STK4080/9080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 13: Aalen's additive regression model

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

Survival regression

Assume that we have a sample of n individuals, and let $N_i(t)$ count the observed occurrences of an event of interest for individual i as a function of (study) time t.

We assume that the intensity process of $N_i(t)$ may be given as

$$\lambda_i(t) = Y_i(t)\alpha(t|\mathbf{x})$$

Earlier we have considered *relative risk* regression models where the hazard rate for individual i takes the form

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

with $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \ldots, x_{ip}(t))^T$ a vector of (possibly) time-dependent covariates.

Aalen's additive regression model

We will now consider the non-parametric additive regression model (or *excess risk regression model*) due to Aalen, where the hazard rate for individual i takes the form

$$\begin{aligned}\alpha(t|\mathbf{x}_i) &= \beta_0(t) + \beta_1(t)x_{i1}(t) + \cdots + \beta_p(t)x_{ip}(t) \\ &= \beta_0(t) + \sum_{q=1}^{p} \beta_q(t)x_{iq}(t)\end{aligned}$$

The $\beta_q(t)$ are *regression functions*.

- The additive regression model is a completely **nonparametric** model that allows the effect of covariates to *change over time*.
- *Note that the model does not make assumptions that constrain the hazard to be non-negative. This is practical, but for small data sets the estimated hazard may take negative values.*
- For estimation, we focus on the **cumulative regression functions**

$$B_q(t) = \int_0^t \beta_q(s)ds$$

Aalen's additive regression model

Generally, at each time t we have for $i = 1, 2, \ldots, n$,

$$\begin{aligned}
dN_i(t) &= \lambda_i(t)dt + dM_i(t) \\
&= Y_i(t)\alpha(t|\mathbf{x}_i)dt + dM_i(t) \\
&= Y_i(t)\left\{\beta_0(t) + \sum_{q=1}^{p} \beta_q(t)x_{iq}(t)\right\} dt + dM_i(t) \\
&\equiv Y_i(t)dB_0(t) + \sum_{q=1}^{p} Y_i(t)x_{iq}(t)dB_q(t) + dM_i(t)
\end{aligned}$$

We may then estimate the increments $dB_q(t) = \beta_q(t)dt$ by **ordinary least squares** at each time $t = T_j$ when an event occurs (*next page*).

The estimate of $B_q(t)$ is then obtained by adding together the estimated increments at all event times $T_j \leq t$.

The regression at an event time $t^* = T_j$

Recall $dN_i(t) = Y_i(t)dB_0(t) + \sum_{q=1}^{p} Y_i(t)x_{iq}(t)dB_q(t) + dM_i(t) \qquad (\#)$

Let t^* be the time of an event, experienced by individual i^*, say. Let

$$V_i = dN_i(t^*) \qquad \text{for } i = 1, \ldots, n$$

Then $V_{i^*} = 1$ and $V_i = 0$ for $i \neq i^*$. Now(#) can be written on linear regression form

$$V_i = \sum_{q=0}^{p} u_{iq}\delta_q + \epsilon_i$$

where $\delta_q = dB_q(t^*)$, $u_{iq} = Y_i(t^*)x_{iq}(t^*)$ (with $x_{i0} = 1$) and $\epsilon_i = dM_i(t^*)$, or on vector/matrix form

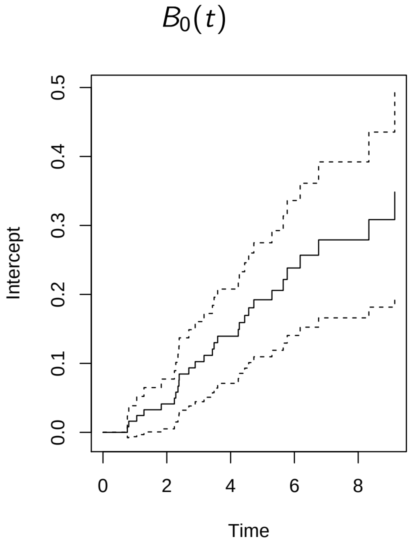
$$\mathbf{V} = \mathbf{U}\boldsymbol{\delta} + \boldsymbol{\epsilon}$$

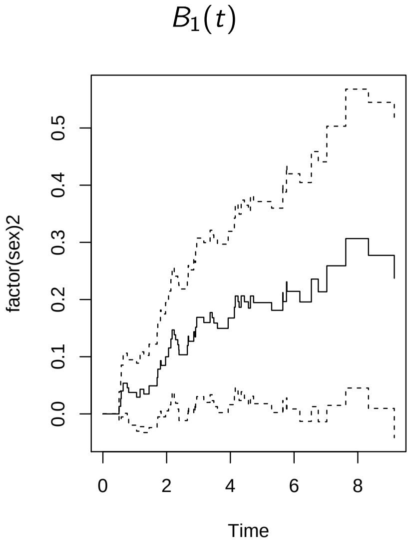
The least squares solution for $d\mathbf{B}(t^*) \equiv \boldsymbol{\delta}$ is hence (well-known!)

$$d\hat{\mathbf{B}}(t^*) = \hat{\boldsymbol{\delta}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{V}$$

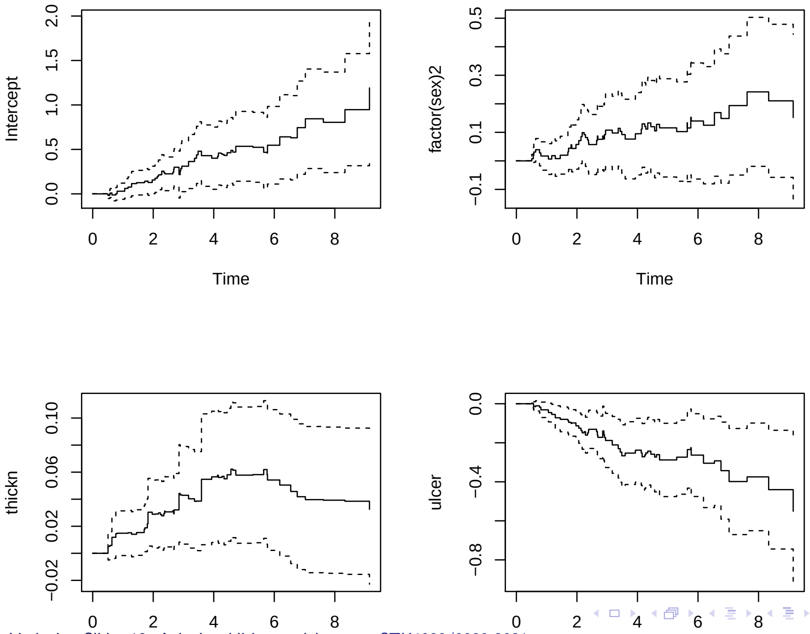
leading to $\qquad \hat{\mathbf{B}}(t) = \sum_{T_j \leq t} d\hat{\mathbf{B}}(T_j).$

Example: Melanoma-data with `sex` as only covariate

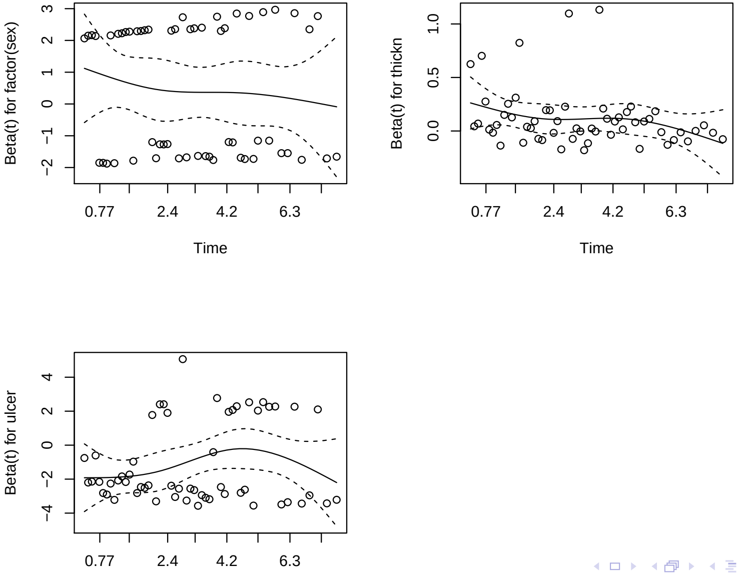




Melanoma-data with `sex`, `thickn`, `ulcer`



For comparison:
Scaled Schoenfeld residuals via Cox-regression



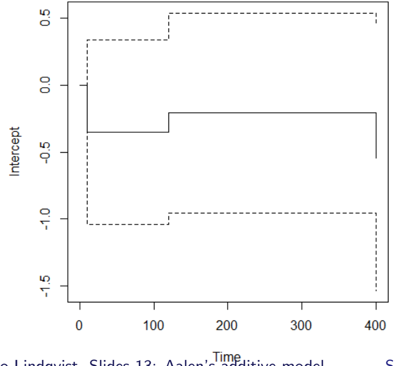
R-code for previous plots

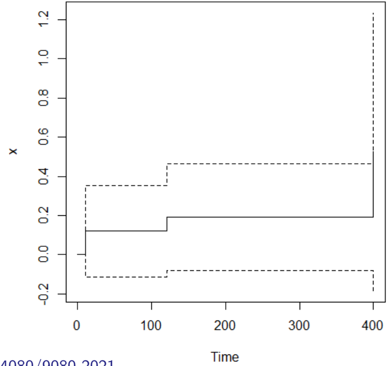
```
# Read data:
path="http://www.uio.no/studier/emner/matnat/math/STK4080/h14/
melanoma.txt"
melanoma=read.table(path,header=T)
# With sex as the only covariate:
fit.s.aal=aareg(Surv(lifetime,status==1)∼sex,data=melanoma)
par(mfrow=c(1,2))
plot(fit.s.aal)
# Model with sex, thickness and ulceration:
fit.stu.aal=aareg(Surv(lifetime,status==1)∼sex
+ thickn+ulcer, data=melanoma)
par(mfrow=c(2,2))
plot(fit.stu.aal)
# Comparison with Schoenfeld residuals
fit.stu.cox =
coxph(Surv(lifetime,status==1) sex+thickn+ulcer,data=melanoma)
plot(cox.zph(fit.stu.cox))
```

Simple example for handcalculation

i	1	2	3	4	5	6	7
$\tilde{T}_i$	5	10	40	80	120	400	600
x_i	12	10	3	5	3	4	1
D_i	0	1	0	0	1	1	0

The following plot is from using R – see calculations next pages





Simple example for hand calculation

i	1	2	3	4	5	6	7
$\tilde{T}_i$	5	10	40	80	120	400	600
x_i	12	10	3	5	3	4	1
D_i	0	1	0	0	1	1	0

At $t^* = 10$:

$$\begin{aligned} dN_i(10) &= Y_i(10)dB_0(10) + Y_i(10)x_i dB_1(10) + dM_i(10) \\ V_i &= u_{i1}\delta_0 + u_{i2}\delta_1 + \epsilon_i \end{aligned}$$

$$\mathbf{V} = \mathbf{U} \quad \boldsymbol{\delta} + \boldsymbol{\epsilon}$$

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 10 \\ 1 & 3 \\ 1 & 5 \\ 1 & 3 \\ 1 & 4 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \delta_0 \\ \delta_1 \end{pmatrix} + \boldsymbol{\epsilon}$$

Hence incr. at $t^* = 10$: $d\hat{\mathbf{B}}(10) = \hat{\boldsymbol{\delta}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{V} = \begin{pmatrix} -0.3521 \\ 0.1197 \end{pmatrix}$

Simple example for hand calculation

i	1	2	3	4	5	6	7
$\tilde{T}_i$	5	10	40	80	120	400	600
x_i	12	10	3	5	3	4	1
D_i	0	1	0	0	1	1	0

$$t^* = 120: dN_i(120) = Y_i(120)dB_0(120) + Y_i(120)x_i dB_1(120) + dM_i(120)$$

$$V_i = u_{i1}\delta_0 + u_{i2}\delta_1 + \epsilon_i$$

$$\mathbf{V} = \mathbf{U} \quad \boldsymbol{\delta} + \boldsymbol{\epsilon}$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 3 \\ 1 & 4 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \delta_0 \\ \delta_1 \end{pmatrix} + \boldsymbol{\epsilon}$$

Hence incr. at $t^* = 120$: $d\hat{\mathbf{B}}(120) = \hat{\boldsymbol{\delta}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{V} = \begin{pmatrix} 0.1429 \\ 0.0714 \end{pmatrix}$

Simple example for hand calculation

i	1	2	3	4	5	6	7
$\tilde{T}_i$	5	10	40	80	120	400	600
x_i	12	10	3	5	3	4	1
D_i	0	1	0	0	1	1	0

$$
\begin{aligned}
t^* = 400 : dN_i(400) &= Y_i(400)dB_0(400) + Y_i(400)x_i dB_1(400) + dM_i(400) \\
V_i &= u_{i1}\delta_0 + u_{i2}\delta_1 + \epsilon_i \\
\mathbf{V} &= \quad \mathbf{U} \quad \boldsymbol{\delta} + \boldsymbol{\epsilon}
\end{aligned}
$$

$$
\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 4 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \delta_0 \\ \delta_1 \end{pmatrix} + \boldsymbol{\epsilon}
$$

Hence incr. at $t^* = 400$: $\quad d\hat{\mathbf{B}}(400) = \hat{\boldsymbol{\delta}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{V} = \begin{pmatrix} -0.3333 \\ 0.3333 \end{pmatrix}$

Simple example for hand calculation using R ...

```
> fit.aal=aareg(Surv(Time,Status==1)~x, nmin=2,data=coxdata)
> par(mfrow=c(1,2))
> plot(fit.aal)
> print(fit.aal)
Call:
aareg(formula = Surv(Time, Status == 1) ~ x, data = coxdata,
    nmin = 2)

  n= 7
    3 out of 3 unique event times used

             slope   coef se(coef)     z     p
Intercept -0.00391 -0.255    0.230 -1.11 0.268
x          0.00477  0.133    0.104  1.28 0.201

Chisq=1.63 on 1 df, p=0.201; test weights=aalen
```

Note that the inclusion of `nmin=2` *was necessary in the use of* `aareg`*. This is because the default is* `nmin=3`*, which requires an 'at risk' number of at least 3 times the dimension of x. At $t = 400$, there are just 2 at risk in these data.*

Vector-valued counting processes, martingales, and stochastic integrals (cf. appendix B)

Consider first a univariate counting process martingale

$$M(t) = N(t) - \int_0^t \lambda(s)ds$$

The stochastic integral $\int_0^t H(s)dM(s)$ is a mean zero martingale with predictable variation processes:

$$\left\langle \int HdM \right\rangle (t) = \int_0^t H(s)^2 \lambda(s)ds$$

and optional variation processes:

$$\left[\int HdM \right] (t) = \int_0^t H(s)^2 dN(s)$$

By the martingale central limit theorem, a sequence of stochastic integrals converge in distribution to a Gaussian martingale (when properly normalized)

k-variate counting process

Now consider a k-variate counting process:

$$\mathbf{N}(t) = (N_1(t), \ldots, N_k(t))^T$$

There are k univariate counting processes, where we assume that two or more component processes do not jump at the same time.

The intensity (given history) of the multivariate counting process is the corresponding collection of the univariate intensity processes:

$$\boldsymbol{\lambda}(t) = (\lambda_1(t), \ldots, \lambda_k(t))^T$$

The vector-valued martingale associated with the multivariate counting process is

$$\mathbf{M}(t) = \mathbf{N}(t) - \int_0^t \boldsymbol{\lambda}(u)du$$

Multivariate stochastic integrals

For a $p \times k$ matrix $\mathbf{H}(u)$ of predictable processes, we define the p-variate vector-valued stochastic integral

$$\int_0^t \mathbf{H}(u) d\mathbf{M}(u)$$

The hth element of this vector is a sum of stochastic integrals:

$$\sum_{j=1}^{k} \int_0^t H_{hj}(u) dM_j(u)$$

The predictable variation process of $\int_0^t \mathbf{H}(u) d\mathbf{M}(u)$ is the $p \times p$ matrix:

$$\left\langle \int \mathbf{H} d\mathbf{M} \right\rangle (t) = \int_0^t \mathbf{H}(u) \mathrm{diag}\{\boldsymbol{\lambda}(u) du\} \mathbf{H}(u)^T$$

while the optional variation process is given by:

$$\left[\int \mathbf{H} d\mathbf{M} \right] (t) = \int_0^t \mathbf{H}(u) \mathrm{diag}\{d\mathbf{N}(u)\} \mathbf{H}(u)^T$$

Multivariate martingale central limit theorem

Consider a sequence of counting process martingales indexed by n (typically the number of individuals):

$$\mathbf{M}^{(n)}(t) = \mathbf{N}^{(n)}(t) - \int_0^t \boldsymbol{\lambda}^{(n)}(u)du$$

and a sequence of stochastic integrals

$$\int_0^t \mathbf{H}^{(n)}(u)d\mathbf{M}^{(n)}(u)$$

where the predictable processes $\mathbf{H}^{(n)}(t)$ have dimension $p \times k_n$.

Multivariate martingale central limit theorem

Let $\mathbf{V}(t) = \mathrm{E}\{\mathbf{U}(t)\mathbf{U}(t)^T\}$ be the covariance matrix for a p-variate mean zero Gaussian martingale $\mathbf{U}(t)$.

Provided that

$$\int_0^t \mathbf{H}^{(n)}(u)\mathrm{diag}\{\boldsymbol{\lambda}^{(n)}(u)du\}\mathbf{H}^{(n)}(u)^T \to \mathbf{V}(t)$$

and the "jumps disappear in the limit", we have that the p-variate stochastic process

$$\int_0^t \mathbf{H}^{(n)}(u)d\mathbf{M}^{(n)}(u)$$

converges in distribution to the stochastic process $\mathbf{U}(t)$.

In particular for a given value of t we have that $\int_0^t \mathbf{H}^{(n)}(u)d\mathbf{M}^{(n)}(u)$ is approximately multivariate normal.

The additive regression model – theory

We introduce the vectors

$$\begin{aligned}
\mathbf{N}(t) &= (N_1(t), N_2(t), \ldots, N_n(t))^T \\
\mathbf{M}(t) &= (M_1(t), M_2(t), \ldots, M_n(t))^T \\
\mathbf{B}(t) &= (B_0(t), B_1(t), \ldots, B_p(t))^T
\end{aligned}$$

and the $n \times (p+1)$ "design matrix"

$$\mathbf{X}(t) = \begin{pmatrix} Y_1(t) & Y_1(t)x_{11}(t) & \cdots & Y_1(t)x_{1p}(t) \\ Y_2(t) & Y_2(t)x_{21}(t) & \cdots & Y_2(t)x_{2p}(t) \\ \vdots & \vdots & \ddots & \vdots \\ Y_n(t) & Y_n(t)x_{n1}(t) & \cdots & Y_n(t)x_{np}(t) \end{pmatrix}$$

The additive regression model – theory

The additive regression model may be written on matrix form as

$$d\mathbf{N}(t) = \mathbf{X}(t)d\mathbf{B}(t) + d\mathbf{M}(t)$$

For each time t, this is a linear regression model on matrix form (conditional on the past).

Ordinary least squares gives

$$d\hat{\mathbf{B}}(t) = (\mathbf{X}(t)^T\mathbf{X}(t))^{-1}\mathbf{X}(t)^T d\mathbf{N}(t)$$

provided $\mathbf{X}(t)$ has full rank.

The estimator $\hat{\mathbf{B}}(t)$

Introduce the indicator:

$$J(t) = I\{\mathbf{X}(t) \text{ has full rank}\}$$

and the least squares generalized inverse

$$\mathbf{X}^{-}(t) = (\mathbf{X}(t)^T\mathbf{X}(t))^{-1}\mathbf{X}(t)^T$$

Then

$$\begin{aligned}
\hat{\mathbf{B}}(t) &= \int_0^t J(u)\mathbf{X}^{-}(u)d\mathbf{N}(u) \\
&= \sum_{T_j \le t} J(T_j)\mathbf{X}^{-}(T_j)\Delta\mathbf{N}(T_j)
\end{aligned}$$

where $T_1 < T_2 < \cdots$ are the event times.

Expected value of $\hat{\mathbf{B}}(t)$

To study the statistical properties of the estimator in the additive model recall that

$$\hat{\mathbf{B}}(t) = \int_0^t J(u)\mathbf{X}^-(u)d\mathbf{N}(u)$$

Here $d\mathbf{N}(u) = \mathbf{X}(u)d\mathbf{B}(u) + d\mathbf{M}(u)$, so

$$\begin{aligned}\hat{\mathbf{B}}(t) &= \int_0^t J(u)d\mathbf{B}(u) + \int_0^t J(u)\mathbf{X}^-(u)d\mathbf{M}(u) \\ &\equiv \mathbf{B}^*(t) + \int_0^t J(u)\mathbf{X}^-(u)d\mathbf{M}(u)\end{aligned}$$

Thus

$$\hat{\mathbf{B}}(t) - \mathbf{B}^*(t) = \int_0^t J(u)\mathbf{X}^-(u)d\mathbf{M}(u)$$

so

$$\mathrm{E}\{\hat{\mathbf{B}}(t) - \mathbf{B}^*(t)\} = 0$$

Covariance matrix of $\hat{\mathbf{B}}(t)$

We have

$$\hat{\mathbf{B}}(t) - \mathbf{B}^*(t) = \int_0^t J(u)\mathbf{X}^-(u)d\mathbf{M}(u)$$

Thus

$$\left\langle \hat{\mathbf{B}} - \mathbf{B}^* \right\rangle (t) = \int_0^t J(u)\mathbf{X}^-(u)\text{diag}\{\boldsymbol{\lambda}(u)du\}\mathbf{X}^-(u)^T$$

$$\left[\hat{\mathbf{B}} - \mathbf{B}^* \right] (t) = \int_0^t J(u)\mathbf{X}^-(u)\text{diag}\{d\mathbf{N}(u)\}\mathbf{X}^-(u)^T$$

We may estimate the covariance matrix of $\hat{\mathbf{B}}(t)$ either by inserting an estimate

$$\widehat{\boldsymbol{\lambda}(u)du} = \mathbf{X}(u)d\hat{\mathbf{B}}(u)$$

for $\boldsymbol{\lambda}(u)du$ in the predictable variation,

... or use the optional variation (the choice in R).

Estimated covariance matrix and confidence interval

This leads to the following estimators of the covariance matrix of $\hat{\mathbf{B}}(t)$:

$$\begin{aligned}
\hat{\Sigma}(t) &= \sum_{T_j \leq t} J(T_j)\mathbf{X}^{-}(T_j)\text{diag}\{\Delta\mathbf{N}(T_j)\}\mathbf{X}^{-}(T_j)^T \\
\tilde{\Sigma}(t) &= \sum_{T_j \leq t} J(T_j)\mathbf{X}^{-}(T_j)\text{diag}\{\mathbf{X}(T_j)\Delta\hat{\mathbf{B}}(T_j)\}\mathbf{X}^{-}(T_j)^T
\end{aligned}$$

where the first option is the one used in R

Martingale central limit theorem gives that $\hat{\mathbf{B}}(t)$ is approximately multivariate normally distributed.

Confidence intervals (included in earlier plots)

$$\hat{B}_q(t) \pm z_{1-\alpha}\sqrt{\hat{\sigma}_{qq}(t)}$$

where $\hat{\sigma}_{qq}(t)$ is the qth diagonal element of $\hat{\Sigma}(t)$.

Testing in the additive regression model

We want to test the null hypothesis

$$H_0: \ \beta_q(t) = 0 \text{ for all } t \in (0, t_0]$$

for a given t_0.

We may base a test on the stochastic integral

$$Z_q(t_0) = \int_0^{t_0} L_q(t) d\hat{\mathbf{B}}_q(t) = \sum_{T_j \le t_0} L_q(T_j) \Delta \hat{\mathbf{B}}_q(T_j)$$

where $L_q(t)$ is a predictable non-negative process (weight function).

It can be shown that $Z_q(t_0)$ is a mean zero martingale under H_0.

Testing in the additive regression model

Predictable variation process

$$\langle Z_q \rangle (t_0) = \int_0^{t_0} L_q^2(t) d \left\langle \hat{B}_q \right\rangle (t)$$

Variance estimator

$$\begin{aligned} V_{qq}(t_0) &= \int_0^{t_0} L_q^2(t) d\hat{\sigma}_{qq}(t) \\ &= \sum_{T_j \le t_0} L_q^2(T_j) \Delta \hat{\sigma}_{qq}(T_j) \end{aligned}$$

$\hat{\sigma}_{qq}(t)$ is the qth diagonal element in the estimator of the covariance matrix of $\hat{\mathbf{B}}(t)$.

The test statistic

Using the martingale central limit theorem we may show that

$$\frac{Z_q(t_0)}{\sqrt{V_{qq}(t_0)}}$$

is approximately standard normally distributed under H_0.

A possible choice of weight process $L_q(t)$ may be based on the matrix

$$\mathbf{K}(t) = \left\{ \text{diag}\left[(\mathbf{X}(t)^T \mathbf{X}(t))^{-1} \right] \right\}^{-1}$$

If we chose $L_q(t)$ as the qth diagonal element of this matrix we get a "logrank type" test. (*Motivation:* in ordinary least squares, the variances of the estimators are proportional to the diagonal elements of $(\mathbf{X}^T\mathbf{X})^{-1}$).

For a model with a single binary covariate, we obtain exactly the logrank test if we use the estimator $\tilde{\Sigma}(t)$ (see earlier slide) for estimating variances (Exercise 4.5).

Example: Testing in melanoma data

Consider the model with `sex`, `cthick` (centered thickness) and `ulcer`:

```
fit.stu=aareg(Surv(lifetime,status==1)∼sex +cthick
+ulcer, data=melanoma)
print(fit.stu)
```

```
Call:
aareg(formula = Surv(lifetime, status == 1) ~ factor(sex) + cthick +
    factor(ulcer), data = melanoma)

  n= 205
    57 out of 57 unique event times used

                  slope     coef se(coef)     z        p
Intercept        0.1240  0.01010 0.001920  5.27 1.35e-07
factor(sex)2     0.0412  0.00299 0.001930  1.55 1.21e-01
cthick           0.0229  0.00117 0.000535  2.19 2.89e-02
factor(ulcer)2 -0.0890 -0.00706 0.002100 -3.37 7.49e-04

Chisq=25.96 on 3 df, p=9.7e-06; test weights=aalen
```

The columns `z` and `p` give the test statistics and *p*-values. `slope` is a kind of average slope in the plots, while `coef` is proportional to `z`.

Example: Testing in melanoma data (cont.)

```
> summary(fit.stu)

$test.statistic
    Intercept     factor(sex)2        cthick     factor(ulcer)2
    25.042901         5.792022     57.099605         -12.267495
```

This gives the test statistics $Z_q(t_0)$

```
$test.var
           b0
b0  22.563202    -5.00541104    3.042205    -14.72388974
    -5.005411    13.96642652   -4.550560      0.01291638
     3.042205    -4.55056035  682.628937     29.72199395
   -14.723890     0.01291638   29.721994     13.24414379
```

This diagonal of the matrix give the estimated variances $V_{qq}(t_0)$