Frailty Models: Theory & Practice

Shared Frailty Models

Theodor Balan & Hein Putter

Department of Biomedical Data Sciences
Leiden University Medical Center

Statistics in Practice session
IBC Barcelona, 9 July 2018

Omitted covariates in paired data

Paired life times

- Paired individuals, with life times T_1 and T_2
- Covariate values are shared for individuals from the same pair
- T_1 and T_2 have the same distribution with hazard function given by

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

- T_1 and T_2 are assumed independent, given X

Paired life times

- Paired individuals, with life times T_1 and T_2
- Covariate values are shared for individuals from the same pair
- T_1 and T_2 have the same distribution with hazard function given by
$$\lambda(t \mid x) = \lambda_0(t) \exp(\beta x)$$
- T_1 and T_2 are assumed independent, given X
- x is realization of X with density $f_X(x)$
- Denote by $f(t \mid x)$ and $S(t \mid x)$ the density and survival function of T, given $X = x$
- Marginal survival function of T is then given by
$$\overline{S}(t) = \int S(t \mid x) f_X(x) dx$$

Positive correlation

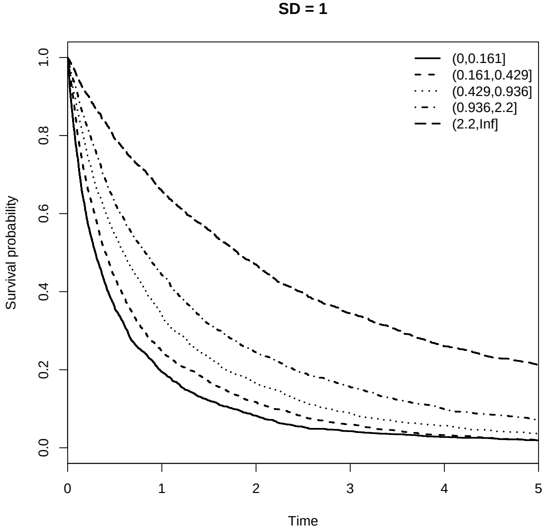
Heuristic explanation

- Consider one pair, with life times T_1 and T_2
- Marginal survival function of T_1 and T_2 given by $\overline{S}(t)$
- If t_1 is observed, distribution of T_2 (given t_1) will change
- If t_1 is large, then it is likely that T_1 has low hazard
- This in turn makes it more likely that value of x in the pair is low if $\beta > 0$
- Since x is shared, then T_2 likely also has low hazard
- So we expect a large life time t_2 as well
- Result: T_1 and T_2 are positively correlated

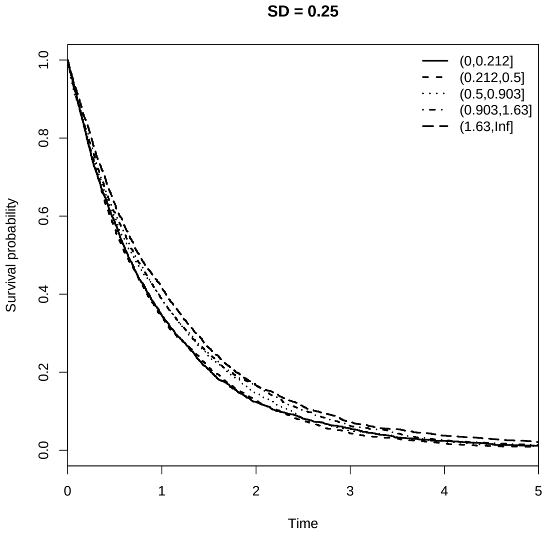
Illustration

- Simulated 10 000 pairs, where
- $X \sim N(0, \sigma^2)$ (shared between individuals in the same pair)
- Given $X = x$, T_1 and T_2 generated independently according to exponential distribution with rate $\lambda \exp(\beta x)$, with $\lambda = 1$ and $\beta = 1$
- Divide t_1 into quintiles, and plotted Kaplan-Meiers for these quintiles
- Standard deviation of X taken to be $\sigma = 0.25, 1, 4$

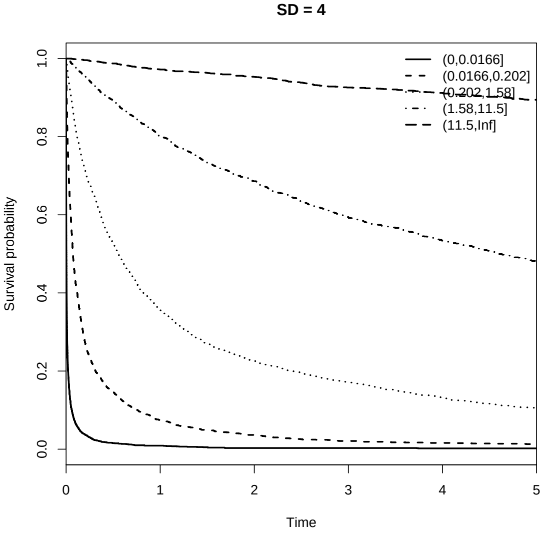
Plots


SD = 1
Survival probability
Time
(0,0.161]
(0.161,0.429]
(0.429,0.936]
(0.936,2.2]
(2.2,Inf]
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5

Plots


SD = 0.25
Survival probability
Time
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
(0,0.212]
(0.212,0.5]
(0.5,0.903]
(0.903,1.63]
(1.63,Inf]

Plots


SD = 4
Survival probability
Time
(0,0.0166]
(0.0166,0.202]
(0.202,1.58]
(1.58,11.5]
(11.5,Inf]
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5

Conclusion

- A shared covariate X affecting the hazard of T_1 and T_2 induces positive correlation
- The correlation increases when the variance of X increases

Shared frailty models

Frailty models & correlated data

With maximum one event / individual, the frailty model explained *unobserved individual risk*.

In the "shared frailty" case, the frailty explaines *unobserved common risk*.

Frailty models & correlated data

With maximum one event / individual, the frailty model explained *unobserved individual risk*.

In the “shared frailty” case, the frailty explaines *unobserved common risk*.

Suitable for two types of outcome:

- **clustered failures**, where several individuals belong to the same cluster (eg. in families, patients in centers, etc)
- **recurrent events**, where an individual may experience several events (e.g. infections, hospital admissions)

Clustered failures

The hazard of individual j in cluster i is:

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

Here:

- $\log Z_i$ plays the role of a random intercept (on the linear predictor scale)
- Clusters are a random sample from a population (of clusters)
- Variation between Z_i describes a variation *between* clusters (and dependence *within* clusters, as Hein showed)

Recurrent events

Gap-time

- based on a renewal process where the time scale resets after each event
- the "hazard" of the *j*-th event of individual *i* is

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

- the frailty describes the correlation of the gap times
- close in spirit to the clustered failure case

Recurrent events

Calendar time

- based on a non-homogeneous Poisson process (or "Andersen-Gill" formulation)
- the event history of an individual is described by a counting process N_i
- the intensity of N_i is specified as:

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

- the frailty describes an overall higher or lower intensity of the Poisson process
- closer in spirit to the univariate survival data

Recurrent events

A few comments

- Recurrent events are a complex topic of their own (Cook & Lawless, 2007)
- The choice of time scale and stochastic model is typically based on mechanistic or biological considerations
- In general, there is no easy correspondence between the two models except for easy cases (e.g. homogeneous Poisson process and exponentially distributed gap-times)

Distributions

- In shared frailty models, different frailty distributions emphasize different types of dependence:
  - the gamma frailty weighs late dependence more
  - the positive stable weighs early dependence more
  - the inverse Gaussian is somewhere in the middle
  - this dependence is usually only tractable for bivariate survival data

Distributions

- In shared frailty models, different frailty distributions emphasize different types of dependence:
  - the gamma frailty weighs late dependence more
  - the positive stable weighs early dependence more
  - the inverse Gaussian is somewhere in the middle
  - this dependence is usually only tractable for bivariate survival data
- In univariate frailty with covariates, we saw that different frailty distributions "shrink" the covariate effects in different ways
- With shared frailty models, this is not so clear any more
  - as the cluster size grows, the marginal correlation between event times becomes more prominent

Empirical Bayes estimates

Remember that for the gamma distribution, we looked at the frailty of the survivors at time t:

$$
\begin{aligned}
f(z|T \geq t) &= \frac{P(T \geq t|Z = z)f(z)}{P(T \geq t)} \\
&\propto S(t|Z = z)f(z) \\
&= \exp(-zH_i(t))f(z) \\
&\propto z^{\alpha-1}e^{-(\gamma+H_i(t))z}
\end{aligned}
$$

Empirical Bayes estimates

Remember that for the gamma distribution, we looked at the frailty of the survivors at time t:

$$\begin{aligned} f(z|T \geq t) &= \frac{P(T \geq t|Z = z)f(z)}{P(T \geq t)} \\ &\propto S(t|Z = z)f(z) \\ &= \exp(-zH_i(t))f(z) \\ &\propto z^{\alpha-1}e^{-(\gamma+H_i(t))z} \end{aligned}$$

Now, instead of $P(T \geq t|Z = z)$, the history of a cluster at time t includes D events and aggregated cumulative hazard $H_{i.}(t) = \sum_j H_{ij}(t)$. We call this information $\mathrm{data}_i(t)$.

Empirical Bayes estimates

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

$$f(z|T \geq t) = \frac{P(\mathrm{data}_i(t)|Z_i = z_i)f(z_i)}{P(\mathrm{data}_i(t))}$$

$$\propto z^{D_i(t_-)} \exp(-zH_{i\cdot}(t))f(z)$$

$$\propto z^{\alpha + D_i(t_-) - 1} e^{-(\gamma + H_{i\cdot}(t))z}$$

Empirical Bayes estimates

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

$$f(z|T \geq t) = \frac{P(\mathrm{data}_i(t)|Z_i = z_i)f(z_i)}{P(\mathrm{data}_i(t))}$$

$$\propto z^{D_i(t_-)} \exp(-zH_i.(t))f(z)$$

$$\propto z^{\alpha + D_i(t_-) - 1} e^{-(\gamma + H_i.(t))z}$$

This is a gamma distribution where $\alpha \to \alpha + D_i(t)$ and $\gamma \to \gamma + H_{i.}(t)$. The marginal hazard for individual j from cluster i is then given by

$$h_{ij}(t) = \frac{\theta + D_i(t_-)}{\theta + H_i(t_-)} h_0(t) \exp(\beta^\top x_{ij})$$

Empirical Bayes estimates

With D_i and $H_{i.}$ the total number of events and aggregated cumulative hazard in cluster i at the end of follow-up, the **empirical Bayes** estimate of the frailty is given by

$$E[Z_i|\mathrm{data}_i] = \frac{\theta + D_i}{\theta + H_{i.}}$$

Empirical Bayes estimates

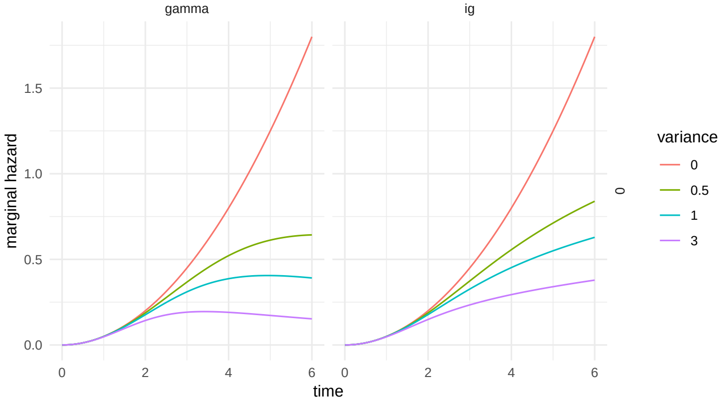
With D_i and $H_{i.}$ the total number of events and aggregated cumulative hazard in cluster i at the end of follow-up, the **empirical Bayes** estimate of the frailty is given by

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

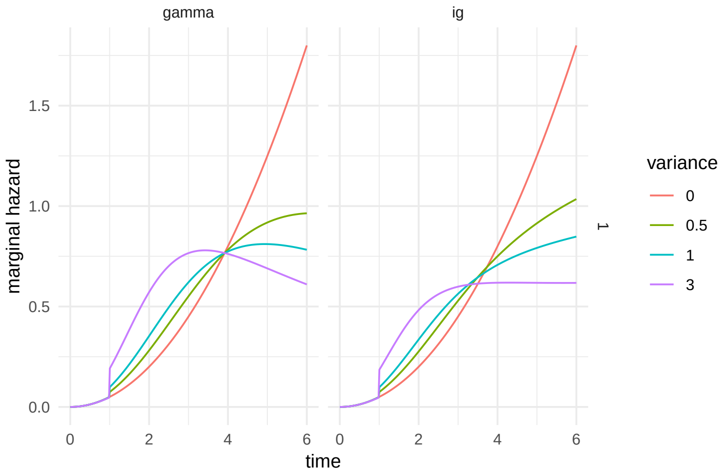
In the general case, this involves taking successive derivatives of the Laplace transform:

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

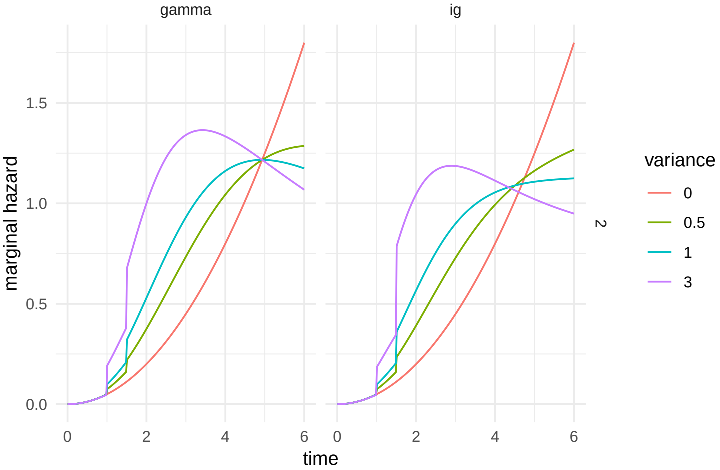
Marginal hazard - no events



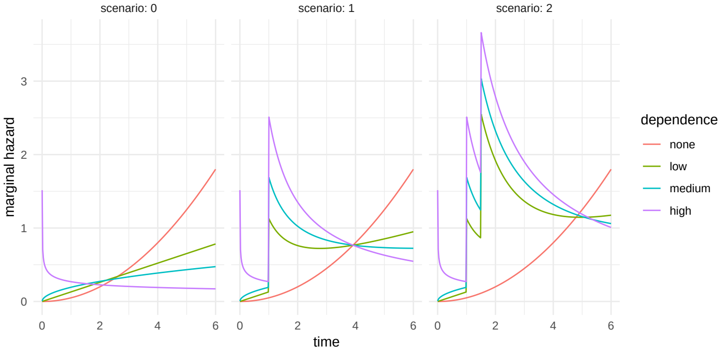
Marginal hazard - one event at t = 1



Marginal hazard - one event at t = 1, one at t = 1.5



Marginal hazard - one event at t = 1, one at t = 1.5



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:

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

This can't be estimated directly because:

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

EM Algorithm

Assume that the frailty of the distribution is defined by the parameter θ.

The marginal likelihood can be written as

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

The "complete data likelihood" is

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

EM algorithm

- The EM algorithm for the gamma frailty was proposed by Nielsen et al. (1992) and Klein (1992)
- The E step gets more difficult for distributions except gamma (still somewhat tractable for the PVF distributions)
- An adaptation is implemented in the `R` package `frailtyEM`

Penalized likelihood

- mathematically equivalent to the EM algorithm, implemented in the `survival::coxph()`, SAS, Stata
- works only with gamma and log-normal frailty, and is a bit limited in features (e.g. confidence interval for $\mathrm{Var}[Z]$)
- idea: consider that $\log Z_i$ are unknown parameters, but penalize them in such a way that the frailty distribution corresponds to the gamma or log-normal (Therneau & Grambsch, 2000)

Other approaches

- take a parametric baseline hazard, so that h_0 depends on a small number of parameters (e.g. Weibull, Gompertz, spline-based)
- technically the EM algorithm can be used, but in general the maximization can be carried out directly
- `parfm` package supports more distributions but has fewer features, `frailtypack` for gamma and log-normal distributions, but very mature and rich in features

Other

- `frailtySurv` implements a pseudo-maximum likelihood estimation technique; idea is to avoid the EM algorithm by maximizing the baseline hazard values in the marginal likelihood in a succession instead of simultaneously.
- `phmm` and h likelihood approach `frailtyHL`

Software and data representation

Software overview (from the `frailtyEM` documentation)

	frailtyEM	survival	coxme	frailtySurv	frailtyHL	frailtypack	parfm	phmm
Distributions								
Gamma	yes	yes	no	yes	no	yes	yes	no
Log-normal	no	yes	yes	yes	yes	yes	yes	yes
PS	yes	no	no	no	no	no	yes	no
IG	yes	no	no	yes	no	no	yes	no
Compound Poisson	yes	no	no	no	no	no	no	no
PVF	yes	no	no	yes	no	no	no	no
Data								
Clustered failures	yes	yes	yes	yes	yes	yes	yes	yes
Recurrent events (AG)	yes	yes	yes	no	no	yes	no	no
Left truncation	yes	no	no	no	no	yes	yes	no
Correlated structure	no	no	yes	no	no	yes	no	yes
Estimation								
Semiparametric	yes	yes	yes	yes	yes	no	no	yes
Posterior frailties	yes	yes	no	no	no	yes	no	no
Conditional Λ_0, S_0	yes	limited	no	yes	no	yes	yes	no
Marginal Λ_0, S_0	yes	no	no	no	no	no	no	no

Syntax - `survival` package

Cox model for failure time data:

```
coxph(Surv(tstop, status) ~ x1 + x2 + ... )
```

with `tstop` is an event time if `status==1` and censoring time otherwise,

Syntax - `survival` package

Cox model for failure time data:

```
coxph(Surv(tstop, status) ~ x1 + x2 + ... )
```

with `tstop` is an event time if `status==1` and censoring time otherwise,

- If the individuals are clustered, this will give wrong standard errors.
- We could use a **marginal** proportional hazards model with:

```
coxph(Surv(tstop, status) ~ cluster(id) + x1 + x2 + ... )
```

where `id` is a column that identifies the cluster.

Syntax - survival package

Marginal proportional hazards model:

```
coxph(Surv(tstop, status) ~
        cluster(id) + x1 + x2 + ... )
```

A gamma and a log-normal frailty model:

```
coxph(Surv(tstop, status) ~
        frailty(id) + x1 + x2 + ... )
coxph(Surv(tstop, status) ~
        frailty(id, "gaussian") + x1 + x2 + ... )
```

cluster() vs frailty()

`cluster`:

- fits a proportional hazards model without frailty
- uses a sandwich estimator for standard errors (conceptually similar to a GEE approach)
- does not estimate within-cluster correlation structures
- covariate estimates have a population-averaged interpretation

`frailty`:

- proportional hazards conditional on the frailty
- the frailty model is not compatible with the marginal proportional hazards model (unless the frailty follows a positive stable distribution)
- extra information with between-cluster independence

Recurrent events representation

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

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

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

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

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

Rats data set

```r
head(rats)
```

```
##   litter rx time status sex
## 1      1  1  101      0   f
## 2      1  0   49      1   f
## 3      1  0  104      0   f
## 4      2  1   91      0   m
## 5      2  0  104      0   m
## 6      2  0  102      0   m
```

Rats data set

```
library(survival)
m1 <- coxph(formula = Surv(time, status) ~
              rx + sex + frailty(litter),
            data = rats)
summary(m1)
```

```
Call:
coxph(formula = Surv(time, status) ~ rx + sex + frailty(litter),
    data = rats)

  n= 300, number of events= 42

                coef    se(coef) se2    Chisq DF    p
rx               0.7947 0.3138   0.3104  6.41  1.00 0.01100
sexm            -3.1437 0.7374   0.7264 18.18  1.00 0.00002
frailty(litter)                          17.59 14.62 0.26000

     exp(coef) exp(-coef) lower .95 upper .95
rx     2.21370     0.4517   1.19674     4.095
sexm   0.04312    23.1892   0.01016     0.183

Iterations: 6 outer, 29 Newton-Raphson
     Variance of random effect= 0.4664846   I-likelihood = -199.5
Degrees of freedom for terms=  1.0  1.0 14.6
Concordance= 0.873  (se = 0.046 )
Likelihood ratio test= 79.72  on 16.56 df,   p=3e-10
```

cluster()

```r
m2 <- coxph(formula = Surv(time, status) ~
              rx + sex + cluster(litter),
            data = rats)
summary(m2)
```

```
Call:
coxph(formula = Surv(time, status) ~ rx + sex + cluster(litter),
    data = rats)

  n= 300, number of events= 42

         coef exp(coef) se(coef) robust se      z Pr(>|z|)
rx    0.79100   2.20559  0.30936   0.29271  2.702  0.00689 **
sexm -3.06769   0.04653  0.72480   0.72066 -4.257 2.07e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     exp(coef) exp(-coef) lower .95 upper .95
rx     2.20559     0.4534   1.24271     3.915
sexm   0.04653    21.4923   0.01133     0.191

Concordance= 0.764  (se = 0.044 )
Rsquare= 0.154   (max possible= 0.777 )
Likelihood ratio test= 50.04  on 2 df,   p=1e-11
Wald test            = 21.23  on 2 df,   p=2e-05
Score (logrank) test = 42.68  on 2 df,   p=5e-10,   Robust = 25.74  p=3e-06
```

Results from frailtyEM

```
library(frailtyEM)
m_gamma <- emfrail(Surv(time, status) ~
                     rx + sex + cluster(litter),
                   data =  rats)
summary(m_gamma)
```

```
Call:
emfrail(formula = Surv(time, status) ~ rx + sex + cluster(litter),
    data = rats)

Regression coefficients:
        coef exp(coef) se(coef) adj. se       z    p
rx    0.7873    2.1974   0.3135  0.3135  2.5112 0.01
sexm -3.1341    0.0435   0.7385  0.7409 -4.2298 0.00
Estimated distribution: gamma / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.201
no-frailty Log-likelihood: -200.426
Log-likelihood: -199.73
LRT: 1/2 * pchisq(1.39), p-val 0.119

Frailty summary:
                   estimate lower 95% upper 95%
Var[Z]                0.445     0.000     1.678
Kendall's tau         0.182     0.000     0.456
Median concordance    0.179     0.000     0.464
E[logZ]              -0.239    -1.038     0.000
Var[logZ]             0.559     0.000     3.678
theta                 2.245     0.596       Inf
Confidence intervals based on the likelihood function
```

frailtyEM, positive stable distribution

```
library(frailtyEM)
m_stable <- emfrail(Surv(time, status) ~
                        rx + sex + cluster(litter),
                    distribution = emfrail_dist("stable"),
                    data =  rats)
summary(m_stable)
```

```
Call:
emfrail(formula = Surv(time, status) ~ rx + sex + cluster(litter),
    data = rats, distribution = emfrail_dist("stable"))

Regression coefficients:
        coef exp(coef) se(coef) adj. se       z    p
rx    0.8031    2.2325   0.3138  0.3147  2.5522 0.01
sexm -3.1821    0.0415   0.7647  0.8088 -3.9343 0.00
Estimated distribution: stable / left truncation: FALSE

Fit summary:
Commenges-Andersen test for heterogeneity: p-val  0.201
no-frailty Log-likelihood: -200.426
Log-likelihood: -200.229
LRT: 1/2 * pchisq(0.394), p-val 0.265

Frailty summary:
                   estimate lower 95% upper 95%
Kendall's tau         0.050     0.000     0.251
Median concordance    0.049     0.000     0.248
E[logZ]               0.031     0.000     0.194
Var[logZ]             0.179     0.000     1.288
Attenuation           0.950     0.749     1.000
theta                18.847     2.983       Inf
Confidence intervals based on the likelihood function
```

frailtypack

```r
library(frailtypack)
frailtyPenal(Surv(time, status) ~
                rx + sex + cluster(litter),
                    n.knots=12,kappa=10000,
                    data =  rats)
```

```
Be patient. The program is computing ...
The program took 0.05 seconds
Call:
frailtyPenal(formula = Surv(time, status) ~ rx + sex + cluster(litter),
    data = rats, n.knots = 12, kappa = 10000)


  Shared Gamma Frailty model parameter estimates
  using a Penalized Likelihood on the hazard function

          coef exp(coef) SE coef (H) SE coef (HIH)        z           p
rx    0.817562 2.2649721    0.313388      0.313388  2.60879 9.0863e-03
sexm -3.134454 0.0435235    0.741004      0.741004 -4.23001 2.3368e-05

    Frailty parameter, Theta: 0.467426 (SE (H): 0.456666 ) p = 0.15302

      penalized marginal log-likelihood = -255.5
      Convergence criteria:
      parameters = 8.89e-05 likelihood = 2.61e-05 gradient = 1.03e-07

      LCV = the approximate likelihood cross-validation criterion
            in the semi parametrical case     = 0.907644

      n= 300
      n events= 42  n groups= 100
      number of iterations:  11
```

About the R packages

- each offers slightly different functionality and was built with a certain purpose in mind
- `frailtyEM` aims to be quite complete for shared frailty models, for example
- `frailtypack` aims to fit a lot of extended frailty models, such as nested frailty models and joint frailty models

Some resources:

- CRAN Task View: Survival Analysis at https://cran.r-project.org/view=Survival
- Therneau & Grambsch, Modeling Survival Data: Extending the Cox Model (2000) for general survival analysis in `R`
- For example, the vignette & documentation of `frailtyEM` has several examples and describes the theory behind estimation and inference with the EM algorithm

Conclusion

- 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 by Hougaard, but in practice it is rarely used

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.