

Complete reference:

Computational Statistics and Data Analysis, 124: 151–167.
doi:10.1016/j.csda.2018.02.007
https://doi.org/10.1016/j.csda.2018.02.007

Fast Bayesian inference using Laplace approximations in a flexible promotion time cure model based on P-splines [☆]

Gressani Oswaldo^{a,*}, Philippe Lambert^{a,b}

^a*Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, B-1348, Louvain-la-Neuve, Belgium*

^b*Faculté des Sciences Sociales, Méthodes Quantitatives en Sciences Sociales, Université de Liège, Place des Orateurs 3, B-4000, Liège, Belgium*

Abstract

Bayesian methods for flexible time-to-event models usually rely on the theory of Markov chain Monte Carlo (MCMC) to sample from posterior distributions and perform statistical inference. These techniques are often plagued by several potential issues such as high posterior correlation between parameters, slow chain convergence and foremost a strong computational cost. **A novel methodology is proposed to overcome** the inconvenient facets intrinsic to MCMC sampling with the major advantage that posterior distributions of latent variables can rapidly be approximated with a high level of accuracy. This can be achieved by exploiting the synergy between Laplace's method for posterior approximations and P-splines, a flexible tool for nonparametric modeling. **The methodology is developed in** the class of cure survival models, a useful extension of standard time-to-event models where it is assumed that an unknown proportion of unidentified (cured) units will never experience the monitored event. An attractive feature of this new approach is that point estimators and credible intervals can be straightforwardly constructed even for complex functionals of latent model variables. The properties of the proposed methodology are evaluated using simulations and illustrated on two real datasets. **The fast computational speed and accurate results suggest** that the combination of P-splines and Laplace approximations can be considered as a serious competitor of MCMC to make inference in semi-parametric models, as illustrated on survival models with a cure fraction.

Keywords: Promotion time cure model, P-splines, Survival Analysis, Laplace Approximation, Approximate Bayesian Inference

1. Introduction

There is a growing interest for cure rate models in survival analysis as witnessed by the number of recently published papers on that topic in statistical journals. These models have gained in popularity as they intrinsically account for long-term survivors that will never experience the event of interest even when followed-up for an extended time period. The promotion time (cure) model introduced by Yakovlev et al. (1996) is motivated by cancer tumor kinetics, the biological mechanism underlying the proliferation and growth of carcinogenic cells. In particular, let $N \sim \text{Poisson}(\phi(\mathbf{x}))$ be the number of carcinogenic cells affecting a given subject with mean $\phi(\mathbf{x}) = \exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta})$. To the i^{th} cell is associated a latent

[☆]R code is available as supplementary material in the electronic version of the paper.

*Corresponding author.

Email addresses: oswaldo.gressani@uclouvain.be (Gressani Oswaldo), p.lambert@ulg.ac.be (Philippe Lambert)

event time $T_i \geq 0$ representing the duration necessary for the cell to grow to a detectable tumor mass. Latent event times $\{T_1, \dots, T_N\}$ are assumed to be independently and identically distributed with common cumulative distribution function $F(t)$ and the **observed survival time is defined as $T = \min\{T_1, \dots, T_N\}$** . When a Cox proportional hazards model [Cox \(1972\)](#) is used to model the N conditional latent distributions $F(t_i|\mathbf{z}) = 1 - S_0(t_i)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})}$, $i = 1, \dots, N$ one can show that **the resulting survival function of T** is given by ([Tsodikov 1998](#); [Chen et al. 1999](#))

$$\begin{aligned} S_p(t|\mathbf{x}, \mathbf{z}) &= \exp(-\phi(\mathbf{x})F(t|\mathbf{z})) \\ &= \exp(-\exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta})(1 - S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})})). \end{aligned} \quad (1)$$

In this model, a subject is cured when $N = 0$, an event arising with a probability given by $P(N = 0 | \mathbf{x}, \mathbf{z}) = \lim_{t \rightarrow \infty} S_p(t|\mathbf{x}, \mathbf{z}) = \exp(-\phi(\mathbf{x}))$. Alternative specifications are proposed in the literature to model the distribution of latent event times $F(t_i)$, for example [Ibrahim et al. \(2001\)](#) propose a semi-parametric form for the latent distribution involving a smoothing parameter controlling the degree of parametricity in the right tail of the population survival function, while [Zeng et al. \(2006\)](#) introduce a semi-parametric class of cure models taking into account a subject-specific frailty. Model (1) can be estimated by maximum likelihood methods in a frequentist setting (see [Tsodikov 2002, 2003](#)). From a Bayesian perspective, [Yin and Ibrahim \(2005\)](#) assume a piecewise exponential model for the baseline survival function with a tradeoff between model flexibility and the number of partitions of the time axis. More recently, [Bremhorst and Lambert \(2016\)](#) use a large number of B-splines to specify the baseline hazard and, following [Eilers and Marx \(1996\)](#), counterbalance the flexibility of the model by using a roughness penalty based on finite differences of adjacent B-spline coefficients.

The rather complex structure of the posterior distributions in the latter Bayesian frameworks requires the use of MCMC techniques. For such models, the MCMC toolbox is usually accompanied by a large computational burden and challenging convergence problems under the original parameterization. A crucial component explaining the inefficiency of rejection sampling techniques is a strong posterior correlation appearing firstly among latent variables and secondly between latent variables and hyperparameters of the model, thus having a global impact on convergence speed and autocorrelation. Integrated Nested Laplace Approximations (INLA) is a sampling-free Bayesian methodology recently introduced in the literature that allows to obtain marginal posteriors in the class of latent Gaussian models and has been recognized to be an interesting alternative to standard MCMC methods. In this dimension, [Rue et al. \(2009\)](#) and [Martino \(2007\)](#) are the pioneering references showing how to perform approximate Bayesian inference in latent Gaussian models via Laplace approximations.

While INLA has been shown to work well in a large variety of applications like stochastic volatility models [Martino et al. \(2011a\)](#), generalized dynamic linear models [Ruiz-Cárdenas et al. \(2012\)](#) and spatio-temporal disease mapping models [Schrödle and Held \(2011\)](#), there seems to be little work related to survival analysis or penalized B-spline models. Among the contributions on the subject, we can cite [Fong et al. \(2010\)](#) who combine INLA and O’Sullivan splines in a nonparametric smoothing setting. [Martino et al. \(2011b\)](#) investigate the use of INLA with the R-INLA package (www.r-inla.org) by considering a Cox model where the baseline hazard has a parametric or semi-parametric specification. Also, [Jiang et al. \(2014\)](#) study the effect of environmental radiation on cancer by using a cure fraction mixture survival model with a Weibull distribution for event times.

We investigate how Laplace approximations can be extended and combined with penalized B-splines in the context of a semi-parametric promotion time cure model. Bridging the gap between Laplace’s method and regression splines brings a twofold advantage. First, it provides a fast computational approach to approximate posterior distributions and second, the spline dimension allows for a flexible specification of

the baseline distribution yielding smooth estimates of survival quantities. Another crucial point is that in contrast to the classic INLA approach which focuses mainly on posterior marginal univariate distributions, our methodology permits to compute reliable approximations to the posterior joint distributions of latent variables including regression parameters, with the implication that set estimators can be derived even for complicated functions of latent variables such as the baseline or conditional population survival functions.

Accordingly, the end user will be endowed with a powerful and rapid tool for making inference in the promotion time cure model. Furthermore, while the code design underlying INLA assumes a one-to-one connection between data points and a subset of the latent field, implying that the dimension of the latter grows with the sample size n , our modeling strategy choice is more efficient as it involves a latent field of a dimension unaffected by the number of observations. Hence, given that the number of B-splines is fixed (to a large value and counterbalanced by a roughness penalty) in the P-spline approach [Eilers and Marx \(2010\)](#), the latent field dimension grows only with the number of regressors in the model and not with n .

This manuscript is organized as follows. In Section 2, the Laplace-P-spline promotion time cure model is defined and the gradient and Hessian of the log-likelihood are computed to obtain a Gaussian approximation of the conditional posterior distribution of the latent field. A strategy is proposed to explore the posterior distribution of the hyperparameter vector and the joint posterior of latent field elements are derived. The construction of credible intervals for the baseline and population survival functions is also addressed here. In Section 3, the merits of the proposed methodology will be assessed by extensive simulations with different scenarios regarding the percentages of cured individuals and right censored subjects. Coverage properties of credible intervals will also be considered. In Section 4, we apply the model to two real datasets and Section 5 concludes with a discussion.

2. Laplace-P-spline promotion time model

2.1. Flexible modeling of the baseline hazard

Following [Rosenberg \(1995\)](#), the log-hazard corresponding to the baseline survival function $S_0(t)$ in (1) is specified as a linear combination of cubic B-splines $h_0(t) = \exp(\theta^T b(t))$, where $b(\cdot) = (b_1(\cdot), \dots, b_K(\cdot))^T$ is a cubic B-spline basis obtained by taking equidistant knots on the compact set $[0, t_u]$, with t_u the upper bound of the follow-up and $\theta = (\theta_1, \dots, \theta_K)^T$ is the vector of B-spline coefficients. Under this specification, the baseline survival function in (1) can be approximated as follows

$$\begin{aligned} S_0(t) &= \exp\left(-\int_0^t \exp(\theta^T b(s)) ds\right) \\ &\approx \exp\left(-\sum_{j=1}^{j(t)} \exp(\theta^T b(s_j)) \Delta_j\right), \end{aligned} \quad (2)$$

where the integral (having no analytic solution) is approximated by the rectangle method with $[0, t_u]$ partitioned into J (say 300) small equal width intervals, i.e. $J_j = [\varphi_{j-1}, \varphi_j]$ with $0 = \varphi_0 < \varphi_1 < \dots < \varphi_J = t_u$, where s_j and Δ_j respectively denote the midpoint and width of J_j and $j(t)$ is an index returning the interval containing t .

2.2. Latent field and priors

The latent field $\xi = (\theta_1, \dots, \theta_K, \beta_0, \dots, \beta_p, \gamma_1, \dots, \gamma_l) \in \mathbb{R}^H$ has dimension $H = K + (p + 1) + l$ and gathers all the latent variables of the model: it contains the B-spline coefficients $\{\theta_k : k = 1, \dots, K\}$, the

regression coefficients $\{\beta_m : m = 0, \dots, p\}$ used to model the expected number of carcinogenic cells and the regression parameters $\{\gamma_s : s = 1, \dots, l\}$ in the Cox model describing the incubation time of a given cell. The key idea behind P-splines [Eilers and Marx \(1996\)](#) is to use a fixed large number of B-spline basis functions and to counterbalance the flexibility by introducing a roughness penalty on finite differences of adjacent B-spline coefficients. The Bayesian analogue [Lang and Brezger \(2004\)](#) translates the roughness penalty into a multivariate normal prior distribution for the spline coefficients $\boldsymbol{\theta}|\lambda \sim \mathcal{N}_K(0, \lambda^{-1}P^{-1})$, with $P = D_r^T D_r + \varepsilon I_K$ where D_r is a $(K-r) \times K$ matrix yielding r^{th} order differences when applied on a K -vector, and λ is a non-negative roughness penalty parameter. For an arbitrary small ε (say $\varepsilon = 10^{-6}$), the diagonal perturbation εI_K ensures that P is full rank. Then, the prior for the full latent field given λ can be written as

$$\boldsymbol{\xi}|\lambda \sim \mathcal{N}_H(\boldsymbol{\mu}_\xi, \Sigma_\xi(\lambda)), \quad \Sigma_\xi(\lambda) = \begin{pmatrix} \lambda^{-1}P^{-1} & 0 \\ 0 & \Sigma_{\beta,\gamma} \end{pmatrix},$$

where $\mathcal{N}_H(\cdot, \cdot)$ denotes a multivariate Gaussian distribution of dimension H , the vector $\boldsymbol{\mu}_\xi$ attributes a zero mean to the B-spline coefficients and a potential informative prior mean on the regression coefficients with (prior) positive-definite covariance matrix $\Sigma_{\beta,\gamma}$. Whenever a priori knowledge on central tendency or correlation measures is available for the regression coefficient vector, it can be incorporated into the prior $\boldsymbol{\xi}|\lambda$ through the mean and covariance structure. The hyperparameters of the model are given by $\boldsymbol{\eta} = (\lambda, \delta)$ as, following [Jullion and Lambert \(2007\)](#), we use a robust specification for the roughness penalty prior $\lambda|\delta \sim \text{Gamma}(\nu/2, (\nu\delta)/2)$ with an uninformative proper distribution on parameter $\delta \sim \text{Gamma}(a_\delta, b_\delta)$. The latter reference shows that when $a_\delta = b_\delta$ are set to a small value (say 10^{-4}), the estimated curve is not sensitive to the choice of ν (here set equal to 3).

2.3. Posterior and Laplace approximation

Let $\mathcal{D}_i = (t_i, \tau_i, \mathbf{x}_i, \mathbf{z}_i)$ denote the observables for unit i , with t_i the failure or censoring time, τ_i a dichotomous event indicator and $\mathbf{x}_i, \mathbf{z}_i$ the covariates. The log-likelihood function of the promotion time cure model is $\ell(\boldsymbol{\xi}; \mathcal{D}) = \sum_{i=1}^n \{\tau_i \log h_p(t_i|\mathbf{x}_i, \mathbf{z}_i) + \log S_p(t_i|\mathbf{x}_i, \mathbf{z}_i)\}$, where $\mathcal{D} = \bigcup_{i=1}^n \mathcal{D}_i$ and $h_p(\cdot|\mathbf{x}, \mathbf{z})$ is the conditional population hazard function, $h_p(t|\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x}) \exp(\mathbf{z}^T \boldsymbol{\gamma}) S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} h_0(t)$. Using the B-spline specification of the baseline hazard, we can write more compactly $\ell(\boldsymbol{\xi}; \mathcal{D}) \approx \sum_{i=1}^n g_i(\boldsymbol{\xi})$. The scalar-valued function $g_i : \mathbb{R}^H \rightarrow \mathbb{R}$ gives the contribution of the i^{th} unit to the log-likelihood and is given by

$$\begin{aligned} g_i(\boldsymbol{\xi}) &= \tau_i \left(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma} + \boldsymbol{\theta}^T b(t_i) - \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^T b(s_j)) \Delta_j \right) \\ &\quad - \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \left(1 - \exp \left(- \sum_{j=1}^{j(t_i)} \exp(\boldsymbol{\theta}^T b(s_j)) \Delta_j \right)^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \right). \end{aligned}$$

The first step of our procedure is to derive the Laplace approximation of the conditional posterior distribution of the latent variables, namely $p(\boldsymbol{\xi}|\lambda, \mathcal{D}) \propto \exp\left(\sum_{i=1}^n g_i(\boldsymbol{\xi}) - \frac{1}{2} \boldsymbol{\xi}^T Q(\lambda) \boldsymbol{\xi} + \boldsymbol{\xi}^T Q(\lambda) \boldsymbol{\mu}_\xi\right)$, where $Q(\lambda) = \Sigma_\xi^{-1}(\lambda)$ is the precision matrix. One major difference with the theoretical set-up described in [Rue et al. \(2009\)](#) is the dimension of the latent field assumed there to be larger than the number of observations; it is usually much smaller here with $H \ll n$. With non-Gaussian responses, $p(\boldsymbol{\xi}|\lambda, \mathcal{D})$ is non-Gaussian and unknown. To make it tractable, we compute a second-order Taylor expansion of $g_i(\boldsymbol{\xi})$ around some point $\boldsymbol{\xi}^{(0)} \in \mathbb{R}^H$. As shown in [Appendix A](#), this enables us to obtain a quadratic form in $\boldsymbol{\xi}$ for the log-likelihood term and thus a Gaussian approximation of the form

$$\tilde{p}_G(\boldsymbol{\xi}|\lambda, \mathcal{D}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\xi}^T \left(Q(\lambda) - \nabla^2 g_{\boldsymbol{\xi}^{(0)}}\right) \boldsymbol{\xi} + \boldsymbol{\xi}^T \left(\nabla g_{\boldsymbol{\xi}^{(0)}} - \nabla^2 g_{\boldsymbol{\xi}^{(0)}} \boldsymbol{\xi}^{(0)} + Q(\lambda) \boldsymbol{\mu}_\xi\right)\right), \quad (3)$$

where $\nabla g_{\xi^{(0)}}$ and $\nabla^2 g_{\xi^{(0)}}$ are the gradient and the Hessian of the log-likelihood function **evaluated at $\xi^{(0)}$** . The posterior mean and precision matrix in that approximation are shown to be $\xi^{(1)} = (Q(\lambda) - \nabla^2 g_{\xi^{(0)}})^{-1} (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)} + Q(\lambda) \mu_\xi)$ and $Q(\lambda)^{(1)} = (Q(\lambda) - \nabla^2 g_{\xi^{(0)}})$ respectively. Next, we repeat the Taylor expansion around $\xi^{(1)}$ and continue to implement this iterative process in a Newton-Raphson type algorithm to converge towards a Gaussian approximation centered around the posterior mode of $p(\xi|\lambda, \mathcal{D})$. Note that in each iteration, the inversion of a large dimensional matrix of the form $Q(\lambda) - \nabla^2 g$ is required. This is achieved by Householder transformations (Householder, 1958; Golub and Van Loan, 2012), a numerically stable tool for matrix inversion that achieves a QR decomposition through a sequence of orthogonal transformations of the input matrix.

2.4. Exploring the posterior of the hyperparameter vector

The next step consists in exploring the posterior distribution of the hyperparameter vector

$$p(\eta|\mathcal{D}) \propto \frac{\mathcal{L}(\xi; \mathcal{D}) p(\xi|\lambda) p(\lambda|\delta) p(\delta)}{p(\xi|\lambda, \mathcal{D})}, \quad (4)$$

where $\mathcal{L}(\xi; \mathcal{D})$ is the likelihood function. In order to avoid identifiability issues, we follow Bremhorst and Lambert (2016) and fix the last B-spline coefficient $\xi_K = \theta_K$ **to a large value (say 10), denoted c** . This forces the baseline survival function $S_0(\cdot)$ to be virtually zero at the end of the follow-up. Taking this constraint into account and using the Gaussian approximation scheme proposed in Section 2.3, we can approximate (4) as follows

$$\tilde{p}(\eta|\mathcal{D}) = \frac{\mathcal{L}(\xi; \mathcal{D}) p(\xi|\lambda) p(\lambda|\delta) p(\delta)|_{\xi=\xi_{cc}^*(\lambda)}}{\tilde{p}_G(\xi_{-K}|\xi_K = c, \lambda, \mathcal{D})|_{\xi_{-K}=\xi_{cc}^*(\lambda)}}, \quad (5)$$

where $\xi_{cc}^*(\lambda) \in \mathbb{R}^{H-1}$ is the conditional posterior mean of the Gaussian approximation given $\xi_K = c$, and $\xi_{cc}^*(\lambda) \in \mathbb{R}^H$ corresponds to the vector $\xi_{cc}^*(\lambda)$ to which we add the constraint c at position K . These two quantities are derived in Appendix B. All the factors in (5) have mathematically closed forms and can be extensively written as

$$\begin{aligned} \tilde{p}(\eta|\mathcal{D}) \propto & \exp\left(\sum_{i=1}^n g_i(\xi_{cc}^*(\lambda)) - \frac{1}{2}(\xi_{cc}^*(\lambda) - \mu_\xi)^T Q(\lambda)(\xi_{cc}^*(\lambda) - \mu_\xi)\right) |Q(\lambda)|^{\frac{1}{2}} |\Sigma_c^*(\lambda)|^{\frac{1}{2}} \\ & \times \lambda^{\frac{\nu}{2}-1} \delta^{\frac{\nu}{2}+a_\delta-1} \exp(-\delta(b_\delta + \nu\lambda/2)). \end{aligned} \quad (6)$$

Note that δ can be integrated out from (6) to obtain the following approximated marginal posterior density of the penalty parameter

$$\tilde{p}(\lambda|\mathcal{D}) \propto \exp\left(\sum_{i=1}^n g_i(\xi_{cc}^*(\lambda)) - \frac{1}{2}(\xi_{cc}^*(\lambda) - \mu_\xi)^T Q(\lambda)(\xi_{cc}^*(\lambda) - \mu_\xi)\right) |Q(\lambda)|^{\frac{1}{2}} |\Sigma_c^*(\lambda)|^{\frac{1}{2}} \lambda^{\frac{\nu}{2}-1} (b_\delta + \nu\lambda/2)^{-(\nu/2+a_\delta)}. \quad (7)$$

In addition, the conditional posterior of δ is given by $\delta|\lambda, \mathcal{D} \sim \text{Gamma}(\nu/2 + a_\delta, b_\delta + \nu\lambda/2)$ and does not directly depend on the data. Next, our aim is to find a sub-region in the domain of $\tilde{p}(\eta|\mathcal{D})$ that supports most of the posterior probability mass. **In that endeavor, we use an equidistant grid $\mathbf{N}_\lambda = \{\lambda_j\}_{j=1}^{m_1}$ of size $m_1 = 10$ in the domain of $\tilde{p}(\lambda|\mathcal{D})$ that supports approximately 95% of the posterior mass. Then, for each point $\lambda_j \in \mathbf{N}_\lambda$, we construct a regular grid of length $m_2 = 5$ with starting and ending values corresponding to the 2.5th and 97.5th percentiles respectively of the $\text{Gamma}(\nu/2 + a_\delta, b_\delta + \nu\lambda_j/2)$ distribution. This enables us to construct a grid $\mathbf{N}_{\lambda,\delta} = (\lambda^{(m)}, \delta^{(m)})_{m=1}^M \in \mathbb{R}_{++}^2$ with $M = m_1 \times m_2$ points that will be used to compute approximations to the posterior distributions of the latent field elements.**

2.5. Posterior multivariate distribution of the latent field

In Sections 2.3 and 2.4, we have seen that the conditional posterior distribution of the latent field for a given $\xi_K = c$ and λ can be approximated by a Gaussian density. By abuse of notation we have $\tilde{p}_G(\xi_{-K}|\xi_K = c, \lambda, \mathcal{D}) = \tilde{p}_G(\xi_{-K}|\lambda, \mathcal{D}) = \mathcal{N}_{H-1}(\xi_c^*(\lambda), \Sigma_c^*(\lambda))$.

The posterior joint distribution of the latent field ξ_{-K} can be written as

$$\begin{aligned} p(\xi_{-K}|\mathcal{D}) &= \int_0^{+\infty} \int_0^{+\infty} p(\xi_{-K}, \lambda, \delta|\mathcal{D}) d\lambda d\delta \\ &= \int_0^{+\infty} \int_0^{+\infty} p(\xi_{-K}|\lambda, \mathcal{D}) p(\lambda, \delta|\mathcal{D}) d\lambda d\delta. \end{aligned} \quad (8)$$

Given our grid coordinates and their associated weights $\Delta_m = \Delta_{\lambda^{(m)}} \times \Delta_{\delta^{(m)}}$ being the area of the parallelograms in the grid, we can approximate (8) by numerical integration

$$\tilde{p}(\xi_{-K}|\mathcal{D}) = \sum_m \tilde{p}_G(\xi_{-K}|\lambda^{(m)}, \mathcal{D}) \tilde{p}(\lambda^{(m)}, \delta^{(m)}|\mathcal{D}) \Delta_m. \quad (9)$$

The weights of the Gaussian densities in the sum can be normalized

$$\omega_m = \frac{\tilde{p}(\lambda^{(m)}, \delta^{(m)}|\mathcal{D}) \Delta_m}{\sum_m \tilde{p}(\lambda^{(m)}, \delta^{(m)}|\mathcal{D}) \Delta_m},$$

to improve the approximation to the approximate joint posterior distribution of ξ_{-K} , yielding

$$\hat{p}(\xi_{-K}|\mathcal{D}) = \sum_m \omega_m \mathcal{N}_{H-1}(\xi_c^*(\lambda^{(m)}), \Sigma_c^*(\lambda^{(m)})). \quad (10)$$

Equation (10) is a multivariate Gaussian mixture density with mean and covariance matrix analytically known to be (see e.g. Frühwirth-Schnatter 2006) $E(\xi_{-K}|\mathcal{D}) = \sum_m \omega_m \xi_c^*(\lambda^{(m)})$ and $V(\xi_{-K}|\mathcal{D}) = \sum_m \omega_m \Sigma_c^*(\lambda^{(m)}) + \sum_m \omega_m (\xi_c^*(\lambda^{(m)}) - E(\xi_{-K}|\mathcal{D}))(\xi_c^*(\lambda^{(m)}) - E(\xi_{-K}|\mathcal{D}))^T$, respectively. These quantities can be used to compute pointwise estimates and approximate credible intervals for latent field elements. In the next section we show how to derive credible intervals for complex functionals of the latent field.

2.6. Credible intervals

The flexibility of the Laplace-P-spline model can be exploited to compute pointwise credible intervals for complicated functions of the latent field and thus go beyond a marginal analysis. Construction of joint credible bands for subsets of the latent field is also discussed in Sørbye and Rue (2011). In this section, we focus on the derivation of pointwise credible intervals for the baseline survival function $S_0(t)$ and for the population survival function given in (1). The “Delta method” will serve as the main mechanism to derive approximate credible intervals. Using a $\log(-\log(\cdot))$ transform of the baseline survival function, we obtain

$$G_0(\theta^c|t) = \log(-\log S_0(t)) = \log\left(\sum_{j=1}^{j(t)} \exp(\theta^T b(s_j)) \Delta_j\right), \quad (11)$$

where $\theta^c = (\theta_1, \dots, \theta_{K-1})$ and $\theta_K = c$ as the last B-spline coefficient is fixed for identifiability purposes in the cure promotion time model. Using the strategy presented in Section 2.3, one has a normal approximation to the conditional posterior of ξ , namely $\tilde{p}_G(\xi|\lambda, \mathcal{D}) = \mathcal{N}_H(\xi^*(\lambda), \Sigma^*(\lambda))$. Taking into account the constraint on the last B-spline coefficient $\theta_K = c$, we recover the following conditional posterior distribution for the vector θ^c ,

$$\tilde{p}_G(\theta^c|\lambda, \mathcal{D}) = \mathcal{N}_{K-1}(\mu_{\theta^c}(\lambda), \Sigma_{\theta^c}(\lambda)), \quad (12)$$

where $\mu_{\theta^c}(\lambda) = (\xi_{c,1}^*(\lambda), \dots, \xi_{c,K-1}^*(\lambda))$ and $\Sigma_{\theta^c}(\lambda)$ is a $K-1$ dimensional square matrix corresponding to the first $K-1$ rows and columns of $\Sigma_c^*(\lambda)$.

Using similar techniques as in Section 2.5, we can show that the approximated joint posterior distribution for the spline coefficients is

$$\hat{p}(\theta^c|\mathcal{D}) = \sum_m \omega_m \mathcal{N}_{K-1}(\mu_{\theta^c}(\lambda^{(m)}), \Sigma_{\theta^c}(\lambda^{(m)})). \quad (13)$$

We thus recover in (13) a multivariate Gaussian mixture for which the mean and covariance matrix are analytically known (see Section 2.5). Let us denote by $\theta_{c,0}^m = \mu_{\theta^c}(\lambda^{(m)})$ the mean of mixture component m . Using a first-order Taylor expansion of $G_0(\cdot|t)$ around $\theta_{c,0}^m$, one gets

$$G_{0,m}(\theta^c|t) \approx G_0(\theta_{c,0}^m|t) + (\theta^c - \theta_{c,0}^m)^T \nabla_{\theta^c} G_0(\theta_{c,0}^m|t), \quad (14)$$

where $\nabla_{\theta^c} G_0(\cdot|t)$ denotes the gradient of G_0 with respect to θ^c . Combining (13) with (14) suggests to approximate the marginal posterior of $G_{0,m}(\theta^c|t)$ by the following univariate normal distribution $(G_{0,m}(\theta^c|t)|\mathcal{D}) \sim \mathcal{N}_1(G_0(\theta_{c,0}^m|t), \nabla_{\theta^c} G_0(\theta_{c,0}^m|t)^T \Sigma_{\theta^c}(\lambda^{(m)}) \nabla_{\theta^c} G_0(\theta_{c,0}^m|t))$. Accordingly, the posterior density of $G_0(\theta^c|t)$ across all mixture components **can in turn be approximated by the mixture** of Gaussian densities $(G_0(\theta^c|t)|\mathcal{D}) \approx \sum_m \omega_m \mathcal{N}_1(G_0(\theta_{c,0}^m|t), \nabla_{\theta^c} G_0(\theta_{c,0}^m|t)^T \Sigma_{\theta^c}(\lambda^{(m)}) \nabla_{\theta^c} G_0(\theta_{c,0}^m|t))$. A $(1 - \alpha) \times 100\%$ credible interval can be obtained numerically by finding C such that

$$\int_C p(G_0(\theta^c|t)|\mathcal{D}) dG_0(\theta^c|t) = 1 - \alpha. \quad (15)$$

To construct credible intervals for the population survival function given in (1) with a given profile of covariates, the procedure is the same except that the function of interest has the following form after a $\log(-\log(\cdot))$ transform

$$G_0(\xi_c|\mathbf{x}, \mathbf{z}, t) = \beta_0 + \mathbf{x}^T \boldsymbol{\beta} + \log \left(1 - \exp \left(- \sum_{j=1}^{j(t)} \exp(\theta^T b(s_j)) \Delta_j \right)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} \right),$$

where $\xi_c = (\theta_1, \dots, \theta_{K-1}, \beta_0, \dots, \beta_p, \gamma_1, \dots, \gamma_l)$. Using a first-order Taylor expansion of G_0 about the mean of each mixture component $\xi_{c,0}^m = \xi_c^*(\lambda^{(m)})$, it can be shown using the same arguments as before that

$$(G_0(\xi_c|\mathbf{x}, \mathbf{z}, t)|\mathcal{D}) \approx \sum_m \omega_m \mathcal{N}_1(G_0(\xi_{c,0}^m|\mathbf{x}, \mathbf{z}, t), \nabla_{\xi_c} G_0(\xi_{c,0}^m|\mathbf{x}, \mathbf{z}, t)^T \Sigma_c^*(\lambda^{(m)}) \nabla_{\xi_c} G_0(\xi_{c,0}^m|\mathbf{x}, \mathbf{z}, t)). \quad (16)$$

2.7. Cure prediction

Another quantity of interest in the promotion time cure model is the probability that a subject is cured given that (s)he has survived until a certain point in time t , say. Mathematically, one has

$$P(T = +\infty | T \geq t, \mathbf{x}, \mathbf{z}) = \exp \left(- \exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta}) S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} \right). \quad (17)$$

Again, taking a $\log(-\log(\cdot))$ transform and using a first-order Taylor expansion, it can be shown that the resulting approximation to the posterior distribution is analogous to (16), the only difference being in the gradient. All the gradients required to compute the credible intervals given above have been computed with analytic forms provided in Appendix C.

3. Simulation study

The aim of this section is to implement a simulation study to assess the statistical performance of the Laplace-P-spline approach (cf. Section 2) in the promotion time cure model. The simulation setting is exactly the same as in Bremhorst and Lambert (2016) when the follow-up is sufficiently long except that we choose different cure and censoring rates, as well as more B-splines in the basis as enabled by the numerical efficiency of our method. **Our methodology can also be applied when the follow-up period is not sufficiently long, provided that we account for identifiability issues. Indeed, as suggested in Bremhorst and Lambert (2016), when the follow-up of any susceptible subject is not long enough to observe its failure, then covariate effects are identifiable under the condition that the covariates are not simultaneously present in the probability to be cured and in the proportional hazards model parts.**

The regressors consist in normal variates $x_{i1} = z_{i1} \sim \mathcal{N}(0, 1)$, $i = 1, \dots, n$ and discrete covariates following a Bernoulli distribution $x_{i2} = z_{i2} \sim \text{Bern}(0.5)$, $i = 1, \dots, n$ to which we subtract 0.5 to obtain mean-centered covariates. The baseline distribution to generate latent event times is chosen to be a Weibull with mean 8 and variance 17.47. The regression coefficients in the Cox PH model are set to $\gamma_1 = 0.40$ and $\gamma_2 = -0.40$, while the coefficients β_0, β_1 and β_2 are calibrated in order to get two different percentages for the proportion of cured subjects, namely around 20% and 30%. Finally, censoring is either governed by a uniform distribution on $[20, 25]$ or by a Weibull with shape parameter 3 and scale parameter 25. We redirect the reader to [Bremhorst and Lambert \(2016\)](#), Section 5.1 for more details concerning the generation of latent event times and censoring times. We use the Laplace-P-spline model with the above covariates and 25 cubic B-splines in $[0, t_u]$ where the upper bound of the follow-up is fixed to $t_u = 25$. A third order penalty on the coefficients of adjacent B-splines is used to counterbalance their flexibility. Furthermore, the last B-spline coefficient is fixed to $\theta_K = 10$ to translate the “sufficiently long follow-up hypothesis” in cure models, thereby avoiding identifiability problems. The simulations were performed on $S = 500$ replicates of sample size $n = 300$ and $n = 600$ with results reported in [Tables 1](#) and [2](#). We report the integrated mean square error (IMSE), the bias, the empirical standard error (ESE) and the root mean square error (RMSE) of the posterior (mixture) mean taken as a pointwise estimator of the regression coefficients in the cure probability and survival parts. The coverage probabilities of the 90% and 95% credible intervals are also given. We observe a negligible bias across the different cure and censoring settings. In addition, the estimated coverage probabilities are reasonably close to the nominal values of 90% and 95% in each setting. We also notice that, as expected, the ESE and RMSE decrease with sample size.

Table 1. Simulation results for $S = 500$ replicates and sample size $n = 300$. Setting 1: Censoring times generated from a uniform $\mathcal{U}(20, 25)$ distribution; Setting 2: Censoring times generated from a $\text{Weibull}(3, 25)$ distribution. Cure fraction is in column 1, true parameter values in column 4.

Cure	Setting	IMSE	Parameters	Bias	CV _{90%}	CV _{95%}	ESE	RMSE
20%	1	0.023	$\beta_0 = 0.75$	0.022	90.4	95.4	0.101	0.103
			$\beta_1 = 0.80$	0.035	89.8	94.0	0.119	0.124
			$\beta_2 = -0.50$	-0.039	89.4	93.8	0.175	0.179
			$\gamma_1 = 0.40$	-0.056	89.0	93.6	0.146	0.156
			$\gamma_2 = -0.40$	0.050	89.0	93.2	0.218	0.223
	2	0.030	$\beta_0 = 0.75$	0.016	90.8	95.4	0.112	0.113
			$\beta_1 = 0.80$	0.045	91.8	96.0	0.137	0.144
			$\beta_2 = -0.50$	-0.036	93.0	97.2	0.200	0.203
			$\gamma_1 = 0.40$	-0.071	90.4	93.6	0.173	0.187
			$\gamma_2 = -0.40$	0.047	92.4	97.0	0.248	0.252
30%	1	0.017	$\beta_0 = 0.30$	0.010	89.4	94.6	0.092	0.092
			$\beta_1 = 1.00$	0.034	89.6	95.0	0.123	0.127
			$\beta_2 = -0.75$	-0.015	89.8	94.2	0.173	0.173
			$\gamma_1 = 0.40$	-0.057	88.6	94.6	0.143	0.154
			$\gamma_2 = -0.40$	0.042	88.8	95.8	0.210	0.214
	2	0.025	$\beta_0 = 0.30$	-0.001	91.2	94.8	0.103	0.103
			$\beta_1 = 1.00$	0.047	91.2	95.6	0.136	0.143
			$\beta_2 = -0.75$	-0.032	91.2	96.8	0.194	0.197
			$\gamma_1 = 0.40$	-0.072	89.0	94.0	0.175	0.189
			$\gamma_2 = -0.40$	0.038	92.6	95.8	0.242	0.245

Table 2. Simulation results for $S = 500$ replicates and sample size $n = 600$. Setting 1: Censoring times generated from a uniform $\mathcal{U}(20,25)$ distribution; Setting 2: Censoring times generated from a $Weibull(3,25)$ distribution. Cure fraction is in column 1, true parameter values in column 4.

Cure	Setting	IMSE	Parameters	Bias	CV _{90%}	CV _{95%}	ESE	RMSE
20%	1	0.010	$\beta_0 = 0.75$	0.016	88.4	94.2	0.069	0.071
			$\beta_1 = 0.80$	0.029	91.0	95.4	0.076	0.081
			$\beta_2 = -0.50$	-0.016	91.0	94.2	0.119	0.119
			$\gamma_1 = 0.40$	-0.054	87.4	93.2	0.099	0.112
			$\gamma_2 = -0.40$	0.039	92.0	95.8	0.143	0.148
	2	0.014	$\beta_0 = 0.75$	0.009	91.6	96.6	0.074	0.074
			$\beta_1 = 0.80$	0.033	89.2	94.8	0.099	0.104
			$\beta_2 = -0.50$	-0.020	89.6	95.6	0.140	0.141
			$\gamma_1 = 0.40$	-0.054	88.0	94.4	0.120	0.131
			$\gamma_2 = -0.40$	0.037	89.6	95.6	0.173	0.177
30%	1	0.010	$\beta_0 = 0.30$	0.002	90.0	95.0	0.064	0.064
			$\beta_1 = 1.00$	0.021	90.6	94.0	0.087	0.089
			$\beta_2 = -0.75$	-0.013	90.0	94.6	0.123	0.123
			$\gamma_1 = 0.40$	-0.037	88.6	93.8	0.104	0.110
			$\gamma_2 = -0.40$	0.028	90.2	95.4	0.147	0.149
	2	0.015	$\beta_0 = 0.30$	0.001	90.6	94.8	0.074	0.074
			$\beta_1 = 1.00$	0.030	90.0	95.2	0.099	0.104
			$\beta_2 = -0.75$	-0.014	90.8	95.8	0.140	0.140
			$\gamma_1 = 0.40$	-0.055	85.8	92.6	0.125	0.137
			$\gamma_2 = -0.40$	0.030	89.4	94.0	0.179	0.181

Coverage estimates of 90% credible intervals for the baseline survival function are reported in Table 3 with an asterisk as superscript when the estimated coverage is incompatible with the nominal value at the 95% level. Globally, the estimated coverage probability across all the quantiles is close to the nominal value of 90% in the different settings. Furthermore, the poor coverage that is observed in the 5% quantile when $n = 300$ improves with growing sample size.

Table 3. Coverage estimates of 90% credible intervals using first-order Taylor approximations for the baseline survival function at selected quantiles (5%, 15%, 35%, 50%, 65%, 75%, 85%, 95%) of T under the promotion time cure model. Setting 1: Censoring times generated from a uniform $\mathcal{U}(20,25)$ distribution; Setting 2: Censoring times generated from a $Weibull(3,25)$ distribution.

n=300	Cure	Cens.	Setting	5%	15%	35%	50%	65%	75%	85%	95%
	20%	20%	1	90.6	91.6	89.8	88.2	89.0	88.4	88.6	88.0
	20%	23%	2	87.8	91.0	88.4	89.0	89.4	91.2	90.4	92.0
	30%	30%	1	87.8	89.0	91.0	91.4	90.6	90.0	88.2	87.2*
	30%	33%	2	82.6*	88.6	88.4	89.0	88.6	89.4	88.0	90.8
n=600	Cure	Cens.	Setting	5%	15%	35%	50%	65%	75%	85%	95%
	20%	20%	1	90.4	92.8*	88.4	90.4	91.2	91.4	88.8	88.4
	20%	23%	2	87.6	89.0	88.8	89.8	87.0*	87.8	85.2*	88.0
	30%	30%	1	91.8	92.6	90.0	89.8	91.0	91.8	90.2	90.4
	30%	33%	2	86.6*	91.0	88.8	87.2*	86.4*	87.6	88.2	88.6

In Fig.1, the solid line is the target baseline survival distribution for the susceptible corresponding to the Weibull with mean 8 and standard deviation 4.18. The gray curves are estimates of $S_0(t)$ under each replicate and the dashed curve corresponds to the pointwise median of the 500 estimated baseline survival functions. Globally, we can say that the Laplace-P-spline approach provides accurate estimates of the baseline distribution with little variability around the target.

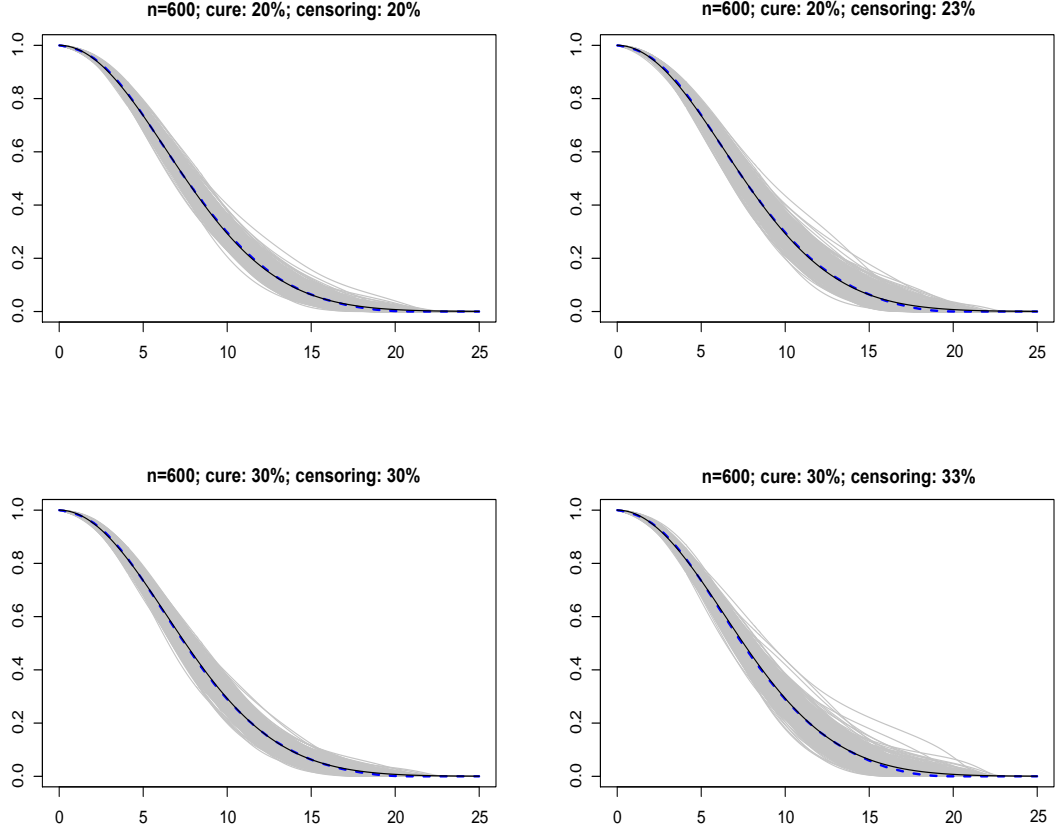


Fig. 1. Estimation of the baseline distribution $S_0(t)$ for $S = 500$ replications, (one gray curve per dataset) and sample size $n = 600$. In the left column the censoring rate is governed by a $\mathcal{U}(20, 25)$ distribution and in the right column it is governed by a $Weibull(3, 25)$ distribution. The solid line is the true function and the dashed line is the pointwise median of the 500 estimated curves.

In Table 4, we report the coverage estimates of 90% credible intervals for the population survival function at selected quantiles with the continuous covariate fixed to 0.1 and the binary covariate to 0.5. Again, the constructed credible intervals show good performances even for the 5% and 95% quantiles.

Table 4. Coverage estimates of 90% credible intervals using first-order Taylor approximations for the population survival function at quantiles (5%, 15%, 35%, 50%, 65%, 75%, 85%, 95%) of T under the promotion time cure model when $x = 0.1$ and $z = 0.5$. Setting 1: Censoring times generated from a uniform $\mathcal{U}(20, 25)$ distribution; Setting 2: Censoring times from a $Weibull(3, 25)$ distribution.

n=300	Cure	Cens.	Setting	5%	15%	35%	50%	65%	75%	85%	95%
	20%	20%	1	89.8	88.4	90.2	91.8	91.8	89.8	88.6	89.6
	20%	23%	2	87.6	89.2	90.0	90.8	91.2	90.8	90.6	92.8*
	30%	30%	1	90.8	90.2	89.0	89.4	90.8	92.2	91.0	90.8
	30%	33%	2	90.2	89.4	89.6	89.6	91.0	90.6	90.2	91.4

For the sake of assessing the algorithmic performance of our approach, we implement a computational speed comparison with a MCMC algorithm. More precisely, the competitor is taken to be a Metropolis-within-Gibbs algorithm with blockwise sampling for which we compute a chain of length 23000 and a burnin of length 3000 to explore the joint posterior of latent field variables. Under the same simulation settings, we observe a computational speed-up of a factor 15 with the Laplace-P-spline approach. It is also worth noting that most of the computational intensive tasks in our MCMC algorithm are written in Fortran language and called via R, while our Laplace-P-spline algorithm is exclusively coded in R language, such that the mentioned computational gain is conservative and under-evaluated.

4. Real data analysis

4.1. Application to malignant melanoma data

In this section, we illustrate the Laplace-P-spline methodology with the analysis of a malignant melanoma survival dataset [Andersen et al. \(1993\)](#). The dataset concerns 205 patients affected by skin cancer and operated for malignant melanoma at Odense University Hospital in Denmark during 1962-1977. The response of interest is the time (in years) elapsed between operation and death from malignant melanoma. The covariates are age at operation (in years), gender (1=M, 0=F), tumor thickness (in mm) and a dichotomous factor indicating presence of ulceration (1=presence, 0=absence) at baseline. Among the 205 patients, 57 died from malignant melanoma while the remaining 148 are right censored. This dataset was first investigated using single-factor analysis techniques ([Drzewiecki et al. 1980b,a](#)) and a Cox regression model [Drzewiecki and Andersen \(1982\)](#). More recently, [Li and Lin \(2009\)](#) used the melanoma dataset to illustrate a semi-parametric mixture model, while [Chyong-Mei and Chen-Hsin \(2016\)](#) implemented it to highlight a heteroscedastic transformation cure model. We propose to use the promotion time approach in which the covariates will simultaneously affect the probability of being cured as well as the time to event for susceptible subjects. The use of the same covariates in the two parts of the model is not problematic when it comes to inference as a plateau is observed in the Kaplan-Meier curve, suggesting that the follow-up is sufficiently long.

We use 50 B-splines on $[0, t_u]$ and follow a common choice in the literature to specify t_u as the largest observed survival time (here $t_u = 15.236$). The algorithm translated in pure R code takes approximately 15 seconds to obtain estimates for all B-spline coefficients, the standard deviation (sd_{post}) and 95% quantile-based credible intervals for the regression coefficients. The first estimation results (not detailed here) suggest that ulceration has a significant effect on the probability to be cured, while tumor thickness significantly affects the time to event for susceptible subjects. The model is then estimated a second time (see [Table 5](#)) by omitting Age and Gender as they have no significant effect in the model (conditionally on treatment and ulceration). The results suggest that ulceration has a negative effect on the probability to be cured. Furthermore, tumor thickness at time of surgery is an important factor affecting the time necessary to detect a new tumor. In fact, a large tumor at baseline may already be a sign of metastatic occurrence such that after an incomplete removal of cancer cells, a relapse is more likely to occur in a shorter period of time.

Our analysis also investigates to what extent ulceration affects the probability that a patient is cured given that (s)he has survived until a given time reference t . This conditional probability is estimated in [Table 6](#) for a median value of tumor thickness (1.94 mm) and approximate 90% credible intervals are also provided. [Fig. 2](#) gives a graphical representation of the pointwise and set estimates for these probabilities. We see that in presence of an ulcer, the estimated probability that a patient is cured given that (s)he has survived until t is smaller than the estimate corresponding to ulcer absence, regardless of the reference time

values. In addition, we see that the estimated probabilities increase with t , simply corroborating the idea that the longer a patient has survived (with or without an ulcer), the larger his/her chances of being cured.

Table 5. Posterior mixture mean for each regression parameter using 50 B-splines for the baseline log-hazard in the reduced model, the 95% quantile-based approximate credible intervals (CI) and the posterior standard deviation. $\phi(\mathbf{x})$ is minus the log of the probability to be cured and $1 - S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})}$ represents the time necessary for a cell to produce a detectable tumor mass.

	Parameters	Estimates	CI 95%	sd _{post}
$\phi(\mathbf{x})$	Intercept	-1.589	[-2.226; -0.948]	0.326
	Thickness	0.067	[-0.010; 0.142]	0.039
	Ulcer	1.096	[0.370; 1.819]	0.370
$1 - S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})}$	Thickness	0.111	[0.017; 0.201]	0.047
	Ulcer	0.327	[-0.619; 1.278]	0.484

Table 6. Pointwise estimates and approximate 90% credible intervals for the conditional probability to be cured given that $T \geq t$ for $t \in \{2, 4, 6, 8\}$ (in years) with and without ulceration and for a median value of tumor thickness.

Probability to be cured given that $T \geq t$				
t	No Ulceration		Ulceration	
	Estimates	CI 90%	Estimates	CI 90%
2	0.812	[0.697; 0.887]	0.538	[0.404; 0.676]
4	0.855	[0.735; 0.924]	0.631	[0.491; 0.799]
6	0.904	[0.773; 0.961]	0.745	[0.596; 0.912]
8	0.944	[0.793; 0.986]	0.849	[0.690; 0.974]

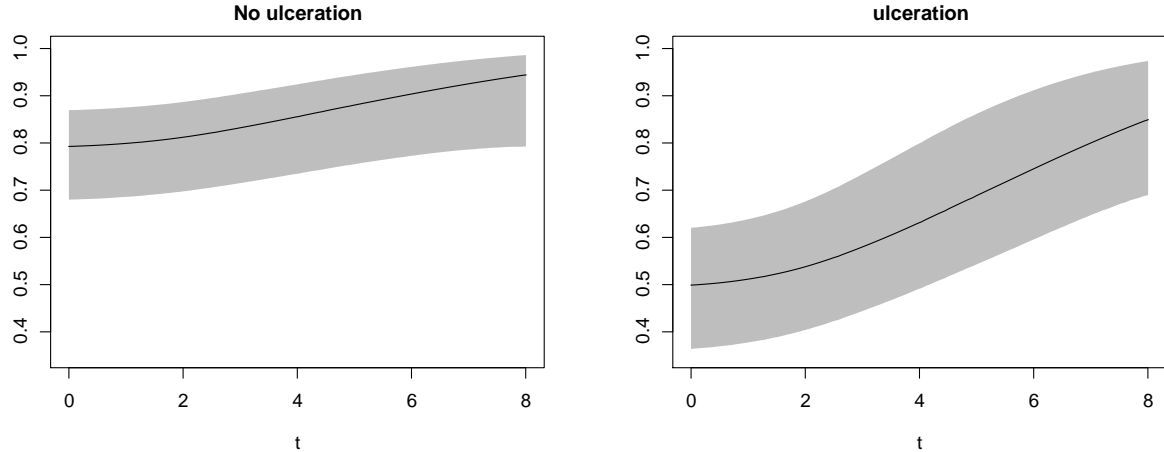


Fig. 2. Evolution over time t of the probability to be cured $P(T = +\infty | T \geq t, TT = 1.94)$ for a median tumor thickness (TT) represented by the solid line for two scenarios, no ulceration (left) and ulceration (right). The gray surface represents the approximate 90% pointwise credible intervals.

4.2. Application to oropharynx carcinoma data

We implement a second data analysis using data from [Kalbfleisch and Prentice \(2012\)](#) on oropharynx carcinoma. The dataset comes from a clinical trial achieved by the Radiation Therapy Oncology group involving patients from six clinics suffering from squamous cell carcinoma located in different sites of the mouth and throat. There are 195 patients randomly assigned in two arms at the moment of entry in the study: (1) radiation therapy alone (standard) or (2) radiation therapy with a chemotherapeutic agent

(special). To highlight the use of our model, we focus on 130 patients (among which 38 are censored) with cancer located in the pharyngeal tongue and tonsillar fossa part of the mouth. We retain the covariates *Age*, *Sex* (1=M, 0=F), *Treatment* (1=special or 0=standard) and tumor staging (*Tumor*) for explaining survival times. For tumor staging, we follow [Lopes and Bolfarine \(2012\)](#) and categorize the variable as *Tumor*=0 if primary tumor and *Tumor*=1 if massive tumor. As in the previous application, we use 50 B-splines in the interval ranging from 0 to the largest observed survival time measured in years (4.99).

The main objective of this analysis is to assess the effect of the two types of treatments on survival times of patients accounting for tumor staging. The estimated Kaplan-Meier curve given in [Fig. 3](#) (left panel) shows a plateau, indicating the presence of a cured fraction and thus justifying our choice to let the covariates influence jointly the probability to be cured and the time to event for susceptible patients. In [Table 7](#), we report the posterior mixture mean, the 90% quantile-based approximate credible interval and the posterior standard deviation. We see that *Tumor* is the only variable having a negative and significant effect on the probability to be cured, such that presence of a massive tumor decreases the chances of being cured from oropharynx cancer. In addition, *Treatment* has a significant impact on a recurrence timing, but not on the probability of its occurrence.

Table 7. Posterior mixture mean, 90% quantile-based approximate credible interval (CI) and posterior standard deviation for each regression parameter of the promotion time model.

	Parameters	Estimates	CI 90%	sd_{post}
$\phi(\mathbf{x})$	Intercept	-0.323	[-1.436; 0.788]	0.676
	Age	0.008	[-0.010; 0.025]	0.011
	Sex	0.291	[-0.148; 0.727]	0.266
	Tumor	0.510	[0.020; 0.998]	0.297
	Treatment	-0.315	[-0.733; 0.101]	0.253
$1 - S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})}$	Age	0.006	[-0.012; 0.022]	0.010
	Sex	-0.704	[-1.253; -0.156]	0.334
	Tumor	0.356	[-0.321; 1.031]	0.411
	Treatment	0.763	[0.231; 1.292]	0.323

In [Fig. 3](#) (right panel), we show the estimated population survival functions when the model is estimated without *Age* and *Sex* and by only accounting for the effects of *Tumor* and *Treatment* on, respectively, the cancer recurrence probability and on its timing for susceptible subjects. Whether we consider a standard or special treatment, we see that the risk of cancer recurrence only changes with tumor status, with a higher risk when a massive tumor is present as compared to a primary tumor. In addition, we see that the type of treatment mainly impacts the speed at which the recurrence arises for susceptible patients.

Finally, [Fig. 4](#) compares the estimated population survival functions obtained with the Laplace-P-spline model (blue curve) against Kaplan-Meier curves (in black) for each tumor staging and treatment configuration. In each situation, the Laplace-P-spline model provides survival curves that appear to be appropriate smoothed versions of the Kaplan-Meier estimates.

5. Discussion

In this paper, we introduced a novel methodology for fast Bayesian inference in semi-parametric cure survival models by coupling P-splines with Laplace approximations. Our approach opens up promising perspectives for inference in cure survival models as it enables to obtain pointwise and set estimators for non-trivial functions of latent variables with a drastic computational speed-up as compared to existing MCMC methods.

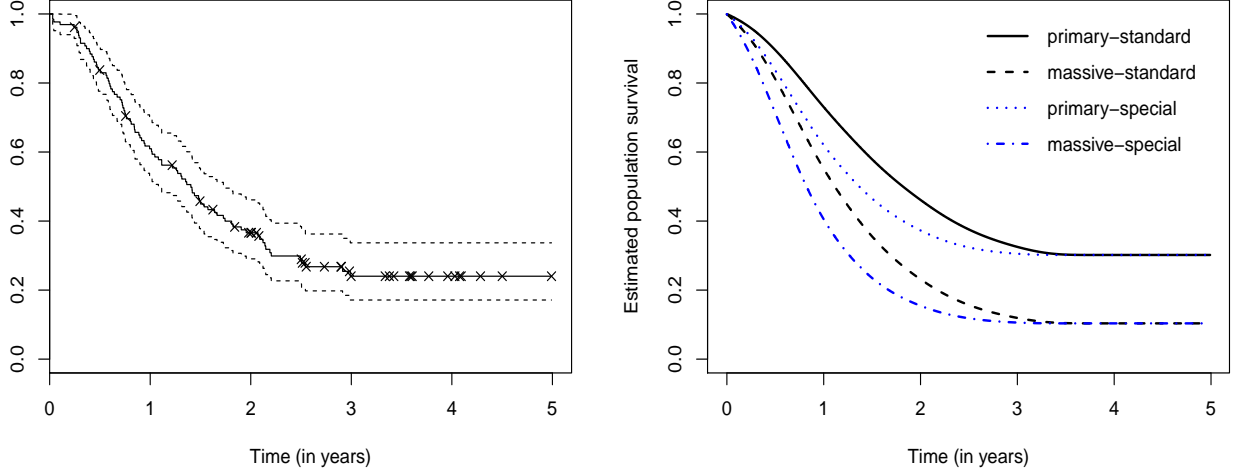


Fig. 3. (Left panel) Kaplan-Meier estimated curve from the oropharynx dataset. A cross indicates a censored patient. (Right panel) Estimated population survival functions for different tumor-treatment configurations.

Even though the Laplace approximation mechanisms presented in this work share some similarities with the classic INLA approach [Rue et al. \(2009\)](#), our methodology is sharply contrasted with the latter in many respects. In particular, our modeling strategy involves a specification of the prior of the roughness penalty parameter that is robust to the choice of hyperparameters [Jullion and Lambert \(2007\)](#). In the standard INLA approach, that concern is not addressed with the implication that posterior estimation can be sensitive towards the hyperparameter prior chosen by the user. In addition, our work goes beyond the treatment of univariate posterior marginal distributions by deriving reliable approximations to the joint posterior distributions of latent variables for which the mean and covariance matrix have known analytic forms.

Another major difference is that the dimension of our latent field only grows with the number of regressors and not with sample size, impacting directly the underlying algorithmic efficiency when dealing with large datasets. A practical limitation may arise when dealing with a hyperparameter vector of large dimension. This might be the case for instance in additive regression models, where the number of roughness penalty parameters is equal to the number of smooth functions to be estimated, implying a much larger computational cost for the grid strategy recommended in [Section 2.4](#). However, even for a large number of hyperparameters, we expect our approach to be much faster than existing MCMC techniques, which would **require long computation times** in such situations. A future research direction would be to explore efficient techniques to deal with models involving a large-dimensional hyperparameter vector. We also plan to extend our methodology to the class of mixture cure survival models, to additive models and to multi-dimensional smoothing.

Acknowledgments

The authors acknowledge the Actions de Recherche Concertées (ARC 11/16-039) grant and the Pôles d’attraction interuniversitaires (PAI P7/06) for funding the project. The authors also thank Vincent Bremhorst for his perceptive and valuable comments.

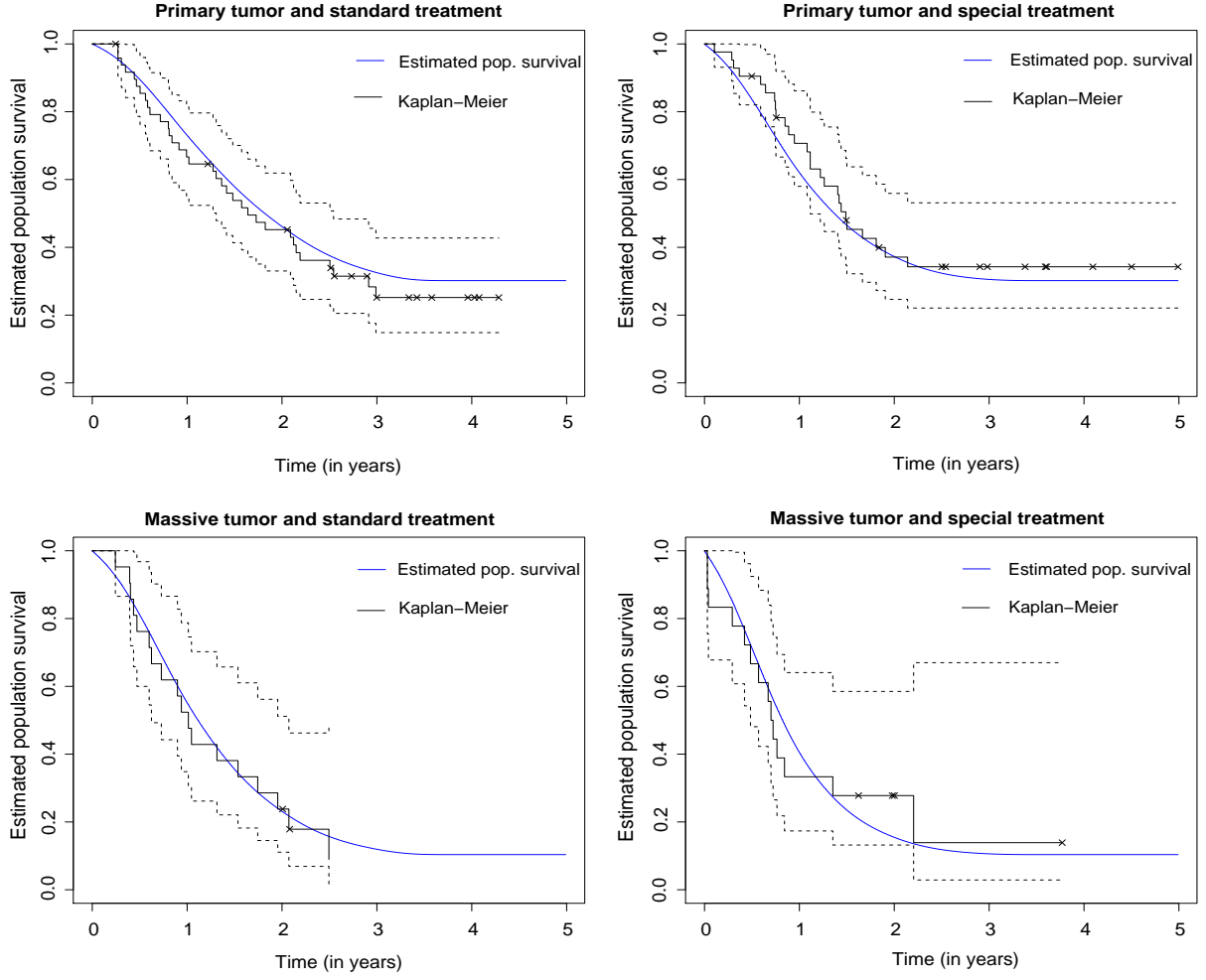


Fig. 4. Estimated population survival functions from the Laplace-P-spline model (blue) versus Kaplan-Meier curves (black) and their 95% confidence interval (dashed) for different tumor status and treatment.

Conflict of interest

The authors declare no conflict of interests.

Appendix A. Taylor expansion of $g_i(\xi)$ for Laplace approximation

The first step is to derive the Gaussian approximation of $p(\xi|\lambda, \mathcal{D})$ around its mode. The latter posterior density is written as

$$\begin{aligned}
 p(\xi|\lambda, \mathcal{D}) &\propto \mathcal{L}(\xi; \mathcal{D})p(\xi|\lambda) \\
 &\propto \exp\left(\ell(\xi; \mathcal{D}) - \frac{1}{2}\xi^T Q(\lambda)\xi + \xi^T Q(\lambda)\mu_\xi\right) \\
 &\propto \exp\left(\sum_{i=1}^n g_i(\xi) - \frac{1}{2}\xi^T Q(\lambda)\xi + \xi^T Q(\lambda)\mu_\xi\right).
 \end{aligned} \tag{18}$$

When the data are Gaussian, $p(\xi|\lambda, \mathcal{D})$ is also Gaussian but in practice we may not be restricted to Gaussian data and more complex distribution patterns can appear for the conditional posterior of the latent field. To make $p(\xi|\lambda, \mathcal{D})$ tractable, the idea is to use a second-order Taylor approximation of $g_i(\xi)$ around a point $\xi^{(0)} \in \mathbb{R}^H$ as follows

$$g_i(\xi) \approx g_i(\xi^{(0)}) + (\xi - \xi^{(0)})^T \nabla g_i(\xi)|_{\xi=\xi^{(0)}} + \frac{1}{2}(\xi - \xi^{(0)})^T \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} (\xi - \xi^{(0)}), \quad (19)$$

where the gradient and Hessian of $g_i(\xi)$ are given by

$$\nabla g_i(\xi)|_{\xi=\xi^{(0)}} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_i(\xi) \\ \vdots \\ \frac{\partial}{\partial \theta_K} g_i(\xi) \\ \frac{\partial}{\partial \beta_0} g_i(\xi) \\ \vdots \\ \frac{\partial}{\partial \beta_p} g_i(\xi) \\ \frac{\partial}{\partial \gamma_1} g_i(\xi) \\ \vdots \\ \frac{\partial}{\partial \gamma_l} g_i(\xi) \end{bmatrix}_{\xi=\xi^{(0)}} \quad \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} = \begin{bmatrix} \underbrace{\frac{\partial^2}{\partial \theta \partial \theta^T} g_i(\xi)}_{K \times K} & \underbrace{\frac{\partial^2}{\partial \theta \partial \beta^T} g_i(\xi)}_{K \times (p+1)} & \underbrace{\frac{\partial^2}{\partial \theta \partial \gamma^T} g_i(\xi)}_{K \times l} \\ \underbrace{\frac{\partial^2}{\partial \beta \partial \theta^T} g_i(\xi)}_{(p+1) \times K} & \underbrace{\frac{\partial^2}{\partial \beta \partial \beta^T} g_i(\xi)}_{(p+1) \times (p+1)} & \underbrace{\frac{\partial^2}{\partial \beta \partial \gamma^T} g_i(\xi)}_{(p+1) \times l} \\ \underbrace{\frac{\partial^2}{\partial \gamma \partial \theta^T} g_i(\xi)}_{l \times K} & \underbrace{\frac{\partial^2}{\partial \gamma \partial \beta^T} g_i(\xi)}_{l \times (p+1)} & \underbrace{\frac{\partial^2}{\partial \gamma \partial \gamma^T} g_i(\xi)}_{l \times l} \end{bmatrix}_{\xi=\xi^{(0)}}.$$

5.1. Computation of the gradient

For the sake of avoiding heavy notation, we define the following scalar quantities

$$\sum_{j=1}^{j(t_i)} h_0(s_j) \Delta_j := \omega_{0i}; \quad \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta_j := \omega_{0i}^k; \quad \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) b_l(s_j) \Delta_j := \omega_{0i}^{kl}.$$

Deriving with respect to the B-spline coefficients gives us

$$\begin{aligned} \frac{\partial}{\partial \theta_k} g_i(\xi) &= \tau_i \left(b_k(t_i) - \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta_j \right) \\ &\quad + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \exp \left(- \sum_{j=1}^{j(t_i)} h_0(s_j) \Delta_j \right)^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma}) - 1} \\ &\quad \times \exp \left(- \sum_{j=1}^{j(t_i)} h_0(s_j) \Delta_j \right) \left(- \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta_j \right) \\ &= \tau_i \left(b_k(t_i) - \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta_j \right) \\ &\quad - \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp \left(- \sum_{j=1}^{j(t_i)} h_0(s_j) \Delta_j \right)^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \left(\sum_{j=1}^{j(t_i)} h_0(s_j) b_k(s_j) \Delta_j \right), \end{aligned}$$

so finally we have

$$\frac{\partial}{\partial \theta_k} g_i(\xi) = \tau_i \left(b_k(t_i) - \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \omega_{0i}^k \right) - \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \omega_{0i}^k, \quad k = 1, \dots, K.$$

The derivatives with respect to the $\boldsymbol{\beta}$ coefficients are

$$\frac{\partial}{\partial \beta_m} g_i(\xi) = \tau_i x_{im} - \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \left(1 - \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \right) x_{im}, \quad m = 0, \dots, p \quad \text{with } x_{i0} = 1.$$

To obtain the derivatives with respect to the $\boldsymbol{\gamma}$ coefficients, we will use the rule

$$\frac{d}{dx} a^{u(x)} = a^{u(x)} \log(a) \frac{d}{dx} u(x), \quad a > 0.$$

$$\frac{\partial}{\partial \gamma_s} g_i(\xi) = \tau_i \left(z_{is} - \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is} \omega_{0i} \right) + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} (-\omega_{0i}) \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is}$$

and more compactly

$$\frac{\partial}{\partial \gamma_s} g_i(\xi) = \tau_i z_{is} (1 - \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \omega_{0i}) - \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \omega_{0i} z_{is}, \quad s = 1, \dots, l.$$

5.2. Computation of the Hessian

To compute the Hessian, we require the block matrices given below. Blocks 21, 31 and 32 are obtained by transposing blocks 12, 13 and 23.

$$\text{Block11} : \frac{\partial^2}{\partial \theta_k \partial \theta_l} g_i(\xi) \quad k = 1, \dots, K \quad l = 1, \dots, K.$$

$$\text{Block12} : \frac{\partial^2}{\partial \theta_k \partial \beta_m} g_i(\xi) \quad k = 1, \dots, K \quad m = 0, \dots, p.$$

$$\text{Block13} : \frac{\partial^2}{\partial \theta_k \partial \gamma_s} g_i(\xi) \quad k = 1, \dots, K \quad s = 1, \dots, l.$$

$$\text{Block22} : \frac{\partial^2}{\partial \beta_m \partial \beta_l} g_i(\xi) \quad l = 0, \dots, p \quad m = 0, \dots, p.$$

$$\text{Block23} : \frac{\partial^2}{\partial \beta_m \partial \gamma_s} g_i(\xi) \quad m = 0, \dots, p \quad s = 1, \dots, l.$$

$$\text{Block33} : \frac{\partial^2}{\partial \gamma_s \partial \gamma_v} g_i(\xi) \quad s = 1, \dots, l \quad v = 1, \dots, l.$$

Block 11

$$\begin{aligned} \frac{\partial^2}{\partial \theta_k \partial \theta_l} g_i(\xi) &= \tau_i \left(-\exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \omega_{0i}^{kl} \right) - \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \\ &\quad \times \left(\exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})-1} \exp(-\omega_{0i}) (-\omega_{0i}^l) \omega_{0i}^k + \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \omega_{0i}^{kl} \right) \end{aligned}$$

and more compactly for $k = 1, \dots, K$ and $l = 1, \dots, K$ we have

$$\frac{\partial^2}{\partial \theta_k \partial \theta_l} g_i(\xi) = -\tau_i \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \omega_{0i}^{kl} + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} (\exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \omega_{0i}^l \omega_{0i}^k - \omega_{0i}^{kl}).$$

Block 12

$$\frac{\partial^2}{\partial \theta_k \partial \beta_m} g_i(\xi) = -\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \omega_{0i}^k x_{im}, \quad k = 1, \dots, K \quad m = 0, \dots, p, \quad x_{i0} = 1.$$

Block 13

$$\begin{aligned} \frac{\partial^2}{\partial \theta_k \partial \gamma_s} g_i(\xi) &= \tau_i \left(-\exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is} \omega_{0i}^k \right) - \omega_{0i}^k \left(\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) z_{is} \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \right. \\ &\quad \left. + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} (-\omega_{0i}) \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is} \right) \end{aligned}$$

and more compactly for $k = 1, \dots, K$ and $s = 1, \dots, l$ we have

$$\frac{\partial^2}{\partial \theta_k \partial \gamma_s} g_i(\xi) = -\tau_i \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is} \omega_{0i}^k - \omega_{0i}^k \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} (z_{is} - \omega_{0i} \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is}).$$

Block 22

$$\frac{\partial^2}{\partial \beta_m \partial \beta_l} g_i(\xi) = -\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \left(1 - \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \right) x_{im} x_{il}, \quad m, l = 0, \dots, p \quad x_{i0} = 1.$$

Block 23

$$\frac{\partial^2}{\partial \beta_m \partial \gamma_s} g_i(\xi) = \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} (-\omega_{0i}) \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{is} x_{im}, \quad \text{in short}$$

$$\frac{\partial^2}{\partial \beta_m \partial \gamma_s} g_i(\xi) = -\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \omega_{0i} z_{is} x_{im}, \quad m = 0, \dots, p \quad s = 1, \dots, l \quad x_{i0} = 1.$$

Block 33

$$\begin{aligned} \frac{\partial^2}{\partial \gamma_s \partial \gamma_v} g_i(\xi) &= -\tau_i z_{is} \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{iv} \omega_{0i} - \omega_{0i} z_{is} \left(\exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) z_{iv} \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \right. \\ &\quad \left. + \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} (-\omega_{0i}) \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{iv} \right) \end{aligned}$$

and more compactly

$$\begin{aligned} \frac{\partial^2}{\partial \gamma_s \partial \gamma_v} g_i(\xi) &= -\tau_i \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) \omega_{0i} z_{is} z_{iv} - \omega_{0i} z_{is} \exp(\beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \boldsymbol{\gamma}) \exp(-\omega_{0i})^{\exp(\mathbf{z}_i^T \boldsymbol{\gamma})} \\ &\quad \times (z_{iv} - \omega_{0i} \exp(\mathbf{z}_i^T \boldsymbol{\gamma}) z_{iv}), \quad s, v = 1, \dots, l. \end{aligned}$$

The Taylor expansion in (19) can be written extensively as follows

$$\begin{aligned} g_i(\xi) &\approx \left[g_i(\xi^{(0)}) + \frac{1}{2} \xi^{(0)T} \nabla^2 g_i(\xi) |_{\xi=\xi^{(0)}} \xi^{(0)} - \xi^{(0)T} \nabla g_i(\xi) |_{\xi=\xi^{(0)}} \right] + \xi^T \nabla g_i(\xi) |_{\xi=\xi^{(0)}} \\ &\quad + \frac{1}{2} \xi^T \nabla^2 g_i(\xi) |_{\xi=\xi^{(0)}} \xi - \xi^T \nabla^2 g_i(\xi) |_{\xi=\xi^{(0)}} \xi^{(0)} \\ &\approx \text{constant} + \xi^T (\nabla g_i(\xi) |_{\xi=\xi^{(0)}} - \nabla^2 g_i(\xi) |_{\xi=\xi^{(0)}} \xi^{(0)}) + \frac{1}{2} \xi^T \nabla^2 g_i(\xi) |_{\xi=\xi^{(0)}} \xi \end{aligned}$$

and defining the short notation $\sum_{i=1}^n \nabla g_i(\xi)|_{\xi=\xi^{(0)}} := \nabla g_{\xi^{(0)}}$ and $\sum_{i=1}^n \nabla^2 g_i(\xi)|_{\xi=\xi^{(0)}} := \nabla^2 g_{\xi^{(0)}}$, we obtain the following expression for the sum of the functions $g_i(\cdot)$ omitting the constant

$$\sum_{i=1}^n g_i(\xi) = \xi^T (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)}) + \frac{1}{2} \xi^T \nabla^2 g_{\xi^{(0)}} \xi. \quad (20)$$

Introducing (20) into (18), we recover expression (3)

$$\tilde{p}_G(\xi|\lambda, \mathcal{D}) \propto \exp\left(-\frac{1}{2} \xi^T (Q(\lambda) - \nabla^2 g_{\xi^{(0)}}) \xi + \xi^T (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)} + Q(\lambda) \mu_\xi)\right). \quad (21)$$

The above expression is a Gaussian density (up to a multiplicative constant) with mean and covariance matrix that can be derived as follows. First take the logarithm of (21)

$$\log \tilde{p}_G(\xi|\lambda, \mathcal{D}) \doteq -\frac{1}{2} \xi^T (Q(\lambda) - \nabla^2 g_{\xi^{(0)}}) \xi + \xi^T (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)} + Q(\lambda) \mu_\xi), \quad (22)$$

where the symbol \doteq denotes equality up to an additive constant. To obtain the mean, we solve

$$\begin{aligned} \nabla_\xi \log \tilde{p}_G(\xi|\lambda, \mathcal{D}) &= 0 \\ \Leftrightarrow -(Q(\lambda) - \nabla^2 g_{\xi^{(0)}}) \xi + (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)} + Q(\lambda) \mu_\xi) &= 0, \end{aligned}$$

so the mean is

$$\xi^{(1)} = (Q(\lambda) - \nabla^2 g_{\xi^{(0)}})^{-1} (\nabla g_{\xi^{(0)}} - \nabla^2 g_{\xi^{(0)}} \xi^{(0)} + Q(\lambda) \mu_\xi).$$

The precision is obtained as the negative of the Hessian matrix

$$Q(\lambda)^{(1)} = -\nabla_\xi^2 \log \tilde{p}_G(\xi|\lambda, \mathcal{D}) = (Q(\lambda) - \nabla^2 g_{\xi^{(0)}}).$$

Appendix B. Conditional mean

The vector $\xi_c^*(\lambda) \in \mathbb{R}^{H-1}$ is the conditional posterior mean of the Gaussian approximation for a given $\xi_K = c$ and should not be confused with $\xi_{cc}^*(\lambda)$. To obtain $\xi_c^*(\lambda)$, we compute the Gaussian approximation around the posterior mode of $p(\xi|\lambda, \mathcal{D})$ as described in Section 2.3 and find a multivariate (H -dimensional) Gaussian distribution with mean $\xi^*(\lambda)$ and covariance matrix $\Sigma^*(\lambda)$. Then, using classic properties of the Normal density, we derive the distribution of $\xi_{-K} = (\xi_1, \dots, \xi_{K-1}, \xi_{K+1}, \dots, \xi_H) \in \mathbb{R}^{H-1}$ given the constraint $\xi_K = c$. The resulting distribution is a normal with mean vector $\xi_c^*(\lambda) = \xi_{-K}^*(\lambda) + \tilde{\Sigma}_{2,1}(\lambda) \tilde{\Sigma}_{1,1}^{-1}(\lambda) (c - \xi_K^*(\lambda))$ and covariance matrix $\Sigma_c^*(\lambda) = \tilde{\Sigma}_{2,2}(\lambda) - \tilde{\Sigma}_{2,1}(\lambda) \tilde{\Sigma}_{1,1}^{-1}(\lambda) \tilde{\Sigma}_{1,2}(\lambda)$, where $\tilde{\Sigma}_{1,1}(\lambda) = \Sigma_{K,K}^*(\lambda)$, $\tilde{\Sigma}_{1,2}(\lambda) = (\tilde{\Sigma}_{2,1}(\lambda))^T = (\Sigma_{K,1}^*(\lambda), \dots, \Sigma_{K,K-1}^*(\lambda), \Sigma_{K,K+1}^*(\lambda), \dots, \Sigma_{K,H}^*(\lambda))$ and $\tilde{\Sigma}_{2,2}(\lambda)$ is the matrix $\Sigma^*(\lambda)$ without row and column K . The vector $\xi_{cc}^*(\lambda) \in \mathbb{R}^H$ corresponds to $\xi_c^*(\lambda)$ to which we add $\xi_K = c$ at position K , i.e. $\xi_{cc}^*(\lambda) = (\xi_{c,1}^*(\lambda), \dots, \xi_{c,K-1}^*(\lambda), c, \xi_{c,K}^*(\lambda), \dots, \xi_{c,H-1}^*(\lambda))$, where $\xi_{c,i}^*(\lambda)$ denotes the i^{th} entry of $\xi_c^*(\lambda)$.

Appendix C. Gradients to compute credible intervals

Gradient associated to the baseline survival function

$$\nabla_{\theta_c} G_0(\theta_{c,0}^m | t) = \begin{bmatrix} \left(\sum_{j=1}^{j(t)} h_0(s_j) \Delta_j \right)^{-1} \sum_{j=1}^{j(t)} \exp\{ \sum_{k=1}^K \theta_k b_k(s_j) \} b_1(s_j) \Delta_j \\ \vdots \\ \left(\sum_{j=1}^{j(t)} h_0(s_j) \Delta_j \right)^{-1} \sum_{j=1}^{j(t)} \exp\{ \sum_{k=1}^K \theta_k b_k(s_j) \} b_{K-1}(s_j) \Delta_j \end{bmatrix}_{\theta_c = \theta_{c,0}^m}.$$

Gradient associated to the population survival function

$$\nabla_{\xi_c} G_0(\xi_{c,0}^m | \mathbf{x}, \mathbf{z}, t) = \begin{bmatrix} v(\boldsymbol{\theta}, \boldsymbol{\gamma})^{-1} \exp(\mathbf{z}^T \boldsymbol{\gamma}) S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} \left(\sum_{j=1}^{j(t)} h_0(s_j) b_1(s_j) \Delta_j \right) \\ \vdots \\ v(\boldsymbol{\theta}, \boldsymbol{\gamma})^{-1} \exp(\mathbf{z}^T \boldsymbol{\gamma}) S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} \left(\sum_{j=1}^{j(t)} h_0(s_j) b_{K-1}(s_j) \Delta_j \right) \\ 1 \\ x_1 \\ \vdots \\ x_p \\ v(\boldsymbol{\theta}, \boldsymbol{\gamma})^{-1} \exp(\mathbf{z}^T \boldsymbol{\gamma}) S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} \left(\sum_{j=1}^{j(t)} h_0(s_j) \Delta_j \right) z_1 \\ \vdots \\ v(\boldsymbol{\theta}, \boldsymbol{\gamma})^{-1} \exp(\mathbf{z}^T \boldsymbol{\gamma}) S_0(t)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})} \left(\sum_{j=1}^{j(t)} h_0(s_j) \Delta_j \right) z_l \end{bmatrix}_{\xi_c = \xi_{c,0}^m},$$

with $v(\boldsymbol{\theta}, \boldsymbol{\gamma}) = 1 - \exp\left(-\sum_{j=1}^{j(t)} \exp\left(\sum_{k=1}^K \theta_k b_k(s_j)\right) \Delta_j\right)^{\exp(\mathbf{z}^T \boldsymbol{\gamma})}$.

Gradient associated to the conditional probability $P(T = +\infty | T \geq t, \mathbf{x}, \mathbf{z})$

$$\nabla_{\xi_c} G_0(\xi_{c,0}^m | \mathbf{x}, \mathbf{z}, t) = \begin{bmatrix} -\exp(\mathbf{z}^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t)} h_0(s_j) b_1(s_j) \Delta_j \\ \vdots \\ -\exp(\mathbf{z}^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t)} h_0(s_j) b_{K-1}(s_j) \Delta_j \\ 1 \\ x_1 \\ \vdots \\ x_p \\ -z_1 \exp(\mathbf{z}^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t)} h_0(s_j) \Delta_j \\ \vdots \\ -z_l \exp(\mathbf{z}^T \boldsymbol{\gamma}) \sum_{j=1}^{j(t)} h_0(s_j) \Delta_j \end{bmatrix}_{\xi_c = \xi_{c,0}^m}.$$

References

- Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer Series in Statistics. Springer New York, <https://dx.doi.org/10.1007/978-1-4612-4348-9>.
- Bremhorst, V., Lambert, P., 2016. Flexible estimation in cure survival models using Bayesian P-splines. Computational Statistics & Data Analysis 93, 270–284, <https://doi.org/10.1016/j.csda.2014.05.009>.
- Chen, M.-H., Ibrahim, J. G., Sinha, D., 1999. A new Bayesian model for survival data with a surviving fraction. Journal of the American Statistical Association 94, 909–919, <http://www.jstor.org/stable/2670006>.
- Chyong-Mei, C., Chen-Hsin, C., 2016. Heteroscedastic transformation cure regression models. Statistics in Medicine 35, 2359–2376, <http://dx.doi.org/10.1002/sim.6896>.
- Cox, D. R., 1972. Regression models and life tables. Journal of the Royal Statistical Society, Series B (Methodological) 34, 187–220. <http://www.jstor.org/stable/2985181>.
- Drzewiecki, K. T., Andersen, P., 1982. Survival with malignant melanoma: A regression analysis of prognostic factors. Cancer 49, 2414–2419, [http://dx.doi.org/10.1002/1097-0142\(19820601\)49:11<2414::AID-CNCR2820491132>3.0.CO;2-V](http://dx.doi.org/10.1002/1097-0142(19820601)49:11<2414::AID-CNCR2820491132>3.0.CO;2-V).
- Drzewiecki, K. T., Christensen, H. E., Ladefoged, C., Poulsen, H., 1980a. Clinical course of cutaneous malignant melanoma related to histopathological criteria of primary tumour. Scandinavian Journal of Plastic and Reconstructive Surgery 14, 229–234, <http://dx.doi.org/10.3109/02844318009106715>.
- Drzewiecki, K. T., Ladefoged, C., Christensen, H. E., 1980b. Biopsy and prognosis for cutaneous malignant melanomas in clinical stage I. Scandinavian Journal of Plastic and Reconstructive Surgery 14, 141–144, <http://dx.doi.org/10.3109/02844318009106699>.

- Eilers, P. H. C., Marx, B. D., 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11, 89–102, <http://www.jstor.org/stable/2246049>.
- Eilers, P. H. C., Marx, B. D., 2010. Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics* 2, 637–653, <http://dx.doi.org/10.1002/wics.125>.
- Fong, Y., Rue, H., Wakefield, J., 2010. Bayesian inference for generalized linear mixed models. *Biostatistics* 11 (3), 397–412, <https://doi.org/10.1093/biostatistics/kxp053>.
- Frühwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer New York, <https://dx.doi.org/10.1007/978-0-387-35768-3>.
- Golub, G. H., Van Loan, C. F., 2012. *Matrix computations*. Vol. 3. John Hopkins University Press.
- Householder, A. S., 1958. Unitary triangularization of a nonsymmetric matrix. *Journal of the ACM (JACM)* 5, 339–342, <https://doi.org/10.1145/320941.320947>.
- Ibrahim, J. G., Chen, M.-H., Sinha, D., 2001. Bayesian semiparametric models for survival data with a cure fraction. *Biometrics* 57, 383–388, <http://dx.doi.org/10.1111/j.0006-341X.2001.00383.x>.
- Jiang, H., Brown, P. E., Rue, H., Shimakura, S., 2014. Geostatistical survival models for environmental risk assessment with large retrospective cohorts. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 177, 679–695, <http://dx.doi.org/10.1111/rssa.12041>.
- Jullion, A., Lambert, P., 2007. Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics & Data Analysis* 51, 2542–2558, <https://doi.org/10.1016/j.csda.2006.09.027>.
- Kalbfleisch, J. D., Prentice, R. L., 2012. *The statistical analysis of failure time data*, second edition. John Wiley & Sons.
- Lang, S., Brezger, A., 2004. Bayesian P-splines. *Journal of Computational and Graphical Statistics* 13, 183–212, <http://www.jstor.org/stable/1391151>.
- Li, G., Lin, C., 2009. Analysis of two-sample censored data using a semiparametric mixture model. *Acta Mathematicae Applicatae Sinica, English Series* 25, 389–398, <https://doi.org/10.1007/s10255-008-8804-4>.
- Lopes, C. M. C., Bolfarine, H., 2012. Random effects in promotion time cure rate models. *Computational Statistics & Data Analysis* 56 (1), 75–87, <https://doi.org/10.1016/j.csda.2011.05.008>.
- Martino, S., 2007. Approximate Bayesian inference for latent Gaussian models. *Fakultet for Informasjonsteknologi, Matematikk og Elektroteknikk* <http://hdl.handle.net/11250/258135>.
- Martino, S., Aas, K., Lindqvist, O., Neef, L. R., Rue, H., 2011a. Estimating stochastic volatility models using Integrated Nested Laplace Approximations. *The European Journal of Finance* 17, 487–503, <http://dx.doi.org/10.1080/1351847X.2010.495475>.
- Martino, S., Akerkar, R., Rue, H., 2011b. Approximate Bayesian inference for survival models. *Scandinavian Journal of Statistics* 38, 514–528, <http://dx.doi.org/10.1111/j.1467-9469.2010.00715.x>.
- Rosenberg, P., 1995. Hazard function estimation using B-splines. *Biometrics* 51, 874–887, <http://www.jstor.org/stable/2532989>.
- Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society, Series B (Methodological)* 71, 319–392, <http://dx.doi.org/10.1111/j.1467-9868.2008.00700.x>.
- Ruiz-Cárdenas, R., Krainski, E., Rue, H., 2012. Direct fitting of dynamic models using Integrated Nested Laplace Approximations-INLA. *Computational Statistics & Data Analysis* 56, 1808–1828, <https://doi.org/10.1016/j.csda.2011.10.024>.
- Schrödle, B., Held, L., 2011. Spatio-temporal disease mapping using INLA. *Environmetrics* 22, 725–734, <http://dx.doi.org/10.1002/env.1065>.
- Sørbye, S. H., Rue, H., 2011. Simultaneous credible bands for latent gaussian models. *Scandinavian Journal of Statistics* 38, 712–725, <http://dx.doi.org/10.1111/j.1467-9469.2011.00741.x>.
- Tsodikov, A. D., 1998. A proportional hazards model taking account of long-term survivors. *Biometrics* 54, 1508–1516, <http://www.jstor.org/stable/2533675>.
- Tsodikov, A. D., 2002. Semi-parametric models of long-and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Statistics in Medicine* 21, 895–920, <http://dx.doi.org/10.1002/sim.1054>.
- Tsodikov, A. D., 2003. Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society: Series B (Methodologica)* 65, 759–774, <http://dx.doi.org/10.1111/1467-9868.00414>.
- Yakovlev, A. Y., Tsodikov, A. D., Asselain, B., 1996. *Stochastic models of tumor latency and their biostatistical applications*. Vol. 1 of *Mathematical Biology and Medicine*. World Scientific, Singapore.
- Yin, G., Ibrahim, J. G., 2005. Cure rate models: a unified approach. *Canadian Journal of Statistics* 33, 559–570, <http://dx.doi.org/10.1002/cjs.5550330407>.
- Zeng, D., Yin, G., Ibrahim, J. G., 2006. Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association* 101, 670–684, <http://dx.doi.org/10.1198/016214505000001122>.