STK4080/9080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 12: Cox regression

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.edu/employees/bo.lindqvist
bo.lindqvist@ntnu.no
boli@math.uio.no

University of Oslo, Spring 2021

Cox model and partial likelihood

The Cox model is given by the hazard specification

$$\alpha(t|\mathbf{x}) = \alpha_0(t) r(\boldsymbol{\beta}, \mathbf{x}(t)) = \alpha_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}(t)\}$$

Partial likelihood: Let *event* times be $T_1 < T_2 < \cdots$,

$$L(\boldsymbol{\beta}) = \prod_j \frac{r(\boldsymbol{\beta}, \mathbf{x}_{i_j}(T_j))}{\sum_{\ell \in \mathcal{R}_j} r(\boldsymbol{\beta}, \mathbf{x}_\ell(T_j))} = \prod_j \frac{\exp\{\boldsymbol{\beta}^T \mathbf{x}_{i_j}(T_j)\}}{\sum_{\ell \in \mathcal{R}_j} \exp\{\boldsymbol{\beta}^T \mathbf{x}_\ell(T_j)\}}$$

Here i_j is the index of the individual who experiences the event at T_j, while $\mathcal{R}_j = \{\ell \mid Y_\ell(T_j) = 1\}$ is the *risk set* at T_j.

The maximum partial likelihood estimator $\hat{\boldsymbol{\beta}}$ is the maximizer of $L(\boldsymbol{\beta})$, or the solution of the equation $U(\boldsymbol{\beta}) = 0$, where

$$U(\boldsymbol{\beta}) = \frac{\partial \log L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$$

Cox regression with left truncation and right censoring

Recall from earlier that *left truncation* can be included in the models by letting the 'at risk' function be

$$Y_i(t) = I(L_i < t \leq \tilde{T}_i)$$

where L_i is the time of entry of the ith individual.

Try the following example in R:

```
library(survival)
library(KMsurv)
data(psych)
attach(psych)
psych
my.surv.object <- Surv(age, age+time, death)
my.surv.object
fit.left = coxph(Surv(age, age+time, death)~ sex)
summary(fit.left)
detach(psych)
```

Here $L_i = \text{age}$, $\tilde{T}_i = \text{age} + \text{time}$, $D_i = \text{death}$

R-output

```
> data(psych)
> attach(psych)
> my.surv.object <- Surv(age, age+time, death)
> my.surv.object
 [1] (51,52]  (58,59]  (55,57]  (28,50]  (21,51+] (19,47]  (25,57]  (48,59]
 [9] (47,61]  (25,61+] (31,62+] (24,57+] (25,58+] (30,67+] (33,68+] (36,61]
[17] (30,61+] (41,63]  (43,69]  (45,69]  (35,70+] (29,63+] (35,65+] (32,67]
[25] (36,76]  (32,71+]
> |
```

```
> fit.left = coxph(Surv(age, age+time, death)~sex)
> summary(fit.left)
Call:
coxph(formula = Surv(age, age + time, death) ~ sex)

  n= 26, number of events= 14

      coef exp(coef) se(coef)     z Pr(>|z|)
sex 0.3900    1.4770   0.6102 0.639    0.523

    exp(coef) exp(-coef) lower .95 upper .95
sex     1.477      0.677    0.4466     4.884

Concordance= 0.58  (se = 0.082 )
Rsquare= 0.016   (max possible= 0.926 )
Likelihood ratio test= 0.43  on 1 df,   p=0.5141
Wald test            = 0.41  on 1 df,   p=0.5227
Score (logrank) test = 0.41  on 1 df,   p=0.5203
```

Time dependent covariates in R

R only allows for *time dependent covariates that are constant on intervals*, i.e. step functions.

Suppose for simplicity that $p = 1$, so there is a single covariate.

Assume for individual i that $x_i(t) = x_\ell$ on the interval $(L_{i\ell}, U_{i\ell}]$ for $\ell = 1, 2, \ldots, J_i$.

One then represents this individual J_i times in the data file as left truncated data with

- $L_{i\ell}$ as left truncation time
- $U_{i\ell}$ as right censoring time
- $D_{i\ell} = D_i \cdot I(\text{event for individual } i \text{ in interval } (L_{i\ell}, U_{i\ell}])$
- x_ℓ as covariate value

For an example, see ASAUR Chapter 8.1:
Stanford Heart Transplant Data

Stratified Cox-regression

Assume that the individuals are divided into k strata, so that for an individual in stratum s with covariate $\mathbf{x}_i(t)$ we have the hazard

$$\alpha(t|\mathbf{x}_i, \text{stratum } s) = \alpha_{s0}(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}$$

Note that the effects of the covariates are here assumed to be the same accross strata, while the baseline hazard may vary between strata.

We now estimate $\boldsymbol{\beta}$ by maximizing the partial likelihood

$$\prod_{s=1}^{k} \prod_{T_{sj}} \frac{\exp\{\boldsymbol{\beta}^T \mathbf{x}_{i_j}(T_{sj})\}}{\sum_{\ell \in \mathcal{R}_{sj}} \exp\{\boldsymbol{\beta}^T \mathbf{x}_\ell(T_{sj})\}}$$

where $T_{s1} < T_{s2} < \cdots$ are the observed event times in stratum s and $\mathcal{R}_{sj}$ is the risk set in this stratum at time T_{sj}.

The maximum partial likelihood estimator has similar properties as for the situation without stratification and statistical tests may be performed as before.

Why stratified Cox-regression?

Recall that

$$\alpha(t|\mathbf{x}_i, \text{stratum } s) = \alpha_{s0}(t) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(t)\}$$

- The stratified Cox model is useful when the proportional model does not hold for a categorical variable.
- Stratify on this variable and keep the regression model for other covariates

Stratified Cox regression: Melanoma data

Consider the melanoma data where we stratify on the variable 'grouped tumor thickness' (grthick).

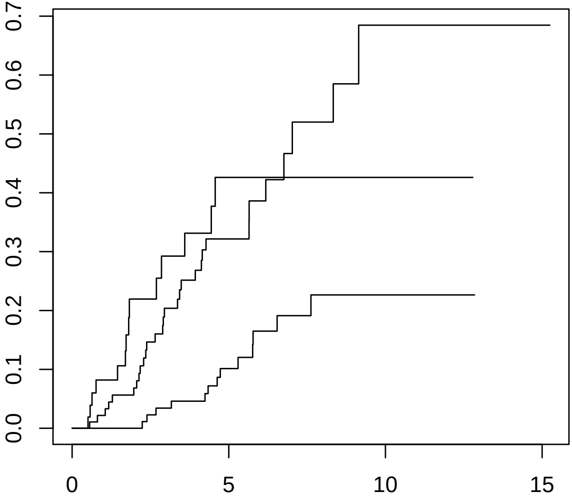
```
path="http://www.uio.no/studier/emner/matnat/math/STK4080/h14/
melanoma.txt"
melanoma=read.table(path,header=T)
# Use 'grthick' as a stratum variable:
coxph(Surv(lifetime,status==1)~ulcer+sex+age+strata(grthick),
data=melanoma)
#
# Use 'grthick' as a factor variable:
coxph(Surv(lifetime,status==1)~
ulcer+sex+age+factor(grthick),data=melanoma)
# Use 'grthick' as a stratum variable and plot the three baseline
hazards:
cox.strat =
coxph(Surv(lifetime,status==1)~ulcer+sex+age+strata(grthick),
data=melanoma)
plot(survfit(cox.strat),fun="cumhaz")
```

R-output

```
> coxph(Surv(lifetime,status==1)~ulcer+sex+age+strata(grthick),data=melanoma)
Call:
coxph(formula = Surv(lifetime, status == 1) ~ ulcer + sex + age +
    strata(grthick), data = melanoma)

          coef exp(coef) se(coef)     z      p
ulcer -0.94796   0.38753  0.32572 -2.91 0.0036
sex    0.40740   1.50291  0.27351  1.49 0.1363
age    0.00630   1.00632  0.00837  0.75 0.4517

Likelihood ratio test=13.2  on 3 df, p=0.00426
n= 205, number of events= 57
> coxph(Surv(lifetime,status==1)~ulcer+sex+age+factor(grthick),data=melanoma)
Call:
coxph(formula = Surv(lifetime, status == 1) ~ ulcer + sex + age +
    factor(grthick), data = melanoma)

                     coef exp(coef) se(coef)     z      p
ulcer            -0.95621   0.38435  0.32407 -2.95 0.0032
sex               0.34157   1.40716  0.27127  1.26 0.2080
age               0.01028   1.01033  0.00845  1.22 0.2240
factor(grthick)2  1.04401   2.84058  0.36538  2.86 0.0043
factor(grthick)3  1.12071   3.06704  0.41641  2.69 0.0071

Likelihood ratio test=45.3  on 5 df, p=1.27e-08
n= 205, number of events= 57
```

The results are only marginally different (also standard errors).

The estimated baseline hazard curves



Stratification is most efficient when these baseline hazards are not proportional.

How can the Cox-model fail?

$$\text{Recall model:} \quad \alpha(t|\mathbf{x}) = \alpha_0(t)\exp\{\boldsymbol{\beta}^T\mathbf{x}(t)\}$$

The Cox-model is flexible w.r.t the baseline $\alpha_0(t)$, but otherwise strict with respect to how the hazard depends on covariates:

- We may have specified a covariate x in a wrong way, where the correct alternative may be, e.g., $\log x$, $x^{1/2}$, etc.
- We may not have a proportional model, so that the effect of a covariate may vary with time, e.g.,

$$\alpha(t|\mathbf{x}) = \alpha_0(t)\exp(\boldsymbol{\beta}(t)^T\mathbf{x}(t))$$

where $\boldsymbol{\beta}(t)$ depends on t.

Model diagnostic tools

- **Martingale residuals.**
  - See ABG Section 4.1.3 for definition. (But we will not cover the treatment after equation (4.21) on p. 144.)
  - See instead ASAUR Section 7.1.1.

- **Checking the proportional hazards assumption.**
  - Log cumulative hazard plots. ASAUR 7.2.1
  - Schoenfeld residuals. ASAUR 7.2.2

Martingale residuals

- Martingale residuals are in some sense similar to the residuals we use in linear regression.
- They are, however, not as useful as the linear regression residuals, because there is no natural distribution to compare them to.
- One of their main applications is to estimate appropriate modifications to the proportional hazards model by way of *covariate transformation*

Martingale residuals: definition

We start with the *counting process martingale* for Cox regression:

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \exp\{\boldsymbol{\beta}^T \mathbf{x}_i(s)\} dA_0(s)$$

which can be interpreted as *"observed minus expected"* for the ith individual.

It becomes the **martingale residual** by plugging in estimators and considering the maximum time τ:

$$\hat{M}_i = N_i(\tau) - \int_0^\tau Y_i(s) \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i(s)\} d\hat{A}_0(s)$$

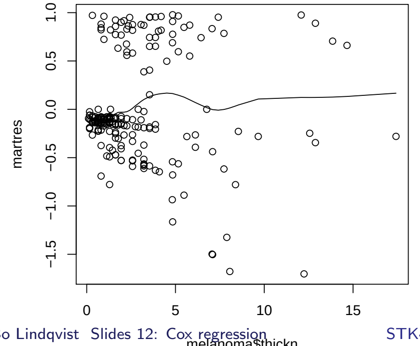
For time-constant covariates we have

$$\hat{M}_i = D_i - \exp\{\hat{\boldsymbol{\beta}}^T \mathbf{x}_i\} \hat{A}_0(\tilde{T}_i)$$

Example: Martingale residuals for the melanoma data

It has been considered that log(`thickn`) is a "better" covariate than `thickn` itself. We will check if this can be discovered from martingale residuals.

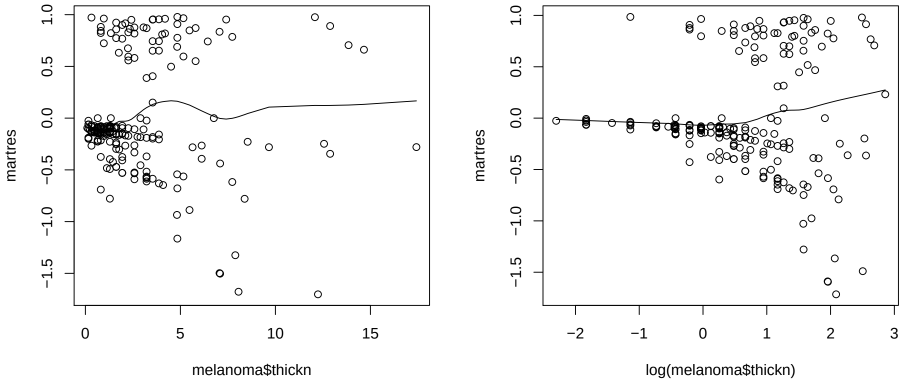
```
# Martingale residual plot against 'thickn':
coxfit<-coxph(Surv(lifetime,status==1)~sex+ulcer+thickn,
data=melanoma)
martres = coxfit$residuals
plot(melanoma$thickn,martres)
lines(lowess(melanoma$thickn,martres))
```



Martingale residuals for melanoma data

To the left is the plot from the previous page using the covariate `thickn`. Tp the right is instead used `log(thickn)`

Note the lowess smooth of the martingale residuals which has been added to the plots (see R-code previous page).



Using martingale residuals for estimating covariate transforms

Consider one component, z, of a covariate vector $(\mathbf{x}, z)$. The question is whether, instead of a *hazard ratio* of $e^{\beta z}$, it might be better to use $e^{f(z)}$ for some suitable function $f(z)$, e.g., $\beta \log(z)$, $\beta\sqrt{z}$, etc.

Consider then the model

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp\{\beta^T \mathbf{x} \ + f(z)\}$$

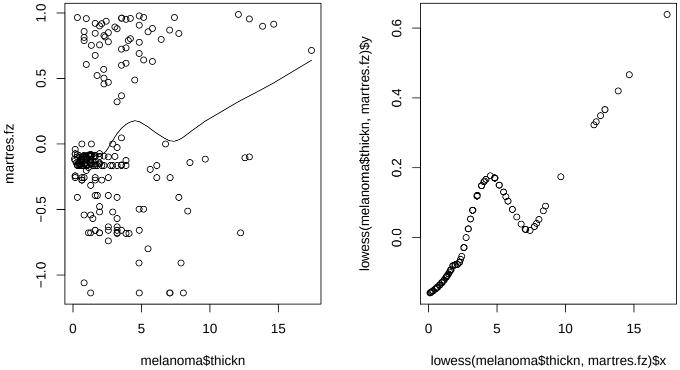
Martingale residuals from the estimated model

$$\alpha(t|\mathbf{x}) = \alpha_0(t) \exp\{\beta^T \mathbf{x}\}$$

i.e., **without including the covariate** *z can then be used to infer the form of $f(z)$.* (see ASAUR 7.1.1)

Should one use log(thickn) instead of thickn?

Below are martingale residual plots for the model **without** $z(=$ thickn$)$. *(R-code is given on the next page.)* **The smoothed plot estimates the underlying** $f(z)$ up to a linear transformation.



Note that a linear plot would correspond to using z itself. The flattening/decreasing tendency around 5 mm, and the fact that few observations have thick > 10 might favor of a log-transform.

R-code for estimating $f(z)$ (previous page)

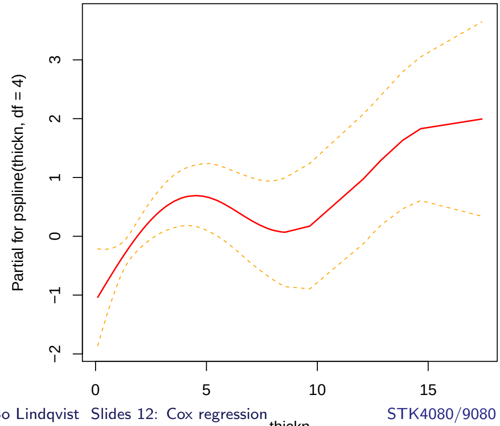
```
# To estimate f(z):
coxfit.fz = coxph(Surv(lifetime,status==1)~sex+ulcer,
data=melanoma)
martres.fz = coxfit.fz$residuals
plot(lowess(melanoma$thickn,martres.fz))
plot(melanoma$thickn,martres.fz)
lines(lowess(melanoma$thickn,martres.fz))
```

See also Section 7.1.1 in ASAUR for an interesting example.

Including smooth estimates of continuous covariates

See Section 6.5 in ASAUR, where splines are used to estimate $f(z)$. This method is based on maximizing a penalized log-partial likelihood, i.e. $\log(\text{partial likelihood}) - \lambda \int (f''(z))^2 dz$.

```
coxfit.spl = coxph(Surv(lifetime,status==1)~sex+ulcer +
pspline(thickn,df=4),data=melanoma)
termplot(coxfit.spl,se=T,terms=3)
```



Checking for non-proportional hazards: *Log cumulative hazard plots*

The classical way of checking *departure from proportionality* is based on the following:

With fixed covariates we have the cumulative hazard

$$A(t|\mathbf{x}) = A_0(t)\exp\{\boldsymbol{\beta}^T\mathbf{x}\}$$

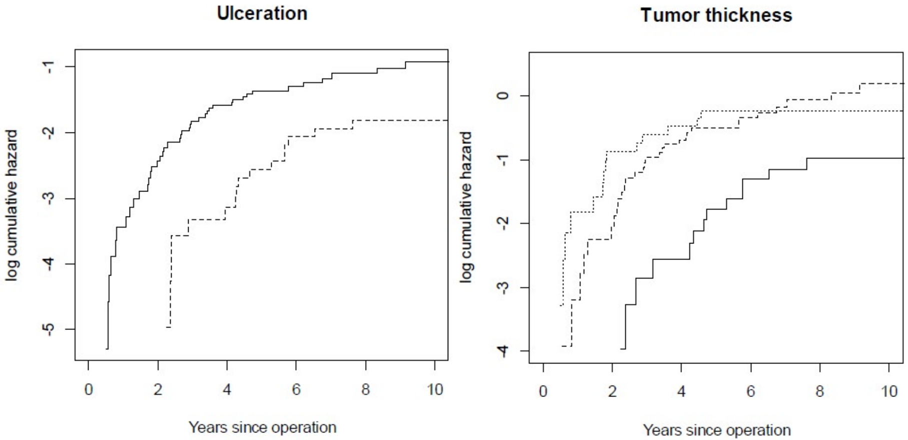
Thus

$$\log(A(t|\mathbf{x})) = \boldsymbol{\beta}^T\mathbf{x} + \log(A_0(t))$$

i.e. $t \mapsto \log(A(t|\mathbf{x}))$, for different choices of $\mathbf{x}$, are **parallel** curves.

- Thus if x is a single *categorical* covariate, we may plot log of Nelson-Aalen estimates for $A(t|x)$ for every level of x.
- Approximately parallel curves then support the use of a proportional hazards model.
- For non-categorical covariates we may group the values of the covariate into a finite number of categories.

Example: tumor-thickness and ulceration in melanoma data



Plot of $\log(\hat{A}(t|\mathbf{x}))$ against t for `ulcer` $= 1,2$ and `grthick` $= 1,2,3$.
(Parallelism seems to be OK?)

Check of proportionality for multivariate covariate vectors

If we have included the covariates $x_1, x_2, \ldots, x_p$ in our model, and want to check if the *categorical* covariate x_{p+1} satisfies the proportionality requirement, we may

- Fit a **stratified** Cox-model with the levels s of x_{p+1} as strata and $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ as covariates.
- Plot $\log(\hat{A}_s(t|\mathbf{x}))$ against t for different levels s of x_{p+1}.
- Check if lines are (approximately) parallel.

Checking for non-proportional hazards: *Schoenfeld residuals*

Consider the Cox-model with time-fixed covariates:

$$\alpha(t|\mathbf{x}) = \alpha_0(t)\exp\{\boldsymbol{\beta}^T\mathbf{x}\}$$

As we have seen, the effect of increasing, say, covariate number 1 by one unit, is to multiply the hazard rate by e^{β_1}, independently of time t.

In practice one might imagine, however, that β_1 could depend on t as a function $\beta_1(t)$.

The **Schoenfeld residual** compares, for each event time T_j, the values of the covariates of the unit that fails, with what would be expected if the Cox-model with constant $\boldsymbol{\beta}$ is correct.

Schoenfeld residuals

"...compares, for each event time T_j, the values of the covariates of the unit that fails, with what would be expected if the Cox-model with constant β is correct."

Let the covariate vector for unit i be $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$.

For each failure time T_j, with individual i_j failing, and with risk set $\mathcal{R}_j$, we compute for each coordinate $k = 1, \ldots, p$,

$$\begin{aligned}
s_{jk} &= x_{i_j k} - \sum_{\ell \in \mathcal{R}_j} x_{\ell k} \hat{P}(\text{unit } \ell \text{ fails at } T_j) \\
&= x_{i_j k} - \sum_{\ell \in R_j} x_{\ell k} \frac{\exp\{\hat{\beta}^T \mathbf{x}_\ell\}}{\sum_{v \in \mathcal{R}_j} \exp\{\hat{\beta}^T \mathbf{x}_v\}} \\
&= x_{i_j} - \frac{\sum_{\ell \in R_j} x_{\ell k} \exp\{\hat{\beta}^T \mathbf{x}_\ell\}}{\sum_{\ell \in \mathcal{R}_j} \exp\{\hat{\beta}^T \mathbf{x}_\ell\}} \equiv x_{i_j k} - \bar{x}_k(T_j)
\end{aligned}$$

If the model is correct, then the s_{jk} are supposed to vary around 0.

Schoenfeld residuals with R: the melanoma data

```
fit.logtu=coxph(Surv(lifetime,status==1)~sex+age+factor(ulcer)+
log(thickn), data=melanoma)
resid.schoen.melanoma = residuals(fit.logtu,type="schoenfeld")
head(resid.schoen.melanoma)
```

	sex	*age*	*factor*(*ulcer*)2	*log*(*thickn*)
0.5068493	0.4339136	−6.327544	−0.2035499	1.1225127
0.5589041	0.4421115	−30.447089	−0.2073955	0.2290838
0.5753425	0.4459645	18.287558	−0.2092030	0.2951022
0.6356164	0.4531361	−9.418362	−0.2125672	1.2145868
0.7643836	−0.5418191	9.405579	−0.2183473	0.6846807
0.8082192	−0.5487693	−5.473772	−0.2211481	0.1233337

The first column is here the failure times T_j. The original idea of Schoenfeld (Biometrika 1982) was to plot the s_{jk} versus T_j. Departures from the proportional hazards model would then be revealed by trends away from 0 in the plot.

Scaling the Schoenfeld residuals

Grambsch and Therneau (Biometrika, 1994) suggested to scale the Schoenfeld residuals at each T_j by the inverse of the covariance matrix corresponding to the distribution of $\mathbf{x}$ in the risk set $\mathcal{R}_j$, thereby obtaining scaled residuals s_{jk}^*.

Then, assuming that the true model is of the form

$$\alpha(t|\mathbf{x}) = \alpha_0(t)\exp\{\boldsymbol{\beta}(t)^T\mathbf{x}\},$$

they showed that $\boldsymbol{\beta}(t)$ can be estimated at each T_j by the components

$$\hat{\beta}_k(T_j) = s_{jk}^* + \hat{\beta}_k; \quad k = 1, 2, \ldots, p$$

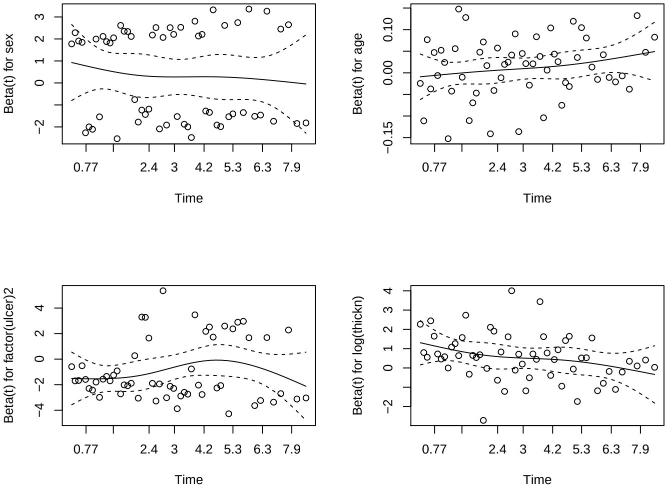
where $\hat{\beta}_k$ is the estimate from an ordinary Cox-regression.

Plots of $\hat{\beta}_k(T_j)$ versus T_j will then show the form of $\boldsymbol{\beta}[t]$ and will indicate possible deviations from the proportional hazards model.

Note: *The equation for $E(r_i^*)$ on p. 98 in ASAUR should in accordance with the above be corrected to $E(r_i^*) + beta \approx \beta(t)$.*

Scaled and smoothed Schoenfeld residuals for estimation of $\beta(t)$ for the melanoma data

```
par(mfrow=c(2,2))
plot(cox.zph(fit.logtu))
```



Compare to $\hat{\beta} = (0.36, 0.01, -0.94, 0.55)$

Proportionality testing by R

The formal test for proportionality of the covariates (Therneau and Grambsch, 1994) is based on defining

$$\beta_k(t) = \beta_k + \theta_k g(t)$$

for a given function $g(t)$, e.g. $\log(t)$, and testing (by score tests) the hypotheses $\theta_k = 0$ for each covariate x_k.

Here is R-code for the melanoma data:

```
fit.logtu=coxph(Surv(lifetime,status==1)~sex+age+factor(ulcer)+
log(thickn), data=melanoma)
cox.zph(fit.logtu,transform="log")
```

The output is on the next page. Note that the function cox.zph is the same as was used in the plotting of scaled Schoenfeld residuals.

Proportionality testing by R

```
cox.zph(fit.logtu,transform="log")
```

	chisq	*df*	*p*
sex	1.13	1	0.2887
age	1.71	1	0.1904
factor(ulcer)	4.17	1	0.0410
log(thickn)	6.76	1	0.0093
GLOBAL	10.86	4	0.0282

There is hence an indication for coefficients for ulcer and (log)thickness to depend on time. Compare to the previous plots!

Some strategies when proportional hazard fails

- Stratified Cox-regression
- Separate analyses on disjoint time intervals
- Time-dependent covariates
- Alternative regression models
  - Accelerated failure time models
  - Additive models

Large sample distribution of the maximum partial likelihood estimator

For simplicity, we restrict attention to Cox regression with a single covariate ($p = 1$).:

$$\alpha(t|x_i) = \alpha_0(t) \exp\{\beta x_i(t)\}$$

$\hat{\beta}$ is the maximizer of the partial likelihood

$$L(\beta) = \prod_j \frac{\exp\{\beta x_{i_j}(T_j)\}}{\sum_{\ell \in \mathcal{R}_j} \exp\{\beta x_\ell(T_j)\}} = \prod_j \frac{\exp\{\beta x_{i_j}(T_j)\}}{\sum_{\ell=1}^{n} Y_\ell(T_j) \exp\{\beta x_\ell(T_j)\}}$$

We will show (only main steps) that $\hat{\beta}$ is approximately normally distributed around the true value β_0 of β with a variance that can be estimated by the inverse information.

The log partial likelihood

The logarithm of the partial likelihood can be written

$$
\begin{aligned}
\ell(\beta) &= \log L(\beta) \\
&= \sum_j \left\{ \beta x_{i_j}(T_j) - \log \left(\sum_{\ell=1}^{n} Y_\ell(T_j) \exp\{\beta x_\ell(T_j)\} \right) \right\} \\
&= \sum_{i=1}^{n} \int_0^\tau \left\{ \beta x_i(u) - \log S^{(0)}(\beta, u) \right\} dN_i(u)
\end{aligned}
$$

where

$$
S^{(0)}(\beta, u) = \sum_{i=1}^{n} Y_i(u) \exp\{\beta x_i(u)\}
$$

Score function

Recall log likelihood:

$$\ell(\beta) = \sum_{i=1}^{n} \int_0^\tau \left\{ \beta x_i(u) - \log S^{(0)}(\beta, u) \right\} dN_i(u)$$

where

$$S^{(0)}(\beta, u) = \sum_{i=1}^{n} Y_i(u) \exp\{\beta x_i(u)\}$$

The score function is then

$$U(\beta) = \ell'(\beta) = \sum_{i=1}^{n} \int_0^\tau \left\{ x_i(u) - \frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right\} dN_i(u)$$

where

$$S^{(1)}(\beta, u) = \sum_{i=1}^{n} Y_i(u) x_i(u) \exp\{\beta x_i(u)\}$$

Then $\hat{\beta}$ solves $U(\beta) = 0$.

Observed information

The observed information may be written

$$I(\beta) = -\ell''(\beta) = -U'(\beta) = \int_0^\tau V(\beta, u) dN_\bullet(u)$$

where

$$V(\beta, u) = \frac{S^{(2)}(\beta, u)}{S^{(0)}(\beta, u)} - \left(\frac{S^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right)^2$$

and

$$S^{(2)}(\beta, u) = \sum_{i=1}^{n} Y_i(u) x_i(u)^2 \exp\{\beta x_i(u)\}$$

Property of score function

We will now look at the score function $U(\beta)$ when evaluated at the true value β_0 of β.

Note that

$$\begin{aligned} dN_i(t) &= \lambda_i(t)dt + dM_i(t) \\ &= Y_i(t)\exp\{\beta_0 x_i(t)\}\alpha_0(t)dt + dM_i(t) \end{aligned}$$

Inserting this in the expression for the score we obtain

$$U(\beta_0) = \sum_{i=1}^{n} \int_0^{\tau} \left\{ x_i(u) - \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} \right\} dM_i(u)$$

(see next page ....)

It follows that the score evaluated at β_0 is a mean zero martingale, and in particular $E\{U(\beta_0)\} = 0$.

....

.... the last equation follows since ...

$$\sum_{i=1}^{n} \int_0^{\tau} \left\{ x_i(u) - \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} \right\} Y_i(u) \exp\{\beta_0 x_i(u)\} \alpha_0(u) du$$

$$= \int_0^{\tau} \left[\sum_{i=1}^{n} \left\{ x_i(u) - \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} \right\} Y_i(u) \exp\{\beta_0 x_i(u)\} \right] \alpha_0(u) du$$

$$= \int_0^{\tau} \left[S^{(1)}(\beta_0, u) - \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} S^{(0)}(\beta_0, u) \right] \alpha_0(u) du$$

$$= 0$$

Predictable variation of the score function

The predictable variation of the score function may be written

$$\langle U(\beta_0)\rangle(\tau) = \sum_{i=1}^{n}\int_0^{\tau}\left\{x_i(u) - \frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}\right\}^2 \lambda_i(u)du$$

$$= \sum_{i=1}^{n}\int_0^{\tau}\left\{x_i(u)^2 - 2x_i(u)\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)} + \left(\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}\right)^2\right\}$$

$$\times \quad Y_i(u)\exp\{\beta_0 x_i(u)\}\alpha_0(u)du$$

$$= \sum_{i=1}^{n}\int_0^{\tau}\left\{S^{(2)}(\beta_0,u) - 2S^{(1)}(\beta_0,u)\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)} + \left(\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}\right)^2 S^{(0)}(\beta_0,u)\right\}$$

$$\times \quad \alpha_0(u)du$$

$$= \sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{S^{(2)}(\beta_0,u)}{S^{(0)}(\beta_0,u)} - \left(\frac{S^{(1)}(\beta_0,u)}{S^{(0)}(\beta_0,u)}\right)^2\right\} S^{(0)}(\beta_0,u)\alpha_0(u)du$$

$$= \int_0^{\tau} V(\beta_0,u)S^{(0)}(\beta_0,u)\alpha_0(u)du$$

Use of martingale central limit theorem

Recall that $U(\beta_0)$ is a mean zero martingale (in τ). Then if we assume that

$$\frac{1}{n}\int_0^\tau V(\beta_0,u)S^{(0)}(\beta_0,u)\alpha_0(u)du \to \sigma^2$$

we have that $\langle (1/\sqrt{n})U(\beta_0)\rangle \to \sigma^2$. It follows by the martingale central limit theorem that

$$\frac{1}{\sqrt{n}}U(\beta_0) \to Z \sim N(0,\sigma^2) \quad \text{as } n \to \infty$$

Further using

$$dN_\bullet(u) = S^{(0)}(\beta_0,u)\alpha_0(u)du + dM_\bullet(u)$$

we get

$$\begin{aligned}\frac{1}{n}I(\beta_0) &= \frac{1}{n}\int_0^\tau V(\beta_0,u)dN_\bullet(u)\\ &\approx \frac{1}{n}\int_0^\tau V(\beta_0,u)S^{(0)}(\beta_0,u)\alpha_0(u)du \approx \sigma^2\end{aligned}$$

Final result on $\sqrt{n}(\hat{\beta} - \beta_0)$

We have that $U(\hat{\beta}) = 0$.

By a Taylor expansion this gives

$$\begin{aligned} 0 = U(\hat{\beta}) &\approx U(\beta_0) + U'(\beta_0)(\hat{\beta} - \beta_0) \\ &= U(\beta_0) - I(\beta_0)(\hat{\beta} - \beta_0) \end{aligned}$$

It follows that

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta_0) &\approx \left(\frac{1}{n} I(\beta_0)\right)^{-1} \frac{1}{\sqrt{n}} U(\beta_0) \\ &\approx \frac{1}{\sigma^2} \frac{1}{\sqrt{n}} U(\beta_0) \to \frac{1}{\sigma^2} Z \sim N(0, 1/\sigma^2) \end{aligned}$$

Recall that $(1/n)I(\beta_0) \approx \sigma^2$ and estimate $I(\beta_0)$ by $I(\hat{\beta})$. It follows that $1/\sigma^2$ can be estimated by $nI(\hat{\beta})^{-1}$ and hence that $Var(\hat{\beta})$ can be estimated by $I(\hat{\beta})^{-1}$.