We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Dependent Dirichlet Processes for Analysis of a Generalized Shared Frailty Model

Chong Zhong, Zhihua Ma, Junshan Shen and Catherine Liu

Abstract

Bayesian paradigm takes advantage of well-fitting complicated survival models and feasible computing in survival analysis owing to the superiority in tackling the complex censoring scheme, compared with the frequentist paradigm. In this chapter, we aim to display the latest tendency in Bayesian computing, in the sense of automating the posterior sampling, through a Bayesian analysis of survival modeling for multivariate survival outcomes with the complicated data structure. Motivated by relaxing the strong assumption of proportionality and the restriction of a common baseline population, we propose a generalized shared frailty model which includes both parametric and nonparametric frailty random effects to incorporate both treatment-wise and temporal variation for multiple events. We develop a survival-function version of the ANOVA dependent Dirichlet process to model the dependency among the baseline survival functions. The posterior sampling is implemented by the No-U-Turn sampler in Stan, a contemporary Bayesian computing tool, automatically. The proposed model is validated by analysis of the bladder cancer recurrences data. The estimation is consistent with existing results. Our model and Bayesian inference provide evidence that the Bayesian paradigm fosters complex modeling and feasible computing in survival analysis, and Stan relaxes the posterior inference.

Keywords: ANOVA DDP, dependent treatments, multivariate survival outcomes, recurrence, Stan

1. Introduction

The shared frailty model, coined by [1], has been widely used in the analysis of multivariate survival outcomes that might be associated with subgroups or clusters. Enormous work has been devoted to the development of the shared frailty model in both Bayesian and frequency paradigms, and the reviews can be found in [2, 3]. As an extension of the well-known Cox's proportional hazard model, *conditional on the frailty effect*, the traditional shared frailty model assumes a proportional hazards structure, that is, the hazard ratio between two sets of covariate values is proportional to their difference in relative risk scores over time [4]. Meanwhile, it fixes the baseline hazard function among all clusters.

Traditional shared frailty models provide a good framework for expediently mathematical tractability of the heterogeneity among multivariate observations,

whereas in practice it needs modification and adaption to tolerate complex structure so as to incorporate cross information owing to the intra- and inter-subject variability [5, 6]. Take the renowned data on recurrences of bladder cancer, for instance [7]. There are three treatment arms, placebo, thiotepa, and pyridoxine. Patients had multiple recurrences of tumors which were sparse beyond the fourth recurrence. **Figure 1** displays the Kaplan-Meier estimators of the survival-function for the times of the first and the second recurrences under three treatments. One observes that in **Figure 1(a)**, the estimated survival curves at the first recurrence are crossed, indicating a crossed hazard, and therefore, the proportional hazard assumption is suspected [8]; in **Figure 1(b)**, the survival curve of pyridoxine drops below that of placebo from the fifth month at the second recurrence rather than above from the 10th month on at the first recurrence. This indicates the functional form of the survival curves varies between recurrences. Neglecting such characteristics of non-proportionality and stratification of recurrences may yield inefficiency by encumbering borrowing strength from potentially related information sources, and consequently may jeopardize the prediction of the global survival times. Moreover, dependency might exist among the treatment strata and the stratification of recurrences [5, 9].

Consequently, more complex modeling is needy to characterize the dependence among the baseline hazard functions and treatment strata due to the temporal effects of recurrences. Frequentist inference and computing are pretty challenging and even infeasible within the most complex model setting. In Bayesian literature, there exists work that allows the baseline survival/hazard function to vary on a single level such as subgroups or the time axis ([10, 11]; among others). Nevertheless, rare work has taken bi-level varying baseline survival/hazard function into account [12], not to mention that dependence among treatment strata [5].

We propose a generalized shared frailty model (GSFM) for multiple events time data that allows the baseline hazard function to change dually along with the types of events and treatment strata, so as to strengthen the ability to borrow information from many sources. The proposed GSFM postulates multiplier frailties including both parametric and nonparametric ones, where the parametric frailty random effect accounts for the within-subject association by treating each subject as a cluster; and a nonparametric frailty effect represents dependency among treatment strata and temporal recurrences. For the GSFM, we suggest a Bayesian solution to estimate the regression coefficient vector, the variance parameter of the frailty
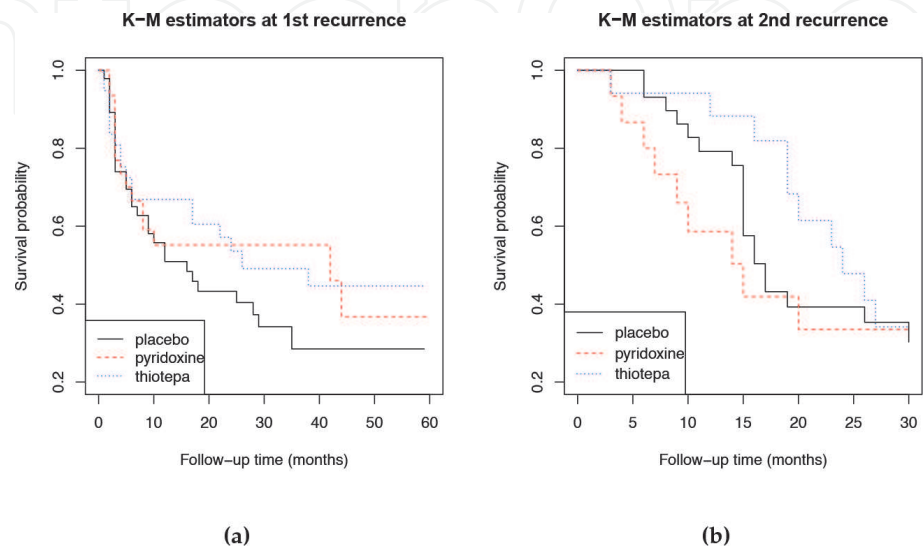

Figure 1.
The Kaplan-Meier estimator of survival functions for first recurrence time (a) and second recurrence (b) in the bladder cancer data.

term, and baseline survival functions stratified by treatments and recurrences. In a Bayesian workflow, the posterior distribution is determined by the combination of observational data in the form of the likelihood function and the prior distribution represented based on the background knowledge. From a Bayesian perspective, we model the dependent nonparametric prior through transferring the data context aforementioned into the ANOVA dependent Dirichlet process (ANOVA DDP), which will be further reviewed in Section 2. The construction of the No-U-Turn sampler for Markov chain Monte Carlo (MCMC) sampling is automated by Stan [13] with its R interface [14]. The posterior inference is conducted by Stan as well.

The rest of this chapter is organized as follows. In Section 2, under typical data scenarios of dependence structure, we summarize several modified versions of the dependent Dirichlet process (DDP) initiated from MacEachern's regression spirit that nests dependent predictors into the traditional Dirichlet process (DP). In Section 3, we postulate the GSFM and transform the dependent dual-stratified multiple events to the survival-function version of the ANOVA DDP. We have a short comparison between Stan and Nimble, two contemporary Bayesian computing tools based on our user experience. In Section 4, we demonstrate the validity of the GSFM, its Bayesian inference, and analysis of the data on recurrences of bladder cancer. A brief conclusion is contained in Section 5.

2. Review of MacEachern's DDP

The DP is the most popular Bayesian nonparametric prior since the seminal work of [15]. The belief in data background that there exists some kind of dependence structure stimulates construction and selection of proper dependent prior. Some dependent DPs are constructed for unsupervised purposes such as clustering [16, 17]. The DDP prior adopted in our proposed model is supervised and predictor-dependent, originated from [18, 19], named as MacEachern's DDP in two recent review papers, which are interpretive and comprehensive [20, 21]. The key idea behind MacEachern's DDP is that the distributions of the random measures are marginally DP distributed, validated by in our Subsections 3.2 and 3.3. Therefore, we here confine how MacEachern's DDP (henceforth we use the DDP to denote the MacEachern's DDP if the context is clear) came into being expanded from the DP; and compare various modified versions of the DDP under various dependency structures. We focus on two fundamental elements of Sethuraman's construction of DP [22], the weight and the atom, following the insight of [21].

2.1 DP vs. DDP

The DP is a distribution on distributions whereas the DDP aims to construct a prior for a collection of distributions $\mathscr{F} = \{F_x | x \in \mathcal{X}\}$ indexed by covariate x. In general, there are several representations of the DP such as Polya Urn, Levy measure, and stick-breaking representations [23]. Here, we use Sethuraman's stick-breaking construction to connect the DP with the DDP. The stick-breaking construction is a kind of infinite sum representation that divides the DP into two countable series, the stick-breaking weights (SBW) and the atoms. Generally, a DP is expressed as a process with two components, the mass parameter determining the weights and the base measure to generate atoms. Through the stick-breaking construction, the DDP can be easily extended from the DP. We list their comparison in **Table 1**, where we can find that the dependency among the covariates set $\mathcal{X}$ is realized by indexing the mass parameter and base measure with the covariate $x \in \mathcal{X}$.

More specifically, the dependency can be characterized through the dependency among the weights and atoms in the DDP.

The DDP can be widely applied to scenarios of various dependence data structures. We review modification versions of the DDP from three categories depending on which part it modifies in the stick-breaking representation, weights, atoms, or both. The first is to impose the dependency on the atoms but keep common weights, leading to two typical representatives, ANOVA and Spatial [9, 24, 25]. The ANOVA type DDP encoded the covariate dependence in the form of regression for the atom processes. The Spatial DDP models for nonstrationary spatial random fields with heterogeneous variance. The second category is to modify the weights to be dependent but keep the common atoms. The early and typical work is the time series DDP [26]. They introduced a Markov Beta process on the weights to account for the temporal dependency. The third category is to impose dependency on both weights and atoms [27]. They constructed vector autoregressive and autoregressive models for atoms and weights, respectively. We summarize the aforementioned types of typical modifications in **Figure 2**.

3. Model and Bayesian inference

Consider a clinical trial with multiple event types, for example, the time of the kth recurrence of a certain disease. In the trial, n subjects are divided into G strata of treatment. Our goal is to describe the relationship between the time to the kth recurrence of a subject, and its treatment stratum, as well as its vector of covariates Z. For a certain subject, the time of recurrences, may be dependent since they occur on the same individual and thus we assume an unobservable independent shared-frailty random effect W to account for this dependence. On the other hand, we may allow the conditional hazard affiliated with the script pair kj to imply distinct survival distributions along with the temporal order of the recurrences of the disease and for specific treatment. For the ith subject in the jth treatment stratum, at the kth recurrence, given the value of frailty variable w_i and its covariate vector z_{kji}, we propose the following frailty model,

$$\lambda_{kj}(t|w_i, z_{kji}) = w_i \lambda_{0kj}(t) \exp\left(\beta^T z_{kji}\right), k = 1, \cdots, K, j = 1, \cdots, G, i = 1, \cdots, n_j. \quad (1)$$

Model (1) is called the *generalized* shared frailty model in the sense that non-proportionality among k-varying recurrences is allowed by the fact that the right-hand baseline hazard has footnotes k and j. We allow dependency among treatment strata in the model (1). Therefore, the baseline hazard function λ_{0kj} acts as if a nonparametric frailty random measure accounting for the dependency owing to the recurrences and treatment schemes.

	DP	DDP
Random probability measure	$F \sim \mathrm{DP}(M, F_0)$	$\mathcal{F} = \{F_x | x \in \mathcal{X}, M_x, F_{0x}\}$
Sethuraman's construction	$F(\cdot) = \sum_{h=1}^{\infty} p_h \delta_{\theta_h}(\cdot)$	$F_x(\cdot) = \sum_{h=1}^{\infty} p_{xh} \delta_{\theta_{xh}}(\cdot)$
	$p_h \sim \mathrm{SBW}(1, M)$	$p_{xh} \sim \mathrm{SBW}(1, M_x)$
	$\theta_h \sim F_0$	$\theta_{xh} \sim F_{0x}$
Convolution	$H(y) = \int k(y|\theta) dF(\theta)$	$H_x(y) = \int k(y|\theta) dF_x(\theta)$

Table 1.
Comparison of DP and DDP.

| | Dependent atoms | | Dependent weights and atoms | Dependent weights |
	ANOVA DDP — Spatial DDP	Spatial DDP	Discrete-Time-Dependent DPM	tsDDP
DDP	$F_x(\cdot) = \sum_{h=1}^\infty p_h \delta_{\theta_{xh}}(\cdot)$	$G_D(\cdot) = \sum_{h=1}^\infty p_h \delta_{\theta_{h,D}}(\cdot)$	$F_t(\cdot) = \sum_{h=1}^\infty p_{th} \delta_{\theta_{th}}(\cdot)$	$F_d(\cdot) = \sum_{h=1}^\infty p_{dh}\theta_h(\cdot)$
Atoms	$\theta_{xh} = \alpha_h^\top d_x$ $\alpha_h \sim F_0$	$\theta_{h,D} \sim G_0$ $G_0^{(n)} \sim N_n(0,\Sigma)$	$\theta_{th} = (\mu_{th}, \Sigma_h)$ $\mu_{th} \sim N(\mu_0, V_o)$ $\mu_{t+1,h}\vert\mu_{th} \sim N(m + \Theta\mu_{th}, V)$	$\theta_h \sim F_0$
Weights	$p_h \sim SBW(1,M)$	$p_h \sim SBW(1,M)$	$p_{th} = (1-\beta_{th})\Pi_{j=1}^{h=1}\beta_{tj}$ $\beta_{th} = \exp(-(2\alpha)^{-1}(\xi_h + \eta_{th})^2)$ $\xi_h, \eta_{1h} \sim N(0,1)$ $\eta_{t+1,h}\vert\eta_{th} \sim N(\phi\eta_{th}, 1-\phi^2)$	$p_{dh} = v_{dh}\Pi_{j=1}^{h-1}(1-v_{dj})$ $v_{1h} \sim \text{Beta}(1,M)$ $z_{dh}\vert v_{dh} \sim B(m_{dh}, v_{dh})$ $v_{d+1,h}\vert v_{dh} \sim \text{Beta}(1+z_{dh}, M-z_{dh}+m_{dh})$
Convolution	$H_x = \int N(\alpha^\top d_x, s)dF(\alpha)$ $F \sim DP(M, F_0)$	$H_D = \int N(\theta_D, \sigma^2)dG(\theta_D)$ $G \sim DP(M, G_0)$	$H_t = \int N(\mu, \Sigma)dF_t(\mu, \Sigma)$	No convolution
Authors	De Iorio, Müller, Rosner and MacEachern (2004, JASA)	Gelfand, Kottas and MacEachern (2005, JASA)	DeYoreo and Kottas (2018, JASA)	Nieto-Barajas, Müller, Ji, Lu and Mills (2012, Biometrics)

Figure 2.
Workflows of representative expansions of DDP.

Model (1) is an extension of the classical shared frailty model (4.1.1) on page 101 of [4] since the baseline hazard function there does not vary among the recurrences and the treatment strata. Model (1) has the analog spirit to the frailty model (1) in [10], whereas their treatment strata are independent.

3.1 Likelihood

The corresponding survival function of the model is given by: $S_{kj}(t|w_i, z_{kji}) = \{S_{0kj}(t)\}^{\exp(\beta^T z_{kji} + v_i)}$, where $S_{0kj}(t) = \exp\left(-\int_0^t \lambda_{0kj}(s)ds\right)$ denotes the baseline survival function of the kth recurrence for subjects in the jth treatment stratum, and $v_i = \log(w_i)$ denotes logarithm transformation of the frailty effect. Let f_{0kj} be the corresponding baseline density function.

Given the data sample $(Y_{kji}, \delta_{kji}, z_{ji})$, where $Y_{kji} = \min(C_{kji}, T_{kji})$, $\delta_{kji} = I(T_{kji} \leq C_{kji})$, with T_{kji} being the gap time between the $(k-1)$th and kth recurrence of the ith subject in the jth stratum and C_{kij} being the corresponding censoring variable that is independent of T_{kji} given the covariate vector z_{kji}, for $k = 1, \cdots, K, j = 1, \cdots, G, i = 1, \cdots, n_j$, and $\sum_{j=1}^G n_j = n$. In the jth stratum, suppose that there are n_{kj} ($n_{kj} \leq n_j$) subjects suffering from the kth recurrence. Then the likelihood is written as:

$$\prod_{k=1}^K \prod_{j=1}^G \prod_{i=1}^{n_{kj}} \left[\exp(\beta^T z_{kji} + v_i) f_{0kj}(y_{kji}) \left\{ S_{0kj}(y_{kji}) \right\}^{\left(\exp(\beta^T z_{kji} + v_i) - 1 \right)} \right]^{\delta_{kji}}$$
$$\times \left\{ S_{0kj}(y_{kji}) \right\}^{(1 - \delta_{kji}) \exp(\beta^T z_{kji} + v_i)}.$$

3.2 Survival-function version of the ANOVA DDP

In the Bayesian workflow for the estimation, prior distributions are first determined. We here specify appropriate nonparametric priors for S_{0kj} and f_{0kj}. Since they can be easily derived from one to the other, we here only introduce the priors for S_{0kj}.

We divide S_{0kj} into K groups, and the kth group has G baseline survival functions of different treatment strata at the kth time of recurrence. That is, for a fixed k, $\mathcal{S}_k = \{S_{0kj}, j = 1, \cdots, G\}$ is a collection of baseline survival functions with length G indexed by the categorical covariate j denoting the treatment stratum. The next procedures come from the spirit of [9]. As a general example, suppose two dugs A and B will be taken in treatment, with V and U levels of doses, respectively. In this case, $G = VU$ denotes the number of treatment strata and let the level of the jth stratum be (v, w). We write the stick-breaking form of S_{0kj} such that $S_{0kj}(t) = \sum_{h=1}^{\infty} p_h I(t > \theta_{kjh})$. We impose an ANOVA structure on θ_{kjh}:

$$\theta_{kjh} = m_{kh} + A_{kvh} + B_{kwh}, \tag{2}$$

where m_{kh} denotes the ANOVA effect shared by all the strata at the kth recurrence, and the rest terms are the ANOVA effects of the jth stratum at the kth recurrence. Let the three components be independently generated from three distributions, and marginally on j, the baseline survival function S_{0kj} follows a DP. The aforementioned procedure implies that $\mathcal{S}_k$ is a survival-function version of the

ANOVA DDP. If we consider a single event, that is, $K = 1$, with linear effects of a regressor vector z involved in (2) such that $\theta_{jh} = m_h + A_{vh} + B_{wh} + \beta_h^T Z$, then it reduces to the univariate survival regression model in [25].

Since any function in the stick-breaking form is discrete almost surely, we place a convolution through the Dirichlet process mixture (DPM) model [28]. Particularly, since the baseline survival functions are defined on the positive half-real line, the convolution kernel in DPM should be positive such as log-normal, Gamma, and Weibull. In this chapter, a log-normal kernel is considered. For different recurrences, we treat the relationship among $\mathcal{S}_k s$ to be independent.

3.3 One-way ANOVA DDP

Considering the data of our interest, where only one drug and one level of dose is used in each treatment stratum, we introduce the modeling of the survival-function version of one-way ANOVA DDP. In this case, the prior for the $\mathcal{S}_k$ reduces to a one-way ANOVA form since the dependency among the G treatment strata is explained by only one ANOVA effect. Furthermore, if we set $m_{kh} = 0$, $\alpha_{kh} = (\theta_{k1h}, \cdots, \theta_{kGh})^T$ reduces to a G-variate variable denoting the locations of all G baseline distributions and thus $\theta_{kjh} = \alpha_{kh}^T d_j$, where d_j is the design vector of the jth stratum to select the appropriate ANOVA effects corresponding to j.

With the above notations, we summarize the procedure to construct the survival-function version of one-way ANOVA DDP prior in model (1) as follows:

1. Stick-breaking form. For $k = 1, \cdots, K$, let $\mathcal{H}_k$ be the collection of G distribution functions s.t $\mathcal{H}_k = \{H_{kj}, j = 1, \cdots, G\}$. $H_{kj}(\cdot) = \sum_{h=1}^{\infty} p_{kh} \delta_{\theta_{kjh}}(\cdot)$.

2. Convolution step. Let $\alpha_{kh} = (\theta_{k1h}, \cdots, \theta_{kGh})^T$, and d_j be the jth design vector of length G with the jth element being 1 and others being 0. Let $H_{0k} = (H_{0k1}, \cdots, H_{0kG})$ be the collection of base measures, $S_{0kj}(t) = \int S_{\text{LN}}(t|\alpha_k^T d_j, \sigma^2) dH_k(\alpha, \sigma)$, where S_{LN} denotes the survival function of the log-normal distribution, and $H_k \sim \text{DP}(M_k, H_{0k})$.

3. Determine the mass parameter and the base measure. For simplicity, we set $M_k = 1$ for all k, which is a commonly used default value of the mass parameter [29], $H_{0k}(\theta, \sigma) = N(0, I_G) \times \text{Cauchy}(0, 5)^+$, where Cauchy$^+$ denotes the half_Cauchy distribution.

Step 1 is a standard stick-breaking representation for DP. Step 2 is kernel mixture of DP whereas the kernel is a survival function rather than a cumulative distribution function. The realization of Step 2 is quite straightforward in Stan as it provides the function lognormal_lccdf to be used as the kernel of the survival function of the log-normal family.

In Step 3, we specify the base measure as the prior for the location and shape parameters of the log-normal kernel directly rather than adding another hyper prior distribution like [25] did. The main reason is to simplify the computation in Stan. Particularly, inspired by [30, 31], we use the half-Cauchy distribution as the non-informative prior for the variance parameter instead of the inverse Gamma prior. In our practice, the choice of half-Cauchy prior significantly improves the speed of convergence and mixture performance of the MCMC chains in our real data analysis and simulation. Another interesting point we met in numerical studies is that the informativeness of the base measure for θ. Here, we do not assign the non-

informative distribution but a weakly informative one is considered since we find such a weakly informative prior provides better MCMC performance than that of non-informative one with a higher effective sample size and better mixture performance. In our other research experience, the weakly informative prior for the variance parameter in the mixing component of the DPM seems to be more preferable.

3.4 Other priors and MCMC

In terms of the prior for the parametric prior w_i, we choose to log-normal prior that $v_i = \log(w_i)$ and $v_i \sim N(0, \tau^2)$, where $\tau > 0$ is an unknown parameter. We further assign a half Cauchy prior for τ s.t $\tau \sim \text{Cauchy}^+(0, 5)$ as a non-informative prior. The prior for the vector of regression coefficients is $\beta \sim N(0, 1000I)$ as a non-informative prior.

We use the truncated Dirichlet process to replace the infinite summand in the DP. The selection of the truncation point is often ad-hoc. Since in Stan the NUTS cannot sampler discrete parameters, we have to fit the truncation number and the mass parameter before the MCMC procedure. In general, the truncation number is set to be large enough s.t the truncated part is negligible. Gelman et al. [32] suggests using a truncation number L that is greater than $5M + 5$. In our computation, we set $L = 12$.

The MCMC sampling for the posterior distribution is realized in Stan. Stan and its **R** version are widely used in statistical modeling and high-performance statistical computing, especially in Bayesian. Stan realizes the MCMC sampling through the No-U-Turn sampler (NUTS). Stan automates the deriving of the fully conditional posterior distribution and NUTS is able to obtain high effective sample size [33].

3.5 Stan and NIMBLE: Programming styles

The MCMC sampling procedure is implemented in Stan and we also tried to implement the model in NIMBLE, another contemporary Bayesian computing tool in **R**. Stan and NIMBLE are two contemporary Bayesian computing tools that have drawn arising interest for Bayesian analysis but still remain under active development [34, 35]. The main advantage of Stan and NIMBLE is that they provide clear automatic posterior sampling procedures based on their specific sampling algorithms without particular justification. Therefore, users can be released from complicated probabilistic deriving and implementation. There has been a buzz group discussion about the comparison between Stan and NIMBLE in environments like [36, 37]. One comparison of their built-in samplers is demonstrated through implementing weakly informative and informative estimation within the trimmed mean regression model setting [38]. Here we contribute a naive comparison on their programming styles based on the first two authors' experience in coding this project and using Stan and NIMBLE, respectively.

A Bayesian paradigm is made up of three main steps, the prior, likelihood, and the posterior. MCMC generates samples to approximate the posterior distribution. Therefore, what one needs to set in a Bayesian computing tool is the prior and likelihood, let alone Stan or NIMBLE. Nevertheless, Stan and NIMBLE take different programming styles in writing likelihood. In Stan, the default way to present the log-likelihood is the syntax target and users can add log contribution to it freely, which is similar to the natural language and straightforward to users whatever level of mathematical background. In NIMBLE, the default way is to transfer the likelihood into some standard distributions given by NIMBLE, which may not be friendly for users who have a relatively less mathematical background.

We take fitting the finite mixture of the Gaussian model as an example. For a fixed positive integer L, the distribution of Y is given by $F_Y(s) = \sum_{l=1}^{L} p_l N(s|\mu_l, \sigma_l^2)$ and the log-likelihood is $\log L(p, \mu, \sigma|Y) = \sum_{i=1}^{n} \sum_{l=1}^{L} \{ \log(p_l) + \log \phi(y_i|\mu_1, \sigma_l) \}$, where ϕ denotes the density function of normal distribution. The code for Stan and NIMBLE to implement this model is listed in Listing 1.1 and 1.2, respectively. In Listing 1.1 we clearly find that the contribution to the syntax target is just the sum of $\log(p_l)$ and the logarithm of the density of normal distribution denoted by normal_lpdf. The rest is to assign a Dirichlet prior to the weights p_l and other parameters. However, in the NIMBLE code shown in Listing 1.2, we have to transfer the likelihood into some sampling procedures by IMAGING that there are L clusters of random numbers, the random numbers are i.i.d Gaussian within each cluster, and the probability a random number is drawn from the lth cluster is p_l. Thereafter, the Dirichlet prior is assigned to p_ls. Such imagine matches the Bayesian philosophy but when the likelihood function becomes to be quite complicated, to understand this sampling procedure may not be easy anymore, especially for practitioners not coming from a mathematics or statistics background.

Listing 1.1: *Stan code for modeling mixture of Gaussian distribution*

```
1 data {
2 int<lower=1> N;
3 vector[N] y;
4 int<lower=1> L;
5 }
6 parameters {
7 simplex[L] p;
8 vector[L] mu;
9 vector<lower=0>[L] sigma;
10 }
11 model {
12 p ~ dirichlet(rep_vector(1, L));
13 mu ~ normal(0, 100);
14 sigma ~ cauchy(0, 2.5);
15 for(i in 1:N) {
16 vector[L] lp_i;
17 for(l in 1:L) {
18 lp_i[l] = log(p[l]) + normal_lpdf(y[i]|mu[l], sigma[l]);
19 }
20 target += log_sum_exp(lp_i);
21 }
22 }
```

Listing 1.2: *NIMBLE code for modeling mixture of Gaussian distribution*

```
1 NimbleCode <- nimbleCode ({
2     for (i in 1:N) {
3         y[i] ~ dnorm(mu_y[z[i]], tau = tau_y[z[i]])
4         z[i] ~ dcat(p[1,L])
5     }
6     for (j in 1:L) {
```

```
 7        mu_y[j] ~ dnorm(0, 0.01)
 8        tau_y[j] ~ dgamma(0.01, 0.01)
 9     }
10    p[1:L] ~ ddirch(alpha0[1,L])
11  })
12  NimbleData <- list(y = y)
13  NimbleConsts <- list(L = L, N = length(NimbleData$y), alpha0 = rep(1, L))
14  NimbleInits <- list(mu_y = rnorm(NimbleConsts$L), tau_y = rgamma
             (NimbleConsts$L),p = rep(1/NimbleConsts$L, NimbleConsts$L))
```

4. Application: Bladder cancer recurrences

We apply the GSFM to analyze the bladder cancer recurrences data set contained in R package survival. Totally 118 subjects in the clinical trial are divided into three treatment strata including placebo, pyridoxine (vitamin B6), and thiotepa. Each subject may experience k (from 1 to 9) times of recurrences and may die from or not from the recurrence of bladder cancer. We do not discriminate the death from cancer and the recurrence, and the death from other causes is treated as censoring status. Our interest is the gap time between the $(k-1)$th and the kth recurrences. Besides the treatment schemes, two clinical covariates are considered: the number of tumors at the beginning (x_1) and the size of the largest tumor (x_2) within a subject. The values of these two covariates are evaluated at the beginning of each recurrence interval. This data set was once analyzed for the time between the first to the second recurrence as a univariate time-to-event outcome [39]. In this chapter, we consider both the first and the second recurrences and thus $K = 2$ here. The two covariates are scaled by divided by 100. To simplify the computation, the follow-up time is transferred from months to years to get lower scalars.

4.1 Model-checking for baseline survival functions

Before further inference, we need to check whether the proposed model is appropriate. As an alternative, a shared frailty model is fit by R package spBayeSurv. In the shared frailty model, the treatment strata are considered as indicator covariates in the parametric term. We run four independent MCMC chains for 5000 times with the first 2000 times burn-in and aggregate the rest chains together as the posterior samples under the GSFM. All chains are well mixed and convergent under the GSFM. For the shared frailty model, we run the MCMC 16,000 times with the first 6000 times burn-in through R function survregbayes using the "IID" Gaussian frailty under "PH" model name. Other settings are default.

The plots of the estimated baseline survival functions under different models stratified by treatment strata can be viewed in **Figure 3**. From that, we find the baseline survival functions estimated under the GSFM show similar trends as that of the K-M estimator in each recurrence and reflect the crossing survival curves at the first recurrence like the K-M estimator. However, the curves estimated by the shared frailty model are not crossed and cannot change along with recurrences. Therefore, the proposed GSFM is appropriate for the data.

4.2 Parametric estimation I: Real data

We use the mean of posterior samples (median for τ) as the estimator of parameter and we list the estimation of the vector of regression coefficients β and standard deviation parameter τ in **Table 2**.
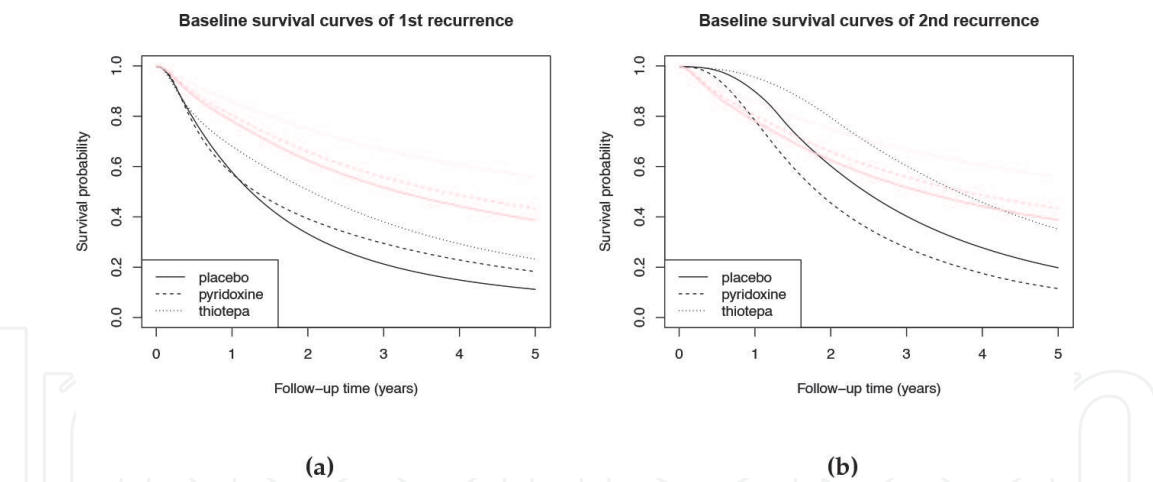
Figure 3.
The estimated baseline survival curves for the first (a) and second (b) recurrence; the black curves are estimated under the proposed generalized shared frailty model, and the pink curves are estimated under the traditional shared frailty model; the real lines, placebo; the dash lines, pyridoxine; the dotted lines, thiotepa.

	Est	SD	ESS	PACE
No. tumors	13.849	11.051	1495	0.145
Tumor size	−14.196	12.341	1114	0.194
τ	1.793	0.383	456	0.474

Est, point estimation; SD, posterior standard deviation; ESS, effective sample size; PACE, the MCMC Pace.

Table 2.
The parametric estimation and the MCMC performance for the bladder cancer recurrences data.

From **Table 2** we find that as the number of tumors at the start point increases, the hazard for recurrences increases as well whereas the larger size of the largest tumor will decrease the hazard. This conclusion is similar to that in [39] who analyzed the first recurrence as univariate time-to-event data by a transformation model.

Besides the parametric estimation result, we also report two metrics about the MCMC performance here. The first one is the effective sample size (ESS), an approximation to the number of "independent" draws in MCMC sampling. It shows that the ESS of all parameters is greater than 400, which is considered to be adequate by [40]. The ESS of τ is significantly lower than that of β, a possible reason is that the frailty random effect might be time-dependent $w_i(t)$ rather than a time-fixed effect. Another metric of interest is the average time needed to generate each effective sample, called MCMC Pace. Stan development team emphasized the importance of MCMC Pace, and the definition is given by the team of NIMBLE in [41] as the time consuming of generating one effective sample. The MCMC Pace to generate τ is much higher than that of β, and we conjecture the possible reason is that the posterior distribution has a long upper tail leading to outliers in posterior samples, which slows down the speed to generate effective samples.

4.3 Parametric estimation II: Simulation

Another simulation study is considered to evaluate the performance of parametric estimation of the MCMC procedure. Our simulation aims to simulate the occurrences of multiple events on the same individual. We take $K = 2$ and $G = 3$ denote the number of types of events and the number of treatment strata,

Parameter	BIAS	RMSE	ESD	SDE	CP
$\beta_1 = 1$	−0.062	0.042	0.222	0.196	96.7
$\beta_2 = 1$	−0.025	0.023	0.148	0.152	92.7
$\tau = 1$	−0.078	0.056	0.213	0.224	96.7

BIAS, averaged bias among the 150 simulations; RMSE, the root of mean square error of the estimation; ESD, *averaged posterior estimated standard deviation; SDE, the standard deviation of point estimate; CP, the coverage probability of 95% credible interval.*

Table 3.
Simulation results for the parametric terms.

respectively. The simulation includes two independent covariates, $x_i \sim \text{Bin}(1, 0.5)$ and $x_2 \sim N(0, 1)$ to incorporate indicator variable and continuous variable as well. For $k = 1, 2, j = 1, 2, 3$, the baseline survival functions S_{0kj} are set as:

- $S_{011} = 1 - 0.5(LN(-0.25, 1) + LN(0.25, 1))$;

- $S_{012} = 1 - 0.5(LN(-0.5, 1) + LN(0.65, 1))$;

- $S_{013} = 1 - 0.5(LN(-0.65, 1) + LN(1.25, 1))$;

- $S_{021} = 1 - LN(0, 1); S_{022} = 1 - LN(-0.5, 1); S_{023} = 1 - LN(0.5, 1)$

When $k = 1$, the three baseline survival functions are crossed whereas when $k = 2$, the three curves are not. The vector of regression coefficients is $\beta = (1, 1)^T$ and the log frailty random effect $v_i \sim N(0, 1)$ independently. The survival time is generated following model (1). The censoring variable of each event is generated from Unif(4,6) independently, leading to a censoring rate of about 28%. We set the number of subjects to be 90 and they are equally divided into three treatment strata. We repeat the simulation for 150 times.

Table 3 summarizes the results for regression parameters β and the standard deviation of frailty effect τ, including the averaged bias (BIAS), the root of mean square error (RMSE), posterior estimated standard deviation (ESD) of each point estimate (posterior mean for β and median for τ), the standard deviation (across 150 replicated simulations) of the point estimate (SDE), and the coverage probability (CP) of the 95% credible interval (given by Wald-type credible interval). The results show that the point estimates of β and τ have quite little bias with low RMSE, ESD values are close to the corresponding SDEs, and the CP values are close to the nominal 95%.

5. Discussion

In this chapter, we show the power of Bayesian computing illustrated by successfully applying the ANOVA DDP model as the nonparametric prior for a relatively complicated shared frailty model. Our survival-function version of the ANOVA DDP, modified based on the ANOVA DDP directly in Subsection 3.3, is constructed for the shared frailty model, but can reduce to modeling the univariate dependent survival functions by involving the continuous covariates into the predictor space of the ANOVA DDP. Hence, our work is an extension of [25] to some extent. However, the proposed GSFM is different from the Linear DDP model for a single group which is a generalization of the accelerated failure time model [42, 43].

Furthermore, although we point out that there exists potentially dual dependence for dual stratification of treatment strata and recurrences, we just simply allow dependence in treatment strata and assume that the recurrences are independent in our methodology demonstration. The dependence across recurrences per subject is dealt with only by the parametric frailty random effect in the proposed shared frailty model. It is more reasonable to be incorporated into the baseline survival functions so that the interaction effects between recurrence and treatment may be accounted for. Under the one-level stratification, Hanson et al. [5] modeled such serial correlation among baseline hazard functions by constructing the so-called dependent tail free process as the prior. It is non-trivial to accommodate dual temporal and stratified dependency as a future research plan.

Acknowledgements

Thanks

The first author Chong Zhong owes deep thanks to his parents. The authors thank Miss. Lulu Zhang for her efficient technical supports in text and figures. The authors thank the service manager Ms. Romina Rovan for her courtesy and professional service. The authors thank the invitation from the editor.

Author details

Chong Zhong[1†], Zhihua Ma[2†], Junshan Shen[3†] and Catherine Liu[1*†]

1 The Hong Kong Polytechnic University, Hong Kong, China

2 Shenzhen University, Guangdong, China

3 Capital University of Economics and Business, Beijing, China

*Address all correspondence to: catherine.chunling.liu@polyu.edu.hk

† These authors contributed equally.

References

[1] Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography. 1979;**16**(3): 439-454

[2] Duchateau L, Janssen P. The Frailty Model. New York: Springer Science & Business Media; 2007

[3] Balan TA, Putter H. A tutorial on frailty models. Statistical Methods in Medical Research. 2020;**29**(11): 3424-3454

[4] Ibrahim JG, Chen M-H, Sinha D. Bayesian Survival Analysis. New York: Springer Science & Business Media; 2001

[5] Hanson TE, Jara A, Zhao L, et al. A bayesian semiparametric temporally-stratified proportional hazards model with spatial frailties. Bayesian Analysis. 2012;**7**(1):147-188

[6] de Castro M, Chen M-H, Zhang Y. Bayesian path specific frailty models for multi-state survival data with applications. Biometrics. 2015;**71**(3): 760-771

[7] Therneau TM. A Package for Survival Analysis in R. R Package Version 3.2-11. 2021

[8] Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. Journal of the Royal Statistical Society, Series B: Statistical Methodology. 2007; **69**(4):507-564

[9] De Iorio M, Müller P, Rosner GL, MacEachern SN. An anova model for dependent random measures. Journal of the American Statistical Association. 2004;**99**(465):205-215

[10] de Castro M, Chen M-H, Ibrahim JG, Klein JP. Bayesian transformation models for multivariate survival data. Scandinavian Journal of Statistics. 2014; **41**(1):187-199

[11] Paulon G, De Iorio M, Guglielmi A, Ieva F. Joint modeling of recurrent events and survival: A bayesian non-parametric approach. Biostatistics. 2020;**21**(1):1-14

[12] Conlon ASC, Taylor JMG, Sargent DJ. Multi-state models for colon cancer recurrence and death with a cured fraction. Statistics in Medicine. 2014; **33**(10):1750-1766

[13] Stan Development Team. The Stan Core Library. Version 2.27. 2018

[14] Stan Development Team. RStan: The R Interface to Stan. R Package Version 2.21.2. 2020

[15] Ferguson TS. Prior distributions on spaces of probability measures. The Annals of Statistics. 1974;**2**(4):615-629

[16] Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical dirichlet processes. Journal of the American Statistical Association. 2006;**101**(476):1566-1581

[17] Rodriguez A, Dunson DB, Gelfand AE. The nested dirichlet process. Journal of the American Statistical Association. 2008;**103**(483):1131-1154

[18] MacEachern SN. Dependent nonparametric processes. In: ASA Proceedings of the Section on Bayesian Statistical Science. Alexandria, VA: American Statistical Association; 1999

[19] MacEachern SN. Dependent Dirichlet Processes. Technical Report. Department of Statistics, The Ohio State University; 2000

[20] MacEachern SN. Nonparametric bayesian methods: A gentle introduction

and overview. Communications for Statistical Applications and Methods. 2016;**23**(6):445-466

[21] Quintana FA, Mueller P, Jara A, MacEachern SN. The dependent dirichlet process and related models. arXiv preprint arXiv:2007.06129. 2020

[22] Sethuraman J. A constructive definition of Dirichlet priors. Statistica Sinica. JSTOR. 1994;**4**:639-650

[23] Phadia EG. Prior Processes and their Applications. New York: Springer; 2015

[24] Gelfand AE, Kottas A, MacEachern SN. Bayesian nonparametric spatial modeling with dirichlet process mixing. Journal of the American Statistical Association. 2005;**100**(471):1021-1035

[25] De Iorio M, Johnson WO, Müller P, Rosner GL. Bayesian nonparametric nonproportional hazards survival modeling. Biometrics. 2009;**65**(3): 762-771

[26] Nieto-Barajas LE, Müller P, Ji Y, Lu Y, Mills GB. A time-series ddp for functional proteomics profiles. Biometrics. 2012;**68**(3):859-868

[27] DeYoreo M, Kottas A. Modeling for dynamic ordinal regression relationships: An application to estimating maturity of rockfish in california. Journal of the American Statistical Association. 2018;**113**(521): 68-80

[28] Lo AY. On a class of bayesian nonparametric estimates: I. Density estimates. The Annals of Statistics. 1984; **12**:351-357

[29] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. Florida: CRC Press; 2013

[30] Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). Bayesian Analysis. 2006; **1**(3):515-534

[31] Gelman A, Jakulin A, Pittau MG, Su Y-S. A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics. 2008;**2**(4):1360-1383

[32] Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using bayesian semi-parametric models: Applications to institutional comparisons. Statistics in Medicine. 2007;**26**(9):2088-2112

[33] Hoffman MD, Gelman A, et al. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research. 2014;**15**(1):1593-1623

[34] Kerioui M, Mercier F, Bertrand J, Tardivon C, Bruno R, Guedj J, et al. Bayesian inference using hamiltonian monte-carlo algorithm for nonlinear joint modeling in the context of cancer immunotherapy. Statistics in Medicine. 2020;**39**(30):4853-4868

[35] Ma Z, Hu G, Chen M-H. Bayesian hierarchical spatial regression models for spatial data in the presence of missing covariates with applications. Applied Stochastic Models in Business and Industry. 2021;**37**(2):342-359

[36] Stan forums. Available from: https://discourse.mc-stan.org

[37] Nimble groups. Available from: https://r-nimble.org/more/issues-and-groups

[38] Zhang L. A bayesian comparison in stan and nimble by trimmed mean regression [M.Phil's thesis]. Hong Kong, China: The Hong Kong Polytechnic University; 2021

[39] Zeng D, Lin DY. Efficient estimation of semiparametric

transformation models for counting processes. Biometrika. 2006;**93**(3): 627-640

[40] Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-normalization, folding, and localization: An improved r for assessing convergence of MCMC. Bayesian Analysis. 2021;**1**(1):1-28

[41] de Valpine P, Paciorek C, Turek D, Michaud N, Anderson-Bergman C, Obermeyer F, Cortes CW, RodrÃguez A, Lang DT, Paganin S. NIMBLE User Manual. R Package Manual Version 0.11.1. 2021

[42] Hanson TE, Jara A. Surviving fully bayesian nonparametric regression models. Bayesian Theory and Applications. 2013:593-615

[43] Riva-Palacio A, Leisen F, Griffin J. Survival regression models with dependent bayesian nonparametric priors. Journal of the American Statistical Association. 2021:1-10