2013

The Frequentist Performance of Some Bayesian Confidence Intervals for the Survival Function

Yingfeng Tao

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCE

THE FREQUENTIST PERFORMANCE OF SOME BAYESIAN CONFIDENCE INTERVALS FOR THE SURVIVAL FUNCTION

By

YINGFENG TAO

A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2013

Yingfeng Tao defended this dissertation on May 13, 2013.

The members of the supervisory committee were:

Fred Huffer
Professor Directing Thesis

Giray Okten
University Representative

Debajyoti Sinha
Committee Member

Xufeng Niu
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with the university requirements.

ACKNOWLEDGMENTS

First of all, I would especially like to thank Dr. Fred Huffer for his guidance and encouragement throughout the preparation of my dissertation.

I would also like to thank Dr. Debajyoti Sinha, Dr. Xufeng Niu and Dr. Giray Okten for serving on my committee and for their valuable contributions to my research.

I would also like to show my gratitude to my parents, wife, and friends for their care and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Estimation of a survival function is a very important topic in survival analysis with contributions from many authors. This dissertation considers estimation of confidence intervals for the survival function based on right censored or interval-censored survival data. Most of the methods for estimating pointwise confidence intervals and simultaneous confidence bands of the survival function are reviewed in this dissertation. In the right-censored case, almost all confidence intervals are based in some way on the Kaplan-Meier estimator first proposed by Kaplan and Meier (1958) and widely used as the nonparametric estimator in the presence of right-censored data. For interval-censored data, the Turnbull estimator (Turnbull (1974)) plays a similar role.

For a class of Bayesian models involving Dirichlet priors, Doss and Huffer (2003) suggested several simulation techniques to approximate the posterior distribution of the survival function by using Markov chain Monte Carlo or sequential importance sampling. These techniques lead to probability intervals for the survival function (at arbitrary time points) and its quantiles for both the right-censored and interval-censored cases. This dissertation will examine the frequentist properties and general performance of these probability intervals when the prior is non-informative. Simulation studies will be used to compare these probability intervals with other published approaches. Extensions of the Doss-Huffer approach are given for constructing simultaneous confidence bands for the survival function and for computing approximate confidence intervals for the survival function based on Edgeworth expansions using posterior moments. The performance of these extensions is studied by simulation.

CHAPTER 1

INTRODUCTION

Kaplan & Meier's (1958) estimator for the survival function was developed in the middle of the last century and has served since then as the foundation of the statistical analysis of censored survival data. Many different pointwise confidence intervals and simultaneous confidence bands for the survival function have been suggested based on the asymptotic distributional results of Breslow & Crowley (1974), Aalen & Johansen (1978) and Gill (1980). Also Thomas & Grunkemeier (1975) derived pointwise confidence intervals and studied their small sample properties by Monte Carlo simulations.

Suggestions for simultaneous confidence bands have been developed by Gillespie & Fisher (1979), Hall & Wellner (1980) and Nair (1981, 1984). Csorgo & Horvath (1986) introduced a class of bands showing the relationship between most of these proposals, and suggested several modified versions of these confidence bands. An extensive study of the small sample properties of the most commonly used simultaneous confidence bands was carried out by Nair (1984). He used the arcsine-square root transformation in connection with simultaneous confidence bands and Kalbfleisch & Prentice (1980) suggested the use of the log-minus-log transformation.

This dissertation is concerned with pointwise confidence intervals and simultaneous confidence bands for the survival function derived from a Bayesian approach in which the prior on the survival function is a mixture of Dirichlet processes. There are two very different types of Bayesian models involving "mixtures" and "Dirichlet processes" which Hanson and Johnson (2004) have abbreviated as DPM and MDP, standing for "Dirichlet process mixture" and "mixture of Dirichlet processes", respectively. In an example of the more common DPM approach, represented by the work of Escobar (1994), Escobar and West (1995), Roeder and Wasserman (1997), and

Ghosal et al. (1999), the data is a random sample from a cdf F which is a mixture of continuous parametric distributions (say, normal distributions) with the mixing measure being a realization from a Dirichlet process. This cdf F has a density so that ties in the data occur with probability zero. The less common MDP approach, represented by the work of Ferguson (1973), Antoniak (1974), Doss & Huffer (2003), and Hanson & Johnson (2004), actually predates DPM. In an example of this approach, the data is a random sample from a cdf F which is a realization of a Dirichlet process whose base measure is random, usually chosen from some standard parametric family of distributions. Here F is discrete so that ties occur with positive probability. A Dirichlet process with a random base measure is called a mixture of Dirichlet processes (Antoniak, 1974).

In this dissertation we adopt the MDP approach; our prior on the distribution of lifetimes F is a mixture of Dirichlet processes. In particular, we follow the work of Doss & Huffer (2003) who developed several simulation techniques to approximate the posterior distribution of the survival function by using Markov chain Monte Carlo or sequential importance sampling. They used these techniques to produce Bayesian confidence intervals (probability intervals) for the survival function and its quantiles.

The first objective of this dissertation is to examine the frequentist properties and general performance of the Bayesian confidence intervals for the survival function and its quantiles given by the Doss-Huffer approach when the prior is non-informative, and to conduct simulation studies comparing these intervals with other methods in the literature. We will then extend the Doss-Huffer methodology to obtain a simultaneous confidence band for the survival function and study its performance via simulation. Finally, we propose approximate confidence intervals for the survival function based on an Edgeworth expansion and conduct simulations to verify the accuracy of this approximation. These goals are described in greater detail below. The emphasis throughout will be to study the frequentist performance of these techniques when the prior is non-informative.

It is well known that Bayes estimates with non-informative priors are often very close to the maximum likelihood estimates. In particular, when a non-informative Dirichlet prior is used for the survival function, the posterior mean of the survival function is generally very close to the nonparametric maximum likelihood estimate (NPMLE), which is the Kaplan-Meier estimator for right-censored data and the Turnbull estimator for interval-censored data. This dissertation will empirically investigate

the corresponding situation for confidence intervals, with the hope that Bayesian confidence intervals produced using non-informative priors will perform well enough to be viable competitors or substitutes for confidence intervals derived by frequentist arguments.

We are interested in estimating a survival function S. Let $P(\cdot \,|\, \texttt{data})$ denote a (Bayesian) posterior probability conditional on the data, and $P_S(\cdot)$ a (frequentist) probability computed under S. In Chapters 3 and 4 we consider the construction of pointwise confidence intervals for the survival function. Let $U(t), L(t)$ denote upper and lower limits for $S(t)$ respectively as functions of time t. We will show that, by choosing an appropriate non-informative prior, Bayesian limits satisfying

$$P(L(t) < S(t) < U(t) \,|\, \texttt{data}) = 1 - \alpha \tag{1.1}$$

will also give

$$P_S(L(t) < S(t) < U(t)) \approx 1 - \alpha \tag{1.2}$$

in the frequentist setting.

Chapter 5 investigates confidence intervals for quantiles of the survival function. In the Bayesian approach, probability limits for quantiles of the survival function are obtained by inverting the limits $U(t)$ and $L(t)$ in (1.1). Suppose these limits are chosen to satisfy

$$P(L(t) < S(t) \,|\, \texttt{data}) = P(S(t) < U(t) \,|\, \texttt{data}) = 1 - \alpha/2\,. \tag{1.3}$$

For $0 < p < 1$, define $S^{-1}(p) = \inf\{t \,:\, S(t) \leq p\}$, and similarly define $U^{-1}(p)$ and $L^{-1}(p)$. It is easily seen that

$$P(L^{-1}(p) < S^{-1}(p) < U^{-1}(p) \mid \texttt{data}) = 1 - \alpha \tag{1.4}$$

so that $(L^{-1}(p), U^{-1}(p))$ supplies a $1 - \alpha$ confidence interval for the quantile $S^{-1}(p)$. (Further details are given in Chapter 5.) We will show that, with an appropriately chosen non-informative prior,

$$P_S(L^{-1}(p) < S^{-1}(p) < U^{-1}(p)) \approx 1 - \alpha\,. \tag{1.5}$$

Chapter 6 studies simultaneous confidence bands for the survival function. In a

frequentist setting, the usual approach for obtaining simultaneous confidence bands for the survival function for times $t \in [a, b]$ is to choose a function $\hat{G}(t)$ and determine a critical value c_α such that

$$P_S\left(\sup_{t\in[a,b]}\left|\frac{S(t)-\hat{S}(t)}{\hat{G}(t)}\right| \le c_\alpha\right) = 1-\alpha\,. \tag{1.6}$$

This leads to the confidence band

$$P_S(S(t) \in (\hat{S}(t) - c_\alpha\hat{G}(t), \hat{S}(t) + c_\alpha\hat{G}(t)), \forall t \in [a, b]) = 1-\alpha\,. \tag{1.7}$$

Here $\hat{S}(t)$ denotes an estimate (usually the MLE) of $S(t)$ and $\hat{G}(t)$ is typically some function of an estimate $\hat{\sigma}^2(t)$ of the variance of $\hat{S}(t)$. This approach is easily adapted to Bayesian estimation, simply replacing $\hat{S}(t)$ and $\hat{\sigma}^2(t)$ by the posterior mean $\mu(t)$ and posterior variance $\sigma^2(t)$ of $S(t)$:

$$\mu = E(S(t)\,|\,\texttt{data}) \quad \text{and} \quad \sigma^2(t) = \mathrm{Var}(S(t)\,|\,\texttt{data})\,. \tag{1.8}$$

Now, with $G(t)$ chosen to be some appropriate function of $\sigma^2(t)$, one must use simulation to find a critical value c_α such that

$$P\left(\sup_{t\in[a,b]}\left|\frac{S(t)-\mu(t)}{G(t)}\right| \le c_\alpha \,\middle|\, \texttt{data}\right) = 1-\alpha \tag{1.9}$$

leading to the confidence band

$$P(S(t) \in (\mu(t) - c_\alpha G(t), \mu(t) + c_\alpha G(t)), \forall t \in [a, b]\,|\,\texttt{data}) = 1-\alpha\,. \tag{1.10}$$

We will show empirically that, with a non-informative prior, this band will also satisfy

$$P_S(S(t) \in (\mu(t) - c_\alpha G(t), \mu(t) + c_\alpha G(t)), \forall t \in [a, b]) \approx 1-\alpha\,. \tag{1.11}$$

The approach to obtaining a pointwise confidence interval for $S(t)$ described in (1.1) uses limits $U(t)$, $L(t)$ chosen to satisfy (1.3). Computing these limits requires a considerable amount of calculation. For each value of t for which limits are desired, we must accumulate a certain Monte Carlo sample (see Section 3.2 for details) whose size is at least several thousands, and then compute the upper and lower $\alpha/2$ quantiles

of the posterior distribution of $S(t)$ by a Newton-Raphson algorithm, each iteration of which requires a pass through this Monte Carlo sample. When the number of time points t is large, this can require prohibitively large amounts of computation time and memory. Therefore, in Chapter 7 we develop an approach that approximates the quantiles in (1.3) using the posterior moments of $S(t)$, which can be easily computed at a large number of time-points t using very little memory. In situations where the posterior distribution of $S(t)$ is approximately normally distributed (which we expect if the sample size n is sufficiently large and the posterior mean $\mu(t)$ is not too close to 0 or 1), we can approximate the posterior distribution of $S(t)$ with a direct Edgeworth expansion [Barndorff-Nielsen and Cox (1989)]

$$P\left(\frac{S(t)-\mu(t)}{\sigma(t)} \leq x \;\middle|\; \texttt{data}\right) \approx \Psi(x) \equiv \Phi(x) - \varphi(x)\left\{\frac{\rho_3}{6}H_2(x) + \frac{\rho_4}{24}H_3(x) + \frac{\rho_3^2}{72}H_5(x)\right\} \tag{1.12}$$

where $\mu(t)$, $\sigma(t)$, ρ_3, and ρ_4 are the mean, standard deviation, and standardized skewness and kurtosis, respectively, of the posterior distribution of $S(t)$; Φ and φ are the standard normal cdf and pdf; and H_i denote Hermite polynomials. Solving $\Psi(x) = 1-\alpha/2$ and $\Psi(x) = \alpha/2$ will give easily computed approximations to $U(t)$ and $L(t)$. Chapter 7 describes in detail how to implement this Edgeworth approximation and studies the performance of the resulting pointwise confidence intervals.

CHAPTER 2

BACKGROUND

In this chapter, we will review several basic definitions which are important to our topic and introduce notation which will be used later in this dissertation.

2.1 Censored Survival Data

The field of survival analysis is widely applied to biological organisms, mechanical systems, public health, and so on. A typical question of survival analysis is to analyze the probability a specified person will survive a certain length of time. The survival function is defined to describe this probability and the time to the event of interest is called either a survival time or failure time.

A survival time (or failure time or event time) is the period elapsing between the completion or institution of any procedure and death or some other event of interest. The observed times in survival analysis can be exact, censored, or truncated. Exact data specifies the precise time until the event of interest and is also called uncensored data. Censored data is a type of missing or incompletely observed data which arises when the time until the event of interest is known only to lie in a certain period of time. It may also happen that the event time occurs within a period of time that is not observed at all. This kind of situation is called truncation. In this dissertation, only censored data are considered. In most situations, observation is conducted only during a finite follow-up period, and survival times exceeding this are censored.

In this dissertation work, the survival times for a sample of n individuals are denoted $X_1, X_2, \ldots, X_n$ and are assumed throughout to be iid from a cumulative distribution function (cdf) F with corresponding survival function $S = 1 - F$. The survival function gives the probability of an individual surviving past a specified time

x:

$$S(x) = P(X_i > x) = 1 - F(x)\,. \tag{2.1}$$

We will consider both right-censored and interval-censored data in this work. A lifetime X is right-censored if we do not observe X but only know that X exceeds some specified value C, i.e., the individual was observed only until time C by which time the event of interest had not yet occurred. The value C is known as the censoring time; it might be the time at which an individual dropped out of the study (was lost to follow-up) or the time at which the study was concluded.

We assume the standard setup for right-censoring: for each individual i, $i = 1, \ldots, n$, there is a lifetime X_i and a censoring time C_i, and we observe whichever occurs first. Thus our data consists of

$$\begin{aligned} &\texttt{data} = \{(Z_i, \delta_i),\ i = 1, \ldots, n\} \\ &\text{where } Z_i = \min(X_i, C_i) \text{ and } \delta_i = I(X_i \le C_i) \end{aligned} \tag{2.2}$$

We assume the lifetimes $X_1, \ldots, X_n$ and censoring times $C_1, \ldots, C_n$ are independent. In all our simulations we take $C_1, \ldots, C_n$ to be iid from some distribution, but this is largely for convenience; the iid assumption is not important in most of the theory.

An event time X is interval-censored if we do not observe X but know only that it lies in some specified interval of time (L, R). For example, suppose that X is the time to recurrence of cancer in an individual who is observed only occasionally (say, at scheduled appointment times). If the individual was free of cancer when observed at time L but not when next observed at time R, we know only that $X \in (L, R)$. With interval-censoring, our data consists of n intervals:

$$\texttt{data} = \{(L_i, R_i),\ i = 1, \ldots, n\} \tag{2.3}$$

representing the knowledge that $X_i \in (L_i, R_i)$ for $i = 1, \ldots, n$. Note that right-censoring and left-censoring are special cases of interval censoring corresponding to $R_i = \infty$ and $L_i = 0$, respectively.

In our later simulations, the intervals (L_i, R_i) are determined as follows. For each individual i, we generate a random sequence of inspection (or appointment) times $0 = V_{i0} < V_{i1} < V_{i2} < \cdots$ and take the interval (L_i, R_i) to be the pair of inspection

times which bracket the event time X_i, that is,

$$L_i = V_{ik} \text{ and } R_i = V_{i\,k+1} \text{ where } k \text{ is such that } V_{ik} < X_i < V_{i\,k+1}. \tag{2.4}$$

The event times $\{X_1, \ldots, X_n\}$ must be independent of the collection of inspection times $\{V_{ij}\}$. We take $\boldsymbol{V_1}, \ldots, \boldsymbol{V_n}$ to be iid, where $\boldsymbol{V_i} = (V_{i0}, V_{i1}, V_{i2}, \ldots)$, but this is largely for convenience.

2.2 Nonparametric Estimators

A parametric estimator is commonly used for estimating the survival function for right censored and interval censored data when the distribution is known (or expected) to be well approximated by a member of some parametric family of distributions. However, in real life the form of the distribution is usually unknown, so that we usually rely on nonparametric methods which do not require the data to come from a specified family of distributions. In this dissertation, we will focus on nonparametric techniques. The Kaplan-Meier estimator (Section 2.2.1) is widely used as the nonparametric estimator in the presence of right-censored data. Similarly, nonparametric estimation of the survival function with interval-censored data is available with the Turnbull estimator (Section 2.2.3).

2.2.1 Kaplan-Meier estimator

Suppose the uncensored deaths (events) occur at k ordered distinct times $t_1 < t_2 < \cdots < t_k$, and d_i is the number of uncensored deaths at time t_i. The standard nonparametric estimator of the survival function, first proposed by Kaplan and Meier (1958), is called the Kaplan-Meier (K-M) estimator, also known as the product-limit estimator. This estimator is defined as:

$$\hat{S}(t) = \prod_{i\,:\,t_i \le t} \left(1 - \frac{d_i}{Y_i}\right) \tag{2.5}$$

where Y_i denotes the number of individuals who are at risk at time t_i, that is, who have not experienced the event of interest or been censored by time t_i.

From equation (2.5), we see that the K-M estimator $\hat{S}(t)$ is a step function which equals 1 for $t < t_1$ and jumps downward at each of the times t_i by an amount which

depends on the number of events observed at each event time and the pattern of the censored observations. The K-M estimator $\hat{S}(t)$ is a commonly used nonparametric estimator and software to compute it is widely available. An estimate of the variance of $\hat{S}(t)$ is given by Greenwood's formula:

$$\widehat{\mathrm{Var}}[\hat{S}(t)] = \hat{S}(t)^2\hat{\sigma}^2(t) \quad \text{where} \quad \hat{\sigma}^2(t) = \sum_{i\,:\,t_i \le t} \frac{d_i}{Y_i(Y_i - d_i)} \tag{2.6}$$

2.2.2 Nelson-Aalen estimator

The K-M estimator can be used to estimate the cumulative hazard function $\Lambda(t) = -\ln[S(t)]$ in the obvious way: $\tilde{\Lambda}(t) = -\ln[\hat{S}(t)]$. An alternate estimator of the cumulative hazard rate was first suggested by Nelson (1972) in a reliability context and rediscovered by Aalen (1978), so that it is commonly referred to as the Nelson-Aalen estimator. Using the same notation as in equation (2.5), it is defined by

$$\hat{\Lambda}(t) = \sum_{i\,:\,t_i \le t} \frac{d_i}{Y_i} \tag{2.7}$$

The variance of the Nelson-Aalen estimator is estimated by :

$$\widehat{\mathrm{Var}}[\hat{\Lambda}(t)] = \hat{\sigma}^2(t) = \sum_{i\,:\,t_i \le t} \frac{d_i}{Y_i^2} \tag{2.8}$$

2.2.3 Turnbull estimator

For interval-censored data, Peto (1973) developed a Newton-Raphson method for obtaining the nonparametric maximum likelihood estimator (NPMLE) of the survival function. Then Turnbull (1974) described an EM algorithm for the NPMLE. This algorithm is now given.

To construct Turnbull's estimator, let $0 = t_0 < t_1 < t_2 < \cdots < t_m$ be a grid of times which includes 0 and all the points L_i and R_i in (2.3) for $i = 1, \ldots, n$. For convenience in describing the algorithm, we suppose the survival function S places all the probability mass at the points $t_1, t_2, \ldots, t_m$ so that these are the only possible event times. For the ith observation, define the weight α_{ij} to be 1 if $(t_{j-1}, t_j]$ is contained in the interval $(L_i, R_i]$ and 0 otherwise. The weight α_{ij} indicates whether t_j is a possible value for the event time X_i known to lie in the interval $(L_i, R_i]$. An

initial guess for $S(t_j)$, $j = 0, \ldots, m$, is made and then Turnbull's algorithm proceeds as follows:

Step 1: Compute the probability of an event occurring at time t_j by:

$$p_j = S(t_{j-1}) - S(t_j), \ j = 1, \cdots, m;$$

Step 2: Estimate the number of events which occurred at t_j by:

$$d_j = \sum_{i=1}^{n} \frac{\alpha_{ij} p_j}{\sum_{k=1}^{m} \alpha_{ik} p_k}, \quad j = 1, \cdots, m;$$

Step 3: Compute the estimated number at risk at time t_j by $Y_j = \sum_{k=j}^{m} d_k$;

Step 4: Compute the updated S to be the "Product-Limit" estimator obtained by applying (2.5) to the pseudo-data d_j, Y_j, $j = 1, \ldots, m$, found in Steps 2 and 3.

Steps 1 to 4 are repeated until S converges, that is, until the change in S from one iteration to the next is sufficiently small. This final version of S is the Turnbull estimate for the survival function.

Software to compute the Turnbull estimator is widely available, e.g., one may use the R package "Interval" [Fay (2010)] or "Survival" [Therneau (2010)] available on the R CRAN website (http://cran.r-project.org/web/packages/).

2.3 Dirichlet Processes

In this section, we will review the basic definitions and propositions concerning Dirichlet processes and mixtures of Dirichlet processes. This material may be found in Ferguson (1973), Antoniak (1974), and Doss & Huffer (2003).

All our Dirichlet processes will be on the real line, so we shall state our definitions only in this case.

Definition 1: Let α be a finite measure on the Borel sets $\mathcal{B}$ of the real line $\mathbb{R}$. We call a random probability measure P on $(\mathbb{R}, \mathcal{B})$ a Dirichlet process (denoted $P \sim \mathcal{D}_\alpha$) with parameter measure α, if for each Borel-measurable partition $G_1, G_2, \ldots, G_k$, $k = 1, 2, \ldots$ of $\mathbb{R}$, the joint distribution of the random probabilities denoted as $(P(G_1), P(G_2), \ldots, P(G_k))$ is Dirichlet with the parameters $(\alpha(G_1), \alpha(G_2), \ldots, \alpha(G_k))$.

Notation: If $P \sim \mathcal{D}_\alpha$ and F is the cdf corresponding to P, we say also that $F \sim \mathcal{D}_\alpha$.

An immediate consequence of **Definition 1** is the following.

Proposition 1: If $P \sim \mathcal{D}_\alpha$, and $A \in \mathcal{B}$, then the expectation $E(P(A)) = \alpha(A)/\alpha(\mathbb{R})$, furthermore the marginal distribution of $P(A) \sim \text{Beta}(\alpha(A), \alpha(\mathbb{R} \setminus A))$.

Much of our theory and Monte Carlo work relies on the next result.

Proposition 2: If $P \sim \mathcal{D}_\alpha$ and conditional given P, the random variables $X_1, \ldots, X_n$ are iid from P, then $P \,|\, X_1, \ldots, X_n \sim \mathcal{D}_{\alpha + \sum_{i=1}^n \delta_{X_i}}$, where δ_{X_i} gives mass 1 to X_i.

If the measure α in **Definition 1** itself is a random measure, we say P is a mixture of Dirichlet processes. All the mixtures of Dirichlet processes that we use have a special form which we now describe.

Definition 2: Let $\{H_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ be a low-dimensional parametric family of absolutely continuous distributions on $\mathbb{R}$, and $M > 0$ be a fixed real value. For each $\theta \in \Theta$, define $\alpha_\theta = M H_\theta$. We say that the distribution of P is a mixture of Dirichlet processes if

$$P \sim \int \mathcal{D}_{\alpha_\theta} \, \nu(d\theta) \tag{2.9}$$

for some probability measure ν on Θ. If θ is taken to be a random vector, then (2.9) is equivalent to the statement:

$$\theta \sim \nu \quad \text{and} \quad P \,|\, \theta \sim \mathcal{D}_{\alpha_\theta} \,. \tag{2.10}$$

Proposition 3: Suppose P has the prior distribution given in (2.9), and conditional on P, the data $X_1, X_2, \ldots, X_n$ is iid from P. Then the posterior distribution of P is the mixture of Dirichlet processes given by

$$P \,|\, X_1, \ldots, X_n \sim \int D_{\alpha_\theta + \sum_{i=1}^n \delta_{X_i}} \, \nu_{\mathbf{X}}(d\theta), \tag{2.11}$$

where

$$\nu_{\mathbf{X}}(d\theta) \;=\; c(\mathbf{X}) \left(\prod^{dist} h_\theta(X_i) \right) \nu(d\theta) \tag{2.12}$$

is an absolutely continuous measure with respect to ν, h_θ is the density of H_θ, "*dist*" indicates the product over the distinct values in $X_1, \ldots, X_n$, and $c_{\mathbf{X}}$ is a normalizing constant.

We use these definitions and propositions in the later sections.

CHAPTER 3

CONFIDENCE INTERVALS FOR THE SURVIVAL FUNCTION

3.1 General Pointwise Confidence Intervals

After the Kaplan-Meier Estimator was developed in the last century, many pointwise confidence intervals for the survival function have been suggested based on the asymptotic distributional results of Breslow & Crowley (1974), Aalen & Johansen (1978), Gill (1980), Thomas & Grunkemeier (1975) for the right-censored case, and Goodall, Dunn & Babiker (2004) for the interval-censored case.

3.1.1 Right censoring

Using the asymptotic distributional results for the Kaplan-Meier estimator (Section 2.2.1), approximate confidence intervals for the survival function may be derived.

The simplest $100(1-\alpha)\%$ confidence interval for $S(t)$ for a fixed t is:

$$\hat{S}(t) \pm z_{\alpha/2}\hat{S}(t)\hat{\sigma}(t) \tag{3.1}$$

where $\hat{\sigma}^2(t)$ is given in (2.6) and $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution. We consider the transformations $g(x) = \log(x)$, $g(x) = \log(-\log x)$, and $g(x) = \arcsin\sqrt{x}$ to derive confidence intervals with better small sample properties. The log transformation gives $100(1-\alpha)\%$ interval:

$$\hat{S}(t)\exp\{\pm z_{\alpha/2}\hat{\sigma}(t)/\hat{S}(t)\} \tag{3.2}$$

and the log-minus-log transformation has the $100(1-\alpha)\%$ interval:

$$\hat{S}(t)^{\exp\{\pm z_{\alpha/2}\hat{\sigma}(t)/\log \hat{S}(t)\}} \tag{3.3}$$

Also the arcsin-square-root transformation gives the interval:

$$\begin{aligned}
&\sin^2\{\max[0, \arcsin(\hat{S}(t)^{1/2}) - 1/2 z_{\alpha/2}\hat{\sigma}(t)[\hat{S}(t)/\{1-\hat{S}(t)\}]^{1/2}]\} \\
&\quad \leq S(t) \leq \\
&\quad \sin^2\{\min[\pi/2, \arcsin(\hat{S}(t)^{1/2}) + 1/2 z_{\alpha/2}\hat{\sigma}(t)[\hat{S}(t)/\{1-\hat{S}(t)\}]^{1/2}]\}
\end{aligned} \tag{3.4}$$

These will be denoted the logarithmic, log-minus-log, and the arcsine-transformed confidence intervals for short. In practice one only computes these intervals for values of t larger than the first observed time of death and less than the largest observed death-time.

Thomas & Grunkemeier (1975) introduced likelihood ratio based pointwise confidence intervals for the survival function. Kaplan and Meier (1958) suggested that the Kaplan-Meier estimator $\hat{S}(t)$, given by (2.5) is a nonparametric maximum likelihood estimate of S since it maximizes the likelihood function

$$L(S) = \prod_{i\,:\,\delta_i=1} [S(Z_i-) - S(Z_i)] \prod_{i\,:\,\delta=0} S(Z_i),$$

where the two products are taken over uncensored and censored times, respectively, and Z_i and δ_i are as defined in (2.2). In terms of this likelihood function, Thomas & Grunkemeier devised a method for testing $H_0 : S(t) = p$ for arbitrary $t > 0$ and $0 < p < 1$. Let Θ denote the family of all discrete survival functions with support on the uncensored lifetimes of the data. Their test is based on the nonparametric log-likelihood ratio statistic given by:

$$\begin{aligned}
R(p,t) &\equiv \log \frac{\sup\{L(S) :\ S(t) = p,\ S \in \Theta\}}{L(\hat{S})} \\
&= \sum_{i\,:\,t_i \leq t} \left\{ (Y_i - d_i) \log\left(1 + \frac{\lambda}{Y_i - d_i}\right) - Y_i \log\left(1 + \frac{\lambda}{Y_i}\right) \right\}
\end{aligned} \tag{3.5}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimator, λ (the Lagrange multiplier) is the unique solution of

$$\prod_{i\,:\,t_i \le t} \left(1 - \frac{d_i}{Y_i + \lambda}\right) = p$$

and the quantities t_i, d_i, Y_i are as defined in (2.5). They showed that

$$\{p : \ -2R(p,t) \le \chi^2_{1,\alpha}\} \tag{3.6}$$

is an asymptotic $100(1-\alpha)\%$ confidence interval for $S(t)$, where $\chi^2_{1,\alpha}$ is the upper α percentile of the chisquare distribution with 1 degree of freedom.

The R packages named "km.ci" [Strobl (2009)] and "survival" [Therneau (2010)] are available to compute various pointwise confidence intervals for the survival function based on the Kaplan-Meier estimator.

3.1.2 Interval censoring

For interval-censored data, research on methods for obtaining confidence intervals for the non-parametric survival function has been more limited. Peto and Turnbull (1973) described the derivation of pointwise confidence intervals based on the observed information matrix. Goodall, Dunn & Babiker (2004) described three methods for calculating pointwise confidence intervals. The first method, called the Wald-1 Method, is based on the full information matrix; the second method (Wald-2) is a modification of this approach involving deletion of rows and columns of the information matrix corresponding to zero estimates prior to inversion; the third is based on likelihood ratio inference. These methods are described more fully below.

Wald-1 method . Assume we have interval-censored data as in (2.3). We follow the notation used to describe the Turnbull estimator (the NPMLE) in Section 2.2.3: Let $0 = t_0 < t_1 < \cdots < t_m$ denote the unique ordered values among 0 and L_i, R_i, $i = 1, \ldots, n$; define $p_j = S(t_{j-1}) - S(t_j)$ for $j = 1, \ldots, m$; and define $\alpha_{ij} = 1$ if $(t_{j-1}, t_j] \subseteq (L_i, R_i]$ and $\alpha_{ij} = 0$ otherwise. Let $\boldsymbol{p} = (p_1, \ldots, p_m)^T$.

The log-likelihood function for interval-censored data is:

$$l(\boldsymbol{p}) = \sum_{i=1}^{n} \log(S(L_i)) - S(R_i)) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} \alpha_{ij} p_j\right)$$
$$\text{subject to} \quad p_j \geq 0 \ \forall j \quad \text{and} \quad \sum_{j=1}^{m} p_j = 1. \tag{3.7}$$

An iterative method such as the EM algorithm is required to compute the maximum likelihood estimate (MLE) $\hat{\boldsymbol{p}} = (\hat{p}_1, \cdots, \hat{p}_m)$ since there is no closed form solution.

Wald type confidence intervals are easily derived from:

$$\text{Var}(\hat{\boldsymbol{p}}) \approx \left[-\frac{\partial^2 l}{\partial p_r \partial p_s}\right]^{-1}_{r,s} = (A^T D A)^{-1} \tag{3.8}$$

where A is the $n \times m$ matrix with elements α_{ij} and D the $n \times n$ diagonal matrix with elements $d_{ii} = 1/(\sum_{j=1}^{m} \alpha_{ij} p_j)^2$. The estimator of the survival function at t_j is $\hat{S}(t_j) = 1 - \sum_{k=1}^{j} \hat{p}_k$, which can be written as $1 - [\mathbf{1}_j^T \mathbf{0}_{m-j}^T]\hat{\boldsymbol{p}}$, where $\mathbf{1}_j$ is the unit vector of length j and $\mathbf{0}_{m-j}$ the null vector of length $m - j$.

By Equation (3.8) the approximate variance of the estimate $\hat{S}(t_j)$ is:

$$\widehat{\text{Var}}(\hat{S}(t_j)) = [\mathbf{1}_j^T \mathbf{0}_{m-j}^T](A^T D A)^{-1} \begin{bmatrix} \mathbf{1}_j \\ \mathbf{0}_{m-j} \end{bmatrix}. \tag{3.9}$$

so that the standard $100(1-\alpha)\%$ confidence interval is given by:

$$\hat{S}(t_j) \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{S}(t_j))} \tag{3.10}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution. This will be referred to as the Wald-1 method.

Wald-2 method . The Wald-1 method is based on asymptotic theory assuming a fixed number of parameters. This assumption is not true here, but the problem of hyper-dimensionality can be reduced as follows:

(1). p_m can be expressed as $p_m = 1 - \sum_{j=1}^{m-1} p_j$ and thus eliminated from the model;

(2). It has been shown that $\hat{p}_j \neq 0$ only if $t_{j-1} \in \{L_i\}$ and $t_j \in \{R_i\}$ [Alioum (1996)]. Elements of $\hat{\boldsymbol{p}}$ which do not satisfy this condition must be zero and can

be eliminated from the model;

(3). Sometimes elements of $\hat{\boldsymbol{p}}$ which do satisfy the condition above in (2) may also be zero. It is suggested that the rows and columns of the information matrix corresponding to these 0 elements should be deleted before performing the matrix inversion.

Incorporating these modifications leads to what is referred to as the Wald-2 method.

Likelihood ratio method . Confidence intervals can also be constructed using the likelihood ratio test. The likelihood ratio statistic for testing

$$H_0 : S(t_j) = S_0 \qquad \text{is} \qquad W(S_0) = 2[l(\hat{\boldsymbol{p}}) - l(\tilde{\boldsymbol{p}})]$$

where $\tilde{\boldsymbol{p}}$ is the MLE of $\boldsymbol{p}$ subject to the constraint $S(t_j) = 1 - [\mathbf{1}_j^T \mathbf{0}_{m-j}^T]\tilde{\boldsymbol{p}} = S_0$. An approximate $100(1-\alpha)\%$ confidence interval is given by the set of values

$$\{S_0 : W(S_0) < \chi_\alpha^2\},$$

where χ_α^2 is the upper α point of the χ^2-distribution with 1 degree freedom. This will be referred to as the likelihood ratio method.

Our simulation studies for all of these three methods use the SAS program ICE.sas which is available in the SAS/IML sample library (under the name Interval Censored Estimation Macro). This fits the non-parametric maximum likelihood estimator (NPMLE) to interval-censored data and was adapted to allow estimation of Wald-1, Wald-2, and likelihood ratio confidence limits.

3.2 Bayesian Pointwise Confidence Intervals

In this section we consider a Bayesian approach to finding confidence intervals for the survival function S, or equivalently the cdf F. The assumptions we make regarding the data remain the same: the lifetimes $X_1, \ldots, X_n$ are iid from F, and the observed data consists of (2.2) or (2.3). We may unify these two cases by saying that our observed data consists of the knowledge that $X_i \in A_i$, $i = 1, \ldots, n$, for particular given sets $A_1, \ldots, A_n$. An uncensored observation corresponds to a singleton set $A_i = \{X_i\}$, a right-censored observation to $A_i = (C_i, \infty)$, and an interval-censored observation to $A_i = (L_i, R_i)$.

In addition to this, our Bayesian model assumes that F is random, being sampled from a prior distribution which we take to be a mixture of Dirichlet processes (section 2.3). For any measure α on the real line $\mathbb{R}$, let $\mathcal{D}_\alpha$ denote a Dirichlet process with parameter measure α. Let $\{H_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ be a low-dimensional parametric family of distributions on $\mathbb{R}$. The mixture of Dirichlet's prior that we use has the following form: we specify a distribution ν on Θ and a scalar $M > 0$, and then assume

$$\theta \sim \nu\,, \quad \text{and given } \theta,\ F \sim \mathcal{D}_{\alpha_\theta} \text{ where } \alpha_\theta = MH_\theta. \tag{3.11}$$

This prior is determined by ν (the prior distribution on θ) and M, which is referred to as the "mass" parameter since it determines the total mass of the measure α_θ ($\alpha_\theta(\mathbb{R}) = M$). Doss & Huffer (2003) consider survival analysis using this general class of priors, and give detailed algorithms and software for implementing the special case where

$$\begin{aligned} &H_\theta \text{ is the exponential distribution with density } h_\theta(x) = \theta e^{-\theta x} \text{ and} \\ &\nu \text{ is the Gamma}(a, b) \text{ distribution with density } \propto \theta^{a-1}e^{-b\theta}. \end{aligned} \tag{3.12}$$

In this special case, the prior is completely determined by the triple (a, b, M).

Doss (1994) suggested that estimators based on mixtures of Dirichlet processes interpolate between the purely parametric and nonparametric models, with the degree of interpolation determined by M. The estimators are essentially equal to the Bayes estimator based on the parametric model for large values of M, and for small M, the estimators are essentially equal to the nonparametric maximum likelihood estimator (NPMLE) except for times t in periods where the data is very sparse or nonexistent. In those regions, the estimators make use of the parametric model.

In general, for the prior in (3.11), the exact computation of the posterior distribution of the survival function is impossible to obtain. Doss & Huffer (2003) consider Monte Carlo methods based on sequential importance sampling [Kong (1994)] or Markov chain Monte Carlo to estimate the posterior distribution. For the unobserved lifetime variables X_i and θ in (3.11), their Monte Carlo methods produce realizations of

$$(X_1, X_2, \ldots, X_n, \theta) \ \sim\ \mathcal{L}(X_1, X_2, \cdots, X_n, \theta \mid \texttt{data}). \tag{3.13}$$

These realizations are iid in the right-censored case where they use sequential importance sampling [Kong (1994)], but correlated for the interval-censored case where

they use a Gibbs sampler [Escobar (1988, 1994)].

Given realizations from (3.13), it is straightforward to estimate the posterior distribution of various quantities of interest. From (section 2.3) proposition 2, we have

$$\mathcal{L}(F \,|\, X_1, \ldots, X_n, \theta) = \mathcal{D}_\beta \quad \text{where } \beta = \alpha_\theta + \sum_{i=1}^{n} \delta_{X_i}$$

which implies for $S(t) = 1 - F(t)$ that

$$\begin{aligned} \mathcal{L}(S(t) \mid X_1, X_2, \ldots, X_n, \theta) \;&\sim\; \text{Beta}\,(a(t), b(t)) \\ &\text{where } a(t) = \beta((t, \infty)) \text{ and } b(t) = \beta((0, t]). \end{aligned} \tag{3.14}$$

For the special case of (3.12), we have

$$a(t) \;=\; M e^{-\theta t} + \sum_{i=1}^{n} I(X_i > t)\,, \quad b(t) = M + n - a(t)\,. \tag{3.15}$$

According to (3.14), if we generate a large number J of realizations from (3.13), and compute $(a(t), b(t))$ for each realization, resulting in $(a^{(j)}(t), b^{(j)}(t))$, $j = 1, \ldots, J$, we may approximate the posterior distribution by the mixture of Beta distributions

$$\mathcal{L}(S(t) \,|\, \texttt{data}) \approx \frac{1}{J} \sum_{j=1}^{J} \text{Beta}(a^{(j)}(t), b^{(j)}(t))\,. \tag{3.16}$$

We can then numerically approximate the upper and lower $\alpha/2$ quantiles of this mixture of Beta's to obtain probability limits $L(t)$ and $U(t)$ approximately satisfying (1.3), leading to the probability interval in (1.1).

For the prior in (3.11) and (3.12), Doss & Huffer (2003) gave examples showing that, for a noninformative prior where all the values in (a, b, M) are small, the posterior mean $E(S(t) \,|\, \texttt{data})$ as a function of t is close to the NPMLE of $S(t)$, which is the Kaplan-Meier estimator (2.5) for right-censored data, and the Turnbull estimator of Section 2.2.3 for interval-censored data. Susarla & Van Ryzin (1976) showed the posterior mean will be the Kaplan-Meier estimator as $\alpha(\mathbb{R}^+) \to 0$ if the prior is a single Dirichlet. It is one goal to show that the corresponding Bayesian probability intervals for $S(t)$ have a similar approximate nonparametric interpretation.

The R programs *ritcen* and *gibbs1* implement the Bayesian probability (confidence) intervals for $S(t)$ described above for right-censored and interval-censored data, respectively. These programs are modifications of the programs (also named *ritcen* and *gibbs1*) described in Doss & Huffer (2003).

CHAPTER 4

SIMULATION PROCEDURE & RESULTS: CONFIDENCE INTERVALS FOR THE SURVIVAL FUNCTION

In this chapter simulation procedures and results are discussed for the pointwise confidence intervals for the survival function introduced in Chapter 3. Simulation studies of the coverage probability are conducted for both right censored data and interval censored data.

As stated earlier, one of the goals of this dissertation is to compare the properties and general performance of Bayesian confidence intervals for the survival function with intervals derived by frequentist arguments. In particular, we compare the Bayesian intervals with intervals based on the the Kaplan-Meier Estimator (right-censored case) and Turnbull Estimator (interval-censored case).

4.1 Right Censored Data

In our simulations we generate samples of iid lifetimes $X_1, X_2, \cdots, X_n$ and iid censoring times $C_1, C_2, \cdots, C_n$ and compute right-censored data $Z_i = \min(X_i, C_i)$ and $\delta_i = I(X_i \leq C_i)$, $i = 1, \ldots, n$, as in (2.2). The simulations use common distributions such as the Exponential, Weibull, and Uniform distribution. Specifically, lifetimes are generated from one of three distributions: Exponential with mean 1 (with survival function $S(t) = e^{-t}$), or Weibull with a shape parameter of $\alpha = 2$ or $\alpha = 0.5$ ($S(t) = e^{-x^{\alpha}}$). The censoring times are Uniform(0,2) leading to censoring percentages of approximately 43%, 48%, and 41% respectively. We use samples of size $n = 50$ or $n = 100$, focusing on relatively small samples since that is the more difficult situation; any

reasonable method will perform well for large samples. For each simulated sample, we use the program *ritcen* (Sec. 3.2) with a noninformative prior $(a, b, M) = (.1, .1, .1)$ to compute the Bayesian 95% confidence interval $(L(t), U(t))$ for particular time points t, and observe whether or not this interval contains the true value $S(t)$. This is repeated many times, leading to estimates of the coverage probability for our intervals. For the simulations in this section, the empirical coverage probabilities are each based on 10,000 simulated data sets. For each data set, *ritcen* uses a Monte Carlo sample size of 5,000 in computing the confidence intervals, that is, there are $J = 5{,}000$ realizations of (3.13) used in (3.16).

The coverage probabilities obtained by the above approach are labeled "Bayes" in our tables. They are compared with the coverage probabilities produced by the intervals in (3.1) to (3.6) which are labeled Plain, Log, Log(-log), Arcsin, and T&G, respectively, in our tables. These intervals are computed using the R packages "Survival" [Therneau (2010)] and "KM.CI" [Strobl (2009)], both available on the R CRAN website http://cran.r-project.org/web/packages/.

The description of our simulations is summarized below:

- 95% confidence intervals;
- 10,000 simulated data sets;
- Sample sizes $n = 50$ and $n = 100$;
- Lifetime distributions: Exponential with mean 1, and Weibull with shape parameters $\alpha = 2$ or $\alpha = 0.5$ (and scale parameter = 1);
- Censoring distribution: Uniform(0,2);
- Prior $(a, b, M) = (.1, .1, .1)$;
- Bayesian confidence intervals (Bayes) based on 5,000 Monte Carlo realizations from the posterior;
- Empirical coverage probabilities reported for "Bayes", "Plain", "Log", "Arcsin", "Log(-log)", and "T&G" confidence intervals.

Priors (a, b, M) where a, b, and M are all small are considered "noninformative" because, at least empirically, with these priors, the posterior mean of the survival function is close to the NPMLE, which for right-censored data is the Kaplan-Meier

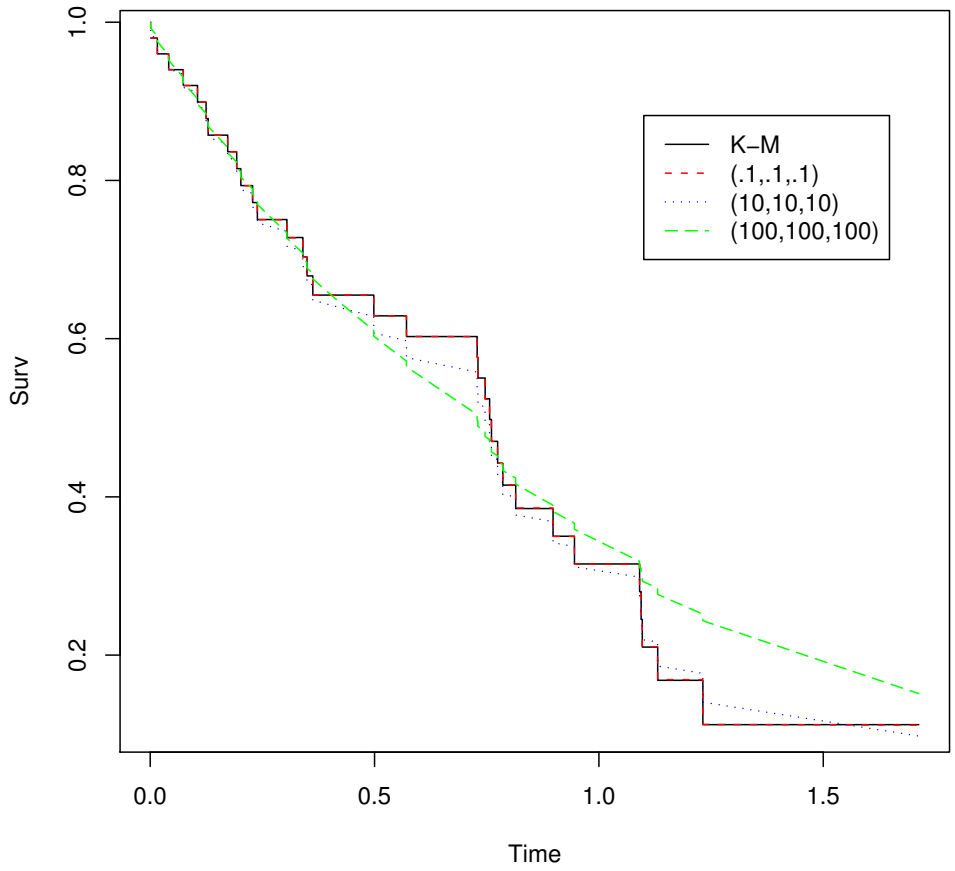

Figure 4.1: The Kaplan-Meier Estimator and three different Bayes estimates obtained using $(a, b, M) = (.1, .1, .1), (10, 10, 10)$ and $(100, 100, 100)$.

estimator. We include one typical numerical illustration of this. Figure 4.1 compares the Kaplan-Meier Estimator with the posterior mean of the survival function (computed using *ritcen* for simulated data with Exp(1) lifetimes, Uniform(0,2) censoring times and sample size $n = 50$) resulting from three different priors with $(a, b, M) = (.1, .1, .1)$, $(10, 10, 10)$ and $(100, 100, 100)$, respectively. In this case, clearly $(a, b, M) = (.1, .1, .1)$ gives the closest agreement with the Kaplan-Meier Estimator. Similar examples can be given for interval-censored data. For survival data with event times having typical magnitudes, the particular values $(a, b, M) = (.1, .1, .1)$ seem to be small enough for our purpose; taking values of a and M which are much smaller

Table 4.1: Values of t and $S(t)$

Exponential with mean 1	t	0.2	0.6	1.0
	$S(t)$	0.8187	0.5488	0.3679
Weibull with $\alpha = 2$	t	0.3	0.8	1.2
	$S(t)$	0.9139	0.5273	0.2369
Weibull with $\alpha = 0.5$	t	0.01	0.5	1.0
	$S(t)$	0.9048	0.4931	0.3679

than this can lead to poor performance of the *ritcen* and *gibbs1* programs (i.e., they crash).

Our simulation results are presented in Table 4.2. For each lifetime distribution, this table gives coverage probabilities for confidence intervals for $S(t)$ at three selected time points. These time points and the corresponding values of $S(t)$ are listed in Table 4.1.

From the results of Table 4.2, we can see the "Plain" and "Log" methods have poor coverage probabilities as expected, but the other methods' results are fairly reasonable, with "Bayes" and "T&G" having excellent agreement with the nominal value of 0.95 in most cases.

4.2 Interval Censored Data

For interval-censored data, in our simulations we generate the data as in Section 4.1 except that now the censoring mechanism is as described in (2.4). Each individual is inspected at times $V_{i1}, V_{i2}, \ldots, V_{i10}$ forming the first 10 arrivals of a Poisson process with a mean time of 0.4 between arrivals. The arrival times bracketing X_i give us the interval (L_i, R_i), with $L_i = 0$ if the event occurs before the first inspection and $R_i = \infty$ if the event occurs after the last inspection. The Bayesian 95% confidence intervals for $S(t)$ (for t in Table 4.1) are computed using the *gibbs1* algorithm of Doss and Huffer (2003), still using a Monte Carlo sample size of 5,000. The empirical coverage probabilities (based on 10,000 simulation replications) are displayed in Table 4.3, along with values from the Wald-1, Wald-2, and Likelihood Ratio methods described in Section 3.1.2 for comparison.

As in the right censoring case, we want all prior parameters to be small enough so that the posterior mean of the survival function is reasonably close to the Turnbull Estimator. Figure 4.2 compares the Turnbull estimator with the posterior mean

Table 4.2: Performance of pointwise confidence intervals for right-censored data.

Dist	Coverage	Bayes			Plain			Log		
	t=	0.200	0.600	1.000	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.947	0.951	0.946	0.932	0.943	0.936	0.937	0.946	0.946
Exp(1)	n=100	0.946	0.945	0.946	0.937	0.942	0.942	0.936	0.944	0.948
Unif(0,2)										
Cen(43%)	Coverage	Arcsin			Log(-log)			T & G		
	t=	0.200	0.600	1.000	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.939	0.951	0.950	0.958	0.953	0.956	0.949	0.952	0.950
	n=100	0.946	0.946	0.948	0.951	0.947	0.951	0.947	0.947	0.949
	Coverage	Bayes			Plain			Log		
	t=	0.200	0.600	1.000	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.955	0.943	0.940	0.907	0.935	0.937	0.909	0.940	0.951
Wei(2)	n=100	0.948	0.951	0.952	0.926	0.941	0.927	0.925	0.941	0.946
Unif(0,2)										
Cen(48%)	Coverage	Arcsin			Log(-log)			T & G		
	t=	0.200	0.600	1.000	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.932	0.951	0.951	0.955	0.950	0.962	0.963	0.951	0.956
	n=100	0.935	0.947	0.943	0.953	0.954	0.947	0.940	0.947	0.944
	Coverage	Bayes			Plain			Log		
	t=	0.200	0.600	1.000	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.937	0.945	0.943	0.869	0.937	0.933	0.856	0.942	0.932
Wei(.5)	n=100	0.946	0.947	0.949	0.909	0.945	0.947	0.915	0.938	0.947
Unif(0,2)										
Cen(41%)	Coverage	Arcsin			Log(-log)			T & G		
	t=	0.200	0.600	1.000	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.937	0.946	0.949	0.950	0.949	0.956	0.937	0.948	0.953
	n=100	0.952	0.948	0.950	0.968	0.952	0.952	0.946	0.949	0.951

(computed using *gibbs1*) for three different choices of the prior: $(a, b, M) = (.1, .1, .1)$, $(10, 10, 10)$, and $(100, 100, 100)$. This is done for simulated interval-censored data with sample size $n = 50$, Exp(1) lifetimes, and inspection times consisting of the first 10 arrivals of a Poisson process with mean time 0.4 between arrivals. As in the right-censored example given earlier, the prior $(.1, .1, .1)$ gives the closest match with the Turnbull estimator, although the agreement is not as good as in the earlier case. one of the possible reasons is the NPMLE and the nonparametric Bayes estimator are not exactly the same for the interval censored data. Mai Zhou (2004) showed several examples for this case.

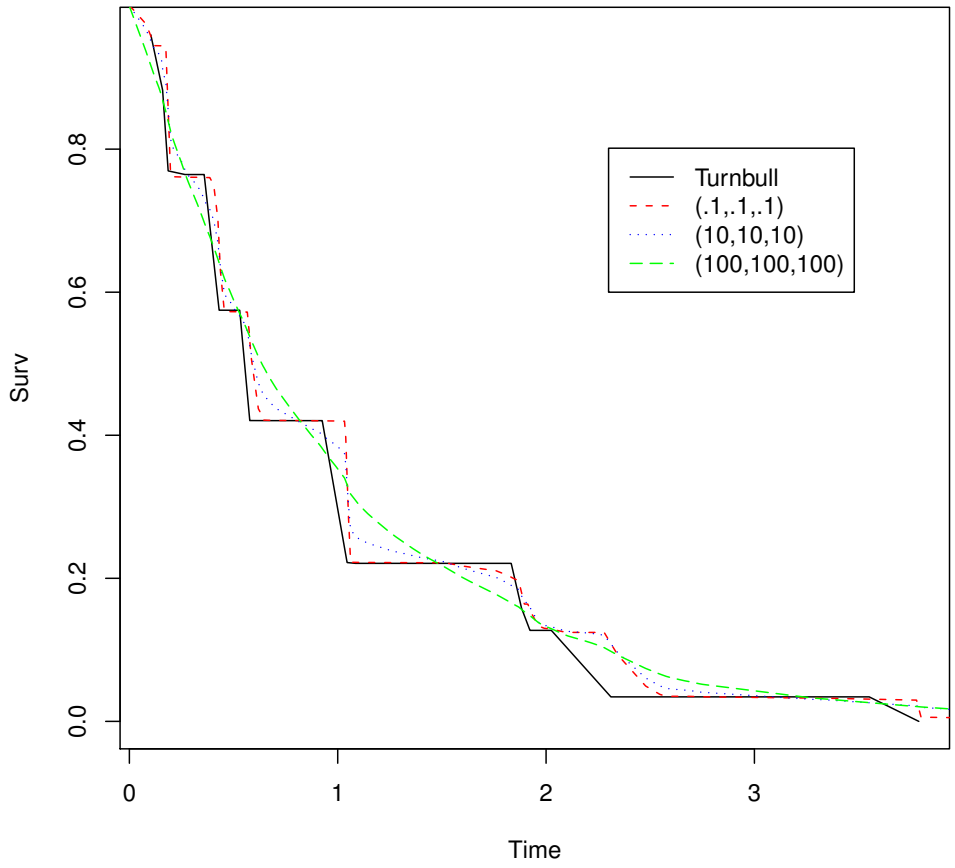

Figure 4.2: The Turnbull Estimator and three different Bayes estimates obtained using $(a, b, M) = (.1, .1, .1), (10, 10, 10)$ and $(100, 100, 100)$.

Table 4.3: Performance of pointwise confidence intervals for interval-censored data.

Dist	Coverage	Bayes			Wald-1		
	t=	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.938	0.945	0.947	0.921	0.957	0.939
	n=100	0.948	0.955	0.945	0.936	0.932	0.933
Exp(1)							
Pois(0.4)	Coverage	Wald-2			Likelihood		
	t=	0.200	0.600	1.000	0.200	0.600	1.000
	n=50	0.936	0.945	0.961	0.953	0.948	0.944
	n=100	0.944	0.956	0.935	0.948	0.945	0.949
	Coverage	Bayes			Wald-1		
	t=	0.300	0.800	1.200	0.300	0.800	1.200
	n=50	0.934	0.947	0.951	0.911	0.939	0.959
	n=100	0.935	0.946	0.948	0.919	0.935	0.933
Wei(2)							
Pois(0.4)	Coverage	Wald-2			Likelihood		
	t=	0.300	0.800	1.200	0.300	0.800	1.200
	n=50	0.928	0.944	0.935	0.939	0.952	0.949
	n=100	0.938	0.944	0.929	0.947	0.945	0.952
	Coverage	Bayes			Wald-1		
	t=	0.010	0.500	1.000	0.010	0.500	1.000
	n=50	0.928	0.949	0.955	0.891	0.933	0.941
	n=100	0.939	0.951	0.946	0.885	0.957	0.957
Wei(.5)							
Pois(0.4)	Coverage	Wald-2			Likelihood		
	t=	0.010	0.500	1.000	0.010	0.500	1.000
	n=50	0.910	0.944	0.943	0.929	0.943	0.948
	n=100	0.913	0.951	0.941	0.932	0.955	0.948

From the results of Table 4.3, we see that "Wald-1" does poorly in most cases, but the other methods perform reasonably well, with the Likelihood Ratio and "Bayes" method giving the best agreement with 0.95.

CHAPTER 5

CONFIDENCE INTERVALS FOR QUANTILES OF THE SURVIVAL FUNCTION

In clinical trials and survival analysis, confidence intervals for quantiles of the survival function $S(t)$ are often of interest to summarize the survival in the group of subjects; confidence intervals for the median survival time are of particular importance. For right-censored data Hutson (2001), and Strawderman et al (1997) developed non-parametric bootstrap-type confidence intervals for these quantiles. Most current methods are based in some way on the Kaplan-Meier or the Nelson-Aalen estimator (section 2.2). For interval-censored data, very little work has been done. Doss and Huffer (2003), by inverting the pointwise confidence intervals for $S(t)$, developed the Bayesian confidence intervals in equation (1.4), which apply in both the right-censored and interval-censored case.

5.1 Right Censoring

First, we review several types of bootstrap based confidence intervals for quantiles of S. We will then compare these methods with our Bayesian confidence intervals by simulation.

5.1.1 Strawderman, Parzen and Wells method

Strawderman, Parzen and Wells (1997), using the theoretical results in Strawderman and Wells (1997), suggested non-parametric confidence intervals based on the

Nelson-Aalen estimator $\hat{\Lambda}(t)$ in (2.7). Define

$$\hat{\Lambda}_{SW}(\alpha;t) = \hat{\Lambda}(t) - \hat{\sigma}(t)\left(z_{1-\alpha} + \left[\frac{\hat{\sigma}(t)}{4} - \frac{\hat{\rho}(t)}{3}\right] z_{1-\alpha}^2 - \left[\frac{\hat{\sigma}(t)}{4} + \frac{\hat{\rho}(t)}{6}\right]\right) \tag{5.1}$$

where $\hat{\sigma}(t)$ is defined in (2.8), z_α is the α percentile from a standard normal distribution, and

$$\hat{\rho}(t) = \frac{1}{\hat{\sigma}^3(t)} \sum_{t_i \le t} \frac{d_i}{Y_i^3}\,. \tag{5.2}$$

For any fixed time point t, Strawderman and Wells (1997) showed that $\hat{\Lambda}_{SW}(\alpha;t)$ supplies an asymptotically valid upper α confidence limit for $\Lambda(t)$, that is,

$$P(\Lambda(t) \le \hat{\Lambda}_{SW}(\alpha;t)) \approx \alpha \quad \text{for large } n.$$

Thus an (approximate) $1-\alpha$ confidence interval for $\Lambda(t)$ is given by

$$(C(t), D(t)) \text{ where } C(t) = \hat{\Lambda}_{SW}(\alpha/2;t) \text{ and } D(t) = \hat{\Lambda}_{SW}(1-\alpha/2;t). \tag{5.3}$$

Strawderman, Parzen and Wells (1997) proposed test-based intervals for quantiles based on inverting the interval in (5.3). This leads to a $1-\alpha$ confidence interval for $\Lambda^{-1}(y)$, for any $y > 0$, which is

$$\{\, t \,:\, y \in (C(t), D(t))\,\} = (D^{-1}(y), C^{-1}(y))\,. \tag{5.4}$$

Since $S(t) = e^{-\Lambda(t)}$, we see that the p-th quantile $Q(p) \equiv S^{-1}(p)$ satisfies $Q(p) = \Lambda^{-1}(-\log p)$ so that a $1-\alpha$ confidence interval for $Q(p)$ is obtained by taking $y = -\log p$ in (5.4). In particular, for the median survival time $Q(1/2)$ we take $y = \log 2$. These intervals were designated I_2 by Strawderman, Parzen and Wells (1997), and performed best among the various types of intervals they considered; we refer to them as SPW in our tables and the following discussion.

5.1.2 Thomas and Grunkemeier method

Barber and Jennison (1999) considered confidence intervals for quantiles $Q(p) = S^{-1}(p)$ based on the Thomas & Grunkemeier (1975) log-likelihood ratio test statistic $R(p,t)$ in equation (3.5); those values of t not rejected by the test of $H_0 : S(t) = p$

form the confidence interval:

$$\{t \,:\, -2R(p,t) \le \chi^2_{1,\alpha}\}\,. \tag{5.5}$$

5.1.3 Hutson method

Hutson (2001) developed an easily implemented method to calculate confidence intervals for quantiles of the survival function for right censored data. We briefly describe his approach.

Let $X_1, X_2, \ldots, X_n$ be a sample with distribution function F, survival function $S(x) = 1 - F(x)$, quantile function $Q(u) = F^{-1}(u)$, $0 < u < 1$, and order statistics $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$. Let $n' = n + 1$. The fractional uniform order statistics $\{U_{n'u:n} \,:\, 0 \le u \le 1\}$ are a Dirichlet process with $U_{n'u:n} \sim \text{Beta}(n'u, n'(1-u))$ and $X_{n'u:n} = Q(U_{n'u:n})$ is called the fractional order statistic. The cdf of $X_{n'u:n}$ is $F_{X_{n'u:n}}(x) = F_{U_{n'u:n}}(F(x))$ which may be estimated by

$$\hat{F}_{X_{n'u:n}}(x) = F_{U_{n'u:n}}(\hat{F}(x))$$

where $\hat{F} = 1 - \hat{S}$ and $\hat{S}$ is the Kaplan-Meier estimator. The α-quantile of $X_{n'u:n}$ may be estimated by

$$\hat{Q}_{X_{n'u:n}}(\alpha) = \inf\{x \,:\, \hat{F}_{X_{n'u:n}}(x) > \alpha\}$$

and Hutson takes his $100 \times (1-\alpha)\%$ two-sided confidence interval for the uth quantile $Q(u)$ to be

$$(\hat{Q}_{X_{n'u:n}}(\alpha/2), \hat{Q}_{X_{n'u:n}}(1-\alpha/2))\,.$$

5.1.4 Bayesian confidence intervals for quantiles of survival function

By the Bayesian approach, we obtain probability limits for the quantiles $S^{-1}(p)$ directly from the corresponding limits for $S(t)$ since

$$P(S^{-1}(p) \le t \mid \texttt{data}) = \; P(S(t) \le p \mid \texttt{data}), \text{ for all } p\,, \tag{5.6}$$

so that, if $P(S(t) \le L(t) \,|\, \texttt{data}) = \alpha/2$ for all t, then the value of t which satisfies $L(t) = p$ will be the $\alpha/2$ point of the posterior distribution of $S^{-1}(p)$. Similarly, if

$P(S(t) \leq U(t) \,|\, \texttt{data}) = 1 - \alpha/2$ for all t, then the $1 - \alpha/2$ point of $S^{-1}(p)$ is the solution t of $U(t) = p$.

The R program *ritcen* (Sec. 3.2) computes $L(t)$ and $U(t)$ at a grid of time points which can be specified by the user. If this grid is sufficiently fine, we can approximate the functions between these times by simple linear interpolation to solve the equations $U(t) = p$ and $L(t) = p$.

5.1.5 Simulation results

Here we do a simulation comparison of all the above methods for constructing confidence intervals for quantiles in the special case of 95% intervals for the Median ($p = 1/2$). We use "Bayes", "HUT", "SPW", and "TG" to denote the Bayesian (Sec. 5.1.4), Hutson (Sec. 5.1.3), Strawderman, Parzen and Wells (Sec. 5.1.1), and Thomas and Grunkemeier (Sec. 5.1.2) methods, respectively. The lifetime distribution is taken to be standard exponential (Exp(1)), or the Weibull distribution with shape=2 or shape=0.5 (Wei(2) or Wei(0.5)) and scale=1. The censoring distribution is either Uniform(0,t) with t=2, 5, or 10, or is taken to be the same as the lifetime distribution. Setting the lifetime and censoring distributions equal results in 50% censoring.

Each simulation includes 10,000 replications with sample sizes n=50 and 100. Each confidence interval method involves solving an equation (or inverting a test) and default or fall-back confidence limits are used in situations where no solution exists. For example, in the Bayesian approach it occasionally happens when the sample size n is small that no solution exists for $L(t) = p$ (or $U(t) = p$). In these cases the confidence limit is set to the smallest (or largest) uncensored observation. In each simulation, the number of times the default limits were used is given (in parentheses) for each method.

From the coverage simulation results of Table 5.1, we can see all methods except the Bayesian ("Bayes") and "SPW" methods give poor coverage values in the high censoring percentage cases (43%, 48%, 50%), especially in 48% and 50% cases. For relatively low censoring, all methods have reasonable coverage results.

5.2 Interval Censoring

As the right censoring case, the approach outlined in equations (1.3) to (1.5) along with (Sec. 5.1.4) also works for the interval censored data. The R program *gibbs1*

Table 5.1: Simulation results of coverage probabilities for Median with right censored data.

Cov(def)	Method	Exp(1)	Wei(2)	Wei(.5)	Exp(1)	Wei(2)	Wei(.5)
		Exp(1)	Wei(2)	Wei(.5)	Cen=Unif(0,2)		
		Cen 50%			Cen 43%	Cen 48%	Cen 41%
n=50	Bayes	0.952(0)	0.944(0)	0.945(0)	0.944(0)	0.951(0)	0.944(0)
	HUT	0.889(117)	0.890(93)	0.890(96)	0.918(263)	0.894(6)	0.934(1434)
	SPW	0.952(51)	0.952(39)	0.948(44)	0.944(62)	0.953(60)	0.949(304)
	TG	0.882(3)	0.894(0)	0.888(0)	0.903(0)	0.886(0)	0.920(0)
n=100	Bayes	0.951(0)	0.948(0)	0.951(0)	0.951(0)	0.952(0)	0.946(0)
	HUT	0.887(2)	0.893(0)	0.888(1)	0.923(3)	0.894(0)	0.939(169)
	SPW	0.945(1)	0.956(0)	0.952(0)	0.954(0)	0.954(0)	0.947(41)
	TG	0.911(0)	0.903(0)	0.921(0)	0.910(0)	0.901(0)	0.938(0)
Cov(def)	Method	Exp(1)	Wei(2)	Wei(.5)	Exp(1)	Wei(2)	Wei(.5)
		Cen=Unif(0,5)			Cen=Unif(0,10)		
		Cen 20%	Cen 18%	Cen 26%	Cen 10%	Cen 9%	Cen 17%
n=50	Bayes	0.959(0)	0.946(0)	0.956(0)	0.945(0)	0.944(0)	0.953(0)
	HUT	0.937(0)	0.931(0)	0.944(5)	0.943(0)	0.935(0)	0.949(0)
	SPW	0.955(0)	0.968(0)	0.949(44)	0.957(0)	0.946(0)	0.950(0)
	TG	0.914(0)	0.923(0)	0.911(0)	0.943(0)	0.947(0)	0.941(0)
n=100	Bayes	0.945(0)	0.949(0)	0.955(0)	0.951(0)	0.952(0)	0.949(0)
	HUT	0.935(0)	0.934(0)	0.943(0)	0.947(0)	0.941(0)	0.942(0)
	SPW	0.951(0)	0.958(0)	0.941(0)	0.955(0)	0.957(0)	0.957(0)
	TG	0.941(0)	0.941(0)	0.945(0)	0.949(0)	0.952(0)	0.953(0)

Table 5.2: Simulation results of coverage probabilities for Median with Interval censored data.

Coverage	Exp(1)	Wei(2)	Wei(0.5)
	Poisson (0.4)		
n=50	0.924	0.918	0.936
n=100	0.935	0.941	0.935
n=200	0.939	0.941	0.944

(Sec. 3.2) computes $L(t)$ and $U(t)$ at a grid of sufficiently fine time points so that we can approximate the functions between these times by simple linear interpolation to solve the equations $U(t) = p$ and $L(t) = p$.

Unlike the situation of right censoring case , for interval-censored data we are aware of no published work on confidence intervals for quantiles of the survival function, so there is no comparison in this case and our work will be the only one.

Then we conduct a simulation of confidence intervals for quantiles in the special case of 95% intervals for the Median ($p = 1/2$) with interval censored data. We simulate data as same as in Section 4.2. Each individual is inspected at times $V_{i1}, V_{i2}, \ldots, V_{i10}$ forming the first 10 arrivals of a Poisson process with a mean time of 0.4 between arrivals and still using a Monte Carlo sample size of 5,000. The comparison sample sizes are $n = 50$, $n = 100$ and $n = 200$ three cases along with three lifetime distributions: Exponential with mean 1, and Weibull with shape parameters $\alpha = 2$ or $\alpha = 0.5$ (scale parameter = 1). The empirical coverage probabilities (based on 10,000 simulation replications) are displayed in Table 5.2.

From Table 5.2, we can see that, with relatively large sample size, for instance $n = 200$, the result shows the great performance of the coverage probabilities. (We may need more information about the interval censored data, so the sample size usually larger than in the right censoring case.)

From the results of both right and interval censoring cases, we conclude that the Bayesian approach using a noninformative prior does indeed produce probability intervals with good frequentist properties and can be used as an alternative to the other available methods.

CHAPTER 6

SIMULTANEOUS CONFIDENCE BANDS FOR THE SURVIVAL FUNCTION

A great deal of research on simultaneous confidence bands for the survival function has been developed by Gillespie & Fisher (1979), Hall & Wellner (1980) and Nair (1981, 1984). Csorgo & Horvath (1986) introduced a class of bands showing the relationship between most of these proposals, and suggested several modified versions of these confidence bands. Most of these confidence bands are for right censored data and based on the Kaplan-Meier estimator (Sec. 2.2.1).

With the idea described in equations (1.9) to (1.11), we may extend the pointwise confidence interval approach in Doss and Huffer (2003) to obtain simultaneous confidence bands for the survival function. We will conduct simulation studies for the comparison of this Bayesian approach with other existing simultaneous confidence bands. Unlike other published approaches, our Bayesian method will be useful for both right and interval censored data. And we figure out there is no published work on simultaneous confidence bands for the interval censoring case, so our Bayesian approach will be the first.

6.1 Right Censoring

First of all, we will review several well-known simultaneous confidence bands and then explain the details of our Bayesian method.

6.1.1 General confidence bands for survival function

The well-known simultaneous confidence bands for the survival function are the Hall-Wellner (HW) band (Hall & Wellner, 1980) and the equal precision (EP) band of Nair (1984). Both of these bands are based on the Kaplan-Meier estimator.

Using the notation of Sec. 3.1.1, we consider simultaneous confidence bands for the survival function $S(t)$ on an interval $[a, b]$. In our description of the HW and EP bands, we follow Borgan and Liestohl (1990). Let

$$\hat{c}_a = n\hat{\sigma}^2(a)/[1 + n\hat{\sigma}^2(a)], \quad \text{and} \quad \hat{c}_b = n\hat{\sigma}^2(b)/[1 + n\hat{\sigma}^2(b)]. \tag{6.1}$$

The $100(1 - \alpha)\%$ HW confidence band is:

$$\hat{S}(t) \pm n^{-1/2} e_\alpha(\hat{c}_a, \hat{c}_b)[1 + n\hat{\sigma}^2(t)]\hat{S}(t) \quad \text{for } t \in [a, b] \tag{6.2}$$

where $e_\alpha(c_a, c_b)$ is the upper α percentile in the distribution of

$$\sup_{c_a \le x \le c_b} |W(x)| \tag{6.3}$$

in which W is the standard Brownian bridge [Koziol (1975)]. Nair's EP band is based on the pointwise confidence interval of equation (3.1) and is given as:

$$\hat{S}(t) \pm d_\alpha(\hat{c}_a, \hat{c}_b)\hat{\sigma}(t)\hat{S}(t) \quad \text{for } t \in [a, b] \tag{6.4}$$

where $d_\alpha(c_a, c_b)$ the upper α percentile in the distribution of

$$\sup_{c_a \le x \le c_b} \left| W(x)[x(1 - x)]^{-1/2} \right|. \tag{6.5}$$

6.1.2 Bayesian confidence bands for survival function

As illustrated by the HW and EP bands, the usual approach in the frequentist setting for obtaining simultaneous confidence bands for the survival function for times $t \in [a, b]$ is to choose a function $\hat{G}(t)$ and determine a critical value c_α such that

$$P_S\left(\sup_{t \in [a,b]} \left| \frac{S(t) - \hat{S}(t)}{\hat{G}(t)} \right| \le c_\alpha \right) \approx 1 - \alpha \tag{6.6}$$

which gives the confidence band

$$P_S(S(t) \in (\hat{S}(t) - c_\alpha \hat{G}(t), \hat{S}(t) + c_\alpha \hat{G}(t)), \forall t \in [a, b]) \approx 1 - \alpha, \tag{6.7}$$

where $\hat{S}(t)$ denotes an estimate (usually the MLE) of $S(t)$ and $\hat{G}(t)$ is often some function of an estimate $\hat{\sigma}^2(t)$ of the variance of $\hat{S}(t)$. This approach can be easily adapted to the Bayesian setting, simply replacing $\hat{S}(t)$ and $\hat{\sigma}(t)$ by the posterior mean $\mu(t)$ and posterior standard deviation $\sigma(t)$ of the survival function $S(t)$.

Our idea is that, choosing $G(t)$ to be some appropriate function of $\sigma(t)$, one can use simulation to find a critical value c_α such that

$$P\left(\sup_{t\in[a,b]} \left|\frac{S(t) - \mu(t)}{G(t)}\right| \leq c_\alpha \,\middle|\, \texttt{data}\right) = 1 - \alpha \tag{6.8}$$

which leads to the confidence band

$$P(S(t) \in (\mu(t) - c_\alpha G(t), \mu(t) + c_\alpha G(t)), \forall t \in [a, b] \,|\, \texttt{data}) = 1 - \alpha\,. \tag{6.9}$$

With a non-informative prior, we expect that this band will empirically also satisfy

$$P_S(S(t) \in (\mu(t) - c_\alpha G(t), \mu(t) + c_\alpha G(t)), \forall t \in [a, b]) \approx 1 - \alpha \tag{6.10}$$

in the frequentist setting. In our work, we simply take $G(t) = \sigma(t)$.

In order to implement the band in (6.9), we need to calculate $\mu(t)$ and $\sigma(t) = G(t)$ for (in principle) all $t \in [a, b]$, and then obtain the critical value c_α in (6.8). The posterior mean $\mu(t)$ and standard deviation $\sigma(t)$ are calculated by the Monte Carlo methods in Doss and Huffer (2003). The value c_α is computed by the obvious approach: we simulate many realizations of $S(\cdot)$ from the posterior, and take c_α to be the upper α quantile of the empirical distribution of

$$\sup_{t\in[a,b]} \left|\frac{S(t) - \mu(t)}{\sigma(t)}\right|. \tag{6.11}$$

To make the calculations feasible, we replace the interval $[a, b]$ in (6.11) by a sufficiently fine finite grid of values $\{t_1, t_2, \ldots, t_k\} \subset [a, b]$ with $t_1 < t_2 < \cdots < t_k$. Then we need only simulate realizations of $(S(t_1), S(t_2), \ldots, S(t_k))$. Since $S(t) =$

$1 - F(t)$ and

$$\mathcal{L}(F \,|\, X_1, \ldots, X_n, \theta) = \mathcal{D}_\beta \quad \text{where } \beta = \alpha_\theta + \sum_{i=1}^{n} \delta_{X_i}\,, \tag{6.12}$$

this may be done by using the well-known representation of the Dirichlet distribution in terms of ratios of gamma random variables: generate independent gamma random variables

$$g_i \sim \text{Gamma}(\alpha = \beta(t_{i-1}, t_i], 1)\,, \quad i = 1, 2, \ldots, k+1, \tag{6.13}$$

where $t_0 = 0$ and $t_{k+1} = \infty$, and define

$$S(t_i) = (g_{i+1} + \cdots + g_{k+1}) / \sum_{i=1}^{k+1} g_i\,, \quad i = 1, 2, \ldots, k. \tag{6.14}$$

The resulting random variables $(S(t_1), \ldots, S(t_k))$ have the required joint distribution.

Computing the bands is done in two Monte Carlo stages. The first stage computes $\mu(t_j)$ and $\sigma(t_j)$ for $j = 1, \ldots, k$. In the second stage, we generate a large number of realizations of $(X_1, \ldots, X_n, \theta)$ from the posterior, and for each of them compute $(S(t_1), \ldots, S(t_k))$ (by the method described above) and then

$$\sup_{1 \le j \le k} \left| \frac{S(t_j) - \mu(t_j)}{\sigma(t_j)} \right|. \tag{6.15}$$

The upper α quantile of these values gives us c_α, which is used to calculate the band in (6.9). Implementing this approach required fairly extensive modifications to the programs *ritcen* and *gibbs1*, which did not originally compute realizations of $(S(t_1), \ldots, S(t_k))$.

A simulation was conducted to conducted to determine a reasonable size k for the grid of times $\{t_1, \ldots, t_k\}$ used in computing c_α. As in Sec. 4.1, we simulated 1,000 right censored data sets with Exp(1) lifetimes and Uniform(0,2) censoring times for each combination of the sample sizes $n = 50, 100$ and grid sizes $k = 20, 40, 100$. For each data set, the critical value c_α (for $\alpha = .05$) was computed using k grid points equally spaced between 0 and 2. Using this c_α, we constructed the 95% simultaneous confidence band for $t \in [0, 2]$ and tested whether it contained the true survival curve $S(\cdot)$ by checking $S(t) \in (\mu(t) - c_\alpha\sigma(t), \mu(t) + c_\alpha\sigma(t))$ at 100 equally spaced times t between 0 and 2. (These 100 points are called the test grid and 0 should be excluded.)

Table 6.1: Coverage probabilities of Bayesian simultaneous confidence bands, varying the grid size k and sample size n. Data is right censored with Exp(1) lifetimes and Unif(0,2) censoring times, and coverage probabilities are computed using 100 test points in $[0, 2]$.

$k = 20$		$k = 40$		$k = 100$	
n=50	n=100	n=50	n=100	n=50	n=100
0.883	0.899	0.932	0.943	0.941	0.953

Table 6.1 displays the coverage probabilities obtained in this simulation. The grid size $k = 100$ clearly gives the most accurate coverage probabilities and $k = 20$ does rather poorly. We use $k = 100$ in the relatively large comparison simulation described later.

6.1.3 Simulation studies of confidence bands for right censoring

In this section we conduct simulations to compare the performance of the Bayesian simultaneous confidence bands with that of the HW and EP bands. These various bands are illustrated in Figure 6.1 where they are computed for simulated data with Exp(1) lifetimes, Uniform(0,2) censoring times, and sample size $n = 50$ (as in Sec.4.1) and times $t \in [0, 2]$. For reference, Figure 6.1 also includes the Bayesian pointwise confidence intervals, and both the posterior mean (the center of the Bayesian band) and the Kaplan-Meier estimator (the center of the HW and EP bands). The program we computed HW and EP bands is in the R-package of "km.ci" [Strobl (2009)]. As expected, the posterior mean essentially coincides with the Kaplan-Meier estimator, and all three simultaneous bands are considerably wider than the Bayesian pointwise confidence intervals for most time points in $[0, 2]$.

Our simulation results are summarized in Table 6.2 which reports the empirical coverage probabilities for nominal 95% simultaneous confidence bands. Each coverage probability is based on 10,000 simulated right censored data sets with:

- sample sizes $n = 50$ or $n = 100$;
- lifetime distributions: Exponential with mean 1, or Weibull with shape parameters $\alpha = 2$ or $\alpha = 0.5$ (and scale parameter = 1);

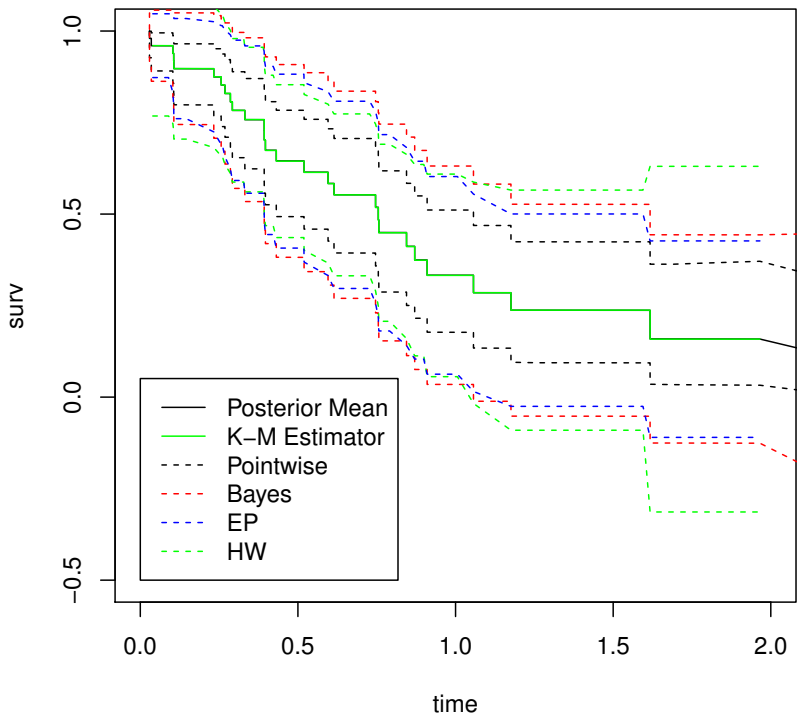

Figure 6.1: Comparison of the Bayesian, HW, and EP simultaneous confidence Bands on $[0, 2]$ with the pointwise Bayesian confidence intervals, with the posterior mean and the Kaplan-Meier estimator included for reference. The data is right-censored with Exp(1) lifetimes, Uniform(0,2) censoring times, and sample size $n = 50$.

- censoring distribution: Uniform(0,2).

The time interval $[a, b]$ for the simultaneous bands was chosen so that a and b are approximately equal to the 0.15 and 0.85 quantiles of the true lifetime distribution unless this led to a value of b exceeding 2, in which case we took $b = 2$. For constructing the Bayesian confidence bands, we took the prior to be $(a, b, M) = (.1, .1, .1)$ and used 5,000 realizations of $(X_1, \ldots, X_n, \theta)$ from the posterior (generated using the algorithm *ritcen*) for estimating $\mu(t)$ and $\sigma(t)$ and another 5,000 for estimating c_α; the computation of c_α used a grid of $k = 100$ equally spaced time points in $[a, b]$. In computing the coverage probabilities for all three bands (denoted Bayes, HW, and EP), we used a test grid of either 100 or 200 equally spaced time points in $[a, b]$; the

Table 6.2: Comparing coverage probabilities of Bayesian simultaneous 95% confidence bands with those of the HW and EP bands using right censored data with Uniform(0,2) censoring times and Exp(1), Weibull(2), or Weibull(.5) lifetimes.

Test grid of 100 points						
	Exp(1)		Weibull(2)		Weibull(.5)	
	$[a,b] = [.15, 1.9]$		$[a,b] = [.4, 1.5]$		$[a,b] = [.03, 2]$	
	n=50	n=100	n=50	n=100	n=50	n=100
Bayes	0.9455	0.9529	0.9425	0.9522	0.9461	0.9501
HW	0.9442	0.9572	0.9511	0.9620	0.9447	0.9558
EP	0.9266	0.9521	0.9232	0.9492	0.9338	0.9514
Test grid of 200 points						
	Exp(1)		Weibull(2)		Weibull(.5)	
	$[a,b] = [.15, 1.9]$		$[a,b] = [.4, 1.5]$		$[a,b] = [.03, 2]$	
	n=50	n=100	n=50	n=100	n=50	n=100
Bayes	0.9441	0.9510	0.9414	0.9516	0.9444	0.9498
HW	0.9411	0.9543	0.9495	0.9599	0.9423	0.9544
EP	0.9242	0.9499	0.9221	0.9476	0.9325	0.9508

true survival curve was considered "covered" if it was inside the band at these time points.

From the results of Table 6.2, we see that the coverage probabilities of the Bayesian bands have good agreement with the nominal value of 0.95. The EP bands have poor coverage performance especially for the smaller sample size $n = 50$. Varying the number of test grid points from 100 to 200 made little difference.

6.2 Interval Censoring

In this section we conduct simulations to study the performance of our Bayesian simultaneous confidence band for interval-censored data. Unlike the HW and EP bands, the Bayesian bands described in Section 6.1.2 work for both right-censored and interval-censored data, with the only difference being that for sampling from the posterior we must use *gibbs1* instead of *ritcen* (see Sec. 3.2) when we have interval-censored data. There is no published work on simultaneous confidence bands for interval-censored data, so our work will be the first.

As in Sec.6.1.3, we have constructed Figure 6.2 to compare the Bayesian simultaneous 95% confidence band and the Bayesian pointwise 95% confidence intervals for interval-censored data. The data is simulated with sample size $n = 50$ and Exp(1) lifetimes with the interval censoring determined by inspection times consisting of the first 10 arrivals of a Poisson process with mean time 0.4 between arrivals as in Section 4.2. As expected, the simultaneous band is considerably wider than the pointwise confidence intervals over most of the time interval.

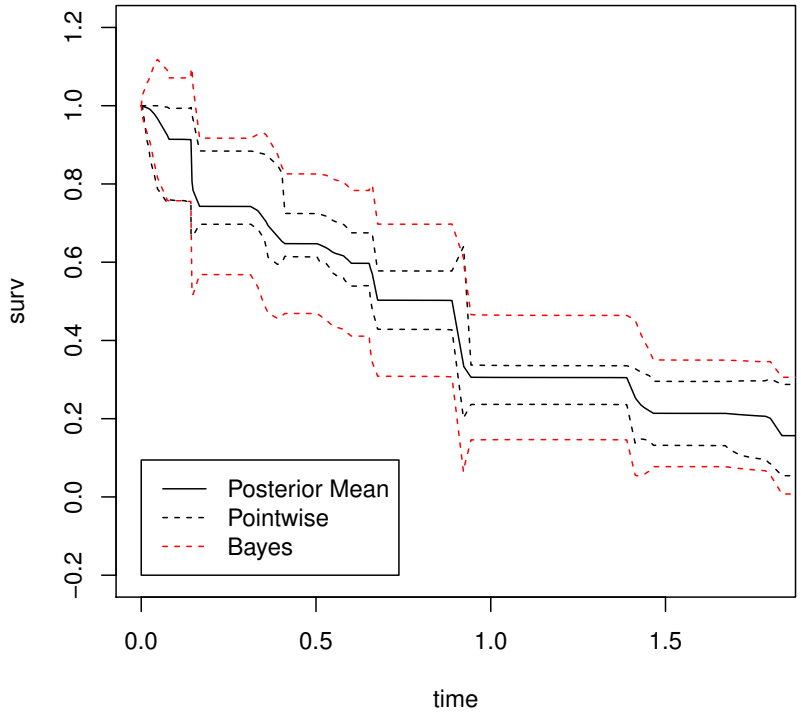

Figure 6.2: The comparison of Bayesian simultaneous confidence bands with pointwise confidence interval.

The results of our simulation are displayed in Table 6.3. The simulation setup is identical to that in Section 6.1.3 except that: (a) the Uniform(0,2) right censoring is replaced by interval censoring determined by inspection times consisting of the first 10 arrivals of a Poisson process with mean time 0.4 between arrivals, (b) the range of sample sizes is expanded to include $n = 200$ in addition to $n = 50$ and 100, and (c) the Monte Carlo sampling is carried out using *gibbs1* instead of *ritcen*.

Table 6.3: Coverage probabilities performance of Bayesian simultaneous 95% confidence bands with interval censored data.

	Exp(1)		Weibull(2)		Weibull(.5)	
	$[a, b] = [.15, 1.9]$		$[a, b] = [.4, 1.5]$		$[a, b] = [.03, 2]$	
	Test 100	Test 200	Test 100	Test 200	Test 100	Test 200
$n = 50$	0.9314	0.891	0.9341	0.9216	0.9333	0.9317
$n = 100$	0.9422	0.9409	0.9423	0.9409	0.9482	0.9470
$n = 200$	0.9552	0.9541	0.9540	0.9528	0.9501	0.9489

From Table 6.3, we see that the Bayesian confidence bands perform well with both 100 and 200 test grid points when the sample size n is sufficiently large. It appears that, to achieve good performance, larger sample sizes are needed with interval censored data than with right censored data. This was also observed in Section 5.1.5.

CHAPTER 7

EDGEWORTH APPROXIMATION APPROACH FOR THE POINTWISE CONFIDENCE INTERVALS

7.1 Direct Edgeworth Expansion Approximation Method

The approach to obtaining the pointwise confidence intervals for the survival function $S(t)$ described in Chapters 1 and 3 computes the quantiles $L(t)$ and $U(t)$ in (1.3) by using the Newton-Raphson algorithm to find the exact quantiles of the mixture of Beta distributions in (3.16). This method will be called the "Exact Method" in this chapter, since it is exact except for the Monte Carlo error. This approach is usually carried out only for a relatively small number of time-points t, and to use it at very many time-points (say, forming a fine grid on the time axis) would require prohibitively large amounts of time and memory.

In this section, we describe an approach that approximates the quantiles based on the posterior moments of $S(t)$, which can be easily computed at a large number of time-points t using very little memory because the computations require only accumulating various totals. Using this approach will speed up the program.

In particular, in situations where the posterior distribution of $S(t)$ is approximately normally distributed (which we expect if the sample size n is sufficiently large and the posterior mean $\mu(t)$ is not too close to 0 or 1), we can approximate the

distribution with a direct Edgeworth expansion [Barndorff-Nielsen and Cox (1989)]

$$P\left(\frac{S(t)-\mu(t)}{\sigma(t)} \le x \;\middle|\; \texttt{data}\right) \approx \Psi(x) \equiv \Phi(x) - \varphi(x)\left\{\frac{\rho_3}{6}H_2(x) + \frac{\rho_4}{24}H_3(x) + \frac{\rho_3^2}{72}H_5(x)\right\} \tag{7.1}$$

where $\mu(t)$, $\sigma(t)$, $\rho_3 = \rho_3(t)$, and $\rho_4 = \rho_4(t)$ are the mean, standard deviation, and standardized skewness and kurtosis of the posterior distribution of $S(t)$. The standardized skewness and kurtosis for a random variable X with mean μ are given by

$$\rho_3 = \text{skew}(X) = \frac{E(X-\mu)^3}{\sigma^3} = \frac{E(X^3) - 3\mu E(X^2) + 2\mu^3}{\sigma^3} \tag{7.2}$$

and

$$\rho_4 = \text{kurt}(X) = \frac{E(X-\mu)^4}{\sigma^4} - 3 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4} - 3\,. \tag{7.3}$$

The other quantities in (7.1) are Φ and φ, the cdf and pdf of the standard normal distribution; and H_2, H_3, and H_5, which are Hermite polynomials given by

$$H_2(x) = x^2 - 1,\; H_3(x) = x^3 - 3x,\; H_5(x) = x^5 - 10x^3 + 15x\,. \tag{7.4}$$

Solving $\Psi(x) = 1 - \alpha/2$ and $\Psi(x) = \alpha/2$ will give easily computed approximations to $U(t)$ and $L(t)$ which are the limits of our confidence interval.

From equation (7.2) and (7.3), if one wants to compute the Edgeworth expansion in (7.1), the first four moments of the posterior distribution of $S(t)$ are required. The obvious way to estimate these moments is to generate a large number of realizations of $S(t)$ from the posterior and then use the empirical moments as estimates. The Monte Carlo sampling carried out by *ritcen* or *gibbs1* generates realizations of $(X_1, \ldots, X_n, \theta)$ from the posterior. These may be used to generate realizations of $S(t)$ since

$$\mathcal{L}(S(t) \mid X_1, X_2, \ldots, X_n, \theta) \;\sim\; \text{Beta}\,(a(t), b(t)) \tag{7.5}$$

where

$$a(t) \;=\; Me^{-\theta t} + \sum_{i=1}^{n} I(X_i > t)\,, \quad b(t) = M + n - a(t) \tag{7.6}$$

as in Section 3.2. But this approach leads to poor accuracy, and generating realizations of $S(t)$ is computationally wasteful.

A better approach directly makes use of the moments of the Beta distribution in (7.5) which are

$$E\left(S(t)^k \,|\, X_1, \ldots, X_n, \theta\right) = \frac{\Gamma(a(t)+K)\cdot\Gamma(a(t)+b(t))}{\Gamma(a(t))\cdot\Gamma(a(t)+b(t)+K)} = \frac{\langle a(t)\rangle_K}{\langle M+n\rangle_K} \tag{7.7}$$

where $\langle\xi\rangle_K = \xi(\xi+1)(\xi+2)\cdots(\xi+K-1)$. This leads to

$$E\left(S(t)^k \,|\, \texttt{data}\right) = \frac{E\left(\langle a(t)\rangle_K \,|\, \texttt{data}\right)}{\langle M+n\rangle_K} \tag{7.8}$$

so that the desired posterior moments of $S(t)$ may be estimated by Monte Carlo averages of $\langle a(t)\rangle_K$, $K = 1, 2, 3, 4$.

Even more accurate estimates of the posterior moments may be obtained by conditioning on the cluster structure (see below). The general technique of achieving variance reduction by conditioning is often called Rao-Blackwellizing. Conditioning on the cluster structure was used by Doss & Huffer (2003) to estimate the first two posterior moments of $S(t)$. They demonstrated that this led to substantial improvements in accuracy over the simpler approach based on (7.8). We now extend their arguments to obtain estimates of the third and fourth moments.

Recall our basic model as described in Section 3.2. For a specified distribution ν on Θ and scalar $M > 0$, we have $\theta \sim \nu$, and given θ, $F \sim \mathcal{D}_{\alpha_\theta}$ where $\alpha_\theta = MH_\theta$. (In all our work, we take H_θ to be an exponential distribution with density $h_\theta(x) = \theta e^{-\theta x}$.) Then conditional on F, the values $X_1, \ldots, X_n$ are iid from F. Our `data` consists of the knowledge that $X_i \in A_i$ for $i = 1, \ldots, n$. Since F is discrete, there is a positive probability of ties among the values $X_1, \ldots, X_n$. We introduce the cluster structure $C = (C_1, C_2, \ldots, C_n)$ to describe the structure of these ties (see Section 2.4 of Doss & Huffer (2003)). Let $Y_1, Y_2, \ldots, Y_{\#(C)}$ be the distinct values among $X_1, \ldots, X_n$, and $C_1, \ldots, C_n$ be integer labels assigned to the clusters of tied values among $X_1, \ldots, X_n$ such that $C_k = i$ if and only if $X_k = Y_i$. Those X_k with $C_k = i$ all take the value Y_i; they form the i-th cluster.

Conditional on $(\texttt{data}, C)$, we must have $Y_i \in B_i$ where we define

$$B_i = \bigcap_{k:C_k=i} A_k$$

for $i = 1, 2, \ldots, \#(C)$. It is well known that, conditional on (θ, C), the distinct

values $Y_1, Y_2, \ldots, Y_{\#(C)}$ are iid from H_θ. Thus, conditional on $(\texttt{data}, \theta, C)$, the values $Y_1, \ldots, Y_{\#(C)}$ are independent with

$$Y_i \sim (H_\theta)_{B_i} \quad \text{for } i = 1, 2, \ldots, \#(C)\,,$$

where $(H_\theta)_{B_i}$ denotes the distribution H_θ restricted to the set B_i and then re-normalized to be a probability measure, i.e., it is the probability distribution of an exponential random variable conditional on its belonging to B_i.

The simplicity of the conditional distribution of $Y_1, \ldots, Y_{\#(C)}$ makes it possible for us to compute the conditional moments of $a(t)$ in (7.6). This in turn allows us to compute the posterior moments of $S(t)$ by using (7.8) and the identity

$$E[\langle a(t)\rangle_K \,|\, \texttt{data}] = E[E(\langle a(t)\rangle_K \,|\, \texttt{data}, \theta, C) \,|\, \texttt{data}]\,. \tag{7.9}$$

We need only keep Monte Carlo averages of $E(\langle a(t)\rangle_K \,|\, \texttt{data}, \theta, C)$ for $K = 1, 2, 3, 4$.

Now our goal becomes the calculation of $E(\langle a(t)\rangle_K | \texttt{data}, \theta, C)$. We condition on $\{\texttt{data}, \theta, C\}$ throughout the following argument, and for convenience use $E(\cdot)$ and $P(\cdot)$ to denote conditional expectations and probabilities given $\{\texttt{data}, \theta, C\}$, and treat all quantities which are fixed under this conditioning as constants.

From (7.6), letting $m = \#(C)$, we have

$$a(t) \;=\; Me^{-\theta t} + \sum_{i=1}^{m} n_j I(Y_j > t)\,, \tag{7.10}$$

where n_j is the size of the jth cluster. Let $w_j = I(Y_j > t)$ and $p_j = P(Y_j > t)$. Then $w_j \sim \text{Bernoulli}(p_j)$ and $w_1, \ldots, w_m$ are conditionally independent given $\{\texttt{data}, \theta, C\}$. Since $Y_j \sim (\text{Exp}(\theta))_{B_j}$, the values p_j are easily obtained. For data sets with right censoring, the set B_j is either a singleton (i.e., $B_j = \{c_j\}$ for some value c_j), or a half-line (i.e., $B_j = (c_j, \infty)$ for some c_j). When $B_j = \{c_j\}$, we have $Y_j = c_j$ so that $p_j = 0$ or 1. When $B_j = (c_j, \infty)$,

$$\begin{aligned} &\text{if } t < c_j, \text{ then } p_j = 1; \\ &\text{if } t > c_j, \text{ then } p_j = \frac{e^{(-\theta t)}}{e^{(-\theta c_j)}}. \end{aligned}$$

For data sets with interval censoring, the sets B_j are intervals: $B_j = (c_j, d_j)$ where $d_j = \infty$ in some cases. This leads to

$$\begin{aligned}
&\text{if } t < c_j, \text{ then } p_j = 1;\\
&\text{if } t > d_j, \text{ then } p_j = 0;\\
&\text{if } c_j < t < d_j, \text{ then } p_j = \frac{e^{(-\theta t)} - e^{(-\theta d_j)}}{e^{(-\theta c_j)} - e^{(-\theta d_j)}}.
\end{aligned}$$

To simplify our expressions, we define $A = a(t)$, $s = Me^{-\theta t}$, and

$$B = \sum_{j=1}^{m} n_j w_j \,.$$

Then (7.10) becomes

$$A = s + B$$

and our goal is to compute the factorial moments $E(\langle A \rangle_K)$, $K = 1, 2, 3, 4$. Since

$$\begin{aligned}
&E\,A\\
&E\,A(A+1) = E\,A^2 + E\,A\\
&E\,A(A+1)(A+2) = E\,A^3 + 3E\,A^2 + 2E\,A\\
&E\,A(A+1)(A+2)(A+3) = E\,A^4 + 6E\,A^3 + 11E\,A^2 + 6E\,A\,,
\end{aligned}$$

it suffices to compute $E\,A^k$, $k = 1, \ldots, 4$. Furthermore, since $A = s + B$, we have

$$\begin{aligned}
E\,A &= s + E\,B\\
E\,A^2 &= s^2 + 2sE\,B + E\,B^2\\
E\,A^3 &= s^3 + 3s^2E\,B + 3sE\,B^2 + E\,B^3\\
E\,A^4 &= s^4 + 4s^3E\,B + 6s^2E\,B^2 + 4sE\,B^3 + E\,B^4\,,
\end{aligned} \tag{7.11}$$

reducing our problem to finding $E\,B^k$, $k = 1, 2, 3, 4$.

In particular, using properties of the Bernoulli distribution, EB and EB^2 are easily seen to be

$$\begin{aligned} E\,B &= \sum_{i=1}^{m} p_i n_i \\ E\,B^2 &= \sum_{i=1}^{m} p_i (n_i)^2 + \left(\sum_{i=1}^{m} p_i n_i \right)^2 - \sum_{i=1}^{m} (p_i)^2 (n_i)^2 \,. \end{aligned} \tag{7.12}$$

We calculate EB^3 and EB^4 using the moment generating function (mgf) of B given by

$$\begin{aligned} M_B(x) &= E\,e^{xB} \\ &= E\,e^{x(\sum_j n_j w_j)} \\ &= \prod_j E\,e^{x n_j w_j} \\ &= \prod_{j=1}^{m} (1 - p_j + p_j e^{n_j x}) \,. \end{aligned}$$

To compute the moments of B, we need the derivatives of $M_B(x)$. For any set $A \subset \{1, 2, \ldots, m\}$, define

$$M_A(x) = \prod_{j \notin A} (1 - p_j + p_j e^{n_j x})$$

so that, for example, $M_{\{i\}}(x) = \prod_{j \neq i} (1 - p_j + p_j e^{n_j x})$. Then we have

$$\begin{aligned} M_B'(x) &= \sum_{i=1}^{m} p_i n_i e^{n_i x} \prod_{j \neq i} (1 - p_j + p_j e^{n_j x}) \\ &= \sum_{i=1}^{m} p_i n_i e^{n_i x} M_{\{i\}}(x) \\ M_B''(x) &= \sum_{i=1}^{m} p_i {n_i}^2 e^{n_i x} M_{\{i\}}(x) + \sum_{i=1}^{m} p_i n_i e^{n_i x} \left(\sum_{j \neq i} p_j n_j e^{n_j x} M_{\{i,j\}}(x) \right) \\ &= \sum_{i=1}^{m} p_i {n_i}^2 e^{n_i x} M_{\{i\}}(x) + \sum_{(i,j), j \neq i} p_i p_j n_i n_j e^{n_i x} e^{n_j x} M_{\{i,j\}}(x) \,. \end{aligned}$$

Similarly

$$M_B'''(x) = \sum_{i=1}^{m} p_i {n_i}^3 e^{n_i x} M_{\{i\}}(x) + 3 \sum_{(i,j),j\neq i} p_i p_j {n_i}^2 n_j e^{(n_i+n_j)x} M_{\{i,j\}}(x) + \sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k n_i n_j n_k e^{(n_i+n_j+n_k)x} M_{\{i,j,k\}}(x) \tag{7.13}$$

and

$$\begin{aligned} {M_B}^{(4)}(x) = & \sum_{i=1}^{m} p_i {n_i}^4 e^{n_i x} M_{\{i\}}(x) + \sum_{i=1}^{m} p_i {n_i}^3 e^{n_i x} \left(\sum_{j\neq i} p_j n_j e^{n_j x} M_{\{i,j\}}(x) \right) \\ & + 3 \sum_{(i,j),j\neq i} p_i p_j {n_i}^2 n_j (n_i + n_j) e^{(n_i+n_j)x} M_{\{i,j\}}(x) \\ & + 3 \sum_{(i,j),j\neq i} p_i p_j {n_i}^2 n_j e^{(n_i+n_j)x} \left(\sum_{k\neq i \mathrm{or} j} p_k n_k e^{n_k x} M_{\{i,j,k\}}(x) \right) \\ & + \sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k n_i n_j n_k (n_i + n_j + n_k) e^{(n_i+n_j+n_k)x} M_{\{i,j,k\}}(x) \\ & + \sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k n_i n_j n_k e^{(n_i+n_j+n_k)x} \left(\sum_{l\neq i,j \mathrm{or} k} p_l n_l e^{n_l x} M_{\{i,j,k,l\}}(x) \right) . \end{aligned} \tag{7.14}$$

Since $E\, B^{(k)} = {M_B}^{(k)}(0)$, using (7.13) we have

$$E\, B^3 = \sum_{i=1}^{m} p_i {n_i}^3 + 3 \sum_{(i,j),j\neq i} p_i p_j {n_i}^2 n_j + \sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k n_i n_j n_k \,, \tag{7.15}$$

in which

$$\begin{aligned} \sum_{(i,j),j\neq i} p_i p_j {n_i}^2 n_j &= \sum_{(i,j)} p_i p_j {n_i}^2 n_j - \sum_{i=1}^{m} {p_i}^2 {n_i}^3 \\ &= \left(\sum_{i=1}^{m} p_i {n_i}^2 \right) \left(\sum_{j=1}^{m} p_j n_j \right) - \sum_{i=1}^{m} {p_i}^2 {n_i}^3 \,, \end{aligned} \tag{7.16}$$

and

$$\begin{aligned}\sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k n_i n_j n_k &= \left(\sum_{j=1}^{m} p_j n_j\right)^3 - 3\sum_{(i,j),j\neq i} p_i {p_j}^2 n_i {n_j}^2 - \sum_{i=1}^{m} {p_i}^3 {n_i}^3 \\ &= \left(\sum_{j=1}^{m} p_j n_j\right)^3 - \sum_{i=1}^{m} {p_i}^3 {n_i}^3 \\ &\quad - 3\left[\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)\left(\sum_{j=1}^{m} p_j n_j\right) - \sum_{i=1}^{m} {p_i}^3 {n_i}^3\right]. \end{aligned} \tag{7.17}$$

Then from equation (7.15) to (7.17),

$$\begin{aligned} E\,B^3 &= \sum_{i=1}^{m} p_i {n_i}^3 + 3\left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right) - 3\sum_{i=1}^{m} {p_i}^2 {n_i}^3 \\ &\quad + \left(\sum_{i=1}^{m} p_i n_i\right)^3 - 3\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right) + 2\sum_{i=1}^{m} {p_i}^3 {n_i}^3 . \end{aligned} \tag{7.18}$$

Similarly, using equation (7.14) we have

$$\begin{aligned} E\,B^4 &= \sum_{i=1}^{m} p_i {n_i}^3 + \sum_{(i,j),j\neq i} p_i p_j {n_i}^3 n_j + 3\sum_{(i,j),j\neq i} p_i p_j {n_i}^2 n_j (n_i + n_j) \\ &\quad + 3\sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k {n_i}^2 n_j n_k + \sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k n_i n_j n_k (n_i + n_j + n_k) \\ &\quad + \sum_{(i,j,k,l),i\neq j\neq k\neq l} p_i p_j p_k n_i n_j n_k n_l \end{aligned} \tag{7.19}$$

which can be simplified as

$$\begin{aligned} E\,B^4 &= \sum_{i=1}^{m} p_i {n_i}^3 + 4\sum_{(i,j),j\neq i} p_i p_j {n_i}^3 n_j + 3\sum_{(i,j),j\neq i} p_i p_j {n_i}^2 {n_j}^2 \\ &\quad + 6\sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k {n_i}^2 n_j n_k + \sum_{(i,j,k,l),i\neq j\neq k\neq l} p_i p_j p_k n_i n_j n_k n_l . \end{aligned} \tag{7.20}$$

Then, using the techniques of (7.16) to (7.17), we have

$$\begin{aligned}\sum_{(i,j),j\neq i} p_i p_j {n_i}^3 n_j &= \sum_{(i,j)} p_i p_j {n_i}^3 n_j - \sum_{i=1}^{m} {p_i}^2 {n_i}^4 \\ &= \left(\sum_{i=1}^{m} p_i {n_i}^3\right)\left(\sum_{j=1}^{m} p_j n_j\right) - \sum_{i=1}^{m} {p_i}^2 {n_i}^4 , \end{aligned} \tag{7.21}$$

$$\begin{aligned}\sum_{(i,j),j\neq i} p_i p_j {n_i}^2 {n_j}^2 &= \sum_{(i,j)} p_i p_j {n_i}^2 {n_j}^2 - \sum_{i=1}^{m} {p_i}^2 {n_i}^4 \\ &= \left(\sum_{i=1}^{m} p_i {n_i}^2\right)^2 - \sum_{i=1}^{m} {p_i}^2 {n_i}^4 , \end{aligned} \tag{7.22}$$

and

$$\begin{aligned}\sum_{(i,j,k),i\neq j\neq k} p_i p_j p_k {n_i}^2 n_j n_k &= \left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{j=1}^{m} p_j n_j\right)^2 - 2\sum_{(i,j),j\neq i} {p_i}^2 p_j {n_i}^3 n_j \\ &\quad - \sum_{(i,j),j\neq i} p_i {p_j}^2 {n_i}^2 {n_j}^2 - \sum_{i=1}^{m} {p_i}^3 {n_i}^4 \\ &= \left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{j=1}^{m} p_j n_j\right)^2 - 2\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^3\right)\left(\sum_{j=1}^{m} p_j n_j\right) \\ &\quad + 2\sum_{i=1}^{m} {p_i}^3 {n_i}^4 - \left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{j=1}^{m} {p_j}^2 {n_j}^2\right) \\ &\quad + \sum_{i=1}^{m} {p_i}^3 {n_i}^4 - \sum_{i=1}^{m} {p_i}^3 {n_i}^4 \\ &= \left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{j=1}^{m} p_j n_j\right)^2 - 2\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^3\right)\left(\sum_{j=1}^{m} p_j n_j\right) \\ &\quad + 2\sum_{i=1}^{m} {p_i}^3 {n_i}^4 - \left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{j=1}^{m} {p_j}^2 {n_j}^2\right) . \end{aligned} \tag{7.23}$$

The last part of equation (7.20) will be

$$\sum_{(i,j,k,l),i\neq j\neq k\neq l} p_i p_j p_k n_i n_j n_k n_l = \left(\sum_{j=1}^{m} p_j n_j\right)^4 - 6 \sum_{(i,j,k,l),i=j\neq k\neq l} {p_i}^2 {n_i}^2 p_k n_k p_l n_l - \sum_{i=1}^{m} {p_i}^4 {n_i}^4 - 3 \sum_{(i,j,k,l),i=j\neq k=l} {p_i}^2 {n_i}^2 {p_k}^2 {n_k}^2 - 4 \sum_{(i,j,k,l),i=j=k\neq l} {p_i}^3 {n_i}^3 p_l n_l \,. \tag{7.24}$$

In fact

$$\begin{aligned}\sum_{(i,k,l),i\neq k\neq l} {p_i}^2 {n_i}^2 p_k n_k p_l n_l &= \left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right)^2 - 2 \sum_{(i,k),i\neq k} {p_i}^3 p_k {n_i}^3 n_k \\ &\quad - \sum_{(i,l),i\neq l} {p_i}^2 {p_l}^2 {n_i}^2 {n_l}^2 - \sum_{i=1}^{m} {p_i}^4 {n_i}^4 \\ &= \left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right)^2 + 2\sum_{i=1}^{m} {p_i}^4 {n_i}^4 \\ &\quad - 2\left(\sum_{i=1}^{m} {p_i}^3 {n_i}^3\right)\left(\sum_{i=1}^{m} p_i n_i\right) - \left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)^2 ,\end{aligned} \tag{7.25}$$

and

$$\sum_{(i,j,k,l),i=j\neq k\neq l} {p_i}^2 {n_i}^2 p_k n_k p_l n_l = \left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)^2 - \sum_{i=1}^{m} {p_i}^4 {n_i}^4 \tag{7.26}$$

$$\sum_{(i,j,k,l),i=j=k\neq l} {p_i}^3 {n_i}^3 p_l n_l = \left(\sum_{i=1}^{m} {p_i}^3 {n_i}^3\right)\left(\sum_{i=1}^{m} p_i n_i\right) - \sum_{i=1}^{m} {p_i}^4 {n_i}^4 \,. \tag{7.27}$$

Then

$$\begin{aligned}\sum_{(i,j,k,l),i\neq j\neq k\neq l} p_i p_j p_k n_i n_j n_k n_l &= \left(\sum_{j=1}^{m} p_j n_j\right)^4 - 6\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right)^2 \\ &\quad + 8\left(\sum_{i=1}^{m} {p_i}^3 {n_i}^3\right)\left(\sum_{i=1}^{m} p_i n_i\right) + 3\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)^2 - 6\sum_{i=1}^{m} {p_i}^4 {n_i}^4 \,.\end{aligned} \tag{7.28}$$

So, by equation (7.19) to (7.28), we have

$$
\begin{aligned}
E\,B^4 \;=& \sum_{i=1}^{m} p_i {n_i}^4 + 4\left(\sum_{i=1}^{m} p_i {n_i}^3\right)\left(\sum_{i=1}^{m} p_i n_i\right) + 3\left(\sum_{i=1}^{m} p_i {n_i}^2\right)^2 \\
& - 7\sum_{i=1}^{m} {p_i}^2 {n_i}^4 + 6\left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right)^2 - 12\left(\sum_{i=1}^{m} {p_i}^3 {n_i}^3\right)\left(\sum_{i=1}^{m} p_i n_i\right) \\
& + 12\sum_{i=1}^{m} {p_i}^3 {n_i}^4 - 6\left(\sum_{i=1}^{m} p_i {n_i}^2\right)\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right) + \left(\sum_{i=1}^{m} p_i n_i\right)^4 \\
& - 6\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)\left(\sum_{i=1}^{m} p_i n_i\right)^2 + 8\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^3\right)\left(\sum_{i=1}^{m} p_i n_i\right) \\
& + 3\left(\sum_{i=1}^{m} {p_i}^2 {n_i}^2\right)^2 - 6\sum_{i=1}^{m} {p_i}^4 {n_i}^4 \,. \qquad (7.29)
\end{aligned}
$$

The formulas for EB, EB^2, EB^3, EB^4 in equations (7.12), (7.18), (7.29) and for EA, EA^2, EA^3, EA^4 in equation (7.11) have been incorporated into the Fortran programs implementing the samplers *ritcen* and *gibbs1*, allowing us to compute the posterior moments of $S(t)$ via (7.8) and (7.9).

7.2 Simulation Studies For Edgeworth Approximation Approach

In this section, we conduct simulations comparing the exact method with the Edgeworth approximation approach for computing pointwise confidence intervals for the survival function [Chapter 4]. This is done for both right and interval censored data.

For convenience of comparison, the simulation settings in this chapter are identical to those in Sections 4.1 and 4.2 except that we now use 20,000 Monte Carlo realizations from the posterior instead of 5,000. In each simulation setting, we generate $N = 10{,}000$ data sets and compute confidence intervals $(L(t), U(t))$ for $S(t)$ at three different times t. In our tables, we report the empirical coverage probabilities for both approaches. In addition, to judge the accuracy of the Edgeworth approximation, we report the mean absolute difference between the endpoints $L(t)$ produced by the two

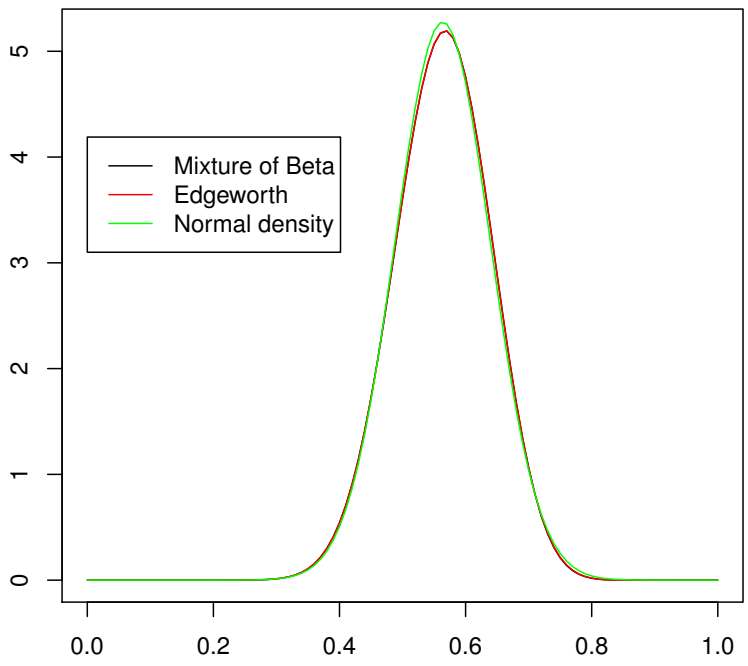

Figure 7.1: Comparing the exact posterior density of $S(t)$ with the Edgeworth approximation and a simple Normal approximation. The posterior is computed from right censored data with Exp(1) lifetimes and sample size $n = 50$, at time $t = 0.6$.

methods (and similarly for $U(t)$). This is defined by

$$\text{Mean Absolute Difference of } L = \frac{1}{N}\sum_{i=1}^{N} |L_i^{\text{Exact}} - L_i^{\text{Edgeworth}}|$$

$$\text{Mean Absolute Difference of } U = \frac{1}{N}\sum_{i=1}^{N} |U_i^{\text{Exact}} - U_i^{\text{Edgeworth}}| \,.$$

In our tables, these two values are reported as a parenthesized pair in units of 10^{-4}.

Differentiating the Edgeworth approximation to the posterior cdf of $S(t)$ given in (7.1) leads to an approximation to the density. A simple normal approximation to the distribution of $S(t)$ is given by $S(t) \approx N(\mu(t), \sigma(t))$. It is instructive to compare

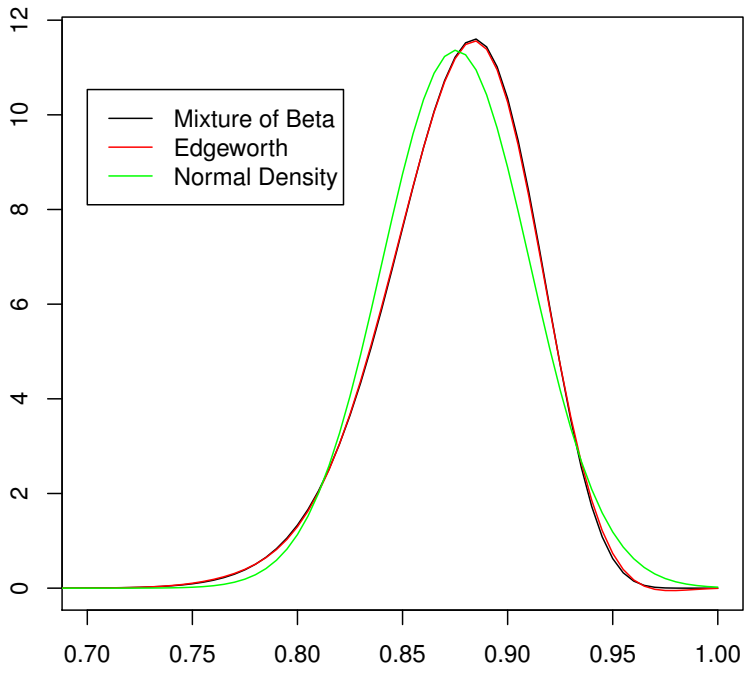

Figure 7.2: Comparing the exact posterior density of $S(t)$ with the Edgeworth approximation and a simple Normal approximation. The posterior is computed from right censored data with Weibull(2) lifetimes and sample size $n = 50$, at time $t = 0.3$.

the exact density of $S(t)$ given by the mixture of Beta distributions in (3.16) with the Edgeworth approximation and the simple normal approximation, and we shall do this before presenting the simulation results. We computed the posterior distribution of $S(t)$ for five data sets generated from different simulation settings used in this chapter, displaying the exact density and the two approximations in Figures 7.1 to 7.5, with the captions describing the data which was used.

Figures 7.1 and 7.2 use right censored data with sample size $n = 50$. Figure 7.1 shows a situation where all three curves are close, and the Edgeworth approximation is an almost prefect match with the exact density. (The curves are drawn in the order black, then red, then green, so that, if the curves match closely, one sees mainly green.) But in Figure 7.2, there is a minor difference between the exact density and

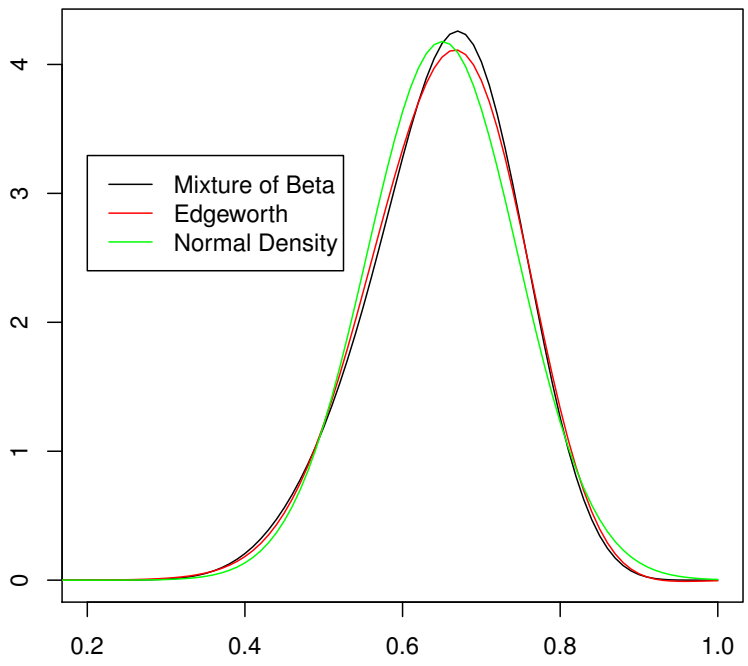

Figure 7.3: Comparing the exact posterior density of $S(t)$ with the Edgeworth approximation and a simple Normal approximation. The posterior is computed from interval censored data with Exp(1) lifetimes and sample size $n = 50$, at time $t = 0.6$.

the Edgeworth expansion which may affect the coverage probabilities a little bit in this simulation setting. This is consistent with the simulation results in Table 7.1. Also in Figure 7.2 we see a relatively big difference between the Normal density and the other two density lines which shows that the skewness and kurtosis corrections in equation (7.1) play an important role in this situation.

Figures 7.3 to 7.5 use interval censored data. In Figure 7.3 we see that, unlike the right censored case, with a small sample size ($n = 50$) the density of the Edgeworth expansion gives a relatively poor fit to the exact density, although it does improve upon the Normal approximation. Figures 7.4 and 7.5 both use sample size $n = 100$ (with 7.4 using the same simulation setting as Figure 7.3) and give much better agreement among the three curves. It seems reasonable to expect the three curves to agree

more closely as n increases since the posterior distribution of $S(t)$ is asymptotically normal.

From the simulation results in Tables 7.1 and 7.2, we can see, for both right and interval censored data, that the two methods give us quite close agreement of the coverage probabilities as desired, in some cases even exactly the same. The mean absolute differences between the $L(t)$ and $U(t)$ of the two methods are also quite small, indicating that the Edgeworth approximation to the posterior quantiles $L(t)$ and $U(t)$ is typically quite accurate. In our simulations, the Edgeworth approximation approach was at least three times faster than the Exact method. This difference in speed can be quite important in large simulations or when finding confidence intervals for $S(t)$ at many time points.

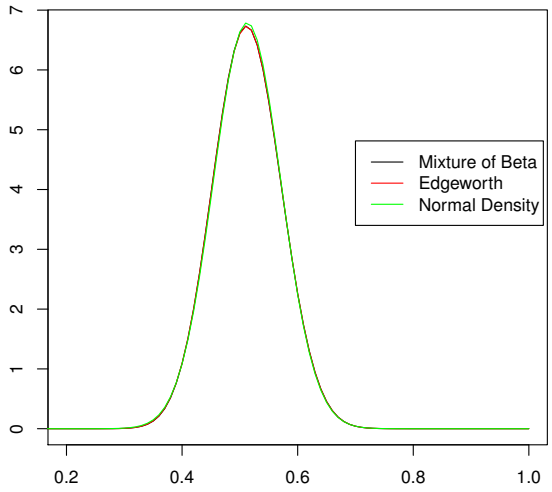

Figure 7.4: Similar to Figure 7.3 but with sample size $n = 100$

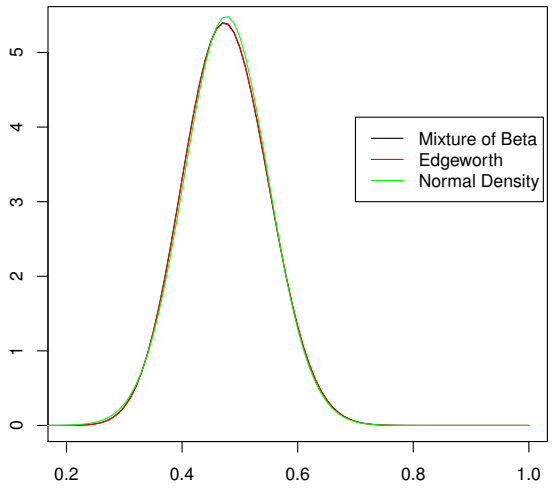

Figure 7.5: Comparing the exact posterior density of $S(t)$ with the Edgeworth approximation and a simple Normal approximation. The posterior is computed from interval censored data with Weibull(2) lifetimes and sample size $n = 100$, at time $t = 0.8$.

Table 7.1: Comparing the Exact method and Edgeworth Approximation approach for right-censored data.

Dist	Cov	Exact			Edgeworth			Mean Abs Diff ($\times 10^{-4}$)		
Exp(1)	t=	0.2000	0.6000	1.0000	0.2000	0.6000	1.0000	0.2000	0.6000	1.0000
Unif(0,2)	n=50	0.9397	0.9448	0.9451	0.9404	0.9445	0.9449	(5.1,3.1)*	(.78,4.3)	(7.1,7.5)
	n=100	0.9468	0.9467	0.9473	0.9469	0.9467	0.9474	(1.6,1.8)	(3.7,2.3)	(5.9,4.7)
	Cov	Exact			Edgeworth			Mean Abs Diff ($\times 10^{-4}$)		
Wei(2)	t=	0.3000	0.8000	1.2000	0.3000	0.8000	1.2000	0.3000	0.8000	1.2000
Unif(0,2)	n=50	0.9384	0.9413	0.9423	0.9393	0.9408	0.9418	(24,1.1)	(6.2,7.1)	(2,5.1)
	n=100	0.9406	0.9450	0.9463	0.9402	0.9455	0.9465	(4.9,4.1)	(5.2,3.5)	(7.7,9.3)
	Cov	Exact			Edgeworth			Mean Abs Diff ($\times 10^{-4}$)		
Wei(.5)	t=	0.0100	0.5000	1.0000	0.0100	0.5000	1.0000	0.0100	0.5000	1.0000
Unif(0,2)	n=50	0.9512	0.9460	0.9432	0.9515	0.9453	0.9424	(7.5,1.5)	(.4,.3)	(8.6,.5)
	n=100	0.9435	0.9484	0.9460	0.9430	0.9482	0.9464	(2.1,2.3)	(.2,.2)	(.4,.3)
*The actual number is (0.00051, 0.00031).										

Table 7.2: Comparing the Exact method and Edgeworth Approximation approach for interval-censored data.

Dist	Cov	Exact			Edgeworth			Mean Abs Diff ($\times 10^{-4}$)		
	t=	0.2000	0.6000	1.0000	0.2000	0.6000	1.0000	0.2000	0.6000	1.0000
Exp(1)	n=50	0.9280	0.9413	0.9431	0.9317	0.9398	0.9413	(7.5,12)*	(2.9,3.4)	(2.2,1.9)
Pois(.4)	n=100	0.9326	0.9418	0.9392	0.9401	0.9413	0.9411	(35,5.2)	(2.7,2.9)	(1.6,1.9)
	n=200	0.9356	0.9511	0.9404	0.9371	0.9502	0.9421	(3.3,1.4)	(6.7,3.1)	(1.2,2.2)
	Cov	Exact			Edgeworth			Mean Abs Diff ($\times 10^{-4}$)		
	t=	0.3000	0.8000	1.2000	0.3000	0.8000	1.2000	0.3000	0.8000	1.2000
Wei(2)	n=50	0.9338	0.9355	0.9370	0.9315	0.9352	0.9369	(1.5,1.5)	(4.8,6.2)	(5.4,6.1)
Pois(.4)	n=100	0.9372	0.9421	0.9380	0.9340	0.9430	0.9383	(6.8,8.2)	(3.8,3.7)	(3.9,6.1)
	n=200	0.9425	0.9433	0.9486	0.9401	0.9449	0.9483	(4.5,8.1)	(6.8,4.2)	(3.4,7)
	Cov	Exact			Edgeworth			Mean Abs Diff ($\times 10^{-4}$)		
	t=	0.0100	0.5000	1.0000	0.0100	0.5000	1.0000	0.0100	0.5000	1.0000
Wei(.5)	n=50	0.9350	0.9443	0.9371	0.9344	0.9446	0.9379	(1.5,.2)	(.05,.5)	(.7,.6)
Pois(.4)	n=100	0.9520	0.9434	0.9399	0.9524	0.9436	0.9398	(.2,1.8)	(.1,.06)	(.3,.6)
	n=200	0.9465	0.9421	0.9405	0.9458	0.9419	0.9413	(.2,1.1)	(.02,.1)	(.03,.07)
*The actual number is (0.00075, 0.0012).										

BIBLIOGRAPHY

[1] Allison, Paul D., 1995: Survival Analysis Using SAS. *Cary, NC: SAS Publishing.*

[2] Alioum, A., Commenges, D., 1996: A proportional hazards model for arbitrarily censored and truncated data *Biometrics*, **52**, 512–524.

[3] Antoniak, C., 1974, Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems. *The Annals of Statistics*, **2**, 1152–1174.

[4] Athreya, K. B., Doss, H., and Sethuraman, J., 1996: On the Convergence of the Markov Chain Simulation Method. *The Annals of Statistics*, **24**, 69–100.

[5] Barber, S., Jennison, C., 1999: Symmetric tests and confidence intervals for survival probabilities and quantiles of censored survival data. *Biometrics*, **55**, 430–436.

[6] Beadle, G., Come, S., Henderson, C., Silver, B., and Hellman, S., 1984a: The Effect of Adjuvant Chemotherapy on the Cosmetic Results After Primary Radiation Treatment for Early Stage Breast Cancer. *International Journal of Radiation Oncology, Biology and Physics*, **10**, 2131–2137.

[7] Beadle, G., Harris, J., Silver, B., Botnick, L., and Hellman, S., 1984b: Cosmetic Results Following Primary Radiation Therapy for Early Breast Cancer. *Cancer*, **54**, 2911–2918.

[8] Breslow, N. and J. Crowley, 1974: A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship. *The Annals of Statistics*, Vol.2, **3**, 437–453.

[9] Borgan, Ornulf and Knut Liestol, 1990: A Note on Confidence Intervals and Bands for the Survival Function Based on Transformations. *Scandinavian Journal of Statistics*, Vol.17, **1**, 35–41.

[10] Cantor, Alan B., SAS Survival Analysis Techniques for Medical Research. *Cary, NC: SAS Publishing.*

[11] Cox, D.R. and O. E. Barndorff-Nielsen, 1989: Asymptotic Techniques for Use in Statistics. Chapman and Hall, London.

[12] Csorgo, S. and Horvath, L., 1986; Confidence bands from censored samples. *Canad. J. Statist.*, **14**, 131-144.

[13] Doss, H., 1994: Bayesian Nonparametric Estimation for Incomplete Data via Successive Substitution Sampling. *The Annals of Statistics*, **22**, 1763–1786.

[14] Doss, H., and Huffer, F., 2003: Monte Carlo Methods for Bayesian Analysis of Survival Data Using Mixtures of Dirichlet Process Priors. *Journal of Computational and Graphical Statistics*, Vol.12, **2**, 282–307.

[15] Efron, B., 1981: Censored data and the bootstrap. *Journal of the American Statistical Association*, **76**, 312–319.

[16] Escobar, M. D., 1988: Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means. Unpublished Ph.D. dissertation, Yale University, Department of Statistics.

[17] Escobar, M. D., 1994: Estimating Normal Means with a Dirichlet Process Prior. *Journal of the American Statistical Association*, **89**, 268–277.

[18] Escobar, M. D., and West, M., 1995: Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, **90**, 577–588.

[19] Fay, M. P., 1999: Comparing Several Score Tests for Interval-Censored Data. *Statistics in Medicine*, **18**, 273–285.

[20] Ferguson, T. S., 1973: A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, **1**, 209–230.

[21] Ghosal, S., Ghosh, J. K. and Ramamoorthi, R. V., 1999: Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics*, **27**, 143–158.

[22] Goodall, Dunn, and Babiker., 2004: Interval-censored survival time data: confidence intervals for the non-parametric survivor function. *Statistics in Medicine*, **23**, 1131–1145.

[23] Hall, W. J. and Jon A. Wellner, 1980: Confidence Bands for a Survival Curve from Censored Data. *Biometrika*, Vol.67, **1**, 133–143.

[24] Hanson, T. E., and Johnson,W. O., 2004: A Bayesian Semiparametric AFT Model for Interval Censored Data. *Journal of Computational and Graphical statistics* Vol.13, **2**, 341–361.

[25] Hollander, M., McKeague, W. and Yang, J., 1997; Likelihood Ratio-Based Confidence Bands for Survival Functions. *Journal of the American Statistical Association* Vol.92, **437**, 215–226.

[26] Hutson, A. D., 2001: Bootstrap-type confidence intervals for quantiles of the survival distribution. *Statistics in Medicine*, **20**, 1693–1702.

[27] Ibrahim, J. G., Chen, M.-H., and Sinha, D., 2001: Bayesian Survival Analysis. New York, Springer-Verlag.

[28] Klein, J., and Moeschberger, M., 1997: Survival Analysis, New York, Springer-Verlag.

[29] Kong, A., Liu, J. S., and Wong, W. H., 1994: Sequential Imputations and Bayesian Missing Data Problems. *Journal of the American Statistical Association*, **89**, 278–288.

[30] Koziol, J. A. and Byar, D. P., 1975: Percentage points of the asymptotic distribution of one and two samples K-S-statistics for truncated or censored data. *Technometrics*, **17**, 507–510.

[31] Kuo, L., and Smith, A. F. M., 1992: Bayesian Computations in Survival Models via the Gibbs Sampler. *Survival Analysis: State of the Art*, 11–24.

[32] Lindsey, Jane C. and Ryan, Louise M., 1998: Tutorial in Biostatistics Methods for Interval-Censored Data. *Statistics in Medicine*, **17**, 219–238.

[33] Liu, J. S., 1996: Nonparametric Hierarchical Bayes via Sequential Imputations. *The Annals of Statistics*, **24**, 911–930.

[34] MacEachern, S. N., 1994: Estimating Normal Means with a Conjugate Style Dirichlet Process Prior. *Communications in Statistics, Part BSimulation and Computation*, **23**, 727–741.

[35] MacEachern, S. N., Clyde, M., and Liu, J. S., 1999: Sequential Importance Sampling for Nonparametric Bayes Models: The Next Generation. *Canadian Journal of Statistics*, **27**, 251–267.

[36] Nair, R. M., 1984: Confidence Bands for Survival Functions with Censored Data: A Comparative Study. *Technometrics*, Vol.26, **3**, 265–275.

[37] Neal, R. M., 2000: Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, **9**, 249–265.

[38] Peto, R., 1973: Experimental Survival Curves for Interval-censored Data. *Journal of the Royal Statistical Society: series C (Applied Statistics)*, Vol.22, **1**, 86–91.

[39] Roeder, K., and Wasserman, L., 1997: Practical Bayesian Density Estimation Using Mixtures of Normals. *Journal of the American Statistical Association*, **92**, 894–902.

[40] Sethuraman, J., 1994: A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, **4**, 639–650.

[41] Sinha, D., and Dey, D., 1997: Semiparametric Bayesian Analysis of Survival Data. *Journal of the American Statistical Association*, **92**, 1195–1212.

[42] Strawderman, RL, Wells MT., 1997: Accurate bootstrap confidence limits for the cumulative hazard and survivor functions under random censoring. *Journal of the American Statistical Society*, **92**, 1356–1374.

[43] Strawderman, RL, Parzen MI, Wells MT., 1997: Accurate confidence limits for quantiles under random censoring. *Biometrics*, **53**, 1399–1415.

[44] Strobl, Ralf, 2009: http://cran.r-project.org/web/packages/km.ci/index.html, R-package "km.ci".

[45] Susarla,V.and Van Ryzin, J., 1976: Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations. *Journal of the American Statistical Association*, **71**, 897–902.

[46] Therneau, T., 2010: http://cran.r-project.org/web/packages/survival/index.html, R-package "Survival".

[47] Thomas, D. and Gary L. Grunkemeier, 1975: Confidence Interval Estimation of Survival Probabilities for Censored Data. *Journal of the American Statistical Association*, Vol.70, **352**, 865–871.

[48] Turnbull, B. W., 1974: Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *Journal of the American Statistical Association*, **69**, 169–173.

[49] Zhou, M., 2004: Nonparametric Bayes estimator of survival functions for doubly/interval censored data. *Statistica Sinica*, **14**, 533–546.

BIOGRAPHICAL SKETCH

Yingfeng Tao was born in a very beautiful city named Wuxi in China. He received a Master's degree in applied mathematics at University of Science and Technology of China. After that, he continued his graduate study in the United States. He obtained Master of Science in Mathematics at State University of New York at Buffalo and transferred to Florida State University to study statistics. Now he got his first PhD degree in Biostatistics in statistics department of FSU.