Computational Statistics and Data Analysis 73 (2014) 177–188

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A Bayesian semiparametric regression model for reliability data using effective age

Li Li, Timothy E. Hanson *

Department of Statistics, University of South Carolina, USA

ARTICLE INFO

Article history:
Received 8 January 2013
Received in revised form 28 October 2013
Accepted 22 November 2013
Available online 12 December 2013

Keywords:
Effective age
Repairable system
Tailfree process
Truncated data

ABSTRACT

A new regression model for recurrent events from repairable systems is proposed. The effectiveness of each repair in Kijima models I and II is regressed on repair-specific covariates. By modeling effective age in a flexible way, the model allows a spectrum of heterogeneous repairs besides "good as new" and "good as old" repairs. The density for the baseline hazard is modeled nonparametrically with a tailfree process prior which is centered at Weibull and yet allows substantial data-driven deviations from the centering family. Linearity in the predictors is relaxed using a B-spline transformation. The method is illustrated using simulations as well as two real data analyses.

1. Introduction

Repairable systems have been widely studied in the literature. Systems fail, get repaired upon failure, and these recurrent events (failures, repairs) are observed. The event process generating the repeated events is closely related to the intensity function, denoted as $\lambda(t|H(t))$ and formally defined in Section 2.1, which describes the probability of an instantaneous new failure, given the history of maintenances and failures $H(t)$. In general, recurrent event modeling methods can be divided into categories based on the type of repairs a system receives. Renewal processes are used if all the repairs bring the system to the "good as new" state and Poisson processes are used if all the maintenances bring the system to a "good as old" state. Kijima (1989) introduced two classes of models using the notion of "effective age" (also known as "virtual age") of the system to allow for a spectrum of repairs between "good as old" and "good as new". Consider a system observed over $[0, \tau]$. Assume the repair times for the system are $0 < t_1 < t_2 < \cdots < t_n$, and denote $\varepsilon(t)$ as the effective age of the system at time t. Suppose that the intensity $\lambda(t|H(t))$ is related to the unknown hazard, or failure rate, of a new system $r(t)$ through $\lambda(t|H(t)) = r\{\varepsilon(t)\}$. Poisson models assume $\varepsilon(t) = t$ and renewal models assume $\varepsilon(t) = t - s_{N(t-)}$ where $s_{N(t-)}$ is the time at which the last repair occurred. Kijima models introduce an age reduction factor D_i for each repair, occurring at calendar time t_i. Define $\varepsilon(t_i) = \varepsilon(t_{i-1}) + [t_i - t_{i-1}]D_i$ for the Kijima type I model and $\varepsilon(t_i) = [\varepsilon(t_{i-1}) + t_i - t_{i-1}]D_i$ for the Kijima type II model. Assume $\varepsilon(t) = \varepsilon(t_{i-1}) + t - t_{i-1}$ for $t \in (t_{i-1}, t_i)$. Note that $D_i = 1$ implies a Poisson process in models I and II, and $D_i = 0$ implies a renewal process in type II.

Lindqvist (2006) provides a review of the modeling of effective age. Dorado et al. (1997) generalize Kijima's models that allow for repairs of varying degree by including known "life supplements" — numbers between zero and one indicating the degrees of the repairs. There is very limited literature dealing with unknown effective age processes. Doyen and Gaudoin (2004) studied a class of Kijima's models where the repairs reduce the effective ages by one overall effectiveness scalar q. Recently, Veber et al. (2008) propose an EM-algorithm to estimate q and use Weibull mixtures for the baseline failure time distribution. Using one scalar is inappropriate for systems where repairs of varying effectiveness occur. For example,

* Correspondence to: 219C LeConte College, University of South Carolina, Columbia, SC 29208, USA. Tel.: +1 803 777 3859; fax: +1 803 777 4048.
E-mail addresses: lil@email.sc.edu (L. Li), hansont@stat.sc.edu (T.E. Hanson).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.11.015

different maintenance types, or levels of experience in those carrying out the repairs, can have drastic differences in repair effectiveness. Very recently, Yuan and Uday (2012) extend the single scalar parameter q to a time-dependent function, e.g. $q(t) = \exp(-et)$ where e is estimated, and assume the baseline distribution to be a parametric power-law distribution. In this work, we regress the effectiveness of each repair on covariates, e.g. materials used or the technician, and relax the parametric assumption of the baseline distribution using nonparametric priors in a Bayesian framework. Time trends in the effectiveness of repairs, characteristics of each repair, and association among repairs within each system can be flexibly coded into the covariate process. Specifically, the effectiveness measure D_i is regressed on a vector of covariates $\mathbf{w}_i$; let $D_i = \exp\{\boldsymbol{\beta}'\mathbf{w}_i\}/\{1+\exp(\boldsymbol{\beta}'\mathbf{w}_i)\}$ or $D_i = \exp(\boldsymbol{\beta}'\mathbf{w}_i)$. The associations between the covariates and the effective age reduction are characterized by $\boldsymbol{\beta}$. When the hazard of the system is monotone increasing, a repair with covariates resulting in a smaller age reduction factor D_i tends to be more effective than other repairs performed at the same effective age of the system.

Other generalizations of renewal and Poisson processes allowing for covariates also assume the effective age process $\varepsilon(t)$ is known, including for example, modulated renewal processes (Cox, 1972), point-process models incorporating renewals and time trends (Lawless and Thiagarajah, 1996), and a general class of semiparametric models (Peña et al., 2007) which simultaneously accommodates the effects of increasing numbers of events, covariates, interventions (repairs), and association among the interevent times within a system. This literature encompasses a rich and widely used family of reliability models. However, it is difficult to assume that the effective age process is known. There might even be interplay among the effective age process and history-dependent covariates and the baseline hazard function, as noted in Peña et al. (2007). Moreover, understanding the performance of repairs is often crucial to decision-making and even predictions.

A parametric analysis of our proposed model can be performed by choosing an appropriate distribution family, e.g. Weibull, for $r(t)$. In this work, we seek a more flexible approach where the entire density, the cumulative hazard function, or the hazard is assigned a nonparametric prior distribution. Bayesian nonparametric priors have achieved prominent success due to their flexibility in modeling unknown distributions; examples include the Dirichlet process (Ferguson, 1973), Polya tree priors (Lavine, 1992), Dirichlet process mixtures (Escobar and West, 1995), etc. However, the use of these nonparametric priors in recurrent event models has been quite limited. Very recently, Taddy and Kottas (2012) used Dirichlet process mixtures for the interfailure density in Poisson process models. Priors on the cumulative hazard $R(t) = \int_0^t r(s)ds$ include the beta and gamma processes (Lo, 1992; Kuo and Ghosh, 1997) which are discrete and not readily used in our context. The weighted gamma process (Ishwaran and James, 2004) is centered at one unique baseline intensity and is also not appropriate for a model that involves a factor in the argument of the intensity. Our proposed framework uses tailfree priors (Freedman, 1963; Ferguson, 1974; Jara and Hanson, 2011), on the space of densities, centered at the Weibull family, but allows for substantial data-driven deviations from the centering families. A special case of the tailfree prior, the Polya tree prior, has been widely used for models that warp the baseline r; see Hanson (2006), Walker and Mallick (1999), and Hanson and Yang (2007) for applications involving the accelerated failure time model and the proportional odds model. Like the Dirichlet process, tailfree priors also have desirable consistency and large support properties (Jara and Hanson, 2011). The general framework proposed herein allows model comparisons using the goodness-of-fit measures LPML and DIC so that comparisons among renewal processes, Poisson processes and Kijima models are readily made. We develop a full, automated MCMC sampling scheme to fit our proposed model and illustrate our method using simulations as well as on real data.

This paper is organized as follows: Section 2 presents a description of our model and an introduction to tailfree priors. Section 3 provides the MCMC algorithm and an approach to relax linearity in the linear predictor, and Section 4 presents simulation results. Section 5 summarizes the results for two real dataset analyses and in Section 6 we provide some concluding remarks.

2. Model development

2.1. Likelihood construction

Consider a system starting from new. Suppose the system gets repaired at times $t_i,\ i = 1, \ldots, n$ and $0 < t_1 < t_2 < \cdots < t_n < \tau$ where τ is the time when data collection stops. We assume τ is independent of the failure process. If a repair is performed without an accompanying failure, the observation of event time is right censored. Let the indicator δ_i take the value 1 if the system fails at time t_i and 0 otherwise. Further we assume a d-dimensional covariate vector for each repair, independent of the failure process, i.e. $\mathbf{w}_i = (w_{i0}, w_{i1}, \ldots, w_{i,d-1})$ for the repair at time t_i. This vector may incorporate information concerning technician skills, repair type, materials used, time trend, etc. Let the counting process $\{N(t), t \geq 0\}$ record the cumulative number of failures over time and $H(t) = \{N(s) : 0 \leq s < t\}$ be the history of the process at time t. The intensity function for an event process is defined as

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

The Kijima models for the event data assume $\lambda(t|H(t)) = r\{\varepsilon(t)\}$ where $\varepsilon(t)$ is the effective age. A Kijima type I model has $\varepsilon(t_i) = \varepsilon(t_{i-1}) + [t_i - t_{i-1}]D_i$ and the type II model has $\varepsilon(t_i) = [\varepsilon(t_{i-1}) + t_i - t_{i-1}]D_i$ where $t_i - t_{i-1}$ is the time since last repair. Denote the effective age right before t_i as $\varepsilon(t_{i-})$. The ith repair at t_i reduces the effective age right before t_i by