BOOTSTRAP CONFIDENCE INTERVALS FOR THE MODE OF LOG-LOGISTIC HAZARD FUNCTION

SITI NORMAH BINTI HASAN

MASTER OF APPLIED STATISTICS
UNIVERSITI PUTRA MALAYSIA
2009

ABSTRACT

By considering hazard function for log-logistic distribution with parameter $\beta > 1$, it is important to perform inferences about the mode of the hazard function with unimodal hazard function. The parameters of this distribution are estimated by maximum likelihood method and they are used to estimate other quantities of interest such as mode of lifetime data and percentile. From the asymptotical normality of the maximum likelihood estimator, confidence intervals can be obtained. However, these results might not be very accurate for the small sample size or large proportion of censored data. In this project, the confidence interval for the mode of the hazard function obtained by asymptotic confidence interval is going to be compared with bootstrap methods. The performance of the procedures is evaluated by simulation with different sample sizes and proportion of censored data.

ABSTRAK

Dengan mempertimbangkan taburan log-logistik sebagai taburan dasar bagi masa kegagalan dengan parameter $\beta > 1$, adalah penting untuk mendapatkan pentakbiran tentang mod bagi masa kegagalan dengan satu puncak. Parameter bagi taburan ini dianggarkan dengan menggunakan kaedah kebolehjadian maksimum dan ini digunakan untuk menganggar penganggar lain seperti mod masa kegagalan dan persentil. Selang keyakinan bagi penganggar kebolehjadian maksimum boleh didapati dari kaedah kenormalan klasik. Tetapi keputusan ini mungkin tidak berapa tepat apabila saiz sampel adalah kecil atau peratusan data tertapis adalah besar. Dalam projek ini, selang keyakinan bagi mod masa kegagalan adalah berdasarkan kaedah klasik akan diperolehi dan dibandingkan dengan kaedah bootstrap. Pencapaian kaedah ini dinilai dengan menjalankan kaedah simulasi dengan saiz sampel berlainan dan peratusan data tertapis yang berbeza.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

In many application of survival data analysis, it is important to perform inferences about the mode of the hazard function in situations of lifetime data modeling with unimodal hazard function. We usually have a hazard function that increases up to a maximum and then decreases after the change point. There were many application of this situation such as medical studies in heart or kidney transplantation where the patients have an increasing hazard during an adaptation period and a decreasing hazard after this adaptation period. There are many probability distributions that have unimodal hazard function such as log-logistic distribution, log-normal distribution and exponentiated-Weibull distribution. Most of the time, we are interested on the mode of the hazard function where the lifetime change-point occurs.

1.2 Problem Statement

In order to deal with the lifetime change-point where the mode of the hazard function occurs, we consider the log-logistic distribution with the shape parameter $\beta > 1$. We will compare the accuracy of asymptotical confidence intervals with two confidence intervals based on bootstrap simulation. We do not discuss the case where the shape parameter $\beta < 1$ since the hazard function is monotone decreasing with maximum at 0.

1.3 Objectives

The objectives of this project are:

1) to perform inferences in the mode of hazard function.
2) to compare the accuracy of asymptotical confidence intervals with two confidence intervals based on bootstrap simulation.
3) to identify the advantages and disadvantages of using asymptotical confidence intervals and bootstrap confidence intervals.

1.4 Significance of the Study

The findings of this project hopefully will help us to improve our knowledge on log-logistic distribution, hazard function and also bootstrap confidence intervals. Nevertheless, this project is also as a guidance for those who are interested to know about simulation using bootstrap confidence intervals.

1.5 Scope and Limitation

Our scope is based on hazard function of log-logistic distribution. We will do simulation by using bootstrap method which will focus only on the type t-bootstrap and percentile-bootstrap. We will only consider the log-logistic distribution for the lifetime data with the shape parameter $\beta > 1$ and compare the accuracy of the asymptotical confidence interval with two confidence intervals based on bootstrap simulation.

1.6 Organization of the Study

The organization of this study is as follows. We will discuss some literature reviews involving this study in Chapter 2 and detailed discussion on the research methodology in Chapter 3. In Chapter 4, findings and result from our analysis are presented. Lastly, conclusions and suggested further research are included in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Lifetime Distributions - Continuous Distribution

We consider the case of a single continuous lifetime variable, T. Specifically, let T be a nonnegative random variable representing the lifetimes of individuals in some population. We defined all functions over the interval $[0,\infty)$.

Let $f(t)$ denote the probability density function (p.d.f.) of T and let the cumulative distribution function (c.d.f.) be $F(t)=\Pr(T\le t)=\int_0^t f(x)dx$. The probability of an individual surviving to time t is given by the survivor function $S(t)=\Pr(T\ge t)=\int_t^\infty f(x)dx$.

In some contexts involving systems or lifetimes of manufactured items, $S(t)$ is referred to as the reliability function (Jerald 2003). Note that $S(t)$ is a monotone decreasing continuous function with $S(0)=1$ and $S(\infty)=\lim_{t\to\infty}S(t)=0$. We may wish to allow $S(\infty) > 0$ to consider settings where some individuals never fail which will be treated as special cases.

A very important concept with lifetime distributions is the hazard function $h(t)$, defined as,

$$h(t) = \mathrm{hm}_{\Delta t \to 0} \frac{\Pr(t < T < t + \Delta t / T > t)}{\Delta t}$$
$$= \frac{f(t)}{S(t)}$$

What is hazard function? We usually have a hazard function that increases up to a maximum and then decreases after this change-point. This is common in medical studies as in heart or kidney transplantation, where the patients have an increasing hazard during an adaptation period and a decreasing hazard after this adaptation period (Collett 1994).

The hazard function specifies the instantaneous rate of death or failure at time t, given that the individual survives up to t. The hazard function is also known as hazard rate and force of mortality.

The function $f(t), F(t), S(t)$ and $h(t)$ give mathematically equivalent specifications of the distribution of T. It is easy to derive expressions for $S(t)$ and $f(t)$ in terms of $h(t)$, since $f(t) = -S'(t)$.

The hazard function implies $h(x) = -\frac{d}{dx}\log S(x)$. Thus, $\log S(x)|_0^t = -\int_0^t h(x)dx$ and since $S(0) = 1$, we find that $S(t) = \exp(-\int_0^t h(x)dx)$. It is useful to define the cumulative hazard function $H(t) = \int_0^t h(x)dx$ which is related to the survival function by $S(t) = \exp[-H(t)]$. If $S(\infty) = 0$, then $H(\infty) = \infty$. Finally, it follows that $f(t) = h(t)\exp(-\int_0^t h(x)dx)$.

2.2 Log-Logistic Distribution

Various parametric families of models are used in the analysis of lifetime data and the modeling of aging or failure processes. Among the univariate models, few distributions occupy a central position because of their demonstrated usefulness in a wide range of situations.

We are interested in the hazard function that increases up to a maximum and then decreases after the change-point. Some common lifetime models like the exponential or Weibull distributions are not appropriate to fit unimodal hazard function of data.

Many existing probability distributions used to analyze lifetime data have unimodal hazard functions such as log-logistic distribution, log-normal distribution, exponentiated-Weibull distribution, inverse-Weibull distribution and many others.

We have interest in the estimation of the lifetime change-point where the mode of the hazard function occurs. Sometimes there is information about the aging of failure process in a population that suggests a particular distribution. This information is rarely specific enough to narrow consideration to just one family of models.

We let T be a random variable representing the lifetime of a unit or patient with a log-logistic distribution with hazard function given by

$$h(t) = \frac{e^{\mu}\beta t^{\beta-1}}{1+e^{\mu}t^{\beta}}$$

Where $t > 0, \beta > 0$ and $-\infty < \mu < \infty$. β is the shape parameter while μ is the scale parameter. The logistic distribution has probability density function of the form

$$f(t) = \frac{e^{\mu}\beta t^{\beta-1}}{[1+e^{\mu}t^{\beta}]^2}, \qquad t > 0$$

and the survival function is given by

$$S(t) = \frac{1}{1+e^{\mu}t^{\beta}}$$

We observe that the hazard function $h(t)$ decreases for $\beta < 1$ with maximum at $t = 0$. For $\beta > 1$, the hazard function is unimodal where $h(t)$ increases for $T \leq \theta$ and decreses for $T > \theta$. The parameter $\theta = [(\beta - 1)e^{-\mu}]^{1/\beta}$ is the hazard change-point.

2.3 Some Remarks on the Hazard Function

According to F.L. Jerald, (2003), the hazard function is particularly important characteristic of a lifetime distribution. It indicates the way the risk of failure varies with age or time, and this is of interest in most applications. Prior information about the shape of the hazard function can help guide model selection. Finally, if factors affected an individual's lifetime vary over time, it is often essential to approach modeling through the hazard function.

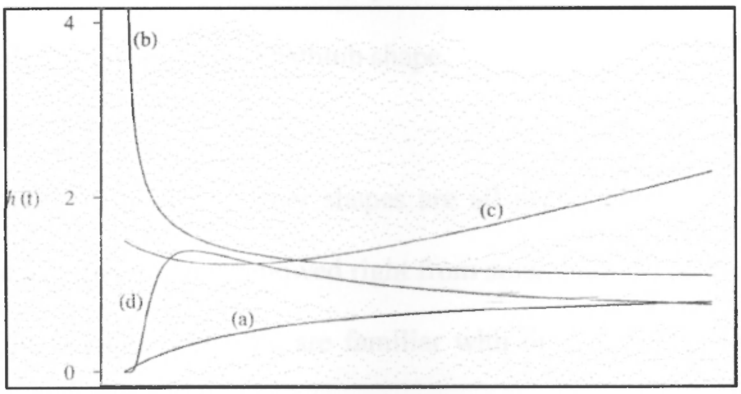

Figure 2.1: Some hazard functions, h(t)

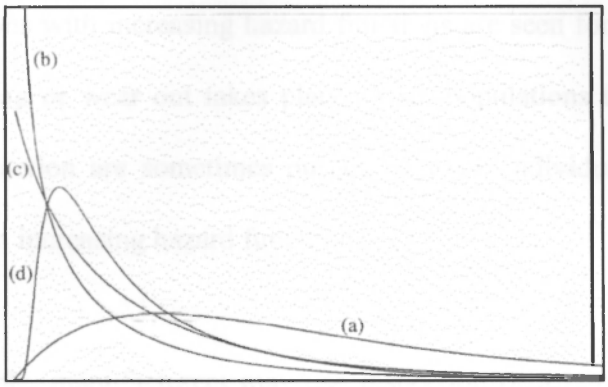

Figure 2.2: Some probability density functions, f(t).

Figure 2.1 shows hazard functions $h(t)$ and figure 2.2 shows probability distribution functions, $f(t)$ for four continuous distributions. The shapes of the hazard functions in figure 2.1 are qualitatively different;

distribution (a) has a monotone increasing hazard function,

distribution (b) has a monotone decreasing hazard function,

distribution (c) has a so-called bathtub-shaped, U-shape hazard function, and

distribution (d) display an inverse bathtub shape.

Models with these and other shapes are all useful in practice. If, for example, individuals in a population are followed right from actual birth to death, a bathtub-shaped hazard function is often seen. We are familiar with this pattern in human populations, after an initial period in which a death result primarily from birth defects of infant diseases, the death rate drops and is relatively constant until the age of 30 or so, after which it increases with age.

Distributions with increasing hazard functions are seen for individuals for whom some kind of aging or wear out takes place. Also, populations that display a bathtub-shaped hazard function are sometimes purged of weak individuals, leaving a reduced population with an increasing hazard function.

For example, manufacturers may use inspection or a burn-in process, in which items are subjected to a brief period of operation before being sent to customer. In this way defective or poor-quality items that would fail early are removed from the population; this frequently leaves a residual population that exhibits an increasing hazard function.

Certain types of electronic devices display a decreasing hazard as items with defects fail and are removed from the population. Roughly constant hazard functions tend to occur in stable settings where failure or death is due to random phenomena such as shocks or accidents, which are external to the individual. Shape (d) in Figure 1.2, where $h(t)$ first increases to a maximum and then decreases, is encountered in many applications, for example, in the case of survival after treatment for cancer where some individuals are cured, and in connection with the duration of marriage.

In many settings factors or covariates affecting an individual's lifetime vary over time. We refer them as "time-varying" or "time-dependent" covariates. For example, in life tests of electrical insulation the voltage level that items are subjected to is sometimes changed over time according to a fixed schedule.

In studies of the age at which smokers develop chronic diseases, the type and level of smoking for each individual can vary over time. The duration of a marriage may be affected by the presence of children or the couple's employment status, both of which can change over time when there are time-varying covariates, it is usually essential to think about the models in terms of their hazard function.

2.4 Bootstrap Interval Estimates

2.4.1 Introduction

Bootstrap is a resampling method for statistical inference. It is commonly used to estimate confidence intervals, but it can also be used to estimate bias and variance of an estimator or calibrate hypothesis tests. There are some illustration of diversity of recent environmetric applications of the bootstrap that can be found in toxicology, fisheries surveys, groundwater and air pollution modeling and others.

Bootstrap is a technique that constitutes a process of statistical inference to assess statistical accuracy. The estimate of bias is the difference between the expectation of an estimator $\hat{\theta}$ and the quantity θ being estimated.

Bootstrap is a computer based method for estimating the standard error of $\hat{\theta}$. Bootstrap estimate of standard error requires no theoretical calculations and it is available no matter how complicated the estimator $\hat{\theta}$ might be.

Below is the bootstrap algorithm for estimating standard error.

1. Select B independent bootstrap samples $y^{*1}, y^{*2}, \ldots, y^{*B}$ each consisting of n data values drawn with replacement from $(y_1, y_2, \ldots, y_n)$. For estimating standard error, the number B will ordinarily be in the range 25-200.

2. Evaluate the bootstrap replication corresponding to each bootstrap sample,

$$\hat{\theta}^*(b) = s(y^{*b}) \qquad b = 1, 2, \ldots, B$$

3. Estimate the standard error $se_F(\hat{\theta})$ by the sample standard deviation of the B replications.

$$se_B = \left\{ \frac{\sum_{b=1}^{B} [\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2}{(B-1)} \right\}^{1/2} \qquad \text{where } \hat{\theta}^*(.) = \sum_{b=1}^{B} \hat{\theta}^*(b) \Big/ B$$

The limit of se_B as B goes to infinity is the ideal bootstrap estimate of $se_F(\hat{\theta})$,

$$\lim_{B\to\infty} se_B = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*)$$

The fact that se_B approaches $se_{\hat{F}}$ as B goes to infinity amounts to say that an empirical standard deviation approaches the population standard deviation as the number of replications grows large.

Now we will look at the bias, which is the difference between the expectation of an estimator $\hat{\theta}$ and the quantity θ being estimated.

The bias of $\hat{\theta} = s(Y)$ as an estimate of θ is defined to be the difference between the expectation of $\hat{\theta}$ and the value of the parameter θ, denoted as

$$bias_F = bias_F(\hat{\theta},\theta) = E_F[s(Y) - t(F)]$$

A large bias is usually an undesirable aspect of an estimator's performance. We are assigned to the fact that $\hat{\theta}$ is a variable estimator of θ, but usually we don't want the variability to be overwhelmingly on the low side or on the high side.

We will use plug-in estimates to find the bias. Although the plug-in estimates $\hat{\theta} = t(\hat{F})$ are not necessarily unbiased, but they tend to have small biases compared to the magnitude of their standard errors. This is the one of the good features of the plug-in principle. Therefore, we can use the bootstrap to assess the bias of any estimator

$\hat{\theta} = s(Y)$. The bootstrap estimate of bias is defined to be the estimate $bias_{\hat{F}}$ that we obtained by substituting $\hat{F}$ for F which is,

$$bias_{\hat{F}} = E_{\hat{F}}[s(Y^*) - t(\hat{F})]$$

Here $t(\hat{F})$, the plug-in estimate of θ may differ from $\hat{\theta} = s(Y)$. In other words, $bias_{\hat{F}}$ is the plug-in estimate of $bias_{F}$. Whether or not $\hat{\theta}$ is the plug-in estimate of θ. Typically a statistics has some bias. However, the $bias_{\hat{F}}$ provides an estimate of this bias.

For most statistics that arise in practice, the ideal bootstrap estimate $bias_{\hat{F}}$ must be approximated by the Monte Carlo simulation. In this study, we will generate independent bootstrap samples $Y^{*1}, Y^{*2}, ..., Y^{*B}$ and evaluate the bootstrap replications, $\hat{\theta}^*(b) = s(y^{*b})$. Then, an approximate of the bootstrap expectation $E_{\hat{F}}[s(X^*)]$ is computed by the average of

$$\hat{\theta}^*(\bullet) = \sum_{b=1}^{B} \frac{\hat{\theta}^*(b)}{B} = \sum_{b=1}^{B} \frac{s(Y^{*b})}{B}$$

The bootstrap estimate of bias that we will be applied based on the number of B replication is,

$$bias_{B} = \hat{\theta}^*(\bullet) - t(\hat{F})$$

2.4.2 Bootstrap-t confidence interval

Bootstrap-t method is motivated by the student's t confidence interval but differs slightly in that it estimates the distribution of the standardized value of the parameter estimate, $\bar{\theta}$ directly form the data instead of approximating using a specific distribution. The table consisting of these standardized values is then used to construct a confidence interval for the parameter, θ in the usual way. It should provide better confidence interval estimates than the interval based on student's t distribution because it takes into account some of the errors such as bias and skewness in the bootstrap samples.

To obtain the bootstrap table, B bootstrap samples will be generated by any bootstrap sampling method. Following that compute $\bar{\theta}^b$ for each bootstrap sample, y^b, where $b = (1,2,...,B)$ and obtain,

$$Z(\bar{\theta}^b) = \frac{\bar{\theta}^b - \bar{\theta}}{se(\bar{\theta}^b)}$$

Here, $se(\bar{\theta}^b)$ is the estimated standard error of $\bar{\theta}^b$ for the bootstrap sample, y^b and $\bar{\theta}$ is the estimate obtained from the original dataset. In the standard normal confidence interval, $Z_{\frac{\alpha}{2}}$ is the $\frac{\alpha}{2}$ quantile of the standard normal distribution. Similarly in bootstrap-t confidence interval the values of $Z_{\frac{\alpha}{2}}$ is based on bootstrap-t percentiles. The easiest way to obtain this is by the following. Obtain $[Z(\bar{\theta}^b)]_K$, where $K = 1,2,...,B$ is the ordered list of the B replication of $Z(\hat{\theta}^b)$. The $\frac{\alpha}{2}$ percentile of $Z(\hat{\theta}^b)$ is the value, $[Z(\hat{\theta}^b)]_{B\frac{\alpha}{2}}$.

Then, the $100(1-\alpha)\%$ confidence interval for θ is,

$$\{\hat{\theta}-[Z(\hat{\theta}^b)]_l.se(\hat{\theta}),\hat{\theta}-[Z(\hat{\theta}^b)]_u.se(\hat{\theta})\}$$

Where $l = B*(1-\frac{\alpha}{2})$ and $u = B*(\frac{\alpha}{2})$. In the case where $B*(\frac{\alpha}{2})$ is not integer, l and u could be rounded to the next highest and lowest integer.

2.4.3 Percentile bootstrap confidence interval

Percentile bootstrap method is a simpler alternative to the bootstrap-t interval and does not require the standard error of an estimate. The method uses appropriate percentiles of the bootstrap distribution to define the confidence level. This procedure requires B independent bootstrap samples, denoted, $y^1, y^2, ..., y^B$. Then, compute $\hat{\theta}^b$ for each of the bootstrap sample, y^B. The percentile $100(1-\alpha)\%$ interval for θ is,

$$\{[(\hat{\theta}^b)]_l,[(\hat{\theta}^b)]_u\},$$

where $l = B*(\frac{\alpha}{2})$ and $u = B*(1-\frac{\alpha}{2})$ and $[\hat{\theta}^b]_k$, where $K = 1,2,...,B$ is the ordered list of the B replication of $\hat{\theta}^b$. Again in the case where $B*(\frac{\alpha}{2})$ is not integer, l and u could be rounded to the next highest and lowest integer. When the bootstrap distribution is roughly normal, the Wald intervals and bootstrap percentile interval should be nearly in agreeable. The advantage of the bootstrap percentile interval is that it automatically incorporates the transformation of the parameter.

CHAPTER 3

METHODS AND PROCEDURES

3.1 Mode of Log-Logistic Hazard Function

In this study, we will consider about the mode of hazard function for log-logistic distribution. As what we know, for log-logistic distribution, hazard function is given by,

$$h(t) = \frac{e^{\mu}\beta t^{\beta-1}}{1+e^{\mu}t^{\beta}}$$

where $t > 0, \beta > 0$ and $-\infty < \mu < \infty$. β is the shape parameter and μ is the scale parameter.

We also have density function and survival function for log-logistic distribution which are given by $f(t) = \frac{e^{\mu}\beta t^{\beta-1}}{[1+e^{\mu}t^{\beta}]^2}$ and $S(t) = \frac{1}{1+e^{\mu}t^{\beta}}$ respectively.

We have interest in the estimation of the lifetime change-point where the mode of the hazard function occurs. Mode of hazard function is define as hazard function that increases up to a maximum and then decreases after this change-point.

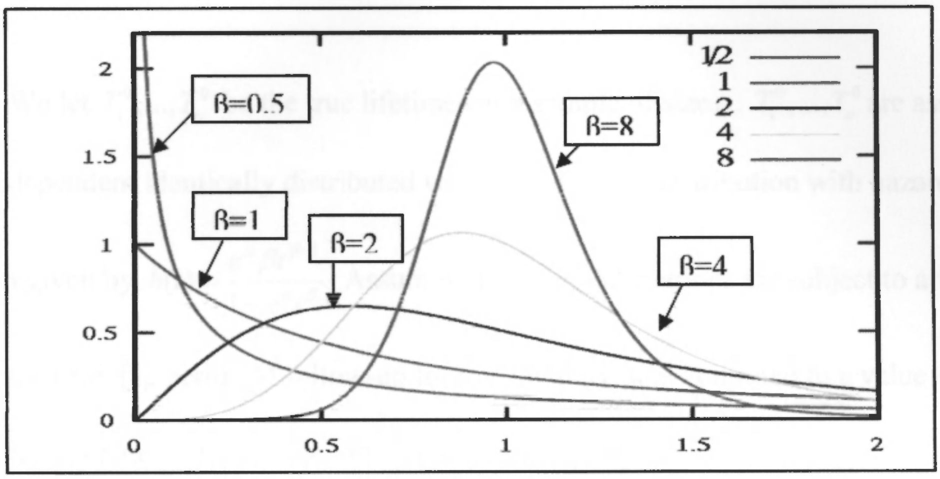

Figure 3.1: The log-logistic hazard functions considering $\mu = 1$, while values of β as shown in legend

Figure 3.1 shows the log-logistic hazard functions considering $\mu = 1$, while values of β as shown in legend. We can see that when β ≤ 1, hazard function decreases with maximum at t=0 while when β >1, hazard function is unimodal where h(t) increases for T≤θ and decreases for T>θ.

In this study, we will only focus on the case of β >1, which will give unimodal hazard function with the change-point at θ. The parameter θ is defined as hazard change-point which is given by $\theta = [(\beta - 1)e^{-\mu}]^{1/\beta}$.

3.2 The Likelihood Function in the Presence of Right Censored Data

We let $T_1^0,...,T_n^0$ be the true lifetimes of a sample of size n. $T_1^0,...,T_n^0$ are assumed to be independent identically distributed with a log-logistic distribution with hazard function given by $h(t) = \frac{e^{\mu}\beta t^{\beta-1}}{1+e^{\mu}t^{\beta}}$. Assuming that the observations are subject to arbitrary right censoring, the period of follow-up for the ith individual is limited to a value C_i. Then, the observed survival time of the ith individual is given by $t_i = \min(T_1^0, C_i)$.

δ_i is defined such that

$$\delta_i = \begin{cases} 0 & \text{if} \quad t_i > T_1^0 \quad \text{(a censored observations)} \\ 1 & \text{if} \quad t_i = T_1^0 \quad \text{(an observed death or failure)} \end{cases}$$

In this study, we also have the likelihood functions for μ and β which is given by

$$L(\mu,\beta \mid t) = \prod_{i=1}^{n} \left(\frac{e^{\mu}\beta t_i^{\beta-1}}{1+e^{\mu}t_i^{\beta}} \right)^{\delta^i} (1+e^{\mu}t_i^{\beta})^{-1}$$

In order to find the Maximum Likelihood Estimation (MLE) for μ and β, we should solve the equations of

$$\begin{cases} \sum_{i=1}^{n} \delta_i S(t_i) - \beta e \mu \sum_{i=1}^{n} t_i (1+e^{\mu} t_i \beta)^{-1} = 0 \\ \frac{1}{\beta} \sum_{i=1}^{n} \delta_i \left[S^{-1}(t_i) + \beta \log t_i \right] S(t_i) \sum_{i=1}^{n} \log(1+e^{\mu} t_i) = 0 \end{cases}$$

where $S(t)$ is given by $S(t) = \frac{1}{1+e^{\mu}t^{\beta}}$.

The maximum likelihood estimator for the change-point θ at the mode of the hazard function assuming $\beta > 1$ is obtained from the maximum likelihood estimators $\hat{\mu}$ and $\hat{\beta}$, that is $\hat{\theta} = [(\hat{\beta}-1)e^{-\hat{\mu}}]^{1/\hat{\beta}}$.

3.3 Asymptotical Confidence Intervals

In this study, we want to compare three methods to find the confidence intervals for μ and β. The first method is asymptotical confidence intervals. Asymptotical confidence intervals for $\theta = g(\mu, \beta)$ are obtained using the delta method, that is, $\hat{\theta} \sim N(\theta; \text{var}(\hat{\theta}))$. Under the delta method, the asymptotical variance of $\hat{\theta} = g(\hat{\mu}, \hat{\beta})$ is given by

$$Var[g(\hat{\mu},\hat{\beta})] = \left[\frac{\partial}{\partial\mu}g(\hat{\mu},\hat{\beta})\right]^2 Var(\hat{\mu}) + \left[\frac{\partial}{\partial\mu}g(\hat{\mu},\hat{\beta})\right]^2 Var(\hat{\beta}) + 2\left[\frac{\partial}{\partial\mu}g(\hat{\mu},\hat{\beta})\right]\left[\frac{\partial}{\partial\mu}g(\hat{\mu},\hat{\beta})\right]Cov(\hat{\mu},\hat{\beta})$$

where

$$\frac{\partial}{\partial\mu}g(\hat{\mu},\hat{\beta}) = -\frac{1}{\beta}\theta\Big|_{\mu=\hat{\mu},\beta=\hat{\beta}} \quad \text{and} \quad \frac{\partial}{\partial\beta}g(\hat{\mu},\hat{\beta}) = -\frac{\theta\{[\log(\beta-1)-\mu](\beta-1)-\beta\}}{\beta^2(\beta-1)}.$$

And the asymptotical variance and covariance for $\hat{\mu}$ and $\hat{\beta}$ are obtained from the inverse of the Fisher information matrix for μ and β.

3.4 Bootstrap Confidence Intervals for μ and β

Alternative to the asymptotical based confidence intervals for θ, we could use parametric bootstrap simulation methods by generating samples of the log-logistic density $f(t) = \frac{e^{\mu}\beta t^{\beta-1}}{[1+e^{\mu}t^{\beta}]^2}$ assuming $\mu = \hat{\mu}$ (or the reparametrization $\hat{\mu} = \log(\beta - 1) - \hat{\beta}\log\hat{\theta}$) and $\beta = \hat{\beta}$ where $\hat{\mu}$ and $\hat{\beta}$ are the obtained maximum likelihood estimates for μ and β.

For the implementation of the parametric bootstrap, we generate psedo-random values from $t = \left(\frac{U}{1-U}e^{-\hat{\mu}}\right)^{1/\hat{\beta}}$ where U has a uniform distribution in the (0,1) interval.

We will introduce the steps for constructing bootstrap confidence intervals. Bootstrap method has an advantage that the joint distribution of the maximum likelihood estimators is not assumed to be normal, unlike in delta method. We consider two types of bootstrap method which are bootstrap-*p* and bootstrap-*t*.

We let $U = (t, \delta)$ be the observed data where $t = (t_1, ..., t_n)$ is the vector of lifetime data and $\delta = (\delta_1, ..., \delta_n)$ is the vector indicators of censored observations.

3.4.1 Bootstrap-p confidence intervals

Steps for constructing bootstrap-p confidence intervals:

[a] Select bootstrap sample, $(t_1^*, \delta_1^*), \ldots, (t_n^*, \delta_n^*)$, randomly (with replacement) from uniform distribution.

[b] From [a], find maximum likelihood estimates of θ denoted as $\hat{\theta}^*$.

[c] Repeat step [a] and [b], B times.

[d] From $\hat{\theta}^* = (\hat{\theta}^*_{(1)} < \hat{\theta}^*_{(2)} < \ldots < \hat{\theta}^*_{(B)})$, find a $100x(1-\alpha)\%$ bootstrap confidence interval given by $(\hat{\theta}^*_{(q1)}, \hat{\theta}^*_{(q2)})$ where $q_1 = [(\alpha/2)B]$ and $q_2 = B - q_1$

3.4.2 Bootstrap-t confidence intervals

Steps for constructing bootstrap-t confidence intervals:

[a] Select bootstrap sample, $(t_1^*, \delta_1^*), \ldots, (t_n^*, \delta_n^*)$, randomly (with replacement) from uniform distribution.

[b] From [a], find maximum likelihood estimates of θ denoted as $\hat{\theta}^*$.

[c] Repeat step [a] and [b], B times.

[d] From $\hat{\theta}^*$ given in step [d] (from bootstrap-p step), find $T^* = (T^*_{(1)}, \ldots, T^*_{(B)})$, where $T^*_{(i)} \leq T^*_{(j)}$ for $i, j = 1, \ldots, B;\ i \neq j$ given by $T_i^* = \dfrac{\hat{\theta}_i^* - \hat{\theta}}{\hat{\sigma}_i^*}$,

where $\hat{\theta}$ is the maximum likelihood estimates for θ and $\widehat{\sigma_i^*}$ is the standard error of $\widehat{\theta_i^*}$. Since $\widehat{\sigma_i^*}$ $(i = 1,..., B)$ can be calculated directly by the inverse of the Fisher information matrix, it is not necessary to resample from the bootstrap sample.

[e] From T^* we find find a $100x(1-\alpha)\%$ bootstrap confidence interval for θ given by $(\hat{\theta} - \hat{\sigma}T^*_{(q_2)}, \hat{\theta} - \hat{\sigma}T^*_{(q_1)})$, where q_1 and q_2 are define in [d] and $\hat{\sigma} = \sqrt{Var(\hat{\theta})}$, ($\hat{\theta}$ and $\hat{\sigma}$ are calculated from the original lifetime data)

REFERENCES

[1] Josmar Mazucheli, et all, Bootstrap Confidence Intervals for the Mode of the Hazard Function, Journal of Computer Methods and Programs in Biomedic, 2004.

[2] Nor Habibah Binti Tarmuji , Alternative Interval Estimation for the Parameters of the Exponential Regression Model, Thesis, 2008.

[3] F.L. Jerald, Statistical Models and Methods for Lifetime Data, Wiley Series in Probability and Statistics, 2003.

[4] D. Collett, Modelling Survival Data in Medical Research, Chapman and Hall, New York, 1994.

[5] P.B.Ng, Bootstrap Confidence Intervals for the Mean Time to Failure, Thesis, 2008.

[6] S. Bennet, Log-Logistic Regression Models for Survival Data, Appl. Stat. 32 (1983) 165-171.

[7] J. Carpenter, J. Bithell, Bootstrap Confidence Intervals: When, Which, What?, Stat. Med. (2000)

[8] N. Veraverbeke, Bootstrapping in Survival Analysis, Stat, 1997.

[9] Format Penulisan Tesis, Manual 2007, UTM, 2007.

[10] Nancy Barker, A Practical Introduction to the Bootstrap Using the SAS System, Oxford Pharmaceutical Sciences, Wallingford, UK, 2006.