Contributions to Reliability and Lifetime Data Analysis

by

Anupap Somboonsavatdee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2007

Doctoral Committee:

Professor Vijayan N. Nair, Co-Chair
Adjunct Associate Professor Ananda Sen, Co-Chair
Associate Professor Moulinath Banerjee
Assistant Professor Douglas E. Schaubel

To my dad and mom,
Chewphin and Chanthana Somboonsavatdee.

ACKNOWLEDGEMENTS

Dr. Vijayan N. Nair and Dr. Ananda Sen have been the ideal thesis advisors. Their invaluable advice, insightful criticisms, patient encouragement and endless kindness have made me who I am now. Thank you!

I also thank the rest of my dissertation committee, Dr. Moulinath Banerjee and Dr. Douglas Schaubel for their helpful comments and suggestions.

Many people in the Statistics Department have made me feel at home. I give my thanks to Aijun, Bibhas, Carles, Ho Keun, Joon Sang, Lili, Nami, Matt, Yan, Youjuan and everybody in the department, for memorable friendship.

More thanks to Aon, Ay, Nong, Nui, Pete and Rin for being there when I need someone to talk to. How long have I known you guys? I lost count. Thanks, Akarin (P Duang), Chetwana (X), Nat, Pantip (P Tim), Paweena (P Por) and Phongphaeth (Duel), for keeping me company during this long journey in Ann Arbor.

Finally but most importantly, I would like to thank my family in Thailand. Thank you, Dad Chewphin, Mom Chanthana, Brother Udomsak, Sister Pornphimon, Brother Phitthaya and Sister Arreerak, for unquestionable and limitless support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER 1

Introduction

This dissertation deals with problems in reliability and lifetime data analysis. The first part focuses on the study of graphical estimators from probability plots with right censored data. The second part deals with reliability inference for repairable systems.

Part I: Probability plots are popular graphical tools for assessing parametric distributional assumptions among reliability engineers and other practitioners. They are particularly well suited for location-scale families or those that can be transformed to such families. When the plot indicates a reasonable conformity to the assumed family, it is common to estimate the underlying location and scale parameters by fitting a line through the plot. This quick-and-easy method is especially useful with censored data. Indeed, the current version of a popular statistical software package uses this as the default estimation method. Part I of the dissertation investigates the properties of graphical estimators with multiply right-censored data and compares their performance to maximum likelihood estimators. Large-sample results on consistency, asymptotic normality, and asymptotic variance expressions are obtained. Small-sample properties are studied through simulation for selected distributions and censoring patterns. The results presented in this study extend the work of Nair (1984) to right-censored data.

Part II: Analysis of failure data arising from repairable systems has received con-

siderable attention in the statistical, engineering, computer software, and medical literature. Data pertaining to a repairable system is viewed as some type of 'recurrent event'. Part II of the dissertation investigates some models and methodologies for analyzing failures from repairable systems with multiple failure modes. We consider the case where the cause-specific failures (from each failure mode) follow some counting processes with an emphasis on nonhomogeneous Poisson processes (NHPPs). Some properties of the data are characterized and estimation methods are studied, both from a single system and multiple systems assuming independence of the failure modes. Some results are also developed when there is partial masking of the failure modes. The NHPP case with a power law intensity function is studied in detail.

CHAPTER 2

Graphical Estimators from Probability Plots with Right Censored Data

2.1 Introduction

The quantile-quantile (Q-Q) plot, often called probability plot, is a very useful graphical tool for assessing distributional assumptions, viz., whether a set of data can be modeled adequately by a hypothesized parametric family. We start with the simple case of uncensored data. Let $X_i,\ i = 1, \ldots, N$ be *iid* observations from a distribution $F(x)$. We will restrict attention in the paper to location-scale families or those which can be transformed to location-scale families such as the Weibull and lognormal. So, $F(x; \mu, \sigma) = F_0(\frac{x-\mu}{\sigma})$ for unknown location parameter μ and scale parameter $\sigma > 0$. For ease of exposition, we shall henceforth suppress the parameters μ, σ while describing either $F(\cdot)$ or $F_0(\cdot)$.

Define $Y_{1N} < \ldots < Y_{NN}$ to be the order statistics (sample quantiles) of the data. Let $p_{iN} = (i - .5)/N,\ i = 1, \ldots, N$, and $F_{iN} = F_0^{-1}(p_{iN})$. Then, $F_{1N} < F_{2N} < \ldots < F_{NN}$ are (approximately) the ordered theoretical quantiles of the baseline hypothesized distribution $F_0(\cdot)$. If the model is true, then $Y_{iN} \approx \mu + \sigma F_{iN}$. Hence, if we plot the sample quantiles against the theoretical quantiles, the data should fall roughly on a line with slope σ and intercept μ.

If the plot looks linear (indicating that hypothesized model is reasonable), a quick-and-easy method of graphical estimation is to fit a line to the data (typically a least-

squares (LS) regression line) and use the slope and intercept to estimate σ and μ. The estimators of scale and location from the ordinary LS (or OLS) line can be written as

$$\hat{\sigma} = \frac{\frac{1}{N}\sum_{i=1}^{N}(F_{iN} - \bar{F})(Y_{iN} - \bar{Y})}{S_F^2} \tag{2.1}$$

and

$$\hat{\mu} = \bar{Y} - \hat{\sigma}\bar{F}, \tag{2.2}$$

where $\bar{F} = \sum_{i=1}^{N} F_{iN}/N$, $\bar{Y} = \sum_{i=1}^{N} Y_{iN}/N$ and $S_F^2 = \sum_{i=1}^{N}(F_{iN} - \bar{F})^2/N$.

Although these estimators are generally viewed as inefficient compared to maximum likelihood estimators (MLEs), they are still popular among reliability engineers primarily due to their ease of computation. [In fact, version 15.0 of Minitab uses the graphical estimators from Q-Q plots as the default option for parametric estimation with Weibull and lognormal distributions.] Practitioners sometimes also use the standard errors from a regression package to obtain the standard errors of these graphical estimators. This is of course inappropriate as the data (the order statistics) are neither independent nor identically distributed

The use of probability plotting extends in a natural way to situations with censored data. With multiple right censoring, the Kaplan-Meier (K-M) or product-limit (PL) estimator provides a basis for the plotting positions. Situations with more complicated censoring can also be handled using the nonparametric MLE of the underlying distribution function (DF). However, we will restrict attention to multiple right censoring.

As we shall see, there are at least two different ways of generalizing the OLS estimator in the uncensored case to censored situations. We consider general weighted least-squares (WLS) estimators that include both of these as special cases. We use the asymptotic properties of the weighted quantile processes of the K-M estimator to obtain the limiting distributions of these estimators. We show that the behavior

of the estimators from the Q-Q plots is equivalent, up to $o_p(N^{-1/2})$, to the estimators obtained as weighted sample mean and standard deviations of the subset of uncensored observations. The finite sample properties of the estimators are studied through an extensive simulation study for the Weibull, lognormal, and log-logistic distributions under various censoring schemes and intensities.

Expressions for standard errors of the graphical estimators in the uncensored case are given in Nair (1984). Those for censored data are quite complicated (See Section 2.3). Version 15.0 of Minitab uses variances of the MLEs to *incorrectly* estimate the variance and covariances of these graphical estimators.

This chapter is organized as follows. Section 2.2 describes the graphical estimators from the Q-Q plot with censored data. Section 2.3 summarizes previous results in the literature for the uncensored cases, discusses the large-sample results, and compares the asymptotic relative efficiencies of OLS estimators compared to MLEs. Section 2.4 summarizes findings on finite-sample behavior from an extensive simulation study.

2.2 Location-Scale Estimation with Multiple Right Censoring

Let (Y_i^0, C_i) be the underlying failure and censoring times for the $i-$th unit, for $i = 1, \ldots, N$. These are assumed to be *iid* from the cumulative distribution functions (CDFs) $F(\cdot)$ and $G(\cdot)$ respectively. We will assume throughout that the failure and censoring times are independent. Further, both $F(\cdot)$ and $G(\cdot)$ are assumed to be continuous and differentiable with densities that are strictly positive.

We observe $Y_i^\diamond = \min\{Y_i^0, C_i\}$ and the indicator variable δ_i which equals 1 or 0 as $Y_i^0 \leq C_i$ or $Y_i^0 > C_i$. The censoring rate (θ) is defined as $Pr(Y^0 > C)$. By independence, the $Y_i^\diamond$'s are *iid* with survival function (SF): $1 - H(\cdot) = (1 - F(\cdot))(1 - G(\cdot))$. Throughout, let $Y_{1N}^\diamond < \ldots < Y_{NN}^\diamond$ be the ordered data, and define δ_{iN} as the δ_i's corresponding to these ordered $Y_{iN}^\diamond$'s. Further, we use the notation $Y_{1M}, \ldots, Y_{MM}$

to define the *ordered uncensored* observations, i.e., ordered subset of the $Y_i^{\diamond}$'s with $\delta_i = 1$, $i = 1, \ldots, N$.

Define

$$k_j = \sum_{\{i\ :Y_{iN}^{\diamond} < Y_{jM}\}} (1 - \delta_{iN}) \tag{2.3}$$

to be the number of censored observations that are less than Y_{jM}. Then, the K-M or PL estimator $\hat{F}_N$ of F can be written as

$$\hat{F}_N(y) = 1 - \prod_{j=1}^{i} \Big(1 - \frac{1}{N - k_j - j + 1}\Big),\ Y_{i-1\ M} < y \leq Y_{iM}. \tag{2.4}$$

We can use the K-M estimator to get the plotting positions for the Q-Q plot as

$$p_{iM} = \Big(\hat{F}_N(Y_{iM}) + \hat{F}_N(Y_{i-1\ M})\Big)/2$$

(see Meeker and Escobar 1998). One can use other plotting positions, but this choice reduces to the commonly used choice of $\frac{i-.5}{N}$ in the uncensored case. Throughout this chapter, we let

$$X_{iM} = F_0^{-1}(p_{iM}).$$

The Q-Q plot with right-censored data is then a plot of $\{X_{iM}, Y_{iM}\}, 1 \leq i \leq M$.

Figure 2.1 is a Weibull Q-Q plot of the shock absorber data in Meeker and Escobar (1998, page 630). There were two competing failure modes in this example, and Figure 2.1 is based on failure data for failure mode M1. The (random) failure times of mode M2 induces right censoring for M1. The plot suggests that the Weibull model is a reasonable fit, so the next step is estimation of the Weibull (or the corresponding smallest extreme-value) parameters. The default method in Minitab Version 15.0 is graphical estimation obtained by fitting an ordinary least-squares line through the plot (see Minitab output in Figure 1). We will revisit this example later in this chapter.

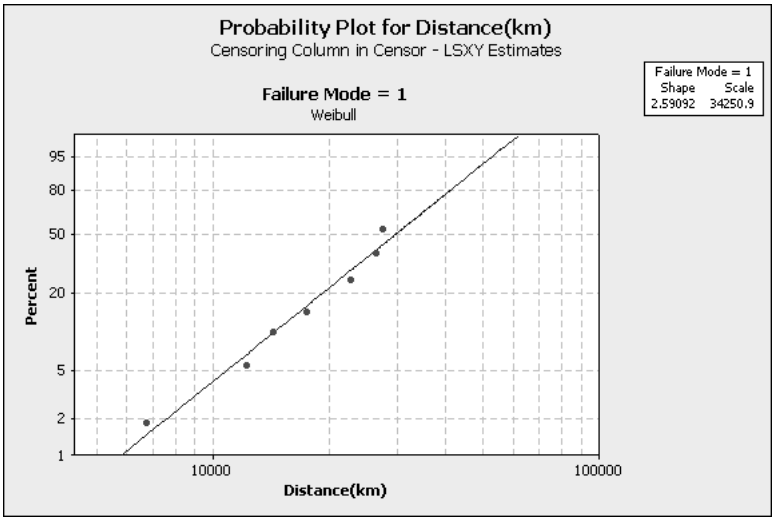


Figure 2.1: Weibull probability plot and fitted line for shock absorber data with failure mode M1 (see Meeker and Escobar 1998, page 630). The line corresponds to OLS estimators.

We will consider a general class of graphical estimators obtained from the slope and intercept of lines that minimize the weighted sum of squared deviations

$$\sum_{i=1}^{M} w_{iM}(Y_{iM} - \mu - \sigma X_{iM})^2 \tag{2.5}$$

with nonnegative weights $w_{iM},\ 1 \le i \le M$, where $\sum_{i=1}^{N} w_{iM} = 1$. We will allow for random weights as well.

We need some additional notations. Define

$$\begin{aligned} \bar{Y}_w &= \textstyle\sum_{i=1}^{M} w_{iM} Y_{iM},\ S^2_{Y,w} = \sum_{i=1}^{M} w_{iM}(Y_{iM} - \bar{Y}_w)^2, \\ \bar{X}_w &= \textstyle\sum_{i=1}^{M} w_{iM} X_{iM},\ S^2_{X,w} = \sum_{i=1}^{M} w_{iM}(X_{iM} - \bar{X}_w)^2 \end{aligned} \tag{2.6}$$

to be the weighted means and variances. Then, the WLS estimators (graphical estimators) of location and scale that minimize (2.5) are given by

$$\hat{\sigma}_w = \frac{\sum_{i=1}^{M} w_{iM}(X_{iM} - \bar{X}_w)(Y_{iM} - \bar{Y}_w)}{S^2_{X,w}} \tag{2.7}$$

and

$$\hat{\mu}_w = \bar{Y}_w - \hat{\sigma}_w \bar{X}_w. \tag{2.8}$$

The most common choice is $w_{iM} = 1/M$. This corresponds to the line shown on

Figure 2.1 for the shock absorber data set. The corresponding Weibull parameters (for the log-location-scale Weibull distribution) are indicated on the plot.

However, there is another estimator that can also be considered a generalization to the censored case. To see this, denote the theoretical quantile function of the baseline CDF by $F_0^{-1}(t) = \inf\{y : F_0(y) \geq t\}$, and the K-M quantile function by $\hat{F}_N^{-1}(t) = \inf\{y : \hat{F}_N(y) \geq t\}$. Then, in situations with no censoring, minimizing (2.5) with constant weights is equivalent (asymptotically) to minimizing the "integral" version

$$\int \left(\hat{F}_N^{-1}(t) - \mu - \sigma F_0^{-1}(t)\right)^2 d\hat{F}_N(t). \tag{2.9}$$

This is so because the empirical CDF has constant jumps of $1/N$ in the uncensored case. But under right censoring, the K-M estimator has jumps $\Delta_{iM} = \hat{F}_N(Y_{iM}) - \hat{F}_N(Y_{i-1\ M})$. So the use of constant weights in (2.9) leads to random weights $w_{iM} = \Delta_{iM}/\hat{F}_N(Y_{MM})$ in the discrete version in (2.5). Thus, there are two ways of generalizing the OLS estimator in the uncensored case, depending on whether we use constant weights in (2.5) or in (2.9). We will consider both of these. The former will be referred to as OLS ($w_{iM} = 1/M$) and the latter ($w_{iM} \propto \Delta_{iM}$) as modified LS (MLS) in this chapter. However, we will focus for the most part on the OLS estimators. The large-sample results in the paper will be developed for general WLS estimators and will include both of these as special cases.

2.3 Theoretical Results

2.3.1 Review of the Uncensored Case

The properties of the estimators from probability plots have been studied in the uncensored case by several authors (see Lloyd 1952, Downton 1954, Blom 1958, Ali and Chan 1964, and Nair 1984). Relevant results from Nair (1984) and references therein are summarized below.

1. If the underlying location-scale family is symmetric, we get $\bar{X}_w = 0$ for any symmetric set of weights w_{iN}'s. In this case, $\hat{\mu}_w = \bar{Y}_w$; in particular, for the OLS case with equal weights, $\hat{\mu}_w$ is just the sample mean.

2. The situation with the scale estimator is more interesting. For the symmetric case, $\hat{\sigma}_w$ is proportional to $\sum_{i=1}^{N} w_{iN} F_{iN} Y_{iN}$. This is an L-estimator (linear-combination of order statistics) of scale.

3. If the underlying distribution is normal, the OLS estimator of location is just the sample mean. The graphical estimator of scale is equal (essentially) to $\sum_{i=1}^{N} \Phi^{-1}(\frac{i-.5}{N}) Y_{iN}$ where $\Phi^{-1}(\cdot)$ is standard normal quantile function (the denominator is close to 1). This is the optimal L-estimator of scale (Chernoff, Gastwirth, and Johns 1967). The results in Nair (1984) show that this L-estimator is equivalent to the sample standard deviation up to order $o_p(N^{-1/2})$.

4. The following more general result was established in Nair (1984). Recall the estimators $S^2_{Y,w}$ and $S^2_{X,w}$ defined in (2.6). The difference $|\hat{\sigma}_w - S_{Y,w}/S_{X,w}|$ goes to zero at the rate $N^{-1/2}$. So the difference between the graphical estimator of scale and the weighted, normalized sample standard deviation is asymptotically very small. In particular, the large-sample distributions of the two estimators of scale are the same for all WLS estimators with their weight functions satisfying Condition 5 and 8 in Nair (1984).

5. Hence, the graphical estimators of scale and location are as good (or as bad) as the weighted, normalized standard deviation $S_{Y,w}$ and weighted and centered mean $\bar{Y}_w$, respectively. In particular, the graphical estimators are efficient in the normal case because they are equivalent to the sample mean and standard deviation which are the MLEs. But in other cases such as the logistic or smallest-extreme value, they are as (in)efficient as the sample mean and SD.

6. The focus in this chapter is on multiple right censoring. Two-sided trimmed and

Winsorized estimators were studied in Nair (1984). One-sided right trimming is equivalent to Type I right censoring.

2.3.2 Large-Sample Results for the Multiply Censored Case

We will summarize the results on asymptotic distributions in this section. The technical results and proofs are deferred to Section 2.7 Before turning to formal statements of results, we provide an intuitive discussion.

Recall that if the hypothesized location-scale model holds, $Y_{iM} = \mu + \sigma X_{iM}+$ error, where $X_{iM} = F_0^{-1}(p_{iM})$ for $i = 1, \ldots, M$. Substituting $Y_{iM} = \mu + \sigma X_{iM}$ into the expressions for the estimators in equations (2.7) and (2.8) and simplifying, we can see that $E(\hat{\sigma}_w) \approx \sigma$ and $E(\hat{\mu}_w) \approx \mu$, suggesting that the WLS graphical estimators will be consistent. This is formalized in Theorem 2.2 under suitable conditions.

To state the results formally, we will need to specify the conditions on the distributions and weight functions. These are stronger than we need but they are easy to state and verify in this form. Besides, they are satisfied for the two classes of estimators considered in the paper. First, define $W_N(t)$ as follows.

(2.10)

$$W_N(t) = \begin{cases} w_{iM}/\Delta_{iM}, & \text{if } \hat{F}_N(Y_{i-1\ M}) < t \leq \hat{F}_N(Y_{iM}) \wedge \hat{F}_N(T)\ (i = 1, \ldots, M) \\ \\ 0, & \text{if } \hat{F}_N(Y_{MM}) \wedge \hat{F}_N(T) < t \leq 1 \end{cases}$$

Conditions:

1. The location-scale family $F_0(y)$ has a density that is continuous and strictly positive on $(-\infty, \infty)$.

2. The CDF $F_0(y)$ has a finite absolute moment of order ν for some $\nu \geq 4$.

3. $W_N(t) \rightarrow^{a.s.} W(t)$ where $W(t)$ is continuous a.e. for $t \in (0, 1)$ with $|W(t)| \leq L$ for some constant L.

Theorem 2.1 below studies the relationship between the graphical scale estimator and the corresponding estimator $S_{Y,w}/S_{X,w}$ based on the normalized weighted standard deviation of the Kaplan-Meier quantiles. It turns out that the differences between these two estimators are very small and hence they have very similar behavior. The latter estimator can also be computed easily and is thus an alternative method of doing quick-and-easy estimation. However, the graphical estimators are more robust as they involve only linear functions of the order statistics and not their squared terms.

Theorem 2.1. *Under Conditions 1-3,* $N^{1/2}[\hat{\sigma}_w - S_{Y,w}/S_{X,w}]$ *converges in probability to zero as* $N \to \infty$.

The formal proof is deferred to Section 2.7, but we provide here a remark about the consequences of this finding. Consider the weighted correlation coefficient between $\{X_{iM}\}$ and $\{Y_{iM}\}$ in the Q-Q plot

$$R_w = \sum_{i=1}^{M} w_{iM}(Y_{iM} - \bar{Y}_w)(X_{iM} - \bar{X}_w)/S_{Y,w}S_{X,w}. \tag{2.11}$$

From (2.7) and (2.11), we can rewrite the graphical WLS estimator of scale as

$$\hat{\sigma}_w = R_w S_{Y,w}/S_{X,w}. \tag{2.12}$$

It turns out that $N^{1/2}(R_w - 1)$ converges to zero in probability , and so the difference $|S_{Y,w}/S_{X,w} - \hat{\sigma}_w| = (R_w - 1)S_{Y,w}/S_{X,w}$ converges in probability to zero at rate $o_p(N^{-1/2})$. This result implies, in particular, that the scale estimator from the Q-Q plot $\hat{\sigma}_w = R_w S_{Y,w}/S_{X,w}$ and the normalized, ratio of weighted sample standard deviations $S_{Y,w}/S_{X,w}$ have the same limiting distribution. But Theorem 2.1 gives us more, the differences in the two classes of estimators are small, so they should be close to each other even in moderate samples.

Theorem 2.2 below establishes that the WLS graphical estimators are consistent and asymptotically normal and obtains explicit expressions for the asymptotic variance of the estimators. This is true only if the hypothesized parametric model holds.

Theorem 2.2. *Under Conditions 1-3,* $N^{1/2}[(\hat{\mu}_w - \mu), (\hat{\sigma}_w - \sigma)]^T$ *has a limiting bivariate normal distribution with mean* $\boldsymbol{0}$ *and covariance matrix* V *with elements*

$$
\begin{aligned}
V_{11} &= \sigma^2(\lambda_{11} + m_1^2\lambda_{22}/4 - m_1\lambda_{12}) \\
V_{12} &= \sigma^2(\lambda_{12} - m_1\lambda_{22}/2)/2 \\
V_{22} &= \sigma^2\lambda_{22}/4
\end{aligned} \tag{2.13}
$$

where the λ_{ij}*'s are given by equation (2.18) and the* m_1 *is given in Lemma 2.2 (see Section 2.7).*

The OLS estimators with censored data correspond to $w_{iM} = 1/M$ and the MLS estimators to $w_{iM} = \Delta_{iM}/\hat{F}_N(Y_{MM})$. Both of these satisfy Conditions 1-3.

In the uncensored case, the graphical estimators satisfy the usual location-scale invariance/equivariance properties. With censoring, however, these properties no longer hold. This is also true for MLE and other methods of estimation with complex censoring. It can, however, be shown that the invariance/equivariance properties continue to hold if the censoring distribution also belongs to the same location-scale family and the censoring proportion is held constant.

Theorem 2.2 provides expressions for the asymptotic variance-covariance matrix of the estimators later in this section. However, these are rather involved and must be evaluated numerically. In practice, it would be simpler to use the bootstrap resampling technique to estimate the variances and also obtain confidence intervals. The use of bootstrap with multiple censored data has been discussed extensively in the literature (Efron and Tibshirani 1993, Efron 1981, and Burr 1994), so we consider only some of the relevant details here and the full details are discussed in Section 2.5. It is reasonable to use the parametric bootstrap for the distribution of interest since the parameterization and the estimation problem make sense only under the model. The censoring distribution G, however, is a nuisance parameter and can be estimated parametrically or nonparametrically. If there is a *low* degree of censoring in

the original problem, the censoring distribution suffers from high *censoring* and may not be estimable in the right tails. In such cases, one has to resort to a parametric model. Of course, this requires knowledge of the model, either from prior information or through graphical methods such as the Q-Q plot of the K-M estimator of G. Of course, with high censoring, the Q-Q plot will have limited information on the right tail and the model selection will be driven by the lower-end of the distributions. This is a practical difficulty in any situation.

2.3.3 Shock Absorber Data Revisited

Consider again the shock absorber data from Meeker and Escobar (1998) discussed earlier. Figure 1 also gives the OLS estimators of the Weibull parameters from Minitab: scale $= \hat{\eta} = 34,250.9$ and shape $= \hat{\beta} = 2.591$. The corresponding estimates of location and scale for the smallest-extreme value distribution are: $\hat{\mu}_{OLS} = \log(\hat{\eta}) = 10.44$ and $\hat{\sigma}_{OLS} = 1/\hat{\beta} = 0.386$ (see Table 2.1). For comparison, the MLEs are also given in the table.

The more relevant issue for our purpose is computation of standard errors (SEs) and using it for inference. Table 2.1 gives SEs of the MLE using the information matrix and of the OLS estimators using bootstrap. The OLS estimators have slightly larger variability, as to be expected. The relative efficiencies are studied in the next section. The table also shows two other incorrect estimates of the SEs of OLS estimators that are sometimes used. *Regression* refers to the naive use of SEs from a regression package as done by some reliability practitioners. This ignores the correlation and unequal variances of the order statistics. In this data set, it grossly underestimates the variability. *Minitab* refers to the output from Minitab Version 15.0. Our understanding is that this is obtained using the information matrix and plugging in the OLS estimators for the unknown parameters. This measures the SE of the MLE rather than OLS and hence is incorrect. These values would normally

Table 2.1: Comparison of estimates and standard errors for the shock absorber data

Parameter	Method	Estimate	$\hat{\text{SE}}$
σ	MLE	0.296	0.084
σ	OLS	0.363	Bootstrap: 0.096 Regression: 0.025 Minitab: 0.161
μ	MLE	10.35	0.148
μ	OLS	10.46	Bootstrap: 0.171 Regression: 0.059 Minitab: 0.243

be closer to the SEs for the MLE. To check this, we used 500 bootstrap samples and computed the 500 SEs. The median of these values were very close to the SEs for the MLEs in Table 2.1.

2.3.4 Asymptotic Relative Efficiency

Table 2.2 gives the asymptotic relative efficiencies (AREs) of the OLS estimators of location and scale (compared to MLEs) for various (log)location-scale distributions and censoring schemes. Table 2.3 gives the corresponding AREs for various quantiles (design life in reliability terminology). These were based on the asymptotic variances of the OLS and ML estimators for various failure-time and censoring DF's and censoring rates. The asymptotic variances were computed by numerical integration using the expression from Theorem 2.2, and were evaluated by using function `dblquad()` in Matlab 7.0.

There are several relevant conclusions from Table 2.2. The results for $\theta = 0$ (the uncensored case) are consistent with those in Nair (1984). In particular, the OLS estimators are asymptotically fully efficient for the lognormal (or equivalently normal location-scale) case. Even with censoring, the OLS estimators remain quite efficient for estimating both location and scale estimators. The AREs are very close to 1 when the censoring distribution is also lognormal. They are lower with uniform censoring (which has a finite end-point) and decrease as the censoring proportion increases (the end-point of the uniform distribution gets smaller) but are still close to 90%. For

Table 2.2: Asymptotic relative efficiencies (AREs) of the OLS estimators. (ARE is defined as the ratio of asymptotic variance of MLE to that of the OLS estimator.)

Underlying Distribution	Censoring Rate (θ)	ARE			
		Location		Scale	
		Censoring Distribution		Censoring Distribution	
		Same Family	Uniform	Same Family	Uniform
Weibull	0.00	0.949	0.949	0.551	0.551
	0.25	0.840	0.802	0.552	0.555
	0.50	0.667	0.560	0.554	0.547
	0.75	0.520	0.434	0.555	0.543
Lognormal	0.00	1.000	1.000	1.000	1.000
	0.25	0.997	0.994	0.993	0.970
	0.50	0.998	0.955	0.997	0.916
	0.75	1.000	0.862	0.999	0.855
Log-logistic	0.00	0.912	0.913	0.911	0.909
	0.25	0.967	0.991	0.854	0.765
	0.50	0.961	0.847	0.819	0.646
	0.75	0.939	0.580	0.814	0.587

the Weibull (which is equivalent to the smallest extreme value location-scale family), in the case with no censoring, the ARE of the scale estimator remains around 55%. This is the same as the ARE of the normalized sample variance, and it is known in the literature that the normalized sample variance is quite inefficient in the smallest extreme-value distribution. The comparative performance of the location estimator is more interesting. In the uncensored case, the graphical estimator of location is a centered version of the usual sample mean and its efficiency is quite good (95%). But as the censoring proportion increases, its ARE efficiency drops considerably, to a low of 45-50% with very high censoring ($\theta = 0.75$). As before, uniform censoring leads to lower AREs than censoring from the same family.

The Weibull and lognormal are the most common parametric distributions among reliability practitioners; the log-logistic is less common. Nevertheless, it is interesting to examine another log-location scale family where the underlying location-scale distribution is symmetric but has heavier tails than the normal. The performance of the graphical estimators are in between those for the Weibull and lognormal cases. The AREs with log-logistic censoring are quite high for both location and scale estimators and closer to the AREs for the lognormal case. The effect of the censoring

Table 2.3: Asymptotic relative efficiencies (AREs) of the OLS estimators of quantiles X_p (or log of design life) for selected values of p. The quantiles are in the scale of the location-scale family (which is the pivot used for log-location-scale distributions).

		ARE				
Censoring Settings	Censoring rate (θ)	$X_{0.10}$	$X_{0.25}$	$X_{0.50}$	$X_{0.75}$	$X_{0.90}$
Weibull/Weibull	0.00	0.685	0.792	0.951	0.869	0.709
	0.25	0.680	0.780	0.883	0.753	0.634
	0.50	0.668	0.732	0.726	0.612	0.556
	0.75	0.621	0.606	0.544	0.507	0.500
Weibull/uniform.	0.00	0.685	0.792	0.951	0.869	0.709
	0.25	0.681	0.774	0.855	0.716	0.609
	0.50	0.647	0.678	0.617	0.521	0.493
	0.75	0.570	0.514	0.446	0.433	0.441
Logormal/Lognormal	0.00	1.000	1.000	1.000	1.000	1.000
	0.25	0.995	0.997	0.996	0.994	0.993
	0.50	0.999	0.999	0.998	0.997	0.996
	0.75	0.999	1.000	1.000	1.000	1.000
Lognormal/uniform	0.00	1.000	1.000	1.000	1.000	1.000
	0.25	0.987	0.994	0.994	0.984	0.977
	0.50	0.955	0.965	0.955	0.931	0.920
	0.75	0.889	0.885	0.862	0.847	0.844
Log-logistic/log-logistic	0.00	0.912	0.912	0.913	0.912	0.911
	0.25	0.887	0.926	0.967	0.945	0.910
	0.50	0.873	0.927	0.961	0.911	0.870
	0.75	0.875	0.927	0.939	0.892	0.859
Log-logistic/uniform	0.00	0.912	0.912	0.913	0.912	0.911
	0.25	0.846	0.924	0.991	0.918	0.845
	0.50	0.767	0.854	0.847	0.724	0.670
	0.75	0.654	0.640	0.580	0.555	0.554

distribution seems to play a much bigger role here. With uniform censoring, the AREs decrease quite a bit more and are less than 60% for $\theta = 0.75$, closer to the Weibull case.

The location and scale parameters themselves are not of direct interest in reliability applications. The primary goal is to estimate quantities such as design life (quantiles), survival probabilities, hazard rates, and so on. Table 2.3 shows the AREs for quantiles in the log-scale, i.e., quantiles of the underlying location-scale family. These are the pivots used for computing confidence intervals, so it is appropriate to compute AREs in this scale. The overall conclusions from Table 2.3 are similar to those from Table 2.2. The graphical estimators do quite well in the lognormal case, quite poorly in the Weibull case, and are in between in the log-logistic case. Again, the AREs for uniform censoring are lower (in some cases much lower) than for censoring from the same family. Note also that the results are not symmetric in the quantiles (X_p vs X_{1-p}) for the symmetric location-scale families (normal and logistic). This is due to the different effects of censoring in the left and right tails. Further, the impact of censoring is greater on larger quantiles; the AREs for smaller quantiles are generally better.

2.4 Finite-Sample Results

This section examines the finite-sample behavior of both the OLS and MLS estimators and compares them with those of the MLE's through simulation. We also studied the RSD (ratio of standard deviations, S_Y/S_X, with constant weights $1/M$). But the performance of this estimator was very similar to OLS, so we omit this in the discussion below.

The design of the simulation study was as follows. We considered sample sizes $N = 25, 50, 75, 100$ and 500 for the following failure time and censoring distribution

combinations: a) Weibull/Weibull, b) Weibull/uniform, c) lognormal/lognormal, d) lognormal/uniform, e) log-logistic/log-logistic, and f) log-logistic/uniform. The censoring proportions θ were chosen as 0, .25, .50, and .75. For censoring proportion .75, we did not consider sample size of 25 due to the high probability that all observations can be censored. The simulation was done in Matlab and the results are based on simulation samples of size $5,000$.

Throughout, μ_F and σ_F of F were set to be 0 and 1. The desired θ, censoring proportion was determined as follows: (i) if F and G are from the same family, set $\sigma_F = \sigma_G$ and choose μ_G to get the desired θ; (ii) if G is from uniform distribution $G \sim$ uniform$[0, a]$, choose a to get the desired θ. For example, for lognormal/lognormal with θ=25%, we let $F \sim$ lognormal$(0, 1)$ and chose $G \sim$ lognormal$(0.954, 1)$. For the lognormal/uniform with θ=25%, we set $F \sim$ lognormal$(0, 1)$ and chose $G \sim$ uniform$[0, 6.066]$. In most scenarios, the values of μ_G are computed numerically to satisfy desire θ. However, when F and G are from lognormal distributions with the same scale parameter $\sigma_F = \sigma_G = \sigma$. Let suppose $Y \sim F$ and $C \sim G$. At fixed θ, by setting $Pr(Y > C) = \theta$, it can be shown that

$$\mu_G = \mu_F - \sqrt{2}\sigma\Phi^{-1}(\theta) \tag{2.14}$$

where $\Phi^{-1}(\cdot)$ is a quantile function for standard normal distribution. The result in (2.14) follows by the fact that $\log C - \log Y \sim N(\mu_2 - \mu_1, 2\sigma^2)$, hence, $Pr(Y > C) = Pr(\log C - \log Y < 0) = \Phi\left(\frac{\mu_F - \mu_G}{\sqrt{2}\sigma}\right) \equiv \theta$ where $\Phi(\cdot)$ is CDF for standard normal. Similarly, when F and G are from the Weibull distribution with the same shape parameter $\beta_F = \beta_G = \beta = 1/\sigma$, then the scale parameter of G, $\eta_G = \log(\mu_G)$, can be written as below,

$$\begin{aligned} \eta_G &= \eta_F \left(\frac{1-\theta}{\theta}\right)^{1/\beta} \text{ or} \\ \mu_G &= \mu_F + \sigma \log\left(\frac{1-\theta}{\theta}\right). \end{aligned} \tag{2.15}$$

By considering, $Pr(Y > C) = \mathbb{E}[Pr(Y > C|C)] = \mathbb{E}[e^{-(C/\eta_F)^\beta}] = \frac{1}{\left(\frac{\eta_G}{\eta_F}\right)^\beta + 1} \equiv \theta$, the result in (2.15) follows. Below is Table 2.4 listing the values of μ_G or a_G used in the simulation under different scenarios.

Table 2.4 shows the values of μ_G or a_G used in the simulations under different scenarios.

Table 2.4: The values of μ_G or a_G for the simulation under different scenarios.

F		G from Same Family	G from uniform$[0, a_G]$
Lognormal	$\theta = .25$	μ_G=0.954	a_G=6.066
	.50	0.000	2.403
	.75	-0.954	1.045
Weibull	θ=.25	μ_G=1.099	a_G=3.921
	.50	0.000	1.594
	.75	-1.099	0.606
Log-logistic	$\theta = .25$	μ_G=1.633	a_G=9.347
	.50	0.000	2.513
	.75	-1.633	0.734

2.4.1 Relative Efficiency

We now turn to the relative efficiency (RE) of the graphical estimators. RE is defined as the ratio of the mean squared error (MSE) of MLE to that of the estimator being compared. The results for the Weibull distributions are given in Figure 2.2 for location and Figure 2.3 for scale. Figures 2.4 and 2.5 summarize the corresponding results for lognormal. And Figures 2.6 and 2.7 are for log-logistic.

There are several conclusions to be made overall:

1. The overall conclusion about the OLS estimators is the same as that from the AREs in the last section: they are quite efficient in the lognormal case even with censored data, inefficient in the Weibull case, and have reasonable performance in the log-logistic case. Furthermore, the performance is poorer with uniform censoring.

2. The MLS estimators are very interesting. They are less efficient than the OLS for lognormal but much more efficient in the Weibull case! The performance in

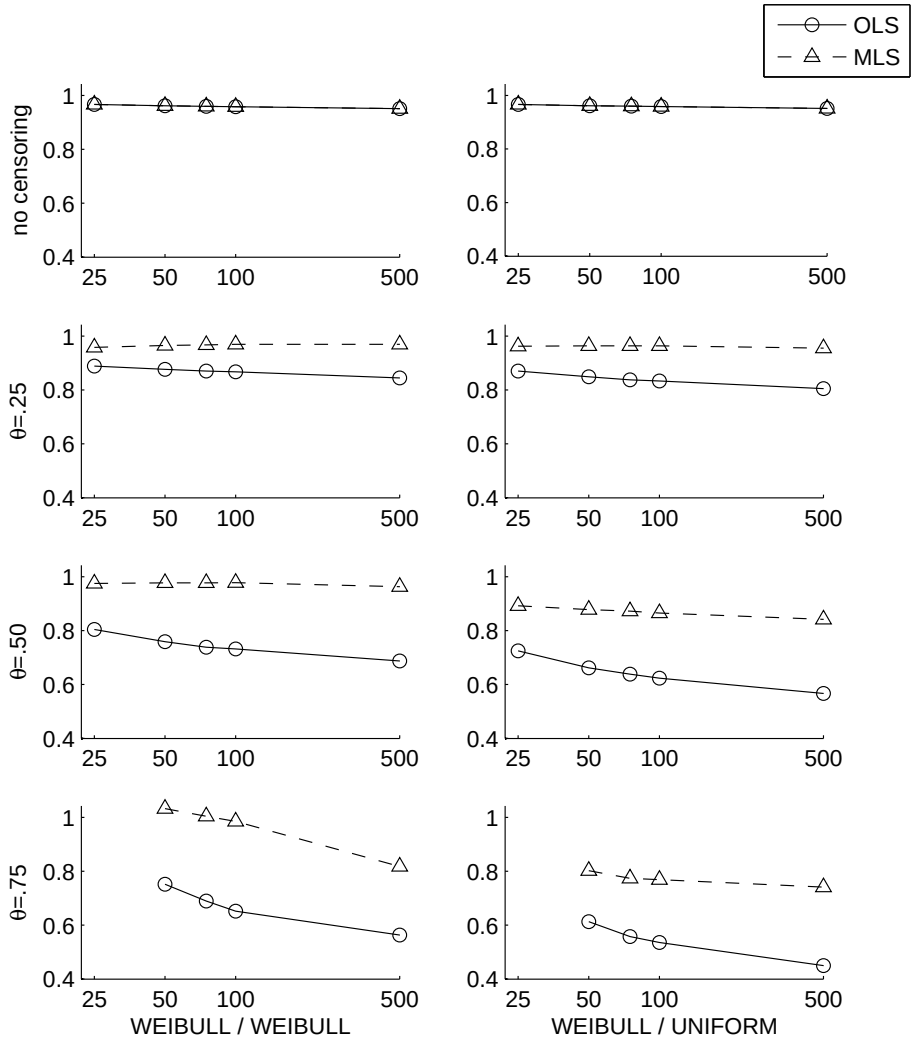


Figure 2.2: Relative efficiencies of OLS and MLS location estimators compared to ML location estimator for Weibull/Weibull (left) and Weibull/uniform (right) censoring. X-axis is the sample size in log-scale.

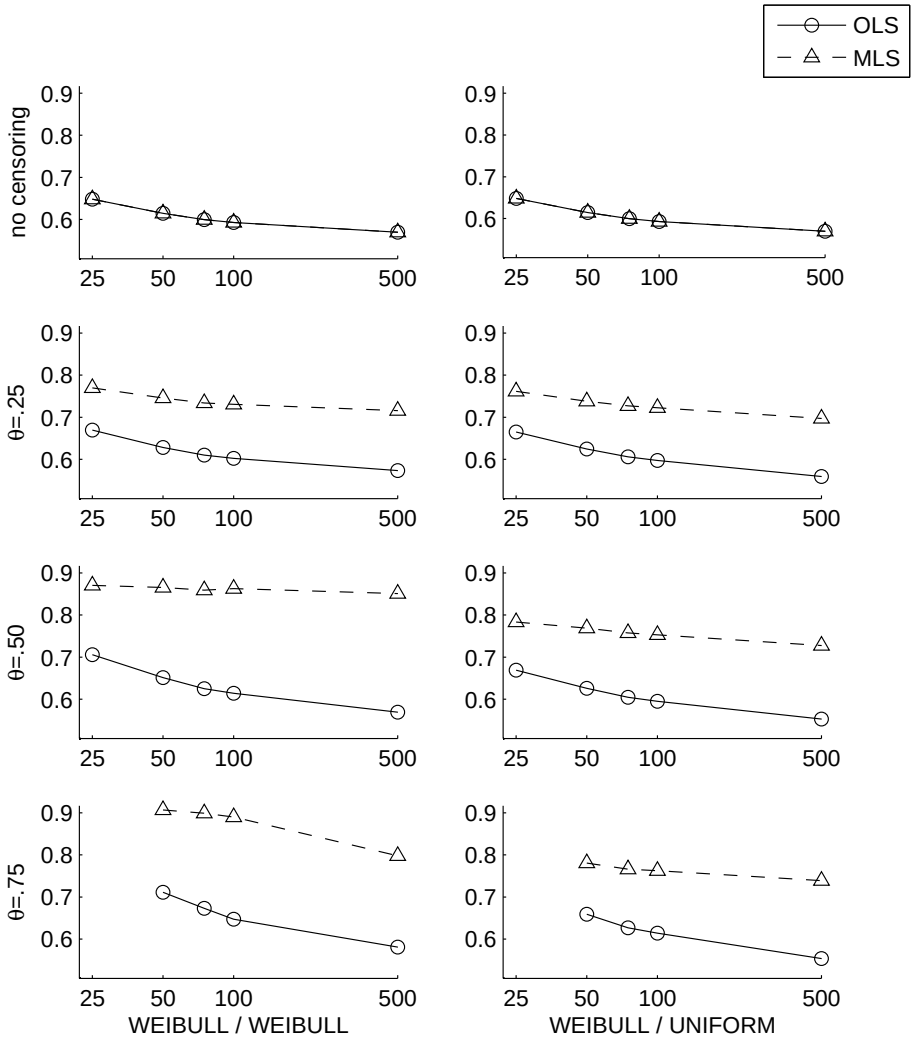


Figure 2.3: Relative efficiencies of OLS and MLS scale estimators compared to ML scale estimator for Weibull/Weibull (left) and Weibull/uniform (right) censoring. X-axis is the sample size in log-scale.

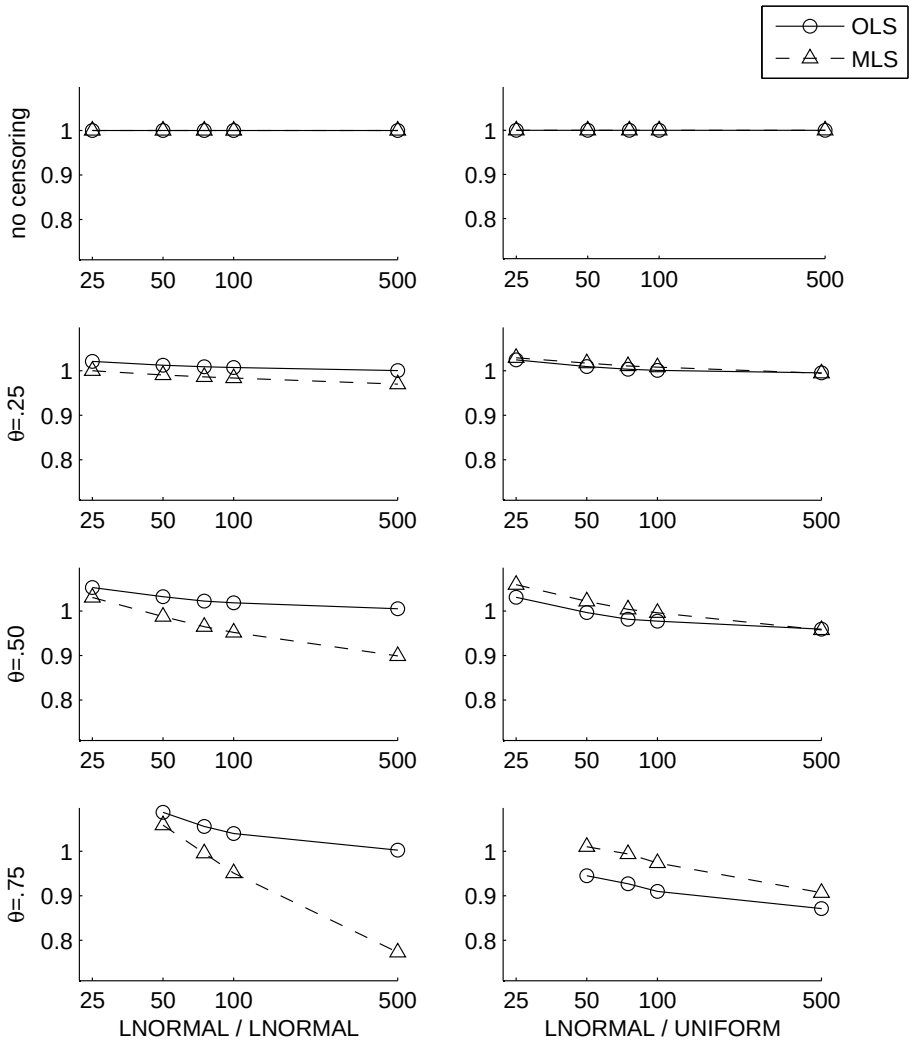


Figure 2.4: Relative efficiencies of OLS and MLS location estimators compared to ML location estimator for lognormal/lognormal (left) and lognormal/uniform (right) censoring settings. X-axis is the sample size in log-scale.

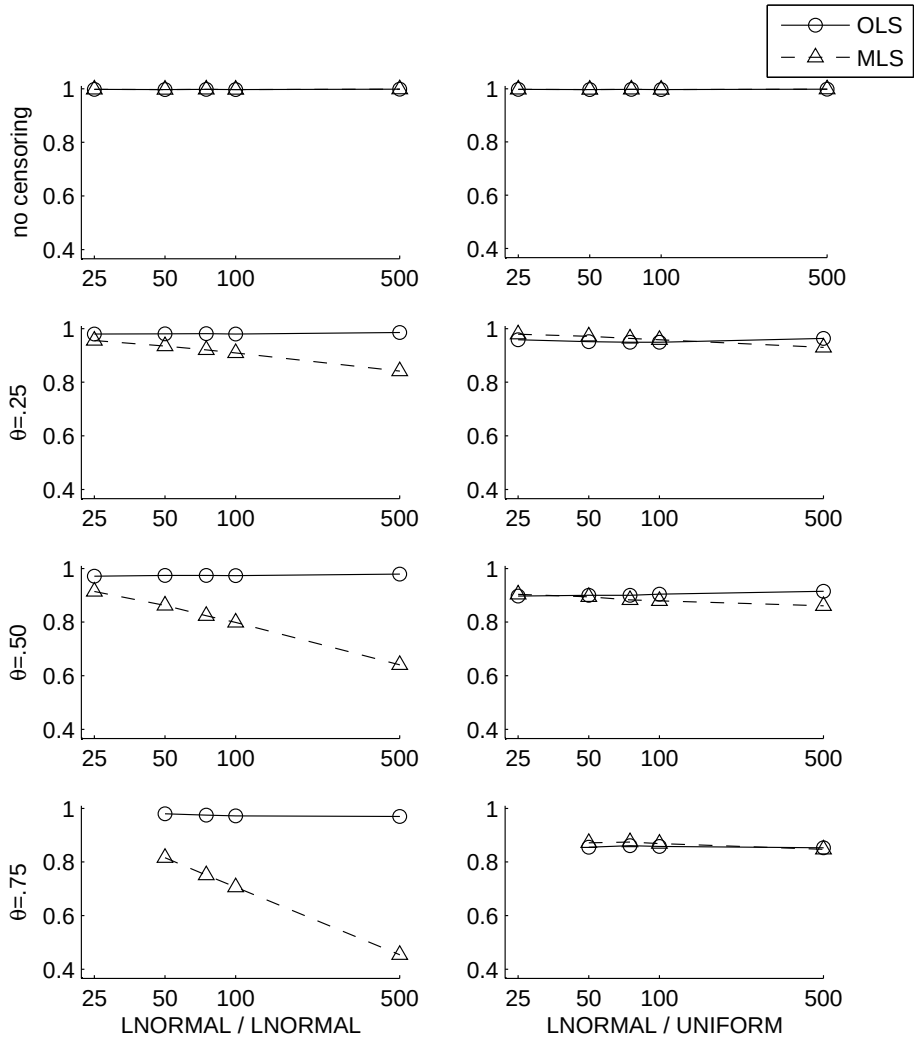


Figure 2.5: Relative efficiencies of OLS and MLS scale estimators compared to ML scale estimator for lognormal/lognormal (left) and lognormal/uniform (right) censoring settings. X-axis is the sample size in log-scale.

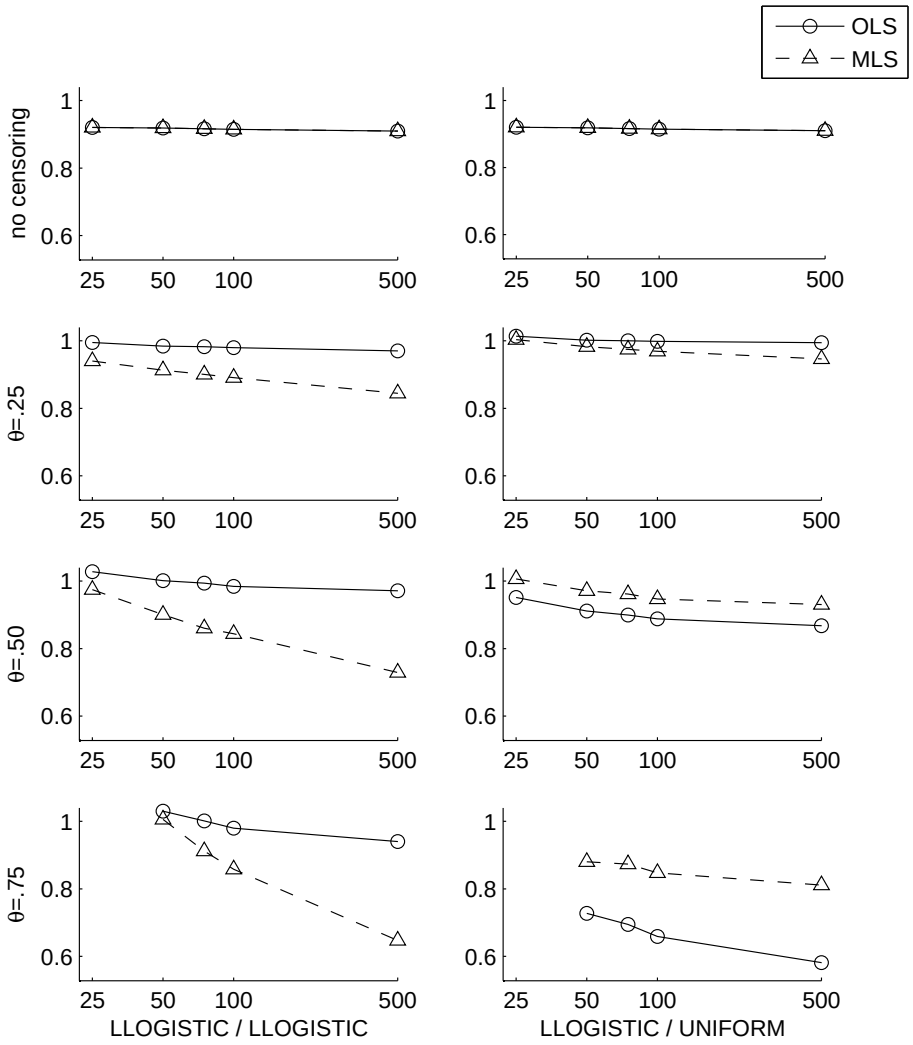


Figure 2.6: Relative efficiencies of OLS and MLS location estimators compared to ML location estimator for log-logistic/log-logistic (left) and log-logistic/uniform (right) censoring settings. X-axis is the sample size in log-scale.

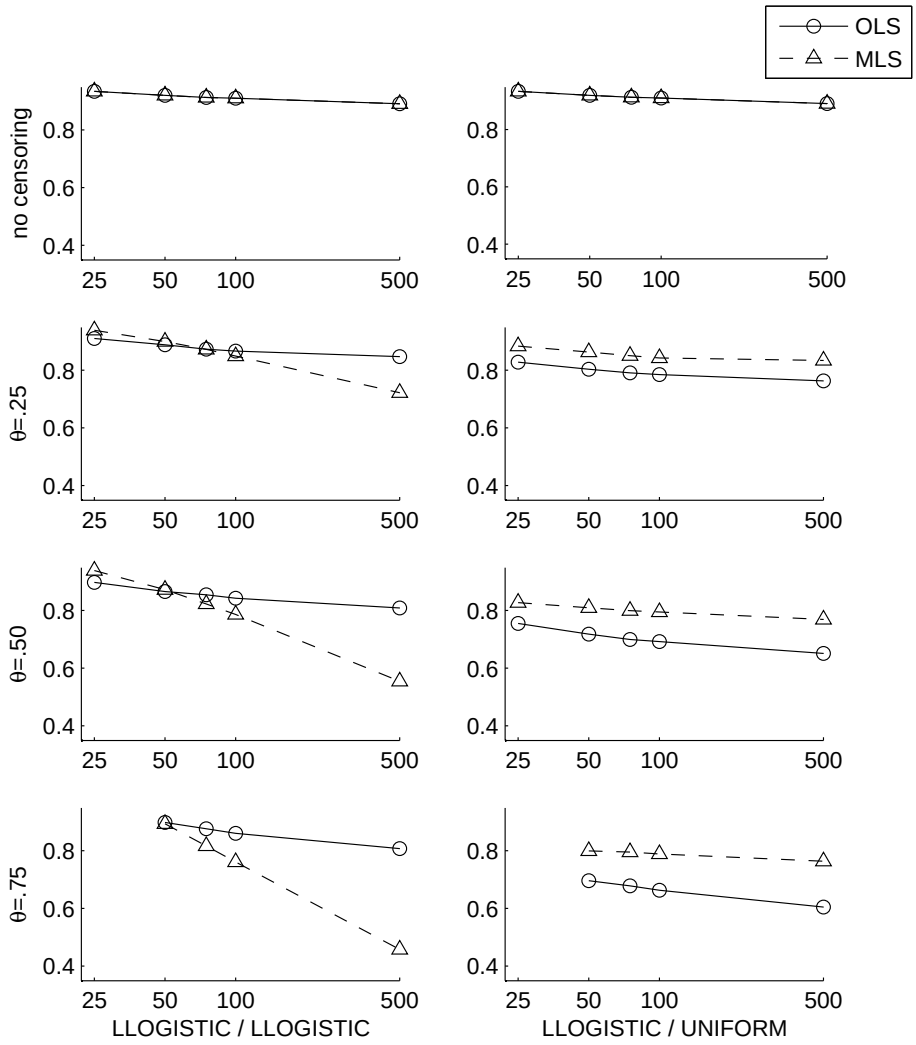


Figure 2.7: Relative efficiencies of OLS and MLS scale estimators compared to ML scale estimator for log-logistic/log-logistic (left) and log-logistic/uniform (right) censoring settings. X-axis is the sample size in log-scale.

the log-logistic case is mixed. Ordinarily, this would merit further investigation to understand and possibly exploit this feature. However, our goal here is not to recommend new estimators over the MLEs but just documenting the behavior of these graphical estimators.

3. All of the curves are decreasing as sample size increases, suggesting that the small sample relative efficiencies are better than the AREs in the last section. This is most striking for the MLS estimators (see Figure 2.4 with lognormal censoring and $\theta = 0.75$). For the OLS estimators, however, most of these differences are relatively small. For example, for the Weibull OLS scale estimators, the REs drop from about 0.7 to 0.6.

4. In some cases, the RE is higher than 1, indicating that the graphical estimators can be more efficient than the MLEs in small samples.

2.5 Use of Bootstrap for Inference

One can do inference (confidence regions or hypothesis tests) using the expressions for the limiting variance-covariance matrix in equation (2.13). However, they are complicated and must be computed numerically. In addition, the large-sample approximations may not work well in finite samples. We briefly review the use of the bootstrap in this section.

The bootstrap is a popular resampling procedure (see Efron and Tibshirani 1993, Efron 1981, and Burr 1994) that has been used quite effectively in a variety of situations to provide standard error estimates. There are several ways of bootstrapping in the presence of censoring. We adopt a version of the method described in Burr (1994). Recall that that we are interested in (log) location-scale distribution F of failure times, and the data are censored by variables from unknown censoring distribution G. We observe only $(Y_i^\diamond, \delta_i)$'s for $i = 1, \ldots, N$ where $Y_i^\diamond = Y_i^0 \wedge C_i$ and $\delta_i = 1$

if $Y_i^{\diamond} = Y_i^0$ or $= 0$ otherwise.

The following steps can be used to obtain the bootstrap estimates of variance and confidence intervals:

1. Assume the TTF distribution follows a (log)location-scale family given by F_0. Obtain $\hat{\sigma}_w$ and $\hat{\mu}_w$ from the Q-Q plot observed as described in (2.7) and (2.8). This yields a parametric estimate $\hat{F} = F_0(\frac{y-\hat{\mu}_w}{\hat{\sigma}_w})$.

2. Estimate G by $\hat{G}$. This can be done nonparametrically using a K-M estimator with $(Y_i^{\diamond}, 1-\delta_i)$'s. If there is a *low* degree of censoring, however, nonparametric estimation can be problematic. In this case, we suggest using a parametric model (based either on prior knowledge or through graphical methods such as the Q-Q plot, but now using the K-M estimator of G).

3. Draw **B** bootstrap samples. Specifically for each bootstrap sample, draw Y_i^{0*}'s from $\hat{F}$ and C_i^*'s from $\hat{G}$ for $i = 1, \ldots, N$. This leads to "bootstrapped observed data" $(Y_i^{\diamond *}, \delta_i^*)$'s where $Y_i^{\diamond *} = Y_i^{0*} \wedge C_i^*$ and $\delta_i^* = 1$ if $Y_i^{\diamond *} = Y_i^{0*}$, $= 0$ otherwise.

4. Calculate the WLS location and scale estimators of F from each bootstrap sample. This leads to $(\hat{\mu}_w^{*(j)}, \hat{\sigma}_w^{*(j)})$'s for $j = 1, \ldots, \mathbf{B}$. Now,compute the standard error from the bootstrap sample. One can also compute the quantile of $(\hat{\mu}_w^{*(j)}, \hat{\sigma}_w^{*(j)})$'s to obtain the confidence interval from WLS estimators.

It is reasonable to re-sample variables parametric from F since the entire estimation process is predicated on the model being correct. However, G is a nuisance parameter, so it is more desirable to estimate G nonparametrically. However, as already noted, if there is a low degree of censoring, one may not be able to estimate the right-tails of the censoring distribution well. We have done extensive comparisons of the use of parametric and nonparametric methods for G. Figures 2.8-2.13 shows the histograms from the parametric and nonparametric methods have comparable in shape including variability and center values. This indicates that the results seems

to be comparable in the most part for parametric and nonparametric bootstrap.

2.6 Summary

We have studied the properties of graphical estimators of location and scale and corresponding estimators of design life (quantiles) from probability plots with censored data. The results include large-sample properties and asymptotic variances as well as finite-sample performance. The relative efficiencies of these estimators, compared to the MLEs, suggest that they do well for lognormal failure-time distributions, reasonably well for log-logistic distributions, and poorly for Weibull distributions.

The discussion of these graphical estimators suggest a related class of scale estimators based on the ratio of weighted sample standard deviations. Their performances are essentially the same as the corresponding scale estimators from a weighted least-squares line fitted to the Q-Q plot. While they are even easier to compute, they are not as robust.

It is not the goal of this chapter to recommend the use of these quick-and-easy estimators over the MLEs. Rather, our intention is to shed light on the behavior of these graphical estimators that appear to be popular among some practitioners. Some have been using incorrect estimates of standard-errors in conjunction with these estimators. The results and discussion in this chapter point out these problems and suggest alternative methods.

2.7 Technical Results and Proofs

The asymptotic distributions of the K-M estimator and the corresponding quantile process have been studied in the literature. We provide a summary before proceeding to our results.

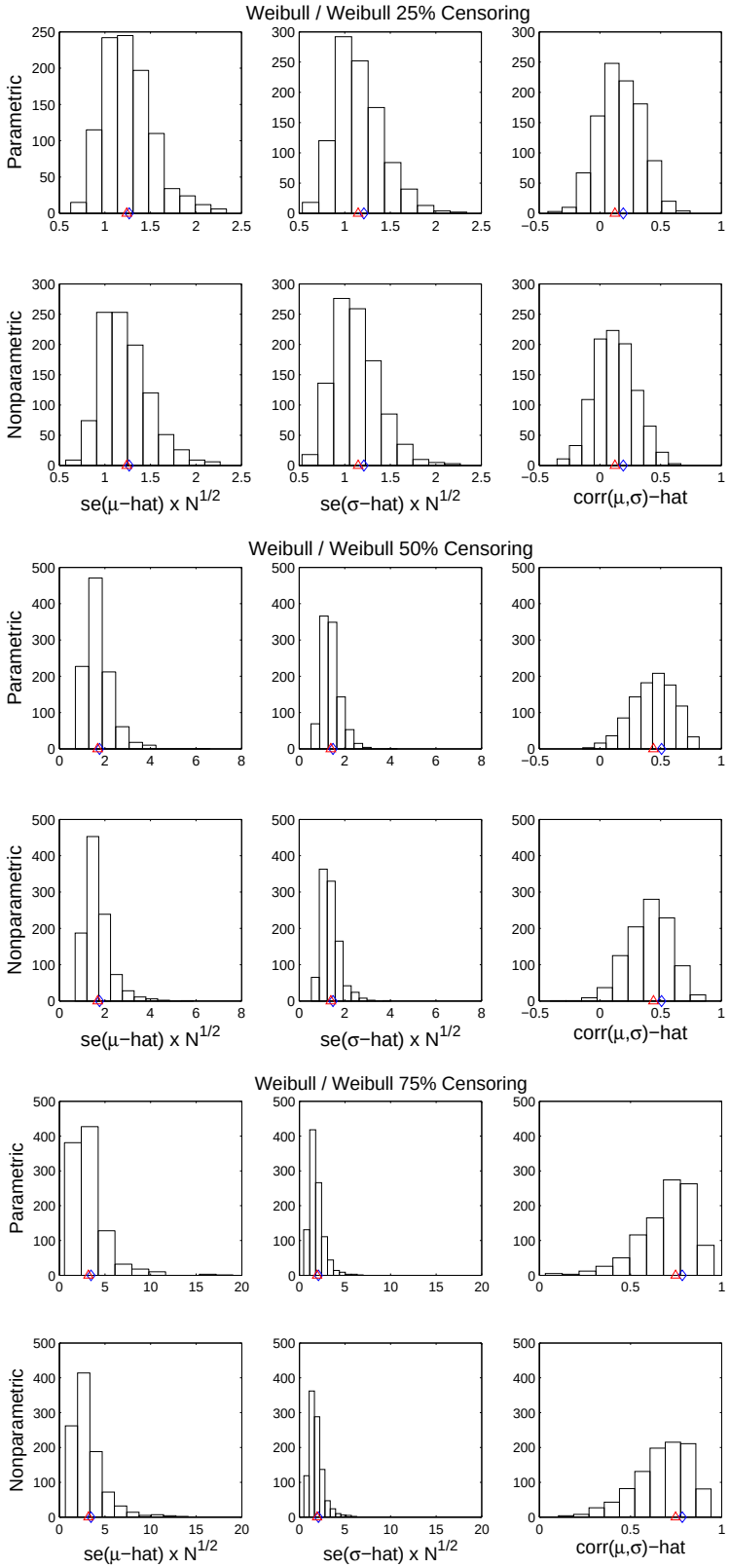


Figure 2.8: Histogram of the $\sqrt{N}\times$ standard errors and correlation for OLS estimators from the bootstrap simulation compare to asymptotic value ($\diamondsuit$) and its true value ($\triangle$) for Weibull/Weibull censoring setting. 'Parametric' and 'Nonparametric' refer to the assumption made on censoring distribution G.

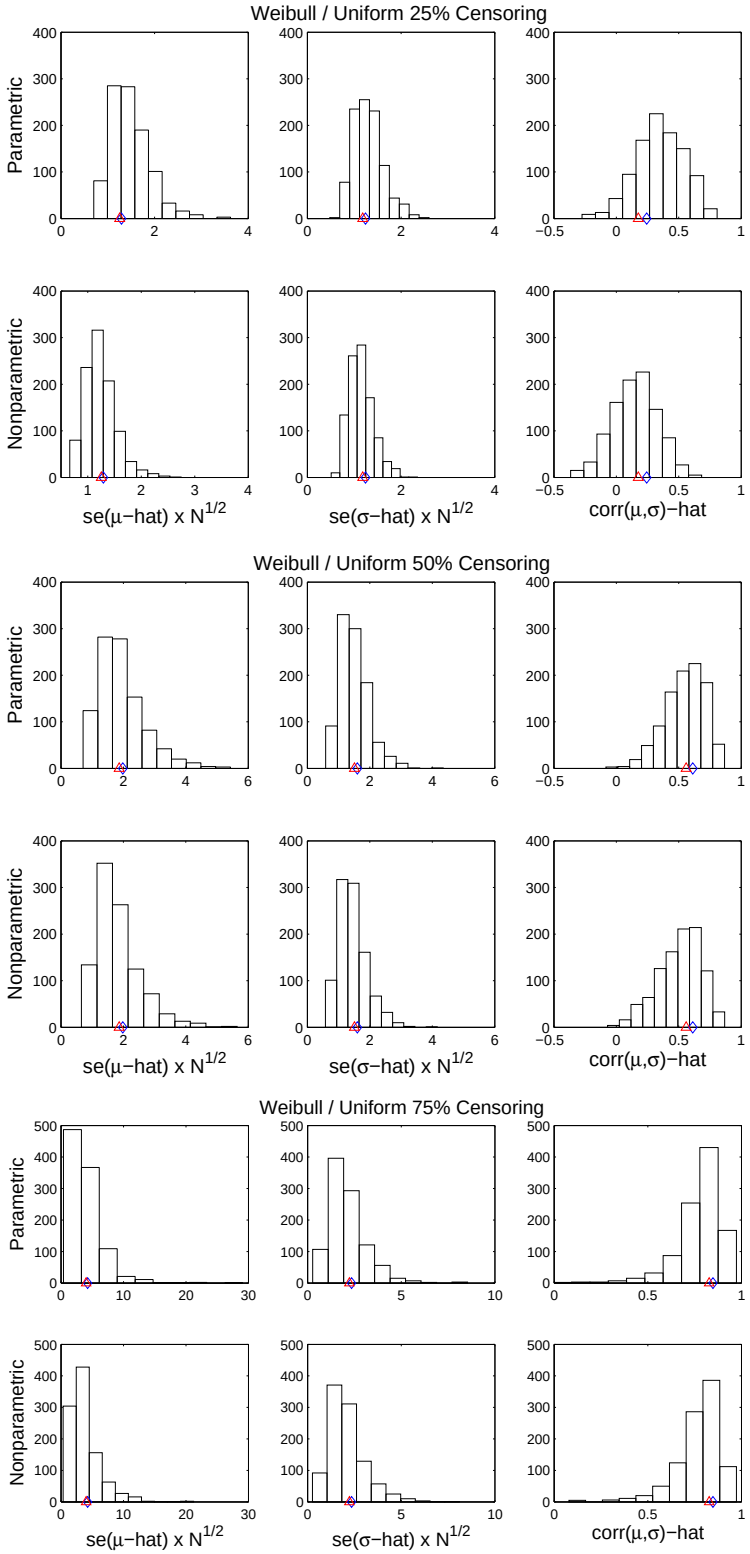


Figure 2.9: Histogram of the $\sqrt{N}\times$ standard errors and correlation for OLS estimators from the bootstrap simulation compare to asymptotic value ($\diamondsuit$) and its true value ($\triangle$) for Weibull/uniform censoring setting. 'Parametric' and 'Nonparametric' refer to the assumption made on censoring distribution G.

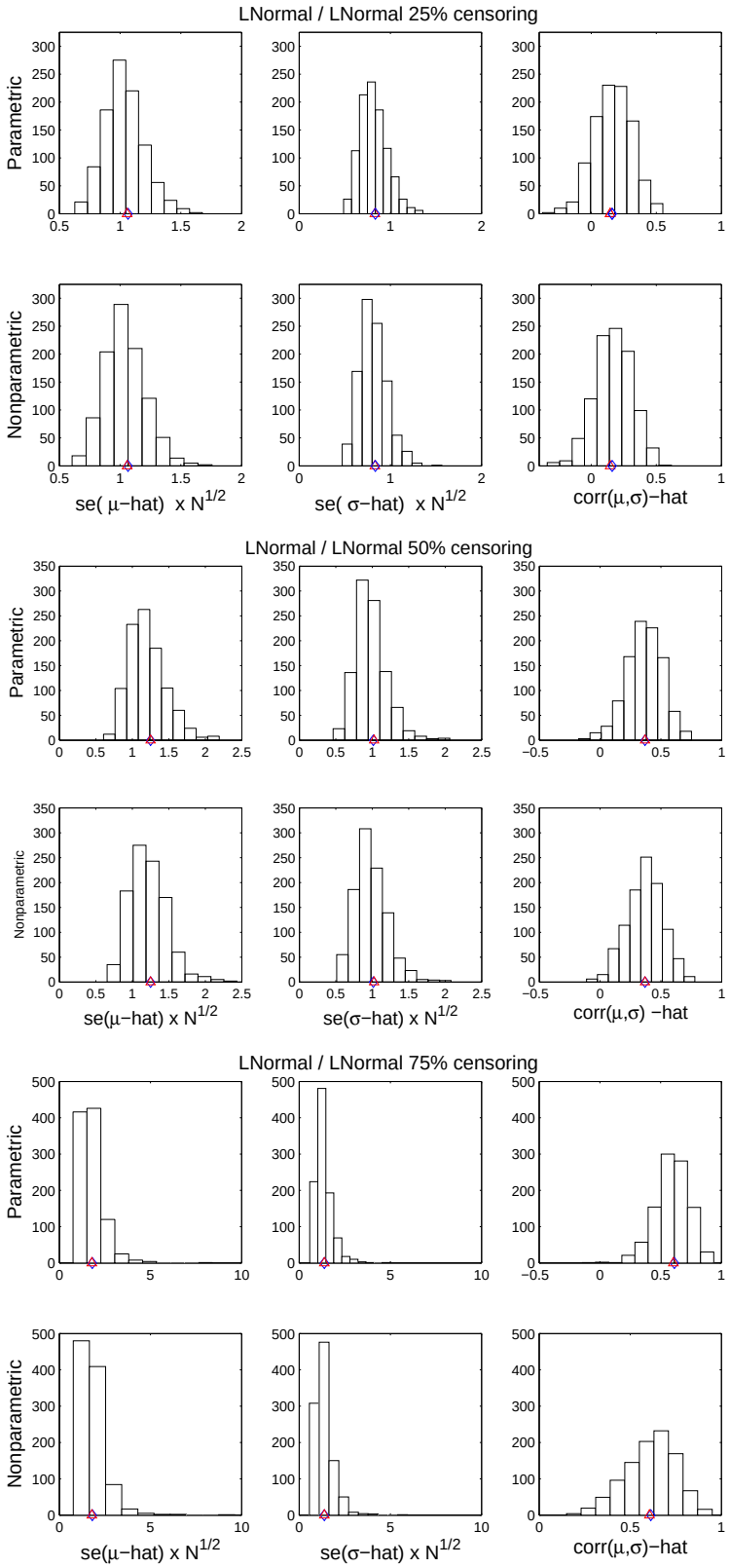


Figure 2.10: Histogram of the $\sqrt{N}\times$ standard errors and correlation for OLS estimators from the bootstrap simulation compare to asymptotic value ($\diamondsuit$) and its true value ($\triangle$) for lognormal/lognormal censoring setting. 'Parametric' and 'Nonparametric' refer to the assumption made on censoring distribution G.

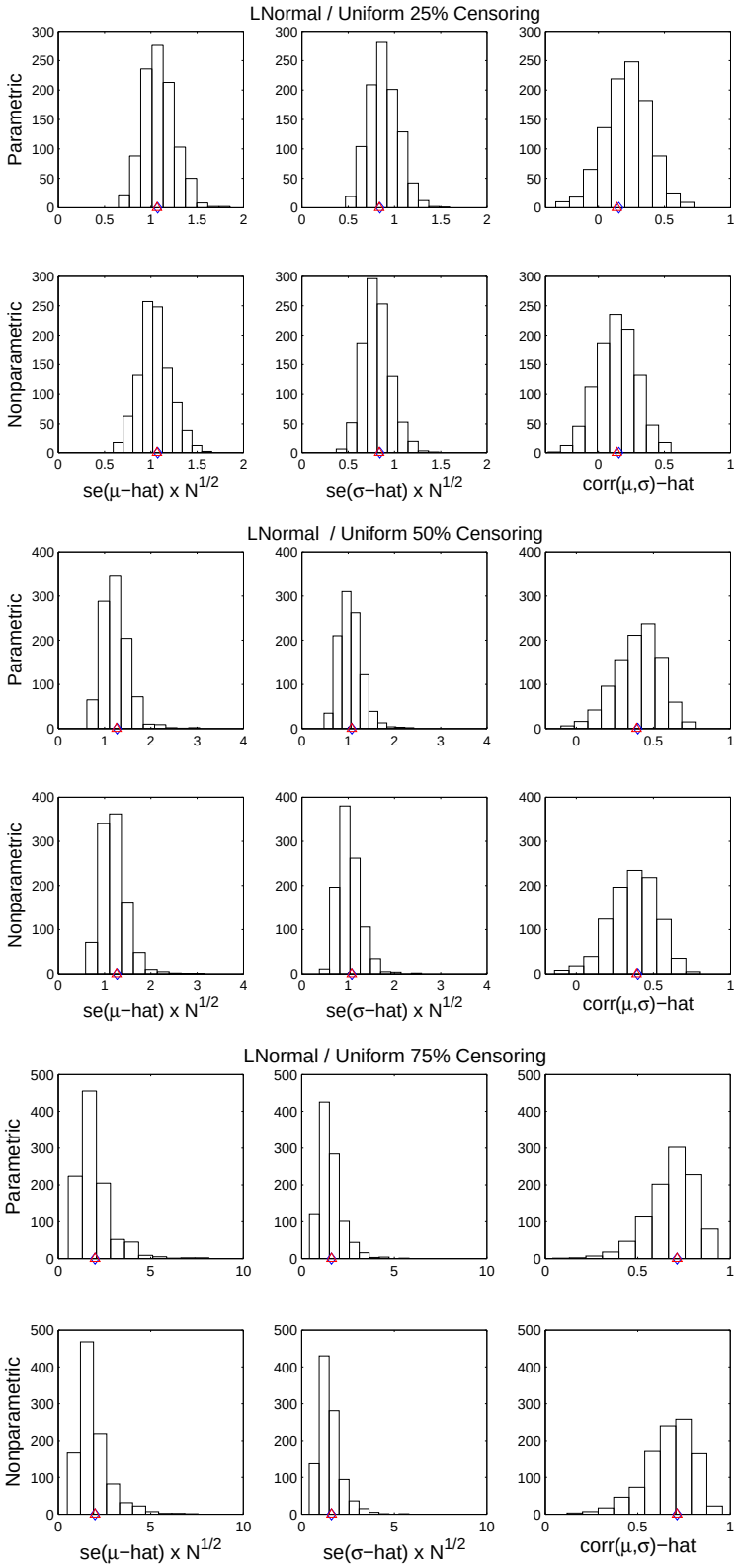


Figure 2.11: Histogram of the $\sqrt{N}\times$ standard errors and correlation for OLS estimators from the bootstrap simulation compare to asymptotic value ($\diamondsuit$) and its true value ($\triangle$) for lognormal/uniform censoring setting. 'Parametric' and 'Nonparametric' refer to the assumption made on censoring distribution G.

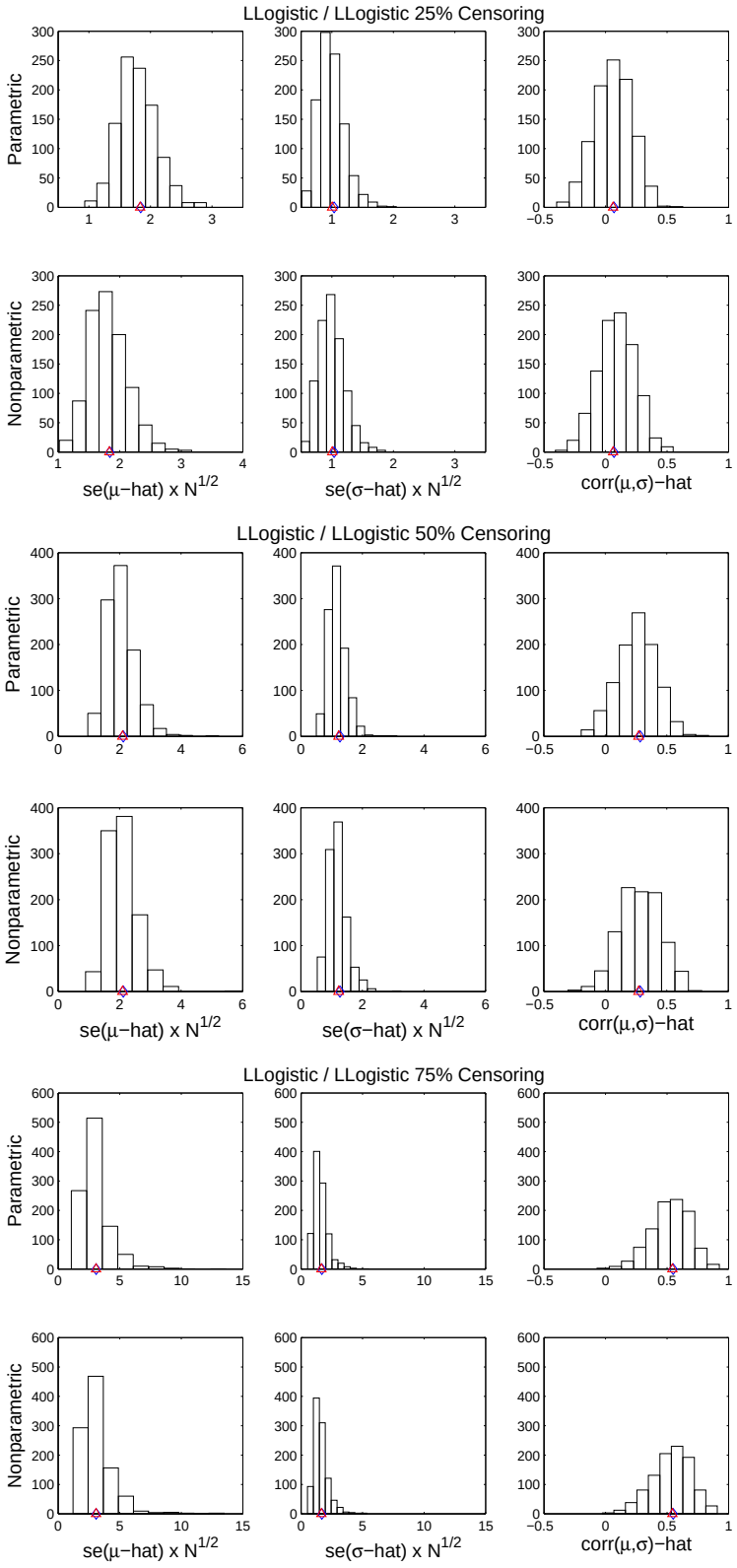


Figure 2.12: Histogram of the $\sqrt{N}\times$ standard errors and correlation for OLS estimators from the bootstrap simulation compare to asymptotic value ($\Diamond$) and its true value ($\triangle$) for log-logistic/log-logistic censoring setting. 'Parametric' and 'Nonparametric' refer to the assumption made on censoring distribution G.

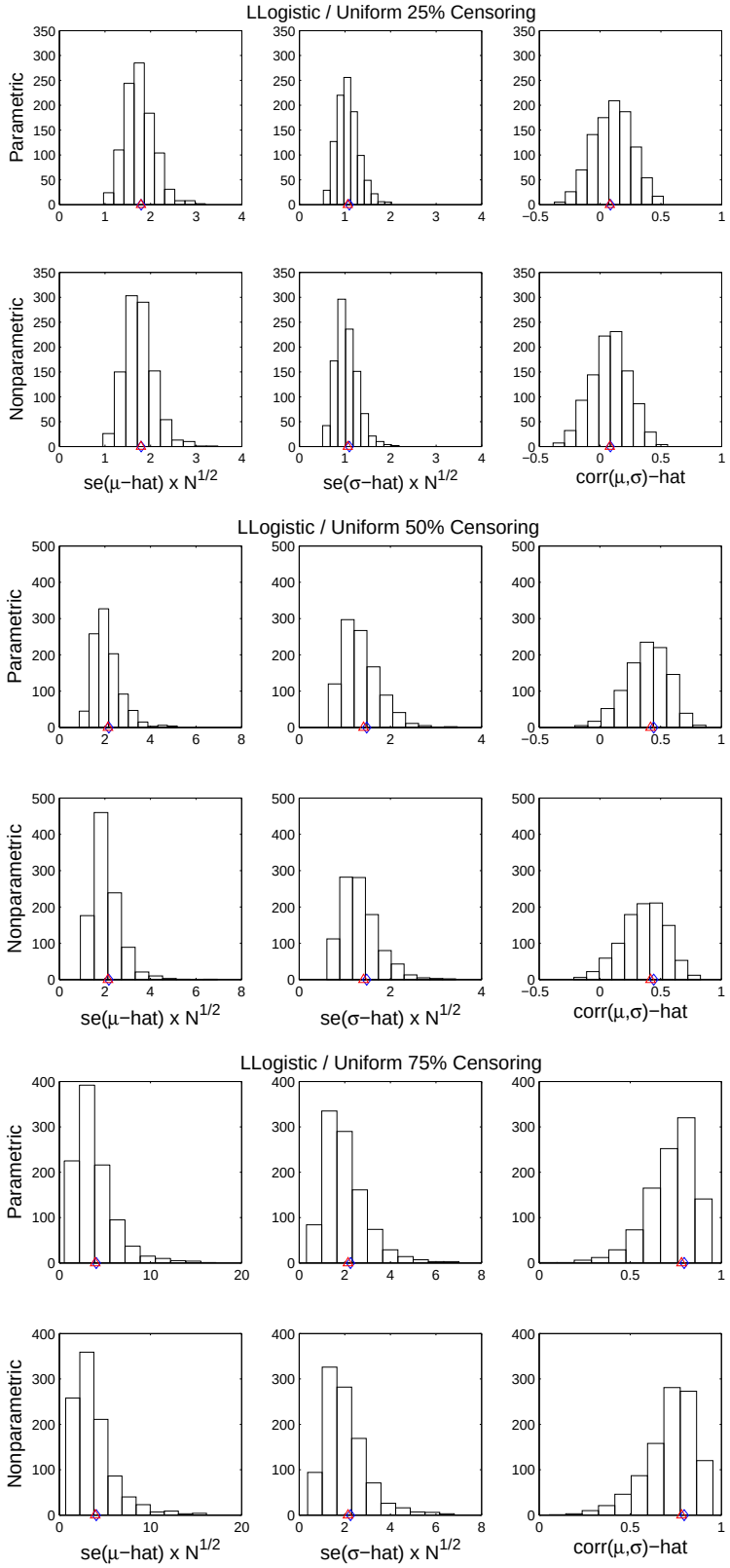


Figure 2.13: Histogram of the $\sqrt{N}\times$ standard errors and correlation for OLS estimators from the bootstrap simulation compare to asymptotic value ($\diamondsuit$) and its true value ($\triangle$) for log-logistic/uniform censoring setting. 'Parametric' and 'Nonparametric' refer to the assumption made on censoring distribution G.

We will assume throughout that the data have been transformed if necessary so that we have an underlying location-scale distribution that is supported on $(-\infty,\ \infty)$. For technical reasons, we have to restrict the support of the K-M estimator and the quantile process on the right tail to a finite value $T < \infty$ such that $y \leq T < H^{-1}(1)$. [Recall that $(1-H) = (1-F)(1-G)$.] The practical implication of this is that the largest order-statistics Y_{iM} used in the Q-Q plot will have to be bounded above by this value of T. It is possible to extend the results to the entire line in special cases, such as when the censoring distribution also has support on the entire line. But we will not pursue such technical issues here.

Let $0 < t < 1$ and define $\sqrt{N}(\hat{F}_N(F^{-1}(t)) - t) = Z_N(t)$. Then, Breslow and Crowley (1974) showed that

$$Z_N(t) \xrightarrow{D} Z(t),\ 0 < t < F(T)$$

where $Z(t)$ is a Gaussian process with

$$\mathbb{COV}[Z(s), Z(t)] = (1-s)(1-t)B(s \wedge t).$$

Here $s \wedge t = \min(s,\ t)$ and $B(t) = \int_0^t [(1-u)^2(1-G(F^{-1}(u)))]^{-1} du$.

This result has been extended to the whole line under suitable conditions (see, for example, Gill 1983). The asymptotic behavior of the K-M estimator can be used to get results about the Kaplan-Meier quantile process.

Fix $\varepsilon > 0$ and let $t \in (\varepsilon, H(T))$. Then, it has been shown that (see, for example. Aly, Csörgő, and Horváth 1985)

$$\sqrt{N}(\hat{F}_N^{-1}(t) - F^{-1}(t)) = Z_N^*(t) \xrightarrow{D} Z^*(t),\ t \in (\epsilon, H(T))$$

where Z^* is a Gaussian process with

$$\mathbb{COV}[Z^*(s), Z^*(t)] = \frac{(1-s)(1-t)B(s \wedge t)}{f(F^{-1}(s))f(F^{-1}(t))},$$

Here the quantile process is restricted to a finite interval on both tails. Relaxing the right-tail involves rather technical conditions. In addition, the variances of the

Kaplan-Meier order statistics in the right-tail can become very large under heavy censoring. So we will keep this restriction. However, we do want to remove small order statistics (left tail) in fitting the least-squares line to the Q-Q plot. Since we are interested only in linear combinations of order statistics or (weighted) integrals of the quantile process, we can let $\varepsilon \to 0$ for such integrals under suitable conditions. We will do this under moment assumptions on F and some mild conditions on the weights w_{iM}. Required Conditions 1-3 are stated in Section 2.3.2.

2.7.1 Proof of Theorems 2.1 and 2.2

The following lemma is the key result we need to establish Theorems 2.1 and 2.2.

Lemma 2.1. *Let $\mu = 0$ and $\sigma = 1$ so that $Y_{iM}(1 \le i \le M \le N)$ denote the censored observations from F_0. Define*

$$D_N = \sum_{i=1}^{M} w_{iM}(Y_{iM} - X_{iM})^2. \tag{2.16}$$

Under Conditions 1-3, $\sqrt{N}D_N \to^P 0$ as $N \to \infty$.

Proof. Recall that $X_{iM} = F_0^{-1}(p_{iM})$. We can express (2.16) as

$$D_N = \int_0^{F(T)} W_N(t)[\hat{F}_N^{-1}(t) - F_0^{-1}(t)]^2 dt + o_p(N^{-1/2}).$$

Define the process $Q_N(t) = F_0(\hat{F}_N^{-1}(t))$. This is a censored quantile process for the uniform distribution. So, we can write

$$D_N = \int_0^{F(T)} W_N(t)[F_0^{-1}(Q_N(t)) - F_0^{-1}(t)]^2 dt + o_p(N^{-1/2}). \tag{2.17}$$

From Shorack and Wellner (Theorem 1, page 657, 1986), $\sup_{0<t<F(T)} |Q_N(t) - t| \to^{a.s.} 0$ as $N \to \infty$. This result and Conditions 1-3 can be used to show that the first term

in (2.17) above is also $o_p(N^{-1/2})$, establishing the result. □

Let $\hat{\sigma}_{RSD,w} = S_{Y,w}/S_{X,w}$ be, respectively, the ratio of sample standard deviations of the observed failures and the corresponding percentiles from the hypothesized distribution. We can now state the main asymptotic results. Define $T_{1N} = (\bar{Y}_w - \mu - \sigma\bar{X}_w), T_{2N} = (\hat{\sigma}^2_{RSD,w} - \sigma^2)$, and $T_N = (T_{1N},\ T_{2N})^T$.

Lemma 2.2. *Under Conditions 1-3, $\sqrt{N}T_N$ has a limiting bivariate normal distribution with mean $\boldsymbol{0}$ and covariance matrix Λ with elements $\lambda^*_{ij} = \sigma^{i+j}\lambda_{ij}$ where the λ_{ij}'s are given by*

$$\lambda_{ij} = \int_0^{F(T)} \int_0^{F(T)} (1-s)(1-t)B(s \wedge t)W(s)W(t)d\tilde{H}_i(s)d\tilde{H}_j(t) \tag{2.18}$$

with $s \wedge t = \min(s,t)$, $\tilde{H}_1(t) = F_0^{-1}(t)$, $\tilde{H}_2(t) = \left(F_0^{-1}(t) - m_1\right)^2/m_2$, $m_1 = \int_0^{F(T)} W(t)F_0^{-1}(t)dt$, and $m_2 = \int_0^{F(T)} W(t)[F_0^{-1}(t) - m_1]^2dt$.

Proof. Suppose $F(y) = F_0((y-\mu)/\sigma)$, then $Y^*_{iM} = (Y_{iM} - \mu)/\sigma \sim F_0$ and $\hat{F}_N^{-1}(t) = \sigma\hat{F}_{0N}^{-1}(t) + \mu$. Define

$$Z^*_N(t) = \sqrt{N}(\hat{F}_{0N}^{-1}(t) - F_0^{-1}(t)). \tag{2.19}$$

Then,

$$\begin{aligned} \sqrt{N}T_{1N} &= \sigma \int_0^{F(T)} W_N(t)\sqrt{N}(\hat{F}_{0N}^{-1}(t) - F_0^{-1}(t))dt + o_p(1) \\ &= \int_0^{F(T)} \sigma W(t)Z^*_N(t)dt + o_p(1). \end{aligned} \tag{2.20}$$

Consider now the limiting distribution of T_{2N}. It is easy to show that $\bar{X}_w \to^P m_1$

and $S^2_{X,w} \to^P m_2$. After some manipulations, we get

$$
\begin{aligned}
(2.21) \qquad \sqrt{N}T_{2N} &= \sqrt{N}[\frac{\sigma^2}{S^2_{X_w}} \sum_{i=1}^{M} w_{iM}(Y^*_{iM} - \bar{Y}^*_w)^2 - \sigma^2] \\
&= \frac{\sigma^2}{S^2_X}\Big[2\int_0^{F(T)} W_N(t)(F_0^{-1}(t) - \bar{X}_w)Z^*_N(t)dt + o_p(1)\Big] \\
&= \int_0^{F(T)} \beta(t)Z^*_N(t)dt + o_p(1),
\end{aligned}
$$

where $\beta(t) = \frac{2\sigma^2}{m_2}W(t)(F_0^{-1}(t) - m_1)$. Using these results, we get

$$
\sqrt{N}T_N \to^D N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2\lambda_{11} & \sigma^3\lambda_{12} \\ \sigma^3\lambda_{12} & \sigma^4\lambda_{22} \end{bmatrix}\right)
$$

which completes the proof. □

We are now ready to prove Theorem 2.1 and 2.2.

Proof of Theorem 2.1.

Again, it suffices to consider the case $\mu = 0, \sigma = 1$. Let $\mathcal{A}_N = \sum_{i=1}^{M} w_{iM}[(Y_{iM} - \bar{Y}_w) - (X_{iM} - \bar{X}_w)]^2$. Then

$$
\begin{aligned}
(2.22) \qquad \mathcal{A}_N &= S^2_{X,w}(\hat{\sigma}^2_{RSD,w} + 1 - 2\hat{\sigma}_w) \\
&= S^2_{X,w}[(\hat{\sigma}_{RSD,w} - 1)^2 + 2(\hat{\sigma}_{RSD,w} - \hat{\sigma}_w)].
\end{aligned}
$$

But $\mathcal{A}_N$ can also be rewritten as

$$
(2.23) \qquad \mathcal{A}_N = \sum_{i=1}^{M} w_{iM}(Y_{iM} - X_{iM})^2 - (\bar{Y}_w - \bar{X}_w)^2.
$$

From Lemma 2.1, the first term on the RHS of (2.23) is $o_p(N^{-1/2})$. The second term is also $o_p(N^{-1/2})$ from Lemma 2.2. Lemma 2.2 further shows that $(\hat{\sigma}_{RSD,w} - 1)^2 = o_p(N^{-1/2})$. Since $S^2_{X,w} \to^P m_2$, the result follows in view of (2.22). □

Proof of Theorem 2.2.

By Theorem 2.1, $(\hat{\sigma}_{RSD,w} - \hat{\sigma}_w)$ is $o_p(N^{-1/2})$. This shows that the limiting distribution of $[\hat{\mu}_w, \hat{\sigma}_w^2]$ is the same as that in Lemma 2.2. Now, we can apply Taylor Series to get the limiting distribution of $\hat{\sigma}_w$ from that of $\hat{\sigma}_w^2$. The details are straightforward. □

2.7.2 Properties of Δ_{iM}

The MLS estimator corresponds to a constant weight function $W_N(t)$, so Condition 3 in the last section is satisfied. The OLS estimator, however, corresponds to $W_N(t) = 1/\Delta_{iM}$, where $\Delta_{iM} = \hat{F}_N(Y_{iM}) - \hat{F}_N(Y_{i-1\ M})$, for $i = 1, \ldots, M$, with $\hat{F}_N(Y_{0M}) = 0$. So we need to show that this is bounded in order for Condition 3 to hold. This requires studying the behavior of the Kaplan-Meier jumps Δ_{iM}.

Define $U_N(t) = [M\Delta_{iM}]^{-1}$ for $\hat{F}_N(Y_{i-1,M}) < t \leq \hat{F}_N(Y_{iM}) \wedge \hat{F}_N(T)$ $(i = 1, \ldots, M)$ and $= 0$ for $\hat{F}_N(Y_{MM}) \wedge \hat{F}_N(T) < t \leq 1$. Further, let $U(t) = \frac{1}{1-\theta}(1 - G(F^{-1}(t)))$ for $0 < t < 1$.

Lemma 2.3.

a). $\frac{1}{N} \leq \Delta_{iM} = (1 - \hat{F}_N(Y_{i-1\ M}))\Big(\frac{1}{N-k_i-i+1}\Big) \leq 1$ *for* $i = 1, \ldots, M$.

b). $U_N(t) \to^P U(t)$ *as* $N \to \infty$.

c). If $Y_{iM} < T$, $\sqrt{N}\Delta_{iM} \to^P 0$ *as* $N \to \infty$.

Proof.

a). The fact that $\Delta_{iM} \leq 1$ is obvious. Now,

$$
\begin{aligned}
\Delta_{iM} &= \hat{F}_N(Y_{iM}) - \hat{F}_N(Y_{(i-1)\ M}) \\
&= (1 - \hat{F}_N(Y_{(i-1)\ M}))\Big(\frac{1}{N - k_i - i + 1}\Big) \\
&= \Big(\frac{1}{N-k_1}\Big)\prod_{j=1}^{i-1}\Big(\frac{N-k_j-j}{N-k_{j+1}-j}\Big).
\end{aligned} \tag{2.24}
$$

Since k_j's form a nondecreasing sequence of nonnegative integers, we must have $\Delta_{iM} \geq 1/N$.

b). First, recall that $H(\cdot)$ is a CDF of $Y^{\diamond}(= \min\{Y^0, C\})$ and that $(1 - H(\cdot)) = (1 - F(\cdot))(1 - G(\cdot))$. If $\hat{H}_N$ and $\hat{G}_N$ are K-M estimators of H and G , then $(1 - \hat{H}_N(Y^{\diamond}_{iN})) = \frac{N-i}{N} = (1 - \hat{F}_N(Y^{\diamond}_{iN}))(1 - \hat{G}_N(Y^{\diamond}_{iN}))$. Recall that k_i is the number of censored observations smaller than Y_{iM} so $Y_{iM} = Y^{\diamond}_{k_i+i\ N}$ and $Y_{i-1\ M} \leq Y^{\diamond}_{k_i+i-1\ N} < Y_{iM}$. Now suppose $Y_{iM} = Y^{\diamond}_{jN}$ (where $j = k_i + i$). Then $\hat{F}_N(Y_{iM}) = \hat{F}_N(Y^{\diamond}_{jN}) > \hat{F}_N(Y^{\diamond}_{j-1\ N}) = \hat{F}_N(Y_{i-1\ M})$ since $\hat{F}_N(\cdot)$ is a nondecreasing step function with a jump at $Y_{i-1\ M}$ and Y_{iM}. $\hat{G}_N(\cdot)$ is also a nondecreasing step function, but since it jumps at the censoring times, so there is no jump at $Y_{iM}(= Y^{\diamond}_{k_i+i\ N})$. Thus, $\hat{G}_N(Y_{iM}) = \hat{G}_N(Y^{\diamond}_{k_i+i-1\ N})$. Hence,

$$\begin{aligned} \frac{N - k_i - i + 1}{N} &= 1 - \hat{H}_N(Y^{\diamond}_{k_i+i-1\ N}) \\ &= (1 - \hat{F}_N(Y_{(i-1)\ M}))(1 - \hat{G}_N(Y_{iM})). \end{aligned}$$

Now, for t such that $\hat{F}_N(Y_{i-1\ M}) < t \leq \hat{F}_N(Y_{iM})$ for $i = 1, \ldots, M$ using the fact that $Y_{iM} = \hat{F}_N^{-1}(t)$, we get

$$\begin{aligned} U_N(t) &= [M(\hat{F}_N(Y_{iM}) - \hat{F}_N(Y_{i-1\ M}))]^{-1} \\ &= \Big(\frac{N}{M}\Big)\Big(1 - \hat{G}_N(Y_{iM})\Big) \\ &= \Big(\frac{N}{M}\Big)\Big(1 - \hat{G}_N(\hat{F}_N^{-1}(t))\Big) \\ &\to^P \frac{1}{1-\theta}\Big(1 - G(F^{-1}(t))\Big), \text{ as } N \to \infty. \end{aligned}$$

Now rewriting $Y_{MM} = \max\{Y_i^{\diamond} \mid \delta_i^{\diamond} = 1\}$, we have $Y_{MM} \to^P T = H^{-1}(1)$ as $N \to \infty$. Hence, $U_N(t) \to^P \frac{1}{1-\theta}\big(1 - G(F^{-1}(t))\big)$ for $0 \leq t \leq F(T)$. Since $(1 - H(\cdot)) =$

Table 2.5: Function $U(t) = \frac{1}{1-\theta}\big(1 - G(\sigma F_0^{-1}(t) + \mu)\big)$ under different censoring scenarios.

Censoring Setting	$U(t)$
Weibull/Weibull	$\frac{1}{1-\theta}(1-t)^{\frac{\theta}{1-\theta}}$
Weibull/uniform	$\frac{1}{1-\theta}[(1 + \frac{\log(1-t)}{a_\theta}) \vee 0]$
Lognormal/lognormal	$\frac{1}{1-\theta}[1 - \Phi(\Phi^{-1}(t) + \sqrt{2}\Phi^{-1}(\theta))]$
Lognormal/uniform	$\frac{1}{1-\theta}[(1 - \frac{e^{\Phi^{-1}(t)}}{b_\theta}) \vee 0]$
Log-logistic/log-logistic	$\frac{1}{1-\theta}[1 - \Phi_{logis}(\Phi^{-1}_{logis}(t) - c_\theta)]$
Log-logistic/uniform	$\frac{1}{1-\theta}[(1 - \frac{e^{\Phi^{-1}_{logis}(t)}}{d_\theta} \vee 0)]$

$(1-F(\cdot))(1-G(\cdot))$ and $H^{-1}(1) = T$, $\min\{(1-F(T))\ ,\ (1-G(T))\} = 0$. If $F(T) < 1$ (i.e., $(1 - F(T)) > 0$), then $(1 - G(T)) = 0$ and $U(t) = \frac{1}{1-\theta}\big(1 - G(F^{-1}(t))\big) = 0$ for $F(T) < t \leq 1$.

Therefore, for $\hat{F}_N(Y_{MM}) < t \leq 1, U_N(t) = 0 \rightarrow^P 0 = U(t)$ for $F(T) < t \leq 1$ as $N \rightarrow \infty$. Hence, $U_N(t) \rightarrow^P \frac{1}{1-\theta}\big(1 - G(F^{-1}(t))\big) = U(t)$ for $0 \leq t \leq 1$.

c). $\sqrt{N}\Delta_{iM} = \frac{\sqrt{N}}{M}(M\Delta_{iM}) = \frac{1}{\sqrt{N}} \cdot \frac{1}{M/N} \cdot \frac{1}{U_N(t)}$. Since $U_N(t) \rightarrow^P U(t) > 0$ for $t \in [0, F(T)]$, the result follows. □

For location-scale family of distributions, $U(t) = \frac{1}{1-\theta}\big(1 - G(\sigma F_0^{-1}(t) + \mu)\big)$. Table 2.5 presents expressions for $U(t)$ at different failure/censoring distribution combinations with $\mu = 0$, and $\sigma = 1$. In the table, a_θ's are 3.9207, 1.5936, .6059 for $\theta =$ .25, .50, and .75, respectively. $\Phi(t)$, and $\Phi^{-1}(t)$ are the CDF and inverse CDF of standard normal distribution. b_θ's are 6.066, 2.403, 1.045 for $\theta = .25, .50$, and .75, respectively. For the same censoring rates, c_θ's are 1.6325, 0, -1.6325, and d_θ's are 9.3466, 2.5128, 0.7336, respectively. $\Phi_{logis}(t)$ and $\Phi^{-1}_{logis}(t)$ are the CDF and quantile function of standard logistic distribution.

CHAPTER 3

Inference for Repairable Systems under Competing Risks

3.1 Introduction

Analysis of failure data arising from repairable systems has received considerable attention in the statistical, engineering, computer software, and medical literature over the past three decades. In many industrial applications, data pertaining to a repairable system is viewed as some type of *recurrent* event. Situations in which individuals or systems experience recurrent events are common in areas such as manufacturing, risk analysis, and clinical trials.

In reliability applications, one area of interest is in the study of reliability growth, which can be described as follows. At the initial stage of many production processes involving complex systems, prototypes are put into life test under a developmental testing program, corrections or design changes are made at the occurrences of failures, and the modified system is tested again. As this test-redesign-retest sequence contributes to an improvement in the system performance, failures become increasingly sparse at the later stages of testing making it more difficult to assess the *current reliability*, a quantity of utmost importance to reliability engineers. A reliability growth (RG) model provides a structure through which the failure data from the current as well as previous stages of testing could be analyzed in an integrated way in order to make efficient inference on system reliability, and other aspects of the underlying failure process.

In a medical application, the 'failures' are translated into *times until the occurrence* of a recurrent event (e.g., appearance of tumor) in individuals. Clinical experiments typically consist of a fairly large number of individuals observed over a relatively short period of time. This is also common with databases of manufactured products generating warranty claims. By contrast, most of the field and bench test data for demonstrating product reliability consist of a very small number of prototypes put under test for a fairly long time. The major objective in either situation is to study the rates of recurrence of the events in question, compare different systems, assess the effect of explanatory variables, or predict a future event.

This chapter investigates some models and methodologies for analyzing failures from repairable systems with multiple failure modes. For non-repairable systems, analysis of failures under multiple failure modes, traditionally studied under the broad umbrella of 'competing risks', is typically undertaken in the framework where occurrence of a system failure is caused by the earliest onset of any of the component failures (a series system). In medical applications, often the censoring mechanism is viewed as a competing risk to the event of interest which could be recurrent or otherwise. There is a substantial body of research in the area of competing risks analysis. Instead of referring to individual articles, we cite here a recent book by Crowder (2001) that provides an excellent overview of models and methodologies pertaining to failure data subject to competing risks. There is also a great deal of research within the reliability community on competing risks models where the exact causes of failure are observed.

In engineering applications, one, however, encounters data where the cause-of-failure is not completely known (c.f. Reiser, Guttman, Lin, Guess, and Usher 1995; Reiser, Flehinger, and Conn 1996). In the statistical literature, such data are termed as *masked* failure data. Masking is often the manifestation of an attempt to expedite the process of repair by replacing the entire subset of components responsible for fail-

ure instead of further investigation towards identifying the specific component which is the culprit. Failure data with masked cause-of-failure is also common in medical applications. The sources of failure (death) in a medical context typically refer to various potential risk factors for a patient observed in a clinical study. Available applications include patients in a heart transplant study (Greenhouse and Wolfe 1984) and breast cancer patients (Cummings et al. 1986) observed longitudinally over several years.

While there has been substantial research in competing risks for non-repairable systems, very little has been done in the area of analyzing failures of systems and associated components that are subjected to multiple recurrent failure processes. Majumdar (1993) documents 262 recurrent failure times of a vertical boring machine spanning a total of 18,285 hours along with the components that are responsible for the failure. Recently, Langseth and Lindqvist (2006) reported cumulative service times of a component spanning over 1600 time units, amplifying each failure with the specific mode causing the failure. The causes were categorized into two broad groups, with several sub-causes specified under each. Both of these examples deal with failures of a single system. Lawless et al. (2001) analyze repeated shunt failures in infants diagnosed with hydrocephalus, where the failures are known to occur due to a variety of causes. All of these examples clearly demonstrate a need for a systematic development of methodologies to analyze recurrent failures under competing risks, which is the focus of this part of the dissertation.

The rest of the chapter is organized as follows. Section 3.2 presents the general framework for analyzing repairable systems subject to multiple failure modes, and some key issues. In the context of analyzing failures of a single repairable system, Section 3.3 discusses statistical inference for a parametric model that extends from a popular single-failure-mode model. Section 3.4 discuses the results for the multiple repairable systems under a nonhomogeneous Poisson process (NHPP) framework.

Section 3.5 discusses an example for a single repairable system case. Section 3.6 discusses the simulation results, section 3.7 summarizes the findings and Section 3.8 discusses possible future research.

3.2 Framework and Results for General Counting Processes

Consider a single repairable system , and let $0 < T_1 < T_2 < \ldots$ be the (cumulative) failure times of the system observed until a fixed time (Type-I censoring) or fixed number of failures (Type-II censoring). Suppose there are J failure modes, and at the $i-$th failure time T_i, we also observe $\delta_i \in \{1, \ldots, J\}$, the failure mode indicator associated with the $i-$th failure. We can view the sequence (T_i, δ_i) as observations of a marked point process. Let $N(a, b]$ denote the number of system failures in time interval $(a, b]$. An essential ingredient in describing recurrent failures under competing risks is the cause-specific intensity associated to failure mode-j is defined as

$$\lambda_j(t) = \lim_{h \to 0} \frac{Pr(\delta(t) = j, N(t, t+h] \geq 1)}{h}.$$

The overall system intensity function equals

$$\lambda(t) = \lim_{h \to 0} \frac{Pr(N(t, t+h] \geq 1)}{h} = \sum_j \lambda_j(t).$$

One can conceptually introduce a counting process N_j with an associated (marginal) intensity function $\lambda^{(j)}$, which simply records the failures due to mode-j for each $j = 1, \ldots, J$. Under that framework, $N = \sum N_j$ is the process associated with all the failures and is simply the superposition of N_j's. It is important to note, however, that under a competing risks framework, one does not observe the marginals, and in general $\lambda_j \neq \lambda^{(j)}$. The only exception is the case where the failure modes can be assumed to act independently of each other. In this case, the cause-specific quantities match the marginals, and one can carry out inference without having to

deal with issues of identifiability.

Under some general conditions, the cause-specific intensities completely determine the likelihood function of the observed data. The following theorem demonstrates that a special structure of the intensity functions provides a necessary and sufficient condition for the time and cause of failure to be stochastically independent. Note that this is simply the repairable system version of an analogous result in Elandt-Johnson (1976) (see also Kochar and Proschan 1991).

Theorem 3.1. *The cumulative failure time and the mode of failure are independent if and only if $\lambda_j(t)$'s are proportional to each other, or equivalently $\lambda_j(t) = p_j\lambda(t)$ where p_j's are non-negative constants with $\sum_{j=1}^{J} p_j = 1$.*

Proof. Conditional on the fact that a failure has occurred at time point t, the probability that the failure is caused by mode-j is given by

$$
\begin{aligned}
(3.1) \qquad p_j(t) &\equiv \lim_{h\to 0} Pr(\delta(t) = j | N(t, t+h] \geq 1) \\
&= \frac{\lim_{h\to 0} Pr(\delta(t) = j, N(t, t+h] \geq 1)/h}{\lim_{h\to 0} Pr(N(t, t+h] \geq 1)/h} = \lambda_j(t)/\lambda(t).
\end{aligned}
$$

The left hand side of the above equation is independent of t if and only if the right hand side is, establishing that the time and cause of failure are independent if and only if $\lambda_j(t)/\lambda(t) = p_j$ for $j = 1, ..., J$. □

Theorem 3.1 describes the marginal distribution of $\delta(t)$. The following theorem goes further to describe the joint distribution of the cause-specific counting processes N_j's. Suppose that the λ_j's are proportional to each other, i.e., $\lambda_j(t) = p_j\lambda(t)$ as described in Theorem 3.1. Let $T_1 < T_2 < \ldots$ be the cumulative failure times for the system, given that $N(T) = n$, and let $\delta_1, \delta_2, \ldots$ be the failure modes corresponding to T_i's. For any fixed time T, let $N_j(T) = \sum_{i:T_i \leq T}\{\delta_i = j\}$ for $j = 1, \ldots, J$ and $X = \{N_1(T), N_2(T), \ldots, N_J(T)\}$.

Theorem 3.2. *$X|N(T) = n \sim Multinomial(n, p_1, \ldots, p_J)$ if $N(t)$, the overall system counting process, has independent increments and all failure modes are mutually independent.*

Proof. Recall that $N_j(T) = \sum_{i=1}^{n} 1\{\delta_i = j\}$ for $j = 1, \ldots, J$. So it is sufficient to show that δ_i's are *iid* with $Pr(\delta_i = j) \equiv p_j$ for $i = 1, \ldots, n$. Consider the case $n = 2$.

$$
\begin{aligned}
&Pr(\delta_1 = k, \delta_2 = l) \\
&= \lim_{h \to 0} Pr(\delta(T_1) = k, \delta(T_2) = l | N(T_1, T_1 + h] \geq 1, N(T_2, T_2 + h] \geq 1]) \\
&= \lim_{h \to 0} \frac{Pr(\delta(T_1) = k, \delta(T_2) = l, N(T_1, T_1 + h] \geq 1, N(T_2, T_2 + h] \geq 1)}{Pr(N(T_1, T_1 + h] \geq 1, N(T_2, T_2 + h] \geq 1)} \\
&= \lim_{h \to 0} \frac{Pr(\delta(T_1) = k, N(T_1, T_1 + h] \geq 1)}{Pr(N(T_1, T_1 + h] \geq 1)} \times \frac{Pr(\delta(T_2) = l, N(T_2, T_2 + h] \geq 1)}{Pr(N(T_2, T_2 + h] \geq 1)} \\
&= \lim_{h \to 0} Pr(\delta(T_1) = k | N(T_1, T_1 + h] \geq 1) \times \lim_{h \to 0} Pr(\delta(T_2) = l | N(T_2, T_2 + h] \geq 1) \\
&= Pr(\delta_1 = k) \times Pr(\delta_2 = l)
\end{aligned}
$$

where the third equality holds when $N(t)$ has independent increments and all failure modes are mutually independent.

Now since λ_j's are proportional, it follows from Theorem 3.1 that $Pr(\delta_i = j) = p_j$ for all $i = 1, 2, \ldots, n$. Therefore, when $N(t)$ has independent increments, it follows that the δ_i's are *iid* with $Pr(\delta_i = j) = p_j$ for $j = 1, 2, \ldots, J$. □

3.2.1 Tail Behavior of the Counting Processes and Failures

This section characterizes the tail behavior of $N_j(t)$ as $t \to \infty$ under the simple condition 3.1 below. To simplify the problem, we consider the counting processes $N_1(t)$ and $N_2(t)$ corresponding to failure modes 1 and 2 respectively. Let $\Lambda_j(t)$ for $j = 1, 2$ be the corresponding cumulative intensity functions, which, under mild

conditions, match $\mathbb{E}[N_j(t)]$. Throughout, we suppose that the $\Lambda_j(t)$'s are monotone increasing.

Condition 3.1. *Suppose* $\Lambda_j(t) \to \infty$ *and* $N_j(t)/\Lambda_j(t) \to^{a.s.} 1$ *as* $t \to \infty$ *for* $j = 1, 2$.

Note that Condition 3.1 does not make any assumption about independent increments or independence among failure modes. An example of a counting process that satisfies Condition 3.1 is a nonhomogeneous Poisson process (NHPP) with a cumulative intensity function that grows as a positive power of time.

Lemma 3.1. *Under Condition 3.1,* $\Lambda(t) \to \infty$ *and* $N(t)/\Lambda(t) \to^{a.s.} 1$ *as* $t \to \infty$.

Proof. With $N(t) = N_1(t) + N_2(t)$, it follows that $\Lambda(t) = \mathbb{E}[N(t)] = \mathbb{E}[N_1(t)] + \mathbb{E}[N_2(t)] = \Lambda_1(t) + \Lambda_2(t) \to \infty$ as $t \to \infty$. Now we consider below,

$$N(t)/\Lambda(t) = \frac{\Lambda_1(t)}{\Lambda(t)} \cdot \frac{N_1(t)}{\Lambda_1(t)} + \frac{\Lambda_2(t)}{\Lambda(t)} \cdot \frac{N_2(t)}{\Lambda_2(t)}.$$

Since $N_j(t)/\Lambda_j(t) \to^{a.s.} 1$ for $j = 1, 2$ as $t \to \infty$, it follows that $N(t)/\Lambda(t) \to^{a.s.} 1$ as $t \to \infty$. □

This result simply states the property of overall counting process under Condition 3.1. One may also be interested in the case where the systems are observed under failure censoring, i.e., they are observed until the $n-$th failure with fixed n. Hence, the observed cumulative failure times are $T_1 < T_2 < \ldots < T_n$. If T_n is the $n-$th failure time, we have $N(T_n) = n$. Let $n_j = N_j(T_n)$ for $j = 1, 2$ be the number of cause-specific failures under failure censoring. We have $n_2 = n - n_1$. Condition 3.1, then needs to be replaced by Condition 3.1′.

Condition 3.1′. $n_j/\Lambda_j(T_n) \to^{a.s.} 1$ *as* $n \to \infty$ *for* $j = 1, 2$.

Now we study the limiting rate of n_j's, for failure censoring with n going to infinity, under two cases of interest: (a) $0 < c \leq 1$ and (b) $c = 0$, where $c = \lim_{t\to\infty} \frac{\Lambda_2(t)}{\Lambda_1(t)}$.

Not that any $c > 1$ situation corresponds to one of (a) or (b) if we interchange the roles of Λ_1, Λ_2.

Case 1: $0 < c \leq 1$

Theorem 3.3. *Suppose Condition 3.1′ is satisfied and* $\lim_{t\to\infty} \frac{\Lambda_2(t)}{\Lambda_1(t)} = c$ *with* $0 < c \leq 1$. *Then,* $n_1/n \to^{a.s.} 1/(1+c)$ *and* $n_2/n \to^{a.s.} c/(1+c)$ *as* $n \to \infty$.

Proof. Since $\lim_{t\to\infty} \frac{\Lambda_1(t)}{\Lambda(t)} = 1/(1+c)$, $T_n \to \infty$ and $n_j/\Lambda_j(T_n) \to^{a.s.} 1$ as $n \to \infty$ for $j = 1, 2$,

$$\frac{n_1}{n} = \frac{N_1(T_n)}{\Lambda_1(T_n)} \times \frac{\Lambda_1(T_n)}{\Lambda(T_n)} \times \frac{\Lambda(T_n)}{n} \to^{a.s.} \frac{1}{1+c} \text{ as } n \to \infty.$$

And similarly,

$$\frac{n_2}{n} = \frac{N_2(t_n)}{\Lambda_2(t_n)} \times \frac{\Lambda_2(t_n)}{\Lambda(t_n)} \times \frac{\Lambda(t_n)}{n} \to^{a.s.} \frac{c}{1+c} \text{ as } n \to \infty.$$

□

Case 2: $c = 0$

When $c = 0$, we have $\frac{\Lambda_1(t)}{\Lambda(t)} \to^{a.s.} 1$ and $\frac{\Lambda_2(t)}{\Lambda(t)} \to^{a.s.} 0$ as $t \to \infty$. This simply implies that as we observe the system failure for the very long period of time, almost all of systems failures would be coming from failure mode 1.

On the other hand, it follows from Conditions 3.1 and 3.1′ that n_2 still goes to infinity. Therefore, it is of interest to study the rate of n_2 as $n \to \infty$. Theorem 3.4 below characterizes this under some situations. The definition of a regularly varying function and the proof are deferred to Section 3.9.

Theorem 3.4. *Suppose Condition 3.1 is satisfied and* $\Lambda_2 \circ \Lambda_1^{-1}(s) = \Lambda_2(\Lambda_1^{-1}(s))$ *with* $\Lambda_1^{-1}(s) = \inf\{t : \Lambda_1(t) = s, t \geq 0\}$. *If* $\lim_{t\to\infty} \frac{\Lambda_2(t)}{\Lambda_1(t)} = 0$, *then*

a). $n_1/n \to^{a.s.} 1$ *as* $n \to \infty$,

b). if $\Lambda_2 \circ \Lambda_1^{-1}$ *is a regularly varying function with exponent* $0 \leq \rho < \infty$*,*

$n_2/\Lambda_2 \circ \Lambda_1^{-1}(n) \rightarrow^{a.s.} 1$ *as* $n \rightarrow \infty$*.*

3.3 Inference for a Single Repairable System

This section discusses statistical analysis of a single repairable system under competing risks. With data only from a single system, the intensity function can be estimated using the natural estimator (Ascher and Feingold 1984),

$$\hat{\lambda}(t) = \frac{\text{Number of failures in } [t, t + \Delta t]}{\Delta t}.$$

This can also be applied to the the cause-specific intensity function,

$$\hat{\lambda}_j(t) = \frac{\text{Number of mode-}j \text{ failures in } [t, t + \Delta t]}{\Delta t},$$

for $j = 1, \ldots, J$.

It is common to assume a parametric model when dealing with a single system, which allows one (a) to assess growth or decay in reliability, and (b) prediction in a formal model based way. The most common model assumed for the system failures is a nonhomogeneous Poisson process (NHPP) with parametric intensity function $\lambda(t; \phi)$ where ϕ is the set of parameter(s) for $\lambda(\cdot)$. For this section and the next, we assume that the failures from each mode follows a NHPP, with power law cumulative intensity function which is also known as Power Law Process (PLP).

3.3.1 Parametric Model under Power Law Process (PLP)

To simplify the problem, we assume that the system has two independent failure modes. The system is observed until the $n-$th failure, analogous to Type-II censoring. We also assume that the failures from mode-j follows a NHPP with intensity function

$$\lambda_j(t) = \mu_j \beta_j t^{\beta_j - 1}, \text{ for } j = 1, 2.$$

This is also called power law process (PLP) and has been popularized by Crow (1974) as a reasonable model for analyzing failures from a repairable system. The linear relation of the intensity (or cumulative intensity) with time in log-log scale conforms to various failure processes encountered in manufacturing or automotive engineering. Further the elegance of the statistical properties and the ease of its implementation have made PLP a popular choice of reliability practitioners over the decades.

It follows that the overall system intensity function is

$$\lambda(t) = \lambda_1(t) + \lambda_2(t) = \mu_1\beta_1 t^{\beta_1 - 1} + \mu_2\beta_2 t^{\beta_2 - 1}$$

which yields a PLP if and only if the shape parameters from both modes are equal ($\beta_1 = \beta_2$). With only two failure modes, we let $\delta_j = 1$ for failure mode-1 and $= 0$ otherwise. The observed data are $\{t_i, \delta_i\}_{i=1,\ldots,n}$ with $0 < t_1 < \ldots < t_n$. The general form of the likelihood function can written as

$$L(Data|\phi) = \prod_{i=1}^{n} \lambda_1(t_i)^{\delta_i} \lambda_2(t_i)^{1-\delta_i} \exp\left[-\int_0^{t_n} \lambda(t)dt\right]$$

with $\phi = (\mu_1, \mu_2, \beta_1, \beta_2)$. The MLEs of the parameters and their properties are discussed next.

Case: Equal Shape Paramters ($\beta_1 = \beta_2$)

We start with the case where equality of β_1 and β_2 is assumed. This is, of course, a testable assumption and should be checked first. The case, nonetheless, is an important one as it is the only case where both the individual and system level failure processes conform to PLP. In this case, the cause-specific intensity for mode-j is $\lambda_j(t) = \mu_j\beta t^{\beta-1}, j = 1, 2$. The cause specific intensities are proportional and the overall system intensity is $\lambda(t) = \lambda_1(t) + \lambda_2(t) = (\mu_1 + \mu_2)\beta t^{\beta-1}$, which is also the intensity from PLP. With $n_1 = \sum_{i=1}^{n} \delta_i$, the likelihood function is

$$L(Data|\mu_1, \mu_2, \beta) = \mu_1^{n_1}\mu_2^{n-n_1}\beta^n \prod_{i=1}^{n} t_i^{\beta-1} \exp\left[-(\mu_1 + \mu_2)t_n^{\beta}\right].$$

By maximizing the likelihood above, we get the MLEs as

$$\hat{\mu}_1 = n_1/t_n^{\hat{\beta}}, \quad \hat{\mu}_2 = (n-n_1)/t_n^{\hat{\beta}}, \quad \hat{\beta} = n/\sum_{i=1}^{n}\log(t_n/t_i). \tag{3.2}$$

Since the overall system has the intensity function from PLP, the following results are readily obtained from the single component case results (see Bain 1978, Theorem 5.2.1):

$$2n\beta/\hat{\beta} \sim \chi^2_{2(n-1)}, \quad 2(\mu_1+\mu_2)t_n^{\beta} \sim \chi^2_{2n}. \tag{3.3}$$

With proportional cause-specific intensity functions and independent increment property of the NHPPs, we can apply Theorems 3.1 and 3.2 to get the following results:

$$\delta_i \overset{iid}{\sim} Bernoulli\left(\tfrac{\mu_1}{\mu_1+\mu_2}\right), \quad n_1 \sim Bin\left(n, \tfrac{\mu_1}{\mu_1+\mu_2}\right), \tag{3.4}$$

where both terms are independent of t.

Note that $2n\beta/\hat{\beta},\ 2(\mu_1+\mu_2)t_n^{\beta}$, and n_1 (or δ_i) are mutually independent. These exact distributions are useful for obtaining the inference results, as we illustrate next. To obtain the confidence region for parameters, it is convenient to reparameterize the parameters to (ρ, μ, β) with $\mu = \mu_1 + \mu_2$ and $\rho = \mu_1/\mu$. It follows that

$$\hat{\rho} = n_1/n, \quad \hat{\mu} = n/t_n^{\hat{\beta}} \tag{3.5}$$

and $\hat{\beta}$ remains the same. Then the confidence region for μ and β can be obtained from the results in Finkelstein (1976) and Bhattacharyya and Ghosh (1991). From these and from (3.3) and (3.4), a $100(1-\alpha)\%$ confidence region for (ρ, μ, β) is provided by the set

$$\mathcal{D} = \{(\rho, \mu, \beta) : a_{1n} < X < a_{2n}, b_{1n} < Y < b_{2n}, c_{1n} < Z < c_{2n}\}$$

where $Pr(\mathcal{D}) = 1-\alpha$ with $X = \hat{\rho}, Y = 2n\beta/\hat{\beta}$ and $Z = 2\mu t_n^{\beta}$. Using the fact that X are independent of Y and Z, we recommend choosing $\mathcal{D} = \mathcal{D}_1 \cap \mathcal{D}_2$ such that

$$\mathcal{D}_1 = \{\rho : a_{1n} < X < a_{2n}\} \ \text{ and } \ \mathcal{D}_2 = \{(\mu, \beta) : b_{1n} < Y < b_{2n}, c_{1n} < Z < c_{2n}\}$$

Note that one can easily find a confidence region $\mathcal{D}_2$ with an exactly specified probability content. For $\mathcal{D}_1$, one can either choose an interval based on normal approximation, or can choose an conservative confidence interval based on the exact 'binomial' distribution of nX. Finally, the overall probability content of $\mathcal{D}$ is obtained as the product of those of $\mathcal{D}_1$ and $\mathcal{D}_2$.

Prediction Intervals for Future Failure Times

Since the overall system has the intensity function of a PLP, the prediction interval for t_{n+k}, the $(n+k)-$th system failure time, follows the result from the single component system in Bain (1978). We can get a lower $(1-\alpha)$ prediction limit for t_{n+k} as

$$T_L(k,\alpha) = t_n \exp\left[\frac{\nu_k F_\alpha(\nu_k, 2(n-1))}{2(n-1)c_k\hat{\beta}}\right]$$

where $c_k = \frac{\psi(n+k)-\psi(n)}{n[\psi'(n)-\psi(n+k)]}$, $\nu_k = 2n[\psi(n+k)-\psi(n)]c_k$, and $F_\alpha(a,b)$ is the $\alpha-$th quantile for F-distribution with degrees of freedom a and b. $\psi(\cdot)$ is the digamma function. For $k=1$, the next system failure time, the formulation is simply reduced to

$$T_L(1,\alpha) = t_n \exp\left[\frac{(1-\alpha)^{-1/(n-1)}-1}{\hat{\beta}}\right].$$

To obtain the prediction interval for the future failure time from a specific failure mode, assume first that ρ is known. Now consider the probability of t_{n+l} being the next mode-1 failure time, with δ being independent of t,

$$\begin{aligned} &Pr(t_{n+l}, \delta_{n+l}=1, \delta_{n+1}=\ldots=\delta_{n+l-1}=0|Data) \\ &\quad = Pr(\delta_{n+l}=1, \delta_{n+1}=\ldots=\delta_{n+l-1}=0) \times Pr(t_{n+l}|Data) \\ &\quad = \rho(1-\rho)^{l-1} Pr(t_{n+l}|Data) \end{aligned}$$

with $Pr(t_{n+l}|Data) \propto t_{n+l}^{\beta-1}\exp(-\mu t_{n+l}^{\beta}) I(t_{n+l} > t_n)$ from Bain (1978). We can see here that, $l \sim Geo(\rho)$. Hence, if we let $t_1^{(M1)}$ be the next mode-1 failure time, then

$$Pr(t_1^{(M1)}|Data) = \sum_{l=1}^{\infty} \rho(1-\rho)^{l-1} Pr(t_{n+l}|Data).$$

Hence, the lower $(1-\alpha)$ prediction limit for $t_1^{(M1)}$ is

$$T_{M1,L}(1,\alpha) = \left\{ T : \sum_{l=1}^{\infty} \rho(1-\rho)^{l-1} Pr(t_{n+l} \geq T | Data) = 1-\alpha \right\}.$$

In the similar fashion, the future $k-$th failure time from mode-1 is

$$T_{M1,L}(k,\alpha) = \left\{ T : \sum_{l=k}^{\infty} \begin{pmatrix} l-1 \\ k-1 \end{pmatrix} \rho^k (1-\rho)^{l-k} Pr(t_{n+l} \geq T | Data) = 1-\alpha \right\}.$$

In practice, ρ is unknown, and it is reasonable to use $\hat{\rho}$ instead. We recommend the following procedure for obtaining the lower $(1-\alpha)$ prediction limit for $t_1^{(M1)}$: for $b = 1, \ldots, B$,

1. draw $n_1^{(b)}$ from $Bin(n, \hat{\rho})$ and let $\rho^{(b)} = n_1^{(b)}/n$ [With large n, one can draw $\rho^{(b)}$ from $N(\hat{\rho}, \hat{\rho}(1-\hat{\rho})/n)$.]

2. draw $l^{(b)}$ from $Geo(\rho^{(b)})$,

3. draw $y^{(b)}$ from $F\left(\nu_{l^{(b)}}, 2(n-1)\right)$.

4. let $t^{(b)} = t_n \exp[\nu_{l^{(b)}} y^{(b)}/2(n-1)c_{l^{(b)}}\hat{\beta}]$.

Then, $T_{M1,L}(1,\alpha) = \frac{1}{B}\sum_{b=1}^{B} t^{(b)}$ can be used as the estimated for the lower bound for of the next failure time caused by mode-1. The $T_{M1,L}(k,\alpha)$ can also be obtained by modifying Step 2 above. Similar arguments can be applied for prediction of future failure time from failure mode-2.

Large-sample Distributions of Estimators

We now study the asymptotic properties of the MLEs. Similar to Theorem 2.1 in Bhattacharyya and Ghosh (1991), the MLEs are consistent but converge at different rates.

Theorem 3.5. *With $\hat{\rho}, \hat{\mu}$, and $\hat{\beta}$ defined in (3.2) and (3.5), let*

$$\begin{aligned} W_{1n} &= \sqrt{n}(\hat{\rho} - \rho), \\ W_{2n} &= \sqrt{n}(\log n)^{-1}(\hat{\mu} - \mu), \\ W_{3n} &= \sqrt{n}(\hat{\beta} - \beta). \end{aligned}$$

Then $W_n = (W_{1n}, W_{2n}, W_{3n})'$ is asymptotically (singular) multivariate normal $N_3(\mathbf{0}, \mathbf{\Sigma})$ where

$$\mathbf{\Sigma} = \begin{bmatrix} \rho(1-\rho) & 0 & 0 \\ 0 & \mu^2 & -\mu\beta \\ 0 & -\mu\beta & \beta^2 \end{bmatrix}.$$

Proof. Since μ, β can be envisioned as the parameters from the PLP governing the system failures, it follows from Theorem 2.1 in Bhattachryya and Ghosh (1991) that

$$(W_{2n}, W_{3n})' \xrightarrow{d} N_2\left(\mathbf{0}, \begin{bmatrix} \mu^2 & -\mu\beta \\ -\mu\beta & \beta^2 \end{bmatrix}\right).$$

And from (3.4), we have $n_1 \sim Bin(n, \rho)$, independently of $\hat{\mu}, \hat{\beta}$. Hence $W_{1n} \xrightarrow{d} N(0, \rho(1-\rho))$ and is independent from W_{2n} and W_{3n}. The result follows. □

We can reformulate the above result in terms of the original parameters. The vector $W_n^* = [\sqrt{n}(\log n)^{-1}(\hat{\mu}_1 - \mu_1),\ \sqrt{n}(\log n)^{-1}(\hat{\mu}_2 - \mu_2),\ \sqrt{n}(\hat{\beta} - \beta)]'$ is asymptotically normal with mean vector $\mathbf{0}$ and variance-covariance matrix

$$\Sigma^* = \begin{bmatrix} \mu_1^2 & \mu_1\mu_2 & -\mu_1\beta \\ \mu_1\mu_2 & \mu_2^2 & -\mu_2\beta \\ -\mu_1\beta & -\mu_2\beta & \beta^2 \end{bmatrix}.$$

The asymptotic result provides some curious insights into the behavior of the MLEs. An interesting observation regarding the large-sample behavior possibly lies

at the root of the pathology we observe here. The joint asymptotic of $\phi = (\mu_1, \mu_2, \beta)$ rests on the crucial Taylor-expansion step, that can be expressed (after suitable scaling) as

$$\begin{pmatrix} n^{-1/2}(\log n)^{-1} l_{1n}(\phi^0) \\ n^{-1/2}(\log n)^{-1} l_{2n}(\phi^0) \\ n^{-1/2} l_{3n}(\phi^0) \end{pmatrix} = (\log n) C_n(\zeta_n^*) W_n \tag{3.6}$$

where $l_n = (l_{1n}, l_{2n}, l_{3n})'$ is the score vector, ϕ^0 is the true parameter value, C_n is the (appropriately scaled) second derivative of the log-likelihood, ζ_n^* is an intermediate random point between $\hat{\phi}$ and ϕ^0. While $W_n = O_p(1)$, C_n can be shown to converge to a nonstochastic singular matrix uniformly in a neighborhood around ϕ^0. In spite of the presence of the growing multiplier of $(\log n)$ on the right of (3.6), the left hand side converges to a (singular) normal, thereby making it a $O_p(1)$ term. This is in stark contrast to the classical *iid* asymptotics.

Case: Unequal Shape Parameters ($\beta_1 \neq \beta_2$)

The likelihood function is now

$$L(Data|\mu_1, \mu_2, \beta_1, \beta_2) = \prod_{i=1}^{m} \left[\mu_1 \beta_1 t_i^{\beta_1 - 1}\right]^{\delta_i} \left[\mu_2 \beta_2 t_i^{\beta_2 - 1}\right]^{1-\delta_i} \exp\left[-(\mu_1 t_n^{\beta_1} + \mu_2 t_n^{\beta_2})\right].$$

The likelihood above is simply the product of likelihood contributed from each independent component. By maximizing the likelihood, we get the MLEs as

$$\hat{\mu}_1 = n_1 / t_n^{\hat{\beta}_1}, \quad \hat{\beta}_1 = n_1 / \textstyle\sum_{i=1}^{n} \log(t_n/t_i)\delta_i, \tag{3.7}$$

$$\hat{\mu}_2 = (n - n_1) / t_n^{\hat{\beta}_2}, \quad \hat{\beta}_2 = (n - n_1) / \textstyle\sum_{i=1}^{n} \log(t_n/t_i)(1 - \delta_i).$$

Note that the MLEs for each failure mode are simply the functions of t_i's corresponding to their own failure mode.

We now study the exact distribution of estimators. The overall intensity is $\lambda(t) = \mu_1 \beta_1 t^{\beta_1 - 1} + \mu_2 \beta_2 t^{\beta_2 - 1}$. Now using the property of NHPP, it follows that

$$\mu_1 t_n^{\beta_1} + \mu_2 t_n^{\beta_2} \quad \sim \quad Gamma(n, 1). \tag{3.8}$$

And by using (3.1), we have

$$\delta_i|t_i \sim Bernoulli\left(\frac{\mu_1\beta_1 t_i^{\beta_1-1}}{\mu_1\beta_1 t_i^{\beta_1-1}+\mu_2\beta_2 t_i^{\beta_2-1}}\right), \tag{3.9}$$

which now depends on t_i. Using the NHPP property of each failure mode, we have $n_1|t_n = N_1(0,t_n] \mid \{N_1(0,t_n]+N_2(0,t_n] = n, t_n\}$ where $N_j(a,b]$, for $j=1,2$, is the number of failures from failure mode-j in the interval $(a,b]$. Since $N_j(0,t_n]|t_n \sim Poisson\left(\int_0^{t_n}\lambda_j(t)dt\right)$, it follows that

$$n_1|t_n \sim Bin\left(n, \frac{\mu_1 t_n^{\beta_1}}{\mu_1 t_n^{\beta_1}+\mu_2 t_n^{\beta_2}}\right). \tag{3.10}$$

Now by conditioning on n_1 and the cause of the last failure δ_n, we can use the results from a single-failure-mode system with failure censoring to get

$$\begin{aligned} 2\mu_1 t_n^{\beta_1}|n_1,\delta_n=1 &\sim \chi^2_{2n_1}, \\ 2\mu_2 t_n^{\beta_2}|n_1,\delta_n=0 &\sim \chi^2_{2(n-n_1)}, \\ 2n_1\beta_1/\hat{\beta}_1|n_1,\delta_n &\sim \chi^2_{2(n_1-\delta_n)}, \\ 2(n-n_1)\beta_2/\hat{\beta}_2|n_1,\delta_n &\sim \chi^2_{2(n-n_1-(1-\delta_n))}. \end{aligned} \tag{3.11}$$

Further, all the terms in (3.11) are mutually independent. Combining the last two terms in (3.11), we get the exact (unconditional) distribution below,

$$2n_1\beta_1/\hat{\beta}_1 + 2(n-n_1)\beta_2/\hat{\beta}_2 \sim \chi^2_{2(n-1)}. \tag{3.12}$$

Based on the results in (3.11), one can obtain exact conditional inference. The unconditional distributions, however, are complicated and do not lend themselves to tractable inference.

The asymptotic distribution for the estimators is studied next. We require a lemma to prepare the groundwork.

Lemma 3.2. *Suppose $\beta_1 > \beta_2$, as $n \to \infty$,*

a). $n_1/n \to^{a.s.} 1$,

b). $\log n_1 / \log n \to^{a.s.} 1$,

c). $(n - n_1)/\mu_2(n/\mu_1)^{\beta_2/\beta_1} \to^{a.s.} 1$, *and*

d). $\log(n - n_1)/\log(\mu_2(n/\mu_1)^{\beta_2/\beta_1}) \to^{a.s.} 1$.

Proof. The results simply follow Theorems 3.4 and 3.11. □

Lemma 3.2 simply shows that with $\beta_1 > \beta_2$, both n_1 and $n - n_1$ go to infinity almost surely at rates, n and n^{β_2/β_1}, respectively.

Theorem 3.6. *With $\hat{\mu}_1, \hat{\mu}_2, \hat{\beta}_1$, and $\hat{\beta}_2$ defined as in (3.7), and $\beta_1 > \beta_2$, let*

$$\begin{aligned} U_{1,n} &= \sqrt{n}(\log n)^{-1}(\hat{\mu}_1 - \mu_1) \\ U_{2,n} &= \sqrt{n}(\hat{\beta}_1 - \beta_1) \\ V_{1,n} &= \sqrt{n^{\beta_2/\beta_1}}(\log n)^{-1}(\hat{\mu}_2 - \mu_2) \\ V_{2,n} &= \sqrt{n^{\beta_2/\beta_1}}(\hat{\beta}_2 - \beta_2). \end{aligned}$$

Then, $U_n = (U_{1,n}, U_{2,n})' \to^D U \sim N_2(\mathbf{0}, \Sigma_1)$ and $V_n = (V_{1,n}, V_{2,n})' \to^D V \sim N_2(\mathbf{0}, \Sigma_2)$ where U and V are independent and

$$\Sigma_1 = \begin{pmatrix} \mu_1^2 & -\mu_1\beta_1 \\ -\mu\beta_1 & \beta_1^2 \end{pmatrix}, \quad \Sigma_2 = \mu_1^{\beta_2/\beta_1}\beta_2^2 \begin{pmatrix} \mu_2/\beta_1^2 & -1/\beta_1 \\ -1/\beta_1 & 1/\mu_2 \end{pmatrix}.$$

Proof. Let

$$S_{1n} = \beta_1 \sum_{i=1}^{n} \log(t_n/t_i)\delta_i \text{ and } S_{2n} = \beta_2 \sum_{i=1}^{n} \log(t_n/t_i)(1 - \delta_i).$$

Now we define

$$\upsilon_{1n} = (S_{1n} - n_1)/\sqrt{n_1} \quad \text{and} \quad \upsilon_{2n} = (S_{2n} - (n - n_1))/\sqrt{n_2}.$$

We need to show that $(\upsilon_{1n}, \upsilon_{2n})' \rightarrow^D N_2(\mathbf{0}, \mathbf{I_2})$, as $n \rightarrow \infty$ with $\mathbf{I_2}$ denotes 2x2 identity matrix. This can be shown by using moment generating functions of υ_{1n} and υ_{2n}. First, we assume $\delta_n = 0$, then by (3.11), we have $S_{1n}|n_1 \sim Gamma(n_1, 1)$. Hence,

$$\mathbb{E}[e^{\upsilon_{1n}t}] = \mathbb{E}_{n_1}\left[\mathbb{E}\left(e^{\frac{1}{\sqrt{n_1}}(S_{1n}-n_1)} \mid n_1\right)\right] = \mathbb{E}_{n_1}\left[e^{\sqrt{n_1}t}\left(\frac{1}{1-\frac{t}{\sqrt{n_1}}}\right)^{n_1}\right].$$

By applying Taylor series expansion, we get

$$\log\left[e^{\sqrt{n_1}t}\left(\frac{1}{1-\frac{t}{\sqrt{n_1}}}\right)^{n_1}\right] = t^2/2 + o(n_1^{-1/2}).$$

With Lemma 3.2a, it follows that $\mathbb{E}[e^{\upsilon_{1n}t}] = \exp\left[t^2/2 + o_p\left(n^{-1/2}\right)\right]$. The result remains the same if $\delta_n = 1$. Thus, $\upsilon_{1n} \rightarrow^d N(0,1)$. Apply the same strategy and Lemma 3.2c, $\upsilon_{2n} \rightarrow^d N(0,1)$. Furthermore, given t_n, S_{1n} and S_{2n} are independent because they are functions of failure times corresponding to two independent failure modes. Additionally, by Lemma 3.2a, $t_n/(n/\mu_1)^{1/\beta_1} \rightarrow^{a.s.} 1$ as $n \rightarrow \infty$, implying that S_{1n} and S_{2n} are asymptotically independent unconditional on t_n. Hence, $(\upsilon_{1n}, \upsilon_{2n})' \rightarrow^d N_2(\mathbf{0}, \mathbf{I_2})$.

Let $V_{1n}^* = \sqrt{g(n)}(\log g(n))^{-1}(\hat{\mu}_2 - \mu_2)$ and $V_{2n}^* = \sqrt{g(n)}(\hat{\beta}_2 - \beta_2)$ with $g(n) = \mu_2(n/\mu_1)^{\beta_2/\beta_1}$. From (3.7), we have $\hat{\beta}_1 = n_1\beta_1/S_{1n}$ and $\hat{\beta}_2 = (n-n_1)\beta_2/S_{2n}$. Hence, with Lemma 3.2, it follows that

$$U_{2n} = -\beta_1\upsilon_{1n} + o_p(1) \text{ and } V_{2n}^* = -\beta_2\upsilon_{2n} + o_p(1). \tag{3.13}$$

Now, we consider the following,

$$\log\hat{\mu}_1 = (\log n_1)\left(1 - \frac{\hat{\beta}_1}{\beta_1}\right) - \frac{\hat{\beta}_1}{\beta_1}\log\left(\frac{\mu_1 t_n^{\beta_1}}{n_1}\right) + \frac{\hat{\beta}_1}{\beta_1}\log\mu_1.$$

With Lemma 3.2, it follows that $\mu_1 t_n^{\beta_1}/n_1 \rightarrow^{a.s.} 1$ as $n \rightarrow \infty$. Hence,

$$\sqrt{n}(\log n)^{-1}(\log \hat{\mu}_1 - \log \mu_1) = \upsilon_{1n} + o_p(1).$$

By Taylor series expansion, it implies that $U_{1n} = \mu_1 \upsilon_{1n} + o_p(1)$. Similarly, $V_{1n}^* = \mu_2 \upsilon_{2n} + o_p(1)$. It can be shown that

$$V_{1n} = \sqrt{\frac{\mu_1^{\beta_2/\beta_1}}{\mu_2}} \cdot \frac{\beta_2}{\beta_1} V_{1n}^* + o_p(1) \quad \text{and} \quad V_{2n} = \sqrt{\frac{\mu_1^{\beta_2/\beta_1}}{\mu_2}} V_{2n}^*. \tag{3.14}$$

The results follows. □

Theorem 3.6 implies that when $\beta_1 > \beta_2$, MLEs are consistent converging to true parameters with rates, $O_p(1/\sqrt{n})$ for $\hat{\beta}_1$, $O_p(\log n/\sqrt{n})$ for $\hat{\mu}_1$, $O_p(1/\sqrt{n^{\beta_2/\beta_1}})$ for $\hat{\mu}_2$, and $O_p(\log n/\sqrt{n^{\beta_2/\beta_1}})$ for $\hat{\beta}_2$.

Statistical Inference from Data with Partial Masking

Masked failures arise in the context of competing risks when, in order to save time and cost in pinpointing the real cause of failure, the entire subset of the components responsible for failure is replaced. We now consider the simple case where the masking probability is independent of the failure process. For a system with two failure modes, only complete masking is possible. For a system with more than two failure modes, one can encounter partial masking where a subset of the components could be identified to be the potential cause of failure. We will concentrate, however, on a 2-mode system. First, we let $\delta_i = 1$ for failure mode-1, $= 2$ for failure mode-2, and $= (*)$ when failure mode is masked.

$\underline{\beta_1 = \beta_2}$

We first consider the case where $\beta_1 = \beta_2 = \beta$. The likelihood function is

$$\mu_1^{n_1}\mu_2^{n_2}(\mu_1+\mu_2)^{n-n_1-n_2}\beta^n \prod_{i=1}^{n} t_i^{\beta-1} \exp[-(\mu_1+\mu_2)t_n^{\beta}] \times \gamma^{n-n_1-n_2}(1-\gamma)^{n_1+n_2}$$

where $n_1 = \sum_{i=1}^{n} 1\{\delta_i = 1\}, n_2 = \sum_{i=1}^{n} 1\{\delta_i = 2\}$ and $\gamma = Pr(\delta = (*))$ the probability of the data being masked. By maximizing the likelihood function above, the MLEs from masked data are

$$\hat{\mu}_1 = \left(\frac{n}{n_1+n_2}\right) n_1/t_n^{\hat{\beta}} \ , \ \hat{\mu}_2 = \left(\frac{n}{n_1+n_2}\right) n_2/t_n^{\hat{\beta}}, \tag{3.15}$$
$$\hat{\beta} = n/\sum_{i=1}^{n} \log(t_n/t_i) \text{ and } \hat{\gamma} = (n - n_1 - n_2)/n.$$

By comparing to the expressions of the MLEs in (3.2), we see that $\hat{\beta}$ does not change. However, $\hat{\mu}_1$ and $\hat{\mu}_2$ are different from the ones in (3.2) by a factor of $\left(\frac{n}{n_1+n_2}\right)$. Note that $n/(n_1 + n_2)$ is precisely the masking adjustment used to inflate estimators of μ_1, μ_2 based solely on the unmasked failures. Moreover,

$$(n_1, n_2, n - n_1 - n_2) \sim Multinomial\left(n, \frac{\mu_1}{\mu_1+\mu_2}(1-\gamma), \frac{\mu_2}{\mu_1+\mu_2}(1-\gamma), \gamma\right).$$

The inference results are similar to the unmasked case, and are not detailed here.

$\underline{\beta_1 \neq \beta_2}$

The likelihood function is

$$\begin{aligned} &L(Data|\mu_1, \mu_2, \beta_1, \beta_2) \\ &= \prod_{i=1}^{n} \left[\mu_1\beta_1 t_i^{\beta_1-1}\right]^{1\{\delta_i=1\}} \left[\mu_2\beta_2 t_i^{\beta_2-1}\right]^{1\{\delta_i=2\}} \left[\mu_1\beta_1 t_i^{\beta_1-1} + \mu_2\beta_2 t_i^{\beta_2-1}\right]^{1\{\delta_i=(*)\}} \\ &\quad \times \exp\left[-(\mu_1 t_n^{\beta_1} + \mu_2 t_n^{\beta_2})\right] \times \gamma^{n-n_1-n_2}(1-\gamma)^{n_1+n_2}. \end{aligned} \tag{3.16}$$

The MLEs based on (3.16) are not available in closed form. By considering $\delta_i = (*)$ as a missing value, we can use the E-M algorithm to estimate parameters. Note first that by (3.9), $\delta_i|t_i \sim Bernoulli\left(\frac{\mu_1\beta_1 t_i^{\beta_1-1}}{\mu_1\beta_1 t_i^{\beta_1-1}+\mu_2\beta_2 t_i^{\beta_2-1}}\right)$. Hence (E-Step), for $k \in \{\delta_k = (*)\}$, let $\delta_k^{(b)} = \frac{\mu_1^{(b)}\beta_1^{(b)} t_k^{\beta_1^{(b)}-1}}{\mu_1^{(b)}\beta_1^{(b)} t_k^{\beta_1^{(b)}-1}+\mu_2^{(b)}\beta_2^{(b)} t_k^{\beta_2^{(b)}-1}}$ and for $l \in \{\delta_l \neq (*)\}$, $\delta_l^{(b)} = 1\{\delta_l = 1\}$. The complete-data likelihood function is in the same form as in (3.7). However, it is important to note that $\delta_i^{(b)}$ now could be a fractional number. Then (M-Step), the

MLE's from the complete-data likelihood are

$$\mu_1^{(b+1)} = \sum_{i=1}^{n} \delta_i^{(b)} / t_n^{\beta_1^{(b+1)}}, \quad \beta_1^{(b+1)} = \sum_{i=1}^{n} \delta_i^{(b)} / \sum_{i=1}^{n} \log(t_n/t_i)\delta_i^{(b)},$$

$$\mu_2^{(b+1)} = \sum_{i=1}^{n} (1-\delta_i^{(b)}) / t_n^{\beta_2^{(b+1)}}, \quad \text{and} \quad \beta_2^{(b+1)} = \sum_{i=1}^{n} (1-\delta_i^{(b)}) / \sum_{i=1}^{n} \log(t_n/t_i)(1-\delta_i^{(b)}).$$

By picking the initial values, and repeating E-M Steps, the solution for MLEs is guaranteed to converge. By Theorem 1 of Dempster et al. (1977), the likelihood function (3.16) increases at each EM iteration. Thus, above algorithm converges to a local maximum. To guarantee the convergence to global maximum, the likelihood (3.16) needs to be a concave function of $(\mu_j, \beta_j)'s, j = 1, 2$. It can be shown that this condition is met when $n_j \geq 2$, for both $j = 1$ and 2.

Interestingly, the above results can be recast in a Bayesian framework with a non-informative choice of prior. For the equal shape parameter case, using the (ρ, μ, β) parameterization, the likelihood function is

$$L(Data|\rho, \mu, \beta) = \rho^{n_1}(1-\rho)^{n-n_1}(\mu\beta)^n \prod_{i=1}^{n} t_i^{\beta-1} \exp[-\mu t_n^{\beta}].$$

The conjugate prior for ρ is a $Beta(\alpha_1, \alpha_2)$ distribution. We choose the noninformative joint prior for (μ, β) in the form $\pi(\mu, \beta) \propto (\mu\beta^{1+\epsilon})^{-1}$ where ϵ is usually 1, but not necessarily. The motivation for such a prior choice can be found in the argument leading to Jeffrey's prior (see Bar-lev, Lavi, and Reiser 1992). This gives the following results:

$$\begin{aligned} \rho|Data &\sim Beta(n_1+\alpha_1, n-n_1+\alpha_2), \\ \beta|Data &\sim (\hat{\beta}/2n)\chi^2_{2(n-\epsilon)}, \text{ and} \\ \mu|\beta, Data &\sim (2t_n^{\beta})^{-1}\chi^2_{2n}. \end{aligned} \tag{3.17}$$

Note that the posterior inference is analogous to the frequentist analysis. In particular, with the choice $\epsilon = 1$, the posterior distribution of β is identical to the sampling distribution of $\hat{\beta}$.

For the case where $\beta_1 \neq \beta_2$, using the independence between failure modes, we can write the likelihood function as the product of the likelihood from each component such as

$$L(Data|\mu_1,\mu_2,\beta_1,\beta_2) = \underbrace{\prod_{i=1}^{n}[\mu_1\beta_1 t_i^{\beta_1-1}]^{\delta_i}\exp(-\mu_1 t_n^{\beta_1})}_{L_1(Data|\mu_1,\beta_1)} \times \underbrace{\prod_{i=1}^{n}[\mu_2\beta_2 t_i^{\beta_2-1}]^{1-\delta_i}\exp(-\mu_2 t_n^{\beta_2})}_{L_2(Data|\mu_2,\beta_2)}.$$

Now we can apply the Bayesian inference on each component separately. We can treat each component like a system with single failure mode. Hence, we can choose the noninformative joint prior for (μ_1,β_1) and (μ_2,β_2) to be

$$\pi_1(\mu_1,\beta_1) \propto (\mu_1\beta_1^{1+\epsilon_1})^{-1} \quad \text{and} \quad \pi_2(\mu_2,\beta_2) \propto (\mu_2\beta_2^{1+\epsilon_2})^{-1}$$

independently of each other. With these priors, we get the following,

$$\beta_1|Data \sim (\hat{\beta}_1/2n_1)\chi^2_{2(n_1-\epsilon_1)}, \quad \mu_1|\beta_1,Data \sim (2t_n^{\beta_1})^{-1}\chi^2_{2n_1}, \tag{3.18}$$
$$\beta_2|Data \sim (\hat{\beta}_2/2(n-n_1))\chi^2_{2(n-n_1-\epsilon_2)}, \quad \text{and} \quad \mu_2|\beta_2,Data \sim (2t_n^{\beta_2})^{-1}\chi^2_{2(n-n_1)}.$$

The posterior inference is straightforward and is based on distributions, in view of (3.18).

Goodness of Fit Tests

It is important to assess the fit of the PLP for the failures from each mode. This can be graphically checked by Duane plots (Duane 1964) for the failure times of each mode separately by plotting $\{\log t_i|\delta_i = j\ ,\ \log(N_j(t_i)/t_i)\}$ for $j = 1,2$. The linear forms of the plots indicate a good fit and the slopes are rough estimates for β_j's.

There are several tests that we can use. Most of these are using the fact that the transformation of the data would follow random variables from uniform or exponential distribution. Conditional on t_n, we can test the goodness of fit for each mode separately. The details of the test can be found in Section 4.6 and 4.7.4 in Rigdon and Basu (2000) which we will not go in details here. Here, we list several well known

tests; Lilliefors', Kuiper's V, Cramér-von Mises C^2, Watson's U^2, Anderson-Darling A^2, Shapiro-Wilk W, and Stepens' W^* tests.

Test equality of β_j's

To test the equality of the β_j's, with independence between two modes, we can consider the failure times from each mode are from different systems. Therefore, we can use the F-test discussed Section 5.5.1 in Rigdon and Basu (2000). Given δ_n and n_1, we have

$$\frac{2n_1\beta_1}{\hat{\beta}_1} \sim \chi^2_{2(n_1-\delta_n)} \text{ and } \frac{2(n-n_1)\beta_2}{\hat{\beta}_2} \sim \chi^2_{2(n-n_1-1+\delta_n)}.$$

Since both terms are independent, the statistic $F = \frac{\frac{2n_1\beta_1}{\hat{\beta}_1}/2(n_1-\delta_n)}{\frac{2(n-n_1)\beta_2}{\hat{\beta}_2}/2(n-n_1-1+\delta_n)}$. Under null hypothesis $H_0 : \beta_1 = \beta_2$,

$$F = \frac{n_1(n-n_1-1+\delta_n)}{(n-n_1)(n_1-\delta_n)}\frac{\hat{\beta}_2}{\hat{\beta}_1} \tag{3.19}$$

and $F \sim F\big(2(n_1-\delta_n), 2(n-n_1-1+\delta_n)\big)$. Therefore, for a size α test of $H_0 : \beta_1 = \beta_2$ against the two-sided alternative $H_a : \beta_1 \neq \beta_2$ is to reject H_0 if

$$F < F_{\alpha/2}\big(2(n_1-\delta_n), 2(n-n_1-1+\delta_n)\big) \text{ or } F > F_{1-\alpha/2}\big(2(n_1-\delta_n), 2(n-n_1-1+\delta_n)\big).$$

When there are more than two failure modes, one can test the equality of β_j's by a conditional χ^2-test, that is analogous to Bartlett's test of homogeneity of variance for independent normal distributions.

3.4 Statistical Inference for Multiple Repairable Systems

We now turn to the analysis for multiple repairable systems. To avoid confusion, we redefine the notation below.

Notation:

k : system index, $k = 1, \ldots, m$.

i : recurrent event index, $i = 1, 2, \ldots, n_k$.

j : failure mode index, $j = 1, \ldots, J$.

τ : censoring time.

δ : exact cause of failure (may be masked).

T_{ki} : time to the $i-$th event for the $k-$th system.

S_{ki} : Masking subset of causes for the $i-$th failure in the $k-$th system.

X_k : covariate associated with the $k-$th system.

The observed data are $\{S_{ki}, T_{ki}, X_k, \delta_{ki}\}_{k=1,\ldots,m;i=1\ldots,n_k}$. Note that $0 < T_{k1} < \ldots < T_{k,n_k} \leq \tau_k$ for all k. Suppose the failure process of each mode is governed by a NHPP with some intensity function $\lambda_{kj}(t; X_k)$. Then, the likelihood contribution of the $k-$th system is

$$\begin{aligned} L_k &= \prod_{i=1}^{n_k} \prod_{j=1}^{J} \left\{ \sum_{l \in S_{ki}} Pr(S_{ki} | T_{ki}, X_k, \delta_{ki} = l) \lambda_{kl}(T_{ki}, X_k) \right\}^{I(\delta_{ki}=j)} \\ &\quad \times \exp\left[-\sum_{j=1}^{J} \int_0^{\tau_k} \lambda_{kj}(t, X_k) dt \right]. \end{aligned}$$

Throughout this section, we assume the following:

- the censoring times for the $k-$th system, τ_k, is non-informative and independent of the failure process, and
- $\lambda_{kj}(t; X_k) = \lambda_{0j}(t) \exp(\gamma_j^T X_k)$ (proportional hazards).

Further, we shall drop the masking and covariate context for simplification and assume that all causes of failures are known exactly. Treatment of masking can be carried out under some additional assumption as in Dewanji and Sengupta (2003), or under a general Bayesian framework.

3.4.1 Parametric Estimators and Theoretical Results under PLP

Now we consider the parametric estimators, especially when the intensity function of each failure mode follows PLP such that

$$\lambda_j(t) = \mu_j \beta_j t^{\beta_j},\ j = 1, \ldots, J.$$

All systems are assumed independent and identically distributed. The observed data are T_{ki} and δ_{ki} where T_{ki} is the $i-$th failure time of the $k-$th system, and δ_{ki} is the failure mode indicator corresponding to T_{ki} for $k = 1, \ldots, m$ and $i = 1, \ldots, n_k$.

Under the assumption that the failure modes are independent, one can consider the (cumulative) failure times of each mode as the data from different independent systems. Then the known results for the multiple repairable systems under PLP can be applied here (see Rigdon and Basu 2000, Section 5).

For brevity, we focus our study on the case where all systems are observed starting from the same time and stopping at the same censoring time $\tau_k = T$ for all k. T could be T_n^* where T_n^* is the $n-$th superposition cumulative failure times from all m systems so that all m systems are being observed under failure censoring. Or, T could be a predetermined valued so that all m systems are being observed under time censoring. Both cases are studied and discussed here. Analogous to the single system case, we assume that there are only two failure modes. Therefore, $\delta_{ki} = 1$ if the failure of the $k-$th system at T_{ki} is caused by failure mode-1 and $=0$ otherwise. We first study the case when we assume that failure modes are independent. Later on, we would consider the case where the failure modes are dependent under the frailty framework.

Independent Failure Modes

Here, we assume that the failure modes are independent. Therefore, if $\lambda_j(t),\ j = 1, 2$ is the intensity function corresponding to each failure mode, then the likelihood

function from the data is

$$(3.20) \qquad L(Data|\lambda_k(\cdot), j = 1, 2) = \left[\prod_{k,i} \lambda_1(t_{ki})^{\delta_{ki}} \lambda_2(t_{ki})^{1-\delta_{ki}}\right] \times \exp\left[-m \int_0^T \lambda_1(t) + \lambda_2(t) dt\right].$$

Note that the likelihood function in (3.20) remains unchanged whether the systems is being observed under time censoring or failure censoring (by setting $T = T_n^*$). We would consider the case where the equality of β_1 and β_2 is assumed and when it is not.

Inference under Failure Censoring

In this case, m systems are observed until n total failures occurs. Since all systems are iid. We can superpose the cumulative failure times from m systems so we observe $t_1^* < t_2^* < \ldots < t_n^*$ and δ_i^* where t_i^* are the $i-$th superposed cumulative failure time from m systems and δ_i^* is the failure mode indicator corresponding to t_i^*. It follows that the intensity function of the superposition counting process is multiplied by factor of m. Therefore, we have the likelihood function below.

$$(3.21) \qquad L(Data|\lambda_k(\cdot), j = 1, 2) = \left[\prod_{k,i} \lambda_1(t_i^*)^{\delta_i^*} \lambda_2(t_i^*)^{1-\delta_i^*}\right] \times \exp\left[-m \int_0^{t_n^*} (\lambda_1(t) + \lambda_2(t))\, dt\right].$$

We can view this case as a single system with $\mu_j^* = m\mu_j$ and $\beta_j^* = \beta_j$. Hence, the results for a single systems in the previous chapter can be applied very easily. The asymptotic results from the previous chapter also hold here when we consider the number of systems, m, fixed and let the number of failures, n goes to infinity. Hence, in the case with equal shape parameters, the MLEs are consistent and converge to true parameter at rate $O_p(1/\sqrt{n})$ for $\hat{\beta}$ and $O_p(\log n/\sqrt{n})$ for $\hat{\mu}_1$, $\hat{\mu}_2$. In the case with unequal shape parameters with $\beta_1 > \beta_2$, the MLEs are consistent converge to the true parameters with rates, $O_p(1/\sqrt{n})$ for $\hat{\beta}_1$, $O_p(\log n/\sqrt{n})$ for $\hat{\mu}_1$, $O_p(1/\sqrt{n^{\beta_2/\beta_1}})$, for $\hat{\beta}_2$, and $O_p(\log n/\sqrt{n^{\beta 2/\beta_1}})$ for $\hat{\mu}_2$.

Inference under Time Censoring

We now come back to the case where the censoring time T is predetermined.

Case: Equal Shape Parameters

With $N = \sum_k n_k$ and $N_1 = \sum_{k,i} \delta_{ki}$, the likelihood function is

$$L(Data|\mu_1, \mu_2, \beta) = \mu_1^{N_1} \mu_2^{N-N_1} \beta^N \prod_{k,i} t_{ki}^{\beta-1} \times \exp\left[-m(\mu_1+\mu_2)T^\beta\right].$$

The corresponding MLEs are

$$(3.22)\ \hat{\mu}_1 = N_1/mT^{\hat{\beta}}, \quad \hat{\mu}_2 = (N-N_1)/mT^{\hat{\beta}}, \quad \text{and} \quad \hat{\beta} = N/\sum_{k,i} \log(T/t_{ki}).$$

With NHPP properties, it follows that

$$(3.23) \qquad n_k \sim Poisson\left((\mu_1+\mu_2)T^\beta\right),\ n_{kj} \sim Poisson(\mu_j T^\beta),$$

$$N \sim Poisson\left(m(\mu_1+\mu_2)T^\beta\right) \quad \text{and} \quad N_j \sim Poisson\left(m\mu_j T^\beta\right)$$

for $k = 1, \cdots, m$, $j = 1, 2$ where $n_{k1} = \sum_{i=1}^{n_k} \delta_{ki}$, $n_{k2} = n_k - n_{k1}$, and $N_2 = N - N_1$. The results in (3.1) and Theorem 3.1 are still applied here, so we have

$$\delta_{ki} \overset{iid}{\sim} Bernoulli\left(\frac{\mu_1}{\mu_1+\mu_2}\right)$$

which independent of t_{ki}. Furthermore, given N, it follows that

$$2N\beta/\hat{\beta}|N \sim \chi^2_{2N}.$$

One may interest in the limiting distributions of these estimators in (3.22) when we let $m \to \infty$. It is described in Theorem 3.7 below.

Theorem 3.7. *With estimators in (3.22), let* $X_{1m} = \sqrt{m}(\hat{\mu}_1 - \mu_1), X_{2m} = \sqrt{m}(\hat{\mu}_2 - \mu_2)$, *and* $X_{3m} = \sqrt{m}(\hat{\beta} - \beta)$. *Then,* $X_m = (X_{1m}, X_{2m}, X_{3m})' \to^d N_3(0, \Xi)$ *such that*

$$\Xi = \frac{1}{T^\beta} \begin{pmatrix} \frac{(\mu_1\beta\log T)^2}{\mu_1+\mu_2} + \mu_1 & \frac{\mu_1\mu_2(\beta\log T)^2}{\mu_1+\mu_2} & -\frac{\mu_1\beta^2\log T}{\mu_1+\mu_2} \\ \frac{\mu_1\mu_2(\beta\log T)^2}{\mu_1+\mu_2} & \frac{(\mu_2\beta\log T)^2}{\mu_1+\mu_2} + \mu_2 & -\frac{\mu_2\beta^2\log T}{\mu_1+\mu_2} \\ -\frac{\mu_1\beta^2\log T}{\mu_1+\mu_2} & -\frac{\mu_2\beta^2\log T}{\mu_1+\mu_2} & \frac{\beta^2}{\mu_1+\mu_2} \end{pmatrix}.$$

Proof. First, we Let $S_m = \beta \sum_{k,i} \log(T/t_{ki})$. Given N, we have $S_m \sim Gamma(N, 1)$. By (3.23), it follows that $N/m(\mu_1 + \mu_2)T^\beta \to^{a.s.} 1$ as $m \to \infty$. Hence, $N \to^{a.s.} \infty$ as $m \to \infty$. Furthermore if we let $\nu_m = (S_m - N)/\sqrt{N}$, then $\mathbb{E}[e^{\nu_m t}] = \exp\left[t^2/2 + o_p(m^{-1/2})\right]$ by a similar argument as in the proof of Theorem 3.6. Hence, $\nu_m \to^d N(0, 1)$ as $m \to \infty$. It is important to mention that ν_m is dependent on $N(= N_1 + N_2)$, however, $\lim_{m\to\infty} \nu_m$ is *not* dependent on N. Now consider the following,

$$\begin{aligned}(3.24)\qquad \sqrt{m}(\hat{\beta} - \beta) &= \frac{1}{\sqrt{(\mu_1 + \mu_2)T^\beta}} \sqrt{\frac{m(\mu_1 + \mu_2)T^\beta}{N}} \sqrt{N}(\hat{\beta} - \beta) \\ &= \frac{\beta}{\sqrt{(\mu_1 + \mu_2)T^\beta}} \left(\frac{N - S_m}{\sqrt{N}}\right) \times \frac{N}{S_m} \\ &= \frac{-\beta}{\sqrt{(\mu_1 + \mu_2)T^\beta}} \nu_m + o_p(1).\end{aligned}$$

Now let $u_{jm} = \sqrt{m\mu_j T^\beta} \log(N_j/m\mu_j T^\beta), j = 1, 2$, then by (3.23), the central limit theorem, and the delta method, $u_{jm} \to^d N(0, 1)$. Furthermore, u_{jm}'s are mutually independent since the two failure modes are assumed independent and also independent to $\lim_{m\to\infty} \nu_m$.

Now let consider

$$\begin{aligned}(3.25)\log(\hat{\mu}_1) &= \log(N_1/mT^{\hat{\beta}}) = \log(\mu_1) + \log(N_1/m\mu_1 T^\beta) + (\beta - \hat{\beta}) \log(T) \\ &= \log(\mu_1) + (\beta - \hat{\beta}) \log(T) + u_{1m}/\sqrt{m\mu_1 T^\beta}.\end{aligned}$$

It follows that

$$\sqrt{m}(\log \hat{\mu}_1 - \log \mu_1) = \frac{\beta \log T}{\sqrt{(\mu_1 + \mu_2)T^\beta}} \nu_m + \frac{u_{1m}}{\sqrt{\mu_1 T^\beta}} + o_p(1).$$

By Taylor expansion, we get

$$(3.26)\qquad \sqrt{m}(\hat{\mu}_1 - \mu_1) = \frac{\mu_1 \beta \log T}{\sqrt{(\mu_1 + \mu_2)T^\beta}} \nu_m + \sqrt{\frac{\mu_1}{T^\beta}} u_{1m} + o_p(1).$$

By similar argument we also get

$$\sqrt{m}(\hat{\mu}_2 - \mu_2) = \frac{\mu_1 \beta \log T}{\sqrt{(\mu_1+\mu_2)T^\beta}}\nu_m + \sqrt{\frac{\mu_2}{T^\beta}}u_{2m} + o_p(1). \tag{3.27}$$

By (3.24), (3.26) and (3.27), the results follow. □

Case: Unequal Shape Parameters

When the equality of β_1 and β_2 is not assumed, we can view the likelihood function as the product of the likelihood from failure mode-1 and -2,

$$\begin{aligned} &L(Data|\mu_1,\mu_2,\beta_1,\beta_2) \\ &\quad = \underbrace{\prod_{k,i}[\mu_1\beta_1 t_{ki}^{\beta_1-1}]^{\delta_{ki}}\exp(-m\mu_1 T^{\beta_1})}_{L_1(Data|\mu_1,\beta_1)} \times \underbrace{\prod_{k,i}[\mu_2\beta_2 t_{ki}^{\beta_2-1}]^{1-\delta_{ki}}\exp(-m\mu_2 T^{\beta_2})}_{L_2(Data|\mu_2,\beta_2)}. \end{aligned}$$

The MLEs are

$$\hat{\mu}_1 = N_1/mT^{\hat{\beta}_1}, \quad \hat{\beta}_1 = N_1/\textstyle\sum_{k,i}\log(T/t_{ki})\delta_{ki}, \tag{3.28}$$

$$\hat{\mu}_2 = (N-N_1)/mT^{\hat{\beta}_2}, \quad \text{and} \quad \hat{\beta}_2 = (N-N_1)/\textstyle\sum_{k,i}\log(T/t_{ki})(1-\delta_{ki}).$$

These estimators are the functions of failure times corresponding to their specific failure modes.

With NHPP properties, we get

$$n_k \sim Poisson\left((\mu_1 T^{\beta_1} + \mu_2 T^{\beta_2})\right), \quad n_{kj} \sim Poisson\left(\mu_j T^{\beta_j}\right), \tag{3.29}$$

$$N \sim Poisson\left(m(\mu_1 T^{\beta_1} + \mu_2 T^{\beta_2})\right) \quad \text{and} \quad N_j \sim Poisson\left(m\mu_j T^{\beta_j}\right) \quad \text{for } j = 1, 2.$$

And

$$\delta_{ki}|t_{ki} \sim Bernoulli\left(\frac{\mu_1\beta_1 t_{ki}^{\beta_1-1}}{\mu_1\beta_1 t_{ki}^{\beta_1-1} + \mu_2\beta_2 t_{ki}^{\beta_2-1}}\right)$$

which depends on t_{ki}. Moreover, we have

$$2N_1\beta_1/\hat{\beta}_1|N_1 \sim \chi^2_{2N_1}, \quad 2(N-N_1)\beta_2/\hat{\beta}_2|N-N_1 \sim \chi^2_{2(N-N_1)}$$

and

$$2N_1\beta_1/\hat{\beta}_1 + 2(N-N_1)\beta_2/\hat{\beta}_2|N \sim \chi^2_{2N}.$$

Below is Theorem 3.8 that describes the limiting distribution of estimators when we let $m \to \infty$.

Theorem 3.8. *With estimators in (3.28), let* $A_{1m} = \sqrt{m}(\hat{\mu}_1 - \mu_1), A_{2m} = \sqrt{m}(\hat{\beta}_1 - \beta_1), B_{1m} = \sqrt{m}(\hat{\mu}_2 - \mu_2)$ *and* $B_{2m} = \sqrt{m}(\hat{\beta}_2 - \beta_2)$. *If* $A_m = (A_{1m}, A_{2m})'$ *and* $B_m = (B_{1m}, B_{2m})'$, *then* $A_m \to^d A$ *and* $B_m \to^d B$ *as* $m \to \infty$ *with* A *and* B *are independent and* $A \sim N_2(\mathbf{0}, \Xi_1)$ *and* $B \sim N_2(\mathbf{0}, \Xi_2)$ *with*

$$\Xi_j = \frac{1}{T^{\beta_j}} \begin{pmatrix} \mu_j(\beta_j \log T)^2 + \mu_j & -\beta_j^2 \log T \\ -\beta_n^2 \log T & \frac{\beta_j^2}{\mu_j} \end{pmatrix} \text{ for } j = 1, 2.$$

Proof. Note that the MLEs for the parameters of the failure mode-j ($j = 1, 2$) are the function of t_{ki}'s corresponding to the mode-j. With assumption on independence of failure modes, the MLEs for the failure mode-1 are independent from the MLEs for the failure mode-2. Now we can apply the result from Crow (1974) for each failure mode to get the limiting distribution of A_m and B_m. This concludes the proof. □

3.4.2 Dependent Failure Modes Under A Frailty Framework

We now consider the case where the failure modes are dependent using a frailty framework. Let Z_k be the unobserved frailty variables. We consider the case where $Z_k \overset{iid}{\sim} Gamma(\eta^{-1}, \eta)$ so that Z_k has mean 1 and variance η. Under frailty framework, given $Z_k = z_k$, two failure modes from the $k-$th system are independent with the cumulative intensity function $z_k \mu_j t^{\beta_j}, j = 1, 2$. [Note that unconditionally, $N_j(t) \sim NegBin(r, p_j(t))$, $j = 1, 2$ and $N(t) \sim NegBin(r, p(t))$ with $r = \eta^{-1}, p_j(t) = (1 + \eta\mu_j t^{\beta_j})^{-1}, p(t) = (1 + \eta(\mu_1 t^{\beta_1} + \mu_2 t^{\beta_2}))^{-1}$. Moreover, $Pr(N_1(t) = x, N_2(t) = y) = \frac{\Gamma(1/\eta+x+y)(\mu_1 t^{\beta_1})^x(\mu_2 t^{\beta_2})^y}{\Gamma(1/\eta)\eta^{1/\eta}(1/\eta+\mu_1 t^{\beta_1}+\mu_2 t^{\beta_2})x!y!}$.]

Estimator for η when Z is Gamma-frailty for NHPPs

We now consider the case when $N_{kj}(t)$'s follows some NHPPs and the frailty variables are from gamma distribution with mean of 1 and variance of η. All m systems are observed until fixed time T. We starts by considering the conditional likelihood function from the data.

$$L_k(Data|Z_k = z_k) = z_k^{n_k} \prod_{i=1}^{n_k} \left[\lambda_1(t_{ki})^{\delta_{ki}} \lambda_2(t_{ki})^{1-\delta_{ki}}\right] \times \exp\left[-z_k \Lambda(T)\right].$$

By taking expectation on Z, we get the unconditional likelihood below,

$$L_k(Data) = \prod_{i=1}^{n_k} \left[\lambda_1(t_{ki})^{\delta_{ki}} \lambda_2(t_{ki})^{1-\delta_{ki}}\right] \times \frac{\Gamma(1/\eta + n_k)}{\Gamma(1/n)\eta^{1/\eta}[\Lambda(T) + 1/\eta]^{1/\eta+n_k}}.$$

Hence, the likelihood from m systems is

$$L(Data) = \prod_{k=1}^{m} \prod_{i=1}^{n_k} \left[\lambda_1(t_{ki})^{\delta_{ki}} \lambda_2(t_{ki})^{1-\delta_{ki}}\right] \times \frac{\prod_{k=1}^{m} \Gamma(1/\eta + n_k)}{\Gamma(1/n)^m \eta^{m/\eta}[\Lambda(T) + 1/\eta]^{m/\eta+N}}.$$

It can be shown that the MLE for $\Lambda(T)$ is $\hat{\Lambda}(T) = N/m$ where $N = \sum_{k=1}^{m} n_k$, then it follows that

$$\hat{\eta} = \arg \max_{\eta} \frac{\prod_{k=1}^{m} \Gamma(1/\eta + n_k)}{\Gamma(1/n)^m \eta^{m/\eta}[N/m + 1/\eta]^{m/\eta+N}}. \tag{3.30}$$

It is important to note that $\hat{\eta}$ does not depend on the form of $\Lambda(t)$ or $\Lambda_j(t)$'s. It can be shown the the variance of $\hat{\eta}$ is

$$\frac{1}{m}\eta^4 \left[\psi'(1/\eta) - \eta + \frac{1}{\Lambda(T) + 1/\eta} - \mathbb{E}\psi'(1/\eta + n_1)\right]^{-1} \tag{3.31}$$

where $\psi(\cdot)$ is digamma function such that $\psi(z) = \frac{d \log \Gamma(z)}{dz}$ and $\psi'(\cdot)$ is the derivation of digamma function. The proof of this is deferred to Section 3.9.

Inference under Time censoring

We now consider the case where the failures follows PLPs under time censoring. First let us consider the case where the equality of β_j's are assumed. The conditional likelihood function from the $k-$th system is

$$L_k(Data|Z_k = z_k) = z_k^{n_k} \mu^{n_{k1}} \mu_2^{n_k - n_{k1}} \beta^{n_k} \prod_{i=1}^{n_k} t_{ki}^{\beta-1} \exp[-z_k(\mu_1 + \mu_2)T^{\beta}].$$

With $Z_k \sim Gamma(\eta^{-1}, \eta)$, we get

$$L_k(Data) = \mu_1^{n_{k1}} \mu_2^{n_k - n_{k1}} \beta^{n_k} \prod_{k=1}^{n_1} t_{ki}^{\beta-1} \times \frac{\Gamma(1/\eta + n_k)}{\Gamma(1/\eta)\eta^{1/\eta}[(\mu_1+\mu_2)T^\beta + 1/\eta]^{1/\eta+n_k}}.$$

Hence the likelihood function contributed from all systems is

$$\begin{aligned} L(Data) &= \prod_{k=1}^{m} L_k(Data) = \mu_1^{N_1} \mu_2^{N-N_1} \beta^N \prod_{k,i} t_{ki}^{\beta-1} \\ &\quad \times \frac{\prod_{k=1}^m \Gamma(1/\eta + n_k)}{\Gamma(1/\eta)^m \eta^{m/\eta}[(\mu_1+\mu_2)T^\beta + 1/\eta]^{m/\eta+N}}. \end{aligned}$$

By maximizing the likelihood, we get MLEs,

$$\hat{\mu}_1 = N_1/mT^{\hat{\beta}},\ \hat{\mu}_2 = N_2/mT^{\hat{\beta}},\ \hat{\beta} = N/\sum_{k,i} \log(T/t_{ki}), \quad \text{and} \tag{3.32}$$

$$\hat{\eta} = \arg\ \max_{\eta} \frac{\prod_{k=1}^m \Gamma(1/\eta + n_k)}{\Gamma(1/\eta)^m \eta^{m/\eta}[N/m + 1/\eta]^{m/\eta+N}}.$$

Note that $\hat{\eta}$ above is the same $\hat{\eta}$ in (3.30) where it is not depend on the forms of $\lambda_j(\cdot)$.

Below is Lemma 3.3 describing the limiting distribution of N_1 and N_2. Note that the results from Lemma 3.3 are not restricted to PLP and distribution of frailty variables.

Lemma 3.3. *With the nonnegative frailty variables Z such that $\mathbb{E}(Z) = 1$ and $\mathbb{VAR}(Z) = \eta$. Furthermore, given $Z_k = z_k$, $n'_{kj}s$ are independent and $n_{kj}|z_k \sim Poisson(z_k\Lambda_j)$ for $0 < \Lambda_j < \infty$, $k = 1, \ldots, m$, $j = 1, 2$ with $N_j = \sum_{k=1}^m n_{kj}$. Then the following are true.*

a). $\sqrt{m}\begin{pmatrix} N_1/m - \Lambda_1 \\ N_2/m - \Lambda_2 \end{pmatrix} \to^d N_2\left(\mathbf{0}, \begin{bmatrix} \Lambda_1 + \eta\Lambda_1^2 & \eta\Lambda_1\Lambda_2 \\ \eta\Lambda_1\Lambda_2 & \Lambda_2 + \eta\Lambda_2^2 \end{bmatrix}\right)$, *as* $m \to \infty$.

b). For $N = N_1 + N_2$ *and* $\Lambda = \Lambda_1 + \Lambda_2$, $\sqrt{m}(N/m - \Lambda) \to^d N(0, \Lambda + \eta\Lambda^2)$ *as* $m \to \infty$.

Proof. a). It suffice to show that $\mathbb{E}(n_{kj}) = \Lambda_j$, $\mathbb{VAR}(n_{kj}) = \Lambda_j + \eta\Lambda_j^2$, and $\mathbb{COV}(n_{k1}, n_{k2}) = \eta\Lambda_1\Lambda_2$ for $k = 1, \ldots, m,\ j = 1, 2$. Then we can apply the central limit theorem to get the results. For $j = 1, 2$, we consider below,

$$\begin{aligned}
\mathbb{E}(n_{kj}) &= \mathbb{E}[\mathbb{E}(n_{kj}|z_k)] = \mathbb{E}(z_k\Lambda_j) = \Lambda_j, \\
\mathbb{VAR}(n_{kj}) &= \mathbb{E}[\mathbb{VAR}(n_{kj}|z_k)] + \mathbb{VAR}[\mathbb{E}(n_{kj}|z_k)] \\
&= \mathbb{E}(z_k\Lambda_j) + \mathbb{VAR}(z_k\Lambda_j) = \Lambda_j + \eta\Lambda_j^2, \text{ and} \\
\mathbb{COV}(n_{k1}, n_{k2}) &= \mathbb{E}[\mathbb{COV}(n_{k1}, n_{k2}|z_k)] + \mathbb{COV}[\mathbb{E}(n_{k1}|z_k), \mathbb{E}(n_{k2}|z_k)] \\
&= 0 + \mathbb{COV}(z_k\Lambda_1, z_k\Lambda_2) = \eta\Lambda_1\Lambda_2.
\end{aligned}$$

b). It is easy to see that $n_k|z_k \sim Poisson(z_k\Lambda)$. By applying a), we get b). □

Theorem 3.9 below describes the asymptotic distributions of estimators in (3.32). Similar to Lemma 3.3, the frailty variable Z in Theorem 3.9 is *not* restricted to gamma distribution.

Theorem 3.9. *With the nonnegative frailty variables Z such that $\mathbb{E}(Z) = 1$ and $\mathbb{VAR}(Z) = \eta$ and the estimators in (3.32), let $W_{1m} = \sqrt{m}(\hat{\mu}_1 - \mu_1)$, $W_{2m} = \sqrt{m}(\hat{\mu}_2 - \mu_2)$, $W_{3m} = \sqrt{m}(\hat{\beta} - \beta)$ and $W_m = (W_{1m}, W_{2m}, W_{3m})'$. Then $W_m \to^d N_3(\mathbf{0}, \Delta)$ as $m \to \infty$ where*

$$\Delta = \frac{1}{T^\beta}\begin{pmatrix}
\frac{(\mu_1\beta\log T)^2}{\mu_1+\mu_2} + \mu_1(1+\eta\mu_1 T^\beta) & \frac{\mu_1\mu_2(\beta\log T)^\beta}{\mu_1+\mu_2} + \eta\mu_1\mu_2 T^\beta & -\frac{\mu_1\beta^2\log T}{\mu_1+\mu_2} \\
\frac{\mu_1\mu_2(\beta\log T)^\beta}{\mu_1+\mu_2} + \eta\mu_1\mu_2 T^\beta & \frac{(\mu_2\beta\log T)^2}{\mu_1+\mu_2} + \mu_2(1+\eta\mu_2 T^\beta) & -\frac{\mu_2\beta^2\log T}{\mu_1+\mu_2} \\
-\frac{\mu_1\beta^2\log T}{\mu_1+\mu_2} & -\frac{\mu_2\beta^2\log T}{\mu_1+\mu_2} & \frac{\beta^2}{\mu_1+\mu_2}
\end{pmatrix}$$

Proof. The proof here is similar to the proof of Theorem 3.7 with some modification. First let $S_m = \beta\sum_{k,i}\log(T/t_{ki})$. Given N, z_k's, we have $S_m \sim Gamma(N, 1)$. By Lemma 3.3b with $\Lambda = (\mu_1 + \mu_2)T^\beta$, it follows that $N/m(\mu_1 + \mu_2)T^\beta \to^{a.s.} 1$ as

$m \to \infty$. Hence $N \to^{a.s.} \infty$ as $m \to \infty$. Furthermore if we let $\nu_m = (S_m - N)/\sqrt{N}$, then $\mathbb{E}[e^{\nu_m t}] = \exp[t^2/2 + o_p(m^{-1/2})]$ by a similar argument as in the proof of Theorem 3.6. Hence $\nu_m \to^d N(0,1)$ as $m \to \infty$. It is important to mention that though ν_m is dependent on N and z_k's, $\lim_{m\to\infty} \nu_m$ is *not* dependent on N and z_k's. Then similar to (3.24), we have

$$\sqrt{m}(\hat{\beta} - \beta) = \frac{-\beta}{\sqrt{(\mu_1 + \mu_2)T^\beta}}\nu_m + o_p(1). \tag{3.33}$$

Now let $u_{jm} = \sqrt{m}\log(N_j/m\mu_j T^\beta), j = 1, 2$ and $u_m = (u_{1m}, u_{2m})'$, then by using the result from Lemma 3.3a with $\Lambda_j = \mu_j T^\beta$ and applying the delta method, we get

$$u_m \to^d N_2\left(\mathbf{0}, \begin{bmatrix} \frac{1+\eta\mu_1 T^\beta}{\mu_1 T^\beta} & \eta \\ \eta & \frac{1+\eta\mu_2 T^\beta}{\mu_2 T^\beta} \end{bmatrix}\right), \text{ as } m \to \infty.$$

Note the u_m and ν_m are asymptotically independent. $\lim_{m\to\infty} \nu_m$.

Now similar to (3.26), we get

$$\log\hat{\mu}_j = \log\mu_j + (\beta - \hat{\beta})\log T + u_{jm}/\sqrt{m},\ j = 1, 2.$$

It follows that

$$\sqrt{m}(\log\hat{\mu}_j - \log\mu_j) = \frac{\beta\log T}{\sqrt{(\mu_1 + \mu_2)T^\beta}}\nu_m + u_{jm} + o_p(1),\ j = 1, 2.$$

And by Taylor series expansion, we have

$$\sqrt{m}(\hat{\mu}_j - \mu_j) = \frac{\mu_j\beta\log T}{\sqrt{(\mu_1 + \mu_2)T^\beta}}\nu_m + \mu_j u_{km} + o_p(1),\ j = 1, 2. \tag{3.34}$$

By (3.33) and (3.34), the results follow. □

Now without the equality assumption of β_j's, the likelihood function is

$$\begin{aligned} L(Data) &= (\mu_1\beta_1)^{N_1}(\mu_2\beta_2)^{N-N_1}\prod_{k,i} t_{ki}^{\beta_1\delta_{ki}+\beta_2(1-\delta_{ki})-1} \\ &\quad \times \frac{\prod_{k=1}^m \Gamma(1/\eta + n_k)}{\Gamma(1/\eta)^m \eta^{m/\eta}[\mu_1 T^{\beta_1} + \mu_2 T^{\beta_2} + 1/\eta]^{m/\eta+N}}. \end{aligned}$$

The MLEs are

$$\hat{\mu}_1 = N_1/mT^{\hat{\beta}_1},\ \hat{\beta}_1 = N_1/\sum_{k,i}\log(T/t_{ki})\delta_{ki},$$

$$\hat{\mu}_2 = (N-N_1)/mT^{\hat{\beta}_2},\ \hat{\beta}_2 = (N-N_1)/\sum_{k,i}\log(T/t_{ki})(1-\delta_{ki}).$$

and

$$\hat{\eta} = \arg\ \max_{\eta} \frac{\prod_{k=1}^{m}\Gamma(1/\eta+n_k)}{\Gamma(1/\eta)^m\eta^{m/\eta}[N/m+1/\eta]^{m/\eta+N}}. \tag{3.35}$$

We can see that regardless of the equality assumption of β_j's, $\hat{\eta}$ is the same and the MLEs for other parameters are exactly the same as when the independence among failure modes is assumed which is the result from (3.30). The form of $\hat{\eta}$ would depend on the choice of distribution of Z. Below is Theorem 3.10 describing the limiting distribution of MLEs in (3.35).

Theorem 3.10. *With the nonnegative frailty variables Z such that $\mathbb{E}(Z)=1$ and $\mathbb{VAR}(Z)=\eta$ and the estimators in (3.35), let $Q_{1m}=\sqrt{m}(\hat{\mu}_1-\mu_1),\ Q_{2m}=\sqrt{m}(\hat{\beta}_1-\beta_1),\ Q_{3m}=\sqrt{m}(\hat{\mu}_2-\mu_2),\ Q_{4m}=\sqrt{m}(\hat{\beta}_2-\beta_2)$ and $Q_m=(Q_{1m},Q_{2m},Q_{3m},Q_{4m})'$. Then $Q_m \to^d N_4(\mathbf{0},\Omega)$ where*

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12}' & \omega_{22} \end{pmatrix}$$

with

$$\omega_{11} = \frac{1}{T^{\beta_1}}\begin{pmatrix} \mu_1(\beta_1\log T)^2+\mu_1(1+\eta\mu_1T^{\beta_1}) & -\beta_1^2\log T \\ -\beta_1^2\log T & \beta_1^2/\mu_1 \end{pmatrix},$$

$$\omega_{22} = \frac{1}{T^{\beta_2}}\begin{pmatrix} \mu_2(\beta_2\log T)^2+\mu_2(1+\eta\mu_2T^{\beta_2}) & -\beta_2^2\log T \\ -\beta_2^2\log T & \beta_2^2/\mu_2 \end{pmatrix},$$

and

$$\omega_{12} = \begin{pmatrix} \mu_1\mu_2\eta & 0 \\ 0 & 0 \end{pmatrix}$$

Proof. Let First, let $S_{jm} = \beta_j \sum_{k,i} \log(T/t_{ki})$ for $j = 1, 2$. Given N_j and z_k's, we have $S_{jm} \sim Gamma(N_j, 1)$ with S_{1m} and S_{2m} are mutually independent. By Lemma 3.3a with $\Lambda_j = \mu_j T^{\beta_j}$, it follows that $N_j/m\mu_j T^{\beta_j} \rightarrow^{a.s.} 1$ as $m \rightarrow \infty$. Hence, $N_j \rightarrow^{a.s.} \infty$ as $m \rightarrow \infty$. Furthermore, if we let $\nu_{jm} = (S_{jm} - N_j)/\sqrt{N_j}$, then $\mathbb{E}[e^{\nu_{km}t}] = \exp[t^2/2 + o_p(m^{-1/2})]$ by a similar argument as in the proof of Theorem 3.6. Hence, $(\nu_{1m}, \nu_{2m})' \rightarrow^d N_2(\mathbf{0}, \mathbf{I_2})$ as $m \rightarrow \infty$. It is important to note that though ν_{1m} and ν_{2m} are dependent, they are asymptotically independent - - $\mathbb{COV}[\lim_{m\to\infty} \nu_{1m}, \lim_{m\to\infty} \nu_{2m}] = 0$. Moreover, $\lim_{m\to\infty} \nu_{1m}$ and $\lim_{m\to\infty} \nu_{2m}$ are *not* dependent on N_1 and N_2. By following a similar argument as in (3.33), we get

$$\sqrt{m}(\hat{\beta}_j - \beta_j) = \frac{-\beta_j}{\sqrt{\mu_j T^{\beta_j}}}\nu_{jm} + o_p(1),\ j = 1, 2. \tag{3.36}$$

Now let $u_{jm} = \sqrt{m}\log(N_j/m\mu_j T^{\beta_j})$ for $j = 1, 2$ and $u_m = (u_{1m}, u_{2m})'$, then by Lemma 3.3a and applying the delta method, we have

$$u_m \rightarrow^d N\left(\mathbf{0}, \begin{bmatrix} \frac{1+\eta\mu_1 T^{\beta_1}}{\mu_1} & \eta \\ \eta & \frac{1+\eta\mu_2 T^{\beta_2}}{\mu_2} \end{bmatrix}\right), \text{ as } m \rightarrow \infty.$$

Similar to argument in (3.34), we get

$$\sqrt{m}(\hat{\mu}_j - \mu_j) = \frac{\mu_j\beta_j \log T}{\sqrt{\mu_j T^{\beta_j}}}\nu_{jm} + \mu_j u_{jm} + o_p(1),\ j = 1, 2. \tag{3.37}$$

By (3.36) and (3.37), the results follow. □

3.5 Example: Vertical Boring Machine (Majumdar, 1993)

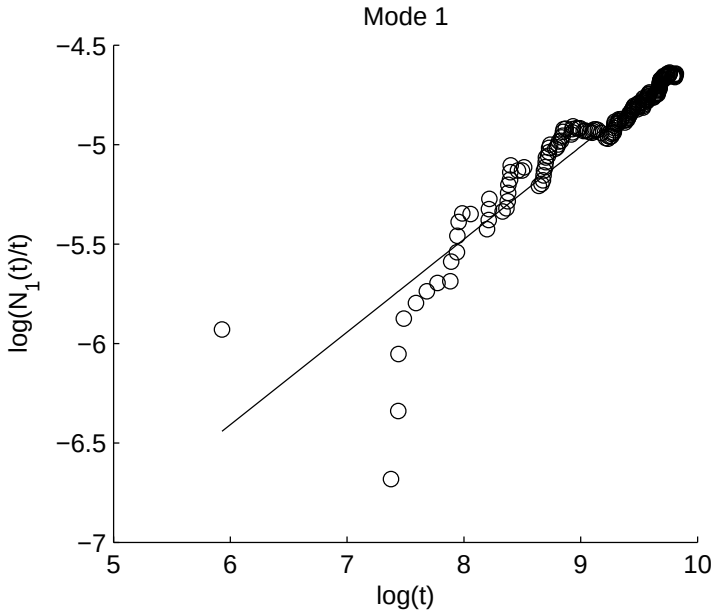
Majumdar (1993) studied recurrent failure times from a vertical boring machine that is subjected to multiple failure modes. The data consists of 262 recurrent failure times spanning a total of 18,285 hours. There are 11 failure times with masked failure causes. For simplicity of the exposition, are combine failure causes to present a 2-mode analysis. By combining groups of failure causes from the classification of Deshpande et al. (2000), we let failure mode-1 be failures from group 1,4,5 and 6 and mode-2 be the failures from groups 2 and 3. Moreover, 11 failure times with unknown caused are considered masked. Therefore, there 176, 75, and 11 failure times are from mode-1, mode-2, and masked, respectively.

Now that the failure modes have been defined, we start by checking if PLPs are good fits for the failure times for both failure modes. We first check the fits graphically by the Duane plots in Figure 3.1 below where the plots (a) and (b) are for mode-1 and mode-2 respectively. Although the plots do not exhibit strong linearity, a PLP assumption may not be too unreasonable, especially for mode-1 failures. This is further confirmed by a Cramér-von Mises test. Primarily, for the purposes of illustration, we shall assume PLP's to govern the underlying failure process for both modes.

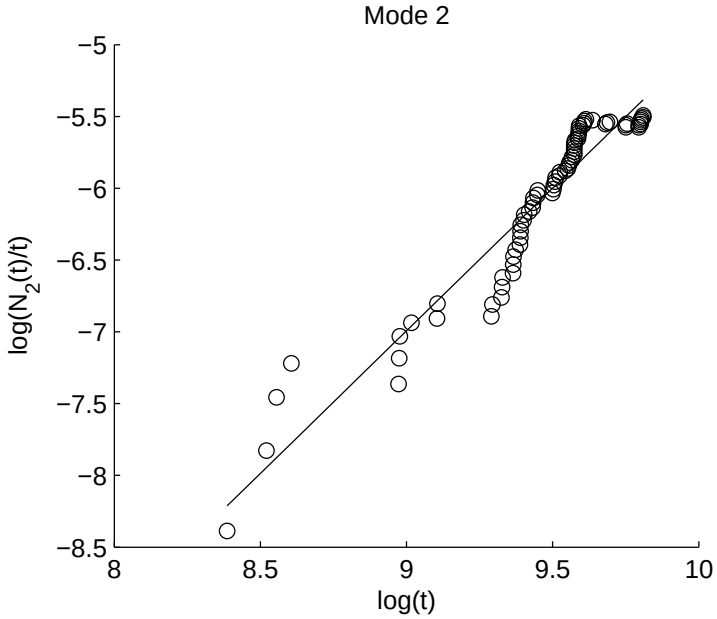
We test the equality of the shape parameters using the F-test in (3.19). Since δ_i's for masked failure times are unknown, we estimate them using $Pr(\delta_i = 1) = \frac{\hat{\mu}_1\hat{\beta}_2 t_i^{\hat{\beta}_1-1}}{\hat{\mu}_1\hat{\beta}_1 t_i^{\hat{\beta}_1-1}+\hat{\mu}_2\hat{\beta}_2 t_i^{\hat{\beta}_2-1}}$. So, we get $n_1 = 185.409$ for calculating F-statistics. The F-statistics is 2.132 which greater than the critical value 1.316 ($0.975-$th quantile) at significant level 0.05 from F-distribution with degrees of freedom 368.82 and 153.18. Therefore, we do *not* assume the equality of β_1 and β_2.

The MLEs are

$$\hat{\mu}_1 = 3.294 \times 10^{-4}, \quad \hat{\beta}_1 = 1.347, \quad \hat{\mu}_2 = 5.391 \times 10^{-11}, \text{ and } \hat{\beta}_2 = 2.856.$$



(a)



(b)

Figure 3.1: Duane plots for the recurrent failure times excluding masked data for failure mode-1 (a) and failure mode-2 (b).

By using the formula $\hat{\beta}_j \times \frac{1}{2n_j}[\chi^2_{2(n_1-\delta_n),.025}, \chi^2_{2(n_1-1+\delta_n),.975}], j = 1, 2$ with $n_2 = n - n_1$ and $\chi^2_{\nu,\alpha}$ is the $\alpha-$th quantile from χ^2-distribution with degree of freedom ν, the (conditional) exact 95% confidence interval for β_1 is $[1.1534, 1.5399]$, and for β_2 is $[2.2527, 3.5306]$. Again, n_1 is estimated by $\hat{\delta}_i$'s of masked data. Since we do not assume equality of β_1 and β_2, we cannot get the exact confidence intervals of μ_1 and μ_2. However, we can get their confidence intervals by using the asymptotic distribution. By using the formula $\hat{\mu_j} \times \exp([\pm z_{.975} * \log n_j/\sqrt{n_j}]), j = 1, 2$ with $z_{.975}$ is the $0.975-$th quantile from $N(0, 1)$, the 95% confidence interval for μ_1 is $[1.5531, 6.9846] \times 10^{-4}$, and for μ_2 is $[2.0402, 14.2432] \times 10^{-11}$. Note that we use n_j's instead of $g(n)$ and n in Theorem 3.6 (Note that $\hat{\beta}_2 > \hat{\beta}_1$ in this example.) since the sample size is small especially when the converging rate of $\hat{\mu}_j$'s are $O_p(\log n/\sqrt{n^{\beta_1/\beta_2}})$ and $O_p(\log n/\sqrt{n})$ respectively.

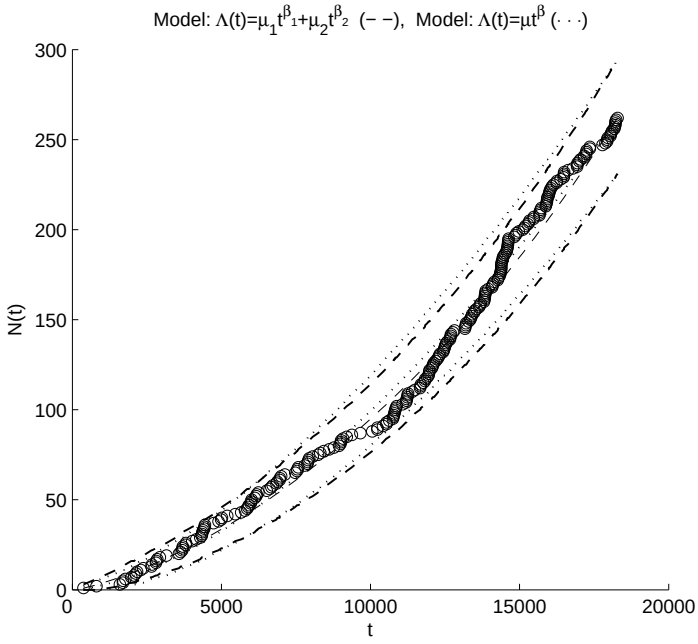


Figure 3.2: The plot of operating times against # of failures with their 95% confidence intervals under different models with single failure mode (- - -) and two failure modes ($\cdots$).

We also use the MLEs to get the 95% confidence intervals for the number of the system failures at time t as shown in Figure 3.2 above. This plot is under the assumption that $N(t) \sim Poisson(\Lambda(t))$ where we estimate $\Lambda(t)$ by using the

$\hat{\Lambda}(t) = \hat{\mu}_1 t^{\hat{\beta}_1} + \hat{\mu}_2 t^{\hat{\beta}_2}$. Figure 3.2 shows the intervals under two models; two failure modes and single failure mode [$\hat{\Lambda}(t) = \hat{\mu} t^{\hat{\beta}}$]. In general, the intervals from both models are similar. However, under the single failure mode model, the observed failure times at around 4000 and 11000 do *not* lie inside the confidence interval, while they *do* under two failure modes model. This shows that the model with two failure modes is doing a better job, even though not by much. We also need to keep in mind that it is expected that the model with two failure modes would do better since it has more parameters and it is the generalized form of the single failure mode model.

3.6 Simulation Results

In this section, the simulations for single system and multiple systems are studied. All the simulations here are based on the simulation size of 5000.

3.6.1 Single System

For single system, the simulation results are divided into two parts, namely, (a) where the equality of β_1 and β_2 is assumed and (b) where it is not.

Case: Equal Shape Parameters

In this part, we will study the behavior of log-MLEs instead of MLEs since asymptotically the bias and their variances and covariances should not depend on values of true parameters. It is confirmed from the simulations that this is true for β, however, it is not for μ_j's, $j = 1, 2$. This is due to the fact the $\hat{\mu}_j$'s ($\log \hat{\mu}_j$'s) converge at a slower rate of $\sqrt{n}/\log n$ than $\sqrt{n}$ of $\hat{\beta}$ ($\log \hat{\beta}$) . For example, for $n = 10000$, $\sqrt{n}/\log n = 10.86$ but $\sqrt{n} = 100$. More specifically, it is due to the exact distribution of $\log \hat{\mu}_j$ which can be written as

$$\log \hat{\mu}_1 =^d \log \mathcal{A}_{n,\rho} + \frac{2n}{\mathcal{B}_n}[\log \mathcal{C}_n - \log \mu]$$

Table 3.1: Biases and covariance matrices of log-MLEs from the simulations at different settings where only the size of $\mu_1 + \mu_2$ changes.

a). $(\mu_1, \mu_2, \beta)' = (.125, .125, 1)'$; $\mu_1 + \mu_2 = .25$ and $\mu_1/(\mu_1 + \mu_2) = 0.5$.

n	25	100	1000
Bias	$\begin{pmatrix} -0.42 \\ -0.41 \\ 0.06 \end{pmatrix}$	$\begin{pmatrix} -0.12 \\ -0.12 \\ 0.02 \end{pmatrix}$	$\begin{pmatrix} -0.02 \\ -0.02 \\ 0.00 \end{pmatrix}$
Covariance Matrix	$\begin{pmatrix} 2.85 & 2.67 & -1.63 \\ 2.67 & 2.91 & -1.65 \\ -1.63 & -1.65 & 1.04 \end{pmatrix}$	$\begin{pmatrix} 1.85 & 1.75 & -1.30 \\ 1.75 & 1.84 & -1.30 \\ -1.30 & -1.30 & 0.97 \end{pmatrix}$	$\begin{pmatrix} 1.46 & 1.42 & -1.18 \\ 1.42 & 1.58 & -1.18 \\ -1.18 & -1.18 & 0.99 \end{pmatrix}$

b). $(\mu_1, \mu_2, \beta)' = (.5, .5, 1)'$; $\mu_1 + \mu_2 = 1$ and $\mu_1/(\mu_1 + \mu_2) = 0.5$.

n	25	100	1000
Bias	$\begin{pmatrix} -0.29 \\ -0.28 \\ -0.06 \end{pmatrix}$	$\begin{pmatrix} -0.10 \\ -0.09 \\ 0.02 \end{pmatrix}$	$\begin{pmatrix} -0.02 \\ -0.02 \\ 0.00 \end{pmatrix}$
Covariance Matrix	$\begin{pmatrix} 1.59 & 1.37 & -1.19 \\ 1.37 & 1.58 & -1.19 \\ -1.19 & -1.19 & 1.07 \end{pmatrix}$	$\begin{pmatrix} 1.14 & 1.04 & -1.01 \\ 1.04 & 1.14 & -1.01 \\ -1.01 & -1.01 & 0.98 \end{pmatrix}$	$\begin{pmatrix} 1.08 & 1.04 & -1.03 \\ 1.04 & 1.08 & -1.04 \\ -1.03 & -1.04 & 1.03 \end{pmatrix}$

c). $(\mu_1, \mu_2, \beta)' = (2, 2, 1)'$; $\mu_1 + \mu_2 = 4$ and $\mu_1/(\mu_1 + \mu_2) = 0.5$.

n	25	100	1000
Bias	$\begin{pmatrix} -0.15 \\ -0.16 \\ 0.06 \end{pmatrix}$	$\begin{pmatrix} -0.07 \\ -0.07 \\ 0.02 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ -0.01 \\ 0.00 \end{pmatrix}$
Covariance Matrix	$\begin{pmatrix} 0.66 & 0.45 & -0.66 \\ 0.45 & 0.67 & -0.67 \\ -0.66 & -0.67 & 1.06 \end{pmatrix}$	$\begin{pmatrix} 0.63 & 0.53 & -0.74 \\ 0.53 & 0.63 & -0.74 \\ -0.74 & -0.74 & 1.03 \end{pmatrix}$	$\begin{pmatrix} 0.68 & 0.64 & -0.80 \\ 0.64 & 0.68 & -0.80 \\ -0.80 & -0.80 & 0.98 \end{pmatrix}$

and

$$\log \hat{\mu}_2 =^d \log \mathcal{A}_{n,1-\rho} + \frac{2n}{\mathcal{B}_n}[\log \mathcal{C}_n - \log \mu]$$

where $\mu = \mu_1 + \mu_2$, $\rho = \mu_1/\mu$, $\mathcal{A}_{n,\rho} \sim Bin(n, \rho)$, $\mathcal{B}_n \sim \chi^2_{2n-2}$, and $\mathcal{C}_n \sim Gamma(n, 1)$ with $\mathcal{A}_{n,*}$, $\mathcal{B}_n$, and $\mathcal{C}_n$ are mutually independent. This can be shown by letting $\mathcal{A}_{n,\rho} = n_1, \mathcal{A}_{n,1-\rho} = n_2, \mathcal{B}_n = 2n\beta/\hat{\beta}$, and $\mathcal{C}_n = \mu t_n^\beta$. We can see that the exact distributions of $\log \hat{\mu}_j$'s depend of the values of μ and ρ, however, they will no longer depend on μ and ρ as $n \to \infty$.

The results for different values of $\mu_1 + \mu_2$ and $\mu_1/(\mu_1 + \mu_2)$ are discussed here where we set $\beta = 1$ throughout. Note that according to Theorem 3.5 and the delta

Table 3.2: Biases and covariance matrices of log-MLEs from the simulations at different settings where only the size of $\mu_1/(\mu_1+\mu_2)$ changes.

a). $(\mu_1, \mu_2, \beta)' = (.65, .35, 1)'$; $\mu_1 + \mu_2 = 1$ and $\mu_1/(\mu_1+\mu_2) = 0.65$.

n	25	100	1000
Bias	$\begin{pmatrix} -0.28 \\ -0.31 \\ 0.06 \end{pmatrix}$	$\begin{pmatrix} -0.08 \\ -0.09 \\ 0.01 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ -0.01 \\ 0.00 \end{pmatrix}$
Covariance Matrix	$\begin{pmatrix} 1.57 & 1.41 & -1.22 \\ 1.41 & 1.73 & -1.22 \\ -1.22 & -1.22 & 1.10 \end{pmatrix}$	$\begin{pmatrix} 1.12 & 1.06 & -1.03 \\ 1.06 & 1.20 & -1.04 \\ -1.03 & -1.04 & 1.02 \end{pmatrix}$	$\begin{pmatrix} 1.02 & 1.00 & -1.00 \\ 1.00 & 1.06 & -1.00 \\ -1.00 & -1.00 & 1.00 \end{pmatrix}$

b). $(\mu_1, \mu_2, \beta)' = (.8, .2, 1)'$; $\mu_1 + \mu_2 = 1$ and $\mu_1/(\mu_1+\mu_2) = 0.8$.

n	25	100	1000
Bias	$\begin{pmatrix} -0.29 \\ -0.33 \\ 0.07 \end{pmatrix}$	$\begin{pmatrix} -0.10 \\ -0.12 \\ 0.02 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ -0.01 \\ 0.00 \end{pmatrix}$
Covariance Matrix	$\begin{pmatrix} 1.56 & 1.45 & -1.22 \\ 1.45 & 1.93 & -1.23 \\ -1.22 & -1.23 & 1.09 \end{pmatrix}$	$\begin{pmatrix} 1.14 & 1.08 & -1.04 \\ 1.08 & 1.34 & -1.04 \\ -1.04 & -1.04 & 1.01 \end{pmatrix}$	$\begin{pmatrix} 1.02 & 0.99 & -0.99 \\ 0.99 & 1.10 & -0.99 \\ -0.99 & -0.99 & 0.98 \end{pmatrix}$

c). $(\mu_1, \mu_2, \beta)' = (.95, .05, 1)'$; $\mu_1 + \mu_2 = 1$ and $\mu_1/(\mu_1+\mu_2) = 0.95$.

n	25	100	1000
Bias	$\begin{pmatrix} -0.31 \\ 0.39 \\ 0.06 \end{pmatrix}$	$\begin{pmatrix} -0.08 \\ -0.13 \\ 0.01 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ -0.02 \\ 0.00 \end{pmatrix}$
Covariance Matrix	$\begin{pmatrix} 1.43 & 1.41 & -1.66 \\ 1.41 & 1.60 & -1.17 \\ -1.66 & -1.17 & 1.06 \end{pmatrix}$	$\begin{pmatrix} 1.10 & 1.07 & -1.02 \\ 1.07 & 1.96 & -1.04 \\ -1.02 & -1.04 & 1.00 \end{pmatrix}$	$\begin{pmatrix} 1.02 & 1.02 & -0.99 \\ 1.02 & 1.47 & -1.02 \\ -0.99 & -1.02 & 0.99 \end{pmatrix}$

method,

$$\begin{bmatrix} \sqrt{n}(\log n)^{-1}(\log\hat{\mu}_1 - \log\mu_1) \\ \sqrt{n}(\log n)^{-1}(\log\hat{\mu}_2 - \log\mu_2) \\ \sqrt{n}(\log\hat{\beta} - \log\beta) \end{bmatrix} \rightarrow^d N\left(\mathbf{0}, \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}\right) \quad \text{as } n \rightarrow \infty.$$

Therefore, we will use the covariance matrix above for comparison of covariance matrices in Table 3.1 and 3.2. Furthermore, the biases in Table 3.1 and 3.2 are defined as $\log\hat{\mu}_1 - \log\mu_1$ and so on.

First we examine the case where the size of $\mu_1+\mu_2$ changes where β and $\mu_1/(\mu_1+\mu_2)$ are held fixed at values of 1 and 0.5 respectively. The summary of the simulation are shown in Table 3.1. Table 3.1 shows that the sizes of biases and the

variances/covariances corresponding to $\hat{\mu}_1$ ($\log \hat{\mu}_1$) and $\hat{\mu}_2$ ($\log \hat{\mu}_2$) are affected by the size of the $\mu_1 + \mu_2$ where they increase as the size $\mu_1 + \mu_2$ increases. However, it is not the case for $\hat{\beta}$ ($\log \hat{\beta}$). This effect does not diminish even at sample size of 1000. As mentioned before, this may be due to the rate $\sqrt{n}/\log n$ of $\hat{\mu}_j$'s ($\log \hat{\mu}_j$'s).

Now we examine the case where the ratio of μ_1 and μ_2 varies. WLOG, we let $\mu_1 > \mu_2$. Table 3.2 shows that as the ratio $\left(\frac{\mu_1}{\mu_1+\mu_2}\right)$ increases, it affects the sizes of biases and variances/covariances corresponding to $\hat{\mu}_2$ ($\log \hat{\mu}_2$), a little to $\hat{\mu}_1$ ($\log \hat{\mu}_1$) and virtually none to $\hat{\beta}$ ($\log \hat{\beta}$). For $\hat{\mu}_2$ ($\log \hat{\mu}_2$), the sizes of biases and variances/covariaces increase as the ratio increases. However, this effect seems to diminish at large sample size. This due to the fact that as the ratio increase, the number of failures from mode-2 decreases and that would affect the efficiency of $\hat{\mu}_2$ ($\log \hat{\mu}_2$).

Case: Unequal Shape Parameters

Now let use consider the case where the equality of β_1 and β_2 are *not* assumed. We also study the behaviors of $\log \hat{\mu}_j$'s and $\log \hat{\beta}_j$ for $j = 1, 2$ since they asymptotically do depend on the choices of true parameters. According to Theorem 3.6 and the delta method,

$$\begin{bmatrix} \sqrt{n}(\log n)^{-1}(\log \hat{\mu}_1 - \log \mu_1) \\ \sqrt{n}(\log \hat{\beta}_1 - \log \beta_1) \\ \sqrt{g(n)}(\log g(n))^{-1}(\log \hat{\mu}_2 - \log \mu_2) \\ \sqrt{g(n)}(\log \hat{\beta}_2 - \log \beta_2) \end{bmatrix} \rightarrow^d N\left(\mathbf{0}, \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\right),$$

as $n \rightarrow \infty$. We will use the covariance matrix above to compare covariance matrices from Table 3.3. Furthermore, the biases in Table 3.3 are defined as $\log \hat{\mu}_1 - \log \mu_1$ and so on.

First let us look at Table 3.3a. With equal μ_j's but $\beta_1 = 2 > 1 = \beta_2$, the sizes of biases and variances/covariances from mode-2 are larger than those from mode-1. But, they all converge to asymptotic values as sample size increases. In Table 3.3b,

Table 3.3: Biases and covariance matrices of log-MLEs from the simulations at different settings where the equality of β_1 and β_2 are *not* assumed.

a). $(\mu_1, \beta_1, \mu_2, \beta_2)' = (1, 2, 1, 1)'$.

n	25	100	1000
$\bar{n}_1/n$	0.814	0.905	0.969
Bias	$\begin{pmatrix} -0.30 \\ 0.07 \\ -0.78 \\ 0.16 \end{pmatrix}$	$\begin{pmatrix} -0.10 \\ 0.02 \\ -0.38 \\ 0.07 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ 0.00 \\ -0.15 \\ 0.02 \end{pmatrix}$
Cov. Matrix	$\begin{pmatrix} 1.70 & -1.41 & -0.13 & 0.05 \\ -1.41 & 1.34 & 0.15 & -0.04 \\ -0.13 & 0.15 & 26.18 & -4.16 \\ 0.05 & -0.04 & -4.16 & 1.71 \end{pmatrix}$	$\begin{pmatrix} 1.21 & -1.13 & 0.02 & -0.01 \\ -1.13 & 1.12 & -0.02 & 0.02 \\ 0.02 & -0.02 & 2.49 & -1.57 \\ -0.01 & 0.02 & -1.57 & 1.26 \end{pmatrix}$	$\begin{pmatrix} 1.05 & -1.03 & -0.02 & 0.01 \\ -1.03 & 1.03 & 0.02 & -0.01 \\ -0.02 & 0.02 & 1.36 & -1.15 \\ 0.01 & -0.01 & -1.15 & 1.08 \end{pmatrix}$

b). $(\mu_1, \beta_1, \mu_2, \beta_2)' = (1, 4, 1, 1)'$.

n	25	100	1000
$\bar{n}_1/n$	0.881	0.963	0.994
Bias	$\begin{pmatrix} -0.28 \\ 0.07 \\ -0.29 \\ 0.23 \end{pmatrix}$	$\begin{pmatrix} -0.09 \\ 0.02 \\ -0.55 \\ 0.18 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ 0.00 \\ -0.52 \\ 0.10 \end{pmatrix}$
Cov. Matrix	$\begin{pmatrix} 1.58 & -1.30 & -0.03 & 0.04 \\ -1.30 & 1.21 & 0.05 & -0.03 \\ -0.03 & 0.05 & 11.78 & -2.64 \\ 0.04 & -0.03 & -2.64 & 1.09 \end{pmatrix}$	$\begin{pmatrix} 1.14 & -1.07 & -0.01 & 0.00 \\ -1.07 & 1.06 & 0.00 & 0.00 \\ -0.01 & 0.00 & 12.77 & -2.91 \\ 0.00 & 0.00 & -2.91 & 1.24 \end{pmatrix}$	$\begin{pmatrix} 1.02 & -0.99 & -0.04 & 0.01 \\ -0.99 & 0.99 & 0.04 & -0.01 \\ -0.04 & 0.04 & 4.83 & -2.04 \\ 0.01 & -0.01 & -2.04 & 1.28 \end{pmatrix}$

c). $(\mu_1, \beta_1, \mu_2, \beta_2)' = (4, 2, .25, 1)'$.

n	25	100	1000
$\bar{n}_1/n$	0.912	0.975	0.996
Bias	$\begin{pmatrix} -0.18 \\ 0.07 \\ 0.22 \\ 0.30 \end{pmatrix}$	$\begin{pmatrix} -0.06 \\ 0.01 \\ -0.58 \\ 0.23 \end{pmatrix}$	$\begin{pmatrix} -0.01 \\ 0.00 \\ -1.09 \\ 0.14 \end{pmatrix}$
Cov. Matrix	$\begin{pmatrix} 0.57 & -0.70 & -0.09 & 0.00 \\ -0.70 & 1.16 & 0.00 & 0.00 \\ -0.09 & 0.00 & 181.72 & 3.41 \\ 0.00 & 0.00 & 3.41 & 0.41 \end{pmatrix}$	$\begin{pmatrix} 0.56 & -0.72 & 0.37 & -0.01 \\ -0.72 & 1.01 & -0.43 & 0.01 \\ 0.37 & -0.43 & 453.50 & -12.78 \\ -0.01 & 0.01 & -12.78 & 0.67 \end{pmatrix}$	$\begin{pmatrix} 0.68 & -0.83 & -0.04 & 0.00 \\ -0.83 & 1.03 & 0.03 & 0.01 \\ -0.04 & 0.03 & 27.31 & -4.77 \\ 0.00 & 0.01 & -4.77 & 1.25 \end{pmatrix}$

we increase β_1 to 4, the sizes of biases and variances/covariances from mode-2 are larger than those from mode-1. However, the biases of $\hat{\mu}_2$ ($\log \hat{\mu}_2$) at sample size 100 and 1000 are larger than that at sample size of 25. By observing the values $\bar{n}_1/n$, we can see that the proportion failure from mode-2 drops as sample size increases. With converging rate of $\sqrt{g(n)}/\log g(n)$ (see Theorem 3.8), the simulation results for $\hat{\mu}_2$ ($\log \hat{\mu}_2$) can be very unstable even at sample size of 1000. In Table 3.3c, β_j's are as same as in Table 3.3a, however, $\mu_1 = 4 > .25 = \mu_2$. The results of Table 3.3c compared to Table 3.3a are similar to those of Table 3.3b to Table 3.3a. This implies that both ratio of β_j's and of μ_j's do affect of efficiency of estimators especially for the failure mode with smaller proportion of failures.

3.6.2 Multiple Systems

The simulation for multiple systems are also studied but the details are omitted. However, we summarize the result here. The simulation results are consistent with Theorem 3.7-3.10 even at small m if the time censoring T is at moderate size. As long as, there are moderated number of failures fall into each failure mode, the estimators seems to do very well. Under frailty dependence, the size of frailty variance does have the negative impact the to the efficiency of the estimators. Moreover, the choice of frailty distribution also affects the efficiency of estimator especially at small sample.

3.7 Summary

We have laid down a general framework for repairable systems under competing risks. Under the assumption that the number of failures from each failure mode follows some general counting process, some properties and necessary condition are studied. These properties are studied so that it could be applied directly but not limited to nonhomogeneous Poisson process (NHPP).

In the latter part of this chapter, we have studied the inferences for a single repairable system and multiple repairable systems under PLP. For a single system, we restrict our study to data from the system under (system) failure censoring. When the shape parameters are equal, some results can be derived from known results of a single-failure-mode system. For the case with unequal shape parameters, as we let n go to infinity, the system failures are dominated by the failures from the mode with the bigger shape parameter. However, the number of failures from the smaller shape parameter still goes to infinity but with slower rate. We have also studied the case where failure modes of the data are partially masked. The problem can be easily handled if it is viewed under bayesian framework.

With data from multiple repairable systems, we focus our study mainly under time censoring scenario where the systems are being observed until a fixed time T. We have mentioned that under the failure censoring, the results from a single repairable systems can be applied. The study are mainly divided into two cases when the independence among failure modes are assumed and it is not. For the case with the independence assumed, the MLEs converges in distribution to the true parameters with rate of $\sqrt{m}$ since all systems are *iid*. Likewise, when the dependence among failure modes assumed, the MLEs also converges weakly with rate of $\sqrt{m}$, however, the covariance matrix has the frailty parameter involved in it especially the terms relating to the scale parameter μ_1 and μ_2. The estimators and their variances of shape parameters β_1 and β_2 are not affected by frailty parameter η.

3.8 Future Research

In this section, we discuss the possible future research for the inferences for repairable systems under competing risks. It is also of interest to study some properties when the recurrent failures follows some counting processes under different assump-

tions not included in this study. The assumptions in this study for the general counting processes are made so that it could be applied to NHPP and PLP framework easily. One may want to study the case where the failures follow different kinds of counting process. The different parametric forms of the cumulative intensive functions are also at interest such as when $\Lambda(t) = \mu[1 - \exp(\beta t)]$. For the case when the failure mode are dependent, one may be interested to study the inferences under the dependent that does not follow frailty framework as it does in our study.

3.9 Technical Results and Proofs

In this section, we summary some technical results needed for some results stated in the earlier sections. We start by introducing regularly and slowly varying function. (for formal definitions, see Feller 1996).

Definition 3.1. *A nonnegative function L defined on $[0, \infty]$ is a regularly varying function with exponent $-\infty < \rho < \infty$ iff*

$$\frac{L(xt)}{L(t)} \to x^{\rho} \text{ as } t \to \infty.$$

And L is a slowly varying function iff L is a regularly varying function with exponent 0 such that

$$\frac{L(xt)}{L(t)} \to 1 \text{ as } t \to \infty.$$

Examples for regularly, slowly, and non-regularly functions are power function (t^{ρ}), logarithm function ($\log(t)$) and exponential function ($\exp(t)$), respectively. Note that logarithm function is also regularly varying function, however, power function is not slowly varying function. Now suppose L is a nonnegative differentiable monotone increasing function, and L^{-1} exists as a nonnegative differentiable monotone increasing function. We have Lemma 3.4 below which is needed for the proof of Theorem 3.4 later on.

Lemma 3.4. *Suppose L is a nonnegative differentiable monotone increasing function on $[0,\infty]$ such that*

$$\frac{L(xt)}{L(t)} \to x^{\rho} \text{ as } t \to \infty,$$

then

$$\frac{L^{-1}(ys)}{L^{-1}(s)} \to y^{1/\rho} \text{ as } s \to \infty, \text{ for } 0 \le \rho \le \infty.$$

Proof. Since L is a nonnegative differentiable monotone increasing function on $[0,\infty]$, L^{-1} exists and also is a nonnegative differentiable monotone increasing function on $[0,\infty]$. By L'Hospital rule,

$$\lim_{t\to\infty} \frac{L(xt)}{L(t)} = \frac{\lim_{t\to\infty} \frac{d}{dt}[L(xt)]}{\lim_{t\to\infty} L'(t)} = \frac{x \lim_{t\to\infty} L'(xt)}{\lim_{t\to\infty} L'(t)} \equiv x^{\rho}$$

implying that

$$h(x) \equiv \frac{\lim_{t\to\infty} L'(xt)}{\lim_{t\to\infty} L'(t)} = x^{\rho-1}$$

where $L'(t) = \frac{d}{dt}L(t)$. Note that $\frac{d}{ds}[L^{-1}(s)] = \frac{1}{L'(L^{-1}(s))}$. Let $Z = \lim_{s\to\infty} \frac{L^{-1}(ys)}{L(s)}$. Now suppose Z exists, consider below,

$$Z = \frac{\lim_{s\to\infty} \frac{d}{ds}[L^{-1}(ys)]}{\lim_{s\to\infty} \frac{d}{ds}[L^{-1}(s)]} = \frac{\lim_{s\to\infty} \frac{y}{L'(L^{-1}(ys))}}{\lim_{s\to\infty} \frac{1}{L'(L^{-1}(s))}} = y \times \left[\frac{\lim_{s\to\infty} L'\left(\frac{L^{-1}(ys)}{L^{-1}(s)} \times L^{-1}(s)\right)}{\lim_{s\to\infty} L'(L^{-1}(s))}\right]^{-1}.$$

Since $L^{-1}(s) \to \infty$ as $s \to \infty$, the above equation can be rewritten as,

$$Z = y \times [h(Z)]^{-1} = y \times Z^{1-\rho}$$

Hence, $Z = y^{1/\rho}$. This completes the proof. □

Lemma 3.4 can be applied to a nonnegative differentiable monotone increasing function, as shown in Corollary 3.1 below.

Corollary 3.1. *Suppose L is a nonnegative differentiable monotone increasing function.*

a). If L is a regularly varying function with exponent $\rho \neq 0$ then L^{-1} is also a regularly varying function with exponent $1/\rho$.

b). L is a slowly varying function iff L^{-1} is not regularly varying function.

Proof. The proof simply follows from the results from Lemma 3.4. □

Now we are ready to state the theorem needed for the proof of Theorem 3.4

Theorem 3.11. *Suppose X_n and Y_n are nonnegative random variables nondecreasing in n such that $X_n \to^{a.s.} \infty$ and $Y_n \to^{a.s.} \infty$, as $n \to \infty$, and f, a nonnegative differentiable monotone increasing function, regularly varying with exponent $0 \leq \rho < \infty$ such that $\lim_{t\to\infty} \frac{f(xt)}{f(t)} = x^\rho$.*

a). If $X_n/Y_n \to^{a.s.} 1$ as $n \to \infty$, then $f(X_n)/f(Y_n) \to^{a.s.} 1$, as $n \to \infty$.

b). Suppose $\rho \neq 0$, $X_n/Y_n \to^{a.s.} 1$ iff $f(X_n)/f(Y_n) \to^{a.s.} 1$ as $n \to \infty$.

Proof.

a). With $\frac{X_n}{Y_n} \to^{a.s.} 1$ as $n \to \infty$, we have

$$\frac{f(X_n)}{f(Y_n)} = \frac{f\left(\frac{X_n}{Y_n} \times Y_n\right)}{f(Y_n)} \to \left(\frac{X_n}{Y_n}\right)^\rho \to^{a.s.} 1 \text{ as } n \to \infty.$$

b). Now suppose ρ is nonzero . Then by Corollary 3.1a, f^{-1} is also regularly varying function. Now, we can write $\frac{X_n}{Y_n} = \frac{f^{-1}(f(X_n))}{f^{-1}(f(Y_n))}$, then applying the result from a) to complete the proof. □

With Theorem 3.11 now proved, we state here the proof of Theorem 3.4.

3.9.1 Proof of Theorem 3.4

Proof. a). Since $\lim_{t\to\infty}\frac{\Lambda_1(t)}{\Lambda(t)} \to 1$, $T_n \to \infty$ and $n_j/\Lambda_j(T_n) \to^{a.s.} 1$ as $n \to \infty$ for $j = 1, 2$, we have $\frac{n_1}{n} = \frac{N_1(T_n)}{\Lambda_1(T_n)} \times \frac{\Lambda_1(T_n)}{\Lambda(T_n)} \times \frac{\Lambda(T_n)}{n} \to^{a.s.} 1$ as $n \to \infty$.

b). Suppose $\Lambda_2 \circ \Lambda_1^{-1}$ is a regularly varying function with $0 \leq \rho < \infty$. Under Condition 3.1, it follows that $\Lambda_2 \circ \Lambda_1^{-1}$ is a nonnegative differential monotone increasing function. Therefore, by applying Theorem 3.11a and result from a), we have $\frac{n_2}{\Lambda_2\circ\Lambda_1^{-1}(n)} = \frac{N_2(T_n)}{\Lambda_2(T_n)} \times \frac{\Lambda_2\circ\Lambda_1^{-1}(\Lambda_1(T_n))}{\Lambda_2\circ\Lambda_1^{-1}(N_1(T_n))} \times \frac{\Lambda_2\circ\Lambda_1^{-1}(n_1)}{\Lambda_2\circ\Lambda_1^{-1}(n)} \to^{a.s.} 1$ as $n \to \infty$. This completes the proof. □

Theorem 3.4 shows that n_1 goes to infinity with rate n, however, n_2 goes to infinity at rate slower than n where the rate for n_2 can only be specified when $\Lambda_2 \circ \Lambda_1^{-1}$ is regularly varying function. In the case where $\Lambda_2 \circ \Lambda^{-1}$ is non-regularly varying function, we only be able to say that n_2 will go to infinity at some rate less than n.

3.9.2 Proof of (3.31)

Proof. For convenient, we let $\theta = 1/\eta$. Hence, we have

$$l_k \equiv \log L_k(Data) = \log\Gamma(\theta+n_k) - \log\Gamma(\theta) + \theta\log(\theta) - (\theta+n_k)\log(\Lambda(T)+\theta) + \text{Constant}.$$

Then,

$$\begin{aligned}\frac{\partial l_k}{\partial\theta} &= \psi(\theta+n_k) - \psi(\theta) + 1 + \log(\theta) - \left[\log(\Lambda(T)+\theta) + \frac{\theta+n_k}{\Lambda(T)+\theta}\right] \\ &= \psi(\theta+n_k) - \psi(\theta) + 1 + \log(\theta) - \left[\log(\Lambda(T)+\theta) + 1 + \frac{n_k-\Lambda(T)}{\Lambda(T)+\theta}\right].\end{aligned}$$

And,

$$\frac{\partial^2 l_k}{\partial\theta^2} = \psi'(\theta+n_k) - \psi'(\theta) + \frac{1}{\theta} - \left[\frac{1}{\Lambda(T)+\theta} - \frac{n_k-\Lambda(T)}{(\Lambda(T)+\theta)^2}\right]$$

Hence, the Fisher information for the data from the $k-$th system is

$$-\mathbb{E}\left[\frac{\partial^2 l_k}{\partial\theta^2}\right] = \psi'(\theta) - \frac{1}{\theta} + \frac{1}{\Lambda(T)+\theta} - \mathbb{E}[\psi'(\theta + n_k)]$$

since $\mathbb{E}[n_k] = \Lambda(T)$. Hence, the variance of $\hat{\theta} = 1/\hat{\eta}$ is

$$\frac{1}{m}\left[\psi'(1/\eta) - \eta + \frac{1}{\Lambda(T)+1/\eta} - \mathbb{E}[\psi'(1/\eta + n_k)]\right]^{-1}.$$

By applying the delta method with $\eta = 1/\theta$, the result follows. □

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Ali, M. M., and Chan, L. K. (1964), "On Gupta's Estimates of the Parameters of the Normal Distribution," *Biometrika*, 51, 498-501.

[2] Aly, E. A. A., Csörgő, M., and Horváth, L. (1985), "Strong approximations of the quantile process of the product-limit estimator (Corr: V19 p366)," *Journal of Multivariate Analysis*, 16 , 185-210.

[3] Ascher, H. and Feingold, H. (1984), *Repairable Systems Reliability*, New York: Marcel Dekker.

[4] Bain, L. J. (1978), *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods*, Volume 24, New York: Marcel Dekker.

[5] Bar-Lev, S. K, Lavi, I. and Reiser, B. (1992), "Bayesian inference for the power law process," *Annals of the Institute of Statistical Mathematics*, 44, 623-639.

[6] Bhattacharyya, G. K. and Ghosh, J. K. (1991), "Asymptotic properties of estimators in a binomial reliability growth model and its continuous-time analog," *Journal of Statistical Planning and Inference*, 29, 43-53.

[7] Blom, G. (1958), *Statistical Estimates and Transformed Beta Variables*, New York: John Wiley.

[8] Breslow, N., and Crowley, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimate Under Random Censorship," *Annals of Statistics*, 2, 437-453.

[9] Burr, D. (1994), "A comparison of certain bootstrap confidence intervals in the Cox model," *Journal of the American Statistical Association*, 89 , 1290-1302.

[10] Chernoff, H., Gastwirth, J. L., and Johns, M. V. (1967), "Asymptotic Distribution of Linear Combinations of Fuctions of Order Statistics With Applications to Estimation," *Annals of Mathematical Statistics*, 38, 52-72.

[11] Crow, L. H. (1974), "Reliability analysis for complex, repairable systems," *Reliability and Biometry*, F. Proschan and R.J. Serfling, eds., SIAM, Philadelphia, 379- 410.

[12] Crowder, M. (2001), *Classical Competing Risks*, New York: Chapman & Hall/CRC.

[13] Cummings, F. J., Gray, R., Davis, T. E., Tormey, D. C., Harris, J. E., Falkson, G. G., and Arsenequ, J. (1986), "Tamoxifen versus placebo: double blind adjuvant trial in elderly woman with stage II breast cancer," *National Cancer Institute Monographs*, 1, 119-23.

[14] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1-38.

[15] Deshpande, J. V., Mukhopadhyay, M. and naik-Nimbalkar, U. V. (2000), "Tests for equality of tntensities of failures of a repairable system under two competing risks," in *Recent Advances in Reliability Theory: Methodology, Practice and Inference (Statistics for Industry and Technology)*, eds. by N. Limnios and M. Nikulin , Birkhäuser, pp. 391-404.

[16] Dewanji, A. and Sengupta, D. (2003), "Estimation of competing risks with general missing pattern in failure types," *Biometrics*, 59, 1063-1070

[17] Downton, F. (1954), "Least-Squares Estimates Using Ordered Observations," *Annals of Mathematical Statistics*, 25, 303-316.

[18] Duane, J. T. (1964), "Learning curve approach to reliability monitoring," *IEEE Transactions on Aerospace*, 2, 563-566.

[19] Efron, B. (1981), "Censored data and the bootstrap," *Journal of the American Statistical Association*, 76 , 312-319.

[20] Efron, B., and Tibshirani, R. (1993), *An introduction to the bootstrap*, Chapman & Hall Ltd., NY.

[21] Elandt-Johnson, R. (1976), "Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates," *Scandinavian Actuarial Journal*, 59, 37-51.

[22] Feller, W. (1996), *Introduction to Probability Theory and Its Applications: Volume 2*, New York: Wiley.

[23] Finkelstein, J. M. (1976), "Confidence bounds on the parameters of the Weibull process," *Technometrics*, 18, 115-117.

[24] Gill, R. (1983), "Large sample behaviour of the product-limit estimator on the whole line," *The Annals of Statistics*, 11 , 49-58.

[25] Greenhouse, J. B. and Wolfe, R. A. (1984) "A competing risks derivation of a mixture model for the analysis of survival data," *Communications in Statistics: Theory and Methods*, 13, 3133-3154.

[26] Gupta, A. K. (1952), "Estimation of the Mean and Standard Deviation of the Normal Population From a Censored Sample," *Biometrika*, 39, 260-273.

[27] Kochar, S. C. and Proschan, F. (1991), "Independence of time and cause of failure in the multiple dependent competing risks model," *Statistica Sinica*, 1, 295-299.

[28] Langseth, H. and Lindqvist, Bo H. (2006), "Competing risks for repairable systems: a data study," *Journal of Statistical Planning and Inference*, 136, 1687-1700.

[29] Lawless, J. F., Wigg, M. B., Tuli, S., Drake, J. and Lamberti-Pasculli, M. (2001), "Analysis of repeated failures or durations, with application to shunt failures for patients with pediatric hydrocephalus," *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 50, 449-465.

[30] Lloyd, E.H. (1952), "Least-Squares Estimation of Location and Scale Parameters Using Order Statistics," *Biometrika*, 39, 88-95.

[31] Majumdar, S. K. (1993), "An optimum maintenance strategy for a vertical boring machine system," *Operational Research Society of India*, 30, 344-365.

[32] Meeker, W. Q., and Escobar, L. A. (1998), *Statistical methods for reliability data*, Wiley, NY.

[33] Nair, V. N. (1984), "On the behavior of some estimators from probability plots," *Journal of the American Statistical Association*, 79, 823-831.

[34] Reiser, B., Guttman, I., Lin, D. K. J., Guess, F. M. and Usher, J. S. (1995), "Bayesian inference for masked system lifetime data," *Applied Statistics*, 44, 79-90.

[35] Reiser, B., Flehinger, B. J. and Conn, A. R. (1996), "Estimating component-defect probability from masked system success/failure data," *IEEE Transactions on Reliability*, 45, 238-243.

[36] Rigdon, S. E. and Basu, A. P. (2000), *Statistical Methods for the Reliability of Repairable Systems*, New York: Wiley.

[37] Shorack, G. R., and Wellner, J. A. (1986), *Empirical processes with applications to statistics*, Wiley, NY.

[38] Thomas, D. R., and Grunkemeier, G. L. (1975), "Confidence interval estimation of survival probabilities for censored data," *Journal of the American Statistical Association*, 70 , 865-871.