Shared Frailty Models

Theodor Balan & Hein Putter

Department of Medical Statistics and Bioinformatics
Leiden University Medical Center

Frailty Models: Theory & Practice
November 16, 2017, Prague

Shared frailty models

Recurrent events

Estimation

Software overview

Shared frailty models

Dependence structures

In time-to-event data, dependence between observations may arise from:

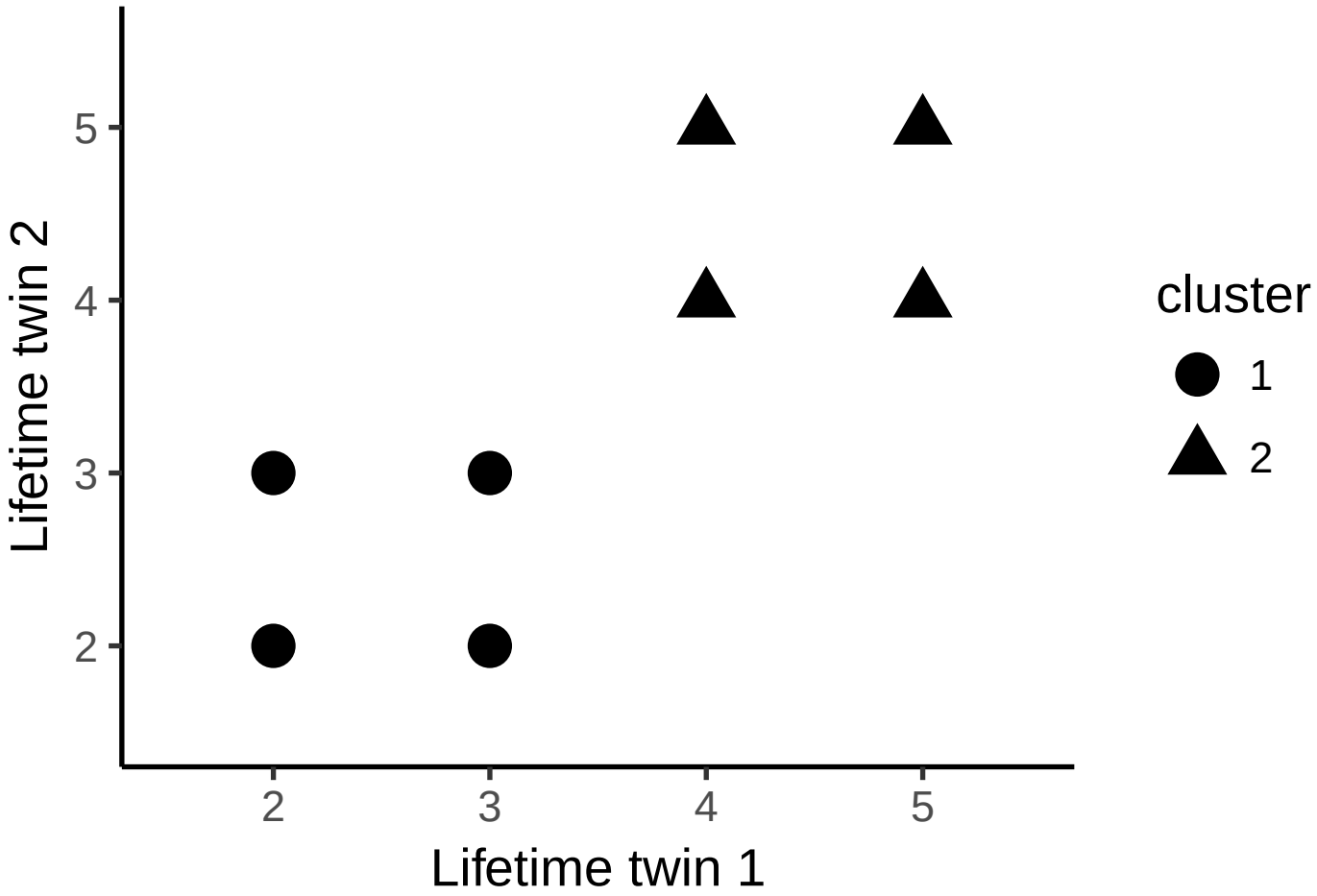
- Common events (several events happening at the same time)
  - competing risks, multi-state models
- Common (unobserved) risk (e.g. hidden covariates, genetics)
  - (shared) frailty models
- Event-related dependence (e.g. one infection may lead to more infections)
  - time-dependent covariates

The goal of a model is to be *biologically plausible*.

Common risk

- Unobserved latent factors that are constant in time may affect lifetimes
- Conditional on the common risk factors, we may assume that observations are independent
- Since they are unobserved, they are treated as random
- An event within a cluster does not change the risk, but it changes our **knowledge** about the risk

Common risk



Event-related dependence

Events might changes future risk

- lose one kidney, this might shorten the life of the second kidney
- after a car accident, it is likely that the driver is more careful
- one myocardial infarction might lead to a more healty lifestyle

Event-related dependence

Events might changes future risk

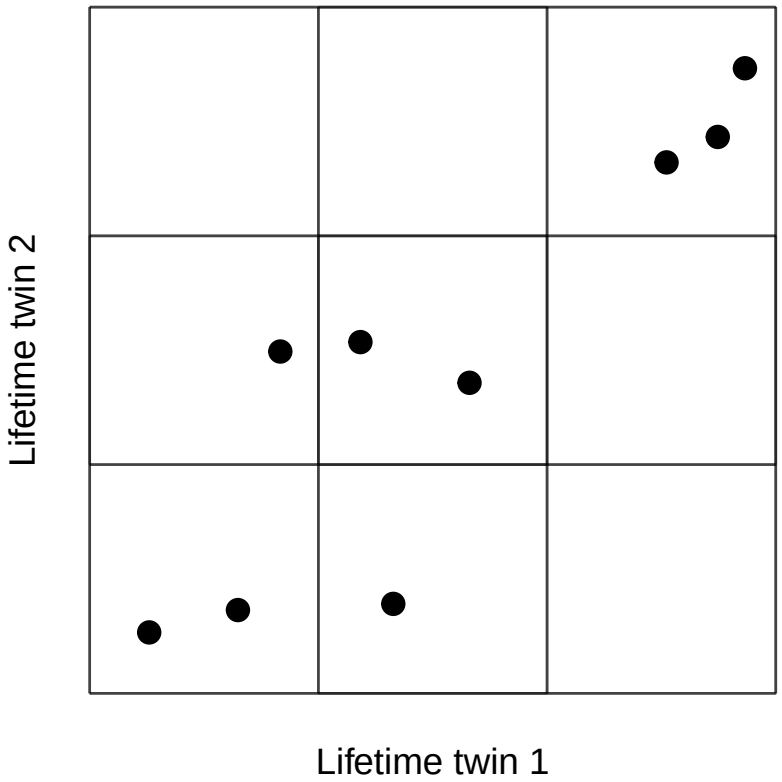
- lose one kidney, this might shorten the life of the second kidney
- after a car accident, it is likely that the driver is more careful
- one myocardial infarction might lead to a more healty lifestyle

More often, mixed causes

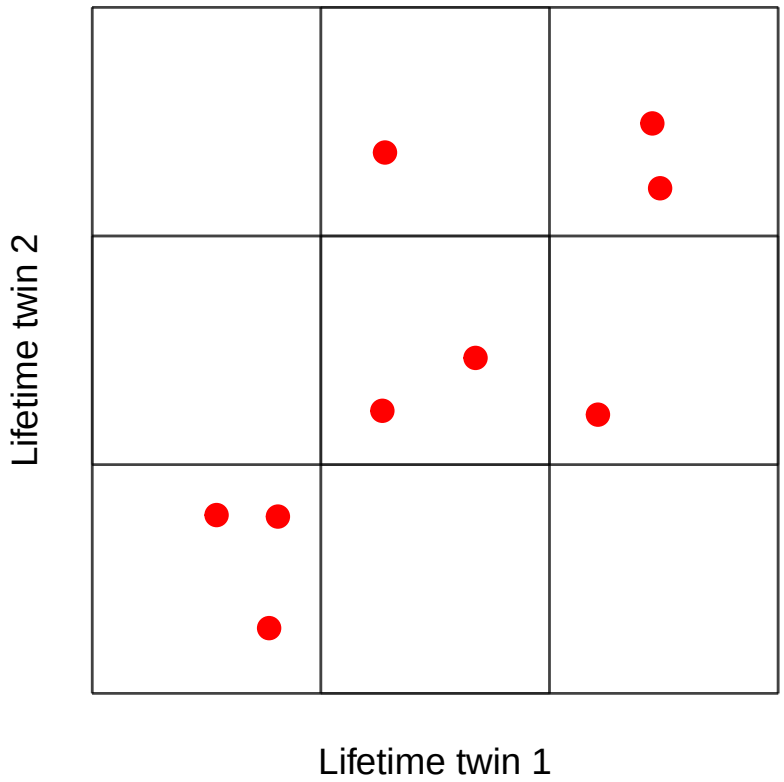
- bad driving skills are an increased risk (common risk), but an accident decreases future risk (event-related dependence)
- infections: infection decreases future risk for an individual but increases the risk for other individuals from the same family

Timing of dependence

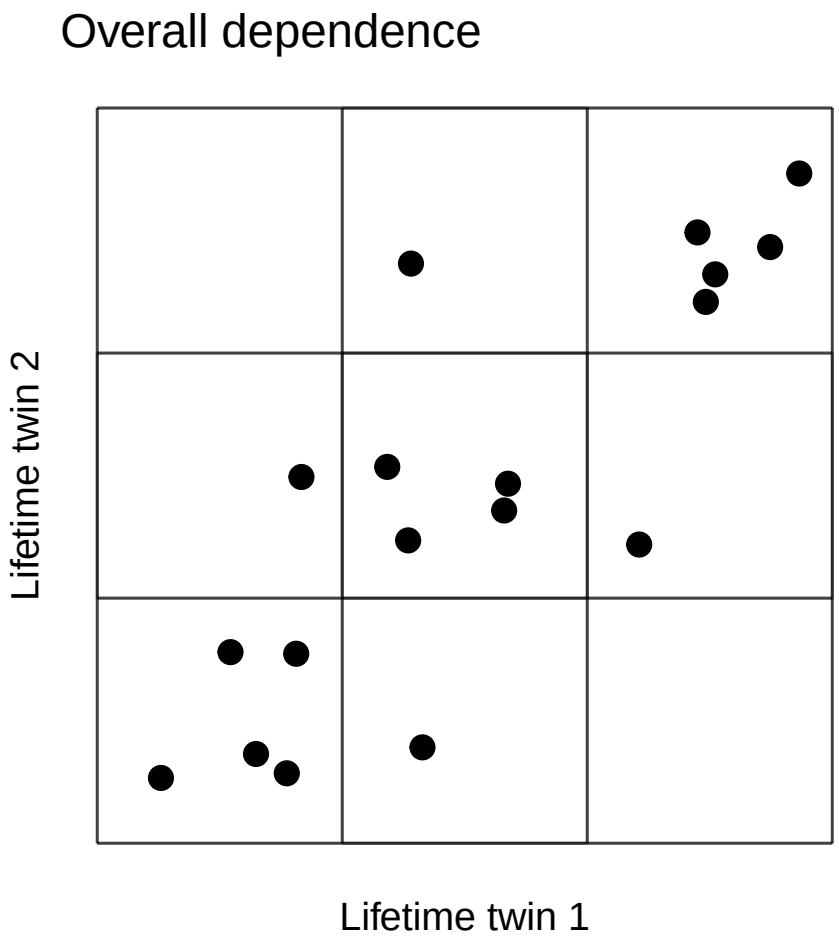
Late dependence



Early dependence



Timing of dependence



About shared frailty

Shared frailty models are models for **common risk**.

- usually clustered failures and recurrent events data
- The frailty Z_i is said to be *shared* between the observations of cluster or individual i
- Dependence within group i is induced by sharing Z_i
- If Z_i's show no variation, then the observations are independent

About shared frailty

Shared frailty models are models for **common risk**.

- usually clustered failures and recurrent events data
- The frailty Z_i is said to be *shared* between the observations of cluster or individual i
- Dependence within group i is induced by sharing Z_i
- If Z_i's show no variation, then the observations are independent

Points of interest

- How different are the clusters? (variability of Z)
- Regression coefficients (when dependence is just a nuisance)
- Prediction for a group after some events have been observed

Clustered failures

Cluster i, observation j: hazard $h_{ij}(t)$.

Approaches

- Stratify on the cluster (each cluster has a different baseline hazard)
- Add a fixed effect for each cluster (proportional hazards)
- Add a random effect for each cluster (*shared* frailty model)
- Try to avoid the correlation structure (*marginal* models)

Fixed effects & stratification

Fixed effects (FE) model:

$$h_{ij}(t) = h_0(t)\exp(\beta^\top x_{ij} + \zeta_i)$$

with $\zeta_1 = 0$ or **stratified** (S) model:

$$h_{ij}(t) = h_i(t)\exp(\beta^\top x_{ij})$$

- In (FE) the hypothesis of no heterogeneity is $H_0 : \zeta_2 = \zeta_3 = ... = \zeta_I = 0$.

Clustered failures

Problems:

- (**FE**) Must assume that number of individuals / group grows to ∞ for the asymptotics to work (in linear models we use exact sum of squares for small samples for this reason)
- (**FE** + **S**) Cannot use covariates that are constant within groups
- (**FE** & **S**) Cannot quantify the strength of the dependence
- (**S**) Loss of information: Need more *events* / group in order to estimate $h_i(t)$; otherwise, it is diffcult to distinguish cluster effects from hazard effects.

Works well when cluster sizes are large, such as hospitals in multicenter studies (Hein will talk later about that).

Clustered failures

Random effects model:

$$h_{ij}(t|Z_i) = h_0(t)Z_i \exp(\beta^\top x_{ij})$$

Here:

- clusters are a random sample from a population (of clusters)
- parameters of Z_i describe variation *between* clusters
- the hypothesis of no variation may be based on testing $\mathrm{Var} Z_i = 0$ (or Z_i having a degenerate distribution with $P(Z_i = 1) = 1$)
- only positive dependence within cluster

Joint survivor function

Assume that there are J_i individuals $j \in \{1, 2...J_i\}$ from the same cluster i, with conditional hazards

$$h_{ij}(t|Z_i) = Z_i h_{ij}(t)$$

and cumulative hazards

$$H_{ij}(t|Z_i) = Z_i \int_0^t h_{ij}(s)ds.$$

Joint survivor function

Assume that there are J_i individuals $j \in \{1, 2...J_i\}$ from the same cluster i, with conditional hazards

$$h_{ij}(t|Z_i) = Z_i h_{ij}(t)$$

and cumulative hazards

$$H_{ij}(t|Z_i) = Z_i \int_0^t h_{ij}(s)ds.$$

The **joint survivor function** may be written as

$$S(t_1, ..., t_{J_i}|Z) = P(T_{i1} > t_1, ..., T_{iJ_i} > t_{J_i}|Z)$$

Joint survivor function

Assume that there are J_i individuals $j \in \{1, 2...J_i\}$ from the same cluster i, with conditional hazards

$$h_{ij}(t|Z_i) = Z_i h_{ij}(t)$$

and cumulative hazards

$$H_{ij}(t|Z_i) = Z_i \int_0^t h_{ij}(s)ds.$$

The **joint survivor function** may be written as

$$S(t_1, ..., t_{J_i}|Z) = \exp\left(-Z(H_{i1}(t_1) + ... + H_{iJ_i}(t_{J_i}))\right)$$

Joint survivor function

The marginal joint survivor function is then:

$$\bar{S}(t_1, ..., t_{J_i}) = \mathrm{E}\left[\exp\left(-Z(H_{i1}(t_1) + ... + H_{iJ_i}(t_{J_i}))\right)\right]$$

Joint survivor function

The marginal joint survivor function is then:

$$\bar{S}(t_1, ..., t_{J_i}) = \mathrm{E}\left[\exp\left(-Z(H_{i1}(t_1) + ... + H_{iJ_i}(t_{J_i}))\right)\right]$$

- the likelihood contribution of cluster i is equal to the joint density of $(T_{i1}...T_{iJ_i})$
- For a positive random variable, $S(t) = P(X \geq t)$; the density is $P(T = t) = \frac{d}{dt}S(t)$.
- Assuming no censoring, the marginal density for times $t_1, ..., t_{J_i}$:

$$f(t_1, ...t_{J_i}) = \frac{\partial^{J_i}}{\partial t_1 ... \partial t_{J_i}} \bar{S}(t_1, ...t_{J_i})$$

Likelihood

With censoring, the partial derivatives are taken only at *event* time points. Denote $D_j = 1$ if t_j is an event. Then the marginal density of the *observed* event time points is:

$$f(t_1 \ldots t_{J_i}; D_1 \ldots D_{J_i}) = (-1)^{D_\cdot} \prod_j h_j(t_j)^{D_j} \mathcal{L}^{(D_\cdot)}(H_1(t_1) + \ldots + H_{iJ_i}(t_{J_i}))$$

- Individual hazards contribute at event time points.
- The censored observations contribute to the total cumulative hazard, that is in the Laplace transform argument.

Quantifying dependence

An identifiability assumption is made for the frailty distribution (usually, $E[Z] = 1$).

Different measures[1]

- the variability of Z (e.g. $Var[Z]$ - most popular, or $Var[\log Z]$)
  - $Var[Z]$ - most popular, or $Var[\log Z]$
- correlation-type measure
  - bivariate measures such as Kendall's τ or median concordance
- the predicted survival or cumulative hazard
  - difficult to visualize for more than a few clusters

[1] Hougaard (2000)

General notes

- Early or late dependence? This is where different distributions for Z play a role
  - Gamma frailty emphasizes late dependence
  - Positive stable emphasizes early dependence
  - log-normal / inverse Gaussian somewhere in between

Covariates

- covariates should usually be included, since the frailty explains unobserved variance due to common sources
- cluster-specific take out from the between-cluster (frailty) variation
- When cluster sizes are small (e.g. of size 2), marginal non-proportional hazards may appear as frailty! (as we saw in the univariate frailty case)

Goodness-of-fit

Likelihood ratio test

- Under no frailty (i.e. frailty variance 0) the limiting model is a Cox model
- The likelihood ratio test (LRT) can be used to test the null hypothesis $H_0 : \mathrm{Var}[Z] = 0$ vs $H_A : \mathrm{Var}[Z] > 0$
- Because the hypothesis is one sided, the asymptotic distribution of the likelihood ratio statistic follows a $\frac{1}{2}\chi^2(1) + \frac{1}{2}\chi^2(0)$

Commenges-Andersen score test

- A score test based on martingale residulas for $H_0 : \mathrm{Var}[Z] = 0$ was proposed by Commenges and Andersen (1995).
- implemented in `frailtyEM`

Marginal modeling

The coordinate-wise approach

- stratfied Cox proportional hazards model with strata 1 the first observation from each cluster, strata 2 the second observation from each cluster, etc
- strata by covariate interaction
- for example: the Wei-Lin-Weissfeld (WLW) model for recurrent events data
- heavy assumptions on the censoring, not model based

Marginal modeling

The coordinate-wise approach

- stratfied Cox proportional hazards model with strata 1 the first observation from each cluster, strata 2 the second observation from each cluster, etc
- strata by covariate interaction
- for example: the Wei-Lin-Weissfeld (WLW) model for recurrent events data
- heavy assumptions on the censoring, not model based

The working independence approach

- assume proportional hazards on the marginal hazards
- neglect dependence and fit a Cox proportional hazards model
- obtain standard errors from a sandwich estimator

Recurrent events

Models

Consider λ_i the *intensity* of a process that generates recurrent events. Specification is the same as the hazard:

$$\lambda_i(t|Z_i) = \lambda_0(t) Z_i \exp(\beta^\top x_i)$$

Calendar time

- time origin: time since the start of the recurrent events process
- Conditional on Z_i, individual i is describe by a Poisson process
- also known as Andersen-Gill formulation

Gap time

- time origin: time since previous event
- Conditional on Z_i, individual i is described by a renewal process

Recurrent events representation

tstart	tstop	gap	event
0	1	1	1
1	3	2	1
3	7	4	0

```
Surv(tstart, tstop, event) # calendar time
```

```
[1] (0,1]  (1,3]  (3,7+]
```

```
Surv(gap, event) # gap time
```

```
[1] 1  2  4+
```

Recurrent vs clusters

Similarities

- the construction of the likelihood and estimation are technically the same as in the cluster case

Differences

- the frailty represents *within-individual* homoegeneity and *between-individual* heterogeneity
- event-related dependence can be accommodated using time-dependent covariates (e.g. time since previous event, number of previous events)[2]
- time aspect not as straight forward as before

[2] Cook & Lawless (2007)

Estimation

Likelihood construction

The goal is to maximize the **observed data** likelihood (the marginal likelihood). The marginal contribution of an individual can be obtained by taking derivatives of the marginal survivor function:

$$L_i = (-1)^{D_{i\cdot}} \prod_j h_{ij}(t_{ij})^{D_{ij}} \mathcal{L}^{(D_{i\cdot})}(H_{i1}(t_{i1}) + \ldots H_{in_i}(t_{in_i}))$$
$$= (-1)^{D_{i\cdot}} \prod_j h_{ij}(t_{ij})^{D_{ij}} \mathcal{L}^{(D_{i\cdot})}(H_{i\cdot})$$

This can't be estimated directly because:

- the derivatives of the Laplace transform are usually difficult to calculate (except for the gamma frailty)
- the hazard involves one parameter at every event time point in the data

EM Algorithm

The marginal likelihood can be written as

$$P(\text{data}_i) = \int_\theta P(\text{data}_i|Z_i) f_\theta(Z_i) dZ_i$$

The "complete data likelihood" is

$$P(\text{data}_i|Z_i) f_\theta(Z_i)$$

Empirical Bayes estimates

The "posterior" distribution of Z_i can be obtained from Bayes' theorem:

$$f(Z_i|\text{data}_i) = \frac{P(\text{data}_i|Z_i)f(Z_i)}{P(\text{data}_i)}$$

By replacing with the Laplace transforms, we obtain that

$$E[Z_i|\text{data}_i] = -\frac{\mathcal{L}^{(D_i+1)}(H_i.)}{\mathcal{L}^{(D_i)}(H_i.)}$$

Not available in closed form, except for the gamma frailty model:

$$E[Z_i|\text{data}_i] = \frac{\theta + D_i.}{\theta + H_i.}$$

EM Algorithm

E step

- Calculate the expectation of the complete-data log-likelihood:

$$E_{Z_i} \log P(\mathrm{data}_i|Z_i) + \log f_\theta(Z_i)$$

- in practice, calculate $E[Z_i|\mathrm{data}]$ and $E[\log Z_i|\mathrm{data}]$

M step

- Maximize the complete data likelihood by treating the frailty as fixed
- A Cox-type likelihood and a term involving θ.

EM Algorithm

Disadvantages of the EM algorithm:

- $E[Z_i|data]$ can be calculated by taking derivatives of the Laplace transform, but $E[\log Z_i|data]$ is in general difficult to calculate
- Need f_θ in closed form, that doesn't exist for PVF or postive stable

We would like to

- not have to express the density in closed form
- no have to calculate $E[\log Z_i|\mathrm{data}_i]$ (can't be expressed as a function of derivatives of the Laplace transform)

Profile EM algorithm

General idea

- Apply the EM algorithm for fixed values of θ, obtaining $\widehat{\beta}_\theta$, $\widehat{h_0(\cdot)}_\theta$ that maximize the likelihood with θ fixed, $L_\theta(\beta, h_0(\cdot))$
- Maximize $L_\theta(\beta, h_0(\cdot))$ in θ

Advantages

- like the EM algorithm, we can use any distribution with a known Laplace transform
- implemented in the `R` package `frailtyEM`

Other methods

Penalized likelihood (`survival`)

- consider that Z_i are unknown variables (similar to fixed effects)
- then we could use a Cox model to estimate all the parameters, including Z_i
- penalize the Cox partial likelihood so that Z_i follow either a gamma or log-normal distribution
- very fast, but only works in a limited number of scenarios

Other methods

Parametric models (`parfm` and `frailtypack`)

- take a parametric baseline hazard, so that h_0 depends on a small number of parameters (e.g. Weibull, Gompertz, spline-based)
- attemt to directly maximize the marginal log-likelihood using numeric integration
- in practice works well for gamma and log-normal.

Empirical Bayes estimates revisited

- In the univariate case, we used $E[Z_i|T \geq t_i]$, where t_i is the event or censoring time of individual i.
- In the multivariate case we have $E[Z_i|\mathrm{data}_i]$, where data_i contains information about the whole survival and event history from cluster i at the end of follow-up
- In the univariate case, the marginal hazard of individual i is

$$\bar{h}_i(t) = E[Z_i|T \geq t]h_i(t)$$

- In the multivariate case, we can write the marginal hazard of indiviudual (i, j) as

$$\bar{h}_{ij}(t) = E[Z_i|\mathrm{data}_i(t)]h_{ij}(t)$$

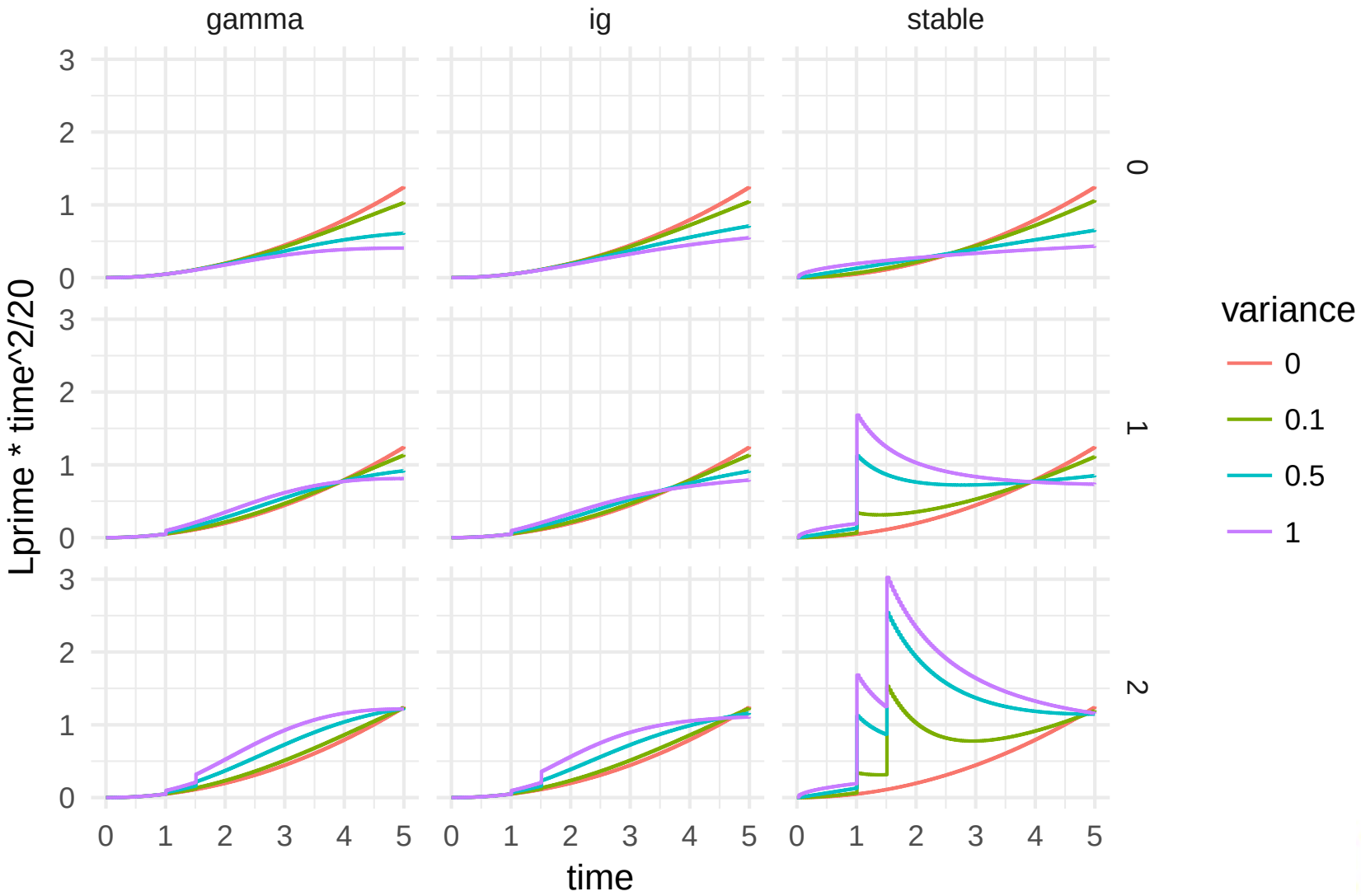
where $\mathrm{data}_i(t)$ is the information accumulated in the whole cluster i up to time t.

Small example

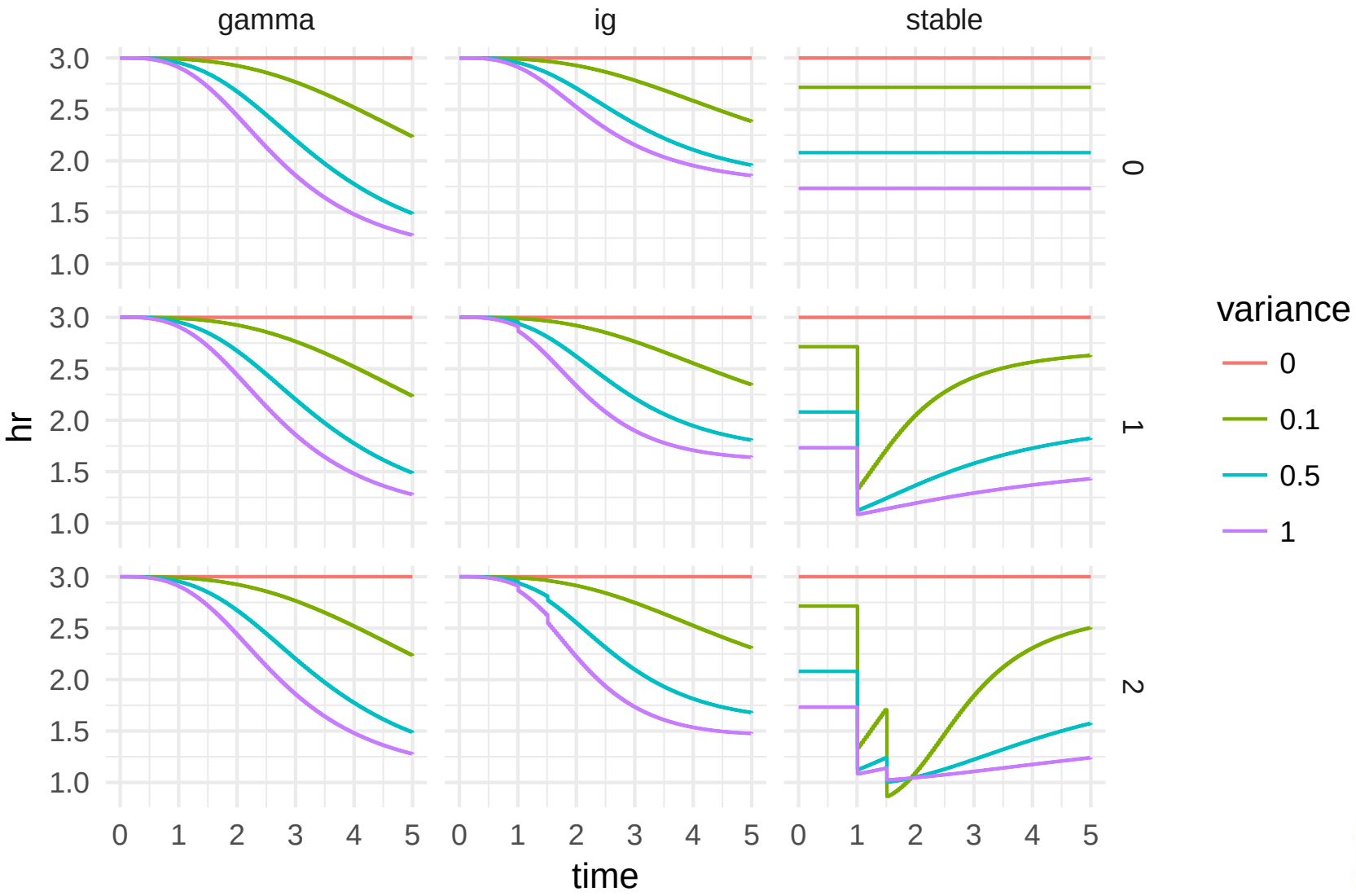
We take two individuals in the same logic as before, and 3 scenarios:

- Scenario 1: no events
- Scenario 2: one event at time 1
- Scenario 3: one event at time 1, one event at time 1.5

Marginal hazard - baseline individual



Marginal hazard ratios



Final words

Before we go to software and applications. . .

- The PVF family of distributions is great, but in practice it is rarely used
- The standard are the gamma frailty and the log-normal frailty models (the latter is *not* in the PVF family)
  - gamma frailty is mostly for mathematical convenience, but there are also some theoretical arguments (Abbring & van den Berg 2007)
  - log-normal allows for negatively correlated frailty models (more about this, later)
- The positive stable distribution is advocated for by Hougaard, but in practice it is difficult to justify its use (and difficult to estimate)

Software overview

In short...

- In `SAS` and `Stata`, shared frailty models with a gamma or log-normal distributions can be estimated using the penalized likelihood method

In R

Semiparametric

- `survival::coxph` is the *de facto* standard for gamma and log-normal frailty models. It is very fast, but it lacks some features
- `coxme::coxme` fits log-normal frailty models including correlated frailty models.
- `frailtyEM::emfrail` (by Balan & Putter, 2017, wonder what they are up to?) aims to be a complete implementation of the EM algorithm with numerous methods for the user. However, it is slow when there are many events per individual.

In R

Parametric

- `frailtypack` has a lot of features for gamma and log-normal frailty modes, including spline-approximated baseline hazard;
- `parfm` is a simple package for parametric frailty models with more distributions, but it lacks some features

Experimental

- `frailtySurv` implements a pseudo-maximum likelihood estimation technique, supports many distributions
- `phmm`, `frailtyHL`

References

- Balan, T. A., & Putter, H. `frailtyEM`: an `R` Package for Estimating Semiparametric Shared Frailty Models.
- Cook, R. J., & Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.
- Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer Science & Business Media.
- Kalbfleisch, J. D., & Prentice, R. L. (2011). *The statistical analysis of failure time data*. John Wiley & Sons.