國立交通大學

統計學研究所

碩 士 論 文

混合治癒模型之文獻回顧

Literature Review on Cure Models - The Mixture Approach

研 究 生：何字卿

指導教授：王維菁　教授

中 華 民 國 九 十 八 年 六 月

混合治癒模型之文獻回顧

Literature Review on Cure Models

- The Mixture Approach

研 究 生：何字卿　　Student：Tzu-Chin He

指導教授：王維菁 教授　　Advisor：Dr. Wei-jing 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

混合治癒模型之統計推論

研究生：何宇卿　　指導教授：王維菁 教授

國立交通大學理學院
統計學研究所

摘要

臨床醫學研究的資料常可見某些患者並不會經歷感興趣的事件。即使有充裕的觀察時間，但“免疫者”依舊存在的存活分析稱為治癒模式。混和模型是最常見的建模方法，母體被區分為“可致病”和“免疫”的兩個群體，並分別對於“免疫比例”與 “可致病者之發病時間”做進一步假設。本論文中回顧此領域部份重要文獻，探討母數與半母數的架構下之推論方法，包含最大概似估計法與動差法。我們提供系統化的角度，透析不同方法背後的建構原則，希望對未來延伸的研究有所助益。

關鍵詞：治癒模型; 可致病的; 轉換模型

、

Statistical Inference for Cure Models By A Mixture Approach

Student：Tzu-Chin He Advisor：Dr. Wei-jing Wang

Institute of Statistics
National Chiao Tung University
Hsinchu, Taiwan

Abstract

Cure models are suitable for analyzing survival data when some people never experience the event of interest despite of long-term follow-up. The most popular modeling approach is the so-called mixture model in which the population is divided into a susceptible group and a group of cure. In the thesis, we review important literature on the cure mixture model. Both parametric and semi-parametric inference methods are considered. In particular, the likelihood approach and methods based on some moment relationship are examined under different model assumptions. We aim to provide a systematic way of studying and comparing these methods.

Keywords: cure model; susceptibility; transformation models

`

致　　謝

此篇碩士論文的完成代表著我將離開校園生活，踏入人生的另一段旅程，在碩士兩年的求學生涯中，我學習到許多，有關乎統計上的知識，亦有人生處事的道理，也經歷了低潮、快樂、滿足。

首先，我要感謝王維菁教授，跟著老師做研究的過程中，老師不斷教導我正確的研究態度與方法，如何將問題回到最原始最基本的架構上作探討進而能推廣到不同情況或假設下，如何將事物用言簡意賅的方式表達，十分感謝老師這一年多來的教誨；也感謝所上所有教授這兩年來的教導，讓我了解到統計更廣更深的面向；感謝郭姐在生活或課業上給予許多叮嚀與關心；在研究室裡一起努力的夥伴們，不論是課業上的交流討論、閒暇時閒扯談天，這些都為我的生活增添許多歡樂。

即將邁入另一階段，我期以自我能將所學有所發揮，保有高度求知慾與熱情。感謝所有給予我幫助與關心的人，我將完成此論文的喜悅與榮耀與你們一同分享。

何字卿 僅誌

中華民國 九十八年夏

于交通大學統計學研究所

、

Contents

Chapter 1 Introduction

1.1 Background

Let T be the failure time of interest and $S(t) = \Pr(T > t)$ be the survival function of T. Traditional survival models assume that every subject in the study will eventually experience the event of interest. This assumption implies that $\lim_{t\to\infty} S(t) = 0$. In practical survival data are often subject to censoring. Let C be the censoring time. Under right censoring, one observes $\{(X_i, \delta_i),\ \ i = 1, ..., n\}$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$. Assume that T and C are independent, the Kaplan-Meier estimator of $S(t)$ can be written as

$$\hat{S}(t) = \prod_{u\leq t}\{1 - \frac{\sum_{i=1}^{n} I(X_i = u, \delta_i = 1)}{\sum_{i=1}^{n} I(X_i \geq u)}\}.$$

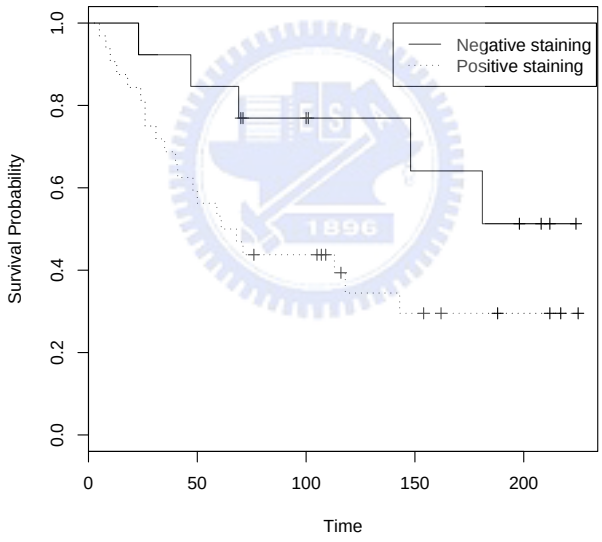

Figure 1: K-M Survival Function

Figure 1.1 shows two Kaplan-Meier survival functions for two groups of patients. Notice that the two curves level off at the right tail and exhibit a stable plateau. Such a phenomenon is commonly seen in clinical trials and cancer studies. Whether there exists a group of subject who will never experience the event of interest despite long-term follow-up is an important scientific problem that requires subject-matter knowledge.

1.2 Mixture Model

Survival models in presence of immune or cured subjects have been extensively studied in statistical literature. The most popular approach is perhaps the mixture model formulation. Denote a binary variable ζ, where $\zeta = 1$ indicates that the subject will experience the event of interest and $\zeta = 0$ indicates that the subject will never experience the event no matter how large C is. For those susceptible ones with $\zeta = 1$, $T < \infty$ with $\tilde{S}(t) = \Pr(T > t|\zeta = 1)$. For those cured individuals with $\zeta = 0$, $T = \infty$. The survival function of the whole population can be written as the following mixture form,

$$\begin{aligned} S(t) &= Pr(T > t|\zeta = 1)Pr(\zeta = 1) + Pr(\infty > t|\zeta = 0)Pr(\zeta = 0) \\ &= \tilde{S}(t) \cdot p + (1 - p), \end{aligned}$$

where $1 - p = \Pr(\zeta = 0)$ is the cure rate. Under this model, the cumulative distribution of T becomes

$$F(t) = \Pr(T \leq t) = p(1 - \tilde{S}(t)) = p\tilde{F}(t),$$

and the hazard function can be written as

$$\lambda(t) = \frac{p\tilde{f}(t)}{1 - p + p\tilde{S}(t)},$$

where $\tilde{f}(t)$ is the density of $T|\zeta = 1$. Usually $\tilde{S}(t)$ is called as the "latency" survival function for the susceptible group.

Cure models with the right censored observations suffer from an inherent identifiability problem. Specifically the observation period should be long enough to make a judgment on the existence of cure. The book of Maller and Zhou (1996) contains detailed discussions on this issue under a non-parametric setting.

In most applications, covariate information is also available which would release the assumption on identifiability. Denote Z as a $p \times 1$ vector of covariates. The data are of the form, (X_i, δ_i, Z_i), $i = 1, ..., n$, where Z_i is the covariate vector for the ith subject. Hence the mixture model can be written as

$$\begin{aligned} S(t|Z) &= Pr(\zeta = 1|Z)Pr(T > t|\zeta = 1, Z) + Pr(\zeta = 0|Z)Pr(\infty > t|\zeta = 0, Z) \\ &= \pi(Z)\tilde{S}(t|Z) + (1 - \pi(Z)). \end{aligned}$$

The cure rate $1-\pi(Z)$ is often analyzed under a logistic regression assumption. In the landmark papers by Farewell (1982,1986), the Weibull distribution is imposed on the latency distribution. Other parametric models such generalized F distributions have also been proposed by Peng,

Dear and Denham (1998). Alternatively semi-paramertric regression models on the latency variable haven been studied. For example, the Cox proportional hazard (PH) model has been assumed by Kuk and Chen (1992), Peng and Dear (2000) and Sy and Taylor (2000), just to name a few. A more general class of semi-parametric transformation cure models has been considered by Lu and Ying (2004). It is suggested that, besides applying statistical methods to verify the identifiability condition, applications of cure models should be done with caution. One should check whether the result is consistent with medical or biological interpretation.

1.3 Outline of the Thesis

In the thesis, we will review some literature on survival models in presence of cure under the mixture formulation. We hope to provide a systematic way of studying different inference methods. Parametric analysis will be discussed in Chapter 2. Chapter 3 considers the likelihood approach for analyzing the Cox proportional hazard model and the linear transformation model. Methods developed based on some moment properties are discussed in Chapter 4. Concluding remarks are given in Chapter 5.

Chapter 2 Parametric Analysis

In this chapter, we examine the likelihood approach for estimating unknown parameters under a parametric setting.

2.1 Homogenous Data without Covariates

Suppose that observed data can be written as $\{(X_i, \delta_i),\ \ i = 1, ..., n\}$, which are independently and identically distributed replications of (X, δ), where $X = T \wedge C$ and $\delta = I(T \wedge C)$. Assume that $S(t)$ and $f(t)$ can be indexed by $S_\theta(t)$ and $f_\theta(t)$, where θ is the parameter of interest. In absence of cure, the parametric likelihood of θ can be written as

$$\prod_{i=1}^{n}[f_\theta(x_i)G(x_i)]^{\delta_i}[S_\theta(x_i)g(x_i)]^{1-\delta_i} \propto \prod_{i=1}^{n} f_\theta(x_i)^{\delta_i} S_\theta(x_i)^{1-\delta_i}, \tag{2.1}$$

where $G(t)$ and $g(t)$ is the survival and density functions of C respectively. The latter expression in (2.1) can also be represented by the hazard and survival functions as follows

$$L(\theta) = \prod_{i=1}^{n}[\lambda_\theta(x_i)S_\theta(x_i)]^{\delta_i}[S_\theta(x_i)]^{1-\delta_i} = \prod_{i=1}^{n} \lambda_\theta(x_i)^{\delta_i} S_\theta(x_i). \tag{2.2}$$

In presence of cure, denote the latency density $\tilde{f}(t)$ as $\tilde{f}_\theta(t)$. The likelihood becomes a function of θ and p which can be written as

$$L(\theta, p) = \prod_{i=1}^{n}[p\tilde{f}_\theta(x_i)]^{\delta_i}[p\tilde{S}_\theta(x_i) + 1 - p]^{1-\delta_i}. \tag{2.3}$$

The maximum likelihood estimator of $(\theta, p)^T$ can be obtained by solving the two equations

$$\frac{\partial logL(\theta, p)}{\partial \theta} = 0$$

and

$$\frac{\partial logL(\theta, p)}{\partial p} = 0$$

given that the functions are differentiable with respect to (θ, p). It follows that

$$\frac{\partial logL(\theta, p)}{\partial p} = \sum_{i=1}^{n} \frac{\delta_i}{p} + \sum_{i=1}^{n}(1 - \delta_i)\frac{\tilde{S}_\theta(x_i) - 1}{p\tilde{S}_\theta(x_i) + 1 - p},$$
$$\frac{\partial logL(\theta, p)}{\partial \theta} = \sum_{i=1}^{n} \frac{\delta_i}{\tilde{f}_\theta(x_i)}\tilde{f}'_\theta(x_i) + \sum_{i=1}^{n} \frac{p(1 - \delta_i)}{p\tilde{S}_\theta(x_i) + 1 - p}\tilde{S}'_\theta(x_i),$$

where $\tilde{f}'_\theta(x_i)$ and $\tilde{S}'_\theta(x_i)$ are partial derivatives of $\tilde{f}_\theta(x_i)$ and $\tilde{S}_\theta(x_i)$ with respect to θ. Notice that the score equations are complicated functions of p and θ since the two types of parameters need to be estimated jointly.

2.2 Parametric Analysis with Covariates

Suppose that observed data can be written as (X_i, δ_i, Z_i), $i = 1, ..., n$. To model $\zeta_i = 1|Z_i$, the following logistic regression model is often assumed:

$$p_i(\gamma) = \Pr(\zeta_i = 1|Z_i) = \frac{exp(Z_i'\gamma)}{1 + exp(Z_i'\gamma)},$$

where $\gamma : q \times 1$ is the parameter vector. Suppose a parametric form is imposed on $\tilde{f}(t|Z)$ as $\tilde{f}_\theta(t|Z)$. The likelihood function can be written as

$$L(\theta, \gamma) = \prod_{i=1}^{n} [p_i(\gamma) f\tilde{|}z_{i\theta}(x_i)]^{\delta_i} [p_i(\gamma) S\tilde{|}z_{i\theta}(x_i) + 1 - p_i(\gamma)]^{1-\delta_i}. \tag{2.4}$$

We will illustrate the likelihood analysis for two parametric models.Maximization of the likelihood function in (2.4) under the two models sometimes are quite complicated.

Farewell (1982, 1986) assumed that $T|\zeta = 1, Z$ follows a Weibull model such that its density function and survival function can be written as

$$\tilde{f}_\theta(t|Z) = \alpha\lambda(\lambda t)^{\alpha-1} \exp(-(\lambda t)^\alpha),$$

where $\theta = (\alpha, \lambda)^T$, $\lambda = exp(-Z'\beta)$ and

$$\tilde{S}_\theta(t|Z) = \Pr(T > t|\zeta = 1, Z) = exp(-(\lambda t)^\alpha).$$

Alternatively Peng, Dear and Denham (1998) applied the generalized F distribution to model the latency distribution. This parametric model is a flexible distribution which contains some useful distributions as special cases such as the Weibull and beta distributions. It is assumed that for those with $\zeta = 1$, $Y = \log(T)$ follows the generalized F distribution with the survival function

$$\tilde{S}_\theta(y|Z) = \Pr(Y > y|\zeta = 1, Z) = \int_0^{s_2(s_2 + S_1 \exp(\frac{y_i - \mu}{\sigma}))^{-1}} \frac{x^{s_2-1}(1-x)^{s_1-1}}{B(s_1, s_2)} dx,$$

where $\infty < \mu = Z'\beta < \infty$, $s_1 > 0$, $s_2 > 0$, β is a vector of coefficients and $B(s_1, s_2)$ is the beta function with parameters s_1 and s_2. The resulting density function of Y is

$$\tilde{f}_\theta(y|Z) = \frac{(s_1 exp(\frac{y_i - \mu}{\sigma})/s_2)^{s_1} (1 + s_1 exp(\frac{y_i - \mu}{\sigma})/s_2)^{-(s_1+s_2)}}{B(s_1, s_2)}.$$

Farewell(1982) suggested to use Newton-Raphson method to maximize the log-likelihood. It requires computing inverse matrix of the second partial derivatives evaluated at the maximum likelihood estimates. However for the generalized F distribution, it is impossible to derive the derivatives as closed-form expressions. This makes it difficult to apply the Newton-Raphson algorithm for solving the MLE. Peng, Dear and Denham (1998) suggested to combine a derivative-free maximization approach and the Newton-Raphson algorithm. Specifically they adopt the simulated annealing algorithm as the derivative-free maximization method to estimate the shape parameters. For fixed values of s_1 and s_2, the Newton-Raphson method is employed to maximize the likelihood function over σ, β, γ.

Chapter 3 Semi-parametric Regression Analysis - Likelihood Approach

In this chapter, we consider modeling $T|\zeta = 1, Z$ by semi-parametric regression models that contain un-specified functions as nuisance parameters.

3.1 Regression Models Without Cure

For illustration, we first introduce traditional regression models in absence of cure. The most well-known regression model in survival analysis is perhaps the proportional hazard model proposed by Cox (1972). Specifically given Z, the model be written as

$$\lambda(t|Z) = \lambda_0(t)\exp(Z'\beta), \tag{3.1a}$$

where $\beta : p \times 1$ is the unknown regression parameter of interest and $\lambda_0(t)$ is the baseline hazard function. Another expression of the Cox model is given by

$$S(t|Z) = S_0(t)^{exp(Z'\beta)} = \exp(-\int_0^t \lambda(s|Z)ds) = \exp(-\Lambda_0(t)\exp(Z'\beta)), \tag{3.1b}$$

where $\Lambda_0(t)$ is the cumulative hazard function and $S_0(t)$ is the survival function for the baseline group.

In recent years there is a trend to generalize different regression models under a unified framework. Notice that the Cox model can be written as

$$\log(-\log(S(t|Z))) = \log(\Lambda_0(t)) + Z'\beta. \tag{3.1c}$$

We see that the right-hand side of (3.1c) shows a linear structure in the parameters $log(\Lambda_0(t))$ and β. Similarly the proportional odds model can be written as

$$logit(S(t|Z)) = \log\frac{S(t|Z)}{1-S(t|Z)} = \log[\exp(Z'\beta)\frac{S_0(t|Z)}{1-S_0(t|Z)}].$$

It follows that

$$logit(S(t|Z)) = logit(S_0(t)) + Z'\beta. \tag{3.2}$$

The two different models in equations (3.1c) and (3.2) are special cases of the following transformation models:

$$\varphi(S(t|Z)) = h(t) + Z'\beta, \tag{3.3a}$$

where $h(\cdot)$ is an unknown increasing function. Another expression of the model is given by

$$h(t) = -Z'\beta + \varepsilon, \tag{3.3b}$$

where ε is the error term with a known continuous distribution. The relationship between equations (3.3a) and (3.3b) is specified by

$$F_\varepsilon(t) = \Pr(\varepsilon \leq t) = 1 - \varphi^{-1}(t)$$

since

$$\begin{aligned} F_\varepsilon(.) &= \Pr(\varepsilon \leq t) = \Pr(\varphi(S(T|Z)) \leq t) \\ &= \Pr(S(T|Z) \leq \varphi^{-1}(t)) = \Pr(1 - F(T|Z) \leq \varphi^{-1}(t)) \\ &= \Pr[(1 - \varphi^{-1}(t)) \leq F(T|Z)] = 1 - \varphi^{-1}(t). \end{aligned}$$

Transformation models form a general class of regression models that have attracted substantial attention in recent years.

3.2 Regression Models In Presence of Cure

In presence of cure, recall that $\zeta|Z$ is modeled by

$$p(\gamma) = \Pr(\zeta = 1|Z) = \frac{\exp(Z'\gamma)}{1 + \exp(Z'\gamma)}.$$

Here we consider to model $T|\zeta = 1, Z$ by a transformation model such that

$$\varphi(\tilde{S}(t|Z)) = h(t) + Z'\beta,$$

or equivalently

$$\tilde{S}(t|Z) = \varphi^{-1}(h(t) + Z'\beta).$$

Combining the two components, the survival function of T can be written as

$$\begin{aligned} S(t|Z) &= p(\gamma)\tilde{S}(t|Z) + (1 - p(\gamma)) \\ &= \frac{\exp(Z'\gamma)}{1 + \exp(Z'\gamma)}\tilde{S}(t|Z) + \frac{1}{1 + \exp(Z'\gamma)}. \end{aligned}$$

Notice that the cumulative hazard function of $T|\zeta = 1, Z$ can be written as

$$\tilde{\Lambda}(t|Z) = -\log[\tilde{S}(t|Z)] = -log[\varphi^{-1}(h(t) + Z'\beta)] = H(h(t) + Z'\beta).$$

We can write

$$S(t|Z) = \frac{\exp(Z'\gamma)\exp[-H(h(t) + Z'\beta)] + 1}{1 + \exp(Z'\gamma)} = \frac{\exp[Z'\gamma - H(h(t) + Z'\beta)] + 1}{1 + \exp(Z'\gamma)}.$$

Consider the example that $T|\zeta = 1, Z$ follows the Cox proportional hazard(ph) model with the hazard function and survival function

$$\tilde{\lambda}(t|Z) = \tilde{\lambda}_0(t)\exp(Z'\beta),$$
$$\tilde{S}(t|Z) = \Pr(T > t|\zeta = 1, Z) = \tilde{S}_0(t)^{\exp(Z'\beta)}.$$

Accordingly the density function and survival function of T given Z can be written as

$$\begin{aligned} f(t|Z) &= p(\gamma)\tilde{\lambda}(t|Z) \times \tilde{S}(t|Z), \\ &= p(\gamma)[\tilde{\lambda}_0(t)\exp(Z'\beta)] \times \tilde{S}_0(t)^{\exp(Z'\beta)}, \\ S(t|Z) &= \tilde{S}(t|Z)p(\gamma) + 1 - p(\gamma) \\ &= \exp[-\tilde{\Lambda}_0(t)\exp(Z'\beta)]p(\gamma) + 1 - p(\gamma). \end{aligned}$$

The major interest is to estimate β and γ while leaving $\tilde{\lambda}_0(t)$ unspecified.

3.3 Likelihood Representations

The general form of likelihood function can be written as

$$\prod_{i=1}^{n}[p_i(\gamma)\tilde{f}(x_i|z_i)]^{\delta_i}[\tilde{S}(x_i|z_i)p_i(\gamma) + 1 - p_i(\gamma)]^{1-\delta_i}, \tag{3.4a}$$

or equivalently

$$\prod_{i=1}^{n}[p_i(\gamma)\tilde{\lambda}(x_i|z_i)\tilde{S}(x_i|z_i)]^{\delta_i}[\tilde{S}(x_i|z_i)p_i(\gamma) + 1 - p_i(\gamma)]^{1-\delta_i}. \tag{3.4b}$$

The second expression is useful when the model is expressed in terms of the hazard function such as the Cox model.

The likelihood approach has been applied to the Cox PH model in presence of cure by several authors. We have seen that the likelihood function is already quite complicated for analyzing parametric regression models. Now the challenge is to handle the infinite-dimensional nuisance function $\tilde{\lambda}_0(t)$ in estimation. Specifically under the Cox model, we can write

$$\begin{aligned} L(\gamma, \beta, \lambda_0(.)) \;=\; & \prod_{i=1}^{n}[p_i(\gamma)\tilde{\lambda}_0(x_i)\exp(z_i'\beta)\tilde{S}_0(x_i)^{\exp(z_i'\beta)}]^{\delta_i} \\ & [\tilde{S}_0(x_i)^{\exp(z_i'\beta)}p_i(\gamma) + 1 - p_i(\gamma)]^{1-\delta_i}. \end{aligned}$$

Besides the complexity of the likelihood function, the major challenge is to deal with the baseline hazard function which is an infinite-dimensional nuisance parameter.

3.4 Principle of the EM Algorithm

To apply the EM algorithm, we first need to construct the likelihood function based on "complete data" $\{(x_i, \delta_i, z_i, \zeta_i), i = 1, \ldots, n\}$ assuming that ζ_i can be observed. Specifically the likelihood function can be written as

$$L_C = \prod_{i=1}^{n} [p_i(\gamma)\tilde{f}(x_i|z_i)]^{\delta_i\zeta_i}[\tilde{S}(x_i|z_i)p_i(\gamma)]^{(1-\delta_i)\zeta_i}[1-p_i(\gamma)]^{1-\zeta_i}.$$

It follows that

$$\begin{aligned} L_C &= \prod_{i=1}^{n}(p_i(\gamma)^{\zeta_i}(1-p_i(\gamma))^{1-\zeta_i} \prod_{i=1}^{n}(\tilde{f}(x_i|z_i)^{\delta_i\zeta_i}\tilde{S}(x_i|z_i)^{(1-\delta_i)\zeta_i}) \\ &= L_1 L_2, \end{aligned} \tag{3.5}$$

where the likelihood function is divided into two parts. Notice that the first part L_1 is a function of γ which is the parameter for the cure rate model and the second part involves on the parameters for the latency distribution. This implies that the two types of parameters can be dealt with separately.

Since the value of ζ_i may be missing due to censoring, in the E-step, we calculate its expected value given the observed data. It follows that

$$E[\zeta_i|x_i, \delta_i, z_i] = \delta_i + (1-\delta_i)\Pr(\zeta_i = 1|x_i, \delta_i = 0, z_i). \tag{3.6}$$

Furthermore

$$\begin{aligned} Pr(\zeta_i = 1|x_i, \delta_i = 0, z_i) &= Pr(\zeta_i = 1|T_i > x_i, z_i) \\ &= \frac{p_i(\gamma)\tilde{S}(x_i|z_i)}{p_i(\gamma)\tilde{S}(x_i|z_i) + 1 - p_i(\gamma)}. \end{aligned} \tag{3.7}$$

Detailed calculations of the above equation is given in Appendix 1. We see that equation (3.7) is a function of γ and $\tilde{S}(x|z)$ contains the nuisance function. In the M-step, we can maximize $logL_1$ with respect to γ and $logL_2$ with respect to the parameters in $\tilde{S}_0(.)$ by replacing ζ_i by an estimate of $E(\zeta_i|x_i, \delta_i = 1, z_i)$ which is treated as a fixed constant by plugging in the parameter estimates obtained previously. Numerical algorithms such as the Newton-Raphson method may be used in the maximization.

3.5 The EM Approach under the Cox Model

Now assume that $T|\zeta = 1, Z$ follows Cox PH model. The likelihood function is the product of the following two components:

$$L_1(\gamma) = \prod_{i=1}^{n}(p_i(\gamma)^{\zeta_i}(1-p_i(\gamma))^{1-\zeta_i} \tag{3.8}$$

and

$$L_2(\beta, \tilde{\lambda}_0(.)) = \prod_{i=1}^{n} [\tilde{\lambda}_0(x_i) \exp(z_i'\beta)]^{\delta_i \zeta_i} exp[-\int_0^{x_i} \tilde{\lambda}_0(t) \exp(z_i'\beta) dt]^{\zeta_i}. \tag{3.9}$$

Recall that ζ_i will be replaced by the estimated value of its conditional expectation. The maximization step of l_1 is straightforward. To handle the nuisance function $\tilde{\lambda}_0(.)$ in L_2, we introduce two approaches proposed by Sy and Taylor(2000) and Peng and Dear(2000).

3.5.1 Baseline estimation by Sy and Taylor

Actually two methods are proposed in the paper of Sy and Taylor (2000). The logarithm of equation (3.9) can be written as

$$l(\beta, \tilde{\lambda}_0(.)) = \sum_{i=1}^{n} \delta_i \zeta_i [\log \tilde{\lambda}_0(x_i) + z_i'\beta] + \sum_{i=1}^{n} \zeta_i [-\tilde{\Lambda}_0(x_i)\ exp(z_i'\beta)].$$

Assume that $\tilde{\Lambda}_0(t)$ only jumps at observed failure points. Hence we can write

$$\tilde{\Lambda}_0(t) = \sum_{i=1}^{n} I(t_j \leq t) \tilde{\Lambda}_0(\Delta t_j),$$

where t_j is the jth observed failure point. Differentiating the log-likelihood with respect to $\tilde{\Lambda}_0(\Delta t_j)$ for $j = 1, \ldots, k$ and setting the function as zero gives the following equation:

$$\hat{\Lambda}_0(\Delta t_j) = \frac{d_j}{\sum_{i \in R(t_j)} \zeta_i \exp(Z_i'\beta)},$$

where d_j is the number of failure at t_j and $R(t)$ is the risk set at time t. Accordingly

$$\hat{\Lambda}_0(t) = \sum_{j=1}^{n} I(t_j \leq t) (\frac{d_j}{\sum_{l \in R(t_j)} \zeta_l \exp(Z_l'\beta)}) \tag{3.10}$$

which can be viewed as a modified version of the Breslow estimator if the value of β is known. Substituting $\hat{\Lambda}_0(\Delta t_j)$ into $L_2(\beta, \tilde{\Lambda}_0(.))$ leads to the following partial likelihood of β:

$$L_3(\beta) = \prod_{i=1}^{n} (\frac{\exp(z_i'\beta)}{\sum_{l \in R(x_i)} \zeta_l exp(z_l'\beta)})^{\delta_i}. \tag{3.11}$$

We can find the estimators of $\Lambda_0(t)$ and β are similar to the estimators obtained without cure except that the weight ζ_l or its estimated conditional expectation is added to. The function $L_3(\beta)$ does not contain any other nuisance function so that β can be estimated easily using the traditional approach.

The above Breslow-type estimation separates the estimation of β and $\Lambda_0(\Delta t_j)$. Sy and Taylor (2000) proposed a different alternative based on joint estimation. Assume that baseline survival function takes jumps on observed failure points. Recall that

$$L_2 = \prod_{i=1}^{n} \tilde{f}(x_i|z_i)^{\delta_i \zeta_i} \tilde{S}(x_i|z_i)^{(1-\delta_i)\zeta_i}.$$

An important technique is to express L_2 based on ordered observed times $t_1 < \ldots < t_k$. Consider the interval $[t_j, t_{j+1})$. If $x_l \in [t_j, t_{j+1})$ and $\delta_l = 1$, the observation contributes the probability $\tilde{f}(x_l|z_l)$ to the likelihood for the latency distribution. If $x_l \in [t_j, t_{j+1})$ and $\delta_l = 0$, the observation contributes the probability $\tilde{S}(x_l|z_l)$ to the likelihood. Using the property that $\tilde{f}(x) = \tilde{\lambda}(x)\tilde{S}(x)$, we have

$$\begin{aligned} L_2 &= \prod_{j=1}^{k}\{\prod_{l\in D_j}(\tilde{f}(x_l|z_l)^{\zeta_l}\}\{\prod_{l\in C_j}(\tilde{S}(x_l|z_l)^{\zeta_l}\} \\ &= \prod_{j=1}^{k}\{\prod_{l\in D_j}(\tilde{\lambda}(x_l|z_l)^{\zeta_l}\}\{\prod_{l\in E_j}(\tilde{S}(x_l|z_l)^{\zeta_l}\}, \end{aligned}$$

where D_j is the set of subjects experiencing a failure event at time t_j and C_j is the set of subjects censored in $[t_j, t_{j+1})$ and E_j is the union of the two sets.

Define $\alpha_i = \Pr(T > t_i | T \geq t_i, Z = 0, \zeta = 1)$ which is the conditional survival probability at time t_i for the baseline susceptible subject. Then the baseline survival function $\tilde{S}_0(t)$ can be written as a product form of α_i. Specifically

$$\tilde{S}_0(t) = \prod_{j:t_j \leq t} \alpha_j.$$

Hence under the proportional hazard model $\tilde{S}(x_l|z_l) = \tilde{S}_0(x_l)^{\exp(z_l'\beta)}$. Also

$$1 - \tilde{\Lambda}(\Delta t_j|z) = \Pr(T > t_j | T \geq t_j, z, \zeta = 1) = (\alpha_j)^{\exp(z_l'\beta)}.$$

The function $L_2(\beta, \tilde{\Lambda}_0(.))$ can be written as

$$L_3(\beta, \alpha) = \prod_{j=1}^{k}[\prod_{l\in D_j}(1 - \alpha_j^{\exp(z_l'\beta)})^{\zeta_l} \prod_{l\in E_j} \alpha_j^{\zeta_l \exp(z_l'\beta)}]. \tag{3.12}$$

Treating β as a constant and assuming that data have no ties, the MLE of α_j can be written as

$$\hat{\alpha}_j = (1 - \frac{\exp(z_j'\beta)}{\sum_{l\in R(t_j)} \zeta_j \exp(z_l'\beta)})^{exp(-z_j'\beta)}. \tag{3.13}$$

Substituting $\hat{\alpha}_i$ into the equation (3.12), a nonparametric profile likelihood of β can be obtained and them maximized to get the MLE of β.

3.5.2 Baseline Estimation by Peng and Dear

Peng and Dear (2000) also suggested to use $L_3(\beta)$ for estimating β without dealing with the baseline hazard rate. For estimating the baseline survival function, Peng and Dear also adopted

the idea of Kalbfleisch and Prentice (1973) such that the survival function is expressed as the product of discrete hazard rates. The MLE of α_j given β has the form,

$$\hat{\alpha_j} \approx 1 - \frac{d_j}{\sum_{l \in R(t_j)} \zeta_l \exp(z_l \beta)}.$$

3.6 Likelihood for Transformation Models

The likelihood function for linear transformation models with cure can be written as

$$\begin{aligned} L(\gamma, \beta, h(.)) &= \prod_{i=1}^{n} [p_i(\gamma) \frac{\partial \varphi^{-1}[h(x) + z_i'\beta]}{\partial x}|_{x=x_i}]^{\delta_i} \\ &\quad \times [p_i(\gamma)\varphi^{-1}[h(x_i) + z_i'\beta] + (1 - p_i(\gamma))]^{1-\delta_i}. \end{aligned}$$

Note that for the Cox model, $h(\cdot)$ is related to the baseline hazard function which has clear properties that can be utilized to simplify the likelihood function. However for other transformation models, the role of $h(\cdot)$ is less clear. Therefore we will discuss methods constructed based on moment properties for transformation models.

Chapter 4 Semi-parametric Regression Analysis - Moment-based Approach

As discussed in Chapter 3, a linear transformation regression model without cure has two equivalent expressions

$$\begin{aligned} h(T) &= -Z'\beta + \varepsilon \\ \varphi(S(t|Z)) &= h(t) + Z'\beta, \end{aligned}$$

where $h(.)$ is an unknown increasing function and $\varphi(.)$ is a known decreasing function related to the distribution of ε such that

$$F_\varepsilon(t) = \Pr(\varepsilon \leq t) = 1 - \varphi^{-1}(t).$$

When cure exists, we will assume that $\zeta = 1|Z$ follows a logistic regression model and $T|\zeta = 1, Z$ follows a transformation model. We have

$$S(t|Z) = \frac{\exp(Z'\gamma)}{1+\exp(Z'\gamma)}\varphi^{-1}(h(t) + Z'\beta) + \frac{1}{1+\exp(Z'\gamma)}.$$

In this chapter, we discuss existing inference methods which utilize the moment principles.

4.1 Moment-based Inference: An Overview

The method of moment is attractive because it does not require strong distributional assumption and hence usually provides more robust results. Although the nonparametric likelihood approach is also robust, it is often very complicated for flexible models. Transformation models contain an un-specified function $h(\cdot)$ which complicates statistical inference.

Here we review some inference methods that are constructed using the idea of the method of moment. For example, denote O_i as a chosen response variable and $E_i(\theta) = E(O_i|Z_i)$ as its expected value, where θ is the parameter of interest. A moment-based estimating function has the form,

$$U(\theta) = \sum_{i=1}^{n} W_i(O_i - E_i(\theta)) = 0, \tag{4.1}$$

where W_i is a weight function for ith subject. How to choose a suitable response O_i is the key. If the response is well chosen, its expected value will be a nice function of the unknown parameters. For transformation models, we shall first pay attention to the form of E_i which may or may not involve the nuisance function $h(\cdot)$.

4.2 Response and its Expectation in Absence of Cure

For illustration, we first review exiting choices in absence of cure. There are three response variables. Specifically $I(T_i \geq T_j), i \neq j$ was proposed by Cheng et al.(1995), $I(T \geq t)$ by Cai et al.(2000) and $N(t) = I(T \leq t)$ by Chen et al.(2002). We first examine their moment properties.

4.2.1 Pairwise Order Indicator

Temporarily assume there is no censoring and the data can be denoted as $(T_i, Z_i), i = 1...n$. Since $h(.)$ is an unknown increasing function, it follows that

$$I(T_i \geq T_j) = I[h(T_i) \geq h(T_j)] = I(-Z_i'\beta + \varepsilon_i \geq -Z_j'\beta + \varepsilon_j) = I(\varepsilon_i - \varepsilon_j \geq Z_i'\beta - Z_j'\beta).$$

Then

$$E[I(T_i \geq T_j)|Z_i, Z_j] = E[I(\varepsilon_i - \varepsilon_j \geq Z_i'\beta - Z_j'\beta)] = \Pr(\varepsilon_i - \varepsilon_j \geq Z_{ij}'\beta) = \Phi(Z_{ij}'\beta),$$

where $Z_{ij} = Z_i - Z_j$, $\Phi(s) = \int_{-\infty}^{\infty}[1 - F_\varepsilon(t+s)]dF_\varepsilon(t)$ and $F_\varepsilon(.)$ is cumulative distribution function for ε. More detailed derivations are given in Appendix 2. We have seen that $h(\cdot)$ disappear in the moment expression.

When censoring is present, observed data can be denoted as $(X_i, \delta_i, Z_i), i = 1...n$. Cheng et al. (1995) suggested to use $I(X_i \geq X_j, \delta_j = 1)$ as a proxy of $I(T_i \geq T_j)$. Note that as long as the smaller observation corresponds to a failure point, the order relationship of the pair is known. However the proxy is biased since

$$E[I(X_i \geq X_j, \delta_j = 1)] = E[E[T_i \geq T_j, C_i \geq T_j, C_j \geq T_j|T_i, T_j]] = E[I(T_i \geq T_j)G^2(T_j)],$$

where $G(.)$ is the survival function of the censoring time. Weighting by the inverse-probability-of-censoring is often used to correct the bias and hence

$$E[\frac{\delta_j I(X_i \geq X_j)}{G^2(X_j)}|Z_i, Z_j] = E[I(T_i \geq T_j)|Z_i, Z_j] = \Phi(Z_{ij}'\beta).$$

In summary $\dfrac{\delta_j I(X_i \geq X_j)}{G^2(X_j)}$ is an unbiased proxy for $I(T_i \geq T_j)$.Detailed derivations are given in Appendix 3.

4.2.2 The At-risk Indicator

Consider $Y(t) = I(T \geq t)$ which indicates whether a subject is at risk. In absence of censoring, its expectation under the transformation model is given by

$$E(I(T \geq t)|Z) = \Pr(T \geq t|Z) = S(t|Z) = \varphi^{-1}[h(t) + Z'\beta].$$

Notice that $h(\cdot)$ still exists. When censoring is present, the corresponding response variable is $I(X \geq t)$. It follows that

$$E(I(X \geq t)|Z) = \Pr(X \geq t|Z) = \varphi^{-1}[h(t) + Z'\beta]G(t).$$

To utilize this variable for further inference, how to construct high-dimensional estimating functions is crucial since the expected involves both β and $h(\cdot)$.

4.2.3 The Counting Process

Define the counting process $N(t) = I(T \leq t)$ and $dN(t) = N(t) - N(t-) = I(T = t)$. The expected value of $dN(t)$ conditional on the filtration F_{t-} is

$$E[dN(t)|F_{t-}]I(T \geq t)\lambda(t)dt = Y(t)\lambda(t)dt = Y(t)d\Lambda(t).$$

Under the transformation model, we can write

$$\Lambda(t|Z) = -\log S(t|Z) = -\log\{\varphi^{-1}[h(t) + Z'\beta]\},$$
$$E(N(t)|F_{t-}) = \int_0^\infty Y(s)d[-\log\varphi^{-1}(h(s) + Z'\beta)].$$

In presence of censoring, the counting process is modified as $N(t) = I(X \leq t, \delta = 1)$ and the at-risk process as $Y(t) = I(X \geq t)$. The conditional expectation of $dN(t)$ given F_{t-} is similar

$$E(dN(t)|F_{t-}) = I(X \geq t)\lambda(t)dt = Y(t)d\Lambda(t).$$

Under the transformation model, we have

$$E(N(t)|F_{t-}) = \int_0^\infty Y(s)d[-\log\varphi^{-1}(h(s) + Z'\beta)].$$

which involves $h(\cdot)$ but does not contains $G(\cdot)$

4.3 Estimating Functions in Absence of Cure

Now we illustrate how to construct estimating functions based on the moment properties. Recall that there are two types parameters, namely β and $h(.)$. The former is of major interest.

4.3.1 Pairwise Order Indicator

Since the mean of $I(T_i \geq T_j)$ for $i \neq j$ does not depend on the nuisance parameter, we have the uncensored version

$$U(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} W(Z_{ij}'\beta)Z_{ij} \times \{I(T_i \geq T_j) - \Phi(Z_{ij}'\beta)\} = 0, \tag{4.2}$$

and the censored version

$$U(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} W(Z'_{ij}\beta)Z_{ij} \times \{\frac{\delta_j I(X_i \geq X_j)}{\hat{G}^2(X_j)} - \Phi(Z'_{ij}\beta)\} = 0, \tag{4.3}$$

where $\hat{G}(.)$ is the Kaplan-Meier estimator of $G(.)$. The solution of $U(\beta) = 0$ yields an estimator of β.

Although this approach is appealing since the nuisance parameter disappears, it has a crucial drawback when censoring is present. Notice that (4.3) has to be valid if $Pr(G(T) = 0) = 0$. This requirement however depends on the censoring mechanism. Define the following support points:

$$\begin{aligned}
\tau_C &= \sup_t t : G(t) = \Pr(C > t) > 0, \\
\tau_T &= \sup_t t : \Pr(T > t) > 0, \\
\tau_X &= \sup_t t : \Pr(X > t) > 0.
\end{aligned}$$

It follows that $\tau_X = min(\tau_T, \tau_C)$. Problem occurs when $\tau_C < \tau_T$ which implies the study period is too short to observe large failure time. In this case the mean condition does not hold and hence the estimating functions is no longer unbiased.

4.3.2 The At-risk Event

Without censoring and from the previous moment derivation, a natural estimating equation is given by

$$\sum_{i=1}^{n}(I(T_i \geq t) - \varphi^{-1}[h(t) + Z'_i\beta]) = 0, \tag{4.4}$$

for $t \in [\tau_a, \tau_b]$, where τ_a, τ_b are pre-specified constants such that $\Pr(T < \tau_a) > 0$ and $\Pr(T < \tau_b) > 0$. A set of equations for t being the observed values of T_i $(i = 1, \ldots, n)$ can be constructed which however are not enough since there are $n + p$ unknown parameters since β is $p-$dimensional. Cai et al.(2000) suggested the following additional estimating equations

$$\sum_{i=1}^{n}\{(I(T_i \geq t) - E[I(T_i \geq t|Z_i)])\} = 0,$$

$$\sum_{i=1}^{n}\int_{\tau_a}^{\tau_b} Z_i(I(T_i \geq t) - \varphi^{-1}[h(t) + Z'_i\beta])dw(t) = 0, \tag{4.5}$$

where $w(.)$ is a known increasing weight function which may be time-dependent. By choosing an initial value of β, $h(T_i)$ $(i = 1, \ldots, n)$ are estimated first and then plugged into equation (4.5) to estimate β. The two types of equations are performed iteratively until the convergence criteria reached.

For right censoring data, we can construct two types of estimating equations:

$$\sum_{1=1}^{n}(I(X_i \geq t) - \varphi^{-1}[h(t) + Z_i'\beta]\hat{G}(t)) = 0, \tag{4.6}$$

and

$$\sum_{i=1}^{n}\int_{\tau_a}^{\tau_b} Z_i\{I(X_i \geq t) - \varphi^{-1}[h(t) + Z_i'\beta]\hat{G}(t)\}dw(t) = 0. \tag{4.7}$$

Numerical operation is same as mentioned above.

4.3.3 The Counting Process

Let $(\beta_0, h_0(.))$ be the true values of $(\beta, h(.))$. It is easy to see that

$$M(t) = N(t) - \int_0^{\infty} Y(s)d[-\log\varphi^{-1}(h_0(s) + Z'\beta_0)$$

is a mean-zero martingale. This property can be used to construct estimating functions. The following is the proposal of Chen, Jin and Ying (2002):

$$\sum_{i=1}^{n} dN_i(t) - Y_i(t)d[-\log\varphi^{-1}(h(t) + Z_i'\beta)] = 0, \tag{4.8}$$

$$\sum_{i=1}^{n}\int_0^{\infty} Z_i[dN_i(t) - Y_i(t)d[-\log\varphi^{-1}(h(t) + Z_i'\beta)]] = 0. \tag{4.9}$$

The first equation is for estimating $h(.)$ at observed failure points and the second equation is for estimating β. The expressions of the estimating equations are similar when censoring is present.

4.4 Semi-parametric Transformation Cure Model

4.4.1 Estimating Functions

Semi-parametric inference for transformation models in presence of cure has been considered by Lu and Ying (2004) who constructed the estimating equations based on the counting process $N(t)$ mentioned above. In presence of cure, we have to modify

$$E(dN(t)|F_{t-}) = I(X \geq t, \delta = 1)\Pr(T = t|T \geq t) = I(X \geq t, \delta = 1)d\Lambda(t|Z)$$

, where $\Lambda(t|Z)$ is the cumulative hazard function of T. Now we first derive the general form of $\Lambda(t|Z)$ based on the model definitions and then discuss how to transform it into a more

tractable form. Recall that $\Lambda(t|Z) = -\log \Pr(T > t|Z)$ and

$$\begin{aligned}\Pr(T > t) &= \Pr(\zeta = 0|Z) + \Pr(\zeta = 1|Z) \Pr(T \geq t|\zeta = 1, Z) \\ &= \frac{1}{1 + exp(Z'\gamma)} + \frac{exp(Z'\gamma)}{1 + exp(Z'\gamma)} \tilde{S}(t|Z).\end{aligned}$$

To create a linear pattern in unknown parameters, additional variable transformation is suggested. It follows that

$$\tilde{S}(t|Z) = exp[-\tilde{\Lambda}(t|Z)] = \varphi^{-1}(h(t) + Z'\beta),$$

where $F_\varepsilon(t) = \Pr(\varepsilon \leq t) = 1 - \varphi^{-1}(t)$ and $\tilde{\Lambda}(t|Z)$ is the cumulative hazard function for susceptible ones. Hence we let

$$\tilde{\Lambda}(t|Z)] = -\log\{\varphi^{-1}(h(t) + Z'\beta)\} = H(h(t) + Z'\beta).$$

Thus we can write

$$\begin{aligned}\Pr(T > t) &= \frac{1}{1 + exp(Z'\gamma)} + \frac{exp(Z'\gamma)}{1 + exp(Z'\gamma)} \tilde{S}(t|Z) \\ &= \frac{1 + \exp(Z'\gamma - H(h(t) + Z'\beta))}{1 + exp(Z'\gamma)} \\ &= \frac{1}{1 + exp(Z'\gamma)} \frac{1 + \exp(Z'\gamma - H(h(t) + Z'\beta))}{1} \\ &= \bar{\psi}(Z'\gamma) \frac{1}{\bar{\psi}\{Z'\gamma - H[h(t) + Z'\beta]\}},\end{aligned}$$

where $\psi(x) = \exp(x)/\{1 + \exp(x)\}$ is the logistic function and $\bar{\psi} = 1 - \psi$. Using the fact that $\psi(-x) = \bar{\psi}(x)$. We have

$$\Lambda(t|Z) = -\log \Pr(T > t|Z) = -\log\{\bar{\psi}(Z'\gamma)\} + \log(\psi\{-Z'\gamma + H[h(t) + Z'\beta]\}). \tag{4.10}$$

It follows that $M(t) = dN(t) - Y(t)d\log\psi\{-Z'\gamma + H[h(t) + Z'\beta]\}$ is a martingale when the parameters are evaluated at their true values. The suggested estimating functions have the forms

$$\sum_{i=1}^{n} dN_i(t) - Y_i(t)dlog[\psi(-Z_i'\gamma + H(h(t) + Z_i'\beta))] = 0, \tag{4.11}$$

and

$$\sum_{i=1}^{n} \int_0^\infty Z_i \left\{dN_i(t) - Y_i(t)dlog(\psi(-Z_i'\gamma + H(h(t) + Z_i'\beta)))\right\}. \tag{4.12}$$

Note that equation (4.11) is used to estimate transformation function $h(.)$ given β and γ eveluated at observed failure points of $h(.)$ while equation (4.12) is for estimating β given $h(.)$ and γ.

It should be noted that in presence of cure, we have additional unknown parameters γ but the above estimating functions are not sufficient to estimate the parameters in the logistic model. In Section 3.5, we have presented the log-likelihood function of γ given ζ is known.It follows that

$$\frac{\partial l(\gamma)}{\partial \gamma} = \sum_{i=1}^{n} Z_i[\zeta_i - \frac{exp(Z_i'\gamma)}{1 + exp(Z_i'\gamma)}] = 0. \tag{4.13}$$

The imputation method is suggested to handle possible missingness of ζ. Specifically we impute ζ by an estimator of $E(\zeta|X, \delta, Z)$. The conditional probability that a subject belongs to the susceptible group given that the subject with covariate Z is censored at X is $\bar{\psi}\{-Z'\gamma + H[h(t) + Z'\beta]\}$. So we get

$$E(\zeta|X, \delta, Z) = \delta + (1 - \delta)\bar{\psi}\{-Z'\gamma + H[h(t) + Z'\beta]\}. \tag{4.14}$$

The equation (4.13) can be modified as

$$\sum_{i=1}^{n} Z_i\{\delta + (1 - \delta)\bar{\psi}\{-Z'\gamma + H[h(t) + Z'\beta]\}] - \psi(Z_i'\gamma)\} = 0. \tag{4.15}$$

4.4.2 Estimation Algorithm

Solving (4.11), (4.12) and (4.15) jointly is not an easy task. Lu et al.(2004) proposed an iterative approach to obtain the solution. Recall that $t_1 < ... < t_k$ be ordered failure points and $t_0 = 0, \quad t_{k+1} = \infty$. Let d_i be the number of failure events at time t_i. Denote the initial value of β and γ as $\hat{\beta}^{(0)}$ and $\hat{\gamma}^{(0)}$ respectively which may be chosen as $\hat{\gamma}^{(0)} = 0$ and $\hat{\beta}^{(0)}$ being the maximum partial likelihood estimator under the ordinary Cox model as suggested by Lu et al.(2004). Denote $\hat{\beta}^{(m)}$, $\hat{\gamma}^{(m)}$ and $\hat{h}^{(m)}(t)$ are the mth estimated values. Let $\hat{h}(\cdot)$ be a step function taking jump at t_i.

Step 1: This step involves solving of $\hat{h}(t_j)$ for $j = 1, ..., k$ based on equation (4.11). With $\hat{\beta}^{(m-1)}$ and $\hat{\gamma}^{(m-1)}$, we obtain $\hat{h}^{(m)}(t_k)$ by solving

$$\sum_{i=1}^{n} dN_i(t_k) - Y_i(t_k)dlog[\psi(H(\hat{h}^{(m)}(t_k) + Z_i'\hat{\beta}^{(m-1)}) - Z_i'\hat{\gamma}^{(m-1)}] = 0.$$

Step 2: Then based on equation (4.12), we update $\hat{\beta}^{(m)}$ by plugging $\hat{h}^{(m)}(t)$ and $\hat{\gamma}^{(m-1)}$ into the equation:

$$\sum_{i=1}^{n} \int_0^{\infty} Z_i dN_i(t_k) - Y_i(t_k) dlog[\psi(H(\hat{h}^{(m)}(t_k) + Z_i'\beta) - Z_i'\hat{\gamma}^{(m-1)})] = 0.$$

Step 3: Update $\hat{\gamma}^{(m)}$ based on the equation (4.15) such that

$$\sum_{i=1}^{n} Z_i\{\delta_i + (1-\delta_i)\bar{\psi}\{-Z_i'\hat{\gamma}^{(m-1)} + H[\hat{h}^{(m)}(t_k) + Z_i'\hat{\beta}^{(m)}]\} - \psi(Z_i'\gamma)\} = 0.$$

The procedure is implemented iteratively between the step 1 to step 3 until the convergence criteria is reached.

Chapter 5 Concluding Remarks

In this thesis, we review literature on cure mixture models. The likelihood can be simplified by assuming complete data, which contains the indicator of susceptibility, are available. Then the EM algorithm can be applied. When the model assumption on the latency distribution becomes more flexible, unknown parameters often involve infinite-dimensional functions. Likelihood-based inference is usually constructed by expressing these functions as step-functions and hence the jump size at different points becomes the target in estimation. If the Cox model is assumed, existing partial likelihood method can be modified to account for the cure. However for transformation models, the idea of method-of-moment is more attractive since the estimation procedure is more simple. Three forms of estimating functions are examined. The one based on the counting process has been extended to cure mixture models by Lu and Ying (2004). Despite its simplicity, efficiency of moment-type methods requires further investigation.

References

[1] Farewell, V. T. (1982), The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, pp.1041-1046.

[2] Lawless, J. F. (1982), Statistical models and methods for lifetime data. *Wiley:New York.*

[3] Farewell, V. T. (1986), Mixture models in survival analysis: are they worth the risk? *Can. J. Statist*, **14**, pp.257-262.

[4] Fleming, T. R. and Harrington, D. P. (1991), Counting processes and survival analysis. *Wiley:New York.*

[5] Kuk, A. Y. C. and Chen, C. H. (1992), A mixture model combining logistic regression with proportional hazards regression. *Biometrika*, **79**, pp. 531-541.

[6] Maller, R.A. and Zhou, S. (1992), Estimating the proportion of immunes in a censored sample. *Biometrika*, **79**, pp. 731-739.

[7] Chen, S. C., Wei, L. J. and Ying, Z. (1995), Analysis of transformation models with censored data. *Biometrika*,**82, 4**, pp. 835-845.

[8] Maller, R.A. and Zhou, S. (1996), Survival analysis with long-term survivors. *Wiley:New York.*

[9] Peng, Y. W., Keith B. G. Dear and Denham, J.W. (1998), A generalized F mixture model for cure rate estimation. *Statistics in Medicine*,**17**, pp. 813-830.

[10] Sy, J. P. and Taylor, J. M. G. (2000), Estimation in a Cox proportional hazards cure model. *Biometrics*,**56**, pp. 227-236.

[11] Cai, T., Wei, L. J. and Wilcox, M. (2000), Semiparametric regression analysis for clustered failure time data. *Biometrika*,**87**, 4, pp. 867-878.

[12] Chen, K., Jin, Z. and Ying, Z. (2002), Semiparametric analysis of transformation models with censored data. *Biometrika*,**89, 3**, pp. 659-668.

[13] Lu, W. and Ying, Z. (2004), On semiparametric transformation cure models. *Biometrics*,**91, 2**, pp. 331-343.

Appendix

A1. Derivation of $E(\zeta_i|(x_i, \delta_i = 0, z_i))$

$$\begin{aligned}
E(\zeta_i|(x_i, \delta_i = 0, z_i)) &= Pr(\zeta_i = 1|T_i > x_i, z_i) \\
&= \frac{Pr(\zeta_i = 1, T_i > x_i|z_i)}{Pr(T_i > x_i|z_i)} \\
&= \frac{Pr(\zeta_i = 1|z_i)Pr(T_i > x_i|\zeta_i = 1, z_i)}{Pr(\zeta_i = 1|z_i)Pr(T_i > x_i|\zeta_i = 1, z_i) + Pr(\zeta_i = 0|z_i)Pr(\infty = T_i > x_i|\zeta_i = 1, z_i)} \\
&= \frac{Pr(\zeta_i = 1|z_i)Pr(T_i > x_i|\zeta_i = 1, z_i)}{Pr(\zeta_i = 1|z_i)Pr(T_i > x_i|\zeta_i = 1, z_i) + Pr(\zeta_i = 0|z_i)} \\
&= \frac{\pi(z_i)\tilde{S}(x_i|z_i)}{\pi(z_i)\tilde{S}(x_i|z_i) + 1 - \pi(z_i)}.
\end{aligned}$$

A2. Derivation of $\Phi(s) = Pr(\varepsilon_i - \varepsilon_j \geq s)$

Let $V = \varepsilon_i - \varepsilon_j$, $\varepsilon \overset{i.i.d}{\sim} f_\varepsilon$,

$$\begin{aligned}
\Phi(s) = Pr(\varepsilon_i - \varepsilon_j \geq s) &= \int_s^\infty \int_{-\infty}^\infty f_{\varepsilon_i,\varepsilon_j}(v + \varepsilon_j, \varepsilon_j) d\varepsilon_j dv = \int_s^\infty \int_{-\infty}^\infty f_\varepsilon(v + \varepsilon_j) f_\varepsilon(\varepsilon_j) d\varepsilon_j dv \\
&= \int_s^\infty \int_{-\infty}^\infty f_\varepsilon(v + \varepsilon_j) dF_\varepsilon(\varepsilon_j) dv = \int_{-\infty}^\infty \int_s^\infty f_\varepsilon(v + \varepsilon_j) dv dF_\varepsilon(\varepsilon_j) \\
&= \int_{-\infty}^\infty (1 - \int_{-\infty}^s f_\varepsilon(v + \varepsilon_j) dv) dF_\varepsilon(\varepsilon_j) \\
&= \int_{-\infty}^\infty (1 - F_\varepsilon(s + \varepsilon_j)) dF_\varepsilon(\varepsilon_j) \\
&= \int_{-\infty}^\infty (1 - F_\varepsilon(s + t)) dF_\varepsilon(t).
\end{aligned}$$

A3. Derivation of $E[\frac{\delta_i I(X_i \geq X_j)}{G^2(X_j)}|Z_i, Z_j]$

$$\begin{aligned}
E[\frac{\delta_i I(X_i \geq X_j)}{G^2(X_j)}|Z_i, Z_j] &= E[\frac{I(C_j \geq T_j)I(X_i \geq X_j)}{G^2(X_j)}|Z_i, Z_j] \\
&= EE[\frac{I(min(C_j, X_i) \geq T_j)}{G^2(T_j)}|Z_i, Z_j, T_j] \\
&= EE[\frac{I(min(C_j, C_i) \geq T_j)I(T_i \geq T_j)}{G^2(T_j)}|Z_i, Z_j, T_j] \\
&= EE[\frac{I(min(C_j, C_i) \geq T_j)}{G^2(T_j)}E[I(T_i \geq T_j)]|Z_i, Z_j, T_j] \\
&= E[E[I(T_i \geq T_j)]|Z_i, Z_j, T_j] \\
&= E[I(T_i \geq T_j)|Z_i, Z_j] = \Phi(Z'_{ij}\beta).
\end{aligned}$$