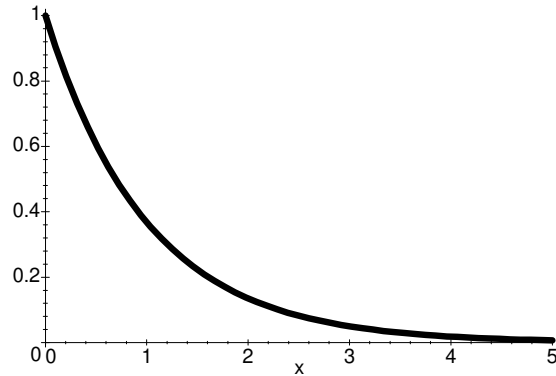


Heterogeneity in survival analysis

- Differences between individuals are of two kinds:
 - those observed through covariates
 - those that are unobserved
- The unobserved differences may be due to:
 - Environment
 - Life style
 - Genes (a few or many)
- The unobserved differences are often disregarded when analyzing survival data
- As explained in the **frailty** theory, we need to keep the unobserved differences in mind when interpreting the results of an analysis

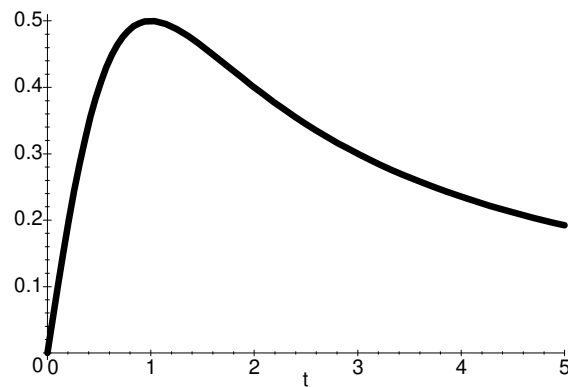
Typical shapes of the hazard rate



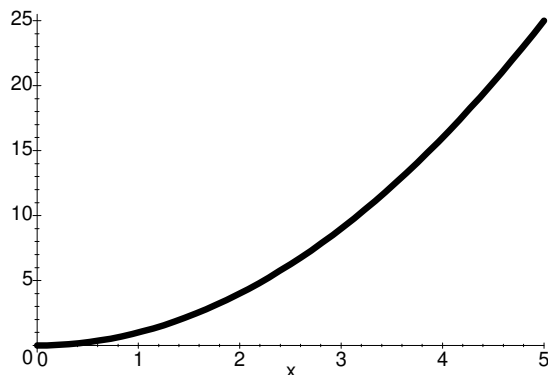
Mortality of
patients with
myocardial
infarction.

How can the
different patterns
be interpreted?

We need to
distinguish
between the
individual hazards
and the
population hazard



Mortality of cancer
patients.
Incidence of
childhood
leukemia.



General
mortality.
Incidence
of most
cancers.

Population hazard and individual hazards

- We will distinguish between the **population hazard rate** and the **individual hazard rates**
- The population hazard rate is influenced by selection: those with highest risk experience the event early
- The shape of the population hazard may be entirely different from that of the individual hazards
- Hence the population hazard can not generally be interpreted as giving information on individual development in risk

The proportional frailty model

We assume that the heterogeneity between individuals may be described by a frailty variable Z

The frailty variable is a non-negative random variable, with large values of Z corresponding to "frail" individuals

It is common to assume that the frailty has a multiplicative effect on the hazard, i.e.

$$\alpha(t \mid Z) = Z \cdot \alpha(t)$$

Here $\alpha(t)$ is the individual baseline hazard (corresponding to $Z=1$) and $\alpha(t \mid Z)$ is the individual hazard for an individual with frailty Z

Note that the frailty is not observed

The most common choice of frailty distribution is the gamma distribution with density ($z > 0$)

$$f(z) = \frac{\nu^\eta}{\Gamma(\eta)} z^{\eta-1} \exp(-\nu z)$$

We restrict attention to gamma frailty in this short course (but note that other frailty models may be useful as well)

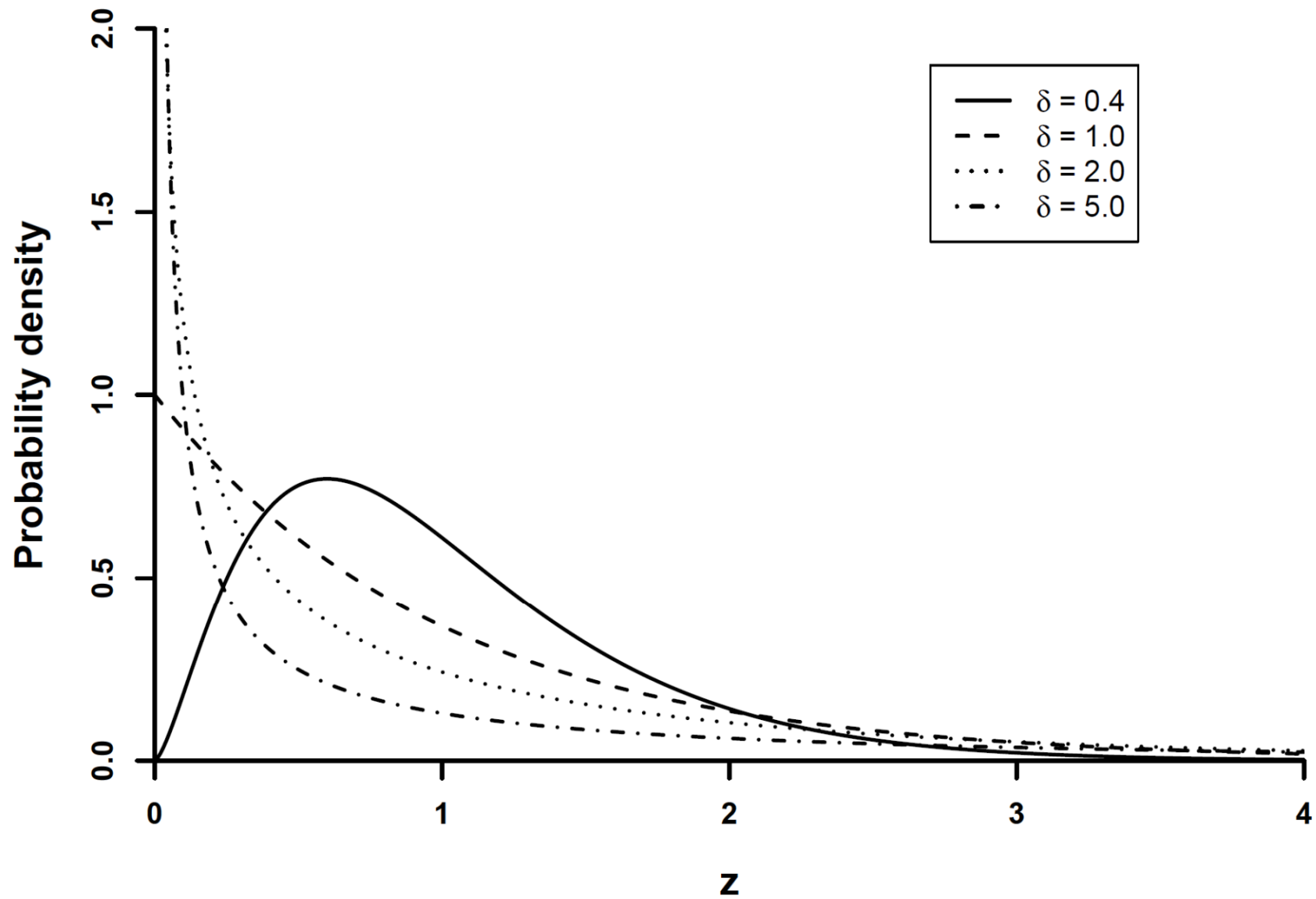
It is well known that the gamma distribution has mean η / ν and variance η / ν^2

We will often assume that the frailties have mean 1

For the gamma distribution this implies that $\eta = \nu$

Then the variance becomes $\delta = \nu^{-1}$

Gamma distributions with mean 1 for some values of the variance δ



Laplace transform

The Laplace transform is a convenient tool to study the proportional frailty model

For a positive random variable Z the Laplace transform is given by

$$\mathcal{L}(c) = \mathbb{E}(e^{-cZ})$$

The Laplace transform is closely related to the moment generating function

$$\mathcal{M}(s) = \mathbb{E}(e^{sZ})$$

It is well known that for the gamma distribution with mean η/ν and variance η/ν^2 , the moment generating function takes the form

$$\mathcal{M}(s) = \left(\frac{1}{1 - s/\nu} \right)^\eta$$

Thus the Laplace transform becomes

$$\mathcal{L}(c) = \mathcal{M}(-c) = \left(\frac{1}{1 + c/\nu} \right)^\eta$$

In particular for the gamma distribution with mean 1 (i.e. $\eta = \nu$) and variance $\delta = \nu^{-1}$ the Laplace transform takes the form

$$\mathcal{L}(c) = \left(\frac{1}{1 + c/\nu} \right)^\nu = (1 + \delta c)^{-1/\delta}$$

Population survival function

Consider a population where the heterogeneity is described by the proportional frailty model

Let T^* and Z be the survival time and the frailty of a randomly selected individual from the population

Then $S(t|Z) = P(T^* > t|Z) = \exp\{-A(t|Z)\} = \exp\{-Z A(t)\}$

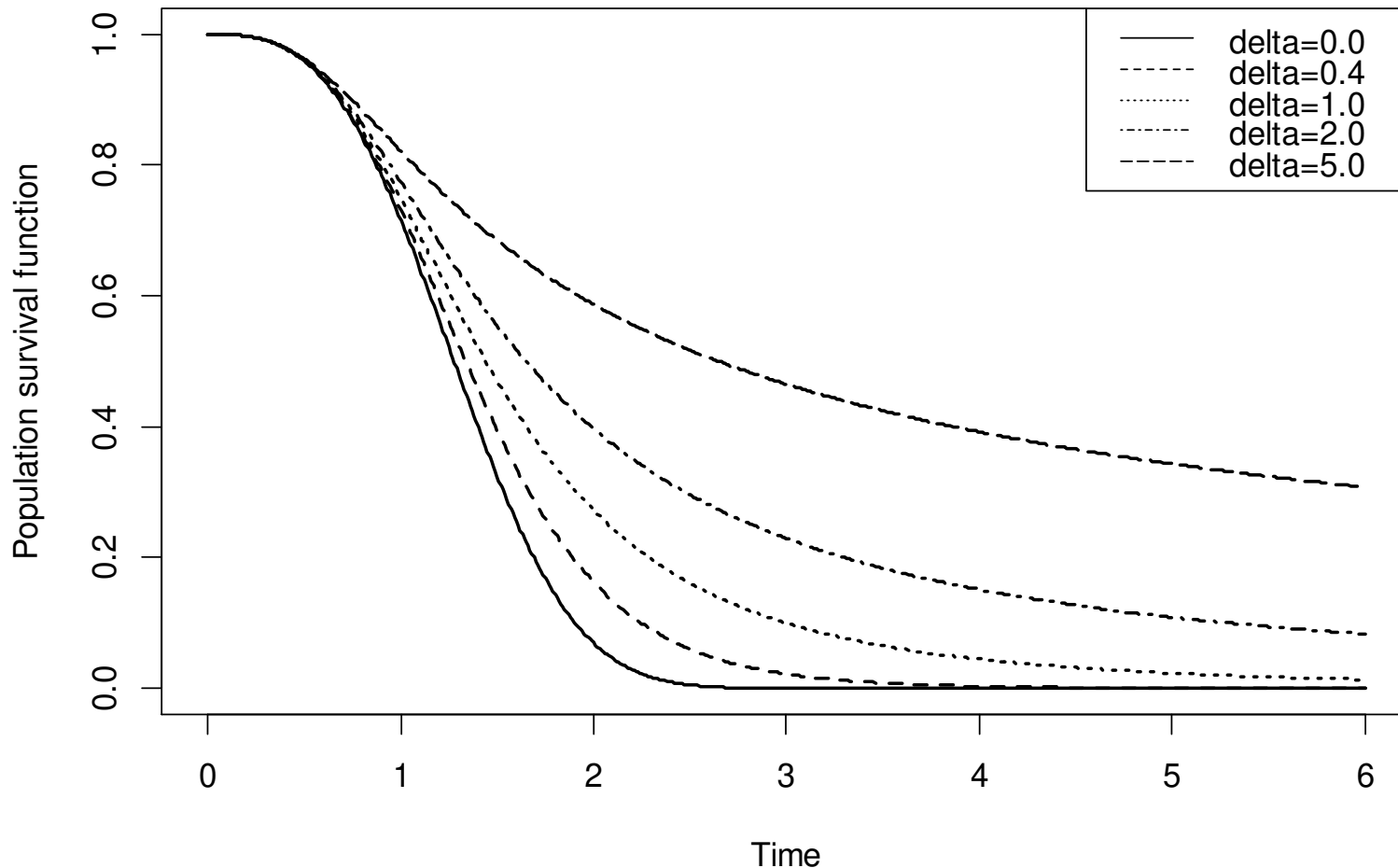
where $A(t) = \int_0^t \alpha(u) du$

The population survival function is given by

$$\begin{aligned} S(t) &= P(T^* > t) = E\{I(T^* > t)\} \\ &= E(E\{I(T^* > t)|Z\}) = E\{P(T^* > t|Z)\} \\ &= E(e^{-ZA(t)}) = \mathcal{L}(A(t)) \end{aligned}$$

If frailty is gamma distribution with mean 1 and variance δ the population survival function becomes

$$S(t) = \mathcal{L}(A(t)) = \{1 + \delta A(t)\}^{-1/\delta}$$



$$\alpha(t) = t^2$$

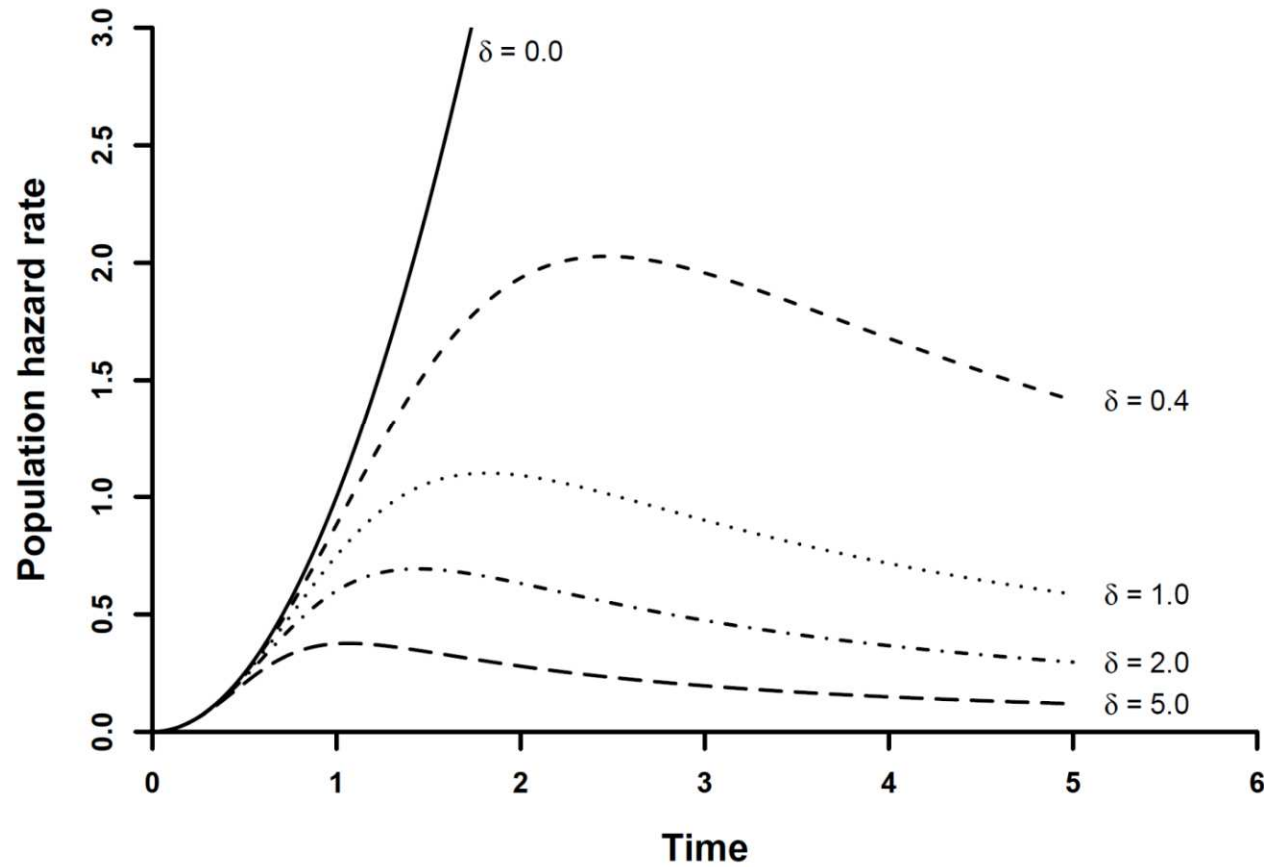
Population hazard

The population hazard becomes

$$\mu(t) = \frac{-S'(t)}{S(t)} = \frac{-\frac{d}{dt} \mathcal{L}(A(t))}{\mathcal{L}(A(t))} = \alpha(t) \frac{-\mathcal{L}'(A(t))}{\mathcal{L}(A(t))}$$

If frailty is gamma distribution with mean 1 and variance δ the population hazard rate is given by

$$\begin{aligned} \mu(t) &= \alpha(t) \frac{-\mathcal{L}'(A(t))}{\mathcal{L}(A(t))} = \alpha(t) \frac{(1 + \delta A(t))^{-\frac{1}{\delta} - 1}}{(1 + \delta A(t))^{-\frac{1}{\delta}}} \\ &= \frac{\alpha(t)}{1 + \delta A(t)} \end{aligned}$$



$$\alpha(t) = t^2$$

One sees that the population hazard is "pulled down" with a strength determined by δ

- For survival data, where each individual can experience at most one event, the frailty effect is not identifiable unless we assume a specific form of the individual baseline hazard rate $\alpha(t)$
- So frailty models for survival data may be speculative, but they offer one possible explanation why the population hazard may have different shapes
- As we will see, estimation of frailty is more relevant for **clustered survival data**
- But first we will take a look at maximum likelihood estimation for parametric survival models

Maximum likelihood estimation for parametric survival models

We start out with a recap of some basic relations

Consider an uncensored survival time T^*
(assumed to have an absolutely continuous distribution).

Survival function: $S(t) = P(T^* > t)$

Hazard rate: $\alpha(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} P(t \leq T^* < t + dt | T^* \geq t)$

We have the relations :

$$\alpha(t) = -\frac{S'(t)}{S(t)} \qquad S(t) = \exp\left\{-\int_0^t \alpha(s) ds\right\}$$

Density:

$$f(t) = -S'(t) = \alpha(t)S(t) = \alpha(t) \exp\left\{-\int_0^t \alpha(s) ds\right\}$$

Likelihood censored survival data

Consider independent survival times $T_1^*, T_2^*, \dots, T_n^*$

For T_i^* we have:

Hazard rate: $\alpha_i(t; \boldsymbol{\theta})$

Survival distribution: $S_i(t; \boldsymbol{\theta}) = \exp\left\{-\int_0^t \alpha_i(u; \boldsymbol{\theta}) du\right\}$

Density: $f_i(t; \boldsymbol{\theta}) = \alpha_i(t; \boldsymbol{\theta})S_i(t; \boldsymbol{\theta})$

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ is a q -vector of parameters

Let $c_1, c_2, \dots, c_n$ be **fixed** censoring times (type I censoring)

We observe ($i = 1, 2, \dots, n$)

$$\tilde{T}_i = \min(T_i^*, c_i) \quad \text{and} \quad D_i = I\{\tilde{T}_i = T_i^*\}$$

The likelihood becomes

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n f(\tilde{T}_i; \boldsymbol{\theta})^{D_i} \cdot S(\tilde{T}_i; \boldsymbol{\theta})^{1-D_i} \\ &= \prod_{i=1}^n \alpha(\tilde{T}_i; \boldsymbol{\theta})^{D_i} \cdot S(\tilde{T}_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \alpha(\tilde{T}_i; \boldsymbol{\theta})^{D_i} \cdot \exp\left\{-\int_0^{\tilde{T}_i} \alpha(t; \boldsymbol{\theta}) dt\right\} \end{aligned}$$

We will assume that we have **independent censoring**

Then it is more complicated to derive the likelihood

But one may show that the likelihood takes the same form as for type I censoring (strictly speaking one then obtain a *partial likelihood* since we do not include the randomness due to censoring)

By using results on martingales and stochastic integrals one may prove that the **MLE** $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_q)$ enjoys "the usual" large sample properties

Thus $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_q)$ is approximately multivariate normally distributed around true value of $\boldsymbol{\theta}$ with a covariance matrix that may be estimated by the inverse of the observed (or expected) information matrix

Also likelihood ratio, score and Wald tests apply as usual for censored survival data

Clustered survival data

We then turn to clustered survival data. First two examples

Example: Litter-matched rats

The data consist of 50 litters of female rats with 3 rats in each litter.

One rat in each litter received a potentially tumorigenic treatment, the other two were controls.

The time (in weeks) until tumor occurrence was observed for each rat.

Censoring was due to death without tumor or end of study (at 104 weeks when the rats still alive were sacrificed)

Litter no.	Drug-treated	Control 1	Control 2
01	101-D	49-T	104-D
03	104-D	102-D	104-D
05	104-D	104-D	104-D
07	77-D	97-D	79-D
09	89-D	104-D	104-D
11	88-T	96-T	104-D

Data for 6 of the 50 litters
(T is occurrence of tumor,
D is death or sacrifice,
i.e. censoring)

Example: Eye data

A study has been conducted to assess the effect of laser photocoagulation in delaying onset of severe visual loss ("blindness") in patients with diabetic retinopathy.

One eye of each of 197 patient was randomly selected for treatment and the other was control.

The patients were followed over several years for the occurrence of blindness in their left and right eyes.

Censoring is caused by death, dropout, or end of the study.

id	trteye	ageonset	typediab	status	time	trt
5	2	28	2	0	46.23	1
14	1	12	1	1	31.30	1
16	1	9	1	0	42.27	1
.....						
5	2	28	2	0	46.23	2
14	1	12	1	0	42.50	2
16	1	9	1	0	42.27	2

A shared frailty model for clustered survival data

Assume that we have data from m independent clusters, and that there are n_i units in the i -th cluster

The **uncensored** survival times of the units in the i -th cluster are denoted T_{ij}^* ($j = 1, \dots, n_i$)

$T_{i1}^*, \dots, T_{in_i}^*$ are **dependent** survival times

To handle the dependence, we will assume that the units in the i -th cluster share the same frailty Z_i

So **given** Z_i , the survival times $T_{i1}^*, \dots, T_{in_i}^*$ are independent and T_{ij}^* has hazard rate

$$\alpha_{ij}(t \mid Z_i) = Z_i \cdot \alpha_{ij}(t)$$

where $\alpha_{ij}(t)$ may depend on covariates (for the clusters and/or the units)

In the i -th cluster we observe censored survival times $\tilde{T}_{ij}$ and censoring indicators D_{ij} ($j = 1, \dots, n_i$)

So the data for the i -th cluster is given by (in addition there may be covariates)

$$H_i = \{\tilde{T}_{ij}, D_{ij} ; j = 1, \dots, n_i\}$$

The frailty variables $Z_1, \dots, Z_m$ for the m clusters are assumed to be independent and identically distributed (usually gamma distributed)

It is common to assume $E(Z_i) = 1$

We will derive the marginal likelihood for the data (i.e. the likelihood based on the distribution of the observable data)

Conditional on frailty Z_i , the likelihood contribution for the i -th cluster is given by

$$\begin{aligned} P(H_i | Z_i) &= \prod_{j=1}^{n_i} \left\{ \left(\alpha_{ij}(\tilde{T}_{ij} | Z_i) \right)^{D_{ij}} \exp \left(- \int_0^{\tilde{T}_{ij}} \alpha_{ij}(u | Z_i) du \right) \right\} \\ &= \prod_{j=1}^{n_i} \left\{ \left(Z_i \alpha_{ij}(\tilde{T}_{ij}) \right)^{D_{ij}} \exp \left(- Z_i A_{ij}(\tilde{T}_{ij}) \right) \right\} \end{aligned}$$

where $A_{ij}(t) = \int_0^t \alpha_{ij}(u) du$

Note that

$$P(H_i | Z_i) = \prod_{j=1}^{n_i} \left\{ \left(\alpha_{ij}(\tilde{T}_{ij}) \right)^{D_{ij}} \right\} \cdot Z_i^{D_{i\cdot}} \exp(-Z_i V_i)$$

where $D_{i\cdot} = \sum_{j=1}^{n_i} D_{ij}$ and $V_i = \sum_{j=1}^{n_i} A_{ij}(\tilde{T}_{ij})$

We obtain the contribution to the marginal likelihood from the i -th cluster by integrating over the frailty distribution:

$$P(H_i) = \prod_{j=1}^{n_i} \left\{ \left(\alpha_{ij}(\tilde{T}_{ij}) \right)^{D_{ij}} \right\} \cdot E_{Z_i} \left\{ Z_i^{D_{i\cdot}} \exp(-Z_i V_i) \right\}$$

Remember the Laplace transform: $\mathcal{L}(c) = E(e^{-cZ})$

Note that the r -th derivative becomes:

$$\mathcal{L}^{(r)}(c) = (-1)^r E(Z^r e^{-cZ})$$

Hence we may write

$$P(H_i) = \prod_{j=1}^{n_i} \left\{ \left(\alpha_{ij}(\tilde{T}_{ij}) \right)^{D_{ij}} \right\} \cdot (-1)^{D_{i\cdot}} \mathcal{L}^{(D_{i\cdot})}(V_i)$$

The total likelihood is given by

$$L = \prod_{i=1}^m P(H_i)$$

so the log-likelihood becomes

$$\log L = \sum_{i=1}^m \left[\sum_{j=1}^{n_i} D_{ij} \log \left(\alpha_{ij}(\tilde{T}_{ij}) \right) + \log \left\{ (-1)^{D_{i\cdot}} \mathcal{L}^{(D_{i\cdot})}(V_i) \right\} \right]$$

If $\alpha_{ij}(t)$ is given a parametric specification, the log-likelihood may be maximized with respect to the parameters of $\alpha_{ij}(t)$ (which may include parameters for the baseline hazard and regression coefficients) and the parameters of the frailty distribution

Standard results for MLE apply

The R library **parfm** may be used to fit frailty models with parametric baseline hazard

Default is a Weibull baseline parameterized as $\alpha(t) = \lambda \rho t^{\rho-1}$

When using a gamma frailty, it is assumed to have mean 1 and variance θ (corresponding to δ above)

Example: Litter-matched rats

Each litter is a cluster, and the rats are the units

We fit a gamma frailty model (where the frailties have mean 1 and variance θ) with treatment as covariate, assuming a Weibull baseline hazard:

$$\alpha_{ij}(t | Z_i) = \mathbf{Z}_i \cdot \boldsymbol{\alpha}_{ij}(t) = \mathbf{Z}_i \cdot \lambda \rho t^{\rho-1} \exp(\beta x_{ij})$$

Here $x_{ij} = 1$ if rat j in litter i is treated, $x_{ij} = 0$ otherwise

ML-estimates (with standard errors):

$$\hat{\rho} = 3.93(0.56) \quad \hat{\lambda} = 0.020(0.008)$$

$$\hat{\beta} = 0.91(0.32) \quad \hat{\theta} = 0.49(0.47)$$

Example: Eye data

Each patient is a cluster, and the eyes are the units

We fit a gamma frailty model with treatment and type of diabetes as covariates, assuming a Weibull baseline hazard:

$$\alpha_{ij}(t \mid Z_i) = \mathbf{Z}_i \cdot \boldsymbol{\alpha}_{ij}(t) = \mathbf{Z}_i \cdot \lambda \rho t^{\rho-1} \exp(\beta_1 x_{ij1} + \beta_2 x_{ij2})$$

$x_{ij1} = 1$ if eye j for patient i is treated, $x_{ij1} = 0$ otherwise

$x_{ij2} = 1$ if age at onset for patient i is at least 20 years,

$x_{ij2} = 0$ otherwise

ML-estimates (with standard errors):

$$\hat{\rho} = 0.95 \text{ (0.08)} \quad \hat{\lambda} = 0.028 \text{ (0.008)} \quad \hat{\theta} = 1.04 \text{ (0.33)}$$

$$\hat{\beta}_1 = -0.96 \text{ (0.18)} \quad \hat{\beta}_2 = 0.04 \text{ (0.23)}$$

R-commands that give results on the previous slides

We attach the survival and parfm libraries:

```
library(survival); library(parfm)
```

We first look at the rat data

The rat data (also for male rats) are available in the survival library

We extract the data for female rats

```
frats=rats[rats$sex=="f", ]
```

We fit a frailty model with gamma frailty (with mean 1 and variance theta) and Weibull baseline. Treatment (rx) is a covariate, and we use years as time scale:

```
fit.frats=parfm(Surv(time/52,status)~factor(rx),cluster="litter", frailty="gamma", data=frats)  
print(fit.frats)
```

We then look at the eye data

Read the data:

```
eyedata="https://www.mn.uio.no/math/english/people/aca/borgan/iwsm2026/retinopathy.txt"  
eye=read.table(eyedata, header=T)
```

Fit a frailty model with gamma and Weibull baseline.

Covariates: treatment (trt) and type diabetes (typediab):

```
fit.eye=parfm(Surv(time,status)~factor(trt)+factor(typediab), cluster="id",  
              frailty="gamma",data=eye)  
print(fit.eye)
```

Testing $H_0 : \delta = 0$ vs $H_A : \delta > 0$

We will use the likelihood ratio test, to test if there is an effect of frailty

This corresponds to testing if the frailty variance δ (or θ) is 0

The usual properties for the likelihood ratio test **do not** apply in this situation, since the null hypothesis is at the boundary of the parameter space

One may show that in such situations two times the difference in log-likelihoods is approximately distributed as

$$\frac{1}{2} \chi_0^2 + \frac{1}{2} \chi_1^2$$

To get the correct P-value, we should therefore simply halve the P-value we obtain from the usual likelihood ratio test

Example: Litter-matched rats

The log-likelihood for the model with frailty is -83.423

The model without frailty has log-likelihood -84.277

The likelihood ratio statistic takes the value

$$2*(-83.423 +84.227) =1.608$$

The P-value becomes

$$0.5*0.2048=0.102$$

Thus there is not a significant litter effect for the rat data

Example: Eye data

The likelihood ratio statistic takes the value

$$2*(-827.73 +836.30) =17.14$$

The P-value becomes $0.5*0.00035=0.000017$

There is a significant patient effect for the eye data

R-commands that give results on the previous slide

We first look at the rat data

Fit a frailty model with and without frailty and compute likelihood ratio statistic

with corrected P-value:

```
fit.frats=parfm(Surv(time/52,status)~factor(rx),cluster="litter",frailty="gamma",data=frats)
```

```
fit.frats0=parfm(Surv(time/52,status)~factor(rx), data=frats)
```

Compute likelihood ratio statistic

```
LR=2*(attributes(fit.frats)$loglik- attributes(fit.frats0)$loglik)
```

Compute corrected P-value:

```
(1-pchisq(LR,1))/2
```

We then do similar calculations for the eye data

```
fit.eye=parfm(Surv(time,status)~factor(trt)+factor(typediab), cluster="id",  
              frailty="gamma",data=eye)
```

```
fit.eye0=parfm(Surv(time,status)~factor(trt)+factor(typediab), data=eye)
```

```
LR=2*(attributes(fit.eye)$loglik- attributes(fit.eye0)$loglik)
```

```
(1-pchisq(LR,1))/2
```

A semi-parametric Cox frailty model

When the baseline hazard is parametric, we have a fully parametric model and the usual likelihood results apply

We may alternatively consider a Cox frailty model with a non-parametric baseline:

$$\alpha_{ij}(t \mid Z_i) = \mathbf{Z}_i \cdot \alpha_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij})$$

where $\mathbf{x}_{ij}$ is a vector of covariates (for clusters and/or units)

One then considers an extended model where the cumulative baseline $A_0(t)$ has jumps at the observed event times (only), and maximizes the log-likelihood

$$\log L = \sum_{i=1}^m \left[\sum_{j=1}^{n_i} D_{ij} \log \left(\Delta A_0(\tilde{T}_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij}) \right) + \log \left\{ (-1)^{D_{i\cdot}} \mathcal{L}^{(D_{i\cdot})}(V_i) \right\} \right]$$

The asymptotic theory becomes quite involved, but it has been worked out. We will not consider this any further