Supplement for Semiparametric frailty model for clustered interval-censored data

Aysun Çetinyürek Yavuz [1], and Philippe Lambert [1,2]

[1] Faculté des Sciences Sociales, Université de Liège, Liège, Belgium
[2] Institut de Statistique, Université catholique de Louvain, Louvain la Neuve, Belgium

Address for correspondence: Faculté des sciences sociales, Université de Liège, Place des orateurs, 3, Quartier Agora (Bat. B31) B-4000 Liège, Belgium.
E-mail: cetinyurek@yahoo.com.

Abstract: This document contains supplementary materials for the article of the same title submitted for publication to the Statistical Modelling: An international journal.

Key words: Interval-censoring; shared frailty; Bayesian; P-splines

1 Introduction

This document contains supplementary materials for the article of the same title submitted for publication to Statistical Modelling: An international journal. Notation not defined in this document is defined as in the main paper. Section 2 presents some details about the MCMC, e.g. the frequentist density estimation procedure from interval-censored data and some strategies for improving mixing and convergence of the chains, that are not addressed in the main paper. Section 3 provides further information on an extensive simulation study while Section 3.1 presents some sensitivity analysis for the considered prior distributions. Section 4 provides some details a data set from Signal Tandmobiel® study.

2 MCMC

MCMC is an important tool for estimating statistical models. However, especially with complex problems, MCMC can require massive computing resources and converge too slowly. In the following sections, we present various useful approaches in order to improve mixing and convergence of the chains.

2.1 Initialization

Starting the chains at good initial values fasten convergence. Usually, these could be obtained from restricted frequentist models as described in the following. The initial state of the chain $\boldsymbol{\vartheta_{(0)}} = \left(\boldsymbol{\phi}_{(0)}, \tau_{(0)}, \xi_{(0)}, \beta_{(0)}, \boldsymbol{z}_{(0)}, \alpha_{(0)}\right)^T$ is chosen as follows: First, a value for $\boldsymbol{\phi}_{(0)}$ can be obtained using the frequentist density estimation procedure (described in detail in Section 2.2) (Çetinyürek-Yavuz and Lambert, 2011). Then, we define $\tau_{(0)}$ as the value of τ yielding the smallest BIC for different values taken in a grid. $\zeta_{(0)}$ is taken as the proportion of pseudo-counts corresponding to small bins located below t_{cens}. In accordance with the estimation of the spline coefficients, we start by ignoring possible covariate effects: $\boldsymbol{\beta}_{(0)} = 0$. We can obtain $\boldsymbol{z}_{(0)}$ from a gamma frailty model applied on data resulting from mid-point imputation, where the estimated variance is used for $\alpha_{(0)}$.

For the nonparametric specification of the frailty, the initial values of the chain $\boldsymbol{\vartheta}$ are chosen along the same lines as described for the parametric frailty model, except for $(\boldsymbol{\phi}^*_{(0)}, \tau^*_{(0)})$. For obtaining an initial estimate of the frailty density, we also applied the same frequentist procedure as described in Section 2.2. However, as frailty is not observed directly, initial values for the latent frailty terms need to be obtained. It is made in two steps: we first fit a parametric (gamma or log-normal) frailty model using midpoint imputation for the interval-censored data. Then, using the estimated frailty terms as if they were actually observed, a value for $\boldsymbol{\phi}^*_{(0)}$ can be obtained from the frequentist procedure after obtaining an approximation for the density of $\boldsymbol{z^*}_{(0)}$. Similarly, $\tau^*_{(0)}$ is defined as the value of τ^* evaluated on a grid yielding the smallest BIC.

2.2 Frequentist estimation of baseline density

Initial values of the spline coefficients and penalty parameters can be obtained using the naive frequentist models. In this spirit, we shall explain a frequentist density estimation procedure from time-to-event data when the covariates and the possible heterogeneity are ignored ($\boldsymbol{\beta} = 0$, $\mathbf{b_g} = 1$ (or $\boldsymbol{z} = 0$)). We start by partitioning the support of t into small bins (more than 100 small bins of equal width) for obtaining an accurate estimate of the density for time-to-event data. Then, following an approach similar to Eilers and Marx (1996) we calculate the number of observations in each small bin, namely the pseudo-counts. These pseudo-counts, which are calculated from the C matrix of the composite link model defined in Eilers and Marx (1996), are later used to build the density estimate. The relationship between the intervals (l_{gj}, r_{gj}) and the small bins is provided by a GxIxn_g array $C = [c_{gji}]$ such that $c_{gji} = 1$ if the ith small bin $I_i \subset (l_{gj}, r_{gj})$ and 0 otherwise. In this spirit, for cluster g, each element of a row in the C matrix is divided by the sum of the elements in that row ($C_{gj.} = \sum_i C_{gji}$). The so-obtained numbers, $W_{gji} = C_{gji}/C_{gj.}$, provide the contribution of the concerned observation in cluster g (e.g. a patient for a multicenter clinical trial) for each small bin partitioning (a, t_{cens}). Then, the contributions of each observation for the i^{th} small bin, I_i, are summed over all observations ($W_{..i} = \sum_g \sum_j W_{gji}$) and rounded to the nearest integer value y_i in order to get the pseudo-count for that small bin. Note that π_i denotes the probability to have an event time in I_i, then the likelihood for these pseudo-

counts is proportional to $\prod_i^I \pi_i^{y_i}$. Alternatively, the well known link between the Poisson and the multinomial distributions suggests to assume that the pseudo-counts, y_i, have a Poisson distribution with mean $\mu_i = \pi_i y_+$ conditional on the total number of observations $y_+ = \sum_i^I y_i$. Using a rich B-spline basis as regressors in a log-linear model for the mean, one obtains the likelihood

$$\mathcal{L}(\boldsymbol{\phi}|y) = \sum_{i=1}^{N} y_i \log(\mu_i) - \sum_{i=1}^{N} \mu_i,$$

where $\log(\mu_i) = \eta_i = \sum_{k=1}^{K} \phi_k b_k(u_i)$. Then, by subtracting the 2^{nd} order penalty (say) and a small ridge penalty from $\mathcal{L}(\boldsymbol{\phi}|y)$, one obtains the penalized log likelihood function

$$\mathcal{L}_p(\boldsymbol{\phi}|y,\tau) = \mathcal{L}(\boldsymbol{\phi}|y) - \frac{\tau}{2}\boldsymbol{\phi'P\phi},$$

where $\boldsymbol{\phi'D'D\phi} = \sum_k (\phi_k - 2\phi_{k-1} + \phi_{k-2})^2$ and $\boldsymbol{P} = \boldsymbol{D'D} + \epsilon I$. The function $\mathcal{L}_p$ can be optimized by solving the score equations $B^T(y-\mu) = \tau \boldsymbol{P\phi}$, using iteratively reweighted least squares (IRWLS): iteratively solve (for ϕ)

$$(B^T\tilde{W}B + \tau P)\phi = B^T\tilde{W}(y - \tilde{\mu}) + B^T\tilde{W}B\tilde{\phi},$$

where $\widetilde{W}$ is a diagonal matrix with elements $\mu_i(\widetilde{\phi})$ and $\tilde{\phi}$ and $\tilde{\mu}$ are current approximations to the solution. The variance-covariance matrix for the estimated spline coefficients $\boldsymbol{\phi}$ is given by (at convergence),

$$\Sigma_0 = (B^T W B + \tau \boldsymbol{P})^{-1}. \tag{2.1}$$

More detailed information can be found in Eilers and Marx (1996). Information criteria such as AIC or BIC could be used for choosing the initial optimal (plausible) value of the penalty parameter τ. In our experience, BIC is preferable to AIC as AIC tends to undersmooth the target curve, which was also mentioned by other authors (Strasak et al., 2009).

2.3 Automatic tuning of the algorithm

Good acceptance rates can be achieved via a careful choice of the standard deviation δ_h in the generation of proposals in a Metropolis algorithm. For an optimal use of Metropolis algorithm, it is recommended to tune the acceptance probability to approximately 0.44 in one dimensional space decreasing to 0.23 in high dimensional spaces (Gelman et al., 1996; Roberts and Rosenthal, 2001). Let δ denote the tuning parameter of interest. The value of δ at iteration $m+1$ can be adjusted using the value at iteration m using (with $\overline{\eta} = 0.44$)

$$\sqrt{\delta_{m+1}} = h\left(\sqrt{\delta_m} + \gamma_m\left(\alpha(\boldsymbol{\theta_{(h)}}, \boldsymbol{\theta_{(h-1)}}) - \overline{\eta}\right)\right)$$

with

$$h(x) = \begin{cases} \epsilon & \text{if } x < \epsilon \\ \text{x} & \text{if } x \in (\epsilon, A) \\ \text{A} & \text{if } x > A \end{cases}$$

where ϵ is a very small number (say 0.0001) and A a large one (say 10000). If the targeted acceptance level is not achieved, these constants should be changed. The series $\{\gamma_m\}$ is a non-increasing sequence of positive real numbers such that $|\gamma_m - \gamma_{m-1}| \leq m^{-1}$. Possible choices for γ_m are $\frac{10}{m}$ or $\frac{1}{m}$. Practically, the MCMC algorithm is run for a few hundred iterations with the δ_m's automatically updated to achieve the targeted acceptance rate (Atchadé and Rosenthal, 2005). Then, the last value of δ_m in the so-generated chain can be used in a non-adaptive version of the modified Metropolis algorithm to produce the long chain(s) that will be used for inference.

2.4 Reparametrizing the posterior

The mixing of the chain could be improved by using a Metropolis algorithm on a re-parametrized posterior (Lambert, 2007). In this sense, one can use an approximation to the 2^{nd} order dependence structure of the conditional posterior. The variance covariance matrix, $\mathbf{\Sigma_0}$, of the penalized maximum likelihood estimator of the spline parameters $\boldsymbol{\phi}$ could be calculated for a fixed and reasonably chosen value of the roughness penalty parameter τ. Then, the posterior can be re-parametrized using φ (Equation 2.1) with $\phi = \phi_0 + L\varphi$ where L denotes the lower triangular matrix obtained from the Cholesky decomposition of $\mathbf{\Sigma_0}$. Then, the univariate Metropolis algorithm described before can be employed on the re-parametrized posterior. This also fastens convergence.

3 Simulation Study

Our data generation and simulation strategy contain the following steps:

1. Firstly, we generate the log-frailty terms z_g from one of the specified frailty distributions.
2. Then the values of the covariate, x_{gj} are generated.
3. Afterwards, given the values of frailty terms and the covariate, the observations t_{gj} $(g = 1, ..., G; j = 1, ..., n_g)$ are generated using the selected proportional hazards frailty model.
4. Each observation t_{gj} is converted into an interval of width w_{gj}, where w_{gj} is generated from a Gamma distribution with a mean equal to the targeted mean width (0.5σ, 1.0σ and 1.3σ) and a variance equal to one fifth of the mean. The interval corresponding to t_{gj} was finally defined as (L_{gj}, R_{gj})=$(t_{gj} - u_{gj} \ . \ w_{gj}, t_{gj} + u_{gj} \ . \ w_{gj})$ where u_{gj} is randomly generated from a uniform distribution on (0,1).

5. For each simulated data set, initial values for the spline parameters were obtained using the strategy described in Section 2.2.
6. We sample the posterior for the parameters of interest using MCMC (see Section 3.3).
7. Steps 1-6 were repeated for all data sets ($S = 300$ times) to obtain the Monte Carlo estimates for the quantities of interest.

3.1 Sensitivity analysis

Following the advice from the referees, some of the simulation studies have been run again with different prior specifications, namely Gamma (2,0.01) for the penalty parameters, and inverse-gamma(1,1) and inverse-gamma(2,1) for the standard deviation of the frailty. The detailed results are presented in the tables below. The results in Tables 1 and 2 can be compared to the results in the main paper. Table 1 and Table 2 presents two different sensitivity analysis with different prior distribution specifications, namely Gamma (2,0.01) prior for the penalty parameters, and inverse-gamma(1,1) and inverse-gamma(2,1) priors for the standard deviation of the frailty.

- The unimodal setting with α=0.8 can be compared to left half of Table 1 (main paper) for the given sample sizes.
- The bimodal setting with α=0.8 can be compared to left half of Table 2 (main paper) for the given sample sizes.
- The bimodal setting with α=1.2 can be compared to right half of Table 2 (main paper) for the given sample sizes.
- The skewed setting with α=1.2 can be compared to Table 3 (main paper) for the given sample sizes.

It reveals that changes in prior specification have a limited impact on bias and coverage of credible intervals for the regression parameters, with occasionally a slight improvement (over our standard prior) in the estimation of the standard deviation of the log-frailty. It should also be stressed here that sensitivity analyses were performed for the smallest sample sizes where the chosen prior has the biggest potential impact.

4 Application: Signal Tandmobiel® Study

The Signal Tandmobiel® data set results from a longitudinal prospective dental study performed in Flanders (northern Belgium) between 1996 and 2001, using 4468 randomly selected children attending the first year of primary school at the beginning of the study. Then annual dental examinations were performed on the selected cohort by one of 16 trained dentists.

Table 1: Sensitivity analysis I: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α in S=300 replications using Gamma(2,0.001) prior for penalty parameters τ and τ^*, Inverse-gamma(1,1) prior for standard deviation of frailty α

					Semiparametric					Gaussian				
					β=0.693					β=0.693				
Distribution	α	n_{cl}	n_g	N	Mean	90% CI	Rbias(%)	EC90	EC80	Mean	90% CI	Rbias(%)	EC90	EC80
Unimodal	0.8	20	10	200	0.75	(0.51-0.98)	8.4	90	80	0.75	(0.51-0.99)	8.6	91	80
		50	4	200	0.75	(0.46-1.03)	8.5	89	77	0.75	(0.46-1.04)	8.7	89	77
		50	6	300	0.74	(0.52-0.95)	6.7	91	78	0.74	(0.52-0.95)	6.8	89	77
		100	4	400	0.71	(0.51-0.91)	2.4	89	78	0.71	(0.51-0.91)	2.7	90	78
Bimodal	0.8	20	10	200	0.76	(0.51-1.06)	9.9	86	80	0.76	(0.50-1.05)	9.1	86	80
		50	4	200	0.75	(0.49-1.04)	8.3	89	79	0.76	(0.48-1.05)	8.0	88	79
		50	10	500	0.71	(0.55-0.88)	3.1	93	84	0.71	(0.55-0.87)	2.9	92	84
Bimodal	1.2	20	10	200	0.71	(0.42-1.02)	2.9	90	81	0.71	(0.40-1.01)	2.1	90	81
		50	4	200	0.72	(0.37-1.00)	4.5	88	75	0.73	(0.37-1.02)	4.7	89	75
		50	6	300	0.72	(0.48-0.96)	3.6	91	82	0.71	(0.46-0.95)	2.4	90	82
		100	4	400	0.71	(0.51-0.93)	2.4	90	80	0.71	(0.49-0.93)	2.9	90	80
Skewed	1.2	20	10	200	0.73	(0.45-1.02)	5.2	92	81	0.73	(0.45-1.01)	5.4	91	81
		50	4	200	0.72	(0.39-1.01)	3.5	90	77	0.72	(0.39-1.02)	3.5	90	79
		50	6	300	0.72	(0.48-0.96)	3.4	91	77	0.72	(0.47-0.96)	3.4	90	77
		100	4	400	0.70	(0.48-0.91)	0.4	88	81	0.69	(0.46-0.92)	0.1	89	80
					α					α				
Distribution	α	n_{cl}	n_g	N	Mean	90% CI	Rbias(%)	EC90	EC80	Mean	90% CI	Rbias(%)	EC90	EC80
Unimodal	0.8	20	10	200	0.87	(0.67-1.12)	8.3	94	87	0.85	(0.65-1.11)	6.2	93	86
		50	4	200	0.84	(0.68-1.06)	4.5	94	86	0.83	(0.68-1.07)	4.0	93	85
		50	6	300	0.83	(0.65-1.01)	3.5	95	85	0.82	(0.65-1.00)	2.7	94	84
		100	4	400	0.80	(0.65-0.95)	0.1	94	84	0.80	(0.65-0.95)	0.5	92	82
Bimodal	0.8	20	10	200	0.76	(0.51-1.06)	-5.5	86	80	0.88	(0.70-1.06)	9.8	96	86
		50	4	200	0.69	(0.55-0.88)	-7.7	84	76	0.84	(0.66-1.04)	5.4	95	86
		50	10	500	0.77	(0.65-0.87)	-3.3	95	93	0.82	(0.70-0.94)	2.9	99	91
Bimodal	1.2	20	10	200	1.28	(0.89-1.47)	7.8	91	83	1.27	(0.91-1.57)	6.2	93	83
		50	4	200	1.27	(0.98-1.57)	5.8	92	83	1.25	(0.98-1.55)	4.1	92	82
		50	6	300	1.12	(0.95-1.30)	-6.3	94	85	1.23	(1.03-1.45)	2.9	96	90
		100	4	400	1.23	(1.04-1.43)	2.4	92	86	1.23	(1.04-1.43)	2.5	93	84
Skewed	1.2	20	10	200	1.27	(0.96-1.65)	5.5	94	88	1.23	(0.93-1.61)	2.9	94	87
		50	4	200	1.25	(0.96-1.60)	4.0	92	84	1.22	(0.94-1.56)	1.9	92	83
		50	6	300	1.23	(0.95-1.52)	2.4	91	81	1.20	(0.93-1.49)	0.3	88	81
		100	4	400	1.21	(1.00-1.45)	1.0	92	83	1.20	(0.99-1.42)	-0.3	92	81

Table 2: Sensitivity analysis II: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α in S=300 replications using Gamma (2,0.001) prior for penalty parameters τ and τ^*, Inverse-gamma(2,1) prior for standard deviation of frailty α

					Semiparametric					Gaussian				
					β=0.693					β=0.693				
Distribution	α	n_{cl}	n_g	N	Mean	90% CI	Rbias(%)	EC90	EC80	Mean	90% CI	Rbias(%)	EC90	EC80
Unimodal	0.8	20	10	200	0.75	(0.49-1.03)	8.6	87	77	0.75	(0.49-1.02)	8.8	86	77
		50	4	200	0.74	(0.46-1.02)	7.0	89	80	0.74	(0.46-1.02)	7.1	88	80
		50	6	300	0.72	(0.52-0.95)	4.1	91	82	0.72	(0.52-0.96)	4.3	91	83
		100	4	400	0.71	(0.53-0.91)	1.9	89	81	0.71	(0.53-0.91)	2.1	90	80
Bimodal	0.8	20	10	200	0.74	(0.46-1.01)	7.3	89	73	0.74	(0.45-1.01)	6.7	87	73
		50	4	200	0.74	(0.45-1.02)	6.2	88	76	0.74	(0.45-1.04)	6.6	88	76
		50	10	500	0.72	(0.55-0.89)	3.4	89	80	0.71	(0.55-0.90)	3.2	89	80
Bimodal	1.2	50	4	200	0.72	(0.40-1.05)	3.3	89	79	0.72	(0.39-1.05)	3.4	90	80
		50	6	300	0.72	(0.47-0.99)	3.4	89	78	0.71	(0.46-0.99)	2.2	88	78
		100	4	400	0.69	(0.49-0.90)	-0.1	91	84	0.69	(0.48-0.90)	-0.2	91	83
Skewed	1.2	20	10	200	0.72	(0.45-1.00)	4.1	92	82	0.72	(0.45-1.00)	4.2	91	82
		50	4	200	0.70	(0.41-1.01)	1.7	91	83	0.70	(0.42-1.02)	1.7	91	84
		50	6	300	0.70	(0.47-0.95)	1.1	91	82	0.70	(0.47-0.95)	1.1	91	81
		100	4	400	0.68	(0.46-0.90)	-1.5	89	77	0.68	(0.46-0.89)	-1.6	88	78
					α					α				
Distribution	α	n_{cl}	n_g	N	Mean	90% CI	Rbias(%)	EC90	EC80	Mean	90% CI	Rbias(%)	EC90	EC80
Unimodal	0.8	20	10	200	0.82	(0.60-1.07)	2.3	96	86	0.81	(0.59-1.05)	1.6	92	85
		50	4	200	0.76	(0.59-1.01)	-4.4	93	78	0.76	(0.59-1.01)	-4.5	90	77
		50	6	300	0.78	(0.62-0.96)	-2.0	93	89	0.79	(0.63-0.97)	-1.7	95	83
		100	4	400	0.78	(0.65-0.94)	-2.6	96	89	0.78	(0.66-0.93)	-2.0	95	87
Bimodal	0.8	20	10	200	0.71	(0.57-0.86)	-11.6	91	82	0.81	(0.64-1.01)	1.4	99	92
		50	4	200	0.63	(0.34-0.83)	-21.6	72	59	0.81	(0.64-0.99)	1.1	97	90
		50	10	500	0.76	(0.66-0.87)	-5.2	96	90	0.81	(0.70-0.92)	0.8	98	94
Bimodal	1.2	20	10	200	1.20	(0.94-1.47)	0.4	95	89	1.19	(0.95-1.46)	-0.5	93	88
		50	4	200	1.09	(0.90-1.26)	-9.3	88	76	1.19	(0.98-1.40)	-0.6	94	89
		50	6	300	1.21	(1.03-1.41)	0.6	94	83	1.21	(1.03-1.40)	0.8	94	88
Skewed	1.2	20	10	200	1.16	(0.87-1.54)	-3.1	94	86	1.14	(0.86-1.51)	-4.6	93	83
		50	4	200	1.14	(0.88-1.43)	-5.0	94	80	1.13	(0.88-1.40)	-6.1	88	79
		50	6	300	1.17	(0.91-1.40)	-2.6	93	85	1.15	(0.90-1.36)	-4.1	90	81
		100	4	400	1.18	(0.97-1.39)	-2.7	90	78	1.17	(0.97-1.39)	-2.7	90	78

References

Atchadé, Y.F. and Rosenthal, J.S. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, **11 (5)**, 815–828.

Çetinyürek-Yavuz, A. and Lambert, P. (2011). Smooth estimation of survival functions and hazard ratios from interval-censored data using Bayesian penalized B-splines. *Statistics in Medicine*, **30**, 75–90.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89–121.

Gelman, A., Roberts, G.O., Gilks, W.R. (1996). Efficient Metropolis jumping rules. *Bayesian Statistics*, **5**, 599–607.

Lambert, P. (2007). Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics and Data Analysis*, **51**, 6307–6320.

Roberts, G.O. and Rosenthal, J.S. (2001). Optimal scaling for various Metropolis- Hastings algorithms. *Statistical Science*, **16**, 351–367.

Strasak, A.M., Lang, S., Kneib, T., Brant, L.J., Klenk, J., Hilbe, W., Oberaigner, W., Ruttmann, E., Kaltenbacha, L., Concin, H., Diem, G., Pfeiffer, K.P., Ulmer, H., The VHM PP Study Group. (2009) Use of penalized splines in extended Cox-type additive hazard regression to flexibly estimate the effect of time-varying serum uric acid on risk of cancer incidence: A prospective, population-based study in 78,850 men. *Annals of Epidemiology*, **19 (1)**, 15–24.