Convergent stochastic algorithm for parameter estimation in frailty models using integrated partial likelihood

Oodally Ajmal, Luc Duchateau, Estelle Kuhn

Convergent stochastic algorithm for parameter estimation in frailty models using integrated partial likelihood

Oodally Ajmal[1], Duchateau Luc[2], Estelle Kuhn[1]

[1] MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
[2]Ghent University, Faculty of Veterinary Medicine Department of Nutrition, Genetics and Ethology, Belgium.

Abstract

Frailty models are often the model of choice for heterogeneous survival data. A frailty model contains both random effects and fixed effects, with the random effects accommodating for the correlation in the data. Different estimation procedures have been proposed for the fixed effects and the variances of and covariances between the random effects. Especially with an unspecified baseline hazard, i.e., the Cox model, the few available methods deal only with a specific correlation structure. In this paper, an estimation procedure, based on the integrated partial likelihood, is introduced, which can generally deal with any kind of correlation structure. The new approach, namely the maximisation of the integrated partial likelihood, combined with a stochastic estimation procedure allows also for a wide choice of distributions for the random effects. First, we demonstrate the almost sure convergence of the stochastic algorithm towards a critical point of the integrated partial likelihood. Second, numerical convergence properties are evaluated by simulation. Third, the advantage of using an unspecified baseline hazard is demonstrated through application on cancer clinical trial data.

1 Introduction

Survival analysis consists in the analysis of the time of occurence of an event of interest. The Cox model introduced in [5] is often used in this area. It allows us to model the risk of occurence of the event of interest, also called hazard, as the product of a baseline hazard function and a function of the covariates. The regression coefficients are usually estimated by maximisation of the partial likelihood which does not depend on the baseline hazard

function. The good asymptotic properties of the estimator, namely the consistency, asymptotic normality and efficiency, based on partial likelihood are detailed and proved in [3]. However, an underlying assumption, called proportional hazard assumption, of this model is that the ratio of the hazards of two individuals is constant over time. This assumption is quite strong and often not fulfilled in practice due to a lack of homogeneity in real data. For example, in a clinical study, the data may be clustered into groups based on the location of the clinics or on different medical staves involved in collecting samples. Frailty models introduced by Vaupel et al in [22] allow to do away with this assumption by taking into account heterogeneity through non observed random effects. For more details on frailty models, we refer to [8], [23].

The literature on parameter estimation in frailty models is quite rich. Maximum likelihood estimation based on an Expected Maximization (EM) algorithm has been studied by [16] with frailties following a Gamma distribution in both non and semi-parametric models. The asymptotic properties of the maximum likelihood estimates with a plug-in estimator for the baseline hazard from a Gamma frailty model without covariates have been studied by [14], [15] and for a correlated Gamma frailty distribution by [17]. The choice of the Gamma distribution is here motivated by its mathematical convenience. Indeed, a closed form of the marginal likelihood can be calculated when the frailties are assumed to follow a Gamma distribution.

An approach based on the maximization of a penalized partial likelihood with a Laplace approximation of the marginal likelihood has been proposed by [18]. Also, [7] implemented an approach based on penalized partial likelihood using an iterative algorithm based on the marginal and penalized log likelihoods. A semi-parametric approach where the baseline hazard is estimated with a splines basis in the Gamma frailty model is implemented in the R package *frailtypack* developed by [19]. An estimation method based on the first and second order Laplace approximations of the complete partial likelihood has been proposed by [9] and implemented in the R package *frailtyHL*. However to the best of our knowledge, none of these existing algorithms has been proved to be convergent theoretically.

The aim of this paper is to propose an efficient stochastic algorithm to maximize the integrated partial likelihood and to prove its theoretical convergence property. We consider the criteria given by the integrated partial likelihood for the frailty model and the estimator that maximizes this criteria. We present an efficient stochastic EM algorithm to calculate its value. Then we establish its theoretical almost sure convergence to a critical point of the integrated partial likelihood. Moreover, we highlight the benefit of using the integrated partial likelihood through simulation studies and real data analysis.

The paper is organized as follows. Section 2 deals with the frailty model. The integrated partial likelihood and the estimator associated are presented

in Section 3. The algorithmic method for inference is presented in Section 4. An extended frailty model and the corresponding stochastic estimation procedure whose convergence property is established are detailed in Section 5. The simulation and real data studies are presented in Section 6. The paper ends with a conclusion and a discussion.

2 The Frailty Model

2.1 Description of the model

We consider a population of individuals clustered into N groups. We denote by n_i the size of the i-th group for $1 \leq i \leq N$. We denote the event time and censoring time of the individual j in group i by T_{ij} and C_{ij} respectively for $1 \leq i \leq N$ and $1 \leq j \leq n_i$. We observe the variable $X_{ij} = \min(T_{ij}, C_{ij})$ and the censoring indicator defined as $\Delta_{ij} = \mathbb{1}_{\{T_{ij} \leq C_{ij}\}}$. We denote by $\mathbf{X} = (X_{ij})_{1 \leq i \leq N, 1 \leq j \leq n_i}$ and by $\mathbf{\Delta} = (\Delta_{ij})_{1 \leq i \leq N, 1 \leq j \leq n_i}$ the observations.

We consider the following frailty model where the hazard for the individual j of group i is expressed as follows :

$$\forall t \geq 0 \; h_{ij}(t|b_i) = h_0(t) \; \exp(Z_{ij}^t \beta + W_{ij}^t b_i), \tag{1}$$

where $h_0(t)$ denotes the baseline hazard function at time t, Z_{ij} and W_{ij} the covariates of individual j of group i, $\beta \in \mathbb{R}^b$ the unknown regression parameter vector and $b_i \in \mathbb{R}^f$ the common frailty shared by individuals of group i. We assume that the probability density function of the unobserved frailty is parametric and denote by γ its parameter taking values in $\mathbb{R}^c$.

Therefore the model parameters are h_0, β and γ. The parameter of interest is usually β, enabling the quantification of the effects of the covariates which is often the main objective of real data analysis.

2.2 Assumptions on the model

We introduce the following usual assumptions on the frailty model:
(F1) The censoring times (C_{ij}) are independent of the event times (T_{ij}) and of the frailties (b_i).
(F2) Conditionally to the frailties (b_i), the event times (T_{ij}) are independent.
(F3) The frailties (b_i) are identically and independently distributed having common density g_γ.
(F4) The function h_0 belongs to the set of functions defined on $\mathbb{R}^+$ taking values in $\mathbb{R}^+$.
(F5) The probability density function of the frailties denoted by g_γ belongs to the set of curved exponential family of probability density functions where γ takes values in $\mathbb{R}^c$.

Remark 1 *We note here that **(F4)** is required only for the construction of the partial likelihood. The regularity condition is weaker than the one in [10] where a choice of parametric structure is made on the baseline hazard function.*

3 Estimation by maximisation of the integrated partial likelihood

We consider the criteria defined by the integrated partial likelihood for the frailty model following the idea of [5]. Then we define the estimator as the parameter value that maximises this criteria.

3.1 Definition of the integrated partial likelihood

Following the idea used in the Cox model, we consider the conditional partial likelihood defined as follows:

$$L^p_{\text{cond}}(\theta; \mathbf{X}, \mathbf{\Delta} \mid \mathbf{b}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} \left(\frac{\exp(Z_{ij}^t \beta + W_{ij}^t b_i)}{\sum_{(l,k) \in R(X_{(ij)})} \exp(Z_{lk}^t \beta + W_{lk}^t b_l)} \right)^{\Delta_{ij}} \tag{2}$$

where $\theta = (\beta, \gamma) \in \mathbb{R}^b \times \mathbb{R}^c$, $R(X_{(ij)}) = \left\{ 1 \leq l \leq N, 1 \leq k \leq n_l : X_{lk} \geq X_{(ij)} \right\}$ is the set of individuals still at risk at time $X_{(ij)}$ and $\mathbf{b} = (b_i)_{1 \leq i \leq N}$.

We then easily deduce the complete partial likelihood expression:

$$\begin{aligned} L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) &= \textstyle\prod_{i=1}^{N} g_\gamma(b_i) \\ &\times \textstyle\prod_{i=1}^{N} \prod_{j=1}^{n_i} \left(\frac{\exp(Z_{ij}^t \beta + W_{ij}^t b_i)}{\sum_{(l,k) \in R(X_{(ij)})} \exp(Z_{lk}^t \beta + W_{lk}^t b_l)} \right)^{\Delta_{ij}} \end{aligned} \tag{3}$$

We emphasize that this partial likelihood no longer involves the baseline h_0 as the partial likelihood in the Cox model.

Finally we define the integrated partial likelihood as defined in [21], also called marginal partial likelihood, obtained by integrating the complete partial likelihood over the unobserved frailties $\mathbf{b}$:

$$L^p_{\text{marg}}(\theta; \mathbf{X}, \mathbf{\Delta}) = \int L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) d\mathbf{b} \tag{4}$$

Remark 2 *We recall that as in the Cox model, this integrated partial likelihood is not a likelihood function, but acts as one as explained in [18].*

3.2 Definition of the maximum integrated partial likelihood estimate

Following [21], we define the estimator $\hat{\theta}$ for the parameters vector as the value that maximizes the integrated partial likelihood:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L^p_{\text{marg}}(\theta; \mathbf{X}, \mathbf{\Delta}) \tag{5}$$

If there exists an analytical expression of the integrated partial likelihood, it can be maximized directly. When the computation of the integrated partial likelihood is difficult, an EM type algorithm can be implemented for the maximization procedure. Therefore, we propose to calculate $\hat{\theta}$ by using a stochastic Expectation Maximization (EM) algorithm following in the footsteps of [10].

4 Algorithmic methods for inference

4.1 Description of the algorithm for parameter estimation

We consider the stochastic EM algorithm introduced by [11] to evaluate the estimator of the parameters defined in Equation (5). It is an extension of the stochastic approximation EM algorithm developed by [6] where the EM algorithm is coupled with a Markov Chain Monte Carlo (MCMC) procedure to simulate the unobserved frailties.

Each iteration of the algorithm is composed of three steps detailed below. We start with initial values θ_0,$\mathbf{b}_0$ and Q_0 arbitrarily chosen.

Repeat until convergence for $k \geq 1$:

1. **Simulation step**: draw realizations $\mathbf{b}_k$ of the unobserved frailties

$$\mathbf{b}_k \sim \Pi_{\theta_{k-1}}(\mathbf{b}_{k-1}, .)$$

where $\Pi_{\theta_{k-1}}$ is a transition probability of a convergent Markov chain having as stationary distribution the conditional distribution $\pi^p_{\theta_{k-1}}(.|\mathbf{X}, \mathbf{\Delta})$ defined by

$$\pi^p_\theta(\mathbf{b}|\mathbf{X}, \mathbf{\Delta}) = \frac{L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b})}{\int L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) d\mathbf{b}}$$

2. **Stochastic approximation step**: compute for all θ

$$Q_k(\theta) = Q_{k-1}(\theta) + \mu_k \left(\log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}_k) - Q_{k-1}(\theta)\right) \tag{6}$$

where the sequence (μ_k) satisfies

$$0 \leq \mu_k \leq 1,\ \sum \mu_k = +\infty,\ \sum \mu_k^2 < +\infty$$

3. **Maximisation step**: update the parameter estimate

$$\theta_k = \underset{\theta}{\operatorname{argmax}}\ Q_k(\theta)$$

This algorithm will be called Algorithm 1 below. Further practical details on the algorithm, namely the simulation procedure used to sample the realizations of the unobserved frailties $\mathbf{b}_k$, the computation of the quantity Q_k and the update of parameter estimates θ_k can be found in Appendices A and B. The choice of the stepsize sequence $(\mu_k)_k$ is detailed in Section 6.

4.2 Estimation of the Fisher Information Matrix

We consider the usual estimate of the Fisher Information Matrix, namely the observed Fisher information matrix $I_{obs}(\theta) = -\partial^2_\theta \log L^p_{\text{marg}}(\theta; \mathbf{X}, \mathbf{\Delta})$ (see [2]). Using Louis's missing information principle (see [12]), we express the matrix $I_{obs}(\theta)$ as:

$$I_{obs}(\theta) = -\mathbb{E}_\theta\big(\partial^2_\theta \log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) \mid \mathbf{X}, \mathbf{\Delta}\big) - \operatorname{Cov}_\theta\big(\partial_\theta \log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) \mid \mathbf{X}, \mathbf{\Delta}\big)$$

where $\mathbb{E}_\theta$ and $\operatorname{Cov}_\theta$ denote respectively the expectation and the covariance under the posterior distribution π^p_θ of the frailty.

We approximate the quantity $I_{obs}(\theta)$ by a Monte Carlo sum based on the realizations of the Markov chain generated in the algorithm having as stationary distribution the posterior distribution π^p_θ. After a burn-in period, we use the remaining M realizations $(\mathbf{b}_m)_{1 \le m \le M}$ of the Markov chain to compute the following quantity:

$$\begin{aligned}\hat{I}_M(\hat{\theta}) = &-\frac{1}{M}\sum_{m=1}^{M} \partial^2_\theta \log L^p(\hat{\theta}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}_m) \\ &- \frac{1}{M}\sum_{m=1}^{M} \big(\partial_\theta \log L^p(\hat{\theta}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}_m) \times \partial_\theta \log L^p(\hat{\theta}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}_m)^t\big) \\ &+ \frac{1}{M^2}\Big(\sum_{m=1}^{M} \partial_\theta \log L^p(\hat{\theta}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}_m)\Big) \times \Big(\sum_{m=1}^{M} \partial_\theta \log L^p(\hat{\theta}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}_m)\Big)^t\end{aligned}$$

The ergodic theorem in [13] guarantees the convergence of the quantity $\hat{I}_M(\hat{\theta})$ to the observed Fisher information matrix $I_{obs}(\hat{\theta})$ as M goes to infinity.

5 Extended frailty model and convergent estimation algorithm

Most of the theoretical convergence properties of stochastic EM like algorithms have been established in the case of curved exponential families as for examples in [6, 11, 1]. Since the complete partial likelihood defined in (3) does not belong to the curved exponential family of probability density functions, we introduce an extended frailty model.

5.1 Extended frailty model

We consider an extended frailty model where the regression parameter β is considered as a population random variable. The extended latent variables are denoted by $\xi = (b_i, i = 1, \ldots, n, \beta)$. Moreover we assume that the population variable β follows a Gaussian distribution with unknown expectation β_0 and fixed variance σ_β^2. We denote the new parameters to be estimated by $\eta = (\beta_0, \gamma)$. The complete likelihood corresponding to the model can be written as follows:

$$\begin{aligned} L^e(\eta; \mathbf{X}, \mathbf{\Delta}, \xi) &= \prod_{i=1}^{N} g_\gamma(b_i) f_{\beta_0}(\beta) \\ &\prod_{i=1}^{N} \prod_{j=1}^{n_i} \left(\frac{\exp(Z_{ij}^t \beta + W_{ij}^t b_i)}{\sum_{(l,k) \in R(X_{(ij)})} \exp(Z_{lk}^t \beta + W_{lk}^t b_l)} \right)^{\Delta_{ij}} \end{aligned} \tag{7}$$

where f stands for the Gaussian probability density function. This likelihood function belongs to the curved exponential family as soon as the frailty probability density function g_γ belongs to the curved exponential family. Sufficient statistics are explicit and can be expressed as $S(\xi) = \left(\sum_{i=1}^{N} S_f(b_i), \beta \right)$ where $S_f(b_i)$ are sufficient statistics corresponding to the frailties (b_i).

By assumption **(F5)**, the complete partial likelihood defined in (7) can be written as follows:

$$L^e(\eta; \mathbf{X}, \mathbf{\Delta}, \xi) = \exp(-\Psi(\eta) + \langle S(\xi), \Phi(\eta) \rangle)$$

where S, Ψ and Φ are Borel functions .

5.2 Description of the stochastic EM Algorithm with truncation on random boundaries

Following [1], we detail a seoncd algorithm, called Algorithm 2, based on the extended likelihood.

Let $(\mathcal{K}_q)_q \geq 0$ be a sequence of increasing compact subsets of S such as $\bigcup_{q\geq 0} \mathcal{K}_q = S$ and $\mathcal{K}_q \subset \text{int}(\mathcal{K}_{q+1})$, $\forall q \geq 0$.

Initialize η_0 in Θ, ξ_0 and s_0 in two fixed compact sets $\mathbf{K}$ and $\mathcal{K}_0$ respectively.

Repeat until convergence for $k \geq 1$:

1. **Simulation step:** Draw $\bar{\xi}$ from a kernel $\Pi_{\eta_{k-1}}$ of a convergent Markov chain having as stationary distribution the conditional distribution with the current parameters:

$$\bar{\xi} \sim \Pi_{\eta_{k-1}}(\xi_{k-1}, .)$$

2. **Stochastic approximation step:** Compute

$$\bar{s} = s_{k-1} + \mu_k(S(\bar{\xi}) - s_{k-1})$$

3. **Truncation step:** If $\bar{s}$ is outside the current compact set $\mathcal{K}_{\kappa_{k-1}}$, where κ is the index of the current active truncation set, or too far from the previous value s_{k-1} then restart the stochastic approximation in the initial compact set, extend the truncation boundary to $\mathcal{K}_{\kappa_k}$ and start again with a bounded value of the missing variable. Otherwise, set $(\xi_k, s_k) = (\bar{\xi}, \bar{s})$ and keep the truncation boundary to $\mathcal{K}_{\kappa_{k-1}}$.

4. **Maximization step:**

$$\eta_k = \underset{\eta}{\text{argmax}}\{-\Psi(\eta) + \langle s_k, \Phi(\eta)\rangle\}$$

In this second algorithm, we construct a sequence (ξ_k, s_k) while satisfying two conditions at each iteration k. Namely we check whether the stochastic approximation wanders outside the current compact set and whether the current value is not too far from the previous value. The latter can be expressed as follows:

$$\|s_k - s_{k-1}\| \leq \epsilon_k$$

where $\epsilon = (\epsilon_k)_{k\geq 0}$ is a monotone non-increasing sequence of positive numbers. A more detailed description of the truncation step can be found in [4].

5.3 Convergence property of Algorithm 2 in the extended frailty model

We consider classical assumptions required to prove the convergence of EM like algorithms as following those of [6].

(M3) The function $\bar{s} : \Theta \to S$ defined as:

$$\bar{s}(\eta) = \int_{\mathbb{R}^l} |S(\xi)| \pi_s^e(\xi) d\xi$$

where

$$\pi_s^e(\xi) = \frac{\exp(-\Psi(\eta(s)) + \langle s, \Phi(\eta(s))\rangle)}{\int \exp(-\Psi(\eta(s)) + \langle s, \Phi(\eta(s))\rangle) d\xi}$$

is continuously differentiable on Θ.

(M4) The function $l^e : \Theta \to \mathbb{R}$ defined as the marginal extended log-likelihood

$$l^e(\eta) = \log \int_{\mathbb{R}^l} L^e(\eta; \mathbf{X}, \mathbf{\Delta}, \xi) d\xi$$

is continuously differentiable on Θ and

$$\partial_\eta \int_{\mathbb{R}^l} L^e(\eta; \mathbf{X}, \mathbf{\Delta}, \xi) d\xi = \int_{\mathbb{R}^l} \partial_\eta \ L^e(\eta; \mathbf{X}, \mathbf{\Delta}, \xi) d\xi$$

(M5) There exists a function $\hat{\eta} : S \to \Theta$ s.t:

$$\forall \ s \in S, \forall \ \eta \in \Theta, \ L(\hat{\eta}(s); s) \geq L(\eta; s)$$

where $L : S \times \Theta \to \mathbb{R}$ is defined as

$$L(\eta; s) = \exp(-\Psi(\eta) + \langle s, \Phi(\eta)\rangle) \tag{8}$$

Moreover, the function $\hat{\eta}$ is continuously differentiable on S.

Following in the lines of [4], we state a first assumption **(A1')** that guarantees the existence of a global Lyapunov function denoted by w defined as:

$$w(s) = -\log \int L^e(\hat{\eta}(s); \mathbf{X}, \xi) d\xi \tag{9}$$

for the mean field h defined as:

$$h(s) = \int (S(\xi) - s) \pi_s^e(\xi) d\xi \tag{10}$$

(A1') The functions w and h are such that

(i) there exists an $M_0 > 0$ such that

$$\mathcal{S} \stackrel{\Delta}{=} \{s \in S, \langle \nabla w(s), h(s)\rangle = 0\} \subset \{s \in S, w(s) < M_0\}$$

where w is defined in (9) and h is defined in (10).

(ii) there exists $M_1 \in]M_0, \infty]$ such that $\{s \in S, w(s) < M_1\}$ is a compact set.

(iii) the closure of $w(\mathcal{L})$ has an empty interior.

(A4) The sequences $\mu = (\mu_k)_{k\geq 0}$ and $\epsilon = (\epsilon_k)_{k\geq 0}$ are non-increasing, positive and satisfy $\sum_{k=0}^{\infty} \mu_k = \infty$, $\inf_{k\to\infty} \epsilon_k = 0$ and $\sum_{k=1}^{\infty} \{\mu_k^2 + \mu_k \epsilon_k^a + (\mu_k \epsilon_k^{-1})^p\} < \infty$, where $a \in]0, 1]$ and $p \geq 2$.

Finally we consider the usual drift assumption **(DRI)** which are detailed in [4].

Theorem 1 *Assume that **(F1-F5)**, **(M3-M5)**, **(A1')**, **(A4)** and **(DRI)** are fulfilled. Then we have with probability 1*

$$\lim_{k\to\infty} d(\eta_k, \mathcal{L}) = 0$$

where $(\eta_k)_k$ is generated by Algorithm 2, $d(x, A)$ denotes the distance from x to any closed subset A and $\mathcal{L} = \{\eta \in \Theta, \partial_\eta \ log \ L^e_{marg}(\eta; \mathbf{X}, \mathbf{\Delta}) = 0\}$.

The assumption **(A1')** corresponds to the assumptions **(A1)** (i), (ii), (iv) of [4] respectively. Assumption **(A4)** deals with the conditions on the step-size sequences involved in the stochastic approximation and truncation steps of Algorithm 2.

Proof of Theorem 1: We will first apply Theorem 5.5 of of [4] to prove the convergence of the sequence (s_k) and checked therefore the assumptions required. To prove that assumption **(A1)(iii)** of [4] is fulfilled in our case, we establish the following lemma following the lines of the proof of Lemma 2 of [6] using in our case the partial likelihood instead of the likelihood:

Lemma 2 *Assuming **(F1–F5)** and **(M3-M5)**, we have for any $s \in S \setminus \mathcal{S}$ $\langle \nabla w(s), h(s) \rangle < 0$*

Proof of Lemma 2 Assumption **(M1)** of [6] is implied by **(F5)**. To fulfill assumption **(M2)** of [6], it suffices to show that Ψ and Φ are twice continuously differentiable. This is a straight consequence of assumptions **(F1–F5)**. The end of the proof follows the same lines as Lemma 2 of [6].

Thereby assumption **(A1)(iii)** of [4] is fulfilled in our case. As detailed in [4], assumptions **(DRI)** imply assumptions **(A2-A3)** by Proposition 6.1. Thus we can apply Theorem 5.5 of [4]. We get that the sequence (s_k) generated by Algorithm 2 satisfayes $\lim_k d(s_k, \mathcal{S}) = 0$. Following the lines of the proof of Lemma 2 of [6], we get that $\lim_k d(\eta_k, \mathcal{L}) = 0$. The proof of Theorem 1 is therefore complete.

6 Numerical studies

All numerical studies have been done using R version 3.3.1 on an Intel Core i7-8550U CPU @ 1.99 GHz, 16 GB RAM.

The aim of our numerical experiments is to compare the performances of the Maximum Integrated Partial Likelihood (MIPL) estimator defined in Section 3.2 to those of other estimators existing in the literature. We also analyse a real dataset of bladder cancer.

We run both algorithms in the numerical studies. Since we get results of the same order, we only present the ones obtained using Algorithm 1, the main motivation of the extended model and of Algorithm 2 being the theoretical convergence result.

1. The decreasing positive step size (μ_k) is taken as follows:

$$\begin{aligned} \forall k > K_0, \quad \mu_k &= \frac{1}{(k - K_0)} \\ \forall k \leq K_0, \quad \mu_k &= 1 \end{aligned}$$

   where K_0 is a number to be specified. The algorithm is said to have no memory during the first K_0 iterations. After this burn-in time which allows for the algorithm to visit the parameter space, the sequence $(\mu_k)_k$ decreases and converges to zero as $k \to \infty$.

2. The transition kernel used for simulating the unobserved frailty is chosen as a transition kernel of a Metropolis Hastings algorithm with proposal distribution q equal to a Gaussian distribution centered at the current value $\mathbf{b_{k-1}}$ at the k^{th} iteration.

3. We define a stopping criterion based on the relative difference between the values of the parameters for two consecutive iterations. Let us fixed a positive threshold $\epsilon > 0$. If for some $k > 1$:

$$\frac{\|\theta_k - \theta_{k-1}\|}{\|\theta_{k-1}\|} < \epsilon$$

   holds true for for three consecutive iterations, the algorithm is stopped. We set $\epsilon = 10^{-4}$ in the simulation study.

6.1 Simulated data

We consider the following setting. The frailties (b_i) are drawn from a centered normal distribution with variance $\gamma = 0.7$. The regression parameter

β used to simulate the data is chosen equal to the vector $(2, 3)$ of size 2. The covariates $((W_{ij}, Z_{ij}))$ are generated independently according to a Bernoulli distribution. We consider varying number N of clusters and $n_i = 4$ observations per cluster.

6.1.1 Study of the consistency property of $\hat{\theta}$

We begin by studying numerically the consistency of the estimate $\hat{\theta}$. The Weibull baseline hazard defined as $h_0(t) = \lambda\rho t^{\rho-1}$ for $t > 0$ is considered in this section using the parameter values $\lambda = 0.01$ and $\rho = 1.5$. There is no censoring.

$$\forall t \geq 0 \; h_{ij}(t|b_i) = \lambda\rho t^{\rho-1} \; \exp(Z_{ij}^t\beta + b_i), \quad \lambda > 0, \rho > 0 \tag{11}$$

The estimate $\hat{\theta}$ is evaluated using the algorithm described in Section 4.1.

Table 1: Mean of parameter estimates $\hat{\theta}$ and standard deviation in parenthesis obtained from 500 repetitions with different group sizes. The number of observations per group is fixed at $n_i = 4$.

Parameters	True values	N=10	N=20	N=50
β_1	2	1.794 (0.385)	1.996 (0.353)	2.002 (0.320)
β_2	3	2.652 (0.427)	2.995 (0.390)	2.999 (0.339)
γ	0.7	0.490 (0.656)	0.649 (0.477)	0.707 (0.287)

The results supporting the numerical consistency of $\hat{\theta}$ are displayed in Table 1. N refers to the number of groups. As the the number of groups N progressively increases, the corresponding estimates get closer to the true values and the standard deviation decreases.

6.1.2 Comparing the maximum integrated partial likelihood estimate with a parametric estimate

We consider a parametric estimate defined in the model with a Weibull baseline hazard function defined as $h_0(t) = \lambda\rho t^{\rho-1}$, $\lambda > 0, \rho > 0$. We denote the vector of parameters by $\theta_{\text{weibull}} = (\lambda, \rho, \beta, \gamma)$. The expression of the complete likelihood is given by:

$$L^{\text{weibull}}(\theta_{\text{weibull}}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) = \prod_{i=1}^{N} g_\gamma(b_i) \prod_{i=1}^{N} \prod_{j=1}^{n_i} \big(\lambda\rho X_{ij}^{\rho-1} \times \exp(Z_{ij}^t\beta + b_i)\big)^{\Delta_{ij}} \exp(-\lambda X_{ij}^{\rho} \exp(Z_{ij}^t\beta + b_i)) \tag{12}$$

The marginal likelihood is obtained by integrating over the frailties b :

$$L_{\text{marg}}^{\text{weibull}}(\theta; \mathbf{X}, \mathbf{\Delta}) = \int L^{\text{weibull}}(\theta_{\text{weibull}}; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) d\mathbf{b} \tag{13}$$

We denote by $\hat{\theta}_{\text{weibull}}$ the estimator of the maximum of the marginal likelihood :

$$\hat{\theta}_{\text{weibull}} = \underset{\theta_{\text{weibull}}}{\text{argmax}}\ L_{\text{marg}}^{\text{weibull}}(\theta_{\text{weibull}}; X, \Delta) \tag{14}$$

The value of $\hat{\theta}_{\text{weibull}}$ is computed using the MCMC-SAEM algorithm proposed in [10]. The event times are first simulated according to (11) with Weibull parameters $\lambda = 0.01$ and $\rho = 1.5$. The number of groups N is fixed at a value of 250. There is no censoring. The results are presented in Table 2. We conclude that both methods give good estimates in this example.

We then consider event times simulated from the model using a Gompertz baseline hazard function. The modeling equation is as follows:

$$h_{ij}(t|b_i) = \lambda \exp(\alpha t)\ \exp(Z_{ij}^t\beta + b_i), \quad \lambda > 0, \alpha > 0 \tag{15}$$

The event times are simulated according to (15) with Gompertz parameters $\lambda = 0.08$ and $\alpha = 2$. There is no censoring. The results are presented in Table 3. The estimate $\hat{\theta}$ which does not require any modeling assumption of h_0 proves to be a good estimator where as $\hat{\theta}_{\text{weibull}}$ does not give good results as it can be seen in Table 3. The wrong specification of h_0 for the latter introduces bias in the estimation of the parameters. These results therefore show the advantages of not having to model the baseline hazard h_0 in the estimation procedure.

6.1.3 Comparison of the maximum integrated partial likelihood estimate with other estimates

The aim of the following simulation study is to compare the performances of the maximum integrated partial likelihood estimate with those of other estimates. Therefore we consider the estimate based on penalized partial likelihood implemented in the R package *coxme* based on [18] and two estimates based on the h-likelihood implemented in the R package *frailtyHL* detailed in [9].

Table 2: Mean of parameter estimates and model-based standard error in parentheses obtained from 500 repetitions with the event times following a Weibull distribution to compare the parametric estimate to the partial integrated likelihood estimate. $N = 250$ and $n_i = 4$.

Method	β_1	β_2	γ
True values	2	3	0.7
$\hat{\theta}$	2.033 (0.133)	3.056 (0.121)	0.702 (0.106)
$\hat{\theta}_{\text{weibull}}$	1.982 (0.133)	2.944 (0.145)	0.701 (0.111)

Table 3: Mean of parameter estimates and model-based standard error in parentheses obtained from 500 repetitions with the event times following a Gompertz distribution to compare the parametric estimate to the integrated partial likelihood estimate. $N = 250$ and $n_i = 4$.

Method	β_1	β_2	γ
True values	2	3	0.7
$\hat{\theta}$	2.031 (0.126)	3.041 (0.134)	0.732 (0.129)
$\hat{\theta}_{\text{weibull}}$	1.380 (0.112)	2.029 (0.146)	0.270 (0.126)

The estimation in the *coxme* package is based on the maximisation of a penalized partial likelihood. This estimator is denoted by $\hat{\theta}_{\text{coxme}}$ later.

The h-likelihood methods implemented in *frailtyHL* are based on a Laplace approximation of the marginal partial likelihood which is then maximised. The two estimators based on h-likelihood chosen differ in the order of the Laplace approximations. They are denoted by $\hat{\theta}_{\text{HL(0,1)}}$ and $\hat{\theta}_{\text{HL(1,2)}}$ with the first one based on the first order Laplace approximation and the second one based on the second order Laplace approximation.

6.1.4 Effect of censoring level on parameter estimation

We first investigate the effect of censoring when comparing the different estimation procedures. We recall that there was no censoring in the previous simulation settings. The event times were simulated according to (11) with Weibull parameters $\lambda = 0.01$ and $\rho = 1.5$. The number of groups N is fixed at a value of 250. Data are simulated under two different censoring levels, low (Table 4) and moderate (Table 5).

In Table 4, in the low censoring level case, we observe that the MIPL estimate $\hat{\theta}$ and the estimate $\hat{\theta}_{\text{HL(1,2)}}$ seem to be closer to the true values as opposed to the estimates $\hat{\theta}_{\text{coxme}}$ and $\hat{\theta}_{\text{HL(0,1)}}$. We make the same observation in Table 5 with the moderate censoring. It seems that $\hat{\theta}$ and $\hat{\theta}_{\text{HL(1,2)}}$ have the same performance level in the estimation of β and give better estimates than $\hat{\theta}_{\text{coxme}}$ and $\hat{\theta}_{\text{HL(0,1)}}$. We note however that $\hat{\theta}$ gives slightly better estimates than $\hat{\theta}_{\text{HL(1,2)}}$ for the variance γ for both low and moderately censored settings.

6.1.5 Robustness to misspecification of the frailty distribution

We investigate in this section the case where the frailty distribution is misspecified in the estimating procedure. For example, assuming a normal frailty as done previously when the frailties instead follow a mixture of normal distributions might introduce bias in the estimates. We study the effects of a misspecification of the frailty distribution on the four estimators presented above. We first consider data simulated with a multiplicative Gamma frailty term. We observe that all estimating procedures give good estimations when a normal frailty is assumed for the estimation task (results non presented). Then we consider frailties drawn from a mixture of normal distributions as follows:

$$b \sim \frac{1}{2}\mathcal{N}(-10, 2) + \frac{1}{2}\mathcal{N}(10, 2)$$

Table 4: Mean of parameter estimates and model-based standard error in parentheses obtained from 500 repetitions with the event times following a Weibull distribution and a low censoring level of 20 %. Comparsion of MIPL with *coxme* and *frailtyHL*. $N = 250$ and $n_i = 4$.

Method	β_1	β_2	γ
True Values	2	3	0.7
$\hat{\theta}$	1.968 (0.123)	2.968 (0.156)	0.672 (0.116)
$\hat{\theta}_{\text{coxme}}$	1.922 (0.120)	2.901 (0.151)	0.606 (0.107)
$\hat{\theta}_{\text{HL(0,1)}}$	1.930 (0.118)	2.939 (0.155)	0.607 (0.107)
$\hat{\theta}_{\text{HL(1,2)}}$	1.954 (0.120)	2.976 (0.158)	0.647 (0.117)

Table 5: Mean of parameter estimates and model-based standard error in parentheses obtained from 500 repetitions with the event times following a Weibull distribution and a moderate censoring level of 40 %. Comparsion of MIPL with *coxme* and *frailtyHL*. $N = 250$ and $n_i = 4$.

Method	β_1	β_2	γ
True Values	2	3	0.7
$\hat{\theta}$	1.896 (0.133)	2.859 (0.153)	0.641 (0.120)
$\hat{\theta}_{\text{coxme}}$	1.850 (0.125)	2.791 (0.149)	0.575 (0.084)
$\hat{\theta}_{\text{HL(0,1)}}$	1.847 (0.125)	2.808 (0.150)	0.576 (0.113)
$\hat{\theta}_{\text{HL(1,2)}}$	1.873 (0.126)	2.846 (0.151)	0.615 (0.121)

Table 6: Mean of parameter estimates and model-based standard error in parentheses obtained from 500 repetitions with the event times following a Weibull distribution. Comparison of MIPL estimate with *coxme* and *frailtyHL* estimates when the frailty distribution is misspecified. A mixture of Gaussian frailties is used to simulate the dataset whereas a Gaussian frailty is assumed in the estimation procedure. $N = 250$ and $n_i = 4$.

Method	β_1	β_2	γ
$\hat{\theta}$	2.036 (0.163)	3.040 (0.201)	25.5 (0.743)
$\hat{\theta}_{\text{coxme}}$	1.531 (0.124)	2.304 (0.133)	6.079 (0.566)
$\hat{\theta}_{\text{HL(1,2)}}$	2.022 (0.110)	3.019 (0.128)	23.0 (2.96)

In all estimating procedures, a normal frailty is assumed. The event times were simulated according to (11) with Weibull parameters $\lambda = 0.01$ and $\rho = 1.5$. The number of groups N is fixed at a value of 250 and there are 4 observations per cluster. All event times are non-censored. The results are presented in Table 6. We observe that the estimates obtained with our method denoted by $\hat{\theta}$ and with *frailtyHL* are close to the true value where as the one obtained by *coxme* does not adjust well to the misspecification of the frailty distribution leading to some bias in the estimation of β in this example.

6.1.6 Correlated frailties

In all of the previous simulation studies, the shared frailty model with the frailty acting only on the group level has been considered. In order to apply the MIPL estimating procedure on a real cancer dataset detailed in Section 6.2, we consider the modeling of the hazard function as follows:

$$h_{ij}(t|b_i) = h_0(t)\ \exp(b_{0i} + Z_{ij}^t(\beta + b_{1i})) \tag{16}$$

with $b = (b_0, b_1) \sim \mathcal{N}(0, \Sigma)$ where $\Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}$

We estimate the parameters $\theta = (\beta, \sigma_0^2, \sigma_1^2, \sigma_{01})$ by maximising the integrated partial likelihood. The event times are simulated following (16) with a Weibull baseline hazard parametrized by $\lambda = 0.01$, $\rho = 1.5$ and the regression paramater $\beta = (2, 3)$. The frailty variances σ_0^2 and σ_1^2 are set to 0.8 and 0.4 respectively and the covariance term σ_{01} is set to 0.226. The number of

Table 7: Mean of parameter estimates and standard deviation of estimates obtained from 500 repetitions using Weibull baseline with $\lambda = 0.01$ and $\rho = 1.5$ in the corrrelated frailty model.

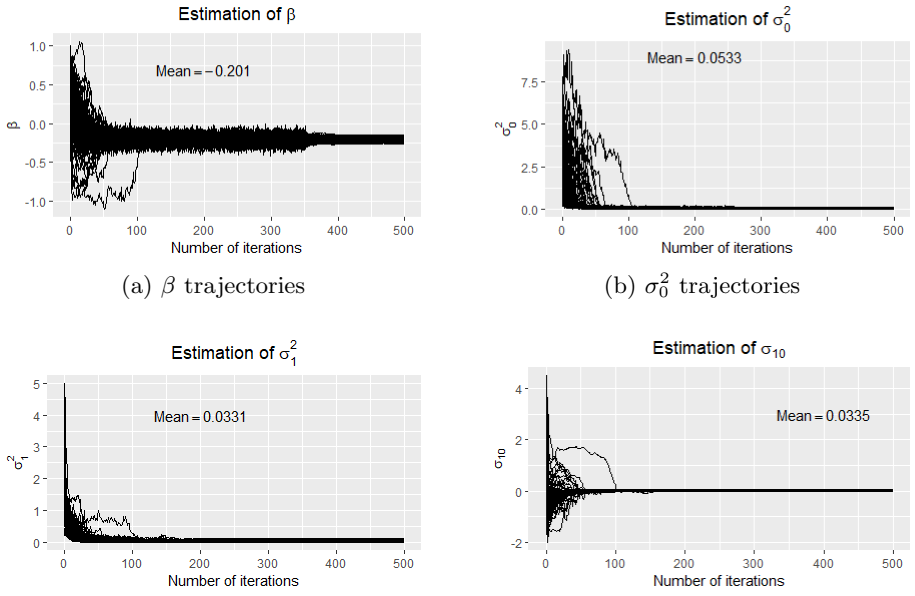
Method	β_1	β_2	σ_0^2	σ_1^2	σ_{01}
True Values	2	3	0.8	0.4	0.226
$\hat{\theta}$	2.020 (0.009)	3.016 (0.010)	0.805 (0.008)	0.403 (0.006)	0.218 (0.00004)

observations per group is not the same for all groups. The group sizes are fixed so as to be close to the configuration of the groups in the real dataset. The results are presented in Table 7. We observe that the estimate $\hat{\theta}$ is close to the true values for all the parameters. The model standard errors based on the estimation of the observed Fisher information matrix are also very small. This conclusive simulation study allows for the estimating procedure to be applied for analysing the real dataset.

6.2 Real data analysis

We consider a bladder cancer dataset from the EORTC. A combined analysis was carried out of individual patient data from 2596 superficial bladder cancer patients included in seven European Organization for Research and Treatment of Cancer trials 30781, 30782, 30791, 30831, 30832, 30845, and 30863 (Genito-Urinary tract cancer Group). Only the groups with more than 20 patients were included in the dataset to be analyzed. After data processing, we are left with 39 groups of patients of different sizes. The censoring level is about 51 % and about 80 % of the patients follow an intravesical treatment (see [20]) which is the only covariate considered. The studies conducted on this dataset suggests that the treatment effect b_{1i} might be correlated to the center effect b_{0i}. This leads us to model the hazard function as detailed in (16). We estimate the parameters $\theta = (\beta, \sigma_0^2, \sigma_1^2, \sigma_{01})$ by maximising the integrated partial likelihood. We run the algorithm using a grid of initial values and the mean of the obtained estimates is computed. The results are then compared with the estimates, which we denote by $\hat{\theta}_{cst}$, obtained in [10] where a constant baseline hazard is assumed.

The trajectories of all parameters estimated are shown in Figure 1. We observe that whatever the initial conditions, all trajectories seem to lead to more or less the same values. The algorithm is therefore not sensible to initial conditions. The estimates obtained in [10] are presented in the second column of Table 7. The estimates obtained with $\hat{\theta}$ and $\hat{\theta}_{cst}$ are not close, especially the parameter of interest β and the variance σ_1^2 which takes into account the effect of the treatment. Thus, in this example, choosing a



(a) β trajectories

(b) σ_0^2 trajectories

(c) σ_1^2 trajectories

(d) σ_{01} trajectories

Figure 1: Integrated partial likelihood estimates for EORTC bladder cancer dataset.

parametric constant form for the baseline affects strongly the estimation of the parameter of interest β, leading to possible wrong interpretation of the effect of the covariates. On the other hand, the estimate $\hat{\theta}$ does not rely on any parametric assumption on the baseline, leading to a robust estimation procedure with respect to any parametric choice for the baseline.

7 Conclusion and discussion

We consider as estimation criteria the integrated partial likelihood and the corresponding estimate which maximizes this criteria. The main advantage of this criteria is that it does not depend on any choice of the baseline function. We propose a stochastic approximation EM algorithm coupled with a MCMC procedure for calculating the parameter estimates in practice. The almost sure convergence of this algorithm to a critical point of the integrated partial likelihood is established under classical assumptions. We then validate the performance of the estimation procedure through simulation studies which highlight the good properties for finite sample size. In cases where the baseline hazard function is misspecified, the proposed estimate performs better than the parametric one. When the hazard function is correctly specified, we perform just as good. The proposed estimate called MIPL is then compared to existing estimates in the literature namely

Table 8: Mean of the estimates and mean model-based standard error in parentheses obtained with $\hat{\theta}_p$ and $\hat{\theta}_{cst}$ on the EORTC bladder cancer dataset.

Parameters	$\hat{\theta}_p$	$\hat{\theta}_{cst}$
β	-0.206 (0.007)	-0.254 (0.070)
σ_0^2	0.0712 (0.0001)	0.0306 (0.0002)
σ_1^2	0.0435 (0.0002)	0.107 (0.0006)
σ_{01}	0.0428 (0.000001)	0.0573 (0.000003)

given by the *coxme* and *frailtyHL* packages where the intractable integral is approximated through a Laplace approximation. The robustness of all estimates to a misspecification of the frailty distribution is analyzed. The simulation setting also takes into account light to heavy censoring to see how the different estimates perform. Finally, we analyse a real bladder cancer dataset and compare our results with a parametric estimating procedure from the literature.

Since we have proposed an efficient convergent algorithm to compute the MIPL estimate, it would be now of great interest to study its asymptotic properties as consistency, asymptotic normality and efficiency.

A Description of the simulation procedure used to sample realizations for the unobserved frailties

We usually construct Π_θ as a step of a Metropolis Hastings algorithm with proposal distribution q. Sample a candidate b^c:

$$b^c \sim q(.|b_{k-1}; \theta_{k-1})$$

We then calculate the acceptance ratio :

$$\alpha(b_{k-1}, b^c) = \min\left(1, \frac{\pi_{\theta_{k-1}}(b^c|\mathbf{X}, \boldsymbol{\Delta}) q(b_{k-1}|b^c; \theta_{k-1})}{\pi_{\theta_{k-1}}(b_{k-1}|\mathbf{X}, \boldsymbol{\Delta}) q(b^c|b_{k-1}; \theta_{k-1})}\right)$$

The simulated candidate is accepted with probability $\alpha(b_{k-1}, b^c)$.

$$b_k = \begin{cases} b^c & \text{with probability } \alpha(b_{k-1}, b^c) \\ b_{k-1} & \text{otherwise} \end{cases}$$

B Estimation equations for the MIPL estimate

We can rewrite (6) for an easier computation of the derivatives. By induction on k, we obtain:

$$Q_k(\theta) = Q_0 \prod_{i=1}^{k}(1-\mu_i) + \sum_{i=1}^{k}\Bigg(\mu_i \log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, b_i) \times \prod_{\substack{j=i+1 \\ i<k}}^{k}(1-\mu_j)\Bigg)$$

This expression of $Q_k(\theta)$ makes the computation of the derivative with respect to θ relatively straightforward :

$$\frac{\partial Q_k(\theta)}{\partial \theta} = \sum_{i=1}^{k}\left(\mu_i \frac{\partial \log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, b^i)}{\partial \theta} \prod_{\substack{j=i+1 \\ i<k}}^{k}(1-\mu_j)\right)$$

The log complete partial likelihood required to compute the quantity Q can be expressed as follows:

$$\log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b}) = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\Delta_{ij}\Bigg(Z_{ij}^t\beta + W_{ij}^t b_i - \log\Bigg(\sum_{(l,k)\in R(X_{(ij)})}\exp(Z_{lk}^t\beta + W_{lk}^t b_l)\Bigg)\Bigg) + \sum_{i=1}^{N}\log(g_\gamma(b_i)) \tag{17}$$

Differentiating (17) with respect to β, we obtain the following equations :

$$\frac{\partial \log L^p(\theta; \mathbf{X}, \mathbf{\Delta}, \mathbf{b})}{\partial \beta} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\Delta_{ij}\Bigg(Z_{ij} - \frac{\sum_{(l,k)\in R(X_{(ij)})} Z_{lk}\exp(Z_{lk}^t\beta + W_{lk}^t b_l)}{\sum_{(l,k)\in R(X_{(ij)})}\exp(Z_{lk}^t\beta + W_{lk}^t b_l)}\Bigg)$$

The parameter γ is easily updated as it is found only in the last term of (3). In many cases, we can update by direct computation the parameter estimate of γ and use a classic gradient descent algorithm to update β.

References

[1] S. Allassonnière, E. Kuhn, and A. Trouvé. Construction of bayesian deformable models via a stochastic approximation algorithm: A convergence study. *Bernoulli*, 16:641–678, 2010.

[2] P. Andersen, J. Klein, K. Knudsen, and R. Tabanera y Palacios. Estimation of variance in cox's regression model with shared gamma frailties. *Biometrics*, 53:1475–1484, 1997.

[3] P.K. Andersen and R.D. Gill. Cox's regression model for counting processes : a large sample study. *Annals of Statistics*, 10:1100–1120, 1982.

[4] C. Andrieu, E. Moulines, and P. Priouret. Stability of stochastic approximation under verifiable conditions. *SIAM J. Control Optim*, 44:283–312, 2005.

[5] D.R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society*, 34:187—-220, 1972.

[6] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1):94–128, 1999.

[7] L. Duchateau and P. Janssen. Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows. *Biometrics*, 60:608–614, 2004.

[8] L. Duchateau and P. Janssen. *The Frailty Model*. Springer-Verlag, New York, 2008.

[9] Il Do Ha, John-Hyeon Jeong, and Youngjo Lee. *Statistical Modelling of Survival Data with Random Effects*. Springer, Singapore, 2017.

[10] E. Kuhn and C. El-Nouty. On a convergent stochastic estimation algorithm for frailty models. *Statistics and Computing*, 23:413–423, 2013.

[11] E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probability and Statistics*, 8:115–131, 2004.

[12] T.A. Louis. Finding the observed information matrix when using the em algorithm. *J. Roy. Statist. Soc. Ser. B*, 44:226–233, 1982.

[13] S. P. Meyn and R.L. Tweedie. Markov chains and stochastic stability. communications and control engineering series. *Springer-Verlag London Ltd*, 1993.

[14] S. A. Murphy. Consistency in a proportional hazards model incorporating a random effect. *Annals of Statistics*, 22:712–731, 1994.

[15] S. A. Murphy. Asymptotic theory for the frailty model. *Annals of Statistics*, 23:182–198, 1995.

[16] G. G. Nielsen, R. D. Gill, P. K. Andersen, and T. I. A. Sorensen. A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.*, 19:25–44, 1992.

[17] E. Parner. Asymptotic theory for the correlated gamma-frailty model. *Annals of Statistics*, 26:183–214, 1998.

[18] S. Ripatti and J. Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56:1016—-1022, 2000.

[19] V. Rondeau, Y. Mazroui, and J. Gonzalez. frailtypack: An r package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47:1–28, 2012.

[20] R. J. Sylvester, A. P. van der Meijden, W. Oosterlinck, J. A. Witjes, C. Bouffioux, L. Denis, D. W. Newling, and K. Kurth. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur. Urol.*, 49(3):466–465, Mar 2006.

[21] T Therneau. Coxme and the laplace approximation. *Technical report*, 2018.

[22] J. Vaupel, K. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439—-454, 1979.

[23] A. Wienke. *Frailty Models in Survival Analysis*. Chapman & Hall/CRC Biostatistics Series. CRC Press, 2010.