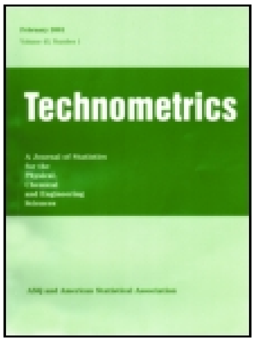

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://www.tandfonline.com/loi/utch20

A Bayesian Nonparametric Test for Minimal Repair

Li Li, Timothy Hanson, Paul Damien & Elmira Popova

To cite this article: Li Li, Timothy Hanson, Paul Damien & Elmira Popova (2014) A Bayesian Nonparametric Test for Minimal Repair, Technometrics, 56:3, 393-406, DOI: 10.1080/00401706.2013.842934

To link to this article: http://dx.doi.org/10.1080/00401706.2013.842934

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

A Bayesian Nonparametric Test for Minimal Repair

Li LI and Timothy HANSON

Department of Statistics
University of South Carolina
Columbia, SC 29201
(lil@email.sc.edu; hansont@stat.sc.edu)

Paul DAMIEN

McCombs School of Business
University of Texas at Austin
Austin, TX 78712
(paul.damien@mccombs.utexas.edu)

Elmira POPOVA

Department Mechanical Engineering
University of Texas at Austin
Austin, TX 78712
(elmira@mail.utexas.edu)

The article develops a Bayesian nonparametric reliability model for recurrent events where failure and truncated time-to-failure density shape is regressed on past maintenance decisions: perfect repair and minimal repair. By comparing the system interfailure lifetime distributions after minimal and perfect repair, we are able to test the minimal repair assumption of "good as old." Interfailure hazard functions after perfect and minimal repairs are estimated, shedding light on departures from minimal repair. The method is illustrated both on simulated data as well as failure time data from air-conditioning units at the South Texas Nuclear Operating Company near Bay City, Texas. This article has supplementary material online.

KEY WORDS: Effective age; Repairable system; Tailfree process; Truncated data.

1. INTRODUCTION

Repairable systems have been widely studied in the reliability literature. Systems fail and upon each failure, a system gets repaired. The distribution of interfailure times between system failures is commonly of interest. In general, recurrent event modeling methods can be divided into categories based on the type of maintenance a system receives. Renewal processes are commonly used if all the maintenance repairs bring the system to a "good as new" state (commonly known as perfect repair). One example of this kind of repair would be a complete overhaul of the system. Nonhomogenous Poisson processes are used if all repairs bring the system to the "good as old" state (commonly known as minimal repair), for example, replacing a failed sub-component of a system. Some authors have proposed models that allow for a combination of perfect and minimal repairs, see Block, Borges, and Savits (1985) and Whitaker and Samaniego (1989). However, the basic assumption of a consistently "minimal" repair is questionable; usually several types of maintenance, with varying degrees of repair, are undertaken throughout the lifetime of the system. Kijima (1989) proposed a general model that includes perfect, minimal, and in-between repairs by introducing the "effective age" of the system after each repair, essentially measuring how successful the repair was. Following Kijima (1989), Dorado, Hollander, and Sethuraman (1997) allowed for repairs of varying degree by including so-called "life supplements"—numbers between zero and one indicating the degree of the repair between perfect and minimal—and assumed the life supplements are *known*. Recently, Veber, Nagode, and Fajdiga (2008) assumed one life supplement that is unknown, that is, each repair reduces the effective age of the system by the same fraction q, and proposed an expectation–maximization (EM) algorithm to estimate q and the unknown failure distribution F. As an extension to a common q, Rigdon and Pan (2009) allowed the repair effectiveness parameter to vary from system to system. Presnell, Hollander, and Sethuraman (1994) proposed a test for the minimal repair assumption in a particular model that Block, Borges, and Savits (1985) proposed; if the null hypothesis where minimal repair assumption holds is rejected, the question remains as to whether "minimal repair" brings the system better or worse than minimal; in many application scenarios this distinction is crucial. If one ignores maintenance decisions, Cooper, de Mello, and Kleywegt (2006) pointed out that decisions based on the incorrect assumption of minimal repairs can lead to a so-called "spiral down" effect, where system reliability gets worse than the predicted level after repair cycles (i.e., more failures than predicted); this happens because the assumed minimal repairs are actually often worse than "good as old."

Consider a brand new system; denote the cumulative distribution function (CDF) for the first failure s as $F_0(s)$. After a perfect repair at failure time s, which sets the system clock back to zero, the distribution governing the next failure is $F_0(t - s)$, where $t > s$. A minimal repair brings the system to the exact

TECHNOMETRICS, AUGUST 2014, VOL. 56, NO. 3
DOI: 10.1080/00401706.2013.842934

state it was in right before failure; this implies that the intensity function, formally defined below in (1), does not change after a minimal repair. After successive minimal repairs, if a minimal repair is newly performed at failure time s (time since new condition), the CDF for the next failure is F_0, but truncated to be larger than s, that is, $[F_0(t) - F_0(s)]/S_0(s)$, where $S_0(s) = 1 - F_0(s)$ is the reliability (also termed survival) function and $t > s$. We propose to relax this assumption by allowing the intensity to change after the minimal repair: the distribution for the next failure is instead $[F_1(t) - F_1(s)]/S_1(s)$. The assumption of a static intensity function is given by $H_0 : F_0 = F_1$, providing an intuitive test of the minimal repair assumption. If H_0 is rejected in favor of $H_1 : F_0 \neq F_1$, estimated hazard functions $h_0(t) = f_0(t)/S_0(t)$ and $h_1(t) = f_1(t)/S_1(t)$ enable us to find when system performs actually worse (or better) than the expected condition under the minimal repair assumption. Our framework can be easily generalized to include known life supplements (as in Dorado, Hollander, and Sethuraman 1997) and subsequently test for this assumption.

The hypothesis testing in this article involves two unknown distributions. A parametric approach assumes particular distribution families for F_0 and F_1, for example, Weibull is commonly used for nonhomogenous Poisson or renewal process models. We propose a Bayesian nonparametric model that generalizes the Weibull assumption on F_0 and F_1, termed a "tailfree prior." The tailfree approach we use augments the standard Weibull family indexed by $\boldsymbol{\theta}$ with additional parameters $\{\pi(\boldsymbol{\epsilon})\}$ that change the shape of the Weibull density in successive layers. These additional parameters add flexibility beyond the Weibull shape, much like adding detail to an initially washed canvas; each new layer or "level" allows more refined detail to be accommodated. The Bayesian approach simply places a prior on the additional parameters $\{\pi(\boldsymbol{\epsilon})\}$.

In general, Bayesian nonparametric methods model distributions as random CDFs $F(s)$, either directly or indirectly (e.g., through the hazard). Technically, a random CDF $F(s)$ is a stochastic process indexed by s, so $\{F(s) : s > 0\}$ describes a random function from $\mathbb{R}^+$ to [0, 1], and for any fixed s, $F(s)$ is a random variable. These processes include the Dirichlet process (Ferguson 1973), Polya tree priors (Lavine 1992), Dirichlet process mixtures (Escobar and West 1995), and neutral to the right processes (Ferguson and Phadia 1979). Taddy and Kottas (2012) used Dirichlet process mixtures for the interfailure density in Poisson process models. Priors on the space of cumulative hazard functions include gamma processes, weighted gamma processes, beta processes; see Lo (1992), Kuo and Ghosh (1997), and Ishwaran and James (2004). Often, the random CDF $F(s)$ is centered at a parametric distribution $G_{\boldsymbol{\theta}}$ in the sense that $E\{F(s)\} = G_{\boldsymbol{\theta}}(s)$ for all $s > 0$, that is, $G_{\boldsymbol{\theta}}$ is the "prior mean" of F. Our proposed framework uses tailfree priors (Fabius 1964; Ferguson 1974; Jara and Hanson 2011) to model F centered at the Weibull family, $E\{F(s)\} = 1 - e^{-(s/\gamma)^{\alpha}}$ given (α, γ), but allows for substantial data-driven deviations from Weibull. Our approach naturally tests whether Weibull is adequate, as well as incorporating maintenances where no failure has actually occurred (i.e., censored failures). Few existing nonparametric approaches make use of information from censored system failure times, although in practice, maintenance schedules are common.

Table 1. Counts of perfect/minimal by response to "failure"/"censored" for the air conditioners

	Failure	Censored
Perfect	86	1175
Minimal	1085	14

The proposed estimation procedure is applied to historical data from the South Texas project nuclear operating company located in Bay City, Texas. The system of interest is the essential chillers, which is a group of six 300-ton air conditioners, three for each nuclear reactor unit. They provide chilled water for air handling units to provide a suitable environment for personnel and equipment located in the electrical auxiliary building, mechanical auxiliary building, and fuel handling building. An essential chiller provides chilled water for the cooling coils of various safety related air handling units during normal, faulted, and upset conditions. All three chilled water system trains are automatically started up if particular emergency situations are detected, such as safety injection signal, loss of offsite power from the switchyard, or a combination of both, to supply cooling to many essential safety systems. Those air conditioners are repairable systems. Maintenances to the air conditioners include replacement of subcomponents (oil pump, vane controller, solenoid valves, etc.) in response to failures and overhauls, typically upon inspection, which involves a major rework on parts, for example, compressor vane and renewing soft materials (gaskets, refrigerant, lubrication—grease, oil) when excessive wear or other degraded conditions are noted. In our analysis, overhauls are categorized into perfect repairs while replacement of subcomponents is grouped into minimal repairs. For repairs that are not in response to a failure, yielding right-censored failure times, we do not differentiate scheduled repairs and responses to apparent degradation (but not failure), and further assume that the times for those repairs are independent of the system failure processes. The dataset is comprised of two groups of observations for the two nuclear reactor units with the first group of 1274 events and the second group of 1092 events. All air conditioners are assumed to work independently. Each observation consists of an event time, associated maintenance decision, and indicator of censoring for whether a failure occurred at the event time, that is, (t_i, d_i, δ_i). Most minimal repairs were in response to failure, and perfect repairs were performed without an accompanying failure (Table 1). It is assumed in this data analysis that those perfect repairs bring the system to the "good as new" condition and our main interest is that whether those minimal repairs bring the system to the "good as old" state.

Section 2 describes the model, introduces tailfree priors, and outlines the Markov chain Monte Carlo (MCMC) algorithm used to fit the model. Section 3 presents simulation results for testing the minimal repair assumption and accompanying density estimation. Section 4 applies the method to the South Texas project data. Section 5 concludes the article with a discussion.

2. MODEL DEVELOPMENT

Consider a general repairable system framework: up to the present time $t_{\max}$ we observe a series of repairs and maintenance decisions made at each repair. The times for repairs are recorded as $0 = t_0 < t_1 < t_2 < \cdots < t_n = t_{\max}$. The corresponding repair at event time t_i is denoted as d_i with $d_i = 1$ if minimal repair was performed and $d_i = 0$ if perfect repair was performed; we assume $d_0 = 0$. Denote the last perfect repair time prior to decision d_i as $t_i^* = \max\{t_j : j < i,\ d_j = 0\}$. If maintenance (random or planned) is performed at time t_i *without* an accompanying failure, the failure time stemming from the previous decision is censored, indicated by $\delta_i = 0$ and 1 otherwise. For simplicity, we assume that δ_i is independent of the failure process. Since repairs must occur after failures, and can also occur without a failure event, the set of failure times is a subset of $\{t_1, \ldots, t_n\}$. Full data are $\mathcal{D} = \{(t_i, \delta_i, d_i)\}_{i=1}^n$. For data observed over the window $[0, t_{\max}]$, the event time $t_n = t_{\max}$ is the time at which data collection stops and $\delta_n = 0$. We assume that maintenance decisions are observable where perfect repairs bring the system to "good as new" state and minimal repairs otherwise. Furthermore, we assume that the time for repair is negligible, that is, there is no "down" time during the repair.

Let the counting process $\{N(t), t \geq 0\}$ record the cumulative number of failures over time and $H(t) = \{N(s) : 0 \leq s < t\}$ denote the history of the process at time t. Then the intensity function is defined as

$$\phi(t|H(t)) = \lim_{\triangle \to 0^+} \frac{\Pr\{N(t+\triangle) - N(t) = 1|H(t)\}}{\triangle}. \tag{1}$$

The intensity function describes the instantaneous probability of a failure occurring at t, conditioning on the process history. Let $F_0(t)$ be the probability that the system lasts less than t time units since a perfect repair and $S_0(t) = 1 - F_0(t)$ be the survival probability; denote the density as $f_0(t)$ and hazard as $h_0(t)$. A previous perfect repair $d_{i-1} = 0$ brings the system to "good as new" status, that is, resets the system clock to zero. A failure t_i right after perfect repair $d_{i-1} = 0$ at t_{i-1} has likelihood contribution $f_0(t_i - t_{i-1})$. If instead, a minimal repair $d_{i-1} = 1$ restores the system to the exact state it was in right before failure at t_{i-1}, then the system has aged $t_i - t_i^*$ units since the last perfect repair, truncated at $t_{i-1} - t_i^*$, yielding the likelihood contribution $f_0(t_i - t_i^*)/S_0(t_{i-1} - t_i^*)$. This above assumption is commonly referred to as "minimal repair assumption" and it implies that the underlying intensity function for the recurrent events does not change after minimal repairs, $\phi(t|H(t)) = h_0(t - t_i^*)$ over $[t_i^*, t_i)$, regardless of the minimal repairs preceding t_i. In our framework, we do not make the minimal repair assumption, and simply allow the intensity function to change after the first minimal repair on the renewed system. The intensity function is then $\phi(t|H(t)) = h_1(t - t_i^*)$, for hazard $h_1(t) = f_1(t)/S_1(t)$ and $S_1(t) = \int_t^\infty f_1(s)ds$, and the likelihood contribution is $f_1(t_i - t_i^*)/S_1(t_{i-1} - t_i^*)$. Note that subsequent minimal repairs do not further change the intensity function. The resulting model can be viewed as a generalization to a two-state Poisson process. Given substantially more data, a multistate Poisson process (Cook and Lawless 2007) could be fit, assuming $\phi(t|H(t)) = h_k(t)$, $N(t-) = k$, $k \in \{0, 1, 2, \ldots\}$ following each perfect repair.

Denote $H_0 : F_0 = F_1$ as the hypothesis assuming the minimal repair assumption holds and $H_1 : F_0 \neq F_1$ as the hypothesis allowing departure from this assumption. Under H_0 we put one tailfree prior (to be elaborated in Section 2.1) on F_0; under H_1 we place two conditionally independent priors on F_0 and F_1. Under H_0, the likelihood is

$$\mathcal{L}(f_0) = \prod_{i=1}^n [f_0(t_i - t_{i-1})^{\delta_i} S_0(t_i - t_{i-1})^{1-\delta_i}]^{1-d_{i-1}} \times \left[\frac{f_0(t_i - t_i^*)^{\delta_i} S_0(t_i - t_i^*)^{1-\delta_i}}{S_0(t_{i-1} - t_i^*)}\right]^{d_{i-1}}. \tag{2}$$

Under H_1, the likelihood is

$$\mathcal{L}(f_0, f_1) = \prod_{i=1}^n [f_0(t_i - t_{i-1})^{\delta_i} S_0(t_i - t_{i-1})^{1-\delta_i}]^{1-d_{i-1}} \times \left[\frac{f_1(t_i - t_i^*)^{\delta_i} S_1(t_i - t_i^*)^{1-\delta_i}}{S_1(t_{i-1} - t_i^*)}\right]^{d_{i-1}}. \tag{3}$$

In terms of interfailure hazard functions h_0, the likelihood under H_1 is

$$\mathcal{L}(h_0, h_1) = \prod_{i=1}^n \left[h_0(t_i - t_{i-1})^{\delta_i} \exp\left\{-\int_0^{t_i - t_{i-1}} h_0(s)ds\right\}\right]^{1-d_{i-1}} \times \left[h_1(t_i - t_i^*)^{\delta_i} \exp\left\{-\int_{t_{i-1}}^{t_i} h_1(s - t_i^*)ds\right\}\right]^{d_{i-1}}.$$

It is straightforward to interpret F_0 as the failure time distribution of a new system. Let t^* be the time at which the last perfect repair was made. Since our alternative model assumes $\phi(t|H(t)) = h_1(t - t^*)$ after minimal repairs, an estimate of h_1 averages the intensities over time after the first minimal repair in each renewed cycle. When the system performs better (or worse) than an expected level at time t under the minimal repair assumption, then $h_1(t)$ will be lower (higher) than $h_0(t)$. We perform a simulation in Section 3 to illustrate this point.

2.1 Tailfree Process Priors on F_0 and F_1

We place tailfree process priors on F_0 and F_1. The use of the term "tailfree" dates to Freedman (1963), who considered conditions for consistency of Bayesian probability measures on the positive integers; the main condition has to do with the shape of the tail of the density, that is, the integers stretching off to infinity. Fabius (1964) extended Freedman's notion of "tailfree" to continuous measures with densities; however the requirements for consistency no longer deal with the tails of the distribution. The construction that follows is a modest reworking of Fabius (1964).

Let $G_{\boldsymbol{\theta}}$ denote the Weibull CDF parameterized as $G_{\boldsymbol{\theta}}(t) = 1 - \exp\{-(t/\gamma)^\alpha\}$ for $t \geq 0$, where $\boldsymbol{\theta} = (\log(\alpha), \log(\gamma))'$. Let $g_{\boldsymbol{\theta}}(t)$ be the corresponding density. The tailfree prior augments the Weibull family indexed by $\boldsymbol{\theta}$ with additional parameters $\{\pi(\boldsymbol{\epsilon})\}$ ($\boldsymbol{\epsilon}$ is a binary number, described below), which change the shape of the Weibull density on successive levels; this "more flexible Weibull" CDF is denoted as $F(t)$. Before delving into the definition, we can get the flavor of the approach through a preliminary look at Figure 1. Panel (a) shows a Weibull density for a particular $\boldsymbol{\theta}$; let T be drawn from this Weibull density. Panel (b) adjusts the shape of the density by adding one parameter $\pi(0)$ that changes the probability of T being less than the median of the Weibull density from 0.5 to 0.45, but leaves the shape

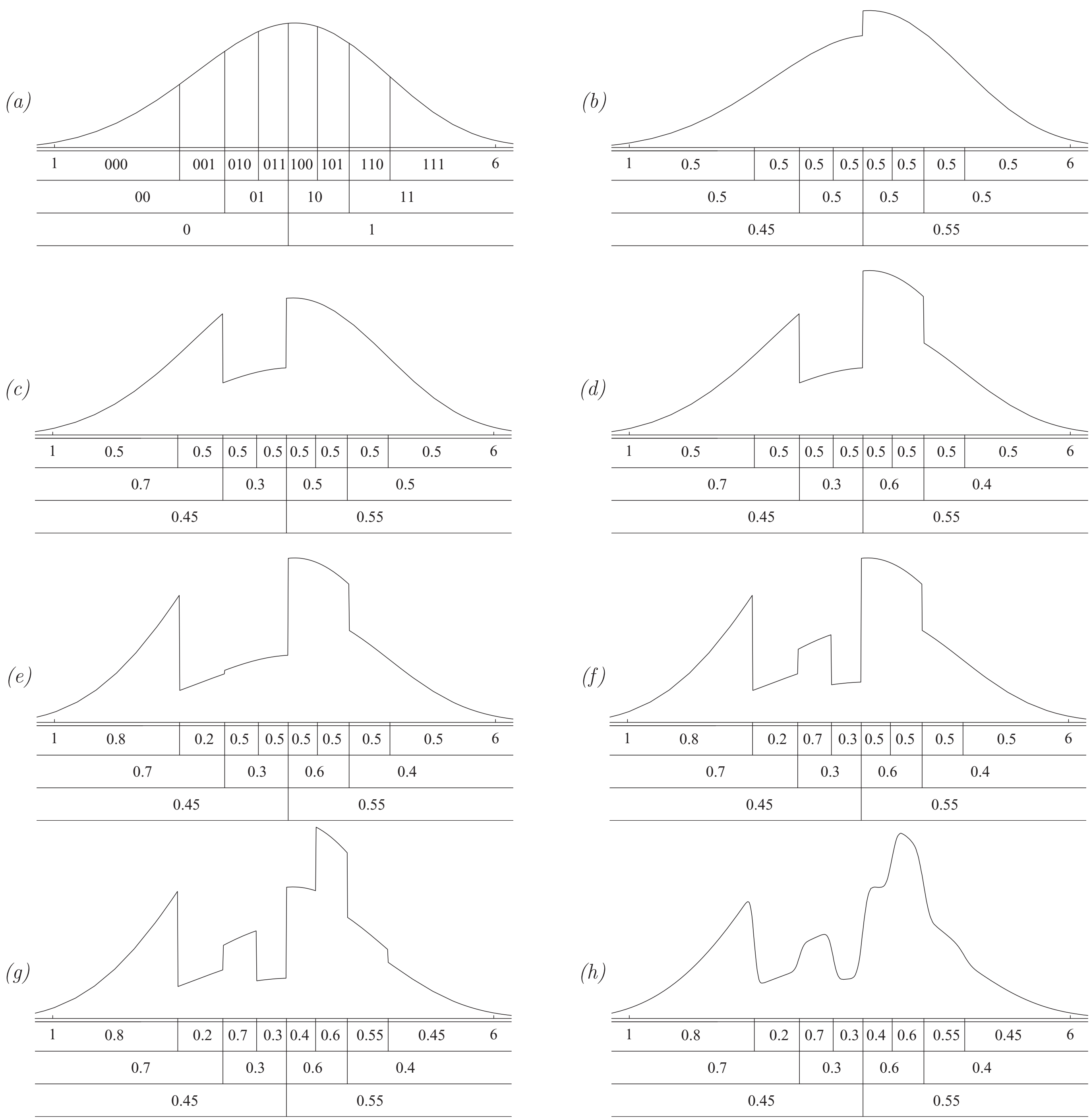

Figure 1. (a) Weibull (α, γ) with $\alpha = 4$ and $\gamma = 4$; (b)–(g) tailfree densities, centered at (a) with conditional probabilities specified up to $J = 3$; and (h) mixture of tailfree processes assuming $\alpha, \gamma \overset{\text{ind.}}{\sim} N(4, 0.05^2)$.

of the density the same—this is the first level. Panels (c) and (d) add two more parameters, $\pi(00)$ and $\pi(10)$ successively, which modify the shape of the density on smaller sets in the second level, but leave the density shape the same on these smaller sets. Panels (e)–(g) add four more parameters on the third level. The Bayesian approach simply places priors on the parameters $\{\pi(\boldsymbol{\epsilon})\}$, in addition to $\boldsymbol{\theta}$, yielding a random CDF F and corresponding density f. Let $T \sim F$. The prior is chosen so that, given $\boldsymbol{\theta}$, the probability $P_F(a < T < b) = \int_a^b f(s)ds$ has expectation $\int_a^b g_{\boldsymbol{\theta}}(s)ds$, for example, $E\{F(s)\} = G_{\boldsymbol{\theta}}(s)$ for any $s > 0$. In this sense, the "prior mean" of F is $G_{\boldsymbol{\theta}}$.

We now present a technical specification of the nonparametric prior. Let $\epsilon_1 \ldots \epsilon_j$ be a j-digit binary number where $\epsilon_i \in \{0, 1\}$ for $i = 1, 2, \ldots, j$. Each $\boldsymbol{\epsilon} = \epsilon_1 \ldots \epsilon_j$ indexes a set $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}) \subset [0, \infty)$. Following Lavine (1992), these sets are intervals with endpoints that are quantiles of the centering family: if m is the base-10 representation of the binary number $\boldsymbol{\epsilon} = \epsilon_1 \ldots \epsilon_j \in \{0, 1\}^j$, then $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon})$ is the interval $(G_{\boldsymbol{\theta}}^{-1}(m/2^j), G_{\boldsymbol{\theta}}^{-1}((m+1)/2^j)]$. Note then that at each level j, the class $\{B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}) : \boldsymbol{\epsilon} \in \{0, 1\}^j\}$ forms a partition of the positive reals and furthermore $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}) = B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}0) \cup B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}1)$. Figure 1(a) shows the first three partitions for a Weibull(4,4) centering distribution, for example, $\boldsymbol{\theta} = (\log(4), \log(4))$. Note that $[0, \infty) = B_{\boldsymbol{\theta}}(0) \cup B_{\boldsymbol{\theta}}(1)$, $[0, \infty) = B_{\boldsymbol{\theta}}(00) \cup B_{\boldsymbol{\theta}}(01) \cup B_{\boldsymbol{\theta}}(10) \cup B_{\boldsymbol{\theta}}(11)$, etc. For a specific $\boldsymbol{\epsilon}0$, the parameter $\pi(\boldsymbol{\epsilon}0)$ approximately follows a beta(cj^2, cj^2) density, where j is the number of digits in $\boldsymbol{\epsilon}0$; more details are presented below. Walker et al. (1999) suggested thinking of a "... *particle cascading through these partitions.*" The particle, say $T \sim F$, initially moves into $B_{\boldsymbol{\theta}}(0)$ with probability $\pi(0)$ or into $B_{\boldsymbol{\theta}}(1)$ with probability $\pi(1) = 1 - \pi(0)$. From then on, at any level j with index $\boldsymbol{\epsilon} = \epsilon_1 \ldots \epsilon_j$, if the particle is in $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon})$, it moves into $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}0)$ with probability $\pi(\boldsymbol{\epsilon}0)$ or into $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon}1)$ with probability $\pi(\boldsymbol{\epsilon}1) = 1 - \pi(\boldsymbol{\epsilon}0)$. When the particle finally makes its way into a set $B_{\boldsymbol{\theta}}(\epsilon_1 \ldots \epsilon_J)$ in the finest partition at level J, it simply follows the base CDF $G_{\boldsymbol{\theta}}$ restricted to $B_{\boldsymbol{\theta}}(\epsilon_1 \ldots \epsilon_J)$—this does not depend on the $\{\pi(\boldsymbol{\epsilon})\}$. That is, for an interval $(a, b) \subset B_{\boldsymbol{\theta}}(\epsilon_1 \ldots \epsilon_J)$ and $T \sim F$,

$$P\{a < T < b | T \in B_{\boldsymbol{\theta}}(\epsilon_1 \ldots \epsilon_J)\} = \frac{\int_a^b g_{\boldsymbol{\theta}}(s) ds}{\int_{B_{\boldsymbol{\theta}}(\epsilon_1 \ldots \epsilon_J)} g_{\boldsymbol{\theta}}(s) ds}. \quad (4)$$

The definition of a tailfree prior uses a binary partitioning tree. Although most authors have used binary splits, other partitioning schemes could be implemented, for example, Mauldin, Sudderth, and Williams (1992).

If all of the conditional probabilities are equal to one-half, that is, $\pi(\boldsymbol{\epsilon}) = 0.5$ for all $\boldsymbol{\epsilon}$, then the density $f(s)$ is simply $g_{\boldsymbol{\theta}}(s)$, the corresponding Weibull density. The tailfree prior simply takes the *expectation* of these conditional probabilities to be one-half, $E\{\pi(\boldsymbol{\epsilon})\} = 0.5$ for all $\boldsymbol{\epsilon}$; then $E\{f(s)\} = g_{\boldsymbol{\theta}}(s)$. For a given set of conditional probabilities $\{\pi(\boldsymbol{\epsilon})\}$, this construction builds a density $f(s)$ that has jumps at the quantiles of $G_{\boldsymbol{\theta}}$, $G_{\boldsymbol{\theta}}^{-1}(m/2^J)$, and the values of $\{\pi(\boldsymbol{\epsilon})\}$ determine the jump size. Figure 1(a) takes all $\pi(\boldsymbol{\epsilon}) = 0.5$. Figure 1(b) then sets $\pi(0) = 0.45$. Figure 1(c) further sets $\pi(00) = 0.7$, then Figure 1(d) sets $\pi(10) = 0.6$. Panels (e)–(g) successively set $\pi(000) = 0.8$, $\pi(010) = 0.7$, and $\pi(100) = 0.4$ and $\pi(110) = 0.55$. Already with only three levels, we obtain quite interesting possibilities. Typically, the number of levels is higher, usually $5 \le J \le 8$, allowing for more refined shapes. The original Fabius (1964) construction deals with $J = \infty$. Figure 1(h) averages tailfree densities that have these conditional probabilities over the prior $\alpha, \gamma \overset{\text{ind.}}{\sim} N(4, 0.05^2)$, yielding a smooth mixture of tailfree densities.

Tailfree prior densities are essentially a weighted average between a parametric density and a histogram, with bin locations coming from the parametric density. The histogram takes the shape of the parametric density over bin intervals, and there are jumps at the bin endpoints as usual. By taking $\boldsymbol{\theta}$ to be random, as in Figure 1(h), the bin locations are "jittered" or shifted, and the resulting density is smoothed, and is in fact differentiable (Hanson 2006, Proposition 1). The resulting density model is similar to the "averaged shifted histogram" of Scott (1985). However, Scott's approach does not make use of a parametric family. The tailfree density has a pronounced nonparametric flavor where data are plentiful and unlike a Weibull density (e.g., multimodal), but retains the shape of the centering Weibull density where data are sparse and/or data approximately follow a Weibull distribution.

Define $\mathbf{p} = (p(1), \ldots, p(2^J))'$ to be the vector of random probabilities of the 2^J sets in the finest partition at level J. Pairs of conditional probabilities $\{(\pi(\boldsymbol{\epsilon}0), \pi(\boldsymbol{\epsilon}1))\}$ are assumed to be mutually independent, implying

$$p(l+1) = P\{T \in B_{\boldsymbol{\theta}}(\epsilon_1 \ldots \epsilon_J)\} = \prod_{i=1}^{J} \pi(\epsilon_1 \ldots \epsilon_i), \quad (5)$$

where $\epsilon_1 \ldots \epsilon_J$ is the base-2 representation of $l, l = 0, \ldots, 2^J - 1$. For example, say $J = 3$. Then to obtain $P\{T \in B_{\boldsymbol{\theta}}(110)\}$, one computes

$$\begin{aligned} P\{T \in B_{\boldsymbol{\theta}}(110)\} &= P\{T \in B_{\boldsymbol{\theta}}(110) | T \in B_{\boldsymbol{\theta}}(11)\} \\ &\quad \times P\{T \in B_{\boldsymbol{\theta}}(11) | T \in B_{\boldsymbol{\theta}}(1)\} P\{T \in B_{\boldsymbol{\theta}}(1)\} \\ &= \pi(110)\pi(11)\pi(1). \end{aligned}$$

We require the survival function $S(t) = 1 - F(t)$. Let $T \sim F$. For a given $t > 0$, let t_l and t_r be the left and right endpoints of the partition interval at level J that contains t. That is, $t_l < t < t_r$, where $t_l = G_{\boldsymbol{\theta}}^{-1}(m/2^J)$, $t_r = G_{\boldsymbol{\theta}}^{-1}((m+1)/2^J)$, and m is such that $G_{\boldsymbol{\theta}}^{-1}(m/2^J) < t < G_{\boldsymbol{\theta}}^{-1}((m+1)/2^J)$. Then $P(T > t) = P(t < T \le t_r) + P(T > t_r)$. Using (4) and (5), $P(t < T \le t_r) = p_m \int_t^{t_r} g_{\theta}(s) ds / \int_{t_l}^{t_r} g_{\theta}(s) ds = p_m [G_{\theta}(t_r) - G_{\theta}(t)]/2^{-J}$ and $P(T > t_r) = \sum_{j=s_{\theta}(t)+1}^{2^J} p(j)$, where $s_{\boldsymbol{\theta}}(t) = m = \lceil 2^J G_{\boldsymbol{\theta}}(t) \rceil$ and $\lceil \cdot \rceil$ is the ceiling function. These results imply that the survival function with respect to F is

$$S(t) = 1 - F(t) = p\{s_{\boldsymbol{\theta}}(t)\}\{s_{\boldsymbol{\theta}}(t) - 2^J G_{\boldsymbol{\theta}}(t)\} + \sum_{l=s_{\boldsymbol{\theta}}(t)+1}^{2^J} p(l), \quad (6)$$

where $p(l)$ is given by (5). By differentiating (6), the density with respect to F is given by

$$f(t) = \sum_{l=1}^{2^J} 2^J p(l) g_{\theta}(t) I_{B_{\theta}\{\epsilon_J(l-1)\}}(t) = 2^J p\{s_{\boldsymbol{\theta}}(t)\} g_{\boldsymbol{\theta}}(t), \quad (7)$$

where $\epsilon_J(i)$ is the binary representation $\epsilon_1 \ldots \epsilon_J$ of the integer i. Recall that Figure 1(b)–1(g) plots the density (7) centered at $G_{\boldsymbol{\theta}} =$Weibull(4,4), $J = 3$, for different sets of $\{\pi(\boldsymbol{\epsilon})\}$.

Now introduce the subscript k to make clear we are defining two tailfree processes F_k where $k = 0, 1$ for perfect and minimal repair, respectively. Let the random variable $\lambda_k(\boldsymbol{\epsilon}0)$ be the logit transformation of $\pi_k(\boldsymbol{\epsilon}0)$, that is,

$$\lambda_k(\boldsymbol{\epsilon}0) = \text{logit}\{\pi_k(\boldsymbol{\epsilon}0)\}. \quad (8)$$

The priors on $\{(\lambda_0(\boldsymbol{\epsilon}0), \lambda_1(\boldsymbol{\epsilon}0))\}$ are given by

$$\lambda_0(\boldsymbol{\epsilon}0), \lambda_1(\boldsymbol{\epsilon}0) \overset{\text{ind.}}{\sim} N\left(0, \frac{2}{c\rho(j)}\right), \quad (9)$$

where j is the number of digits in $\boldsymbol{\epsilon}0$. The $N(0, 2/c\rho(j))$ prior on $\lambda_k(\boldsymbol{\epsilon}0)$ mimics a beta$(c\rho(j), c\rho(j))$ prior for Polya tree conditional probabilities $\{\pi_k(\boldsymbol{\epsilon}0)\}$ (Jara and Hanson 2011). A common choice that we adopt is $\rho(j) = j^2$. The parameter c acts much like the precision in a Dirichlet process (Ferguson 1973). As $c \to 0^+$, $E\{F_k(\cdot)\}$ tends to the empirical CDF of the data (Hanson and Johnson 2002); as $c \to \infty$, all conditional probabilities $\pi_k(\boldsymbol{\epsilon})$ go to 0.5 and hence $F_k(t) \to G_{\boldsymbol{\theta}_k}(t)$ with probability one for all $t > 0$. We assign c a gamma prior $c \sim \Gamma(a, b)$; typically $a = 10$ or 5 and $b = 1$; motivation for these priors is provided in Hanson, Kottas, and Branscum (2008) using the prior

L_1 distance between F_k and $G_{\boldsymbol{\theta}_k}$. For $c \sim \Gamma(5,1)$, the median L_1 distance of the random tailfree density from the centering distribution is 0.28 with 95% probability interval (0.11,0.76); for $c \sim \Gamma(10,1)$ these values are 0.19 and (0.08,0.51). So $\Gamma(5,1)$ typically allows about 30% more mass to be moved than $\Gamma(10,1)$, as we would expect. The model under the alternative hypothesis H_1 is summarized in terms of interfailure times as

$$\begin{aligned} t_i - t_{i-1}|d_{i-1} = 0 &\overset{\text{ind.}}{\sim} F_0(\cdot), \\ t_i - t_i^*|d_{i-1} = 1 &\overset{\text{ind.}}{\sim} \frac{F_1(\cdot)}{S_1(t_{i-1} - t_i^*)}, \\ F_0|\boldsymbol{\theta}_0, c &\sim \mathrm{TF}^J(c, \rho, G_{\boldsymbol{\theta}_0}), \\ F_1|\boldsymbol{\theta}_1, c &\sim \mathrm{TF}^J(c, \rho, G_{\boldsymbol{\theta}_1}), \end{aligned}$$

where $\mathrm{TF}^J(c, \rho, G_{\boldsymbol{\theta}_k})$ is shorthand for the random tailfree F_k given through (6)–(9) up to level J. The model under null hypothesis H_0 simply replaces F_1 by F_0 and $\boldsymbol{\theta}_1$ by $\boldsymbol{\theta}_0$ above. The two models are referred to as M_1 and M_0 with respect to H_1 and H_0.

The model with a common Weibull centering distribution $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$ is a linear dependent tailfree process (Jara and Hanson 2011) regressed on a binary predictor (maintenance decisions), albeit with a likelihood involving truncated observations, for example, $F_1(\cdot)/S_1(t_{i-1} - t_i^*)$ for $d_{i-1} = 1$ under H_1. This model generalizes the Polya tree in the same spirit as De Iorio et al. (2004) generalized the celebrated Dirichlet process through an analysis of variance (ANOVA)-type structure. Under this model, the $e^{\lambda_1(\boldsymbol{\epsilon})-\lambda_0(\boldsymbol{\epsilon})}$ are interpreted as how the odds of failing in the time interval $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon})$ change from minimal to perfect repair; this information can be useful for finding time intervals $B_{\boldsymbol{\theta}}(\boldsymbol{\epsilon})$ where minimal repair fixes the problem in a manner substantially worse than F_0 would allow. Under the model where $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$, if each pair of $\lambda_0(\boldsymbol{\epsilon}0)$, $\lambda_1(\boldsymbol{\epsilon}0)$ are assigned identical and independent priors, then $E\{F_0(t)\} = E\{F_1(t)\} = G_{\boldsymbol{\theta}}(t)$ for all $t > 0$ and hence the null model M_0 is formally nested in the alternative model M_1. From many simulations (beyond what is included in this article), allowing distinct $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ increases discriminatory ability, but also inflates Type I error.

2.2 Testing H_0 Versus H_1

As stated in the introduction, the test for the assumption of "minimal repair" is of interest, and so it is important to choose a measure to compare the models. As H_0 is formally nested in H_1, a likelihood-ratio-type test could be considered, or a Bayes' factor (the Bayesian equivalent). However, computing the Bayes' factor with the truncated data likelihoods (2) and (3) is challenging and existing methods are unstable (Hanson 2006). Instead we consider an alternative measure, termed the log pseudo-marginal likelihood (LPML; Geisser and Eddy 1979), a measure of a model's predictive ability. The LPML is easy to compute based on MCMC output (Gelfand and Dey 1994). By definition,

$$\mathrm{LPML} = \sum_{i=1}^{n} \log\{f_i(t_i|\mathbf{t}_{-i})\}.$$

Here, $f_i(t_i|\mathbf{t}_{-i})$ is the predictive density for t_i based on the remaining data $\mathbf{t}_{-i} = \{t_j : j \neq i\}$, $f_i(\cdot|\mathbf{t}_{-i})$, evaluated at t_i. This is called the ith conditional predictive ordinate (CPO) statistic, and measures how well t_i is predicted from the remaining $\mathbf{t}_{-i}$ through the model. In our context, we compute the predictive density ($\delta_i = 1$) or survival ($\delta_i = 0$) at t_i based on the failure times, repair times, and repair decisions during $[0; t_{i-1}]$ and $[t_{i+1}, t_{\max}]$, plus partial information during (t_{i-1}, t_{i+1}) that a certain type of repair was performed at t_i. The LPML simply aggregates the log of these. The difference in LPML measures between H_1 and H_0 can be exponentiated giving the pseudo Bayes' factor BF_{10} for the two models. Common interpretations for Bayes' factors apply, for example, $3 < \mathrm{BF}_{10} < 20$ indicates "positive" evidence toward H_1; $20 < \mathrm{BF}_{10} < 150$ indicates "strong" evidence, and $\mathrm{BF}_{10} > 150$ indicates "very strong" evidence (Kass and Raftery 1995). Under mild conditions, the LPML converges to the posterior score and so the pseudo Bayes' factor is related to Aitkin's posterior Bayes' factor (Aitkin 1991) as well. In simulations, we find the LPML to work well in differentiating H_1 from H_0.

The LPML is approximated by

$$\mathrm{LPML} = -\sum_{i=1}^{n} \log\left\{\frac{1}{s}\sum_{k=1}^{s}\frac{1}{p_i(t_i|\mathcal{D}, \boldsymbol{\tau}^k)}\right\}, \tag{10}$$

where p_i is the likelihood contribution of event at time t_i, $\mathcal{D}$ is the observed data, and (2) and (3) define the model under H_0 and H_1, respectively, $\{\boldsymbol{\tau}^k,\ k = 1, 2, \ldots, s\}$ are iterates from MCMC outputs of all the parameters, that is, $\{\boldsymbol{\lambda}_0^k, \boldsymbol{\lambda}_1^k, c^k,\ k = 1, 2, \ldots, s\}$ under H_1.

2.3 Model Fitting

Following the discussion at the end of Section 2.1, for the purposes of testing $H_0 : F_0 = F_1$ we suggest that F_0 and F_1 have the same prior mean Weibull distribution in fitting M_1, for example, $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \boldsymbol{\theta}$. In simulations, we fix $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \widehat{\boldsymbol{\theta}}$, where $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimate (MLE) assuming minimal repair holds, and Weibull reliability, that is, $F_0(t) = F_1(t) = G_{\boldsymbol{\theta}}(t)$ (the Weibull is obtained under the tailfree prior when $c \to \infty$). A similar practice is recommended by Berger and Guglielmi (2001) and Hanson, Branscum, and Gardner (2008) in simpler situations.

Upon rejecting H_0, we suggest refitting M_1, allowing distinct $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. It is known in the literature that fixing $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ results in "jumpy" densities as each of f_0 and f_1 has discontinuities at each partition interval endpoint, as Figure 1 shows. For the purposes of estimating the interfailure hazard functions $\hat{h}_0(t)$ and $\hat{h}_1(t)$, we suggest an empirical Bayes' approach: place normal priors on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ derived from their large-sample asymptotic distributions under the underlying Weibull assumption ($c \to \infty$):

$$\boldsymbol{\theta}_k \overset{\text{ind.}}{\sim} N_2(\widehat{\boldsymbol{\theta}}_k, \boldsymbol{\Sigma}_k), \tag{11}$$

where $\boldsymbol{\Sigma}_k$ is the large-sample covariance matrix under a frequentist Weibull fit ($\widehat{\boldsymbol{\theta}}_k$ and $\boldsymbol{\Sigma}_k$ are easily obtained using optimization procedures in R or SAS); noninformative priors ($p(\boldsymbol{\theta}_k) \propto 1$) for $\boldsymbol{\theta}_k$ can also be used. Placing priors on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ smooths out the estimated density and hazard curves, yielding a mixture of tailfree processes for F_0 and F_1.

MCMC computing requires full specification of the likelihoods and priors. The likelihoods (2) and (3) under H_0 and H_1 are functions of f_0 and (f_0, f_1), respectively. Let $\mathcal{E} = \{\boldsymbol{\epsilon} = \epsilon_1 \ldots \epsilon_j, j = 1, \ldots, J-1\}$. Conditioning on $\{\boldsymbol{\pi}_k(\boldsymbol{\epsilon}0)\}$ for $\boldsymbol{\epsilon} \in \mathcal{E}$, the densities ($f_0$ and f_1) and reliability functions (S_0 and S_1) are given by (7) and (6) in terms of probability vectors $\mathbf{p}_k$, functions of $\boldsymbol{\pi}_k(\boldsymbol{\epsilon}0)$ defined in (5). Note that $\boldsymbol{\pi}_k(\boldsymbol{\epsilon}0)$ is a function of $\boldsymbol{\lambda}_k$ through (8). The posterior under H_1 is proportional to

$$p(\boldsymbol{\lambda}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, c|\mathcal{D}) \propto \mathcal{L}(f_0, f_1) p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) \Gamma(c|a, b) \times \prod_{k=0}^{1} \prod_{\boldsymbol{\epsilon} \in \mathcal{E}} N\left(\lambda_k(\boldsymbol{\epsilon}0)|0, \frac{2}{cj^2}\right),$$

where $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_0(\boldsymbol{\epsilon}0), \boldsymbol{\lambda}_1(\boldsymbol{\epsilon}0)\}$, $\mathcal{L}(f_0, f_1)$ is defined in Equation (3) and $p(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$ are product of independent priors for $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. The posterior under H_0 is similar.

Parameters $\{\boldsymbol{\lambda}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$ are updated using random-walk Metropolis–Hastings updates (Tierney 1994). Gaussian random-walk proposals are used for each element of $\{\lambda_k(\boldsymbol{\epsilon}0) : k = 0, 1; \boldsymbol{\epsilon} \in \mathcal{E}\}$,

$$\lambda_k(\boldsymbol{\epsilon}0)^* \sim N(\lambda_k(\boldsymbol{\epsilon}0), v_k(\boldsymbol{\epsilon}0)),$$

where $\lambda_k(\boldsymbol{\epsilon}0)^*$ is the latest accepted value for $\lambda_k(\boldsymbol{\epsilon}0)$, and $v_k(\boldsymbol{\epsilon}0)$ is tuned to get acceptance rates in the 20%–50% range. Similarly, $\boldsymbol{\theta}_k \sim N_2(\boldsymbol{\theta}_k^*, V_k)$, where V_k needs to be tuned. We have found automatic tuning of $v_k(\boldsymbol{\epsilon})$ and V_k to proceed quickly (Haario, Saksman, and Tamminen 2005). Specifically, let the sequence $\lambda_k^1(\boldsymbol{\epsilon}0), \lambda_k^2(\boldsymbol{\epsilon}0), \ldots$ be the states of the Markov chain for $\lambda_k(\boldsymbol{\epsilon}0)$. When deciding the tth state $\lambda_k^t(\boldsymbol{\epsilon}0)$, we sample $\lambda_k(\boldsymbol{\epsilon}0)^* \sim N(\lambda_k^{t-1}(\boldsymbol{\epsilon}0), v_k^t(\boldsymbol{\epsilon}0))$ with

$$v_k^t(\boldsymbol{\epsilon}0) = \begin{cases} v^0(\boldsymbol{\epsilon}0), & t < t_0 \\ s\,\mathrm{var}\{\lambda_k^1(\boldsymbol{\epsilon}0), \ldots, \lambda_k^{t-1}(\boldsymbol{\epsilon}0)\} + s_0, & t > t_0 \end{cases},$$

where s is recommended to be 2.4, s_0 is a small constant, and $v_k^0(\boldsymbol{\epsilon}0)$ is an initial variance of the proposal distribution. A similar automatic procedure applies to $\boldsymbol{\theta}_k$ with V_k^t being the empirical covariance matrix after t_0. The parameter c is updated through posterior

$$p(c|\boldsymbol{\lambda}, \boldsymbol{\theta}, \mathcal{D}) \sim \Gamma\left\{(a + 2^J - 1), b + \sum_{\epsilon_1\epsilon_2\ldots\epsilon_j \in \mathcal{E}} \sum_{k=0}^{1} \boldsymbol{\lambda}_k(\epsilon_1\epsilon_2\ldots\epsilon_j)^2 j^2/4\right\}.$$

FORTRAN 90 code for fitting the data analysis in Section 4 is included in the online supplementary material for this article.

3. SIMULATIONS

We conducted four simulations to see how well the pseudo Bayes' factor can discriminate between H_0 and H_1 and one simulation to illustrate the estimation of the reliability functions. Simulation I involves a sequence of increasing departures of f_1 from f_0 according to our alternative model M_1. Simulation II considers a sequence of departures from H_0 using the effective age models. Simulation III investigates Type I error. Simulation IV examines how the prior on c affects the test. Simulation V estimates reliability functions F_0 and F_1. Sample sizes for simulated data are the total number of interfailure times after perfect or minimal repair. Each dataset is comprised of one-third interfailure times after perfect repair and two-thirds interfailure times after minimal repair truncated from the accumulated age since the most recent perfect repair. For simplicity, all repairs occur in response to failures. For the hypothesis tests, the unknown distributions are assigned finite tailfree priors with the following specifications. We fix $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ for F_0 under M_0 and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \widehat{\boldsymbol{\theta}}$ for F_0, F_1 under M_1, where $\widehat{\boldsymbol{\theta}}$ is an estimate for $\boldsymbol{\theta}$ under the Weibull null model; $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ each contain the log of the Weibull shape and scale parameters; the level of the partition tree is fixed at $J = 5$, and c is considered with prior $\Gamma(5, 1)$ for simulations I–III and priors $\Gamma(5, 1)$ and $\Gamma(10, 1)$ for simulation IV. Based on our simulation experience, $J = 5$ is sufficient for the sample sizes in our simulations and increasing J changes the LPML negligibly. For each dataset, we run 4000 MCMC iterations and use the last 3000 MCMC samples for inferences. We reject H_0 in favor of H_1 if the LPML for H_1 is greater than that for H_0 by 3.5, otherwise we choose H_0. For estimating the reliability functions in simulation V, we place the empirical Bayes' priors as detailed in Section 2.3 on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, and assume $c \sim \Gamma(10, 1)$. After a burn-in of 10,000 iterates, 400,000 iterates were thinned to a sample of 4000 for inference.

Simulation I. Let $W(w, \alpha_1, \gamma_1, \alpha_2, \gamma_2)$ be a mixture of two Weibull distributions with weights w and $1 - w$, shape parameters α_1, α_2, and scale parameters γ_1, γ_2. Table 2 reports the results of simulation I testing H_0 versus increasing departures H_1 according to our alternative model; $f_0 = W(0.5, 2, 3, 2, 6)$, $f_1 = W(0.5, 2, 3, 2, \gamma_2)$ with nine densities corresponding to $\gamma_2 = \{6, 5.5, 5, 4.5, 4, 3.5, 3, 2.5, 2\}$. The simulation involves two sample sizes $n = 200$ or $n = 500$. For each condition, 200 datasets are simulated. The hazard functions for f_0 (solid thick line) and f_1 (dashed thin lines) are plotted in the left panel of Figure 2; Table 2 values are proportions rejecting H_0. The power is reasonably good in detecting this sequence of departures.

Simulation II. We use the notion of imperfect repairs and effective age to introduce increasing departures from the minimal repair assumption but using our proposed models M_0 and M_1 for hypothesis testing. Effective age modeling in reliability has received a lot of attention since introduced by Kijima (1989). A spectrum of imperfect repairs can be modeled through effective ages, generating event processes that include renewal processes and Poisson processes as special cases. Following the notation introduced in Section 2.1, the times for repairs are recorded as $0 = t_0 < t_1 < t_2 < \cdots < t_n = t_{\max}$. The repair at t_i is denoted as d_i with $d_i = 0$ if perfect repair was performed and $d_i = 1$ otherwise. Define $z(t)$ as the effective age of the system at time t. We still assume perfect repairs reset the effective age to zero but now

Table 2. Type I error and power for testing H_0 versus H_1 for simulation I; 6.0–2.0 are nine values of γ_2 in defining f_1; tabled values are the proportion out of 200 replications where H_0 is rejected

γ_2	6.0	5.5	5.0	4.5	4.0	3.5	3.0	2.5	2.0
Sample event									
$n = 200$	0.00	0.01	0.04	0.14	0.39	0.58	0.85	0.94	0.96
$n = 500$	0.04	0.05	0.14	0.50	0.93	1.00	1.00	1.00	1.00

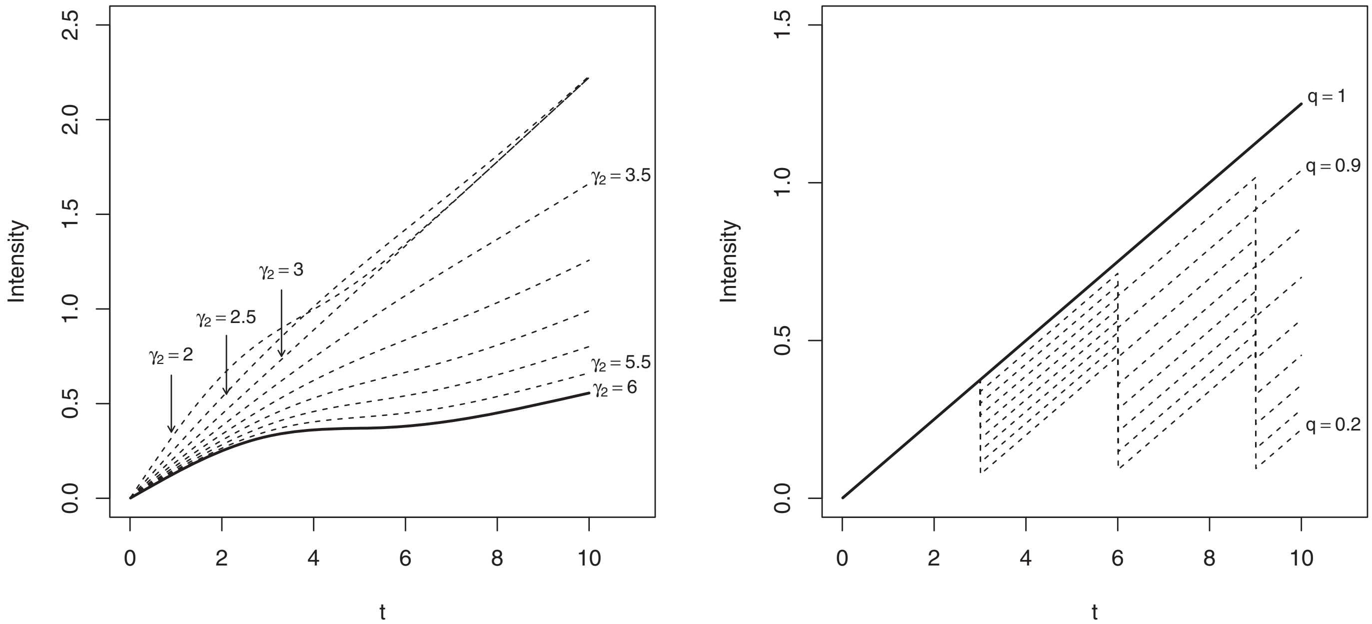

Figure 2. Left panel plots the hazard h_0 (solid and thick line) and eight choice for h_1 (dashed and thin lines) versus time t for simulation I; right panel plots the intensity of the system versus time t when failures occur at $\{3, 6, 9\}$ for all q from 1 (solid and thick line) and 0.2–0.9 (dashed and thin lines) for simulation II.

repairs recorded as $d_i = 1$ multiply the effective age right before the repair by a fraction q (known as Kijima Type II model). That is, $z(t_i) = 0$ if $d_i = 0$, $z(t_i) = \{z(t_{i-1}) + t_i - t_{i-1}\}q$ if $d_i = 1$ and $z(t) = z(t_i) + t - t_i$ for $t_i < t < t_{i+1}$. This is only one departure from the null model M_0 that is different from our alternative model M_1 (Presnell, Hollander, and Sethuraman 1994). Note that $q = 1$ implies that the minimal repair assumption holds and $q < 1$ indicates repairs being better than "good as old." Suppose F_0 defines the CDF of the first failure time for a system; after repair at t_i, the distribution for the time to next failure is $F_0(z(t_i) + t)/S(z(t_i))$, $t > 0$. In the following simulation, we examine the power of the proposed test for a sequence of q using two sample sizes $n = 200$ or $n = 500$. For each condition, 200 datasets are simulated.

Table 3 reports the results of testing H_0 versus H_1; q takes values from 1 to 0.2 by 0.1; $f_0 = \text{Weibull}(2, 4)$; the intensities for a system with failures at $\{3, 6, 9\}$ are plotted in the right panel of Figure 2; the solid thick line corresponds to $q = 1$ and dashed lines correspond to $q < 1$ from 0.9 to 0.2; the tabled values are percentages of times rejecting H_0. Even though the Kijima departure is not in the realm of our model, our test performs satisfactorily with power increasing to one as q gets small for $n = 500$. According to additional simulations, not included here, the power of our test increases when the slope of the hazard increases in the Kijima models.

Simulation III. We perform a simulation to investigate Type I error using three sample sizes $n \in \{1000, 1500, 2000\}$, and three choices for f_0. Table 4 reports the results of the third simulation based on 200 datasets where data are simulated only from M_0; the three densities are $W(0.5, 2, 3, 2, 6)$, Weibull$(2, 4)$, and Weibull$(1, 4)$; the tabled values are proportions rejecting H_0. The Type I errors appear to be stable and are always less than 0.05 for these distributions and sample sizes. Tables 1 and 2 show that the LPML cutoff of 3.5 may be conservative for smaller sample sizes.

Simulation IV. We also conduct a simulation to see how the prior on c affects the test. One hundred different datasets for each of three sample sizes $n = 200, 500, 1000$ (300 datasets total) were generated from model M_0 with $f_0 = W(0.5, 2, 3, 2, 5)$ where the minimal repair assumption holds, as well as model M_1 where F_0 and F_1 are different with $f_0 = W(0.2, 2, 0.7, 2, 5)$, $f_1 = W(0.5, 2, 0.7, 2, 3)$. Now c is considered with two priors, $\Gamma(5, 1)$ and $\Gamma(10, 1)$. Table 5 reports the results of testing H_0 versus H_1; the tabled values are percentages of times rejecting H_0. The $\Gamma(5, 1)$ prior favors

Table 3. Type I error and power for testing H_0 versus H_1 for simulation II; 0.2–1.0 represent nine choices of q and 1.0 represents no departure of minimal repair; tabled values are the proportion out of 200 replications where H_0 is rejected

q	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
Sample event									
$n = 200$	0.01	0.02	0.03	0.04	0.11	0.20	0.29	0.53	0.72
$n = 500$	0.03	0.04	0.06	0.21	0.53	0.81	0.94	0.99	1.00

Table 4. Type I error for testing H_0 versus H_1 for simulation III; 1–3 represents three choices of f_0 described in the text

Density type	1	2	3
Sample event			
$n = 1000$	0.03	0.04	0.04
$n = 1500$	0.03	0.05	0.03
$n = 2000$	0.04	0.05	0.04

Table 5. Type I error and power for testing H_0 versus H_1 with two sets of prior for c for simulation IV; tabled values are the proportion out of 100 simulated datasets where H_0 is rejected

Sample event	$H_0: F_0 = F_1$ $c \sim \Gamma(5, 1)$	$H_0: F_0 = F_1$ $c \sim \Gamma(10, 1)$	$H_1: F_0 \neq F_1$ $c \sim \Gamma(5, 1)$	$H_1: F_0 \neq F_1$ $c \sim \Gamma(10, 1)$
$n = 200$	0.00	0.00	0.75	0.60
$n = 500$	0.00	0.00	0.81	0.70
$n = 1000$	0.05	0.02	1.00	1.00

smaller values of c, yielding more modeling flexibility, and hence increasing differentiability. This effect is more obvious for smaller sample sizes.

Simulation V. The last simulation illustrates our approach by estimating reliability functions after perfect and minimal repairs for three simulated datasets. One dataset of 1000 events was simulated from the above setting of M_1. Two datasets of 1000 events were simulated from the Kijima Type II effective age model with $q = 0.2, 0.5$ and $f_0 = \text{Weibull}(2, 4)$. For the first dataset, the true reliability functions S_0 and S_1, and hazards h_0 and h_1 are displayed in the left panels of Figure 3 with F_0 plotted using solid lines and F_1 plotted using short-dashed lines. The estimated survival and hazard (pointwise posterior means) are plotted in the right panels of Figure 3, along with the 95% credible intervals for the estimates (long-dashed lines); we can see that local features of the distributions are well captured. For the other two datasets, we plot the estimated hazard functions for h_0 (solid black) and h_1 (dashed black) in Figure 4, overlaid with intensities (dashed gray) of 10 systems. We plot each intensity function over time since the first minimal repair for the corresponding system. For both datasets, our model's h_1 estimates essentially average the true intensities of the 10 systems. We can see a larger difference between h_1 and h_0 estimates for $q = 0.2$ than that for $q = 0.5$, indicating better performances of system when $q = 0.2$.

The computing time for running the above simulations mainly depends on the number of MCMC iterates, sample size, and the level of the partition tree J. For $J = 5$, 4000 MCMC iterates, and a 3.00 GHz processor, it may take a few seconds for small

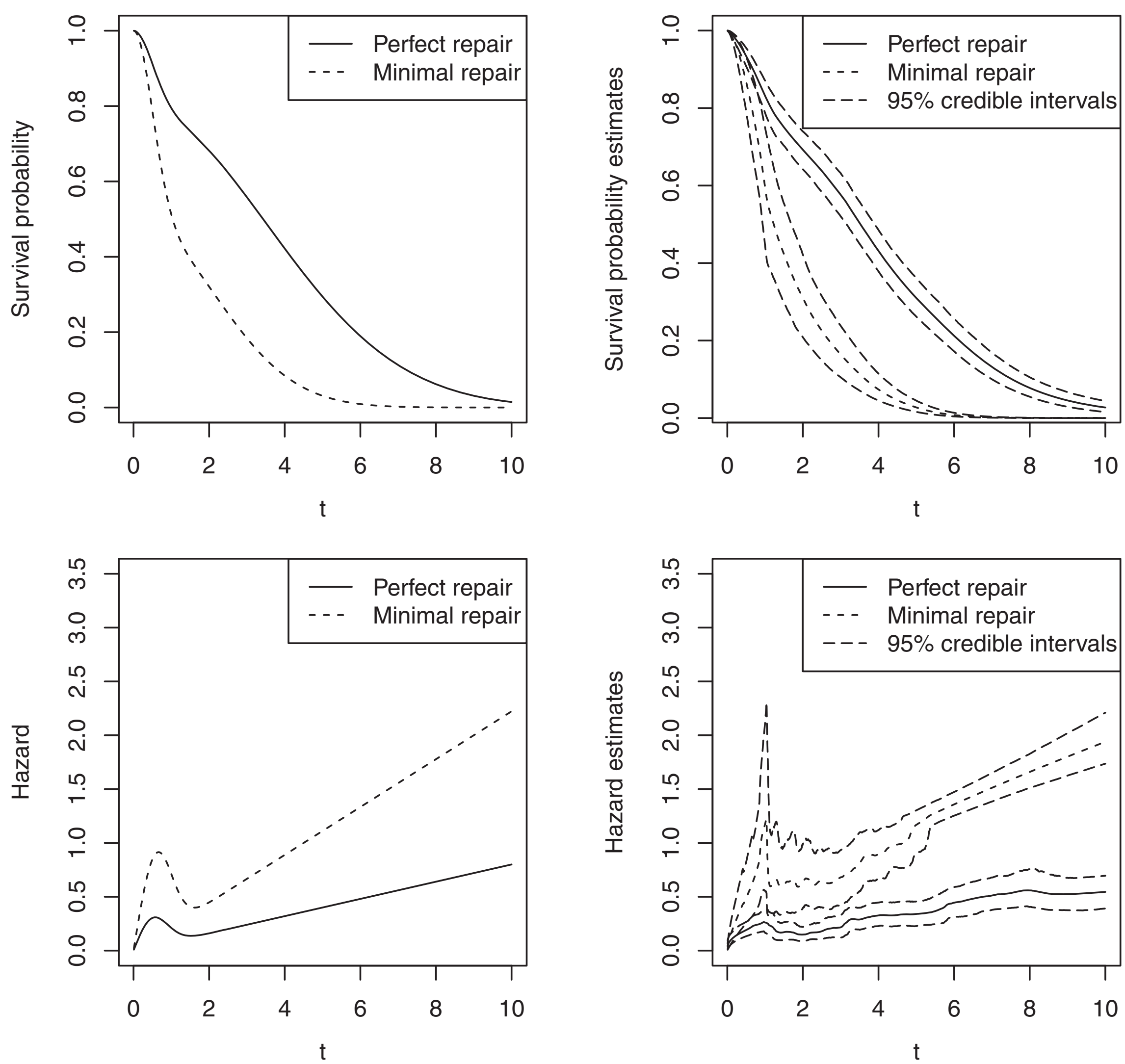

Figure 3. Results of a simulated sample of $n = 1000$ events under H_1; true (left) and estimated (right) survival and hazard estimates versus time t; solid lines correspond to F_0 and short-dashed lines correspond to F_1; long-dashed lines correspond to 95% credible intervals.

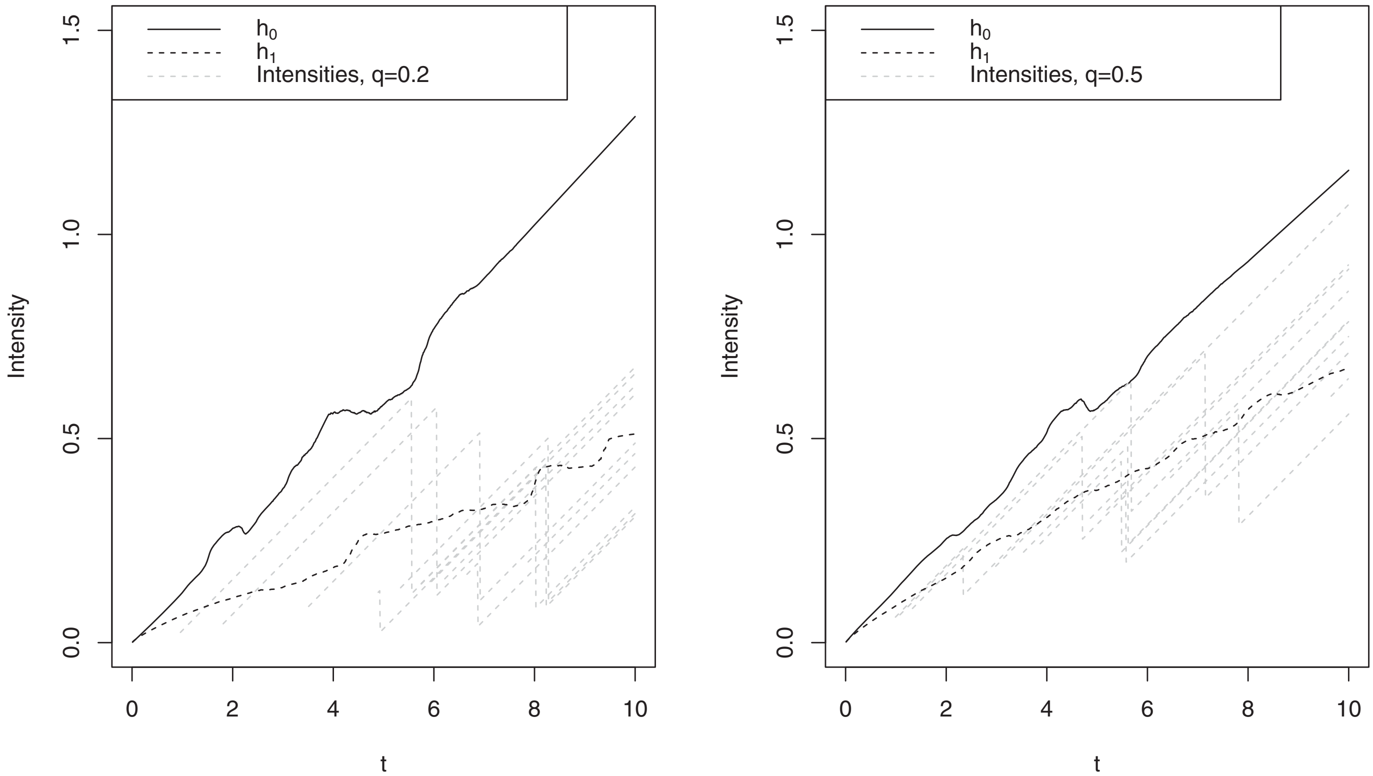

Figure 4. Results of two simulated samples of $n = 1000$ interfailure times under Kijima Type II model with $q = 0.2$ (left) and 0.5 (right); hazard estimates of h_0 (solid black) and h_1 (dashed black) versus time t; dashed gray lines are intensities of 10 systems.

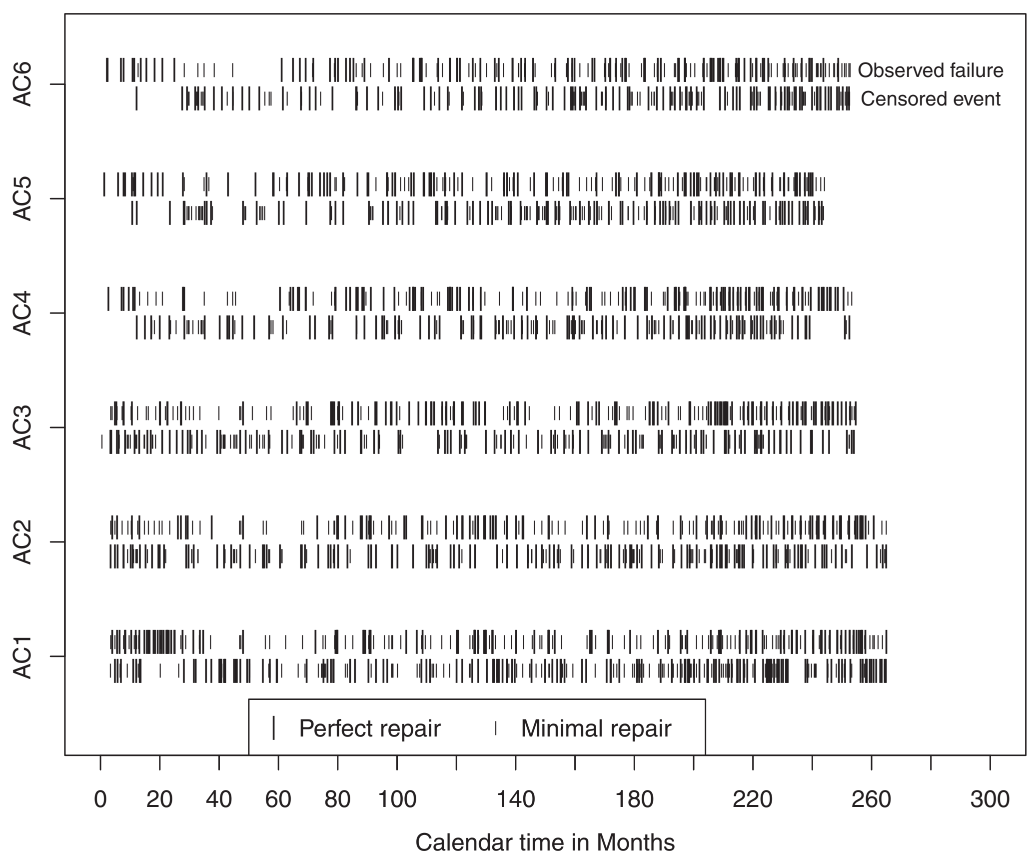

Figure 5. Calendar times of events for the six chillers (AC); for each chiller, vertical bars represent observed failures on the top line and censored events on the bottom line; big (small) vertical bar denotes perfect (minimal) repair at the event time.

sample sizes (e.g., 200, 500) to a couple of minutes for large sample sizes (i.e., 2000). The computing times for longer chains or higher levels of J are longer.

4. DATA ANALYSIS

We studied the dataset provided by South Texas Project Electric Generating Station for the essential chiller system. Details on the chillers and maintenances are presented in the introduction. The original calendar time was recorded in days and we divided the time by 30.4 to transform the units to months. Calendar times of maintenance events in months are plotted in Figure 5 by chillers where vertical bars represent censored failure times on the top line and observed failure times on the bottom line. The size of the vertical bar indicates perfect or minimal repair and the number of each type of repair per chiller is given in Table 6. We first investigate whether the six chillers are identical in new condition. Chillers (AC 4–6) in group 2 tend to last longer than Chillers (AC 1–3) in group 1 based on the Kaplan–Meier estimates (Figure 6) using the first failures after perfect repairs.

Table 6. Counts of perfect/minimal by response to "failure"/"censored" for each chiller

	AC1		AC2		AC3	
	Failure	Censored	Failure	Censored	Failure	Censored
Perfect	17	217	18	199	15	202
Minimal	232	4	179	1	184	3
	AC4		AC5		AC6	
	Failure	Censored	Failure	Censored	Failure	Censored
Perfect	14	189	9	184	13	184
Minimal	160	1	167	2	163	3

Within each group, there appears no significant difference. The log-rank tests for homogeneity of survival curves for the first failures give a significant p-value of 0.03 across the two groups, but nonsignificant 0.26 for chillers within the first group, and 0.78 for chillers within the second group. Therefore, we pool

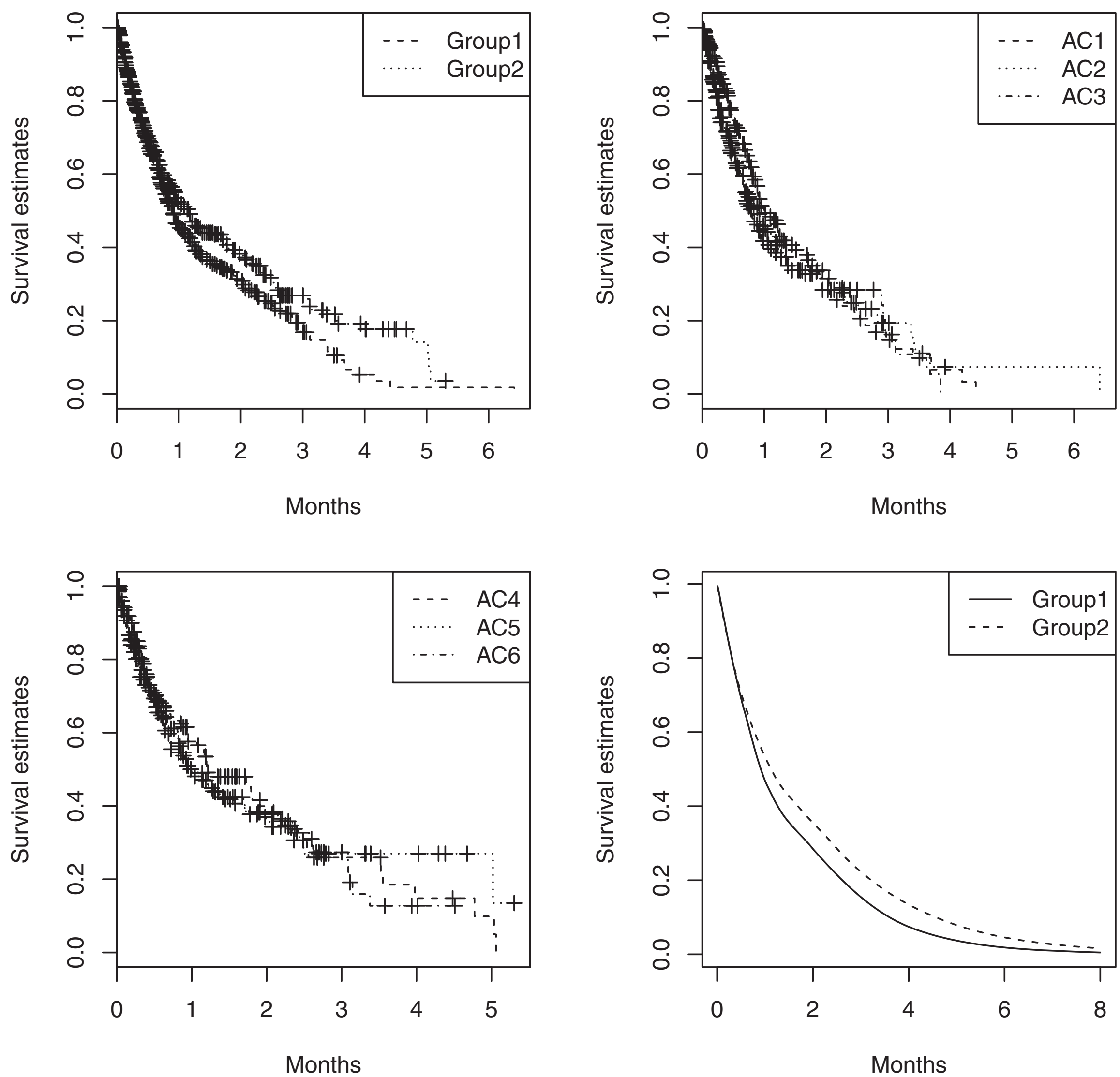

Figure 6. Kaplan–Meier estimates for survival functions using first failures after perfect repairs of the essential chillers system; groups 1 and 2 (top left); AC 1–3 in group 1 (top right); AC 4– 6 in group 2 (bottom left); bottom right panel plots M_1 estimates for S_0 for groups 1 and 2.

observations across chillers within each group and perform separate analyses for the two groups. From now on, the first group is referred to as "group 1" and the second group as "group 2."

It is of interest to test the minimal repair assumption, that is, whether there is a significant difference between the reliability distributions for the two types of maintenance decisions. We first fit the proposed nonparametric test. For group 1, the LPML for H_0 is -770 and for H_1 is -755. For group 2, the LPML for H_0 is -713 and for H_1 is -695. Exponentiating the LPML differences in the two groups (> 150) leads to strongly rejecting H_0 in both. We also fit parametric tests with Weibull family assumption for F_0 and F_1. For group 1, the LPML for H_0 is -770 and for H_1 is -760. For group 2, the LPML for H_0 is -714 and for H_1 is -694. Note that the nonparametric method yields greater difference in LPMLs for group 1 than parametric method does. For group 2, there is not a significant difference between the parametric and nonparametric method. For estimating F_0 and F_1, we refit M_1 using the nonparametric method presented in Section 2.3, place noninformative priors on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ ($p(\boldsymbol{\theta}_k) \propto 1$), and assume $c \sim \Gamma(10, 1)$ for the two groups. After a burn-in of 50,000 iterates, 4000 MCMC samples were thinned from a total of 400,000 iterates. The computing time was about 30 sec for hypotheses testing and a few minutes for estimation in M_1. We plot the estimated pointwise posterior mean survival functions for F_0 (solid lines) and F_1 (short-dashed lines) on the left panel of Figure 7. The 95% credible intervals for the survival functions are plotted with long-dashed lines. The right panel of Figure 7 is the estimated pointwise posterior mean hazard functions for F_0 (solid lines) and F_1 (short-dashed lines) from both the nonparametric (less smooth) and parametric (smooth) approach. The nonparametric estimates for h_0 and h_1 for group 1 (top right panel) are close to each other in the first 1 month, but after h_1 is larger than h_0. This implies that the system performs "good as old" after minimal repairs at a young cumulative age but worse at an older cumulative age (the cumulative age is the time since the latest perfect repair). The posterior mean estimate for h_1 is always greater than h_0 for group 2 (bottom right panel) indicating the system

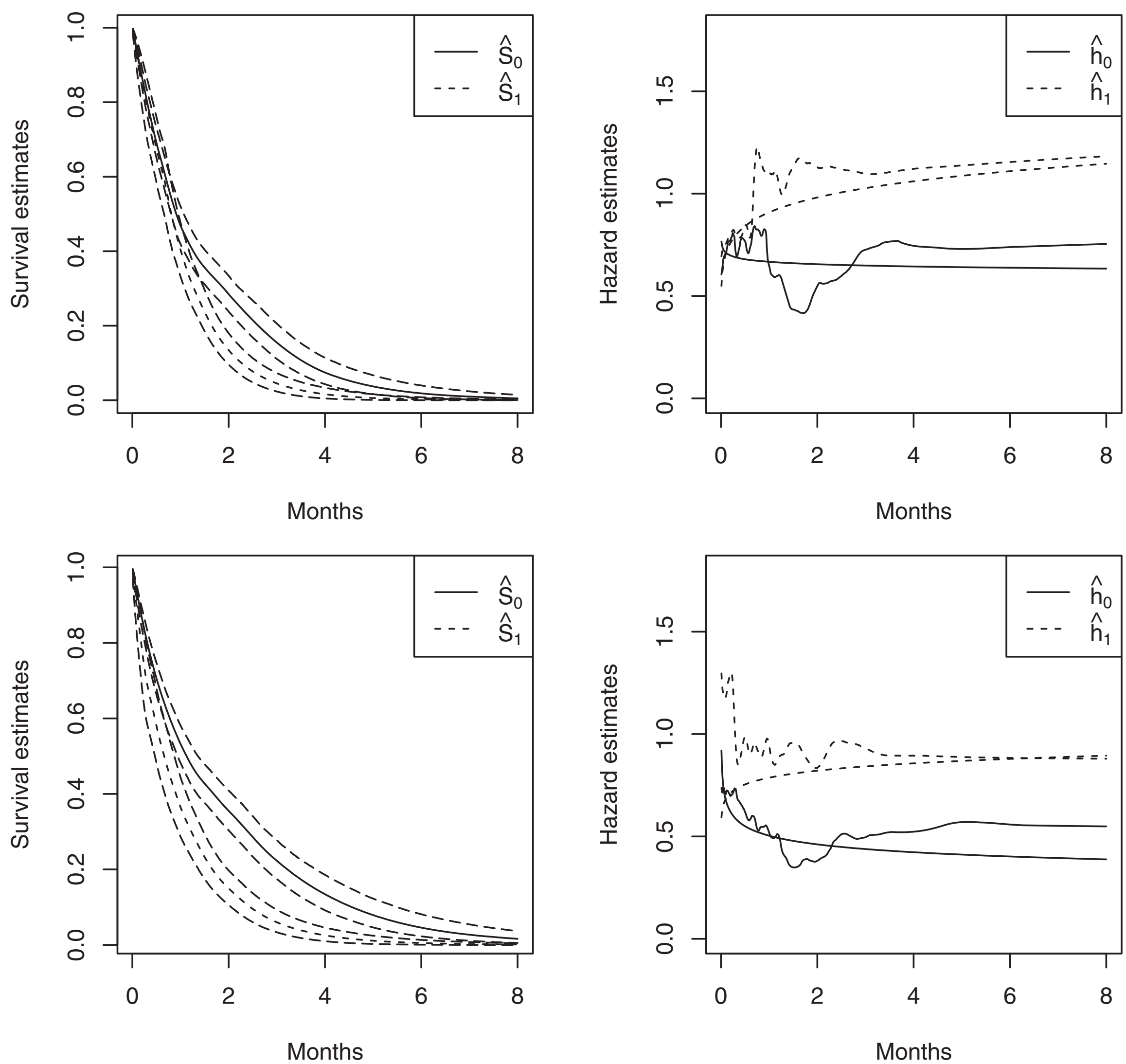

Figure 7. This figure contains estimates of the survivor and hazard functions for groups 1 (top panels) and 2 (bottom panels) essential chiller system when both parametric and nonparametric models are fitted for M_1. Left panels plot nonparametric estimates of the survivor functions corresponding to F_0 (solid) and F_1 (short-dashed) and their 95% credible intervals (long-dashed). Right panels plot the parametric (smooth) and nonparametric (less smooth) estimates of the hazard functions corresponding to F_0 (solid) and F_1 (short-dashed).

performs worse than expected after minimal repairs. Compared to the estimates of h_0 and h_1 from the parametric method, the nonparametric estimates exhibit much more flexibility in the shape during the first few months where data are more plentiful, but follow a Weibull shape as time increases and data are scarcer.

5. DISCUSSION

We proposed a flexible Bayesian nonparametric framework to model recurrent events in a repairable system for the purpose of generalizing and testing the common "minimal repair" assumption. Upon system failure either a perfect or a minimal repair is performed. Tailfree priors are assumed for the unknown distributions F_0 and F_1 centered at the Weibull distribution. The Weibull serves to anchor inference and guide density shape where data are scarce, but tailfree probabilities change the Weibull shape when necessary in locations where data are plentiful. The typical assumption that a minimal repair brings the system back to the exact state it was in right before failure is tested via pseudo Bayes' factors. In simulations, the test was found to have good power, and appropriate Type I error. If the alternative model $H_1 : F_0 \neq F_1$ is preferred, we further compare the estimated hazard functions, shedding light on how minimal repairs perform relative to perfect repairs at different ages of the system. This is particularly useful for managers to schedule maintenance. If the null model is preferred, our model becomes a Bayesian nonparametric generalization of Weibull for modeling the failure times from nonhomogenous Poisson processes. It is then typically of interest to obtain smooth estimates for the density, hazard, and survival function. With slight changes in the likelihood, our method can also be used to test other repair assumptions, for example, a known life supplement for a type of repair. We note that it is straightforward to include time-dependent covariates into the model, such as operating settings and the identity of the person making repairs.

We stress that perfect repairs are indeed assumed to bring the system to as "good as new." In practice, there may be several types of maintenances pooled together, which are close to "perfect repairs." If several identical systems are maintained in the same way, maintenance records may be combined since identical systems have the same contribution as old systems that have just received an overhaul. However, combining the records from systems, which are very different, could result in confounding between the actual effects of the maintenances and the reliability of the system.

We also assume that the system after each minimal repair depends on the preceding minimal repairs only through the accumulated age (time since last perfect repair). The minimal repairs are "good as old" repairs with respect to F_1. That is, the hazard function remains $h_1(t)$ over time after the first minimal repair in each cycle. This simplification facilitates the testing of H_0 versus H_1 and also allows comparison of maintenance decisions over time. However, if $H_1 : F_0 \neq F_1$ is concluded, the estimated h_1 may not be the dynamic hazard for the system after minimal repairs. The reason is that when minimal repairs are not "good as old" repairs, the actual effects of maintenances could aggregate, changing the hazard function after each repair. The "effective age" modeling of the system (Kijima 1989) captures a dynamically changing hazard. However, there is difficulty in determining the degree of each repair and hence the effective age for the system. It is one of our interests to model the dynamic hazard using Bayesian nonparametric methods.

SUPPLEMENTARY MATERIALS

Data and code: The dataset studied in Section 4, with R code for obtaining Weibull MLEs and fortran code implementing the MCMC algorithm (zip file).

ACKNOWLEDGMENTS

The work of the first and second authors was supported by NSF grant CMMI-0855329. The work of the third and fourth authors has been partially supported by NSF grant CMMI-0855577 and STPNOC grant B02857. The authors are grateful to both referees, the AE, and the editor for constructive comments that greatly improved the readability of this article.

[Received January 2012. Revised August 2013.]

REFERENCES

Aitkin, M. (1991), "Posterior Bayes Factors," *Journal of the Royal Statistical Society,* Series B, 53, 111–142. [398]

Berger, J. O., and Guglielmi, A. (2001), "Bayesian and Conditional Frequentist Testing of a Parametric Model Versus Nonparametric Alternatives," *Journal of the American Statistical Association*, 96, 174–184. [398]

Block, H. W., Borges, W. S., and Savits, T. H. (1985), "Age–Dependent Minimal Repair," *Journal of Applied Probability*, 22, 370–385. [393]

Cook, R. J., and Lawless, J. F. (2007), *The Statistical Analysis of Recurrent Events*, New York: Springer. [395]

Cooper, W. L., de Mello, T. H., and Kleywegt, A. J. (2006), "Models for the Spiral–Down Effect in Revenue Management," *Operations Research*, 54, 968–987. [393]

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205–215. [398]

Dorado, C., Hollander, M., and Sethuraman, J. (1997), "Nonparametric Estimation for a General Repair Model," *The Annals of Statistics*, 25, 1140–1160. [393]

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 1301–1312. [394]

Fabius, J. (1964), "Asymptotic Behavior of Bayes Estimates," *Annals of Mathematical Statistics*, 35, 846–856. [394,395,397]

Ferguson, T. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230. [394,397]

——— (1974), "Prior Distribution on the Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629. [394]

Ferguson, T., and Phadia, E. G. (1979), "Bayesian Nonparametric Estimation Based on Censored Data," *The Annals of Statistics*, 7, 163–186. [394]

Freedman, D. (1963), "On the Asymptotic Distribution of Bayes' Estimates in the Discrete Case," *Annals of Mathematical Statistics*, 34, 1386–1403. [395]

Geisser, S., and Eddy, W. (1979), "A Predictive Approach to Model Selection," *Journal of the American Statistical Association*, 74, 153–160. [398]

Gelfand, A. E., and Dey, D. K. (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society,* Series B, 56, 501–514. [398]

Haario, H., Saksman, E., and Tamminen, J. (2005), "Componentwise Adaptation for High Dimensional MCMC," *Computational Statistics*, 20, 265–273. [399]

Hanson, T. (2006), "Inference for Mixtures of Finite Polya Tree Models," *Journal of the American Statistical Association*, 101, 1548–1565. [397,398]

Hanson, T., Branscum, A., and Gardner, I. (2008), "Multivariate Mixtures of Polya Trees for Modelling ROC Data," *Statistical Modelling*, 8, 81–96. [398]

Hanson, T., and Johnson, W. (2002), "Modeling Regression Error With a Mixture of Polya Trees," *Journal of the American Statistical Association*, 97, 1020–1033. [397]

Hanson, T., Kottas, A., and Branscum, A. (2008), "Modelling Stochastic Order in the Analysis of Receiver Operating Characteristic Data: Bayesian

Nonparametric Approaches," *Journal of the Royal Statistical Society,* Series C, 57, 207–225. [397]
Ishwaran, H., and James, L. F. (2004), "Computational Methods for Multiplicative Intensity Models Using Weighted Gamma Processes," *Journal of the American Statistical Association*, 99, 175–190. [394]
Jara, A., and Hanson, T. (2011), "A Class of Mixtures of Dependent Tailfree Processes," *Biometrika*, 98, 553–566. [394,397,398]
Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795. [398]
Kijima, M. (1989), "Some Results for Repairable Systems With General Repair," *Journal of Applied Probability*, 26, 89–102. [393,399,405]
Kuo, L., and Ghosh, S. K. (1997), "Bayesian Nonparametric Inference for Nonhomogeneous Poisson Processes," Technical Report 9718, Department of Statistics, University of Connecticut. [394]
Lavine, M. (1992), "Some Aspects of Polya Tree Distributions for Statistical Modeling," *The Annals of Statistics*, 20, 1222–1235. [394,396]
Lo, A. Y. (1992), "Bayesian Inference for Poisson Process Models With Censored Data," *Journal of Nonparametric Statistics*, 2, 71–80. [394]
Mauldin, R. D., Sudderth, W. D., and Williams, S. C. (1992), "Polya Trees and Random Distributions," *The Annals of Statistics*, 20, 1203–1221. [397]
Presnell, B., Hollander, M., and Sethuraman, J. (1994), "Testing the Minimal Repair Assumption in an Imperfect Repair Model," *Journal of the American Statistical Association*, 89, 289–297. [393,400]
Scott, D. W. (1985), "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions," *The Annals of Statistics*, 13, 1024–1040. [397]
Taddy, M., and Kottas, A. (2012), "Mixture Modeling for Marked Poisson Processes," *Bayesian Analysis*, 7, 335–362. [394]
Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762. [399]
Veber, B., Nagode, M., and Fajdiga, M. (2008), "Generalized Renewal Process for Repairable Systems Based on Finite Weibull Mixture," *Reliability Engineering and System Safety*, 93, 1461–1472. [393]
Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999), "Bayesian Nonparametric Inference for Random Distributions and Related Functions," *Journal of the Royal Statistical Society,* Series B, 61, 485–527. [397]
Whitaker, L. R., and Samaniego, F. J. (1989), "Estimating the Reliability of Systems Subject to Imperfect Repair," *Journal of the American Statistical Association*, 84, 301–309. [393]