Statistica Sinica **25** (2015), 1613-1635
doi:http://dx.doi.org/10.5705/ss.2012.351

BAYESIAN INFERENCE OF HIDDEN GAMMA WEAR PROCESS MODEL FOR SURVIVAL DATA WITH TIES

Arijit Sinha[1], Zhiyi Chi[2] and Ming-Hui Chen[2]

[1]*Novartis Healthcare Private Ltd. and* [2]*University of Connecticut*

Abstract: Survival data often contain tied event times. Inference without careful treatment of the ties can lead to biased estimates. This paper develops the Bayesian analysis of a stochastic wear process model to fit survival data that might have a large number of ties. Under a general wear process model, we derive the likelihood of parameters. When the wear process is a Gamma process, the likelihood has a semi-closed form that allows posterior sampling to be carried out for the parameters, hence achieving model selection using Bayesian deviance information criterion. An innovative simulation algorithm via direct forward sampling and Gibbs sampling is developed to sample event times that may have ties in the presence of arbitrary covariates; this provides a tool to assess the precision of inference. An extensive simulation study is reported and a data set is used to further illustrate the proposed methodology.

Key words and phrases: Direct forward sampling, Gibbs sampling, jump process, latent variables, proportional hazards model, tied event times.

1. Introduction

Tied event times are a common phenomenon in time-to-event studies. For events that only happen at specific points in time, ties occur naturally, and for events that can happen at any point in time, ties may arise when a coarse time scale is used to record data (cf., Rossi, Berk, and Lenihan (1980)). Even when continuous event times are recorded at a fine time scale, ties can occur. Thus, machines in a workshop can stop working instantaneously from a power outage, an abrupt worsening of air quality can cause multiple emergency calls in a short time period, and in a sudden natural or man-made disaster casualties tend to happen at the same time (cf., Gold et al. (2007)). It can be more useful to model observed ties as the outcome of a certain mechanism underpinning the events than to either account for them as artifacts or to ignore them altogether.

Under the Proportional Hazards (PH) model of Cox (1972, 1975), the survival function of a subject with time-invariant covariates is expressed as

$$P(T > t \,|\, x) = \exp\{-H(t)\exp(x'\beta)\}, \tag{1.1}$$

where T is the failure time, x is a vector of covariates, H is a completely unspecified baseline cumulative hazard function, and β is a vector of regression coefficients. In the non-Bayesian setting, inference from the model uses the partial likelihood of β, which implicitly assumes that the baseline hazard rate $H'(t)$ exists, hence ruling out ties between independent failure times. In practice, in the presence of ties, approximation can be made by applying the formula for the no-tie case or by discretizing time (Cox (1972); Peto (1972); Breslow (1974); Efron (1977)).

Taking H as modeled as a stochastic process provides a powerful way to handle ties. In reliability analysis, Gaver (1963) took H to be a process with independent increments. Reynolds and Savage (1971) studied Gaver's model in detail and obtained a likelihood function of its parameters and, for the case of Gamma process, several closed form results. However, as Gaver's model sets $\beta = 0$ in (1.1), its primary concern is different from the PH model. Within the Bayesian setting, Dirichlet, Gamma, Beta, beta-Stacy and, more generally, neutral-to-the-right processes have been introduced as priors on H or its transformations (Ferguson (1973); Doksum (1974); Kalbfleisch (1978); Ferguson and Phadia (1979); Hjort (1990); Walker and Muliere (1997); Epifani, Lijoi, and Prünster (2003)), and inference on β can be based on maximum likelihood estimation or Bayesian posterior analysis (Kalbfleisch (1978); Hjort (1990); Damien, Laud, and Smith (1996); Laud, Damien, and Smith (1998); Kim and Lee (2003); Lee and Kim (2004)). Under these priors, tied event times occur with positive probability. On the other hand, a standard Bayesian approach that rules out ties imposes priors on the baseline hazard rate function H', which leads to continuous H (Antelman and Savage (1965); Reynolds and Savage (1971); Dykstra and Laud (1981); Lo and Weng (1989); Clayton (1991); Ibrahim, Chen, and Sinha (2001); Nieto-Barajas and Walker (2004); James (2005, 2006); Lijoi, Prünster, and Walker (2008b); Peccati and Prünster (2008); Kim and Kim (2009); Kim, Park, and Kim (2011)). In data analysis, tie-breaking has been used to artificially transform tied event times into distinct ones (Kalbfleisch (1978); Chen, Ibrahim, and Shao (2006)). However, the resulting estimate can be seriously biased if the proportion of ties is large (Burridge (1981)).

Following Gaver (1963), we regard H as a hidden stochastic wear process underlying the failures, rather than a parameter with a certain prior distribution. Based on the joint likelihood of β and the parameters of the process H, we establish Bayesian inference and model selection to analyze survival data with ties. The joint likelihood is obtained under a general multivariate process model that associates with each subject a possibly different H. This model unifies the PH model as a limiting case and others, such as the Lévy copula model (Epifani and Lijoi (2010)). We obtain the likelihood of the parameters using

an argument similar in spirit to those for several special cases (Reynolds and Savage (1971); Lijoi, Prünster, and Walker (2008a); Epifani and Lijoi (2010)). For homogeneous Gamma wear processes, we derive the likelihood in a semi-closed form. By imposing suitable noninformative priors on β and the parameters of the Gamma process, we can sample from their joint posterior distribution efficiently using Gibbs sampling. While similar methods have been used for posterior sampling of β (Damien, Laud, and Smith (1996); Laud, Damien, and Smith (1998)), the joint sampling appears to be new. With this in place, we propose to use the Bayesian deviance information criterion (DIC) (Spiegelhalter et al. (2002)) to guide the selection of Gamma process models.

We develop a Gibbs sampling-based simulation algorithm, termed the Direct Forward Sampling (DFS), to sample multiple failure times allowing for ties from a homogeneous Gamma process H in the presence of arbitrary values of β and covariates. The sampling is clearly different from posterior sampling, which has failure times already observed, and it does not rely on truncating the Lévy measure or sampling the path of H at pre-selected time points (Damien, Laud, and Smith (1996); Laud, Damien, and Smith (1998); Lee and Kim (2004)). Except for the approximation error of the Gibbs sampling of H just before and at failure times, as in posterior Gibbs sampling (Laud, Damien, and Smith (1998)), our sampling method is precise. In fact, by replacing Gibbs sampling with rejection sampling, exact sampling can be achieved (Chi (2012)).

The rest of the paper is as follows. Section 2 sets up notation. In Section 3 we propose a general multivariate additive process model, derive the likelihood function for the model, and apply it to Gamma wear processes. In Section 4 we describe the DFS algorithm. Section 5 details a Gibbs sampling algorithm for posterior computation. In Section 6 we report on an extensive simulation study to examine the empirical properties of the Gamma wear process model. In this section, we use DIC to guide the choice of parameters. In Section 7 we analyze a prostate cancer data set with our methodology. Section 8 ends with a discussion and potential future research work on this topic. Proofs and a discussion on possible extension to processes other than Gamma processes are given in Supplementary Material.

2. Basic Setup

Suppose n subjects are observed in a time-to-event study. Denote by T_i, C_i, and $Y_i = \min(T_i, C_i)$, the random failure time, right-censoring time, and endpoint of the i^{th} subject, respectively. We use the corresponding lower-case letters to denote the actual values of the random variables. Thus $y_i = \min(t_i, c_i)$ is the observed endpoint for the i^{th} subject, with t_i observable if and only if $t_i \le c_i$. Let $\delta_i = I\{y_i = t_i\} = I\{t_i \le c_i\}$. The observed data is $D_{\text{obs}} = \{y_i, \delta_i, x_i; i \le n\}$,

where x_i is the vector of covariates of the i^{th} subject. Let $\mathscr{D} = \{i : \delta_i = 1\}$, $\mathscr{N} = \{i : \delta_i = 0\}$, so $\mathscr{D}$ consists of those subjects that fail before censoring, and $\mathscr{N}$ those that are censored. Denote by $0 < \tau_1 < \tau_2 < \cdots < \tau_N$ the *distinct* values of $y_1, \ldots, y_n$ and $\tau_0 = 0$. For $j \leq N$, let

$$\begin{aligned} &\mathscr{D}_j = \{i \in \mathscr{D} : y_i = \tau_j\}, \quad \mathscr{N}_j = \{i \in \mathscr{N} : y_i = \tau_j\}, \\ &\mathscr{R}_j = \bigcup_{i \geq j} (\mathscr{D}_i \cup \mathscr{N}_i), \quad \mathscr{R}'_j = \mathscr{R}_j \setminus \mathscr{D}_j, \end{aligned} \tag{2.1}$$

so $\mathscr{D}_j$ consists of subjects that fail at time τ_j, $\mathscr{N}_j$ those censored at τ_j, $\mathscr{R}_j$ those at risk in time interval $(0, \tau_j)$, and $\mathscr{R}'_j$ those at risk in time interval $(0, \tau_j]$. Let $n_T = |\{j : \mathscr{D}_j \neq \emptyset\}|$ be the number of endpoints where failures occur. For $A \subset \{1, \ldots, n\}$, take $\kappa_A = (a_1, \ldots, a_n)$ with $a_i = I\{i \in A\}$. For brevity, write $\kappa_i = \kappa_{\{i\}}$. Let

$$\varrho_j = \kappa_{\mathscr{R}_j}, \quad \omega_j = \kappa_{\mathscr{R}'_j}, \quad j = 1, \ldots, N \tag{2.2}$$

and $\varrho_{N+1} = \omega_{N+1} = (0, 0, \ldots, 0)$. All analyses are conditional on $C_1, \ldots, C_n$.

A stochastic process $\mathcal{W} = (\mathcal{W}(t) : t \geq 0)$ is said to be additive if it has independent increments, is stochastically continuous, and with probability 1, the function $t \to \mathcal{W}(t)$ is right-continuous in $t \geq 0$ with $\mathcal{W}(0) = 0$ and has left limit in $t > 0$. In this paper, $\mathcal{W} = (H_1, \ldots, H_n)$ is an additive process taking values in $\mathbb{R}^n_+$ with $\mathbb{R}_+ := [0, \infty)$, and we refer to $\mathcal{W}$ as a pure jump process. It is well known that each H_i in $\mathcal{W}$ is nondecreasing and, for $a \in \mathbb{R}^n_+$,

$$E[e^{-a'\mathcal{W}(t)}] = e^{-\Psi(a,t)} \text{ with } \Psi(a,t) = \int_0^t \mathrm{d}v \int (1 - e^{-a's})\, \varphi(\mathrm{d}s \,|\, v), \tag{2.3}$$

where given $t > 0$, $\varphi(\mathrm{d}s \,|\, t)$ is a Lévy measure on $\mathbb{R}^n_+$ with $\int \min(1, |s|)\, \varphi(\mathrm{d}s \,|\, t) < \infty$ (Sato (1999)). We refer to Ψ as the characteristic exponent of $\mathcal{W}$. By Ferguson and Phadia (1979), $\mathcal{W}$ is homogeneous if $\Psi(a,t) = \Psi_1(a)\Psi_2(t)$.

Denote by $U(0,1)$ the uniform distribution on $(0,1)$, Gamma(a,b) the distribution with density $I\{x > 0\}\, b^{-a} x^{a-1} e^{-x/b}$, Exp$(c)$ = Gamma$(1,c)$, and δ the unit mass concentrated at 0. If F is a nondecreasing function on $\mathbb{R}_+$, then let $F^*(z) = \inf\{t > 0 : F(t) \geq z\}$ with the convention $\inf \emptyset = \infty$.

3. Joint Likelihood for Wear Process Model

3.1. N-variate wear process model

Assume that each of n subjects is exposed to a type of environmental fluctuation characterized by a nondecreasing stochastic process H_i with $H_i(0) = 0$ and $H_i(\infty) = \infty$, such that $\mathcal{W} = (H_1, \ldots, H_n)$ is a pure jump process and,

conditional on $\mathcal{W}$, the failure times $T_1, \ldots, T_n$ of the subjects are independent, with

$$P(T_i > t \,|\, \mathcal{W}) = e^{-H_i(t)}, \quad i \le n. \tag{3.1}$$

We assume $\mathcal{W}$ is unobservable. We also assume the right-censoring times C_1, $\ldots$ C_n are independent of $\mathcal{W}$ and $T_1, \ldots, T_n$. The process $\mathcal{W}$ is referred to as a (cumulative) wear process (Gaver (1963); Reynolds and Savage (1971)).

Example 1 (PH model)**.** In a Bayesian analysis of the PH model, typically there is a univariate pure jump process H such that, conditional on H, $T_1, \ldots, T_n$ are independent with $P(T_i > t \,|\, H) = e^{-\gamma_i H(t)}$, where γ_i is a constant that may incorporate covariates of the i^{th} subject. Here, H is often referred to as the baseline cumulative hazard function. To account for possible changes over time of the covariates, one might take $P(T_i > t \,|\, H) = \exp\{-\int_0^t \gamma_i(v)\,\mathrm{d}H(v)\}$, where $\gamma_i \ge 0$ is a bounded nonrandom function such that $\int_0^\infty \gamma_i(v)\,\mathrm{d}H(v) = \infty$ with probability 1. Let $H_i(t) = \int_0^t \gamma_i(v)\,\mathrm{d}H(v)$ and $\mathcal{W} = (H_1, \ldots, H_n)$. Since $\mathcal{W}$ is H-measurable,

$$\begin{aligned} P(T_1 > t_1, \ldots, T_n > t_n \,|\, \mathcal{W}) &= E[P(T_1 > t_1, \ldots, T_n > t_n \,|\, H) \,|\, \mathcal{W}] \\ &= E\Big[\prod_{i=1}^n e^{-H_i(t_i)} \,\Big|\, \mathcal{W}\Big] = \prod_{i=1}^n e^{-H_i(t_i)}. \end{aligned}$$

The PH model can thus be formulated as an n-variate model with $\mathcal{W}$ the wear process. Let $\varphi_0(\mathrm{d}x \,|\, t)$ be the Lévy measure of H. For $a \in \mathbb{R}_+^n$ and $t > 0$, since $E[e^{-a'\mathcal{W}(t)}] = E[e^{-\int_0^t \lambda(v)\,\mathrm{d}H(v)}] = \exp\{-\int_0^t \mathrm{d}v \int_0^\infty [1 - e^{-\lambda(v)x}]\varphi_0(\mathrm{d}x \,|\, v)\}$, with $\lambda(t) = a_1\gamma_1(t) + \cdots + a_n\gamma_n(t)$, the characteristic exponent of $\mathcal{W}$ is

$$\Psi(a, t) = \int_0^t \mathrm{d}v \int_0^\infty [1 - e^{-a_1\gamma_1(v)x - \cdots - a_n\gamma_n(v)x}]\, \varphi_0(\mathrm{d}x \,|\, v).$$

Consequently, the Lévy measure $\varphi(\mathrm{d}s \,|\, t)$ of $\mathcal{W}$, where $s = (s_1, \ldots, s_n)$, is as follows. Given $t > 0$, if all $\gamma_i(t) = 0$, then $\varphi(\mathrm{d}s \,|\, t) = 0$. On the other hand, if $\gamma_i(t) > 0$ for some i, then

$$\varphi(\mathrm{d}s \,|\, t) = \varphi_0\Big(\frac{\mathrm{d}s_i}{\gamma_i(t)} \,|\, t\Big) \prod_{j \ne i} \delta\Big(\mathrm{d}s_j - \frac{\gamma_j(t)s_i}{\gamma_i(t)}\Big). \tag{3.2}$$

Clearly, φ is determined by both φ_0 and γ_i. In Bayesian analysis, often only the parameters in γ_i are estimated, while the parameters of φ_0 are regarded as hyperparameters. However, under the n-variate model, this distinction between φ_0 and γ_i disappears, as both become parameters of the wear process $\mathcal{W}$.

Example 2 (Lévy copula)**.** A Lévy copula survival model was studied by Epifani and Lijoi (2010), in which the subjects are divided into two nonempty groups and a bivariate pure jump process $Z = (Z_1, Z_2)$ is used as the wear process such that, conditional on Z, the failure times are independent, and for each $i = 1, \ldots, n$ and $j = 1, 2$, if the i^{th} subject is in the j^{th} group, then $P(T_i > t \,|\, Z) = e^{-Z_j(t)}$. By letting $H_i = Z_j$, the model becomes an n-variate model. Suppose subject 1 belongs to group 1, subject 2 belongs to group 2 and, for each $i > 2$, j_i is the index of the group subject i belongs to. Then the characteristic exponent and Lévy measure of $\mathcal{W} = (H_1, \ldots, H_n)$ are

$$\Psi(a,t) = \int_0^t \mathrm{d}v \int (1 - e^{-a_1 s_1 - a_2 s_2 - \sum_{i>2} a_i s_{j_i}})\, \varphi_0(\mathrm{d}s_1, \mathrm{d}s_2 \,|\, v),$$
$$\varphi(\mathrm{d}s \,|\, t) = \varphi_0(\mathrm{d}s_1, \mathrm{d}s_2 \,|\, t) \prod_{i>2} \delta(\mathrm{d}s_j - s_{j_i}),\ s = (s_1, \ldots, s_n),$$

respectively, where φ_0 is the Lévy measure of Z.

Example 3 (Independent failure times)**.** In the above examples, the H_i are dependent processes, making T_i dependent random variables. If the H_i are independent, then the T_i are independent. If the Lévy measure of each H_i is $\varphi_i(\mathrm{d}x \,|\, t)$, then the characteristic exponent and Lévy measure of $\mathcal{W}$ are

$$\Psi(a,t) = \sum_{i=1}^n \Psi_i(a_i, t), \quad \varphi(\mathrm{d}s \,|\, t) = \sum_{i=1}^n \varphi_i(\mathrm{d}s_i \,|\, t) \prod_{j \neq i} \delta(\mathrm{d}s_j),$$

respectively, where $\Psi_i(a_i, t)$ is the characteristic exponent of H_i.

Under the n-variate model, when the H_i are dependent, the probability of ties among T_i is positive. From (3.1),

$$P(T_i > t) = \exp\left\{ - \int_0^t f_i(v)\, \mathrm{d}v \right\}, \quad \text{with } f_i(v) = \int (1 - e^{-s_i})\, \varphi(\mathrm{d}s \,|\, v) \geq 0,$$

so each T_i has a probability density. Then the T_i are dependent. It is noteworthy to mention that T_i are pairwise locally *independent* as defined by Oakes (1989). Let $X, Y > 0$ be random variables. For $t = (t_1, t_2)$, let $S(t) = P(X > t_1, Y > t_2)$ and $D_\alpha = \partial / \partial t_\alpha$. Then $\theta^*_{XY}(t) = S(t) D_1 D_2 S(t) / [D_1 S(t) \times D_2 S(t)]$ is the ratio of the conditional hazard rate of X at t_1 given $Y = t_2$, to that of X at t_1 given $Y > t_2$. X and Y are called locally independent if $\theta^*_{XY}(t) \equiv 1$.

Proposition 1. *$T_1, \ldots, T_n$ are pairwise locally independent.*

3.2. Likelihood function

The Lévy measure φ can be regarded as the only parameter of $\mathcal{W}$.

Theorem 1. *The likelihood function of φ based on D_{obs} is*

$$L(\varphi \,|\, D_{\text{obs}}) = \prod_{j=1}^{N} e^{-\Psi(\varrho_j,\tau_j)+\Psi(\varrho_j,\tau_{j-1})} \times \prod_{\mathscr{D}_j \neq \emptyset} \int e^{-\omega_j' s} \prod_{i \in \mathscr{D}_j} (1-e^{-s_i}) \varphi(\mathrm{d}s \,|\, \tau_j). \quad (3.3)$$

Kalbfleisch (1978) observed that in the setting of Example 1, if H is a Gamma process, then depending on its variability a spectrum of likelihoods can be obtained. To characterize this in general, write $\varphi(\mathrm{d}s \,|\, t) = c\nu(c\,\mathrm{d}s \,|\, t)$, with $c > 0$, where $\nu(\mathrm{d}s \,|\, t)$ is a Lévy measure with support in $\mathbb{R}_+^n$. Suppose for all $i = 1, \ldots, n$ and $t > 0$,

$$m_i(t) := \int_0^t \mathrm{d}v \int s_i \nu(\mathrm{d}s \,|\, v) < \infty, \quad (3.4)$$

and $\sigma_{ii}(t) < \infty$, where $\sigma_{ij}(t) = \int_0^t \mathrm{d}v \int s_i s_j \, \nu(\mathrm{d}s \,|\, v)$, $j = 1, \ldots, n$. Let $m(t) = (m_1(t), \ldots, m_n(t))$ and $\Sigma(t) = (\sigma_{ij}(t))$. Then

$$E[\mathcal{W}(t)] = \int_0^t \mathrm{d}v \int s \, \varphi(\mathrm{d}s \,|\, v) = \int_0^t \mathrm{d}v \int cs \, \nu(c\,\mathrm{d}s \,|\, v) = m(t),$$
$$\mathrm{Var}[\mathcal{W}(t)] = \int_0^t \mathrm{d}v \int ss' \, \varphi(\mathrm{d}s \,|\, v) = \int_0^t \mathrm{d}v \int css' \, \nu(c\,\mathrm{d}s \,|\, v) = c^{-1}\Sigma(t).$$

Here, c is called a precision parameter; the larger c is, the less variable $\mathcal{W}$ is. For us, c is fixed, its value will be determined via model selection, so it is not a part of the parameter to be estimated; see Sections 6−7 for more detail. Whenever c is involved, we rewrite the likelihood as $L(\nu \,|\, c, D_{\text{obs}})$.

Proposition 2. *If we fit D_{obs} to the model $\mathcal{W}$, with Lévy measure $c\,\nu(c\,\mathrm{d}s \,|\, t)$ satisfying* (3.4)*, then, as $c \to \infty$,*

$$L(\nu \,|\, c, D_{\text{obs}}) \to I\{\textit{all } |\mathscr{D}_j| = 0 \textit{ or } 1\} \times \prod_{i=1}^{n} e^{-m_i(y_i)} [m_i'(y_i)]^{\delta_i}.$$

Consider Example 1 again. Suppose $\varphi_0(\mathrm{d}x \,|\, t) = ch(cx, t)\,\mathrm{d}x$ for $x > 0$. Letting $g(t) = \int_0^t sh(s, v)\,\mathrm{d}v$, it can be seen that $m_i(t) = \int_0^t \gamma_i(v) g'(v)\,\mathrm{d}v$. As a result, as $c \to \infty$, the likelihood tends to

$$I\{\text{all } |\mathscr{D}_j| = 0 \text{ or } 1\} \times \prod_{i=1}^{n} e^{-\int_0^t \gamma_i(v) g'(v)\,\mathrm{d}v} [\gamma_i(y_i) g'(y_i)]^{\delta_i},$$

and hence it behaves similarly to the one under the PH model. However, whereas the likelihood based on the n-variate model automatically discriminates against ties in this case, the one based on the PH model cannot.

The result implicitly assumes that m is differentiable at every τ_j with $\mathscr{D}_j \neq 0$. Following the argument for the existence of probability density of T_i, this indeed holds with probability 1. Here then is some information on the probability of ties for the case of most interest to us.

Proposition 3. *Let $\mathcal{W}$ be homogeneous such that, for any $a \in \mathbb{R}^n_+$, $\Psi(a,t) = \Psi_1(a)\Psi_2(t)$. Then*

$$P(T_i = T_j) = \frac{\Psi_1(\kappa_i) + \Psi_1(\kappa_j)}{\Psi_1(\kappa_i + \kappa_j)} - 1, \quad i \neq j. \tag{3.5}$$

If $\mathcal{W}$ has Lévy measure $\varphi(\mathrm{d}s \,|\, t) = c\nu(c\mathrm{d}s \,|\, t)$ with $\nu(\mathrm{d}s \,|\, t) = h(t)\lambda(\mathrm{d}s)$ satisfying (3.4), where λ is a Lévy measure on $\mathbb{R}^n_+$, then as $c \to \infty$, $P(T_i$ all different$) \to 1$.

3.3. A Gamma wear process model

Let $\mathcal{W} = \gamma H$, where $\gamma = (\gamma_1, \ldots, \gamma_n)$ is a constant vector with $\gamma_i > 0$ and H is a homogeneous Gamma process with Lévy measure

$$\varphi_0(\mathrm{d}s \,|\, t) = cf(t) I\{s > 0\} s^{-1} e^{-cs} \,\mathrm{d}s,$$

with $c > 0$ being the precision parameter and $f = F'$. Denote $H \sim \mathcal{GP}(cF, c)$. We refer to the corresponding n-variate model as the Gamma Process (GP) model. The parameters of the model are γ, F, and c, but c will be fixed via model selection and only γ and F will be estimated.

Corollary 1. *The likelihood function for the GP model is*

$$L(\gamma, F \,|\, c, D_{\mathrm{obs}}) = \prod_{j=1}^{N} \left(\frac{c}{c + \varrho_j' \gamma} \right)^{c[F(\tau_j) - F(\tau_{j-1})]} \times \prod_{j : \mathscr{D}_j \neq \emptyset} cf(\tau_j) \int_0^\infty s^{-1} e^{-(c + \omega_j'\gamma)s} \prod_{i \in \mathscr{D}_j} (1 - e^{-\gamma_i s}) \,\mathrm{d}s. \tag{3.6}$$

The proof of (3.6) is quick. As $a'\mathcal{W}(t) = a'\gamma H(t) \sim \mathrm{Gamma}(cF(t), a'\gamma/c)$ for $0 \neq a \in \mathbb{R}^n_+$, $e^{-\Psi(a,t)} = (1 + a'\gamma/c)^{-cF(t)}$. Then the first factor on the right hand side in (3.6) follows from that in (3.3); the second factor follows from that in (3.3) and (3.2).

When the data have no ties, (3.6) can be shown to coincide with (14) in Kalbfleisch (1978). To take ties into account, Kalbfleisch (1978) derived a likelihood of regression coefficients in his (23) which, if expressed in integral form, is a part of the likelihood in (3.6) but with the factor $\prod_{\mathscr{D}_j \neq \emptyset} cf(\tau_j)$ missing. From Proposition 2, we get the following when $c \to \infty$.

Corollary 2. *Given D_{obs}, let the GP model be fit with $\mathcal{W} = \gamma H$, where $\gamma = (\gamma_1, \ldots, \gamma_n)$ with $\gamma_i > 0$ and $H \sim \mathcal{GP}(cF, c)$. Fixing γ and F, as $c \to \infty$,*

$$L(\gamma, F \,|\, c, D_{\mathrm{obs}}) \to I\{\text{all } |\mathscr{D}_j| = 0 \text{ or } 1\} \times \prod_{i=1}^{n} e^{-\gamma_i F(y_i)} [\gamma_i f(y_i)]^{\delta_i}.$$

Corollary 3. *Suppose there is no censoring. Then for any $i \neq j$,*

$$P(T_i = T_j) = \frac{\ln(1+\gamma_i/c) + \ln(1+\gamma_j/c)}{\ln(1+\gamma_i/c+\gamma_j/c)} - 1.$$

Thus, as $c \to 0$, $P(T_i$ all equal$) \to 1$, and as $c \to \infty$, $P(T_i$ all different$) \to 1$.

It should be noted that in general, as the wear process becomes more variable, it is not necessarily true that $P(T_i \text{ all equal}) \to 1$. For example, let H be a generalized Gamma process with time-independent Lévy density $c^2 h_1(cs)$, where $h_1(s) = s^{-\alpha-1}e^{-s}$, $0 < \alpha < 1$ (Hougaard (1986); Brix (1999); Epifani, Lijoi, and Prünster (2003); Lijoi, Mena, and Prünster (2007); Argiento, Guglielmi, and Pievatolo (2010)). Then $\Psi_1(\lambda) = (1+\lambda/c)^\alpha - 1$ and, by Proposition 3, as $c \to 0$,

$$P(T_i = T_j) = \frac{(\gamma_i + c)^\alpha + (\gamma_j + c)^\alpha - 2c^a}{(\gamma_i + \gamma_j + c)^\alpha - c^\alpha} - 1 \to \frac{\gamma_i^\alpha + \gamma_j^\alpha}{(\gamma_i + \gamma_j)^\alpha} - 1 \in (0,1).$$

4. Sampling of Survival Data from Gamma Process

Proposition 4. *Let F be strictly increasing and $\alpha(t) = cF(t)$. Let $G \sim \mathscr{GP}(t,1)$ be a standard Gamma process and η_i be i.i.d.* Exp(c) *random variables independent of G. Then $(T_1, \ldots, T_n) \sim (\alpha^{-1}(G^*(\eta_1/\gamma_1)), \ldots, \alpha^{-1}(G^*(\eta_n/\gamma_n)))$.*

Here, since η_i and G are independent, if we can sample $G^*(\theta_i)$ for an arbitrary *fixed* set of $\theta_i > 0$, then we can sample $T_1, \ldots, T_n$. The result follows from the inversion formula for univariate distributions (Devroye (1986)). The inversion is used by Bender, Augustin, and Blettner (2005) to sample failure times with no ties for the PH model. We next describe how to jointly sample $G^*(\theta_i)$ forwardly. For convenience, suppose θ_i are already sorted in increasing order.

Theorem 2. *Let $G \sim \mathscr{GP}(t,1)$. Given a single $\theta > 0$, the distribution function of $G^*(\theta)$ is given by*

$$P(G^*(\theta) \le t) = \frac{1}{\Gamma(t)} \int_\theta^\infty u^{t-1} e^{-u} \, \mathrm{d}u \tag{4.1}$$

and, given $G^*(\theta) = \tau$, the conditional distribution of $G(\tau)$ is

$$P(G(\tau) \le r \,|\, G^*(\theta) = \tau) = \frac{M_{\theta,\tau}(r)}{M_{\theta,\tau}(\infty)}, \quad \theta \le r < \infty, \tag{4.2}$$

where

$$M_{\theta,\tau}(r) = \int_\theta^r e^{-s} \left(\int_0^\theta \frac{u^{\tau-1}\,\mathrm{d}u}{s-u} \right) \mathrm{d}s.$$

This enables us to sample $G^*(\theta_i)$ for $0 < \theta_1 < \cdots < \theta_n$, as follows. First sample $\tau_1 = G^*(\theta_1)$ from (4.1) and $r_1 = G(\tau_1)$, conditional on τ_1, from (4.2). If $r_1 \geq \theta_n$, then all $G^*(\theta_i) = \tau_1$. Otherwise, with s the number with $\theta_s \leq r_1 < \theta_{s+1}$, $G^*(\theta_1) = \cdots = G^*(\theta_s) = \tau_1 < G^*(\theta_{s+1}) \leq \cdots \leq G^*(\theta_n)$. In general, if (τ_1, r_1), $\ldots$, (τ_k, r_k) have been sampled but there is $s < n$ such that $\theta_s \leq r_k < \theta_{s+1}$, then τ_{k+1} and $r_{k+1} = G^*(\tau_{k+1})$ are sampled as follows. First, independently from all (τ_j, r_j), $j \leq k$, sample $\tilde{\tau}_{k+1} \sim G^*(\theta_{s+1} - r_k)$ from (4.1) and $\tilde{r}_{k+1} \sim G(\tilde{\tau}_{k+1})$, conditional on $G^*(\theta_{s+1} - r_k) = \tilde{\tau}_{k+1}$, from (4.2). Then $\tau_{k+1} = G^*(\theta_{s+1}) = \tau_k + \tilde{\tau}_{k+1}$ and $r_{k+1} = r_k + \tilde{r}_{k+1}$. If $r_{k+1} \geq \theta_n$, then all $G^*(\theta_{s+1}) = \cdots = G(\theta_n) = \tau_k$, otherwise, sample the next distinct failure time τ_{k+2} and r_{k+2}. The procedure continues until all $G^*(\theta_i)$ are sampled.

Gibbs sampling can be applied to the conditional distribution (4.2). Introduce two latent variables U and V such that $0 < U < \theta$, $V > 0$ and, conditional on $G^*(\theta) = \tau$, $G(\tau)$, U, and V have joint density

$$m_\theta(r, u, v) \propto e^{-r} u^{\tau - 1} e^{-v(r-u)}, \quad \theta < r < \infty,\ 0 < u < \theta,\ 0 < v < \infty.$$

Let $\zeta = \ln[U/(\theta - U)]$ and denote the conditional joint density of $G(\tau)$, ζ, and V by $m_\theta(r, z, v)$. Using the collapsed Gibbs method (Liu (1994); Chen, Shao, and Ibrahim (2000)), we then sample from, in turn: (i) $m_\theta(z \,|\, r)$, (ii) $m_\theta(v \,|\, r, z)$, and (iii) $m_\theta(r \,|\, v)$. For (i), we have

$$m_\theta(z \,|\, r) \propto \frac{e^{\tau z}}{(1 + e^z)^\tau} \times \frac{1}{r + (r - \theta)e^z}, \quad -\infty < z < \infty;$$

this can be shown to be a log-concave density with the conditional mode

$$z_{\text{mod}} = \ln \left[\frac{(\tau - 1)(r - \theta) + \{[(\tau - 1)(r - \theta)]^2 + 4(r - \theta)\tau r\}^{1/2}}{2(r - \theta)} \right],$$

thus allowing the application of the adaptive-rejection algorithm of Gilks and Wild (1992) to sample ζ conditional on $G(\tau) = r$. For (ii), we have

$$m_\theta(v \,|\, r, z) \propto \exp\left\{ -v \left(r - \frac{\theta e^z}{1 + e^z} \right) \right\}, \quad 0 < v < \infty,$$

which is an exponential density with mean $[r - \theta e^z/(1 + e^z)]^{-1}$. For (iii), $m_\theta(r \,|\, v)$ $\propto e^{-r(1+v)}$, $r > \theta$, and hence sampling r is also straightforward. We use the following algorithm to generate failure times that may have ties.

Direct Forward Sampling (DFS) Algorithm

1. Set n (number of failure times), c (precision parameter), and $\gamma_1, \ldots, \gamma_n$ (coefficients).

2. Generate η_i i.i.d. $\sim \mathrm{Exp}(c)$ and set $\theta_i = \eta_i/\gamma_i$ for $i \le n$.
3. Rearrange (γ_i, θ_i) so that $0 < \theta_1 < \cdots < \theta_n$.
4. Initialize $k = 0$, $t = 0$, and $h = 0$.
5. Generate τ, which is a realization of $G^*(\theta_{k+1} - h)$.
6. Generate r, which is a realization of $G(\tau)$ conditional on $G^*(\theta_{k+1} - h) = \tau$.
7. Update t to $t + \tau$.
8. Update h to $h + r$.
9. If $h \ge \theta_n$, then $G^*(\theta_{k+1}) = \cdots = G^*(\theta_n) = t$. Go to Step 14.
10. If $\theta_{k+s} \le h < \theta_{k+s+1}$ for some $1 \le s < n - k$, then $G^*(\theta_{k+1}) = \cdots = G^*(\theta_{k+s}) = t$.
11. Update k to $k + s$.
12. Go to Step 5.
13. Follow Steps 8 through 12 until all $G^*(\theta_1), \ldots, G^*(\theta_n)$ are generated.
14. Return $c^{-1}F^{-1}(G^*(\theta_i))$, the failure times from $\mathscr{GP}(cF, c)$.

5. Bayesian Posterior Inference

Henceforth, we assume that in the GP model $\gamma_i = \exp(x_i'\beta)$, where x_i is the vector of covariates of the i^{th} subject and β the vector of regression coefficients. Our goal is to develop posterior inference for (β, F) given c.

5.1. Prior

We assume a piecewise linear model for F as follows. Partition the time axis into K intervals $(a_0, a_1]$, $(a_1, a_2]$, $\ldots$, $(a_{K-1}, a_K]$, where $a_0 = 0$ and $a_K \ge \tau_N$. Then let $f(t) = F'(t) = \lambda_k$ for $a_{k-1} < t \le a_k$. Under our model,

$$F(\tau_j) - F(\tau_{j-1}) = \sum_{k=1}^{K} \lambda_k \, d_{jk}, \quad \text{with } d_{jk} = |(a_{k-1}, a_k) \cap (\tau_{j-1}, \tau_j)|. \tag{5.1}$$

Let $\lambda = (\lambda_1, \ldots, \lambda_K)'$, and for $j \le N$, $\nu(j)$ be the unique index with $a_{\nu(j)-1} < \tau_j \le a_{\nu(j)}$. Then, the likelihood function in (3.6) can be rewritten as

$$L(\beta, \lambda \,|\, c, D_{\text{obs}}) = \prod_{j=1}^{N} \left(\frac{c}{c + \varrho_i'\gamma} \right)^{c[F(\tau_j) - F(\tau_{j-1})]} \times \left[c\lambda_{\nu(j)} \int_0^\infty s^{-1} e^{-(c+\omega_j'\gamma)s} \prod_{i \in \mathscr{D}_j} (1 - e^{-\gamma_j s}) \, \mathrm{d}s \right]^{I\{\mathscr{D}_j \neq \emptyset\}}. \tag{5.2}$$

We assume the prior $\pi(\beta,\lambda) \propto \exp(-\beta'\Sigma_0^{-1}\beta/2)\prod_{k=1}^K \lambda_k^{\alpha_0-1}e^{-\alpha_1\lambda_k}$. Under the prior, β, and $\lambda_1,\ldots,\lambda_K$ are independent, with $\beta \sim N_p(0,\Sigma_0)$ and $\lambda_k \sim \text{Gamma}(\alpha_0, 1/\alpha_1)$. In Sections 6 and 7, we specify $\Sigma_0 = 10^4 I_p$ and $\alpha_0 = \alpha_1 = 10^{-2}$, which lead to a relatively vague prior for (β,λ).

5.2. Posterior computation

To sample the joint posterior distribution

$$\pi(\beta,\lambda \,|\, c, D_{\text{obs}}) \propto L(\beta,\lambda \,|\, c, D_{\text{obs}})\pi(\beta,\lambda),$$

introduce the latent variable $s = (s_j : \mathscr{D}_j \neq \emptyset, j \le N)$ and define an augmented joint posterior distribution $\pi(\beta,\lambda,s \,|\, c, D_{\text{obs}}) \propto L(\beta,\lambda,s \,|\, c, D_{\text{obs}})\pi(\beta,\lambda)$, where

$$L(\beta,\lambda,s \,|\, c, D_{\text{obs}}) = \prod_{j=1}^N \left(\frac{c}{c+\varrho_j'\gamma}\right)^{c[F(\tau_j)-F(\tau_{j-1})]} \left[c\,\lambda_{\nu(j)} s_j^{-1} e^{-s_j(c+\omega_j'\gamma)} \prod_{i\in\mathscr{D}_j}(1-e^{-\gamma_i s_j})\right]^{I\{\mathscr{D}_j\neq\emptyset\}}.$$

Since $\int \pi(\beta,\lambda,s \,|\, c, D_{\text{obs}})\,\mathrm{d}s = \pi(\beta,\lambda \,|\, D_{\text{obs}})$ by Corollary 1, $\pi(\beta,\lambda \,|\, c, D_{\text{obs}})$ can be sampled by applying Gibbs sampling to $\pi(\beta,\lambda,s \,|\, c, D_{\text{obs}})$. We sample (β,λ,s) from the following, in turn: (i) $\pi(\beta \,|\, \lambda, s, c, D_{\text{obs}})$, (ii) $\pi(\lambda \,|\, \beta, s, c, D_{\text{obs}})$, and (iii) $\pi(s \,|\, \beta, c, D_{\text{obs}})$. For (i),

$$\pi(\beta \,|\, \lambda, s, c, D_{\text{obs}}) = \prod_{j=1}^N \left(\frac{c}{c+\varrho_j'\gamma}\right)^{c[F(\tau_j)-F(\tau_{j-1})]} \left[e^{-s_j(c+\omega_j'\gamma)} \prod_{i\in\mathscr{D}_j}(1-e^{-\gamma_i s_j})\right]^{I\{\mathscr{D}_j\neq\emptyset\}} \pi(\beta).$$

Since $\gamma_i = \exp(x_i'\beta)$, it is easy to show that $\pi(\beta \,|\, \lambda, s, D_{\text{obs}})$ is log-concave in each component of β, and so we can use the adaptive rejection algorithm of Gilks and Wild (1992) to sample β. For (ii), given β and s, $\lambda_1,\ldots,\lambda_K$ are conditionally independent and, for each k, the conditional posterior distribution of λ_k is

$$\pi(\lambda_k \,|\, \beta, s, c, D_{\text{obs}}) \propto \lambda_k^{\alpha_0+\sum_{j=1}^N I\{\mathscr{D}_j\neq\emptyset, a_{k-1}<\tau_j\le a_k\}} \exp\left\{-\lambda_k\left[\alpha_1 - c\sum_{j=1}^N d_{jk}\ln\left(\frac{c}{c+\varrho_j'\gamma}\right)\right]\right\}.$$

Thus, λ_k follows a Gamma distribution that is easy to sample. For (iii), given β, $s_1,\ldots,s_N$ are conditionally independent and, for each j with $\mathscr{D}_j \neq \emptyset$,

$$\pi(s_j \,|\, \beta, c, D_{\text{obs}}) \propto s_j^{-1} e^{-(c+\omega_j'\gamma)s_j} \prod_{i\in\mathscr{D}_j}(1-e^{-\gamma_i s_j}).$$

Let $u_j = \ln s_j$. Then the conditional posterior density of u_j is

$$\pi(u_j \,|\, \beta, c, D_{\text{obs}}) \propto \exp\left\{-(c + \omega_j'\gamma)e^{u_j}\right\} \prod_{i \in \mathscr{D}_j} (1 - \exp\{-\gamma_i\, e^{u_j}\}).$$

It is easy to show that $\pi(u_j \,|\, \beta, D_{\text{obs}})$ is log-concave. Then we again can use the adaptive rejection algorithm to sample u_j and set $s_j = \exp(u_j)$.

6. A Simulation Study

We conducted a simulation study to compare the PH model and the GP model with $H \sim \mathscr{GP}(cF, c)$. As the value of c is unknown in practice, to guide the choice of c in fitting the GP model, we use deviance information criterion (DIC) (Spiegelhalter et al. (2002)). Define the deviance function

$$D(\boldsymbol{\psi}) = -2 \ln L(\beta, \lambda \,|\, c, D_{\text{obs}}),$$

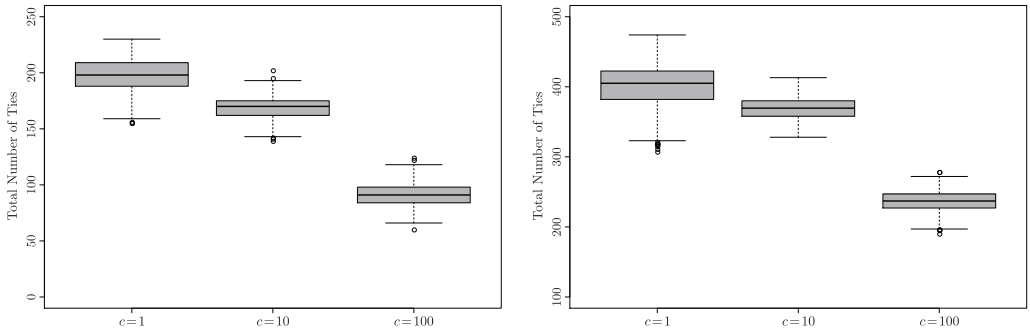
where $\boldsymbol{\psi} = (\beta', \lambda')'$ and $L(\beta, \lambda \,|\, c, D_{\text{obs}})$ is given in (5.2). Then

$$\text{DIC} = D(\overline{\boldsymbol{\psi}}) + 2p_D, \tag{6.1}$$

where $\overline{\boldsymbol{\psi}} = E[\boldsymbol{\psi} \,|\, D_{\text{obs}}]$ and $p_D = \overline{D(\boldsymbol{\psi})} - D(\overline{\boldsymbol{\psi}})$ with $\overline{D(\boldsymbol{\psi})} = E[D(\boldsymbol{\psi}) \,|\, D_{\text{obs}}]$. In (6.1), $D(\overline{\boldsymbol{\psi}})$ measures the goodness-of-fit, and p_D is the effective number of model parameters. The DIC is a Bayesian measure of fit or adequacy with $2p_D$ being the dimensional penalty term. The smaller the DIC value, the better the model fits the data. In this simulation study, our second goal was to examine the performance of DIC in correctly identifying c in the fitted the GP model.

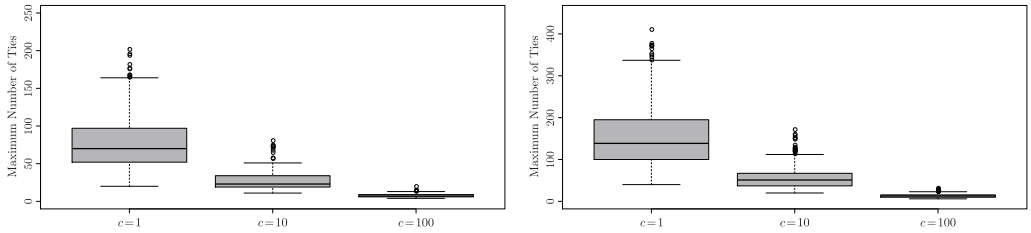
In the simulation study, the data were generated as follows. We generated $x_i = (x_{i1}, x_{i2})'$, $i \le n$, where $x_{i1} \sim N(0, 1)$, $x_{i2} \sim$ Bernoulli(0.7) were all independent. We set $\beta = (\beta_1, \beta_2) = (1, -0.5)$, and $F(t) = t$, and considered sample sizes $n = 250$ and 500. We used the DFS algorithm in Section 4 to generate failure times from the GP model with $\gamma_i = \exp(x_i'\beta)$, and for c as 1, 10, and 100. We independently generated n censored times from a rescaled beta distribution such that $C_i = 38q_i$ with $q_i \sim \text{beta}(1, 3)$, which yielded approximately 15% of censored observations for each simulated data set. We independently generated 500 data sets under each simulation setting.

For each data set, we let $N_{\text{total}} = \sum_j |\mathscr{D}_j| I\{|\mathscr{D}_j| > 1\}$ and $N_{\max} = \max_j |\mathscr{D}_j|$, where $\mathscr{D}_j$ is defined as (2.1). Figure 1 shows the boxplots of $N_{\text{total}}^{(1)}, \ldots, N_{\text{total}}^{(500)}$ and Figure 2 shows the boxplots of $N_{\max}^{(1)}, \ldots, N_{\max}^{(500)}$ for the 500 simulated data sets under the six simulation settings. From Figure 1, we can see that as c increases from 1 to 100, $N_{\text{total}}^{(1)}, \ldots, N_{\text{total}}^{(500)}$ in the simulated data sets decrease as a whole, and their median drops substantially.



(a) $n = 250$ (b) $n = 500$

Figure 1. Boxplots of the total numbers of ties in 500 simulated data sets of sizes $n = 250$ and $n = 500$ generated from the GP models with $c = 1, 10,$ and 100.



(a) $n = 250$ (b) $n = 500$

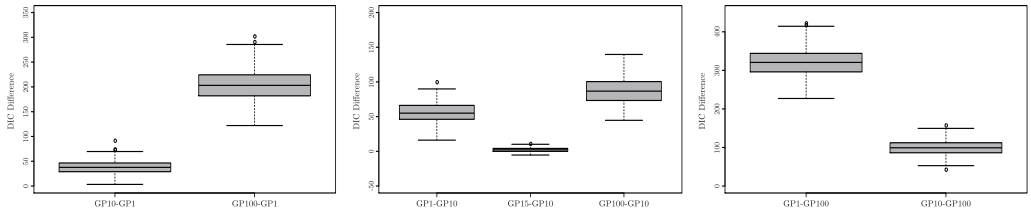
Figure 2. Boxplots of the maximum numbers of ties in 500 simulated data sets of sizes $n = 250$ and $n = 500$ generated from the GP model with $c = 1, 10,$ and 100.

For each simulated data set, we fit the PH model with a constant baseline hazard rate function and the GP model with $F(t) = t$; the true value of c was used in the simulation. For each simulated data set, we implemented the Gibbs sampling algorithm of Section 5.2 and used 5,000 Gibbs iterations after a burn-in of 500 iterations to compute the posterior estimates. Let $\hat{\beta}_{j\ell}$ and $\mathrm{sd}_\ell(\beta_j)$ denote the posterior mean and the posterior standard deviation of β_j computed from the ℓ^{th} simulated data set for $\ell = 1, \ldots, 500$. The simulation posterior estimate (Est), the simulation posterior standard deviation (SD), the simulation error (SE), and the mean squared error (MSE) for β_j are, respectively, $\overline{\hat{\beta}}_j = (1/500)\sum_{\ell=1}^{500} \hat{\beta}_{j\ell}$, $\overline{\mathrm{sd}}(\beta_j) = (1/500)\sum_{\ell=1}^{500} \mathrm{sd}_\ell(\beta_j)$, $\mathrm{SE}(\beta_j) = \Big[(1/499)\sum_{\ell=1}^{500}(\hat{\beta}_{j\ell} - \overline{\hat{\beta}}_j)^2\Big]^{1/2}$, and $\mathrm{MSE}(\beta_j) = (1/500)\sum_{\ell=1}^{500}(\hat{\beta}_{j\ell} - \beta_j)^2$, where β_j is the true value. We define the same simulation summary statistics for λ. We let CP denote the coverage probability of the 95% highest posterior density (HPD) intervals that contain the true parameter value in the 500 simulated data sets, using the Monte Carlo method developed by Chen and Shao (1999). Table 1 shows these

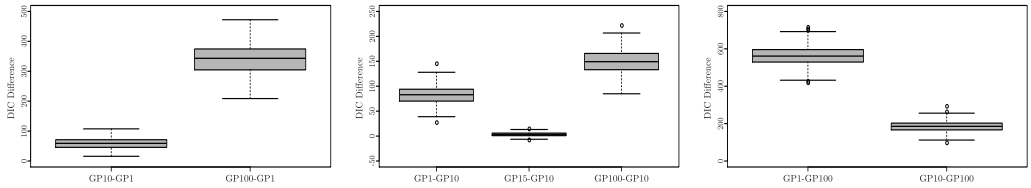
Table 1. Summary of posterior estimates for the GP and the PH Models in simulation studies.

n	Parameter		True $c = 1$		True $c = 10$		True $c = 100$	
			GP	PH	GP	PH	GP	PH
250	β_1	True	1		1		1	
		Est	1.012	0.827	1.013	0.965	1.002	0.995
		SD	0.096	0.070	0.085	0.072	0.078	0.073
		SE	0.094	0.242	0.085	0.127	0.079	0.085
		MSE	0.009	0.089	0.007	0.017	0.006	0.007
		CP	0.944	0.300	0.950	0.682	0.942	0.906
	β_2	True	-0.5		-0.5		-0.5	
		Est	-0.504	-0.407	-0.514	-0.482	-0.498	-0.494
		SD	0.178	0.158	0.152	0.156	0.139	0.156
		SE	0.157	0.176	0.142	0.158	0.128	0.144
		MSE	0.025	0.040	0.020	0.025	0.016	0.021
		CP	0.974	0.894	0.962	0.956	0.966	0.970
	λ	True	1		1		1	
		Est	1.055	0.903	1.020	0.964	1.011	1.007
		SD	0.225	0.116	0.135	0.123	0.113	0.128
		SE	0.234	0.782	0.136	0.254	0.109	0.145
		MSE	0.057	0.619	0.019	0.066	0.012	0.021
		CP	0.940	0.262	0.944	0.654	0.962	0.924
500	β_1	True	1		1		1	
		Est	1.010	0.824	0.999	0.964	1.003	0.995
		SD	0.068	0.050	0.061	0.051	0.057	0.051
		SE	0.071	0.244	0.063	0.129	0.055	0.062
		MSE	0.005	0.091	0.004	0.018	0.003	0.004
		CP	0.938	0.184	0.948	0.532	0.964	0.892
	β_2	True	-0.5		-0.5		-0.5	
		Est	-0.501	-0.400	-0.511	-0.489	-0.508	-0.500
		SD	0.129	0.113	0.113	0.111	0.098	0.110
		SE	0.117	0.156	0.098	0.112	0.094	0.108
		MSE	0.014	0.034	0.010	0.013	0.009	0.012
		CP	0.966	0.754	0.972	0.952	0.948	0.952
	λ	True	1		1		1	
		Est	1.026	0.860	1.020	0.986	1.013	1.000
		SD	0.188	0.078	0.108	0.089	0.081	0.090
		SE	0.190	0.597	0.102	0.263	0.079	0.112
		MSE	0.037	0.375	0.011	0.070	0.006	0.013
		CP	0.940	0.200	0.970	0.498	0.956	0.870

simulation summary statistics. We see that the GP model generally performed well and the posterior estimates were very close to the true values of β and λ, and the coverage probabilities were close to 95%. Meanwhile, the PH model performed poorly and there were substantial biases in the posterior estimates,



(a1) $n = 250$, true $c = 1$ (a2) $n = 250$, true $c = 10$ (a3) $n = 250$, true $c = 100$



(b1) $n = 500$, true $c = 1$ (b2) $n = 500$, true $c = 10$ (b3) $n = 500$, true $c = 100$

Figure 3. Boxplots of DIC differences in 500 simulations for the data of sizes $n = 250$ and $n = 500$ generated from the GP model with $c = 1$, $c = 10$, and $c = 100$.

especially when c was small. When $c = 100$, the performance of the PH model improved and the biases of the posterior estimates under the PH model were reduced considerably, but the coverage probabilities were still smaller than the expected 95%, especially for β_1 when $n = 500$.

To examine the performance of DIC, for each simulated data set, we fit the GP model with $c = 1$, 10, and 100 when the true $c = 1$ or 100. When the true $c = 10$, we fit the GP model with $c = 1$, 10, 15, and 100. In the DIC computation, we used (3.6) to compute $L(\beta, \lambda, c \,|\, D_{\text{obs}})$. Our simulation codes were written in FORTRAN 95 with double precision. The IMSL subroutine DQDAGI was used for evaluating all one-dimensional integrals involved in the likelihood function. For $\ell = 1, \ldots, 500$, let $\text{DIC}_{c,\ell}$ denote the DIC computed for the ℓ^{th} simulated data set. The boxplots of the DIC differences $\Delta_\ell(c, c') = \text{DIC}_{c',\ell} - \text{DIC}_{c,\ell}$ for different values of c' and c in Figure 3 show that DIC could identify the true GP model correctly for most of the simulated data sets and the DIC differences were quite large when the value of c in the fitted GP model was far from the true c. Even when c in the fitted GP model was close to the true one, for example, when $c = 15$ in the fitted GP model and the true $c = 10$, the boxplot shown in Figure 3 is nearly above zero, but with much smaller DIC differences.

7. Analysis of Prostate Cancer Data

We considered a subset of the data from a prostate cancer study published by D'Amico et al. (2010), which consisted of 558 patients with high risk prostate

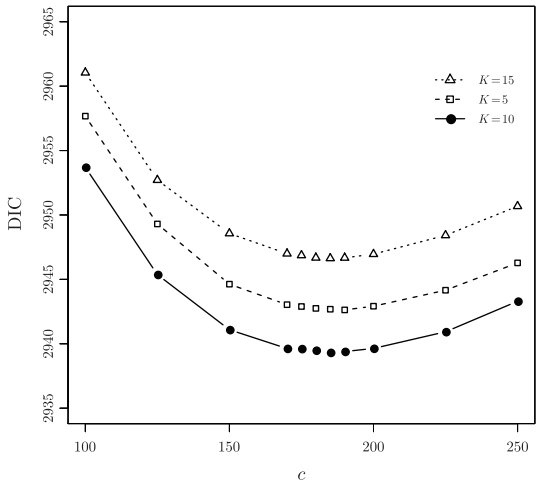


Figure 4. Plots of DIC values versus c with $K = 5, 10, 15$ for the prostate cancer data.

cancer, namely, prostate specific antigen (PSA) > 20, clinical Gleason score ≥ 8, or clinical stage T3 or higher. All patients in the subset were treated with radical prostatectomy (RP) between 1989 and 2008. In these data, the response is time to PSA failure or time to the last follow-up from the time of RP, whichever is smaller. The time of PSA failure is the time of prostate cancer recurrence after RP. The clinical implication of PSA recurrence is that men are offered salvage therapy (second treatment), which may prolong life or cure the patient but may have side effects. The covariates include age in years at the date of RP, the logarithm of PSA (logpsa), pathological Gleason score (pGS7 and pGS8H), pathological stage (pT3H), positive surgical margin (Margin), and year of RP. Among these covariates, age, logpsa, and year of RP are continuous, while pGS7 $= 1$ and pGS8H $= 0$ if pathological Gleason score was 7, pGS7 $= 0$ and pGS8H $= 1$ if pathological Gleason score was 8 or higher, and pGS7 $= 0$ and pGS8H $= 0$ if pathological Gleason score was 6 or less; pT3H $= 1$ if pathological stage was T3 or higher and 0 otherwise; Margin $= 1$ if the surgical margin was positive and 0 if surgical margin was negative. There are 216 censored and 342 failed patients in the data. The total number of ties is 215 and the maximum size of tied group is 16.

We fit the GP model with the seven covariates (age, logpsa, pGS7, pGS8H, pT3H, Margin, year of RP) to the data. In all the posterior computations, the covariates were standardized. A piecewise linear model was assumed for $F(t)$ in the GP model. The intervals $(a_{k-1}, a_k]$ were chosen to be the $(100k/K)^{\text{th}}$ percentile of the ordered distinct failure times for $k \leq K$. The model parameters included $\beta = (\beta_1, \ldots, \beta_7)'$ and $\lambda = (\lambda_1, \ldots, \lambda_K)'$. We computed DIC and p_D

under various values of c and K. The values of DIC are plotted in Figure 4. For all the values of c considered, the values of p_D range from 12.03 to 12.16 for $K = 5$, 17.03 to 17.13 for $K = 10$, and 22.12 to 22.23 for $K = 15$; these are almost the same as those corresponding numbers of parameters. The GP model with $c = 185$ and $K = 10$ attained the smallest DIC value among all of the combinations of (c, K) considered. However, as seen from Figure 4, the DIC values were very close for $170 \leq c \leq 200$. In fact, for $K = 10$, the DIC values were 2,939.63, 2,939.61, 2,939.49, 2,939.32, 2,939.40, 2,939.64 for $c = 170$, 175, 180, 185, 190, 200, respectively. To further verify this finding, we simulated 500 data sets of size $n = 558$ from the GP model with $c = 185$ under the simulation setting discussed in Section 6; the resulting median and IQR of N_{total} were 219 and (208, 229), closely matching $N_{\text{total}} = 215$ in the prostate cancer data. In addition, as shown in Figure 4, the GP model with $K = 10$ clearly outperforms those with $K = 5$ and $K = 15$ according to the DIC measure.

Under the best DIC GP model with $c = 185$ and $K = 10$, we computed the posterior means, posterior standard deviations (SD), and 95% HPD intervals of β. We also fit the PH model with the piecewise linear baseline hazard function with $K = 10$. Table 2 shows the maximum partial likelihood estimates (MPLEs) and Bayes estimates of β. From Table 2, we see that (i) under the PH model, the Bayes estimates were very close to the MPLEs; (ii) the estimates of β_1 and β_7 were very similar under the PH and GP models; (iii) the estimates of β_2, β_5, and β_6 under the PH model were slightly smaller than those under the GP model; and (iv) the estimates of β_3 and β_4 under the PH model were much smaller than those under the GP model. The difference in the estimates of β is expected as there were a large number of ties in the data and the best GP model was the one with $c = 185$ according to the DIC measure. Also, due to the large value of c, the difference in the estimates of β should not be too large as shown in our simulation study. When the regression coefficients were underestimated, the effects of the covariates could not be accurately assessed, which may lead to an incorrect conclusion regarding the impact of important clinical factors, such as a pathological Gleason score, on the risk of PSA failure.

In all the Bayesian computations in this section, we used 50,000 Gibbs iterations after a burn-in of 1,000 iterations to compute the posterior estimates, including DICs, posterior means, posterior standard deviations, and 95% HPD intervals. The convergence of the Gibbs sampling algorithm was checked and the autocorrelations for all model parameters disappeared before lag 5.

8. Discussion

We have carried out an in-depth investigation of the GP model and its properties. Our results are obtained as special cases of a general multivariate wear

Table 2. Estimates of β under the PH and GP Models for the prostate cancer data.

Method	Variable	Parameter	Estimate	SD*	95% Interval†
MPLE	age	β_1	0.003	0.008	(-0.012, 0.019)
	logpsa	β_2	0.282	0.066	(0.152, 0.411)
	pGS7	β_3	0.461	0.178	(0.113, 0.809)
	pGS8H	β_4	0.916	0.177	(0.570, 1.263)
	pT3H	β_5	0.533	0.145	(0.248, 0.818)
	Margin	β_6	0.532	0.121	(0.296, 0.769)
	year of RP	β_7	-0.054	0.014	(-0.082, -0.026)
Bayes	age	β_1	0.003	0.008	(-0.013, 0.019)
Based on	logpsa	β_2	0.270	0.066	(0.144, 0.403)
PH Model	pGS7	β_3	0.468	0.179	(0.114, 0.818)
	pGS8H	β_4	0.913	0.178	(0.569, 1.267)
	pT3H	β_5	0.534	0.146	(0.249, 0.823)
	Margin	β_6	0.527	0.121	(0.289, 0.762)
	year of RP	β_7	-0.052	0.014	(-0.079, -0.024)
Bayes	age	β_1	0.003	0.008	(-0.012, 0.019)
Based on	logpsa	β_2	0.309	0.067	(0.175, 0.437)
GP Model	pGS7	β_3	0.507	0.177	(0.163, 0.854)
	pGS8H	β_4	0.992	0.175	(0.639, 1.329)
	pT3H	β_5	0.553	0.144	(0.278, 0.840)
	Margin	β_6	0.566	0.120	(0.337, 0.806)
	year of RP	β_7	-0.053	0.014	(-0.081, -0.025)

* For MPLE, the values under the SD column are the standard errors of the estimates.

† For MPLE, the 95% intervals are the 95% confidence intervals while for Bayes, those intervals are the 95% HPD intervals.

process model. A novel DFS algorithm and a new Gibbs sampling algorithm have been developed that allow us to generate the tied failure times from the GP model, and to carry out posterior computations. The simulation study of Section 6 revealed some empirical properties of the GP model and the degree of biases of the parameter estimates when fitting the PH model to the data generated from the GP model.

One potential limitation of our analysis is its use of homogeneous Gamma process to model the baseline wear process H. This choice allows us to obtain the joint likelihood in a form explicit enough to achieve several goals, including posterior sampling of parameters and model selection. To our best knowledge, processes such as the Dirichlet and the Beta, do not yield such formulas for the joint likelihood. Furthermore, in S2 of the Supplementary Material, we argue that, under mild conditions, a Beta process is a homogeneous Gamma process plus an independent compound Poisson process with bounded Lévy density. This suggests that our sampling algorithm of failure times can be extended to Beta

processes. It also implies that Beta and Gamma processes have similar behavior at small jumps, while the former have more large jumps. Therefore, for data sets that exhibit few large jumps, these two should have similar performance as models for H. In S2, we also comment on how to extend the sampling algorithm to other types of pure jump processes.

In our simulation study and data analysis, we used DIC to determine the value of c when we fit the GP model to survival data with ties. Our simulation study showed that DIC was an effective measure in determining the true value of c. As an extension of this research, one can assume that c is an unknown parameter. With a prior distribution for c, posterior inference needs to be carried out. An unknown c may pose a computational challenge in sampling from its conditional posterior distribution. Theoretically, when there is a large number of ties, the PH model is not appropriate because under the model the probability of tied failure times is zero. When there are no ties in failure times, as shown in Proposition 2, the likelihood function under the GP model converges to the one under the PH model when the fitted wear process is concentrated. In practice, one can fit a GP model to survival data and then determine the "best" value of c according to DIC. When c is large, the PH model might be appropriate for fitting such survival data. Other extensions of the proposed methodology include time-dependent covariates, multivariate failure times, and non-proportional hazards models. These extensions are currently under investigation.

Supplementary Materials

The online supplementary material has two sections. Section S1 contains proofs of the theoretical results of the paper. Section S2 is a discussion on possible extension to wear processes other than the Gamma processes considered in the paper.

Acknowledgement

The authors thank the referees for their careful reviews, an associate editor, and the Editor. A suggestion of one of the referees inspired our development of the multivariate wear process model. The authors thank Dr. Anthony V. D'Amico of Brigham and Women's Hospital and Dana-Farber Cancer Institute for providing the prostate cancer data and advice on the clinical implication of PSA failure. Dr. Chen's research was partially supported by NIH grants #GM70335 and #P01CA142538.

References

Antelman, G. and Savage, I. R. (1965). Characteristic functions of stochastic integrals and reliability theory. *Naval Res. Logist. Quart.* **12**, 199-222.

Argiento, R., Guglielmi, A. and Pievatolo, A. (2010). Bayesian density estimation and model selection using nonparametric hierarchical mixtures. *Comput. Statist. Data Anal.* **54**, 816-832.

Bender, R., Augustin, T. and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statist. Med.* **24**, 1713-1723.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-100.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Probab.* **31**, 929-953.

Burridge, J. (1981). Empirical Bayes analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **43**, 65-75.

Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comput. Graph. Statist.* **8**, 69-92.

Chen, M.-H., Ibrahim, J. G. and Shao, Q.-M. (2006). Posterior propriety and computation for the Cox regression model with applications to missing covariates. *Biometrika* **93**, 791-807.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.

Chi, Z. (2012). On exact sampling of the first passage event of Lévy process with infinite Lévy measure and bounded variation. Technical Report 28, Department of Statistics, University of Connecticut. Available at arxiv.org with article id 1207.2495.

Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467-485.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269-276.

D'Amico, A. V., Chen, M.-H., Sun, L., Lee, W. R., Mouraviev, V., Robertson, C. N., Walther, P. J., Polascik, T. J., Albala, D. M. and Moul, J. W. (2010). Adjuvant versus salvage radiation therapy for prostate cancer and the risk of death. *BJU International* **106**, 1618-1622.

Damien, P., Laud, P. W. and Smith, A. F. M. (1996). Implementation of Bayesian nonparametric inference based on beta processes. *Scand. J. Statist.* **23**, 27-36.

Devroye, L. (1986). *Nonuniform Random Variate Generation*. Springer-Verlag, New York.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2**, 183-201.

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Stat.* **9**, 356-367.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72**, 557-565.

Epifani, I. and Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statist. Sinica* **20**, 1455-1484.

Epifani, I., Lijoi, A. and Prünster, I. (2003). Exponential functionals and means of neutral-to-the-right priors. *Biometrika* **90**, 791-808.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1**, 209-230.

Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Ann. Stat.* **7**, 163-186.

Gaver, Jr., D. P. (1963). Random hazard in reliability problems. *Technometrics* **5**, 211-226.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

Gold, L. S., Kane, L. B., Sotoodehnia, N. and Rea, T. (2007). Disaster events and the risk of sudden cardiac death: a Washington State investigation. *Prehosp Disaster Medicine* **22**, 313-317.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Stat.* **18**, 1259-1294.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-396.

Ibrahim, J. G., Chen, M.-H. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.

James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Stat.* **33**, 1771-1799.

James, L. F. (2006). Poisson calculus for spatial neutral to the right processes. *Ann. Stat.* **34**, 416-440.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **40**, 214-221.

Kim, Y. and Kim, D. (2009). Bayesian partial likelihood approach for tied observations. *J. Statist. Plann. Inference* **139**, 469-477.

Kim, Y. and Lee, J. (2003). Bayesian analysis of proportional hazard models. *Ann. Stat.* **31**, 493-511.

Kim, Y., Park, J. K. and Kim, G. (2011). Bayesian analysis for monotone hazard ratio. *Lifetime Data Anal.* **17**, 302-320.

Laud, P. W., Damien, P. and Smith, A. F. M. (1998). Bayesian nonparametric and covariate analysis of failure time data. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, 213-225. Springer, New York.

Lee, J. and Kim, Y. (2004). A new algorithm to generate beta processes. *Comput. Statist. Data Anal.* **47**, 441-453.

Lijoi, A., Mena, R. H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B* **69**, 715-740.

Lijoi, A., Prünster, I. and Walker, S. G. (2008a). Investigating nonparametric priors with Gibbs structure. *Statist. Sinica* **18**, 1653-1668.

Lijoi, A., Prünster, I. and Walker, S. G. (2008b). Posterior analysis for some classes of nonparametric models. *J. Nonparametr. Statist.* **20**, 447-457.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J. Amer. Statist. Assoc.* **89**, 958-966.

Lo, A. Y. and Weng, C.-S. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41**, 227-245.

Nieto-Barajas, L. E. and Walker, S. G. (2004). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statist. Sinica* **14**, 1127-1146.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84**, 487-493.

Peccati, G. and Prünster, I. (2008). Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Ann. Appl. Probab.* **18**, 1910-1943.

Peto, R. (1972). Contribution to the discussion of paper by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34**, 205-207.

Reynolds, D. S. and Savage, I. R. (1971). Random wear models in reliability theory. *Adv. Appl. Probab.* **3**, 229-248.

Rossi, P. H., Berk, R. A. and Lenihan, K. J. (1980). *Money, Work and Crime: Experimental Evidence*. Academic Press, New York.

Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Roy. Statist. Soc. Ser. B* **64**, 583-639.

Walker, S. and Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Stat.* **25**, 1762-1780.

Novartis Healthcare Private Ltd., Raheja Mindspace Building #6, Madhapur Hyderabad 500081, Andhra Pradesh, 500081, India.

E-mail: arijit.uconn@gmail.com

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, CT 06269, U.S.A.

E-mail: zhiyi.chi@uconn.edu

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, CT 06269, U.S.A.

E-mail: ming-hui.chen@uconn.edu

(Received November 2012; accepted November 2014)