# frailtyEM: an **R** Package for Estimating Semiparametric Shared Frailty Models

**Theodor Adrian Balan**
Leiden University Medical Center

**Hein Putter**
Leiden University Medical Center

### Abstract

The software support for fitting so-called "frailty" (random effect) models for time-to-event data has grown considerably in the previous years. Such models are attractive to use when modeling recurrent event data or clustered failures. The usual problem specific to mixed models, which is integrating over the random effects, is further complicated by the presence of a non-parametric "baseline" intensity function. So far, the support for such semi-parametric models was limited, both in terms of the choice for the random effect distribution and in terms of the type of data that the model can be fitted on. We propose a new R package that estimates shared frailty models using the full likelihood, based on the Expectation-Maximization algorithm. The software supports a large number of distributions for the random effect from the Power Variance Family (PVF). Left truncated clustered failures and recurrent events in Andersen-Gill or gaptime formulation are also supported, and conditional and marginal estimates of the survival and cumulative hazard are provided.

*Keywords*: shared frailty, EM algorithm, recurrent events, clustered failures, left truncation, survival analysis, R.

## 1. Introduction

Time-to-event data is very common in medical applications. Often, these data are marked by incomplete observations. For example, the phenomena of right censoring occurs when the actual event time is not observed, but all is known is that the event did not take place by the end of follow-up. Sometimes, individuals enter the data set only if they have not experienced the event until a certain time point. This is known as left truncation, which, if not accounted for correctly, leads to bias. Regression models for such data have been developed in the field of survival analysis. The most popular is the Cox proportional hazards model (Cox 1972), which is semi-parametric in nature: the effect of the covariates is assumed to be time-constant and fully parametric, while the time-dependency arises from the non-parmetric baseline hazard. Cox regression has been the standard in survival analysis for a few reasons: the non-parametric baseline is the best one can do if this function is not known in advance, and the estimation is not computationally intensive. For a long time, this has been implemented in all major statistical software, such as R (R Core Team 2016) (from S-PLUS times), Stata, SAS, SPSS.

When individuals belong to clusters, or may experience recurrent events, the observations are correlated, and in this case the Cox model is not appropriate. Random effect "shared frailty" models have been developed for dealing with such situations. Originating from the field

of demographics (Vaupel, Manton, and Stallard 1979), these models traditionally assume that the proportional hazards model holds conditional on the frailty. The most popular distribution for the random effect is the gamma distribution, chosen mostly for computational convenience. In Hougaard (2000) a variety of distributions that have desirable properties have been proposed. These include the gamma, positive stable (PS), inverse gaussian (IG) and the Power Variance Family (PVF), which includes compound Poisson distributions with mass at 0.

Several surprising results have been demonstrated with frailty models. For example, if proportional hazards are assumed *conditional* on the frailty, then this assumption does not hold on the *marginal* level, for most distributions except the PS. In this case, the model fits proportional hazards on both conditional and marginal levels, with the marginal effect always being an attenuated version of the conditional. This implies that only the semi-parametric PS frailty model can be compared to a marginal Cox model. In fact, the choice of the distribution for the frailty implies a different marginal model, with some emphasizing early depence of the observations (IG) and others the late dependence (gamma). Therefore, it is of great interest to be able to compare a number of different distributions of the random effect.

For Cox model, the computational advantage comes from the fact that the semi-parametric (infinitely dimensional) baseline hazard is not directly estimated, and this is due to the proportional hazards assumption. This simplicity does not carry over to shared frailty models. In this paper we present **frailtyEM** , a R package which uses the general Expectation-Maximization (EM) algorithm for fitting shared frailty models. This implementation comes to complete the landscape of libraries that may be used for such models. At the time of writing this manuscript, in R, semi-parametric shared frailty models can also be fitted in other ways. The first is via a penalized likelihood method with the **survival** (Therneau and Grambsch 2000; Therneau 2015b) and **coxme** (Therneau 2015a) packages. The second way is via h-likelihood with the **frailtyHL** (Do Ha, Noh, and Lee 2012) package, and the the third way is via a pseudo full likelihood approach **frailtySurv** package (Monaco, Gorfine, and Hsu 2017; Gorfine, Zucker, and Hsu 2006). Finally, a Monte Carlo EM-type estimation is available in the **phmm** Donohue and Xu (2013); Vaida and Xu (2000); Donohue, Overholser, Xu, and Florin (2011). Several other options are available for parametric modeling: **frailtypack** (Rondeau, Mazroui, and Gonzalez 2012; Rondeau and Gonzalez 2005) may be used where the baseline is fully parametric or spline-approximated and **parfm** (Munda, Rotolo, Legrand *et al.* 2012) for fully parametric models.

The **frailtyEM** estimates semi-parametric shared frailty models that may be used for recurrent events data in Andersen-Gill and gaptime formulation, clustered failures and clustered failures with left truncation. The supported family of distributions for the random effect includes gamma, IG, PS and the PVF family. The results of the estimation can be easily visualized. Point estimates for regression coefficients are provided with confidence intervals which take into account the estimation of the frailty distribution, and plot methods may be used to visualize both conditional and marginal survival and cumulative hazard curves wiht 95% confidence bands, marginal covariate effects, and empirical Bayes estimates of the random effects. A comparison between **frailtyEM** and other R packages is provided in Table 1.

The rest of this paper is structured as follows. In section 2 we present a brief overview the semi-parametric shared frailty model, and the implications of left truncation. In section 3 we discuss the estimation method used and how this is implemented. In section 4 we illustrate the usage of the functions from the **frailtyEM** package on two classical data sets available in

R.

# 2. Model

We consider the following scenario: there are $I$ clusters and $J_i$ individuals within clusters. The outcome from each individual is represented by a realization of a counting process $N_{ij}$. We consider that the intensity of $N_{ij}$ takes the form

$$\lambda_{ij}(t|z_i) = z_i \exp(\beta^\top \mathbf{x}_{ij}(t))\lambda_0(t) \tag{1}$$

where $z_i$ is an unobserved random effect common to all individuals from cluster $i$ (the "shared frailty"), $\mathbf{x}_{ij}(t)$ a vector of possibly time-dependent covariates, $\beta$ a vector of unknown regression coefficients and $\lambda_0(t) > 0$ an unspecified baseline intensity function. We consider the general case where the $z_i$ follows a distribution with positive support from the infinitely divisible family, i.e., they are i.i.d. realizations of a random variable described by the Laplace transform

$$\mathcal{L}_Z(c; \alpha, \gamma) \equiv \mathrm{E}\left[\exp(-zc)\right] = \exp(-\alpha\psi(c; \gamma)) \tag{2}$$

with $\alpha > 0$ and $\gamma > 0$. This formulation includes several distributions, such as the gamma, PS, IG, PVF. These distributions have been extensively studied in Hougaard (2000). Denote $\theta = (\alpha, \gamma)$ as the parameter vector that describes the distribution. The parametrizations used are described in Appendix A1.

## 2.1. Likelihood

The maximum likelihood problem is to maximize the marginal likelihood, based only on the observed data. This is obtained by integrating over the random effects. Now denote the at-risk indicator of $N_{ij}$ as $Y_{ij}(t)$. With the specification (1), the marginal likelihood is obtained as the product over clusters of expected marginal contributions, i.e.,

$$L(\theta, \beta, \lambda_0(\cdot)) = \prod_i \mathrm{E}_\theta \left[ \prod_j \int_0^\infty \left\{ Y_{ij}(t) z_i \exp(\beta^\top \mathbf{x}_{ij}(t)\lambda_0(t)) \right\}^{dN_{ij}(t)} \right.$$

$$\left. \times \exp(-\sum_j \int_0^\infty Y_{ij}(t) z_i \exp(\beta^\top \mathbf{x}_{ij}(t))\lambda_0(t)dt) \right]$$

To make the connection with the data representation, we consider that $(i, j, k)$ is the indicator for the $k$-th observation from the $j$-th individual in the $i$-th cluster, and $\delta_{ijk}$ is the event indicator for this observation. We write the value of the covariate vector for this observation as $\mathbf{x}_{ijk}$. In the most basic case of clustered failures, $k \equiv 1$, while in the case of recurrent events $j \equiv 1$. More observations for one individual may also arise in the case of clustered failures when the covariates are time-dependent. Nevertheless, the $(i, j, k)$ pair refers to a certain cluster, individual, and period of time where the covariate vector does not change.

The baseline cumulative hazard for this observation is denoted as $\Lambda_{0,ijk}$. Also, let $\tilde{\Lambda}_i = \sum_{jk} \exp(\beta' \mathbf{x}_{ijk})\Lambda_{0,ijk}$. Then, the marginal likelihood can be written as

$$L(\theta, \beta, \lambda_0(\cdot)) = \prod_i \mathrm{E}_\theta \left[ \prod_j \left\{ \prod_k (z_i \exp(\beta^\top \mathbf{x}_{ijk})\lambda_0(t_k))^{\delta_{ijk}} \right\} \exp(-z_i \tilde{\Lambda}_i) \right].$$

| | frailtyEM | survival | coxme | frailtySurv | frailtyHL | frailtypack | parfm | phmm |
|---|---|---|---|---|---|---|---|---|
| **distributions** | | | | | | | | |
| gamma | yes | yes | no | no | no | yes | yes | no |
| log-normal | no | yes | yes | yes | yes | yes | yes | yes |
| PS | yes | no | no | no | no | no | no | no |
| IG | yes | no | no | yes | no | no | yes | no |
| compound Poisson | yes | no | no | no | no | no | no | no |
| PVF | yes | no | no | yes | no | no | no | no |
| **data** | | | | | | | | |
| clustered failures | yes | yes | yes | yes | yes | yes | yes | yes |
| recurrent events (AG) | yes | yes | yes | no | ? | yes | no | no |
| left truncation | yes | no | no | no | no | yes | no | yes |
| correlated structure | no | no | yes | no | no | yes | no | yes |
| **estimation** | | | | | | | | |
| semi-parametric | yes | yes | yes | yes | yes | no | no | yes |
| posterior frailties | yes | yes | no | no | no | ? | no | no |
| conditional $\Lambda_0$, $S_0$ | yes | limited | no | yes | no | yes | ? | no |
| marginal $\Lambda_0$, $S_0$ | yes | no | no | no | no | no | no | no |

Table 1: Comparsion of R packages for frailty models. Versions: **frailtyEM** 0.4.8, **survival** 2.40-1, **coxme** 2.2-5, **frailtyHL** 1.1, **frailtypack** 2.10.5, **parfm** 2.7.1, **phmm** 0.7-5

We consider the Breslow estimator for the baseline hazard, i.e., $\lambda_0(t) \equiv \lambda_{0t}$ for $t$ an event time, and 0 otherwise. By using (2), the marginal likelihood can be rewritten as

$$L(\theta, \beta, \lambda_0(\cdot)) = \prod_i \left[ \prod_j \left\{ \prod_k (z_i \exp(\beta^\top \mathbf{x}_{ijk}) \lambda_0(t_k))^{\delta_{ijk}} \right\} (-1)^{n_i} \mathcal{L}_Z^{(n_i)}(\tilde{\Lambda}_i) \right], \qquad (3)$$

where $\mathcal{L}_Z^{(k)}$ is the $k$-th derivative of the Laplace transform.

## 2.2. Ascertainment

The problem of ascertainment with random effect time-to-event data is usually a difficult one. Consider that the event of observing the cluster $i$ in the data set is $A_i$. Then, the distribution of the random effect in cluster $i$ is described by the Laplace transform of $Z|A_i$, which follows from Bayes' rule as

$$\mathcal{L}_{Z|A_i}(c) = \frac{\mathrm{E}\left[P(A_i|Z)\exp(-cZ)\right]}{\mathrm{E}\left[P(A_i|Z)\right]}. \qquad (4)$$

Expressing $P(A_i|Z = z)$ depends on the type of the study at hand and on the way the data were collected. In **frailtyEM** an option is included to deal with the classical scenario of left truncation, i.e., where

$$P(A_i|Z_i = z_i) = P(T_{i1} > t_{L,i1}, T_{i2} > t_{L,i2}...T_{J_i} > t_{L,iJ_i}|Z_i = z_i)$$

Assume that, given $z_i$, the left truncation times $\mathbf{t}_{L,i}$ are independent and the cluster size is not informative. In this case,

$$P(A_i|Z_i = z_i) = \prod_{j=1}^{J_i} \exp\left(-z_i \int_0^{t_{L,ij}} \exp(\beta^\top \mathbf{x}_{ij}(t))\lambda_0(t)dt\right). \qquad (5)$$

A difficulty here is that the values of the covariate vector and of the baseline intensity must be known prior to the entry time in the study. The consequences of the semi-parametric model, where $\lambda_0 > 0$ only at event time points, are that the $P(T > t) = 1$ for every $t$ before the first event time point. Also, to assign a value for $\mathbf{x}$ before the entry time is speculative. Therefore, we only consider this case when $\mathbf{x}_i$ is time constant.

With the previous notation, denote the risk accumulated before each of the entry times of cluster $i$ as

$$\tilde{\Lambda}_{L,i} = \sum_j \exp(\beta^\top \mathbf{x}_{ij})\Lambda_{0L,ij}$$

where $\Lambda_{0L,ij} = \int_0^{t_{L,ij}} \lambda_0(t)dt$. Then, it results from (4) and (2) that the Laplace transform can be written as

$$\mathcal{L}_{Z|A_i}(c; \alpha, \gamma) = \frac{\exp(-\alpha\psi(c + \Lambda_{L,i}; \gamma))}{\exp(-\alpha\psi(\Lambda_{L,i}; \gamma))} = \exp(-\alpha\tilde{\psi}(c; \Lambda_{L,i}, \gamma)) \qquad (6)$$

where $\tilde{\psi}(c; \Lambda_{L,i}, \gamma) = \psi(c + \Lambda_{L,i}; \gamma) - \psi(\Lambda_{L,i}; \gamma)$. Thus, the random effect stays in the same infinitely divisible family of distributions under this ascertainment scheme.

Note that, in general, the ascertainment scheme does not have a simple description and $P(A_i|Z_i = z_i)$ may or may not be available in closed form. For example, in family studies, the families may be selected only when a number of individuals live long enough (Rodríguez-Girondo, Deelen, Slagboom, and Houwing-Duistermaat 2016). In this case, (5) does not hold. In the case of registry data on recurrent events, individuals (clusters) may be selected only if they have at least one event during a certain time window (Balan, Jonker, Johannesma, and Putter 2016b).

### 2.3. Goodness of fit

The first question when fitting random effect models is whether there is evidence for hetero-geneity. To answer this *a priori* , the score test of Commenges and Andersen (1995) may be used, and is refered in **frailtyEM** as the Commenges-Andersen test. In general, score tests are attractive because they do not require estimating the model. In this spirit, the assumption of the censoring being non-informative on the frailty (Nielsen, Gill, Andersen, and Sørensen 1992) may be easily tested as described in Balan, Boonk, Vermeer, and Putter (2016a).

After fitting the model, the likelihood ratio test may be used to assess whether the model with the frailty is a better fit than a model without frailty. For the gamma and PVF distributions, this is equivalent to testing the hypothesis that $Var(Z) = 0$ versus the alternative $Var(Z) > 0$. Simulation studies for the gamma frailty suggest that the asymptotic distribution of the likelihood ratio test statistic follows a $\chi^2(0) + \chi^2(1)$ distribution (Zhi, Grambsch, and Eberly 2005).

Several measures of dispersion such as the estimated variance of the random effect, Kendall's tau or the $c$-index are discussed in Hougaard (2000) and are implemented in **frailtyEM**.

## 3. Estimation

We propose a general full-likelihood estimation procedure for the gamma, positive stable and PVF frailty models, based on a profile likelihood method and making use of the expectation-maximization (EM) algorithm Dempster, Laird, and Rubin (1977).

For fixed parameters of the frailty distribution $\theta$, we define the profile maximum likelihood

$$\widehat{L}(\theta) = \max_{\beta, \lambda_0} L(\beta, \lambda_0|\theta).$$

For each $\theta$, denote $\hat{\beta}(\theta)$ and $\hat{\lambda}_0(\theta)$ the value of the parameters that maximize $L(\beta, \lambda_0|\theta)$. A first observation is that, if $\hat{\theta}$ maximizes $L(\theta)$, then $(\hat{\theta}, \hat{\beta}(\hat{\theta}), \hat{\lambda}_0(\hat{\theta}))$ maximize $L(\theta, \beta, \lambda_0)$. Thus, we split the problem of maximizing the likelihood into two: obtaining $\hat{\beta}(\theta), \hat{\lambda}_0(\theta)$ for a fixed $\theta$ (the "inner problem") and maximizng $L(\theta)$ over $\theta$ (the "outer problem").

### 3.1. The inner problem

For the inner problem the EM algorithm can be used. This has been first proposed for the gamma frailty model in Nielsen *et al.* (1992) and Klein (1992), and a generalization is discussed in Hougaard (2000).

Most ideas from Nielsen *et al.* (1992) are used here. The crucial observations are that the E step involves calculating the empirical Bayes estimates $\hat{\mathbf{z}} = E[\mathbf{z}|data]$. The expectation is

taken with respect to the "posterior" distribution of the random effect. Afterwards, the M step is essentially a proportional hazards model with log $\hat{\mathbf{z}}$ as offset.

**The E step**  For the E step $\beta$ and $\lambda_0$ are fixed, either at their initial values or at the values from the previous M step. The conditional distribution of $Z_i$ given the data this distribution has Laplace transform

$$\mathcal{L}(c) = \frac{\mathrm{E}\left[z_i^{n_i}\exp(-z_i\tilde{\Lambda}_i)\exp(-zc)\right]}{\mathrm{E}\left[z_i^{n_i}\exp(-z_i\tilde{\Lambda}_i)\right]} = \frac{\mathcal{L}^{(n_i)}(c+\tilde{\Lambda}_i)}{\mathcal{L}^{(n_i)}(\tilde{\Lambda}_i)}. \tag{7}$$

The E step reduces to calculating the derivative of (7) in 0, i.e.,

$$\widehat{z_i} = -\frac{\mathcal{L}^{(n_i+1)}(\tilde{\Lambda}_i)}{\mathcal{L}^{(n_i)}(\tilde{\Lambda}_i)}. \tag{8}$$

The marginal (log-)likelihood is also calculated at this point, $L_\theta(\beta, \lambda_0)$ to keep track of convergence. It can be seen that (3) involved only the denominator of (7) in addition to a simple expression of $\beta$ and $\lambda_0$.

The E step is generally the expensive operation of the EM algorithm. In very few scenarios can (8) be expressed in a closed form: for the gamma and the inverse gaussian distributions. In these scenarios, the E step is calculated with the `fast_estep()` routine. For all other cases, the E step is calculated via a recursive algorithm with an internal routine `estep()`, which is described in Appendix A2. For efficiency and speed, this function was written in C++ and is interfaced with R via the **Rcpp** library (Eddelbuettel and François 2011; Eddelbuettel 2013).

**The M step**  With the same argument as made in Nielsen *et al.* (1992), the M step is equivalent to a regular proportional hazards model with log $\widehat{z_i}$ added as an offset for all the cases in $z_i$. This is done via the `agreg.fit()` function in the **survival** package. Estimates of $\beta$ are directly obtained from this, while estimates for $\lambda_0$ and the subsequent calculations of $\tilde{\Lambda}_i$ (and, eventually $\tilde{\Lambda}_{L,i}$) require a careful calculation of subjects at risk and the respective linear predictors at every event and entry time point. The ordering required for determining these "at risk" sets is cached in `emfrail()`.

The EM algorithm stops after the marginal log-likelihood has converged.

### 3.2. Outer problem

The "outer" problem refers to finding $\widehat{\theta}$ which maximizes $\widehat{L}(\theta)$. The resulting $\widehat{\theta}$ is the maximum likelihood estimator and the maximum likelihood is obtained at $\widehat{L}(\widehat{\theta})$. All the infinite divisible distributions in this library involve only a one-dimensional $\theta$. The maximizer of choice may be one of those from the **optimx** package (Nash and Varadhan 2011; Nash 2014), and it defaults to `bobyqa`. The results of this maximization are returned in the final object and are accessible to the user.

### 3.3. Return object and standard errors

After the maximizer has converged and the outer maximization is done, a number of calls to the internal `em_fit()` are performed at $\hat{\theta}$ and $\hat{\theta} \pm \varepsilon$, this time returning a complete list of

results including the variance-covariance matrix of $(\beta, \lambda_0)$. These results are used for general output, and also for calculating an adjusted variance-covariance matrix for the uncertainty in estimating $\theta$. The idea is described in Appendix A3.

In general, the return object type is `emfrail`, which is essentially a list that contains the results of the "outer" maximization as returned by `optimx()`, the results of the "inner" maximization at the estimate, and a few other fields which are used for different methods. The object type is documented in `?emfrail`. Some options to obtain only part of the results are available via the `.control` argument.

### 3.4. Output, summary and prediction

By itself, an `emfrail` object prints the call, a summary of "outer" optimization, the estimates of the covariates and the *p*value of the Commenges-Andersen test. A more user-readable summary of an `emfrail` object is provided by the `summary.emfrail()` method. This returns an object of the class `emfrail_summary` that contains general fit information, covariate estimates and several distribution-specific measures of fit and dispersion. Since most of these depend on estimated parameters, the Delta method as implemented in `deltamethod` from the package **msm** (Jackson 2011) is used.

A method for predicting cumulative hazard and survival curves, both conditional and marginal, exists in `predict.emfrail()`. Confidence bands are based on the asymptotic normality of the estimated $\lambda_0$, and available both for adjusted and un-adjusted for the uncertainty of $\theta$. The user can specify which quantities to obtain and values of the linear predictor where to calculate these curves. The function returns a data frame from which several plots can be easily created.

A few simple plot functions have been created for convenience, although more flexibility may be achieved with **ggplot2** (this is described in the documentation of `predict.emfrail`). An overview of the plots is available with `?plot_emfrail`. For obtaining a histogram of the empirical Bayes estimates of the frailties, `hist_frail()` may be used. For plotting predicted cumulative hazard or survival curves `plot_pred()` may be used, and to plot the marginal and conditional estimated hazard ratios `plot_hr()` may be used.

An additional function, `emfrail_pll()` is provided to calculate the marginal log-likelihood for a vector of values of $\theta$, without actually performing the outer optimizaion. This may be useful for visualizing the profile log-likelihood or when debugging (e.g.,if the maximum likelihood estimate of $\theta$ lies on the boundary).

## 4. Illustration

For a quick start, one can run this:

```
> library(frailtyEM)
```

The features of the package are now illustrated with two well-known data sets available in R.

### 4.1. bladder

The data on recurrences of bladder cancer has been used several times to demonstrate methodology for recurrent events and is part of the **survival** package. The data set in Andersen-Gill

format is present in `bladder2`.

```
> bladder2$rx <- as.factor(bladder2$rx)
```

The variables of interest here are: `start`, `stop` and `event` determine the outcome, while `rx` is a factor variable for treatment, `number` is the number of initial tumors and `size` is the size of the largest initial tumor.

A basic `emfrail` model can be fitted like this:

```
> m1 <- emfrail(.data = bladder2,
+               .formula = Surv(start, stop, event) ~ rx + number + size + cluster(id))
> m1

Call:
emfrail(Surv(start, stop, event) ~ rx + number + size + cluster(id))
          theta    loglik fevals gevals niter convcode kkt1 kkt2 xtimes
bobyqa 1.075755 -442.6776     19     NA    NA        0 TRUE TRUE  0.426


           coef exp(coef) se(coef) adjusted se        z      p
rx2    -0.582849  0.558306 0.317177    0.317502 -1.837613 0.0661
number  0.224087  1.251180 0.088881    0.089335  2.521193 0.0117
size   -0.023309  0.976961 0.107086    0.107213 -0.217662 0.8277


Commenges-Andersen test for heterogeneity: p-val 0.00768
```

The arguments of `emfrail` visible above are `.data` and `.formula`. The `.control` and `.distribution` are taken as defaults; for the latter, that is the gamma frailty distribution. The `.formula` argument contains a `Surv` object at the left hand side and a `+cluster()` statement on the right hand side (essentially as `+frailty()` in coxph). The `.distribution` and `.control` arguments must be objects of the type `emfrail_distribution` and `emfrail_control`, which are created by calls to functions with the same names. For example, the default choice for the distribution is:

```
> str(emfrail_distribution())

List of 4
 $ dist           : chr "gamma"
 $ theta          : num 2
 $ pvfm           : num -0.5
 $ left_truncation: logi FALSE
 - attr(*, "class")= chr "emfrail_distribution"
```

The `emfrail_distribution` objects have 4 fields: `dist` describes the distribution of the frailty (here, a gamma distribution), `theta` is the frailty parameter and the starting value for the optimization. The parametrizations are described in Appendix A1. The `pvfm` field only plays a role when `dist=="pvf"`, and describes which PVF family distribution should be used. Finally, `left_truncation` is a logical variable, on whether to treat an observation as

left truncated or not. For example, in the case of recurrent events in Andersen-Gill format, this should be `FALSE`, because the `start` column does not refer to ascertainment, and the frailty must not be taken conditional on not having an event before that time point. The adjustment that happens if left truncation is present is described in section 2.2.

The `emfrail` object prints a slightly modified version of the output from the "outer" maximization, which is the same as one from `optimx`, with a few additions.

```
> m1


Call:
emfrail(Surv(start, stop, event) ~ rx + number + size + cluster(id))
           theta     loglik fevals gevals niter convcode kkt1 kkt2 xtimes
bobyqa 1.075755 -442.6776      19     NA    NA         0 TRUE TRUE  0.426


              coef exp(coef)  se(coef) adjusted se        z      p
rx2     -0.582849  0.558306  0.317177    0.317502 -1.837613 0.0661
number   0.224087  1.251180  0.088881    0.089335  2.521193 0.0117
size    -0.023309  0.976961  0.107086    0.107213 -0.217662 0.8277


Commenges-Andersen test for heterogeneity: p-val 0.00768
```

A more interesting output is obtained by calling `summary()`:

```
> sm1 <- summary(m1)
> sm1


Summary of emfrail fit
Regression coefficients:
              coef exp(coef)  se(coef) adjusted se        z       p
rx2     -0.582849  0.558306  0.317177    0.317502 -1.8376 0.06612 .
number   0.224087  1.251180  0.088881    0.089335  2.5212 0.01170 *
size    -0.023309  0.976961  0.107086    0.107213 -0.2177 0.82769
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.00768
(marginal) no-frailty Log-likelihood: -453.243
(marginal) Log-likelihood: -442.678
LRT: 1/2 * pchisq(21.1), p-val 2.15e-06

Frailty summary:
theta = 1.076 (0.39) / 95% CI: [0.531, 2.178]
variance = 0.93 (0.33) / 95% CI: [0.459, 1.882]
Kendall's tau: 0.317 (0.08) / 95% CI: [0.187, 0.485]
```

The first two parts of this output, about regression coefficients and fit summary, exist regardless of the frailty distributions. The last part, "frailty summary", depends on the distribution at hand. The frailty variance is estimated as 0.93 in this case. Both the Commenges-Andersen test for heterogeneity and the one-sided likelihood ratio test deems the random effect highly significant. This is also suggested by the confidence interval for the frailty variance, which is far from 0.

The results are almost identical with a gamma frailty fit from `coxph`. The marginal log-likelihood in the `emfrail` object is slightly higher, that is because the estimation of the frailty distribution is more precise. In addition, `emfrail` also provides a 95% confidence interval for the frailty variance.

```
> m_cph <- coxph(Surv(start, stop, event) ~ rx + number + size + frailty(id),
+             data = bladder2,
+             ties = "breslow")
> m_cph
```

```
Call:
coxph(formula = Surv(start, stop, event) ~ rx + number + size +
    frailty(id), data = bladder2, ties = "breslow")

             coef se(coef)     se2   Chisq   DF      p
rx2       -0.5839   0.3176  0.2174  3.3802  1.0 0.0660
number     0.2249   0.0882  0.0563  6.4973  1.0 0.0108
size      -0.0233   0.1083  0.0711  0.0465  1.0 0.8293
frailty(id)                        73.8156 42.2 0.0019

Iterations: 6 outer, 28 Newton-Raphson
     Variance of random effect= 0.93   I-likelihood = -442.7
Degrees of freedom for terms=  0.5  0.4  0.4 42.2
Likelihood ratio test=130  on 43.5 df, p=1.58e-10  n= 178
```
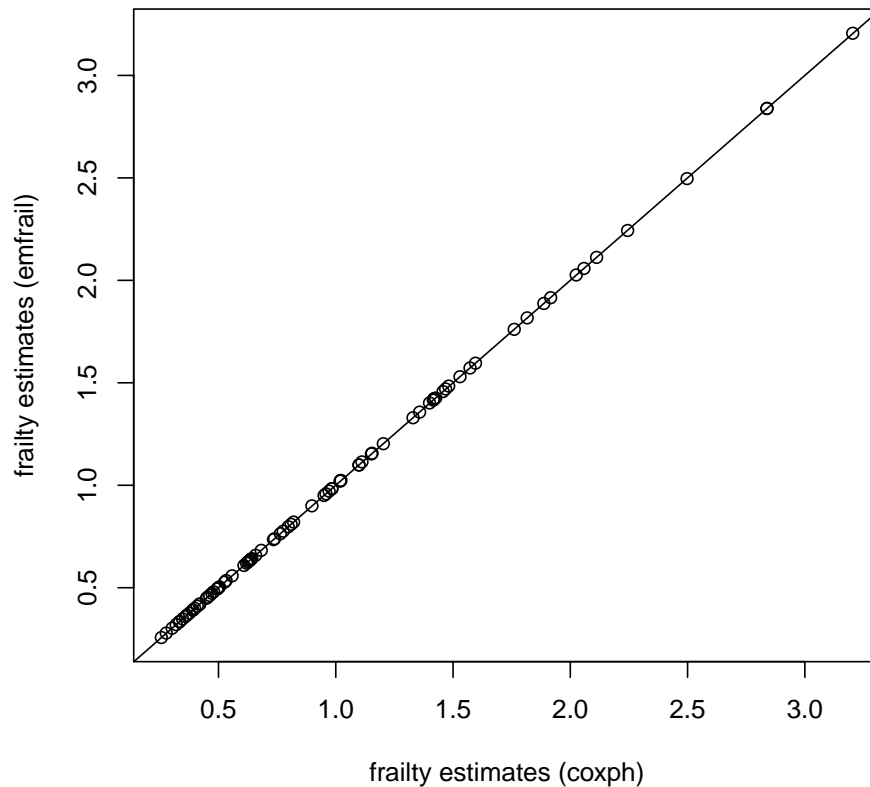
The empirical Bayes frailty estimates of $\widehat{\mathbf{z}}$ are also identical for the two ways of fitting the model.

```
> plot(exp(m_cph$frail),
+      sm1$frail$z,
+      xlab = "frailty estimates (coxph)",
+      ylab = "frailty estimates (emfrail)")
> abline(0,1)
```

To look at predicted cumulative hazard curves we take two individuals, one from the treatment arm and one from the placebo arm, both with three recurrent tumor at baseline and with size 3.

```
> par(mfrow=c(1,2))
> plot_pred(m1,
+          newdata = data.frame(rx = "2", number = 3, size = 3),
+          ylim = c(0,4),
+          main = "treatment")
> plot_pred(m1,
+          newdata = data.frame(rx = "1", number = 3, size = 3),
+          ylim = c(0,4),
+          main = "placebo")
```
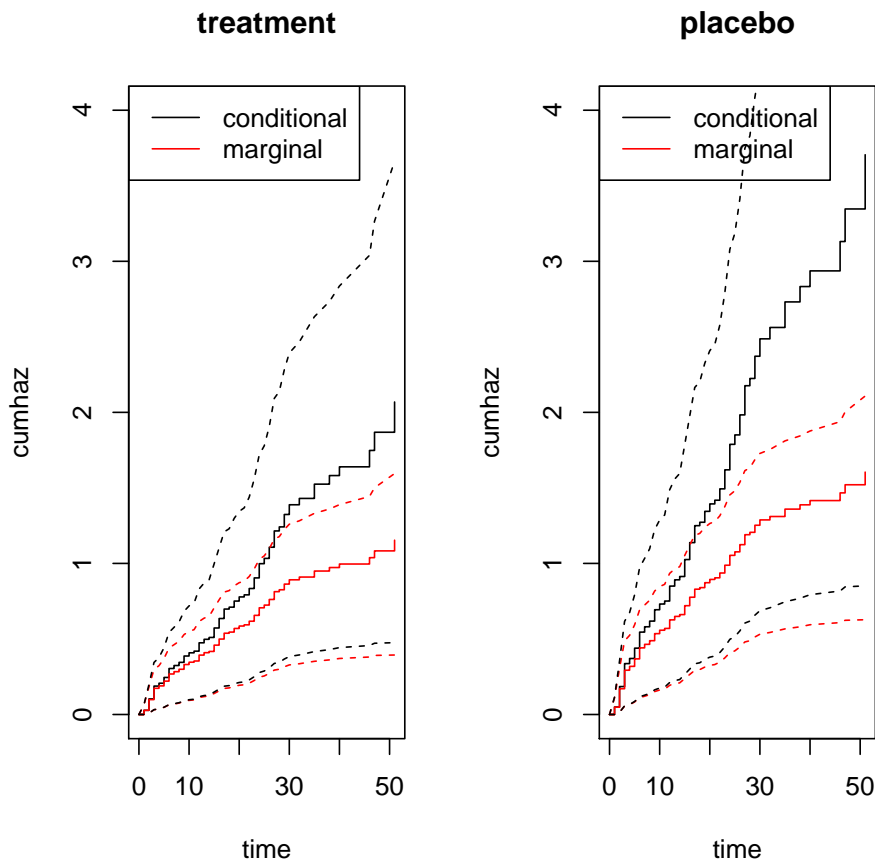
The cumulative hazard in this case can be interpreted as the expected number of events at a certain time. It can be seen that the frailty "drags down" the marginal hazard. This is a well-known effect observed in frailty models, as described in Aalen, Borgan, and Gjessing (2008, ch. 7).

A similar model can be fitted with the positive stable distribution:

```
> m2 <- emfrail(.data = bladder2,
+               .formula = Surv(start, stop, event) ~ rx + number + size + cluster(id),
+               .distribution = emfrail_distribution(dist = "stable"))
> summary(m2)


Summary of emfrail fit
Regression coefficients:
            coef  exp(coef)  se(coef)  adjusted se        z        p
rx2     -0.578448  0.560768  0.309813     0.312524  -1.8671  0.061889 .
number   0.218502  1.244212  0.070128     0.073333   3.1157  0.001835 **
size    -0.032408  0.968111  0.101512     0.102164  -0.3193  0.749535
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Estimated distribution: stable / left truncation: FALSE
```

```
Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.00768
(marginal) no-frailty Log-likelihood: -453.243
(marginal) Log-likelihood: -448.185
LRT: 1/2 * pchisq(10.1), p-val 0.000735

Frailty summary:
theta = 0.22 (0.11) / 95% CI: [0.085, 0.569]
Kendall's tau: 0.18 (0.07) / 95% CI: [0.078, 0.363]
Attenuation factor: 0.78 / Var[log(Z)] = 0.163
```

The coefficient estimates are similar with those of `m1`. The "frailty summary" part is quite different though. The positive stable distribution has infinite expectation. However, Kendall's tau is easily obtained, and in this case it is smaller than in the gamma frailty model. Unlike the gamma or PVF distributions, the positive stable frailty predicts a marginal model with proportional hazards. This is discussed at length in Hougaard (2000). This can be easily visualized with `emfrail`.

```
> par(mfrow=c(1,2))
> plot_hr(m1,
+         newdata = data.frame(rx = c("1", "2"), number = 3, size = 3),
+         main = "gamma")
> plot_hr(m2,
+         newdata = data.frame(rx = c("1", "2"), number = 3, size = 3),
+         main = "stable")
>
```
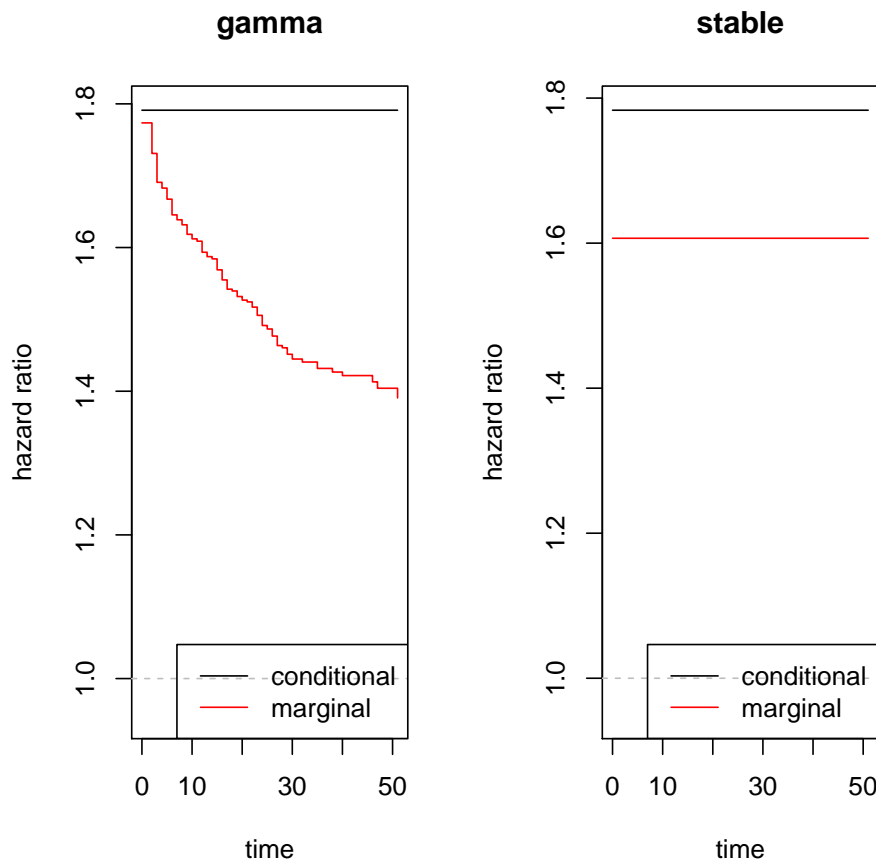
The plot shows that the marginal hazard ratio of the gamma frailty model is not time-constant, while the one from the positive stable frailty model is. This is discussed in Aalen *et al.* (2008, ch. 7). In Hougaard (2000) this is seen as a strength of the stable frailty model.

## 4.2. kidney

The kidney data set is also available in the **survival** package. The data, presented originally in McGilchrist and Aisbett (1991), contains the time to infection, at the point of insertion of the catherer, for kidney patients using portable dialysis equipment. If the catheters are removed for other reasons, then the observation is censored. Each of the 38 patients has exactly 2 observations. There are 3 covariates: sex, age and a disease (a factor with 4 levels). This data are analized in Therneau and Grambsch (2000, ch. 9.5.2). The authors note that, when disease is included in the model, a gamma frailty model offers no evidence of heterogeneity. When disease is not included in the model, then there seems to be moderate evidence for heterogeneity; this is an example where the frailty may be interpreted as a missing covariate.

```
> data(kidney)
> kidney$sex <- ifelse(kidney$sex == 1, "male", "female")
> m_gam_d <- emfrail(.data = kidney,
+                    .formula = Surv(time, status) ~ age + sex + disease + cluster(id))
> summary(m_gam_d)
```

```
Summary of emfrail fit
Regression coefficients:
                coef   exp(coef)  se(coef) adjusted se      z        p
age        0.0034309  1.0034368  0.0111484   0.0111484  0.3077  0.75827
sexmale    1.4715944  4.3561751  0.3579169   0.3579220  4.1116 3.93e-05 ***
diseaseGN  0.0894202  1.0935401  0.4068262   0.4068271  0.2198  0.82603
diseaseAN  0.3518318  1.4216694  0.4002232   0.4002232  0.8791  0.37935
diseasePKD -1.4276348  0.2398756  0.6309795   0.6309844 -2.2626  0.02366 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.879
(marginal) no-frailty Log-likelihood: -179.394
(marginal) Log-likelihood: -179.394
LRT: 1/2 * pchisq(0), p-val 0.5

Frailty summary:
theta = 10878.12 (73909.27) / 95% CI: [0.018, 6606662793]
variance = 0 (0) / 95% CI: [0, 55.831]
Kendall's tau: 0 (0) / 95% CI: [0, 0.965]


> m_gam <- emfrail(.data = kidney,
+                  .formula = Surv(time, status) ~ age + sex + cluster(id))
> summary(m_gam)


Summary of emfrail fit
Regression coefficients:
           coef exp(coef)  se(coef) adjusted se      z        p
age    0.0054372 1.0054520 0.0115813   0.0116976 0.4695 0.6387255
sexmale 1.5528409 4.7248739 0.4451768   0.4995213 3.4881 0.0004864 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.0245
(marginal) no-frailty Log-likelihood: -184.657
(marginal) Log-likelihood: -182.053
LRT: 1/2 * pchisq(5.21), p-val 0.0112

Frailty summary:
theta = 2.517 (1.49) / 95% CI: [0.791, 8.012]
variance = 0.397 (0.23) / 95% CI: [0.125, 1.264]
Kendall's tau: 0.166 (0.08) / 95% CI: [0.059, 0.387]
```

Therneau and Grambsch discuss these models and they conclude that an outlier case is at the source of this surprising result. This raises the question on what the frailty actually means in this model. With the **frailtyEM** package, the stable frailty model may also be fitted. Unlike the gamma frailty model, the positive stable does not attemt to "correct" non-proportional hazards.

```
> m_stab <- emfrail(.data = kidney,
+                   .formula = Surv(time, status) ~ age + sex + cluster(id),
+                   .distribution = emfrail_distribution(dist = "stable"))
> summary(m_stab)


Summary of emfrail fit
Regression coefficients:
           coef exp(coef)  se(coef) adjusted se      z        p
age     0.0021815 1.0021839 0.0092246   0.0092246 0.2365 0.813054
sexmale 0.8209953 2.2727608 0.2987197   0.2987197 2.7484 0.005989 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Estimated distribution: stable / left truncation: FALSE


Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.0245
(marginal) no-frailty Log-likelihood: -184.657
(marginal) Log-likelihood: -184.657
LRT: 1/2 * pchisq(0), p-val 0.5


Frailty summary:
theta = 0 (0) / 95% CI: [0, Inf]
Kendall's tau: 0 (0) / 95% CI: [0, NaN]
Attenuation factor: 1 / Var[log(Z)] = 0
```

The Commenges-Andersen test for heterogeneity shows the same evidence as before, as it does not depend on the frailty distribution. However, the positive stable parameter lies at the edge of the parameter space ($\theta$ is between 0 and 1 for the PS distribution). Therefore, the LRT is not significant. The major difference with the gamma frailty fit is that the regression coefficient for sex is much smaller. To untangle this effect, one can check the (marginal) proportional hazards assumption. This reveals that `sex` has a significant non-proportional effect on the hazards:

```
> zph1 <- cox.zph(coxph(Surv(time, status) ~ age + sex + cluster(id), data = kidney))
> zph1


          rho    chisq        p
age      0.0214   0.0231 8.79e-01
sexmale -0.4390  29.2598 6.33e-08
GLOBAL      NA   29.3325 4.27e-07
```

In small samples, the gamma frailty model implicitly fits a marginal non-proportional hazards model, and in this case it succeeds. The PS distribution fits proportional hazards both conditional and marginal, and in this case it fails. To untangle this effect, we can perform a proportional hazards test with the log-estimated frailties as an offset:

```
> s_gam <- summary(m_gam)
> off_z <- log(s_gam$frail$z)[match(kidney$id, s_gam$frail$id)]
> zph2 <- cox.zph(coxph(Surv(time, status) ~
+                       age + sex + offset(off_z) + cluster(id),
+                  data = kidney))
> zph2

           rho   chisq     p
age     -0.0145 0.00427 0.948
sexmale -0.2170 1.39043 0.238
GLOBAL       NA 1.41146 0.494
```

In this case, this is evidence that the gamma frailty corrects for proportionality rather than heterogeneity.

# 5. Conclusion

We have shown that the EM based approach has certain advantages in the context of frailty models. First of all, it is semiparametric, which means that it is an extension of the Cox proportional hazards model. In this way, classical results from semiparametric frailty models (for example, the data sets in section 4) can be replicated and further insight may still be obtained. Until now, the Commenges-Andersen test, PS, PVF family, have not all been implemented in a consistent way in an R package.

Several options not discussed in this paper include the left truncation adjustment. There is no available data set to illustrate this option, however the peroforming of a larger simulation study to assess the effects of left truncation in clustered failure data is now possible.

Other possible extensions of this software are posible, since all that is needed is to specify the Laplace transform and the corresponding derivatives for the E step. An interesting extension would be to choose discrete distributions for the random effect. The newest features will be implemented in the development version of the package at https://github.com/teddybalan/frailtyEM.

In the current landscape for modeling random effects in survival analysis, **frailtyEM** is a contribution that focuses on implementing classical methodology in an efficient way. This comes to aid researches, as well as clinicians, facilitating the analysis of previous and future studies.

# Appendix A1: Results for the Laplace transforms

We consider distributions from the infinitely divisible family (Ash 2014, ch 8.5) with the Laplace transform

$$\mathcal{L}(c) = \exp(-\alpha\psi(c;\gamma)).$$

**The gamma distribution** For the gamma distribution, $\psi(c;\gamma) = \log(\gamma + c) - \log(\gamma)$. We have

$$\psi^{(k)}(c;\gamma) = (-1)^{k-1}(k-1)!(\gamma + c)^{-k}.$$

For identifiability, the natural parametrization would be $\alpha = \gamma$. Then, we have for a $\theta > 0$, $\theta = \alpha = \gamma$, which implies mean 1 and variance $\theta^{-1}$.

**The positive stable distribution** For the positive stable distribution, $\psi(c;\gamma) = c^\gamma$ with $\gamma \in (0,1)$.

$$\psi^{(k)}(c;\gamma) = \frac{\Gamma(k-\beta)}{\Gamma(1-\gamma)}(-1)^{k-1}c^{\gamma-1}.$$

For identifiability, we take $\alpha = 1$. Then, we have for a $\theta > 0$, $\gamma = 1 - \frac{\theta}{\theta+1}$. This parametrization is equivalent to that from Munda *et al.* (2012). In Hougaard (2000), the parameter $\alpha_H = \gamma$. Kendall's $\tau$ is then $\tau = \frac{\theta}{\theta+1}$ and the median concordance is $\kappa = 2^{2-2^{1-\frac{\theta}{\theta+1}}} - 1$

**The PVF distributions** For the PVF distribution, with fixed parameter $m \in \mathbb{R}$, $m > -1$ and $m \neq 0$,

$$\psi(c;\gamma) = \text{sign}(m)(1 - \gamma^m(\gamma + c)^{-m})$$

where sign denotes the sign. This is the same parametrizaion as in Aalen *et al.* (2008). To exhibit the tie-in with that of Hougaard (2000), it would be $m = -\alpha_H$, $\gamma = \theta_H$, $\alpha = |\delta_H/m\gamma^{-m}|$. We have

$$\psi^{(k)}(c;\gamma) = \text{sign}(m)(-\gamma)^m(\gamma + c)^{-m-k}(-1)^{k+1}\frac{\Gamma(m+k)}{\Gamma(m)}.$$

It can be seen that the expectation of this distribution is minus the first derivative of the Laplace transform calculated in 0, i.e.,

$$E(Z) = \alpha\psi'(0;\gamma)\mathcal{L}(0;\alpha,\gamma) = \frac{\alpha}{\gamma}m.$$

The second moment of the distribution can be calculated as the second derivative of the Laplace transform at 0,

$$EZ^2 = \alpha^2\psi'^2(0) - \alpha\psi''(0) = \frac{\alpha^2}{\gamma^2}m^2 + \frac{\alpha}{\gamma^2}m(m+1).$$

For identifiability, we set the mean to 1. For a $\theta > 0$, we write $\gamma = (m+1)\theta$ and $\alpha = \frac{m+1}{m}\theta$. This results in the variance of the frailty equal to $\theta^{-1}$.

## Left truncation

To determine the Laplace transform under left truncation, we determine $\tilde{\psi}$ from 6.

For the gamma distribution, we have

$$\tilde{\psi}(c;\gamma, \Lambda_L) = \log(\gamma + \Lambda_L + c) - \log(\gamma + \Lambda_L)$$

which implies that the frailty of the survivors is still gamma distributed, but with a change in the parameter $\gamma$.

For the positive stable we have

$$\tilde{\psi}(c; \gamma, \Lambda_L) = (c + \Lambda_L)^\gamma - \Lambda_L^\gamma$$

which is not a positive stable distribution any more.

For the pvf distribution, we have

$$\tilde{\psi}(c; \gamma, \Lambda_L) = \text{sign}(m) \left( \gamma^m (\gamma + \Lambda_L)^{-m} - (\gamma + \Lambda_L)^m (\gamma + \Lambda_L + c)^{-m} \right)$$

which is not pvf any more.

### Closed forms

The gamma distribution leads to a Laplace transform for which the derivatives can be calculated in closed form. It can be seen that

$$\mathcal{L}(c; \alpha, \gamma) = \gamma^\alpha (\gamma + c)^{-\alpha}.$$

The $k$-th derivative of this expression is

$$\mathcal{L}^{(k)}(c; \alpha, \gamma) = \gamma^\alpha (\gamma + c)^{-\gamma - k} \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)}.$$

This can be exploited also in the case of left truncation, since the gamma frailty is preserved, as shown in the previous section.

The inverse gaussian distribution is obtained when the pvf parameter is $m = -\frac{1}{2}$. Under the current parametrization, we have $\beta = \theta/2$ and $\alpha = \theta$. In this case, the Laplace transform is

$$\mathcal{L}(c; \theta) = \exp \left\{ \theta \left( 1 - \sqrt{1 + 2c/\theta} \right) \right\}.$$

The $k$-th derivative of this can be written as

$$\mathcal{L}^{(k)}(c; \theta) = (-1)^k \left( \frac{2}{\theta} c + 1 \right)^{-k/2} \frac{\mathcal{K}_{k-1/2} \left( \sqrt{2\theta \left( c + \frac{\theta}{2} \right)} \right)}{\mathcal{K}_{1/2} \left( \sqrt{2\theta \left( c + \frac{\theta}{2} \right)} \right)}$$

where $\mathcal{K}$ is the modified Bessel function of the second kind.

The `emfrail()` uses the closed form formulas when possible, by default.

## Appendix A2: A general E step

As shown in (7), the calculation of the E step for the general case involves taking derivatives of Laplace transforms of the form

$$\mathcal{L}(c) = \exp(g(c))$$

where for simplicity we denote $g(c) = -\alpha\psi(c; \gamma)$. The expression for the $k$-th derivative of $\mathcal{L}(c)$ can be obtained with a classical calculus result, di Bruno's formula, i.e.,

$$\mathcal{L}^{(n)}(c) = \sum_{\mathbf{m} \in \mathcal{M}_n} \frac{n!}{m_1! m_2! ... m_n!} \prod_{j=1}^{n} \left( \frac{g^{(j)}(c)}{j!} \right)^{m_j} \mathcal{L}(c) \tag{9}$$

where $\mathcal{M}_n = \{(m_1, ..., m_n)| \sum_{j=1}^{n} j \times m_j = n\}$. For example, for $n = 3$,

$$\mathcal{M}_3 = \{(3, 0, 0), (1, 1, 0), (0, 0, 1)\}.$$

This corresponds to the "partitions of the integer" 3, i.e., all the integers that sum up to 3:

$$\{(1, 1, 1), (1, 2, 0), (3, 0, 0)\}.$$

We implemented a recursive algorithm in C++ which resides in the `emfrail_estep.cpp` which loops through these partitions, calculates the corresponding derivatives of $\psi$ and the coefficients.

## Appendix A3: Standard errors

Denote the vector of parameters $\eta = (\beta, \lambda_0(\cdot))$. The information matrix for $(\theta, \eta)$ can be written as follows:

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_{\theta,\theta} & \mathcal{I}_{\theta,\eta} \\ \mathcal{I}_{\eta,\theta} & \mathcal{I}_{\eta,\eta} \end{bmatrix}.$$

From this, $\mathcal{I}_{\theta,\theta}$ is approximated numerically as the Hessian of the "outer" maximization. The part corresponding to $\eta$, $\mathcal{I}_{\eta,\eta}$ is calculated using Louis' formula, which has been commonly employed to obtain this quantity from EM algorithms Louis (1982). It only requires the first and second derivatives of $\log \widehat{L}(\eta)$.

By inverting $\mathcal{I}_{\theta,\theta}$ and $\mathcal{I}_{\eta,\eta}$, two variance-covariance matrices are obtained: $Var(\theta)$ and $Var(\eta|\theta = \widehat{\theta})$. The latter matrix is the variance-covariance for $\eta$ under fixed $\theta$; from this, the standard errors are calculated, for example, in the **survival** package. This leads to an under-estimation of the standard errors, the extent of which, to our knowledge, has not been yet studied.

The calculation of the variance-covariance matrix $\mathcal{I}^{-1}$ in this case has been described in Hougaard (2000, Appendix B.3) and Putter and Van Houwelingen (2015).

## References

Aalen O, Borgan O, Gjessing H (2008). *Survival and event history analysis: a process point of view.* Springer-Verlag.

Ash RP (2014). *Real Analysis and Probability: Solutions to Problems.* Academic press.

Balan TA, Boonk SE, Vermeer MH, Putter H (2016a). "Score test for association between recurrent events and a terminal event." *Statistics in medicine*, **35**(18), 3037–3048.

Balan TA, Jonker MA, Johannesma PC, Putter H (2016b). "Ascertainment correction in frailty models for recurrent events data." *Statistics in medicine*, **35**(23), 4183–4201.

Commenges D, Andersen PK (1995). "Score test of homogeneity for survival data." *Lifetime data analysis*, **1**(2), 145–156.

Cox DR (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society B*, **34**(2), 187–220. ISSN 00359246. URL http://www.jstor.org/stable/2985181.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society B*, pp. 1–38.

Do Ha I, Noh M, Lee Y (2012). "**frailtyHL**: a package for fitting frailty models with h-likelihood." *RJ*, **4**(2), 28–36.

Donohue MC, Overholser R, Xu R, Florin V (2011). "Conditional Akaike information under generalized linear and proportional hazards mixed models." *Biometrika*, (98, 3), 685–700.

Donohue MC, Xu R (2013). ***phmm***: *Proportional Hazards Mixed-effects Models*. R package version 0.7-5.

Eddelbuettel D (2013). *Seamless R and C++ Integration with Rcpp*. Springer-Verlag, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel D, François R (2011). "**Rcpp**: Seamless R and C++ Integration." *Journal of Statistical Software*, **40**(8), 1–18. URL http://www.jstatsoft.org/v40/i08/.

Gorfine M, Zucker DM, Hsu L (2006). "Prospective survival analysis with a general semiparametric shared frailty model: A pseudo full likelihood approach." *Biometrika*, pp. 735–741.

Hougaard P (2000). *Analysis of multivariate survival data*. Springer-Verlag.

Jackson CH (2011). "Multi-State Models for Panel Data: The **msm** Package for R." *Journal of Statistical Software*, **38**(8), 1–29. URL http://www.jstatsoft.org/v38/i08/.

Klein JP (1992). "Semiparametric estimation of random effects using the Cox model based on the EM algorithm." *Biometrics*, pp. 795–806.

Louis TA (1982). "Finding the observed information matrix when using the EM algorithm." *Journal of the Royal Statistical Society B*, pp. 226–233.

McGilchrist C, Aisbett C (1991). "Regression with frailty in survival analysis." *Biometrics*, pp. 461–466.

Monaco JV, Gorfine M, Hsu L (2017). ***frailtySurv***: *General Semiparametric Shared Frailty Model*. R package version 1.3.2, URL https://CRAN.R-project.org/package=frailtySurv.

Munda M, Rotolo F, Legrand C, *et al.* (2012). "**parfm**: Parametric frailty models in R." *Journal of Statistical Software*, **51**(1), 1–20.

Nash JC (2014). "On Best Practice Optimization Methods in R." *Journal of Statistical Software*, **60**(2), 1–14. URL http://www.jstatsoft.org/v60/i02/.

Nash JC, Varadhan R (2011). "Unifying Optimization Algorithms to Aid Software System Users: **optimx** for R." *Journal of Statistical Software*, **43**(9), 1–14. URL http://www.jstatsoft.org/v43/i09/.

Nielsen GG, Gill RD, Andersen PK, Sørensen TI (1992). "A counting process approach to maximum likelihood estimation in frailty models." *Scandinavian Journal of Statistics*, pp. 25–43.

Putter H, Van Houwelingen HC (2015). "Dynamic frailty models based on compound birth–death processes." *Biostatistics*, **16**(3), 550–564.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rodríguez-Girondo M, Deelen J, Slagboom EP, Houwing-Duistermaat JJ (2016). "Survival analysis with delayed entry in selected families with application to human longevity." *Statistical methods in medical research*, p. 0962280216648356.

Rondeau V, Gonzalez JR (2005). "**frailtypack**: A computer program for the analysis of correlated failure time data using penalized likelihood estimation." *Computer Methods and Programs in Biomedicine*, **80**(2), 154–164.

Rondeau V, Mazroui Y, Gonzalez JR (2012). "**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**(4), 1–28. URL http://www.jstatsoft.org/v47/i04/.

Therneau TM (2015a). *coxme: Mixed Effects Cox Models*. R package version 2.2-5, URL https://CRAN.R-project.org/package=coxme.

Therneau TM (2015b). *A Package for Survival Analysis in S*. Version 2.38, URL https://CRAN.R-project.org/package=survival.

Therneau TM, Grambsch PM (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York. ISBN 0-387-98784-3.

Vaida F, Xu R (2000). "Proportional hazards model with random effects." *Statistics in Medicine*, (19), 3309–3324.

Vaupel JW, Manton KG, Stallard E (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography*, **16**(3), 439–454.

Zhi X, Grambsch PM, Eberly LE (2005). "Likelihood ratio test for the variance component in a semi-parametric shared gamma frailty model." *Research Report 2005-5*.

**Affiliation:**

Theodor Adrian Balan
Department of Medical Statistics and Bioinformatics
Leiden University Medical Center
2300 RC Leiden, The Netherlands
E-mail: t.a.balan@lumc.nl