Survival Analysis Based on the Information Criterion EIC

Toru Kaise

Doctor of Philosophy

Department of Statistical Science
School of Mathematical and Physical Science
The Graduate University for Advanced Studies

Abstract

In this thesis applications of the information criterion EIC for survival analysis are proposed. It is shown that our EIC procedures based on model-based bootstraps make it possible to compare goodnesses of nonparametric models with those of parametric models. It is also shown that the EIC procedures are effective to search for factors affecting survival probabilities.

The EIC procedures are explored as follows: firstly, a linearly interpolated Kaplan-Meier model, Weibull and gamma models are handled. Fits of models to survival data are evaluated by using EIC. Factors affecting survival probabilities are also evaluated by using EIC. Secondly, a linearly interpolated Cox survival probability model and accelerated models based on Weibull and gamma distributions are handled. These models are used to search for multiple factors affecting survival probabilities. The goodnesses of these models are evaluated by using EIC. Finally, we demonstrate the use of the EIC procedures for survival analysis of surgical outcomes in cancer patients. This example shows that our EIC procedures are effective for analyzing survival data.

Acknowledgements

I would like to express my sincerest appreciation to my advisor Professor Makio Ishiguro. During the past four years, he has been a model for me to copy behavior and attitude. He introduced me to see "real-world" with statistical eyes. Without his insights and criticisms, I would not have been able to continue and finish this thesis.

I would like to thank M.D., Ph.D. Toshiki Matsubara of Cancer Institute Hospital for his helpful comments as a surgical expert, and for offering the survival data. I would like to thank M.D., Ph.D. Akifumi Yafune of Kitasato Institute & University of Tokyo for his advices and discussions. I would like to thank Professor Genshiro Kitagawa and Professor Toshiya Sato for their valuable comments and brilliant seminars. I would like to thank Professor Toshinari Kamakura of Chuo University for his advices and discussions.

Professor Yoshihiko Kawazoe of Saitama Institute of Technology, and Professor Michio Horigome of Tokyo University of Marchantil Marine introduced me into the Graduate University for Advanced Studies. I would like to thank them for their brilliant advices.

I would like to thank my parents for their financial support. Finally, I would like

to express my dearest appreciation to my wife Yayoi.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Survival analysis is used to estimate survival probabilities and reliabilities in many fields of medical, engineering and economic researches. For example surgical outcomes are evaluated by survival probabilities estimated from survival data after operations. Therefore studies and developments for the survival analysis are widely required. In this thesis, the survival analysis based on the information criterion EIC (Ishiguro et al., 1991, 1994) is proposed. Particularly the survival analysis of surgical outcomes in cancer is handled. In biomedical science, many authors, including Kalbfleisch and Prentice (1980), Miller (1981), Lawless (1982), Cox and Oakes (1984), Lee (1992) and Collet (1994) have written fundamental books of the survival analysis. Many papers in survival analysis have been also given. However, studies and developments of method for survival analysis based on the information criterion have not yet been discussed.

A basic method of survival analysis is to estimate survival probabilities based on survival probability models. The survival probability models are classified into simple models and conditional models of covariates. Each survival probability model

belongs to either a nonparametric model or a parametric model.

Kaplan-Meier (KM) method was proposed for estimating survival probability function nonparametrically (Kaplan and Meier, 1958). It is obtained by maximizing a nonparametric likelihood function under existence of censored data (Kaplan and Meier, 1958, Johansen, 1978). The KM model is fundamental, and is widely used for survival analysis in biomedical science. This nonparametric model is convenient to analyze the survival data, because this model is a distribution free model. On the other hand, some parametric models are applicable to the survival analysis. In particular reliability analysis often employs the parametric models. Weibull distribution is widely used as a model of survival probability for electrical equipments (Nelson, 1982). Gamma distribution is also employed as a survival probability model. The estimate of the 50 percentile point, for example, is dependent on the choice of a model to approximate true survival probability. In this case, the comparisons between nonparametric and parametric approaches are important. For example, Miller (1983) showed that the maximum likelihood estimator of parametric model is more asymptotically efficient rather than the estimator of KM method for large sample size. Cantor (1992) showed the comparison between the parametric and nonparametric models for estimating cure rate based on paediatric cancer data.

Survival probabilities are influenced by conditions of individuals. Therefore the relations between the survival probabilities and the conditions are analyzed. In those cases, survival data are categorized according to conditions of individuals. Each set of categorized data is analyzed to estimate the survival probability. Nonparametric

two-sample tests are used to search for differences between survival probabilities for data divided into groups of categories. The generalized Wilcoxon test proposed by Gehan (1965) and the log-rank test proposed by Peto and Peto (1972) are widely used for the search. Matsubara et al. (1994) proposed the use of Akaike information criterion AIC (Akaike, 1973) for comparisons of the Weibull models fitted to strata of cancer invasions. The AIC procedures are convenient for estimating relations between conditions and survival probabilities. However it is not clear whether AIC is reliable for comparing models when sample size is relatively small. AIC is applicable only to parametric models and it does not evaluate nonparametric models.

An approach to the survival analysis with covariates, Cox (1972) proposed a proportional hazard model (Cox model). The covariates have quantities about the conditions of individuals. The Cox model has the linear regression for the covariates, and the parameters of the regression model are estimated by maximizing a partial or marginal likehood function. The survival probability model derived from the Cox model is nonparametric. This model is called the Cox survival probability model. Now, the Cox model is one of the most frequently used tools for the survival analysis in biomedical science (Robin, 1995). On the other hand, regression survivor models based on parametric distributions, which are called accelerated models, can be applied to the survival analysis. Weibull and gamma accelerated models are often used for the survival analysis. However, the following are required for the analysis of survival data by using the regression survivor models: firstly, which of the survival probability models, the nonparametric Cox survival probability model

or the parametric accelerated models, is better. Secondly, what models for the covariates is good. AIC can be used to choose a better model for the parametric accelerated models, however in the Cox model, AIC is available for choosing only the regression model for the covariates (Collet, 1994). Therefore AIC can not make the comparison between the parametric accelerated models and the Cox survival probability model.

In this thesis we propose applications of the information criterion EIC based on bootstrap procedures for the survival analysis. EIC was proposed to enlarge the range of the use of the information criterion in statistical analysis. We apply EIC for not only to the parametric survival probability models but also to the nonparametric models. Then to enable the application of EIC, we use a linearly interpolated KM (LIKM) model which is derived from the KM model, and a model-based bootstrap method instead of a bootstrap method for survival data given by Efron (1981). It is found that the EIC procedure makes it possible to compare the goodness of the LIKM model with those of the Weibull and gamma models, and the EIC procedure is more suitable for choice of model fitted to data of small sample size than AIC. The EIC procedure is used to compare the goodness of the survival probability models fitted to strata of categories, then the results by EIC give the insight into the relations between the conditions and the survival probabilities. We propose also the use of EIC to compare the Cox survival probability model with parametric accelerated models and to search for the factors affecting the survival probabilities. In this case, a linearly interpalated Cox (LIC) survival probability

model is used. The LIC survival probability model is derived from the Cox survival probability model for applying EIC. The model-based bootstrap method is used for the EIC procedures of the regression survivor models. The EIC procedure for the LIC survival probability model is the extension of the method proposed for the LIKM model.

In Chapter 2 the basic definitions of the survival probability models and estimation procedures, including the LIKM model derived from the KM model, the LIC survivor model derived from the Cox survivor model, the Weibull and gamma distributions, the accelerated models, are described.

In Chapter 3 we review the derivation of information criteria. The model-based bootstrap methods are given. The applications of EIC to the LIKM and LIC models are proposed for analyzing the survival data based on the model-based bootstrap methods.

In Chapter 4 it is shown through simulation studies that EIC is more suitable for survival data of small sample size than AIC. The comparisons of the parametric models and the LIKM models by EIC are shown through the simulation studies. Furthermore the comparisons of the parametric accelerated models and the LIC models by EIC are also given.

In Chapter 5 surgical outcomes in cancer of the thoracic esophagus are analyzed based on our EIC procedures. The observations of patient are the survival time after the surgical operation and the status of invasions at the surgical operation. The EIC procedures are used to search for the factors affecting the surgical outcomes based on

the parametric and nonparametric models. We get the best model for the survival data based on the results by EIC.

Chapter 2

Survival Analysis

This chapter introduces notations and basics of survival analysis. Firstly we describe types of survival data. Secondly survival probability models for simple survival data are described. Thirdly survival probability models for survival data with the multivariate covariates are described. Linearly interpolated nonparametric models are also provided.

2.1 Survival data

In medical researches, survival times of patients are defined as their times between initial points and the deaths. Censoring times are times observed before the deaths. Covariates are defined as quantities about the conditions of individuals. In this thesis, the survival time of ith patient is denoted by t_i, and the covariate vector is denoted by $\boldsymbol{z}_i$. We consider two types of data set, the simple survival data set $x_i = \{t_i, \delta_i\}$ and the survival data set with covariate $x_i = \{t_i, \delta_i, \boldsymbol{z}_i\}$, where $\delta_i = 1$ if the ith patient is dead or $\delta_i = 0$ otherwise. The sample of size n is denoted by vector $\boldsymbol{x} = (x_1, \ldots, x_n)$.

In many cases, the survival analysis deals with the censored data which are classified into three types, type I (fixed time censored data), type II (fixed number censored data) and type III (randomly censored data). In particular, random censoring arises in medical applications with animal studies or clinical trials (Miller, 1981). In this thesis we assume that the survival data can be treated as the randomly censored data.

2.2 Survival analysis of simple survival data

2.2.1 Survival probability

Let T be a positive random variable of survival time with the probability density function $f(t)$. Let $S(t)$ denotes the survival probability function,

$$S(t) = Pr\{t \leq T\} = \int_t^\infty f(x)dx. \tag{2.1}$$

The hazard rate function $h(t)$ is defined as

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{Pr\{t \leq T < t + \Delta t | t \leq T\}}{\Delta t} \\ &= \frac{f(t)}{S(t)}, \end{aligned} \tag{2.2}$$

which often has one of three properties, IHR (increasing hazard rate), DHR (decreasing hazard rate) or CHR (constant hazard rate). The complicated hazard rates other than these patterns are considered. $S(t)$ is also written as follows:

$$\begin{aligned} h(t) = \frac{-S(t)'}{S(t)} &= -\frac{d \log S(t)}{dt}, \\ \log S(t) &= -\int_0^t h(u)du, \\ S(t) &= \exp[-\int_0^t h(u)du]. \end{aligned} \tag{2.3}$$

2.2.2 Models and estimation procedures

According to Kaplan and Meier (1958), the nonparametric likelihood function of the survival probability model for the survival data $\boldsymbol{x}$ is given by

$$L = \prod_{i=0}^{k} p(t_{(i)})^{d_i} S(t_{(i)})^{\lambda_i}, \tag{2.4}$$

where k is the number of knot points, $t_{(i)}$ is the ordered failure time $(0 = t_{(0)} < \ldots < t_{(i-1)} < t_{(i)} < \ldots < t_{(k+1)} = \infty)$, $p(t_{(i)}) = S(t_{(i-1)}) - S(t_{(i)})$, $S(t_{(0)}) = 1$, $p(t_{(0)}) = 0$, d_i is the number of the failures at $t_{(i)}$, $d_0 = 0$, λ_i is the number of the censored data in $[t_{(i)}, t_{(i+1)})$, and the sample size n satisfies $n = \sum_{j=0}^{k}(d_j + \lambda_j)$. In (2.4), if $S(t_{(k)}) > 0$, i.e., $\lambda_k \neq 0$, then the number of free parameters is k because $\sum_{j=1}^{k+1} p(t_{(j)}) = 1$. if $S(t_{(k)}) = 0$, i.e., $\lambda_k = 0$, then the number of free parameters is $k-1$ because $\sum_{j=1}^{k} p(t_{(j)}) = 1$. By maximizing (2.4), the KM estimate $\hat{S}(t)$ for the survival probability at t is obtained as

$$\hat{S}(t) = \prod_{t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right), \tag{2.5}$$

where r_i is a risk set at $t_{(i)}$, i.e., the number of the survivors just before $t_{(i)}$.

Proof. Setting

$$s_j = S(t_{(j)})/S(t_{(j-1)}), \;\; q_j = 1 - s_j, \quad j = 1, 2, \ldots, k, \;\; s_0 = 1, \;\; , d_0 = 0, \;\; , \lambda_0 = 0,$$

then

$$\begin{aligned} s_1 s_2 \cdots s_{j-1} q_j &= \frac{S(t_{(1)})}{S(t_{(0)})} \frac{S(t_{(2)})}{S(t_{(1)})} \cdots \frac{S(t_{(j-1)})}{S(t_{(j-2)})} \left(1 - \frac{S(t_{(j)})}{S(t_{(j-1)})}\right) \\ &= S(t_{(j-1)}) - S(t_{(j)}) \\ &= p(t_{(j)}), \end{aligned}$$

furthermore

$$s_1 s_2 \cdots s_j \;=\; S(t_{(j)}).$$

Therefore the nonparametric likelihood function (2.4) is given by

$$\begin{aligned} L &= \prod_{j=0}^{k} (s_1 s_2 \cdots s_{j-1} q_j)^{d_j} (s_1 s_2 \cdots s_j)^{\lambda_j} \\ &= \prod_{j=0}^{k} q_j^{d_j} (s_1 s_2 \cdots s_{j-1})^{d_j+\lambda_j} s_j^{\lambda_j}. \end{aligned}$$

Here

$$r_j - r_{j+1} = \lambda_j + d_j,$$

$$\lambda_j = r_j - r_{j+1} - d_j.$$

Therefore L is given by

$$\begin{aligned} L &= \prod_{j=0}^{k} q_j^{d_j} (s_1 s_2 \cdots s_{j-1})^{d_j+\lambda_j} s_j^{r_j - r_{j+1} - d_j} \\ &= \prod_{j=0}^{k} q_j^{d_j} s_j^{r_j - d_j} s_j^{-r_{j+1}} (s_1 s_2 \cdots s_{j-1})^{d_j+\lambda_j}. \end{aligned}$$

Setting

$$\begin{aligned} L_p &= \prod_{j=0}^{k} s_j^{-r_{j+1}} (s_1 s_2 \cdots s_{j-1})^{d_j+\lambda_j} \\ &= [s_0^{-r_1} s_0^{d_1+\lambda_1}][s_1^{-r_2} s_1^{\sum_{j=2}^{k}(d_j+\lambda_j)}][s_2^{-r_3} s_2^{\sum_{j=3}^{k}(d_j+\lambda_j)}] \cdots [s_k^{-r_{k+1}}], \end{aligned}$$

here

$$s_0 = 1, \quad r_j \;=\; \sum_{l=j}^{k} d_l + \lambda_l, \quad r_{k+1} = 0,$$

then $L_p = 1$. Thus

$$L = \prod_{j=0}^{k} q_j^{d_j} s_j^{r_j - d_j}.$$

Solving equations for $j = 0, \ldots, k$,

$$\begin{aligned}
&\frac{\partial}{\partial s_j}(q_j^{d_j} s_j^{r_j-d_j}) = 0, \\
&\frac{\partial}{\partial s_j}\{(1-s_j)^{d_j} s_j^{r_j-d_j}\} = 0, \\
&-d_j(1-s_j)^{d_j-1} s_j^{r_j-d_j} + (1-s_j)^{d_j}(r_j-d_j)s_j^{r_j-d_j-1} = 0, \\
&-d_j(1-s_j)^{-1} + (r_j-d_j)s_j^{-1} = 0, \\
&s_j = \frac{r_j-d_j}{r_j}.
\end{aligned}$$

Thus we get

$$\hat{S}(t_{(i)}) \;=\; \prod_{j=1}^{i} s_j = \prod_{j=1}^{i}\left(1 - \frac{d_j}{r_j}\right).$$

■

(2.5) defines the corresponding survival probability at $t_{(i-1)} \leq t < t_{(i)}$. $\hat{p}(t_{(i)})$ is given by

$$\begin{aligned}
\hat{p}(t_{(i)}) &= \hat{S}(t_{(i-1)}) - \hat{S}(t_{(i)}) \\
&= \frac{d_i}{n}\prod_{j=0}^{i-1}(1 + \frac{\lambda_j}{r_{j+1}}).
\end{aligned} \tag{2.6}$$

Proof.

$$\hat{p}(t_{(i)}) = \hat{S}(t_{(i-1)})(1 - \frac{\hat{S}(t_{(i)})}{\hat{S}(t_{(i-1)})}).$$

Here

$$\frac{\hat{S}(t_{(i)})}{\hat{S}(t_{(i-1)})} = 1 - \frac{d_i}{r_i}.$$

Therefore

$$\begin{aligned}
\hat{p}(t_{(i)}) &= \hat{S}(t_{(i-1)})\frac{d_i}{r_i} \\
&= \frac{d_i}{r_i}\prod_{j=0}^{i-1}\left(1-\frac{d_j}{r_j}\right) \\
&= \frac{d_i}{r_i}\left(1-\frac{d_0}{r_0}\right)\left(1-\frac{d_1}{r_1}\right)\left(1-\frac{d_2}{r_2}\right)\cdots\left(1-\frac{d_{i-1}}{r_{i-1}}\right) \\
&= d_i\left(\frac{r_0-d_0}{r_0}\right)\left(\frac{r_1-d_1}{r_1}\right)\cdots\left(\frac{r_{i-1}-d_{i-1}}{r_{i-1}}\right)\frac{1}{r_i} \\
&= \frac{d_i}{r_0}\left(\frac{r_0-d_0}{r_1}\right)\left(\frac{r_1-d_1}{r_2}\right)\cdots\left(\frac{r_{i-1}-d_{i-1}}{r_i}\right).
\end{aligned}$$

Here

$$r_j - d_j = r_{j+1} + \lambda_j, \quad r_0 = n, \quad d_0 = 0.$$

Therefore

$$\hat{p}(t_{(i)}) = \frac{d_i}{n}\prod_{j=0}^{i-1}(1+\frac{\lambda_j}{r_{j+1}}).$$

■

When $\lambda_j = 0$, $j = 0, \ldots, i-1$, $\hat{p}(t_{(i)})$ is given by d_i/n.

The estimate of variance for KM estimate $\hat{S}(t_{(i)})$ is given by

$$\hat{Var}(\hat{S}(t_{(i)})) = \hat{S}(t_{(i)})^2 \sum_{j=1}^{i}\left(\frac{d_j}{r_j(r_j-d_j)}\right), \tag{2.7}$$

which corresponds to Greenwood's formula (Miller, 1981).

Proof. Setting

$$\hat{v}_j = (r_j - d_j)/r_j.$$

We assume that $r_j - d_j$ has a binomial distribution with parameters r_j and $\hat{v}_j$. The variance of $r_j - d_j$ is given by

$$\hat{Var}(r_j - d_j) = r_j \hat{v}_j (1 - r_j).$$

Therefore

$$\hat{Var}(\hat{v}_j) = \frac{\hat{Var}(r_j - d_j)}{r_j^2} = \frac{\hat{v}_j(1 - \hat{v}_j)}{r_j}. \tag{2.8}$$

The delta method is as follows: we assume that $\hat{v}_j$ is distributed by

$$\hat{v}_j \sim (\mu, \sigma^2),$$

where $\mu = E\{\hat{v}_j\}$, $\sigma^2 = Var\{\hat{v}_j\}$. $g(\hat{v}_j)$ denotes the function of $\hat{v}_j$, and the Taylor series approximation gives

$$g(\hat{v}_j) = g(\mu) + (\hat{v}_j - \mu)\frac{dg(\hat{v}_j)}{d\hat{v}_j}.$$

Therefore $g(\hat{v}_j)$ is distributed by

$$g(\hat{v}_j) \sim (g(\mu), \sigma^2(dg(\hat{v}_j)/d\hat{v}_j)^2),$$

then

$$Var(g(\hat{v}_j)) = \sigma^2 \left(\frac{dg(\hat{v}_j)}{d\hat{v}_j}\right)^2.$$

Setting $g(\hat{v}_j) = \log \hat{v}_j$, then according to (2.8),

$$\begin{aligned} \hat{Var}(\log \hat{v}_j) &= \hat{Var}(\hat{v}_j)\left(\frac{d \log \hat{v}_j}{d\hat{v}_j}\right)^2 \\ &= \frac{\hat{Var}(\hat{v}_j)}{\hat{v}_j^2} = \frac{1 - \hat{v}_j}{r_j \hat{v}_j} = \frac{d_j}{r_j(r_j - d_j)}. \end{aligned}$$

Therefore

$$\hat{Var}\{\log \hat{S}(t_{(i)})\} = \hat{Var}\{\sum_{j=1}^{i} \log \hat{v}_j\}$$
$$= \sum_{j=1}^{i} \frac{d_j}{r_j(r_j - d_j)}. \tag{2.9}$$

The delta method gives

$$\hat{Var}\{\log \hat{S}(t_{(i)})\} = \frac{1}{\hat{S}(t_{(i)})^2} \hat{Var}\{\hat{S}(t_{(i)})\},$$

then according to (2.9),

$$\hat{Var}\{\hat{S}(t_{(i)})\} = \hat{S}(t_{(i)})^2 \sum_{j=1}^{i} \frac{d_j}{r_j(r_j - d_j)}.$$

■

The estimated discrete probability density $\hat{f}(t)$ is given by $\hat{f}(t) = \hat{p}(t_{(i)})\xi_i$, where if $t = t_{(i)}$, then $\xi_i = 1$, otherwise $\xi_i = 0$. In this thesis the following are required:

1. The estimated probability density function $\hat{f}(t)$ is continuous to compare with the parametric distribution models.

2. The KM model is parametrized to apply EIC introduced in the next section.

Therefore, we define a linearly interpolated KM (LIKM) estimate $\hat{S}(t)$ derived from (2.5) as

$$\hat{S}(t) = \hat{S}(t_{(i-1)}) - \hat{\theta}(t_{(i)})(t - t_{(i-1)}), \quad t_{(i-1)} \leq t < t_{(i)}, \tag{2.10}$$

where $\hat{\theta}(t_{(i)})$ are the estimated parameters of the LIKM model defined by $\hat{\theta}(t_{(i)}) = \hat{p}(t_{(i)})/(t_{(i)} - t_{(i-1)})$, $i = 1, \ldots, k$. The estimated continuous probability density is

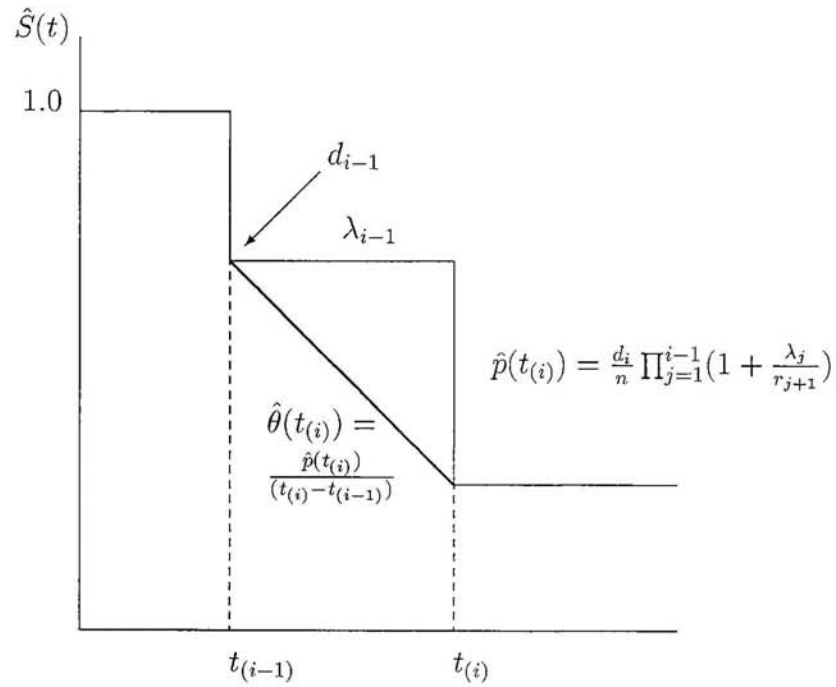

Figure 2.1: Illustration of linearly interpolated KM (LIKM) model.

given by $\hat{f}(t) = \hat{\theta}(t_{(i)})$, $t_{(i-1)} \leq t < t_{(i)}$. The log likelihood of the LIKM model is given by

$$\log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) = \sum_{i=0}^{k} \left[d_i \log \hat{\theta}(t_{(i)}) + \lambda_i \log \hat{S}(t_{(i)}) \right], \tag{2.11}$$

where $\hat{\boldsymbol{\theta}}(\boldsymbol{x}) = (\hat{\theta}(t_{(i)}), \ldots, \hat{\theta}(t_{(k)}))$ is the parameter vector estimated from the data $\boldsymbol{x}$, and $d_0 \log \hat{\theta}(t_{(0)}) = 0$. Figure 2.1 illustrates the LIKM model. Since the number of the parameters k depends on the data $\boldsymbol{x}$, the LIKM model is a flexible model for estimating survival probabilities.

Next we introduce the parametric survival probability models. The survival

probability function of Weibull distribution is given by

$$S(t) = \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}, \quad \alpha, \beta > 0, \tag{2.12}$$

where α and β are scale and shape parameters, respectively. In case of $\beta = 1$, S(t) corresponds to the survivor function of the exponential distribution. The hazard rate function $h(t)$ of the Weibull distribution is given by

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}. \tag{2.13}$$

In case of $\beta > 1$ ($\beta < 1$), $h(t)$ has IHR (DHR). It is known that the Weibull distribution is the asymptotic distribution of the least extreme value. Therefore if the failure is caused by damage to the weakest point of the individual, the probability of failure time will be expressed by the Weibull distribution.

The survival probability function of gamma distribution is given by

$$S(t) = \frac{\Gamma(a, \frac{t}{b})}{\Gamma(a)}, \quad a, b > 0, \tag{2.14}$$

where $\Gamma(\cdot)$ and $\Gamma(\cdot, \cdot)$ are gamma and incomplete gamma functions, respectively, a and b are shape and scale parameters, respectively. The hazard rate function $h(t)$ is given by

$$h(t) = \frac{t^{a-1}e^{-\frac{t}{b}}}{b^a\Gamma(a, \frac{t}{b})}. \tag{2.15}$$

When $a > 1$ ($a < 1$), $h(t)$ is IHR (DHR). In case of $a = 1$, the gamma distribution is reduced to the exponential distribution. It is known that if the failure is caused by s times Poisson shocks, the survival probability of life time follows the gamma distribution with the shape parameter $a = s$.

We use the maximum likelihood method to estimate the parameters of the parametric models. We assume the random censoring mechanism. Let Y be a random variable of the failure which has the distribution function $F(y|\boldsymbol{\theta})$ with parameter $\boldsymbol{\theta}$. Let C be a random variable of censoring which has the distribution function $F_c(c|\boldsymbol{\theta}_c)$ with parameter $\boldsymbol{\theta}_c$. The random variable of observation T is given by $T = min(Y, C)$. The likelihood function for the survival data is given by

$$\prod_{i=1}^{n} [f(t_i|\boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta})^{1-\delta_i}][f_c(t_i|\boldsymbol{\theta}_c)^{1-\delta_i} S_c(t_i|\boldsymbol{\theta}_c)^{\delta_i}], \tag{2.16}$$

where $S(t_i|\boldsymbol{\theta}) = 1 - F(t_i|\boldsymbol{\theta})$ and $S_c(t_i|\boldsymbol{\theta}_c) = 1 - F_c(t_i|\boldsymbol{\theta}_c)$. Only $\boldsymbol{\theta}$ is of our interest, hence we maximize the following likelihood function:

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta})^{\delta_i} S(t_i|\boldsymbol{\theta})^{1-\delta_i}. \tag{2.17}$$

The log likelihood of parametric models is given by

$$\log f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \sum_{i=1}^{n} [\delta_i \log f(t_i|\boldsymbol{\theta}) + (1 - \delta_i) \log S(t_i|\boldsymbol{\theta})]. \tag{2.18}$$

The maximum likelihood estimates are numerically obtained by DALL (Ishiguro and Akaike, 1989). The parameter vector estimated from the data $\boldsymbol{x}$ is denoted by $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$. According to the delta method, the variance of $\hat{S}(t)$ is given by

$$Var(\hat{S}(t)) = \left[\left(\frac{\partial S(t)}{\partial \boldsymbol{\theta}} \right)^T I^{-1} \left(\frac{\partial S(t)}{\partial \boldsymbol{\theta}} \right) \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(\boldsymbol{x})}, \tag{2.19}$$

where I is Fisher's information matrix and $(\partial S(t)/\partial \boldsymbol{\theta})$ is the column vector. The elements I_{ij} of I is given by

$$I_{ij} = -E_X \left\{ \frac{\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\}. \tag{2.20}$$

2.3 Analysis for survival data with covariates

2.3.1 Survival probability

Let $f(t;\boldsymbol{z})$ be the conditional probability density function of T given covariate $\boldsymbol{z}$. Let $S(t;\boldsymbol{z})$ denotes the survivor function,

$$S(t;\boldsymbol{z}) = Pr\{t \leq T;\boldsymbol{z}\} = \int_t^\infty f(x;\boldsymbol{z})dx. \tag{2.21}$$

The hazard rate function $h(t;\boldsymbol{z})$ is defined as

$$\begin{aligned} h(t;\boldsymbol{z}) &= \lim_{\Delta t \to 0} \frac{Pr\{t \leq T < t+\Delta t | t \leq T;\boldsymbol{z}\}}{\Delta t} \\ &= \frac{f(t;\boldsymbol{z})}{S(t;\boldsymbol{z})}. \end{aligned} \tag{2.22}$$

$h(t;\boldsymbol{z})$ has also one of following three properties: IHR (increasing hazard rate), DHR (decreasing hazard rate) and CHR (constant hazard rate). $S(t;\boldsymbol{z})$ is also written as follows:

$$\begin{aligned} h(t;\boldsymbol{z}) &= \frac{-S(t;\boldsymbol{z})'}{S(t;\boldsymbol{z})} = -\frac{d\log S(t;\boldsymbol{z})}{dt}, \\ \log S(t;\boldsymbol{z}) &= -\int_0^t h(u;\boldsymbol{z})du, \\ S(t;\boldsymbol{z}) &= \exp[-\int_0^t h(u;\boldsymbol{z})du]. \end{aligned} \tag{2.23}$$

2.3.2 Models and estimation procedures

Cox (1972) proposed a proportional hazard model for the survival data with covariate $\boldsymbol{z}$. It is given by

$$h(t;\boldsymbol{z}) = h_o(t)\exp(\boldsymbol{p}\boldsymbol{z}^T), \tag{2.24}$$

where $h_o(t)$ is an arbitrary function called the baseline hazard rate function. Let $h(t;\boldsymbol{z})$ be called Cox model. $S(t;\boldsymbol{z})$ is also written as follows:

$$\begin{aligned} S(t;\boldsymbol{z}) &= \exp[-\int_0^t h(u;\boldsymbol{z})du] \\ &= \exp[-\int_0^t h_o(u)\exp(\boldsymbol{p}\boldsymbol{z}^T)]du \\ &= S_o(t)^{\exp(\boldsymbol{p}\boldsymbol{z}^T)}, \end{aligned} \tag{2.25}$$

where $S_o(t)$ denotes the baseline survival probability function.

Several methods for estimating the parameter $\boldsymbol{p}$ are known (Kalbfleisch and Prentice, 1980, Miller, 1981). Cox (1972) proposed the method of conditional likelihood for the estimation. Kalbfleisch and Prentice (1973) proposed the method based on a marginal likelihood. Cox (1975) also proposed the method of partial likelihood. We use the maximization of the approximated marginal likelihood given by Breslow (1974). The marginal likelihood is given by

$$L_p = \prod_{i=1}^{k} e^{\boldsymbol{p}\boldsymbol{S}_i^T} / (\sum_{l\in R_i} e^{\boldsymbol{p}\boldsymbol{z}_l^T})^{d_i}, \tag{2.26}$$

where $t_{(i)}$, $i = 1,\ldots,k$ are the ordered failure times $0 = t_{(0)} < \ldots < t_{(i-1)} < t_{(i)} < \ldots < t_{(k+1)} = \infty$, k is the number of ordered failure times, d_i is the number of failures at $t_{(i)}$, $\boldsymbol{S}_i$ is the sum of covariate vectors for d_i individuals observed to die at $t_{(i)}$. R_i is the risk set at $t_{(i)}$, i.e., the set of survivors just before $t_{(i)}$.

Some methods for estimating the baseline survival probability $\hat{S}_o(t)$ are also proposed (Collet, 1994). Basically $\hat{S}_o(t)$ is derived from the maximization of the following nonparametric likelihood function with the estimated parameter $\hat{\boldsymbol{p}}$:

$$L_{S_o} =$$

$$\prod_{i=1}^{k}\left[\prod_{j\in D_i}\{\hat{S}_o(t_{(i-1)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_j^T)}-\hat{S}_o(t_{(i)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_j^T)}\}\prod_{l\in C_i}\hat{S}_o(t_{(i)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_l^T)}\right], \quad (2.27)$$

where D_i and C_i denote the sets of failures and censored patients at $t_{(i)}$, respectively. An approximated estimate of $\hat{S}_o(t)$ is given by

$$\hat{S}_o(t)=\prod_{t_{(i)}<t}\exp\left(\frac{-d_i}{\sum_{l\in R_i}exp(\hat{\boldsymbol{p}}\boldsymbol{z}_l^T)}\right). \quad (2.28)$$

The estimate of the survival probability with $\boldsymbol{z}$ is given by $\hat{S}(t;\boldsymbol{z})=\hat{S}_o(t)^{exp(\hat{\boldsymbol{p}}\boldsymbol{z}^T)}$.

In this thesis, the probability density function of baseline $f_o(t)$, $t_{(i-1)}\le t<t_{(i)}$ is continuous. We define a linearly interpolated Cox (LIC) estimate of the baseline $\hat{S}_o(t)$ derived from (2.28) as

$$\hat{S}_o(t)=\hat{S}_o(t_{(i-1)})-\hat{\theta}_o(t_{(i)})(t-t_{(i-1)}), \quad t_{(i-1)}\le t<t_{(i)}, \quad (2.29)$$

where $\hat{\theta}_o(t_{(i)})$ are the estimated parameters given by

$$\hat{\theta}_o(t_{(i)})=(\hat{S}_o(t_{(i-1)})-\hat{S}_o(t_{(i)}))/(t_{(i)}-t_{(i-1)}), \quad i=1,\ldots,k. \quad (2.30)$$

The estimated probability density function of the baseline is given by $\hat{f}_o(t)=\hat{\theta}_o(t_{(i)})$, $t_{(i-1)}\le t<t_{(i)}$. The likelihood of the LIC survival probability model is defined as

$$f(\boldsymbol{x}\mid\hat{\boldsymbol{\theta}}(\boldsymbol{x}))=$$
$$\prod_{i=1}^{k}\left[\prod_{j\in D_i}\left\{\frac{\hat{S}_o(t_{(i-1)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_j^T)}-\hat{S}_o(t_{(i)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_j^T)}}{t_{(i)}-t_{(i-1)}}\right\}\prod_{l\in C_i}\hat{S}_o(t_{(i)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_l^T)}\right],$$

where $\hat{\boldsymbol{\theta}}(\boldsymbol{x})=(\hat{\boldsymbol{p}},\hat{\theta}_o(t_{(1)}),\ldots,\hat{\theta}_o(t_{(k)}))$ is the parameter vector estimated from the data $\boldsymbol{x}$. The log likelihood of the LIC survival probability model is given by

$$\log f(\boldsymbol{x}\mid\hat{\boldsymbol{\theta}}(\boldsymbol{x}))=$$
$$\sum_{i=1}^{k}\left[\log\prod_{j\in D_i}\left\{\frac{\hat{S}_o(t_{(i-1)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_j^T)}-\hat{S}_o(t_{(i)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_j^T)}}{t_{(i)}-t_{(i-1)}}\right\}+\log\prod_{l\in C_i}\hat{S}_o(t_{(i)})^{\exp(\hat{\boldsymbol{p}}\boldsymbol{z}_l^T)}\right]. \quad (2.31)$$

Next, we introduce accelerated models based on parametric distributions. Assume that the random variable of survival time T satisfies

$$\log T = \boldsymbol{p}\boldsymbol{z}^T + W, \tag{2.32}$$

where $\boldsymbol{p}$ and $\boldsymbol{z}$ are parameter and covariate vectors, respectively, and W is an error variable depending on a distribution. Defines the random variable of survival time of baseline as $T_o = e^W$, then the regression model $T = T_o \exp(\boldsymbol{p}\boldsymbol{z}^T)$ is obtained. In parametric approaches to survival analysis, the Weibull and gamma distributions are widely used. The accelerated models are derived from assuming the parametric distributions for W in (2.32).

The survivor function $S(t; \boldsymbol{z})$ of the accelerated model based on the Weibull distribution is given by

$$S(t; \boldsymbol{z}) = \exp\left\{-\left(\frac{t}{\exp(\boldsymbol{p}\boldsymbol{z}^T)}\right)^{\beta}\right\}, \quad \beta > 0, \tag{2.33}$$

where β is the shape parameter. In the case of $\beta = 1$, $S(t; \boldsymbol{z})$ corresponds to the survival function of the exponential regression model. The hazard rate function $h(t; \boldsymbol{z})$ of Weibull accelerated model is given by

$$h(t; \boldsymbol{z}) = \beta t^{\beta-1} exp(-\beta \boldsymbol{p}\boldsymbol{z}^T). \tag{2.34}$$

In the case of $\beta > 1$ ($\beta < 1$), $h(t; \boldsymbol{z})$ has IHF (DHF), and $h(t; \boldsymbol{z})$ is a proportional hazard model. The probability density function of the baseline is given by

$$f_o(t) = \beta t^{\beta-1} \exp(-t^{\beta}). \tag{2.35}$$

The survival probability function $S(t;\boldsymbol{z})$ of gamma accelerated model is given by

$$S(t;\boldsymbol{z}) = \frac{\Gamma(a, \frac{t}{\exp[\boldsymbol{p}\boldsymbol{z}^T]})}{\Gamma(a)}, \quad a > 0, \tag{2.36}$$

where a is the shape parameter. The hazard rate function $h(t;\boldsymbol{z})$ is given by

$$h(t;\boldsymbol{z}) = \frac{t^{a-1}\exp\left(\frac{t}{\exp[\boldsymbol{p}\boldsymbol{z}^T]}\right)}{\Gamma(a, \frac{t}{\exp[\boldsymbol{p}\boldsymbol{z}^T]})}\exp(a\boldsymbol{p}\boldsymbol{z}^T). \tag{2.37}$$

When $a > 1$ ($a < 1$), $h(t;\boldsymbol{z})$ is IHR (DHR). The probability density function of baseline is given by

$$f_o(t) = \frac{t^{a-1}\exp(-t)}{\Gamma(a)}. \tag{2.38}$$

The maximum likelihood method is used to estimate the parameters of the parametric accelerated models. We assume the random censoring mechanism. Let Y be the random variable of failure which has the distribution function $F(y;\boldsymbol{z}|\boldsymbol{\theta})$ with the parameter $\boldsymbol{\theta}$. Let C be the random variable of the censoring which has the distribution function $F_c(c|\boldsymbol{\theta}_c)$ with the parameter $\boldsymbol{\theta}_c$. The random variable of observation T is given by $T = min(Y, C)$. The likelihood function for the survival data is given by

$$\prod_{i=1}^{n}[f(t_i;\boldsymbol{z}_i|\boldsymbol{\theta})^{\delta_i}S(t_i;\boldsymbol{z}_i|\boldsymbol{\theta})^{1-\delta_i}][f_c(t_i|\boldsymbol{\theta}_c)^{1-\delta_i}S_c(t_i|\boldsymbol{\theta}_c)^{\delta_i}], \tag{2.39}$$

where $S(t_i;\boldsymbol{z}_i|\boldsymbol{\theta}) = 1 - F(t_i;\boldsymbol{z}_i|\boldsymbol{\theta})$ and $S_c(t_i|\boldsymbol{\theta}_c) = 1 - F_c(t_i|\boldsymbol{\theta}_c)$. Only $\boldsymbol{\theta}$ is of our interest, hence we maximize the following likelihood function:

$$f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i;\boldsymbol{z}_i|\boldsymbol{\theta})^{\delta_i}S(t_i;\boldsymbol{z}_i|\boldsymbol{\theta})^{1-\delta_i}. \tag{2.40}$$

The log likelihood of accelerated models is given by

$$\log f(\boldsymbol{x} \mid \boldsymbol{\theta}) = \sum_{i=1}^{n} [\delta_i \log f(t_i; \boldsymbol{z}_i | \boldsymbol{\theta}) + (1 - \delta_i) \log S(t_i; \boldsymbol{z}_i | \boldsymbol{\theta})]. \tag{2.41}$$

The maximum likelihood estimates are numerically obtained by DALL (Ishiguro and Akaike, 1989). The variance of $\hat{S}(t; \boldsymbol{z})$ is given by

$$Var(\hat{S}(t; \boldsymbol{z})) = \left[\left(\frac{\partial S(t; \boldsymbol{z})}{\partial \boldsymbol{\theta}} \right)^T I^{-1} \left(\frac{\partial S(t; \boldsymbol{z})}{\partial \boldsymbol{\theta}} \right) \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(\boldsymbol{x})}, \tag{2.42}$$

where I is Fisher's information matrix and $(\partial S(t)/\partial \boldsymbol{\theta})$ is the column vector. The elements I_{ij} of I is given by

$$I_{ij} = -E_X \left\{ \frac{\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\}. \tag{2.43}$$

Chapter 3

Information Criteria and Bootstrap Methods

This chapter provides bootstrap methods and information criteria for the survival analysis. Firstly, standard bootstrap methods are described. Secondly, the information criteria, AIC and EIC, are introduced. Furthermore we propose EIC procedures based on model-based bootstrap methods for the nonparametric survival probability models. The detailed descriptions of the information criteria are referred by Sakamoto et al. (1986), Ishiguro et al. (1991, 1994), Kitagawa et al. (1995), Konishi and Kitagawa (1995).

3.1 Bootstrap methods

A data set of size n, $\boldsymbol{t} = (t_1, \ldots, t_n)$ is drawn from the true distribution $G(t)$. The empirical distribution $G_*(t)$ is given by

$$G_*(t) = \frac{1}{n}\sum_{i=1}^{n} I(t, t_i), \tag{3.1}$$

where $I(t, a)$ is defined by $I(t, a) = 0$ if $t < a$ and $I(t, a) = 1$ otherwise. Efron (1979) proposed a set of bootstrap samples of size n, $\boldsymbol{t}^* = (t_1^*, \ldots, t_n^*)$ derived from

the empirical distribution G_*. In the bootstrapping, the true distribution $G(t)$ is replaced by the empirical distribution $G_*(t)$. Therefore variances of estimators can be estimated using the bootstrap methods instead of the Monte Carlo procedures.

Incidentally, bootstrap procedures play important roles in survival analysis. Efron (1981) proposed the bootstrap procedure for estimating interval estimates of KM model for survival data. He showed that the bootstrap estimate of the variance of $\hat{S}(t)$ corresponds to Greenwood's formula given by (2.7). In this case the bootstrap samples $x_i^* = \{t_i^*, \delta_i^*\}$, $i = 1, \ldots, n$ were derived from $t_i^* = min(y_i^*, c_i^*)$, $\delta_i^* = 1$ if $y_i^* \leq c_i^*$ or $\delta_i^* = 0$ otherwise, where y_i^* is given by the KM estimate $\hat{S}(y_i^*)$ in (2.5) for the data $\boldsymbol{x}$ and c_i^* is given by the KM estimate $\hat{S}(c_i^*)$ for $\boldsymbol{x}^{-1}$. $\boldsymbol{x}^{-1}$ is a set of the data given by translations from δ_l of $\boldsymbol{x}$, 1 into 0 and 0 into 1, $l = 1, \ldots, n$. The probability of $t_i^* = t_i$ is $1/n$, therefore this procedure is the same as the n times replacement from $x_i = \{t_i, \delta_i\}$, $i = 1, \ldots, n$.

Proof. Let the bootstrap random variables be denoted by T^*, Y^* and C^*, where $T^* = min(Y^*, C^*)$. The KM estimate of $S(t) = Pr\{Y^* > t\}$ is given by

$$\hat{S}(t) = \prod_{j=1}^{l} \left(\frac{n-j}{n-j+1} \right)^{\delta_j},$$

where $t \in [t_{(l)}, t_{(l+1)})$, n is the sample size, and $t_{(j)}$, $j = 1, \ldots, k$ are ordered observations. The KM estimate of $S_c(t) = Pr\{C^* > t\}$ is given by

$$\hat{S}_c(t) = \prod_{j=1}^{l} \left(\frac{n-j}{n-j+1} \right)^{1-\delta_j}.$$

The estimate of probability $Pr\{T^* > t\} = S(t)S_c(t)$ is given by

$$S(t)S_c(t) = \prod_{j=1}^{l} \left(\frac{n-j}{n-j+1} \right) = 1 - \frac{l}{n}.$$

$1 - S(t)S_c(t)$ means the empirical distribution. Then $Pr\{T^* = t\}$ is $1/n$.

■

A similar bootstrap procedure for survival data, $S_c(t)$ having the empirical distribution, was proposed by Reid (1981). The comparison of asymptotic behaviors about these bootstrap methods was offered by Akritas (1986).

Burr (1994) showed bootstrap intervals of parameters in the Cox model. He considered three types of bootstrap methods based on the following basics of generating survival data. An individual i has a covariate vector $\boldsymbol{z}_i$ and a data set $x_i = \{t_i, \delta_i, \boldsymbol{z}_i\}$. The random variable of observation T_i is given by $T_i = min(Y_i, C_i)$, where Y_i is a random variable of failure time and C_i is a random variable of censoring time. Y_i, $i = 1, \ldots, n$ are assumed to be $iid \sim \hat{S}(t; \boldsymbol{z}_i)$ based on (2.28) for the data $\boldsymbol{x}$. C_i are also assumed to be $iid \sim \hat{S}_c(t)$ derived from the KM estimates for the data $\boldsymbol{x}^{-1}$. The bootstrap methods proposed by Burr are as follows: (1) the replacement of the original data based on the empirical distribution, (2) the direct generation of T_i^*, Y_i^*, C_i^* derived from the above mechanism under resampling conditional on covariates, (3) the conditional generation of Y_i^*, C_i^* derived from the above distributions under resampling conditional on covariates.

3.2 Information criteria AIC and EIC for survival analysis

The Kullback-Leibler information number $I(g; f)$ is given by

$$I(g; f) = E_X \left\{ \log \frac{g(X)}{f(X)} \right\} = \int \left\{ \log \frac{g(x)}{f(x)} \right\} g(x) dx$$

$$= \int g(x) \log g(x) dx - \int g(x) \log f(x) dx, \tag{3.2}$$

where $g(x)$ and $f(x)$ are the true distribution and the model distribution, respectively. The characteristics of $I(g; f)$ are as follows:

1. $I(g; f) \geq 0$.

2. $I(g; f) = 0 \Longleftrightarrow g(x) = f(x)$.

The entropy $B(g; f)$ is defined as $B(g; f) = -I(g; f)$. A model which maximizes $B(g; f)$ is considered to be the most appropriate model to the true distribution. (3.2) shows that the model which maximizes the second term is a good model. The second term is the expected log likelihood. From the point of view of the entropy maximization principle proposed by Akaike (1973), the information criteria estimate the sample size times the mean expected log likelihood (ELL). ELL for the random censoring mechanism is given by

$$\begin{aligned} E_X\{E_U\{\log f(\boldsymbol{u} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}))\}\} = nE_X \Big\{ & \int_0^\infty \log f(u|\hat{\boldsymbol{\theta}}(\boldsymbol{x})) \cdot f(u|\boldsymbol{\theta}_o) S_c(u|\boldsymbol{\theta}_{co}) du \\ & + \int_0^\infty \log S(u|\hat{\boldsymbol{\theta}}(\boldsymbol{x})) \cdot f_c(u|\boldsymbol{\theta}_{co}) S(u|\boldsymbol{\theta}_o) du \Big\}, \end{aligned}$$

where $\boldsymbol{\theta}_o$ and $\boldsymbol{\theta}_{co}$ are the true parameters of the failure and censoring distributions, respectively. It is shown that a model which maximizes ELL is considered to be the most appropriate model. A natural estimate of ELL is provided by a log likelihood $\log f(\boldsymbol{x}|\hat{\boldsymbol{\theta}}(\boldsymbol{x}))$. The bias of the information criterion means the expectation of the difference between ELL and the log likelihood (LL). The bias C is given by

$$C = E_X\{\log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) - E_U\{\log f(\boldsymbol{u} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})\}\}. \tag{3.3}$$

Therefore, the unbiased estimate of ELL is given by $\log f(\boldsymbol{x}|\hat{\boldsymbol{\theta}}(\boldsymbol{x})) - C$.

We assume that $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is the maximum likelihood estimate. Akaike (1973) showed that C can be asymptotically approximated by the dimension of the parameter vector $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$. AIC is defined as

$$AIC = -2 \times \log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) + 2 \times V,$$

where V is the number of free parameters. A model which minimizes AIC is considered to be the most appropriate model. The derivation of AIC is as following. The bias C is also given by

$$\begin{aligned} C &= E_X\{\log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) - \log f(\boldsymbol{x} \mid \boldsymbol{\theta}_o)\} \\ &+ E_X\{\log f(\boldsymbol{x} \mid \boldsymbol{\theta}_o) - E_U\{\log f(\boldsymbol{u} \mid \boldsymbol{\theta}_o)\}\} \\ &+ E_X\{E_U\{\log f(\boldsymbol{u} \mid \boldsymbol{\theta}_o)\} - E_U\{\log f(\boldsymbol{u} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}))\}\} \\ &\equiv C_1 + C_2 + C_3 \end{aligned} \tag{3.4}$$

According to the Taylor series approximation, C_1 is given by

$$\begin{aligned} C_1 &= E_X\{\log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) - \log f(\boldsymbol{x} \mid \boldsymbol{\theta}_o)\} \\ &= E_X\left\{\log f(\boldsymbol{x} \mid \boldsymbol{\theta}_o) + (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_o)\left[\frac{\partial \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}\right. \\ &+ \left.\frac{1}{2}(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_o)\left[\frac{\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_o)^T - \log f(\boldsymbol{x} \mid \boldsymbol{\theta}_o)\right\}, \end{aligned} \tag{3.5}$$

where $[\partial \log f(\boldsymbol{x} \mid \boldsymbol{\theta})/\partial \boldsymbol{\theta}]$ is V dimensional column vector, and $[\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^T]$ is $V \times V$ dimensional matrix. Because $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is the maximum likelihood estimate, the

second term of (3.5) is 0. Furthermore $\Delta\hat{\boldsymbol{\theta}}(\boldsymbol{x}) = E_X\{(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta})\}$ is distributed by the normal distribution with the mean 0 and the variance I^{-1}, i.e., $\Delta\hat{\boldsymbol{\theta}}(\boldsymbol{x}) \sim N(0, I^{-1})$. Therefore C_1 is given by

$$\begin{aligned} C_1 &= \frac{1}{2}E_X\left\{(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_o)\left[\frac{\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}_o)^T\right\} \\ &= \frac{1}{2}\Delta\hat{\boldsymbol{\theta}}(\boldsymbol{x})I\Delta\hat{\boldsymbol{\theta}}(\boldsymbol{x})^T \\ &\approx \frac{V}{2}. \end{aligned} \tag{3.6}$$

Note that $\Delta\hat{\boldsymbol{\theta}}(\boldsymbol{x})I\Delta\hat{\boldsymbol{\theta}}(\boldsymbol{x})^T \sim \chi^2$ distribution with the degree of freedom V, and the mean of the χ^2 distribution is V. C_2 is given by

$$\begin{aligned} C_2 &= E_X\{\log f(\boldsymbol{x} \mid \boldsymbol{\theta}_o) - E_U\{\log f(\boldsymbol{u} \mid \boldsymbol{\theta}_o)\}\} \\ &= 0. \end{aligned} \tag{3.7}$$

According to the Taylor series approximation, C_3 is given by

$$\begin{aligned} C_3 &= E_X\{E_U\{\log f(\boldsymbol{u} \mid \boldsymbol{\theta}_o)\} - E_U\{\log f(\boldsymbol{u} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}))\}\} \\ &= E_X\left\{E_U\left\{\log f(\boldsymbol{u} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) + (\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}(\boldsymbol{x}))\left[\frac{\partial \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{x})}\right.\right. \\ &+ \left.\left.\frac{1}{2}(\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}(\boldsymbol{x}))\left[\frac{\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{x})}(\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}(\boldsymbol{x}))^T\right\} - E_U\{\log f(\boldsymbol{u} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}))\}\right\}. \end{aligned} \tag{3.8}$$

Because of $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is the maximum likelihood estimate, the second term of (3.8) is 0, $\Delta\boldsymbol{\theta}_o = E_X\{(\boldsymbol{\theta}_o - \hat{\boldsymbol{\theta}}(\boldsymbol{x}))^T\} \sim N(0, I^{-1})$ and $E_X\{[\partial^2 \log f(\boldsymbol{x} \mid \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}(\boldsymbol{x})}\}$ $\approx I$. According to the mean of the χ^2 distribution, C_3 is given by

$$\begin{aligned} C_3 &= \frac{1}{2}\Delta\boldsymbol{\theta}_o I\Delta\boldsymbol{\theta}_o^T \\ &\approx \frac{V}{2}. \end{aligned} \tag{3.9}$$

According to (3.6), (3.7) and (3.9), The bias C is given

$$\begin{aligned} C &= C_1 + C_2 + C_3 \\ &\approx V. \end{aligned} \tag{3.10}$$

AIC is restricted to the models with maximum likelihood estimates. Ishiguro et al. (1991, 1994) proposed the information criterion EIC which is an estimation of ELL and constructed on the bootstrap method. EIC is applicable to the wide models with not only maximum likelihood estimates but also other estimates. Now we assume that $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is not restricted to the maximum likelihood estimate. EIC is defined as

$$\begin{aligned} EIC &= -2 \times \log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) + 2 \times \hat{C}^* \\ \hat{C}^* &= E_{X^*}\{\log f(\boldsymbol{x}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*)) - E_{U^*}\{\log f(\boldsymbol{u}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*)\}\}, \end{aligned} \tag{3.11}$$

where $\boldsymbol{x}^*$ and $\boldsymbol{u}^*$ represent the bootstrap samples from original data $\boldsymbol{x}$, and $\hat{C}^*$ is the estimate of the bias C defined by (3.3). A model which minimizes EIC is considered to be the most appropriate model. According to (3.4), C^* is given by

$$\begin{aligned} \hat{C}^* &= E_{X^*}\{\log f(\boldsymbol{x}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*)) - \log f(\boldsymbol{x}^* \mid \boldsymbol{\theta}_o)\} \\ &+ E_{X^*}\{\log f(\boldsymbol{x}^* \mid \boldsymbol{\theta}_o) - E_{U^*}\{\log f(\boldsymbol{u}^* \mid \boldsymbol{\theta}_o)\}\} \\ &+ E_{X^*}\{E_{U^*}\{\log f(\boldsymbol{u}^* \mid \boldsymbol{\theta}_o)\} - E_{U^*}\{\log f(\boldsymbol{u}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*))\}\} \\ &\equiv C_1^* + C_2^* + C_3^*. \end{aligned} \tag{3.12}$$

Because of $C_2^* = 0$ and $\boldsymbol{\theta}_o \approx \hat{\boldsymbol{\theta}}(\boldsymbol{x})$, the bootstrap gives the estimate of C as follows:

$$\hat{C^{**}} = C_1^* + C_3^*$$

$$
\begin{aligned}
&= E_{X^*}\{\log f(\boldsymbol{x}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*)) - \log f(\boldsymbol{x}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})\} \\
&+ E_{X^*}\{E_{U^*}\{\log f(\boldsymbol{u}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})\} - E_{U^*}\{\log f(\boldsymbol{u}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*))\}\}.
\end{aligned} \tag{3.13}
$$

Kitagawa et al. (1995) showed that the variance of $\hat{C}^{**}$ is less than that of $\hat{C}^{*}$ given by (3.11), because the variance of C_2^* is the largest and depends on the sample size. In this thesis we define EIC as follows,

$$
EIC = -2 \times \log f(\boldsymbol{x} \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x})) + 2 \times \hat{C}^{**}. \tag{3.14}
$$

EIC can estimate goodness for not only the maximum likelihood estimator of the parametric distribution model with the asymptotic normality but also wide estimator with non-asymptotic normality. When sample sizes are small in the maximum likelihood estimation, the bias correction of AIC might be unsuitable because the asymptotic theory is not applicable. EIC might be more suitable for the model selection in such a case.

3.3 EIC for nonparametric models

EIC can be used to evaluate goodnesses of nonparametric models. Firstly, the EIC procedure for the KM survival probability model is discussed. Secondly, we consider the application of the EIC procedure to the LIC survival probability model.

3.3.1 EIC procedure for KM survival probability model

The KM estimate has the asymptotic normality in large sample situations (Breslow and Crowley, 1974). Fleming and Harrington (1991), Andersen et al. (1992) gave

proofs of the asymptotic normality of the KM estimate by using martingale theory. EIC does not require the asymptotic normality of the LIKM estimate $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ in (2.11).

Model-based bootstrap method

k and k^* are the numbers of parameters of the LIKM model fitted to the original data $\boldsymbol{x}$ and the bootstrap data $\boldsymbol{x}^*$, respectively. In general it is not satisfied $k = k^*$. When $\boldsymbol{x}^*$ is obtained by resampling from $\boldsymbol{x}$, the tie samples are involved in the resampled bootstrap samples. Thus it is satisfied $k \geq k^*$ in the bootstrapping proposed by Efron (1981). It is happened by using the the resampled bootstrap samples that the bias correction of EIC is estimated on the reduced model under few tied data involved in the original data. Thus the estimated bias is smaller than the true bias.

We propose a bootstrap method based on a model fitted to $\boldsymbol{x}$. We call this method the model-based bootstrap. The model-based bootstrap method has little possibility of making tie. The samples $\boldsymbol{x}^*$ of the model-based bootstrap based on the random censoring mechanism are given by as follows:

1. The bootstrap failure sample y_j^* is generated by using the model $S(y_j \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}))$.

2. The bootstrap censored sample c_j^* is generated by using the model $S_c(c_j \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^{-1}))$.

3. The bootstrap sample t_j^* is defined as $t_j^* = min(y_j^*, c_j^*)$, $\delta_j^* = 1$ $(y_j^* \leq c_j^*)$ or $\delta_j^* = 0$ otherwise.

4. The steps $1 \sim 3$ are repeated n times.

We assume $S_c(\cdot)$ is the same model as $S(\cdot)$. The model-based bootstrap is also used for the parametric models in order to make a comparison with the LIKM model.

The model-based bootstrap procedure for the LIKM model is as follows: let W_j be a uniform random number. Bootstrap samples y_j^* for failure under the original data $\boldsymbol{x}$ are generated from

$$y_j^* = \frac{\hat{S}(t_{(i-1)}) - W_j}{\hat{\theta}(t_{(i)})} + t_{(i-1)}, \quad \hat{S}(t_{(i-1)}) \geq W_j > \hat{S}(t_{(i)}), \quad i = 1, \ldots, k, \quad (3.15)$$

according to the definition of the LIKM model in (2.10). When $W_j \leq \hat{S}(t_{(k)})$, the survival probability $\hat{S}(t)$, $t \geq t_{(k)}$ must be defined. This case is caused by being the censored data $t_l > t_{(k)}$, $\delta_l = 0$, $l = 1, \ldots, c$ in $\boldsymbol{x}$. We propose the following models for $\hat{S}(t)$, $t \geq t_{(k)}$:

1. $\hat{S}(t) = \hat{S}(t_{(k)}) \exp[-(t - t_{(k)})/r]$, $r = M(t_c - t_{(k)})$,

2. $\hat{S}(t) = \hat{S}(t_{(k)})[(r - t)/(r - t_{(k)})]$, $r = Mt_c$,

3. $\hat{S}(t) = \hat{S}(t_{(k)})$,

where M is a constant. The first model means the probability density function $f(t)$, $t \geq t_{(k)}$ is given by a exponential distribution. The second model means $f(t)$, $t_{(k)} \leq t < Mt_c$ is given by a uniform distribution. The third model means $f(t)$, $t \geq t_{(k)}$ is truncated. The bootstrap samples c_j^* for censoring are generated by the same method as the failure under $\boldsymbol{x}^{-1}$.

Bootsprapping log likelihood

According to (2.11), we propose that the bootstrapping log likelihood in (3.13)

for the LIKM model is given by

$$\log f(\boldsymbol{u}^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^*)) = \sum_{i=0}^{k_u^*} \left[d_{ui}^* \log \hat{\theta}_{x^*}(t_{u(i)}^*) + \lambda_{ui}^* \log \hat{S}_{x^*}(t_{u(i)}^*) \right], \tag{3.16}$$

where the notations $\ddagger^*$, $\ddagger_u^*$ imply $\ddagger$ under $\boldsymbol{x}^*$, $\boldsymbol{u}^*$, respectively, and

$$\hat{\theta}_{x^*}(t_{u(i)}^*) = \frac{\hat{S}_{x^*}(t_{u(i-1)}^*) - \hat{S}_{x^*}(t_{u(i)}^*)}{t_{u(i)}^* - t_{u(i-1)}^*}, \tag{3.17}$$

$$\hat{S}_{x^*}(t_{u(i)}^*) = \hat{S}(t_{(j-1)}^*) - \delta_{ui}^*[\hat{\theta}(t_{(j)}^*)(t_{u(i)}^* - t_{(j-1)}^*)],$$

$$t_{(j-1)}^* \leq t_{u(i)}^* < t_{(j)}^*, \quad j = 1, \ldots, k^*. \tag{3.18}$$

This means that the estimates $\hat{\theta}_{x^*}(t_{u(i)}^*)$ are derived from using the parameters $\hat{\theta}(t_{(j)}^*)$ estimated from $\boldsymbol{x}^*$, $i = 1, \ldots, k_u^*, j = 1, \ldots, k^*$. If $\hat{\theta}_{x^*}(t_{u(i)}^*) = \hat{\theta}(t_{(j)}^*)$, $t_{(j-1)}^* \leq t_{u(i)}^* < t_{(j)}^*$ is supposed on (3.16), then the LIKM model is restricted to be $k_u^* = k^*$. When $t_{u(i)}^* \geq t_{(k^*)}^*$, the tail of the LIKM model must be defined as before,

1. $\hat{S}_{x^*}(t_{u(i)}^*) = \hat{S}(t_{(k^*)}^*) \exp[-(t_{u(i)}^* - t_{(k^*)}^*)/r]$, $r = M(t_{c^*} - t_{(k^*)}^*)$,

2. $\hat{S}_{x^*}(t_{u(i)}^*) = \hat{S}(t_{(k^*)}^*)[(r - t_{u(i)}^*)/(r - t_{(k^*)}^*)]$, $r = Mt_{c^*}$,

3. $\hat{S}_{x^*}(t_{u(i)}^*) = \hat{S}(t_{(k^*)}^*)$.

Choice of model for tail

It is important to know which model of tail is good. We choose a good model of tail through results by simulation studies. The failure data y_i and the censored data c_i are generated from the Weibull distributions given by (2.12) with $(\alpha, \beta) = (30.0, 0.9)$ and $(10.0, 2.0)$, respectively. The set of simulation data $\boldsymbol{x}$ is given by $t_i = min(y_i, c_i)$. $\delta_i = 1$ if $y_i \leq c_i$ or $\delta_i = 0$ otherwise, $i = 1, \ldots, n$. Table 3.1 shows

the ordered simulation data ($n = 50$) and survival probabilities estimated by using the LIKM model. In this example, the last failure is $i = 34$. There are 16 censored data after the last failure data. The survival curves of the models for tails are plotted in Figures 3.1 and 3.2. The simulation results by these models are summarized in Table 3.2. The bootstrap procedures for EIC are repeated 1000 times for X^* and 100 times for U^* in (3.13). The models 1 and 2 have similar goodness. Note that the choice of the constant M is not very critical. In the result by the model 3, $E\{n_1\}$, $E\{n_2\}$ are not close to n_1, n_2, respectively, where n_1 and n_2 denote the numbers of the failures and censored data, respectively. In this paper, we adopt the model 2 ($M = 10$), since the probability density function of the LIKM model $f(t)$ is defined as the uniform value interval $t_{(i-1)} \leq t < t_{(i)}$, $i = 1, \ldots, k$, so that the model 2 is similar to the definition of $f(t)$.

3.3.2 EIC procedure for LIC survival probability model

Model-based bootstrap method

The LIC survival probability model is nonparametric. Because of the same reason as the LIKM model, it is happened by using resampled bootstrap samples that the bias correction of EIC is estimated on the reduced model under few tied data involved in the original data. Therefore we use the model-based bootstrap for the EIC procedure of the LIC survival probability model. The samples $\boldsymbol{x}^*$ of the model-based bootstrap based on the data with covariates are obtained as follows:

1. The bootstrap covariate $\boldsymbol{z}_j^*$ is generated by resampling $\boldsymbol{x}$.

2. The bootstrap failure sample y_j^* is generated by using the model

   $S(y_j; \boldsymbol{z}_j^* \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}))$.

3. The bootstrap censoring sample c_j^* is generated by using the model

   $S_c(c_j \mid \hat{\boldsymbol{\theta}}(\boldsymbol{x}^{-1}))$.

4. The bootstrap sample t_j^* is defined as $t_j^* = min(y_j^*, c_j^*)$, $\delta_j^* = 1$ $(y_j^* \leq c_j^*)$ or $\delta_j^* = 0$ otherwise.

5. The steps $1 \sim 4$ are repeated n times.

When $S(\cdot)$ is the LIC survival probability model, $S_c(\cdot)$ is the LIKM model. This model-based bootstrap method is similar to the second procedure of Burr's methods (1994) which we have detailed in Section 3.1. However our method for the LIC survival probability model is constructed on the linearly interpolated nonparametric models. The model-based bootstrap is also used for the parametric models in order to make a comparison with the LIC survival probability model.

The model-based bootstrap procedure for the LIC survival probability model is as follows: let W_j be a uniform random number, and $W_{oj} = W_j^{\exp(-\hat{\boldsymbol{p}}\boldsymbol{z}_j^{*T})}$. The bootstrap samples y_j^* for failure under the original data $\boldsymbol{x}$ are generated from

$$y_j^* = \frac{\hat{S}_o(t_{(i-1)}) - W_{oj}}{\hat{\theta}_o(t_{(i)})} + t_{(i-1)}, \quad \hat{S}_o(t_{(i-1)}) \geq W_{oj} > \hat{S}_o(t_{(i)}), \quad i = 1, \ldots, k, \quad (3.19)$$

according to (2.29). When $W_{oj} \leq \hat{S}_o(t_{(k)})$, the survival probability $\hat{S}_o(t)$, $t \geq t_{(k)}$ must be defined. This case is caused by being the censored data $t_l > t_{(k)}$, $\delta_l = 0$, $l = 1, \ldots, c$ in $\boldsymbol{x}$. The following models for the $\hat{S}_o(t)$, $t \geq t_{(k)}$ are proposed:

1. $\hat{S}_o(t) = \hat{S}_o(t_{(k)})[(r-t)/(r-t_{(k)})]$, $r = Mt_c$,

2. $\hat{S}_o(t) = \hat{S}_o(t_{(k)})$,

where M is a constant. The bootstrap samples for censored c_j^* are generated by the same method as the LIKM model under $\boldsymbol{x}^{-1}$. Note that a model of the tail based on a exponential function is not suitable, because $\hat{S}_o(t) \approx 0$ ($W_{oj} = W_j^{\exp(-\hat{\boldsymbol{p}}\boldsymbol{z}_j^{*T})} \approx 0$) is often supposed.

Bootsprapping log likelihood

According to (2.31), we propose that the bootstrapping log likelihood in (3.13) for the LIC survival probability model is given by

$$\log f(\boldsymbol{u}^*|\hat{\boldsymbol{\theta}}(\boldsymbol{x}^*)) = \sum_{i=1}^{k_u^*}\left[\log \prod_{j\in D_{ui}^*} \frac{1}{t_{u(i)}^* - t_{u(i-1)}^*}\left(\hat{S}_{ox^*}(t_{u(i-1)}^*)^{exp(\hat{\boldsymbol{p}}(\boldsymbol{x}^*)\boldsymbol{z}_{uj}^{*T})} - \right.\right.$$
$$\left.\left.\hat{S}_{ox^*}(t_{u(i)}^*)^{exp(\hat{\boldsymbol{p}}(\boldsymbol{x}^*)\boldsymbol{z}_{uj}^{*T})}\right) + \log \prod_{l\in C_{ui}^*} \hat{S}_{ox^*}(t_{u(i)}^*)^{exp(\hat{\boldsymbol{p}}(\boldsymbol{x}^*)\boldsymbol{z}_{ul}^{*T})}\right],$$

where the notations of $\ddagger^*$, $\ddagger_u^*$ imply $\ddagger$ under $\boldsymbol{x}^*$, $\boldsymbol{u}^*$, respectively, $\hat{\boldsymbol{p}}(\boldsymbol{x}^*)$ means the estimate of $\boldsymbol{p}$ derived from $\boldsymbol{x}^*$,

$$\hat{S}_{ox^*}(t_{u(i)}^*) = \hat{S}_o(t_{j-1}^*) - \delta_{ui}^*[\hat{\theta}_o(t_{(j-1)}^*)(t_{u(i)}^* - t_{(j-1)}^*)],$$
$$t_{(j-1)}^* \le t_{u(i)}^* < t_{(j)}^*, \quad j = 1, \ldots, k^*. \tag{3.20}$$

When $t_{u(i)}^* \ge t_{(k^*)}^*$, the tail of the LIC survival probability model must be defined as before,

1. $\hat{S}_{ox^*}(t_{u(i)}^*) = \hat{S}_o(t_{(k^*)}^*)[(r - t_{u(i)}^*)/(r - t_{(k^*)}^*)]$, $r = Mt_{c^*}$,

2. $\hat{S}_{ox^*}(t_{u(i)}^*) = \hat{S}_o(t_{(k^*)}^*)$.

Choice of model for tail

It is important to know which the model for tail good. We choose a good model for tail through results by simulation studies. The failure data y_i and the censored data c_i are generated from the Weibull accelerated model given by (2.33) with $(\alpha, \boldsymbol{p})= (0.9, 5.6, 0.9, 0.2)$ and the Weibull distribution given by (2.12) with $(\alpha, \beta)= (1.7, 2.6)$, respectively. The set of simulation data $\boldsymbol{x}$ is given by $t_i = min(y_i, c_i)$, $\delta_i = 1$ $(y_i \leq c_i)$ or $\delta_i = 0$ $(y_i > c_i)$, $\boldsymbol{z}_i= 1$ or 0 at random, $i = 1, \ldots, n$. Table 3.3 shows the ordered simulation data $(n = 50)$ and survival probabilities of baseline estimated by using the LIC survival probability model. In this example, the last failure is $i = 46$. There are 4 censored data after the last failure data. The simulation results by two models are summarized in Table 3.4. The bootstrap procedures for EIC are repeated 1000 times for X^* and 100 times for U^* in (3.13). The results by the model 1 shows that the choice of the constant M is not very critical. In this thesis, we adopt the model 1 $(M = 10)$. This model is the same as the LIKM model.

Table 3.1: Simulation data and survival probabilities estimated by using LIKM model (n_1=13, n_2=37).

i	*data* (t_i)	δ_i	*tie*	$\hat{S}(t_i)$	i	*data* (t_i)	δ_i	*tie*	$\hat{S}(t_i)$
1	0.0125	1	1	0.9800	26	6.2740	0	1	0.7371
2	0.0284	0	1	0.9800	27	6.5210	0	1	0.7371
3	0.1242	0	1	0.9800	28	6.8620	1	1	0.7050
4	0.6229	0	1	0.9800	29	7.4484	0	1	0.7050
5	0.7932	1	1	0.9587	30	8.1210	0	1	0.7050
6	0.9133	1	1	0.9374	31	8.1958	0	1	0.7050
7	1.0491	0	1	0.9374	32	8.8261	0	1	0.7050
8	1.1704	1	1	0.9156	33	8.9002	0	1	0.7050
9	1.2131	0	1	0.9156	34	9.0612	1	1	0.6636
10	1.6935	0	1	0.9156	35	9.2363	0	1	0.6636
11	1.7387	0	1	0.9156	36	9.3247	0	1	0.6636
12	2.0109	1	1	0.8921	37	9.8025	0	1	0.6636
13	2.0775	1	1	0.8686	38	10.3080	0	1	0.6636
14	2.5816	1	1	0.8452	39	10.8660	0	1	0.6636
15	2.8346	0	1	0.8452	40	11.5156	0	1	0.6636
16	3.1878	0	1	0.8452	41	11.6794	0	1	0.6636
17	3.2172	0	1	0.8452	42	11.7941	0	1	0.6636
18	3.5944	1	1	0.8196	43	11.9682	0	1	0.6636
19	3.6148	1	1	0.7939	44	12.8587	0	1	0.6636
20	3.7407	0	1	0.7939	45	13.4564	0	1	0.6636
21	4.2218	0	1	0.7939	46	15.2471	0	1	0.6636
22	4.6604	1	1	0.7666	47	19.3702	0	1	0.6636
23	5.5339	0	1	0.7666	48	22.2027	0	1	0.6636
24	5.5467	0	1	0.7666	49	23.3178	0	1	0.6636
25	5.7425	1	1	0.7371	50	28.6372	0	1	0.6636

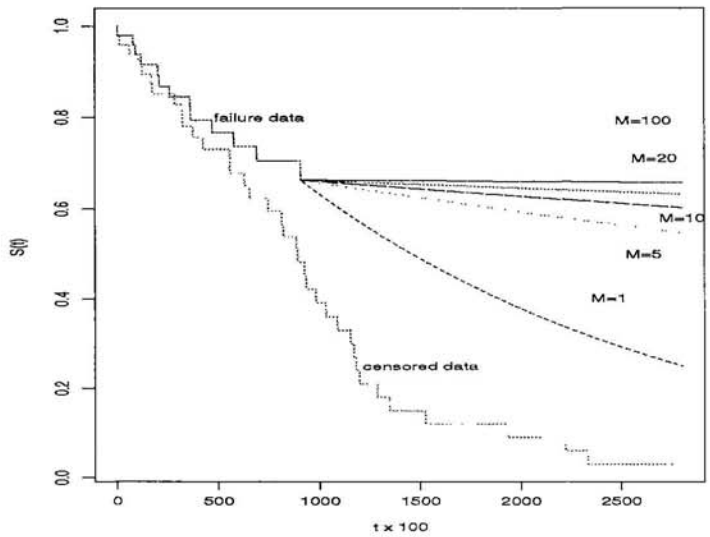

Figure 3.1: KM survival curves of failure and censoring estimated from simulation data, and survival curves of tail based on model 1, $\hat{S}(t) = \hat{S}(t_{(k)}) \exp[-(t - t_{(k)})/r]$, $r = M(t_c - t_{(k)})$, where $t_{(k)}$ =9.0612, t_c =28.6372, $\hat{S}(t_{(k)})$ =0.6636.

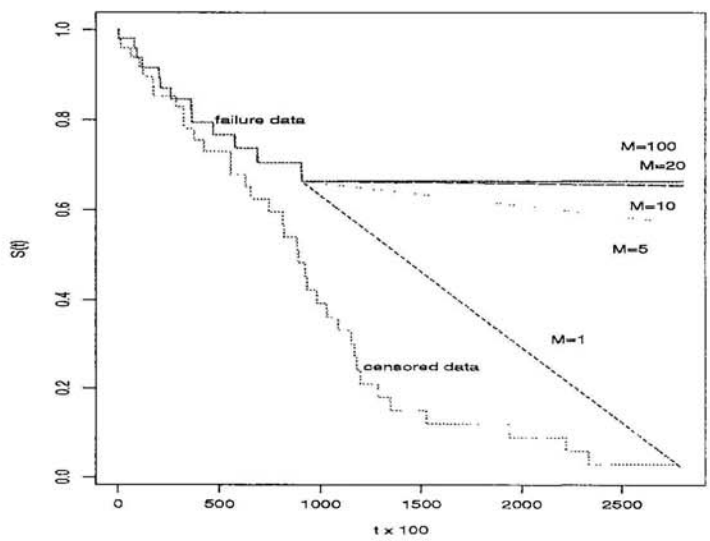

Figure 3.2: KM survival curves of failure and censoring estimated from simulation data, and survival curves of tail based on model 2, $\hat{S}(t) = \hat{S}(t_{(k)})[(r - t)/(r - t_{(k)})]$, $r = Mt_c$, where $t_{(k)}$ =9.0612, t_c =28.6372, $\hat{S}(t_{(k)})$ =0.6636.

Table 3.2: Simulation results by models for tails (n_1=13, n_2=37).

model	M	$E\{n_1^*\}$	$E\{n_2^*\}$	*bias*
	1	17.70	32.30	16.93
1	5	14.46	35.54	14.87
	10	14.09	35.91	15.17
	20	13.62	36.38	15.28
	100	13.65	36.35	16.03
	1	18.73	31.27	17.31
2	5	13.99	36.01	14.77
	10	13.80	36.20	15.22
	20	13.60	36.40	15.51
	100	13.53	36.47	16.34
3		31.84	18.16	21.74

Table 3.3: Simulation data and survival probabilities of baseline estimated by using LIC survival probability model (n_1=17, n_2=33).

i	*data* (t_i)	δ_i	*tie*	z_1	z_2	z_3	$\hat{S}(t_i)$	i	*data* (t_i)	δ_i	*tie*	z_1	z_2	z_3	$\hat{S}(t_i)$
1	0.0089	1	1	0	1	1	0.9617	26	1.4037	0	1	1	1	0	0.4424
2	0.0754	1	1	0	1	1	0.9234	27	1.4193	0	1	1	1	1	0.4424
3	0.0805	1	1	0	0	0	0.8851	28	1.5105	1	1	0	1	0	0.3798
4	0.1234	1	1	0	0	0	0.8467	29	1.6160	1	1	0	0	1	0.3207
5	0.1661	0	1	1	0	0	0.8467	30	1.8562	0	1	1	1	1	0.3207
6	0.2799	0	1	0	1	1	0.8467	31	1.8624	0	1	1	1	1	0.3207
7	0.3040	1	1	0	0	1	0.8065	32	2.0726	0	1	1	0	1	0.3207
8	0.3346	1	1	0	1	1	0.7650	33	2.2086	0	1	1	0	0	0.3207
9	0.3455	0	1	1	0	0	0.7650	34	2.3015	0	1	0	0	0	0.3207
10	0.5293	0	1	1	1	1	0.7650	35	2.3165	0	1	1	0	1	0.3207
11	0.5381	1	1	0	0	1	0.7235	36	2.5814	0	1	1	1	0	0.3207
12	0.5662	1	1	0	0	1	0.6805	37	2.6012	0	1	1	0	0	0.3207
13	0.6725	1	1	0	1	0	0.6359	38	2.6051	0	1	1	0	0	0.3207
14	0.7423	1	1	0	0	0	0.5923	39	2.6569	0	1	0	1	0	0.3207
15	0.7439	1	1	0	1	0	0.5488	40	2.7122	1	1	0	0	0	0.2207
16	0.7827	0	1	1	1	0	0.5488	41	2.7919	0	1	1	1	0	0.2207
17	0.9961	0	1	1	1	1	0.5488	42	3.0047	0	1	0	1	0	0.2207
18	1.0314	0	1	1	0	0	0.5488	43	3.0739	0	1	1	0	0	0.2207
19	1.1303	0	1	1	0	0	0.5488	44	3.0913	0	1	1	1	1	0.2207
20	1.1305	0	1	0	0	1	0.5488	45	3.1635	0	1	1	0	0	0.2207
21	1.2736	0	1	0	1	1	0.5488	46	3.2236	1	1	0	1	1	0.0834
22	1.3200	1	1	0	1	1	0.4957	47	3.2762	0	1	1	1	1	0.0834
23	1.3641	1	1	0	0	1	0.4424	48	3.5798	0	1	1	0	0	0.0834
24	1.3713	0	1	0	1	0	0.4424	49	4.3349	0	1	1	0	1	0.0834
25	1.3720	0	1	1	0	1	0.4424	50	4.6633	0	1	1	1	0	0.0834

Table 3.4: Simulation results by models for tails (n_1=17, n_2=33).

model	M	$E\{n_1^*\}$	$E\{n_2^*\}$	*bias*
	1	15.99	34.01	20.57
1	5	16.03	33.97	22.54
	10	16.03	33.97	22.47
	20	15.99	34.01	23.38
	100	15.93	34.07	24.83
2		20.04	29.96	21.42

Chapter 4

Simulation studies

In this chapter, it is shown through simulation studies that our EIC procedures are effective for analyzing the survival data. Firstly, the EIC procedures for simple survival data are handled by using the LIKM and parametric models. Secondly, the EIC procedures for survival data with covariates are handled by using the LIC survival probability model and parametric accelerated models. These EIC procedures are used to analyze the survival data of the esophageal cancer in the next chapter.

4.1 EIC procedures for simple survival data

It is shown that the EIC procedures given in the previous chapter can evaluate goodnesses of the survival probability models and handle comparisons between the LIKM model and parametric models. Furthermore, an explanation of the difference in the goodness between the LIKM model and the parametric models is given. It is also shown that the EIC procedures can evaluate goodnesses of models fitted to strata of categories. In the present study, the bootstrap procedures for EIC are repeated 1000 times for X^* and 100 times for U^* in (3.13).

4.1.1 Simple survival data

We try to know which survival probability model fits better to the simulation data derived from a log normal distribution. The failure time y_i is generated from the log normal distribution given by

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2 y_i^2}} \exp\left[-\frac{(\log y_i - \mu)^2}{2\sigma^2}\right], \tag{4.1}$$

where μ and σ^2 are mean and variance parameters, respectively. The censoring time c_i is generated from the exponential distribution given by

$$f_c(c_i) = \frac{1}{\lambda} \exp\left[-\frac{c_i}{\lambda}\right], \tag{4.2}$$

where λ is the scale parameter. The simulation data $\boldsymbol{x}$ are given by $t_i = min(y_i, c_i)$, $\delta_i = 1$ if $y_i \leq c_i$ or $\delta_i = 0$ otherwise, $i = 1, \ldots, n$. We assume $\mu = 1.5$, $\sigma^2 = 1.0$ and $\lambda = 8.0$.

Figure 4.1 and Figure 4.2 show the estimated survival curves of the Weibull and KM models for a set of the simulation data ($n = 50$), respectively. The error bands of the estimated curves are derived from the standard deviations of mutually independent bootstrap samples. Table 4.1 shows the mean values of information criteria for the Weibull, gamma and LIKM models for two cases of the simulation data ($n = 50$, $n = 15$). In this section, the mean values are obtained by 100 times repeating. In the table, $E\{diff.\}$ denotes the mean value of the difference between ELL and LL obtained by the Monte Carlo procedure instead of the bootstrap procedure in (3.13). $E\{bias\}$ denotes the mean value of the bias correction obtained by the EIC procedure. The bands of E{diff.} and E{bias} are derived from the

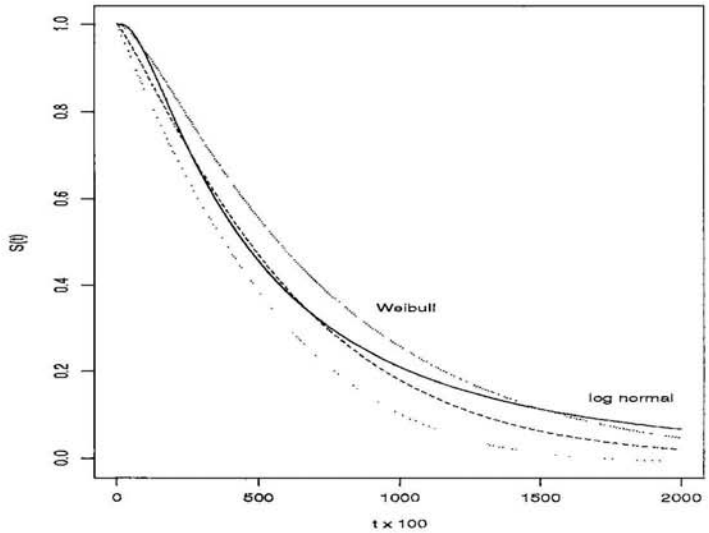

Figure 4.1: True log normal survival curve (solid line) and Weibull survival curve (middle dotted line) with band estimated from simulation data. Band between upper and lower dotted line indicates variance of estimation derived from bootstrap procedure.

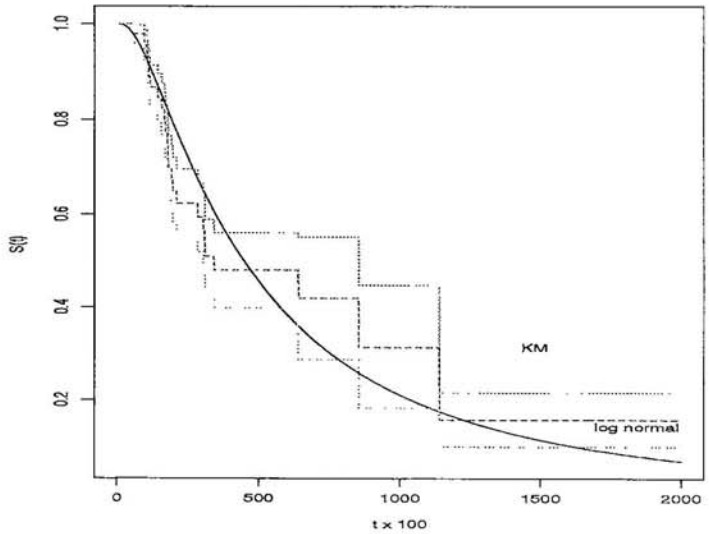

Figure 4.2: True log normal survival curve (solid line) and KM survival curve (middle dotted line) with band estimated from simulation data. Band between upper and lower dotted line indicates variance of estimation derived from bootstrap procedure.

standard deviations of estimates. The value of $E\{bias\}$ for each model is close to the value of $E\{diff.\}$ for each model. Therefore it is shown that EIC is suitable for the estimator of -2 times ELL under the existence of the censored data. When the sample size is small, $E\{diff.\}$ of the Weibull model is larger than the number of free parameters 2. In such a case, EIC is more suitable than AIC. It is shown by EIC for the simulation data ($n = 50$) that the LIKM model is worse than the Weibull and gamma models. The information criteria say the LIKM model with many parameters increases the bias correction. Then the LIKM model is worse for fitting to large sample.

However, other cases can be considered. The simulation data $\boldsymbol{x}_1$ are generated from the failure data and the censored data of the Weibull distributions with (α, β) =(20.0, 5.0) denoted by $S_1(t)$ and (50.0, 3.0), respectively, and (n_1, n_2) =(48, 2). n_1 and n_2 denote the numbers of the failures and censored data, respectively. The simulation data $\boldsymbol{x}_2$ are generated from the failure data and the censored data of the Weibull distributions with (α, β) =(1.0, 3.0) denoted by $S_2(t)$ and (8.0, 1.0), respectively, (n_1, n_2) =(27, 3). $\boldsymbol{x}_{mix}$ denotes the mixtured data of those with sample size $(n_1, n_2) = (75, 5)$. EIC of the Weibull and LIKM models for $\boldsymbol{x}_{mix}$ are 530.47 and 454.21, respectively. In this case the LIKM model is better. Figure 4.3 shows the survival curve of true structure derived from $(S_1(t) + S_2(t))/2$, the KM and Weibull survival curves estimated from $\boldsymbol{x}_{mix}$. This example demonstrates that the LIKM model is better for fitting to the data derived from a complex structure.

In practice mixture data are handled to estimate the survival probability for

cured patients. The cured patients mean the long-term survivors as well as the general population. Farewell (1983), Greenhouse and Wolfe (1984) proposed the mixture models for cured data.

4.1.2 Grouped survival data

The simulation data $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are generated from the failure data of the Weibull distributions with (α, β)= $(5.0, 2.0)$, $(3.0, 1.5)$, $(2.0, 0.9)$, respectively, and the censored data of the exponential distributions with $\lambda = 8.0$. The sample sizes of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are $(n_1, n2)$ =$(30, 20)$, $(25, 5)$, $(19, 1)$, respectively. $\boldsymbol{x}_{mix}$ denotes the mixtured data of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. We define a grading system as a set of strata in data classified from $\boldsymbol{x}_{mix}$. One grading system is denoted by $\{\boldsymbol{x}_1\}$, $\{\boldsymbol{x}_2\}$, $\{\boldsymbol{x}_3\}$, where $\{\cdot\}$ means a stratum of classified data. EIC makes it possible to compare the goodness of the survival probability model for the grading system with that of the survival probability model fitted to $\boldsymbol{x}_{mix}$. The value of EIC for the model fitted to the grading system is the summation of the values of EIC for models fitted to $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$, since is is assumed that each model for the stratum is independent of each other.

Table 4.2 shows EIC of the Weibull and LIKM models for the grouped simulation data. EIC_L and EIC_U are the low and upper estimates which are derived from the standard deviation of the bias correction, respectively. The values of EIC for the models for the grading system $\{\boldsymbol{x}_1\}$, $\{\boldsymbol{x}_2\}$, $\{\boldsymbol{x}_3\}$ are less than those for $\boldsymbol{x}_{mix}$. Therefore, the model for the grading system is more suitable than the model for $\boldsymbol{x}_{mix}$. Other grading systems of the simulation data can be assumed. The values

Table 4.1: Results of simulation studies of Weibull, gamma and LIKM models, where error bands of E{diff.} and E{bias} are derived from the standard deviations of estimates.

n	*model*	$E\{n_1\}, E\{n_2\}$	$E\{AIC\}$	E{EIC}	$E\{diff.\}$	$E\{bias\}$
	Weibull		158.37	158.76	-1.16 2.53 6.21	-0.21 2.20 4.61
50	Gamma	26.88,23.12	157.46	158.31	-0.73 2.26 5.25	-2.22 2.42 7.07
	LIKM			172.96	14.52 22.85 31.18	20.02 27.07 34.12
	Weibull		49.59	51.95	-7.64 4.15 15.94	-3.35 3.18 9.71
15	Gamma	8.29,6.71	49.44	50.35	-0.80 2.60 6.00	-2.31 2.45 7.22
	LIKM			50.56	1.89 6.53 11.17	4.05 8.14 12.22

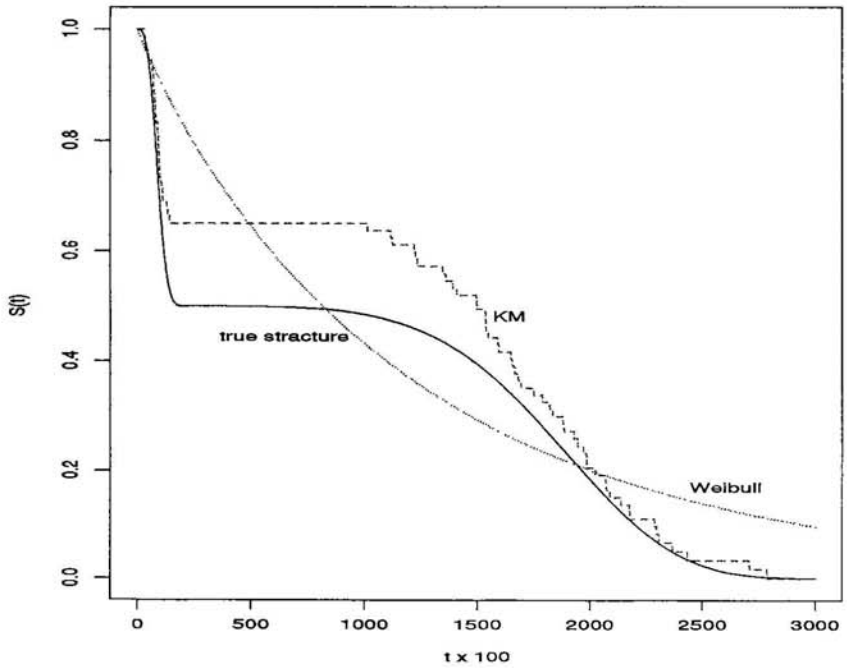

Figure 4.3: Survival curve derived from true structure $(S_1(t) + S_2(t))/2$, KM and Weibull survival curves estimated from simulation data $\boldsymbol{x}_{mix}$. Values of EIC for Weibull and LIKM models are 530.47 and 454.21, respectively.

of EIC for the other grading systems are summarized in Figures 4.4 and 4.5. It is shown that the grading system $\{\boldsymbol{x}_1\}$, $\{\boldsymbol{x}_2\}$, $\{\boldsymbol{x}_3\}$ is the best. The Weibull model for the grading system is better than the LIKM model. In this case, the true structure of generating data is estimated by the EIC procedures.

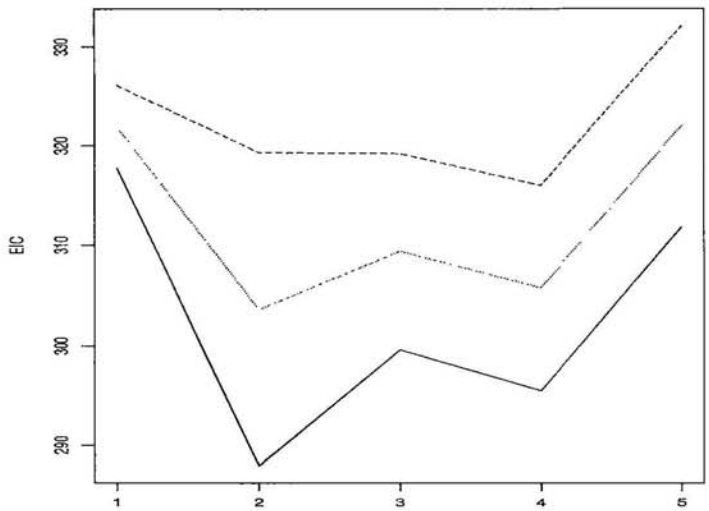

Figure 4.4: EIC_U, EIC, EIC_L of Weibull models for grading systems based on simulation data, where 1:$\boldsymbol{x}_{mix}$, 2:$\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}$, 3:$\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}$, 4:$\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3\}$, 5:$\{\boldsymbol{x}_1, \boldsymbol{x}_3\}, \{\boldsymbol{x}_2\}$.

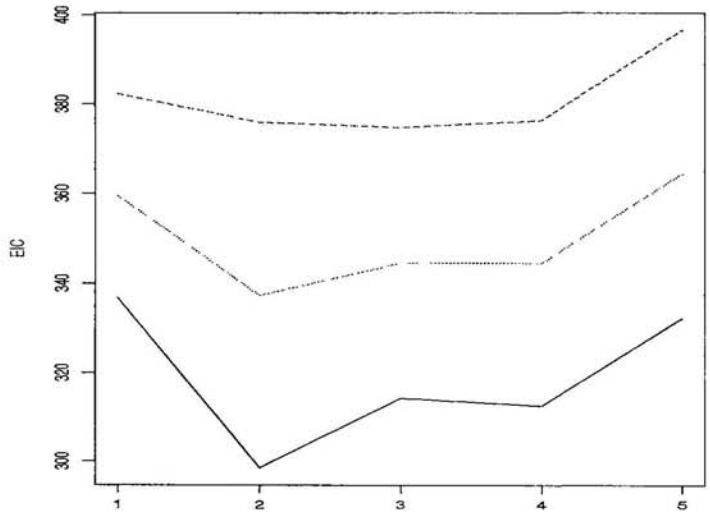

Figure 4.5: EIC_U, EIC, EIC_L of LIKM models for grading systems based on simulation data, where 1:$\boldsymbol{x}_{mix}$, 2:$\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}$, 3:$\{\boldsymbol{x}_1, \boldsymbol{x}_2\}, \{\boldsymbol{x}_3\}$, 4:$\{\boldsymbol{x}_1\}, \{\boldsymbol{x}_2, \boldsymbol{x}_3\}$, 5:$\{\boldsymbol{x}_1, \boldsymbol{x}_3\}, \{\boldsymbol{x}_2\}$.

4.2 EIC procedures for survival data with covariates

It is shown that the EIC procedures given in the previous chapter can evaluate goodnesses of the survival probability models with covariates and handle comparisons between the LIC survival probability model and parametric accelerated models. Furthermore, an explanation of the difference in the goodness between the LIC survival probability model and the parametric accelerated models is given. It is also shown that the EIC procedures can evaluate goodnesses of models fitted to strata of categories with covariates. The bootstrap procedures for EIC are repeated 1000 times for X^* and 100 times for U^* in (3.13).

4.2.1 Survival data with covariate

The baseline failure time y_{oi} is generated from the log normal distribution given by

$$f(y_{oi}) = \frac{1}{\sqrt{2\pi\sigma^2 y_{oi}^2}} \exp\left[-\frac{(\log y_{oi} - \mu)^2}{2\sigma^2}\right], \tag{4.3}$$

where μ and σ^2 are mean and variance parameters. The failure time y_i is given by

$$y_i = y_{oi} \exp(-\boldsymbol{p}\boldsymbol{z}_i^T). \tag{4.4}$$

The censoring time c_i is generated from the exponential distribution given by $f_c(c_i) = e^{(-c_i/\lambda)}/\lambda$, where λ is the scale parameter. The simulation data $\boldsymbol{x}$ are given by $t_i = min(y_i, c_i)$, $\delta_i = 1$ if $y_i \geq c_i$ or $\delta_i = 0$ otherwise, $i = 1, \cdots, n$. We assume $\mu = 0.5$, $\sigma^2 = 1.0$, $\lambda = 8.0$ $\boldsymbol{p}$= $(0.3, 0.2, 0.5)$ and $\boldsymbol{z}_i$= 1 or 0 at random.

Figures 4.6, 4.7 and 4.8 show the estimated survival curves of the Weibull, gamma accelerated model and the Cox survival probability models for a simulation data

set ($n = 50$) under $\boldsymbol{z} = (1, 1, 1)$, respectively. The error bands of the estimated curves are derived from the standard deviations of mutually independent bootstrap samples. Note that the true structure does not have a proportional hazard. Then the estimated curves might be over with the true curve.

Table 4.3 shows the mean values of information criteria for the Weibull, gamma accelerated models and the LIC survival probability model under the simulation data. In this section, the mean values are obtained by 100 times repeating. In the table, $E\{diff.\}$ denotes the mean value of the difference between ELL and LL derived from the Monte Carlo procedure instead of the bootstrap procedure in (3.13), and $E\{bias\}$ denotes the mean value of the bias correction in EIC. The bands of E{diff.} and E{bias} are derived from the standard deviations of estimates. The value of $E\{bias\}$ for each model is close to the value of $E\{diff.\}$ for each model. Therefore it is shown that EIC is suitable for the estimator of -2 times ELL under the existence of the censored data. When the sample saize is small, $E\{diff.\}$ of the Weibull accelerated model is larger than the number of the free parameters 4, thus EIC is more suitable than AIC. It is shown by EIC that when the sample size is large, the Weibull and gamma accelerated models are better than the LIC survival probability model, since the nonparametric model with many parameters increases the bias correction.

Next the mixture data are considered. The failure times of simulation data are generated from the Weibull distribution $S_o(t_o) = \exp[-(t/\alpha)^\beta]$, $t = t_o \exp(-\boldsymbol{p}\boldsymbol{z}^T)$. The censoring times of the simulation data are generated from the Weibull distri-

Table 4.2: EIC of Weibull and LIKM models for grouped simulation data. EIC_L and EIC_U are low and upper estimates which are derived from standard deviation of bias correction, respectively.

n	*model*	n_1, n_2	*Weibull*				*LIKM*			
			EIC_L	EIC	EIC_U	*bias*	EIC_L	EIC	EIC_U	*bias*
100	$\boldsymbol{x_{mix}}$	74,26	317.69	321.85	326.02	2.10	336.74	359.50	382.26	76.48
50	$\boldsymbol{x_1}$	30,20	136.86	141.90	146.94	2.23	143.31	158.28	173.24	30.72
30	$\boldsymbol{x_2}$	25,5	88.86	94.18	99.50	2.31	95.17	108.36	121.55	21.45
20	$\boldsymbol{x_3}$	19,1	62.16	67.51	72.85	2.26	60.08	70.67	81.26	17.43
	grading system		287.88	303.59	319.29		298.56	337.31	376.05	

Table 4.3: Results of simulation studies of Weibull, gamma accelerated models and LIC survival probability models, where bands of E{diff.} and E{bias} are derived from the standard deviations of estimates.

model	$E\{n_1\}, E\{n_2\}$	$E\{AIC\}$	E{EIC}	$E\{diff.\}$	$E\{bias\}$
Weibull		69.30	70.64	0.17 4.42 8.68	0.42 4.67 8.92
Gamma	23.86, 26.14	66.61	65.60	-0.55 3.42 7.03	0.43 3.50 6.88
LIC			86.06	16.26 24.75 33.23	21.75 29.28 36.81
Weibull		32.53	38.19	-1.33 5.77 12.86	-2.62 6.83 16.28
Gamma	10.52,11.48	32.65	35.28	-0.79 4.15 9.09	-2.37 5.31 12.99
LIC			46.70	4.76 15.15 25.54	7.34 18.80 30.26

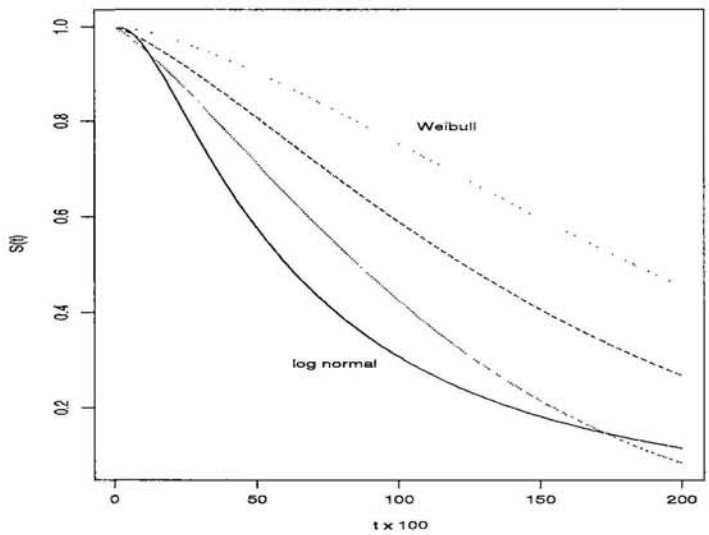

Figure 4.6: True log normal survival curve and Weibull survival curve with band estimated from simulation data under $\boldsymbol{z} = (1, 1, 1)$.

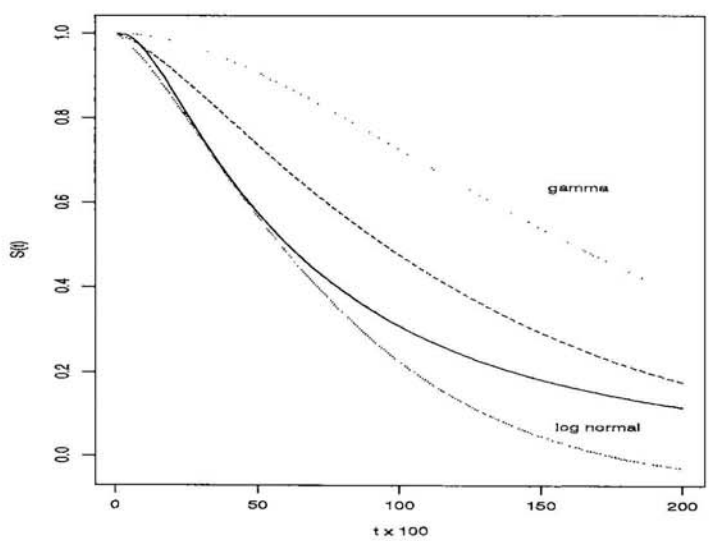

Figure 4.7: True log normal survival curve and gamma survival curve with error band estimated from simulation data under $\boldsymbol{z} = (1, 1, 1)$.

bution with (β_c, α_c). $\boldsymbol{x}_1$ are given by $(\beta, \alpha, \boldsymbol{p}) = (1.0, 3.0, 0.5, 0.7, 0.3)$ and (β_c, α_c) $= (1.0, 8.0)$. $\boldsymbol{x}_2$ are given by $(\beta, \alpha, \boldsymbol{p}) = (5.0, 20.0,\ 0.5, 0.7, 0.3)$ and $(\beta_c, \alpha_c) = (3.0, 30.0)$. The true baselines of the survival probabilities for $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are denoted by $S_{o1}(t)$ and $S_{o2}(t)$, respectively. Figure 4.9 shows the estimated survival curves derived from the mixture simulation data $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ ($n_1 = 68, n_2 = 12$). The values of EIC for the Weibull accelerated model and the LIC survival probability model are 503.00 and 433.02, respectively. When the true structure is complex, the nonparametric model is better the the parametric model.

4.2.2 Grouped survival data with covariates

The simulation data $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are generated from the failure data of the Weibull distributions with the regression models (2.33) under $(\beta, \boldsymbol{p}) = (1.0, 0.2, 0.2, 0.4)$, $(0.8, 0.8,\ 0.2, 0.5)$, $(3.5, 0.2, 0.2, 0.1)$, respectively, and the censored data of the exponential distributions with $\lambda = 3.0$. $\boldsymbol{x}_{mix}$ is the mixture of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$. We define a grading system as a set of strata of data classified from $\boldsymbol{x}_{mix}$. A grading system is denoted by $\{\boldsymbol{x}_1\}$, $\{\boldsymbol{x}_2\}$, $\{\boldsymbol{x}_3\}$. EIC makes it possible to compare the goodness of the survival probability model for the grading system with that of the survival probability model fitted to $\boldsymbol{x}_{mix}$. The value of EIC for the grading system is the summation of the values of EIC for models fitted to the strata. Table 4.4 shows EIC of the Weibull accelerated model and LIC survival probability model for the simulation data, where EIC_L and EIC_U are the low and upper error estimates which are derived from the standard deviation of the bias correction, respectively. The values of EIC for the grading system are less than those for $\boldsymbol{x}_{mix}$. It is suggested that the

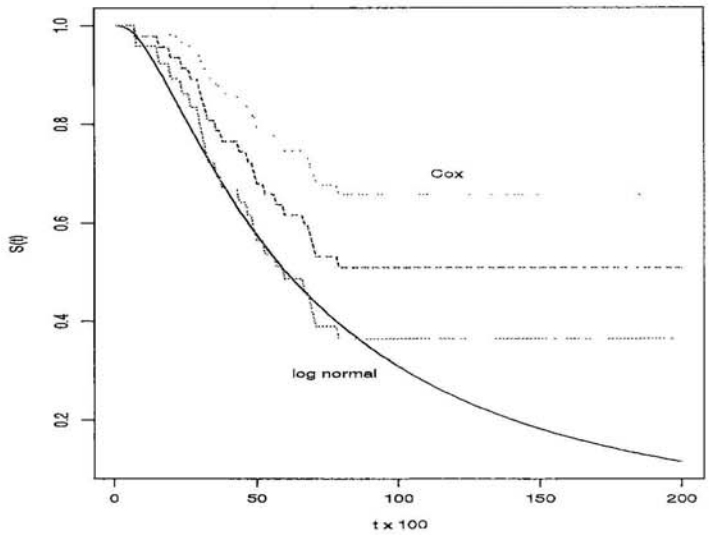

Figure 4.8: True log normal survival curve and LIC survival curve with error band estimated from simulation data under $\boldsymbol{z} = (1, 1, 1)$.

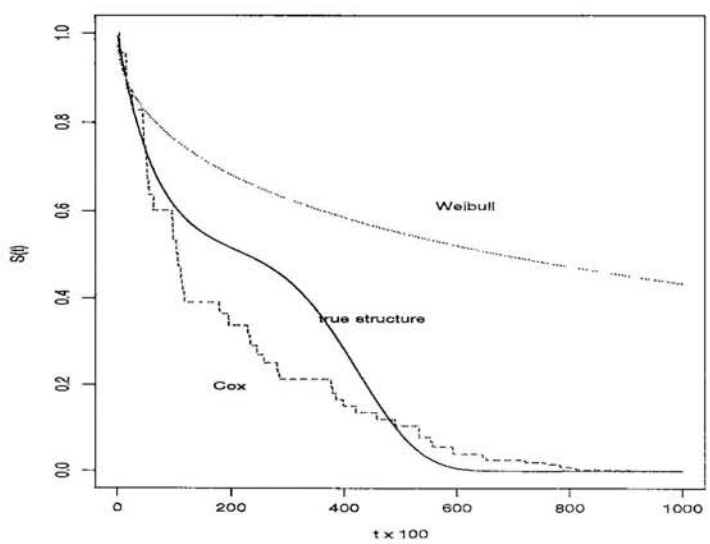

Figure 4.9: True survival curve $(S_{o1}(t)^{\exp(\boldsymbol{p}\boldsymbol{z}^T)} + S_{o2}(t)^{\exp(\boldsymbol{p}\boldsymbol{z}^T)})/2$, Cox and Weibull accelerated survival curve estimated from mixture simulation data under $\boldsymbol{z} = (1, 1, 1)$.

Table 4.4: EIC of Weibull accelerated models and LIC survival probability models for grouped simulation data. EIC_L and EIC_U are low and upper estimates which are derived from standard deviation of bias correction, respectively.

n	*model*	*n1,n2*	*Weibull*			*LIC*		
			EIC_L EIC EIC_U	*bias*		EIC_L EIC EIC_U	*bias*
105	$\boldsymbol{x}_{mix}$	64,41	211.32 220.06 228.80	4.78		247.17 267.63 288.08	67.05
50	$\boldsymbol{x}_1$	33,17	94.44 101.94 109.43	4.18		107.91 126.21 144.51	38.58
30	$\boldsymbol{x}_2$	15,15	61.01 78.91 96.80	6.21		60.60 83.17 105.75	24.80
25	$\boldsymbol{x}_3$	16,9	-3.06 22.85 48.77	7.33		10.39 47.23 84.06	31.24
	grading system		152.39 203.70 255.00			178.90 256.61 334.32	

grading system is more suitable than the model for $\boldsymbol{x}_{mix}$.

Next we consider the following case. The simulation data $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are generated from the failure data of the Weibull distributions with the regression models (2.33) under $(\beta, \boldsymbol{p}) = (1.0, 0.2, 0.2, 0.4)$, $(0.8, 0.2, 0.2, 0.4)$, $(3.5, 0.2, 0.2, 0.4)$, respectively. These failure data are generated from the models of common regression parameter $\boldsymbol{p}$. The censored data are generated from the exponential distributions with $\lambda = 3.0$. Table 4.5 shows EIC of the Weibull accelerated model and LIC survival probability model for the simulation data. The values of EIC for the grading system are less than those for $\boldsymbol{x}_{mix}$. The EIC procedure can estimate the true structure.

We consider the following case. The failure data of $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are generated from the Weibull distributions with the regression models (2.33) under $(\beta, \boldsymbol{p})$ $= (1.2, 0.2, 0.2, 0.4)$, $(1.2, 0.8, 0.2, 0.5)$, $(1.2, 0.2, 0.2, 0.1)$, respectively. These failure

data are generated from the models of common shape parameter β. The censored data are generated from the exponential distributions with $\lambda = 3.0$. Table 4.6 shows EIC of the Weibull accelerated model and LIC survival probability model for the simulation data. The values of EIC for the grading system are not less than those for $\boldsymbol{x}_{mix}$. In this case, the EIC procedure can not estimate the true structure.

The baseline hazard rate function of the Weibull accelerated model is given by $h_o(t) = \beta t^{\beta-1}$, then the mode of baseline depends on the shape parameter β. The EIC procedure may be sensitive about the mode of baseline.

Table 4.5: EIC of Weibull accelerated models and LIC survival probability models for grouped simulation data (common regression parameter). EIC_L and EIC_U are low and upper estimates which are derived from standard deviation of bias correction, respectively.

n	*model*	*n1,n2*	*Weibull*			*LIC*		
			EIC_L EIC EIC_U	*bias*		EIC_L EIC EIC_U		*bias*
105	$\boldsymbol{x_{mix}}$	66,39	206.38 214.89 223.40	4.61		255.38 275.94 296.51		66.87n
50	$\boldsymbol{x_1}$	33,17	94.44 101.94 109.43	4.18		107.91 126.21 144.51		38.58
30	$\boldsymbol{x_2}$	18,12	58.13 81.33 104.53	6.78		72.22 94.17 116.12		24.42
25	$\boldsymbol{x_3}$	15,10	1.48 28.66 55.84	7.61		9.28 45.57 81.85		29.80
	grading system		154.05 211.93 269.80			189.41 265.95 342.57		

Table 4.6: EIC of Weibull accelerated models and LIC survival probability models for grouped simulation data (common shape parameter). EIC_L and EIC_U are low and upper estimates which are derived from standard deviation of bias correction, respectively.

	model	*n1,n2*	*Weibull*			*LIC*		
			EIC_L EIC EIC_U	*bias*		EIC_L EIC EIC_U		*bias*
105	$\boldsymbol{x_{mix}}$	63,42	202.97 210.35 217.73	4.21		242.96 264.47 285.97		67.82
50	$\boldsymbol{x_1}$	33,17	92.91 100.18 107.45	4.06		111.94 130.81 149.69		37.34
30	$\boldsymbol{x_2}$	16,14	53.11 77.38 101.64	7.42		70.35 96.11 121.87		26.83
25	$\boldsymbol{x_3}$	14,11	31.88 52.14 72.40	7.08		43.59 68.27 92.94		22.77
	grading system		177.90 229.70 281.49			225.88 295.19 364.50		

Chapter 5

Real data analysis

In this chapter we analyze survival data in cancer of the thoracic esophagus based on the EIC procedures. A set of cancer data involving 143 patients (56 failures and 87 censored patients) was obtained in the department of surgery in Cancer Hospital from Jan. 1985 through Oct. 1992, (Matsubara, 1992., Yafune et al., 1993.). The observations of patient are the survival time after the surgical operation and the status of invasions at the surgical operation, the number of invaded lymph nodes (N factor), the depth of cancer invasion (T factor) and the distribution pattern of lymphatic invasion. We assume that the surgical operations are assigned at random during the period of investigation and then the survival data can be treated as the randomly censored data.

Surgical outcomes are evaluated by survival probabilities estimated from survival data after surgical operations. It is known that the surgical outcomes are influenced by malignant invasions. Classifications into several groups using grading systems of invasions are used for therapies. The grading systems mean the grouping patterns of invasions. It is quite important, therefore, that how good the grading system of

the factors for evaluating the surgical outcomes is. But it is difficult to know which of grading systems is the best, because the patterns of lymphatic metastasis in the esophageal cancer are complicated. Matsubara (1992) searched for the factors affecting the surgical outcomes in cancer of the thoracic esophagus by using the generalized Wilcoxon test and the KM survival probability models. As another method, Matsubara et al. (1994) proposed the use of AIC for the comparisons of the strata of the invasions categorized by the factors affecting surgical outcomes using the Weibull distribution. They showed that the AIC procedures of searching grading systems are convenient for estimating relations between the factors and the surgical outcomes. Then the better grading system in cancer of the thoracic esophagus was proposed based on results by AIC. However the AIC procedure can not evaluate the goodness of nonparametric model.

In the first section, the EIC procedures are used to compare the goodnesses of the LIKM model with those of the Weibull and gamma models, and then the EIC procedures compare the goodnesses of the models fitted to strata of categories. Then the results by EIC give the better grading system of the factors affecting the surgical outcomes.

However the complicated relations between the factors and the surgical outcomes are not shown enough by the comparisons of models fitted to strata categorized by the factors. The analysis of survival data with the multivariate covariates is useful for the detailed search. In the second section, we use the EIC procedures to compare the LIC survival probability model with parametric accelerated models and to search

for the multiple factors affecting the surgical outcomes. Then the EIC procedures give the better survival model with the covariates of the factors.

5.1 Grading systems of simple factors

The number of invaded lymph nodes (N factor) is thought to be one of the most important factors affecting surgical outcomes in cancer of the thoracic esophagus. Clinically the N factor is convenient for the classification of the phases of lymphatic invasion. However several other systems such as the depth of cancer invasion (T factor) and the distribution pattern of lymphatic invasion may be important for the classification. Then we compare the goodness of grading systems based on the above factors by the EIC procedures. Note that the survival data (day) are divided by 100.

Table 5.1 shows EIC of the Weibull, gamma and LIKM models for the full data of the esophageal cancer. The full data means the data which are not classified into the groups. Figure 5.1 shows the survival curves of the models estimated from the full data. Table 5.2 shows the estimated parameters and 50 percentile points of the survival models for the full data. It is found in Table 5.1 that the parametric models are better than the LIKM model, and the values of EIC for the Weibull and gamma models are close. The estimated shape parameters of the parametric models in Table 5.2 are close to 1, which means the similarity to the exponential distributions. Then we accept the Weibull model as the parametric model in the following analysis.

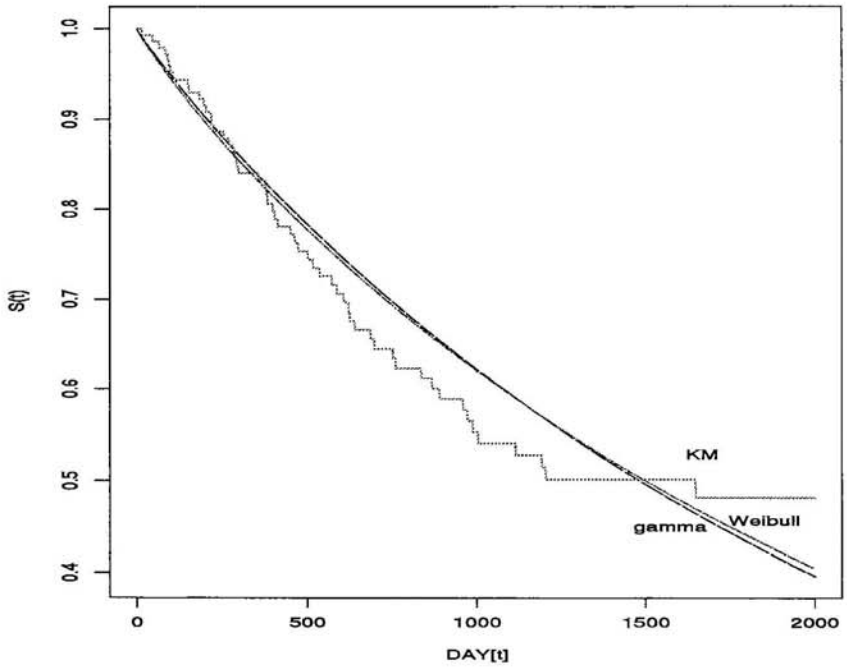

Figure 5.1: Weibull, gamma and KM survival curves estimated from full cancer data.

5.1.1 N factor

Table 5.3 shows the numbers of patients grouped by the N factor. Table 5.4 shows the values of EIC for the Weibull and LIKM models for the grading system of the N factor, $\{0,1\}$,$\{2 \sim 7\}$, $\{8 \sim\}$. The value of EIC for the grading system is derived from adding the value of each stratum. In this case, the values of EIC for the grading system are less than those for full data. Therefore the model fitted to the grading system is better than the model fitted to the full data. The value of EIC of the Weibull models for the grading system is less than that of the LIKM model. The estimated survival curve of the Weibull model for each group is plotted in Figure 5.2. The estimated parameters and 50 percentile points of the Weibull models are summarized in Table 5.5.

It is possible to divide the data set into another groups by using the N factor. Figures 5.3 and 5.4 show the values of EIC for the another grading systems. It is found that the grading system of Table 5.4 is the best.

Table 5.6 shows the results of the generalized Wilcoxon tests about the N factor. The table shows that the grading system $\{0,1\}$,$\{2 \sim\}$ is proposed. However, it is difficult for nonparametric tests to compare the grading system of one factor with that of the other factor.

Table 5.1: EIC of Weibull, gamma and LIKM models for full cancer data.

Model	n_1, n_2	EIC_L	*EIC*	EIC_U	*bias*
Weibull		452.81	457.02	461.24	2.04
Gamma	56,87	432.71	458.89	485.06	2.78
LIKM		488.27	507.68	527.09	55.77

Table 5.2: Estimates of Weibull, gamma and LIKM models for full cancer data.

		estimates	hazard	*50 % point (day)*
Weibull	$(\hat{\beta}, \hat{\alpha})$=	(0.92, 22.31)	DHR	1499
Gamma	$(\hat{a}, \hat{b})$=	(0.95, 23.16)	DHR	1482
LIKM				1317

Table 5.3: Numbers of patients grouped by N factor.

N factor	*patients*	*failures*	*censored patients*
0	44	6	38
1	25	7	18
2,3	24	11	13
4∼7	24	13	11
8∼	26	19	7

Table 5.4: EIC of Weibull and LIKM models for data divided by N factor.

grade	n_1, n_2	*Weibull*				*LIKM*			
		EIC_L	*EIC*	EIC_U	*bias*	EIC_L	*EIC*	EIC_U	*bias*
full data	56,87	452.81	457.02	461.24	2.04	488.27	507.68	527.09	55.77
0,1	13,56	130.30	134.72	139.14	1.99	135.15	144.26	153.37	12.93
2∼7	24,24	173.72	178.41	183.10	2.14	176.66	189.70	202.74	25.74
8∼	19,7	97.43	104.89	112.34	2.60	94.16	105.79	117.41	17.70
grading system		401.45	418.02	434.58		405.97	439.75	473.52	

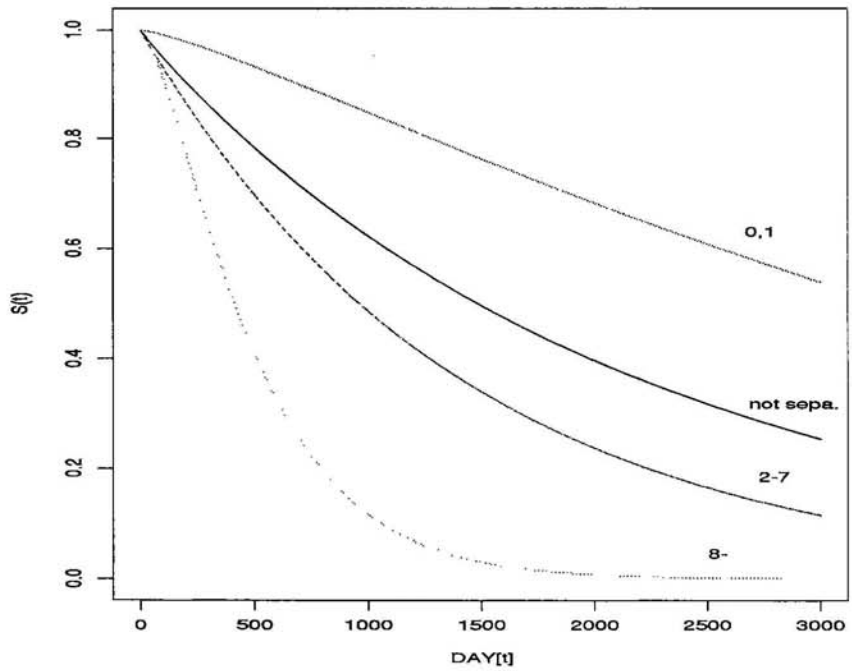

Figure 5.2: Weibull survival curves estimated from data divided by N factor.

Table 5.5: Estimates of Weibull models for data divided by N factor.

	$\hat{\beta}$	$\hat{\alpha}$	hazard	*50 % point (day)*
0,1	1.18	45.89	IHR	3365
2∼7	0.95	14.15	DHR	963
8∼	1.41	5.52	IHR	426

Table 5.6: Results of generalized Wilcoxon tests about N factor, where S: to be significant (significant level $p < 0.05$), N.S: not to be significant.

null hypothesis	*result*
H_0:{0}={1}	N.S
H_0:{0}={2,3}	S
H_0:{0}={4∼7}	S
H_0:{0}={8∼}	S
H_0:{1}={2,3}	S
H_0:{1}={4∼7}	S
H_0:{1}={8∼}	S
H_0:{2,3}={4∼7}	N.S
H_0:{2,3}={8∼}	N.S
H_0:{4∼7}={8∼}	N.S

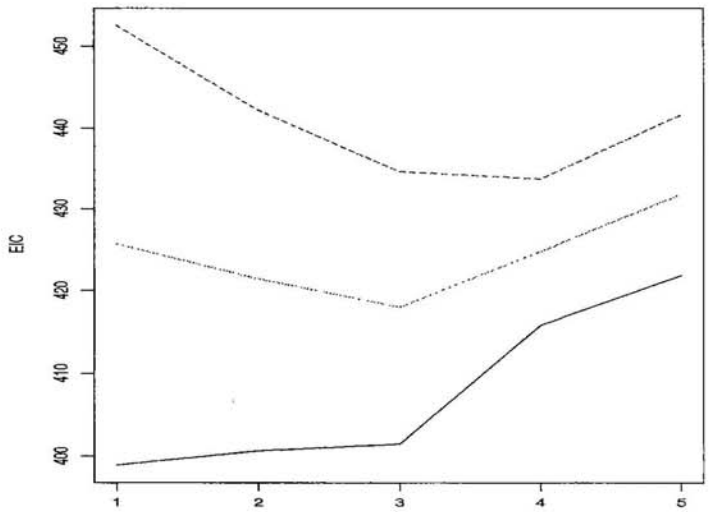

Figure 5.3: EIC_U, EIC, EIC_L of Weibull models for grading systems of N factor, where 1:$\{0\}$, $\{1\}$, $\{2,3\}$, $\{4 \sim 7\}$, $\{8 \sim\}$, 2:$\{0\}$, $\{1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$, 3:$\{0,1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$, 4:$\{0,1\}$, $\{2 \sim\}$, 5:$\{0 \sim 7\}$, $\{8 \sim\}$.

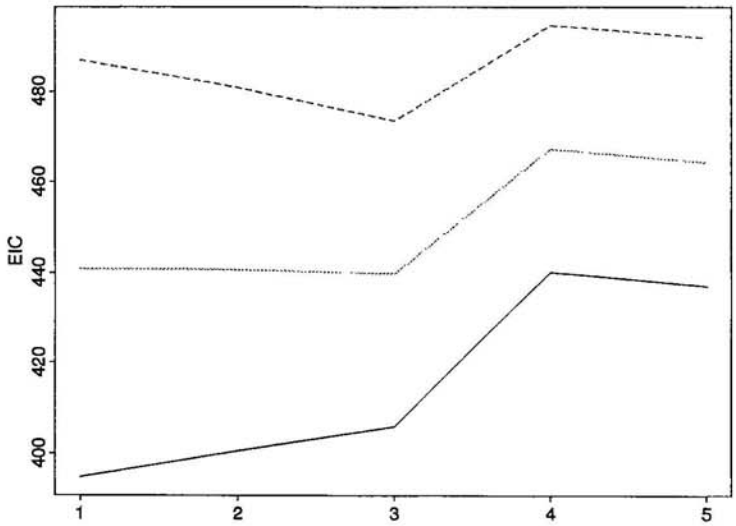

Figure 5.4: EIC_U, EIC, EIC_L of LIKM models for grading systems of N factor, where 1:$\{0\}$, $\{1\}$, $\{2,3\}$, $\{4 \sim 7\}$, $\{8 \sim\}$, 2:$\{0\}$, $\{1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$, 3:$\{0,1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$, 4:$\{0,1\}$, $\{2 \sim\}$, 5:$\{0 \sim 7\}$, $\{8 \sim\}$.

5.1.2 T factor

Clinically the notations of the T factor are represented as ep, m, sm, pm, $a1$, $a2$, $a3$. Table 5.7 shows the numbers of patients grouped by the T factor. Several grading systems of the T factor can be considered. Figures 5.5 and 5.6 show the values of EIC for the grading systems of the T factor. The value of EIC for the grading system 10:$\{m, sm, pm\}$, $\{a1, a2, a3\}$ of the Weibull model is the least in Figure 5.5. The above grading system is not the best in Figure 5.6, but the grading system 4:$\{m, sm\}$, $\{pm\}$, $\{a1\}$, $\{a2\}$, $\{a3\}$ is the best. However the value of EIC for 10 of the Weibull model is less than that for 4 of the LIKM model. Table 5.8 shows the values of EIC for the Weibull and LIKM models about the grading system 10. The values of EIC for the grading system are smaller than those of the models for the full data. The estimated survival curves of the Weibull models for the grading system are plotted in Figure 5.7. The estimated parameters and 50 percentile points of the Weibull models are summarized in Table 5.9.

5.1.3 Distribution pattern of lymphatic invasion

In the esophageal cancer, the surgical outcomes are affected by the distribution pattern of lymphatic invasion, since the lymph nodes around the thoracic esophagus are complicated. It is important, therefore, to know how the distribution pattern of lymphatic invasion is useful for evaluating the surgical outcomes. We denote the lymph nodes as follows:

Table 5.7: Numbers of patients grouped by T factor.

T factor	*patients*	*failures*	*censored patients*
ep	0		
m	6	1	5
sm	42	8	34
pm	10	3	7
a1	14	11	3
a2	65	30	35
a3	6	3	3

Table 5.8: EIC of Weibull and LIKM models for data divided by T factor.

grade	n_1, n_2	*Weibull*				*LIKM*			
		EIC_L	EIC	EIC_U	*bias*	EIC_L	EIC	EIC_U	*bias*
full data	56,87	452.81	457.02	461.24	2.04	488.27	507.68	527.09	55.77
m, sm, pm	12,46	115.98	120.65	125.32	2.09	117.10	126.86	136.62	12.77
a1, a2, a3	44,41	318.53	323.00	327.48	2.13	339.57	355.89	372.22	45.38
grading system		434.51	443.65	452.80		456.67	482.75	508.84	

Table 5.9: Estimates of Weibull models for data divided by T factor.

	$\hat{\beta}$	$\hat{\alpha}$	hazard	*50 % point (day)*
m, sm, pm	0.85	60.49	DHR	3937
a1, a2, a3	1.01	13.73	IHR	954

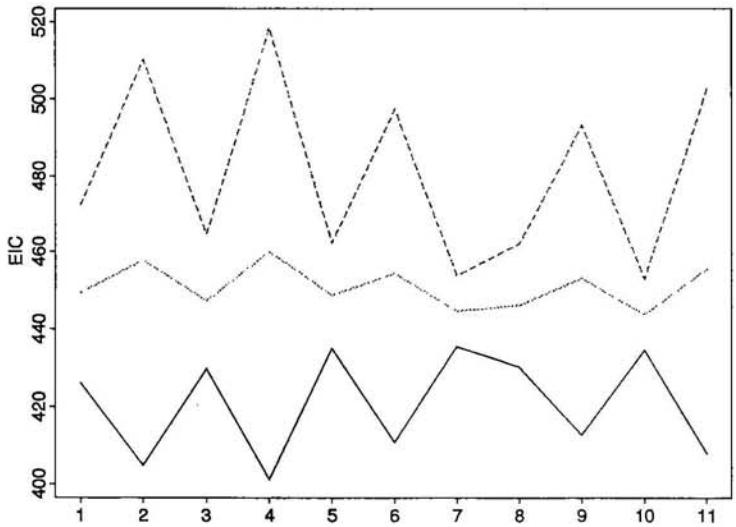

Figure 5.5: EIC_U, EIC, EIC_L of Weibull models for grading systems of T factor, where 1:$\{m, sm\}$, $\{pm\}$, $\{a1\}$, $\{a2, a3\}$, 2:$\{m, sm\}$, $\{pm\}$, $\{a1, a2\}$, $\{a3\}$, 3:$\{m, sm\}$, $\{pm\}$, $\{a1, a2, a3\}$, 4:$\{m, sm\}$, $\{pm\}$, $\{a1\}$, $\{a2\}$, $\{a3\}$, 5:$\{m, sm\}$, $\{pm, a1\}$, $\{a2, a3\}$, 6:$\{m, sm\}$, $\{pm, a1, a2\}$, $\{a3\}$, 7:$\{m, sm\}$, $\{pm, a1, a2, a3\}$, 8:$\{m, sm, pm\}$, $\{a1\}$, $\{a2, a3\}$, 9:$\{m, sm, pm\}$, $\{a1, a2\}$, $\{a3\}$, 10:$\{m, sm, pm\}$, $\{a1, a2, a3\}$, 11:$\{m, sm, pm\}$, $\{a1\}$, $\{a2\}$, $\{a3\}$.

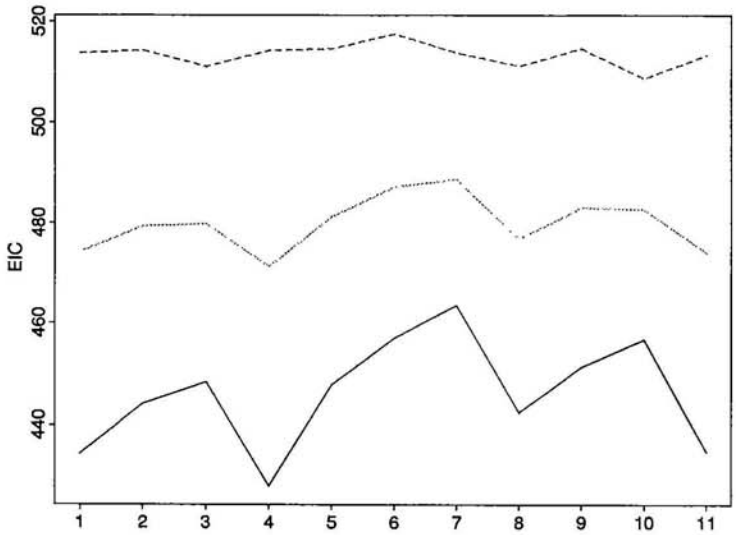

Figure 5.6: EIC_U, EIC, EIC_L of LIKM models for grading systems of T factor, where 1:$\{m, sm\}$, $\{pm\}$, $\{a1\}$, $\{a2, a3\}$, 2:$\{m, sm\}$, $\{pm\}$, $\{a1, a2\}$, $\{a3\}$, 3:$\{m, sm\}$, $\{pm\}$, $\{a1, a2, a3\}$, 4:$\{m, sm\}$, $\{pm\}$, $\{a1\}$, $\{a2\}$, $\{a3\}$, 5:$\{m, sm\}$, $\{pm, a1\}$, $\{a2, a3\}$, 6:$\{m, sm\}$, $\{pm, a1, a2\}$, $\{a3\}$, 7:$\{m, sm\}$, $\{pm, a1, a2, a3\}$, 8:$\{m, sm, pm\}$, $\{a1\}$, $\{a2, a3\}$, 9:$\{m, sm, pm\}$, $\{a1, a2\}$, $\{a3\}$, 10:$\{m, sm, pm\}$, $\{a1, a2, a3\}$, 11:$\{m, sm, pm\}$, $\{a1\}$, $\{a2\}$, $\{a3\}$.

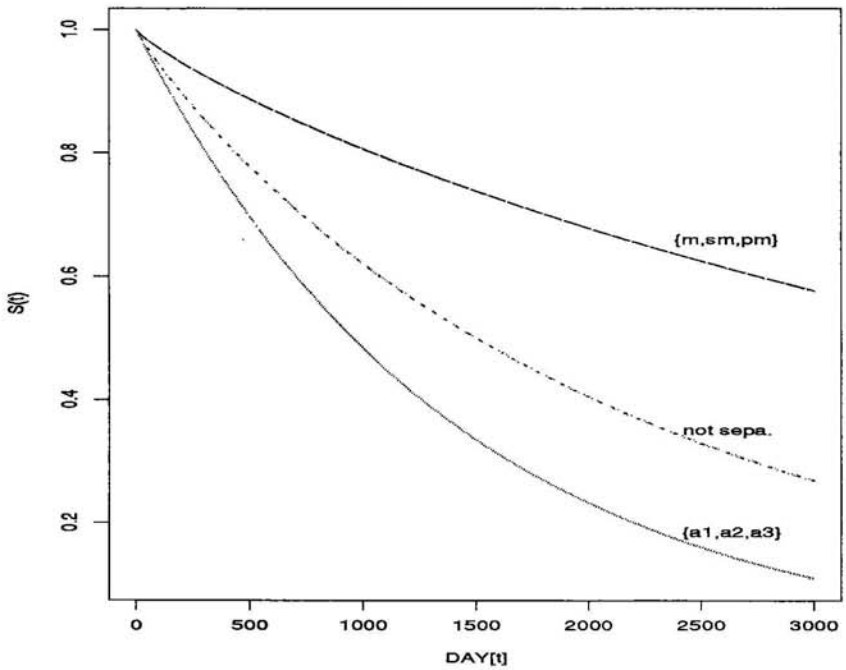

Figure 5.7: Weibull survival curves estimated from data divided by T factor.

$lp1$ that deep cervical lymph nodes.

$lp2$ that para-tracheal lymph nodes.

$lp3$ that middel and lower mediastinal lymph nodes.

$lp4$ that upper gastic lymph nodes.

$lp5$ that abdominal left para-aortic lymph node.

We assume that the lymph nodes are divided into two or three parts. For example a distribution pattern of lymphatic invasion divided into two parts is denoted by $\{lp1, lp2, lp3\}$, $\{lp4, lp5\}$, and (p_1, p_2) denotes the pattern of lymphatic invasion. When the first set $\{lp1, lp2, lp3\}$ is invaded, $p_1 = 1$, otherwise $p_1 = 0$. p_2 denotes the same notation for the second set $\{lp4, lp5\}$.

Figures 5.8 and 5.9 show the results by EIC for the grading systems of distribution patterns. The results by EIC of the LIKM models show that the grading system of the distribution pattern 5:$\{lp1, lp2\}$, $\{lp3, lp4\}$, $\{lp5\}$ is the best in Figure 5.9. However, the value of EIC of the Weibull model for the grading system of the distribution pattern 2:$\{lp1, lp2\}$, $\{lp3, lp4, lp5\}$ is less than that of the above LIKM model. The values of EIC of the distribution pattern system $\{lp1, lp2\}$, $\{lp3, lp4, lp5\}$ are summarized in Table 5.10. The estimated survival curves of the Weibull models for the the best system are plotted in Figure 5.10. The estimated parameters and 50 percentile points of the Weibull models are summarized in Table 5.11.

5.1.4 Comments on results

It was found by the EIC procedures that the goodness of the Weibull models are better than that of the LIKM models for the esophageal cancer data. According to

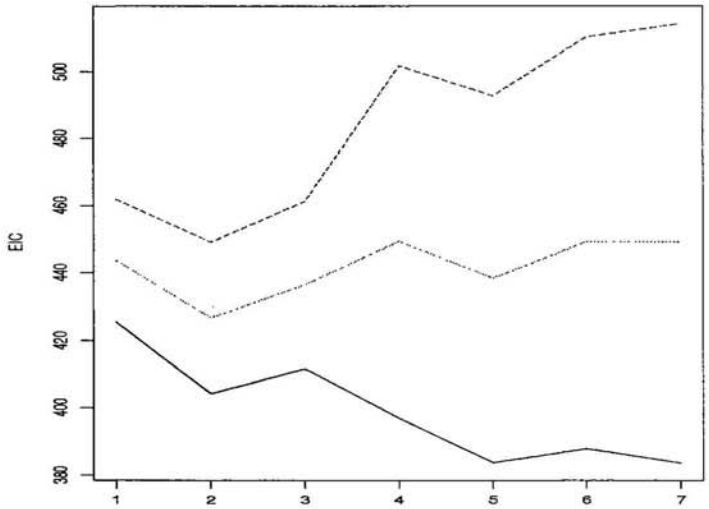

Figure 5.8: EIC_U, EIC, EIC_L of Weibull models for grading systems of distribution pattern of lymphatic invasion, where 1:$\{lp_1\}$, $\{lp_2, lp_3, lp_4, lp_5\}$, 2:$\{lp_1, lp_2\}$, $\{lp_3, lp_4, lp_5\}$, 3:$\{lp_1, lp_2, lp_3\}$, $\{lp_4, lp_5\}$, 4:$\{lp_1, lp_2, lp_3, lp_4\}$, $\{lp_5\}$, 5:$\{lp_1, lp_2\}$, $\{lp_3, lp_4\}$, $\{lp_5\}$, 6:$\{lp_1, lp_2, lp_3\}$, $\{lp_4\}$, $\{lp_5\}$, 7:$\{lp_1\}$, $\{lp_2, lp_3\}$, $\{lp_4, lp_5\}$.

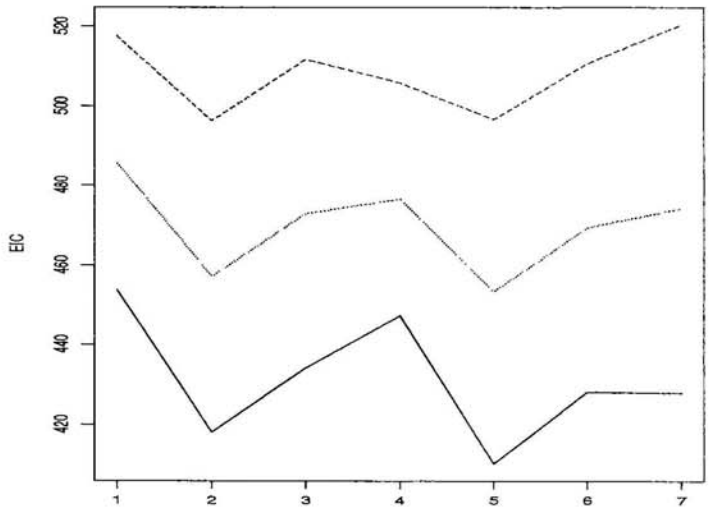

Figure 5.9: EIC_U, EIC, EIC_L of LIKM models for grading systems of distribution pattern of lymphatic invasion, where 1:$\{lp_1\}$, $\{lp_2, lp_3, lp_4, lp_5\}$, 2:$\{lp_1, lp_2\}$, $\{lp_3, lp_4, lp_5\}$, 3:$\{lp_1, lp_2, lp_3\}$, $\{lp_4, lp_5\}$, 4:$\{lp_1, lp_2, lp_3, lp_4\}$, $\{lp_5\}$, 5:$\{lp_1, lp_2\}$, $\{lp_3, lp_4\}$, $\{lp_5\}$, 6:$\{lp_1, lp_2, lp_3\}$, $\{lp_4\}$, $\{lp_5\}$, 7:$\{lp_1\}$, $\{lp_2, lp_3\}$, $\{lp_4, lp_5\}$.

Table 5.10: EIC of Weibull and LIKM models for data divided by distribution pattern of lymphatic invasion $\{lp_1, lp_2\}, \{lp_3, lp_4, lp_5\}$.

grade	n_1, n_2	*Weibull*			*LIKM*		
		EIC_L	EIC	EIC_U	*bias*	EIC_L EIC EIC_U	*bias*
full data	56,87	452.81	457.02	461.24	2.04	488.27 507.68 527.09	55.77
(1,0)	9,11	73.82	79.40	84.97	2.43	73.38 82.00 90.62	10.36
(1,1)	34,16	204.97	209.78	214.58	2.26	220.69 236.12 251.56	36.75
(0,1)	7,22	64.11	70.29	76.48	2.33	60.33 68.12 75.90	8.38
(0,0)	6,38	61.37	67.20	73.02	2.31	63.69 70.95 78.22	6.69
grading system		404.27	426.67	449.05		418.09 457.19 496.30	

Table 5.11: Estimates of Weibull models for data divided by distribution pattern of lymphatic invasion $\{lp_1, lp_2\}, \{lp_3, lp_4, lp_5\}$.

grade	β	α	hazard	*50 % point (day)*
(1,0)	1.11	21.95	IHR	1579
(1,1)	1.05	7.52	IHR	531
(0,1)	0.98	40.88	DHR	2813
(0,0)	1.39	45.18	IHR	3468

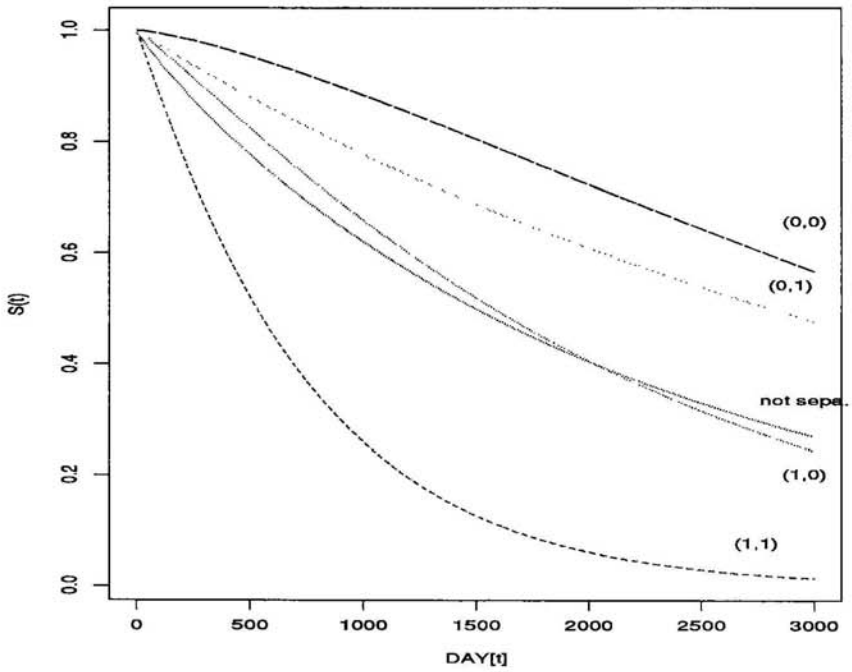

Figure 5.10: Weibull survival curves estimated from data divided by distribution pattern of lymphatic invasion $\{lp_1, lp_2\}, \{lp_3, lp_4, lp_5\}$.

Tables 5.4, 5.8 and 5.10, the grading system of the N factor $\{0, 1\}$, $\{2-7\}$, $\{8-\}$ is the best for evaluating the surgical outcomes in the analyzed systems. The system of the distribution pattern of lymphatic invasion $\{lp1, lp2\}$, $\{lp3, lp4, lp5\}$ is better than that of the T factor $\{m, sm, pm\}$, $\{a1, a2, a3\}$. These results correspond to the results by Matsubara et al. (1994).

5.2 Grading systems of multiple factors

In this section we analyze the survival data with the multiple factors affecting the surgical outcomes based on the EIC procedures. The results by EIC give the relations between the factors and the surgical outcomes. We accept the better system searched by the EIC procedures. The multiple factors are the degree of the number of invaded lymph nodes (N factor), the degree of the depth of cancer invasion (T factor) and the distribution pattern of lymphatic invasion. In this section the surgical outcomes are analyzed based on the results by the previous section.

5.2.1 Models of full factors

We suppose the covariate vector $\boldsymbol{z}= (z_1, z_2, z_3, z_4)$. The elements z_1 and z_2 take 1 or 0 according to be involved or not in the distribution patterns of lymphatic invasion $\{lp1, lp2\}$ and $\{lp3, lp4, lp5\}$, respectively. z_3 takes either 1, 1/2 or 1/3 according to the N factor $\{0, 1\}$, $\{2 \sim 7\}$ or $\{8 \sim\}$. z_4 takes either 1 or 1/2 according to the T factor $\{m, sm, pm\}$ or $\{a1, a2, a3\}$. Table 5.12 shows the values of EIC of the parametric accelerated models and LIC survival probability model for the cancer data. It is shown that the parametric accelerated models are better than the LIC survival probability model. Table 5.12 also shows that the values of EIC for the Weibull and gamma accelerated models are close. Table 5.13 shows the estimated parameters and the half percentile points of baselines ($\boldsymbol{z}= \mathbf{0}$). The baselines estimated from the models are plotted in Figure 5.11. Figure 5.12 shows the survival curves estimated from the models with $\boldsymbol{z}$=(0, 1, 0.5, 0.5). The estimated

survival curves of the Weibull accelerated model and the Cox survival probability model are plotted in Figure 5.13 and Figure 5.14, respectively.

We suppose the other grading system of the distribution pattern of lymphatic invasion with covariate vector $\boldsymbol{z}= (z_1, z_2, z_3, z_4, z_5)$. z_1, z_2 and z_3 are the elements of the grading system of $\{lp1, lp2\}$, $\{lp3, lp4\}$, $\{lp5\}$. z_4 and z_5 are the elements of the degree of the N factor and the T factor, respectively. In this case, the value of EIC for the Weibull accelerated model is 408.32. On the other hand, z_1 and z_2 are elements of the grading system of $\{lp1, lp2\}$, $\{lp3, lp4, lp5\}$. z_3 is the cross term given by $z_1 \times z_2$. z_4 and z_5 are the same as before. EIC of the Weibull accelerated model is 408.29. Furthermore, when z_6 is the cross term given by $z_4 \times z_5$, EIC of the Weibull accelerated model is 410.67. When $z_7 = z_4 \times z_3$, $z_8 = z_5 \times z_3$, $z_9 = z_4 \times z_5 \times z_6$, EIC of the Weibull accelerated model is 424.59. It is shown by the results that the grading system of Table 5.12 is better than those of above case.

5.2.2 Strata of N factor

The strata of the N factor with the other covariates can be considered. We assume the covariate vector $\boldsymbol{z}= (z_1, z_2, z_3)$. z_1 and z_2 are the elements of the distribution pattern of lymphatic invasion. z_3 is the degree of the T factor. Table 5.14 shows EIC of the Weibull accelerated model and LIC survival probability model. It is shown in the table that the models for the grading system of the N factor with the multiple factors are not more suitable than the models of the multiple factors for the full data. Table 5.15 shows the estimated parameters and the 50 percentile points of baselines ($\boldsymbol{z}= \mathbf{0}$) about the Weibull accelerated models. The models of multiple

Table 5.12: EIC of Weibull, gamma accelerated models and LIC survival probability model for full factors.

Models	*n1, n2*	EIC_L	EIC	EIC_U	*bias*
Weibull		398.29	405.61	412.94	5.02
Gamma	56,87	398.33	405.52	412.71	5.22
LIC		441.91	463.58	485.25	62.05

Table 5.13: Estimates of Weibull, gamma accelerated models and LIC survival probability model for full factors.

		estimates	hazard	*50 % point (day)* $\boldsymbol{z} = \mathbf{0}$
Weibull	$(\hat{\beta}, \hat{\boldsymbol{p}})=$	(1.16,-0.65,0.67,3.08,1.27)	IHR	73
Gamma	$(\hat{a}, \hat{\boldsymbol{p}})=$	(1.26,-0.70,0.57,2.87,1.17)	IHR	95
LIKM	$\hat{\boldsymbol{p}} =$	(0.88,-0.44,-2.68,-1.13)		200

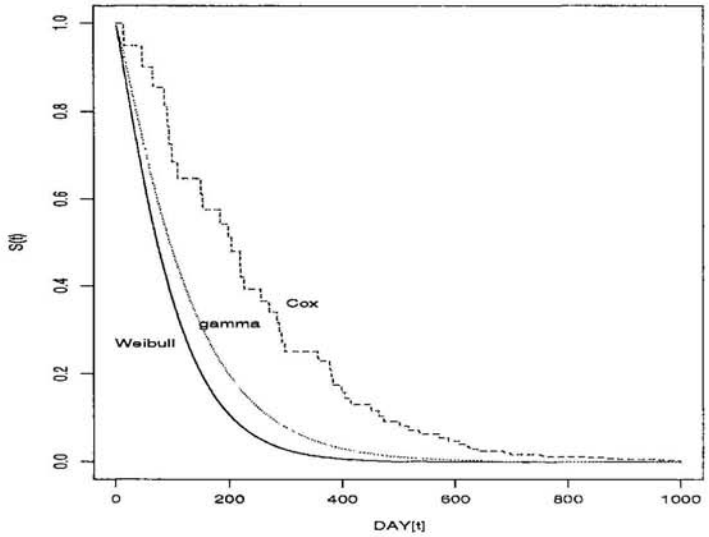

Figure 5.11: Baselines of survival probabilities estimated from Weibull, gamma accelerated models and Cox survival probability model.

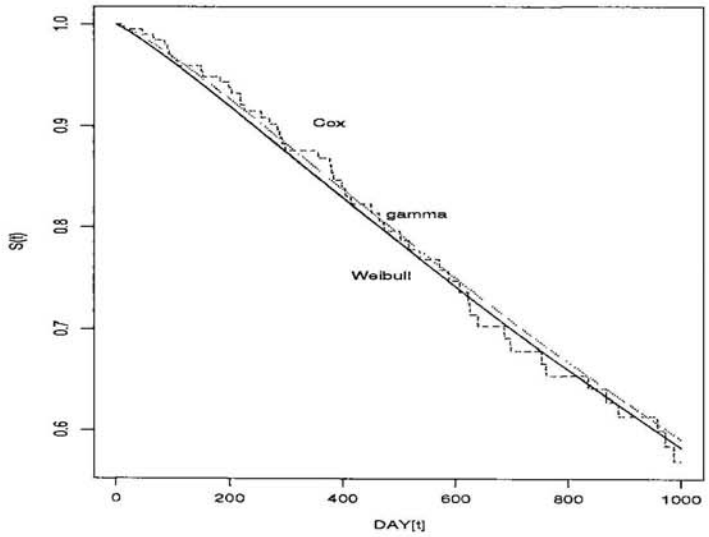

Figure 5.12: Survival curves estimated from Weibull, gamma accelerated models and Cox survival probability model, where $\boldsymbol{z}$=(0, 1, 0.5, 0.5).

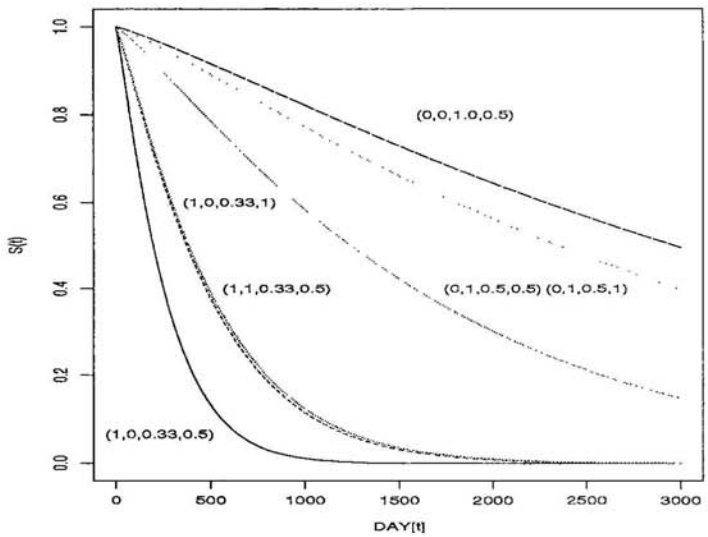

Figure 5.13: Survival curves estimated from Weibull accelerated models with covariates of all factors.

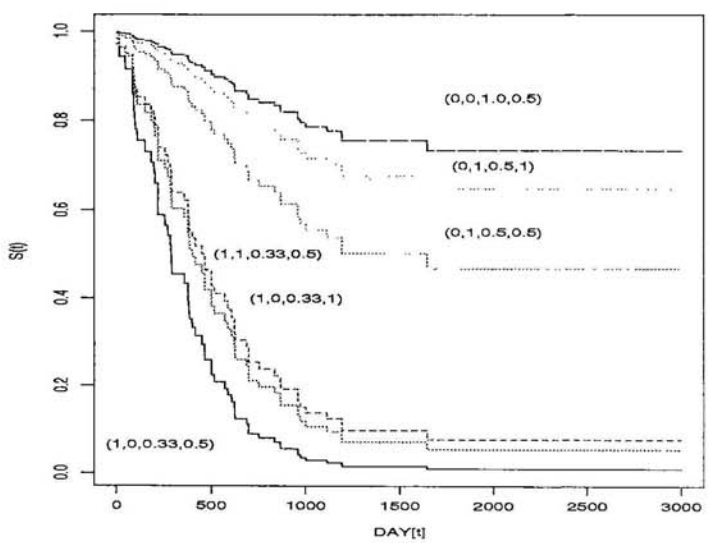

Figure 5.14: Survival curves estimated from Cox models with covariates of all factors.

factors for the full data in Table 5.14 are worse than the models in Table 5.12.

5.2.3 Strata of T factor

The strata of the T factor with the other covariates can be considered. We assume the covariate vector $\boldsymbol{z}= (z_1, z_2, z_3)$. z_1 and z_2 are the same as before. z_3 is the degree of the N factor. Table 5.16 shows EIC of the Weibull accelerated models and LIC survival probability models. It is shown in the table that the grading system of the T factor with the multiple factors is not suitable than the model of the multiple factors for the full data in the Weibull accelerated model. Table 5.17 shows the estimated parameters and the half percentile points of baselines ($\boldsymbol{z}= \mathbf{0}$). The estimated survival baselines of the Weibull accelerated models and the Cox models for the grading system are plotted in Figure 5.15. The models of the T factor with the multiple factors for the full data are worse than the models in Table 5.12.

5.2.4 Comments on results

It was found by the the EIC procedures that the goodness of the Weibull accelerated models are better than that of the LIC survival probability models for the esophageal cancer data. The model of the multiple factors, the N factor, the T factor and the distribution pattern of lymphatic invasion, for the full data is the best for evaluating the surgical outcomes in the analyzed systems. It is suggested by the above result that the multiple factors affect the scale of survival probability distribution.

The value of EIC for this model is less than that of the grading system of the N factor $\{0, 1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$ in the previous section. Matsubara et al. (1994) found

Table 5.14: EIC of Weibull accelerated models and LIC survival probability models for data divided by N factor.

grade	*n1,n2*	*Weibull*				*LIC*			
		EIC_L	EIC	EIC_U	*bias*	EIC_L	EIC	EIC_U	*bias*
full data	56,87	434.86	443.35	451.83	5.04	450.86	471.06	491.27	58.18
0,1	13,56	138.04	150.36	162.67	4.33	138.45	155.48	172.51	20.16
2∼7	24,24	166.60	182.47	198.34	5.76	170.82	194.41	218.01	31.22
8∼	19,7	84.93	134.46	183.98	18.40	152.40	251.09	349.77	20.93
grading system		389.57	467.29	544.99		461.67	600.98	740.29	

Table 5.15: Estimates of Weibull accelerated models for data divided by N factor.

grade	*estimates*	hazard	*50 % point (day)* $\boldsymbol{z} = \mathbf{0}$
full data	$(\hat{\beta}, \hat{\boldsymbol{p}})$=(0.86, -0.90, 0.07, 5.65)	DHR	65
0,1	$(\hat{\beta}, \hat{\boldsymbol{p}})$= (0.73, -0.14, 2.28, 5.83)	DHR	61
2∼7	$(\hat{\beta}, \hat{\boldsymbol{p}})$= (0.91, -0.63, 1.09, 3.31)	DHR	67
8∼	$(\hat{\beta}, \hat{\boldsymbol{p}})$= (1.45, -0.20, 0.98, 1.88)	IHR	78

Table 5.16: EIC of Weibull accelerated models and LIC survival probability models for data divided by T factor.

grade	*n1,n2*	*Weibull*				*LIC*			
		EIC_L	EIC	EIC_U	*bias*	EIC_L	EIC	EIC_U	*bias*
full data	56,87	402.87	408.94	415.02	3.99	445.83	467.60	489.37	60.63
m, sm, pm	12,46	112.84	121.77	130.70	4.35	103.63	132.54	161.46	19.15
a1, a2, a3	44,41	284.47	291.44	298.41	4.44	301.10	324.85	348.60	52.90
grading system		397.31	413.21	429.11		404.73	457.39	510.06	

Table 5.17: Estimates of Weibull accelerated models for data divided by T factor.

grade	*estimates*	hazard	*50 % point (day)* $\boldsymbol{z} = \mathbf{0}$
full data	$(\hat{\beta}, \hat{\boldsymbol{p}})$=(1.12, -0.57, 0.88, 4.12)	IHR	72
m, sm, pm	$(\hat{\beta}, \hat{\boldsymbol{p}})$= (0.81, -0.95, 1.31, 5.02)	DHR	64
a1, a2, a3∼7	$(\hat{\beta}, \hat{\boldsymbol{p}})$= (1.25, -0.38, 0.84, 3.64)	IHR	75

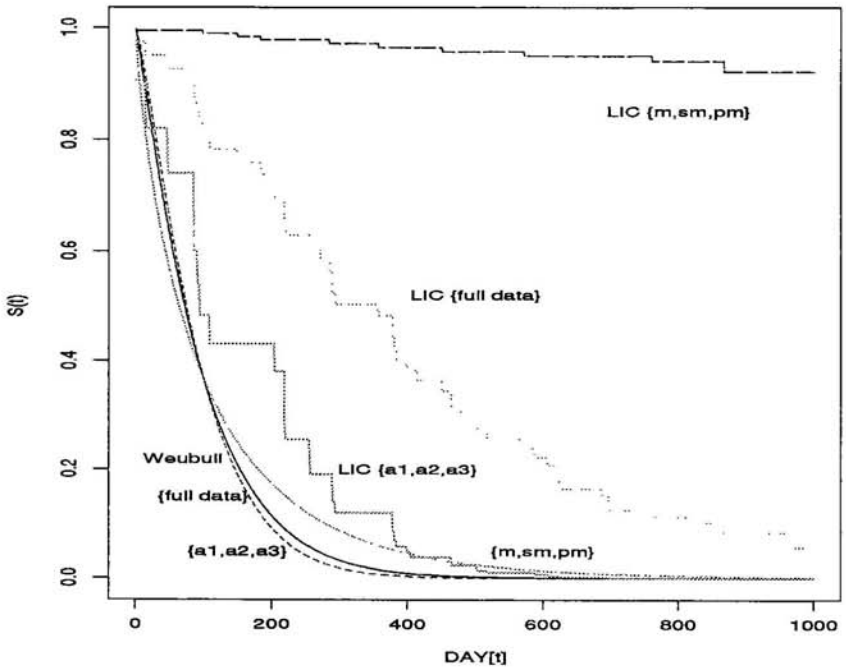

Figure 5.15: Baselines of Weibull and LIC survival probability models estimated from data divided by T factor.

by the AIC procedure that the grading system which was designed by combining the N factor with the distribution pattern of lymphatic invasion was better than the simple system categorized according to the N factor. The result of this section corresponds to their result.

Chapter 6

Conclusions

In this thesis, we proposed the applications of the information criterion EIC to the survival analysis. Then we tried to use EIC for not only the parametric survival probability models but also the nonparametric models. In particular, the survival analysis for medical field was handled.

Firstly, the linearly interpolated KM (LIKM) model, the Weibull and gamma models were handled. The model-based bootstrap method was used for applying EIC to these simple survival probability models. It was found that the EIC procedure made it possible to compare the goodness of the LIKM model with those of the Weibull and gamma models, and the EIC procedure was more suitable for the choice of the model fitted to small sample. The EIC procedure was used to compare the goodness of the survival probability models fitted to strata of categories.

Secondly, we proposed the applications of EIC to compare the linearly interpalated Cox (LIC) survival probability model with parametric accelerated models and to search for the multiple factors affecting the survival probabilities. The model-based bootstrap method was used to evaluate these models by EIC. This

EIC procedures were the extensions of the methods proposed for the simple survival probability models.

Finally, we demonstrated the analysis of the surgical outcomes in cancer of the thoracic esophagus. It was found by the EIC procedures that the goodnesses of the Weibull models were better than that of the LIKM models to estimate the survival probabilities for the esophageal cancer data, and the the system of the N factor $\{0,1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$ was the best for evaluating the surgical outcomes in the systems of simple factors. The system of the distribution pattern of lymphatic invasions $\{lp1, lp2\}$, $\{lp3, lp4, lp5\}$ was better than that of the T factor $\{m, sm, pm\}$, $\{a1, a2, a3\}$. Furthermore we searched for the relations between the multiple factors and the surgical outcomes in the esophageal cancer. The EIC procedures made it possible to compare the LIC survival probability models with the Weibull and gamma accelerated models. The results by EIC showed the parametric accelerated models were better than the LIC survival probability models for the cancer data. The results by the EIC procedures that the grading system with the covariates of all factors, the N factor, the T factor and the distribution pattern of lymphatic invasions, was the best. This model was better than the model for the system of the N factor $\{0,1\}$, $\{2 \sim 7\}$, $\{8 \sim\}$.

Through the modeling of real survival data based on EIC, we showed that our EIC procedures were effective for the survival analysis.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, 2nd Inter. Symp. on Information Theory (Petrov, B. N. and Csaki, F. eds.), Akademiya Kiado, Budapest, 267-281.

Akritas, M. G. (1986). Bootstrapping the Kaplan-Meier estimator, Journal of American Statistical Association 81, 1032-1038.

Andersen, P. K., Borgan, Ø., Gill, R. D., Keidig, N. (1994). Statistical models based on counting processes, Springer-Verlag.

Breslow, N., Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship, The Annals of Statistics 2, 437-453.

Breslow, N. (1974). Covariance analysis of censored survival data, Biometrics 30, 89-99.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model, Journal of the American Statistical Association 86, 1290-1302.

Cantor, A. B. (1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data, Statistics in Medicine 11, 931-937.

Collet, D. (1994). Modelling survival data in medical research, Chapman & Hall.

Cox, D. R. (1972). Regression models and life tables (with discussion). J. R. Stat. Soc. B 34, 187-220.

Cox, D. R. (1975). Partial likelihood, Biometrika 62, 269-276.

Cox, D. R., Oakes, D. (1984). Analysis of survival data, Chapman & Hall.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, Annals of Statistics 17, 1-26.

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

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

Fleming, T. R., Harrington, P. D. (1991). Counting processes and survival analysis, Wiley, New York.

Gehan, E. D. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples, Biometrika 52, 203-223.

Greenhouse, J. B., Wolfe, R. A. (1994), A competing risks derivation of a mixture model for the analysis of survival data, Communications in Statistics: Theory and Method, 13, 3133-3154.

Ishiguro, M., Akaike, H. (1989). DALL: Davidon's algorithm for log likelihood maximization, Computer Science Monographs No. 25, The Institute of Statistical Mathematics, Tokyo.

Ishiguro, M., Sakamoto, Y. (1991). WIC: an estimation-free information criterion,

Research Memorandum No. 410, The Institute of Statistical Mathematics, Tokyo.

Ishiguro, M., Sakamoto, Y., Kitagawa, G. (1994). Bootstrapping log likelihood and EIC, an extension AIC, Research Memorandum No. 532, The Institute of Statistical Mathematics, Tokyo.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator, Scandinavian Journal of Statistics 5, 195-199.

Kaplan, E. L., Meier, P. (1958). Nonparametric estimation from incomplete observations, Journal of American Statistical Association 53, 457-481.

Kalbfleisch, J., Prentice, R. L. (1973). Marginal likelihoods based on Cox's regression and life model, Biometrika 60, 267-278.

Kalbfleisch, J., Prentice, R. L. (1980). The statistical analysis of failure time data, Wiley, New York.

Kitagawa, G., Ishiguro, M., Sakamoto, Y. (1995). Model selection by EIC, a bootstrap version of AIC, Research Memorandum No. 540, The Institute of Statistical Mathematics, Tokyo.

Konishi, S., Kitagawa, G. (1995). Generalized information criterion and the bootstrap, Research Memorandum No. 549, The Institute of Statistical Mathematics, Tokyo.

Lawless, J. F., (1982). Statistical models and methods for lifetime data, Wiley, New York.

Lee, E. T., (1992). Statistical methods for survival data analysis, Wiley, New York.

Matsubara, T. (1992). Pattern of lymphatic spreading in cancer of the thoracic esophagus, Journal of Japan Surgical Society 93, 377-387 (in Japanese).

Matsubara, T., Kaise, T., Ishiguro, M., Nakajima, T. (1994). Better grading system for evaluating the degree of lymph node invasion in cancer of the thoracic esophagus, Surgery Today 24, 500-505.

Miller, R. G. (1981). Survival analysis, Wiley, New York.

Miller, R. G. (1983). What price Kaplan-Meier ?, Biometrics 39, 1077-1081.

Nelson, W. (1982). Applied life data analysis, Wiley, New York.

Peto, R., Peto, J. (1972). Asymptotically efficient rank invariant test procedures, Journal of the Royal Statistical Society, Series A 135, 185-207.

Reid, N. (1981). Estimating the median survival time, Biometrika 68, 601-608.

Robin, H. (1995). Problems and prediction in survival-data analysis, Statistical in Medicine 14, 161-184.

Sakamoto, Y., Ishiguro, M., Kitagawa, G. (1986). Akaike information criterion statistics, D. Reidel, Dordrecht, Holland.

Yafune, A., Matsubara, T., Ishiguro, M. (1993). Bayesian analysis of lymphatic spreading patterns in cancer of the thoracic esophagus, Annals of the Institute of Statistical Mathematics 45, 401-418.