parfm: Parametric Frailty Models in R

Package vignette, V. 1.4 (January 25th, 2017)

Marco Munda
Arlenda

Federico Rotolo
Gustave Roussy

Catherine Legrand
Université catholique de Louvain

Abstract

Frailty models are getting more and more popular to account for overdispersion and/or clustering in survival data. When the form of the baseline hazard is somehow known in advance, the parametric estimation approach can be used advantageously. Nonetheless, there is no unified widely available software that deals with the parametric frailty model. The new **parfm** package remedies that lack by providing a wide range of parametric frailty models in R. The available baseline hazard failies are: exponential, Weibull, inverse Weibull (Fréchet), Gompertz, lognormal, log-kewNormal, and loglogistic. The gamma, positive stable, inverse Gaussian, and lognormal frailty distributions can be specified, together with five different baseline hazards. Parameter estimation is done by maximising the marginal log-likelihood, with right-censored and possibly left-truncated data. In the multivariate setting, the inverse Gaussian may encounter numerical difficulties with a huge number of events in at least one cluster. The positive stable model shows analogous difficulties but an ad-hoc solution is implemented, whereas the gamma model is very resistant due to the simplicity of its Laplace transform.

Keywords: parametric frailty models, survival analysis, gamma, positive stable, inverse Gaussian, Weibull, inverse Weibull, Fréchet, exponential, Gompertz, loglogistic, lognormal, logskewnormal, skew-normal, R, **parfm**.

This vignette is an up-to-date versuib of the paper published in the Journal of Statistical Software in 2012: Munda, Rotolo, and Legrand (2012).

1. Introduction

Survival data, or time-to-event data, measure the time elapsed from a given origin to the occurrence of an event of interest. The observation of survival data is very common in the medical fields where, for instance, the clinician is interested in the time to relapse of a pathology after the therapy. However, the researcher cannot always observe the event due to censoring. Right-censoring occurs when the time of interest cannot be observed but only a lower bound is available. Particular techniques are therefore required as described by a number of textbooks, e.g., Klein and Moeschberger (2003).

Most commonly, survival data are handled by means of the proportional hazards regression model popularised by Cox (1972). But correct inference based on those proportional hazards models needs independent and identically distributed samples. Nonetheless, subjects may be exposed to different risk levels, even after controlling for known risk factors; this is because some relevant covariates are often unavailable to the researcher or even unknown (univariate

case). Also, the study population may be divided into clusters so that subjects from the same cluster behave more cohesively than subjects from different clusters (multivariate case). Lots of examples of clustered survival data arise from large-scale clinical trials in which patients are recruited at several hospital centres (Duchateau, Janssen, Lindsey, Legrand, Nguti, and Sylvester 2002; Glidden and Vittinghoff 2004). Another classical example is the analysis of lifetimes of matched human organs such as eyes or kidneys.

The frailty model, introduced in the biostatistical literature by Vaupel, Manton, and Stallard (1979), and discussed in details by Hougaard (2000), Duchateau and Janssen (2008), and by Wienke (2010), accounts for this heterogeneity in baseline. It is an extension of the proportional hazards model in which the hazard function depends upon an unobservable random quantity, the so-called frailty, that acts multiplicatively on it.

The gamma frailty model assumes a gamma distribution for the frailties. Arguably, this is the most popular frailty model due to its mathematical tractability. The lognormal frailty model is also well-liked for its strong link with generalised linear mixed models. Other frailty distributions include the positive stable and the inverse Gaussian. All of these are reviewed by Duchateau and Janssen (2008, Chapter 4).

Of particular interest in the multivariate case is the association between related event times. Indeed, different dependence structures result from different frailty distributions (Hougaard 1995). In particular, positive stable frailties typically generate very strong dependence initially while, at equal global dependence, gamma frailties lead to stronger dependence at late times, and inverse Gaussian frailties are in between the two. These three distributions therefore cover a wide range of association structures in the data.

Estimation of the frailty model can be parametric or semi-parametric. In the former case, a parametric density is assumed for the event times, resulting in a parametric baseline hazard function. Estimation is then conducted by maximising the marginal log-likelihood (see Section 2). In the second case, the baseline hazard is left unspecified and more complex techniques are available to approach that situation (Cortinas Abrahantes, Legrand, Burzykowski, Janssen, Ducrocq, and Duchateau 2007). Even though semi-parametric estimation offers more flexibility, the parametric estimation will be more powerful if the form of the baseline hazard is somehow known in advance. Further, the estimation technique is much simpler.

Slowly but surely, a variety of estimation procedures becomes available in standard statistical software. In R (R Development Core Team 2012), the `coxph()` function from the **survival** package (Therneau 2012b) handles the semi-parametric model with gamma and lognormal frailties. Important options supported by `coxph()` and its output are described in details by Therneau and Grambsch (2000, Chapter 9). Recently, the **frailtypack** package (Gonzalez, Rondeau, Mazroui, and Diakite 2012) by Rondeau and Gonzalez (2005) and Rondeau, Mazroui, and Gonzalez (2012) has been updated and it stands now for gamma frailty models with a semi-parametric estimation but also with a parametric approach using the Weibull baseline hazard. Other R packages include **coxme** (Therneau 2012a) and **phmm** (Donohue and Xu 2012). These two perform semi-parametric estimation in the lognormal frailty model. SAS (SAS Institute Inc. 2011) also deals with the lognormal distribution. On the one hand, `proc phreg` can now fit the semi-parametric lognormal frailty model. On the other hand, `proc nlmixed` deals with the parametric version by using Gaussian quadrature to approach the marginal likelihood; see, e.g., Duchateau and Janssen (2008, Example 4.16). In the parametric setting, STATA (StataCorp 2011) provides some flexibility. The `streg` command (Gutierrez 2002)

is able to perform maximum likelihood estimation with various choices of baselines: exponential, Weibull, Weibull, Gompertz, lognormal, loglogistic, and generalised gamma. Take notice, however, that STATA fits the accelerated failure time model. Still, with exponential or Weibull baselines, both the proportional hazards and the accelerated failure time representations are allowed. As for the frailty distribution, the gamma and the inverse Gaussian are the only two that are supported. On a side note, Bayesian analyses can be conducted in WinBUGS (Spiegelhalter, Thomas, Best, and Lunn 2003); see, e.g., Duchateau and Janssen (2008, Example 6.4). For a deeper overview of who supports what, and for a comparison of some of the aforementioned functions, see Hirsch and Wienke (2012).

Hereinbelow, we illustrate **parfm** (Rotolo and Munda 2012), a new R package that fits the gamma, the positive stable, the inverse Gaussian, and lognormal proportional hazards frailty models with either exponential, Weibull, inverse Weibull (Fréchet), Gompertz, lognormal, log-skewNormal, or loglogistic baseline. The main advantage of **parfm** therefore relies on the large choice of frailty distributions and parametric baseline hazards it supports. Parameter estimation is done by maximising the marginal log-likelihood.

The model and the marginal log-likelihood are shown in Section 2. There, we also outline the estimation method, while Sections 2.1–2.4 provide details for the three frailty distributions supported by **parfm**. In Section 3, we apply **parfm** to a real dataset in order to illustrate its use and its output. Section 4 concludes with remarks.

2. Model estimation

From a modelling point of view, the multivariate model includes the univariate. Because of this, we shall mainly refer to the first. However, they are used in two different contexts: in the former case, the frailty distribution variability is related to a measure of dependence between clustered subjects, whereas it is rather interpreted as a measure of overdispersion in the latter.

Model. The frailty model is defined in terms of the conditional hazard

$$h_{ij}(t \mid u_i) = h_0(t) u_i \exp\left(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}\right),$$

with $i \in I = \{1, \ldots, G\}$ and $j \in J_i = \{1, \ldots, n_i\}$, where $h_0(\cdot)$ is the baseline hazard function, u_i the frailty term of all subjects in group i, $\boldsymbol{x}_{ij}$ the vector of covariates for subject j in group i, and $\boldsymbol{\beta}$ the vector of regression coefficients.

If the number of subjects n_i is 1 for all groups, then the univariate frailty model is obtained (Wienke 2010, Chapter 3), otherwise the model is called the shared frailty model (Hougaard 2000, Chapter 7; Duchateau and Janssen 2008) because all subjects in the same cluster share the same frailty value u_i.

Baseline hazard. Under the parametric approach, the baseline hazard is defined as a parametric function and the vector of its parameters, say $\boldsymbol{\psi}$, is estimated together with the regression coefficients and the frailty parameter(s). A bunch of possibilities are considered in the literature; in the **parfm** package the exponential, Weibull, inverse Weibull (Fréchet), Gompertz, lognormal, logSkewNormal (Azzalini 1985), and loglogistic distributions are available. Table 1 shows the hazard and cumulative hazard functions for each of these distributions.

Table 1: Parametric distributions available in **parfm** for the baseline hazard. With the log(skew)normal, $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote the probability density and the cumulative distribution functions of a standard normal random variable. With the log-SkewNormal, $SN(\cdot)$ denotes the cumulative distribution functions of a skewNormal random variable with parameters ξ, ω, and α (Azzalini 1985).

distribution	$\boldsymbol{h_0(t)}$	$\boldsymbol{H_0(t)} = \int_0^t h_0(s)\mathrm{d}s$	$\boldsymbol{S_0(t)} = \exp[-H_0(s)]$	**parameters**
exponential	λ	λt	$\exp(-\lambda t)$	$\lambda > 0$
Weibull	$\lambda\rho t^{\rho-1}$	λt^{ρ}	$\exp(-\lambda t^{\rho})$	$\rho, \lambda > 0$
inverse Weibull (Fréchet)	$\dfrac{\lambda\rho t^{-(\rho+1)}}{\exp(\lambda t^{-\rho}) - 1}$	$-\log\left[1 - \exp(-\lambda t^{-\rho})\right]$	$1 - \exp(-\lambda t^{-\rho})$	$\rho, \lambda > 0$
Gompertz	$\lambda \exp(\gamma t)$	$\frac{\lambda}{\gamma}\left(\exp(\gamma t) - 1\right)$	$\exp\left[-\frac{\lambda}{\gamma}\left(\exp(\gamma t) - 1\right)\right]$	$\gamma, \lambda > 0$
lognormal	$\dfrac{\phi\left(\frac{\log(t)-\mu}{\sigma}\right)}{\sigma t\left[1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)\right]}$	$-\log\left[1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)\right]$	$1 - \Phi\left(\frac{\log(t)-\mu}{\sigma}\right)$	$\mu \in \mathbb{R}$, $\sigma > 0$
log-SkewNormal	$\dfrac{2\phi\left(\frac{\log(t)-\xi}{\omega}\right)\Phi\left(\alpha\frac{\log(t)-\xi}{\omega}\right)}{t\omega\left[1 - SN\left(\log(t); \xi, \omega, \alpha\right)\right]}$	$-\log\left[1 - SN\left(\log(t); \xi, \omega, \alpha\right)\right]$	$1 - SN\left(\log(t); \xi, \omega, \alpha\right)$	$\xi, \alpha \in \mathbb{R}$, $\omega > 0$
loglogistic	$\dfrac{\exp(\alpha)\kappa t^{\kappa-1}}{1 + \exp(\alpha)t^{\kappa}}$	$\log\left[1 + \exp(\alpha)t^{\kappa}\right]$	$\dfrac{1}{1 + \exp(\alpha)t^{\kappa}}$	$\alpha \in \mathbb{R}$, $\kappa > 0$

Frailty distribution. The frailty u_i is an unobservable realisation of a random variable U with probability density function $f(\cdot)$—the frailty distribution. Since u_i multiplies the hazard function, U has to be non-negative. Another constraint is further needed for identifiability reasons, similar to the zero-mean constraint of a random effect in a standard linear mixed model. More specifically, the mean of U is typically restricted to unity when possible (i.e., when $\mathsf{E}(U)$ exists) in order to separate the baseline hazard from the overall level of the random frailties.

Various frailty distributions have been proposed in the literature (Duchateau and Janssen 2008, Chapter 4). Hereinafter, we shall focus on the gamma, the positive stable, and the inverse Gaussian frailty distributions. In all of these three, a single heterogeneity parameter (denoted either θ or ν) indexes the degree of dependence. In the following, ξ is used as a generic notation to denote either θ or ν.

Data. For right-censored clustered survival data, the observation for subject $j \in J_i = \{1, \ldots, n_i\}$ from cluster $i \in I = \{1, \ldots, G\}$ is the couple $\boldsymbol{z}_{ij} = (y_{ij}, \delta_{ij})$, where $y_{ij} = \min(t_{ij}, c_{ij})$ is the minimum between the survival time t_{ij} and the censoring time c_{ij}, and where $\delta_{ij} = I(t_{ij} \leq c_{ij})$ is the event indicator. Covariate information may also have been collected; in this case, $\boldsymbol{z}_{ij} = (y_{ij}, \delta_{ij}, \boldsymbol{x}_{ij})$, where $\boldsymbol{x}_{ij}$ denote the vector of covariates for the ij-th observation. Further, if left-truncation is also present, truncation times τ_{ij} are gathered in the vector $\boldsymbol{\tau}$.

Likelihood. In the parametric setting, estimation is based on the marginal likelihood in which the frailties have been integrated out by averaging the conditional likelihood with respect to the frailty distribution. Under assumptions of non-informative right-censoring and of independence between the censoring time and the survival time random variables, given the covariate information, the marginal log-likelihood of the observed data $\boldsymbol{z} = \{\boldsymbol{z}_{ij}; i \in I, j \in J_i\}$ can be written as (van den Berg and Drepper 2016)

$$\begin{aligned}\ell_{\text{marg}}(\boldsymbol{\psi}, \boldsymbol{\beta}, \xi; \boldsymbol{z} \mid \boldsymbol{\tau}) = \sum_{i=1}^{G} \Bigg\{ & \left[\sum_{j=1}^{n_i} \delta_{ij} \left(\log(h_0(y_{ij})) + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} \right) \right] \\ & + \log \left[(-1)^{d_i} \mathcal{L}^{(d_i)} \left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}) \right) \right] \\ & - \log \left[\mathcal{L} \left(\sum_{j=1}^{n_i} H_0(\tau_{ij}) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}) \right) \right] \Bigg\}, \end{aligned} \tag{1}$$

with $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ the number of events in the i-th cluster, and $\mathcal{L}^{(q)}(\cdot)$ the q-th derivative of the Laplace transform of the frailty distribution defined as

$$\mathcal{L}(s) = \mathsf{E}\big[\exp(-Us)\big] = \int_0^\infty \exp(-u_i s) f(u_i) \mathrm{d}u_i, \qquad s \geq 0.$$

Estimation. Estimates of $\boldsymbol{\psi}$, $\boldsymbol{\beta}$, and $\boldsymbol{\xi}$ are obtained by maximising the marginal log-likelihood 1; this can easily be done if one is able to compute higher order derivatives $\mathcal{L}^{(q)}(\cdot)$

of the Laplace transform up to $q = \max\{d_1, \ldots, d_G\}$. Symbolic differentiation might be performed in R, but is impractical here, mainly because this is very time consuming. Therefore, explicit formulas are rather desirable. Further, they will be used in the calculation of predictions as shown below.

Prediction. Besides parameter estimates, prediction of frailties are sometimes desirable. As an aside, they are needed at each expectation step of the expectation-maximisation (EM) algorithm that fits the semi-parametric frailty model.

The frailty term u_i can be predicted by $\hat{u}_i = \mathsf{E}\left(U \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\beta}}, \hat{\xi}\right)$, with $\boldsymbol{z}_i$ and $\boldsymbol{\tau}_i$ the data and the truncation times of the i-th cluster. This conditional expectation can be achieved as

$$\mathsf{E}\left(U \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi\right) = -\frac{\mathcal{L}^{(d_i+1)}\left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})\right)}{\mathcal{L}^{(d_i)}\left(\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})\right)},$$

which can be seen from Appendix A.2, together with $\mathsf{E}\left[U^q \exp(-Us)\right] = (-1)^q \mathcal{L}^{(q)}(s)$.

Outline. In Sections 2.1–2.4 we illustrate the three frailty distributions which are available in the **parfm** package: the gamma, the positive stable, the inverse Gaussian, and the lognormal. Note that the Laplace transform of a lognormal random variable does not exist in a closed form. Hence, Equation 1 requires numerical approximation in that case, which is not considered here.

2.1. Gamma frailty

A gamma frailty term is a random variable $U \sim \text{Gam}^\star(\theta)$ with probability density function

$$f(u) = \frac{\theta^{-\frac{1}{\theta}} u^{\frac{1}{\theta}-1} \exp\left(-u/\theta\right)}{\Gamma(1/\theta)}, \qquad \theta > 0,$$

where $\Gamma(\cdot)$ is the gamma function. It corresponds to a gamma distribution $\text{Gam}(\mu, \theta)$ with μ fixed to 1 for identifiability. Its variance is then θ.

The associated Laplace transform is given by

$$\mathcal{L}(s) = (1 + \theta s)^{-\frac{1}{\theta}}, \qquad s \geq 0,$$

and it is easy to show that, for $q \geq 1$,

$$\mathcal{L}^{(q)}(s) = (-1)^q (1 + \theta s)^{-q} \left[\prod_{l=0}^{q-1} (1 + l\theta)\right] \mathcal{L}(s).$$

Therefore, in Equation 1, we have

$$\log\left((-1)^q \mathcal{L}^{(q)}(s)\right) = -\left(q + \frac{1}{\theta}\right) \log(1 + \theta s) + \sum_{l=0}^{q-1} \log(1 + l\theta). \tag{2}$$

For the gamma distribution, the Kendall's tau (Hougaard 2000, Section 4.2), which measures the association between any two event times from the same cluster in the multivariate case, can be computed as

$$\tau = \frac{\theta}{\theta + 2} \in (0, 1).$$

2.2. Positive stable frailty

Hougaard (2000, Section A.3.3) introduces the positive stable distributions as a family with two parameters: a scale $\delta > 0$ and the so-called index $\alpha < 1$. Imposing $\delta = \alpha$, the positive stable frailty distribution $\mathrm{PS}^\star(\nu)$ is obtained, with $\nu = 1 - \alpha$.

The associated probability density function is then

$$f(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{\Gamma(k(1-\nu)+1)}{k!} \left(-u^{\nu-1}\right)^k \sin((1-\nu)k\pi), \qquad \nu \in (0, 1).$$

The mean and variance are both undefined. Therefore, the heterogeneity parameter ν does not correspond to the variance of the frailty term. Because of that, we intentionally call it ν instead of θ to avoid misinterpretation.

In contrast to the probability density function, the associated Laplace transform takes a very simple form,

$$\mathcal{L}(s) = \exp\left(-s^{1-\nu}\right), \qquad s \geq 0,$$

and Wang, Klein, and Moeschberger (1995) found that, for $q \geq 1$,

$$\mathcal{L}^{(q)}(s) = (-1)^q \left((1-\nu)s^{-\nu}\right)^q \left[\sum_{m=0}^{q-1} \Omega_{q,m} s^{-m(1-\nu)}\right] \mathcal{L}(s),$$

where the $\Omega_{q,m}$'s are polynomials of degree m, given recursively by

$$\begin{aligned} \Omega_{q,0} &= 1, \\ \Omega_{q,m} &= \Omega_{q-1,m} + \Omega_{q-1,m-1} \left\{\frac{q-1}{1-\nu} - (q-m)\right\}, \qquad m = 1, \ldots, q-2, \\ \Omega_{q,q-1} &= (1-\nu)^{1-q} \frac{\Gamma(q-(1-\nu))}{\Gamma(\nu)}. \end{aligned} \tag{3}$$

It follows that

$$\log\left((-1)^q \mathcal{L}^{(q)}(s)\right) = q\left(\log(1-\nu) - \nu\log(s)\right) + \log\left[\sum_{m=0}^{q-1} \Omega_{q,m} s^{-m(1-\nu)}\right] - s^{1-\nu}. \tag{4}$$

With clustered data, the Kendall's tau for positive stable distributed frailties is

$$\tau = \nu \in (0, 1).$$

2.3. Inverse Gaussian frailty

The inverse Gaussian frailty distribution $\mathrm{IG}^\star(\theta)$ has density

$$f(u) = \frac{1}{\sqrt{2\pi\theta}} u^{-\frac{3}{2}} \exp\left(-\frac{(u-1)^2}{2\theta u}\right), \qquad \theta > 0.$$

The mean and the variance are 1 and θ, respectively. For the Laplace transform, one has

$$\mathcal{L}(s) = \exp\left(\frac{1}{\theta}\left(1 - \sqrt{1+2\theta s}\right)\right), \qquad s \geq 0,$$

and, for $q \geq 1$,

$$\mathcal{L}^{(q)}(s) = (-1)^q (2\theta s + 1)^{-\frac{q}{2}} \frac{K_{q-(1/2)}\left(\sqrt{2\theta^{-1}(s+\frac{1}{2\theta})}\right)}{K_{1/2}\left(\sqrt{2\theta^{-1}(s+\frac{1}{2\theta})}\right)} \mathcal{L}(s), \tag{5}$$

where K is the modified Bessel function of the second kind (Hougaard 2000, Section A.4.2)

$$K_\gamma(\omega) = \frac{1}{2}\int_0^\infty t^{\gamma-1} \exp\left\{-\frac{\omega}{2}\left(t + \frac{1}{t}\right)\right\} \mathrm{d}t, \qquad \gamma \in \mathbb{R}, \omega > 0.$$

The proof of this result, given in Appendix A.1, sketches a general constructive method to obtain the derivatives of the Laplace transform for any distribution for which the moments of $U \mid \boldsymbol{z}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi$, the conditional frailty given the data, are known.

Noting that $K_{1/2}(\omega) = \sqrt{\frac{\pi}{2\omega}}\exp(-\omega)$, we have

$$\log\left((-1)^q \mathcal{L}^{(q)}(s)\right) = -\frac{q}{2}\log(2\theta s+1) + \log\left(K_{q-(1/2)}(z)\right) - \left[\frac{1}{2}\left(\log\left(\frac{\pi}{2z}\right)\right) - z\right] + \frac{1}{\theta}\left(1 - \sqrt{1+2\theta s}\right), \tag{6}$$

with $z = \sqrt{2\theta^{-1}(s+\frac{1}{2\theta})}$.

With multivariate data, an inverse Gaussian distributed frailty yields a Kendall's tau given by

$$\tau = \frac{1}{2} - \frac{1}{\theta} + 2\frac{\exp(2/\theta)}{\theta^2}\int_{2/\theta}^\infty \frac{\exp(-u)}{u}\mathrm{d}u \in (0, 1/2).$$

2.4. Lognormal frailty

The lognormal frailty distribution $\mathrm{LN}^\star(\theta)$ has density

$$f(u) = (2\pi\theta)^{-\frac{1}{2}} u^{-1} \exp\left\{-\frac{(\log u)^2}{2\theta}\right\}, \tag{7}$$

with $\theta > 0$.

If $U \sim \text{LN}(\theta)$, then the Laplace transform does not exist in closed form. Consequently

$$\begin{aligned}\mathcal{L}^{(q)}(x) &= (-1)^q \int_0^\infty u^q \exp(-ux) f(u)\,\mathrm{d}u \\ &= (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_0^\infty u^q \exp(-ux) \frac{1}{u} \exp\Big(-\frac{1}{2\theta}\Big(\log(u)\Big)^2\Big)\,\mathrm{d}u\end{aligned}$$

needs to be approximated ($x \geq 0$). By using the change of variable $w = \log(u)$, we have

$$\begin{aligned}\mathcal{L}^{(q)}(x) &= (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^\infty \Big(\exp(w)\Big)^q \exp\Big(-\exp(w)x\Big) \exp\left(-\frac{w^2}{2\theta}\right)\,\mathrm{d}w \\ &= (-1)^q \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^\infty \exp\left\{qw - \exp(w)x - \frac{w^2}{2\theta}\right\}\,\mathrm{d}w.\end{aligned}$$

We approximate this by means of the Laplace approximation of integrals. Let

$$\begin{aligned}g(w;\, x, \theta) &:= -qw + \exp(w)x + \frac{w^2}{2\theta} \\ g^{(1)}(w;\, x, \theta) &:= \frac{\mathrm{d}g}{\mathrm{d}w}(w;\, x, \theta) = -q + \exp(w)x + \frac{w}{\theta} \\ g^{(2)}(w;\, x, \theta) &:= \frac{\mathrm{d}^2 g}{\mathrm{d}w^2}(w;\, x, \theta) = \exp(w)x + \frac{1}{\theta} > 0\end{aligned}$$

The approximation consists of replacing $g(\cdot)$ by the first three terms of its Taylor series expansion around some $\tilde{w}$,

$$g(w;\, x, \theta) \approx g(\tilde{w};\, x, \theta) + (w - \tilde{w}) g^{(1)}(\tilde{w};\, x, \theta) + \frac{(w-\tilde{w})^2}{2} g^{(2)}(\tilde{w};\, x, \theta)$$

The value of $\tilde{w}$ is chosen such that $g^{(1)}(\tilde{w};\, x, \theta) = 0$, so that $\mathcal{L}^{(q)}(x)$ can be approximated by

$$\begin{aligned}\mathcal{L}^{(q)}(x) \approx& (-1)^q \frac{1}{\sqrt{2\pi\theta}} \exp\left\{-g(\tilde{w};\, x, \theta)\right\} \\ &\times \int_{-\infty}^\infty \exp\left\{-\frac{(w-\tilde{w})^2}{2} g^{(2)}(\tilde{w};\, x, \theta)\right\}\,\mathrm{d}w \\ =& (-1)^q \frac{1}{\sqrt{\theta}} \exp\Big\{-g(\tilde{w};\, x, \theta)\Big\} \Big[g^{(2)}(\tilde{w};\, x, \theta)\Big]^{-1/2}\end{aligned}$$

where the last line follows by recognising the kernel of a normal density with mean $\tilde{w}$ and variance $1/g^{(2)}(\tilde{w};\, x, \theta)$. This is known as the Laplace approximation. The underlying idea is that the main contribution to the integral comes from where $g(\cdot)$ is close to its minimum. We refer to Goutis and Casella (1999) for further motivation and explanation of this kind of approximation.

3. Case study

We illustrate the **parfm** package with the very well-known `kidney` dataset that contains the recurrence times to kidney infection for 38 patients using portable dialysis equipment

(McGilchrist and Aisbett 1991).

```
R> R.Version()[["version.string"]]
```

```
[1] "R version 3.3.1 (2016-06-21)"
```

```
R> library("parfm")
R> packageDescription("parfm", fields="Version")
```

```
[1] "2.7"
```

The dataset is available in **parfm** via the command `data("kidney")` and it looks like the following:

```
R> head(kidney)
```

```
  id time status age sex disease frail
1  1    8      1  28   1   Other   2.3
2  1   16      1  28   1   Other   2.3
3  2   23      1  48   2      GN   1.9
4  2   13      0  48   2      GN   1.9
5  3   22      1  32   1   Other   1.2
6  3   28      1  32   1   Other   1.2
```

Each observation corresponds to a kidney, the variable `id` being the patient's code. The time from insertion of the catheter to infection or censoring is stored in `time` while `status` is 1 when infection has occurred and 0 for censored observations (catheters may be removed for reasons other than infection). Three covariates are available: `age`, the age of the patient in years, `sex`, being `1` for males and `2` for females, and `disease`, the disease type (`GN`, `AN`, `PKD` or `Other`). Finally `frail` is the frailty prediction from the original paper which fits a semi-parametric lognormal frailty model.

First and foremost, `sex` is recoded as a 0–1 indicator for ease of interpretation:

```
R> kidney$sex <- kidney$sex - 1
```

The hazard of infection will be modelled as a function of the patient's age and sex. Clearly, kidneys from the same patient cannot be considered independent. Therefore, the use of a shared frailty model is advisable, with clusters of size 2 corresponding to patients.

The `parfm()` function must have the following inputs. `formula`: a formula with an object of class `Surv` on the left-hand side; `cluster`: the cluster variable's name; `data`: the dataset; `dist`: the baseline hazard, either `exponential`, `weibull`, `gompertz`, `lognormal` or `loglogistic`; `frailty`: the frailty distribution, either `none`, `gamma`, `possta` or `ingau`.

Model estimation. The model with exponential baseline hazard and gamma frailty distribution is first fitted.

```
R> mod <- parfm(Surv(time, status) ~ sex + age, cluster="id",
+               data=kidney, dist="exponential", frailty="gamma")
R> mod
```

```
Frailty distribution: Gamma
Basline hazard distribution: Exponential
Loglikelihood: -333.248

       ESTIMATE SE    p-val
theta   0.301   0.157
lambda  0.025   0.015
sex    -1.485   0.398 <.001 ***
age     0.005   0.011 0.662
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.131
```

Standard errors are computed as the square roots of the diagonal elements of the observed information matrix. According to this model, `sex` has a significant impact on the hazard of infection while it is not affected by `age`. Conditional on the patient's frailty and on the age, the hazard of infection for a female at any time t is estimated to be $\exp(-1.485) \approx 0.227$ times that of a male, with Wald confidence interval

```
> ci.parfm(mod, level=0.05)["sex",]
```

```
  low    up
0.104 0.495
```

As for the heterogeneity parameter, it is estimated to be 0.301 which corresponds to a Kendall's tau equal to 0.131.

Frailty prediction. Prediction of frailties can be obtained via the `predict()` function, with the parametric frailty model object as unique argument. For instance, the predictions for the gamma–exponential model, `mod`, are obtained via the command

```
R> u <- predict(mod)
```

which returns an object of class `predict.parfm`. These predictions can easily be plotted (Figure 1) with the command `plot(u, sort="i")`.

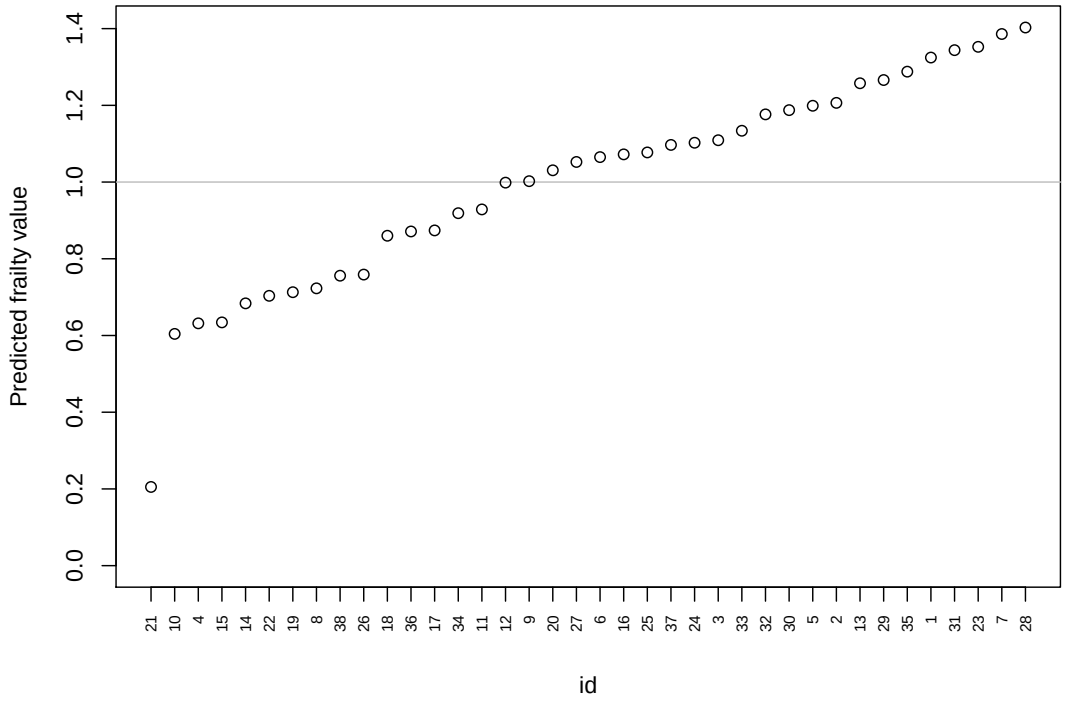

Figure 1: Prediction of frailties for the kidney dataset as given by the parametric gamma–exponential frailty model.

Comparison of different models. In some circumstances, it might be useful to easily obtain AIC and BIC values for a series of candidate models. This can be done using the `select.parfm()` function. Its use is similar to that of the `parfm()` function, but the `dist` and `frailty` values are vectors that contain all the alternatives to try.

```
R> kidney.parfm <- select.parfm(Surv(time, status) ~ sex + age,
+    cluster="id", data=kidney,
+    dist=c("exponential", "weibull", "loglogistic",
+           "lognormal", "logskewnormal"),
+    frailty=c("gamma", "ingau", "possta", "lognormal"))
R> kidney.parfm

 AIC:             gamma  ingau possta lognor
   exponential      674    676    682    675
   weibull          674    677    682    676
   loglogistic      685    685    686    685
   lognormal        679    679    681    679
   logskewnormal    681    681    682    681
```

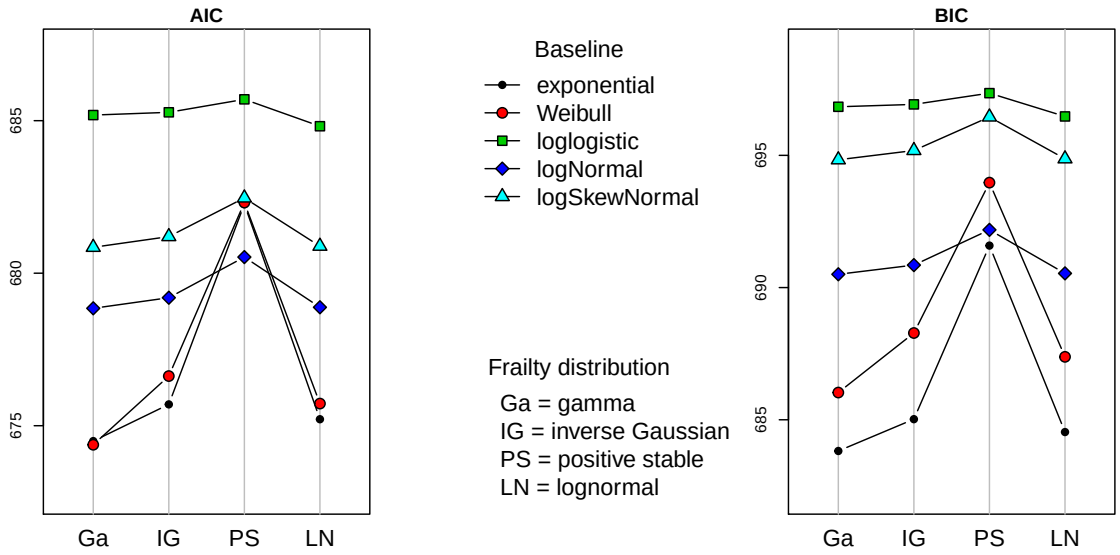

Figure 2: AIC and BIC values of **parfm** models for the kidney dataset.

```
BIC:              gamma  ingau possta lognor
  exponential       684    685    692    685
  weibull           686    688    694    687
  loglogistic       697    697    697    696
  lognormal         691    691    692    691
  logskewnormal     695    695    696    695
```

The results can be plotted (Figure 2) via the command `plot(kidney.parfm)`. In this particular example, the exponential baseline seems to be a good candidate.

As a comparison, the model with inverse Gaussian distributed frailties is fitted by changing the `frailty` argument into `'ingau'`.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="ingau")

Frailty distribution: Inverse Gaussian
Basline hazard distribution: Exponential
Loglikelihood: -333.85

       ESTIMATE SE    p-val
theta   0.375   0.259
lambda  0.022   0.013
sex    -1.310   0.373 <.001 ***
age     0.004   0.011 0.693
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.125
```

In this case, the conclusions drawn from the previous two models are essentially analogous. Consider now the model with the positive stable frailty distribution. In this example, it converges to a solution which is not valid ($\nu = 0$) with the default settings.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="possta")

Frailty distribution: Positive Stable
Basline hazard distribution: Exponential
Loglikelihood: -337.132

       ESTIMATE SE p-val
nu      0.000
lambda  0.012
sex    -0.885
age     0.004
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0
Warning message:
In parfm(Surv(time, status) ~ sex + age, cluster = "id", data = kidney,  :
  Error in solve.default(res$hessian) :
  Lapack routine dgesv: system is exactly singular
```

The default initial value for ν is $1/2$ in the case of positive stable frailties; it can be changed by means of the `iniFpar` option in `parfm()`. Let us try with $\nu = 0.25$.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="possta",
+        iniFpar=0.25)

Execution time: 1.71 second(s)

Frailty distribution: Positive Stable
Basline hazard distribution: Exponential
Loglikelihood: -336.182

       ESTIMATE SE    p-val
nu      0.112   0.084
lambda  0.014   0.008
sex    -0.951   0.348 0.006 **
age     0.004   0.011 0.698
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.112
```

The problem might also be fixed by changing the optimisation method (see `optimx()`; Nash and Varadhan 2011). By default it is set to `'BFGS'`, but it can be changed through the `method` option.

```
R> parfm(Surv(time, status) ~ sex + age, cluster="id",
+        data=kidney, dist="exponential", frailty="possta",
+        method="Nelder-Mead")

Execution time: 1.51 second(s)

Frailty distribution: Positive Stable
Basline hazard distribution: Exponential
Loglikelihood: -336.182

       ESTIMATE SE    p-val
nu      0.112   0.084
lambda  0.014   0.008
sex    -0.951   0.348 0.006 **
age     0.004   0.011 0.694
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Kendall's Tau: 0.112
```

In this example the results obtained by changing the optimisation method are the same as those obtained by changing the initial value of ν. When convergence problems occur, using different starting values and/or different optimisation methods is generally sufficient to find the global maximum of the marginal likelihood function.

Finally we provide a comparison with the semi-parametric model. As an example, we fit the semi-parametric model with gamma frailties via the `coxph()` function.

```
R> coxph(Surv(time, status) ~ sex + age +
+        frailty(id, distribution="gamma", eps=1e-11),
+        outer.max=15, data=kidney)

                          coef     se(coef) se2    Chisq DF   p
sex                       -1.58323 0.4594   0.3515 11.88  1.0 0.00057
age                        0.00522 0.0119   0.0088  0.19  1.0 0.66000
frailty(id, distribution                             22.96 12.9 0.04100

     Variance of random effect= 0.408   I-likelihood = -181.6
```

Estimates of regression parameters are quite similar to those of the exponential–gamma model, while the frailty variance is sensibly different, arguably because of the difference in how the baseline hazard is treated.

4. Discussion

To the best of our knowledge, parametric frailty models are currently especially handled in STATA by means of the `streg` command. With **parfm**, they are now readily fitted in R. Further, **parfm** provides the positive stable frailty distribution which is presently unavailable in STATA. Actually, except for a SAS macro, `ps_frail`, developed by Shu and Klein (1999) in the semi-parametric setting, we are not aware of another package that provides the positive stable frailty distribution.

The **parfm** package is flexible and easy to use. It provides five distributions for the baseline hazard and three frailty distributions. Parameter estimation is done by maximising the marginal log-likelihood given in Equation 1. The `optim()` function is employed, and its `method` option is passed to `parfm()` (with `method="BFGS"` by default). If not specified in the `inip` option, initial values for all but the heterogeneity parameter are obtained by fitting an unadjusted (i.e., without frailty) parametric proportional hazards model. The initial heterogeneity parameter can also be specified by the user via the `iniFpar` option; otherwise it is set to 1 when frailties follow a gamma or an inverse Gaussian distribution, or to 1/2 when they follow the positive stable distribution.

Additionally, when `frailty="none"`, `parfm()` fits the unadjusted parametric proportional hazards model, similar to `survreg()` (from the **survival** package) or to `phreg()` (from the **eha** package; Broström 2012). However, `survreg()` returns the parameter estimates in the log-linear model and `phreg()` uses yet another parametrisation (see the documentation). Often, the user has then to transform back the parameters and to employ the delta method in order to get estimates for the standard errors. The `parfm()` function directly uses the proportional hazards representation.

Nonetheless, **parfm** might reach its limits when at least one d_i, the number of events in the i-th cluster, $i \in \{1, \ldots, G\}$, is very large. First, consider the positive stable distribution and observe that, for a fixed value of $m \in \{1, \ldots, q-1\}$, $\Omega_{q,m}$ rapidly grows as q increases; see Equations 3. At the extreme, some of them might exceed the largest representable number in R. These are then stored as `Inf`. This, in turn, prevents the marginal log-likelihood 1 to be evaluated and hence maximised. On a side note, also the SAS macro `ps_frail` that implements the EM algorithm to fit the semi-parametric positive stable frailty model has analogous difficulties when the number of events is large (or even moderate). The following ad-hoc solution is implemented in **parfm**: in order to keep the polynomials $\Omega_{q,m}$'s reasonably small, they are divided by some factor 10^K which does not change the marginal log-likelihood except for an additive constant (equal to $s \times K \times \log(10)$). The value of K is specified via the `correct` option (default is `correct=0`, i.e., no correction) and `parfm()` returns the readjusted log-likelihood value. That solution serves the purpose for moderately large values of d_i (say up to about 200 events per cluster according to our experience, but it depends on the data, on the other parameters, and on the hardware characteristics). With the inverse Gaussian distribution, the Bessel function $K_{q-1/2}(z)$ in Equation 6 raises the same problem. Indeed, it explodes when z is small relative to q; see Figure 3. Currently, that distribution should, therefore, preferably be avoided when there are very large values of d_i (say above 200 events per cluster according to our experience, but, again, it depends on the data, on the other parameters and on the hardware characteristics). Moreover, $K_{q-1/2}(z)$ rapidly goes to zero as z increases. So, in case of very small apparent heterogeneity, $\theta \to 0$ which implies $z \to \infty$, $K_{q-1/2}(z)$ might be stored as 0 in R and hence $\log(K_{q-1/2}(z))$ cannot be computed. However,

as this problem occurs in the case of very small heterogeneity, this would rather suggest to fit the model with `frailty="none"`. When frailties are gamma distributed, which is by far the

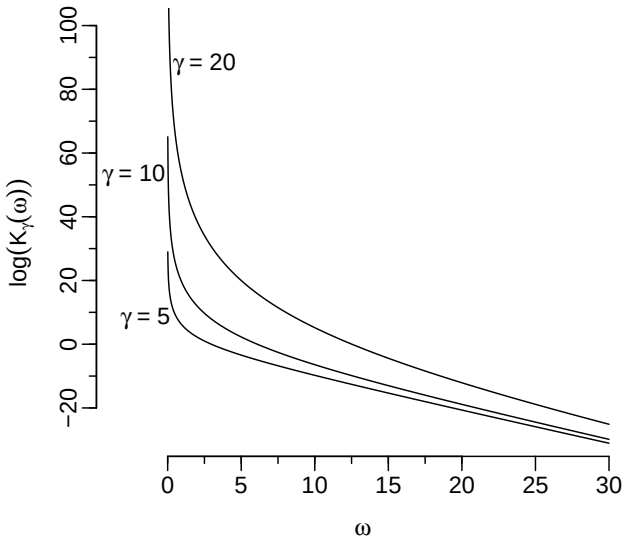

Figure 3: The logarithm of the Bessel function, $\log(K_\gamma(\omega))$, versus ω for different values of γ.

most popular assumption in common practice, the quantities involved in Equation 2 do not raise any worry. In practice, even when dealing with datasets with huge numbers of events per cluster, there is no real risk of exceeding the range of floating-point numbers.

References

Azzalini A (1985). "A class of distributions which includes the normal ones." *Scandinavian Journal of Statistics*, **12**(2), 171–178. URL http://www.jstor.org/stable/4615982.

Broström G (2012). ***eha**: Event History Analysis*. R package version 2.0-7, URL http://CRAN.R-project.org/package=eha.

Cortinas Abrahantes J, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L (2007). "Comparison of Different Estimation Procedures for Proportional Hazards Model with Random Effects." *Computational Statistics & Data Analysis*, **51**(8), 3919–3930. doi:10.1016/j.csda.2006.03.009.

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

Donohue M, Xu R (2012). ***phmm**: Proportional Hazards Mixed-Effects Model*. R package version 0.7-4, URL http://CRAN.R-project.org/package=phmm.

Duchateau L, Janssen P (2008). *The Frailty Model*. Springer-Verlag.

Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R, Sylvester R (2002). "The Shared Frailty Model and the Power for Heterogeneity Tests in Multicenter Trials." *Computational Statistics and Data Analysis*, **40**(3), 603–620. doi:10.1016/S0167-9473(02)00057-9.

Glidden D, Vittinghoff E (2004). "Modelling Clustered Survival Data From Multicentre Clinical Trials." *Statistics in Medicine*, **23**(3), 369–388. doi:10.1002/sim.1599.

Gonzalez JR, Rondeau V, Mazroui Y, Diakite A (2012). ***frailtypack**: General Frailty Models Using a Semi-Parametrical Penalized Likelihood Estimation or a Parametrical Estimation*. R package version 2.2-23, URL http://CRAN.R-project.org/package=frailtypack.

Goutis C, Casella G (1999). "Explaining the saddlepoint approximation." *The American Statistician*, **53**(3), 216–224. doi:10.1080/00031305.1999.10474463.

Gutierrez RG (2002). "Parametric Frailty and Shared Frailty Survival Models." *Stata Journal*, **2**(1), 22–44.

Hirsch K, Wienke A (2012). "Software for Semiparametric Shared Gamma and Log-Normal Frailty Models: An Overview." *Computer Methods and Programs in Biomedicine*, **107**(3). doi:10.1016/j.cmpb.2011.05.004.

Hougaard P (1995). "Frailty Models for Survival Data." *Lifetime Data Analysis*, **1**(3), 255–273.

Hougaard P (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag.

Klein JP, Moeschberger ML (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag.

McGilchrist CA, Aisbett CW (1991). "Regression with Frailty in Survival Analysis." *Biometrics*, **47**(2), 461–466. doi:10.2307/2532138.

Munda M, Rotolo F, Legrand C (2012). "`parfm`: Parametric Frailty Models in `R`." *Journal of Statistical Software*, **51**(1). doi:10.18637/jss.v051.i11.

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

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

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. doi:10.1016/j.cmpb.2005.06.010.

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." *Journal of Statistical Software*, **47**(1), 1–28. doi:10.18637/jss.v047.i04.

Rotolo F, Munda M (2012). ***parfm**: Parametric Frailty Models*. R package version 2.02, URL http://CRAN.R-project.org/package=parfm.

SAS Institute Inc (2011). *SAS/STAT Software, Version 9.3*. Cary, NC. URL http://www.sas.com/.

Shu Y, Klein JP (1999). "A SAS Macro for the Positive Stable Frailty Model." In *Proceedings of the Statistical Computing Section*, pp. 47–52.

Spiegelhalter D, Thomas A, Best N, Lunn D (2003). "WinBUGS User Manual." MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK, URL http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

StataCorp (2011). *STATA 12 Base Reference Manual*. College Station, TX. URL http://www.stata.com/.

Therneau T (2012a). ***coxme**: Mixed Effects Cox Models*. R package version 2.2-3, URL http://CRAN.R-project.org/package=coxme.

Therneau T (2012b). ***survival**: Survival Analysis, Including Penalised Likelihood*. R package version 2.36-14, URL http://CRAN.R-project.org/package=survival.

Therneau TM, Grambsch PM (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.

van den Berg GJ, Drepper B (2016). "Inference for shared-frailty survival models with left-truncated data." *Econometric Reviews*, **35**(6), 1075–1098. doi:10.1080/07474938.2014.975640.

Vaupel JW, Manton KG, Stallard E (1979). "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography*, **16**(3), 439–454. doi:10.2307/2061224.

Wang ST, Klein JP, Moeschberger ML (1995). "Semi-parametric Estimation of Covariate Effects Using the Positive Stable Frailty Model." *Applied Stochastic Models and Data Analysis*, **11**(2), 121–133. doi:10.1002/asm.3150110203.

Wienke A (2010). *Frailty Models in Survival Analysis*. Chapman & Hall/CRC biostatistics series. Taylor and Francis.

A. Proofs

A.1. Derivatives of the Laplace transform of the inverse Gaussian frailty distribution

On the one hand, for any frailty distribution $f(u_i;\xi)$, the α-th moment of U ($\alpha \in \mathbb{N}$), conditional on the data from the i-th cluster and on the parameters, can be written in the form

$$\mathsf{E}(U^\alpha \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi) = \frac{\mathsf{E}\left(U^{d_i+\alpha} \exp\left(-UH_{i\cdot,c}(\boldsymbol{y}_i)\right)\right)}{\mathsf{E}\left(U^{d_i} \exp\left(-UH_{i\cdot,c}(\boldsymbol{y}_i)\right)\right)}, \tag{8}$$

with $H_{i\cdot,c}(\boldsymbol{y}_i) = \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})$.

This is a generalisation of a result found by Wang *et al.* (1995) which follows from Bayes's formula applied to $f(u_i \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi)$ in

$$\mathsf{E}\left(U^\alpha \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi\right) = \int_0^\infty u_i^\alpha f\left(u_i \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi\right) \mathrm{d}u_i$$

(see Appendix A.2 for more details). Now, since the expected values in the right-hand side of Equation 8 can be written in terms of derivatives of the Laplace transform

$$\mathsf{E}\left(U^q \exp(-sU)\right) = (-1)^q \mathcal{L}^{(q)}(s), \qquad q, s \geq 0,$$

we have that

$$\mathcal{L}^{(d_i+\alpha)}\left(H_{i\cdot,c}(\boldsymbol{y}_i)\right) = (-1)^\alpha \mathsf{E}(U^\alpha \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi)\ \mathcal{L}^{(d_i)}\left(H_{i\cdot,c}(\boldsymbol{y}_i)\right). \tag{9}$$

On the other hand, if $U \sim \mathrm{IG}^\star(\theta)$, then it is easy to show (Appendix A.3) that the conditional distribution of U given the data and the parameters is a generalised inverse Gaussian distribution:

$$U \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \theta \sim \mathrm{GIG}(\gamma_{\mathrm{GIG}}, \delta_{\mathrm{GIG}}, \theta_{\mathrm{GIG}})$$

with

$$\gamma_{\mathrm{GIG}} = d_i - \frac{1}{2}, \tag{10}$$

$$\theta_{\mathrm{GIG}} = \frac{1}{2\theta} + H_{i\cdot,c}(\boldsymbol{y}_i), \tag{11}$$

$$\delta_{\mathrm{GIG}} = \frac{1}{\sqrt{2\theta}}. \tag{12}$$

Hence (Hougaard 2000, Section A.3.6)

$$\mathsf{E}(U^\alpha \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi) = \left(\frac{\theta_{\mathrm{GIG}}^{1/2}}{\delta_{\mathrm{GIG}}}\right)^{-\alpha} \frac{K_{\gamma_{\mathrm{GIG}}+\alpha}(2\delta_{\mathrm{GIG}}\theta_{\mathrm{GIG}}^{1/2})}{K_{\gamma_{\mathrm{GIG}}}(2\delta_{\mathrm{GIG}}\theta_{\mathrm{GIG}}^{1/2})}. \tag{13}$$

Combining (9) and (13), Equation 5 is deduced. □

A.2. Conditional expectation of frailty terms

For ease of notation, let $H_{i\cdot,c}(\boldsymbol{y}_i)$ denote $\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})$.
For any frailty distribution $f(u_i;\xi)$ and for any $\alpha \in \mathbb{N}$, we have

$$\begin{aligned}
\mathsf{E}\left(U^\alpha \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi\right) &= \int_0^\infty u_i^\alpha f\left(u_i \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi\right) \mathrm{d}u_i \\
&= \int_0^\infty u_i^\alpha \frac{L_{\text{cond}}\left(\boldsymbol{\psi}, \boldsymbol{\beta} \mid \boldsymbol{\tau}_i, u_i; \boldsymbol{z}_i\right) f\left(u_i \mid \boldsymbol{\tau}_i; \xi\right)}{L_{\text{marg}}\left(\boldsymbol{\psi}, \boldsymbol{\beta}, \xi \mid \boldsymbol{\tau}_i; \boldsymbol{z}_i\right)} \mathrm{d}u_i,
\end{aligned}$$

with

$$\begin{aligned}
L_{\text{cond}}\left(\boldsymbol{\psi}, \boldsymbol{\beta} \mid \boldsymbol{\tau}_i, u_i; \boldsymbol{z}_i\right) &= \left[\prod_{j=1}^{n_i} \left(h_0(y_{ij}) u_i \exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})\right)^{\delta_{ij}}\right] \exp\left(-u_i H_{i\cdot,c}(\boldsymbol{y}_i)\right) \exp\left(u_i H_{i\cdot,c}(\boldsymbol{\tau}_i)\right), \\
f\left(u_i \mid \boldsymbol{\tau}_i; \xi\right) &= \frac{\exp\left(-u_i H_{i\cdot,c}(\boldsymbol{\tau}_i)\right) f(u_i;\xi)}{\mathcal{L}\left(H_{i\cdot,c}(\boldsymbol{\tau}_i)\right)}, \\
L_{\text{marg}}\left(\boldsymbol{\psi}, \boldsymbol{\beta}, \xi \mid \tau_i; \boldsymbol{z}_i\right) &= \int_0^\infty L_{\text{cond}}\left(\boldsymbol{\psi}, \boldsymbol{\beta} \mid \boldsymbol{\tau}_i, u_i; \boldsymbol{z}_i\right) f\left(u_i \mid \boldsymbol{\tau}_i; \xi\right) \mathrm{d}u_i.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathsf{E}\left(U^\alpha \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \xi\right) &= \frac{\displaystyle\int_0^\infty u_i^{d_i+\alpha} \exp\left(-u_i H_{i\cdot,c}(\boldsymbol{y}_i)\right) f(u_i;\xi) \mathrm{d}u_i}{\displaystyle\int_0^\infty u_i^{d_i} \exp\left(-u_i H_{i\cdot,c}(\boldsymbol{y}_i)\right) f(u_i;\xi) \mathrm{d}u_i} \\
&= \frac{E\left[U^{d_i+\alpha} \exp\left(-U H_{i\cdot,c}(\boldsymbol{y}_i)\right)\right]}{E\left[U^{d_i} \exp\left(-U H_{i\cdot,c}(\boldsymbol{y}_i)\right)\right]}.
\end{aligned}$$

□

A.3. Conditional distribution of inverse Gaussian frailty

Let $U \sim \text{IG}^{\star}(\theta)$, then the distribution of $U \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \theta$ is

$$\begin{aligned}
f(u_i \mid \boldsymbol{z}_i, \boldsymbol{\tau}_i; \boldsymbol{\psi}, \boldsymbol{\beta}, \theta) &= \frac{L_{\text{cond}}\left(\boldsymbol{\psi}, \boldsymbol{\beta} \mid \boldsymbol{\tau}_i, u_i; \boldsymbol{z}_i\right) f(u_i \mid \boldsymbol{\tau}_i; \theta)}{L_{\text{marg}}\left(\boldsymbol{\psi}, \boldsymbol{\beta}, \theta \mid \boldsymbol{\tau}_i; \boldsymbol{z}_i\right)} \\
&\propto L_{\text{cond}}\left(\boldsymbol{\psi}, \boldsymbol{\beta} \mid \boldsymbol{\tau}_i, u_i; \boldsymbol{z}_i\right) f(u_i \mid \boldsymbol{\tau}_i; \theta) \\
&\propto \left[\prod_{j=1}^{n_i} \left(h_0(y_{ij}) u_i \exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta})\right)^{\delta_{ij}}\right] \exp\left(-\sum_{j=1}^{n_i} H_0(y_{ij}) u_i \exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta})\right) \\
&\qquad \times \sqrt{\frac{1}{2\pi\theta}} u_i^{-\frac{3}{2}} \exp\left(-\frac{1}{2\theta u_i}(u_i - 1)^2\right) \\
&\propto u_i^{d_i - \frac{3}{2}} \exp\left(-u_i \left[\sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta})\right] - \frac{1}{2\theta} u_i - \frac{1}{2\theta}\frac{1}{u_i}\right) \\
&= u_i^{d_i - \frac{3}{2}} \exp\left(-\left(\frac{1}{2\theta} + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta})\right) u_i - \frac{1}{2\theta}\frac{1}{u_i}\right),
\end{aligned}$$

which is proportional to the density of a generalised inverse Gaussian distribution (Hougaard 2000, Section A.3.6) with parameters given by Equations 10–12. □

Affiliation:

Marco Munda
Arlenda
93 Chaussée verte
4470 Saint-Georges-sur-Meuse, Belgium
E-mail: marco.munda@arlenda.com

Federico Rotolo
Department of Biostatistics and Epidemiology
Gustave Roussy cancer center
114, Rue Edouard Vaillant
94805 Villejuif, France
E-mail: federico.rotolo@gustaveroussy.fr

Catherine Legrand
Institut de Statistique, Biostatistique et Sciences Actuarielles
Université catholique de Louvain
Voie du Roman Pays, 20
1348 Louvain-la-Neuve, Belgium
E-mail: catherine.legrand@uclouvain.be