國立交通大學

統計學研究所

碩士論文

非混合治癒模型之統計推論

Statistical Inference for Cure Models

The Non-mixture Approach

研究生：陳瀅竹

指導教授：王維菁　教授

中華民國九十八年六月

非混合治癒模型之統計推論

Statistical Inference for Cure Models

The Non-mixture Approach

研 究 生：陳瀅竹　　Student：Ying-jhu Chen

指導教授：王維菁　教授　　Advisor：Dr. Weijing Wang

國 立 交 通 大 學
統 計 學 研 究 所
碩 士 論 文

A Thesis
Submitted to Institute of Statistics
College of Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of Master
in
Statistics

June,2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

非混合治癒模型之統計推論

學生：陳瀅竹　　指導教授：王維菁 教授

國立交通大學統計學研究所　碩士班

摘　　要

在存活分析中，若在實驗期間能無限延長下，有部分的資料能夠永久存活或者不會經歷到所感興趣的事件發生，必須考慮使用治癒模型。本論文主要探討的是 Tsodikov (1998) 所提出的非混合治癒模型，討論在數學模型的架構下代表的生物意義，藉模擬討論在有母數與無母數的統計推論方法中對治癒率以及發病時間函數參數估計的影響。最後在有解釋變數對模型的影響中，提供符合比例風險迴歸模型(PH)與加速失敗時間模式(accelerated failure time model)的圖型診斷方法。

關鍵字: 治癒率；非混合治癒模型；模型診斷

Statistical Inference for Cure Models

The Non-Mixture Approach

Student: Ying-jhu Chen Advisor: Dr. Weijing Wang

Institute of Statistics

National Chiao Tung University

Hsinchu, Taiwan

Abstract

Cure models are survival models which allows for the existence of cure despite of long-term follow-up. In this thesis, we consider statistic inference for cure models based on the non-mixture approach proposed by Tsodikov(1998). This model has some interesting biological interpretations. Parametric and semi-parametric methods will be investigated. We propose a model diagnostic tool to verify the form of regression models.

Key Words: Cure rate; Non-mixture model; Model checking.

誌　　謝

兩年的碩士生涯，在此論文完成同時也響起了尾聲。

很幸運的在兩年前能夠進入學術界神聖的殿堂，衷心感謝交大統計所老師這些日子的教導。尤其要感謝我的恩師王維菁教授，您總是細心的帶領著我，無論是在課業或論文遇到困難時，引導我如何剖析問題並一步一步去解決。您給了我一個不一樣的視野，使我看問題的核心更加準確。而對於我的生活及未來規劃，也會適時地給予意見與協助。

感謝研究所班上二十一位同學們。這段全天候打拼的日子，沒有你們的鼓勵還有支持，我不會成長了那麼多。一同討論問題、腦力激盪找原因，過程中偶而參插著歡樂的笑聲，我永遠會記著這每一刻。由衷的感謝你們，讓我在交大統研所的生活更多采多姿。

特別要感謝東華大學應用數學系的曾玉玲教授。是您，我才會踏入統計這個領域。我的統計生涯，因為有您，開始有了起點。在這畢業之際，對於未來，我會繼續努力。

最後，謹將此論文獻給我最愛的家人還有王維菁教授，謝謝你們無私的奉獻，給我這個機會支持我，讓我順利的完成碩士學歷。

感謝這一路上曾經幫助過我的人，你們都是我生命中無法取代的貴人。

陳瀅竹　謹誌于
國立交通大學統計學研究所
中華民國九十八年六月

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

Let T be the failure time of interest which is subject to right censoring by C. Denote $S(t) = \Pr(T > t)$, $h(t) = lim_{\Delta \to 0} \frac{\Pr(T \in [t, t+\Delta) | T \geq t)}{\Delta}$ and $H(t) = \int_{x=0}^{t} h(x)dx$ as the survival hazard and cumulative hazard functions respectively. Observed data can be denoted as $\{(X_i, \delta_i)(i = 1, \ldots, n)\}$, where $X_i = T_i \wedge C_i$ and $\delta_i = I(T_i \leq C_i)$ and (T_i, C_i) are random replications of (T, C). Let $t_{(1)} < t_{(2)} < \cdots < t_{(D)}$ be distinct ordered failure times. Under the assumption that T_i and C_i are independent, $S(t)$ can be estimated by the Kaplan-Meier estimator

$$\hat{S}(t) = \prod_{k: t_{(k)} < t} \left(1 - \frac{d_k}{r_k}\right),$$

where d_k is the number of failures and r_k is the number at risk at $t_{(k)}$.

Figure 1-1 depicts the Kaplan-Meier curve of an example in the paper of Tsodikov(2003). The curve estimates the survival function for patients with Hodgkin's disease treated by radiotherapy. The event of interest was defined as death due to the disease. An interesting

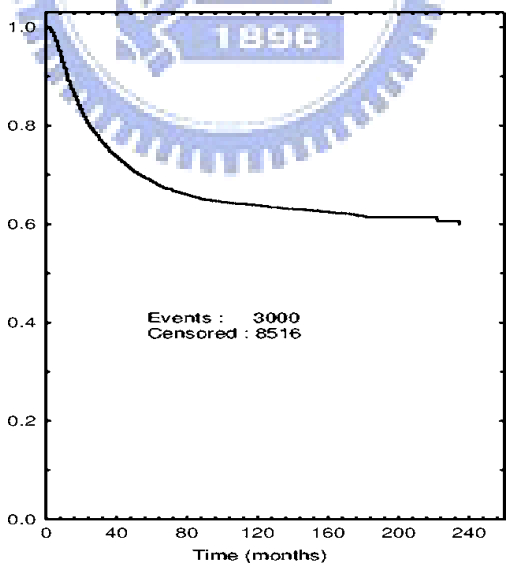


Figure 1-1: K-M Survival function for patients with Hodgkin's disease (Tsodikov, 2003)

feature of the plot is that the tail does not descend to zero. One explanation is that the study period is not long enough to observe large failure points. A different interpretation

is that some patients would never develop the event of interest despite of long-term follow-up. In other words, these patients can be viewed as "cured" and hence free from the event of interest. In many practical applications, such a plateau in the Kaplan-Meier curves is commonly seen. Estimating the proportion of cure may have important medical implication.

Cure models are survival models that allow for cured individuals (Boag, 1949). Here we introduce two types of models which include the possibility of cure in survival analysis.

1.2 The Mixture Approach

It is assumed that the population can be divided into two sub-populations, namely the susceptible group and the cured group. For people in the cured group (with proportion p), the failure time is set to be ∞. Subjects in the susceptible group (with proportion $1-p$) will ultimately experience the event of interest. Denote the corresponding survival function as $S_0(t)$ which is a proper survival function satisfying $\lim_{t\to\infty} S_0(t) = 0$. Hence the survival function of the population can be written as the following mixture form,

$$S(t) = p + (1-p)S_0(t). \tag{1}$$

The corresponding hazard function is given by

$$h(t) = \frac{(1-p)f(t)}{S(t)},$$

where $f(t)$ is the density of T. This model was developed by Berkson and Gage (1952) and later studied extensively by a number of statisticians.

1.3 The Non-Mixture Approach

In the thesis, we will focus on the second type of cure fraction model. Yakovlev et al. (1993) proposed the so-called "non-mixture cure model" or "bounded cumulative hazard" (BCH) model by defining an asymptote for the cumulative hazard function. Specifically it is assumed that the cumulative hazard function can be written as

$$H(t) = \int_0^t h\,(x)\,dx = \theta\tilde{F}(t),$$

where $0 \leq \tilde{F}(t) \leq 1$ satisfies properties of a cumulative distribution function. It follows that

$$\lim_{t\to\infty} H(t) = \theta < \infty.$$

Since

$$S(t) = \exp\left(-\mathrm{H(t)}\right) = \exp\left(-\theta\tilde{\mathrm{F}}(\mathrm{t})\right), \tag{2}$$

it follows that

$$\lim_{t\to\infty} S(t) = \exp(-\theta) = p > 0.$$

1.4 Comparison of the Two Models

The survival function in (2) can be re-expressed as the mathematical form of model (1). Specifically define

$$p = e^{-\theta}$$

and then

$$S_0(t) = \frac{e^{-\theta\tilde{F}(t)} - e^{-\theta}}{1 - e^{-\theta}}. \tag{3}$$

Notice, that in (3), p and $S_0(t)$ are both functions of θ. If the cure fraction p is the only interest, it makes no difference which model formulation is chosen under a nonparametric framework. However, an important feature of the mixture model in (1) is that p and $S_0(t)$ are modeled separately so that $S_0(t)$ does not contain information of p. Therefore, the two approaches become different when additional model assumptions are imposed on the cure rate and the latency distributions.

1.5 Estimation under Censoring

In practice, censoring happens when subjects are lost to follow-up. Observed data are of the form, $\{(X_i, \delta_i), i = 1, 2, \cdots, n\}$. Our main purpose is to estimate θ and $\tilde{F}(\cdot)$. It is easy to see that $\Pr(\delta = 0) \geq \exp(-\theta)$ and $\Pr(\delta = 0) = \exp(-\theta)$ if $C = \infty$. Estimation of θ will be affected by both the distribution of C and $\tilde{F}(t)$. In particular, those outliers are those with extreme values of T which are likely to be censored.

1.6 Outline of the Thesis

The thesis considers statistical inference for the non-mixture model in (2). In Chapter 2, we provide more discussions on this model which include useful biological interpretations. Parametric and non-parametric inference methods are studied in Chapter 3 and Chapter 4 respectively. Regression analysis that models the effect of covariates on p and $\tilde{F}$ is discussed in Chapter 5. We also propose diagnostic plots to verify the proportional hazard (PH) and accelerated failure time (AFT) assumptions. Concluding remarks are given in Chapter 6.

Chapter 2 Model Properties

In this chapter, we will examine properties of the non-mixture model:

$$S(t) = \exp\left(-\theta \tilde{F}(t)\right).$$

In Section 2.1, we introduce a model with interesting biological interpretations which would yield the above representation. More detailed mathematical properties are examined in Section 2.2.

2.1 Biological Interpretation

Consider the following biological mechanism about tumor development. Assume that an individual in the population has N tumor cells where $N = 0, 1, 2 \cdots$. Each tumor cell will develop the event of the metastasis with the failure time $\tilde{T}$. Let $\tilde{T}_j$ be the random time for the j_{th} tumor cell to produce the detectable metastatic disease. Cancer occurs under a competing risk framework such that the time to develop cancer is defined as $T = \min\{\tilde{T}_j, 1 \le j \le N\}$ if $N \ge 1$. If a person has no tumor cell with $N = 0$, he/she is called "immune" or "cured". We will set $\Pr(\tilde{T}_0 = \infty) = 1$.

Assume that N follows a Poisson distribution with mean θ. Also assume that, Given N, the random variables $\tilde{T}_j$ $(j = 1, \ldots, N)$ are independent and identically distributed with a common distribution function $\tilde{F}(t) = \Pr(T_j \le t)$ that does not depend on N. It follows that $S(t)$, the survival function for $T = \min\{\tilde{T}_j, 0 \le j \le N\}$, can be written as

$$\begin{aligned}
S(t) &= \Pr(\text{no metastatic cancer by time t}) \\
&= \Pr(N = 0) + \sum_{k=1}^{\infty} \Pr(T > t, N = k) \\
&= \Pr(N = 0) + \sum_{k=1}^{\infty} \Pr(T > t)\Pr(N = k) \\
&= \exp(-\theta) + \sum_{k=1}^{\infty} (\tilde{S}(t))^k \frac{\exp(-\theta)\theta^k}{k!} \\
&= \exp\left(-\theta + \theta \tilde{S}(t)\right) \\
&= \exp\left(-\theta \tilde{F}(t)\right),
\end{aligned}$$

which is exactly the model considered here. There are two distinct characteristics of tumor growth; namely the initial number of tumor cells and the rate of their progression. Thus parameters in the model have a clear biological meaning. Such a model is reasonable to describe cancer relapse, time to inflection and so forth.

Notice that $\lim_{t\to\infty} S(t) = \Pr(N = 0) = \exp(-\theta)$. As $\theta \to \infty$, the cure rate tends to zero; as $\theta \to 0$, the cure rate reduces to one. Usually we have $0 < p = \exp(-\theta) < 1$. Since $H(t) = \theta\tilde{F}(t)$, the survival function is $S(t) = e^{-\theta\tilde{F}(t)}$ and the hazard function becomes $h(t) = \theta\tilde{f}(t)$, where that $\tilde{f}$ is the density of T_j.

2.2 Mathematical Properties

Based on the biological framework, we discuss how the shape of $S(t)$ is affected by N and $\tilde{F}(t)$. Notice that the cure rate equals $\Pr(N = 0) = \exp(-\theta) = p$ which measures the heaviness of the tail of $S(t)$. The failure time T is determined by both N and the distribution of T_j for $j = 1, \ldots, N$. Larger value of N (i.e. more tumor cells) tends to be associated with smaller value of $T = \min\{\tilde{T}_j, 1 \le j \le N\}$ (i.e. early onset of the event). In other words, if a person has more tumor cells, the time of cancer relapse tends to happen earlier.

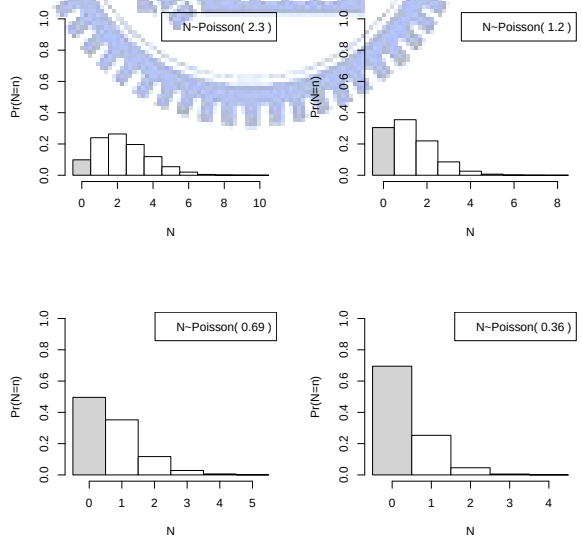


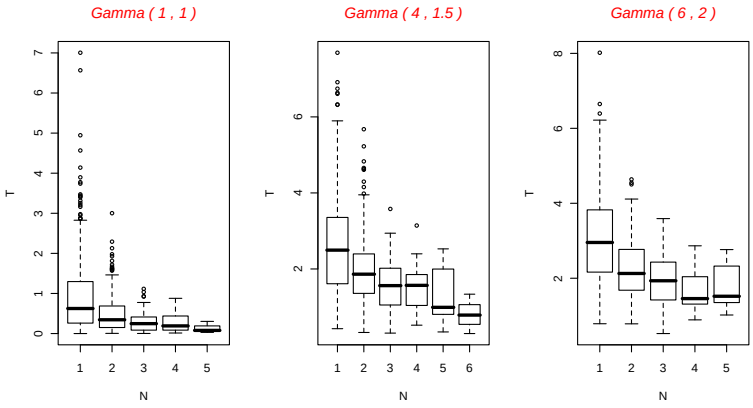
Figure 2-1: The Number of Recurrence tumor

Figure 2-1 plots the probability function of N for selected values of θ. The shaded region corresponds to the cure rate, $\Pr(N = 0)$. Figure 2-2 contains several boxplots of $T|N$

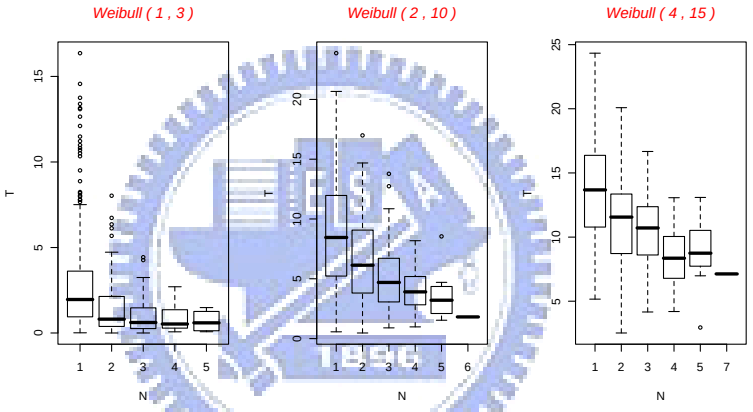
for $N = 0, 1, 2, \cdots$ where $\tilde{T}_j$ follows the Gamma, Weibull and Log-normal distributions respectively. As expected, the median of $T|N$ shows a decreasing pattern as a function of N.

The pattern of outliers in each plot deserves special attention. Outliers of T happen more frequently for those with $N = 1$. Especially in Figure 2-2, there are has a lot of outliers for $T|N = 1$ with $T_j \sim$ Weibull(1,3). If the follow-up period is not very long, large outliers are more likely to be censored. It becomes more difficult to estimate $p = \exp(-\theta)$ if T has a long tail. We will investigate this conjecture via simulations. Among the three Weibull distributions, Var$(T|N)$ for $N \geq 1$ is largest for Weibull(4,15). We will also examine how this characteristic affects parametric inference.

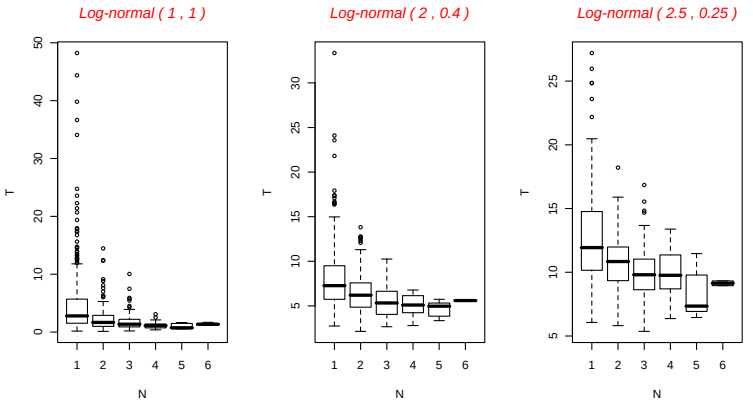
In Figure A-1 of the Appendix, we plot the survival function of T, and the density and the hazard functions of T_j for selected parametric distributions.



(a) Gamma



(b) Weibull



(c) Log-normal

Figure 2-2: The failure time T given tumor number N

Chapter 3 Parametric Inference

Parametric analysis can provide more clear description of the data if the model assumption is correct. In Section 3.1, we derive the likelihood and score functions under a parametric framework. Specific examples are given in Section 3.2. Section 3.3 presents numerical analysis that compares the bias and standard deviation of the maximum likelihood estimates under several parametric settings.

3.1 Likelihood Function

Recall that T denotes the failure time with the survival function $S(t) = \Pr(T > t)$. The censoring variable is denoted by C with the survival function $G(t) = \Pr(C > t)$ and the density function $g(t)$. Assume that T and C are independent. In presence of right censoring, the observed variables are (X, δ), where $X = T \wedge C$ and $\delta = I(T \leq C)$. The sample contains independently and identically distributed observations of (X, δ), denoted as $(X_i, \delta_i)(i = 1, 2, \cdots, n)$.

The likelihood function has the form

$$\prod_{i=1}^{n} [f(x_i)G(x_i)]^{\delta_i} \times [S(x_i)g(x_i)]^{1-\delta_i} \propto \prod_{i=1}^{n} [f(x_i)^{\delta_i} S(x_i)^{1-\delta_i}].$$

Under the non-mixture model, $S(t)$ can be express as

$$S(t) = \exp(-\theta \tilde{F}(t)), \tag{4}$$

and the density function can be expressed as

$$f(t) = -\frac{dS(t)}{dt} = \theta \tilde{f}(t) \exp(-\theta \tilde{F}(t)), \tag{5}$$

where $\tilde{f}(t) = \tilde{F}'(t)$. Accordingly the parametric log-likelihood function can be written as

$$\sum_{i=1}^{n} \{\delta_i [\log(\theta) + \log \tilde{f}(x_i)] - \theta \tilde{F}(x_i)\}. \tag{6}$$

Suppose that $\tilde{F}(\cdot)$ is modeled as a parametric function $\tilde{F}_\eta(\cdot)$ so that the log-likelihood is a function of (θ, η), say $\ell(\theta, \eta)$. The maximum likelihood estimator of (θ, η) can be obtained by solving the equations $\frac{\partial}{\partial \theta}\ell(\theta, \eta) = 0$ and $\frac{\partial}{\partial \eta}\ell(\theta, \eta) = 0$.

3.2 Two Examples

First suppose that $\tilde{f}(\cdot)$ follows an exponential distribution with parameter α. The log-likelihood function in (6) becomes

$$\ell(\theta, \alpha) \propto \sum_{i=1}^{n} \{\delta_i \log(\theta\alpha e^{-\alpha x_i}) - \theta(1 - e^{-\alpha x_i})\}.$$

The score functions are

$$\begin{cases} \theta = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} (1 - e^{-\alpha x_i})}, \\ \sum_{i=1}^{n} \{\delta_i (\frac{1}{\alpha} - x_i) - \theta x_i e^{-\alpha x_i}\} = 0. \end{cases}$$

The solution $\hat{\alpha}$ and $\hat{\theta}$ can be obtained by numerical methods.

In the second example, we assume that $\tilde{f}(\cdot)$ follows a Weibull distribution with the parameters k and λ, where k is the shape parameter and λ is the scale parameter. The decreasing rate of hazard is controlled by k. If $k > 1$, the hazard he hazard has a peak in the λ; if $k \leq 1$, it is decreasing. Note that the shape of hazard function has right tail when coefficient of skewness is larger than zero. The log-likelihood function in (6) becomes

$$\sum_{i=1}^{n} \left\{ \delta_i \left[\log\theta + \log k + (k-1)\log x_i - (k-1)\log\lambda - \left(\frac{x_i}{\lambda}\right)^k \right] - \theta \left(1 - e^{\left(\frac{x_i}{\lambda}\right)^k}\right) \right\}.$$

The score functions are

$$\begin{cases} \theta = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} 1 - e^{\left(\frac{x_i}{\lambda}\right)^k}}, \\ \sum_{i=1}^{n} \left\{ \delta_i \left[\frac{1}{k} + \log x_i - \log\lambda - \left(\frac{x_i}{\lambda}\right)^k \log\left(\frac{x_i}{\lambda}\right) \right] + \theta \left(\frac{x_i}{\lambda}\right)^k \log\left(\frac{x_i}{\lambda}\right) e^{\left(\frac{x_i}{\lambda}\right)^k} \right\} = 0, \\ \sum_{i=1}^{n} \left\{ \delta_i \left[\frac{k-1}{\lambda} + k\lambda^{-k-1} \right] - \theta k \lambda^{-k-1} e^{\left(\frac{x_i}{\lambda}\right)^k} \right\} = 0. \end{cases}$$

The maximum likelihood estimates, $\hat{k}$, $\hat{\lambda}$ and $\hat{\theta}$, can be obtained by numerical methods.

3.3 Generation of a Parametric Distribution

To generate data from a parametric model of the form $S(t) = \exp(-\theta\tilde{F}(t))$, we propose two algorithms. One is based on mathematical properties of the random variables and the other utilizes properties of the biological model as discussed earlier.

3.3.1 Algorithm I

This algorithm is constructed based on the biological model of tumor development.

- Step 1: Generate the number of tumor cell N from a Poisson distribution with mean θ such that $\theta = -\log(p)$, where p is the cure fraction.
- Step 2: If $N = 0$, set T to be a large number;
  if $N > 0$, generate $\tilde{T}_j$ from $\tilde{F}(\cdot)$ independently for $j = 1, \ldots, N$.
- Step 3: Set $T = \min\{\tilde{T}_1, \tilde{T}_2, \cdots, \tilde{T}_N\}$.
- Step 4: Generate C from $Unif(0, K)$.
- Step 5: Set $X = \min\{T, C\}$ and $\delta = \mathrm{I(T \leq C)}$.

3.3.2 Algorithm II

Without the biological setting, we has to handel the challenge that T is not a proper random variable since positive mass is put on $T = \infty$. The second algorithm uses the idea of mixture models to generate random variables. Specifically under the mixture model, we have

$$S(t) = (1 - p)S_0(t) + p,$$

where $S_0(t)$ is the survival function of a proper random variable, say T_0. It follows that $S_0(T_0) \sim U(0, 1)$ so that $T_0 = S_0^{-1}(U)$ where $U \sim U(0, 1)$. Hence we can set $T = T_0$ with probability $1 - p$ and T as a very large number with probability p.

The relationship between T_0 and $(\tilde{F}(\cdot), \theta)$ is given in equation (3). Thus we have

$$S_0(T_0) = \frac{e^{-\theta \tilde{F}(T_0)} - e^{-\theta}}{1 - e^{-\theta}}. \tag{7}$$

Suppose that the parametric form of $\tilde{F}(\cdot)$ is specified. It follows that

$$\tilde{F}(T_0) = \frac{-\log\{U(1 - e^{-\theta}) + e^{-\theta}\}}{\theta}.$$

The second algorithm is described below.

- Step 1: Generate $U \sim U(0, 1)$.

- Step 2: Set
$$T_0 = \tilde{F}^{-1}\left(-\frac{\log(U\{1-e^{-\theta}\}+e^{-\theta}}{\theta}\right).$$
- Step 3: Generate $V \sim U(0,1)$.

  If $V > p = \exp(-\theta)$, set $T = T_0$;

  if $V < p$, set T equal to a very large number.
- Step 4: Generate C from a uniform distribution.
- Step 5: Set $X = T \wedge C$ and $\delta = \mathrm{I}(\mathrm{T} \leq \mathrm{C})$.

3.4 Simulation Results

In the simulations, we evaluate the MLE performance for selected parametric distributions. One set if simulations did not include censoring by setting $C = \infty$ with $p = 0.3$. The other setting was designed such that $p = 0.3$ and $\Pr(\delta = 0) = 0.4$. This means that 10% subjects who are censored but actually susceptible. Since C follows a U(0,k), we need to find the value of k such that $\Pr(T \leq C)$ can achieve the target goal. Table **??** lists the value of k which yields the target value $\Pr(\delta = 0)$ for different parametric distributions.

Data were generated using the algorithm derived from model. Two sample size with $n = 100$ and $n = 300$ were considered. For each estimator, the average bias and standard deviation of parameter estimates are reported base on 500 replications.

Table A-4 $\sim$ A-6 summarize the result of MLE under $C = \infty$ when T_j follows Gamma, Weibulll and log-normal distributions respectively. Note that $p = \exp(-\theta) = 0.3$. Table A-7 $\sim$ A-9 contain the results with same $p = 0.3$ but $\Pr(\delta = 0) = 0.4$. That is 10% subjects who have at least one tumors but are censored due to limited observation period. The estimate of p is poor for Weibull(1,3) which has more outliers than the other two Weibull distributions. The estimator of λ is poor for Weibull(1,3) and Weibull(4,15). Parameter λ controls the spread of the distribution which involves the tail behavior. Weibull(4,15) has large tail values which are likely to be censored.

Chapter 4 Nonparametric Inference

In this chapter, we discuss estimation of θ and $\tilde{F}(\cdot)$ when the distribution of the latter is not specified. The paper of Tsodikov (2001) considered this setting but did not provide the details. Here we give detailed derivations. The likelihood function and score functions are discussed first. Two algorithms for solving the score equations are proposed.

4.1 Likelihood Function

Based on the original data notations, the nonparametric likelihood function can be written as

$$\prod_{i=1}^{n} \left\{ [\Pr(T = x_i, C > x_i)]^{\delta_i} [\Pr(C = x_i, T > x_i)]^{1-\delta_i} \right\}.$$

After taking logarithm and assuming that T_i and C_i are independent, the log-likelihood function can be written as

$$\sum_{i=1}^{n} \left\{ \delta_i \log[\Pr(T = x_i)] + (1 - \delta_i) \log[\Pr(T > x_i)] \right\}.$$

In order to estimate $\tilde{F}(\cdot)$ nonparametrically, we assume that is takes jumps only in observed failure points. Now, we express the log-likelihood function in terms of ordered observations. Let $t_{(1)} < t_{(2)} < \ldots < t_{(D)}$ be observed failure points. Set $t_{(0)} = 0$ and $t_{(D+1)} = \infty$. Denote $m_j = \sum_{i=1}^{n} I(X_i = t_{(j)}, \delta_i = 1)$ which measures the number of failure points at time $t_{(j)}$ and and $n_j = \sum_{i=1}^{n} I(t_{(j)} \leq X_i < t_{(j+1)}, \delta_i = 0)$ which measures the number of censored points in the interval $[t_{(j)}, t_{(j+1)})$. Under the model assumption, it follows that

$$\begin{aligned}
\log L(\theta, \tilde{F}) &\propto \sum_{i=1}^{D} \left\{ m_i \log[S(t_{(i-)}) - S(t_{(i)})] + n_i \log[S(t_{(i)})] \right\} \\
&= \sum_{i=1}^{D} \left\{ m_i \left[-\theta \sum_{k=1}^{i-1} \triangle \tilde{F}_k + \log \left[1 - \exp\left(-\theta \triangle \tilde{F}_i \right) \right] \right] + n_i \left[-\theta \sum_{k=1}^{i} \triangle \tilde{F}_k \right] \right\} \\
&= -\theta \sum_{i=2}^{D} \sum_{k=1}^{i-1} m_i \triangle \tilde{F}_k + \sum_{i=1}^{D} m_i \log \left[1 - \exp\left(-\theta \triangle \tilde{F}_i \right) \right] - \theta \sum_{i=1}^{D} \sum_{k=1}^{i} n_i \triangle \tilde{F}_k
\end{aligned}$$

where $\triangle \tilde{F}_i = \tilde{F}(t_{(i)}) - \tilde{F}(t_{(i-1)})$ subject to the restriction that $\sum_{i=1}^{D} \triangle \tilde{F}_i = 1$. Now we discuss how to maximize $\log L(\theta, \tilde{F})$ with respect to θ under the constraint $\triangle \tilde{F}_j$ $(j = 1, \ldots, D)$.

4.2 Score Functions

In this section, we present two ways to solve the score functions. In Section 4.2.1, since we want to maximize log-likelihood function subject to some constraints, the lagrange multiplier method is natural choice. The other way to solve the variables is presented in Section 4.2.2. We change the form of variables in the log-likelihood function and use the condition to solve the variables.

4.2.1 An algorithm based on the method of Lagrange Multiplier

Applying the Lagrange multiplier method, we can handle the constraint of the parameters in the maximization procedure. Adding the Lagrange multiplier λ, we consider the function

$$h(\theta, \tilde{F}, \lambda) = \log L(\theta, \tilde{F}) + \lambda(1 - \sum_{i=1}^{D} \triangle \tilde{F}_i).$$

Taking the derivatives with respect to the unknown parameters, the score equations can be written as

$$\frac{\partial h(\theta, \tilde{F}, \lambda)}{\partial \triangle \tilde{F}_D} = 0 \Leftrightarrow \lambda = \frac{\partial \log L(\theta, \tilde{F})}{\partial \triangle \tilde{F}_D} = \frac{m_D \theta \exp(-\theta \triangle \tilde{F}_D)}{1 - \exp(-\theta \triangle \tilde{F}_D)} - \theta n_D \tag{8}$$

and

$$\frac{\partial h(\theta, \tilde{F}, \lambda)}{\partial \triangle \tilde{F}_j} = 0, \quad 1 \leq j \leq D - 1.$$

It follows that for each $j = 1, 2, \cdots, D-1$, we have

$$-\theta \sum_{i=j+1}^{D} m_i - \theta \sum_{i=j}^{D} n_i + \frac{m_j \theta \exp(-\theta \triangle \tilde{F}_j)}{1 - \exp(-\theta \triangle \tilde{F}_j)} - \lambda = 0. \tag{9}$$

By taking

$$\frac{\partial h(\theta, \tilde{F}, \lambda)}{\partial \theta} = 0,$$

we obtain

$$-\sum_{i=2}^{D} \sum_{k=1}^{i-1} m_i \triangle \tilde{F}_k - \sum_{i=1}^{D} \sum_{k=1}^{i} n_i \triangle \tilde{F}_k + \sum_{i=1}^{D} m_i \frac{\triangle \tilde{F}_i \exp\left(-\theta \triangle \tilde{F}_i\right)}{1 - \exp\left(-\theta \triangle \tilde{F}_i\right)} = 0 \tag{10}$$

and also

$$\sum_{i=1}^{D} \triangle \tilde{F}_i = 1. \tag{11}$$

The algorithm is summarized below.

- Step 1: Take summation (9)$\times\triangle\tilde{F}_i$ from i =1 to D-1

$$-\theta\sum_{j=1}^{D-1}\sum_{i=j+1}^{D} m_i - \theta\sum_{j=1}^{D-1}\sum_{i=j}^{D} n_i + \sum_{j=1}^{D-1}\triangle\tilde{F}_j\frac{m_j\theta\exp\left(-\theta\triangle\tilde{F}_j\right)}{1-\exp\left(-\theta\triangle\tilde{F}_j\right)} - \sum_{j=1}^{D-1}\lambda = 0 \qquad (12)$$

Then multiply equation((10) by θ

$$-\theta\sum_{i=2}^{D}\sum_{k=1}^{i-1} m_i\triangle\tilde{F}_k - \theta\sum_{i=1}^{D}\sum_{k=1}^{i} n_i\triangle\tilde{F}_k + \theta\sum_{i=1}^{D} m_i\frac{\triangle\tilde{F}_i\exp\left(-\theta\triangle\tilde{F}_i\right)}{1-\exp\left(-\theta\triangle\tilde{F}_i\right)} = 0$$

and subtract from (12). We obtain

$$\triangle\tilde{F}_D\frac{m_D\theta\exp\left(-\theta\triangle\tilde{F}_D\right)}{1-\exp\left(-\theta\triangle\tilde{F}_D\right)} - \theta n_D\triangle\tilde{F}_D + \sum_{j=1}^{D-1}\triangle\tilde{F}_j\lambda = 0.$$

and equation (11). Accordingly we have

$$\lambda = \frac{\theta n_D\triangle\tilde{F}_D - \triangle\tilde{F}_D\frac{m_D\theta\exp\left(-\theta\triangle\tilde{F}_D\right)}{1-\exp\left(-\theta\triangle\tilde{F}_D\right)}}{1-\triangle\tilde{F}_D}. \qquad (13)$$

- Step 2: Comparing (13) and (8), we can get

$$\frac{\theta n_D\triangle\tilde{F}_D - \triangle\tilde{F}_D\frac{m_D\theta\exp\left(-\theta\triangle\tilde{F}_D\right)}{1-\exp\left(-\theta\triangle\tilde{F}_D\right)}}{1-\triangle\tilde{F}_D} = \frac{m_D\theta\exp(-\theta\triangle\tilde{F}_D)}{1-\exp(-\theta\triangle\tilde{F}_D)} - \theta n_D.$$

It follows that

$$\frac{m_D\theta\exp(-\theta\triangle\tilde{F}_D)}{1-\exp(-\theta\triangle\tilde{F}_D)} - \theta n_D = 0.$$

- Step 3: Using the estimation of $\triangle\tilde{F}_D$ in (9), we obtain

$$\frac{m_D\theta\exp(-\theta\triangle\tilde{F}_D)}{1-\exp(-\theta\triangle\tilde{F}_D)} - \frac{m_j\theta\exp(-\theta\triangle\tilde{F}_j)}{1-\exp(-\theta\triangle\tilde{F}_j)} + \sum_{i=j+1}^{D} m_i + \sum_{i=j}^{D-1} n_i = 0,$$

which is used to solve $\triangle\tilde{F}_j$ for $j = 1, 2, \cdots, D-1$.

- Step 4: Finally, obtain the estimate of θ based on equation (10):

$$-\sum_{i=2}^{D}\sum_{k=1}^{i-1} m_i\triangle\tilde{F}_k - \sum_{i=1}^{D}\sum_{k=1}^{i} n_i\triangle\tilde{F}_k + \sum_{i=1}^{D} m_i\frac{\triangle\tilde{F}_i\exp\left(-\theta\triangle\tilde{F}_i\right)}{1-\exp\left(-\theta\triangle\tilde{F}_i\right)} = 0.$$

The above procedure is summarized below.

- Step 1: Obtain $\theta\triangle\tilde{F}_D$ from

$$\frac{m_D\theta\exp(-\theta\triangle\tilde{F}_D)}{1-\exp(-\theta\triangle\tilde{F}_D)}-\theta n_D=0$$

- Step 2: Solve $\theta\triangle\tilde{F}_j$ for $j=1,2,\cdots,D-1$ based on

$$\frac{m_j\theta\exp(-\theta\triangle\tilde{F}_j)}{1-\exp(-\theta\triangle\tilde{F}_j)}-\theta\sum_{i=j+1}^{D}m_i-\theta\sum_{i=j}^{D}n_i=0$$

- Step 3: Since $\sum_{i=1}^{D}\triangle\tilde{F}_i=1$, then set $\sum_{i=1}^{D}\theta\triangle\tilde{F}_i=\theta$ and $\triangle\tilde{F}_i=\frac{\theta\triangle\tilde{F}_i}{\theta}$, $1\le i\le D-1$

4.2.2 An algorithm based on change of variables

Denote $\theta_k=\theta\triangle\tilde{F}_k$ for $k=1,2,\cdots,D$. The log-likelihood function as a function of θ_k can be expressed as the following simpler form:

$$\log L(\theta,\tilde{F})=-\sum_{i=2}^{D}\sum_{k=1}^{i-1}m_i\theta_k+\sum_{i=1}^{D}m_i\log\left[1-\exp\left(-\theta_i\right)\right]-\sum_{i=1}^{D}\sum_{k=1}^{i}n_i\theta_k.$$

Taking the derivatives with respect to θ_j for D first and then $j=1,\ldots,D-1$, we obtain the following score equations:

$$\frac{\partial L(\theta,\tilde{F})}{\partial\theta_D}=0\Leftrightarrow\frac{m_D\exp(-\theta_D)}{1-\exp(-\theta_D)}-n_D=0;$$

and for $1\le j\le D-1$

$$\frac{\partial L(\theta,\tilde{F})}{\partial\theta_j}=0\Leftrightarrow\frac{m_j\exp(-\theta_j)}{1-\exp(-\theta_j)}-\sum_{i=j+1}^{D}m_i-\sum_{i=j}^{D}n_i=0.$$

Notice that θ_j does not depend on other θ_i for $i\neq j$.

Now we summarize the numerical algorithm.

- Step 1: Solve the equation

$$\frac{m_D\exp(-\theta_D)}{1-\exp(-\theta_D)}-n_D=0$$

to obtain the estimator of θ_D.

- Step 2: For $1 \leq j \leq D-1$, the estimator of θ_j solves

$$\frac{m_j \exp(-\theta_j)}{1-\exp(-\theta_j)} - \sum_{i=j+1}^{D} m_i - \sum_{i=j}^{D} n_i = 0,$$

- Step 3: Under the constraint, $\sum_{i=1}^{D} \triangle \tilde{F}_i = 1$, we get

$$\theta_1 + \theta_2 + \cdots + \theta_D = \theta \triangle \tilde{F}_1 + \theta \triangle \tilde{F}_2 + \cdots + \theta \triangle \tilde{F}_D = \theta$$

and $\triangle \tilde{F}_k = \frac{\theta_k}{\theta}$ for $k = 1, 2, \cdots, D$.

4.3 Simulation Results

Here we evaluate the non-parametric approach. Two sample sizes with $n = 100$ and $n = 300$ are evaluated. The censoring variable C by simulations was generated from a uniform distribution in the the interval $[0, k]$, where k is determined based on Table **??**. For θ and p, the average bias and standard deviation are reported based on 500 replications. The result are summarized in Tables A-10 to A-15.

The estimated survival functions using the parametric and non-parametric methods are shown in Figure A-2 to Figure A-10 with $n = 100$ when the parametric assumption is correctly specified. We see that the parametric estimator is closer to the true function and both parametric and non-parametric estimator are poorer with the increasing of curing rate.

In Figure 4-1 below shows the result when the parametric form $\tilde{F}(t)$ is mis-specified. The NPMLE approach is still accurate while the parametric approach fails.

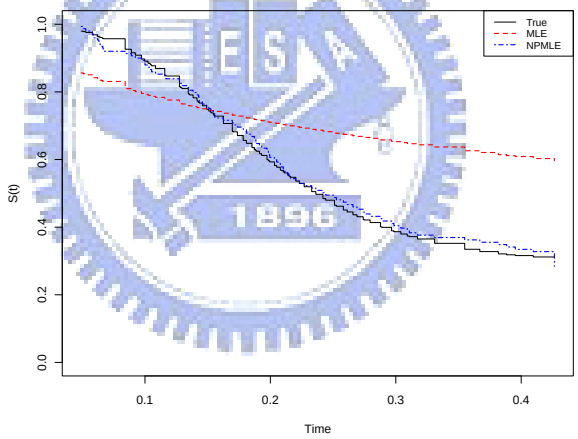


Figure 4-1: Estimated survival funtions under model mis-specification

Chapter 5 Regression Analysis

In practice, data also contain covariate information. Let Z be the covariate so that observed data can be written as $\{(X_i, \delta_i, Z_i)(i = 1, \cdots, n)\}$. The non-mixture cured model in presence of covariate can be written as

$$S(t|Z) = \exp(-\theta(Z)\tilde{F}(t|Z)). \tag{14}$$

5.1 Classification of Covariate Effects

In (14), covariate Z affects both θ and $\tilde{F}(t)$. Here we consider some special cases which are simplified conditions of the general situation. The long-term effect stands for its influence on the tail $\exp(-\theta(Z)) = p(Z)$. The short-term effect is often associated with the timing of tumor occurrence for those with $N \geq 1$. Hence the model in (14) combines both effects.

Example 1: Short-term effect without long-term effect

Here we assume $\theta(Z)$ is independent of Z. For example, if two treatment groups have the same long-term survival rates, and one of them is characterized by more rapidly developing tumors, we could see a significant short-term effect, as shown in Figure 5-1(a). In different groups, the cure rate θ is the same but $\tilde{F}(t|Z)$ depends on Z. Equation (14) can be simplified as

$$S(t|Z) = \exp(-\theta\tilde{F}(t|Z)).$$

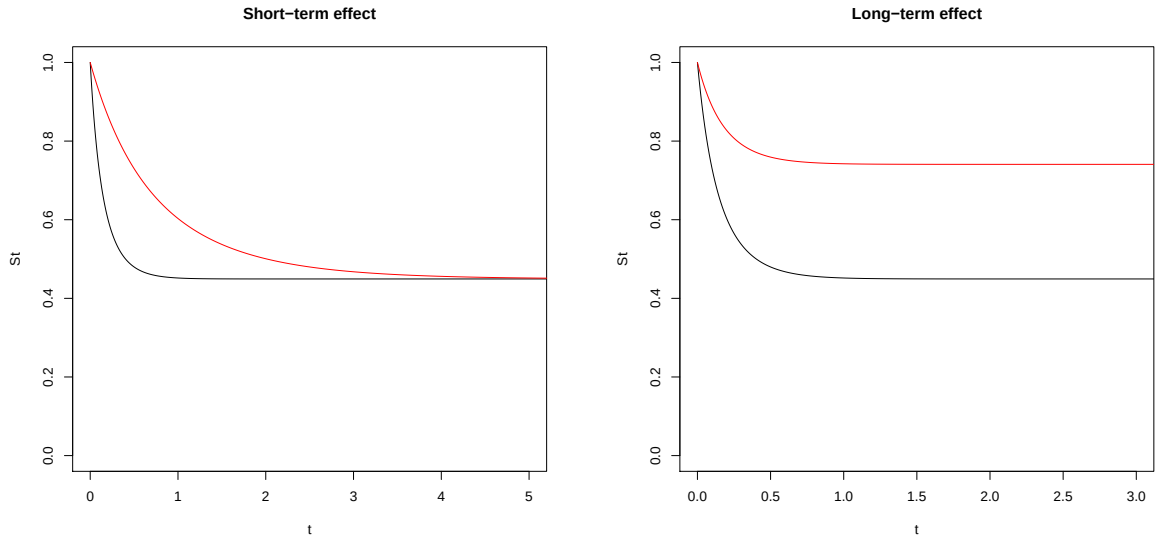
Survival curves for the two groups would diverge initially, and converge again as time passes by.

Example 2: Long-term effect without short-term effect

In contrast, we assume that Z affects the cure rate but not $\tilde{F}(t)$, equation (14) can be rewritten as

$$S(t|Z) = \exp(-\theta(z)\tilde{F}(t)).$$

Figure 5-1(b) depicts the situation for $Z = 0, 1$.



(a) short-term effect (b) Long-term effect

Figure 5-1: The effects of covariate

5.2 Short-term Effect: Proportional Hazards Model

We now illustrate two forms of $\tilde{F}(t|Z)$ which are the most popular choices of regression models in survival analysis. We will also propose model checking procedures to verify the validity of the model assumption. To simplify the presentation, we focus on the two-sample case with $Z = 0, 1$.

Assume the survival function $\tilde{S}(t)$ follows the form of a PH model such that

$$\tilde{S}(t|Z = 1) = \tilde{S}(t|Z = 0)^k, \tag{15}$$

where k is a pre-determined constant. Notice that

$$S(t|Z) = \exp\left(-\theta(z)\tilde{F}(t|Z)\right) = \exp\left(\theta(z)\left(\tilde{S}(t|Z) - 1\right)\right). \tag{16}$$

Thus we have

$$\frac{\log S(t|Z)}{\theta(z)} + 1 = \tilde{S}(t|Z). \tag{17}$$

It follows that

$$\frac{\log S(t|Z = 1)}{\theta(1)} + 1 = \tilde{S}(t|Z = 1) = \left\{\tilde{S}(t|Z = 0)\right\}^k = \left\{\frac{\log S(t|Z = 0)}{\theta(0)} + 1\right\}^k,$$

Accordingly

$$\log\left(\frac{\log S(t|Z=1)}{\theta(1)}+1\right)=k\log\left(\frac{\log S(t|Z=0)}{\theta(0)}+1\right). \tag{18}$$

Notice that equation (18) now involves only estimable quantities. The function $S(t|z)$ can be estimated by the Kaplan-Meier estimator

$$\hat{S}(t|Z=z)=\prod_{0<u\leq t}\left\{1-\frac{\sum_{i=1}^{n}I(X_i=u,\delta_i=1,Z_i=z)}{\sum_{i=1}^{n}I(X_i\geq u,Z_i=z)}\right\}.$$

Denote $\hat{\theta}(z)$ be the nonparametric estimate of $\theta(z)$. Define

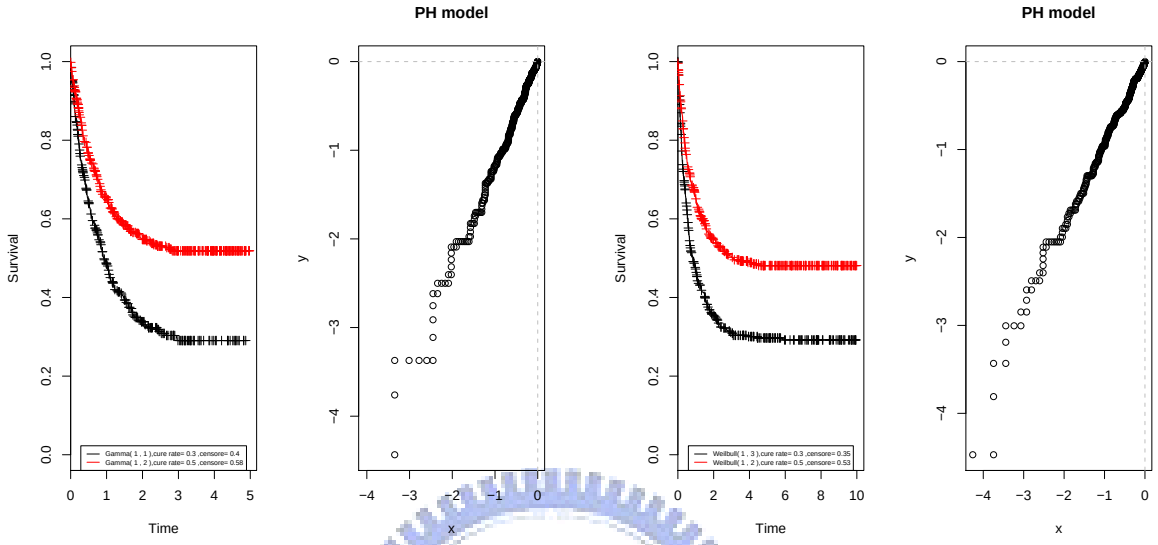
$$X_i=\log\left(\frac{\log\hat{S}(t_i|z=1)}{\hat{\theta}(1)}+1\right)$$

and

$$Y_i=\log\left(\frac{\log\hat{S}(t_i|z=0)}{\hat{\theta}(0)}+1\right),$$

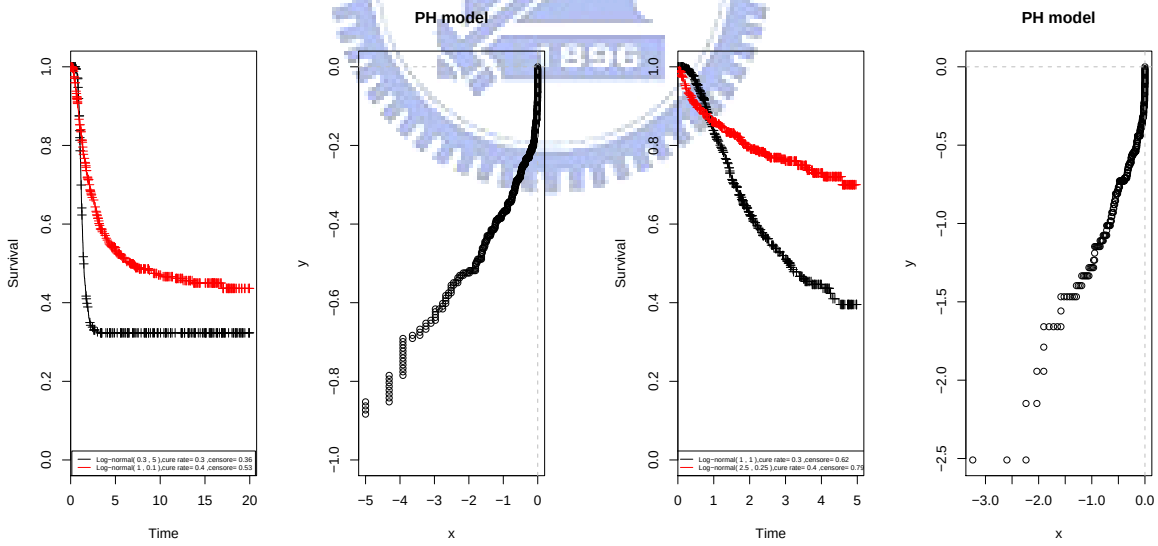
where $t_1<t_2<\cdots<t_D$ are ordered failure points. If the PH assumption holds, equation (18) indicates that Y follow a linear relationship passing through the origin. We present some plots of $\hat{S}(t|Z)$ for $z=0,1$ and the corresponding diagnostic plot of X_i versus Y_i for $i=1,2,\cdots,D$ based on 1000 simulated observations.

The diagnostic plots in Figure 5-2 reveal clear linear pattern for most points. Figure 5-3 presents the plots when the PH assumption is violated. There is a curved relationship between X_i and Y_i.



(a) Case I (b) Case II

Figure 5-2: K-M curves and diagnostic plots when the PH assumption holds.



(a) Case I (b) Case II

Figure 5-3: K-M curves and diagnostic plots when the PH assumption is violated.

5.3 Short Term Effect: Accelerated Failure Time Model

Under the AFT model, the survival function $\hat{S}(t|z)$ for $z = 0, 1$ follows the relationship

$$\tilde{S}(t|Z = 1) = \tilde{S}(kt|Z = 0)$$

for k being a prespecified constant. It follows that

$$S(t|z = 1) = \exp\left\{\theta(1)\left[\tilde{S}(t|z = 1) - 1\right]\right\} = \exp\left\{\theta(1)\left[\tilde{S}(kt|z = 0) - 1\right]\right\}.$$

Accordingly we have

$$\frac{\log S(t|Z = 1)}{\theta(1)} + 1 = \tilde{S}(t|Z = 1) = \tilde{S}(kt|Z = 0) = \frac{\log S(kt|Z = 0)}{\theta(0)} + 1. \tag{19}$$

We want to find a clear relationship from the above equation.

Define $p_1 < \cdots < p_M$ as some constants locating in $(0, 1)$. Then we solve X_i satisfying

$$\frac{\log S(x_i|Z = 1)}{\theta(1)} = p_i,$$

$i = 1, 2, \cdots, M$. Thus, for each p_i we have

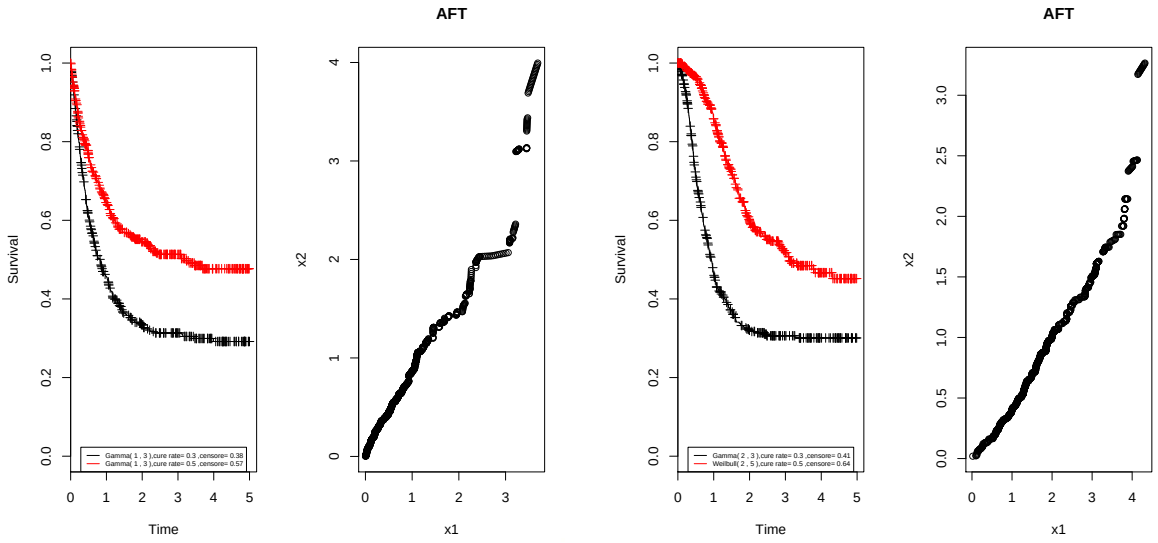
$$X_i = \hat{S^{-1}_{Z=1}}\left\{\exp\left(p_i\hat{\theta}(1)\right)\right\}.$$

Similar steps can be derived based on the right-hand side of equation (19). Set

$$Y_i = \hat{S^{-1}_{Z=0}}\left\{\exp\left(p_i\hat{\theta}(0)\right)\right\}$$

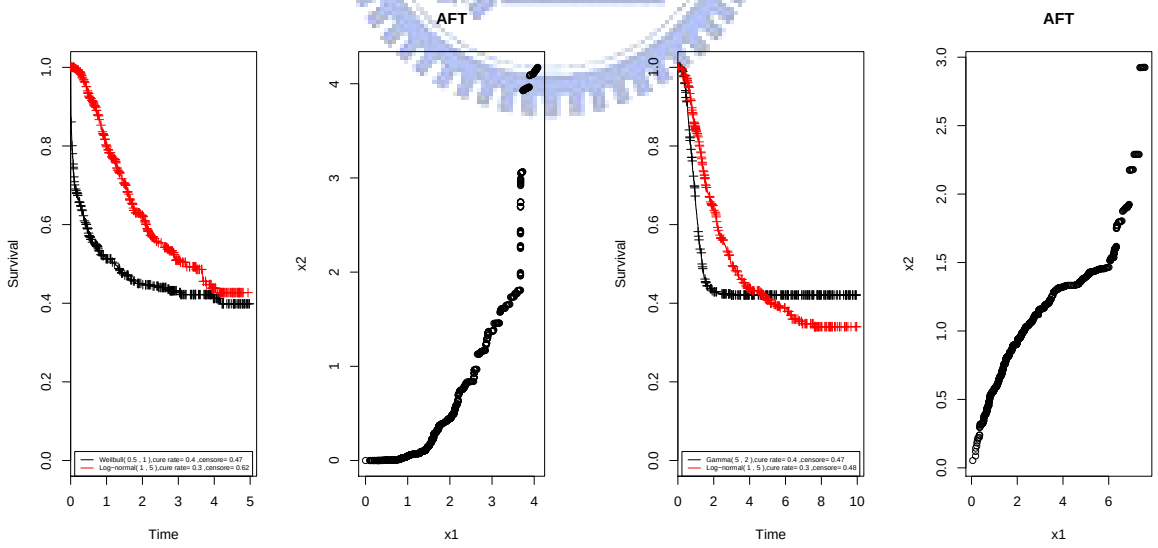
for $i = 1, 2, \cdots, M$. When the AFT model holds, $(X_i,\ Y_i)$ $(i = 1, 2, \cdots, M)$ will follow a straight line through the origin. Plots following the AFT model are presented in Figure 5-4, in which n=1000.

The diagnostic plots in Figure 5-4 reveal clear linear pattern for most points. Figure 5-5 presents the plots when the AFT assumption is violated. There is a curved relationship between X_i and Y_i.



(a) Case I (b) Case II

Figure 5-4: K-M curves and diagnostic plots when the AFT assumption holds.



(a) Case I (b) Case II

Figure 5-5: K-M curves and diagnostic plots when the AFT assumption is violated.

Chapter 6 Conclusion

In the thesis, we study the non-mixture approach for analyzing survival data in presence of cure. This formulation has an interesting biologial interpretation. In parametric analysis, we find that outliers in T which often occur when N = 1 will affect estimation of the cure rate. For nonparametric inference, we propose two algorithms to solve the score funcions of nonparametric MLE. One is the classical Lagrange multiplier method and the other is by change of variables. Two regression models are considered under the simplified two-sample setting. One is the proportional hazard model and the other is the accelerated failure time model. We propose diagnostic plots which can verify the form of regression effect.

References

[1] J. Berkson and R. P. Gage (1952). Survival Curve for Cancer Patients Following Treatment. *J. Amer. Statist. Assoc.* 47, 501-515.

[2] Boag, JW. (1949). Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy. *J. Royal Statist. Soc., Series B* 11:15-44.

[3] Tsodikov, A. D. (1998). A Proportional Hazards Model Taking Account of Long-Term Survivors. *Biometrics* 54, 1508-1516.

[4] Tsodikov, A. D. (2001). Estimating of Survival Based on Proportional Hazards When Cure is a Possibility. *Mathematical and Computer modelling.* 33, 1227-1236.

[5] Tsodikov, A. D., Ibrahim, J. G. and Yakovlev A. Y. (2003). Estimating Cure Rates From Survival Data : An Alternative to Two-Component Mixture Models . *E. J. Am. Statist. Assoc.* 98, 1063-1078.

Appendix: Additional Figures

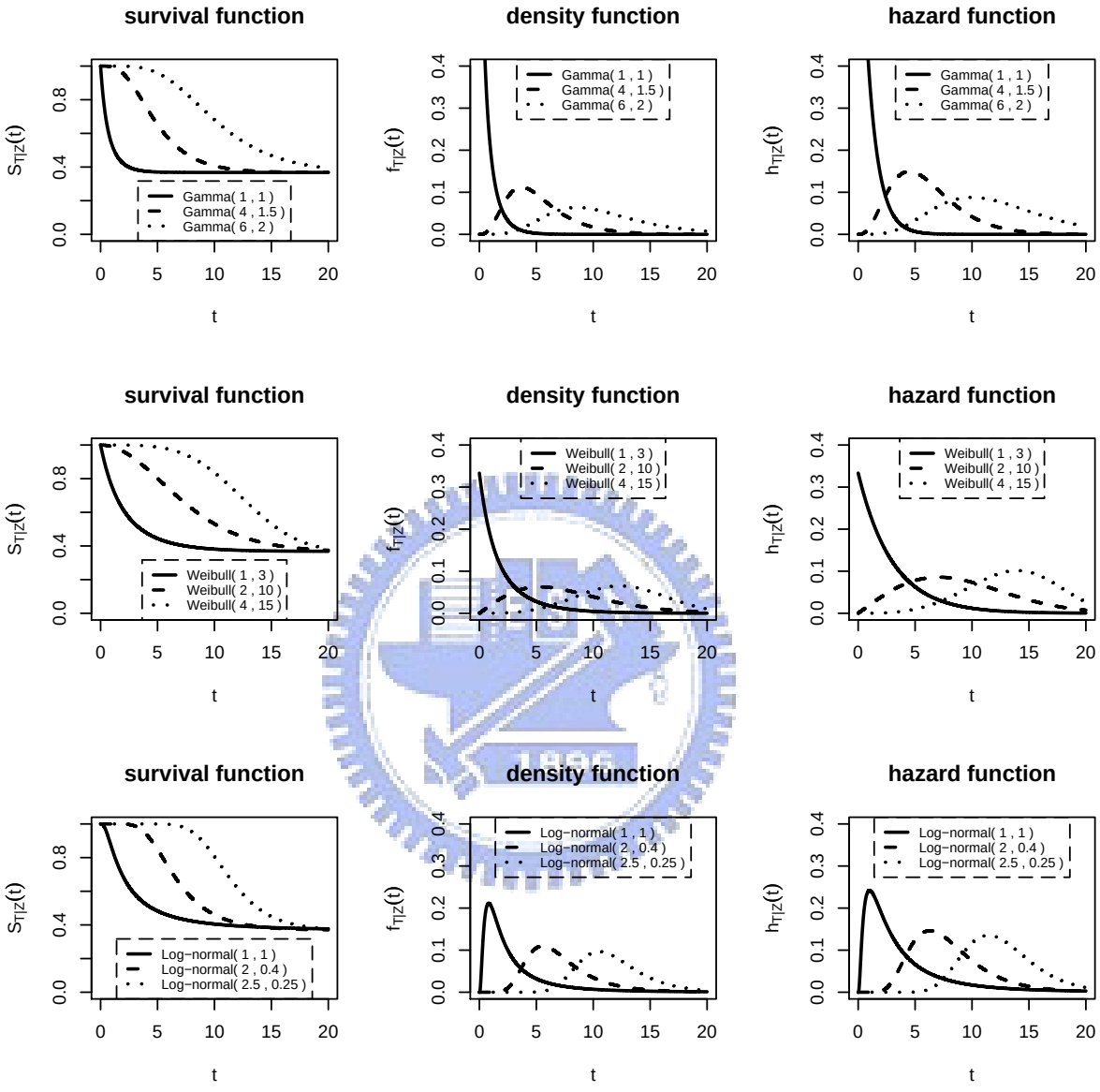


Figure A-1: Survival density and hazard functions for selected parametric distributions.

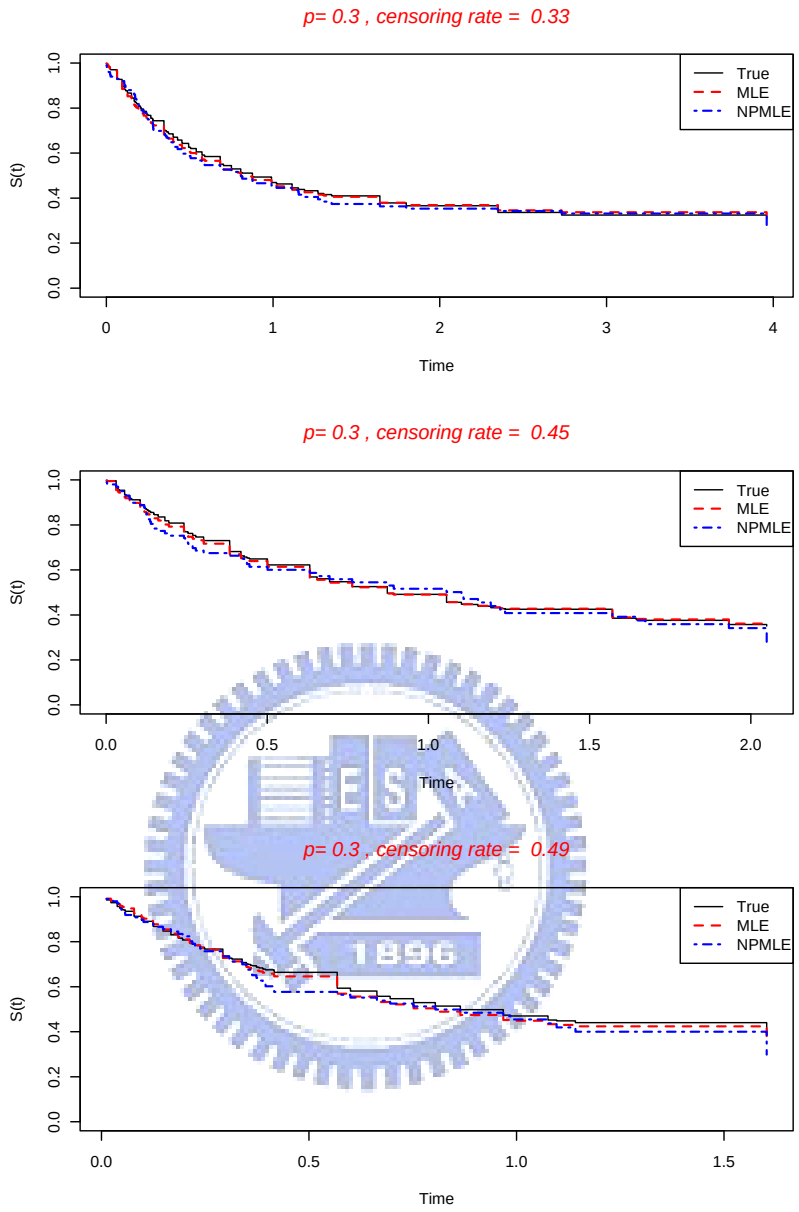


Figure A-2: Estimated survival functions when $\tilde{F}$ is correctly specified as Gamma(1,1)

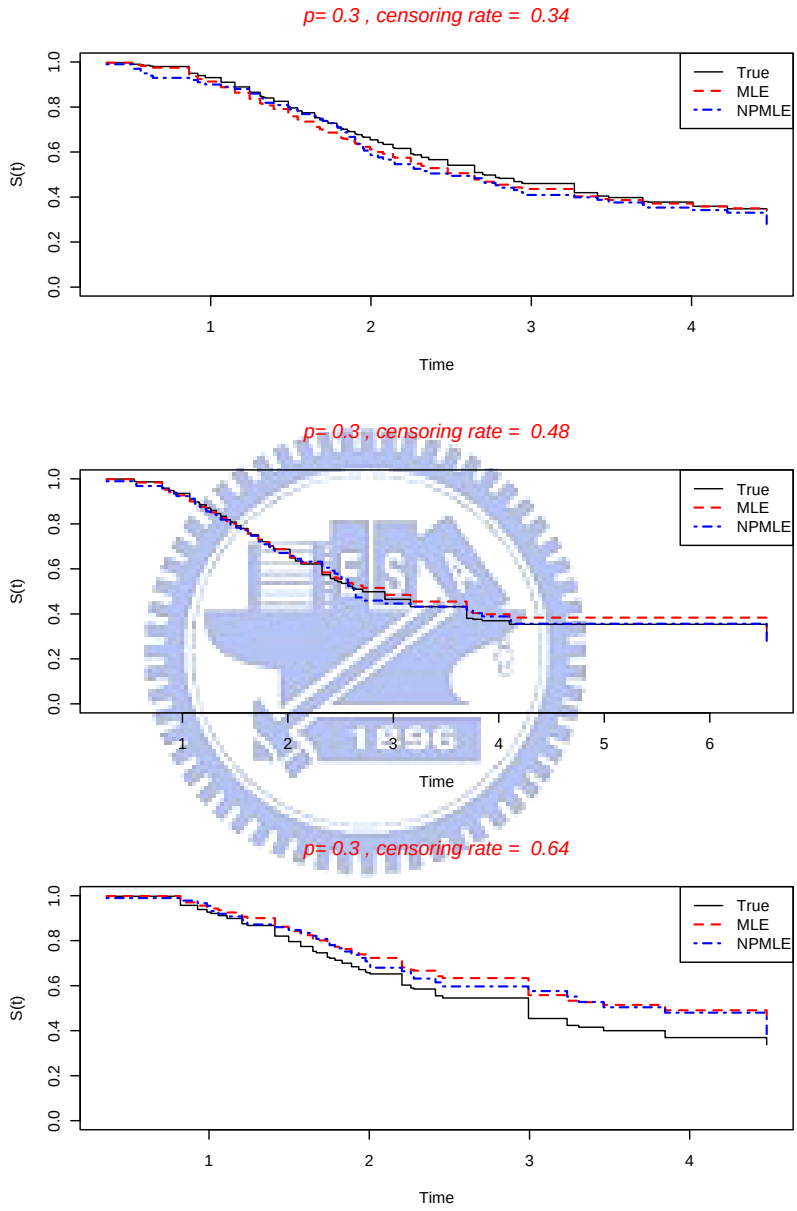


Figure A-3: Estimated survival functions when $\tilde{F}$ is correctly specified as Gamma(4,1.5)

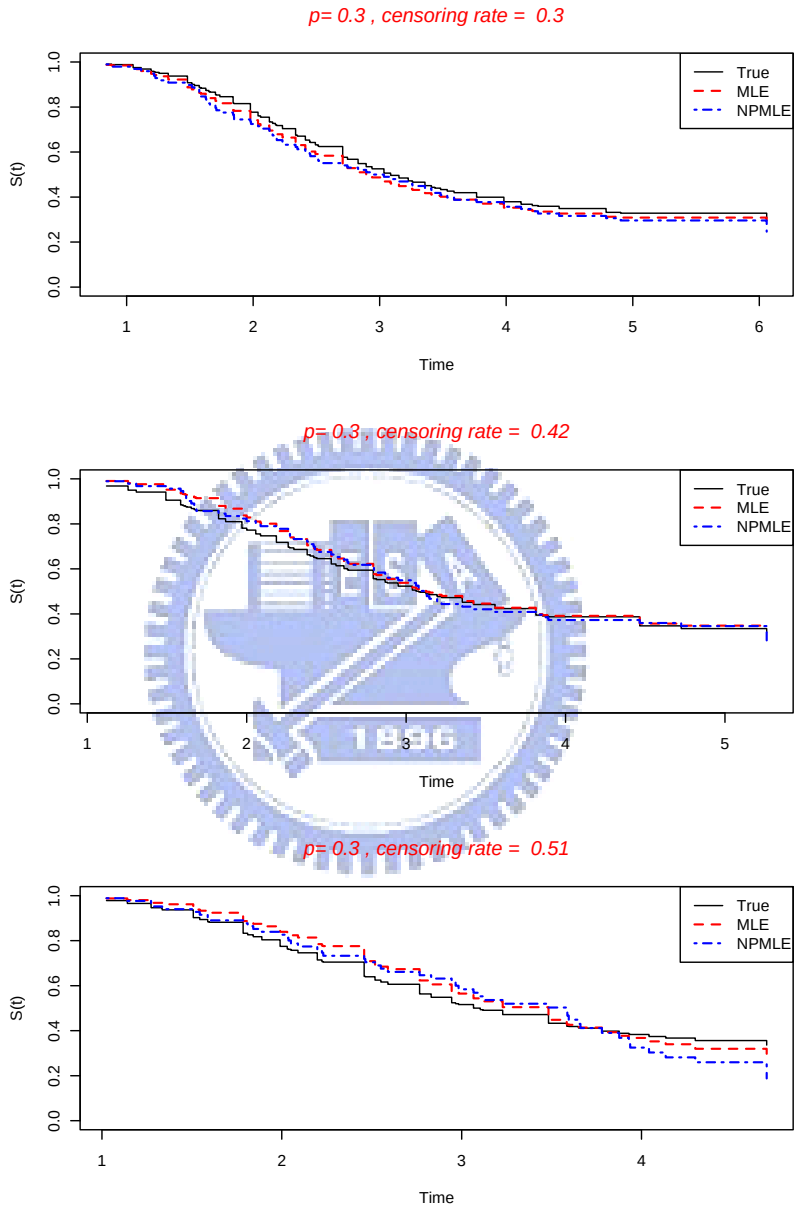


Figure A-4: Estimated survival functions when $\tilde{F}$ is correctly specified as Gamma(6,2)

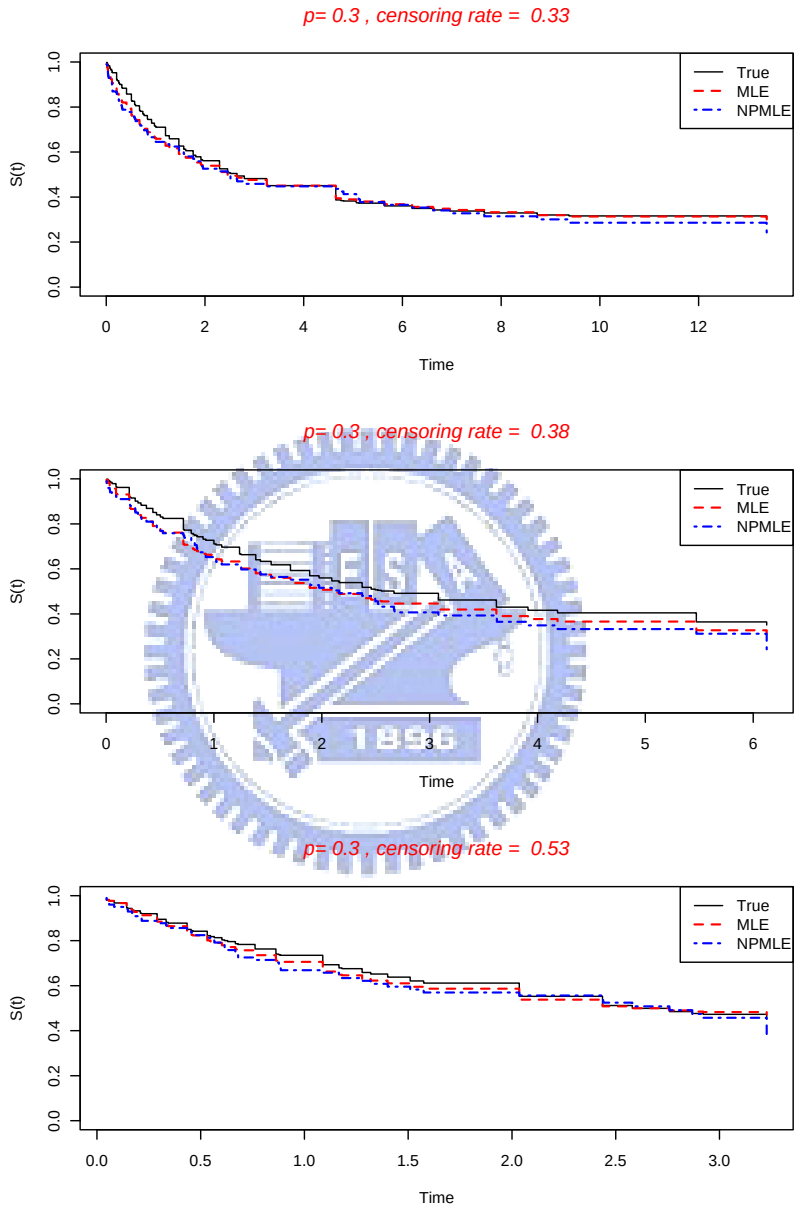


Figure A-5: Estimated survival functions when $\tilde{F}$ is correctly specified as Weibull(1,3)

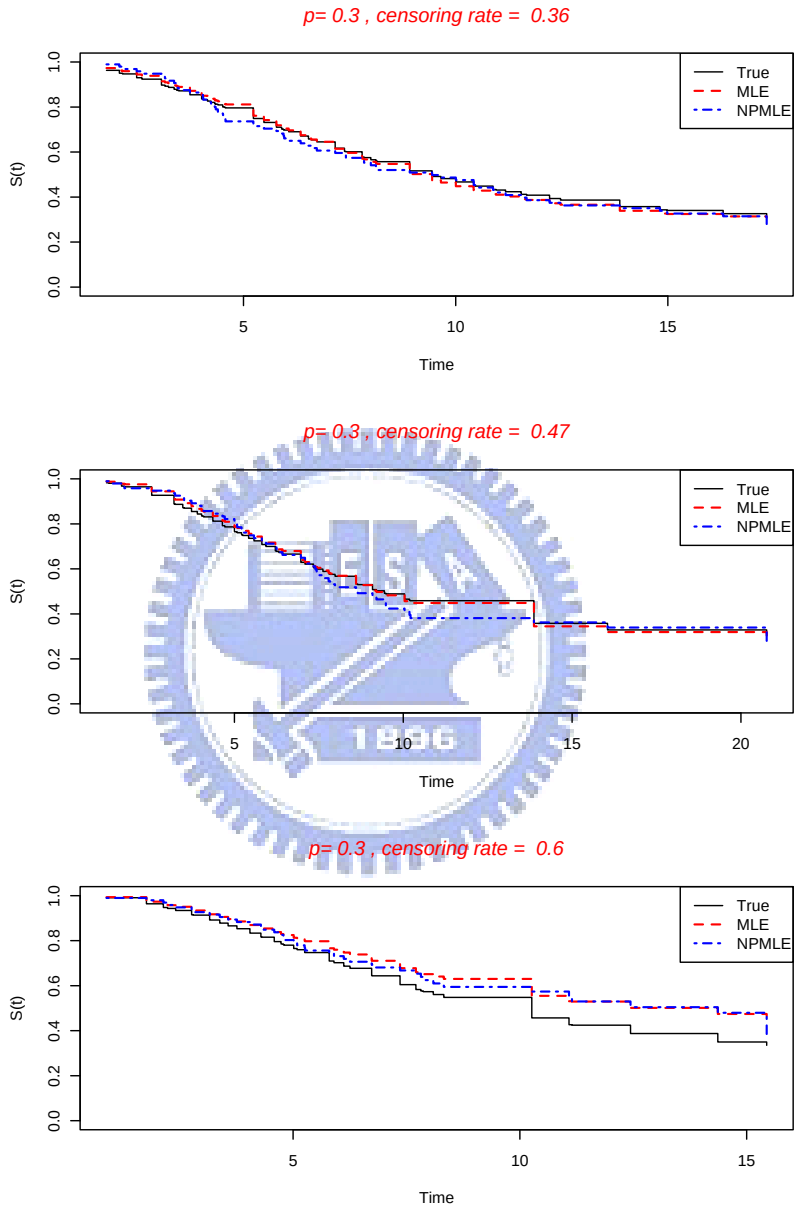


Figure A-6: Estimated survival functions when $\tilde{F}$ is correctly specified as Weibull(2,10)

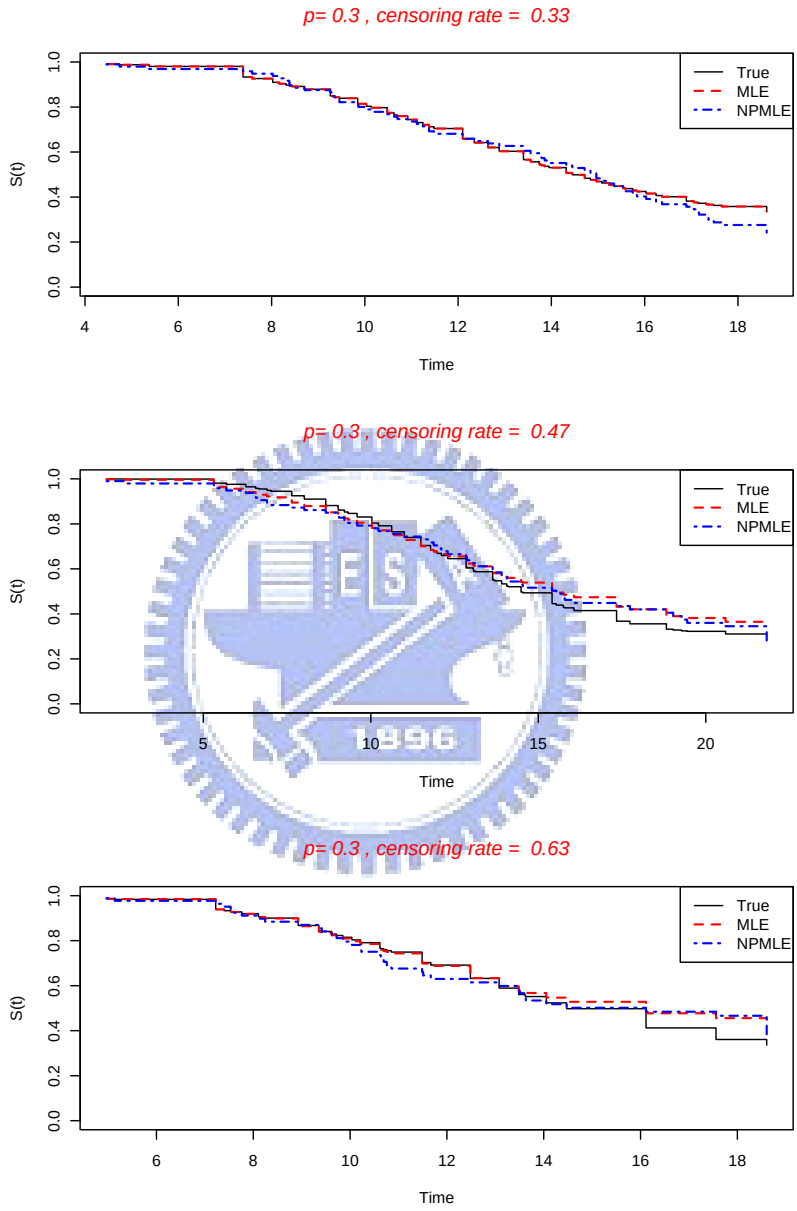


Figure A-7: Estimated survival functions when $\tilde{F}$ is correctly specified as Weibull(4,15)

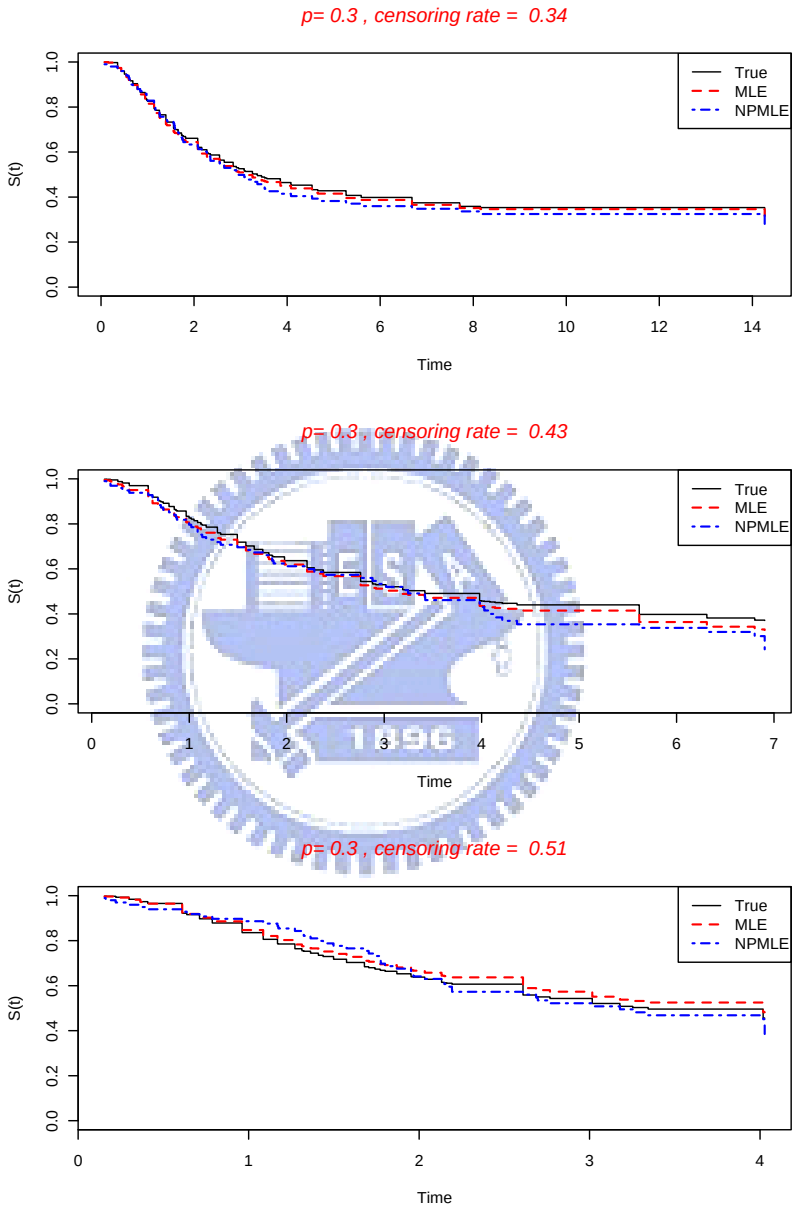


Figure A-8: Estimated survival functions when $\tilde{F}$ is correctly specified as log-normal(1,1)

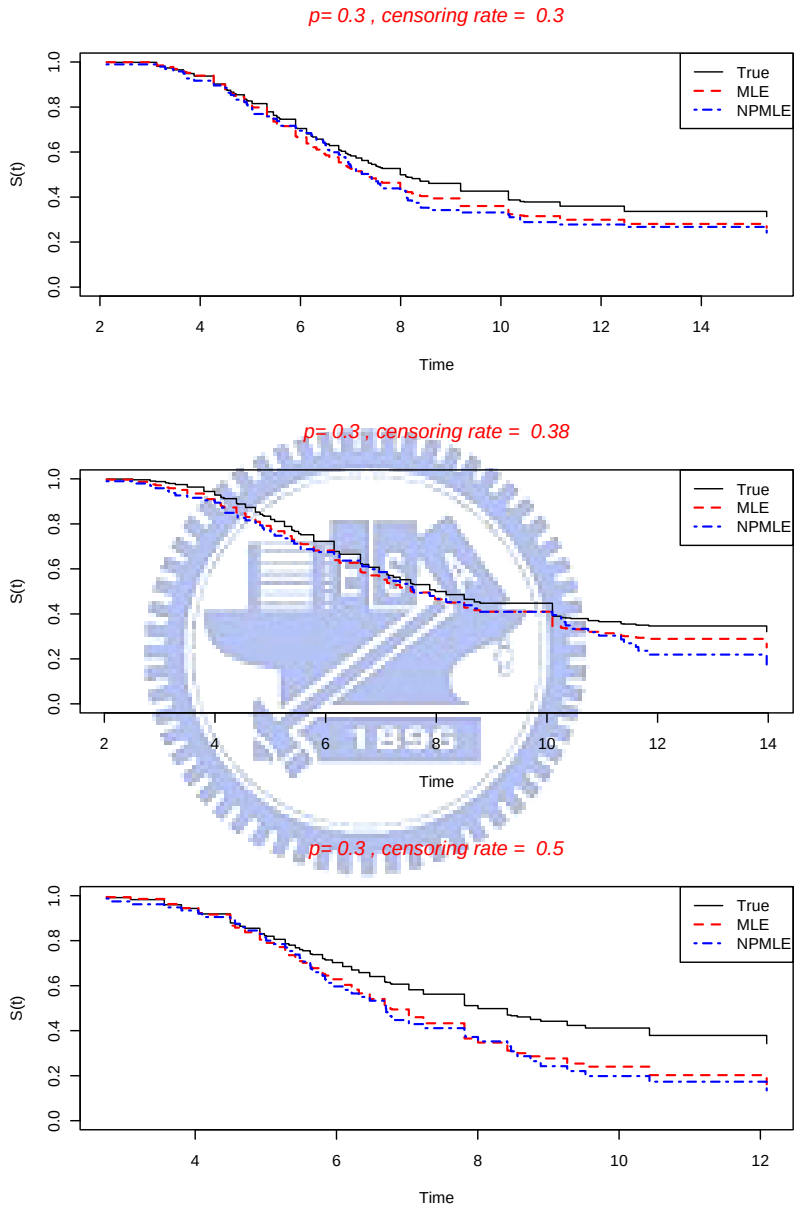


Figure A-9: Estimated survival functions when $\tilde{F}$ is correctly specified as log-normal(2,0.4)

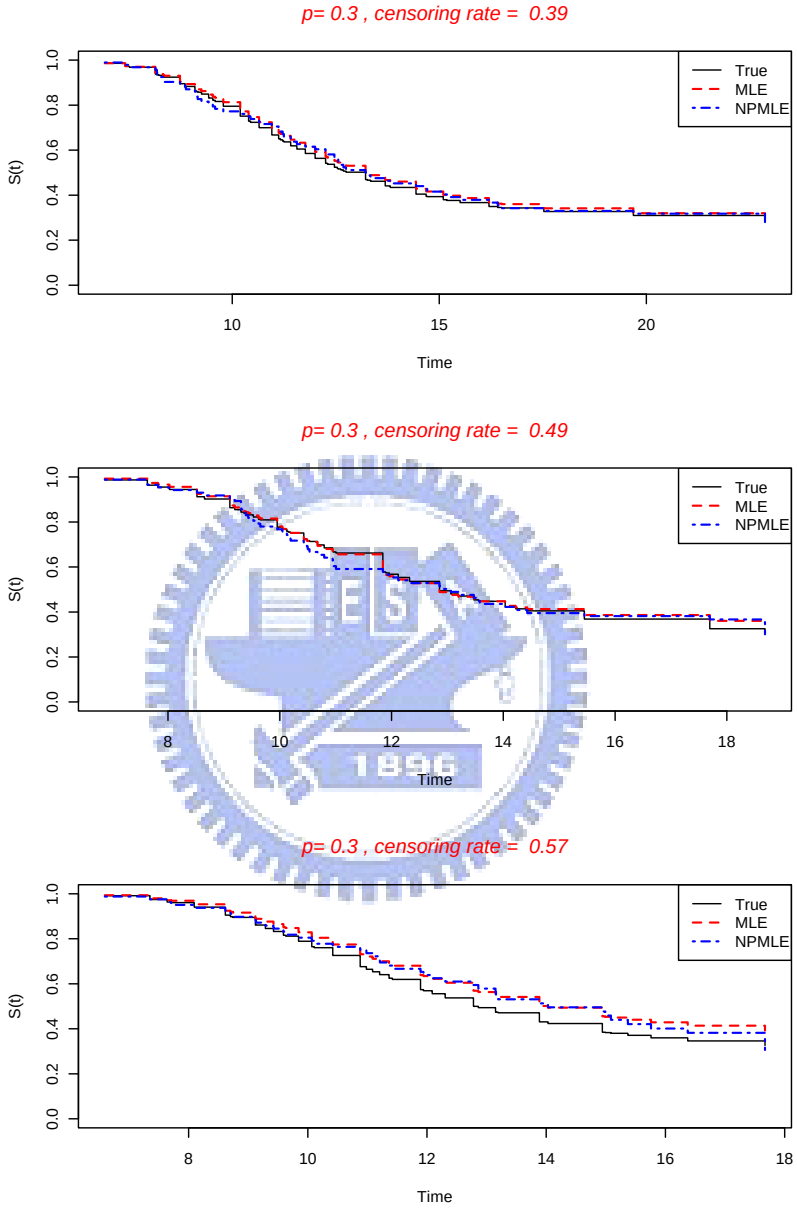


Figure A-10: Estimated survival functions when $\tilde{F}$ is correctly specified as log-normal(2.5,0.25)

Appendix: Simulation Results

Table A-1: Relationship between C∼Unif[0,k] and Pr($\delta = 0$) for Gamma Distributions

	Gamma(1,1)	k	Gamma(4,1.5)	k	Gamma(6,2)	k
p=0.3	Pr(δ=0)=0.3	×	Pr(δ=0)=0.3	×	Pr(δ=0)=0.3	×
	Pr(δ=0)=0.4	4	Pr(δ=0)=0.4	11	Pr(δ=0)=0.4	13
	Pr(δ=0)=0.5	2	Pr(δ=0)=0.5	7	Pr(δ=0)=0.5	7
p=0.5	Pr(δ=0)=0.5	×	Pr(δ=0)=0.5	×	Pr(δ=0)=0.5	×
	Pr(δ=0)=0.6	3	Pr(δ=0)=0.6	8	Pr(δ=0)=0.6	10
	Pr(δ=0)=0.7	2	Pr(δ=0)=0.7	5	Pr(δ=0)=0.7	6

Table A-2: Relationship between C∼Unif[0,k] and Pr($\delta = 0$) for Weibull Distributions

	Weibull(1,3)	k	Weibull(2,10)	k	Weibull(4,15)	k
p=0.3	Pr(δ=0)=0.3	×	Pr(δ=0)=0.3	×	Pr(δ=0)=0.3	×
	Pr(δ=0)=0.4	10	Pr(δ=0)=0.4	33	Pr(δ=0)=0.4	58
	Pr(δ=0)=0.5	6	Pr(δ=0)=0.5	22	Pr(δ=0)=0.5	36
p=0.5	Pr(δ=0)=0.5	×	Pr(δ=0)=0.5	×	Pr(δ=0)=0.5	×
	Pr(δ=0)=0.6	8	Pr(δ=0)=0.6	18	Pr(δ=0)=0.6	45
	Pr(δ=0)=0.7	4	Pr(δ=0)=0.7	17	Pr(δ=0)=0.7	26

Table A-3: Relationship between C∼Unif[0,k] and Pr($\delta = 0$) for Log-normal Distributions

	log-normal(1,1)	k	log-normal(2,0.4)	k	log-normal(2.5,0.25)	k
p=0.3	Pr(δ=0)=0.3	×	Pr(δ=0)=0.3	×	Pr(δ=0)=0.3	×
	Pr(δ=0)=0.4	14	Pr(δ=0)=0.4	33	Pr(δ=0)=0.4	54
	Pr(δ=0)=0.5	8	Pr(δ=0)=0.5	19	Pr(δ=0)=0.5	33
p=0.5	Pr(δ=0)=0.5	×	Pr(δ=0)=0.5	×	Pr(δ=0)=0.5	×
	Pr(δ=0)=0.6	12	Pr(δ=0)=0.6	25	Pr(δ=0)=0.6	40
	Pr(δ=0)=0.7	8	Pr(δ=0)=0.7	15	Pr(δ=0)=0.7	25

Table A-4: Maximized likelihood estimators for Gamma distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.3$

		Gamma(1,1)		Gamma(4,1.5)		Gamma(6,2)	
		bias	sd	bias	sd	bias	sd
n=100	α	0.032	0.159	0.129	0.706	0.252	0.086
	β	0.069	0.285	0.079	0.345	0.082	0.344
	θ	0.014	0.147	0.003	0.165	0.030	0.163
	p	0.001	0.044	0.005	0.049	0.005	0.048
n=300	α	0.001	0.074	0.042	0.359	0.061	0.512
	β	0.014	0.134	0.022	0.175	0.021	0.206
	θ	0.003	0.091	0.004	0.093	0.012	0.102
	p	0.002	0.027	0.003	0.028	0.005	0.031

Table A-5: Maximized likelihood estimators for Weibull distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.3$

		Weibull(1,3)		Weibull(2,10)		Weibull(4,15)	
		bias	sd	bias	sd	bias	sd
n=100	k	0.008	0.107	0.028	0.197	0.079	0.461
	λ	0.028	0.483	0.051	0.746	0.068	0.683
	θ	0.013	0.141	0.012	0.154	0.034	0.190
	p	0.007	0.043	0.007	0.046	0.005	0.053
n=300	k	0.001	0.057	0.024	0.102	0.049	0.240
	λ	0.016	0.287	0.039	0.503	0.048	0.383
	θ	0.007	0.095	0.021	0.101	0.008	0.096
	p	0.001	0.028	0.005	0.030	0.001	0.028

Table A-6: Maximized likelihood estimators for Log-normal distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.3$

		log-normal(1,1)		log-normal(2,0.4)		log-normal(2.5,0.25)	
		bias	sd	bias	sd	bias	sd
n=100	μ	0.007	0.110	0.005	0.058	0.001	0.037
	σ^2	0.014	0.094	0.010	0.031	0.001	0.021
	θ	0.010	0.166	0.039	0.173	0.028	0.178
	p	0.001	0.048	0.007	0.050	0.004	0.005
n=300	μ	0.008	0.091	0.001	0.033	0.001	0.023
	σ^2	0.005	0.057	0.005	0.020	0.001	0.015
	θ	0.035	0.101	0.007	0.094	0.001	0.091
	p	0.009	0.029	0.001	0.028	0.001	0.027

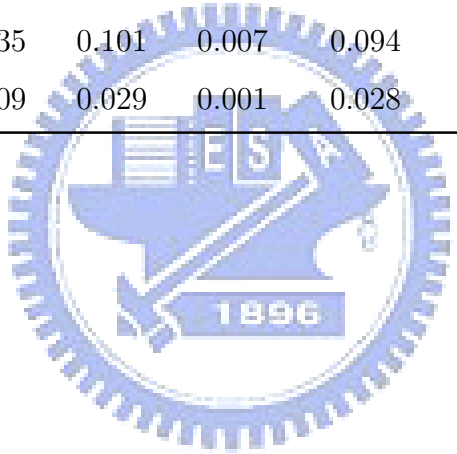


Table A-7: Maximized likelihood estimators for Gamma distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.4$

		Gamma(1,1)		Gamma(4,1.5)		Gamma(6,2)	
		bias	sd	bias	sd	bias	sd
n=100	α	0.036	0.175	0.154	0.768	0.221	1.183
	β	0.095	0.423	0.071	0.384	0.090	0.480
	θ	0.041	0.260	0.035	0.201	0.018	0.194
	p	0.003	0.067	0.005	0.057	<0.001	0.055
n=300	α	0.001	0.091	0.093	0.426	0.117	0.677
	β	0.022	0.237	0.046	0.215	0.051	0.275
	θ	0.016	0.139	0.005	0.107	0.005	0.112
	p	0.002	0.039	<0.001	0.032	0.003	0.021

Table A-8: Maximized likelihood estimators for Weibull distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.4$

		Weibull(1,3)		Weibull(2,10)		Weibull(4,15)	
		bias	sd	bias	sd	bias	sd
n=100	k	0.016	0.131	0.049	0.232	2.979	0.134
	λ	51.455	1111.197	0.060	1.314	85.573	2163.867
	θ	0.535	7.980	0.030	0.226	0.375	6.338
	p	0.016	0.085	0.002	0.058	0.008	0.081
n=300	k	0.001	0.076	0.003	0.127	0.032	0.243
	λ	0.112	0.698	0.001	0.623	0.053	0.377
	θ	0.024	0.169	0.007	0.113	0.002	0.110
	p	0.003	0.047	<0.001	0.033	0.001	0.033

Table A-9: Maximized likelihood estimators for Log-normal distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.4$

		log-normal(1,1)		log-normal(2,0.4)		log-normal(2.5,0.25)	
		bias	sd	bias	sd	bias	sd
n=100	μ	0.035	0.289	0.003	0.063	0.002	0.038
	σ^2	0.001	0.149	0.007	0.041	0.003	0.026
	θ	0.067	0.354	0.022	0.182	0.007	0.179
	p	0.007	0.073	0.002	0.052	0.003	0.053
n=300	μ	0.010	0.152	<0.001	0.037	<0.001	0.023
	σ^2	0.004	0.085	0.002	0.025	0.001	0.014
	θ	0.013	0.144	0.001	0.101	0.001	0.114
	p	0.001	0.042	0.002	0.030	0.001	0.032

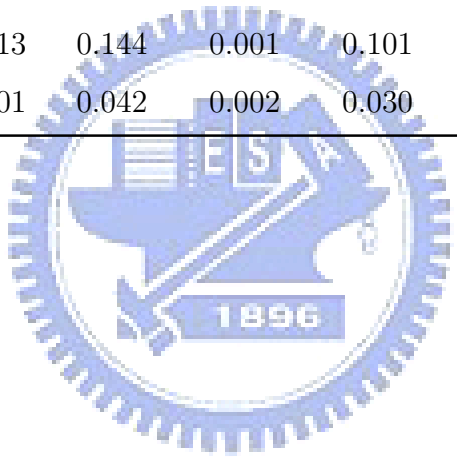


Table A-10: NPMLE of θ and p for Gamma distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.3$.

		Gamma(1,1)		Gamma(4,1.5)		Gamma(6,2)	
		bias	sd	bias	sd	bias	sd
n=100	θ	0.158	0.172	0.131	0.143	0.131	0.139
	p	0.040	0.043	0.034	0.036	0.034	0.036
n=300	θ	0.130	0.142	0.125	0.138	0.12	0.129
	p	0.034	0.036	0.033	0.031	0.032	0.033

Table A-11: NPMLE of θ and p for Weibull distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.3$.

		Weibull(1,3)		Weibull(2,10)		Weibull(4,15)	
		bias	sd	bias	sd	bias	sd
n=100	θ	0.130	0.142	0.048	0.249	0.026	0.266
	p	0.034	0.036	0.004	0.083	0.004	0.090
n=300	θ	0.124	0.134	0.039	0.250	0.033	0.258
	p	0.033	0.034	0.001	0.085	0.001	0.087

Table A-12: NPMLE of θ and p for Log-normal distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.3$.

		log-normal(1,1)		log-normal(2,0.4)		log-normal(2.5,0.25)	
		bias	sd	bias	sd	bias	sd
n=100	θ	0.137	0.309	0.141	0.289	0.056	0.246
	p	0.035	0.037	0.036	0.037	0.007	0.082
n=300	θ	0.136	0.297	0.139	0.279	0.059	0.239
	p	0.035	0.037	0.035	0.037	0.008	0.080

Table A-13: NPMLE of θ and p for Gamma distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.4$.

		Gamma(1,1)		Gamma(4,1.5)		Gamma(6,2)	
		bias	sd	bias	sd	bias	sd
n=100	θ	0.090	0.337	0.148	0.389	0.096	0.327
	p	0.017	0.073	0.036	0.041	0.019	0.071
n=300	θ	0.092	0.331	0.147	0.381	0.097	0.321
	p	0.018	0.072	0.036	0.040	0.020	0.070

Table A-14: NPMLE of θ and p for Weibull distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.4$.

		Weibull(1,3)		Weibull(2,10)		Weibull(4,15)	
		bias	sd	bias	sd	bias	sd
n=100	θ	0.143	0.375	0.079	0.231	0.146	0.405
	p	0.035	0.040	0.015	0.075	0.036	0.041
n=300	θ	0.144	0.413	0.082	0.227	0.145	0.396
	p	0.035	0.040	0.016	0.073	0.036	0.040

Table A-15: NPMLE of θ and p for Log-normal distributions with $p = 0.3$ and $\Pr(\delta = 0) = 0.4$.

		log-normal(1,1)		log-normal(2,0.4)		log-normal(2.5,0.25)	
		bias	sd	bias	sd	bias	sd
n=100	θ	0.140	0.264	0.142	0.274	0.066	0.239
	p	0.036	0.039	0.036	0.038	0.010	0.079
n=300	θ	0.138	0.259	0.141	0.267	0.069	0.234
	p	0.035	0.039	0.036	0.037	0.012	0.077