Nonparametric and Semiparametric Bayesian Reliability Analysis

Kaushik Ghosh[(1)], **Ram C. Tiwari**[(2)]

[(1)] Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, NJ 07102

[(2)] Statistical Reserach and Applications Branch
Division of Cancer Control and Population Sciences
National Cancer Institute
Bethesda, MD

CAMS Report 0607-18, Fall 2006/Spring 2007

Center for Applied Mathematics and Statistics

NJIT

Nonparametric and Semiparametric Bayesian Reliability Analysis

Kaushik Ghosh*
Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, NJ 07102, USA

Ram C. Tiwari
Statistical Research and Applications Branch
Division of Cancer Control and Population Sciences
National Cancer Institute
Bethesda, MD 20892, USA

*Contact person for proofs. Address: Department of Mathematical Sciences, New Jersey Institute of Technology, 323 Martin Luther King Jr. Blvd., Newark, NJ 07102, USA. Email:ghosh@njit.edu. Phone: +1-973-642-4496, Fax: +1-973-596-5591.

Abstract

In this article, we first provide an overview of some nonparametric Bayesian methods of inference using life-history data. These methods include those that use Dirichlet process, gamma process and beta process as the prior. We then present a semiparametric Bayesian method for estimating the reliability of a component in a multi-component system using lifetime data from several systems. This method assumes that the component lifetimes have a parametric distribution with a Dirichlet process prior on the distribution of the parameters. The semiparametric method is illustrated with a simulation study using a data augmentation procedure.

Keywords: Data augmentation, Dirichlet process prior, Gibbs sampler, k-out-of-n system, order statistics.

1 Introduction

Suppose T is a non-negative random variable denoting lifetime. We define the survival function at time t to be $S(t) = P(T > t) = 1 - F(t)$ where F is the cumulative distribution function. If T has density $f(t)$, the hazard rate function $r(t)$ at time t is given by $r(t) = \frac{f(t)}{S(t)}$. This gives the instantaneous failure rate at time t given survival just prior to time t. The corresponding cumulative hazard at time t is $R(t) = \int_0^t r(s)ds$. It is well known that $S(t) = \exp(-R(t))$. Hence, knowing any one of density, cdf, survival function, hazard rate function or cumulative hazard function is equivalent for inference about the properties of the life distribution. Sometimes, the phrase "reliability function" is used to denote the survival function of a component or a system of components connected in some structure.

In reliability and survival analysis, it is often of interest to estimate the reliability of the system/component from the observed lifetime data. Suppose n components are put on test and $T_1, \ldots, T_n$ are the corresponding lifetimes. If all $T_1, \ldots, T_n$ are actually observed, we have complete data. Usually, however, some of the T_i's are censored. Suppose T_i and C_i are the lifetime and (right) censoring times respectively of the ith component. One observes

$X_i = \min(T_i,\ C_i)$ and $\delta_i = I_{T_i \leq C_i}$. Based on observed data $(X_i,\ \delta_i),\ i = 1,\ \ldots,\ n$, one would like to infer about the lifetime distribution as a whole.

No matter whether one has complete or censored data, a host of techniques have been developed for their analysis. They can be broadly classified as parametric and nonparametric approaches. The advantages of parametric methods are their simplicity in the sense that only a handful of parameters can be used to explain the behavior. For example, Kvam and Samaniego (1993) used a parametric approach for analyzing the life-history data from an r-out-of-k system whereby the component lifetimes were assumed to be exponentially distributed. The unknown parameter was estimated using the maximum likelihood method. However, as is well known, parametric models impose certain structural restrictions and are less than optimal if one is unable to verify/justify the model assumptions. An alternative is to use nonparametric methods such as the ones used by Kvam and Samaniego (1993) or Chen (2003) in estimating component reliability.

In this article, we first review some nonparametric and semiparametric Bayesian methods available for analysis of life-history data. The advantage of having Bayesian methods is ability to incorporate prior information in the inferential procedure and semiparametric models allow for partial specification of model structure through parameters. We then present a semiparametric Bayesian approach to estimating the component lifetime distribution based on the system lifetime data along with additional information (such as censoring indicator and number of failed components) for multiple k-out-of-n systems (see Boyles and Samaniego 1987).

In Section 2, we briefly review past work on the nonparametric Bayesian analysis for survival data. In Section 3, we present a theoretical development of a semiparametric procedure. In Section 4, we briefly discuss the sampling procedure that will be used to implement the model. In Section 5, we present the results of a simulation study based on data from Chen (2003). Finally, in Section 6, we close with some concluding remarks.

2 Nonparametric Methods

In Bayesian analysis, it is very important to have as close to correct a likelihood function as possible. Often, the standard parametric models are not rich enough to capture the uncertainty in the observed data. In such situations, nonparametric Bayesian methods provide a more flexible alternative. Nonparametric Bayesian methods in reliability can be broadly classified into 3 groups, depending on the quantity where one places a prior distribution on: prior on the class of all distributions, prior on the class of all hazard rates and prior on the class of all cumulative hazards. In each case, the underlying distribution is assumed to be free of any parameters and the prior information is combined with life-history data to obtain the corresponding posterior. Due to the nonparametric nature, the prior distributions are taken to be stochastic processes. We provide a brief review of the three methods in the following subsections. For a comprehensive review of various nonparametric methods of estimating the survival function, see Ferguson et al. (1992). Also see Sinha and Dey (1997) and Singpurwalla (2006) for detailed discussion on nonparametric approaches to reliability estimation.

2.1 Prior on distributions

Dirichlet processes, introduced by Ferguson (1973) are one of the fundamental concepts in the nonparametric Bayesian analysis literature. A Dirichlet process can be used to provide a nonparametric prior for a distribution function.

Suppose F is a cumulative distribution function (cdf). We say that F has a Dirichlet Process prior i.e. $F \sim \mathcal{D}(M, F_0)$ if the following happens. For any partition $A_1 \cup A_2 \cup \cdots \cup A_k = \mathbb{R}$,

$$(F(A_1),\ F(A_2),\ \ldots,\ F(A_k)) \sim Dirichlet(MF_0(A_1), MF_0(A_2),\ \ldots,\ MF_0(A_k)),$$

where $Dirichlet(\alpha_1,\ \ldots,\ \alpha_k)$ denotes a Dirichlet distribution with parameters $(\alpha_1,\ \ldots,\ \alpha_k)$.

See Wilks (1962) for more on Dirichlet distribution. $M > 0$ is called the precision parameter and F_0 is called the baseline distribution. F_0 can be thought of as the "average value" of F and M as the amount of "concentration" of F around F_0. A high value of M signifies that F is very close to F_0 and a low value of M signifies large dispersion around F_0. However, see Sethuraman and Tiwari (1982) for another interpretation of small values of M.

Ferguson (1973) showed that if $\theta|F \sim F$ and $F \sim \mathcal{D}(M, F_0)$, then $F|\theta \sim \frac{1}{M+1}[MF_0 + \delta_\theta]$. Repeated application of this result shows that if $T_1, \ldots, T_n$ are observations generated by F, and F has a Dirichlet process prior $\mathcal{D}(M, F_0)$, the Bayes estimator of F is

$$\hat{F}(t) = \frac{M}{M+n}F_0(t) + \frac{1}{M+n}F_n(t)$$

where $F_n(t)$ is the empirical distribution function of the sample.

The above requires complete information about the lifetimes (i.e., no censoring). Susarla and VanRyzin (1976) developed the nonparametric Bayes estimator of the cdf in the presence of right censoring and assuming a Dirichlet process prior on the underlying distribution. Let $(X_i, \delta_i), i = 1, \ldots, n$ be the observed data with $X_i = \min(T_i, C_i)$ and $\delta_i = I_{T_i \leq C_i}$. Let $u_1 < \cdots < u_k$ be the distinct values among $X_1, \ldots, X_n$ and λ_j be the number of censored observations at u_j. Let $k(t)$ be the number of u_j's that are less than or equal to u_k and h_k be the number greater than u_k. Then, the Bayes estimator of the Survival function under squared-error loss is given by

$$\frac{M(1-F_0(t)) + h_k(t)}{M+n} \prod_{j=1}^{k(t)} \frac{M(1-F_0(u_j)) + h_j + \lambda_j}{M(1-F_0(u_j)) + h_j}.$$

This estimator reduces to the Kaplan-Meier estimator as $M \to 0$ and under no censoring, reduces to the estimator given earlier.

Dirichlet processes give rise to distributions as priors that are discrete with probability one. In addition, the corresponding Bayes estimator gets complicated for right-censored data. To avoid these difficulties, one may use neutral to the right (NTR) priors. A random

distribution function F on the real line is said to be NTR if for every m and $t_1 < t_2 < \cdots < t_k$, there exist independent random variables $V_1, \ldots, V_m$ such that $(1-F(t_1), \ldots, 1-F(t_m))$ has the same distribution as $(V_1, V_1V_2, \ldots, \prod_{i=1}^{m} V_i)$. Doksum (1974) and Ferguson and Phadia (1979) show that if $X_1, \ldots, X_n$ is a sample from F and F is NTR, then the posterior distribution of F given $X_1, \ldots, X_n$ is also NTR. It turns out that the censored case is simpler to treat than the uncensored case. The implementation of the results for practical applications gets cumbersome. Damien et al. (1995) and Walker and Damien (1998) have proposed simulation based approaches for a full Bayesian analysis involving NTR processes.

2.2 Prior on hazard functions

An alternative approach is to put a nonparametric prior on the class of all hazard functions. Suppose we partition the time axis into $(k+1)$ intervals $0 < t_1 < \cdots < t_k < \infty$. Let $p_1 = F(t_1)$, $p_i = F(t_i) - F(t_{i-1})$, $i = 2, \ldots, k$. Note that $F(0) = 0$ and $F(\infty) = 1$. Define $Z_1 = p_1$, $Z_i = p_i/(1 - p_1 - \cdots - p_{i-1})$ for $i = 1, 2, \ldots, k$. Note that Z_i is the failure rate over the interval $[t_{i-1}, t_i)$ and $(Z_1, \ldots, Z_k)$ gives the piecewise constant hazard function over $[0, t_k)$. Let n_i denote the number of failures in the interval $[t_{i-1}, t_i)$, $\boldsymbol{p} = (p_1, \ldots, p_k)$ and $\boldsymbol{d} = (n_1, \ldots, n_{k+1})$. The observed data is given by $\boldsymbol{d}$.

In such a scenario, one may assume that the prior distribution of the Z_i's is independent Beta(ν_{1i}, ν_{2i}). This results in a generalized Dirichlet prior for $\boldsymbol{p}$ given by the probability function

$$f(p_1, \ldots, p_k) = \prod_{i=1}^{k} \frac{\Gamma(\nu_{1i} + \nu_{2i})}{\Gamma(\nu_{1i})\Gamma(\nu_{2i})} p_i^{\nu_{1i}-1}(1 - p_1 - \cdots - p_k)_i^{\gamma},$$

where $\gamma_i = \nu_{2i} - \nu_{1,i+1} - \nu_{2,i+1}$ for $i = 1, \ldots, k-1$ and $\gamma_k = \nu_{2k} - 1$. See Basu and Tiwari (1982) for the special case when the generalized Dirichlet prior for $\boldsymbol{p}$ reduces to the Dirichlet prior. Given the observed count data $\boldsymbol{d}$, the posterior of the piecewise hazards turns out also to be generalized Dirichlet. Lochner (1975) and Tiwari and Rao (1983) provide further discussion on use of this approach to estimate F. This method has several disadvantages.

First, it uses count data, leading to loss of information by not using the actual failure times. Second, the inferential conclusions depend on whether the cells are combined at the prior or at the posterior stage. In addition, one needs an excessive number of parameters to specify the generalized Dirichlet distribution. See also Wilks (1962) for more on generalized Dirichlet distributions.

The above procedure cannot account for any structural pattern in the hazard rate function. Padgett and Wei (1981) and Arjas and Gasbarra (1994) take the prior on the hazard rate function to be a Poisson process with constant jump size. Mazzuchi and Singpurwalla (1985) take the prior on hazard rate function to be ordered Dirichlet. The former ensures that the hazard rate function is non-decreasing and the latter ensures it to be monotone.

Dykstra and Laud (1981) introduce an extended Gamma process and use it to model the prior distribution of a non-decreasing hazard rate process. They show that when $t_1, \ldots, t_n$ are the right censoring times of n observations and $\Gamma(a(s), \beta(s))$ is the prior on the hazard rate process, the posterior on the hazard rate process is also an extended Gamma process $\Gamma(a(s), \hat{\beta}(s))$ where

$$\hat{\beta}(s) = \frac{\beta(s)}{1 + \beta(s) \sum_{i=1}^{n} (t_i - s)^+}$$

with $a^+ = \max(a, 0)$. When one observes the actual failure times $t_1, \ldots, t_n$, the resulting posterior hazard rate is a mixture of extended gamma processes, which is complicated to calculate. Laud et al. (1996) approximate this posterior by approximating its random independent increments via a Gibbs sampler. Once the posterior hazard rate process is obtained, the corresponding survival function is given by

$$P(T \geq t) = \exp\left[- \int_0^t \log(1 + \hat{\beta}(s)(t - s)) da(s) \right].$$

Other methods of specifying prior on hazard rates include use of Markov Beta and Markov Gamma processes considered by Nieto-Barajas and Walker (2002). See also Tiwari and Rao (1983), Tiwari and Jammalamadaka (1985) and Tiwari and Kumar (1989).

2.3 Prior on cumulative hazard

An alternative to putting prior on the hazard rate function is to put prior on the cumulative hazard function. This is particularly attractive especially when there is no density, since the cumulative hazard still exists in such situations. Kalbfleisch (1978) proposed using a gamma process as a prior for the cumulative hazard function $R(t)$. Take any partition $0 \equiv t_0 < t_1 < t_2 < \cdots < t_{k-1} < t_k \equiv \infty$, of the time points and define $r_i = \log(1 - Z_i)$ where $Z_i = P(T \in [t_{i-1},\ t_i) | T \geq t_{i-1})$ is the hazard rate over the interval $[t_{i-1},\ t_i)$. Assume that r_i's are independent gamma random variables with shape $\alpha_i - \alpha_{i-1}$ and scale c, where $\alpha_i = cR^*(t_i)$ and $R^*(t)$ is interpreted as the best guess of $R(t)$. c is the measure of strength of conviction about the guess and large values indicate a strong conviction.

Given n failure times $\tau_1,\ \ldots,\ \tau_n$, the posterior cumulative hazard will be a process with independent increments. Kalbfleisch (1978) has shown that the posterior cumulative hazard has increment at τ_i which is given by a density $A(c + A_i,\ c + A_{i+1})$ at u, where $A_i = n - i$ and $A(a,\ b)$ is of the form

$$\frac{\exp(-bu) - \exp(au)}{u \log(a/b)}$$

Between τ_{i-i} and τ_i, the increments are prescribed by a gamma process with shape function $cR^*(\cdot)$ and scale $c + A_i$. The survival function is recovered either by simulation or by approximation using expected value of the process.

The above formulation suffers from some difficulties. First, there is a lack of intuition regarding the assumption of gamma distribution on r_i. Second, the independent increments property of $R(t)$ may not be meaningful, since under aging and wear, the successive Z_i's would be judged to be increasing. Finally, presence of ties in the failure time data presents problems in the model fitting.

A different model was proposed by Hjort (1990) to model the randomness in the cumulative hazard rate. Suppose R has a beta process prior with parameters $c(\cdot)$ and $R_0(\cdot)$. The

posterior of R given life-history data $(X_1,\ \delta_1),\ \ldots,\ (X_n,\ \delta_n)$ is also a beta process of the form

$$R|\text{data} \sim Beta[c(\cdot) + Y(\cdot),\ \int_0^{(\cdot)} \frac{cdR_0 + dN}{c + Y}]$$

where

$$N(t) = \sum_{i=1}^{n} I(X_i \leq t, \delta_i = 1)$$

and

$$Y(t) = \sum_{i=1}^{n} I(X_i \geq t)$$

are two counting processes derived from the data. The Bayes estimator of $R(t)$ under squared-error loss is

$$\hat{R}(t) = \int_0^t \frac{cdR_0 + dN}{c + Y}$$

and the corresponding estimator of F is

$$\hat{F}(t) = 1 - \prod_{[0,\ t]} \left[1 - \frac{cdR_0 + dN}{c + Y}\right].$$

As $c(\cdot)$ decreases to zero, $\hat{F}$ tends to the Kaplan-Meier estimator.

3 Semiparametric Method

Often, it may be necessary to impose certain structural restrictions on the underlying survival/reliability model to aid in the physical understanding and interpretation, all the while maintaining considerable generality. In such situations, one may decide to use semiparametric models, whereby parts of the model are parametric (reflecting the desired structural restriction) and the remainder nonparametric. An example of this is the Cox model, introduced by Cox (1972). To perform Bayesian analysis, the nonparametric part is assumed to be a realization of a stochastic process. As before in the nonparametric Bayesian analysis case, one can put nonparametric priors on the hazard rates. Different methods have been discussed in Sinha and Dey (1997) for dealing with these models.

Recently, Merrick et al. (2003) developed a semiparametric Bayesian proportional hazards model for reliability and maintenance of machine tools. Their proposed model uses a mixture of Dirichlet process (MDP) prior for the baseline failure rate in the proportional hazards model. Such priors were introduced by Antoniak (1974) and have been popularized by various authors such as MacEachern (1994) and West et al. (1994). Apart from the above, we were unable to find Bayesian semiparametric models in the context of reliability analysis. Below, we present a new model to estimate component reliability using system reliability data from multiple r-out-of-k systems.

In industrial and biological problems, one often has a multicomponent system (consisting of, say, k components) and observes the time of system failure. Due to the system architecture, failure of the system occurs if and only if at least a certain number of components (say, r) fail. For example, an LCD display may be said to correctly function when at least 80% of its pixels properly function. Such a system is called an r-out-of-k system, special cases of which are a series system ($r = 1$) and a parallel system ($r = k$). r-out-of-k systems have been well studied in the reliability context and are a favorite way of increasing system redundancy. See Hoyland and Rausand (1994) for more details on r-out-of-k systems and associated examples.

Suppose we have an r-out-of-k system which consists of k components with identical life distributions that act independent of each other. Assume that the component life distribution is given by $F(\cdot|\theta)$ where θ is an unknown p-dimensional parameter. Suppose T is the system failure time and C is the censoring time. We observe $X = T \wedge C$ and $\delta = I(X = T)$, the censoring indicator. When $\delta = 1$, X is distributed as the rth order statistic based on a sample of size k from $F(\cdot|\theta)$. When $\delta = 0$ (i.e., the system was alive and censored at X), we may or may not have information on the number of components $s(< r)$ in the system that have failed. Let $\gamma = 1$ indicate that s is observed and $\gamma = 0$ indicate that s is unobserved. Assuming that $F(\cdot|\theta)$ is absolutely continuous with respect to the Lebesgue measure, the

likelihood contribution of the system is

$$\begin{aligned} L(\theta) &= \left[r\binom{k}{r} F^{r-1}(x|\theta) f(x|\theta) \{1 - F(x|\theta)\}^{k-r} \right]^{\delta} \\ &\quad \left[\binom{k}{s} F^{s}(x|\theta) \{1 - F(x|\theta)\}^{k-s} \right]^{(1-\delta)\gamma} \\ &\quad \left[\sum_{s=0}^{r-1} \binom{k}{s} F^{s}(x|\theta) \{1 - F(x|\theta)\}^{(k-s)} \right]^{(1-\delta)(1-\gamma)} . \end{aligned}$$

When $\delta = 1$, we take $\gamma = 1$ and $s = r$.

Suppose for $i = 1, \ldots, n$, we have m_i copies of an r_i-out-of-k_i system. The m_i copies are assumed to be independently distributed with the same component distribution $F(\cdot|\theta_i)$. For $i \neq j$, the systems are also assumed to be conditionally independent of each other and the parameters θ_i and θ_j may or may not be equal. The resulting data $\boldsymbol{X}$ is presented in the following tabular form:

System	r	k	Observations		
1	r_1	k_1	$(X_{11},\ \delta_{11},\ \gamma_{11},\ s_{11})$	$\cdots$	$(X_{1m_1},\ \delta_{1m_1},\ \gamma_{1m_1},\ s_{1m_1})$
2	r_2	k_2	$(X_{21},\ \delta_{21},\ \gamma_{21},\ s_{21})$	$\cdots$	$(X_{2m_2},\ \delta_{2m_2},\ \gamma_{2m_2},\ s_{2m_2})$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n	r_n	k_n	$(X_{n1},\ \delta_{n1},\ \gamma_{n1},\ s_{n1})$	$\cdots$	$(X_{nm_n},\ \delta_{nm_n},\ \gamma_{nm_n},\ s_{nm_n})$

Denoting $\boldsymbol{\theta}_n = (\theta_1, \ldots, \theta_n)$, the likelihood is given by

$$\begin{aligned} & L(\boldsymbol{\theta}_n|\boldsymbol{X}) \\ = & \prod_{i=1}^{n} \prod_{j=1}^{m_i} \left[r_i \binom{k_i}{r_i} \{F(x_{ij}|\theta_i)\}^{r_i-1} f(x_{ij}|\theta_i) \{1 - F(x_{ij}|\theta_i)\}^{k_i-r_i} \right]^{\delta_{ij}} \times \\ & \left[\binom{k_i}{s_{ij}} \{F(x_{ij}|\theta_i)\}^{s_{ij}} \{1 - F(x_{ij}|\theta_i)\}^{k_i-s_{ij}} \right]^{(1-\delta_{ij})\gamma_{ij}} \times \\ & \left[\sum_{s=0}^{r_i-1} \binom{k_i}{s} \{F(x_{ij}|\theta_i)\}^{s} \{1 - F(x_{ij}|\theta_i)\}^{k_i-s} \right]^{(1-\delta_{ij})(1-\gamma_{ij})} . \end{aligned}$$

We assume that the θ_i's are independent identically distributed from a distribution G with a Dirichlet process prior having baseline distribution G_0 and precision M. Hence the

prior distribution of $\theta_1,\ \ldots,\ \theta_n$ assuming M and G_0 are known is given by (see Antoniak 1974; Blackwell and MacQueen 1973):

$$\pi(\boldsymbol{\theta}_n) = \prod_{i=1}^{n} \left[\frac{MG_0(d\theta_i) + \sum_{j<i} \delta_{\theta_j}(d\theta_i)}{M+i-1} \right].$$

As mentioned earlier, under the Dirichlet process set-up, some of the system parameters θ_i may be identical. This is a reflection of the fact that some of the systems may be built using components from the same manufacturer and thus would have similar behavior.

Our goal will be to estimate the reliability of a component using the predictive approach. Assuming that a future system has a parameter θ_{n+1} and that X_{n+1} is the lifetime of a component of the system, we want

$$\begin{aligned}
S(x) &= P(X_{n+1} > x|\boldsymbol{X}) \\
&= \int P(X_{n+1} > x|\theta_{n+1},\ \boldsymbol{X},\ \boldsymbol{\theta}_n) f(\theta_{n+1}|\boldsymbol{\theta}_n,\ \boldsymbol{X}) f(\boldsymbol{\theta}_n|\boldsymbol{X}) d\boldsymbol{\theta}_{n+1} \\
&= \int \overline{F}_{n+1}(x|\theta_{n+1}) \frac{1}{M+n} [MG_0(d\theta_{n+1}) + \sum_{j=1}^{n} \delta_{\theta_j}(d\theta_{n+1})] f(\boldsymbol{\theta}_n|\boldsymbol{X}_n) d\boldsymbol{\theta}_n \\
&= \frac{1}{M+n} \int [M \int \overline{F}(x|\theta) G_0(d\theta) + \sum_{j=1}^{n} \overline{F}(x|\theta_j)] f(\boldsymbol{\theta}_n|\boldsymbol{X}) d\boldsymbol{\theta}_n. \qquad (1)
\end{aligned}$$

Note that the complicated nature of the likelihood precludes any conjugate choice of baseline prior to simplify the posterior distribution calculations.

Ferguson (1983), Kuo (1986), Tiwari and Kumar (1989) have used this type of mixture approach for estimating parameters such as density function and reliability. We will use the Gibbs sampler to sample from the posterior and draw inferences. The Gibbs sampler is difficult to implement as is, since the likelihood contributions involve calculations of the cdf, the pdf and/or sums involving them. While several algorithms have been proposed to deal with such non-conjugate set-ups for sampling from the posterior in a MDP set-up (see Neal 2000), they are computationally intensive.

The problem arises because we have several unobserved lifetimes, essentially making data incomplete. Here, we introduce a data-augmentation technique (see Tanner and Wong

1987; Tanner 1993) whereby the observed data are augmented to get the "complete" data $\widetilde{\boldsymbol{X}} = \{X_{ijl}\}$, i.e., the exact failure times of all the components in a system. The likelihood based on the augmented data can then be written as

$$\widetilde{L}(\boldsymbol{\theta}_n|\widetilde{\boldsymbol{X}}) = \prod_{i=1}^{n}\prod_{j=1}^{m_i}\prod_{l=1}^{k_i} f(x_{ijl}|\theta_i).$$

Similar techniques were also used in Kim and Arnold (1999). This simplifies calculations by taking advantage of the conjugate structure of the likelihood and the baseline prior. Note that the posterior distribution of the number of failed components s_{ij} for those systems that such data are missing arises naturally and can be interpreted as the posterior based on an uniform prior on $\{0,\ \ldots,\ k_i\}$.

4 Sampling procedure

1. Generate a set of starting values of $\theta_1,\ \ldots,\ \theta_n$.

2. If $\delta_{ij} = \gamma_{ij} = 0$, sample s_{ij} from binomial$(k_i,\ F(X_{ij}|\theta_i))$ truncated to $\{0,\ \ldots,\ r_i - 1\}$.

3. If $\delta_{ij} = 1$, generate $\{X_{ij(l)}\}_{l=1}^{r-1}$ as order statistics from

$$g_L(x|X_{ij},\ \theta_i) = \frac{f(x|\theta_i)}{F(X_{ij}|\theta_i)}, \qquad 0 < x < X_{ij}.$$

   Also generate $\{X_{ij(l)}\}_{l=r+1}^{k_i}$ as order statistics from

$$g_U(x|X_{ij},\ \theta_i) = \frac{f(x|\theta_i)}{1 - F(X_{ij}|\theta_i)}, \qquad X_{ij} < x < \infty.$$

   Also, set $X_{ij(r)} \equiv X_{ij}$.

   Note that $g_L(\cdot|X_{ij},\ \theta_i)$ is the density of observations from $F(\cdot|\theta_i)$ truncated-above at X_{ij}. Similarly, $g_U(\cdot|X_{ij},\ \theta_i)$ is the density for observations that are truncated-below.

4. If $\delta_{ij} = 0$, generate $\{X_{ij(l)}\}_{l=1}^{s_{ij}}$ as order statistics from $g_L(\cdot|X_{ij},\ \theta_i)$ and also generate $\{X_{ij(l)}\}_{l=s_{ij}+1}^{k_i}$ as order statistics from $g_U(\cdot|X_{ij},\ \theta_i)$.

5. Having generated the "complete data" $\{X_{ijl}\}$, we update the θ's conditionally as

$$\theta_i|\text{ rest } \sim \prod_{j=1}^{m_i}\prod_{l=1}^{k_i} f(X_{ijl}|\theta_i)\left[\frac{MG_0(d\theta_i)+\sum_{j\neq i}\delta_{\theta_j}(d\theta_i)}{M+n-1}\right].$$

   This is done using the Gibbs sampler. Note that in the above expression, we need the exact failure times X_{ijl}, not the ordered values $X_{ij(l)}$. If the sufficient statistic does not depend on the ordering of the observations, one can replace the exact values by the ordered values. Such is the case, for example, when one is interested in inferring about the mean of a normal or the scale of a Weibull distribution.

6. Go back to 2. and repeat until convergence to the steady state.

Below, we give a special case for illustration:

Lognormal distribution

If the component failure times are distributed as Lognormal($\mu,\ \sigma^2$) ($LN(\mu,\ \sigma^2)$), we can generate samples from the truncated distributions given by $g_L(\cdot)$ and $g_U(\cdot)$ using

$$X_L = \exp\left[\mu+\sigma\Phi^{-1}\left(U\Phi\left(\frac{\log X-\mu}{\sigma}\right)\right)\right]$$

and

$$X_U = \exp\left[\mu+\sigma\Phi^{-1}\left(U+(1-U)\Phi\left(\frac{\log X-\mu}{\sigma}\right)\right)\right],$$

where $U\sim U(0,\ 1)$. Note that the random variables X_L and X_U above are generated by inverting the cdfs corresponding to g_L and g_U respectively.

In this case,

$$\prod_{j=1}^{m_i}\prod_{l=1}^{k_i} f(x_{ij(l)}|\theta_i) = \frac{1}{(2\pi\sigma^2)^{m_ik_i/2}}\frac{1}{\prod_{j,l}x_{ij(l)}}\exp\left[-\frac{1}{2\sigma^2}\sum_{j,l}\left(\log x_{ij(l)}-\mu_i\right)^2\right].$$

Note that $\sum_{j=1}^{m_i}\sum_{l=1}^{k_i}\log x_{ij(l)} = \sum_{j=1}^{m_i}\sum_{l=1}^{k_i}\log x_{ijl}$ is the sufficient statistic and it is not order-dependent.

5 Illustration

We used the system configuration given in Table 1 to generate our data (see Chen 2003).

[Table 1 about here.]

To generate an observation on an r_i-out-of-k_i system, a simple random sample of size k_i is generated from F and another simple random sample of size k_i is generated from G. Let $u_{i(r_i)}$ and $v_{i(r_i)}$ be the r_ith order statistics from the two samples respectively. If $u_{i(r_i)} \leq v_{i(r_i)}$, the generated observation is taken as $(X_i,\ \delta_i,\ \gamma_i,\ s_i) = (u_{i(r_i)},\ 1,\ 1,\ r_i)$. Otherwise, let r be the rank of the largest order statistic from F that is smaller than $v_{i(r_i)}$. With 90% probability, the generated observation is taken as $(X_i,\ \delta_i,\ \gamma_i,\ s_i) = (v_{i(r_i)},\ 0,\ 1,\ r)$ and with 10% probability, the generated observation is taken as $(X_i,\ \delta_i,\ \gamma_i,\ s_i) = (v_{i(r_i)},\ 0,\ 0,\ 0)$. We used lognormal distribution with $\mu_{01} = 1,\ \sigma_{01} = 1$ as the F and a lognormal distribution with $\mu_{02} = \{0.8,\ 2.8\},\ \sigma_{02} = 1$ as G. This procedure was seen to give rise to 68% censoring when $\mu_{02} = 0.8$ and about 0% censoring when $\mu_{02} = 2.8$. As indicated earlier, for a censored lifetime, the number of failed components was noted with probability 0.9 in each of the censoring schemes.

Once the dataset was generated according to the above procedure, we estimated the reliability of a single component $S(x)$ as outlined in Equation (1). We assumed that the underlying population is $LN(\mu, 1)$ with $\mu \sim \mathcal{D}(M,\ G_0)$ and $G_0 \sim N(1,\ 1)$. We also assumed $M \sim Gamma(.1,\ .1)$ and ran a full Bayesian approach. The updating of M was done along the lines of Escobar and West (1995) and for improved mixing, the μ_i's were updated using "Algorithm 3" in Neal (2000). The predictive reliability is based on 5,000 runs of the Gibbs sampler with a thinning of 10, obtained after a burn-in of 5,000 iterations. Convergence was ascertained using CODA (see Best et al. 1995). The results are presented in Figure 1. The precision of the estimate is measured by RMSE, which is the root-mean-square-error of the estimate at selected points is given in Table 2.

[Table 2 about here.]

We see that as the percentage of censoring increases, the estimated values get farther away from the true values. In the case where $\mu_{02} = 0.8$, we also kept track of the true (but unobserved) number of component failures at the data generation stage and the estimated values at the simulation stage. The average number of failures turned out to be $(0,\ 0,\ 2.18,\ 2.77,\ 5.70,\ 5.97)$ while the true values are $(0,\ 0,\ 2,\ 1,\ 6,\ 6)$.

[Figure 1 about here.]

6 Conclusion

We have presented a brief overview of nonparametric and semiparametric Bayes methods in lifetime data analysis. In the semiparametric method outlined here, one can, as a byproduct, infer about the number of failed components for a censored system lifetime when it is unobserved. The method of data augmentation can be used when the sufficient statistic is not order-dependent — otherwise one can always use the original likelihood and use one of the non-conjugate sampling methods outlined in Neal (2000). This discussion is not exhaustive but is intended to give a flavor of the current state of the art. Due to advances in computing, realistic models which were once avoided due to difficulty in implementation will be becoming more popular.

References

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems," *The Annals of Statistics*, 2, 1152–1174.

Arjas, E. and Gasbarra, D. (1994), "Nonparametric Bayesian Inference from Right-censored Survival Data using the Gibbs Sampler," *Statistica Sinica*, 4, 505–524.

Basu, D. and Tiwari, R. C. (1982), "A Note on the Dirichlet Process," in *Statistics and Probability: Essays in Honor of C. R. Rao*, North-Holland, New York, pp. 89–103.

Best, N. G., Cowles, M. K., and Vines, K. (1995), "CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output, Version 0.30," Tech. rep., MRC Biostatistics Unit, University of Cambridge.

Blackwell, D. and MacQueen, J. B. (1973), "Ferguson Distribution Via Polya Urn Schemes," *The Annals of Statistics*, 1, 353–355.

Boyles, R. A. and Samaniego, F. J. (1987), "On Estimating Ccomponent Reliability for Systems with Random Redundncy Levels," *IEEE Transactions in Reliability*, R-36, 403–407.

Chen, Z. (2003), "Component Reliability Analysis of k-out-of-n Systems with Censored Data," *Journal of Statistical Planning and Inference*, 116, 305–315.

Cox, D. R. (1972), "Regresion Models and Life Tables," *Journal of the Royal Statistical Society, Series B*, 34, 187–220.

Damien, P., Laud, P. W., and Smith, A. F. M. (1995), "Approximate Random Variate Generation from Infinitely Divisible Distributions with Applications to Bayesian Inference," *Journal of the Royal Statistical Society, Series B*, 57, 547–563.

Doksum, K. A. (1974), "Tailfree and Neutral Random Probabilities and their Posterior Distributions," *The Annals of Probability*, 2, 183–201.

Dykstra, R. L. and Laud, P. (1981), "A Bayesian Approach to Reliability," *The Annals of Statistics*, 9, 356–367.

Escobar, M. D. and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

— (1983), "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*, New York; London: Academic Press, pp. 287–302.

Ferguson, T. S. and Phadia, E. G. (1979), "Bayesian Nonparametric Estimation Based on Censored Data," *The Annals of Statistics*, 7, 163–176.

Ferguson, T. S., Phadia, E. G., and Tiwari, R. C. (1992), "Bayesian Nonparametric Inference," in *Current Issues in Statistical Inference:Essays in Honor of D. Basu*, Institute of Mathematical Statistics, Hayward, CA, no. 17 in IMS Lecture Notes Monograph Series, pp. 127–150.

Hjort, N. L. (1990), "Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data," *The Annals of Statistics*, 18, 1259–1294.

Hoyland, A. and Rausand, M. (1994), *System Reliability Theory: Models and Statistical Methods*, Wiley, New York.

Kalbfleisch, J. D. (1978), "Non-parametric Bayesian Analysis of Survival Time Data," *Journal of the Royal Statistical Society, Series B*, 40, 214–221.

Kim, Y. and Arnold, B. C. (1999), "Parameter Estimation under Generalized Ranked Set Sampling," *Statistics and Probability Letters*, 42, 353–360.

Kuo, L. (1986), "Computations of Mixtures of Dirichlet Processes," *SIAM Journal of Scientific Statistical Computing*, 7, 60–71.

Kvam, P. H. and Samaniego, F. J. (1993), "On Maximum Likelihood Estimation Based on Ranked Set Samples with Applications to Reliability," in *Advances in Reliability*, Netherlands: Elsevier Science B. V., pp. 215–229.

Laud, P. W., Smith, A. F. M., and Damien, P. (1996), "Monte Carlo Methods for Approximating a Posterior Hazard Rate Process," *Statistics and Computing*, 6, 77–83.

Lochner, R. H. (1975), "A Generalized Dirichlet Distribution in Bayesian Life Testing," *Journal of the Royal Statistical Society, Series B*, 37, 103–113.

MacEachern, S. N. (1994), "Estimating Normal Means with a Conjugate-Style Dirichlet Process Prior," *Communications in Statistics: Simulation and Computation*, 23, 727–741.

Mazzuchi, T. A. and Singpurwalla, N. D. (1985), "A Bayesian Approach to Inference for Monotone Failure Rates," *Statistics and Probability Letters*, 3, 135–142.

Merrick, J. R. W., Soyer, R., and Mazzuchi, T. A. (2003), "A Bayesian Semiparametric Analysis of the Reliability and Maintenance of Machine Tools," *Technometrics*, 45, 58–69.

Neal, R. M. (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249–265.

Nieto-Barajas, L. E. and Walker, S. G. (2002), "Markov Beta and Gamma Processes for Modelling Hazard Rates," *Scandinavian Journal of Statistics*, 29, 413–424.

Padgett, W. J. and Wei, L. J. (1981), "A Bayesian Nonparametric Estimator of Survival Probability Assuming Increasing Failure Rate," *Communications in Statistics: Theory and Methods*, A10, 49–63.

Sethuraman, J. and Tiwari, R. C. (1982), "Convergence of Dirichlet Measures and the Interpretaion of Their Parameter," in *Statistical Decision Theory and Related Topics III*, Academic Press, vol. 2, pp. 305–315.

Singpurwalla, N. D. (2006), *Reliability and Risk: A Bayesian Perspective*, Wiley:NY.

Sinha, D. and Dey, D. K. (1997), "Semiparametric Bayesian Analysis of Survival Data," *Journal of the American Statistical Association*, 92, 1195–1212.

Susarla, V. and VanRyzin, J. (1976), "Nonparametric Bayesian Estimation of Survival Curves from Incomplete Observations," *Journal of the American Statistical Association*, 71, 740–754.

Tanner, M. A. (1993), *Tools for Statistical Inference*, Springer, New York.

Tanner, M. A. and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Tiwari, R. C. and Jammalamadaka, S. R. (1985), "Estimation of Survival Function and Failure Rates with Censored Data," *Statistics*, 16, 535–540.

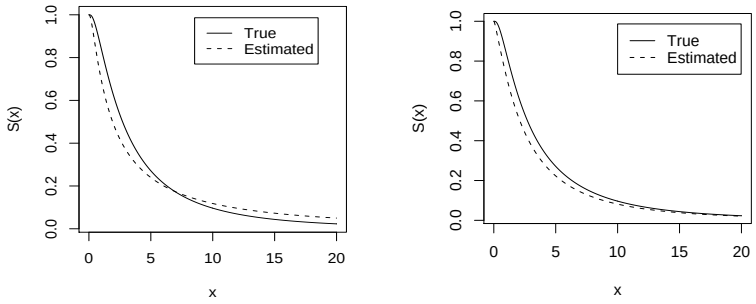
Tiwari, R. C. and Kumar, S. (1989), "Bayes Reliability Estimation under a Random Environment Governed by a Dirichlet Prior," *IEEE Transactions on Reliability*, R-37, 218–223.

Tiwari, R. C. and Rao, J. S. (1983), "Bayesian Nonparametric Estimation of Failure Rates with Censored Data," *Calcutta Statistical Association Bulletin*, 32, 79–90.

Walker, S. and Damien, P. (1998), "A Full Bayesian Non-parametric Analysis Involving Neutral To The Right Process," *Scandinavian Journal of Statistics*, 25, 669–680.

West, M., Muller, P., and Escobar, M. D. (1994), "Hierarchical Priors and Mixture Models with Application in Regression Density Estimation," in *Aspects of Uncertainty: A Tribute to D. V. Lindley*, London: Wiley, pp. 363–368.

Wilks, S. S. (1962), *Mathematical Statistics*, New York: Wiley.



(a) $\mu_{02} = 0.8$ (b) $\mu_{02} = 2.8$

Figure 1: Estimated reliability of a component from the system configuration in Table 1 using different censoring distributions.

i	1	2	3	4	5	6
r_i	1	3	5	1	4	7
k_i	5	5	5	7	7	7
m_i	10	9	11	12	10	13

Table 1: Configuration of several r-out-of-k systems.

μ_{02}	RMSE	Censoring proportion
0.8	0.055	67.69%
2.8	0.048	0%

Table 2: Accuracy of the estimates based on data from various censoring schemes.