<table>
<thead>
<tr>
<th><strong>Statistica Sinica Preprint No:</strong> SS-12-351R2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
</tr>
<tr>
<td><strong>Manuscript ID</strong></td>
</tr>
<tr>
<td><strong>URL</strong></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
BAYESIAN INFERENCE OF HIDDEN GAMMA WEAR PROCESS MODEL FOR SURVIVAL DATA WITH TIES

Arijit Sinha, Zhiyi Chi, and Ming-Hui Chen

Department of Statistics, University of Connecticut

Abstract: Real-life survival data often contain tied event times. Inference without careful treatment of the ties may lead to biased estimates. This paper develops Bayesian analysis of a stochastic wear process model to fit survival data that may 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, which 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, which provides a tool to assess the precision of inference. An extensive simulation study is conducted and a real 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, such as those during an annual sports event, ties occur naturally. For events that can happen at any point in time, ties may arise if a coarse time scale is used to record data (cf. Rossi et al., 1980). Still, even when continuous event times are recorded at a fine time scale, ties are prevalent. For example, 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). Strictly speaking, simultaneity hardly exists in any real-life study. Nevertheless, it can be more useful to model observed ties as the outcome of certain mechanism underpinning the events than to either account 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

\[ P(T > t \mid x) = \exp\{-H(t) \exp(x' \beta)\} \]  

(1.1)

where \( T \) is the failure time, \( x \) a vector of covariates, \( H \) a completely unspecified baseline cumulative hazard function, and \( \beta \) a vector of regression coefficients. In the non-Bayesian setting, inference from the model uses partial likelihood of \( \beta \), 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 may be made by applying the same formula for the no-tie case or by discretization of time (Cox, 1972; Peto, 1972; Breslow, 1974; Efron, 1977).

The perspective that \( H \) may be modeled as a stochastic process provides powerful ways to handle ties. In reliability analysis, Gaver (1963) postulated that \( H \) is a process with independent increments. Reynolds & 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 & Phadia, 1979; Hjort, 1990; Walker & Muliere, 1997; Epifani et al., 2003), and inference on \( \beta \) can be based on maximum likelihood estimation or Bayesian posterior analysis (Kalbfleisch, 1978; Hjort, 1990; Damien et al., 1996; Laud et al., 1998; Kim & Lee, 2003; Lee & Kim, 2004). Under these priors, tied event times occur with positive probability. On the other hand, a standard Bayesian approach that rules out ties is to impose priors on the baseline hazard rate function \( H' \), which leads to continuous \( H \) (Antelman & Savage, 1965; Reynolds & Savage, 1971; Dykstra & Laud, 1981; Lo & Weng, 1989; Clayton, 1991; Ibrahim et al., 2001; Nieto-Barajas & Walker, 2004; James, 2005, 2006; Lijoi et al., 2008; Peccati & Prünster, 2008; Kim & Kim, 2009; Kim et al., 2011). In data analysis, tie-breaking has been used to artificially transform tied event times into distinct ones (Kalbfleisch, 1978; Chen et al., 2006). However, the resulting estimate can be seriously biased if the proportion of ties is large (Burridge, 1981).
Following Gaver (1963), in this paper we regard $H$ as a hidden stochastic wear process underlying the failures, rather than a parameter with certain prior distribution. The contribution of the paper is twofold. First, based on the joint likelihood of $\beta$ 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 which 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 & Lijoi, 2010). We obtain the likelihood of the parameters using an argument similar in spirit to those for several special cases (cf. Reynolds & Savage, 1971; Lijoi et al., 2008a; Epifani & Lijoi, 2010). For homogeneous Gamma wear processes, we derive the likelihood in a semi-closed form. By imposing suitable noninformative priors on $\beta$ 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 $\beta$ (Damien et al., 1996; Laud et al., 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.

Second, we develop a Gibbs sampling based simulation algorithm, which is termed as 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 $\beta$ and covariates. The algorithm provides a useful simulation tool to evaluate the precision of the Bayesian inference. 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 preselected time points (Damien et al., 1996; Laud et al., 1998; Lee & Kim, 2004). Except for the approximation error of the Gibbs sampling of $H$ just before and at failure times as in posterior Gibbs sampling (cf. Laud et al., 1998), our sampling method is precise. In fact, by replacing Gibbs sampling with rejection sampling, truly exact sampling can be achieved (cf. 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 perform an extensive simulation study to examine the empirical properties of the Gamma wear process model. In this section, we propose to use DIC to guide the choice of parameters. In Section 7 we analyze a prostate cancer data set with our proposed methodology. Section 8 ends with a discussion and potential future research work on this topic. Proofs of the theoretical results 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 $t_i \leq c_i$.

Denote $\delta_i = I\{y_i = t_i\} = I\{t_i \leq c_i\}$. The observed data is $D_{\text{obs}} = \{y_i, \delta_i, x_i; i \leq n\}$, where $x_i$ is the vector of covariates of the $i^{th}$ subject. Define $D = \{i : \delta_i = 1\}$, $N = \{i : \delta_i = 0\}$, so $D$ consists of subjects that fail before censoring, and $N$ those that are censored. Denote by $0 < \tau_1 < \tau_2 < \ldots < \tau_N$ the distinct values of $y_1, \ldots, y_n$ and $\tau_0 = 0$. For $j \leq N$, let

\begin{align*}
\mathcal{D}_j &= \{i \in D : y_i = \tau_j\}, \\
\mathcal{A}_j &= \{i \in N : y_i = \tau_j\}, \\
\mathcal{R}_j &= \cup_{i \geq j}(D_i \cup N_i), \\
\mathcal{R}'_j &= \mathcal{R}_j \setminus \mathcal{D}_j, \tag{2.1}
\end{align*}

i.e., $\mathcal{D}_j$ consists of subjects that fail at time $\tau_j$, $\mathcal{A}_j$ those censored at $\tau_j$, $\mathcal{R}_j$ those at risk in time interval $(0, \tau_j]$, and $\mathcal{R}'_j$ those at risk in time interval $(0, \tau_j]$. Let $n_T = |\{j : \mathcal{D}_j \neq \emptyset\}|$ be the number of endpoints where failures occur. For $A \subset \{1, \ldots, n\}$, denote $\kappa_A = (a_1, \ldots, a_n)$ with $a_i = I\{i \in A\}$. For example, if $n = 5$, then $\kappa_{\{1,3,4\}} = (1,0,1,1,0)$. For brevity, write $\kappa_i = \kappa_{\{i\}}$. Denote

\begin{align*}
\varrho_j &= \kappa_{\mathcal{D}_j}, \\
\omega_j &= \kappa_{\mathcal{A}_j}, \\
\varrho_{N+1} &= \omega_{N+1} = (0,0,\ldots,0). \quad \text{All analyses will be conditional on } C_1, \ldots, C_n.
\end{align*}

A stochastic process $W = (W(t) : t \geq 0)$ is said to be additive if it has independent increments, is stochastically continuous, and with probability 1, the function $t \to W(t)$ is right-continuous in $t \geq 0$ with $W(0) = 0$ and has left limit
Bayesian Inference of Hidden Gamma Wear Process Model

5

In this paper, we always assume that \( W = (H_1, \ldots, H_n) \) is an additive process taking values in \( \mathbb{R}_+^n \) with \( \mathbb{R}_+ := [0, \infty) \), and refer to \( W \) as a pure jump process. It is well known that each \( H_i \) in \( W \) is nondecreasing and for \( a \in \mathbb{R}_+^n \),

\[
E[e^{-a'W(t)}] = e^{-\Psi(a, t)} \quad \text{with} \quad \Psi(a, t) = \int_0^t dv \int (1 - e^{-a's}) \varphi(ds | v),
\]

where given \( t > 0 \), \( \varphi(ds | t) \) is a Lévy measure on \( \mathbb{R}_+^n \) with \( \int \min(1, |s|) \varphi(ds | t) < \infty \) (Sato, 1999). We refer to \( \Psi \) as the characteristic exponent of \( W \). By Ferguson & Phadia (1979), \( W \) is said to be 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 \( I\{x > 0\}b^{-a}x^{a-1}e^{-x/b} \), \( \text{Exp}(c) = \text{Gamma}(1,c) \), and \( \delta \) the unit mass concentrated at 0. If \( F \) is a nondecreasing function on \( \mathbb{R}_+ \), then denote \( F^*(z) = \inf\{t > 0 : F(t) \geq z\} \) with the convention \( \inf\emptyset = \infty \).

3. Joint Likelihood for Wear Process Model

3.1. \( N \)-variate wear process model

We introduce an “\( n \)-variate” model as follows. It assumes that the \( n \) subjects each is exposed to a type of environmental fluctuations characterized by a nondecreasing stochastic process \( H_i \) with \( H_i(0) = 0 \) and \( H_i(\infty) = \infty \), such that \( W = (H_1, \ldots, H_n) \) is a pure jump process, and conditional on \( W \), the failure times \( T_1, \ldots, T_n \) of the subjects are independent, with

\[
P(T_i > t | W) = e^{-H_i(t)}, \quad i \leq n.
\]

We assume \( W \) is unobservable. We also assume the right-censoring times \( C_1, \ldots, C_n \) are independent from \( W \) and \( T_1, \ldots, T_n \).

The process \( W \) is referred to as a (cumulative) wear process (Gaver, 1963; Reynolds & Savage, 1971). We next give examples on how the \( n \)-variate model unifies various classes of Bayesian survival models.

Example 3.1 (PH model). In 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 \( \gamma_i \) is a constant that may incorporate covariates of the \( i^{th} \) subject. In this setting, \( H \) is often referred to as the baseline cumulative hazard function. To account for possible changes over time of the
covariates, one can instead assume \( P(T_i > t \mid H) = \exp\{-\int_0^t \gamma_i(v) \, dH(v)\} \), where \( \gamma_i \geq 0 \) is now a bounded nonrandom function such that \( \int_0^\infty \gamma_i(v) \, dH(v) = \infty \) with probability 1. Let \( H_i(t) = \int_0^t \gamma_i(v) \, dH(v) \) and \( W = (H_1, \ldots, H_n) \). Since \( W \) is \( H \)-measurable,

\[
P(T_1 > t_1, \ldots, T_n > t_n \mid W) = E[P(T_1 > t_1, \ldots, T_n > t_n \mid H) \mid W] = E \left[ \prod_{i=1}^n e^{-H_i(t_i)} \right] = \prod_{i=1}^n e^{-H_i(t_i)}.
\]

Therefore, the PH model can be formulated as an \( n \)-variate model with \( W \) being the wear process. Let \( \varphi_0(dx \mid t) \) be the Lévy measure of \( H \). Since \( E[e^{-aW(t)}] = E[e^{-\int_0^t \lambda(v) \, dH(v)}] = \exp\{-\int_0^t dv \int_0^\infty [1 - e^{-\lambda(v)x}] \varphi_0(dx \mid v)\} \), with \( \lambda(t) = a_1 \gamma_1(t) + \cdots + a_n \gamma_n(t) \), the characteristic exponent of \( W \) is

\[
\Psi(a, t) = \int_0^t dv \int_0^\infty [1 - e^{-a_1 \gamma_1(v)x} - \cdots - a_n \gamma_n(v)x] \varphi_0(dx \mid v), \quad a \in \mathbb{R}_+^n, \ t > 0.
\]

Consequently, the Lévy measure \( \varphi(ds \mid t) \) of \( W \), where \( s = (s_1, \ldots, s_n) \), is as follows. Given \( t > 0 \), if all \( \gamma_i(t) = 0 \), then \( \varphi(ds \mid t) = 0 \). On the other hand, if \( \gamma_i(t) > 0 \) for some \( i \), then

\[
\varphi(ds \mid t) = \varphi_0(ds_i/\gamma_i(t) \mid t) \prod_{j \neq i} \delta(ds_j - \gamma_j(t)s_i/\gamma_i(t)). \tag{3.2}
\]

Clearly, \( \varphi \) is determined by both \( \varphi_0 \) and \( \gamma_i \). In Bayesian analysis, given data, often only the parameters in \( \gamma_i \) are estimated. The parameters of \( \varphi_0 \) are regarded as hyperparameters and their estimation is treated as an empirical Bayes problem. However, under the \( n \)-variate model, this distinction between \( \varphi_0 \) and \( \gamma_i \) disappears, as both become parameters of the wear process \( W \).

\[\square\]

**Example 3.2 (Lévy copula).** A Lévy copula survival model was recently studied by Epifani & 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 \( P(T_i > t \mid Z) = e^{-Z_j(t)} \). 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
\[
\Psi(a, t) = \int_0^t dv \int (1 - e^{-a_1 s_1 - a_2 s_2 - \sum_{i > j} a_i s_i}) \varphi_0(ds_1, ds_2 | v),
\]
\[
\varphi(ds | t) = \varphi_0(ds_1, ds_2 | t) \prod_{i > 2} \delta(ds_j - s_j), \quad s = (s_1, \ldots, s_n)
\]
respectively, where $\varphi_0$ is the Lévy measure of $Z$. \hfill \Box

**Example 3.3** (Independent failure times). In the above examples, $H_i$ are dependent processes, making $T_i$ dependent random variables. On the other hand, if $H_i$ are independent, then $T_i$ are independent. If the Lévy measure of each $H_i$ is $\varphi_i(dx | t)$, then the characteristic exponent and Lévy measure of $\mathcal{W}$ are
\[
\Psi(a, t) = \sum_{i=1}^n \Psi_i(a_i, t), \quad \varphi(ds | t) = \sum_{i=1}^n \varphi_i(ds_i | t) \prod_{j \neq i} \delta(ds_j),
\]
respectively, where $\Psi_i(a_i, t)$ is the characteristic exponent of $H_i$. \hfill \Box

Under the $n$-variate model, when $H_i$ are dependent, the probability of ties among $T_i$ is positive. On the other hand, from (3.1)
\[
P(T_i > t) = \exp \left\{ - \int_0^t f_i(v) \, dv \right\}, \quad \text{with} \quad f_i(v) = \int (1 - e^{-s_i}) \varphi(ds | v) \geq 0,
\]
so each $T_i$ has a probability density. As a result, $T_i$ have to be dependent. It is thus interesting that $T_i$ are pairwise locally independent as defined by Oakes (1989). Let $X, Y > 0$ be random variables. For $t = (t_1, t_2)$, denote $S(t) = P(X > t_1, Y > t_2)$ and $D_0 = \partial / \partial t_0$. 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 3.1.** $T_1, \ldots, T_n$ are pairwise locally independent, i.e., for $i \neq j, \theta_{T_iT_j}^*(t) = 1$.

### 3.2. Likelihood function

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

**Theorem 3.2.** The likelihood function of $\varphi$ based on $D_{obs}$ is given by
\[
L(\varphi | D_{obs}) = \prod_{j=1}^N e^{-\Psi(\varphi_j, \tau_j)} \prod_{j \neq \emptyset} e^{-\omega_j^s} \prod_{i \in \emptyset_j} (1 - e^{-s_i}) \varphi(ds | \tau_j).
\]
(3.3)
Kalbfleisch (1978) observed that in the setting of Example 3.1, if $H$ is a Gamma process, then depending on its variability a spectrum of likelihoods can be obtained. To characterize this in general, reparametrize $\varphi(ds \mid t) = c\nu(c ds \mid t)$, with $c > 0$, where $\nu(ds \mid t)$ is a Lévy measure with support in $\mathbb{R}^+_n$. Suppose for all $i = 1, \ldots, n$ and $t > 0$,

$$m_i(t) := \int_0^t dv \int s_i \nu(ds \mid v) < \infty,$$

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

$$E[W(t)] = \int_0^t dv \int s \varphi(ds \mid v) = \int_0^t dv \int c s \nu(c ds \mid v) = m(t),$$

$$\text{Var}[W(t)] = \int_0^t dv \int ss' \varphi(ds \mid v) = \int_0^t dv \int c ss' \nu(c ds \mid v) = c^{-1}\Sigma(t).$$

Based on the identities, $c$ is called a precision parameter. In this paper, $c$ will be fixed and 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. For this reason, whenever $c$ is involved, we rewrite the likelihood as $L(\nu | c, D_{\text{obs}})$.

**Proposition 3.3.** Given $D_{\text{obs}}$, suppose we fit the model using $W$ with Lévy measure $c\nu(c ds \mid t)$ satisfying (3.4). Then, as the fitted process $W$ becomes concentrated, i.e., $c \to \infty$,

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

To see the implication of the result, consider Example 3.1 again. Suppose $\varphi_0(dx \mid t) = ch(cx, t) dx$ for $x > 0$. Letting $g(t) = \int_0^t sh(s, v) dv$, it can be seen that $m_i(t) = \int_0^t \gamma_i(v)g'(v) dv$. As a result, as $c \to \infty$, the likelihood tends to

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

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.
Note that the result implicitly assumes that $m$ is differentiable at every $\tau_j$ with $\mathcal{D}_j \neq 0$. Following the argument for the existence of probability density of $T_i$, this indeed holds with probability 1. Finally, the next result provides some information on the probability of ties for the case of most interest to us.

**Proposition 3.4.** Let $W$ be homogeneous such that for any $a \in \mathbb{R}^n_+$, $\Psi(a, t) = \Psi_1(a) \Psi_2(t)$. Then

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

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

### 3.3. A Gamma wear process model

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

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

with $c > 0$ being the precision parameter and $f = F'$. We refer to the corresponding $n$-variate model as the Gamma Process (GP) model. Clearly, the parameters of the model are $\gamma$, $F$, and $c$. However, as mentioned in Section 3.2, $c$ will be fixed via model selection and only $\gamma$ and $F$ will be estimated.

**Corollary 3.5.** The likelihood function for the GP model is

$$L(\gamma, F | c, D_{obs}) = \prod_{j=1}^N \left( \frac{c}{c + \omega_j \gamma} \right) e^{[F(\tau_j) - F(\tau_{j-1})]} \int_0^{\infty} s^{-1} e^{-(c+\omega_j \gamma)s} \prod_{i \in \mathcal{D}_j} (1 - e^{-\gamma_i s}) ds. \quad (3.6)$$

The proof of (3.6) is quick. As $d'W(t) = a'\gamma H(t) \sim \text{Gamma}(cF(t), a'\gamma/c)$ 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 on the other hand follows from that in (3.3) and (3.2).

When the data has no ties, (3.6) can be shown to coincide with formula (14) of Kalbfleisch (1978). On the other hand, to take into account ties, Kalbfleisch
ARIJIT SINHA, ZHIYI CHI, AND MING-HUI CHEN

(1978) derived a likelihood of regression coefficients in his formula (23), which, if expressed in integral form, is a part of the likelihood in (3.6) but with the factor \( \prod_{D_j \neq \emptyset} cf(\tau_j) \) missing. From Proposition 3.3, we immediately get the following result when the variability of \( H \) is small, i.e., \( c \to \infty \).

**Corollary 3.6.** Given \( D_{\text{obs}} \), let the GP model be fit with \( W = \gamma H \), where \( \gamma = (\gamma_1, \ldots, \gamma_n) \) with \( \gamma_i > 0 \) and \( H \sim \mathcal{GP}(cF, c) \). Fixing \( \gamma \) and \( F \), as \( c \to \infty \),

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

Finally, Proposition 3.4 leads to the following result on probability of ties.

**Corollary 3.7.** 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 \( c \to 0 \), \( P(T_i \text{ all equal}) \to 1 \), and as \( c \to \infty \), \( P(T_i \text{ all different}) \to 1 \).

It should be noted that in general, as the wear process becomes more variable, i.e., \( c \to 0 \), there is not necessarily \( P(T_i \text{ all equal}) \to 1 \). For example, let \( H \) be a generalized Gamma process with time-independent Lévy density \( c^2 h_1(cs) \), where \( h_1(s) = s^{-\alpha - 1}e^{-s} \), \( 0 < \alpha < 1 \) (cf. Hougaard, 1986; Brix, 1999; Epifani et al., 2003; Lijoi et al., 2007; Argiento et al., 2010). Then \( \Psi_1(\lambda) = (1 + \lambda/c)^{\alpha} - 1 \) and by Proposition 3.4, as \( c \to 0 \),

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

4. **Sampling of Survival Data from Gamma Process**

To sample failure times for the GP model, we start with the next result.

**Proposition 4.1.** Let \( F \) be strictly increasing. Denote \( \alpha(t) = cF(t) \). Let \( G \sim \mathcal{GP}(t, 1) \) be a standard Gamma process and \( \eta_i \) i.i.d. \( \text{Exp}(c) \) random variables independent of \( G \). Then \( (T_1, \ldots, T_n) \sim (\alpha^{-1}(G(\eta_1/\gamma_1)), \ldots, \alpha^{-1}(G(\eta_n/\gamma_n))) \).

Since \( \eta_i \) and \( G \) are independent, from the result, 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 (cf. Devroye, 1986). The
inversion is used by Bender et al. (2005) to sample failure times with no ties for the PH model. We next describe how to jointly sample \( G^*(\theta_i) \) for \( i \) such that \( 0 < \theta_1 < \ldots < \theta_n \) as follows. First, sample \( \tau_1 = G^*(\theta_1) \) from the distribution (4.1) and \( r_1 = G(\tau_1) \) conditional on \( \tau_1 \) from the distribution (4.2). If \( r_1 > \theta_n \), then all \( G^*(\theta_i) = \tau_1 \). Otherwise, letting \( s \) be the number with \( \theta_s \leq r_1 < \theta_{s+1} \), \( G^*(\theta_1) = \ldots = G^*(\theta_s) = \tau_1 < G^*(\theta_{s+1}) \leq \ldots \leq G^*(\theta_n) \). In general, if \( (\tau_1, r_1), \ldots, (\tau_k, r_k) \) have been sampled but there is \( s < n \) such that \( \theta_s \leq r_k < \theta_{s+1} \), then the next distinct failure time \( \tau_{k+1} = G^*(\tau_{k+1}) \) are sampled as follows. First, independently from all \( (\tau_j, r_j), j \leq k, \) sample \( \tilde{\tau}_{k+1} \sim G^*(\theta_{s+1} - r_k) \) from the distribution (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 the distribution (4.2). Then \( \tau_{k+1} = G^*(\theta_{s+1}) = \tau_k + \tilde{\tau}_{k+1} + \tilde{r}_{k+1} = r_k + \tilde{r}_{k+1} \). If \( r_{k+1} \geq \theta_n \), then all \( G^*(\theta_{s+1}) = \ldots = G(\theta_n) = \tau_k \), otherwise, sample the next distinct failure time \( \tau_{k+2} \) 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 a joint density

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

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

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

which 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 & Wild (1992) to sample $\zeta$ conditional on $G(\tau) = r$. For (ii), we have

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

which is an exponential density with mean $|r - \theta e^z/(1 + e^z)|^{-1}$. Finally, for (iii), $m_\theta(r \mid v) \propto e^{-r(1+v)}$, $r > \theta$, and hence sampling $r$ is also straightforward.

We thus obtain the following algorithm which can be followed step by step 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 $\eta_i$ i.i.d. $\sim \text{Exp}(c)$ and set $\theta_i = \eta_i / \gamma_i$ for $i \leq n$.
3. Rearrange $(\gamma_i, \theta_i)$ so that $0 < \theta_1 < \ldots < \theta_n$.
4. Generate $\tau_1$, which is a realization of $G^*(\theta_1)$.
5. Generate $r_1$, which is a realization of $G(\tau_1)$ conditional on $G^*(\theta_1) = \tau_1$.
6. If $r_1 \geq \theta_n$, then $G^*(\theta_1) = \ldots = G^*(\theta_n) = \tau_1$, i.e., all failure times are tied.
7. If $\theta_s \leq r_1 < \theta_{s+1}$, then $G^*(\theta_1) = \ldots = G^*(\theta_s) = \tau_1$.
8. Change $\theta_i$ to $\theta_i - r_1$ for $i \geq s + 1$.
9. Go to Step 4 to generate $t^*$, which is a realization of $G^*(\theta_{s+1})$. Note that $\theta_{s+1}$ are the values updated in Step 8.
10. Return the next distinct failure time $\tau_2 = \tau_1 + t^*$.
11. Go to Step 5 to simulate $G(t^*)$.
12. Go to Step 6 or Step 9 to examine the number of ties.
13. Follow Step 8 through Step 12 until all $G^*(\theta_1), \ldots, G^*(\theta_n)$ are generated.

14. Return $c^{-1}F^{-1}(G^*(\theta_i))$, which are the failure times from $\mathcal{G}\mathcal{P}(cF, c)$.

5. Bayesian Posterior Inference

Starting this section, 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 $\beta$ the vector of regression coefficients. Our goal is to develop posterior inference for $(\beta, F)$ 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 $a_K \geq \tau_N$. Then let $f(t) = F'(t) = \lambda_k$ for $a_{k-1} < t \leq a_k$. Under the model,

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

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

$$
L(\beta, \lambda | c, D_{\text{obs}}) = \prod_{j=1}^{N} \left( \frac{c}{c + g'_j \gamma} \right)^{c[F(\tau_j)-F(\tau_{j-1})]} \times \left[ c\lambda_{\nu(j)} \int_{0}^{\infty} s^{-1}e^{-(c+\omega_j \gamma)s} \prod_{i \in D_j} (1 - e^{-\gamma_i s}) \, ds \right]^{I(D_j \neq \emptyset)}. \tag{5.2}
$$

We assume the prior $\pi(\beta, \lambda) \propto \exp(-\beta'\Sigma^{-1}_0 \beta/2) \prod_{k=1}^{K} \lambda_k^{\alpha_0-1}e^{-\alpha_1 \lambda_k}$. Under the prior, $\beta, \lambda_1, \ldots, \lambda_K$ are independent, with $\beta \sim \mathcal{N}_p(0, \Sigma_0)$ and $\lambda_k \sim \text{Gamma}(\alpha_0, 1/\alpha_1)$. In Sections 6 and 7, we specify $\Sigma_0 = 10^4I_p$ and $\alpha_0 = \alpha_1 = 10^{-2}$, which lead to a relatively vague prior for $(\beta, \lambda)$.

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 latent variable \( s = (s_j : \mathcal{D}_j \neq \emptyset, j \leq N) \) and define an augmented joint posterior distribution \( \pi(\beta, \lambda, s \mid c, D_{\text{obs}}) \propto L(\beta, \lambda, s \mid c, D_{\text{obs}})\pi(\beta, \lambda), \) where

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

Since \( \int \pi(\beta, \lambda, s \mid c, D_{\text{obs}}) \, ds = \pi(\beta, \lambda \mid D_{\text{obs}}) \) by Corollary 3.5, \( \pi(\beta, \lambda \mid c, D_{\text{obs}}) \) can be sampled by applying Gibbs sampling to \( \pi(\beta, \lambda, s \mid c, D_{\text{obs}}) \). We sample \( (\beta, \lambda, s) \) from the following conditional posterior distributions in turn: (i) \( \pi(\beta \mid \lambda, s, c, D_{\text{obs}}) \); (ii) \( \pi(\lambda \mid \beta, s, c, D_{\text{obs}}) \); and (iii) \( \pi(s \mid \beta, c, D_{\text{obs}}) \). For (i),

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

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

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

Thus, \( \lambda_k \) follows a Gamma distribution that is easy to sample. Finally, for (iii), given \( \beta, s_1, \ldots, s_N \) are conditionally independent and for each \( j \) with \( \mathcal{D}_j \neq \emptyset \),

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

Let \( u_j = \ln s_j \). Then the conditional posterior density of \( u_j \) is given by

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

It is easy to show that \( \pi(u_j \mid \beta, D_{\text{obs}}) \) 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 $H \sim G(\mathcal{P}(cF,c))$. As the value of $c$ is unknown in practice, to guide the choice of $c$ in fitting the GP model, we propose to use deviance information criterion (DIC) (Spiegelhalter et al., 2002). Define the deviance function

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

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

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

(6.1)

where $\overline{\psi} = E[\psi | D_{\text{obs}}]$ and $p_D = D(\overline{\psi}) - D(\overline{\psi})$ with $D(\overline{\psi}) = E[D(\psi) | D_{\text{obs}}]$. In (6.1), $D(\overline{\psi})$ 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 GP model.

In the simulation study, the data was generated as follows. First, we generated $x_i = (x_{i1}, x_{i2})'$, $i \leq n$, where $x_{i1} \sim N(0,1)$, $x_{i2} \sim \text{Bernoulli}(0.7)$ 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. Second, 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 3 values of $c$, which were 1, 10, and 100. Third, 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. Finally, we independently generated 500 data sets under each simulation setting.

For each data set, we define the total number of ties as

$$N_{\text{total}} = \sum_j |D_j| I\{ |D_j| > 1 \},$$

and the maximum size of tied group as $N_{\text{max}} = \max_j |D_j|$, respectively, where $D_j$ is defined in (2.1). Figure 1 shows the boxplots of $N_{\text{total}}^{(1)}, \ldots, N_{\text{total}}^{(500)}$ and Figure 2 shows the boxplots of $N_{\text{max}}^{(1)}, \ldots, N_{\text{max}}^{(500)}$ of the 500 simulated data sets under the six simulation settings. From Figure 1, we can see that as $c$ increases from
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 $c = 1, 10, \text{ and } 100$.

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 $c = 1, 10, \text{ and } 100$. 
### Table 1: Summary of posterior estimates for the GP and the PH Models in simulation studies

<table>
<thead>
<tr>
<th>$n$</th>
<th>Parameter</th>
<th>True $c = 1$</th>
<th>True $c = 10$</th>
<th>True $c = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GP</td>
<td>PH</td>
<td>GP</td>
</tr>
<tr>
<td>250</td>
<td>$\beta_1$</td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Est</td>
<td>1.012 0.827</td>
<td>1.013 0.965</td>
<td>1.002 0.995</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.096 0.070</td>
<td>0.085 0.072</td>
<td>0.078 0.073</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.094 0.242</td>
<td>0.085 0.127</td>
<td>0.079 0.085</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.009 0.089</td>
<td>0.007 0.017</td>
<td>0.006 0.007</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.944 0.300</td>
<td>0.950 0.682</td>
<td>0.942 0.906</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>Est</td>
<td>-0.504 -0.407</td>
<td>-0.514 -0.482</td>
<td>-0.498 -0.494</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.178 0.158</td>
<td>0.152 0.156</td>
<td>0.139 0.156</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.157 0.176</td>
<td>0.142 0.158</td>
<td>0.128 0.144</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.025 0.040</td>
<td>0.020 0.025</td>
<td>0.016 0.021</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.974 0.894</td>
<td>0.962 0.956</td>
<td>0.966 0.970</td>
</tr>
<tr>
<td>500</td>
<td>$\lambda$</td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Est</td>
<td>1.055 0.903</td>
<td>1.020 0.964</td>
<td>1.011 1.007</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.225 0.116</td>
<td>0.135 0.123</td>
<td>0.113 0.128</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.234 0.782</td>
<td>0.136 0.254</td>
<td>0.109 0.145</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.057 0.619</td>
<td>0.019 0.066</td>
<td>0.012 0.021</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.940 0.262</td>
<td>0.944 0.654</td>
<td>0.962 0.924</td>
</tr>
<tr>
<td></td>
<td>$\beta_2$</td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.5</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>Est</td>
<td>-0.501 -0.400</td>
<td>-0.511 -0.489</td>
<td>-0.508 -0.500</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.129 0.113</td>
<td>0.113 0.111</td>
<td>0.098 0.110</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.117 0.156</td>
<td>0.098 0.112</td>
<td>0.094 0.108</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.014 0.034</td>
<td>0.010 0.013</td>
<td>0.009 0.012</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.966 0.754</td>
<td>0.972 0.952</td>
<td>0.948 0.952</td>
</tr>
<tr>
<td></td>
<td>$\lambda$</td>
<td>True</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Est</td>
<td>1.026 0.860</td>
<td>1.020 0.986</td>
<td>1.013 1.000</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.188 0.078</td>
<td>0.108 0.089</td>
<td>0.081 0.090</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.190 0.597</td>
<td>0.102 0.263</td>
<td>0.079 0.112</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.037 0.375</td>
<td>0.011 0.070</td>
<td>0.006 0.013</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.940 0.200</td>
<td>0.970 0.498</td>
<td>0.956 0.870</td>
</tr>
</tbody>
</table>
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.

For each simulated data set, we fit the PH model with a constant baseline hazard rate function and the GP model with \(F(t) = t\) and the true value of \(c\) used in the simulation. Also, for each simulated data set, we implemented the Gibbs sampling algorithm developed in Section 5.2 and used 5,000 Gibbs iterations after a burn-in of 500 iterations to compute all the posterior estimates. Let \(\hat{\beta}_j\) and \(\text{sd}_\ell(\beta_j)\) denote the posterior mean and the posterior standard deviation of \(\beta_j\) computed from the \(\ell\)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 \(\beta_j\) are defined, respectively, by

\[
\hat{\beta}_j = \frac{1}{500} \sum_{\ell=1}^{500} \hat{\beta}_{j\ell}, \quad \text{sd}(\beta_j) = \frac{1}{500} \sum_{\ell=1}^{500} \text{sd}_\ell(\beta_j),
\]

\[
\text{SE}(\beta_j) = \left[ \frac{1}{499} \sum_{\ell=1}^{500} (\hat{\beta}_{j\ell} - \hat{\beta}_j)^2 \right]^{1/2}, \quad \text{and} \quad \text{MSE}(\beta_j) = \frac{1}{500} \sum_{\ell=1}^{500} (\hat{\beta}_{j\ell} - \beta_j)^2,
\]

where \(\beta_j\) is the true value. Similarly, we define the same simulation summary statistics for \(\lambda\). In addition, 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 & Shao (1999). Table 1 shows these simulation summary statistics. We see that the GP model generally performed well and the posterior estimates were very close to the true values of \(\beta\) and \(\lambda\) and the coverage probabilities are close to 95\%. On the contrary, the PH model performs poorly and there were substantial biases in the posterior estimates especially when \(c\) is 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. However, the coverage probabilities were still smaller than the expected 95\% especially for \(\beta_1\) when \(n = 500\).

To examine the performance of DIC, for each simulated data set, we fit the GP model with \(c = 1, 10, \text{and} 100\) when the true \(c = 1 \text{ or} 100\). When the true \(c = 10\), we fit the GP model with \(c = 1, 10, 15, \text{and} 100\). In the DIC computation, we used the identity (3.6) to compute \(L(\beta, \lambda, c \mid D_{\text{obs}})\). 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 \(\ell\)th simulated data set. The boxplots of the DIC differences \(\Delta_{\ell}(c, c') = \)
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 $c = 100$. 
DIC_{c', \ell} - DIC_{c, \ell} for different values of \( c' \) and \( c \) in Figure 3 empirically 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 apart 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 whole box shown in Figure 3 is nearly above zero. However, the DIC differences in this case were much smaller.

7. Analysis of Prostate Cancer Data

We considered a subset of the data, which will be referred to as the PC data, consisting of 558 patients with high risk prostate cancer, namely, prostate specific antigen (PSA) > 20, or clinical Gleason score of 8 or higher, or clinical stage T3 or higher, from a prostate cancer study published by D’Amico et al. (2010). All patients in this subset were treated with radical prostatectomy (RP) between 1989 and 2008. In the PC data, the response is time to PSA failure or time to the last follow up from the time of RP whichever is minimum. The time of PSA failure is the time of prostate cancer recurrence after RP. The clinical implication of PSA recurrence is that men will be 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; and 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 PC 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 \((100k/K)^{th}\) percentile of the ordered distinct failure times for \(k \leq K\). The model parameters include \(\beta = (\beta_1, \ldots, \beta_7)'\) and \(\lambda = (\lambda_1, \ldots, \lambda_K)'\). For the PC data, we computed DIC and \(p_D\) under various values of \(c\) and \(K\). The values of DIC are plotted in Figure 4. The values of \(p_D\) range from 12.03 to 12.16 for \(K = 5\), 17.03 to 17.13 for \(K = 5\), and 22.12 to 22.23 for \(K = 15\), which are almost the same as those corresponding numbers of parameters, for all the values of \(c\) we considered. The GP model with \(c = 185\) and \(K = 10\) attained the smallest DIC value among all of the combinations of \((c, K)\) we 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 2939.63, 2939.61, 2939.49, 2939.42, 2939.40, and 2939.64 for \(c = 170, 175, 180, 185, 190, 200\), respectively. To further verify this finding, we generated 500 data sets of size \(n = 558\) from the GP model with \(c = 185\) under the simulation setting discussed in Section 6 and the resulting median and IQR of \(N_{\text{total}}\) were 219 and (208, 229). Recall that \(N_{\text{total}}\) in the PC data was 215, which is very close to 219. 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...
We also fit the PH model with the piecewise linear baseline hazard function with $K = 10$. Table 2 shows the maximum partial likelihood estimates (MPL) and Bayes estimates of $\beta$. From Table 2, we see that (i) under the PH model, the Bayes estimates were very close to the MPL; (ii) the estimates of $\beta_1$ and $\beta_7$ were very similar under both the PH and GP models; (iii) the estimates of $\beta_2$, $\beta_5$, and $\beta_6$ under the PH model were slightly smaller than those under the GP model; and (iv) the estimates of $\beta_3$ and $\beta_4$ under the PH model were much smaller than those under the GP model. The difference in the estimates of $\beta$ 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 $\beta$ should not be too large as shown in our simulation study. When the regression coefficients were under estimated, 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 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 1000 iterations to compute all 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

In this paper, we have carried out an in-depth investigation of the GP model and several interesting theoretical results of the GP model. These results are obtained as special cases of a general multivariate wear process model. A novel DFS algorithm and a new Gibbs sampling algorithm have also been developed, which allow us to generate the tied failure times from the GP model and to carry out posterior computation. The simulation study carried out in Section 6 further revealed some interesting empirical properties of the GP model and the degree of biases of the parameter estimates by 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 $H$. However, this choice allows us to
<table>
<thead>
<tr>
<th>Method</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
<th>95% Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MPLE</strong></td>
<td><strong>age</strong></td>
<td>$\beta_1$</td>
<td>0.003</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td><strong>logpsa</strong></td>
<td>$\beta_2$</td>
<td>0.282</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td><strong>pGS7</strong></td>
<td>$\beta_3$</td>
<td>0.461</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td><strong>pGS8H</strong></td>
<td>$\beta_4$</td>
<td>0.916</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td><strong>pT3H</strong></td>
<td>$\beta_5$</td>
<td>0.533</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td><strong>Margin</strong></td>
<td>$\beta_6$</td>
<td>0.532</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td><strong>year of RP</strong></td>
<td>$\beta_7$</td>
<td>$-0.054$</td>
<td>0.014</td>
</tr>
<tr>
<td><strong>Bayes</strong></td>
<td><strong>age</strong></td>
<td>$\beta_1$</td>
<td>0.003</td>
<td>0.008</td>
</tr>
<tr>
<td>Based on</td>
<td><strong>logpsa</strong></td>
<td>$\beta_2$</td>
<td>0.270</td>
<td>0.066</td>
</tr>
<tr>
<td><strong>PH Model</strong></td>
<td><strong>pGS7</strong></td>
<td>$\beta_3$</td>
<td>0.468</td>
<td>0.179</td>
</tr>
<tr>
<td></td>
<td><strong>pGS8H</strong></td>
<td>$\beta_4$</td>
<td>0.913</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td><strong>pT3H</strong></td>
<td>$\beta_5$</td>
<td>0.534</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td><strong>Margin</strong></td>
<td>$\beta_6$</td>
<td>0.527</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td><strong>year of RP</strong></td>
<td>$\beta_7$</td>
<td>$-0.052$</td>
<td>0.014</td>
</tr>
<tr>
<td><strong>Bayes</strong></td>
<td><strong>age</strong></td>
<td>$\beta_1$</td>
<td>0.003</td>
<td>0.008</td>
</tr>
<tr>
<td>Based on</td>
<td><strong>logpsa</strong></td>
<td>$\beta_2$</td>
<td>0.309</td>
<td>0.067</td>
</tr>
<tr>
<td><strong>GP Model</strong></td>
<td><strong>pGS7</strong></td>
<td>$\beta_3$</td>
<td>0.507</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td><strong>pGS8H</strong></td>
<td>$\beta_4$</td>
<td>0.992</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td><strong>pT3H</strong></td>
<td>$\beta_5$</td>
<td>0.553</td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td><strong>Margin</strong></td>
<td>$\beta_6$</td>
<td>0.566</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td><strong>year of RP</strong></td>
<td>$\beta_7$</td>
<td>$-0.053$</td>
<td>0.014</td>
</tr>
</tbody>
</table>

* 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.

Table 2: Estimates of $\beta$ under the PH and GP Models for the PC Data
get joint likelihood in a form explicit enough to achieve several goals, including posterior sampling of parameters and model selection. To our best knowledge, other popular processes, such as Dirichlet processes and Beta processes, do not yield such formulas for joint likelihood. Furthermore, in S2, we will argue that under mild conditions, a Beta process is a homogeneous Gamma process added with 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 has more large jumps. Therefore, for data sets that exhibit few large jumps, these two should have similar performance as models for \( H \). In Section S2 of Supplementary Material, we will comment on how to extend the sampling algorithm to other types of pure jump processes.

In our simulation study and real 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 empirically showed that DIC was an effective measure in determining the true value of \( c \). As an extension of this research, we may assume that \( c \) is an unknown parameter. Thus, a prior distribution needs to be specified for \( c \) and consequently, the posterior inference of \( c \) needs to be carried out. However, an unknown \( c \) may pose a computational challenge in sampling \( c \) 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 3.3, the likelihood function under the GP model converges to the one under the PH model when the fitted wear process is concentrated. Thus, in practice, one may fit a GP model to survival data and then determine the “best” value of \( c \) according to DIC. When \( c \) is large, then the PH model may be appropriate to fit 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.

Acknowledgment. The authors would like to thank the careful reviews of two referees, the Associate editor, and the editor. A suggestion of one of the referees inspired our development of the multivariate wear process model. The authors would also like to 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 #GM 70335 and #CA 74015.

References


Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, CT 06269, U.S.A.
E-mails:
arijit.uconn@gmail.com, zhiyi.chi@uconn.edu, ming-hui.chen@uconn.edu