Cox regression

STK4080 H16

1. Proportional hazards model
2. Partial likelihood
3. Counting process-martingale representation
4. Large sample properties
5. Estimation of cumulative baseline (Breslow-estimator)

Need for regression models

In K-sample situations "$\alpha_k(t)$ hazard in group k" we may plot Kaplan-Meier or Nelson-Aalen estimators in different groups to describe the differences.

With continuous covariats or many categorical covariats this becomes impossible, due to few or no individuals with each covariat value.

Need to use suitable assumptions on how hazards differ for different covariat values, i.e.

Regression models

Important (afterward) to check if the models are adequate for presnt data.

Tests for $\mathbf{H}_0 : \alpha_1(t) = \alpha_2(t) = \cdots = \alpha_K(t)$ corresponds

to oneway ANOVA which again is a special case of multiple regression.

May code ANOVA as linear expression

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{K-1} x_{K-1}$$

where the x_k are indicator variables for belonging to category k.

In regression analysis we correspondingly use linear expressions

$$\beta' x = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

where x_j is a covariat (explanatory variable) no. j and β_j the regression parameter corresponding to x_j.

On vectorial form the linear expression $\beta' x$ where $\beta' = (\beta_1, \beta_2, \ldots, \beta_p)$ and $x' = (x_1, x_2, \ldots, x_p)$.

Common regression models for survival data

Regression models for survival data are often specified by hazards $\alpha(t|x) =$ hazard with covariat x.

Common models:

- Proportional hazards models $\alpha(t|x) = \exp(\beta' x)\alpha_0(t)$
- Accelerated lifetime models
  $\alpha(t|x) = \exp(\beta' x)\alpha_0(\exp(\beta' x)t)$
- Additive hazard models $\alpha(t|x) = \alpha_0(t) + \beta' x$

All these models include $\alpha_0(t)$ = baseline (underlying) hazard and a linear expression $\beta' x$.

Note that we get $\alpha(t|0) = \alpha_0(t) =$ hazard with covariat $x = 0$.

General risk functions

ABG (and other texts) present the proportional hazards model in the more general form hazard for individual i given by

$$\alpha(t|x_i) = r(\beta, x_i(t))\alpha_0(t)$$

where $r(\beta, x_i(t))$ is some non-negative function. An example is

$$r(\beta, x_i(t)) = (1 + \beta_1 x_{i1}(t))(1 + \beta_2 x_{i2}(t))$$

used for an uranium miner data set.

Most of the theory goes through as easy as with the exponential risk function $r(\beta, x_i(t)) = \exp(\beta' x_i(t))$, but some expressions gets a little more messy.

There is some software that allow for other risk functions, but I have not seen this in R.

The proportional hazards model: 1. One covariat

Hazard rate for subject with one covariat x:

$$\alpha_x(t) = \alpha_0(t)\exp(\beta x)$$

where baseline hazard $\alpha_0(t)$ is the hazard for subject with $x = 0$.

Interpretation: Hazard rate ratio (or loosely Relative Risk),

$$\mathrm{HR} = \exp(\beta(x_1 - x_0)) = \frac{\alpha_{x_1}(t)}{\alpha_{x_0}(t)}$$

In particular with x binary

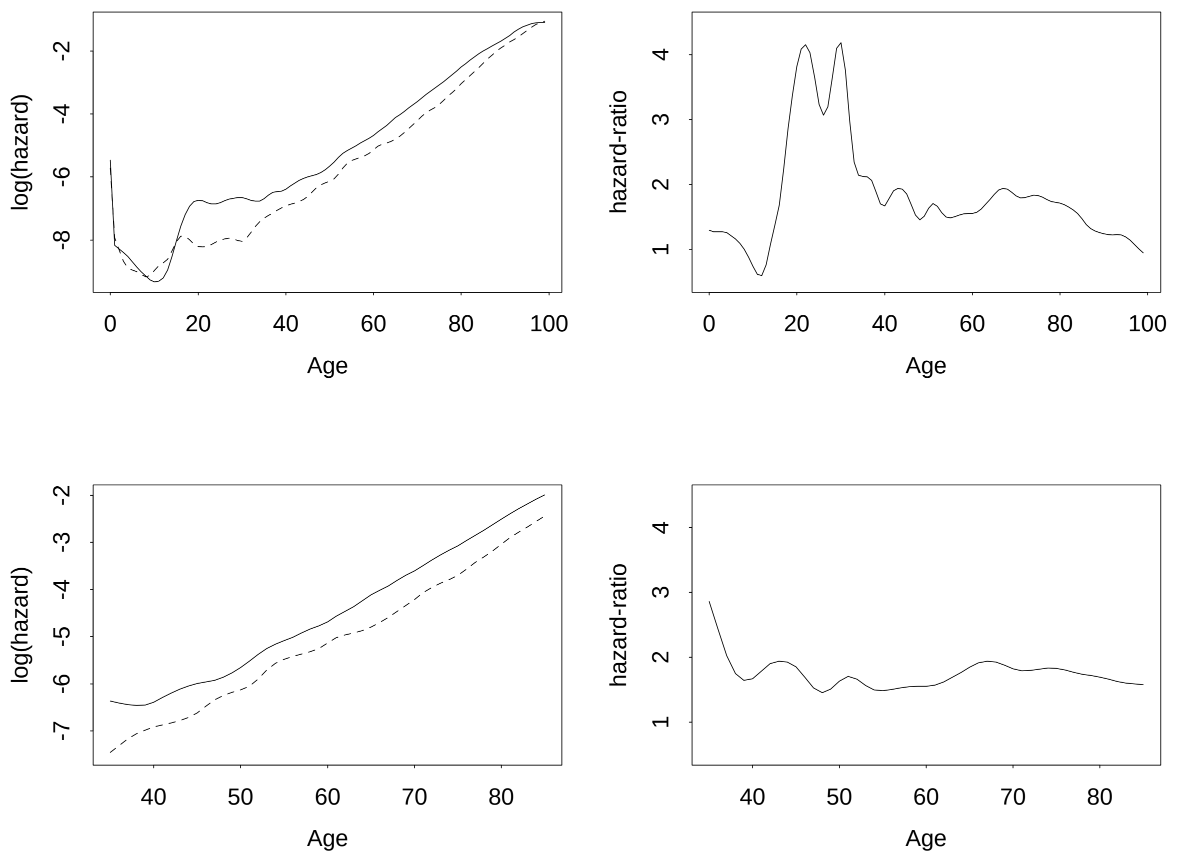
$$\mathrm{HR} = \exp(\beta) = \frac{\alpha_1(t)}{\alpha_0(t)}$$

Example: Mortality rates among men and women, Statistics Norway, 2000, smoothed.

Binary covariate x indicator of men.

Prop. hazard model **not** valid in age interval 0-100 years

Prop. hazard model roughly valid in interval 40-85 years with HR ≈ 1.8.



Example 1: Melanoma data

T = time to death from melanoma

hazard $\alpha_x(t) = \alpha_0(t)\exp(\beta x)$

x = indicator of ulceration,

$HR = \frac{\alpha_1(t)}{\alpha_0(t)} = \exp(\beta) =$ hazard ratio between those with and without ulceration.

x_1 = tumor thickness (mm) subject 1,

$x_2 =$ thickness (mm) subject 2= $x_1 + 1$ mm,

$HR = \exp(\beta)$ = rate ratio w. 1 mm difference.

Proportional hazards model: 2. Several covariates

Hazard rate for individual with covariat vector $x = (x_1, x_2, ..., x_p)$

$$\alpha_x(t) = \alpha_0(t) \exp\{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p\}$$

where baseline hazard $\alpha_0(t)$ is hazard function for individual with all $x_1 = x_2 = ... = x_p = 0$.

Interpretation: Hazard rate ratio (HR)

Another subject with $x' = (x'_1, x'_2, ..., x'_p)$ where $x'_1 = 1$, $x_1 = 0$ and $x'_j = x_j$ otherwise:

$$\mathrm{HR}_1 = \exp\{\beta_1\} = \frac{\alpha_{x'}(t)}{\alpha_x(t)}$$

Example 1: Melanoma data

$$\alpha_x(t) = \alpha_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)$$

x_1 = sex (M=1, F=0)

x_2 = indicator of ulceration,

$x_3 =$ age,

$x_4 =$ thickness (mm)

$x = (x_1, 0, x_3, x_4)$

$x' = (x_1, 1, x_3, x_4)$

$HR = \frac{\alpha_x(t)}{\alpha_{x'}(t)} = \exp(\beta_2) =$ hazard ratio between those with and without ulceration **adjusted** for sex, age and thickness.

Estimation in the proportional hazards model

- With baseline hazard $\alpha_0(t) = \alpha_0(t, \theta)$ parametrically specified
$\rightarrow$ by likelihood for censored data (ABG Ch. 5).

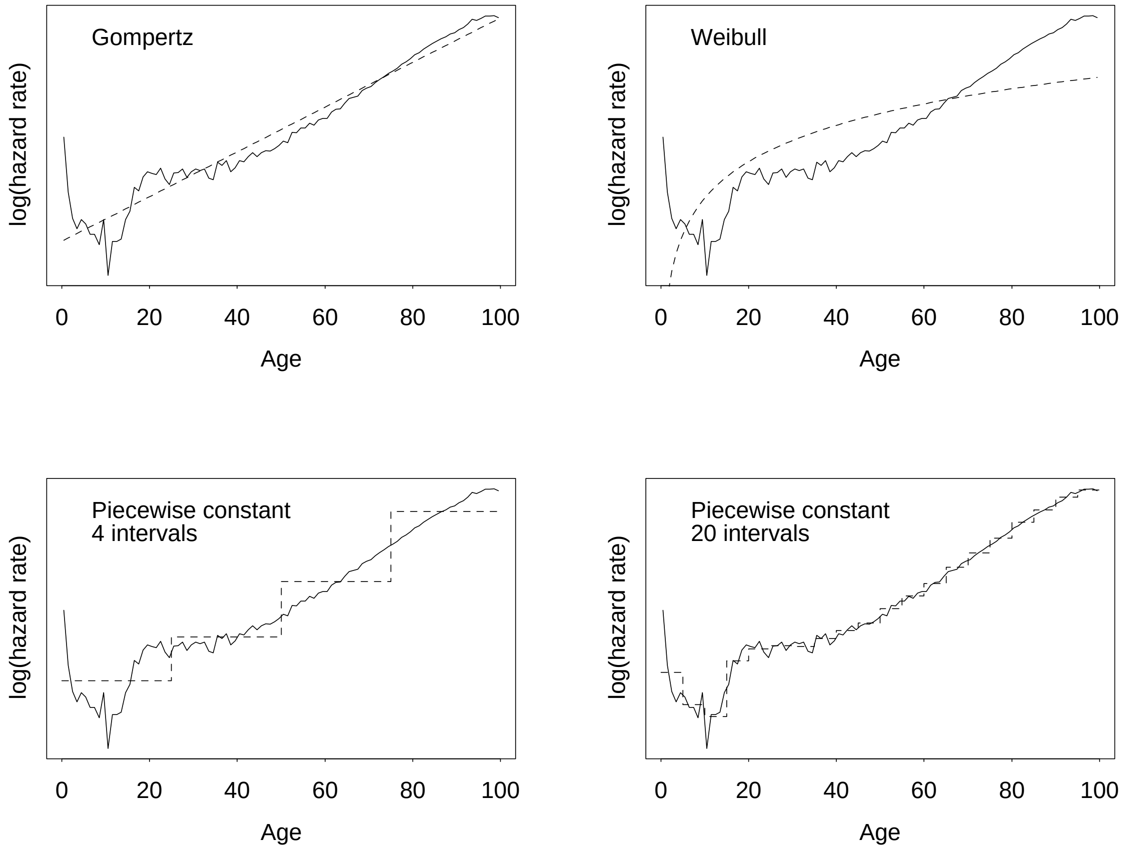
$$\text{Gompertz: } \alpha_0(t, \theta = (\gamma, \lambda)) = \lambda \gamma^t$$

$$\text{Weibull: } \alpha_0(t, \theta = (\gamma, \lambda)) = \lambda^\gamma t^{\gamma - 1}$$

- With baseline $\alpha_0(t) = \alpha_{0j}$ piecewise constant on $(t_{j-1}, t_j]$
$\rightarrow$ by Poisson regression (ABG, 5.2.1).

- With baseline hazard $\alpha_0(t)$ arbitrary function
$\rightarrow$ by Cox-regression (ABG, 4.1).

Comparison of different types of baseline hazards



Cox (1972), Regression models and life-tables, JRSSB:

We may estimate β without assuming anything about $\alpha_0(t)$!

Main interest lies in effect β_j from covariat x_j.

Baseline $\alpha_0(t)$ represent "nuisance parameters" (plageparametre).

References in ISI-database

- Per 12. Oct 2004: 19246 citations
- Per 12. Oct 2010: 24606
- Per 12. Oct 2016 (yesterday): 30343

Cox' Regression:

Death at t_i. Let

$$\begin{aligned} L_i(\beta) &= \text{P(Subject } i \text{ died at } t_i | i \in \mathcal{R}(t_i), \text{ death at } t_i) \\ &= \frac{\alpha_i(t_i)}{\sum_{k\in\mathcal{R}(t_i)} \alpha_k(t_i)} \\ &= \frac{\exp(\beta x_i)\alpha_0(t_i)}{\sum_{k\in\mathcal{R}(t_i)} \exp(\beta x_k)\alpha_0(t_i)} \\ &= \frac{\exp(\beta x_i)}{\sum_{k\in\mathcal{R}(t_i)} \exp(\beta x_k)} \end{aligned}$$

where

- $\alpha_i(t) = \alpha_0(t)\exp(\beta x_i) =$ hazard subject i at t
- $\mathcal{R}(t)$ = subjects under observation at $t-$ = risk set at t.

Note $L_i(\beta)$ depend on β only,
not on the baseline hazard $\alpha_0(t)$.

Cox' Partial likelihood:

Assume that individual i dies at $t_i, i = 1, \ldots, D =$ no. deaths.

Estimates β by maximizing (Cox, 1972)

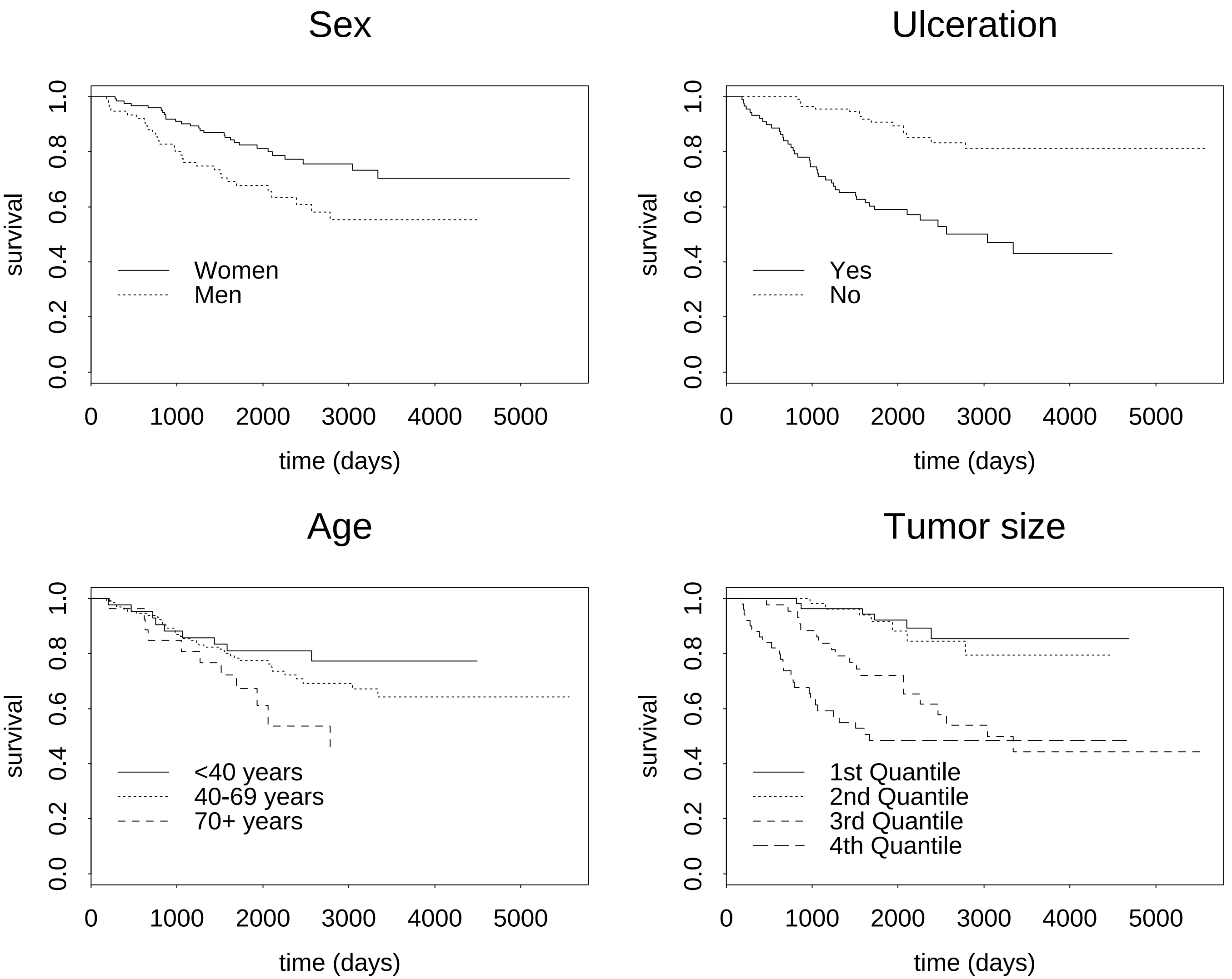
$$L(\beta) = \prod_{i=1}^{D} L_i(\beta) = \prod_{i=1}^{D} \frac{\exp(\beta' x_i)}{\sum_{k \in \mathcal{R}(t_i)} \exp(\beta' x_k)}$$

Note: We my estimate β and $HR_j = \exp(\beta_j)$ without specifying the baseline $\alpha_0(t)$.

In addition: The partial likelihood $L(\beta)$ behaves as a standard likelihood (will show this).

In particular: standard errors for $\hat{\beta}$ as for standard likelihoods.

Example: Melanoma data



Example: Melanoma data

Variable	$\hat{\beta}$	se($\hat{\beta}$)	Z-value	p-value
tumorsize (mm)	0.11	0.04	2.89	0.004
ulceration	1.16	0.31	3.76	0.0002
sex (F=0,M=1)	0.43	0.27	1.62	0.11
age (years/10)	0.12	0.08	1.47	0.14

Variable	$\hat{\text{HR}} = \exp(\hat{\beta})$	$\hat{\text{HR}}_L$	$\hat{\text{HR}}_U$
tumorsize (mm)	1.12	1.04	1.20
ulceration	3.20	1.75	5.88
sex (F=0,M=1)	1.54	0.91	2.60
age (years/10)	1.13	0.96	1.33

R-code and output:

```
> coxph(Surv(lifetime,dead)~sex+ulcer+age+thickn,data=mel)
Call:
coxph(formula = Surv(lifetime, dead) ~ sex + ulcer + age + thickn, data


          coef exp(coef) se(coef)     z       p
   sex  0.4328     1.542   0.2674  1.62 0.11000
 ulcer -1.1645     0.312   0.3098 -3.76 0.00017
   age  0.0122     1.012   0.0083  1.47 0.14000
thickn  0.1089     1.115   0.0377  2.89 0.00390

Likelihood ratio test=41.6  on 4 df, p=2e-08  n= 205
```

More R-code and output:

```
> summary(coxph(Surv(lifetime,dead)~sex+ulcer+age+thickn,data=mel))
coxph(formula = Surv(lifetime, dead) ~ sex + ulcer + age + thickn, data

  n= 205
          coef exp(coef) se(coef)     z       p
   sex  0.4328     1.542   0.2674  1.62 0.11000
 ulcer -1.1645     0.312   0.3098 -3.76 0.00017
   age  0.0122     1.012   0.0083  1.47 0.14000
thickn  0.1089     1.115   0.0377  2.89 0.00390

       exp(coef) exp(-coef) lower .95 upper .95
   sex     1.542      0.649     0.913     2.604
 ulcer     0.312      3.204     0.170     0.573
   age     1.012      0.988     0.996     1.029
thickn     1.115      0.897     1.036     1.201

Rsquare= 0.184   (max possible= 0.937 )
Likelihood ratio test= 41.6  on 4 df,   p=2e-08
Wald test            = 39.4  on 4 df,   p=5.72e-08
Score (logrank) test = 46.7  on 4 df,   p=1.79e-09
```

What kind of "animal" is $L(\beta)$? really

- Cox(1972): Conditional likelihood.
  Problematic with time dependent covariats.
- Cox(1975): Partial likelihood.
  But what exactly is that?
- Johansen (1983): Profile-likelihood over $\alpha_0(t)$.
  Nice, but helpful?
- Andersen & Gill (1982): $U(\beta) = \frac{\partial \log(L(\beta))}{\partial \beta}$ is a martingale.
  Leads to likelihood properties: $U(\beta)$ approx. normal with

$$\mathrm{E}[U(\beta)] = 0$$

$$\mathrm{Var}[U(\beta)] = -\mathrm{E}[\frac{\partial^2 \log(L(\beta))}{\partial \beta^2}]$$

Why - again - is the MLE approx. normal?

Assume $\hat{\theta}$ maximizes likelihood $L(\theta)$ and has score function $U(\theta) = \frac{\partial \log(L(\theta))}{\partial \theta}$ and information $I(\theta) = -\frac{\partial^2 \log(L(\theta))}{\partial \theta^2}$ such that $U(\hat{\theta}) = 0$

- $\mathrm{E}[U(\theta)] = 0$ and $\mathrm{Var}[U(\theta)] = \mathrm{E}[I(\theta)]$
- $\frac{1}{\sqrt{n}} U(\theta) \sim \mathrm{N}(0, \Sigma)$
  where Σ is the limit for $\frac{1}{n} I(\theta)$ when $n \to \infty$.

Then, by 1.order Taylor expansion

$$0 = U(\hat{\theta}) = U(\theta) - (\hat{\theta} - \theta) I(\theta) + \text{ remainder term}$$

which give

$$\sqrt{n}(\hat{\theta} - \theta) \approx \frac{n^{-1/2} U(\theta)}{n^{-1} I(\theta)} \to \mathrm{N}(0, \Sigma^{-1}).$$

Martingale representation for $U(\beta)$. Define

- $Y_k(t) = I($ ind. k under risk at (right before) $t)$
- $N_k(t) =$ counting process for ind. no. k
- $S^{(0)}(\beta, t) = \sum_{k=1}^{n} Y_k(t) \exp(\beta' x_k)$
- $S^{(1)}(\beta, t) = \sum_{k=1}^{n} x_k Y_k(t) \exp(\beta' x_k) = \frac{\partial S^{(0)}(\beta,t)}{\partial \beta}$
- $S^{(2)}(\beta, t) = \sum_{k=1}^{n} x_k x_k^{\top} Y_k(t) \exp(\beta' x_k) = \frac{\partial^2 S^{(0)}(\beta,t)}{\partial \beta^2}$

Then

$$L(\beta) = \prod_{i=1}^{D} \frac{\exp(\beta' x_i)}{S^{(0)}(\beta, t_i)}$$

which leads to the log-partial likelihood

$$\begin{aligned} \log(L(\beta)) &= \sum_{i=1}^{D} [\beta' x_i - \log(S^{(0)}(\beta, t_i))] \\ &= \sum_{i=1}^{n} \int [\beta' x_i - \log(S^{(0)}(\beta, t))] dN_i(t) \end{aligned}$$

$U(\beta)$ martingale, contd.

Since $\frac{\partial \log(S^{(0)}(\beta,t))}{\partial \beta} = \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}$ we get score

$$U(\beta) = \sum_{i=1}^{n} \int [x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] dN_i(t)$$

But

$$dN_i(s) = Y_i(t) \exp(\beta' x_i) \alpha_0(t) dt + dM_i(t)$$

where $\lambda_i(t) = Y_i(t) \exp(\beta' x_i) \alpha_0(t)$ is the intensity process and $M_i(t)$ the martingale related to $N_i(t)$. Inserting gives

$$\begin{aligned} U(\beta) = & \textstyle\sum_{i=1}^{n} \int [x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] Y_i(t) \exp(\beta' x_i) \alpha_0(t) dt \\ & + \textstyle\sum_{i=1}^{n} \int [x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] dM_i(t) \\ = & \textstyle\sum_{i=1}^{n} \int [x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}] dM_i(t) \end{aligned}$$

$U(\beta)$ martingal III

since

$$\sum_{i=1}^{n}\int[x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]Y_i(t)\exp(\beta' x_i)\alpha_0(t)dt = 0.$$

This follows from $\sum_{i=1}^{n}[x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]Y_i(t)\exp(\beta' x_i)$

$$= S^{(1)}(\beta,t) - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}S^{(0)}(\beta,t) = 0$$

and so

$$U(\beta) = \sum_{i=1}^{n}\int[x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]dM_i(t)$$

is a sum of integrals wrt. martingals, thus itself a martingale with $\mathrm{E}[U(\beta)] = 0$.

Will show Var$[U(\beta)] = $ E$[I(\beta)] = $ E $\left[-\frac{\partial U(\beta)}{\partial \beta}\right]$

(but only for one covariat, $p = 1$)

The $M_i(t)$ are uncorrelated martingales with

$$\mathrm{Var}[M_i(t)] = \mathrm{E}[\int_0^t Y_i(s)\exp(\beta x_i)\alpha_0(s)ds].$$

Thus

$$\begin{aligned} \mathrm{Var}(U(\beta)) \quad &= \mathrm{E}\textstyle\sum_{i=1}^n \int [x_i - \frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]^2 Y_i(t)\exp(\beta x_i)\alpha_0(t)dt \\ &\qquad\qquad \text{(after some calculations)} \\ &= \mathrm{E}\textstyle\int \left[S^{(2)}(\beta,t) - \frac{(S^{(1)}(\beta,t))^2}{S^{(0)}(\beta,t)}\right]\alpha_0(t)dt \\ &= \mathrm{E}[I(\beta)] \end{aligned}$$

where the last equality will be derived

Contd. Var$[U(\beta)] = $ E$[I(\beta)]$

But furthermore $\frac{\partial[\frac{S^{(1)}(\beta,t)}{S^{(0)}(\beta,t)}]}{\partial\beta} = \frac{S^{(2)}}{S^{(0)}} - [\frac{S^{(1)}}{S^{(0)}}]^2$ which give

$$\begin{aligned} I(\beta) = -\frac{\partial U(\beta)}{\partial\beta} &= -\sum_{i=1}^n \int \frac{\partial[x_i - \frac{S^{(1)}}{S^{(0)}}]}{\partial\beta} dN_i(t) \\ &= \sum_{i=1}^n \int \{\frac{S^{(2)}}{S^{(0)}} - [\frac{S^{(1)}}{S^{(0)}}]^2\} dN_i(t) \\ &= \int \{\frac{S^{(2)}}{S^{(0)}} - [\frac{S^{(1)}}{S^{(0)}}]^2\} dN_\bullet(t) \end{aligned}$$

where $N_\bullet(t) = \sum_{i=1}^n N_i(t) = \int S^{(0)}(\beta,t)\alpha_0(t)dt + M_\bullet(t)$.
And this leads to

$$\mathrm{E}[I(\beta)] = \mathrm{E}[\int \{\frac{S^{(2)}}{S^{(0)}} - [\frac{S^{(1)}}{S^{(0)}}]^2\} S^{(0)}(\beta,t)\alpha_0(t)dt] = \mathrm{Var}(U(\beta))$$

qed.

Large sample properties Cox-estimator

Furthermore, by the martingale CLT, $\frac{1}{\sqrt{n}}U(\beta) \to \mathrm{N}(0, \sigma^2)$ where $\frac{I(\beta)}{n} \to \sigma^2$, and so

$$\hat{\beta} \sim \mathrm{N}(\beta, I(\hat{\beta})^{-1})$$

for β scalar. We may thus test $\mathrm{H}_0 : \beta = 0$ by statistic

$$Z = \frac{\hat{\beta}}{\mathrm{se}} \sim \mathrm{N}(0, 1)$$

where $\mathrm{se}^2 = I(\hat{\beta})^{-1}$.

But since $L(\beta)$ have the usual likelihood properties we may alternatively teste $\mathrm{H}_0 : \beta = 0$ by (under the null)

- Likelihood-ratio test: $2[\log(L(\hat{\beta})) - \log(L(0))] \sim \chi^2_1$
- Score-test: $U(0) \sim N(0, I(0))$ and $\frac{U(0)^2}{I(0)} \sim \chi^2_1$

Cox Score test and Log-rank test

Assume that x_i is a binary variable and the only covariate.

Then the score test for Cox-regression is the same as the log-rank if there are no ties (Exercise 4.2).

This result generalises directly to dummy variables indicating levels of a categorical covariate

For this reason the score test for Cox-regression is referred to as a generalized log-rank test.

Comparison Cox-regression and Log-rank

```
> summary(coxph(Surv(lifetime,dead)~ulcer,data=mel))

        coef exp(coef) se(coef)     z       p
ulcer -1.47      0.23    0.295 -4.98 6.3e-07

      exp(coef) exp(-coef) lower .95 upper .95
ulcer      0.23       4.36     0.129      0.41

Rsquare= 0.13   (max possible= 0.937 )
Likelihood ratio test= 28.4  on 1 df,   p=9.68e-08
Wald test            = 24.8  on 1 df,   p=6.3e-07
Score (logrank) test = 29.6  on 1 df,   p=5.41e-08

> survdiff(Surv(lifetime,dead)~ulcer,data=mel)

          N Observed Expected (O-E)^2/E (O-E)^2/V
ulcer=1  90       41     21.2      18.5      29.6
ulcer=2 115       16     35.8      10.9      29.6

 Chisq= 29.6  on 1 degrees of freedom, p= 5.41e-08
```

Generalized log-rank test

For general x_i (not necessarily binary) we have

$$U(0) = \sum_{i=1}^{n} \int [x_i - \frac{\sum_{k=1}^{n} x_k Y_k(t)}{Y(t)}] dN_i(t)$$

which has variance under $\mathrm{H}_0 : \beta = 0$ given by

$$I(0) = \int \left\{ \frac{\sum_{i=1}^{n} x_i^2 Y_i(t)}{Y(t)} - \left[\frac{\sum_{i=1}^{n} x_i^2 Y_i(t)}{Y(t)} \right] \right\} dN_{\bullet}(t)$$

The generalized log-rank test is the given by that under H_0:

$$\frac{U(0)^2}{I(0)} \sim \chi_1^2$$

Large sample properties Cox-estimator, II

We showed

$$\text{Var}(U(\beta)) = \text{E}[I(\beta)]$$

for $p = 1$ covariat. The result holds true also for general $p > 1$. Thus

$$\hat{\beta} \sim \text{N}(\beta, I(\hat{\beta})^{-1})$$

when $\hat{\beta}$ is a p-dim. vector and $I(\beta)$ a pxp-matrix.

We may teste the "complete" null hypothesis of no effect of any of the covariates $\text{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ by the 3 (asymptotically equivalent) tests:

- Wald-test: $\hat{\beta}^\top I(\hat{\beta})\hat{\beta} \sim \chi^2_p$
- Likelihood ratio test: $2[\log(L(\hat{\beta})) - \log(L(0))] \sim \chi^2_p$
- Score-test $U(0)^\top I(0)^{-1} U(0) \sim \chi^2_p$

R-output (repeated):

```
> summary(coxph(Surv(lifetime,dead)~sex+ulcer+age+thickn,data=mel))
coxph(formula = Surv(lifetime, dead) ~ sex + ulcer + age + thickn, data

  n= 205
          coef exp(coef) se(coef)     z       p
   sex  0.4328     1.542   0.2674  1.62 0.11000
 ulcer -1.1645     0.312   0.3098 -3.76 0.00017
   age  0.0122     1.012   0.0083  1.47 0.14000
thickn  0.1089     1.115   0.0377  2.89 0.00390

       exp(coef) exp(-coef) lower .95 upper .95
   sex     1.542      0.649     0.913     2.604
 ulcer     0.312      3.204     0.170     0.573
   age     1.012      0.988     0.996     1.029
thickn     1.115      0.897     1.036     1.201

Rsquare= 0.184   (max possible= 0.937 )
Likelihood ratio test= 41.6  on 4 df,   p=2e-08
Wald test            = 39.4  on 4 df,   p=5.72e-08
Score (logrank) test = 46.7  on 4 df,   p=1.79e-09
```

HR with CI

Naturally we estimate hazard-rate-ratios $\text{HR}_j = \exp(\beta_j)$ by

$$\widehat{\text{HR}}_j = \exp(\hat{\beta}_j)$$

The diagonal of $I(\hat{\beta})^{-1}$ gives the variance estimates for the different $\hat{\beta}_j$. So with $se_j =$ square-root of $j-$th diagonal element we get a 95% CI for β_j by

$$\hat{\beta}_j \pm 1.96 se_j.$$

This interval is transformed to a 95% CI for HR_j by

$$\exp(\hat{\beta}_j \pm 1.96 se_j) = \widehat{\text{HR}}_j \exp(\pm 1.96 se_j)$$

Composite hypothese

Ex: Melanoma data. We may be more concerned with some variables, for.ex.

- Ulceration and thickness primary variables
- Sex and age confounds the association (are "lurking variables") and should only be adjusted for.

We may thus want to test, with x_{i1} = ulceration and x_{i2}= tumor thickness,

$$\mathrm{H}_0 : \beta_1 = \beta_2 = 0,$$

while β_3 and β_4 may take arbitrary values.

Composite hypothesis, in general

Let β be p-dimensional. We will for $q < p$ test

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_q = 0,$$

while $\beta_{q+1}, \ldots, \beta_p$ may be aribtrary. Let the Cox-estimator for β be

- Apriori: $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)'$
- Under $\mathrm{H}_0 : \beta^\star = (0, 0, \ldots, 0, \beta^\star_{q+1}, \ldots, \beta^\star_p)'$

and the blocked covariance matrix $\hat{\Sigma} = I(\hat{\beta})^{-1}$ for $\hat{\beta}$:

$$\hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}$$

where $\hat{\Sigma}_{11}$ is the qxq covariance matrix for $(\hat{\beta}_1, \ldots, \hat{\beta}_q)'$ etc.

More composite hypothesis

Let also $\hat{\beta}_{(q)} = (\hat{\beta}_1, \ldots, \hat{\beta}_q)'$. Then we may test the null hypotesis by

- Wald-test: $\hat{\beta}_{(q)}^{\top} \hat{\Sigma}_{11}^{-1} \hat{\beta}_{(q)} \sim \chi_q^2$
- Likelihood ratio: $2[\log(L(\hat{\beta})) - \log(L(\beta^{\star}))] \sim \chi_q^2$

approximately under H_0.

A related score-test for composite hypotheses may be developed.

Composite hypothesis example

Tests H_0 : No effect of ulceration or tumor thickness

1. Likelihood Ratio:

```
> coxfit0<-coxph(Surv(lifetime,dead)~age+sex,data=mel)
> coxfit1<-coxph(Surv(lifetime,dead)~ulcer+thickn+age+sex,data=mel)
> coxfit0$loglik
[1] -283.1992 -278.2284
> coxfit1$loglik
[1] -283.1992 -262.3895
> LR<-2*(coxfit1$loglik[2]-coxfit0$loglik[2])
> LR
[1] 31.67779
> 1-pchisq(LR,2)
[1] 1.322071e-07
```

It turns out that there is a clear effect of these two variables also when we adjust for age and sex.

Composite hypothesis example, II

The Wald-test is also simple:

```
> bmat<-matrix(coxfit1$coef[1:2],ncol=1)

> kjikv<-t(bmat)%*%solve(coxfit1$var[1:2,1:2])%*%bmat
> kjikv
         [,1]
[1,] 30.87181

> 1-pchisq(kjikv,2)
[1] 1.978209e-07
```

Comparable result - but small sample properties of Wald-tests are often poorer than LR and Score-tests.

Estimation of cumulative hazard $A_0(t) = \int_0^t \alpha_0(s)ds$

So we manage to estimate β without saying anything about the baseline. However, often we still want to estimate $\alpha_0(t)$ after having estimated β. The most common estimator for $A_0(t)$ is the Breslow-estimator

$$\hat{A}_0(t) = \sum_{\tilde{T}_i \leq t} \frac{D_i}{\sum_{k \in \mathcal{R}(\tilde{T}_i)} \exp(\hat{\beta}' x_k)} = \int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\hat{\beta}, s)}$$

Note that this estimator is similar to the Nelson-Aalen estimator.

Given $\hat{A}_0(t)$ it is simple to estimate cumulative hazard for an individual with hazard $\alpha(t|x_i) = \exp(\beta' x_i)\alpha_0(t)$ as

$$\hat{A}(t|x_i) = \exp(\hat{\beta}' x_i)\hat{A}_0(t)$$

Estimation of survival

With covariate $x = 0$ the survival function equals $S_0(t) = \exp(-A_0(t))$. Thus it may be estimated by

$$\hat{S}_0(t) = \exp(-\hat{A}_0(t)).$$

With covariate x_i the survival function

$$S(t|x_i) = \exp(-A(t|x_i)) = \exp(-\exp(\beta' x_i) A_0(t))$$

which may be estimated

$$\hat{S}(t|x_i) = \exp(-\hat{A}(t|x_i)) = \hat{S}_0(t)^{\exp(\hat{\beta}' x_i)}.$$

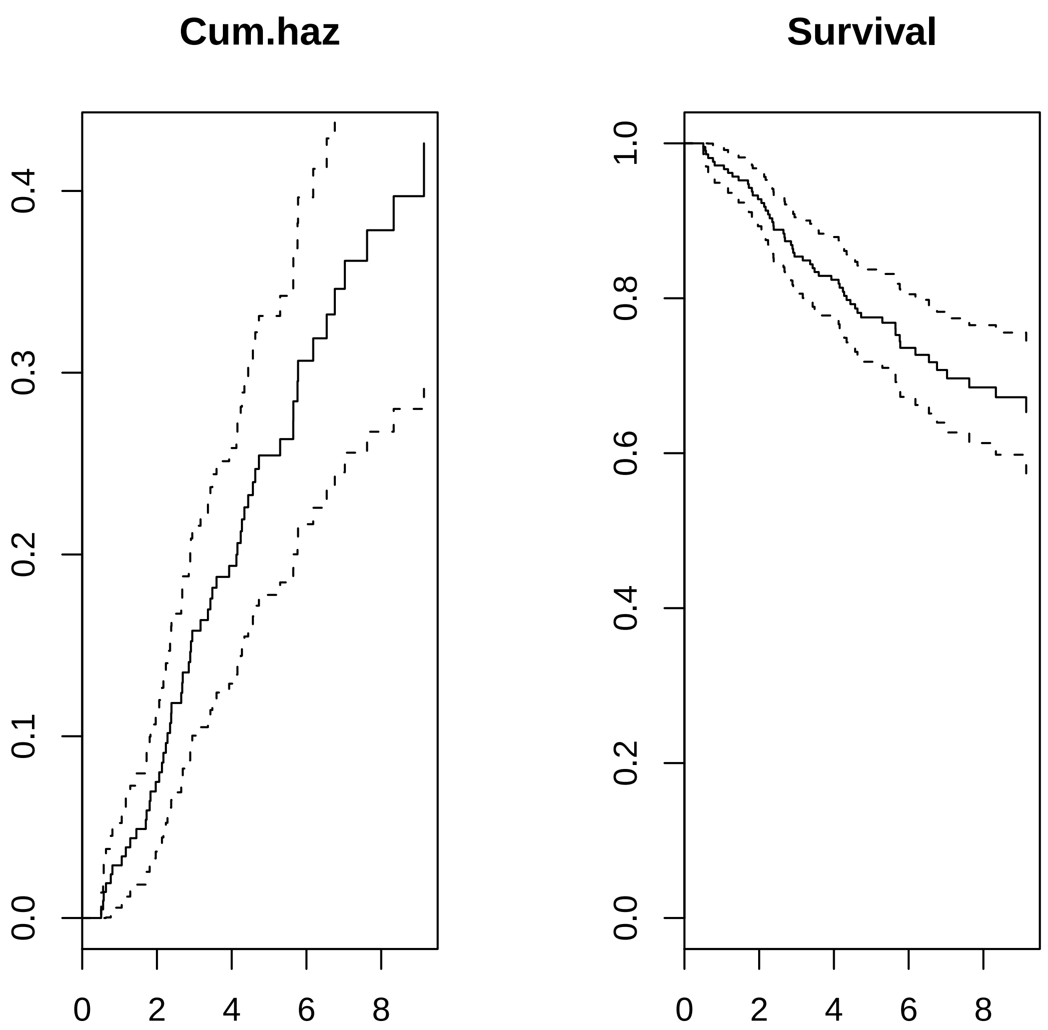
Breslow-estimator in R

- Generate a Cox-regression object:

```
cox.mel<- coxph(Surv(lifetime,status==1)~factor(sex),
data=melanoma)
```

- The Breslow estimator for survival is calculated by `survfit(cox.mel,type="aalen")`
- and may be plotted by `plot(survfit(cox.mel,type="aalen"),fun="cumhaz")`
- The corresponding survival function is obtained by `plot(survfit(cox.mel,type="aalen"))`

Breslow-estimator in R, Figure



Breslow-est. in R, contd.

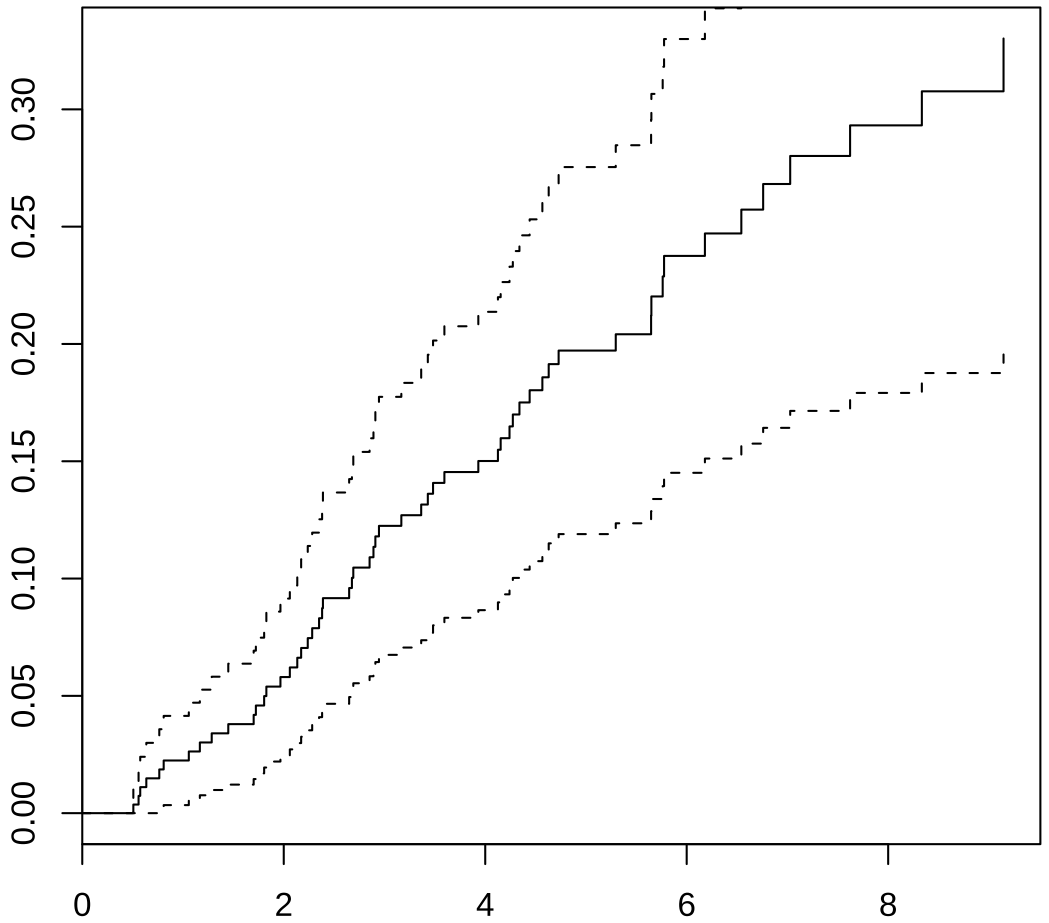
Note. This estimator is calculated for the **average covariate**, in the example `sex`=1.385.

To calculate $\hat{S}(t|x)$ for a specified x we need to make a new "data frame" with the same covariate names as the Cox object and the specified covariate values

```
nydata<-as.data.frame(matrix(rep(1,8),nrow=1))
names(nydata)<-names(melanoma)
nydata
plot(survfit(cox.mel,newdata=nydata,type="aalen"),fun="cumhaz")
```

R & Breslow, last

This gives the plot



Why is $\hat{A}_0(t)$ a sensible estimator?

(when the proportional hazards model is correct)?

Martingale decomposition gives

$$N_\bullet(t) = \int_0^t S^{(0)}(\beta, s)\alpha_0(s)ds + M_\bullet(t)$$

With true value of β (and number at risk > 0) this leads to

$$\int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\beta, s)} = \int_0^t \alpha_0(s)ds + \text{ Martingale}$$

which has expectation $A_0(t)$. Because $\hat{\beta} \approx \beta$ we then get

$$\hat{A}_0(t) = \int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\hat{\beta}, s)} \approx \int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\beta, s)} \approx A_0(t)$$

Variance of $\hat{A}_0(t)$

may also be derived by martingale arguments. Here is an outline.

By Taylor-expansion we get

$$\frac{1}{S^{(0)}(\hat{\beta}, s)} - \frac{1}{S^{(0)}(\beta, s)} = -(\hat{\beta} - \beta)\frac{S^{(1)}(\beta, s)}{S^{(0)}(\beta, s)^2} + \text{ remainder term.}$$

This leads to

$$\begin{aligned}\hat{A}_0(s) &= \int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\beta,s)} + \int_0^t \left[\frac{1}{S^{(0)}(\hat{\beta},s)} - \frac{1}{S^{(0)}(\beta,s)}\right] dN_\bullet(s) \\ &\approx A_0(t) + \int_0^t \frac{dM_\bullet(s)}{S^{(0)}(\beta,s)} - (\hat{\beta} - \beta)\int_0^t \frac{S^{(1)}(\beta,s)}{S^{(0)}(\beta,s)^2} dN_\bullet(s)\end{aligned}$$

Variance of $A_0(t)$ contd.

Furthermore (luckily) $\hat{\beta} - \beta$ and $\int_0^t \frac{dM_\bullet(s)}{S^{(0)}(\beta,s)}$ are asymptotically uncorrelated (arguments omitted).
Thus we may estimate $\text{Var}(\hat{A}_0(t))$ by

$$\widehat{\text{Var}}[\hat{A}_0(t)] = \left[\int_0^t \frac{S^{(1)}(\hat{\beta},s)}{S^{(0)}(\hat{\beta},s)^2} dN_\bullet(s)\right]^\top I(\hat{\beta})^{-1} \int_0^t \frac{S^{(1)}(\hat{\beta},s)}{S^{(0)}(\hat{\beta},s)^2} dN_\bullet(s) + \int_0^t \frac{dN_\bullet(s)}{S^{(0)}(\beta,s)^2}$$

and because $\hat{S}_0(t) - S_0(t) \approx -S_0(t)[\hat{A}_0(t) - A_0(t)]$ an estimator of the variance of $\hat{S}_0(t)$ equals

$$\widehat{\text{Var}}[\hat{S}_0(t)] = \hat{S}_0(t)^2 \widehat{\text{Var}}[\hat{A}_0(t)]$$