

Frailty Models: Theory & Practice

Introduction

Theodor Balan & Hein Putter

Department of Biomedical Data Sciences
Leiden University Medical Center

Statistics in Practice session
IBC Barcelona, 9 July 2018



Outline

This course

Frailty models: Theory & Practice

Basics

Concepts

The Cox model

The model

Selection effects of the Cox model
(Violation of) proportional hazards

Unobserved heterogeneity

Sources of unobserved heterogeneity
Models for unobserved heterogeneity
Effects of unobserved heterogeneity

This course

Frailty models: Theory & Practice

- ▶ Introduction
- ▶ Univariate frailties
- ▶ Shared frailty models
- ▶ Applications and extensions

This course

A little bit on how we will work

- ▶ Emphasis on univariate frailty models and theory in the first half
- ▶ In second half on practice
 - ▶ Shared frailty
 - ▶ Software examples
- ▶ We like to keep things informal
 - ▶ Please interrupt, ask questions

Survival analysis

- ▶ Quote from Odd Aalen: “it takes time to observe time”
- ▶ Most important consequence
 - ▶ (Right) censoring: event has not (yet) occurred

Survival analysis

- ▶ Quote from Odd Aalen: “it takes time to observe time”
- ▶ Most important consequence
 - ▶ (Right) censoring: event has not (yet) occurred

Random variables

- ▶ Two random variables
 - ▶ \tilde{T} : Survival time
 - ▶ C : Censoring time
- ▶ Assumption: \tilde{T} and C are independent
 - ▶ Possibly conditional on explanatory variables

Distribution and survival functions

- ▶ Cumulative distribution functions

$$F(t) = P(\tilde{T} \leq t)$$

$$G(t) = P(C \leq t)$$

- ▶ $F(t)$ often called “(cumulative) incidence function”, “distribution function” or “failure function”
- ▶ More common to use the complimentary survival and censoring functions

$$S(t) = 1 - F(t) = P(\tilde{T} > t),$$

$$\overline{G}(t) = 1 - G(t) = P(C > t).$$

- ▶ Assumption: S and G continuous (only to make notation easier)

Observations

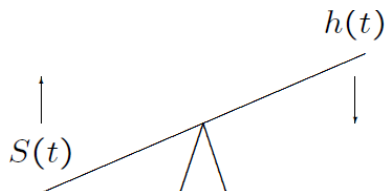
- ▶ For $i = 1, \dots, n$, we observe realizations (t_i, d_i, x_i) of
 - ▶ $T_i = \min(\tilde{T}_i, C_i)$
 - ▶ $D_i = \mathbf{1}\{\tilde{T}_i \leq C_i\}$
 - ▶ X_i : p -vector of covariates
- ▶ Interest is in
 - ▶ Estimating survival probabilities based on these data
 - ▶ Quantifying the effects of covariates on these probabilities

Hazard

- ▶ **Hazard function** $h(t)$; event probability at t , given event-free until t

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq \tilde{T} < t + dt \mid \tilde{T} \geq t)}{dt} \approx \frac{P(\tilde{T} = t)}{P(\tilde{T} \geq t)}$$

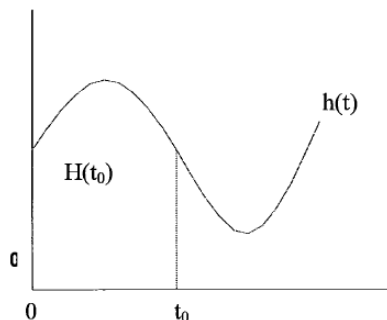
- ▶ Crucial to understand that the hazard function describes the randomness of the survival time



Cumulative hazard

- ▶ The cumulative hazard function is the area under the curve of the hazard function

$$H(t) = \int_0^t h(s) ds$$



Quantitative relation between survival and hazard function

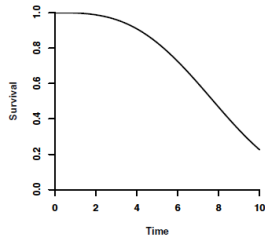
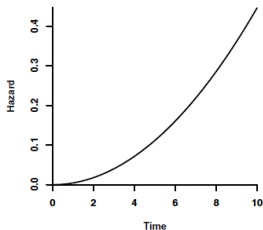
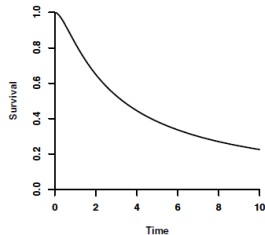
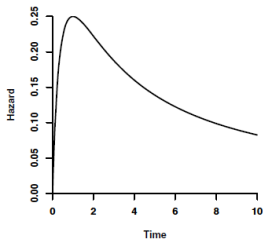
- ▶ The survival and hazard function carry the same information and they can be computed from each other

$$h(t) = -\frac{S'(t)}{S(t)}; H(t) = -\log(S(t))$$

- ▶ Inverse relation between survival and hazard

$$S(t) = \exp(-H(t)) = \exp\left(-\int_0^t h(s)ds\right)$$

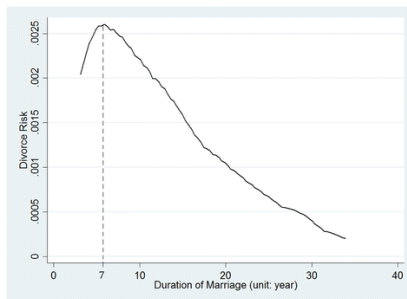
Illustration of the relation



The seven-year itch

Apart from the famous movie with Marilyn Monroe ...

- ▶ The seven-year itch refers to the peak of divorce rate after seven years of marriage



Density

- Formally given by

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq \tilde{T} < t + dt)}{dt}$$

- Informally by

$$f(t) \approx P(\tilde{T} = t)$$

- Remember that

$$h(t) \approx \frac{P(\tilde{T} = t)}{P(\tilde{T} \geq t)} = \frac{f(t)}{S(t)}$$

- In other words, we have the relation

$$f(t) = h(t) \cdot S(t)$$

Likelihood

Contribution to likelihood

- ▶ Subject i contributes
 - ▶ $S(t_i)$, if censored ($d_i = 0$)
 - ▶ $f(t_i)$, if event ($d_i = 1$)
- ▶ Since $f(t_i) = h(t_i)S(t_i)$, we can write this contribution in one formula as

$$h(t_i)^{d_i} S(t_i)$$

- ▶ Likelihood of sample of i independent contributions will be

$$L = \prod_{i=1}^n h(t_i)^{d_i} S(t_i)$$

- ▶ Becomes especially useful when a *model* for the hazard is employed (such as a **proportional hazards** model)

Proportional hazards

- ▶ The Cox model specifies that the hazard, given covariates x , is given by

$$h(t | x) = h_0(t) \exp(\beta^\top x)$$

- ▶ β is a p -vector of regression coefficients
- ▶ $\exp(\beta_j)$ is the **hazard ratio** corresponding to covariate x_j
- ▶ $h_0(t)$ is an unspecified **baseline hazard** function
- ▶ In a sense, Cox's proportional hazards model models **observed heterogeneity**
- ▶ Subjects may have a different hazard, depending on the specific values of their covariates
- ▶ Hazards of subjects with covariates x^* and \tilde{x} are different, except if $\beta^\top x^* = \beta^\top \tilde{x}$

Selection effects of the Cox model

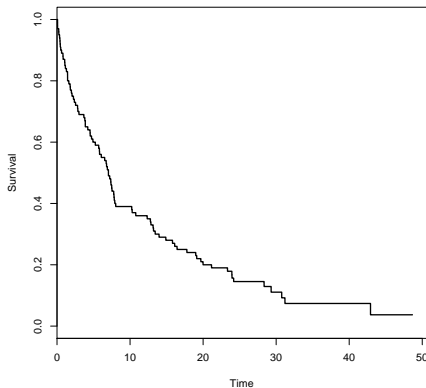
- ▶ The Cox model is well known
- ▶ We assume some familiarity with the Cox model
- ▶ Aim here is to highlight some of the effects of the Cox model on survivors
- ▶ For simplicity start with one covariate x
- ▶ Let us assume
 - ▶ It has a standard normal distribution (mean 0, variance 1)
 - ▶ $\beta > 0$
- ▶ Two questions for you:
 1. What are mean and variance of X at baseline?
 2. What are mean and variance of X among subjects at risk at $t = 1$?
 - ▶ Do you think it will be higher or lower?

Small simulation

- ▶ One replication, with sample size $n = 100$
- ▶ $X \sim N(0, 1)$, $\beta = 1$
- ▶ Baseline exponential with rate 0.1 (mean 10)
- ▶ Censoring uniform on $(20, 50)$

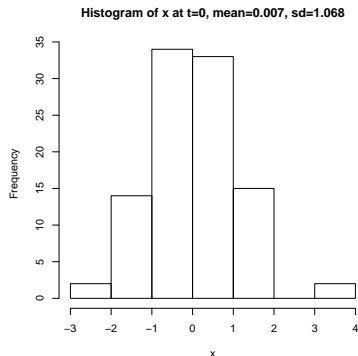
Results

- ▶ X had mean -0.007 and SD 1.068
- ▶ Estimated β : 0.943 (SE 0.127)
- ▶ Estimated overall survival curve

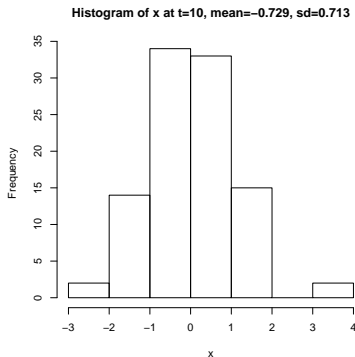
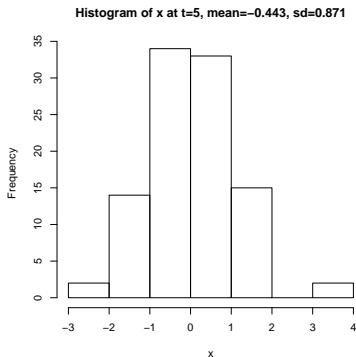


Distribution of X

At baseline



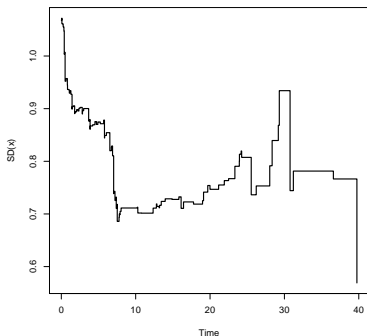
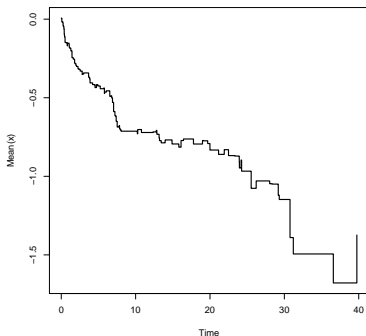
What about the distribution of X later on?



Selection effects of the Cox model

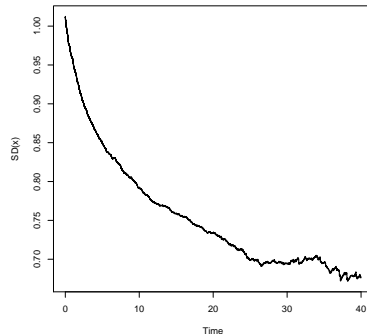
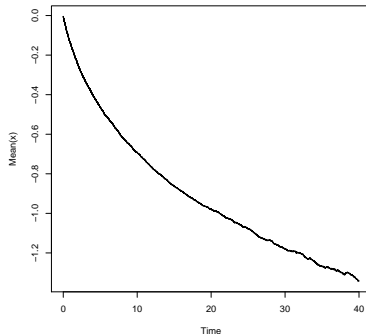
Mean and SD over time

- ▶ Mean at time 0, 5, 10: -0.007, -0.443, -0.729
- ▶ SD at time 0, 5, 10: 1.068, 0.871, 0.713
- ▶ Plots (over time):



Larger sample size

- Same thing, with $n = 10,000$



Take home message

- ▶ If some subjects have higher hazards than others
 - ▶ Those with higher hazards die earlier
 - ▶ The ones that remain tend to have lower hazards
 - ▶ And are more alike
 - ▶ With explained heterogeneity, this means that
 1. The means of x of those at risk decreases over time
 2. The variances of x of those at risk decreases over time

The proportional hazards assumption

For simplicity with one covariate x

- ▶ Cox's proportional hazards model specifies

$$h(t | x) = h_0(t) \exp(\beta x)$$

- ▶ The proportional hazards (PH) assumption specifies that the ratio $h(t | x^*)$ divided by $h(t | \tilde{x})$ equals $\exp(\beta(x^* - \tilde{x}))$ and *does not depend on time*
- ▶ The PH assumption is violated if

$$h(t | x) = h_0(t) \exp(\beta(t)x)$$

and $\beta(t)$ is not constant

Schoenfeld residuals

- ▶ Grambsch & Therneau developed and implemented a very useful way of checking the PH assumption
- ▶ Based on scaled *Schoenfeld residuals*
- ▶ Schoenfeld residuals are covariate specific
- ▶ A smoothed average over time is calculated
- ▶ A flat horizontal line means that the PH assumption is fulfilled
- ▶ The shape suggests how the effect changes through time
- ▶ Formal test associated with Schoenfeld residuals
- ▶ Theory beyond scope of this course, just show you some results when applied to simulated data

Omitting covariates causes non-PH

- ▶ The following slides are meant to show you what happens when you omit covariates

Simulation

- ▶ Single replication, with large sample size $n = 10,000$
- ▶ Two independent covariates X_1 and X_2 , both $\sim N(0, 1)$, $\beta_1 = \beta_2 = 1$
- ▶ Baseline exponential with rate 0.1 (mean 10)
- ▶ Censoring uniform on $(20, 50)$

Cox with two covariates

```
> c12 <- coxph(Surv(time, status) ~ x1 + x2, data=d)
> c12

## Call:
## 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
```

- ▶ Both estimated regression coefficients very close to 1

Results of cox.zph

```
> cz12 <- cox.zph(c12, transform="identity")
> cz12

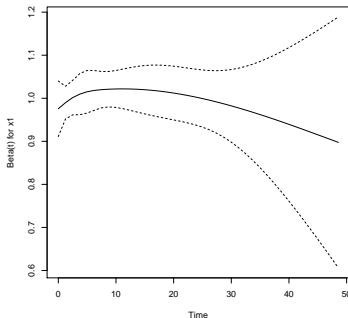
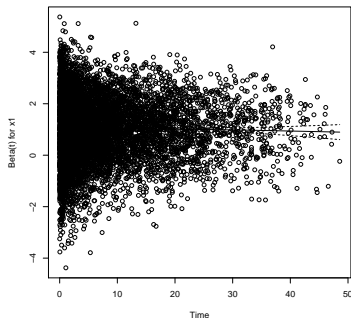
##              rho  chisq      p
## x1          0.00101 0.0081 0.928
## x2         -0.00357 0.1050 0.746
## GLOBAL              NA 0.1510 0.927

> plot(cz12)
```

- ▶ Test indicates no departures from PH
- ▶ Plots shown on next slide

Plots of cox.zph

- ▶ On the left with `resid=TRUE`, on the right with `resid=FALSE`
- ▶ Plots for x_2 very similar



Omitting one covariate

- Same dataset, now omitting x_2

```
> c1 <- coxph(Surv(time, status) ~ x1, data=d)
> c1

## Call:
## 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
```

- Note that the estimated regression coefficient is quite a bit smaller than 1!

Results of cox.zph

```
> cz1 <- cox.zph(c1, transform="identity")
> cz1

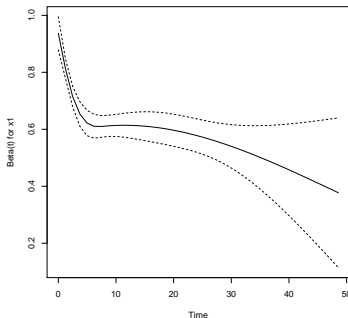
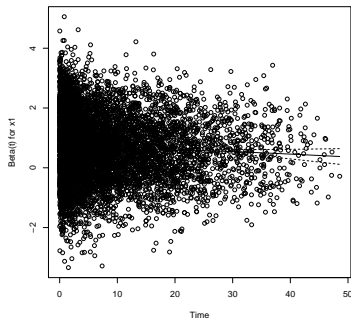
##           rho chisq           p
## x1 -0.0852  55.3 1.06e-13

> plot(cz1)
```

- ▶ Test indicates clear departure from PH
- ▶ Plots shown on next slide

Plots of cox.zph

- On the left with `resid=TRUE`, on the right with `resid=FALSE`



Take home message 2

- ▶ Suppose we have a vector of covariates x
- ▶ Suppose a PH model holds for x
- ▶ Suppose we can split x into $x = (x_{\text{incl}}, x_{\text{omit}})$ and β into $\beta = (\beta_{\text{incl}}, \beta_{\text{omit}})$
- ▶ So the correct model is

$$h(t | x) = h_0(t) \exp(\beta_{\text{incl}}^T x_{\text{incl}} + \beta_{\text{omit}}^T x_{\text{omit}})$$

- ▶ If we fit a PH model using only x_{incl}
- ▶ Then
 1. Estimate of β_{incl} will be biased (attenuated towards 0)
 2. The PH assumption for β_{incl} will be violated

Conditional versus marginal

- ▶ In the model

$$h(t | x) = h_0(t) \exp(\beta_{\text{incl}}^T x_{\text{incl}} + \beta_{\text{omit}}^T x_{\text{omit}})$$

β_{incl} quantifies the effect of x_{incl} , for given values of x_{omit}

- ▶ We say that β_{incl} is the **conditional** effect of x_{incl} , given x_{omit}
- ▶ And $h_0(t) \exp(\beta_{\text{incl}}^T x_{\text{incl}} + \beta_{\text{omit}}^T x_{\text{omit}})$ is the **conditional** hazard, given x_{omit}
- ▶ What we illustrated is the **marginal** effect of x_{incl}
- ▶ It describes the **marginal hazard** $h(t | x_{\text{incl}})$; we just saw that this marginal hazard was *not* given by $h_0(t) \exp(\beta_{\text{incl}}^T x_{\text{incl}})$
- ▶ The marginal hazard here is still conditional on x_{incl}
 - ▶ It describes the hazard given the included covariates
 - ▶ At time t , this is obtained by averaging over the distribution of x_{omit}
 - ▶ To be precise: the conditional distribution given survival until time t

Unobserved heterogeneity

- ▶ We just saw
 1. Selection effects of **observed heterogeneity**, through observed covariates
 2. Bias and violation of PH, when omitting important covariates
- ▶ In reality, we often
 - ▶ Cannot measure all covariates that we know or suspect are important (i.e., have an effect on survival)
 - ▶ Do not **know** all covariates that have an effect on survival
 - ▶ These covariates are omitted from a proportional hazards model
- ▶ This leads to **unobserved heterogeneity**

Unobserved heterogeneity

- ▶ Suppose we have covariates $x = (x_{\text{incl}}, x_{\text{omit}})$
- ▶ Only x_{incl} are included into the model, x_{omit} are omitted, because we haven't measured them, we don't know about them or just did not include them into the model
- ▶ If the correct model is

$$h(t | x) = h_0(t) \exp(\beta_{\text{incl}}^{\top} x_{\text{incl}} + \beta_{\text{omit}}^{\top} x_{\text{omit}})$$

- ▶ Then we can summarize all these unmeasured or unknown covariates and their effects into a latent, unobserved random variable $U = \beta_{\text{omit}}^{\top} x_{\text{omit}}$
- ▶ So we have

$$h(t | x_{\text{incl}}, U) = h_0(t) \exp(\beta_{\text{incl}}^{\top} x_{\text{incl}} + U) = Zh_0(t) \exp(\beta_{\text{incl}}^{\top} x_{\text{incl}})$$

- ▶ The random variable Z is called a **frailty** term
- ▶ It is assumed to act multiplicatively on the hazard

Randomness

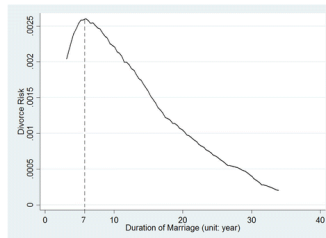
- ▶ Additional level of randomness
- ▶ Hazard itself describes randomness
- ▶ Frailties, random effects, Z , add another level of randomness
- ▶ Hard to distinguish one from the other (impossible without any assumptions)

Effects of unobserved heterogeneity

- ▶ If we think unobserved heterogeneity to be caused by unmeasured or unknown covariates
- ▶ That act multiplicatively on the hazards
- ▶ Then we may expect similar effects as we just observed in the Cox model
- ▶ Among survivors:
 - ▶ Means of decrease over time
 - ▶ Variances decrease over time
- ▶ Omitting unobserved heterogeneity leads to marginal effects of the included covariates being different from conditional effects
- ▶ These and other phenomena will come back in the second lecture
- ▶ Odd Aalen calls this “the odd effects of frailty”
 - ▶ Which is odd, since his first name is Odd

Conditional versus marginal revisited

- ▶ Important to understand distinction
- ▶ One is for individual
- ▶ Second is for population
- ▶ Take the seven-year itch: which hazard is important for *your* marriage?



- ▶ Similar issues play a role for covariate effects
- ▶ All to be further studied and explained in this course