Heterogeneity in survival analysis

In Chapter 3 of ABG we assumed that individuals in a population *share the same survival distribution* (hazard etc.) Or, we considered two or more such populations and compared their hazard functions.

In Chapter 4 we introduced **survival regression**, where differences between individuals in a populations were modeled in terms of hazard functions that are functions of **observable** covariates.

There may, however, be other differences between individuals which we *do not measure* or which we *may not know* about:

- Such unobserved differences may be due to:
  - Environment
  - Life style
  - Genes
- *The unobserved differences are often disregarded when analyzing survival data.*
- **As explained in the "frailty" theory, we should take these unobserved differences into account in our analyses.**

A simple example

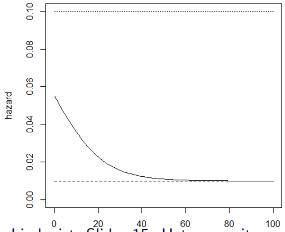
Suppose we have a batch of electronic components where

- Half have a lifetime which is exponential with expected value 100
- Half have a lifetime which is exponential with expected value 10

You draw randomly a number of components from the batch and estimate the hazard function for this sample. What do you see?

You are estimating the hazard function for a randomly drawn component:

$$\alpha(t) = \frac{f(t)}{S(t)} = \frac{(1/2)(1/100)\exp(-t/100) + (1/2)(1/10)\exp(-t/10)}{(1/2)\exp(-t/100) + (1/2)\exp(-t/10)}$$



The proportional frailty model

Instead of assuming that the population has just two kinds of individuals (*high risk* and *low risk*, e.g.), we now assume that the heterogeneity between individuals may be described by a **frailty variable** Z.

Z is a non-negative random variable. Each individual has its "own" Z, where large values of Z corresponding to "frail" individuals.

The assumption is that an individual with frailty Z has a hazard function

$$\alpha(t|Z) = Z \cdot \alpha(t),$$

where $\alpha(t)$ is a baseline hazard (corresponding to $Z = 1$).

It is commonly assumed that $E(Z) = 1$.

Note that the frailty Z is not observable.

Gamma distributed frailties

The most common choice of distribution for the frailty Z is the gamma distribution with density

$$f(z) = \frac{\nu^\eta}{\Gamma(\eta)} z^{\eta-1} e^{-\nu z}; \quad z > 0$$

The expected value is η/ν and the variance η/ν^2.

Assuming $E(Z) = 1$, we have $\eta = \nu$, and the variance then becomes

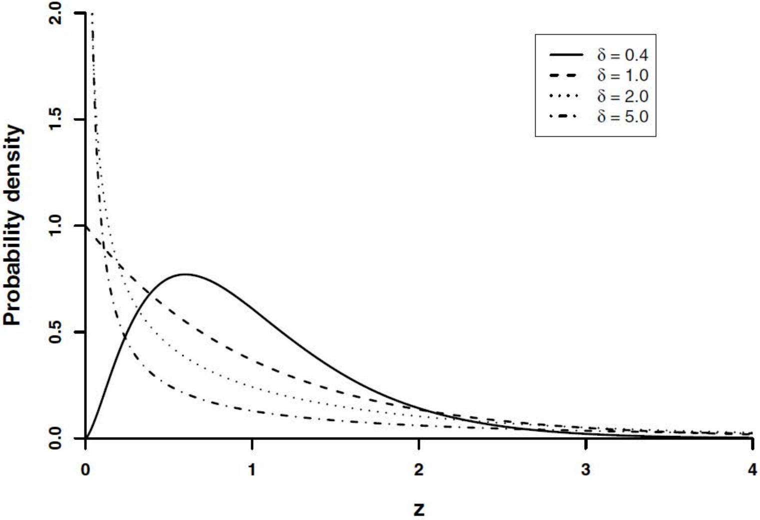
$$Var(Z) = \frac{\eta}{\nu^2} = \frac{1}{\nu} \equiv \delta$$

δ is hence a convenient parameter to be used for a gamma frailty distribution (and is then the only parameter). The density of Z becomes

$$f(z) = \frac{1}{\delta^{1/\delta}\Gamma(1/\delta)} z^{(1/\delta)-1} e^{-z/\delta}$$

Gamma densities with expectation 1 and variance δ

$$f(z) = \frac{1}{\delta^{1/\delta}\Gamma(1/\delta)} z^{(1/\delta)-1} e^{-z/\delta}$$



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

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

The Laplace transform is closely related to the moment generating function

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

Gamma distribution

For the gamma distribution with density

$$f(z) = \frac{\nu^\eta}{\Gamma(\eta)} z^{\eta-1} e^{-\nu z}; \quad z > 0$$

it is well known that the moment-generating function is

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

so 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 $\delta = 1/\nu$, the Laplace transform takes the form

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

Population survival function

Consider a population where the heterogeneity is described by the proportional frailty model, i.e., an individual with frailty Z has the hazard

$$\alpha(t|Z) = Z\alpha(t)$$

and hence the survival function

$$S(t|Z) = e^{-\int_0^t \alpha(u|Z)du} = e^{-Z\int_0^t \alpha(u)du} = e^{-ZA(t)}$$

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

Let now T be the survival time of a randomly selected individual from the population. Then

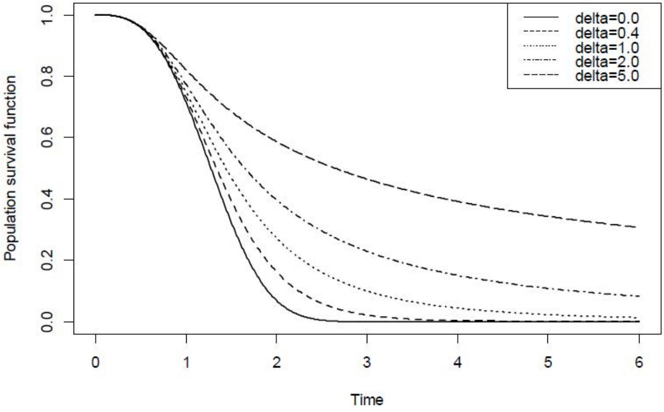
$$\begin{aligned} S(t) &= P(T > t) = E(P(T > t|Z)) \\ &= E(e^{-ZA(t)}) = \mathcal{L}(A(t)) \end{aligned}$$

Gamma distributed frailties

If the frailty Z is gamma distributed with mean 1 and variance δ, the population survival function becomes

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

Example: Let $\alpha(t) = t^2$, so $A(t) = (1/3)t^3$:



Population hazard

The population hazard becomes

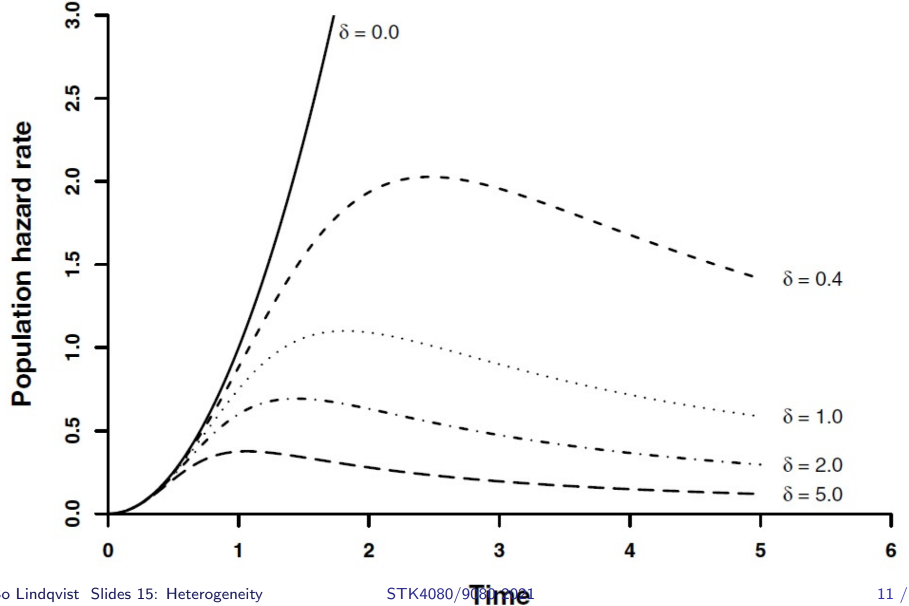
$$\mu(t) = \frac{-S'(t)}{S(t)} = \frac{-(d/dt)\mathcal{L}(A(t))}{\mathcal{L}(A(t))} = \alpha(t)\frac{-\mathcal{L}'(A(t))}{\mathcal{L}(A(t))}$$

If frailty Z is gamma distributed with mean 1 and variance δ, then $\mathcal{L}(c) = (1+\delta c)^{-1/\delta}$, so $\mathcal{L}'(c) = -(1+\delta c)^{(-1/\delta)-1}$. Hence we get

$$\mu(t) = \alpha(t)\frac{(1+\delta A(t))^{-\frac{1}{\delta}-1}}{(1+\delta A(t))^{-\frac{1}{\delta}}} = \frac{\alpha(t)}{1+\delta A(t)}$$

Note: When $\delta = 0$ there is no frailty and $\mu(t) = \alpha(t)$. As δ increases, the denominator becomes larger, and it also increases with time, yielding the typical frailty shape of a hazard function that is “dragged down”.

Population hazard with gamma distributed frailties with $\alpha(t) = t^2$



Estimating frailty

- For survival data where only a single event is available for each individual, the frailty effect is not identifiable unless we assume a specific form of the individual baseline hazard rate $\alpha(t)$.
  [If we take a sample from the population, we can estimate the population distribution only, since the Z are unobserved.]
- Frailty models for survival data may be speculative, but they are useful for understanding why the population hazard may have different shapes.
- Estimation of frailty is more relevant for **clustered survival data** and **recurrent event data** (repeated events). (See Slides 16).