Advances in frailty models

Balan, T.A.

Chapter 1

Introduction: A tutorial in frailty modeling

1.1 Introduction

Cox's proportional hazards model (Cox, 1972) is one of the most popular regression models for time to event outcomes. The hazard function, which may be used to describe the distribution of event times, is defined as the instantaneous probability of an event, given that the individual has not experienced the event previously. The proportional hazards assumption specifies that the ratio of the hazards between any two individuals is constant in time, and the shape of the hazard is given by a non-parametric "baseline hazard". If a model is perfectly specified, so that all possible relevant covariates are accounted for, then the baseline hazard reflects the randomness of the event time, given the value of covariates.

In practice however, it is rarely possible to account for all relevant covariates. Then the explanatory variables account for *observed heterogeneity*, and the unaccounted part is termed *unobserved heterogeneity*. If this is the case, then the estimated hazard for a specific set of covariates does not have an individual interpretation (Woodbury and Manton, 1977). Rather, it represents an average hazard function, where the average is taken at each time point over the individuals still alive at that time point. The effects of unobserved heterogeneity on life times were collectively referred to as *frailty* in demographic research (Vaupel, Manton, and Stallard, 1979). The frailty is an unobserved

This chapter is part of the manuscript under preparation: T.A. Balan, H. Putter. *A tutorial in frailty models: theory and practice*

individual random effect that acts multiplicatively on the hazard. The estimated spread of this random effect (e.g. variance) is an indication of the amount of unobserved heterogeneity. The frailty model quickly gained popularity in econometrics (Heckman and Singer, 1984), demographics (Vaupel and Yashin, 1985) and biostatistics (Aalen, 1988).

The Cox model, developed originally for *univariate* survival data, has been extended to a more general framework based on counting processes (Andersen and Gill, 1982). The resulting "extended Cox model" easily accommodates more complex data, such as correlated event times (*clustered failures*) or multiple events per individual (*recurrent events*). Frailty models based on the extended Cox model are referred to as *shared* frailty models (Nielsen et al., 1992; Andersen, Borgan, et al., 1993), as opposed to *univariate* frailty models in the simpler univariate survival data scenario.

For clustered failures, the estimated frailty variance describes unobserved heterogeneity between clusters. Within a cluster, the event times are assumed to be independent, given the frailty. Therefore, shared frailty models are often used to model the effect of unobserved risk factors that are specific to the clusters. For recurrent events, the estimated frailty variance describes unobserved heterogeneity between individuals, as in the univariate frailty case. Conditional on the frailty, the event history of an individual is typically modeled as a Poisson or renewal process. In all cases, frailty models involve the conditional specification of the hazard or intensity of the event process, as if the frailty would be observed. Consequently, the estimated covariate effects retain the interpretation of an individual effect.

Most theoretical results in frailty models have focused on the gamma frailty model. In particular, maximum likelihood estimators have been shown to be well behaved (Murphy, 1994; Murphy, 1995b). However, numerous other frailty distributions have been proposed in the literature (Hougaard, 1986a; Hougaard, 2000; Paddy Farrington, Unkel, and Anaya-Izquierdo, 2012). The real frailty distribution is almost impossible to be known in advance. It is therefore of interest to compare the characteristics of different frailty models in terms of the resulting population hazards (for univariate survival data) or within cluster correlation patterns (for clustered survival data).

The aim of this chapter is to provide an overview of theory and practice in the field of frailty models, while offering insight into the problems that are addressed by such models. In Section 1.2, we study the effects of unobserved heterogeneity in survival data, univariate frailty models and different frailty distributions. In Section 1.3, we analyze the effect of unobserved heterogeneity in clustered survival data and introduce the shared frailty model. We study different correlation structures and we discuss frailty models for recurrent events data. In Section 1.4, we discuss estimation and inference procedures for frailty models, we compare available software packages and we examine the representation of event history data in the R statistical software. In Section 1.5 we overview different extensions to the frailty models. Finally, in Section 1.6, we conclude with an outline of the rest of this thesis.

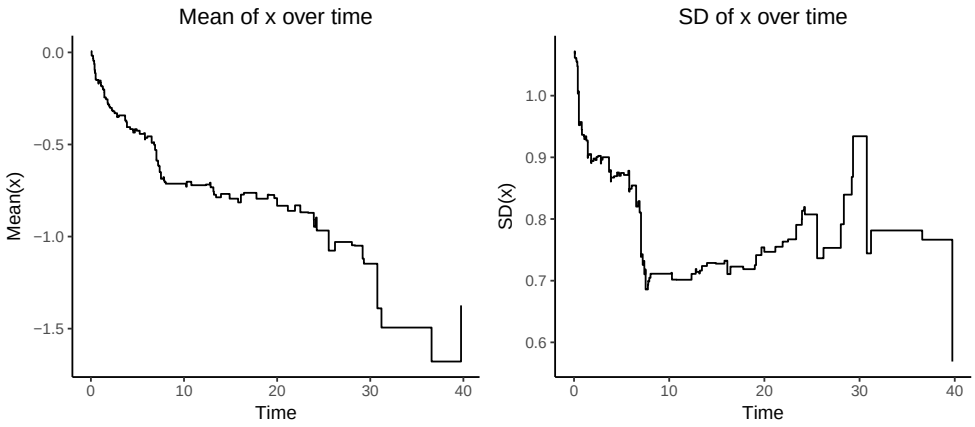

Figure 11: Changes in the mean and variance of a covariate x over time among survivors in a proportional hazards model.

1.2 Univariate frailty models

1.2.1 Heterogeneity in the Cox model

Heterogeneity over time

The Cox model specifies that the hazard as

$$\lambda(t \mid x) = \lambda_0(t) \exp(\beta^\top \mathbf{x}), \tag{1.1}$$

where β is a $p \times 1$ vector of regression coefficients, $\mathbf{x}$, is a $p \times 1$ vector of covariates and $h_0(t)$ is an unspecified baseline hazard function. The hazard functions of two individuals with covariate vectors $\mathbf{x}^*$ and $\tilde{\mathbf{x}}$ are equal only when $\beta^\top \mathbf{x}^* = \beta^\top \tilde{\mathbf{x}}$. The exponent $\exp(\beta_j)$ is the hazard ratio between an individual with $x_j = 1$ and an individual with $x_j = 0$. Time dependent covariates are easily accommodated in (1.1) and are discussed discussed in Section 1.4. Henceforth, we assume that $\mathbf{x}$ is time-constant.

The risk set at time t is composed of individuals that have not yet experienced the event of interest or have not yet been removed for other reasons, such as censoring. The distribution of the covariates among the individuals in the risk set changes in time. We illustrate this by considering the model (1.1) and only one covariate following a standard normal distribution $x \sim N(0, 1)$ and $\beta > 0$, so that individuals with larger values of x have a higher hazard. At time $t = 0$, the mean and variance of x are 0 and 1, respectively. As time passes, the risk set progressively comprises individuals with lower values of x. For this reason, the average value and the sample variance of x among the individuals at risk decreases over time.

This is illustrated by simulating a single sample of size $n = 100$, and a covariate $x \sim N(0, 1)$, with $\beta = 1$, $\lambda_0(t) \equiv 0.1$ and uniform censoring on $(20, 50)$. In this simulated

sample, at time 0, x had mean 0.007 and standard deviation 1.068. The estimated β was 0.943, with a standard error of 0.127. The mean and standard deviation of x among the individuals in the risk set are shown in Figure 11, as a function of time. The message is that, in time, the population of "survivors" (those still at risk) is more homogeneous and of a lower risk than the original risk set at time 0.

Heterogeneity due to missing covariates

The proportional hazards assumption in the Cox model (1.1) specifies that the ratio $\lambda(t \mid \mathbf{x}^*)$ divided by $\lambda(t \mid \tilde{\mathbf{x}})$ equals $\exp(\beta^\top(\mathbf{x}^* - \tilde{\mathbf{x}}))$, which does not depend on time. When this assumption is violated, the covariate effect β is time dependent. The true model is therefore

$$\lambda(t \mid \mathbf{x}) = \lambda_0(t) \exp(\beta(t)\mathbf{x})$$

with $\beta(t)$ not constant.

Assume that the model (1.1) is correct and $p \geq 2$. Then, if important covariates are omitted from the model, the proportional hazards assumption does not usually hold for the remaining covariates. This is illustrated by simulating a single large data set with sample size $n = 10,000$. Two independent covariates x_1 and x_2 are considered, both ~ $N(0,1)$, with $\beta_1 = \beta_2 = 1$, $\lambda_0 \equiv 1$ and uniform censoring on $(20, 50)$. The following output is shown from Cox models fitted with the standard **survival** package in R (Therneau and Grambsch, 2000). When both covariates are included into the model, the results are close to the simulation scenario, with both estimated regression coefficients close to 1:

```
## Call:
## c12 <- coxph(formula = Surv(time, status) ~ x1 + x2, data = d)
##
##       coef exp(coef) se(coef)    z      p
## x1 1.0016    2.7225   0.0138 72.7 <2e-16
## x2 1.0240    2.7843   0.0140 73.2 <2e-16
##
## Likelihood ratio test=9014  on 2 df, p=0
## n= 10000, number of events= 8240
```

No evidence of violation of the proportional hazards assumption is found, when a test based on Schoenfeld residuals is used (Grambsch and Therneau, 1994):

```
## Call: cox.zph(c12, transform = "identity")
##              rho  chisq     p
## x1       0.00101 0.0081 0.928
## x2      -0.00357 0.1050 0.746
## GLOBAL        NA 0.1510 0.927
```

However, if x_2 is omitted, the resulting estimate of the effect of x_1 is visibly smaller than 1:

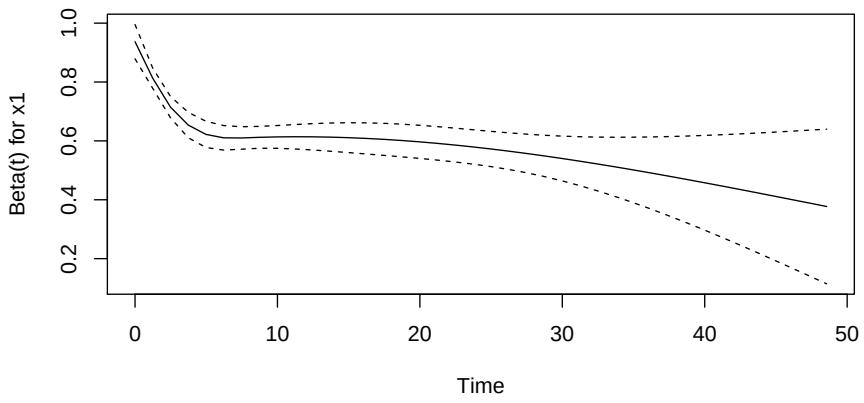

Figure 12: Plot of scaled Schoenfeld residuals based $\beta(t)$ induced by omitting a covariate from a proportional hazards model.

```
## Call:
## c1 <- coxph(formula = Surv(time, status) ~ x1, data = d)
##
##      coef exp(coef) se(coef)    z      p
## x1 0.7028    2.0195   0.0124 56.6 <2e-16
##
## Likelihood ratio test=3271  on 1 df, p=0
## n= 10000, number of events= 8240
```

Moreover, there is clear evidence against the proportional hazards assumption.

```
## Call: cox.zph(c1, transform = "identitiy")
##        rho chisq        p
## x1 -0.0852  55.3 1.06e-13
```

This is also illustrated by the plot of scaled Schoenfeld residuals of $\beta(t)$ in figure 12. It appears that the effect of x starts as close to the true value 1, and then decreases in time. The result given by the Cox model only with x_1 is an "average" effect in this case.

The explanation for the phenomenon illustrated in the simulated example is that the individual hazard is determined by the combined effect of x_1 and x_2. On average, the "high risk" individuals (high x_1, high x_2) are the first to have the event, followed by the "moderate risk" ones (high x_1 and low x_2, or low x_1 and high x_2), and eventually the "low risk" ones (low x_1 and low x_2). In particular, the individuals that survive up to a certain

time are more likely to have lower values of x_2. If x_2 is omitted from the model, this decrease in risk among the survivors must be accounted for only by x_1, thus reducing the perceived effect of the included covariate.

Conditional and marginal hazards

More generally, suppose that the proportional hazards model (1.1) holds for a covariate vector $\mathbf{x} = (\mathbf{x}_{\text{incl}}, \mathbf{x}_{\text{omit}})$ with covariate effects $\beta = (\beta_{\text{incl}}, \beta_{\text{omit}})$, so that the true model is

$$\lambda(t \mid \mathbf{x}) = \lambda_0(t) \exp(\beta_{\text{incl}}^\top \mathbf{x}_{\text{incl}} + \beta_{\text{omit}}^\top \mathbf{x}_{\text{omit}}). \tag{1.2}$$

Imagine that a Cox model is fitted only including x_{incl}. This will result in an estimated effect that is biased towards 0 and, usually, a violation of the proportional hazards assumption. In reality, it is possible that some relevant covariates are not measured (here represented by $\mathbf{x}_{\text{omit}}$). In this case, these omitted covariates are said to induce *unobserved heterogeneity*. The differences between individuals that are explained by $\mathbf{x}_{\text{incl}}$ are referred to as *observed heterogeneity*.

The $\lambda(t \mid \mathbf{x})$, as defined in model (1.2), is referred to as the *conditional* hazard, with β_{incl} summarizing the *conditional* effect of $\mathbf{x}_{\text{incl}}$. When unobserved heterogeneity is present, the resulting $\lambda(t \mid \mathbf{x}_{\text{incl}})$ is referred to as the *marginal* hazard (although it is marginal with respect to $\mathbf{x}_{\text{omit}}$ but still conditional on $\mathbf{x}_{\text{incl}}$). The estimated effect from the marginal model does not have an individual interpretation. Namely, $\lambda(t|\mathbf{x}_{\text{incl}})$ represents a weighted average of the individual hazards corresponding to those individuals in the risk set at time t, where the weighing is determined by the distribution of $\mathbf{x}_{\text{omit}}$ in this risk set.

Since the effect of $\mathbf{x}_{\text{omit}}$ cannot be directly observed, one can define the random variable $Z = \exp(\beta_{\text{omit}}^\top \mathbf{x}_{\text{omit}})$. Z is referred to as a "frailty" term that acts multiplicatively on the hazard.

1.2.2 The frailty model

In the univariate frailty model, the hazard of an individual with frailty Z is specified as

$$\lambda(t \mid Z) = Z\lambda(t). \tag{1.3}$$

For identifiability, Z is assumed to be scaled so that $\mathrm{E}Z = 1$. The second term in (1.3), $\lambda(t) \equiv \lambda(t \mid Z = 1)$, is the conditional hazard for an individual with $Z = 1$. We refer to $\lambda(t)$ simply as the conditional hazard. The conditional cumulative hazard is defined as $\Lambda(t) = \int_0^t \lambda(s)ds$. The conditional survival function for an individual with frailty Z is then given by

$$S(t \mid Z) = \exp(-Z\Lambda(t)).$$

The marginal survival curve associated with $\Lambda(t)$ is obtained by taking the expectation of $S(t|Z)$ with respect to Z,

$$\overline{S}(t) = \mathrm{E}[\exp(-Z\Lambda(t))]. \tag{1.4}$$

Unlike $S(t)$, $\overline{S}$ has a population averaged interpretation. If there are no covariates, $\overline{S}$ may be seen as a weighted average of individual survival curves, where the weighing depends on the distribution of Z. The hazard may be derived from the survival function as $\lambda(t) = d/dt[-\log S(t)]$. Therefore, the marginal hazard may be calculated as

$$\begin{aligned}\overline{\lambda}(t) &= \frac{\mathrm{E}\left[Z\exp(-Z\Lambda(t))\right]}{\mathrm{E}\left[\exp(-Z\Lambda(t)\right]}\lambda(t)\\ &= \mathrm{E}[Z|T \geq t]\lambda(t).\end{aligned}$$

A population averaged interpretation may also be given here: $\overline{\lambda}(t)$ may be seen as a weighted average of individual hazards *of individuals alive at time t*, where the weighing depends on the distribution of Z *among the individuals alive at time t*.

The conditional and marginal hazards are equal only if $\mathrm{E}[Z|T \geq t] = 1$ for all t. In other words, if all frailties Z are equal to 1. Otherwise, the frailty distribution among the survivors at time t behaves in a similar fashion with the distribution of an observed covariate among survivors, as shown in Section 1.2.1.

If observed covariates are also present, then it is usually assumed that the proportional hazards assumption holds conditional on the frailty, with $\lambda(t|Z) = Z\exp(\beta^\top \mathbf{x})\lambda_0(t)$. Then, the population averaged interpretations of $\overline{S}$ and $\overline{h}$ hold conditional on $\mathbf{x}$. In other words, for a hypothetical population of individuals with given covariate values $\mathbf{x}$. This is the same as the interpretation that is given to the marginal hazard in Section 1.2.1.

Regardless of whether the differences between individuals come from observed covariates $\mathbf{x}$ or from the frailty, individuals with higher hazards die earlier. Therefore, the population of survivors is more homogeneous and at a lower risk for events than the general population at time 0. The advantage of frailty models is that this is explicitly modeled. Before we further study the relation between marginal and conditional hazards in Section 1.2.4, we first discuss different frailty distributions in Section 1.2.3.

1.2.3 Frailty distributions

The Laplace transform

The distribution of a random variable $Z > 0$ can also be uniquely specified by its Laplace transform,

$$\mathcal{L}(c) = \mathrm{E}\left[\exp(-Zc)\right].$$

It is immediate that $\mathcal{L}(0) = 1$. The expectation of Z may be obtained as minus the derivative of $\mathcal{L}$ calculated in 0, $\mathrm{E}Z = -\mathcal{L}'(0)$. Furthermore, $\mathcal{L}''(0) = \mathrm{E}Z^2$ and higher order moments of Z can be obtained by taking further derivatives of $\mathcal{L}$. Denote the kth derivative of $\mathcal{L}$ as $\mathcal{L}^{(k)}$. The squared coefficient of variation, defined as $\mathrm{CV}^2 = \mathrm{var}[Z]/(\mathrm{E}[Z])^2$, may be expressed as

$$\mathrm{CV}^2[Z] = \frac{\mathcal{L}''(0)}{(\mathcal{L}'(0))^2} - 1.$$

In terms of the Laplace transform, the marginal survival function from (1.4) may be written as

$$\overline{S}(t) = \mathcal{L}(\Lambda(t)),$$

and the marginal hazard as

$$\overline{\lambda}(t) = \frac{d}{dt}[-\log S(t)] = -\frac{\mathcal{L}'(\Lambda(t))}{\mathcal{L}(\Lambda(t))}\lambda(t).$$

The Laplace transform of the frailty distribution of survivors can be obtained from Bayes' theorem:

$$\begin{aligned}\mathcal{L}_{Z|T\geq t}(c) &= \mathrm{E}\left[\exp(-Z\,c)|T \geq t\right] \\ &= \frac{\mathrm{E}[\exp(-Z(c+\Lambda(t)))]}{\mathrm{E}[\exp(-Z\Lambda(t))]} \\ &= \frac{\mathcal{L}(c+\Lambda(t))}{\mathcal{L}(\Lambda(t))}.\end{aligned} \tag{1.5}$$

The expectation, variance and squared coefficinet of variation of $Z|T \geq t$ follow as

$$\begin{aligned}\mathrm{E}[Z|T \geq t] &= -\frac{\mathcal{L}'(\Lambda(t))}{\mathcal{L}(\Lambda(t))}, \\ \mathrm{var}[Z|T \geq t] &= \frac{\mathcal{L}''(\Lambda(t))}{\mathcal{L}(\Lambda(t))} - \left(\frac{\mathcal{L}'(\Lambda(t))}{\mathcal{L}(\Lambda(t))}\right)^2 \\ \mathrm{CV}^2[Z|T \geq t] &= \frac{\mathcal{L}''(\Lambda(t))\mathcal{L}(\Lambda(t))}{(\mathcal{L}'(\Lambda(t)))^2} - 1.\end{aligned}$$

Infinitely divisible distributions

The *infinitely divisible* distributions are a family of distributions with tractable Laplace transform, specified as $\mathcal{L}(c) = \exp(-\alpha\psi(c;\gamma))$ with $\alpha > 0$ and $\gamma > 0$. The expectation and variance can be expressed as

$$\begin{aligned}&\mathrm{E}[Z|T \geq t] = \alpha\psi'(\Lambda(t);\gamma), \\ &\mathrm{var}[Z|T \geq t] = -\alpha\psi''(\Lambda(t);\gamma), \\ &\mathrm{CV}^2[Z|T \geq t] = -\frac{\psi''(\Lambda(t);\gamma)}{\alpha(\psi'(\Lambda(t)))^2}.\end{aligned} \tag{1.6}$$

The **gamma** distribution is a prominent member of the infinitely divisible family. The density of the gamma distribution with parameters $\theta > 0$ and $\eta > 0$ is given by $f(t;\theta,\eta) = \frac{\theta^\eta}{\Gamma(\eta)}t^{\eta-1}e^{-\theta t}$, where $\Gamma(\eta) = \int_0^\infty s^{\eta-1}e^{-s}ds$ is the gamma function. Its Laplace transform is given by

$$\mathcal{L}(c) = \left(\frac{\theta}{\theta + c}\right)^\eta,$$

which, in terms of (1.6), can be expressed as $\alpha = \eta$, $\theta = \gamma$, and $\psi(c;\gamma) = \log(\gamma + c) - \log(\gamma)$. By convention, the expectation of the frailty is fixed to 1, so the restriction $\theta = \eta$ is applied. In this parmaetrization, Z follows a gamma(θ, θ) distribution, with $\mathrm{E}[Z] = 1$ and $\mathrm{var}[Z] = \theta^{-1} = \xi$. The expectation and variance of the frailty distribution of the survivors is given through (1.6), resulting in

$$\mathrm{E}\left[Z|T \geq t\right] = \frac{\theta}{\theta + \Lambda(t)},$$

$$\mathrm{var}\left[Z|T \geq t\right] = \frac{\theta}{(\theta + \Lambda(t))^2}.$$

Both functions reach their maximum at $t = 0$, with expectation 1 and variance θ^{-1}, and decrease over time. For the gamma frailty, it is immediate that $\overline{\lambda}(t) \leq \lambda(t)$. In other words, the marginal hazard is always smaller than the hazard of an individual with frailty 1.

A more general family of infinitely divisible distributions is the **power variance function (PVF)** family, with the Laplace transform $\mathcal{L}$ described by

$$\mathcal{L}(c;\alpha,\gamma,m) = \exp\left(-\alpha\,\mathrm{sign}(m)\left\{1 - \left(\frac{\gamma}{\gamma + c}\right)^m\right\}\right)$$

where $\mathrm{sign}(m)$ is the sign of m, and $m > -1$ and $m \neq 0$. It was proposed in a series of papers (Hougaard, 1984; Hougaard, 1986a; Hougaard, 1986b) and is studied in detail in Hougaard (2000). To obtain $\mathrm{E}[Z] = 1$ and $\mathrm{var}[Z] = \theta^{-1}$, one can set $\alpha = \theta\mathrm{sign}(m)(m + 1)/m$ and $\gamma = \theta(m + 1)$. Particular cases of include:

- The gamma frailty, obtained as a limiting case when $m \longrightarrow 0$ with $m < 0$;
- The inverse Gaussian distribution, when $m = -1/2$;
- The so-called Hougaard distributions, when $m < 0$;
- The compound Poisson distribution, when $m > 0$, which has probability mass at 0. This is consistent with a scenario where a part of the population is not susceptible for the event of interest;
- The positive stable distribution, obtained as a limiting case when $\gamma \longrightarrow 0$. This distribution cannot be scaled to have $\mathrm{E}[Z] = 1$, so usually the $\alpha = 1$ restruction is imposed. Its expectation is infinite and the variance is not defined. However, the resulting Laplace transform takes the simple form $\mathcal{L}(c) = \exp(-\alpha c^\gamma)$, with $\alpha > 0$ and $\gamma \in (0, 1)$.

The **log-normal** distribution has often been used for frailty models, although it it not part of the PVF family. It is infinitely divisible, but the corresponding expression of ϕ cannot be expressed in closed form. Consequently, its Laplace transform and expressions for the distribution of survivors are not easily obtained. Its popularity stems from the

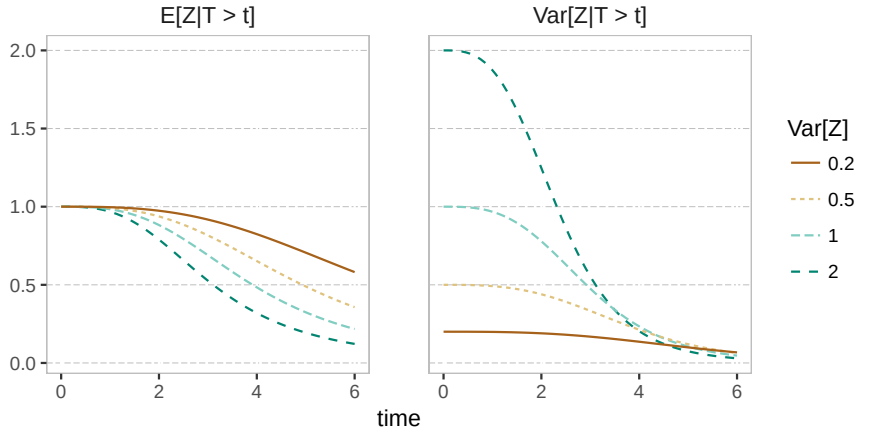

Figure 13: Frailty distribution of survivors, gamma frailty, $\lambda(t) = t^2/20$.

normal random effects in linear models. The log-normal frailty is usually parametrized with the $\mathrm{E}[\log Z] = 0$ and $\mathrm{var}[\log Z] = \sigma^2$, corresponding to a normally distributed random effect on the scale of the covariates. If matched by mean and variance, it is virtually indistinguishable from the inverse Gaussian distribution. Other families of distributions, such as the Addams and Kummer families of distributions were also introduced in the context of frailty models (Aalen, Borgan, and Gjessing, 2008; Paddy Farrington, Unkel, and Anaya-Izquierdo, 2012).

1.2.4 Frailty effects

The different frailty distributions discussed in Section 1.2.3 represent different ways of expressing unobserved heterogeneity. Different frailty distributions lead to different selection effects . Moreover, they impact the marginal effect of the observed covariates in different ways, generalizing the phenomenon illustrated in Section 1.2.1. An advantage of the PVF family of distributions and their closed form Laplace transforms is that it facilitates the study of these phenomenons (Aalen, 1988; Aalen, 1994; Vaupel and Yashin, 1985). An overview may be found in Aalen, Borgan, and Gjessing (2008, ch. 6).

The selection effect In Section 1.2.3, it was shown that, for the gamma frailty model, the expectation and variance of the frailty distribution of the survivors decreases in time. In Figure 13, we show the expectation and the variance of $\mathrm{E}[Z|T \geq t]$, when the conditional hazard is given by $\lambda(t) = t^2/20$, for variances 0.2, 0.5, 1 and 2.

It follows that the marginal hazard appears as a "dragged down" version of the conditional hazard, similar to Figure 11. This selection effect is stronger if the frailty variance is larger. In particular, the marginal hazard may appear to grow, reach a peak and then

decrease beyond a time point, even if the conditional hazard is increasing. As in Section 1.2.1, the explanation is that individuals with a higher hazard die earlier, on average, than individuals with a lower hazard. In particular, this is true for all frailty distributions discussed in Section 1.2.3. For example, for the compound Poisson distribution, when individuals with frailty 0 never experience the event of interest, the marginal hazard will eventually decrease towards 0 after some time point. The point made in Section 1.2.1 is essential here as well: in the presence of unobserved heterogeneity, the marginal hazard has a population averaged rather than an individual interpretation.

The marginal hazard ratio In Section 1.2.1, we illustrated that, when important covariates are omitted in a Cox model, the marginal effect of the remaining covariates is time dependent. The same phenomenon happens with the marginal covariate effect in the case of frailty models. Suppose that only one observed covariate is present, $x \in \{0, 1\}$, and that the frailty model (1.3) is true. Then, e^{β} is the hazard ratio between two individuals with the same frailty, one with $x = 1$, the other with $x = 0$. The marginal hazards for the two groups defined by x are given by

$$\begin{aligned}\overline{\lambda}_0(t) &= \mathrm{E}[Z|T \geq t, x = 0]\, \lambda_0(t),\\ \overline{\lambda}_1(t) &= \mathrm{E}[Z|T \geq t, x = 1]\, e^{\beta}\lambda_0(t).\end{aligned}$$

The marginal effect of x can be quantified by the ratio of these two marginal hazards. This results in

$$e^{\overline{\beta}(t)} = \frac{\overline{\lambda}_1(t)}{\overline{\lambda}_0(t)} = \frac{\mathrm{E}[Z|T \geq t, x = 1]}{\mathrm{E}[Z|T \geq t, x = 0]}\, e^{\beta}.$$

In general, $\overline{\beta}(t)$ is not constant in time. If Z is a gamma frailty with variance θ^{-1}, for example, this is

$$e^{\overline{\beta}(t)} = \frac{\theta + \Lambda_0(t)}{\theta + e^{\beta}\Lambda_0(t)}\, e^{\beta}.$$

If $\beta < 0$, $e^{\overline{\beta}(t)}$ is an increasing function with a minimum of e^{β} and asymptotic maximum of 1. Conversely, if $\beta > 0$, then $e^{\overline{\beta}(t)}$ is a decreasing function with a maximum of e^{β} and asymptotic minimum of 1. The conclusion is that, at the population level, the covariate effect appears to vanish over time. Therefore, the gamma frailty shows a similar behavior with the unobserved covariates scenario that was studied by simulation in Section 1.2.1.

Similar considerations apply for other frailty distributions. For example, for the inverse Gaussian distribution, the marginal hazard ratio is

$$e^{\overline{\beta}(t)} = \left(\frac{\theta + 2\Lambda(t)}{\theta + 2\Lambda_0(t)e^{\beta}}\right)^{1/2}.$$

A peculiar case is that of the positive stable distribution, for which

$$e^{\overline{\beta}(t)} = e^{\gamma\,\beta},$$

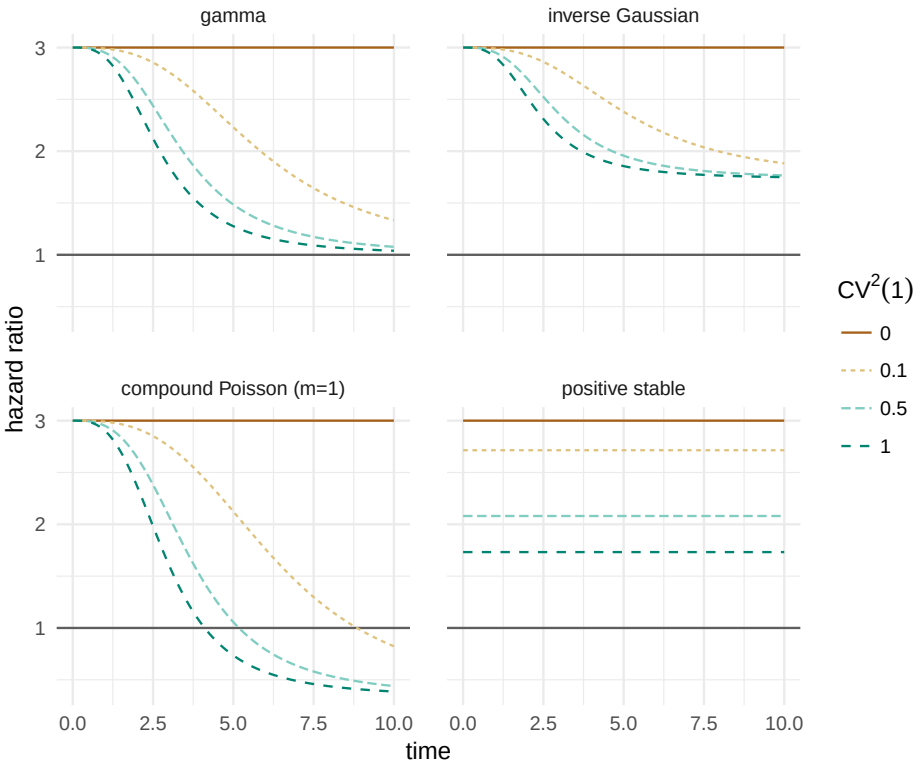

Figure 14: Marginal hazard ratio between two groups of individuals: a high risk one with $\lambda_1(t) = 3\lambda_0(t)$ and a low risk one with $\lambda_0(t) = t^2/20$. For comparability, the distribution are matched by the squared coefficient of variation of the distribution of survivors at time $t = 1$, with $CV^2(1) = \text{var}[Z|T \geq 1]/\text{E}[Z|T \geq 1]^2$.

which does not depend on time, so we have $\overline{\beta}(t) \equiv \overline{\beta} = \gamma\,\beta$. Since $0 < \gamma < 1$, $\overline{\beta}$ is an "attenuated" version of β.

The effect of different frailty distributions on the hazard ratio is illustrated in Figure 14. For the gamma and inverse Gaussian, the marginal hazard ratio approaches 1 with time. For the positive stable distribution, the attenuated marginal effect is observed. For the compound Poisson distribution, a "crossover" is present, where the marginal hazard ratio actually crosses 1. This is the effect of having non-susceptible individuals, represented by the mass at 0 of the distribution. This happens because, in the risk set at large time points, the proportion of non susceptible individuals is higher in the high risk group than in the low risk group.

Implications The shrinking of the hazard ratio in the presence of unobserved heterogeneity has important implications. One is that this might explain moderate familial risks found in clinical studies (Aalen, Valberg, et al., 2014). Moreover, it has an impact for the interpretation of estimated regression coefficients. In the context of a randomized clinical trial with two treatment arms, unobserved heterogeneity induces a loss of balance between the groups. While this may cause an effect as illustrated in Figure 14, it also implies that the estimated marginal hazard ratio does not have a causal interpretation anymore (Aalen, Cook, and Røysland, 2015).

Another phenomenon of interest is the "interruption of treatment", where x may change value at some point, describing the situation where individuals are moved from the treatment to the control group once the treatment does not appear to have any more effect (Aalen, 1994). If the treatment is beneficial, then individuals surviving in the control group will on average have a lower frailty than those in the treatment group. As an artifact, it might seem advantageous to remove individuals from the treatment group after some time, because the control group seems at a lower risk (comprising mostly low-frailty individuals).

1.2.5 Identifiability

In the frailty model, the marginal hazard equals $\overline{\lambda}(t) = \lambda(t)\mathrm{E}[Z \mid T > t]$. If there are no covariates, then only $\overline{\lambda}(t)$ is observed. Without strong parametric assumptions on $\lambda(t)$, is impossible to decide whether frailty is present or not. In other words, the frailty model is not identifiable in this case. The presence of covariates, together with the assumption of proportional hazards conditional on the frailty, make the frailty model identifiable, as long as the frailty distribution has finite expectation. This is due to the marginal non-proportional effect of the observed covariates, as exemplified in Figure 14. This identifiability result has been studied in Elbers and Ridder (1982)

Without further assumptions, observing a time dependent covariate effect of the type shown in Figure 14 is equally compatible with two explanations. One is that the proportional hazards assumption holds in the conditional model, and this effect appears distorted at the marginal level as a result of unobserved heterogeneity. The second is that there is no unobserved heterogeneity, and the observed covariate simply has a time dependent effect. In the first case, the frailty model is the natural choice. In the second case, the modeling strategy would rather include a stratified analysis or an extended Cox model with interactions of covariates with time (Therneau and Grambsch, 2000, ch. 6.5).

In this context, the result of Elbers and Ridder (1982), while theoretically interesting, is of little practical use. Only a firm - and probably naïve - belief in the conditional proportional hazards assumption can substantiate a claim towards the presence of frailty. In principle, this situation changes in the case of clustered survival data, because positive correlation between the event times is induced by the frailty. This is discussed in Section 1.3. The more information on the correlation structure, the easier it is to distinguish the frailty from non proportional hazards. However, when the cluster size is small, the identifiability result, identifying the appropriate model remains a difficult problem.

The positive stable distribution does not have finite expectation, and therefore it does not fall under the Elbers and Ridder (1982) result. As shown in Figure 14, it preserves the proportional hazards assumption at the marginal level. It is not identifiable with univariate survival data, even with covariates. In some sense, this may be seen as an advantage, since it illustrates that the identifiability of univariate frailty models is based on a strong assumption about the mechanism that generated the data. The positive stable distribution does prove useful in the context of clustered failures or recurrent events in Section 1.3.

1.3 Shared frailty models

1.3.1 Missing covariates in paired data

Consider the situation of paired life times, where covariate values are the same for individuals from the same pair. Assume that individuals from a given pair have the same distribution of the event time, denoted as T, with the hazard function $\lambda(t \mid x) = \lambda_0(t)\exp(\beta x)$. Further, assume that x is a realization of a random variable X with density $f_X(x)$. We denote $f(t \mid x)$ and $S(t \mid x)$ as the density and survival function of T, given $X = x$. The marginal survival function of T (where the covariate x is integrated over) is given by $\overline{S}(t) = \int S(t \mid x) f_X(x) dx$.

Consider one pair, with life times T_1 and T_2. The marginal survival function of either T_1 or T_2 is given by $\overline{S}$. However, if $T_1 = t_1$ is observed, the marginal survival function of T_2 will change. Heuristically, if a large life time t_1 is observed, then it is likely that the pair has a low hazard, which in turn makes it more likely that the value of x in that pair is low if $\beta > 0$ (or high if $\beta < 0$). Since x is shared by both individuals, a low hazard for T_1 means that the hazard for T_2 is also low, and that in turn makes it more likely that the corresponding life time t_2 is large as well.

All this leads to positive marginal correlation of the two life times. More specifically, it is straightforward to show that the marginal survival function of T_2, given $T_1 = t_1$, is given by

$$S(t_2 \mid T_1 = t_1) = \int \overline{f}(t_1 \mid x) S(t_2 \mid x) dx,$$

with $\overline{f}(t_1 \mid x) = f(t_1 \mid x) f_X(x) / (\int f(t_1 \mid x) f_X(x) dx)$. Figure 15 shows $S(t_2 \mid T_1 = t_1)$ for $t_1 = 0.1$ and $t_1 = 2$, for the case where the conditional distribution of T_1 and T_2, given $x = 0$, is exponential with mean 1, and $\beta = 1$, and X has a normal distribution with mean 0 and standard deviation σ.

It can be seen that for $t_1 = 2$, the conditional survival curves are higher than the marginal survival curve, while for $t_1 = 0.1$ this is the other way around. For higher standard deviation of the distribution of X, the conditional survival curves are more distinct from the marginal survival function. That means that for higher standard deviation of X, the influence of knowing the value of T_1 is higher, and the correlation between T_1 and T_2 is higher. In fact, one can derive an explicit expression of the correlation between

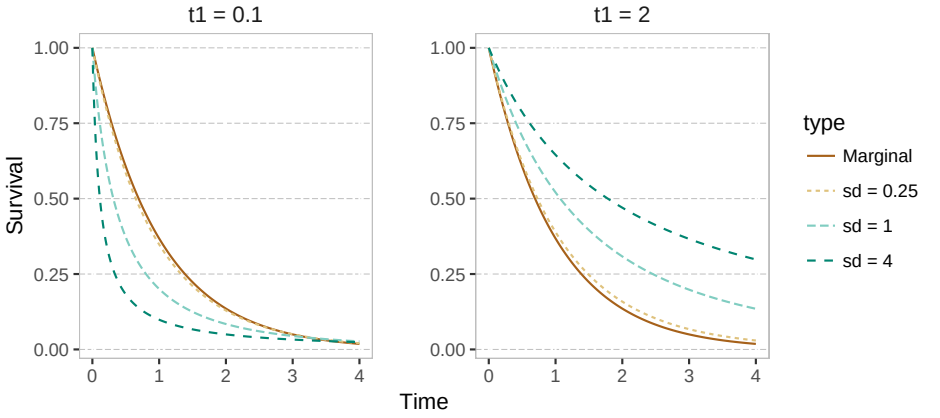

Figure 15: Conditional survival function of T_2, given $t_1 = 0.1$ and given $t_1 = 2$; the conditional distribution of T_1 and T_2 given $X = x$ is exponential with rate $\lambda e^{\beta x}$ and $\beta = 1$, and X has a normal distribution with mean 0 and standard deviation σ^2, with different values of σ.

T_1 and T_2, when the baseline distribution of T_1 is exponential with rate h. It is given by

$$\operatorname{cor}(T_1, T_2) = \frac{e^{2\beta^2\sigma^2} - e^{\beta^2\sigma^2}}{2e^{2\beta^2\sigma^2} - e^{\beta^2\sigma^2}}.$$

A plot of the correlation as a function of σ^2, for $\beta = 1$, is shown in Figure 16.

If the correlation of life times cannot be explained by observed covariates (for example, because x is omitted), then there are two practical approaches. One is marginal modeling, which is in the spirit of general estimating equation (GEE) models. For the Cox model, this involves adjusting the standard errors of the observed covariates (Lin and Wei, 1989). The second is to model the conditional hazard by introducing a "shared" frailty Z, that would take the place of $\exp(\beta x)$ in the previous example. The resulting "shared" frailty model is detailed in Section 1.3.2. The advantage of this approach is that differences between clusters can be quantified, and that the covariate effects have an individual interpretation, as in the case of univariate frailty models.

1.3.2 Clustered failures

The shared frailty model

Assume that there are N clusters and n_i individuals are part of cluster i. The hazard of the jth individual from cluster i is specified as

$$\lambda_{ij}(t|Z_i) = Z_i \exp(\beta^\top \mathbf{x}_{ij})\lambda_0(t). \tag{1.7}$$

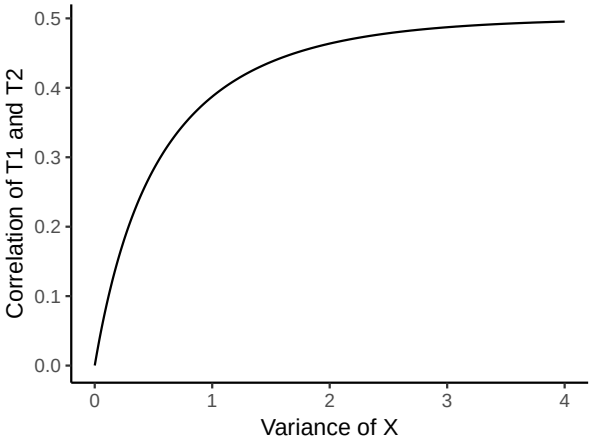

Figure 16: Correlation between T_1 and T_2 as a function of σ^2; the conditional distribution of T_1 and T_2 given $X = x$ is exponential with rate $\lambda e^{\beta x}$ and $\beta = 1$, and X has a normal distribution with mean 0 and variance σ^2.

The individuals in cluster i share the frailty Z_i, and conditional on Z_i their lifetimes are assumed to be independent. While in the univariate case individuals are thought to be a random sample from a larger population of individuals, in the clustered failures case the clusters are thought to be a random sample from a population of clusters.

In the univariate case, the marginal hazard was derived given the individual survival until time t. In the clustered failure case, the marginal hazard is derived given all information about the cluster until time t, including observed events and censorings. This is studied in the following section.

Frailty distributions and clustered failures

Suppose that there are two individuals in a cluster. The conditional cumulative hazard for individuals $j = 1, 2$ is given by

$$\Lambda_j(t) = \int_0^t Y_j(s) \exp(\beta^\top \mathbf{x}_j) \lambda_0(s) ds,$$

where $Y_j(s) = 1$ when individual j is at risk at time s and 0 otherwise. Conditional on Z, the conditional joint survival function of T_1, T_2 is defined as

$$\begin{aligned} S(t_1, t_2|Z) &= P(T_1 > t_1, T_2 > t_2|Z) \\ &= \exp(-Z(\Lambda_1(t_1) + \Lambda_2(t_2))). \end{aligned}$$

The marginal joint survival is obtained by taking the expectation with respect to Z, which results in

$$S(t_1, t_2) = \mathcal{L}(\Lambda_1(t_1) + \Lambda_2(t_2)). \tag{1.8}$$

The Laplace transform of Z, given that individual 1 and 2 are alive at t_1 and t_2, is obtained, with the same arguments as in (1.5), as

$$\mathcal{L}_{Z|T_1>t_1,T_2>t_2}(c) = \frac{\mathcal{L}(c + \Lambda_1(t_1) + \Lambda_2(t_2))}{\mathcal{L}(\Lambda_1(t_1) + \Lambda_2(t_2))}.$$

The only difference from the univariate case is that $\Lambda(t)$ is now replaced by $\Lambda_1(t_1)+\Lambda_2(t_2)$.

Assume now that the event time of the first individual T_1 is observed at t_1. Recall that the density of T is given by $f(t) = \lambda(t)S(t)$. It is then obtained that

$$\begin{aligned}\lim_{dt\downarrow 0} \frac{\mathsf{P}(t_1 \le T_1 < t_1 + dt, T_2 > t_2|Z)}{dt} &= Z\lambda_1(t_1)\exp(-Z(\Lambda_1(t_1) + \Lambda_2(t_2))) \\ &= \frac{\partial}{\partial t_1} S(t_1, t_2|Z).\end{aligned}$$

The Laplace transform of $Z|T_1 = t_1, T_2 > t_2$, defined as

$$\mathcal{L}_{Z\,|\,T_1=t_1,T_2>t_2}(c) = \mathrm{E}[\exp(-cZ)\,|\,T_1 = t_1, T_2 > t_2]$$

can be calculated from Bayes' theorem:

$$\begin{aligned}\mathcal{L}_{Z\,|\,T_1=t_1,T_2>t_2}(c) &= \frac{\mathrm{E}[Z\lambda_1(t_1)\exp(-Z(c + \Lambda_1(t_1) + \Lambda_2(t_2)))]}{\mathrm{E}[Z\lambda_1(t_1)\exp(-Z(\Lambda_1(t_1) + \Lambda_2(t_2)))]} \\ &= \frac{\mathcal{L}'(c + \Lambda_1(t_1) + \Lambda_2(t_2))}{\mathcal{L}'(\Lambda_1(t_1) + \Lambda_2(t_2))}.\end{aligned}$$

More generally, for a cluster of arbitrary size, denote the event history of all its individuals up to some horizon τ as $\mathcal{F}_\tau$. Assume that this comprises $N(\tau)$ observed events, and let

$$\Lambda_\bullet(\tau) = \sum_j \Lambda_j(\tau). \tag{1.9}$$

Denote $\mathcal{L}^{(k)}$ as the k-th derivative of the Laplace transform. The Laplace transform of $Z|\mathcal{F}_\tau$ is given by

$$\mathcal{L}_{Z|\mathcal{F}_\tau}(c) = \frac{\mathcal{L}^{(N(\tau))}(c + \Lambda_\bullet(\tau))}{\mathcal{L}^{(N(\tau))}(\Lambda_\bullet(\tau))}. \tag{1.10}$$

The expectation of this distribution follows as

$$\mathrm{E}[Z\,|\,\mathcal{F}_\tau] = -\frac{\mathcal{L}^{(N(\tau)+1)}(\Lambda_\bullet(\tau))}{\mathcal{L}^{(N(\tau))}(\Lambda_\bullet(\tau))}. \tag{1.11}$$

Therefore, for an individual with covariate vector $\mathbf{x}$ from a cluster where the event history of the cluster is given by $\mathcal{F}_t$, the marginal hazard is

$$\overline{\lambda}(t) = \mathrm{E}[Z|\mathcal{F}_{t-}] \exp(\beta^\top \mathbf{x})\lambda_0(t). \tag{1.12}$$

For the gamma frailty, it is obtained that

$$\mathrm{E}[Z|\mathcal{F}_{t-}] = \frac{\theta + N(t-)}{\theta + \Lambda_\bullet(t-)},$$
$$\mathrm{var}[Z|\mathcal{F}_{t-}] = \frac{\theta + N(t-)}{(\theta + \Lambda_\bullet(t-))^2}.$$

Similar expressions may be derived in a similar fashion for other PVF frailty distributions.

Dependence and the cross-ratio

The estimated frailty variance offers an indication of unobserved heterogeneity between clusters, but it offers little information on the resulting marginal correlation of the event times. Even for paired data, the formulas for the bivariate survival function in (1.8) are difficult to interpret.

One measure of bivariate dependence is Kendall's coefficient of concordance (Kendall's tau). Denote two pairs of individuals as $\{(11), (12)\}$ and $\{(21), (22)\}$, where (ij) refers to individual j of cluster (pair) i. Kendall's tau is defined as

$$\tau_K = \mathrm{E}[\mathrm{sign}(T_{11} - T_{21})(T_{12} - T_{22})],$$

where sign is the sign function. This is proportional to the probability that the order of events are concordant between the two clusters clusters. The median concordance is a similar measure that only involves one pair,

$$\kappa = \mathrm{E}[\mathrm{sign}(T_1 - \mathrm{median}(T_1))(T_2 - \mathrm{median}(T_2))].$$

This is proportional to the probability that the events within the same cluster are concordant, i.e. they occur both before the median survival time or after. In frailty models, both τ_K and κ are positive quantities, since the specification (1.12) only allows for positive dependence. Under independence, both measures would be 0. However, the reverse statement is not usually true.

A more natural way of exploring the within-cluster dependence structure is via the cross-ratio (Oakes, 1989), which compares how the marginal hazard would behave if an event would happen as opposed to an event not happening. Unlike τ_K and κ, this is a local measure of dependence. To illustrate this, we consider one cluster with individuals 1 and 2. Conditional on the frailty, their event times T_1 and T_2 are independent. Denote the marginal hazard of individual 2 if individual 1 is alive at t_1 as

$$\lambda_2(t|T_1 > t_1) = \frac{\mathcal{L}'(\Lambda_1(t_1) + \Lambda_2(t))}{\mathcal{L}(\Lambda_1(t_1) + \Lambda_2(t))}\lambda_2(t),$$

and the marginal hazard of individual 2 if individual 1 had an event at time t_1 as

$$\lambda_2(t|T_1 = t_1) = \frac{\mathcal{L}''(\Lambda_1(t_1) + \Lambda_2(t))}{\mathcal{L}'(\Lambda_1(t_1) + \Lambda_2(t))}\lambda_2(t).$$

These two are marginal hazards under different hypothetical event histories of the other individual in the cluster. They are equal only if there is no dependence between the two individuals. The cross-ratio can be expressed as

$$\begin{aligned}\mathrm{CR}(t) &= \frac{\lambda_2(t|T_1 = t_1)}{\lambda_2(t|T_1 > t_1)} \\ &= \frac{\mathcal{L}''(\Lambda_1(t_1) + \Lambda_2(t))}{\mathcal{L}'(\Lambda_1(t_1) + \Lambda_2(t))}\left(\frac{\mathcal{L}'(\Lambda_1(t_1) + \Lambda_2(t))}{\mathcal{L}(\Lambda_1(t_1) + \Lambda_2(t))}\right)^{-1}.\end{aligned}$$

Intuitively, if there is positive dependence between the two event times, $\mathrm{CR}(t) \geq 1$. Hougaard (2000) suggested that a more interpretable comparison would be to replace the denominator by $\lambda_2(t|T_1 > t)$, to compare the hazard given that "individual 1 died at time t_1" with "individual 1 is alive now". For $t > t_1$, the adjusted cross ratio is defined as

$$\begin{aligned}\mathrm{CR}_a(t) &= \frac{\lambda_2(t|T_1 = t_1)}{\lambda_2(t|T_1 > t)} \\ &= \frac{\mathcal{L}''(\Lambda_1(t_1) + \Lambda_2(t))}{\mathcal{L}'(\Lambda_1(t_1) + \Lambda_2(t))}\left(\frac{\mathcal{L}'(\Lambda_1(t) + \Lambda_2(t))}{\mathcal{L}(\Lambda_1(t) + \Lambda_2(t))}\right)^{-1}.\end{aligned}$$

For $t \leq t_1$, this quantity does not have a direct interpretation.

In Figure 17, we illustrate the unadjusted and adjusted cross-ratio functions for the gamma, inverse Gaussian and positive stable distributions. For comparison purposes, the distributions are matched by Kendall's tau rather than variance. Both unadjusted and adjusted cross-ratio functions show that the marginal hazard of individual 2 is larger if individual 1 has an event. For the unadjusted, the cross-ratio for the gamma frailty is constant, showing that the event of individual 1 affects the hazard in a "proportional" manner. The shape of $\mathrm{CR}(t)$ for the inverse Gaussian and positive stable frailties show that there is a strong immediate dependence that vanished in time.

The adjusted cross-ratio paints a slightly different picture. For the gamma, it implies that, if the partner dies, the hazard for the survivor would appear increasingly larger as compared to the scenario where the partner would still be alive. For the positive stable distribution, the surviving individual is at a perceived high risk right after the partner died, but the differences quickly decreases. This can be interpreted as a large correlation between the life times on the short term. As before, the inverse Gaussian is somewhere in the middle.

$\mathrm{CR}(t)$ may be interpreted as an "instantaneous odds ratio" (Anderson et al., 1992), and for bivariate survival data it may be used for selecting the frailty distribution (Duchateau

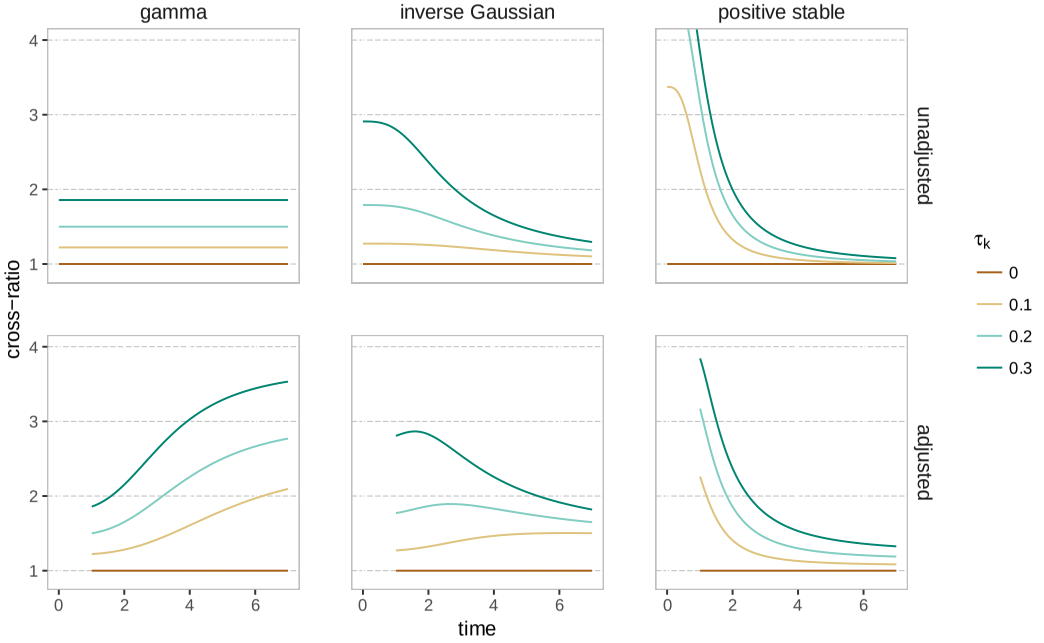

Figure 17: Cross-ratio and adjusted cross-ratio ($t_1 = 1$) for the gamma, inverse Gaussian and positive stable distributions, for different values of Kendall's tau. The individual hazard is taken as $\lambda(t) = t^2/20$.

and Janssen, 2007, ch. 4). One disadvantage is that CR depends on the conditional cumulative hazard; a scaled cross-ratio that overcomes this has been proposed by Paddy Farrington, Unkel, and Anaya-Izquierdo (2012).

The gamma frailty is said to induce "late dependence" (a high probability of events occurring close by at later time points), the positive stable frailty induces "early dependence" (a high probability of event occurring close by early in the follow-up) and the inverse Gaussian is somewhere in the middle. The timing of the dependence can be studied by analyzing the marginal joint distribution of T_1 and T_2 (Hougaard, 2000). A disadvantage of this approach is that the marginal distributions of T_1 and T_2 must be known separately, which is usually not possible.

1.3.3 Frailty model for recurrent events

Recurrent events are most commonly defined in the framework of counting processes. Each individual is described by a process $N(t)$ with the property that $N(t)$ "counts" the number of events experienced by the individual until time t.

The two common frameworks for modeling N are the Poisson process and the renewal process (Cook and Lawless, 2007). If unobserved individual heterogeneity is present, then there are two approaches that may be used in practice. One is the marginal approach, where the unobserved heterogeneity is treated as a nuisance (Cook and Lawless, 1997). In that case, the focus of analysis is the *rate* of N, which is defined as the probability of observing an increase in N in the small interval $(t, t + dt)$.

The second approach is to model the intensity of N. While the hazard is defined as the instantaneous probability of an event given that the individual is alive, the intensity is defined as the instantaneous probability of an event given *the whole previous event history* of the individual. Let Z be the frailty of the individual. The intensity of N can be specified as

$$\lambda(t|Z) = Y(t)Z \exp(\beta^\top \mathbf{x})\lambda_0(t), \tag{1.13}$$

where $Y(t)$ is an indicator function that is 1 if the individual is under observation at time t and 0 otherwise. Similarly to the univariate frailty, the variance of Z describes between-individual unobserved heterogeneity.

The marginal intensity is obtained by replacing Z by $\mathrm{E}[Z|\mathcal{F}_{t-}]$, with $\mathcal{F}_{t-}$ now representing the event history of the individual until just before time t. As in the case of univariate frailty in Section 1.2.4, the effect of the covariates in the marginal intensity is usually time dependent. Similar to the clustered failures scenario, $\mathrm{E}[Z|\mathcal{F}_{t-}]$ includes information on previous event times.

The intensity in (1.13), with t referring to "time since origin of the recurrent event process", is referred to as the *calendar time* or *Andersen-Gill* formulation. Conditional on Z, N is assumed to be a Poisson process, meaning that its intensity conditional on Z does not depend on the history $\mathcal{F}_{t-}$. Alternatively, in the gap-time scale, t refers to "time since the previous event". The intensity may then be expressed as $\lambda(t|Z) = \lambda(t - B(t)|Z)$, where $B(t)$ is the time of the event before time t. From a practical point of view, the gap

time scale has a very similar representation to (1.7), where $\lambda_{ij}(t|Z)$ interpreted as the hazard of the j-th event. Conditional on Z, N is then a renewal process. The resulting marginal intensities are said to define a *modulated* Poisson or renewal processes.

In the case of recurrent events, the frailty mainly expresses that, if two individuals with identical covariates were observed over the same period of time, the expected number of events is larger for the one with the higher frailty. The number of events carries the most information on the frailty (Hougaard, 2000, ch. 9). Therefore, the measures of dependence discussed in Section 1.3.2 are of little interest in this context.

Modeling recurrent events is a complex task and several types of models may be accommodated with counting processes (Therneau and Grambsch, 2000, ch. 8.5). Furthermore, time dependent covariates representing, for example, the number of previous events, may also be added in the model (Aalen, Borgan, and Gjessing, 2008, ch. 8), thus severely complicating the time dependence mechanisms. A comprehensive reference on recurrent event modeling may be found in Cook and Lawless (2007).

1.4 Frailty models in practice

1.4.1 Estimation and inference

Depending on the nature of λ_0, the models may be classified as semiparametric or parametric. In semiparametric models, no assumptions are made on the baseline hazard λ_0 and the maximum likelihood estimate of λ_0 has mass only at the event times, as is the case for the Breslow estimator (Breslow, 1972). In parametric models, λ_0 is determined by a small number of parameters, such as the exponential, Weibull or Gompertz models, or flexible parametric approaches employing spline-based estimators.

Likelihood and EM-based approaches

The likelihood construction for counting processes is detailed in most survival analysis textbooks (Aalen, Borgan, and Gjessing, 2008; Kalbfleisch and Prentice, 2002). To cover all the scenarios described in this chapter, assume that i denotes the cluster, (i, j) the j-th individual within the cluster i and t_{ijk} denotes the k-th event or censoring observed on individual (i, j). We define the event indicator δ_{ijk} as 1 if t_{ijk} is an event time and 0 otherwise. Suppose that the conditional hazard of subject (i, j), conditional on the frailty Z_i is given by $\lambda_{ij}(t|Z_i) = Z_i\lambda_{ij}(t)$ with $\lambda_{ij}(t) = \lambda_0(t)\exp(\beta^\top \mathbf{x}_{ij})$. Denote the at risk indicator of subject j in cluster i by $Y_{ij}(t)$ and let $\Lambda_{i\bullet} = \sum_j \int_0^\infty Y_{ij}(t)\lambda_{ij}(t)dt$ be the sum of conditional cumulative hazards of cluster i, as defined in equation (1.9).

Assuming that the frailties Z_i are observed, the *conditional likelihood* contribution of cluster i is given by

$$L_i(\beta, \lambda_0|Z_i) = \prod_{jk} \left(\lambda_{ij}(t_{ijk}|Z_i)\right)^{\delta_{ijk}} \times \exp\left(-Z_i\Lambda_{i\bullet}\right).$$

and the likelihood for all the individuals is a product of all L_is. The clustered failure data is represented by having only one time point per individual ($k \equiv 1$), while the recurrent events case is represented by having only one individual per cluster ($j \equiv 1$). An implicit assumption here is that censoring is independent. In terms of counting processes, the at-risk process $Y(t)$ is assumed to be independent of $N(t)$, given the covariates and event history up to time t.

In the first part of this expression, Z_i appears to the power N_i, the total number of events from the cluster i. The *marginal likelihood* contribution of cluster i is obtained as taking the expectation over Z_i:

$$L_i(\beta, h_0, \theta) = \prod_{j,k} \lambda_{ij}(t_{ijk})^{\delta_{ijk}} \, \mathrm{E}\left[Z_i^{N_{i\bullet}} \exp\left(-Z_i \Lambda_{i\bullet}\right) \right]. \tag{1.14}$$

For valid inference based on $L(\beta, h_0, \theta)$, the censoring or at-risk process must also not involve the frailty, for reasons outlined in Nielsen et al. (1992). This assumption is similar to that of regular Cox models, where dependent censoring arises, for example, if the censoring process depends on unobserved covariates.

The "posterior" distribution of Z_i, $Z_i|\mathcal{F}_\tau$, has the density kernel

$$f_{Z_i}(z|\mathcal{F}_\tau) \propto z^{N_{i\bullet}} \exp(-z\Lambda_{i\bullet}) f_{Z_i}(z).$$

This is available in closed form only for the gamma frailty. A consequence of this is that the expectation in (1.14) is typically difficult to calculate for other frailty distributions. The expectation of this distribution is also known as the *empirical Bayes* frailty estimate. It can be calculated via the Laplace transform, as discussed in Section 1.3.2. This may involve having to take many derivatives of the Laplace transform, if $N_{i\bullet}$ is large. Another difficulty arises in semiparametric models, where the dimension of λ_0 is usually equal to the total distinct event time points from the data. This prevents a direct maximization of the likelihood.

The Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977) has been proposed for semiparametric gamma frailty models (Nielsen et al., 1992; Klein, 1992), and can be easily extended to the PVF family of distributions (Hougaard, 2000, ch. 8). This involves iterating between two steps:

1. The "E" step, which involves calculating the expected log-likelihood,

$$\mathrm{E}\ell(\beta, \lambda_0, \theta) = \sum_i \mathrm{E}\left[\log L_i(\beta, \lambda_0|Z)\right].$$

   In practice, this involves calculating $\mathrm{E}[Z_i|\mathcal{F}_\tau]$ and $\mathrm{E}[\log Z_i|\mathcal{F}_\tau]$.

2. The "M" step, where β, h_0 and θ are updated, by maximizing $\mathrm{E}\ell(\beta, \lambda_0, \theta)$.

The advantage of this approach is that the M step may be calculated via Cox's partial likelihood (Cox, 1975), effectively eliminating the problem of the high-dimensional

λ_0. However, the E step still requires numerical integration for distributions except the gamma.

Of the two posterior expectations that are calculated in the E step, $\mathrm{E}[Z_i|\mathcal{F}_\tau]$ may be expressed as a ratio of derivatives of the Laplace transform. The calculation of $\mathrm{E}[\log Z_i|\mathcal{F}_\tau]$ can be avoided via a "profile EM" algorithm, which involves performing the EM algorithm described here for fixed values of θ, resulting in a two-stage maximization. Alternatively, the Monte Carlo EM algorithm may be employed, which involves a stochastic approximation of the E step (Vaida and Xu, 2000).

Alternative approaches

The penalized likelihood method (Ripatti and Palmgren, 2000; Therneau, Grambsch, and Pankratz, 2003) is a very popular way of estimating gamma and log-normal semiparametric frailty models. The basic idea behind it is that, for fixed θ, the $\log Z_i$'s may be treated as regular parameters (on the same scale with the regression coefficients β). Afterwards, a penalization of a specific form is imposed upon them. Depending on the penalization, the results are equivalent to those of a gamma or a log-normal distributional assumption. This approach is typically the fastest for semiparametric models. A downside is that it is not immediately possible to extend the estimation to other frailty distributions.

Other approaches include a pseudo-likelihood method (Gorfine, Zucker, and Hsu, 2006), which leads to consistent estimators and may be employed for a larger number of frailty distributions, and the h-likelihood method (Ha, Lee, and Song, 2001; Ha, Jeong, and Lee, 2017). This approach relies on maximizing the joint likelihood of the observed *and* unobserved data. It has been developed for the gamma and log-normal distributions.

Inference

For parametric models, the variance-covariance matrix is typically obtained directly, as the inverse of the numeric Hessian matrix. This is usually provided by directly by an optimization software.

For models estimated with the EM algorithm, Louis' formula may be used (Louis, 1982) to obtain standard errors of the estimates. It has been shown that the h_0 may be regarded, for practical purposes, as an ordinary finite dimensional parameter and the information matrix may be constructed from the matrix of second derivatives (Andersen, Klein, et al., 1997).

For the profile EM algorithm, the variance covariance matrix for (β, h_0) is obtained under the assumption of fixed θ. Similarly to the penalized likelihood methods, the variance covariance matrix for β, based on the partial likelihood, is also obtained under fixed θ. The complete variance-covariance matrix for β, h_0 or for β should then be adjusted for the variability of θ (Hougaard (2000), ch B3; Putter and Van Houwelingen (2015)), although this is often ignored in practice.

Inference regarding the frailty variance is more challenging. The limiting case, when the variance is 0, is a proportional hazards model without frailty. A likelihood ratio test based on a mixture of χ^2 distributions can be employed to test the difference between these two models (Self and Liang, 1987; Claeskens, Nguti, and Janssen, 2008). Another issue is that, since the variance must be positive, symmetric confidence intervals are not very meaningful. An alternative is to calculate likelihood based confidence intervals, as is illustrated in Therneau and Grambsch (2000, ch. 9).

1.4.2 Software

Support for frailty models exists in major statistical packages such as R (R Core Team, 2017), SAS (Inc., 2003) and Stata (StataCorp, 2017). The `PHREG` command in SAS implements the penalized likelihood method for the gamma and log-normal frailty models. The `streg` procedure in Stata implements parametric gamma and inverse Gaussian frailty models. In what follows we will focus on packages for R.

Semiparametric gamma and log-normal frailty models may be estimated via the penalized likelihood method in the **survival** package (Therneau and Grambsch, 2000; Therneau, 2015a). Semiparametric frailty models with the infinitely divisible class of frailty distributions discussed in Section 1.2.3 may be estimated via the profile EM algorithm with the **frailtyEM** package. Log-normal frailty models (including correlated frailties, discussed in Section 1.5) may be estimated with the **coxme** package (Therneau, 2015b). Similar models may be fitted with the Monte Carlo EM algorithm with the **phmm** R package (Donohue and Xu, 2013). Log-normal and gamma frailty models can also be estimated via h–likelihood with the **frailtyHL** package (Do Ha, Noh, and Lee, 2012). The pseudo-likelihood approach is implemented in the **frailtySurv** package (Monaco, Gorfine, and Hsu, 2017), supporting some of the infinitely divisible distributions from the PVF family.

Parametric and flexible parametric frailty models for the gamma and log-normal distributions are supported by the **frailtypack** package (Rondeau and Gonzalez, 2005; Rondeau, Mazroui, and Gonzalez, 2012) (including correlated random effects, nested random effects and numerous other scenarios). Parametric frailty models with support for some of the PVF family distributions are implemented in the **parfm** package (Munda, Rotolo, and Legrand, 2012).

1.4.3 Data representation

In R (R Core Team, 2017), the canonical resources for survival analysis are found in the **survival** package (Therneau, 2015a). Event histories corresponding to survival times or to recurrent events have a very similar representation, as is described in detail in Therneau and Grambsch (2000).

An event history is represented by a collection of *observations*, which are vectors $(t_L, t_R, \delta, \mathbf{x})$ where (t_L, t_R) are two time points that define an "at-risk" interval, δ is equal

to 1 if the interval ended with an event and 0 otherwise, and $\mathbf{x}$ is a vector of covariate values that are constant on this interval. In R, the tuple (t_L, t_R, δ) is referred to as (`tstart`, `tstop`, `status`). Univariate survival times and clustered failures are usually represented by having $t_L = 0$ and a simplified `(tstop, status)` notation. Furthermore, this notation may also be used to express:

- Recurrent events in calendar time (or "Andersen-Gill" representation). In this case, for an individual, t_R are event times and t_L is usually 0 or the time of the previous event. Usually, the last observation is censored with the last t_R being the end of follow-up.
- Recurrent events in gap time. In this case, $t_L = 0$ and t_R are observed gap times. The last observation may be censored, indicating an incomplete gap time at the end of follow-up.
- Left truncated survival times, where t_L is the time point after which the individual enters the study.
- Time dependent covariates. In this case, if the value $\mathbf{x}$ changes at time $\tilde{t} \in (t_L, t_R)$, this results in two observations corresponding to time intervals $(t_L, \tilde{t})$ and $(\tilde{t}, t_R)$, with the first one being artificially censored.

In the presence of frailty, an observation is interpreted as a contribution to the conditional likelihood of the form

$$L(\beta, \lambda_0 \mid Z; t_L, t_R, \delta, \mathbf{x}) = \left\{ Z\lambda_0(t_R)e^{\beta^\top \mathbf{x}} \right\}^{\delta} \cdot \exp\left(-Z(\Lambda_0(t_R) - \Lambda_0(t_L))e^{\beta^\top \mathbf{x}} \right).$$

For a collection of observations sharing the same frailty Z, the software maximizes

$$\mathrm{E}_Z \left[\prod_{\text{intervals}} L(\beta, \lambda_0 \mid Z; t_L, t_R, \delta, \mathbf{x}) \right],$$

which is the contribution of one cluster to the marginal likelihood (1.14). This is appropriate in the case of recurrent events and time dependent covariates, or for clustered survival times without left truncation.

For left truncated survival times however, this is generally incorrect. In the univariate case, the frailty distribution of a left truncated individual is $Z|T \geq t_L$, referred to as the distribution of survivors in Section 1.2.4.

In the case of clustered survival times, the event of observing the whole cluster must be taken into account (Erikson, Martinussen, and Scheike, 2015; Van den Berg and Drepper, 2011; Jensen et al., 2004). If the individuals from the same cluster have truncation times $t_{L,1}, t_{L,2}, \ldots t_{L,J}$ that are independent given Z, then the frailty distribution of the cluster is $Z|T_1 > t_{L,1}, \ldots T_J > t_{L,J}$.

More complicated selection schemes arise when the left truncation times are not independent, even conditional on the frailty (Rodríguez-Girondo et al., 2018). In the

case of recurrent events, such selection schemes may arise when individuals are included into the study only if they experience a certain number of events (Balan, Jonker, et al., 2016). Such scenarios usually require *ad-hoc* estimation procedures and are not generally supported by the main software packages.

In R, one of the reasons why the same notation is used to denote both recurrent events and left truncation is because they lead to the same likelihood in frailty-less models. In the case of frailty models, the treatment depends on the package used. For example, the **survival** package calculates the correct likelihood for the recurrent events case, **parfm** calculates the correct likelihood for the left truncation case. In **frailtypack** and **frailtyEM**, both scenarios are supported.

1.5 Extensions

In the models discussed in Sections 1.2 and 1.3, the frailty plays the role of a random intercept. In certain scenarios, particularly in studies on bivariate outcomes, correlated random effects have been proposed (Yashin, Vaupel, and Iachine, 1995; Yashin, Iachine, et al., 2001; Wienke, 2010). These address the limitation that shared frailty models may only be employed for positively correlated event times.

Furthermore, the random effect Z has been so far assumed to be time constant. This is consistent with the interpretation that Z accounts for individual specific or cluster specific characteristics that are fixed from the time origin, and have an effect that is constant in time. However, the unobserved heterogeneity might be time dependent, thus better explained by an unobserved random processes that unfolds in time. Several approaches based on this idea have been proposed. The frailty may be modeled with diffusion processes (Yashin and Manton, 1997; Aalen and Gjessing, 2004) or Levy processes (Gjessing, Aalen, and Hjort, 2003). More recently, an approach on birth-death Poisson processes has been proposed (Putter and Van Houwelingen, 2015). Simpler, piecewise constant, frailty models have also been considered (Paik, Tsai, and Ottman, 1994; Wintrebert et al., 2004). A limited implementation combining the birth-death processes and the piecewise constant frailty is implemented in the R package **dynfrail** (Balan, 2017). Related approaches include the constructions of auto-regressive frailty processes based on log-normal frailties (Yau and McGilchrist, 1998; Munda, Legrand, et al., 2016) or gamma frailties (Fiocco, Putter, and Van Houwelingen, 2008).

For the models presented in Sections 1.2 and 1.3 are intended for the analysis of one stochastic event process, it has been assumed that the censoring does not depend on the frailty. This assumption may be tested (Balan, Boonk, et al., 2016), or the event and censoring processes can be jointly modeled. An example is when the observation recurrent event process may be stopped by death (Liu, Wolfe, and Huang, 2004) or when the frailty is also associated with the censoring (Huang and Wolfe, 2002).

Moreover, we assumed that the time dependent covariate vector $\mathbf{x}$ is somewhat "external" to the event process, in the sense of (Kalbfleisch and Prentice, 2002). If $\mathbf{x}$ contains internal time dependent covariates, such as repeated individual measurements, the pro-

cesses should be jointly analyzed (Rizopoulos, 2012, ch. 2). In this case, the frailty is shared by the model for the time dependent covariate (or biomarker) and the model for the event process. Software for estimating joint models is also available in R (Rizopoulos, 2016).

1.6 Outline of the thesis

Frailty models account for unobserved individual or cluster characteristics. In the case of gap-time recurrent events and clustered failure times, they relax the usual independence of event times assumption to a conditional independence assumption. In the case of recurrent events in calendar time, the assumption of a Poisson process is relaxed.

In Chapter 2, the topic of identifiability of shared frailty models is analyzed by means of a simulation study. It has been shown that the univariate frailty model is identifiable, as long as the frailty has finite expectation and covariates are present (Elbers and Ridder, 1982). This result implies that, for univariate survival data, it is very difficult to distinguish between the effect of unobserved heterogeneity can and a possible time dependent effect of the covariates. We analyze how this problem extends to shared frailty models.

In Chapter 3, we study the situation where a recurrent event process may be associated with a terminal event, such as death, due to unobserved factors. Because the recurrent event cannot be observed any more after death, this is an example where the observation of the process is not independent of the process itself. We propose a score test for association between the recurrent event process and the terminal event. This test provides evidence against the usual assumption of independent observation.

In Chapter 4, we analyze the phenomenon of ascertainment of patients in observational studies on recurrent events data. More specifically, we study the case when individuals are included in the study only if at least one event is observed in a specific ascertainment time frame. We propose a solution for accounting for this selection scheme.

In Chapter 5, we discuss maximum likelihood estimation for frailty models and present the implementation from the **frailtyEM** package (Balan and Putter, 2017) in R (R Core Team, 2017). The package, which supports semiparametric estimation of frailty models with distributions from the PVF family, employs a profile expectation-maximization algorithm. Advantages and disadvantages of such approach are discussed, together with a practical demonstration of the software.