ARC Centre of Excellence in Population Ageing Research

Working Paper 2023/07

A Dirichlet Process Mixture regression model for the analysis of competing risk events

Francesco Ungolo and Edwin R. van den Heuvel

University of New South Wales

A Dirichlet Process Mixture regression model for the analysis of competing risk events

Francesco Ungolo[1,2,3,*] and Edwin R. van den Heuvel[4]

[1] *School of Risk and Actuarial Studies, University of New South Wales, Kensington NSW 2052, Australia*
[2] *ARC Centre of Excellence in Population Ageing Research, University of New South Wales, Kensington NSW 2052, Australia*
[3] *Chair of Mathematical Finance, Technical University of Munich, 85748 Garching bei München, Germany*
[4] *Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands*
[*] *Corresponding author: Francesco Ungolo, f.ungolo@unsw.edu.au*

April 18, 2023

We develop a regression model for the analysis of competing risk events. The joint distribution of the time to these events is characterized by a random effect following a Dirichlet Process, explaining their variability. This entails an additional layer of flexibility of this joint model, whose inference is robust with respect to the misspecification of the distribution of the random effects. The model is analysed in a fully Bayesian setting, yielding a flexible Dirichlet Process Mixture model for the joint distribution of the time to events. An efficient MCMC sampler is developed for inference. The modelling approach is applied to the empirical analysis of the surrending risk in a US life insurance portfolio previously analysed by Milhaud & Dutang (2018). The approach yields an improved predictive performance of the surrending rates.

Keywords: Competing Risks, Survival Analysis, Dirichlet Processes, Bayesian analysis, Lapse risk, MCMC

1. Introduction

The analysis of competing risk events is commonplace in the statistical and the actuarial fields. It concerns the probabilistic mechanism of the time to the events an individual is simultaneously exposed. The most common example is the case of an individual who is exposed to several causes of death, such as cancer, cardiovascular disease, and so on.

The interest lies in the joint probability distribution of the time to events for each of the M causes of decrement, denoted as T_1, T_2, ..., T_M, whilst the researcher can only observe $T = \min(T_1, \dots, T_M)$, that is the time to the first occurring event The other events are known to occur after T (at least theoretically), hence they can be considered as right censored. However, this joint distribution is not identifiable given the data (see Tsiatis (1975), Crowder (1996) and Crowder (1997) for a more detailed account of this issue).

Therefore, further point identifying assumptions are needed, that is assumptions which cannot be tested in practice. One common example is to assume that $(T_1, \dots, T_M)$ is a vector of pairwise independently distributed lifetimes, which allow for the factorization of their marginal distribution.

Models which focus on the time to specific events gathered a lot of attention in the literature, especially under the assumption of independence. For example, Fine & Gray (1999) focus on a Cox proportional hazard model (Cox 1972) for the cumulative incidence function, while Lunn & McNeil (1995) use the same model on a properly augmented data set to account for competing risks. The work of Dimitrova et al. (2013) includes several references about the inferential aspects of the problem within the medical, biostatistical, demographic and actuarial literature.

Several approaches have been proposed to account for dependence in the times to competing risk events. One of these, consists of adjusting the distribution of a event of interest to account for the potential occurrence of other events, such as the paper of Jackson et al. (2014). They specify a Cox proportional hazard model including a step-change component when the event of interest is subject to informative censoring (the sole competing event), and assess the sensitivity of the resulting inference with respect to the independence assumption.

Scharfstein & Robins (2002) and Rotnitzky et al. (2007), similarly focus on the informative censoring problem, which can be seen as a special case of competing risk events, where $M = 2$. These papers enrich the distribution of the time to event of interest with a hazard function for the censoring time, which depends on the time to event. Other examples of joint models for the distribution of the times to competing events is the bivariate Weibull model of Emoto & Matthews (1990), the bivariate Makeham model of

Arnold & Brockett (1983), and models based on known copula functions, such as Zheng & Klein (1995) and Escarela & Carrière (2003).

The approach proposed in this paper lies within the class of models which models the joint distribution of the times to competing risk events as conditionally independent given a random component which explains their dependence. Among these, we mention the work of Yashin & Iachine (1995) who developed a correlated gamma frailty model, and the work of Huang & Wolfe (2002) who considers the inclusion of covariates within a semi-parametric Cox proportional hazard model alongside a normally distributed log-frailty component. A paper similar to Huang & Wolfe (2002) is the work of Gorfine & Hsu (2011), which considers other parametric functions for the distribution of the individual frailty. We believe that these approaches, although promising due to their flexibility, favoured by the conditional independence assumption, are restrictive since the resulting inference on the model parameters (and thus the survival function) can be affected by the misspecification of the distribution of the random component. For this reason, Ungolo & van den Heuvel (2022) exploited the conditional independence assumption in order to use a bivariate random effect with a discrete distribution with unknown number of levels, thus weakening the distributional assumption over the frailty. As further explained in Section 3, their approach scales poorly to the analysis of models with a large number of parameters, which may be needed for the analysis of large data sets, as can be available within the actuarial field, which can include several thousands of observations.

In this paper we develop a Dirichlet Process Mixture model where the time to competing risk events are independently distributed conditional to a multivariate random effect which we denote as θ. In order to reduce the impact of the misspecification of the distribution of θ over the probabilistic mechanism of $T_1, \ldots, T_M$, we assume that θ is a random draw from a Dirichlet Process. The resulting mixture model allows for various shapes of the density, capturing several features of the data, such as multimodality and overdispersion. Similar to the work of Ungolo & van den Heuvel (2022), the inferential approach will be fully Bayesian, as we use the data to learn the number of levels of the discrete random effect. In this way, we exploit the knowledge of the researcher in the form of prior distribution when making inference on the parameters. The model can easily account for individual covariates, and its inference can easily deal with the case of censored observations. We apply the approach of this work to the empirical analysis of the lapse risk in a portfolio of life insurance policies, with specific emphasis on the case of surrending, and the prediction of the surrending rates. Indeed, an insurance policy can be terminated due to surrending, death, maturity or default on paying the periodic premium. The simultaneous exposure to different causes of termination of an insurance policy suggests an analysis within the competing-risk framework (Milhaud & Dutang 2018).

This paper is structured as follows: in Section 2 we introduce the Dirichlet Process and the Dirichlet Process Mixture, Section 3 describes the joint model for the time to competing risk events, Section 4 describes the empirical data set and the model used for the analysis, Section 5 describes the inferential procedure. Section 6 reports the results of the empirical analysis and finally Section 7 concludes and outlines extensions of the model for future research.

2. A primer on Dirichlet Process Mixture models

A Dirichlet Process (DP, Ferguson (1973)) is a probability distribution over random probability distributions, denoted by P. The DP is characterized by a *base probability measure* P_0, providing an initial guess for P, analogously to the prior specification for a parameter in a Bayesian analysis, and a concentration parameter ϕ, which measures the strength of the prior distribution with respect to P_0. If P has a DP prior, then it is denoted as $P \sim \mathrm{DP}(\phi, P_0)$.

If the random variable $\theta \sim P(\cdot)$, then it follows that (Sethuraman (1994)):

$$P(\cdot) = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}(\cdot), \tag{2.1}$$

where $\theta_1, \theta_2, \ldots$ are i.i.d. samples from P_0 and δ_{θ_k} is the Dirac measure assigning unitary mass if $\theta = \theta_k$ and zero otherwise. The mixture weights $\pi_1, \pi_2, \ldots$ are calculated by means of the so called *stick breaking procedure* (SBP):

$$\pi_k(\psi_{1:k}) = \psi_k \prod_{j=1}^{k-1} (1 - \psi_j), \tag{2.2}$$

where $\psi_{1:k} = (\psi_1, \ldots, \psi_k)$ and $\psi_k \sim \mathrm{Beta}(1, \phi)$. As k becomes larger, then π_k is extremely small. For simplicity, in the remainder of the paper we write $\pi_k(\psi_{1:k}) = \pi_k$.

Therefore, $P(\theta)$ in equation (2.1) is a mixture distribution with an infinite number of components, given by the Dirac functions in this case. Random draws from a Dirichlet Process consist of discrete distributions over countably infinite atoms from the base measure P_0.

A DP is useful to characterize the prior distribution for the density of a random variable Y, generically denoted by f, as a mixture with an unknown number of components, yielding a Dirichlet Process Mixture model with respect to the indexing parameter θ (DPM, see Lo (1984) and Escobar & West (1995)):

$$f(y; P) = \int_{\Omega_\theta} f(y; \theta)\, \mathrm{d}P(\theta) = \sum_{k=1}^{\infty} \pi_k f(y; \theta_k). \tag{2.3}$$

with Ω_θ denoting the sample space of θ In this way, the density of Y can assume various shapes which can account for example for multimodality and heavy tails.

From an additional perspective, the DP induces a dependence structure among the θ parameters. Let $\theta_{1:n} = (\theta_1, \ldots, \theta_n)$ denote a sample of n draws from P. Their joint distribution, marginalized with respect to the random probability measure P, $p(\theta_{1:n}; \phi, P_0)$ allows to derive the probability distribution of θ_n given $\theta_{1:n-1} = (\theta_1, \ldots, \theta_{n-1})$ as a Polya urn distribution (Blackwell & MacQueen (1973)):

$$p(\theta_n \mid \theta_{1:n-1}; \phi, P_0) \propto \frac{1}{\phi + n - 1} \sum_{i=1}^{n-1} \delta_{\theta_i}(\theta_n) + \frac{\phi}{\phi + n - 1} P_0(\theta_n). \tag{2.4}$$

If we write $\theta^*_{1:n-1} = (\theta^*_1, \ldots, \theta^*_J)$, $J < n$, as the set of unique values of $\theta_{1:n-1}$, then equation (2.4) can be rewritten as follows:

$$p(\theta_n \mid \theta_{1:n-1}; \phi, P_0) \propto \frac{1}{\phi + n - 1} \sum_{j=1}^{J} n_j \delta_{\theta^*_j}(\theta_n) + \frac{\phi}{\phi + n - 1} P_0(\theta_n), \tag{2.5}$$

where n_j are such that $\sum_j n_j = n - 1$, and denote the number of observations which are equal to θ^*_j.

Equations (2.4)-(2.5) show how the DP induces a sequential clustering process for θ. As we sample more observations, the subsequent ones are likely either to take one of the values already observed with probability which depends on their current frequency, or a new value with probability which increases with ϕ. In other words, as the number of observations grows these are more likely to be in a class which has been already observed (the so called *richer-by-richer effect*). This sampling process is also known in the literature as the *Chinese restaurant process* (see Aldous (1985), Heinz (2014) and Orbanz (2014)). Despite the nonparametric nature of this approach, the parameter ϕ allows for regularization by favouring the shrinkage of the distribution of θ towards the base distribution, which can have a simple parametric form.

Under a DPM, a sample $y_1, \ldots, y_n$ is generated according to the following hierarchy:

$$\begin{aligned} &P \mid \phi, P_0 \sim \mathrm{DP}(\phi, P_0) \\ &\theta_i \mid P \sim P, \quad i = 1, \ldots, n; \\ &y_i \mid \theta_i \sim f(y_i; \theta_i) \quad i = 1, \ldots, n \end{aligned} \tag{2.6}$$

or equivalently using θ^*

$$\begin{aligned} &\pi_k \mid \phi \sim \mathrm{SBP}(\phi), \quad k = 1, 2, \ldots; \\ &\theta^*_k \mid P_0 \sim P_0, \quad k = 1, 2, \ldots; \\ &s_i \mid \pi_1, \pi_2, \ldots \sim \mathrm{Mult}(\pi_1, \pi_2, \ldots), \quad i = 1, \ldots, n; \\ &y_i \mid \theta^*_{s_i} \sim f\left(y_i; \theta^*_{s_i}\right) \quad i = 1, \ldots, n, \end{aligned} \tag{2.7}$$

where the Multinomial distribution is extended to account for an infinite number of classes. The Dirichlet Process mixture of equation (2.3) is then obtained following the latency of the mixture allocation variables s_i.

3. Dirichlet Process Mixture Regression model for competing risk events

Let $T_1, \ldots, T_M$ denote the vector of random times to M competing events. Their joint distribution is non-identifiable given the data, since for each individual we can only observe $T = \min(T_1, \ldots, T_M)$, that is we can observe only which cause of decrement C occurred, and thus also $T_C = T$. Therefore, the time to the other events are considered to be right censored at time T.

We propose to point-identify this joint distribution by assuming that the time to each event are pairwise conditionally independent given a latent vector θ for the ith unit:

$$\begin{aligned} f\left(t_{1,i},\ldots,t_{M,i}\right) &= \int_{\Omega_{\theta_i}} \prod_{c=1}^{M} f_c\left(t_{c,i};\theta_i\right) \mathrm{d}P\left(\theta_i\right) \\ &= \int_{\Omega_{\theta_i}} \left[\prod_{c=1}^{M} f_c\left(t_{c,i};\theta_{c,i}\right)\right] \mathrm{d}P\left(\theta_{1,i},\ldots,\theta_{M,i}\right) \\ &= \sum_{k=1}^{\infty} \pi_k \left[\prod_{c=1}^{M} f_c\left(t_{c,i};\theta^*_{c,k}\right)\right], \end{aligned} \tag{3.1}$$

hence, $\theta \sim P$ and P is a draw from a Dirichlet Process, yielding the Dirichlet Process Mixture in the last equality of equation (3.1).

The latent multivariate random vector $\theta = (\theta_1,\ldots,\theta_M)$ has the role of inducing a dependence relationship among $T_1,\ldots,T_M$. Its aim is to capture those latent features which are hidden, but affect the joint occurrence of the M risks. More precisely, these unit-specific parameters allow for individual-level heterogeneity as not explained by other variables in the model. The interpretation of θ can be analogous to the frailty in survival analysis (see Wienke (2014) for a review). Indeed, the conditional independence assumption in equation (3.1) is not new: for example, Vaupel & Yashin (1985) assume that each cause has its own independently distributed gamma random component, thus deriving closed form formulas for the cause-specific hazard functions.

Hence, at high values of θ for one cause, there can be an associated high value of the corresponding parameter for another cause, and so on. Furthermore, our approach is more general, because it allows for pairwise independence among $T_1,\ldots,T_M$ if $p(\theta) = p(\theta_1)\cdots p(\theta_M)$, as well as for independence among grouped competing causes of decrement in analogous way. For example, when analysing competing causes of death, there can be unobserved genetic factors or habits, such as smoking, which can increase the likelihood of dying by certain cancers and cardiovascular diseases, making these two causes positively associated.

In actuarial science this type of model can be useful for the analysis of the lifetimes of couples, as these can be positively associated (Frees et al. 1996), since two (or more) individuals can be exposed to similar risks. Or when analysing the lapse risk, an old person in need of money to pay for medical expenses, may be more likely to surrender her policy (Milhaud & Dutang 2018). In this case, time to death and time to surrending may be positively associated, as it will result also in the empirical analysis (Section 6).

Alternative formulations for $P(\theta)$ can be specified: for example, the paper of Huang & Wolfe (2002) assumes a normally distributed univariate parameter θ when analysing survival data subject to informative censoring. However, the limitation of this approach was to force either positive dependence or negative dependence. Our approach is more general for a two-fold reason: first we do not constrain a positive or a negative dependence for all units in the sample, and secondly, we avoid the misspecification of the random effect distribution ensuring the robustness of the estimated joint model.

The identifiability of the joint model of equation (3.1) follows as a straightforward extension of the bivariate version in Ungolo & van den Heuvel (2022).

The work of Ungolo & van den Heuvel (2022), which considers only two dependent competing risk events (time to event and informative censoring) assumes a bivariate discrete distribution for θ with a number of levels chosen by means of appropriate model selection criteria. Their work overcomes the limitation of the frailty model of Huang & Wolfe (2002), but requires the estimation of several models, which may not be feasible in practice for large data sets with a higher number of covariates and parameters.

Fundamental quantities of interest for actuaries can be readily obtained. For example, the *joint survivor function* is given by:

$$S_{1,\dots,M}\left(t_1,\dots,t_M\right)=\Pr\left(T_1>t_1,\dots,T_M>T_m\right)=\sum_{k=1}^{\infty}\pi_k\left[\prod_{c=1}^{M}\left(1-F_c\left(t_{c,i};\theta^*_{c,k}\right)\right)\right] \tag{3.2}$$

where $F_c\left(t_c;\theta_c\right)=\Pr\left(T_c<t_c\mid\theta_c\right)=\int_0^{t_c}f_c\left(s\mid\theta_c\right)\mathrm{d}s$. The calculation of the *overall survivor function* $S_{1,\dots,M}\left(t,\dots,t\right)$ and of the *marginal survivor function* $S_c\left(t_c\right)=S_{1,\dots,M}\left(0,\dots,0,t_c,0,\dots,0\right)$ follows along the same lines. From $S_c\left(t_c\right)$ it can be possible to obtain the *net hazard function*, equal to $f_c\left(t_c\right)/S_c\left(t_c\right)$. Finally, it is possible to obtain also the *crude survivor function*, defined as:

$$\begin{aligned}S'_c\left(t\right)&=\Pr\left(\min\left(T_1,\dots,T_M\right)>t,\min\left(T_1,\dots,T_M\right)=T_c\right) \qquad (3.3)\\&=\sum_{k=1}^{\infty}\pi_k\left[\int_t^{\infty}f_c\left(s\mid\theta^*_{c,k}\right)\prod_{j\neq c}\left(1-F_j\left(s;\theta^*_{j,k}\right)\right)\mathrm{d}s\right]\end{aligned}$$

and the corresponding *crude hazard function.*

This approach allows also for the inclusion of unit cause-specific covariates, denoted by the p_c-dimensional vector $x_{c,i}=\left(x_{c,1,i},\dots,x_{c,p_c,i}\right)$ with regression coefficient β_c. For simplicity, we assume that θ is independently distributed with respect to the covariates, although the approach allows for this possibility. Therefore, the joint density of $\left(T_1,\dots,T_M\right)$ is obtained:

$$f\left(t_{1,i},\dots,t_{M,i}\mid x_{1,i},\dots,x_{M,i};\beta_1,\dots,\beta_M\right)=\sum_{k=1}^{\infty}\pi_k\left[\prod_{c=1}^{M}f_c\left(t_{c,i}\mid x_{c,i};\beta_c,\theta^*_{c,k}\right)\right] \tag{3.4}$$

This approach is general, as it can be applied with any parametric specification for the distribution of T_c. Furthermore, the random time to event for each cause can be characterized by different parametric laws, for example we can have $T_{1,i}\sim\operatorname{Exp}\left(e^{\beta_1 x_{1,i}+\theta_{1,i}}\right)$, and $Y_{2,i}=\log T_{2,i}\sim N\left(\beta_1 x_{2,i}+\theta_{2,i},\sigma^2\right)$, and easily characterize a joint model for $\left(T_1,Y_2\right)$. Hence, for T_1, we can assume a proportional effect of θ, while for Y_2 this can have a linear effect.

Furthermore, the scope of the joint model of this work goes beyond the analysis of competing risks: Frees & Valdez (1998) is a fundamental reference for insurance applications where models for the dependence of multiple random variables are needed.

4. Data and model

We illustrate the modelling approach described in Section 3 through the analysis of the policyholders' surrending behavior. Milhaud & Dutang (2018) emphasize the need of insurance companies to obtain good forecasts of the time to surrending, since this is relevant to recover the initial expenses incurred when issuing the contract. This issue is also relevant to annuity providers, as the occurrence of surrenders lower than predicted is likely to cause an unexpected increase in the value of the liabilities. The policyholders' behavior should be carefully monitored and modelled, since it can affect the pricing of options and guarantees, the solvency capital requirements and the effectiveness of the hedging strategies (Knoller et al. 2016). Escobar et al. (2016) emphasize how an appropriate prediction of the surrending times can be extremely crucial in the light that policyholder behavioral risk cannot be hedged, and it can entail serious liquidity issues.

The data set consists of a portfolio of Whole Life insurance policyholders sold from an insurer operating in the US. This data set, available through the R package `CASdata sets` (Dutang & Charpentier (2020)), contains several information about 29,317 records of policies sold between January 1995 and December 2008. Milhaud & Dutang (2018) extensively analysed this data set by means of a competing-risk approach using a Cox propotional-hazard based approach relying on independence among the causes, as well as using the subdistibution approach of Fine & Gray (1999).

We randomly split the data set into a training set, corresponding to the 75% of the policyholders within the data set (21,988 units), and use the remaining 25% to test the predictive ability of the model (7,329 units).

The three causes of decrement are *surrending*, *death* and *other*. Within the training data set we observe these three events for 8,347 (38%), 968 (4.4%) and 1,846 (8.4%) units respectively. The remaining units are Type I censored, since they reached the end of the observational period without any of the events being occurred. As pointed out by Milhaud & Dutang (2018) we should consider the statistical association between causes of decrement, as mentioned in Section 3.

We consider the three causes simultaneously, and then focus merely on surrending. The set of available covariates used for the three competing causes are as follows:

$$\begin{aligned} x_{\text{surr}} &= (\text{AP}, \text{DJ}, \text{ADR}, \text{G}, \text{PF}, \text{UWA}) \\ x_{\text{death}} &= (\text{G}, \text{UWA}, \text{LP}, \text{RS}) \\ x_{\text{other}} &= (\text{AP}, \text{PF}, \text{ADR})\,, \end{aligned}$$

where:

- AP is the standardized annual premium;
- DJ is the last observed quarterly variation of the Dow Jones index (standardized);
- ADR is a binary covariate equal to 1 if the policy has an accidental death rider and 0 otherwise;

- G indicates the gender (1 if female, 0 if male);
- PF is the payment frequency (0 if infrannual and 1 otherwise);
- UWA is a binary covariate equal to 1 if the policyholder is aged 54 or younger when underwriting the policy, and 0 if older than 54;
- LP denotes the living place in US (LP=0 if the policyholder lives either in the East or West coast, and LP=1 otherwise);
- RS denotes the smoking status (1 if smoker and 0 otherwise).

Compared to the work of Milhaud & Dutang (2018), we include a lower number of covariates when analysing each cause. This is because it can be argued that smoking does not (at least directly) affect the policyholder surrending behavior. On the other hand, the approach does not exclude its statistical association with surrending. This is because we may think of smoking as a mediator effect, which affects death, and in its turn may affect surrending, throughout the statistical association between these two competing causes. A generalization of the method to perform variable selection is discussed in Section 7

To simply illustrate the method, we assume the following model for $Y_c = \ln T_c$ ($c \in \{\text{surr}, \text{death}, \text{other}\}$):

$$Y_c = \ln T_c = \beta_c x_{c,i} + \theta_{c,i} + \epsilon_{c,i} \quad \epsilon_{c,i} \sim N\left(0, \sigma_c^2\right), \tag{4.1}$$

where $\theta_i \sim P$, as from the generating process of equation (2.7), representing the individual random intercept. Furthermore, we assume that P_0 corresponds to the multivariate normal distribution with mean m_θ and covariance Σ_θ, denoted as $\text{MVN}\left(\theta_i \mid m_\theta, \Sigma_\theta\right)$. Hence, $\theta_i = (\theta_{1,i}, \ldots, \theta_{M,i}) \sim P\left(\phi, \text{MVN}\left(\theta_i \mid m_\theta, \Sigma_\theta\right)\right)$. This assumption allows for the specification of conditionally conjugate priors for all model parameters, which speeds up computations.

The model of equation (4.1) corresponds to the Accelerated Failure time model with normal error terms (see for example Collett (2003) or Sha et al. (2006)), which we extend by including a multivariate random component to induce dependence between times to events.

The model for the time to death is far from perfect, because for example, we do not have any information about the age of the policyholder, and the data set is subject to a very large right-censoring for this risk. Indeed, we do not aim at proposing a specific model or developing an alternative theory for the surrending-death-other risk. We remark how this paper has the sole purpose to propose a joint modelling framework for dependent random variables, where the dependence is explained by flexibly distributed random effects, which point identify the joint distribution of mutually censored lifetimes. Nevertheless, this specific model can be useful just to have an indication of the statistical association between the time to event and the covariates, or to have an idea of the statistical association among the times to competing events.

As highlighted in Section 3, other parametric specifications are allowed under the framework hereby proposed, such as the Cox proportional hazard model, or parametric models for the hazard function (see for example Richards (2008) and Richards et al. (2013)). However, in many cases, for some parameters, the analytical formulation of the conditional updates are not possible. Therefore, the inferential task may turn out computationally intensive. A similar model was analyzed by Ungolo et al. (2020) who focus on a Gompertz hazard function with a proportional random component in the hazard function. Furthermore, the use of a log-Normal distribution allows for a clearer analysis of the statistical association among Y by means of the (linear) correlation matrix of the resulting θ.

5. Inference

First of all, we approximate the DPM model of equation (3.4) by setting an upper bound K to the number of mixture components. In this way, we obtain the truncated SBP, see Ishwaran & James (2001). Dunson (2010) suggests that an upper bound of $K = 25$ should be sufficient to represent a DPM. Other alternatives, which avoid setting K upfront are the slice samplers of Walker (2007) and Kalli et al. (2011), or the retrospective sampler of Papaspiliopoulos & Roberts (2008).

Within the fully Bayesian approach of this work, we also let the data to help learning about ϕ, m_θ and Σ_θ, by assuming these are random variables with their own prior distributions.

5.1. Prior distributions

A Bayesian analysis which facilitates the computation of the posterior distribution is possible for the model described in Section 4 by first assuming that all parameters are *a priori* independently distributed, and then specifying (conditionally) conjugate prior distributions. In order to reduce the extent of the researcher choice on the posterior distribution, we specify non-informative and pairwise independent prior distributions for the model parameters β_c and σ_c^2. Therefore, we assume that $\sigma_c^2 \sim \text{Inv-Gamma}\,(1,1)$ and $\beta_{c,p} \sim N\,(0,9)$ for $c = 1,\ldots,M$ and $p = 1,\ldots,p_c$.

For the parameters of the Dirichlet Process, we assume $\phi \sim \text{Gamma}\,(1,1)$, which yields conditional conjugacy (Escobar & West (1995)) of its posterior distribution. The choice of hyperparameters is motivated by the results of a simulation study of Ungolo & van den Heuvel (n.d.), which is also a common choice in applied statistics (Dunson (2010)). Then, we assume that $m_\theta \sim \text{MVN}\,(\mathbf{0}, 9I_M)$, $\Sigma_\theta \sim \text{Inv-Wishart}\,(8, 0.001\lambda_1)$, where I_M denotes the M-dimensional identity matrix and

$$\lambda_1 = \begin{bmatrix} 1 & 0.5 & 0.2 \\ 0.5 & 1 & 0.3 \\ 0.2 & 0.3 & 1 \end{bmatrix} \tag{5.1}$$

In this way, we centered the distribution of the random effect around 0, and specified a weakly informative distribution for Σ_θ in order to embed the available information about the dependence among causes of decrement. For example, consistently with actuarial intuition, we assume a positive correlation between the random effects of the surrending and the death decrements.

5.2. Likelihood

Let $d_{c,i} = 1_{[c_i=c]}$ denote the indicator variable which is equal to 1 if the cth cause occurs for the ith individual and 0 otherwise. Furthermore, we assume that the time to events for each individual are independently distributed, conditional on the covariates and the multivariate random effect. Here, $c = 1$ corresponds to the surrending cause, $c = 2$ to death and $c = 3$ to other.

We simplify the notation and write $\mathbf{y} = (y_1, \dots, y_n)$, where $y_i = \log t_i$, $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{c} = (c_1, \dots, c_n)$, $\beta = (\beta_1, \dots, \beta_M)$, $\sigma^2 = \left(\sigma_1^2, \dots, \sigma_M^2\right)$, $\psi = (\psi_1, \dots, \psi_{K-1})$ and $\theta^* = \left(\theta^*_{1,1}, \dots, \theta^*_{M,K}\right)$

The likelihood function of the parameters conditional on observable $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{c}$ for each individual is given by:

$$
\begin{aligned}
&L\left(\beta, \sigma^2, \psi, \theta^* \mid \mathbf{y}, \mathbf{x}, \mathbf{c}\right) \\
&= \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} \pi_k \left[\prod_{c=1}^{M} f_c\left(y_i \mid x_{c,i}; \beta_c, \theta^*_{c,k}, \sigma_c^2\right)^{d_{c,i}} \left(1 - F_c\left(y_i \mid x_{c,i}; \beta_c, \theta^*_{c,k}, \sigma_c^2\right)\right)^{1-d_{c,i}} \right] \right\}
\end{aligned} \tag{5.2}
$$

where we account for right-censored observations, and marginalize the resulting distribution with respect to the latent variable θ^*.

For convenience, we rewrite this likelihood function by using the indicator variable $s_{i,k}$, which takes value 1 if the ith individual belongs to the kth class and zero otherwise. This formulation is useful when deriving the conditional MCMC updates (Section 5.3):

$$
\begin{aligned}
&L\left(\beta, \theta^*, \sigma^2, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{c}, \mathbf{s}\right) \\
&\propto \prod_{i=1}^{n} \pi_k^{s_{i,k}} \left[\prod_{c=1}^{M} f_c\left(y_i \mid x_{c,i}; \beta_c, \theta^*_{c,k}, \sigma_c^2\right)^{d_{c,i}} \left(1 - F_c\left(y_i \mid x_{c,i}; \beta_c, \theta^*_{c,k}, \sigma_c^2\right)\right)^{1-d_{c,i}} \right]^{s_{i,k}}
\end{aligned} \tag{5.3}
$$

where $\mathbf{s} = (s_{1,1}, \dots, s_{n,K})$.

5.3. Posterior distribution and MCMC updates

The posterior distribution is thus obtained as the product of the likelihood and the prior distribution:

$$\begin{aligned}
&p\left(\beta, \theta^*, \sigma^2, m_\theta, \Sigma_\theta, \phi, \psi \middle| \mathbf{y}, \mathbf{x}, \mathbf{c}, \mathbf{s}\right) \qquad (5.4)\\
&\propto L\left(\beta, \theta^*, \sigma^2, \psi \mid \mathbf{y}, \mathbf{x}, \mathbf{c}, \mathbf{s}\right) \left[\prod_{k=1}^{K} \mathrm{MVN}\left(\theta^*_{\cdot,k} \mid m_\theta, \Sigma_\theta\right)\right] \left[\prod_{c=1}^{M} \prod_{p=1}^{p_c} \mathrm{N}\left(\beta_{c,p} \mid 0, 9\right)\right]\\
&\times \left[\prod_{c=1}^{M} \text{Inv-Gamma}\left(\sigma_c^2 \mid 1, 1\right)\right] \mathrm{MVN}\left(m_\theta \mid \mathbf{0}, 9 I_M\right) \text{Inv-Wishart}\left(\Sigma_\theta \mid 8, \lambda_1\right)\\
&\times \left[\prod_{k=1}^{K-1} \mathrm{Beta}\left(\psi_k \mid 1, \phi\right)\right] \mathrm{Gamma}\left(\phi \mid 1, 1\right)
\end{aligned}$$

In order to efficiently learn the posterior distribution of equation (5.4), we combine the steps of a blocked Gibbs sampler (Ishwaran & James (2001)), consisting of a sequential draws of the parameters, together with two Data Augmentation steps (Tanner & Wong (1987)), where we first draw a value for the censored log-time for the events which were not observed, and then we draw the missing indicator $s_{i,k}$ $(i = 1, \ldots, n,\ k = 1, \ldots, K)$.

Therefore, by using some simple algebra, the following updates can be obtained for the two-Data Augmentation-Blocked Gibbs sampler (the superscript (ℓ) denotes the iteration).

Step 0: Set an initial value for the parameters $\left(\beta^{(0)}, \theta^{*(0)}, \sigma^{2(0)}, m_\theta^{(0)}, \Sigma_\theta^{(0)}, \phi^{(0)}, \psi^{(0)}\right)$;

At the ℓth iteration:

Step 1: For each individual, sample the right-censored log-random time to event $y^*_{c,i}$ for the unobserved causes (that is, for those causes c such that $d_{c,i} = 0$) from a truncated normal distribution with lower truncation level equal to the observed y_i[1]. In this way, we obtain the completed vector of time to event for each individual $y_i^* = \left(y^*_{1,i}, \ldots, y^*_{M,i}\right)$, where:

$$\begin{cases} y^{*\,(\ell)}_{c,i} = y_i = \log t_i & \text{if } d_{c,i} = 1 \\ y^{*\,(\ell)}_{c,i} \sim \text{Trunc-N}\left(y^{*\,(\ell)}_{c,i} \mid \beta_c^{(\ell-1)} x_{c,i} + \theta^*_{c,w_i^{(\ell-1)}}, \sigma_c^{2(\ell-1)}, y_i\right) & \text{if } d_{c,i} = 0 \end{cases} \qquad (5.5)$$

Step 2: Allocate each individual to one of the mixture components by sampling $w_i^{(\ell)}$ $(W_i^{(\ell)} \in \{1, \ldots, K\})$ from a discrete distribution with probabilities:

[1] In our implementation, we also set an upper bound of 6, in order to maintain the draws within reasonable bounds and enhance the convergence of the MCMC sampler. Indeed, $\exp(6)$ corresponds to 403 quarters (more than 100 years).

$$\Pr\left(W_i^{(\ell)} = k \mid y_i, x_i, c_i\right) \tag{5.6}$$
$$= \frac{\pi_k^{(\ell-1)}\left[\prod_{c=1}^{M} f_c\left(y_{c,i}^{*\,(\ell)} \mid x_{c,i}; \beta_c^{(\ell-1)}, \theta_{c,k}^{*(\ell-1)}, \sigma_c^{2(\ell-1)}\right)\right]}{\sum_{j=1}^{K}\pi_j^{(\ell-1)}\left[\prod_{c=1}^{M} f_c\left(y_{c,i}^{*\,(\ell)} \mid x_{c,i}; \beta_c^{(\ell-1)}, \theta_{c,j}^{*(\ell-1)}, \sigma_c^{2(\ell-1)}\right)\right]},$$

thus setting $s_{i,k}^{(\ell)} = 1$ if $w_i^{(\ell)} = k$ and $s_{i,k}^{(\ell)} = 0$ otherwise;

Step 3: Sample stick-breaking weights ψ from a conditionally conjugate Beta distribution, and update the mixture weights π:

Step 3.1: Sample $\psi_k^{(\ell)}$ ($k = 1, \ldots, K-1$, with $\psi_K = 1$):

$$\psi_k^{(\ell)} \sim \text{Beta}\left(\psi_k \middle| 1 + \sum_{i=1}^{n} 1_{\left[w_i^{(\ell)}=k\right]},\ \phi^{(\ell-1)} + \sum_{i=1}^{n} 1_{\left[w_i^{(\ell)}>k\right]}\right) \tag{5.7}$$

Step 3.2: Update π_k:

$$\pi_k^{(\ell)} = \psi_k^{(\ell)} \prod_{j<k}\left(1 - \psi_j^{(\ell)}\right) \tag{5.8}$$

Step 4: Sample $\beta_{c,\cdot}^{(\ell)}$ from a conjugate multivariate p_c-dimensional normal posterior distribution $\text{MVN}\left(\beta_{c,\cdot} \mid B_{1,c}, B_{2,c}\right)$, where

$$\begin{aligned} B_{1,c} &= B_{2,c}\left(x_{c,\cdot}/\sigma_c^{2(\ell-1)}\right)'\left(y_{c,\cdot}^{*(\ell)} - \theta_{c,w^{(\ell)}}^{*(\ell-1)}\right) \\ B_{2,c} &= \left(x_{c,\cdot}' x_{c,\cdot}/\sigma_c^{2(\ell-1)} + \Sigma_{\beta_c}^{-1}\right)^{-1} \end{aligned} \tag{5.9}$$

where $x_{c,\cdot} = (x_{c,1}, \ldots, x_{c,n})'$ is a $n \times p_c$-dimensional matrix, $y_{c,\cdot}^* = \left(y_{c,1}^*, \ldots, y_{c,n}^*\right)'$, $\theta_{c,w}^* = \left(\theta_{c,w_1}^*, \ldots, \theta_{c,w_n}^*\right)'$ (iteration omitted on w for ease of notation) and $\Sigma_{\beta_c} = 9I_{p_c}$ denotes the covariance matrix of β_c;

Step 5: Draw $\theta_k^{*(\ell)} \sim \mathrm{MVN}\left(\theta_{\cdot,k} \mid \Theta_{1,k}, \Theta_{2,k}\right)$, where

$$\begin{aligned}
\Theta_{1,k} &= \Theta_{2,k}\left[\Sigma_y^{(\ell)^{-1}}\left(\sum_{i:s_{i,k}^{(\ell)}=1} y_i^{*(\ell)} - \beta^{(\ell)}x_{\cdot,i}\right) + \Sigma_\theta^{(\ell-1)^{-1}} m_\theta{}^{(\ell-1)}\right] \qquad (5.10)\\
\Theta_{2,k} &= \left(\Sigma_\theta^{(\ell-1)^{-1}} + n_k^{(\ell)}\Sigma_y^{(\ell)^{-1}}\right)^{-1}\\
n_k^{(\ell)} &= \sum_{i=1}^{n} s_{i,k}^{(\ell)}\\
\Sigma_y &= \mathrm{diag}\left(\sigma_1^{2(\ell-1)}, \ldots, \sigma_M^{2(\ell-1)}\right)\\
y_i^{*(\ell)} - \beta^{(\ell)}x_{\cdot,i} &= \begin{bmatrix} y_{1,i}^{*}{}^{(\ell)} - \beta_1^{(\ell)}x_{1,i}\\ \cdots \\ y_{M,i}^{*}{}^{(\ell)} - \beta_M^{(\ell)}x_{M,i}\end{bmatrix}
\end{aligned}$$

Step 6: Sample $m_\theta \sim \mathrm{MVN}\left(m_\theta \mid M_1, M_2\right)$, where:

$$\begin{aligned}
M_1 &= M_2\left(\Sigma_\theta^{(\ell-1)^{-1}}\sum_{k=1}^{K}\theta_k^{*(\ell)}1_{\left[n_k^{(\ell)}>0\right]}\right) \qquad (5.11)\\
M_2 &= \left(\frac{1}{9}I_M + \Sigma_\theta^{(\ell-1)^{-1}}\sum_{k=1}^{K}1_{\left[n_k^{(\ell)}>0\right]}\right)^{-1}
\end{aligned}$$

Step 7: Sample Σ_θ from the conjugate posterior which is the Inverse-Wishart distribution with degrees of freedom Λ_2 and scale matrix Λ_3:

$$\Sigma_\theta^{(l+1)} \sim \text{Inv-Wishart}\left(\Sigma_\theta \mid \Lambda_2, \Lambda_3\right) \qquad (5.12)$$

where

$$\begin{aligned}
\Lambda_2 &= 8 + \sum_{k=1}^{K}1_{\left[n_k^{(\ell)}>0\right]} \qquad (5.13)\\
\Lambda_3 &= 8\lambda_1 + \sum_{k=1}^{K}1_{\left[n_k^{(\ell)}>0\right]}\left(\theta_k^{*(\ell)} - m_\theta^{(\ell)}\right)\left(\theta_k^{*(\ell)} - m_\theta^{(\ell)}\right)'
\end{aligned}$$

Step 8: Sample $\sigma_c^{2(\ell)}$ from the Inverse Gamma distribution with shape $\Gamma_{c,1}$ and rate $\Gamma_{c,2}$:

$$\begin{aligned}
\Gamma_{c,1} &= 1 + 0.5n \qquad (5.14)\\
\Gamma_{c,2} &= 1 + 0.5\sum_{i=1}^{n}\left(y_{c,i}^{*(\ell)} - x_{c,i}\beta_c^{(\ell)} - \theta^{*(\ell-1)}_{c,w_i^{(\ell)}}\right)^2
\end{aligned}$$

Step 9: Sample ϕ. We follow Escobar & West (1995) who devised the following steps for conditional conjugacy:

Step 9.1: Sample $\zeta \sim \text{Beta}\left(\zeta \mid \phi^{(\ell-1)} + 1, n\right)$;

Step 9.2: Sample $Z \sim \text{Bernoulli}\left(Z \mid \pi_\zeta\right)$, where

$$\pi_\zeta = \frac{\sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]}}{n\left(1 - \ln \zeta\right) + \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]}} \tag{5.15}$$

Step 9.3: Sample ϕ:

$$\begin{aligned} \phi^{(\ell)} \sim & 1_{[Z=1]}\text{Gamma}\left(\phi \middle| 1 + \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]}, 1 - \ln \zeta\right) \\ & + 1_{[Z=0]}\text{Gamma}\left(\phi \middle| \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]}, 1 - \ln \zeta\right) \end{aligned} \tag{5.16}$$

In our analysis we implement the Two-Data Augmentation MCMC scheme, by running Step 1-9 for 50,000 iterations. Then, we discard the first 40,000 draws (burn-in) to ensure the sampler converged towards a stationary posterior distribution, and thin the resulting sample by retaining only every 20th simulated draw in order to reduce both the autocorrelation of the parameter draws.

The code ensuring the reproducibility of the results of this work is available on the Github repository `https://github.com/ungolof/AFT-DPM-R`.

6. Results

6.1. Convergence

We run four chains of the two-Data Augmentation sampler outlined in Section 5 with sparse starting values, in order to assess whether different chains convergence towards the same distribution. We observe how the chains converge towards a stationary distribution for all parameters, and these mix very well, except for θ^* and $\pi\left(\psi\right)$ (see Figures A.1-A.4 in Appendix). This result is expected, due to the label switching problem which characterizes mixture distributions (see Betancourt (2017), Ungolo et al. (2020) and Ungolo & van den Heuvel (2022) and the references therein for further details). This is a problem only in terms of interpretation of the groups from the results of one chain compared to another. Nevertheless, this does not represent an issue when making predictions, or when the purpose is to learn the parameters which are common to all mixture components, such as the regression coefficients β. Furthermore, the marginal posterior density of all parameters are unimodal[2].

[2]Plots available upon request to the authors.

As illustrated in Section 5, we learn the posterior distribution of the parameters by setting an upper bound of $K = 25$. Throughout the sampling process, we observe that all 25 mixture components had at least one observation. From a deeper inspection of the output of each MCMC chain we note that by looking at the value of $n_k^{(\ell)}$, a minimum of 8 and a maximum of 10 components (average 8.8) had at least than 550 observations, corresponding to 2.5% of the total number of observations in the training set (Figure 6.1). When using a prior distribution with a larger mean for the elements of Σ_θ, we obtain a similar evidence (minimum of 7, maximum of 9, with an average of 8.5). These results show that the data are informative about the parameters, since their posterior distribution are not sensitive with respect to the prior specification, together with the appropriateness of the choice of K.

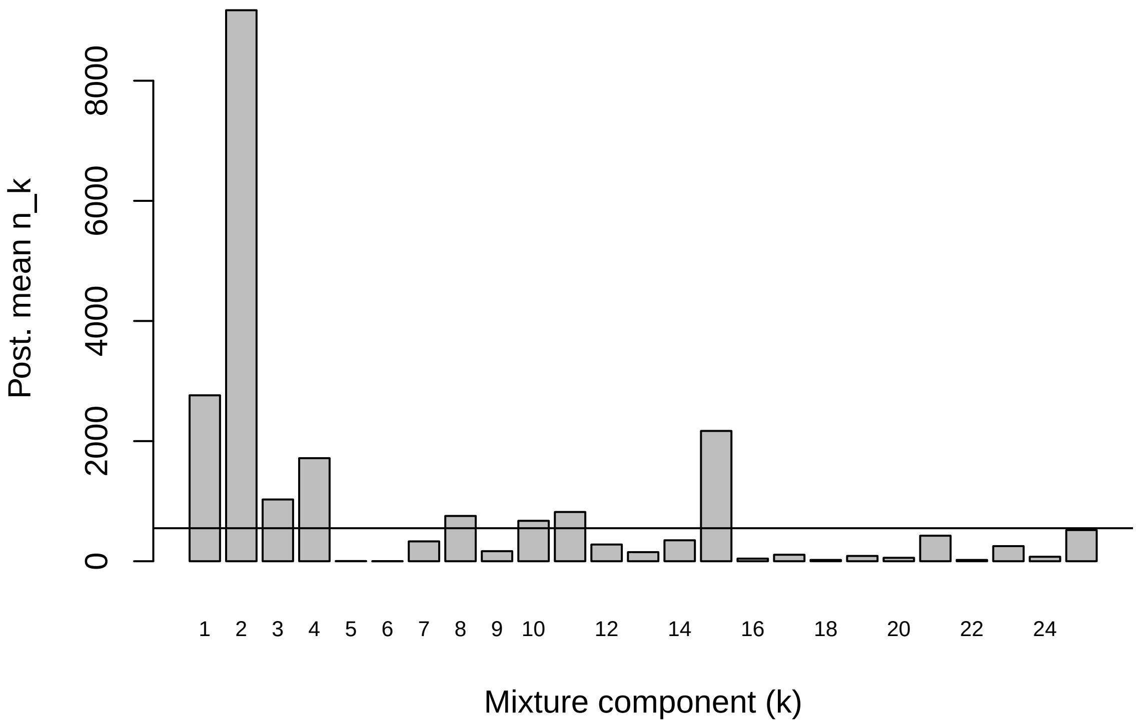

Figure 6.1: Barplot of the posterior average of the mixture component occupancy ($n_k^{(\ell)}$, $k = 1, \ldots, K$) based on the retained posterior sample. The horizontal black line indicates the level of 550 units.

6.2. Model analysis

Regression analysis of surrending risk

Table 6.1 shows the posterior mean and the 95% credible interval of the key parameters for the surrending cause. The marginal posterior distribution of β excludes the value of zero from the 95% credible interval, despite their prior was centered on the null value. This is an indication of the statistical significance of the covariate within the analysis. The time to surrending is negatively associated with respect to the standardized annual premium amount due to the negative coefficient of $\beta_{\text{surr,AP}}$. Similarly, a larger value of the Dow Jones variation is associated with a higher propensity to surrender. As noted in Milhaud & Dutang (2018), the policyholders of participating policies in the data set, which do not benefit from the good financial market conditions, have a higher propensity to surrender, as they can have the possibility to reinvest the *cash value* of the policy in the market at higher returns. Conversely, policyholders with an accidental death rider (ADR=1), females (G=1), and with an annual (or over-annual) premium frequency are associated with a lower propensity to surrender (that is, a "longer" time to surrender their policy). Finally, people younger than 54 years old show a higher propensity to surrender. This result is in line with respect to the theory, since younger people may be more likely to pursue personal projects. All these results are consistent with practitioners' intuition and the numerical evidences in Milhaud & Dutang (2018).

The most relevant mixture component parameters $\theta^*_{\text{surr},\cdot}$ and mixture weights $\pi_\cdot$ show a narrow 95% credible intervals. Their different value, denotes the presence of additional sources of heterogeneity, which may not be explained by the variables included in the linear regression function.

Table 6.1: Posterior summaries of β_{surr}, σ^2_{surr}, $\theta_{\text{surr},k}$, π_k for $k = 1, 2, 3, 4, 15$

Parameter	**Mean**	**95% CI**
$\beta_{\text{surr,AP}}$	-0.0411	$(-0.0514; -0.0306)$
$\beta_{\text{surr,DJ}}$	-0.6450	$(-0.6563; -0.6351)$
$\beta_{\text{surr,ADR}}$	0.0534	(0.0233; 0.0883)
$\beta_{\text{surr,G}}$	0.0310	(0.0095; 0.0501)
$\beta_{\text{surr,PF}}$	0.0744	(0.0540; 0.0964)
$\beta_{\text{surr,UWA}}$	-0.1036	$(-0.1379; -0.0733)$
σ^2_{surr}	0.0669	(0.0621; 0.0722)
π_1	0.1257	(0.1084; 0.1425)
π_2	0.4170	(0.4018; 0.4321)
π_3	0.0468	(0.0342; 0.0555)
π_4	0.0780	(0.0715; 0.0850)
π_{15}	0.0983	(0.0921; 0.1049)
$\theta^*_{\text{surr},1}$	4.8801	(4.8176; 4.9473)
$\theta^*_{\text{surr},2}$	4.1100	(4.0785; 4.1416)
$\theta^*_{\text{surr},3}$	5.2727	(5.2145; 5.3312)
$\theta^*_{\text{surr},4}$	2.5661	(2.5180; 2.6092)
$\theta^*_{\text{surr},15}$	3.1892	(3.1440; 3.2334)

Time to event dependence

In order to obtain an insight about the correlation among the time to events, we sample a large number of log-time to events Y, conditional on sampled mixture allocation component S, using the posterior mean of the parameters. The correlation coefficients are shown in Table 6.2. We note how the model was capable to capture the positive correlation between all causes of decrement. In addition, these coefficients are relatively different from the specification of λ_1 in the prior distribution of the base distribution of the random effect θ, meaning that the data are informative about the statistical association among the time to the three events.

Table 6.2: Correlation matrix of $Y_{\text{Surr}}, Y_{\text{Death}}, Y_{\text{Other}}$

	Surrending	**Death**	**Other**
Surrending	1	0.325	0.297
Death	0.325	1	0.889
Other	0.297	0.889	1

Group analysis

A by-product of the DPM approach of this work is the possibility to investigate *a posteriori* the resulting classes, by creating groups of units in order to obtain additional insights about the similarities of the latter. We carry out this analysis by using the Bayes' rule. Under this method, for the ith individual we calculate her k-class probability, denoted as $q_{i,k}$, conditional on her observable data and the model parameters as follows:

$$
\begin{aligned}
q_{i,k} &= \Pr\left(S_i = k \mid y_i, x_{1,i}, \ldots, x_{M,i};\right) \qquad (6.1)\\
&= \frac{\pi_k \prod_{c=1}^{M} f_c\left(y_i \mid x_{c,i}; \theta^*_{c,k}\right)^{d_{c,i}} \left(1 - F_c\left(y_i \mid x_{c,i}; \theta^*_{c,k}\right)\right)^{1-d_{c,i}}}{\sum_{j=1}^{K}\left[\pi_j \prod_{c=1}^{M} f_c\left(y_i \mid x_{c,i}; \theta^*_{c,j}\right)^{d_{c,i}} \left(1 - F_c\left(y_i \mid x_{c,i}; \theta^*_{c,j}\right)\right)^{1-d_{c,i}}\right]}
\end{aligned}
$$

Hence, each unit i is hard-assigned to a specific class s_i by setting $s_i = k$ if $q_{i,k} > q_{i,j}$ for $j \neq k$. The quantities in equation (6.1) are obtained using the posterior means of the parameters. Table 6.3 illustrates the composition of the largest four classes obtained using the Bayes' rule, and the posterior mean of their corresponding random effect $\theta^*_{\cdot,k}$. These four classes cover almost 80% of the training sample, and do not always to correspond to the five main groups identified in the analysis of the posterior distribution of the parameters.

Table 6.3: Features of the four largest classes resulting from the Bayes' rule.

	Group 2	Group 4	Group 11	Group 15	Train. sample
% Composition	57.7	8.2	4.3	8.3	–
Annual Prem. (mean in $)	536.83	648.02	641.64	650.68	560.88
Accidental D. Rider (Yes in %)	17.4	14.0	13.8	12.5	16.4
Pr. Freq. (Ann+Oth in %)	41.8	30.4	34.9	29.8	38.9
UW Age (0-54 in %)	80.4	84.5	84.1	84.8	81.4
Surrending (in %)	14.7	100	100	92.5	38
θ^*_{Surr}	4.11	2.57	1.76	3.19	–
θ^*_{Death}	5.00	4.48	4.03	2.02	–
θ^*_{Other}	4.90	4.39	3.78	3.91	–

The Bayes' rule creates one group (Group 2) characterized by a lower amount of annual premium, and a slightly lower percentage of people underwriting the policy at a younger age. This group has indeed a significantly lower percentage of policyholders surrending the policy, and a larger value of the random effect parameter $\theta_{\cdot,2}$ for all three competing events. On the other hand, Group 4, 11, and 15 include a pool of policyholders with similar characteristics in terms of composition in terms of the covariates, and surrending behavior. Their different value of the random effect θ may be indicative of the presence of other factors, unobserved to the researcher which explain the heterogeneity among those groups. For example, the policyholders hard-assigned to Group 4 may have a higher propensity to surrender compared to those in Group 15 despite the similar composition in terms of features because they may have a financial advisor who can affect the choices of the policyholder, or else, they may be facing a period of financial distress.

Posterior predictive density of future observations

Figure 6.2 shows the posterior predictive density (PPD) for new observations $\tilde{t}_{\text{surr}} = \exp\left(\tilde{y}_{\text{surr}}\right)$ of the time to surrending for different configurations of the covariates $\tilde{x}$, which is input exogeously. Let $\Delta = \left(\beta, \theta^*, \sigma^2, \pi\left(\psi\right)\right)$ and data $= \{\mathbf{t}, \mathbf{x}, \mathbf{c}\}$. The PPD is defined as follows:

$$f\left(\tilde{t}_{\text{surr}} \mid \tilde{x}; \text{data}\right) = \int f\left(\tilde{t}_{\text{surr}} \mid \tilde{x}; \Delta\right) p\left(\Delta \mid \text{data}\right) \mathrm{d}\Delta \qquad (6.2)$$

This density is computed by sampling several values of $\tilde{y}_{\text{surr}}$ from the density of equation (6.2):

1) Sample Δ from the posterior distribution $p\left(\Delta \mid \text{data}\right)$ obtainable using the MCMC sampler of Section 5;

3) Sample mixture index s from a discrete distribution with parameter vector $\pi\left(\psi\right)$;

3) Sample $\tilde{y}_{\text{surr}} \sim \mathrm{N}\left(\tilde{y}_{\text{surr}} \mid \beta_{\text{surr}}\tilde{x} + \theta^*_{\text{surr},s}, \sigma^2_{\text{surr}}\right)$, where all parameters are from Step 1) and the mixture index s is obtained in Step 2);

4) Get $\tilde{t}_{\text{surr}} = \log \tilde{y}_{\text{surr}}$.

In our analysis we repeat Step 1)-4) for 1,000,000 times.

The top panel of Figure 6.2 shows how for each configuration of $\tilde{x}$ in terms of accidental death rider, payment frequency and underwriting age (with an average annual premium amount and zero variation in the DJIA), the estimated model yields a bimodal mixture density. In this way, it is clearly shown how this model can flexibly account for heterogeneity and dependence among the observations, since this model accommodates for various shapes of the density function. Such flexibility is more evident when inspecting the two modes of the resulting densities: differences due to the covariates values become more evident when looking to the rightmost mode of the posterior density, which may be the key cause of the shift in the mean value of the time to surrending.

The bottom panel of Figure 6.2 shows how the these difference can lead to a significant difference in the posterior predictive expected time to surrending for different level of the annual premium.

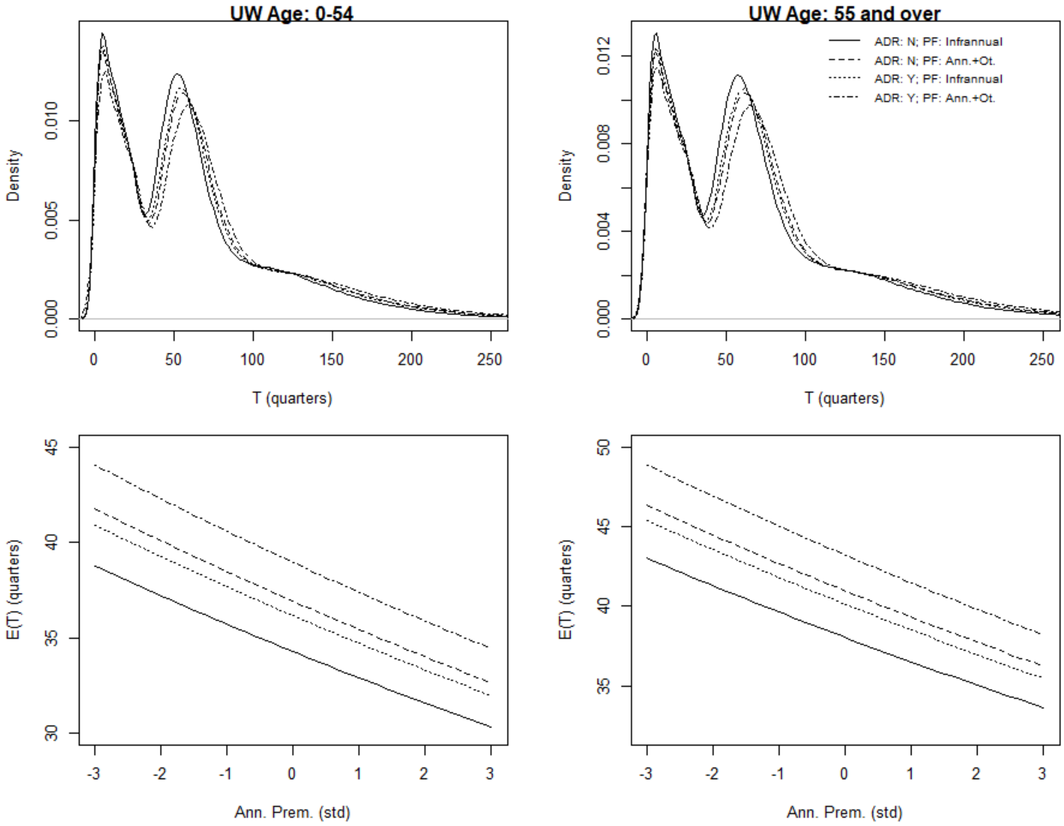

Figure 6.2: Posterior predictive density of the time to surrending (top), and posterior predictive expected time to surrending (bottom) underwritten until age 54 (left) and from age 55 onwards (right).

6.3. Predictive ability of the model

Competing models

We analyse the predictive ability of the DPM-Regression model by analysing the out-of-sample performance of the model using the held-out data set of 7,321 observations. Such results are compared with those obtainable under the Cox Proportional Hazard model (Cox PH, Cox (1972)) and the subdistribution approach (SBD) of Fine & Gray (1999).

Both approaches assume that for each cause the distribution of the time to event is characterized by a semi-parametric form for the hazard function $\mu_c(t)$:

$$\mu_c(t) = \mu_{0,c}(t) \exp(\beta_c x_{c,i}) \tag{6.3}$$

where $\mu_{0,c}$ is the nonparametric hazard function for cause c. The Cox PH and the SBD approaches differ in the definition of the time to event: for the former, the occurrence of any cause, censores the time to event for the others, while the SBD defines the time to event for the cause of interest as follows:

$$T_c^{sbd} = T\mathbb{1}_{[C=c]} + \infty\mathbb{1}_{[C\neq c]} \tag{6.4}$$

The choice of these two models is motivated by their widespread use, their easy understanding among practitioners, and by the availability of software for their estimation, such as the R package `survival` (Therneau (2022)). At different extents, these two models assume that the causes of decrement are independently distributed, while on the other hand are greatly flexible due to their nonparametric specification of the baseline hazard function $\mu_{0,c}(t)$.

In our comparisons, we use the same set of covariates as for the DPM model of this work.

Out-of-sample performance of the models

First, we compute the predicted surrending rate for each quarter $[s_q - s_{q+1})$ using the following conditional probability:

$$\hat{r}_q = \frac{1}{n_{s_q}} \sum_{i \in \mathcal{R}_{s_q}} \widehat{\Pr}\left(s_q < T_{\text{surr},i} \leq s_{q+1}, c_i = 1 \mid T_{\text{surr},i} > s_q, T_{\text{death},i} > T_{\text{surr},i}, T_{\text{other},i} > T_{\text{surr},i}, x_i\right) \tag{6.5}$$

where n_{s_q} is the size of the at-risk population $\mathcal{R}_{s_q}$. The $\widehat{\Pr}$ denotes that such probability is computed under the estimated models. For the DPM model, we use the posterior mean of the parameters as a point estimate. Its computation is detailed in Appendix C.

Figure 6.3 shows the value of r_q, and the Rolling Root Mean Square Error (R-RMSE) by quarter, calculated as:

$$\text{R-RMSE}_Q = \sqrt{\frac{1}{Q}\sum_{q=1}^{Q}\left(\hat{r}_q^{\text{Model}} - r_q^{\text{Empirical}}\right)^2} \tag{6.6}$$

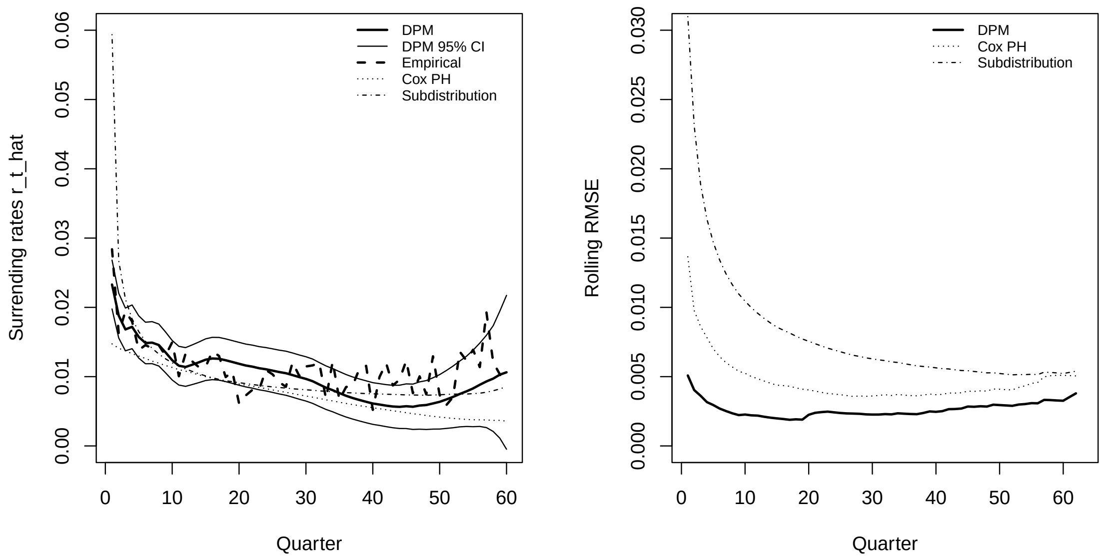

Figure 6.3: Predicted surrending rates by quarter $\hat{r}_q$ (left) and R-RMSE$_Q$ (right).

We observe how the fitted rates using the DPM approach of this work closely resemble those empirically observed, especially in the first 8 years (32 quarters), after which the behavior of surrending rates becomes more erratic due to the lower number of observations. During this period, the corresponding 95% confidence interval of these rates, seem to include those empirically observed, except at the 20th quarter. The subdistribution approach seems instead to overestimate the surrending rates in the first year of the policy, and then to slightly underestimate these rates afterwards. The Cox proportional hazard model instead appears to underestimate the surrending rates throughout the quarters. In general, we observe how the rates obtainable under the DPM approach can have a changing shape throughout the quarter, while the rates under the SBP and the Cox PH seem to have a smoother behavior. The Rolling RMSE shows how the DPM model returns surrending rates closer to those empirically observed, compared to the other two competing approaches. An appropriate prediction of the surrending rates, especially in the first few years since policy issue is of fundamental importance for insurance firms, in order to recover the initial expenses.

We closely inspect this result, by looking at the predicted rates by value of the accidental death rider and the payment frequency (6.4).

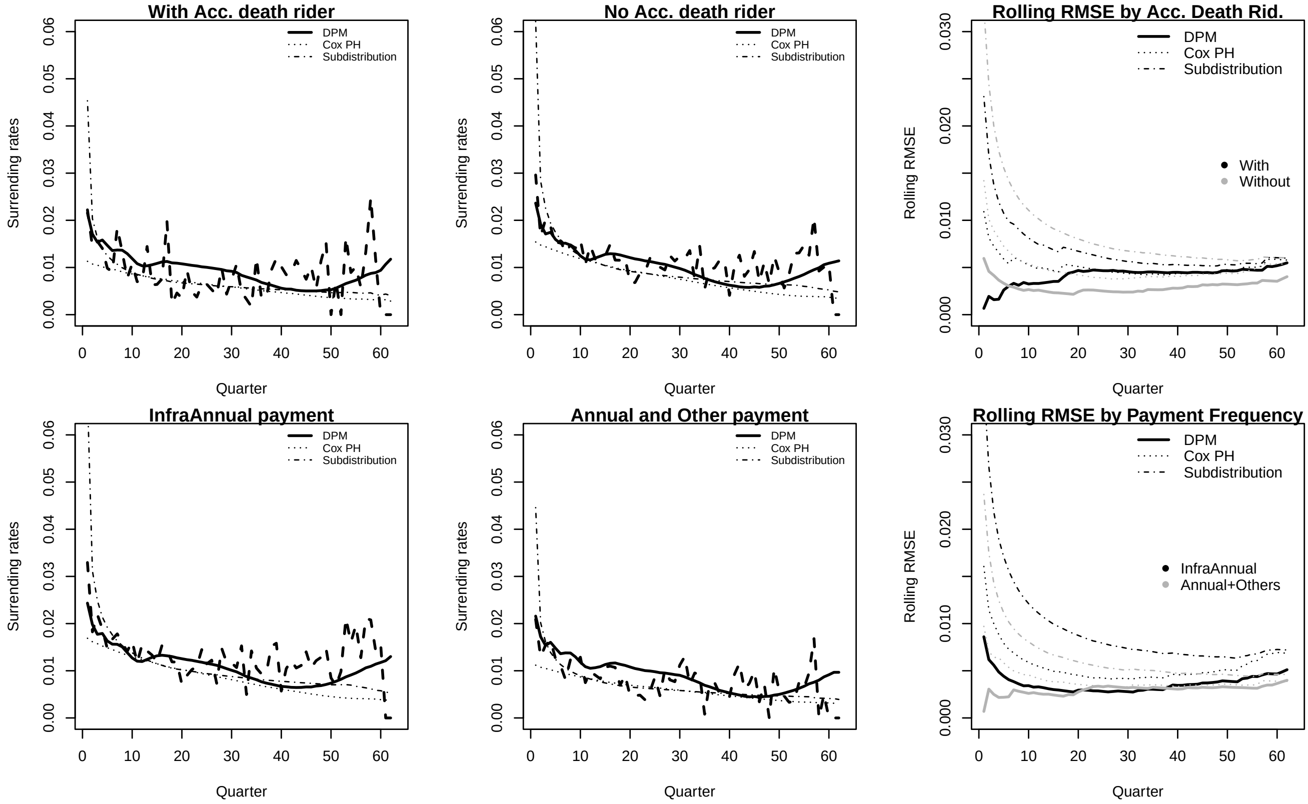

Figure 6.4: Predicted surrending rates by quarter $\hat{r}_q$ (left) and R-RMSE$_Q$ (right) by covariate Accidental death rider (top panel) and Payment frequency (bottom panel).

Again, we observe how the DPM-based surrending rates shows a changing shape throughout the quarters, while for the other models, the shape is more smooth.

When analysing the results by covariate value, we note once again how the DPM model outperforms the Cox PH and the SBD approaches, especially in the first quarters since policy inception. The performance deteriorates however, when looking at those policies with an accidental death rider and to those with a lesser frequent premium payment. Indeed, for these two cases we can observe how the DPM and the Cox PH models show a comparable performance in terms of R-RMSE in the long run.

7. Discussion and future work

This paper outlines a flexible modelling approach using flexibly distributed multivariate random effects, allowing to point-identify the joint distribution of the time to competing risk events. This approach, based on Dirichlet Process Mixture model weakens the sensitivity of the resulting inference with respect to the specification of a parametric distribution for the random effect. The model could easily accommodate for individual-specific covariates. We analyse this approach from a fully Bayesian perspective and outline an efficient MCMC scheme allowing for closed-form updates. The framework hereby developed has been implemented to the empirical analysis of the surrending rates of a US life insurer data set. The model showed an improved prediction of the surrending rates, compared to standard approaches, especially in the first years since policy inception, which are crucial for recovering the initial expenses.

As mentioned in Section 4, other parametric models with a non-parametric random components can be specified. An alternative to the log-normal model of this work consists of a log-t model (Sha et al. 2006), which allows for heavier tails compared to the log-Normal model hereby analysed. This model similarly allows for a fully conditionally conjugate posterior distribution, which gives the possibility to devise an MCMC sampler with parameter updates similar to those described in Section 5.3. Furthermore, the approach offers the possibility to explore the set of covariates which may turn out statistically significant in the analysis of competing risk events. An extension in this direction, consists of enriching the model with a stochastic search variable selection component, in the spirit of George & McCulloch (1993), with a prior favouring sparseness in order to obtain a more parsimonious model (see Lucas et al. (2006)). This additional layer of analysis would sensibly increase the number of parameters we need to learn. Therefore, a faster alternative approach to the MCMC sampler of this work is the use of Variational Bayes methods (Blei et al. 2016), which can be easily applied to the fully conjugate model of this work. Finally, the model can be further expanded by specifying a random component, which depends on common covariates afftecting all competing risks, explaining the dependence among the competing causes. For statistical models outside the competing risk analysis domain, Barcella et al. (2017) reviews a set of approaches which can be developed to obtain Dependent Dirichlet Processes (MacEachern 2000). We leave these extensions for future work on the analysis of more complex data sets.

A. Trace plots of some model parameters

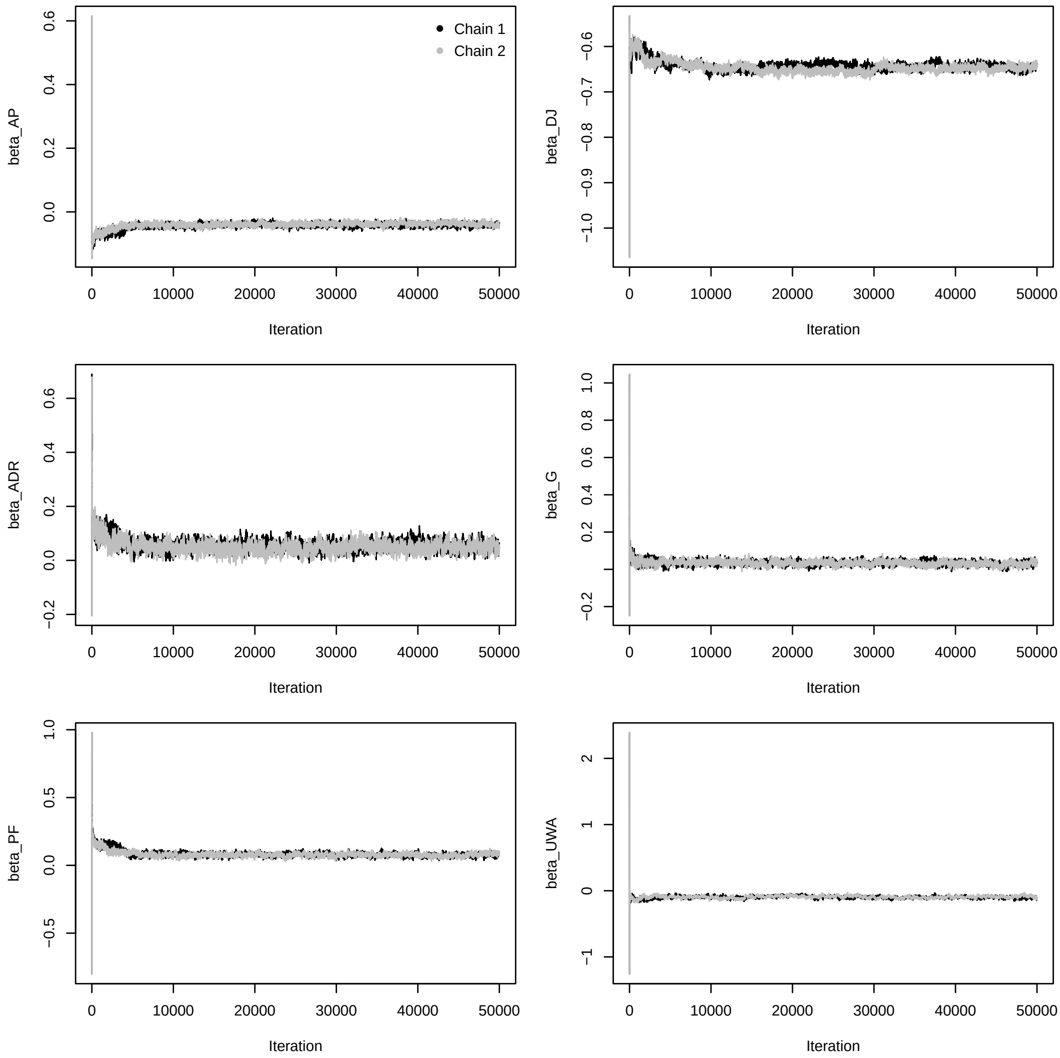

Figure A.1: Trace plots of β_{surr}.

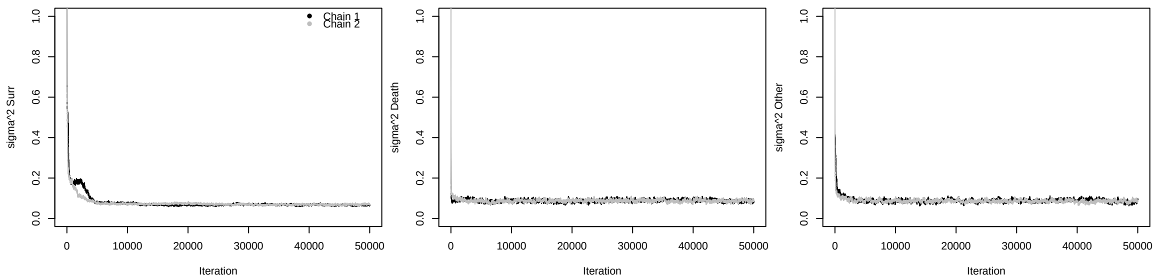

Figure A.2: Trace plots of σ_c^2 ($c \in \{\text{surr}, \text{death}, \text{other}\}$).

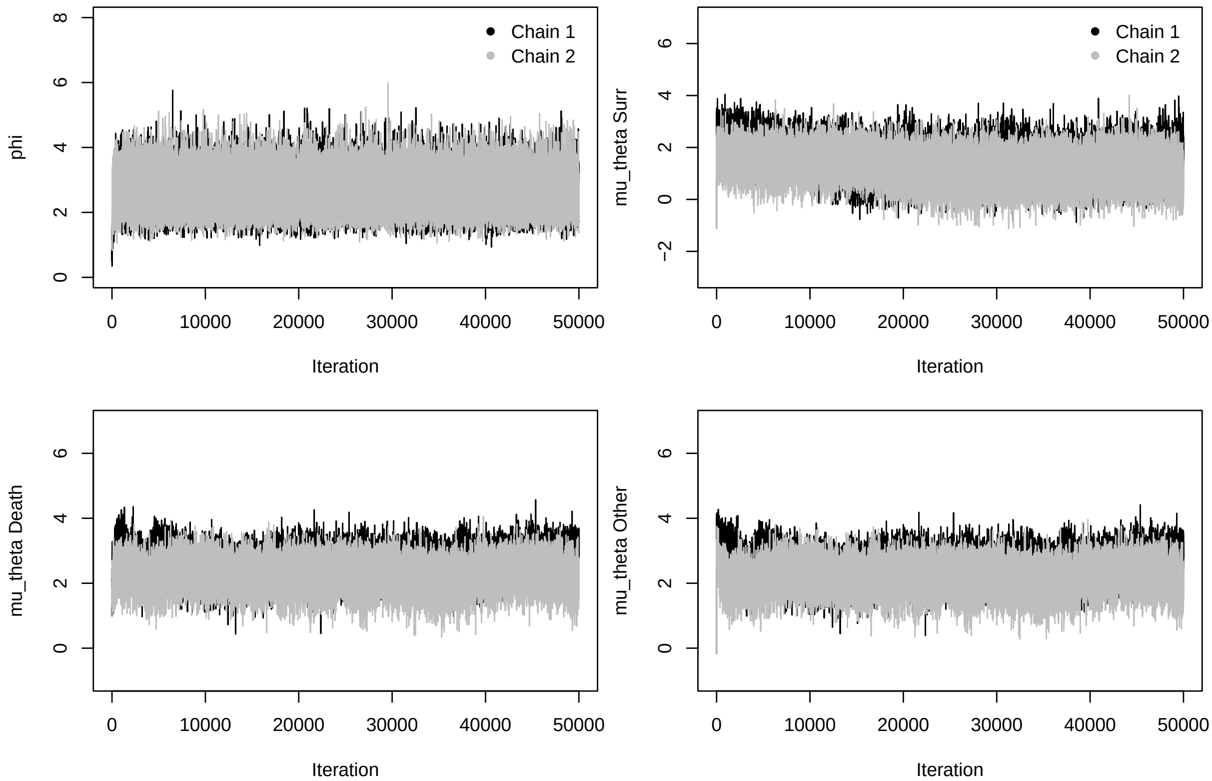

Figure A.3: Trace plots of ϕ (top-left) and of m_θ.

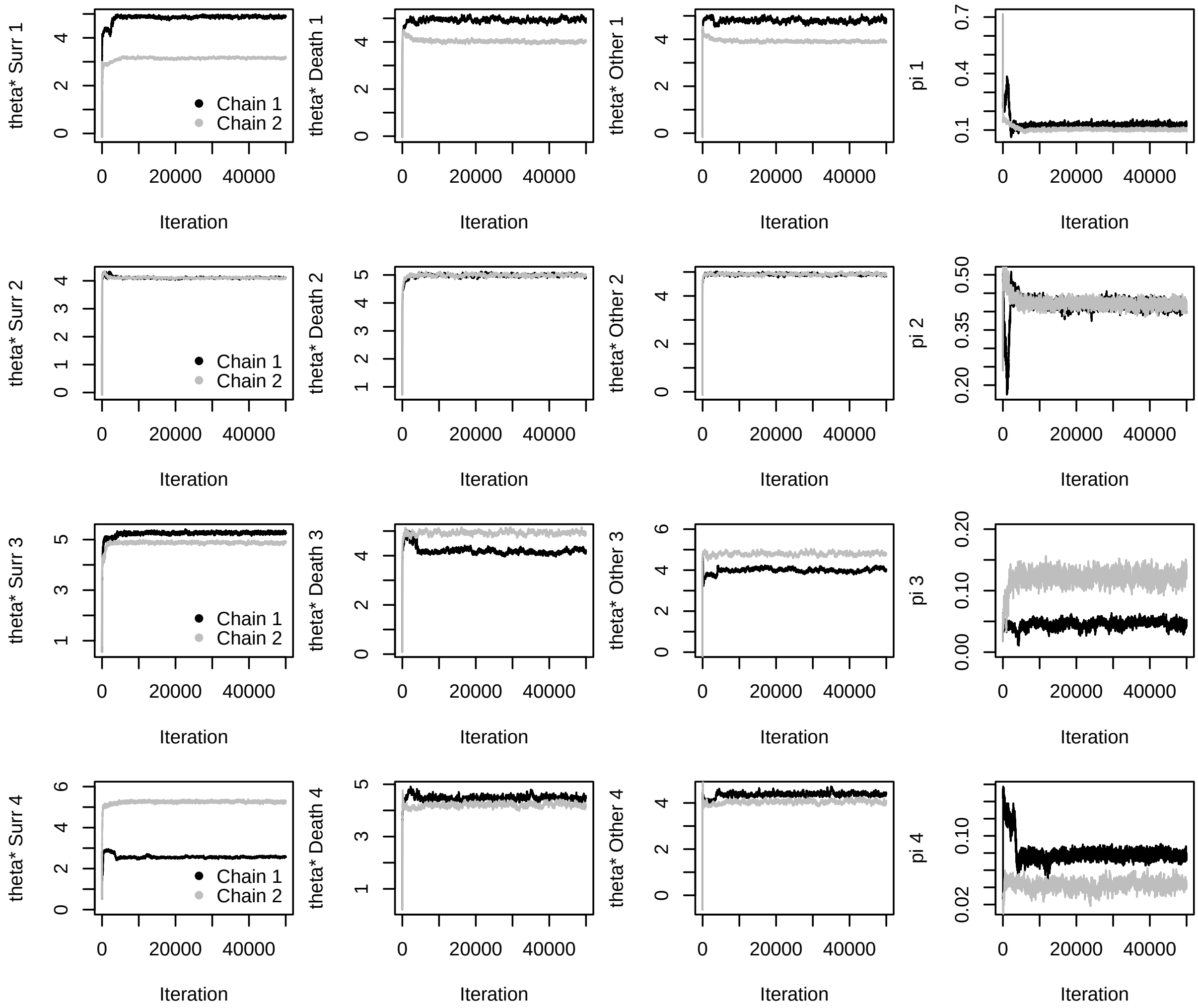

Figure A.4: Trace plots of $\theta^*_{\cdot,k}$ and π_k $(k = 1, \ldots, 4)$.

B. Posterior summaries

Table B.1: Posterior summaries of the model parameters.

	Surrending			Death			Other	
Parameter	**Mean**	**95% Cred. Int.**	**Parameter**	**Mean**	**95% Cred. Int.**	**Parameter**	**Mean**	**95% Cred. Int.**
$\beta_{\text{surr,AP}}$	-0.0411	$(-0.0514; -0.0306)$	$\beta_{\text{death,G}}$	-0.0030	$(-0.0389; 0.0310)$	$\beta_{\text{other,AP}}$	0.0277	(0.0117; 0.0428)
$\beta_{\text{surr,DJ}}$	-0.6450	$(-0.6563; -0.6351)$	$\beta_{\text{death,UWA}}$	0.0078	$(-0.0347; 0.0514)$	$\beta_{\text{other,AP}}$	-0.0133	$(-0.0468; 0.0166)$
$\beta_{\text{surr,ADR}}$	0.0534	(0.0233; 0.0883)	$\beta_{\text{death,LP}}$	0.0089	$(-0.0291; 0.0438)$	$\beta_{\text{other,AP}}$	-0.0261	$(-0.0689; 0.0180)$
$\beta_{\text{surr,G}}$	0.0310	(0.0095; 0.0501)	$\beta_{\text{death,PF}}$	-0.0355	$(-0.0738; 0.0010)$			
$\beta_{\text{surr,PF}}$	0.0744	(0.0540; 0.0964)						
$\beta_{\text{surr,UWA}}$	-0.1036	$(-0.1379; -0.0733)$						
σ^2_{surr}	0.0669	(0.0621; 0.0722)	σ^2_{death}	0.0881	(0.0770; 0.0999)	σ^2_{other}	0.0869	(0.0697; 0.1012)
$\theta^*_{\text{surr},1}$	4.8801	(4.8176; 4.9473)	$\theta^*_{\text{death},1}$	4.9563	(4.8232; 5.1109)	$\theta^*_{\text{other},1}$	4.8192	(4.6623; 4.9739)
$\theta^*_{\text{surr},2}$	4.1100	(4.0785; 4.1416)	$\theta^*_{\text{death},2}$	4.9964	(4.9192; 5.0656)	$\theta^*_{\text{other},2}$	4.8987	(4.8352; 4.9548)
$\theta^*_{\text{surr},3}$	5.2727	(5.2145; 5.3312)	$\theta^*_{\text{death},3}$	4.1762	(4.0287; 4.3215)	$\theta^*_{\text{other},3}$	4.0139	(3.8554; 4.1621)
$\theta^*_{\text{surr},4}$	2.5661	(2.5180; 2.6092)	$\theta^*_{\text{death},4}$	4.4835	(4.3683; 4.6043)	$\theta^*_{\text{other},4}$	4.3870	(4.2736; 4.5037)
$\theta^*_{\text{surr},5}$	2.3551	$(-1.2814; 5.9564)$	$\theta^*_{\text{death},5}$	-2.6156	$(-2.9122; -2.2655)$	$\theta^*_{\text{other},5}$	-2.7766	$(-3.0677; -2.4257)$
$\theta^*_{\text{surr},6}$	-6.4157	$(-6.9300; -5.8661)$	$\theta^*_{\text{death},6}$	2.4684	$(-1.4455; 5.6519)$	$\theta^*_{\text{other},6}$	2.5678	$(-1.4029; 5.7305)$
$\theta^*_{\text{surr},7}$	3.4469	(3.1963; 3.7040)	$\theta^*_{\text{death},7}$	2.0855	(1.9180; 2.2254)	$\theta^*_{\text{other},4}$	1.9504	(1.8039; 2.0799)
$\theta^*_{\text{surr},8}$	4.8966	(4.6735; 5.1694)	$\theta^*_{\text{death},8}$	3.5980	(3.3795; 3.8451)	$\theta^*_{\text{other},4}$	3.4284	(3.2040; 3.6744)
$\theta^*_{\text{surr},9}$	2.8114	(2.5702; 3.0749)	$\theta^*_{\text{death},9}$	1.4097	(1.2651; 1.5459)	$\theta^*_{\text{other},4}$	1.3016	(1.1671; 1.4389)
$\theta^*_{\text{surr},10}$	1.0860	(1.0197; 1.1521)	$\theta^*_{\text{death},10}$	4.8116	(4.5129; 5.0901)	$\theta^*_{\text{other},4}$	4.7591	(4.4514; 5.0781)
$\theta^*_{\text{surr},11}$	1.7583	(1.6242; 1.8905)	$\theta^*_{\text{death},11}$	3.8516	(3.5459; 4.1328)	$\theta^*_{\text{other},4}$	3.7785	(3.4751; 4.0767)
$\theta^*_{\text{surr},12}$	3.1405	(2.8579; 3.5585)	$\theta^*_{\text{death},12}$	2.8908	(2.7538; 3.0486)	$\theta^*_{\text{other},4}$	2.7641	(2.6380; 2.8984)
$\theta^*_{\text{surr},13}$	-0.4326	$(-0.5229; -0.2581)$	$\theta^*_{\text{death},13}$	3.0429	(2.6615; 3.3488)	$\theta^*_{\text{other},4}$	3.0057	(2.6544; 3.3378)
$\theta^*_{\text{surr},14}$	0.3320	(0.2711; 0.4026)	$\theta^*_{\text{death},14}$	4.7482	(4.0971; 5.1682)	$\theta^*_{\text{other},14}$	4.7140	(4.1261; 5.1490)
$\theta^*_{\text{surr},15}$	3.1892	(3.1440; 3.2334)	$\theta^*_{\text{death},15}$	4.0283	(3.9599; 4.1021)	$\theta^*_{\text{other},15}$	3.9103	(3.8602; 3.9654)
$\theta^*_{\text{surr},16}$	-2.9640	$(-3.1656; -2.8376)$	$\theta^*_{\text{death},16}$	3.0520	(0.9789; 5.0482)	$\theta^*_{\text{other},16}$	3.0759	(1.0224; 5.1092)
$\theta^*_{\text{surr},17}$	2.2156	(2.0328; 2.4054)	$\theta^*_{\text{death},17}$	0.7638	(0.6334; 0.8908)	$\theta^*_{\text{other},17}$	0.6256	(0.5141; 0.7298)
$\theta^*_{\text{surr},18}$	2.0652	(0.1821; 4.8315)	$\theta^*_{\text{death},18}$	-1.1368	$(-1.2760; -0.9965)$	$\theta^*_{\text{other},18}$	-1.2589	$(-1.3906; -1.1082)$
$\theta^*_{\text{surr},19}$	-1.1603	$(-1.3027; -0.8878)$	$\theta^*_{\text{death},19}$	3.4972	(1.9551; 5.0119)	$\theta^*_{\text{other},19}$	3.4841	(1.8878; 5.0204)
$\theta^*_{\text{surr},20}$	2.4622	(1.4899; 3.9776)	$\theta^*_{\text{death},20}$	-0.1684	$(-0.2812; -0.0455)$	$\theta^*_{\text{other},20}$	-0.3051	$(-0.4088; -0.1999)$
$\theta^*_{\text{surr},21}$	4.5588	(4.2010; 4.9242)	$\theta^*_{\text{death},21}$	3.1420	(2.6683; 4.1928)	$\theta^*_{\text{other},21}$	2.9856	(2.5406; 4.0499)
$\theta^*_{\text{surr},22}$	-3.9615	$(-4.0983; -3.8109)$	$\theta^*_{\text{death},22}$	2.9278	(0.1497; 5.1619)	$\theta^*_{\text{other},22}$	2.9750	(0.1293; 5.1946)
$\theta^*_{\text{surr},23}$	3.1712	$(-1.8390; 4.5894)$	$\theta^*_{\text{death},23}$	2.7227	(2.3262; 3.8224)	$\theta^*_{\text{other},23}$	2.6129	(2.2003; 3.9229)
$\theta^*_{\text{surr},24}$	-2.0423	$(-2.4403; -1.8631)$	$\theta^*_{\text{death},24}$	3.0610	(1.6868; 4.9038)	$\theta^*_{\text{other},24}$	3.0668	(1.6004; 4.8771)
$\theta^*_{\text{surr},25}$	1.9557	(1.6375; 2.0821)	$\theta^*_{\text{death},25}$	3.3197	(3.0466; 3.9524)	$\theta^*_{\text{other},25}$	3.2369	(2.9331; 3.8969)
ϕ	2.6748	(1.6955; 4.0501)	π_1	0.1257	(0.1084; 0.1425)	π_2	0.417	(0.4018; 0.4321)
π_3	0.0468	(0.0342 0.0555)	π_4	0.0780	(0.0715; 0.0850)	π_5	0.0002	(0.0001; 0.0004)
π_6	0.0001	(0.0000; 0.0003)	π_7	0.0151	(0.0103; 0.0193)	π_8	0.0343	(0.0262; 0.0405)
π_9	0.0076	(0.0059; 0.0094)	π_{10}	0.0307	(0.0269; 0.0346)	π_{11}	0.0373	(0.0246; 0.0497)
π_{12}	0.0127	(0.0072; 0.0174)	π_{13}	0.0070	(0.0056; 0.0084)	π_{14}	0.0159	(0.0137; 0.0180)
π_{15}	0.0983	(0.0921; 0.1049)	π_{16}	0.0020	(0.0013; 0.0028)	π_{17}	0.0050	(0.0037; 0.0063)
π_{18}	0.0011	(0.0007; 0.0015)	π_{19}	0.0040	(0.0029; 0.0054)	π_{20}	0.0026	(0.0020; 0.0034)
π_{21}	0.0193	(0.0017; 0.0319)	π_{22}	0.0010	(0.0006; 0.0015)	π_{23}	0.0114	(0.0006; 0.0182)
π_{24}	0.0034	(0.0014; 0.0047)	π_{25}	0.0235	(0.0069; 0.0403)			

C. Computation of the conditional surrending probability

The probability distribution of the time to surrender can be obtained as follows:

$$\widehat{\Pr}\left(s_q < T_{\text{surr},i} \leq s_{q+1}, c_i = 1 \mid T_{\text{surr},i} > s_q, T_{\text{death},i} > T_{\text{surr},i}, T_{\text{other},i} > T_{\text{surr},i}, x_i\right) = \frac{\widehat{F}\left(s_{q+1}\right) - \widehat{F}\left(s_q\right)}{1 - \widehat{F}\left(s_q\right)}$$

where, by simplifying the notation we obtain

$$\begin{aligned}\widehat{F}(s) &= \widehat{\Pr}\left(T_{\text{surr},i} \leq s, T_{\text{death},i} > T_{\text{surr},i}, T_{\text{other},i} > T_{\text{surr},i} \mid x_i\right) \\ &= \int_0^s \widehat{f}_{\text{surr}}\left(s \mid x_{\text{surr},i}\right)\left(1 - \widehat{F}_{\text{death}}\left(s \mid x_{\text{death},i}\right)\right)\left(1 - \widehat{F}_{\text{other}}\left(s \mid x_{\text{other},i}\right)\right) \mathrm{d}s. \quad \text{(C.1)}\end{aligned}$$

$\widehat{f}_c(s \mid x_{c,i})$ and $\widehat{F}_c(s \mid x_{c,i})$ denote respectively the density and the cumulative distribution function of the cth cause under the estimated models.

For the Cox PH model and the SBD approach we have:

$$\begin{aligned}\widehat{f}_c(s \mid x_{c,i}) &= \exp\left(-\int_0^s \widehat{\mu}_{0,c}(u) \exp\left(\widehat{\beta}_c x_{c,i}\right) \mathrm{d}u\right) \widehat{\mu}_{0,c}(s) \exp\left(\widehat{\beta}_c x_{c,i}\right) \\ \widehat{F}_c(s \mid x_{c,i}) &= \exp\left(-\int_0^s \widehat{\mu}_{0,c}(u) \exp\left(\widehat{\beta}_c x_{c,i}\right) \mathrm{d}u\right) \quad \text{(C.2)}\end{aligned}$$

which we approximate by assuming a piecewise constant baseline hazard function $\mu_{0,c}$.

When using the subdistribution approach we have for the cause of interest c $\widehat{F}(s) = \widehat{F}_c(s \mid x_{c,i})$.

References

Aldous, D. J. (1985), Exchangeability and related topics, *in* P. L. Hennequin, ed., 'École d'Été de Probabilités de Saint-Flour XIII — 1983', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–198.

Arnold, B. C. & Brockett, P. L. (1983), 'Identifiability for dependent multiple decrement/competing risk models', *Scandinavian Actuarial Journal* **1983**(2), 117–127.
URL: *https://doi.org/10.1080/03461238.1983.10408697*

Barcella, W., De Iorio, M. & Baio, G. (2017), 'A comparative review of variable selection techniques for covariate dependent dirichlet process mixture models', *Canadian Journal of Statistics* **45**(3), 254–273.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/cjs.11323*

Betancourt, M. J. (2017), 'Identifying bayesian mixture models'.
URL: *https://betanalpha.github.io/assets/case_studies/identifying_mixture_models.html*

Blackwell, D. & MacQueen, J. B. (1973), 'Ferguson distributions via polya urn schemes', *Ann. Statist.* **1**(2), 353–355.
URL: *https://doi.org/10.1214/aos/1176342372*

Blei, D., Kucukelbir, A. & McAuliffe, J. (2016), 'Variational inference: A review for statisticians', *Journal of the American Statistical Association* **112**.

Collett, D. (2003), *Modelling Survival Data in Medical Research, Second Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.

Cox, D. R. (1972), 'Regression models and life-tables', *Journal of the Royal Statistical Society. Series B (Methodological)* **34**(2), 187–220.

Crowder, M. (1996), 'On assessing independence of competing risks when failure times are discrete', *Lifetime Data Analysis* **2**(2), 195–209.

Crowder, M. (1997), 'A test for independence of competing risks with discrete failure times', *Lifetime Data Analysis* **3**(3), 215.

Dimitrova, D., Haberman, S. & Kaishev, V. (2013), 'Dependent competing risks: Cause elimination and its impact on survival', *Insurance: Mathematics and Economics* **53**.

Dunson, D. B. (2010), *Nonparametric Bayes applications to biostatistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, p. 223–273.

Dutang, C. & Charpentier, A. (2020), *CASdatasets: Insurance datasets*. R package version 1.0-11.

Emoto, S. E. & Matthews, P. C. (1990), 'A weibull model for dependent censoring', *Ann. Statist.* **18**(4), 1556–1577.

Escarela, G. & Carrière, J. (2003), 'Fitting competing risks with an assumed copula', *Statistical methods in medical research* **12**, 333–49.

Escobar, M. D. & West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476550*

Escobar, M., Krayzler, M., Ramsauer, F., Saunders, D. & Zagst, R. (2016), 'Incorporation of stochastic policyholder behavior in analytical pricing of gmabs and gmdbs', *Risks* **4**(4).
URL: *https://www.mdpi.com/2227-9091/4/4/41*

Ferguson, T. S. (1973), 'A bayesian analysis of some nonparametric problems', *Ann. Statist.* **1**(2), 209–230.
URL: *https://doi.org/10.1214/aos/1176342360*

Fine, J. P. & Gray, R. J. (1999), 'A proportional hazards model for the subdistribution of a competing risk', *Journal of the American Statistical Association* **94**(446), 496–509.
URL: *http://www.jstor.org/stable/2670170*

Frees, E. & Valdez, E. (1998), 'Understanding relationships using copulas', *North American Actuarial Journal* **2**, 1–25.

Frees, E. W., Carriere, J. & Valdez, E. (1996), 'Annuity valuation with dependent mortality', *The Journal of Risk and Insurance* **63**(2), 229–261.
URL: *http://www.jstor.org/stable/253744*

George, E. I. & McCulloch, R. E. (1993), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* **88**(423), 881–889.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476353*

Gorfine, M. & Hsu, L. (2011), 'Frailty-based competing risks model for multivariate survival data', *Biometrics* **67**(2), 415–426.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2010.01470.x*

Heinz, D. (2014), Nonparametric mixed membership models, *in* 'Handbook of Mixed Membership Models and Their Applications', Handbook of Modern Statistical Methods, Chapman & Hall, pp. 89–116.

Huang, X. & Wolfe, R. A. (2002), 'A frailty model for informative censoring', *Biometrics* **58**(3), 510–520.

Ishwaran, H. & James, L. F. (2001), 'Gibbs sampling methods for stick-breaking priors', *Journal of the American Statistical Association* **96**(453), 161–173.
URL: *https://doi.org/10.1198/016214501750332758*

Jackson, D., White, I. R., Seaman, S., Evans, H., Baisley, K. & Carpenter, J. (2014), 'Relaxing the independent censoring assumption in the cox proportional hazards model using multiple imputation', *Statistics in Medicine* **33**(27), 4681–4694.

Kalli, M., Griffin, J. & Walker, S. (2011), 'Slice sampling mixture models', *Statistics and Computing* **21**, 93–105.

Knoller, C., Kraut, G. & Schoenmaekers, P. (2016), 'On the propensity to surrender a variable annuity contract: An empirical analysis of dynamic policyholder behavior', *Journal of Risk and Insurance* **83**(4), 979–1006.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jori.12076*

Lo, A. Y. (1984), 'On a Class of Bayesian Nonparametric Estimates: I. Density Estimates', *The Annals of Statistics* **12**(1), 351 – 357.
URL: *https://doi.org/10.1214/aos/1176346412*

Lucas, J., Carvalho, C., Wang, Q., Bild, A., Nevins, J. R. & West, M. (2006), *Sparse Statistical Modelling in Gene Expression Genomics*, Cambridge University Press, p. 155–176.

Lunn, M. & McNeil, D. (1995), 'Applying cox regression to competing risks.', *Biometrics* **51 2**, 524–32.

MacEachern, S. (2000), 'Dependent dirichlet process', *Technical Report, Department of Statistics, The Ohio State University* .

Milhaud, X. & Dutang, C. (2018), 'Lapse tables for lapse risk management in insurance: a competing risk approach', *European Actuarial Journal* **8**, 97–126.

Orbanz, P. (2014), 'Lecture notes on bayesian nonparametrics'.

Papaspiliopoulos, O. & Roberts, G. O. (2008), ‘Retrospective markov chain monte carlo methods for dirichlet process hierarchical models’, *Biometrika* **95**(1), 169–186.
URL: *http://www.jstor.org/stable/20441450*

Richards, S. J. (2008), ‘Applying survival models to pensioner mortality data’, *British Actuarial Journal* **14**(2), 257–303.

Richards, S., Kaufhold, K. & Rosenbusch, S. (2013), ‘Creating portfolio-specific mortality tables: a case study’, *European Actuarial Journal* **3**(2), 295–319.
URL: *https://doi.org/10.1007/s13385-013-0076-6*

Rotnitzky, A., Farall, A., Bergesio, A. & Scharfstein, D. (2007), ‘Analysis of failure time data under competing censoring mechanisms’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(3), 307–327.

Scharfstein, D. O. & Robins, J. M. (2002), ‘Estimation of the failure time distribution in the presence of informative censoring’, *Biometrika* **89**(3), 617–634.

Sethuraman, J. (1994), ‘A constructive definition of dirichlet priors’, *Statistica Sinica* **4**(2), 639–650.
URL: *http://www.jstor.org/stable/24305538*

Sha, N., Tadesse, M. G. & Vannucci, M. (2006), ‘Bayesian variable selection for the analysis of microarray data with censored outcomes’, *Bioinformatics* **22**(18), 2262–2268.
URL: *https://doi.org/10.1093/bioinformatics/btl362*

Tanner, M. A. & Wong, W. H. (1987), ‘The calculation of posterior distributions by data augmentation’, *Journal of the American Statistical Association* **82**(398), 528–540.

Therneau, T. M. (2022), *A Package for Survival Analysis in R*. R package version 3.4-0.
URL: *https://CRAN.R-project.org/package=survival*

Tsiatis, A. (1975), ‘A nonidentifiability aspect of the problem of competing risks’, *Proceedings of the National Academy of Sciences of the United States of America* **72**(1), 20–22.

Ungolo, F., Kleinow, T. & Macdonald, A. S. (2020), ‘A hierarchical model for the joint mortality analysis of pension scheme data with missing covariates’, *Insurance: Mathematics and Economics* **91**, 68 – 84.
URL: *http://www.sciencedirect.com/science/article/pii/S0167668720300032*

Ungolo, F. & van den Heuvel, E. R. (2022), ‘Inference on latent factor models for informative censoring’, *Statistical Methods in Medical Research* **0**(0), 09622802211057290. PMID: 35077263.
URL: *https://doi.org/10.1177/09622802211057290*

Ungolo, F. & van den Heuvel, E. R. (n.d.), ‘A joint bayesian nonparametric frailty model for the analysis of competing risks’, *Working paper* .

Vaupel, J. W. & Yashin, A. I. (1985), 'The deviant dynamics of death in heterogeneous populations', *Sociological Methodology* **15**, 179–211.
URL: *http://www.jstor.org/stable/270850*

Walker, S. G. (2007), 'Sampling the dirichlet mixture model with slices', *Communications in Statistics - Simulation and Computation* **36**(1), 45–54.
URL: *https://doi.org/10.1080/03610910601096262*

Wienke, A. (2014), *Frailty Models*, John Wiley Sons, Ltd.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06941*

Yashin, A. I. & Iachine, I. A. (1995), 'Genetic analysis of durations: Correlated frailty model applied to survival of danish twins', *Genetic Epidemiology* **12**.

Zheng, M. & Klein, J. P. (1995), 'Estimates of marginal survival for dependent competing risks based on an assumed copula', *Biometrika* **82**(1), 127–138.