphical and iterative methods can be used to solve these equations to obtain $\widehat{\lambda}$ and $\widehat{k}$ (Teimouri, Hoseini, and Nadarajah, 2013).

Method of Moments The regular MOM equates the first two sample moments to the corresponding two population moments, $\overline{Y} = E(Y)$ and $S_Y^2 = Var(Y)$. The sample mean is $\overline{Y} = (\sum_{i=1}^{n} y_i)/n$, and sample variance is

$$S_Y^2 = (\sum_{i=1}^{n} (y_i - \overline{Y}^2))/(n-1).$$

After a little algebra, the moment estimator of parameter λ becomes

$$\widetilde{\lambda} = \overline{Y}/\Gamma(1 + 1/k),$$

but the moment estimator of parameter k must be solved by numerically; $\widetilde{k}$ is the solution of the equation

$$S_Y/\overline{Y} = [\Gamma(1+2/k)/(\Gamma(1+1/k))^2 - 1]^{(1/2)}.$$

It can also be seen that $\widetilde{\lambda}$ depends on the parameter k.

2.2.3 Gamma Distribution

Maximum Likelihood Estimation For MLE, there are three steps to obtain the estimators: first, let $z_1, ..., z_n$ be a random sample from a two-parameter gamma distribution with parameter ω and θ. The likelihood function is

$$L(\omega, \theta \mid z_1, z_2, \ldots, z_n) = \prod_{i=1}^{n} z_i^{\omega-1} exp(-\sum_{i=1}^{n} z_i/\theta)\theta^{-n\omega}(\Gamma(\omega)^{-n});$$

second, take the logarithm of the likelihood function. The log-likelihood function is

$$\ell(\omega, \theta \mid z_1, z_2, \ldots, z_n) = (\omega - 1)\sum_{i=1}^{n} \log z_i - \sum_{i=1}^{n} z_i/\theta - n\omega \log\theta - n \log\Gamma(\omega);$$

finally, take the partial derivatives for ω and θ, and then let these two equations equal to 0. Hence, the estimator of the parameter θ is

$$\widehat{\theta} = \sum_{i=1}^{n} z_i/n\omega.$$

The estimator of the parameter ω can be found by using an iterative algorithm (Minka, 2002; Choi and Wette, 1969), the update is

$$1/\omega^{new} = 1/\omega + (\overline{\log z} - \log\overline{z} + \log\omega - \psi(\omega))/\omega^2(1/\omega - \psi'(\omega)),$$

where ψ is the digamma function and $\log\overline{z} \geq \overline{\log z}$ by Jensen's inequality. A good starting point for the iteration is $\omega \approx 0.5/\log\overline{z} - \overline{\log z}$.

Method of Moments The regular MOM equates the first two sample moments to the corresponding two population moments, $\overline{Z} = \mathrm{E}(Z)$ and $S_Z^2 = \mathrm{Var}(Z)$. The sample mean

is $\overline{Z} = (\sum_{i=1}^{n} z_i)/n$ and sample variance is

$$S_Z^2 = (\sum_{i=1}^{n}(z_i - \overline{Z}^2))/(n-1).$$

After a little algebra, the moment estimators of ω and θ are are found as

$$\widetilde{\omega} = (\overline{Z})^2/S_Z^2,$$

$$\widetilde{\theta} = S_Z^2/\overline{Z}.$$

2.3 Right-censored Data

2.3.1 Introduction

In life testing, failure events of some subjects may not be observed. If the actual time of occurrence of an event is unobserved, it is said to be censored. There are many types of censoring, such as right-censoring, left-censoring, and interval-censoring. Since the right-censoring is the most common strategy in the reliability studies, the remaining sections will focus on right-censored data.

Right-censoring happened when some subjects were still "alive" at the last observation. The observation may stop because the study has ended, or the subjects have disappeared for some reason, such as when the patient is discharged from the hospital after completing a course of treatment. There are two special cases for the right-censored data as follows:

- Type I censored data: In the reliability studies, we fixed the experimental time for all subjects or items to run, and some subjects or items still survived at the end of the experimental end time, which is often predetermined due to administrative and other reasons.
- Type II censored data: In the reliability studies, the proportion of the "dead" subjects or items may be fixed in advance. The test is terminated when a predetermined proportion of subjects died.

Type I censored data is the most common type of the right-censored data, and this will be discussed for the next sections.

2.3.2 Estimation with Type I Censored Data

If a suitable lifetime distribution can be assumed for the event time T, right-censored data can also be used to do the parametric estimation, i.e., maximum likelihood estimation (MLE) and moment-method estimation (MOM). However, the construction of the appropriate moment equations is difficult to implement under censoring testing (Zhou and He, 2005). Although MOM can be used (Sirvanci and Yang, 1984; Zhou and He, 2005), MLE is preferred when we have right-censored data.

Suppose that the probability density function of the underlying distribution is $f_T(t;\eta)$, where the η is a vector of parameters. Given a set of lifetime data and the likelihood function is a function of η. In type I censoring, the observations surviving beyond a fixed censoring time c are right-censored. If $T > c$, the censored subjects contribute $P(T > c|\eta) = S_T(c;\eta)$ to the likelihood. The joint likelihood for all observations is

$$L(\eta) = \prod_{i=1}^{n} f_T(t_i;\eta)^{\delta_i} \prod_{i=1}^{n} S_T(c;\eta)^{1-\delta_i},$$

where δ_i is the censoring indicator with $\delta_i = 0$ if $T > c$. Take the logarithm of the likelihood function, and the log-likelihood function can be written as follows:

$$\ell(\eta) = \delta_i \sum_{i=1}^{n} \log(f_T(t_i;\eta)) + (1-\delta_i) \sum_{i=1}^{n} \log(S_T(c;\eta)).$$

2.4 Estimation with Small Samples

For the three lifetime distributions dealt with in this thesis, the parameter estimation can be done by using the R package `fitdistrplus`. In this package, different methods are available to estimate the distribution parameters by using functions `fitdist` and `fitdistcens`. These are:

- maximum likelihood estimation (`mledist`),

- moment matching estimation (`mmedist`),
- quantile matching estimation (`qmedist`),
- maximum goodness-of-fit estimation (`mgedist`).

The maximum likelihood estimation for two-parameter lifetime distributions can deal with both censored and complete data. However, the other three methods can only be applied to complete data. In addition, as we have mentioned in Section 2.3, MLE is preferred for type I censored data. Therefore, MLE is employed in this section to deal with type I censored data when the sample sizes are small.

2.4.1 Maximum Likelihood Estimation with Small Sample Sizes

The maximum likelihood method can satisfactorily estimate the parameters of lifetime distributions for type I censored data, even if the sample size is small. Since the parent distribution of data may be unknown, we try to fit all three lifetime distributions, even if the underlying distribution is misspecified. It is of interest to investigate the effects of misspecifications on estimated distributions.

For illustrative purposes, we compare the inter-distances between the PDF of two different lifetime distributions and plot the density function of the relative errors. We try to find the relative error between the PDF of the fitted and true lifetime distributions. The procedure for finding of the relative error is similar to the search procedure described for the optimal parameters of the lognormal distribution mentioned in Chapter 1. The equations of the relative error are given by

$$\|f_{fitted} - f_{true}\| = \int_0^{+\infty} |f_{fitted}(t) - f_{true}(t)|\, dt.$$

The box plot display is used to summarize the distribution of relative errors. In addition, the smoothed density of relative errors is plotted to visualize the differences. Monte Carlo simulation with 2000 repetitions for each comparison is used to obtain the empirical distribution of the relative errors. In the simulation, the type I censored data is generated from the parent distribution with the censoring time T. Since it is difficult to define what

a small sample size is, we considered a selected set of small sample sizes to perform the analysis.

The relative error is calculated by the following four steps:

- Step 1: Assume the parent distribution and for given values of small sample size n and censoring time T, a random type I censored data can be generated from the assumed parent distribution.
- Step 2: With random type I censored data obtained in Step 1, obtain the fitted parameters of all three lifetime distributions by using the `mledist` function in `R`.
- Step 3: Calculate the fitted probability density function by using the fitted parameters obtained in Step 2, and also calculate the true PDF of the parent distribution.
- Step 4: Find the relative error by calculating the sum of the absolute value between the true PDF and the fitted PDF.

Suppose that the type I censored data is generated from the lognormal distribution. In the study, a random sample with the sample size $n = 30$ is generated from the lognormal distribution with parameters $\mu = 0.15$, $\sigma = 0.90$, and censoring time $T = 2$. Following the above Monte Carlo simulation steps, the distribution of relative errors is presented in Figure 2.1. As shown in Figure 2.1(a), the relative error between the true lognormal PDF and the fitted lognormal PDF is different from another two combinations: true lognormal versus fitted Weibull and true lognormal versus fitted gamma. In addition, the box plot of these two combinations is similar. The smoothed density curves are shown in Figure 2.1(b). The PDF curve of the relative error between the true lognormal PDF and the fitted lognormal PDF is depicted as the green line. It can be seen that this line is different from the other lines. Therefore, we can conclude that the relative error between the true lognormal PDF and the fitted lognormal PDF is smaller than the other two combinations.

Suppose that the type I censored data is generated from the Weibull distribution. In the study, a random sample with the sample size $n = 30$ is generated from the Weibull distribution with shape parameter $k = 1.50$, scale parameter $\lambda = 0.20$, and censoring time $T = 2$. Figure 2.2 describes the distribution of relative errors by following the simulation

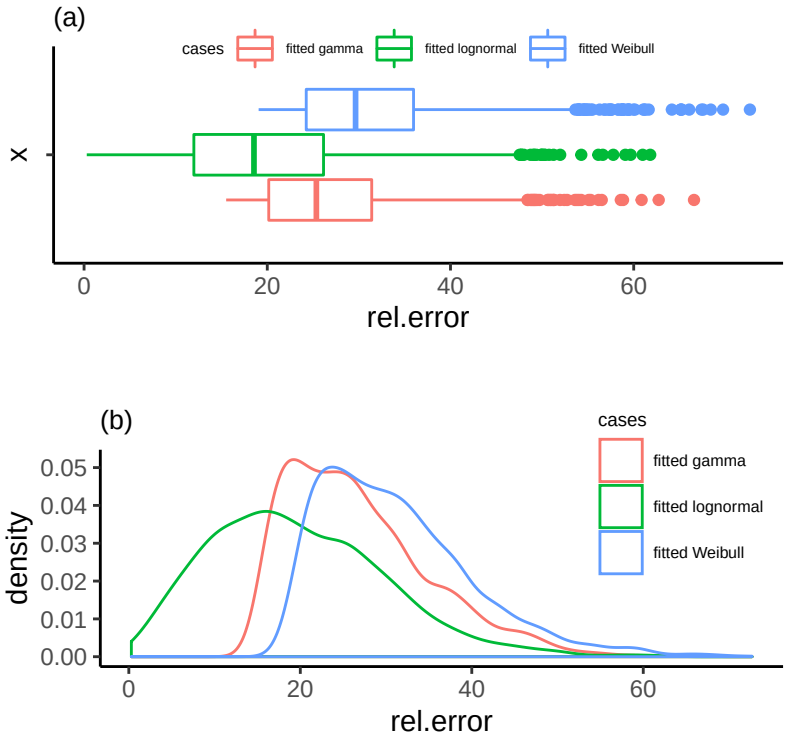

Figure 2.1: *Plots of the relative error when the type I censored data comes from lognormal distribution for n=30: (a) box plot, (b) smoothed density*

steps. As shown in Figure 2.2(a), the relative error between the true Weibull PDF and the fitted lognormal PDF is significantly large when compared to the other two combinations: true Weibull versus fitted Weibull and true Weibull versus fitted gamma. In addition, the box plots of the other two combinations look somewhat similar. The smoothed density curves are shown in Figure 2.2(b). The PDF curve of the relative errors between the true Weibull PDF and the fitted lognormal PDF is depicted as the green line. The PDF curve of the relative error between the true Weibull PDF and the fitted Weibull PDF is depicted as the blue line. The PDF curve of the relative error between the true Weibull PDF and the fitted gamma PDF is depicted as the red line. It is seen that the green line is strikingly different from red and blue lines. Therefore, we can conclude that the relative error between the true Weibull PDF and the fitted lognormal PDF is bigger when compared to the other two options.

Suppose that type I censored data is generated from the gamma distribution. In this study, a random samples of size $n = 30$ is generated from the gamma distribution with

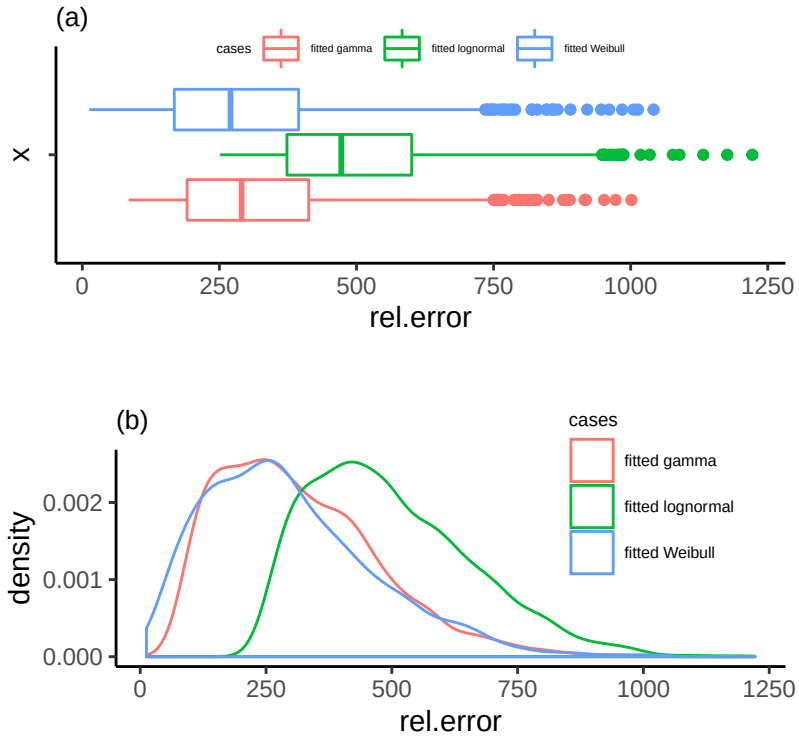

Figure 2.2: *Plots of the relative error when the type I censored data comes from Weibull distribution for n=30: (a) box plot, (b) smoothed density*

shape parameter $\omega = 1.40$, rate parameter $\psi = 5.00$, and censoring time $T = 2$. As shown in Figure 2.3(a), the relative error between the true gamma PDF and the fitted lognormal PDF is distinct from the other two combinations: true gamma versus fitted Weibull and true gamma versus fitted gamma. In addition, the box plots of these two combinations are similar. The smoothed density curves are shown in Figure 2.3(b). The PDF curve of the relative errors between the true gamma PDF and the fitted lognormal PDF is depicted as the green line. The PDF curve of the relative errors between the true gamma PDF and the fitted Weibull PDF is depicted as the blue line. The PDF curve of the relative error between the true gamma PDF and the fitted gamma PDF is depicted as the red line. It can be seen that the green line stands out from the red and blue lines. In addition, the red and blue lines look somewhat similar. Therefore, we can conclude that the relative error between the true gamma PDF and the fitted lognormal PDF is larger than the other two options.

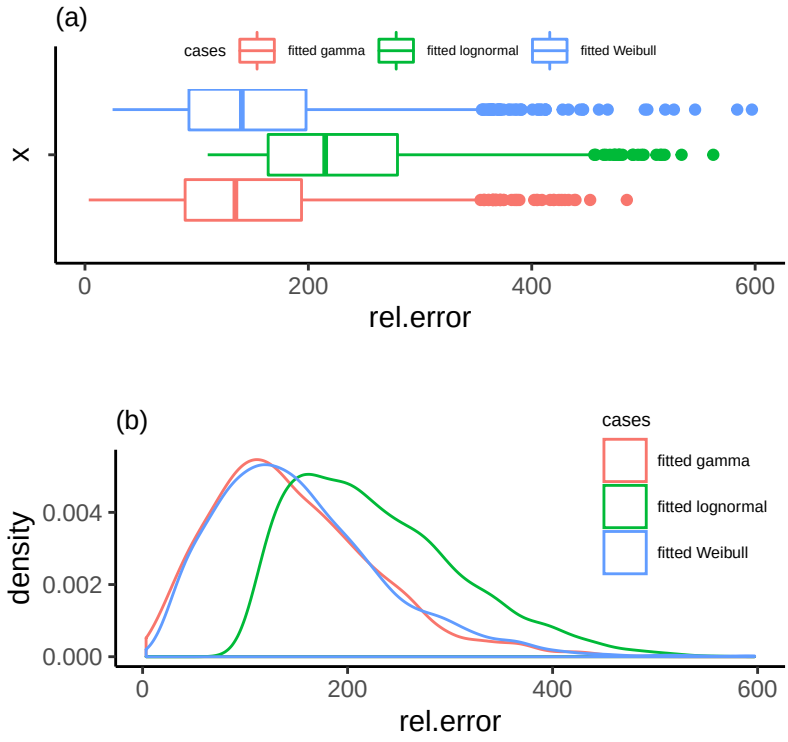

Figure 2.3: *Plots of the relative error when the type I censored data comes from gamma distribution for n=30: (a) box plot, (b) smoothed density*

2.5 Conclusions

In this chapter, we reviewed the estimation methods for three common lifetime distributions and provided a brief discussion about the parameter estimation for type I censored data. In addition, a numerical study of the parameter estimation (MLE) for type I censored data under small sample sizes is also given.

Based on the results presented, the maximum likelihood estimation is found suitable for type I censored data. The lognormal distribution is distinctly different from Weibull and gamma distributions when the fitted distribution is a misspecified one for small sample sizes. In the next chapter, we continue to investigate the case of small type I censored data but design a variables sampling plan suitable for small sample sizes. This sampling plan design continues to employ the maximum likelihood estimation procedure.

Chapter 3

Life Test Sampling Plan for Small Sample Sizes

3.1 Introduction

The ultimate performance of products can be evaluated using a life test sampling plan (Schilling and Neubauer, 2017). When the consumer conducts a sampling inspection, they want to protect their interests even if they have little information about the product. Usually, the sample size in a sampling inspection plan tends to be quite small. Since waiting for all products to fail can be time-consuming, particularly when the expectancy of the product lifetime is long, small size right-censored data are common in life test sampling (Wilrich, 2018). It is of practical interest to design a life test sampling plan with the type I censoring for small sample sizes.

Acceptance sampling is the technique that determines whether to accept or reject a lot based on the results of sample inspection. The acceptance sampling plan is a specific plan that determines the sample size, and resulting acceptance and rejection criteria. When the lifetime of a product is an important quality characteristic, the life test sampling plan prescribes the sampling rules and associated acceptance criteria (Seo, Jung, and Kim, 2009). There are many ways to do life testing, such as truncation life test, type I censoring, type II censoring, hybrid censoring, and progressive censoring (Kim and Yum, 2011). We will

consider the type I censoring in the next few sections since it is the most common type of data in life testing.

Life test sampling plan can be classified into two classes:

- Attribute sampling plan, in which products used for decision making are classified as conforming or nonconforming,
- Variables sampling plan, in which actual lifetimes are used for decision making without classifying each observation as conforming or not.

The attribute sampling plan for different underlying distributions has been discussed in several publications. For lognormal and normal distributions, Gupta (1962) designed an acceptance sampling plan for life testing with type I censoring. This approach was also extended to the underlying gamma distribution with known shape parameters. For example, Gupta and Groll (1961) designed a sampling plan for controlling the mean life time criterion. The Weibull distribution with known shape parameter is investigated by Goode and Kao (1960). They designed a life test sampling plan based on the mean life time criterion. Furthermore, this methodology was extended by Zhang and Meeker (2005) to a life test sampling plan with type II censoring. The attribute sampling plan can also be designed for other distributions, such as exponential distribution (Epstein, 1954; Spurrier and Wei, 1980), Birnbaum Saunders distribution (Baklizi and Masri, 2004), log-logistic distribution (Rao and Kantam, 2010), and inverse-gamma distribution (Al-Masri, 2018).

Although the attribute sampling plan is very simple and convenient to use, it is not making full use of the data that are measured on a numerical scale. The variables sampling plan is preferred in many applications because it utilizes more information in the test data, and hence it needs a smaller sample size than the attribute sampling plan for ensuring the same level of protection to the consumer and producer (Amitava, 2008; Schilling and Neubauer, 2017). The variables sampling plan with various underlying distributions has been discussed in several publications. Studies employing the lognormal distribution can be found in Lina, Wua, and Balakrishnan (2009) for type I censoring, Balasooriya and Balakrishnan (2000) for progressive censoring, and Schneider (1989) for type II censoring.

The Weibull distribution with the known shape parameter was employed for the type II censored sampling plan (Wu and Tsai, 2000; Jun, Balamurali, and Lee, 2006). The Weibull distribution with the unknown shape parameter can be employed for various censored sampling plans, such as type I censoring sampling plan of(Wilrich, 2018), progressive censoring sampling plan of Balasooriya, Saw, and Gadag (2000), and hybrid censoring sampling plan of Chen et al. (2004). The variables sampling plan was also originally based on the exponential distribution Balasooriya and Saw (1998) and then extended to the case of Burr type X distribution (Kantam and Ravikumar, 2016). All these authors designed the variables sampling plans for controlling the fraction or proportion nonconforming criterion.

The life test sampling plan may make a wrong decision, particularly for small sample sizes. The possibility of an incorrect decision can be divided into the following two types:

- Producer's Risk (α): It is the probability of rejecting good or reliable lots. A good lot which has acceptable quality level should be accepted, but in fact, it is rejected.
- Consumer's Risk (β): It is the probability of accepting bad or unreliable lots. A bad lot which has limiting quality level should be rejected, but in fact, it is accepted.

Although an effective life test sampling plan should be able to control both risks, it is difficult to keep them at very low levels due to small samples being taken. In this chapter, we try to control the consumer's risk in the life test sampling plan. Once the sampling plan is determined for the desired level of consumer's risk, the probability of acceptance can be calculated for various lot quality levels to obtain the Operating Characteristic (OC) curve.

Although many studies have focused on the life test sampling plan assuming various underlying distributions, only a few researchers focused on designing sampling plans with small sample sizes under model misspecification. The main aim of this chapter is to design a variables sampling plan which ensures consumer's protection for small sample sizes when the data are the type I censored. This plan is designated as the M-method sampling plan. We show the superiority of the M-method plan over the alternatives. In

this study, we will consider three common lifetime distributions discussed in the previous chapter: the lognormal, Weibull, and gamma distributions.

Section 2 of this Chapter briefly describes the two existing methods for designing a life test sampling plan, namely the Gupta's method and the Wilrich's method. In Section 3, a variables sampling plan under the M-method is designed, which is particularly suitable for small sample sizes. The methods described in Sections 2 and 3 are compared in Section 4 via Monte Carlo simulation. A brief conclusion is provided at the end.

3.2 Existing Methods

In this section, two existing methods for designing life test sampling plans are briefly reviewed: Gupta's method, which is an attribute sampling methodology designed by Gupta (1962), and Wilrich's method, which is a variables sampling methodology designed by Wilrich (2018).

3.2.1 Gupta's Method

Gupta's method is basically an attribute sampling plan and can be used when the lifetime of the product follows the normal distribution or lognormal distribution. For convenience, we just discuss the lognormal distribution in this section. In the life test associated with Gupta's method, one stops at a preassigned time L_G and notes the number of failures d, which leads to type I censored data. The purpose of this test is to set a specific median life with a probability P^* that can protect the interests of consumers. The specifications of Gupta's sampling plan are given as follows:

- n_G is the number of units on the test,
- c is an acceptance number, the lot is accepted if $d \leq c$ during the fixed time L_G,
- $\exp(\mu_0)$ is a desirable lot quality level at the preassigned consumer's risk $\beta = 1 - P^*$,
- A lot with the true median life $\exp(\mu) \geq \exp(\mu_0)$ is acceptable,
- For the lognormal distribution, the statistics $z = (\log L_G - \mu_0)/\sigma$ is used, where the values of σ and μ_0 are assumed to be known.

Gupta's sampling plan can be characterized by (n_G, c). Therefore, the sampling plan can be designed for controlling the consumer's risk, which is the probability of accepting the bad quality lots, at or below $1 - P^*$. Gupta (1962) considered a large sample size along with the binomial distribution. The minimum sample size n_G should be satisfied with the inequality

$$\sum_{i=0}^{c} \binom{n_G}{i} p^i (1-p)^{n-i} \leq 1 - P^*,$$

where $p = \Phi((\log L_G - \mu_0)/\sigma)$. The OC function of Gupta's sampling plan is given by

$$\sum_{i=0}^{c} \binom{n_G}{i} p^i (1-p)^{n-i},$$

where $p = \Phi((\log L_G - \mu)/\sigma)$.

3.2.2 Wilrich's Method

Wilrich's method is a type I censored sampling plan for the Weibull distribution. Let the lifetime of the product $T \sim Wei(k, \lambda)$ with lower specification limit L_W, the parameters k and λ are all unknown. Wilrich (2018) studied the logarithms of T, which follows the Gumbel distribution with the location parameter ν and scale parameter τ, where $\nu = \log \lambda$ and $\tau = 1/k$. It is further denoted by $W = \log T \sim \ Gumbel(\nu, \tau)$. The estimation approach continues to be the MLE for this study. The life testing stops at t_W, which resulted in type I censoring. The specifications of Wilrich's sampling plan are given as follows:

- n_W is the number of units on the test,
- An item is nonconforming if $T < L_W$,
- p_0 is the fraction of nonconforming, p_1 and p_2 are the specified points. If $p_0 < p_1$, a lot of items is acceptable. If $p_0 > p_2$, a lot of items is rejectable,
- The statistics $y = \widehat{\nu} - g\widehat{\tau}$ is used, where g is the acceptance factor. If $y \geq w_L = \log L_W$, accept the lot; otherwise reject the lot.

Wilrich's sampling plan can be characterized by (n_W, g). Therefore, the sampling plan can be made from controlling the consumer's risk, which means the probability of acceptance for given $p_0 > p_2$ does not exceed β. On the other hand, the sampling plan can also be designed for controlling the producer's risk, which means the probability of rejection for given $p_0 < p_1$ does not exceed α. Given the values of p_1, p_2, α, and β, the formulas for n_W and g are given as follows:

$$n_W = f^2(g, z_C)/A^2,$$

$$g = (z_{p_1} u_{1-\beta} - z_{p_2} u_\alpha)/(u_\alpha - u_{1-\beta}),$$

where

$$A = (z_{p_1} - z_{p_2})/(u_\alpha - u_{1-\beta}),$$

$$z_C = \log t_W - \log L_W - g,$$

where z_{p_1} and z_{p_2} are the quantile of the standardized Gumbel distribution, and u_α and $u_{1-\beta}$ are the quantile of the standardized normal distribution. The OC function of the sampling plan (n_W, g) is given by

$$P_a = 1 - \Phi((z_p + g)/A).$$

Gupta's sampling plan is an attribute sampling plan, and Wilrich's sampling plan is a variables sampling plan, but both of them are suitable for type I censored data. Since the test statistic in Wilrich's method relies on the unknown shape parameter, which is difficult to estimate precisely under small sample sizes, his method may not be ideal for small sample sizes. Given the limitation of time and costs, the case of small sample sizes is of greater practical relevance in the design of sampling plans.

3.3 M-method for Acceptance Sampling Plan

As mentioned earlier, the M-method sampling plan is suitable for small sample sizes. The plan fixes a maximum allowable fraction nonconforming, M, for the operation of the plan.

This is similar to fixing the acceptance number c, the maximum allowable number of nonconforming units, for operating the attribute plan.

To design the M-method sampling plan for type I censored data, it is assumed that the test time t_M is fixed in advance. We also assume that the underlying distribution parameters are all unknown and employ the MLE as the estimation approach. A lower specification limit for the lifetime is also assumed to be fixed in order to protect the interests of consumers. The design constants for the M-method sampling plan and its operation are given below:

- n_M is the number of units on the test,
- M is the maximum allowable fraction nonconforming,
- $\widehat{p}$ is the estimated fraction nonconforming,
- L_M is a lower specification limit. In other words, the value of M-quantile is equal to L_M. If $L_0 > L_M$, a lot of items is acceptable, where L_0 is the estimated lifetime quantile for the items.
- If $\widehat{p} \leq M$, accept the lot; otherwise reject the lot.

The M-method sampling plan can be characterized by its design constants (n_M, M). Therefore, the M-method sampling plan can be designed for controlling the consumer's risk, which means the probability of accepting with $L_0 \leq L_M$ does not exceed β.

Monte Carlo simulations can be used to compute the probability of acceptance P_a and median $\widehat{p}$. Since the distribution parameters are all unknown, we have to estimate the parameters, and we adopt the MLE method for it. The simulation steps for obtaining the operating characteristic or the values of the probability of acceptance for the proposed M-method plan are described below:

- Step 1: Given values of small sample size n_M and censoring time t_M, a type I censored data can be generated from lifetime distributions.
- Step 2: With type I censored data obtained in Step 1, obtain the fitted parameters of the lifetime distribution by using the `fitdist` function in R.

- Step 3: Given a value of L_M, $\widehat{p}$ can be computed by estimating the tail area which is below L_M.
- Step 4: Obtain $\widehat{p}$ and compare it with M, where the value of M is obtained by searching via Monte Carlo simulation, and the selection criterion is to ensure that the consumer's risk β is lower at the set level of poor quality under the given parameters, i.e., sample size n_M. Accept the lot if $\widehat{p} \leq M$.
- Step 5: Repeat the steps from Step 1 to Step 4 for many times, median $\widehat{p}$ and the probability of acceptance P_a can be calculated.

3.4 Comparison of Three Methods

To demonstrate the superiority of the M-method sampling plan for small sample sizes under right-censoring, we compare the discriminatory performance of the M-method plan with Gupta's method plan as well as with Wilrich's method plan. This comparison is made in terms of OC curves via Monte Carlo simulation. The comparison is under made the same, i.e., the same censoring time and the same lower specification limit. All the number of Monte Carlo repetition in the comparison is 2000.

For illustrative purposes, we consider the case of fixing one parameter and varying the other parameter in all our two-parameter lifetime distributions. We also consider the cases of equal and unequal sample sizes. The OC curves are estimated by Monte Carlo simulation given above (Steps 1 to 5).

For comparison, the estimated OC curves for two different sampling plans should intersect at one point. The point of intersection ensures the probability of acceptance P_a be the same at one point of the true fraction of nonconforming p. If the point of intersection is selected at the top of the OC curve where P_a is large, then the protection to the producer is assured. On the other hand, consumer protection is ensured when the point of intersection is associated with a small P_a-value. In addition, the chosen criterion for the true fraction of nonconforming p, the estimated OC curve can also be locally smoothed via the Monte Carlo simulation.

For comparing the M-method sampling plan, the simulation steps for searching the appropriate value of the maximum allowable fraction nonconforming M are given below:

- Step 1: Choose an M-value arbitrarily and set an intersection point where the probability of acceptance P_a for the estimated OC curves of two different method sampling plans is equal.
- Step 2: By using the Monte Carlo simulation, plot two estimated OC curves under the set acceptance criteria for given sample size and censoring time, and draw a black vertical line that can indicate the interaction point.
- Step 3: Obtain the appropriate M-value by trial and error, and this process needs to comply with the criteria for the M-value mentioned in the previous section.

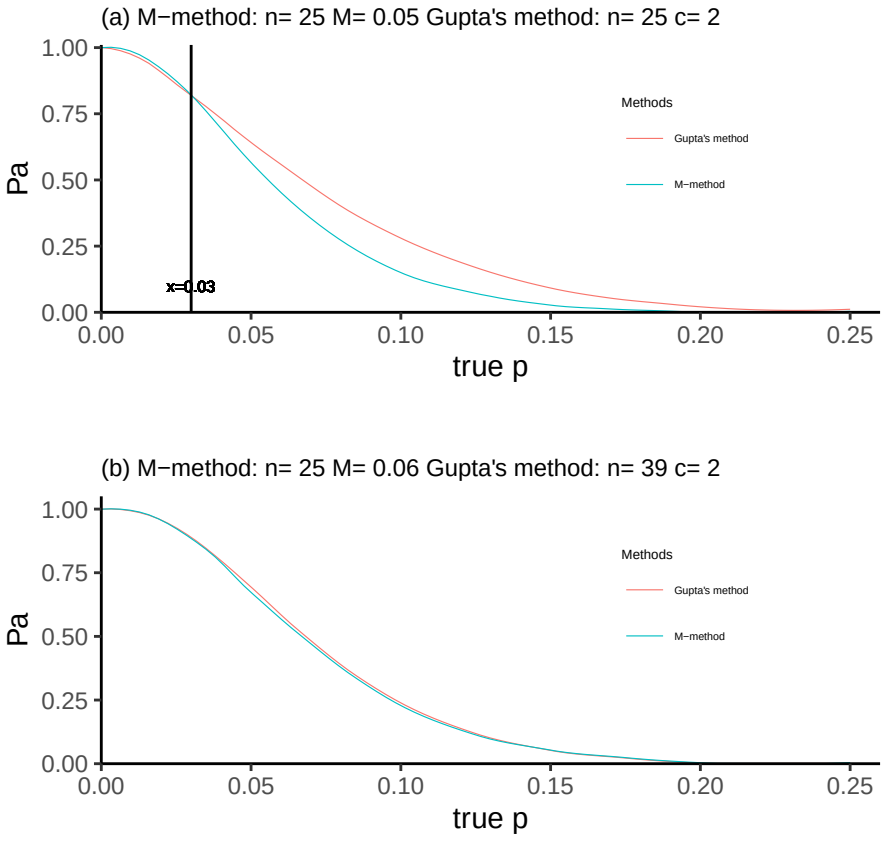

Figure 3.1: *Comparison between M-method and Gupta's method sampling plans.*

To compare the M-method with Gupta's method, we suppose that the sample data comes from the lognormal distribution, and then fit the lognormal distribution by using

MLE. As an example, we fix the parameter $\sigma = 0.90$ and change parameter μ. The root (parameter μ) of the function $\{L_M - \text{qlnorm}(true\ p, \mu, \sigma)\}$ is searched using the `uniroot` function in `R`, where `qlnorm` is the quantile function of the lognormal distribution in `R`, and the searching interval was set as 0 to 100. The chosen values of true p are given in Table 3.1. We set the censoring time $t_M = t_G = 25.5$, and lower specification limit $L_M = L_G = 1.25$. Based on the criteria described in Section 2 and simulation steps in Section 3, the simulated OC curves of the M-method plan and Gupta's method plan are shown in Figure 3.1, where the probability of acceptance P_a is plotted against the true fraction of nonconforming p. It can be seen that the M-method plan is better or more discriminatory than Gupta's method plan for small sample sizes. As shown in Figure 3.1(a), the simulated OC curves of the M-method plan and Gupta's method plan are comparable for equal sample sizes, $n_M = n_G = 25$. In addition, the black vertical line shows that the estimated OC curves of the sampling plan for two methods intersect at the point of true $p = 0.03$. The P_a of the sampling plan for two methods is almost similar at good quality levels. The M-method rejects more than Gupta's sampling plan at poor quality levels. In other words, the M-method sampling plan is sensitive to reject poor quality. Therefore, we can conclude that the M-method is more discriminatory or powerful even when the sample size is small.

As shown in Figure 3.1(b), the simulated OC curves of the M-method and Gupta's method are made approximately the same with unequal sample sizes for the two plans. The simulated OC curve for the M-method with a sample size $n_M = 25$ is depicted as the blue line. The red line gives the simulated OC curve for Gupta's method when the sample size $n_G = 39$. It can be seen that the simulated OC curves are matched when $n_M < n_G$. According to Schilling and Neubauer (2017), the performance of the specific sampling plan becomes better with increasing the sample size. Since the performance of Gupta's method plan with large sample sizes is matched to the M-method plan with small sample sizes in this comparison, we can establish the improved power of the M-method over Gupta's method for small sample sizes. For the second case of matching the M-method and Gupta's method plans with unequal sample sizes, we show the resulting design constants of the M-method and Gupta's method plans under the same censoring time $t_M = t_G = 25.5$ and

Table 3.1: *The true fraction of nonconforming p of different comparisons*

Comparisons	True fraction of nonconforming p
M-method vs Gupta's method	
Equal sample sizes	(0.0001,0.0005,0.005,0.01,0.014,0.016,0.02, 0.025,0.03,0.035,0.04,0.045,0.05,0.055,0.06,0.065, 0.085,0.1,0.125,0.13,0.15,0.17,0.19,0.20,0.25,0.30.)
Unequal sample sizes	(0.0001,0.0005,0.005,0.01,0.015,0.02,0.025, 0.03,0.04,0.05,0.065,0.085,0.1,0.125,0.15, 0.175,0.20,0.25.)
M-method vs Wilrich's method	
Equal sample sizes	(0.0001,0.0005,0.01,0.015,0.02,0.025,0.03,0.035, 0.04,0.041,0.044,0.045,0.05,0.052,0.060,0.063, 0.065,0.070,0.071,0.072,0.074,0.745,0.075,0.0755, 0.077,0.085,0.090,0.1.)
Unequal sample sizes	(0.0001,0.0005,0.005,0.01,0.015,0.02,0.025, 0.03,0.04,0.05,0.055,0.060,0.065,0.085,0.1,0.125, 0.15,0.175,0.20,0.25,0.35.)

lower specification limit $L_M = L_G = 1.25$ with different fixed parameter σ in Table 3.2. This again shows the advantages of the M-method plans. For further comparisons, a web-based Shiny app is given at https://zoe1994.shinyapps.io/shiny_for_comparing_m_with_gupa/, which helps to implement and test the efficacy of matching the M-method and Gupta's method plans with different combinations of parameters.

Table 3.2: *Matched M-method and Gupta's method plans*

Parameter σ	M-method plan		Gupta's method plan	
	n	M	n	c
3.83	10	0.09	15	1
1.33	20	0.08	30	2
0.90	25	0.06	39	2
1.05	30	0.07	47	3
0.90	50	0.05	88	4
0.99	100	0.06	174	10

To compare the M-method and Wilrich's method plans, we suppose that the data comes from the Weibull distribution and then fit the Weibull distribution by using MLE. As an example, we fix the shape parameter $k = 3.02$ and change the scale parameter λ. The scale parameter λ is calculated from the function of $L_M/(-\log(1 - true\ p))^{(1/k)}$. Table 3.1 illustrates these given values of true fraction nonconforming p. We set the censoring time

$t_M = t_W = 61.52$, and lower specification limit $L_M = L_W = 3.29$. Following almost the same steps as in the comparison of the M-method and Gupta's method plans, the simulated OC curves of the M-method and Wilrich's method plans were obtained, and these are shown in Figure 3.2, where P_a is plotted against the true fraction of nonconforming p. As shown in Figure 3.2(a), the simulated OC curves of the M-method and Wilrich's method plans are overlapping for small but unequal sample sizes. The blue line gives the simulated OC curve for Wilrich's method for the sample size $n_W = 21$. The red line gives the simulated OC curve for the M-method when the sample size $n_M = 20$. Two nearly matching simulated OC curves were obtained for many combinations of $n_M < n_W$ we investigated, but n_M was only slightly smaller than n_W. Hence we conclude that the M-method is marginally superior to Wilrich's method.

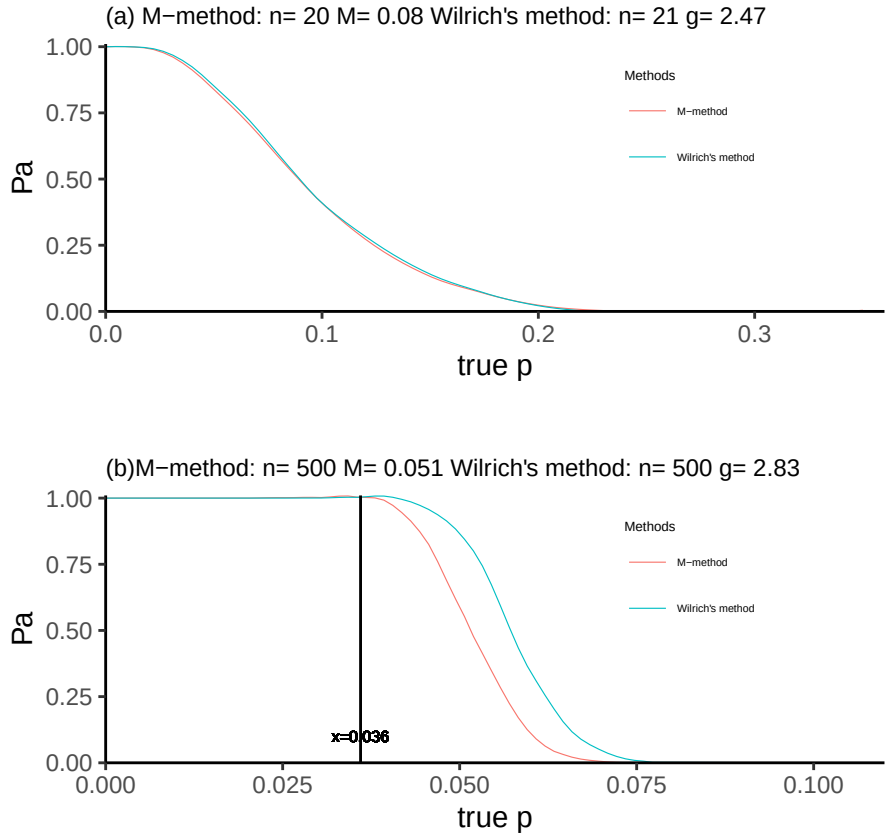

Figure 3.2: *Comparison between M-method and Wilrich's method sampling plans.*

Wilrich (2018) designed a variables sampling plan for Weibull distribution for two designated points of the OC curve. In Wilrich's method, the test statistic requires the

estimated shape and scale parameters. Since the shape parameter is difficult to estimate precisely for small sample sizes, the impact of this uncertainty on the producer's and consumer's risks can be greater. Therefore, it is not fair to make a comparison of the M-method with the Wilrich's method when the sample sizes n_W and n_M are set equal but small. However, we can set the sample sizes n_W and n_M as equal provided these are large. In Figure 3.2(b), the M-method plan is compared with Wilrich's plan for a large sample of size $n_M = n_W = 500$. The performance of the M-method is somewhat better than the Wilrich's plan at poor quality levels. On the other hand, the probability of acceptance P_a for the M-method plan and Wilrich's plan is very similar at good quality levels, except for a slight difference before true $p = 0.036$. In addition, the black vertical line shows that the estimated OC curves of the sampling plan for two methods intersect at the point of true $p = 0.036$. Therefore, we can conclude that the M-method is slightly more discriminatory than Wilrich's method for very large sample sizes.

3.5 Conclusions

In this chapter, we designed a variables sampling plan for small sample sizes, which is called the M-method sampling plan. We also compared the M-method sampling plan with Gupta's method plan for lognormal and Wilrich's method plans for Weibull distribution.

Based on the examples given, the M-method sampling plan is advantageous for small sample sizes. The M-method sampling plan can become more discriminatory when the parent distribution has no shape parameter, i.e., the lognormal distribution. The M-method sampling plan is sensitive to reject at poor quality levels for small sample sizes. However, when the parent distribution has shape parameters, i.e., the Weibull distribution, it is difficult to detect the superiority of the M-method plan with small sample sizes. This is because we are unable to get enough information on the shape of small sample sizes. Hence, the sample size needs to be very large to reflect the effect of the shape on the producer's and consumer's risks.

The M-method sampling plan can be used for small sample sizes as well as for different underlying distributions. In this chapter, we assumed that the parent distribution of the data is known, but the distribution parameters are assumed to be unknown. For lognormal

distribution, we compared the M-method plan with Gupta's method plan. For Weibull distribution, the M-method plan was compared with Wilrich's method plan. In both cases, the M-method plan is beneficial to control or maintain the risks. In the next chapter, the case of the underlying distribution being misspecified is studied for lifetime sampling inspection purposes.

Chapter 4

Life Test Sampling Plans under Distribution Misspecification

In Chapter 3, the life test sampling plan under various distribution specifications for small sample sizes was discussed. The performance of the M-method sampling plan for small sample sizes is satisfactory in many cases. However, there is a possibility that the underlying model becomes misspecified for life testing, particularly when special or assignable causes present in the manufacturing process. The accelerated life test under distribution misspecification was studied by Pascual and Montepiedra (2005). They presented an optimal test plan under the criterion of minimizing the asymptotic mean square error of the maximum likelihood estimator of life quantiles (the lognormal and Weibull distributions).

In this chapter, the efficacy of the M-method to resolve the distribution misspecification in the life test sampling plan is investigated. The basic idea here is to use a shapeless parameter distribution as a surrogate for distributions with a shape parameter, even if the data actually come from a distribution having a shape parameter. We show that the surrogate performs well for small sample sizes, even when the underlying distribution is misspecified. Particularly, we adopt the M-method to design the life test sampling plan to tackle the distribution misspecification. Again, we focus on type I censored data

generated from three common lifetime distributions: the lognormal, Weibull, and gamma distributions.

This chapter is organized as follows. Section 4.1 presents a table to summarize various combinations of working or assumed and the underlying or true models. We call a working model as the distributional model that is used to fit the data. Certain combinations described in Section 4.1 are discussed in Section 4.2 for small sample sizes, and Section 4.3 for large sample sizes. The application of the proposed method is demonstrated in Section 4.4. The conclusion is given in Section 4.5.

4.1 Combination of the Underlying Model and Working Model

The various combinations of the underlying model and the working model are briefly listed in this section. When the underlying model is specified correctly, the correct distribution is used as the working model to fit the data. When the underlying model is misspecified, the misspecified distribution is used as the working model. Since we focus on three different lifetime distributions, we will consider two cases: data comes from the Weibull distribution and gamma distribution.

When the data come from the Weibull distribution, three situations can be considered and given in Table 4.1. In addition, Table 4.1 summarized the combinations of the underlying model and the working model when the data come from the gamma distribution.

Table 4.1: *Combination of the underlying model and working model*

Data comes from the Weibull distribution		Data comes from the gamma distribution	
Underlying model	Working model	Underlying model	Working model
Weibull distribution	Weibull distribution	gamma distribution	gamma distribution
	gamma distribution		Weibull distribution
	lognormal distribution		lognormal distribution

For illustrative purposes, two situations will be considered in the next section: the working model is a misspecified lognormal distribution and a correctly-specified underlying distribution. Hence, the lognormal distribution is distinct from the other two distributions as it is shape parameter-free.

4.2 The Case of Small Sample Sizes

To compare the life test sampling plans under possible distribution misspecifications, we study the performance of the M-method based life test sampling plans with the correctly-specified and misspecified working models via Monte Carlo simulations. We would show that the method of using a shapeless parameter distribution as a surrogate for distributions with a shape parameter is suitable for small sample sizes.

We set up the same censoring time and the same lower specification limit for all the sampling plans. The performance of various sampling plans is summarized in their OC curves. Following almost similar simulation steps as in the comparisons of three method plans, the appropriate values of the maximum allowable fraction nonconforming M can be obtained. Type I censored data is generated from either Weibull or gamma distribution. Since the lognormal distribution is easy to estimate as a shape-free distribution, two working models are considered, i.e., the lognormal distribution as a misspecified model and the Weibull or gamma distribution as a correctly-specified model. In addition, in the case of the correctly-specified distribution, we assume that the shape parameter can be either known, unknown, or misspecified as a wrong value. In the case of small sample sizes, all the number of the Monte Carlo repetition used to simulate the estimated OC curves is $10,000$.

4.2.1 Type I Censored Data from Weibull Distribution

Suppose that the data come from the Weibull distribution, the true underlying distribution should be the Weibull distribution. However, the underlying distribution can be misspecified in practice. In the life test sampling plan, we compare the correctly-specified Weibull distribution with misspecified lognormal distribution as the working model when the sample size is small. Following the simulation steps of the M-method in Chapter 3, we

can obtain the estimated OC curves of two different working models via Monte Carlo simulation.

Unknown Shape Parameter

Assume that the shape parameter of a correctly-specified working model (the Weibull distribution) is unknown. For illustrative purposes, the type I censored data with the sample sizes $n_{Wei} = n_{LN} = 10, 15, 20, 25$ are generated from the Weibull distribution with a shape parameter $k = 0.45$. Let the censoring time $t_M = 8.05$ and the lower specification limit $L_M = 2.00$. The values of the maximum allowable fraction nonconforming M are shown in Table 4.2. The estimated OC curves are shown in Figure 4.1, where the probability of acceptance P_a is plotted against the true fraction of nonconforming p. As shown in Figure 4.1, the estimated OC curves of the misspecified lognormal distribution and the correctly-specified Weibull distribution are matched for small sample sizes. Therefore, we can conclude that the Weibull distribution with the unknown shape parameter can be replaced by the lognormal distribution as the working model when designing a life test sampling plan for small sample sizes.

Table 4.2: *The values of maximum allowable fraction nonconforming M based on the Weibull data for small sample sizes*

Shape parameter k	Sample size n	Maximum allowable fraction nonconforming M	
		Weibull distribution	Lognormal distribution
Unknown	10	0.080	0.075
	15	0.070	0.065
	20	0.033	0.029
	25	0.022	0.020
Known-true/misspecified	10	0.095	0.030
	15	0.095	0.055
	20	0.095	0.075
	25	0.087	0.065

Known Shape Parameter

Assume that the shape parameter of a correctly-specified working model (the Weibull distribution) is either known or misspecified as a wrong value. The type I censored data with the sample sizes $n_{Wei} = n_{LN} = 10, 15, 20, 25$ are generated from the Weibull distribution with a shape parameter $k = 0.45$. Set up the censoring time $t_M = 8.05$ and

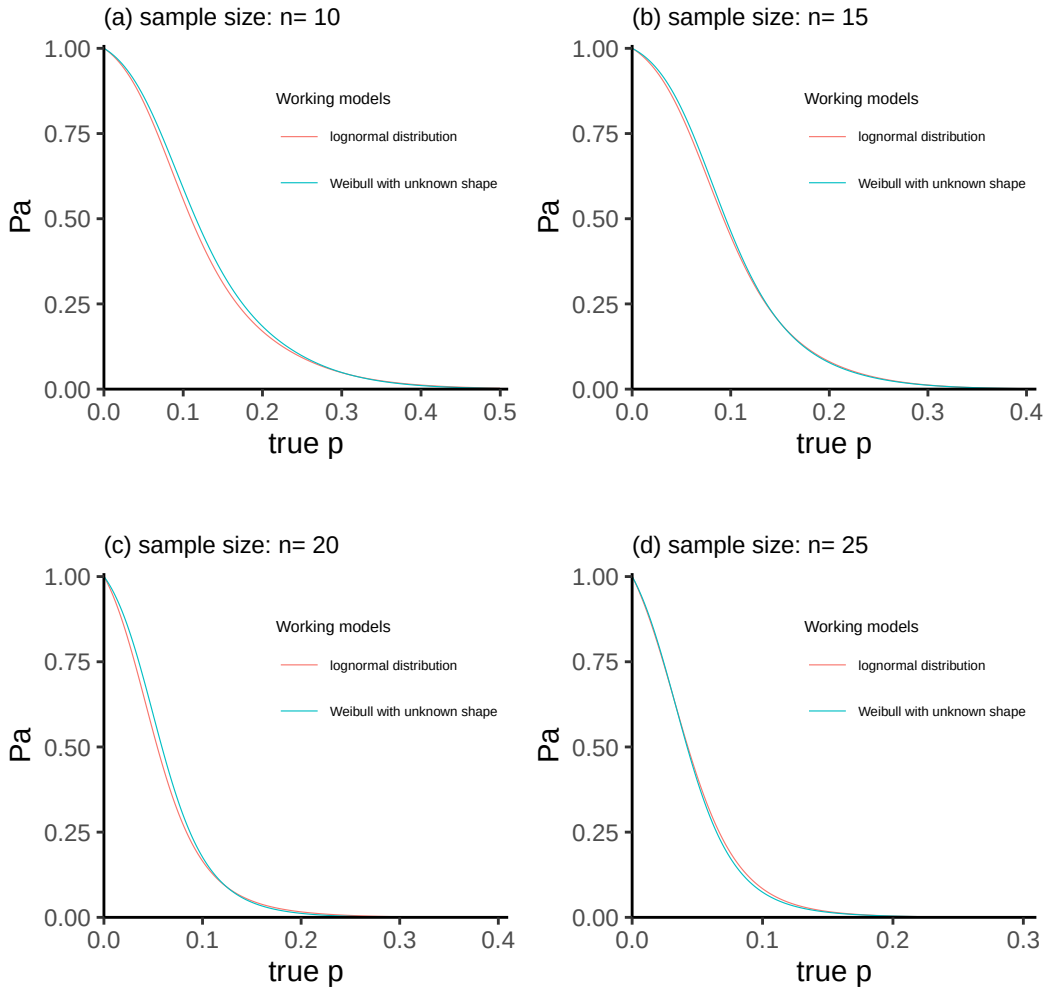

Figure 4.1: *OC curves of the working models with unknown shape parameter for small sample sizes (data comes from the Weibull distribution)*

the lower specification limit $L_M = 2.00$. The values of the maximum allowable fraction nonconforming M are shown in Table 4.2.

The estimated OC curves are shown in Figure 4.2, where the probability of acceptance P_a is plotted against the true fraction of nonconforming p. The point of intersection is through the black vertical line. The estimated OC curve of the misspecified lognormal distribution is depicted as the red line. The estimated OC curve of the correctly-specified Weibull distribution with the true shape parameter is depicted as the blue line. The estimated OC curve of the correctly-specified Weibull distribution with the misspecified shape parameter is depicted as the green line. It can be seen that the green line is strikingly different from red and blue lines. In other words, although the shape parameter is known, the sampling

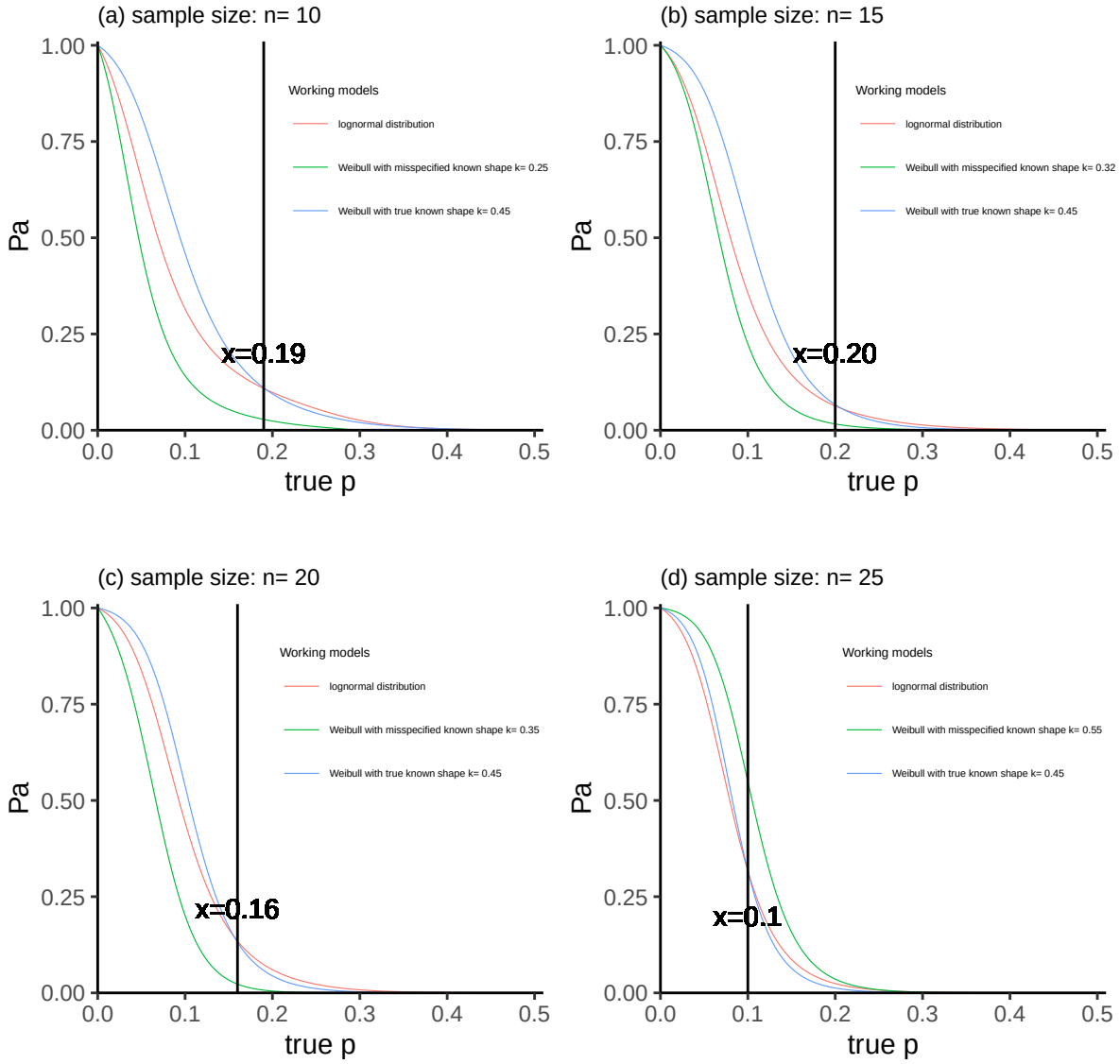

Figure 4.2: *OC curves of the working models with true and misspecified known shape parameter for small sample sizes (data comes from the Weibull distribution)*

plan is ineffective if the shape parameter is misspecified. As the sample size increases, the difference between the red and blue lines is decreasing. In addition, the tails of the red and blue lines look somewhat similar. In designing a life test sampling plan for small sample sizes, we can conclude that the method of using the lognormal distribution to replace the Weibull distribution does not perform well when the shape parameter of the correctly-specified Weibull model is assumed to be known.

As a result, the lognormal distribution can be used instead of the Weibull distribution when the sample size is small. In addition, it works well when the shape parameter of the Weibull model is assumed to be unknown. In other words, for small sample sizes, though the specified underlying distribution should be the Weibull distribution when the data come from the Weibull distribution, we can directly use the lognormal distribution as the underlying distribution and the working model to fit Weibull data.

4.2.2 Type I Censored Data from Gamma Distribution

Suppose that the data come from the gamma distribution, the true underlying distribution is the gamma distribution. The underlying distribution has a chance of being misspecified in practice. We assume the working model is the correctly-specified gamma distribution and the misspecified lognormal distribution, and compare them under small sample sizes. Follow the simulation steps of the M-method in Chapter 3 and calculate the estimated OC curves of two different working models.

Unknown Shape Parameter

Assume that the shape parameter of a correctly-specified working model (the gamma distribution) is unknown. For illustrative purposes, the type I censored data with the sample sizes $n_{Gam} = n_{LN} = 10, 15, 20, 25$ are generated from the gamma distribution with a shape parameter $\omega = 9.00$. We set the censoring time $t_M = 35$ and the lower specification limit $L_M = 1.15$. In addition, Table 4.3 shows the values of the maximum allowable fraction nonconforming M. As shown in Figure 4.3, the estimated OC curves of the misspecified lognormal distribution and the correctly-specified gamma distribution are similar when the sample size is small. Therefore, we suggest that the gamma distribution with the unknown shape parameter can be replaced by the lognormal distribution as the working model when designing a life test sampling plan for small sample sizes.

Known Shape Parameter

Assume that the correctly-specified working model is gamma distribution, and the shape parameter is assumed to be known as a true value and a misspecified value. Type I censored data with the sample sizes $n_{Gam} = n_{LN} = 10, 15, 20, 25$ are generated from the

Table 4.3: *The values of maximum allowable fraction nonconforming M based on the gamma data for small sample sizes*

Shape parameter ω	Sample size n	Maximum allowable fraction nonconforming M	
		Gamma distribution	Lognormal distribution
Unknown	10	0.060	0.046
	15	0.075	0.063
	20	0.087	0.080
	25	0.080	0.070
Known-true/misspecified	10	0.098	0.030
	15	0.081	0.035
	20	0.10	0.060
	25	0.10	0.065

gamma distribution with a shape parameter $\omega = 9.00$. Given the values of the censoring time $t_M = 35$ and the lower specification limit $L_M = 1.15$. In addition, shown the values of the maximum allowable fraction nonconforming M (Table 4.3).

Figure 4.4 illustrates the estimated OC curves of three different working models. The intersection point is going through the black vertical line. The estimated OC curve of the misspecified lognormal distribution is depicted as the blue line. The estimated OC curve of the correctly-specified gamma distribution with the true shape parameter is depicted as the green line. The estimated OC curve of the correctly-specified gamma distribution with the misspecified shape parameter is depicted as the red line. It is observed from Figure 4.4 that the red line is significantly different from green and blue lines. The difference between the tail of the green and blue lines is smaller. In other words, the consumer's risk of two lines is roughly met. The performance of the estimated OC curve does not fare satisfactorily when the known shape parameters are misspecified. Hence, it can be conservatively concluded that if we know the shape parameters, but if these are misspecified and result in the poor performance, we would better of with ignorance of the shape parameters instead.

As a result, the lognormal distribution can be used instead of the gamma distribution when the sample size is small. In addition, it works well when the shape parameter of the gamma model is assumed to be unknown. In other words, for small sample sizes, though the specified underlying distribution should be the gamma distribution when the data come from the gamma distribution, we can directly use the lognormal distribution as the underlying distribution and the working model to fit gamma data.

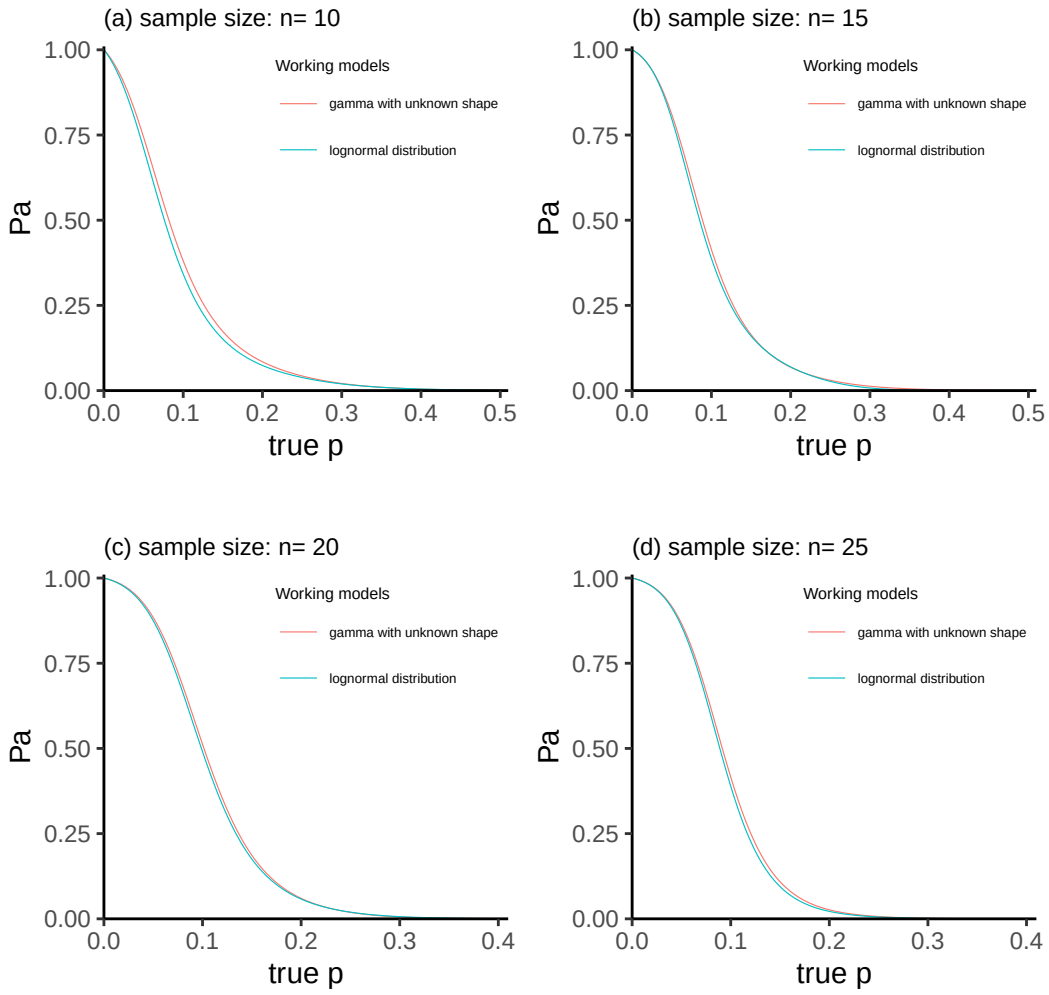

Figure 4.3: *OC curves of the working models with unknown shape parameter for small sample sizes (data comes from the gamma distribution)*

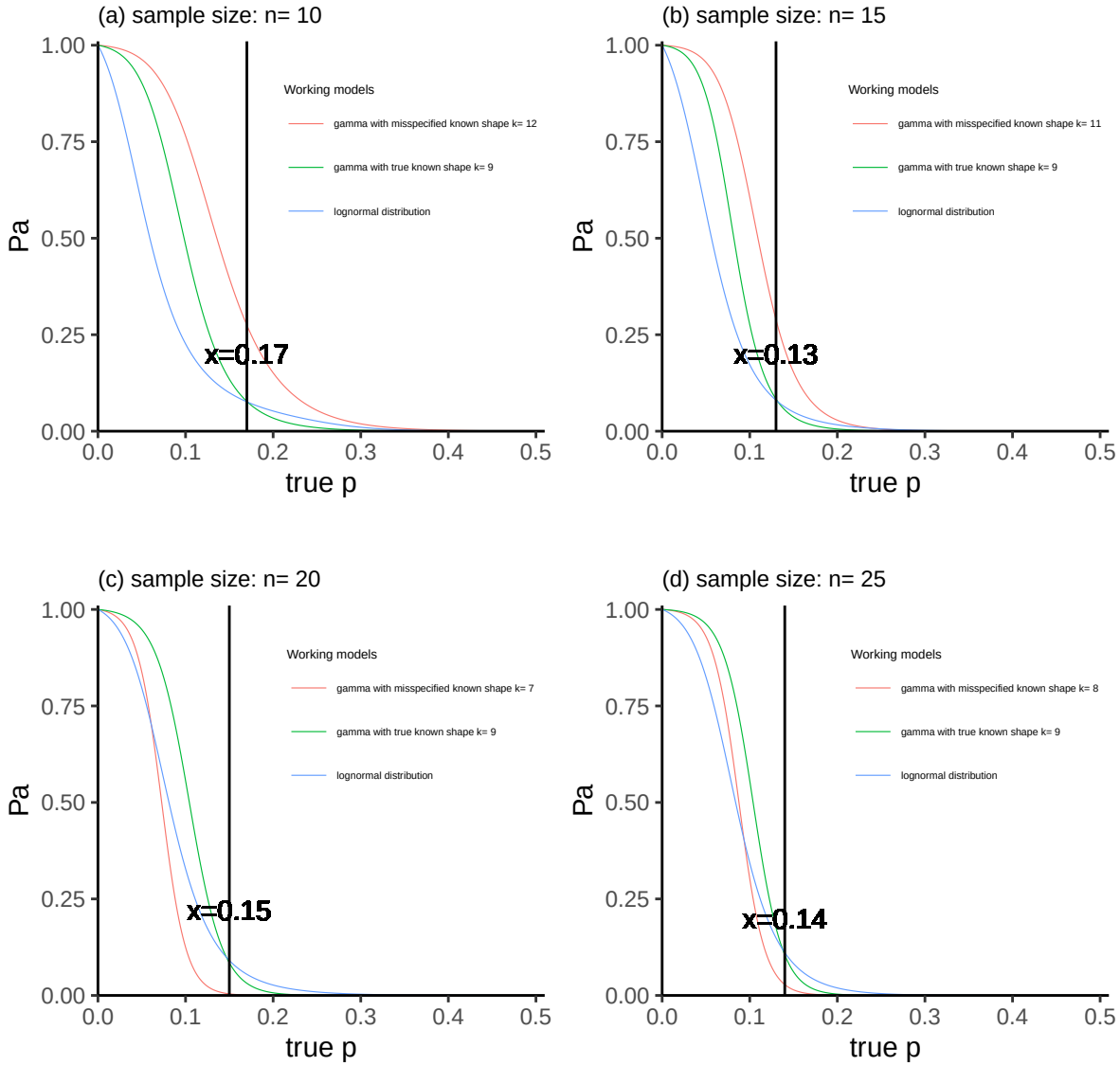

Figure 4.4: *OC curves of the working models with true and misspecified known shape parameter for small sample sizes (data comes from the gamma distribution)*

4.3 The Case of Large Sample Sizes

For large sample sizes, we make the comparisons under possible distribution misspecification and study the performance of the life test sampling plan with the correctly-specified and misspecified working models. We would like to show that the M-method is not suitable for large sample sizes when the lifetime distribution is misspecified, no matter what value of the shape parameter is assumed to be.

Following the same conditions in Section 4.2, and plotting the estimated OC curves to show the performance of various sampling plans at large sample sizes. In addition, the same simulation steps described in section 4.2 was followed to search the appropriate values of the maximum allowable fraction nonconforming M for the comparisons. In the study, the type of data is the type I censored data and still generated from either Weibull or gamma distribution. Do the same repetitions 10,000 for each situation via Monte Carlo simulation. In addition, the working model considered and the shape parameter of the correctly-specified distribution assumed in Section 4.3 are the same as the case of small sample sizes discussed in Section 4.2.

4.3.1 Type I Censored Data from Weibull Distribution

Suppose that the parent distribution of the data is Weibull distribution. For the working model, we assume it is the correctly-specified Weibull distribution and the misspecified lognormal distribution, and compare them in the life test sampling plan. Follow the steps in simulations for the M-method described in Chapter 3 and draw the estimated OC curves of the two working models.

Unknown Shape Parameter

Assume that the shape parameter of the correctly-specified Weibull working model is unknown. For illustrative purposes, type I censored data with the sample sizes $n_{Wei} = n_{LN} = 150, 250, 350, 450$ are generated from the Weibull distribution with a shape parameter $k = 0.45$. Given the values of the censoring time $t_M = 8.05$ and the lower specification limit $L_M = 2.00$. In addition, the values of the maximum allowable fraction nonconforming M are the same as the searching in Section 4.2.1 at the unknown shape

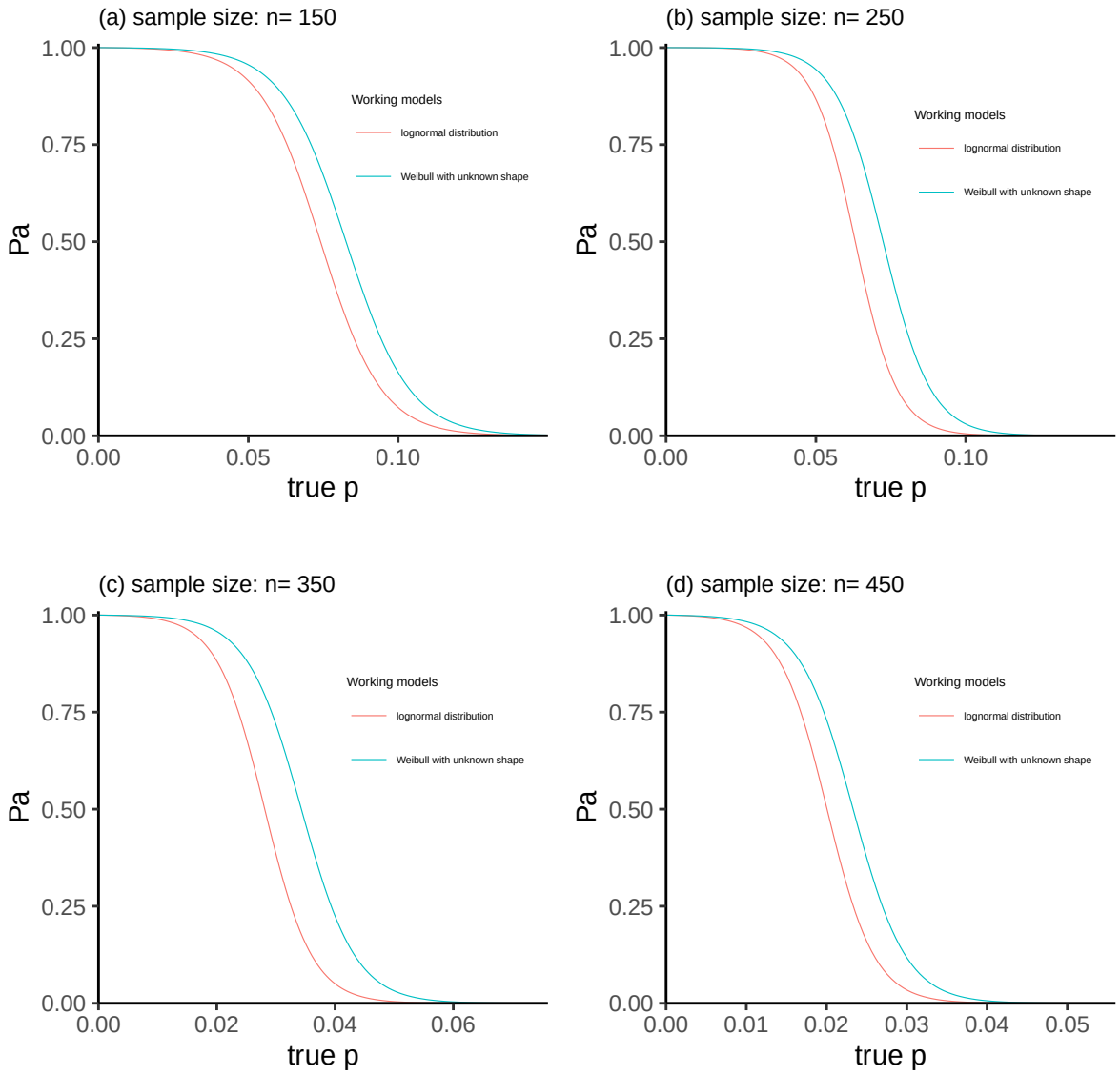

Figure 4.5: *OC curves of the working models with unknown shape parameter for large sample sizes (data comes from the Weibull distribution)*

parameter (Table 4.2). The estimated OC curves are shown in Figure 4.5, where the probability of acceptance P_a is plotted against the true fraction of nonconforming p. As shown in Figure 4.5, the estimated OC curves of the misspecified lognormal distribution and the correctly-specified Weibull distribution are different for large sample sizes.

Known Shape Parameter

Assume that the shape parameter of a correctly-specified Weibull working model is either known or misspecified as a wrong value. Generate the type I censored data with the

sample sizes $n_{Wei} = n_{LN} = 150, 250, 350, 450$ from the Weibull distribution with a shape parameter $k = 0.45$. Again, set up the censoring time $t_M = 8.05$ and the lower specification limit $L_M = 2.00$. In addition, the values of the maximum allowable fraction nonconforming M are shown in Table 4.4. Figure 4.6 describes the estimated OC curves of three different working models. As illustrated, one can clearly distinguish the three estimated OC curves.

Table 4.4: *The values of maximum allowable fraction nonconforming M based on the Weibull data for large sample sizes*

Shape parameter k	Sample size n	Maximum allowable fraction nonconforming M	
		Weibull distribution	Lognormal distribution
Known-true/misspecified	150	0.095	0.080
	250	0.095	0.089
	350	0.095	0.085
	450	0.087	0.075

As a result, the method presented in Section 4.2 that use the lognormal distribution to replace the Weibull distribution is not suitable for large sample sizes under distribution misspecification, even if the shape parameter of the Weibull working model is assumed to be unknown.

4.3.2 Type I Censored Data from Gamma Distribution

Suppose that the parent distribution of the data is gamma distribution. For the working model, we assume it is the correctly-specified gamma distribution and the misspecified lognormal distribution, and compare them in the life test sampling plan. Design the M-method based life test sampling plan and follow the Monte Carlo simulation steps to calculate the estimated OC curves of the two working models.

Unknown Shape Parameter

Assume that the shape parameter of the correctly-specified working model (the gamma distribution) is unknown. For illustrative purposes, type I censored data with the sample sizes $n_{Gam} = n_{LN} = 150, 250, 350, 450$ are generated from the gamma distribution with a shape parameter $\omega = 9.00$. We set the censoring time $t_M = 35.00$ and the lower specification limit $L_M = 1.15$. Table 4.5 notes the values of maximum allowable fraction

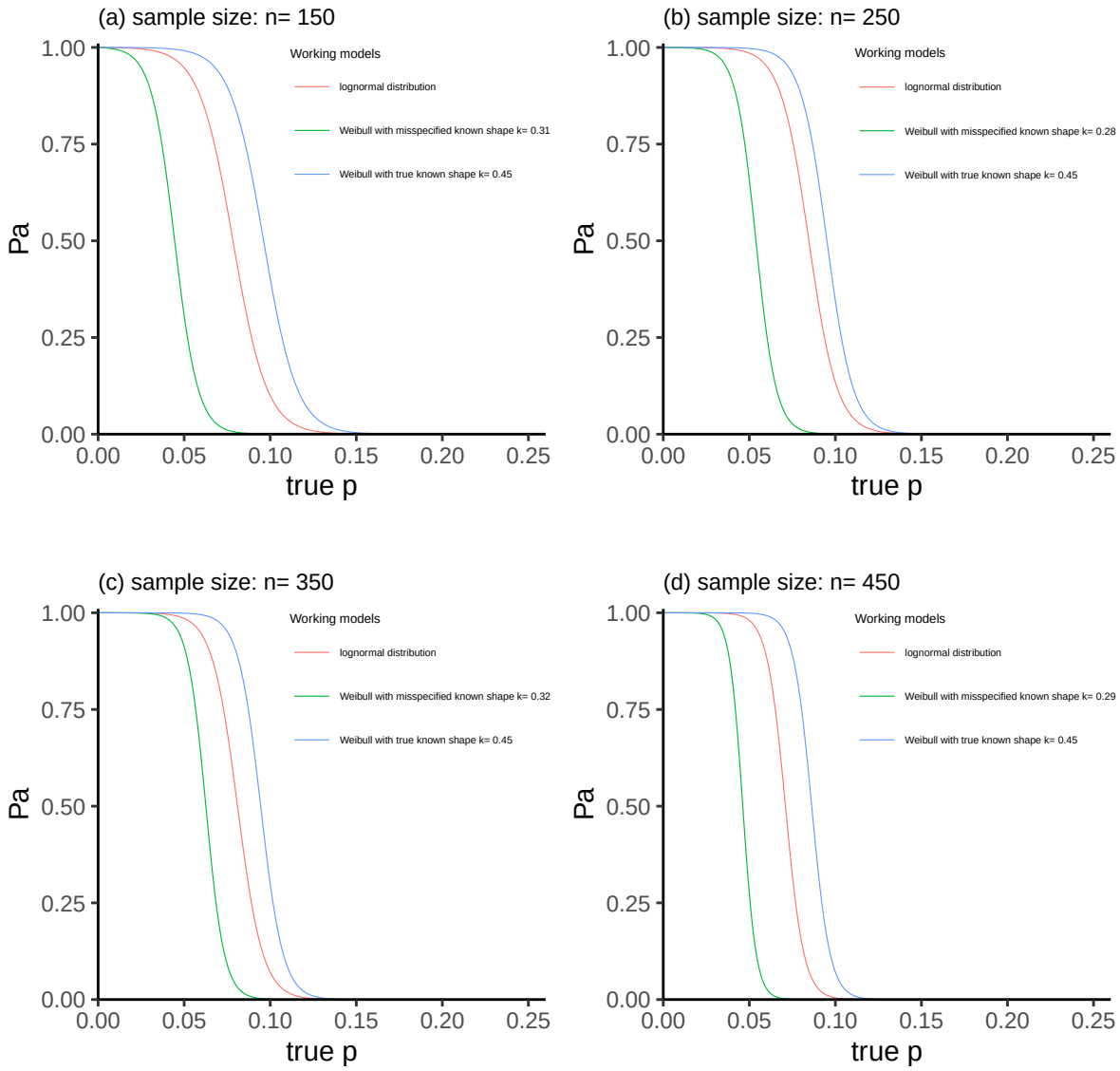

Figure 4.6: *OC curves of the working models with true and misspecified known shape parameter for large sample sizes (data comes from the Weibull distribution)*

nonconforming M. As shown in Figure 4.7, the estimated OC curves of the misspecified lognormal distribution and the correctly-specified gamma distribution are discriminatory when the sample size is large.

Known Shape Parameter

Assume that the shape parameter of a correctly-specified gamma working model is either known or misspecified as a wrong value. Generate the type I censored data with the sample sizes $n_{Gam} = n_{LN} = 150, 250, 350, 450$ from the gamma distribution with a shape

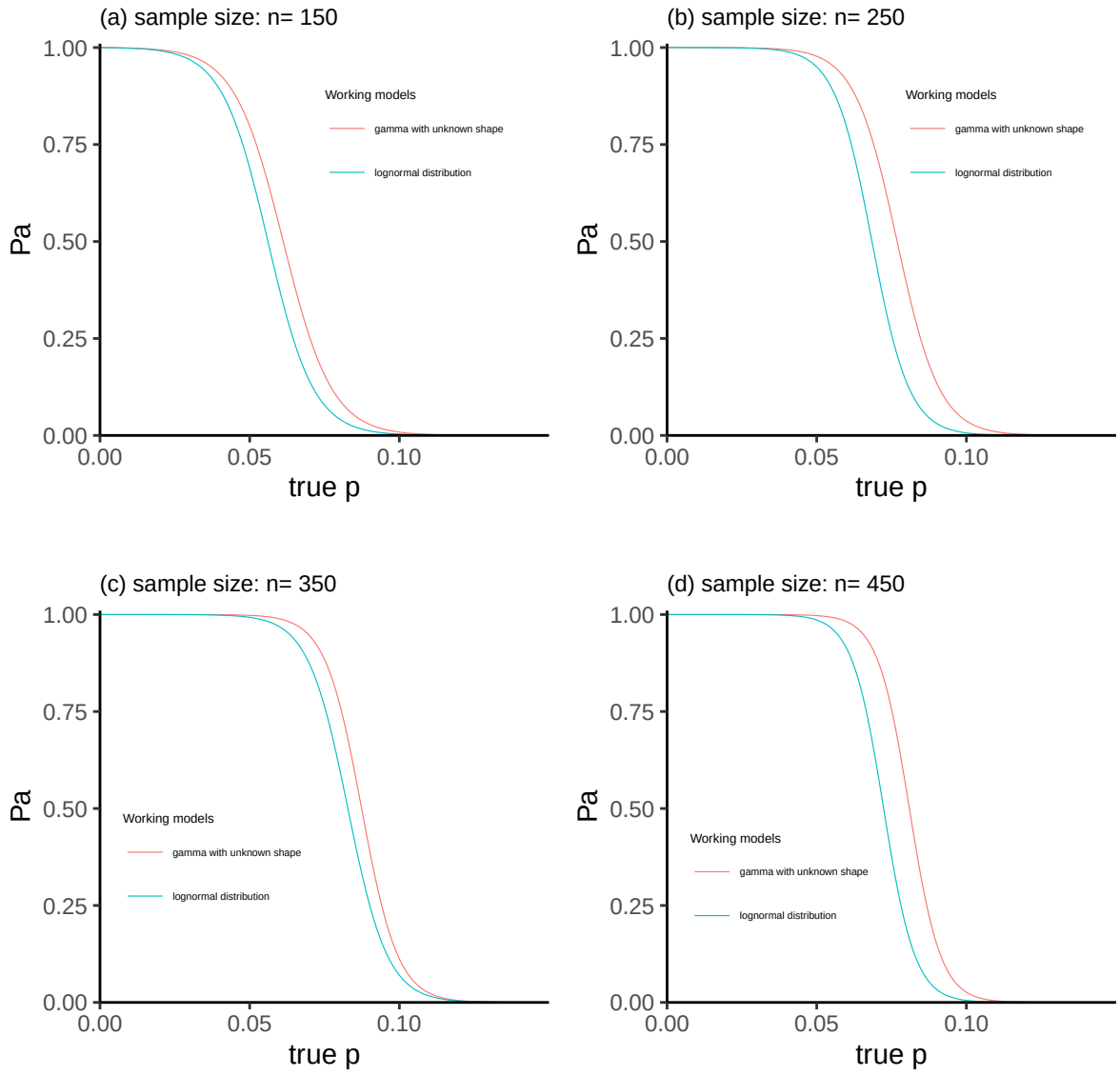

Figure 4.7: *OC curves of the working models with unknown shape parameter for large sample sizes (data comes from the gamma distribution)*

parameter $\omega = 9.00$. Again, let the censoring time $t_M = 35$ and the lower specification limit $L_M = 1.15$. In addition, the values of the maximum allowable fraction nonconforming M are shown in Table 4.5. Figure 4.8 describes the estimated OC curves of three different working models. As illustrated, it can be found that the difference between them is substantial.

Table 4.5: *The values of maximum allowable fraction nonconforming M based on the gamma data for large sample sizes*

Shape parameter ω	Sample size n	Maximum allowable fraction nonconforming M	
		Gamma distribution	Lognormal distribution
Unknown	150	0.060	0.045
	250	0.075	0.060
	350	0.087	0.075
	450	0.080	0.065
Known-true/misspecified	150	0.098	0.030
	250	0.081	0.089
	350	0.10	0.060
	450	0.10	0.065

As a result, under distribution misspecification, the method presented in Section 4.2 cannot be used when the data come from the gamma distribution for large sample sizes, even if the shape parameter of the gamma model is assumed to be unknown.

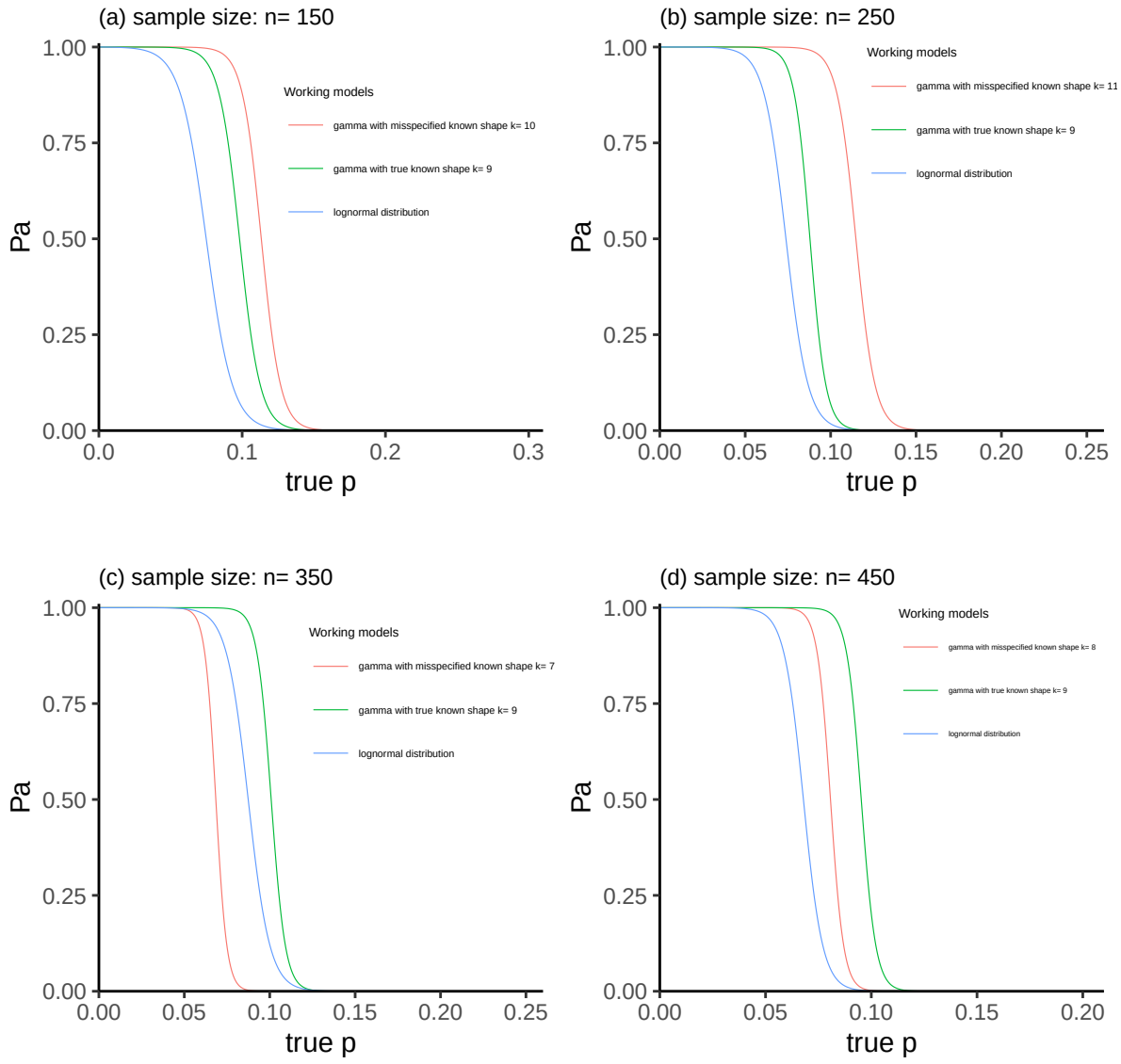

Figure 4.8: *OC curves of the working models with true and misspecified known shape parameter for large sample sizes (data comes from the gamma distribution)*

4.4 Case Study

In the previous section, we design the M-method based life test sampling plan and compare the estimated OC curves of the misspecified working model and the correctly-specified working model. For the life test sampling plan under distribution misspecification, we demonstrate that a shapeless parameter distribution could be used as the underlying distribution instead of the distribution with a shape parameter when the sample size is small. Since we focus on the three common lifetime distributions, we can use the

lognormal distribution as a surrogate to the Weibull distribution and gamma distribution. In this section, we provide an example to illustrate the implementation of the M-method based life test sampling plan under distribution misspecification for small sample sizes.

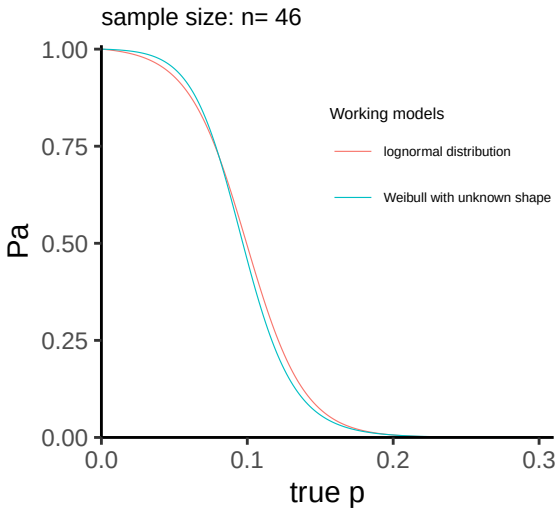

Figure 4.9: *OC curves of the working models with unknown shape parameter (data comes from the Weibull distribution)*

Jun, Balamurali, and Lee (2006) discussed the design of variables sampling plans under Weibull distribution with the known shape parameter $k = 2$ by using the two-point approach. They provided an example for a single sampling plan with the sample size $n_{Wei} = 46$, the lots with acceptable reliability level $p_0 = 0.05$ are accepted with probability $1 - \alpha_{Jun\ et\ al.} = 0.95$, and the lots with lot tolerance reliability level $p_1 = 0.20$ are accepted with probability $\beta_{Jun\ et\ al.} = 0.10$.

By following this example, we assume the working distribution either to be the correctly-specified Weibull distribution or the misspecified lognormal distribution. For illustrative purposes, we compare the estimated OC curves of these two working distributions in designing the M-method based life test sampling plan. In the study, the type I censored data is generated from the Weibull distribution with a shape parameter $k = 2.00$. In addition, the shape parameter is assumed to be unknown for the correctly-specified working model. Given the values of $n_{Wei} = n_{LN} = 46$, $t_M = 8.80$, and $L_M = 1.10$. Search the values of the maximum allowable fraction nonconforming M by following the simulation steps described in Section 3.4, to obtain $M_{Wei} = 0.092$, and $M_{LN} = 0.08$. The

number of Monte Carlo repetitions is 10,000. The estimated OC curves for two working distributions are shown in Figure 4.9. As shown in Figure 4.9, it can be seen that the estimated OC curves look somewhat similar. Therefore, we can conclude that the method of using the lognormal distribution as a surrogate to the Weibull distribution performs well under small sample sizes. In the estimated OC curve of the Weibull distribution, the true fraction of nonconforming $p = 0.05$ corresponding to $1 - \alpha_{M-method} = 0.95$, which is equal to $1 - \alpha_{Jun\ et\ al.} = 0.95$. However, the true fraction of nonconforming $p = 0.20$ corresponding to $\beta_{M-method} = 0.006$, which is less than $\beta_{Jun\ et\ al.} = 0.01$. Hence, the superiority of the M-method sampling plan in small sample sizes is again found valid.

4.5 Conclusions

In this chapter, we studied the M-method to resolve the distribution misspecification in the life test sampling plan and demonstrated that it could be used for small sample sizes. Since the lognormal distribution is functionally different from the Weibull and gamma distributions when the sample size is small, we assumed that the misspecified working model to be the lognormal distribution and compared with the correctly-specified Weibull or gamma working models. In addition, for the correctly-specified working model, we also considered three cases involving the shape parameter: it is assumed to be unknown and either known with a true or misspecified value.

We compared the M-method based life test sampling plan under two situations: small and large sample sizes. Based on the comparison studies done, we assert that the estimated OC curves of the correctly-specified working model and the misspecified working model can be matched well for small sample sizes. In addition, the method performs well under the correctly-specified working model with an unknown shape parameter. Hence, the M-method that uses the lognormal distribution as a surrogate to the Weibull and gamma distributions is suitable for small sample sizes.

Chapter 5

General Conclusions and Future Perspectives

5.1 Conclusions

Lifetime distributions are popular in life testing problems, including but not limited to, biomedical and reliability engineering. In many applications, large data can be collected to identify the underlying lifetime distribution. However, it is difficult to identify the parent distribution for small sample sizes. The M-method sampling plan design proposed in this thesis is suitable for small sample sizes. The M-method plan that uses a simple distribution (without shape parameter) to replace a complicated distribution (with shape parameter) is studied to resolve the distribution-data identification issue for small sample sizes. Particularly, three common lifetime distributions are considered, i.e., the lognormal, Weibull, and gamma distributions.

In Chapters 1 and 2, we presented some common grounds and pertinent differences between the lognormal, Weibull, and gamma distributions. Particularly, the lognormal distribution has no shape parameter, but the Weibull distribution and gamma distribution have the shape parameters. For small sample sizes, the shape parameters of the Weibull and gamma distributions are difficult to estimate precisely. Hence, we considered using the shapeless distribution (lognormal distribution) to match the distributions with the

shape parameters (Weibull and gamma distributions). It should be noted this study is limited to type I censoring, and the estimation approach used for the analysis is the maximum likelihood estimation only.

In Chapter 3 and Chapter 4, we designed the M-method life test sampling plan, which can be applied to the small sample sizes. In addition, the performances of the M-method, Gupta's method, and Wilrich's method are compared under the correctly specified model assumption. It was shown that the M-method is easy to use, and the simulation results provided an appealing control of the consumer's risk. For the life test sampling plan under distribution misspecification, we demonstrated via Monte Carlo simulation that the method of using the lognormal distribution instead of the Weibull distribution and gamma distribution could be efficient for small sample sizes. In other words, the lognormal distribution can be employed to approximate the Weibull and gamma distributions to deal with the distribution-data identification issue for small sample sizes. When life testing is done with small samples, such as 10, this method can be used to protect the consumer's interest because the consumer's risks under these three distributions are similar.

5.2 Future Study

In our future research, we plan to focus on getting into more in-depth theory arising from the current results. There are some possible extensions that are presented in this section.

As shown in Chapter 2, we investigated the parametric methods for three lifetime distributions and focused on type I censored data in our study. However, there are other data types, such as interval-censored data and left-censored data, which can also arise from life testing. It should be useful to investigate other different kinds of data for the design of plans with small sample sizes. In addition, the numeric study of the parameter estimation is done by using the `fitdistrplus` package in R. However, there are other packages that can be chosen in R to accomplish the parameter estimation, i.e., the `likelihood` package. To further investigate, this study can either replace with another package or create a new package in R.

In Chapter 3, we only consider one situation for the shape parameter of the Weibull distribution $k > 1$ to do the simulation. In Chapter 4, we also only consider one situation for the shape parameter of the Weibull and gamma distributions, i.e., $k < 1$ and $\omega > 1$. The choice of shape parameters of the lifetime distribution is limited in our study. In our future work, we do a more in-depth study of the range values of k and ω.

As shown in Chapter 4, we considered the shape parameter to be known as a true value and a misspecified value. In this study, the distance between the true and misspecified values is small. In other words, the M-methods works well when the misspecified value is close to the true value. However, there are other situations that can be considered when we do the simulation, i.e., the distance is large. It is more interesting to study the relationship between the deviation level (distance) and the performance of the M-method, but it is much harder.

Ideally, the design of the life test sampling plan under model uncertainties must combine the ideas presented and established in Chapters 3 and 4 in an aleatory manner. The numeric study was done with random type I censored data with either equal or unequal sample sizes in comparisons. It could be interesting to extend the situation to fix the total sample size and apply model averaging methods, which are common in other fields but not in acceptance sampling.

Appendix A

SHINY APP

The following steps are followed to implement the matching between M-method and Gupta's method plans with the Shiny app hosted at https://zoe1994.shinyapps.io/shiny_for_comparing_m_with_gupa/.

- Step 1: Select parameters for the M-method sampling plan.
- Step 2: Click the "View plot" button, and the result will be displayed on the right-side panel, see Figure A.1.

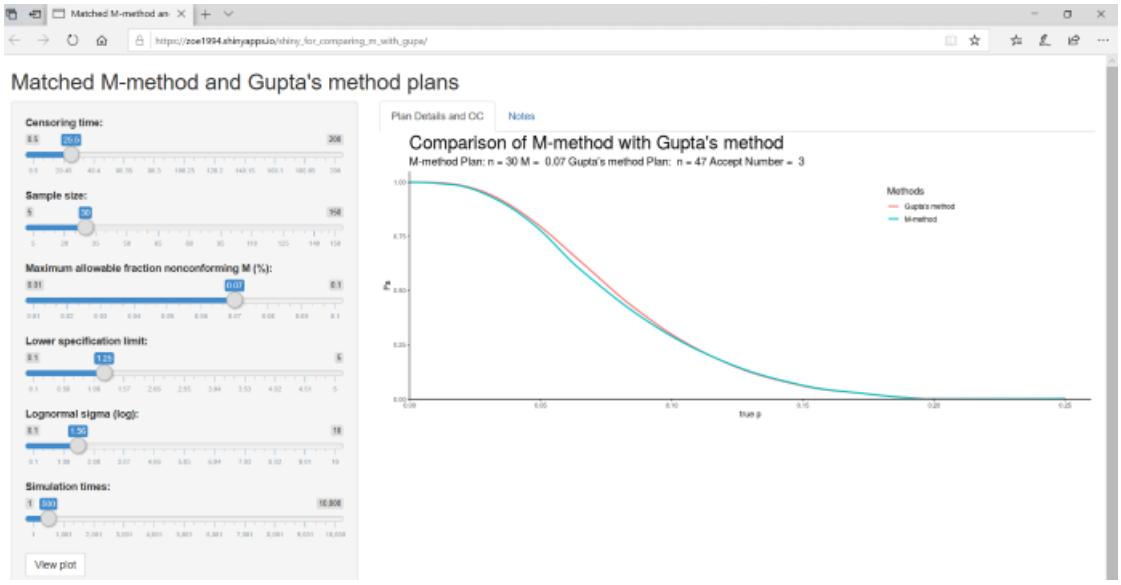

Figure A.1: *Matched M-method and Gupta's method plans*

Bibliography

Aksoy, H (2000). Use of gamma distribution in hydrological analysis. *Turk J Engin Environ Sci* **24**(6), 419–428.

Amitava, M (2008). *Fundamentals of quality control and improvement*. 3rd. Hoboken, New Jersey.

Baklizi, A and AEQE Masri (2004). Acceptance sampling based on truncated life tests in the Birnbaum Saunders model. *Risk Analysis* **24**(6), 1453–1457.

Balasooriya, U and N Balakrishnan (2000). Reliability sampling plans for lognormal distribution, based on progressively-censored samples. *IEEE Transactions on Reliability* **49**(2), 199–203.

Balasooriya, U and SLC Saw (1998). Reliability sampling plans for the two-parameter exponential distribution under progressive censoring. *Journal of Applied Statistics* **25**(5), 707–714.

Balasooriya, U, SL Saw, and V Gadag (2000). Progressively censored reliability sampling plans for the Weibull distribution. *Technometrics* **42**(2), 160–167.

Boland, PJ (2007). *Statistical and probabilistic methods in actuarial science*. 1st. Chapman and Hall/CRC, FL.

Casella, G and RL Berger (2002). *Statistical inference*. 2nd. Pacific Grove, CA.

Chen, J, W Chou, H Wu, and H Zhou (2004). Designing acceptance sampling schemes for life testing with mixed censoring. *Naval Research Logistics* **51**(4), 597–612.

Choi, SC and R Wette (1969). Maximum likelihood estimation of the parameters of the gamma distribution and their bias. *Technometrics* **11**(4), 683–690.

Dey, AK and D Kundu (2009). Discriminating among the log-normal, Weibull, and generalized exponential distributions. *IEEE Transactions on Reliability* **58**(3), 416–424.

Dubey, SD (1967). Normal and Weibull distributions. *Naval Research Logistics Quarterly* **14**(1), 69–79.

Epstein, B (1954). Truncated life test in the exponential case. *The Annals of Mathematical Statistics* **25**(3), 555–564.

Goode, HP and JHK Kao (1960). *Sampling plans based on the Weibull distribution*. 1st. NY: Cornell University, Ithaca. https://apps.dtic.mil/dtic/tr/fulltext/u2/243881.pdf.

Gupta, SS (1962). Life test sampling plans for normal and lognormal distributions. *Technometrics* **4**(2), 151–175.

Gupta, SS and PA Groll (1961). Gamma distribution in acceptance sampling based on life tests. *Journal of the American Statistical Association* **56**(296), 942–970.

Johnson, NL, S Kotz, and N Balakrishman (1994). *Continuous univariate distributions*. 2nd. Wiley, New York.

Julian, H (1932). *Problems of relative growth*. 1st. Methuen, London.

Jun, CH, S Balamurali, and SH Lee (2006). Variables sampling plans for Weibull distributed lifetimes under sudden death testing. *IEEE Transactions on Reliability* **55**(1), 53–58.

Kantam, RRL and M Ravikumar (2016). Limited failure censored life test sampling plan in Burr type X distribution. *Journal of Modern Applied Statistical Methods* **15**(2), 428–454.

Kim, M and BJ Yum (2011). Life test sampling plans for Weibull distributed lifetimes under accelerated hybrid censoring. *Statistical Papers* **52**(2), 327–342.

Lawless, JF (2003). *Statistical models and methods for lifetime data*. 2nd. Wiley-Interscience, New Jersy.

Lina, CT, SS Wua, and N Balakrishnan (2009). Planning life tests with progressively type-I interval censored data from the lognormal distribution. *Journal of Statistical Planning and Inference* **139**(1), 54–61.

Loistl, O (2015). Does mean-variance portfoliomanagement deserve expected utility's approximative affirmation? *European Journal of Operational Research* **247**(2), 676–681.

Marshall, AW, JC Meza, and I Olkin (2001). Can data recognize its parent distribution. *Journal of Computational and Graphical Statistics* **10**(3), 555–580.

Al-Masri, AQS (2018). Acceptance sampling plans based on truncated life tests in the inverse-Gamma model. *Electronic Journal of Applied Statistical Analysis* **11**(2), 397–404.

Minka, TP (2002). *Estimating a gamma distribution*. Microsoft Research Technical Report.

Pascual, FG and G Montepiedra (2005). Lognormal and Weibull accelerated life test plans under distribution misspecification. *IEEE Transactions on Reliability* **54**(1), 43–52.

Rao, GS and RRL Kantam (2010). Acceptance sampling plans from truncated life tests based on the log-logistic distributions for percentiles. *Economic Quality Control* **25**(2), 153–167.

Schilling, EG and DV Neubauer (2017). *Acceptance sampling in quality control*. 3rd. Boca Raton : CRC, FL.

Schneider, H (1989). Failure-censored variables-sampling plans for lognormal and Weibull distributions. *Technometrics* **31**(2), 199–206.

Seo, JH, M Jung, and CM Kim (2009). Design of accelerated life test sampling plans with a nonconstant shape parameter. *European Journal of Operational Research* **197**(2), 659–666.

Sirvanci, M and G Yang (1984). Estimation of the Weibull parameters under type I censoring. *Journal of the American Statistical Association* **79**(385), 183–187.

Spurrier, JD and LJ Wei (1980). A test of the parameter of the exponential distribution in the type I censoring case. *Journal of the American Statistical Association* **75**(370), 405–409.

Teimouri, M, SM Hoseini, and S Nadarajah (2013). Comparison of estimation methods for the Weibull distribution. *A Journal of Theoretical and Applied Statistics* **47**(1), 93–109.

Wang, WB, ZN Wu, CF Wang, and RF Hu (2013). Modelling the spreading rate of controlled communicable epidemics through an entropy-based thermodynamic model. *Science China-Physics Mechanics and Astronomy* **56**(11), 2143–2150.

Wilrich, PT (2018). Sampling inspection by variables under Weibull distribution and type I censoring. *Frontiers in Statistical Quality Control 12*, 307–326.

Wu, JW and WL Tsai (2000). Failure-censored sampling plan for the Weibull distribution. *Information and Management Sciences* **11**(2), 13–25.

Zhang, Y and WQ Meeker (2005). Bayesian life test planning for the Weibull distribution with given shape parameter. *Metrika* **61**(3), 237–249.

Zhao, XJ, R Pan, E Castillo, and M Xie (2019). An adaptive two-stage Bayesian model averaging approach to planning and analyzing accelerated life tests under model uncertainty. *Journal of Quality Technology* **51**(2), 181–197.

Zhou, NX and FL He (2005). Moment-method estimation based on censored sample. *Journal of Systems Science and Complexity* **18**(2), 254–264.