5-1-2003

Modeling Correlated Time-Varying Covariate Effects In A Cox-Type Regression Model

Mourad Tighiouart

Utah State University, mourad@math.usu.edu

Modeling Correlated Time-Varying Covariate Effects
In A Cox-Type Regression Model

Mourad Tighiouart
Department of Mathematics and Statistics
Utah State University

In this paper, I extend the proposed model by McKeague and Tighiouart (2000) to handle time-varying correlated covariate effects for the analysis of survival data. I use the conditional predictive ordinates (CPO's) for model comparison and the methodology is illustrated by an application to nasopharynx cancer survival data. A reversible jump MCMC sampler to estimate the CPO's will be presented.

Key words: Correlated time-varying covariate effects, Right censoring; Reversible Jump MCMC; Pseudo-Bayes factors

Introduction

The proportional hazards model of Cox (1972) is considered to be the most popular approach to the analysis of time-to-event data. In the past three decades, many authors have proposed variants of this model to relax the somehow restrictive proportional hazards assumption and to analyze multivariate survival data, see Andersen et al. (1992) and Ibrahim, et al. (2001).

In this paper, I use the local characteristics of Gaussian Markov random fields to describe the prior information of the conditional hazard function for right-censored survival data. McKeague and Tighiouart (2000) modeled the conditional hazard function (given covariates z) $h(t|z)$ as a product of conditionally independent stochastic processes, corresponding to (1) a baseline hazard function $h_0(t)$, and (2) a regression function $exp(\boldsymbol{b}(t)'z)$ representing the effects of covariates:

Mourad Tighiouart is an Assistant Professor in the Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan UT 84322-3900. E:mail: mourad@math.usu.edu.

$$h_0(t) = \sum_{i \geq 1} I(\boldsymbol{t}_i < t \leq \boldsymbol{t}_{i+1}) \, h_i \qquad (1)$$

$$h(t|z) = \sum_{i \geq 1} I(\boldsymbol{t}_i < t \leq \boldsymbol{t}_{i+1}) \, h_i \, \exp(\boldsymbol{b}_i' z) \qquad (2)$$

A discretized version of model (2) in which the jump times $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, ¼, $\boldsymbol{t}_k$ are fixed and the levels h_1, h_2, ¼, h_{k-1} form a first order auto-regressive process has been considered by Gamerman (1991) and West (1992). Arjas and Gasbarra (1994) and McKeague and Tighiouart (2002) extended model (1) by allowing the jump times to be random and McKeague and Tighiouart (2000) considered a dynamic version of model (2) in which the log-levels $\boldsymbol{l}_i = \log(h_i)$ and covariate effects $\boldsymbol{b}_i$, $i=1,2,\dots$ form a Gaussian Markov random field. A related Markov random field model for the prior intensity of a non-homogenous Poisson process was introduced by Arjas and Heikkinen (1997), but was not studied in the survival analysis context and adjustment for covariate effects was not considered.

The class of priors used by McKeague and Tighiouart (2000) for $\boldsymbol{b}(t)$ implies independence between the covariate effects, an assumption that may not be true in practice. For instance, in a case study of nasopharynx cancer survival data, West (1992) and McKeague and Tighiouart (2000) showed a clear correlation between the posterior

mean effects of the two measures of the extent of the cancer, which was not accounted for in the model.

In this paper, I extend the model proposed by McKeague and Tighiouart (2000) by implementing a correlation structure between some of the covariate effects in the prior. I use the pseudo-Bayes factor for model selection, and calculation of the conditional predictive ordinates (CPO's) are performed using the output from the Metropolis-Hastings-Green (MHG) algorithm (Metropolis et al., 1953; Hastings, 1970; Green (1995). The analysis indicates that the null hypothesis of no correlation between the effects of the two measures of the extent of the cancer is rejected and a correlated prior process should be used to estimate conditional survival probabilities.

Methodology

Let $T_1, ¼, T_n$ be nonnegative independent random variables with associated p-dimensional covariate vectors $z_j, j = 1,\ldots, n$. Assume that the data may be subject to right censoring, i.e., we observe $(X_1, \boldsymbol{d}_1, z_1),\ldots, (X_n, \boldsymbol{d}_n, z_n)$ where $X_j = \min(T_j, U_j)$, U_j being the censoring time for the j-th individual, and $\boldsymbol{d}_j = I\{T_j \leq U_j\}$. The conditional hazard function is given by (2), where $I\{\}$ is the indicator function, $0 = t_1 < t_2 < t_3 < \ldots$ is an increasing sequence of jump times, the h_i's represent the levels of the baseline hazard function $h_0(t)$, and $\{\boldsymbol{b}_i, i \geq 1\} = \{(\boldsymbol{b}_{i1},\ldots,\boldsymbol{b}_{ip})', i\geq 1\}$ is a p-dimensional process describing the effect of covariate vector z.

Let $\boldsymbol{t}_{\max} = \max\{X_j, 1\leq j\leq n\}$. The Bayesian approach consists of putting a prior distribution on the p covariate effects and the unknown baseline hazard function. The jump times $\boldsymbol{t}_2, \boldsymbol{t}_3, \ldots$ form a time-homogeneous Poisson process with rate $\boldsymbol{g}$. The prior distributions of the remaining parts of the model are specified conditionally given the number m of $\boldsymbol{t}_i$, $i \geq 1$ in the interval $[0, \boldsymbol{t}_{\max}]$, as follows.

Covariate Effects Prior

I specify $\boldsymbol{b}_m = \{\boldsymbol{b}_{kj}: k=1,\ldots, m, j=1,\ldots, p\}$ to be a Gaussian Markov random field with a neighborhood system $\{\P(k, j)\}$ of the following form: $\P(k,j) = \{(k\text{-}1,j), (k+1,j), (k, l), l\hat{\boldsymbol{I}} \ \P(j)\}$, where $\{\P(j), j = 1,\ldots, p\}$ is a given neighborhood system for the covariate effects. This means that

interactions in time are only permitted between the same components of the covariate effects. The model then amounts to:

$$\boldsymbol{b}_{kj}|\{ \boldsymbol{b}_{il}, (i,l)\neq(k,j)\}$$
$$\overset{d}{=}\boldsymbol{b}_{kj}|\{ \boldsymbol{b}_{il}, (i,l)\in \partial(k,j)\}. \quad (3)$$
$$\sim N(\boldsymbol{n}_{kj},\boldsymbol{s}_{kj}^2)$$

In addition, I assume only pairwise interaction between the covariate effects. It follows that the conditional mean $\boldsymbol{n}_{kj}$ is given by

$$\boldsymbol{n}_{kj} = \boldsymbol{m}_{kj} + s_{kj}\left(\boldsymbol{b}_{k-1,j} - \boldsymbol{m}_{k-1,j}\right)$$
$$+ r_{kj}\left(\boldsymbol{b}_{k+1,j} - \boldsymbol{m}_{k+1,j}\right) +$$
$$\sum_{l\in\partial(j)} \boldsymbol{r}_{kl}\left(\boldsymbol{b}_{kl} - \boldsymbol{m}_{kl}\right),$$

see Cressie (1993, Ch. 6).

The hyperparameters $\boldsymbol{m}_{kj} = E(\boldsymbol{b}_{kj})$, $k = 1,\ldots, m$ represent the trend in the levels of the j-th covariate effect, s_{kj}, r_{kj} are used to smooth the j-th covariate effect, and $\boldsymbol{r}_{kl}$, $l\in\P(j)$ measure the correlation between $\boldsymbol{b}_{kj}$ and $\boldsymbol{b}_{kl}$, $l\in\P(j)$. The distribution of $\boldsymbol{b}_m$ is completely determined by its local characteristics provided the hyperparameters satisfy the following conditions: $s_{kj}, r_{kj}, \boldsymbol{r}_{kl}, l\hat{I}\P(j)$ are nonnegative with

$$s_{kj} + r_{kj} + \sum_{l\in\partial(j)} \boldsymbol{r}_{kl} <1, \quad r_{kj}\boldsymbol{s}_{k+1,j}^2 = s_{k+1,j}\boldsymbol{s}_{kj}^2, \quad j=1,\ldots,$$
$$p \text{ and } \boldsymbol{r}_{kj}\boldsymbol{s}_{kj}^2 = \boldsymbol{r}_{kl}\boldsymbol{s}_{kl}^2 \text{ for } j\hat{I}\P(l),$$

see, e.g., Besag and Kooperberg (1995). McKeague and Tighiouart (2000) introduced a way of controlling the hyperparameters by the length of adjacent time intervals and I can adapt their approach to the present setting as follows:

$$r_{kj} = \frac{\left(\Delta_k + \Delta_{k+1}\right)c_j}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}}, \quad s_{kj} = \frac{\left(\Delta_{k-1} + \Delta_k\right)c_j}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}},$$
$$\boldsymbol{s}_{kj}^2 = \frac{2\boldsymbol{s}_j^2}{\Delta_{k-1} + 2\Delta_k + \Delta_{k+1}}, \quad \boldsymbol{r}_{kj} = \boldsymbol{r}\boldsymbol{s}_j^2,$$

where $\Delta_k = \Delta_{k+1} - \Delta_k$ is the gap between the k-th and $(k+1)$-st jump times, $2 \leq k \leq m-1$, and the parameters c_j, $s_j > 0$, and $r \geq 0$ satisfy

$$c_j + r \sum_{l \in \partial(j)} s_l^2 < 1, \quad j=1,\ldots,p. \qquad (4)$$

The parameter g controls the rate of jump times, c_j controls the nearest neighbor interaction between the levels of the j-th covariate effect, s_j represents the precision of the prior information of the j-th covariate effect, and r controls the dependency structure between neighboring covariate effects: higher values of r signify greater correlation, and $r = 0$ gives rise to the conditionally independent time-varying covariate effects model analyzed by McKeague and Tighiouart (2000). For simplicity of presentation, I restrict attention to the case $m_{kj} = m_j$ which indicates constant prior levels in the mean of the j-th covariate effect.

The distribution of b_m is Gaussian with mean vector m_{bm} and covariance matrix $(I_{mp}-C_1)^{-1} M_1$, where $m_{bm} = \{m_{kj}: k=1,\ldots, m, j=1,\ldots, p\}$, C_1 is an $mp \times mp$ matrix defined as follows. For $j=1,\ldots, p$ and $i=m(j-1),\ldots, mj$, $c_{i,i+1}=r_{ij}$, $c_{i,i-1}=s_{ij}$, $c_{i,i+ml}=r_{il}$, for $l \hat{I} \P(j)$, $c_{i+ml,i}=r_{i+ml,i}$ for $i \hat{I} \P(j)$, $c_{lk}=0$ otherwise, $M_1=\text{diag}(s^2_{kj}, k=1,\ldots, m, j=1,\ldots, p)$, and I_{mp} is the identity matrix.

Baseline Hazard Prior

Let $l_i = \log(h_i)$. The prior distribution for the levels of the log-baseline hazard $l_1,\ldots, l_m$ is taken to be the same as the prior for the j-th covariate effect when $p=1$. Denote by $m_k = E(l_k)$ the trend in the levels of the baseline hazard function, s_k^2 the conditional variance of l_k given l_i, $i \neq k$, and s_k, r_k the influences of the left and right neighbors of l_k, respectively. The corresponding nearest neighbor interaction and precision of the prior information parameters will be denoted by c and s, respectively.

In what follows, I denote by l_m both the random vector $(l_1,\ldots, l_m)$ and the last log-level of the baseline hazard function. The joint distribution of $l_m=(l_1,\ldots, l_m)$ is Gaussian with mean vector m_{lm} and covariance matrix $(I_m-C)^{-1} M$, where $m_{lm} = (m,\ldots, m)$, $C = (c_{ij})_{1 \leq i,j \leq m}$, $c_{i,i+1} = r_i$, $c_{i,i-1} = s_i$, $M =$

$\text{diag}(s_1^2,\ldots, s_m^2)$, and I_m is the identity matrix. Again, I will assume that $m_k = m$ indicating a constant prior level in the mean of the log-baseline hazard function.

Data Likelihood and Posterior

For $k = 1,\ldots, p$, and $i = 1,\ldots, m$, let N_i be the number of observed deaths in the interval $(t_i, t_{i+1}]$, $W_{ik} = \sum_{\{j: t_i < X_j \leq t_{i+1}, d_j=1\}} z_{jk}$, and $W_i = (W_{i1},\ldots, W_{ip})$ with $t_{m+1} = t_{\max}$. Assuming that the censoring mechanism is non-informative, the likelihood is proportional to the product form

$$\prod_{j=1}^{n} [(h(X_j| z_j)]^{d_j} \prod_{j=1}^{n} \exp\left\{-\int_0^{X_j} h(s| z_j)\, ds\right\}$$

$$= \exp\{\sum_{i=1}^{m} (N_i l_i + b_i' W_i)$$

$$-\int_0^{t_{\max}} [\sum_{j=1}^{n} I(X_j \geq s)\, h(s| z_j)]\, ds\}.$$

Let $t_m = (t_1,\ldots, t_m)$, and $l_m = (l_1,\ldots, l_m)$, then the posterior density of the parameter (t_m, l_m, b_m) is proportional to the product of the prior and likelihood

$$g^m (2p)^{-3m/2} |A|^{1/2} \exp\left\{-\frac{1}{2}(l_m - m_{lm})' A (l_m - m_{lm})\right\}$$

$$|A_1|^{1/2} \exp\left\{-\frac{1}{2}(b_m - m_{bm})' A_1 (b_m - m_{bm})\right\}$$

$$\times \exp\{\sum_{i=1}^{m} (N_i l_i + b_i' W_i)$$

$$-\int_0^{t_{\max}} [\sum_{j=1}^{n} I(X_j \geq s)\, h(s| z_j)]\, ds\},$$

where $A = M^{-1}(I_m - C)$ and $A_1 = M_1^{-1}(I_{2m} - C_1)$.

I use a reversible jump MCMC algorithm to extract features from this posterior distribution, see the appendix.

Model Comparison

In this section, I test the null hypothesis $H_0: r = 0$ against the alternative $H_1: r > 0$. This is equivalent to selecting between the conditionally independent time-varying covariate effects model

M_1 analyzed by McKeague and Tighiouart (2000) and model M_2, in which the covariate effects satisfy (3). Pseudo-Bayes factor is used to select the best model (Gelfand et al. (1992)), and its calculation uses the output of the MHG sampler.

Let $X = (X_1,\ldots, X_n)$ denote the data vector, and $q = (l(t), b(t))$ be the model parameter. The predictive density is $f(X) = \int f(X|\ q\ z)\ p(q)\ dq$, where $p(q)$ denotes the prior density of q and the conditional predictive ordinate (CPO) is given by

$$f(X_i \mid X_{(i)}) = \frac{f(X)}{f(X_{(i)})}$$
$$= \int f(X_i \mid X_{(i)}, q, z)\, p(q \mid X_{(i)})\, dq,$$

where $X_{(i)}$ is the data vector X with X_i deleted. The pseudo-Bayes factor is given by

$$B = \frac{\prod_{i=1}^{n} f(X_i \mid X_{(i)}, M_1)}{\prod_{i=1}^{n} f(X_i \mid X_{(i)}, M_2)}$$

and model selection proceeds by choosing M_1 (M_2) according to $B > (<)\ 1$. For a complete discussion and justification of this technique, see Geisser and Eddy (1979), Box (1980), Gelfand et al. (1992), and Gelfand and Mallick (1995).

Exact calculation of B is not possible, however Monte Carlo estimates of the CPO's can be obtained using the output of the MHG sampler $q_1,\ldots,\ q_N$ and the idea of importance sampling density, see Gelfand and Dey (1994). The approximation is given by

$$f(X_i \mid X_{(i)}) \approx N \left[\sum_{j=1}^{N} \frac{1}{f(X_i \mid q_j, z)} \right]^{-1}.$$

For a censored observation, I compute the conditional survival function $S(X_i|X_{(i)}, M_j), j=1, 2$.

Results

West (1992) and McKeague and Tighiouart (2000) studied data on 181 nasopharynx cancer patients whose cancer careers, culminating in either death (127 cases) or censoring (54 cases) are recorded to the nearest month, ranging from 1 to 177 months.

The analyses were based on five covariates: (1) Sex of the patient (0 for male, 1 for female); (2) Age of the patient at time $t = 0$, the start of monitoring of the cancer career of that patient (standardized to have zero mean and unit standard deviation across all patients in the study); (3) Dosel, an average measure of the extent of radiotherapy treatment to which the patient has been subjected (also standardized, as with age); (4) Tumor1, a measurement of the extent of the cancer (in terms of an estimate of the number of cancerous cells), taking value 1, 2, 3 or 4; (5) Tumor2, a measure similar to Tumor1, taken from a different X-ray section, again taking values 1, 2, 3 or 4.

The right hand side of Figure 1 (following page) shows the posterior mean effects for tumor1 and tumor2 obtained by McKeague and Tighiouart, and the left hand side the estimates obtained by West. The similar pattern of the posterior mean effects of tumor1 and tumor2 suggests that a correlated prior process for the two effects is more realistic. I therefore fitted model M_2 with $r = 1/2$ and compared it with model M_1, fitted by McKeague and Tighiouart, which corresponds to $r = 0$. The remaining hyperparameters were the same for both models, and can be found in McKeague and Tighiouart (2000). The logarithm of the pseudo-Bayes factor is found to be

$$Log(B) = Log \left[\frac{\prod_{i=1}^{n} f(X_i \mid X_{(i)}, M_2)}{\prod_{i=1}^{n} f(X_i \mid X_{(i)}, M_1)} \right] = 4.56$$

and

$$\frac{\prod_{i=1}^{n} S(X_i \mid X_{(i)}, M_2)}{\prod_{i=1}^{n} S(X_i \mid X_{(i)}, M_1)} = 1.53$$

suggesting that a time-varying correlated covariate effect should be used to estimate conditional survival probabilities.

Conclusion

I have presented a complete nonparametric Bayesian approach to inference from right-censored survival data. The methodology is an extension of the model proposed by McKeague and Tighiouart (2000) in the sense that the Bayesian model accounts for any correlation structure between some of the time-varying covariate effects in the prior. Except for the constraints (4), direct specification of the parameter controlling the amount of correlation between the covariate effects is not possible. A second stage prior can easily be placed on the hyperparameter ρ ; I did not pursue this hierarchy here because my goal is to simplify the presentation of this methodology.

The computational method used to extract features of the posterior distribution is similar to the one used in McKeague and Tighiouart (2000). The only difference is the extra term involved in the prior ratio of the correlated covariate effects. This is very convenient when writing the codes of this sampler.

The methodology was illustrated by an analysis of a nasopharynx cancer survival data set. The class of prior processes defining the Bayesian model was flexible enough to detect a correlation structure between some of the time-varying covariate effects; in particular, pseudo-Bayes factors were calculated to support this evidence.

Figure 1.
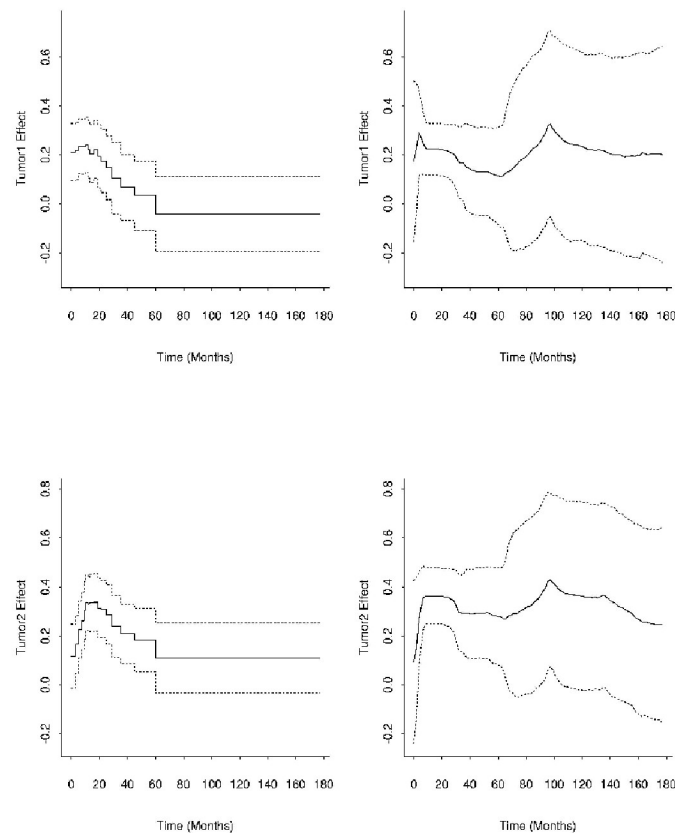
References

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (1992). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Arjas, E., & Gasbarra, D. (1994). Nonparametric Bayesian Inference for Right-Censored Survival Data, Using the Gibbs Sampler. *Statistica Sinica*, 2, 505-524.

Arjas, E., & Heikkinen, J. (1997). An Algorithm for Nonparametric Bayesian estimation of a Poisson intensity. *Journal of Computational Statistics*, 12, 385-402.

Besag, J. E., & Kooperberg, C. (1995). On Conditional and Intrinsic Autoregressions. *Biometrika*, 82, 733-746.

Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). *Journal of the Royal Statistical Society*, B 34, 187-220.

Cressie, N. (1993). *Statistics for Spatial Data.* New York: Wiley.

Gamerman, D. (1991). Dynamic Bayesian Models for Survival Data. *Applied Statistics*, 40, 63-79.

Gelfand, A. E., Dey, D. K., & Chang, H. (1992). Model Determination using Predictive Distribution with Implementation via Sampling-Based Methods. *Bayesian Statistics* 4, 147-167.

Gelfand, A. E., & Dey, D. K. (1994). Bayesian Model Choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society*, B 56, 501-514.

Gelfand, A. E., & Mallick, B. K. (1995). Bayesian Analysis of Proportional Hazards Models Built from Monotone Functions. *Biometrics*, 51, 843-852.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82, 711-732.

Hastings, W. K. (1970). Monte Carlo Sampling Methods using Markov Chains and their Applications. *Biometrika*, 57, 97-109.

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

McKeague, I. W., & Tighiouart, M. (2002). Nonparametric Bayes Estimators for Hazard Functions Based on Right Censored Data. *Tamkang Journal of Mathematics*, 33, No 2, 173-189.

McKeague, I. W., & Tighiouart, M. (2000). Bayesian Estimators for Conditional Hazard Functions. *Biometrics*, 56, 1007-1015.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087-1092.

West, M. (1992). *Modelling Time-Varying Hazards and Covariate Effects.* Survival Analysis: State of the Art, J. P. Klein, P. K. Goel, eds. Kluwer, Boston, 47-62.

Appendix

To simplify the description of the algorithm, I will assume that $p=2$ and will denote by $a(t)$ and $b(t)$ the two time-varying correlated covariate effects and m_a, m_b their constant prior means, respectively. The constant prior mean of the log-baseline hazard function $l(t)$ will be denoted by m_l. The procedure for calculating features of the posterior distribution of (t_m, l_m, a_m, b_m) (note that here m is random) consists of running a reversible Markov chain on the state space $S = \bigcup_{i \geq 1} S_i$, where $S_i = D_i \times \mathbb{R}^{3i}$, and $D_i = \{(x_1, x_2, \ldots, x_i): 0 = x_1 < x_2 < \ldots < x_i < t_{\max}\}$, using the Metropolis-Hastings-Green algorithm.

A transition from (t_m, l_m, a_m, b_m) to $(t'_{m'}, l'_{m'}, a'_{m'}, b'_{m'})$ is accomplished by randomly selecting one of five types of moves (H_0, H_a, H_b, B, D): a change of height of a randomly selected level of the baseline hazard rate, change of height of a randomly selected level of the covariate effect $a(t)$, change of height of a randomly selected level of the covariate effect $b(t)$, birth of a new jump time at a randomly selected location in $(0, t_{\max})$, and death of a randomly selected jump time, respectively.

When selecting moves of type H_0, H_a, H_b, the acceptance probability is the same as in the classical Metropolis-Hastings algorithm:

min {1, (likelihood ratio) × (prior ratio) × (proposal ratio)},

whereas if moves of type B or D are selected, the current state (t_m, l_m, a_m, b_m) is mapped onto $(t'_{m'}, l'_{m'}, a'_{m'}, b'_{m'})$ by a one-to-one transformation t. The acceptance probability then takes the form:

min {1, (likelihood ratio) × (prior ratio) × (proposal ratio) × $J(t)$},

where $J(t)$ is the Jacobian of the transformation t. Except for the expressions of the prior ratios in the moves of type H_a, H_b, B, and D, a complete description of the types of moves, transformation t, expressions of the likelihood and proposal ratios, and

the Jacobian can be found in McKeague and Tighiouart (2000).

Move of type H_α:

An index k is uniformly selected from $\{1,2,\ldots, m\}$ and V is generated uniformly in the interval $(-\delta_\alpha, \delta_\alpha)$, where δ_α is a sampler parameter. The proposed new level for the covariate effect $\alpha(t)$ is $\alpha'_k = \alpha_k + V$. The proposed new point is $(\tau'_m, \lambda'_m, \alpha'_m, \beta'_m)$ with $\tau'_m = \tau_m$, $\lambda'_m = \lambda_m$, $\beta'_m = \beta_m$, and $\alpha'_i = \alpha_i$ for $i \neq k$.

The prior ratio is:
$\exp\{-\Delta_H(A_1, \alpha_m, \alpha'_m, \beta_m)/2\}$,
where for a $2m \times 2m$ matrix A and m-dimensional vectors α, α', β,

$$
\begin{aligned}
\Delta_H(A, \alpha, \alpha', \beta) =\ & a_{kk}(\alpha'_k - \alpha_k)(\alpha_k + \alpha'_k - 2\mu_\alpha) \\
& + 2a_{k,k-1}(\alpha_{k-1} - \mu_\alpha)(\alpha'_k - \alpha_k) \\
& + 2a_{k,k+1}(\alpha_{k+1} - \mu_\alpha)(\alpha'_k - \alpha_k) \\
& + 2a_{k,m+k}(\alpha'_k - \alpha_k)(\beta_k - \mu_\beta).
\end{aligned}
$$

A similar expression holds for a move of type H_β.

Move of type B:

A new jump time τ^* is drawn uniformly in the interval (τ_1, τ_{max}). Suppose that $\tau^* \in (\tau_{k-1}, \tau_k)$. This new jump time induces two new levels for the log-baseline hazard rate λ'_{k-1} and λ'_k, two new levels for the covariate effect $\alpha(t)$, α'_{k-1}, α'_k, and two new levels for the covariate effect $\beta(t)$, β'_{k-1}, β'_k using the transformation τ described in McKeague and Tighiouart.

The prior ratio is :

$$
\gamma(2\pi)^{-3/2}\left(\frac{|A'||A'_1|}{|A||A_1|}\right)^{1/2}
$$

$$
\exp\left\{\frac{1}{2}(\Delta_B(A, A', \lambda_m, \lambda'_m) + \Delta_B(A_{11}, A'_{11}, \alpha_m, \alpha'_m)\right.
$$

$$
\left. + \Delta_B(A_{22}, A'_{22}, \beta_m, \beta'_m) + \Delta_{BC}(A_1, A'_1, \alpha_m, \beta_m, \alpha'_m, \beta'_m))\right\},
$$

where A_{11}, A_{22} are $m \times m$ matrices such that

$$
A_1 = \begin{pmatrix} A_{11} & \times \\ \times & A_{22} \end{pmatrix},
$$

and for an $m \times m$ matrix A and m-dimensional vector λ_m,

$$
\begin{aligned}
\Delta_B(A, A', \lambda_m, \lambda'_m) =\ & (\lambda_{k-2} - \mu_\lambda)^2(a_{k-2,k-2} \\
& - a'_{k-2,k-2}) \\
& + (\lambda_k - \mu_\lambda)^2(a_{kk} - a'_{kk}) \\
& + a_{k-1,k-1}(\lambda_{k-1} - \mu_\lambda)^2 \\
& - a'_{k-1,k-1}(\lambda'_{k-1} - \mu_\lambda)^2 \\
& - a'_{k+1,k+1}(\lambda'_k - \mu_\lambda)^2 \\
& - 2(\lambda_{k-2} - \mu_\lambda)[a'_{k-2,k-1} \\
& (\lambda'_{k-1} - \mu_\lambda) \\
& - a_{k-2,k-1}(\lambda_{k-1} - \mu_\lambda)] \\
& - 2(\lambda_k - \mu_\lambda)[a'_{k,k+1}(\lambda'_k - \mu_\lambda) \\
& - a_{k-1,k}(\lambda_{k-1} - \mu_\lambda)] \\
& - 2a'_{k-1,k}(\lambda'_{k-1} - \mu_\lambda) \\
& (\lambda'_k - \mu_\lambda).
\end{aligned}
$$

Also, for a $2m \times 2m$ matrix A_1, and m-dimensional vectors α_m, β_m,

$$
\begin{aligned}
\Delta_{BC}(A_1, A'_1, \alpha_m, \beta_m, \alpha'_m, \beta'_m) =\ & -2(\alpha_{k-2} - \mu_\alpha)(\beta_{k-2} - \mu_\beta) \\
& (a'_{k-2,m+k-2} - a_{k-2,m+k-2}) \\
& - 2(\alpha_k - \mu_\alpha)(\beta_k - \mu_\beta) \\
& (a'_{k+1,m+k+1} - a_{k,m+k}) \\
& - 2a'_{k-1,m+k-1} \\
& (\alpha'_{k-1} - \mu_\alpha)(\beta'_{k-1} - \mu_\beta) \\
& - 2a_{k,m+k} \\
& (\alpha'_k - \mu_\alpha)(\beta'_k - \mu_\beta) \\
& + 2a_{k-1,m+k-1} \\
& (\alpha_{k-1} - \mu_\alpha)(\beta_{k-1} - \mu_\beta).
\end{aligned}
$$

Move of type D:

An index k is uniformly selected from $\{2,3,\ldots, m\}$ corresponding to the removal of the jump time τ_k. The proposed new point is $(\tau'_m, \lambda'_m, \alpha'_m, \beta'_m)$, with $\lambda'_i = \lambda_i$, $\alpha'_i = \alpha_i$, $\beta'_i = \beta_i$ for $i \leq k-2$, $\lambda'_j = \lambda_{j+1}$, $\alpha'_j = \alpha_{j+1}$, $\beta'_j = \beta_{j+1}$, $\tau'_j = \tau_{j+1}$ for $j \geq k$, and $\tau'_i = \tau_i$ for $i \leq k-1$.

The likelihood, prior, proposal ratios, and the Jacobian for this type of move are the inverse ratios of the ones for the move of type B with the proper labeling of the jump times and the log-hazard levels.