Survival and longitudinal data analysis
Chapter 3: Counting processes and martingales

Agathe Guilloux
Professeure au LaMME - Université d'Évry - Paris Saclay

One example

Marketing: Monetization for free-to-play games

- Times of monetization for players until their giving-ups
- Several hours of game-play history for ~1MM players

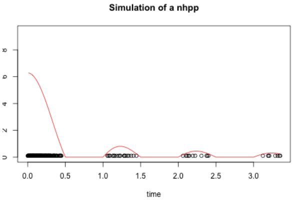
Large number of observations (individuals), time-dependent covariates

$$(n, p) \to \Big(n, p, D\Big)$$

Counting processes and intensity function

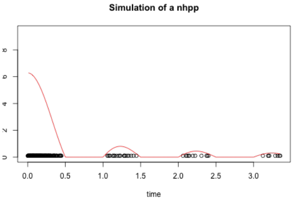
Time(s) to event(s) data

What are we observing ?



Time(s) to event(s) data

What are we observing ?



The higher the intensity, the more points we observe :

$$\lambda^\star(t) = \text{infinitesimal } \mathbb{P}(\text{event} \in [t, t + dt[)$$

Counting process and intensity

Definition

A **counting process** is a stochastic process $N^\star : N^\star(t), t \geq 0$

- with $N^\star(0) = 0$ and $N^\star(t) < \infty$,
- whose sample paths a right-continuous and piecewise constant with jump of size $+1$,
- such that $N^\star(t) =$ number of observed events in $[0, t]$.

To the counting process $N^\star$ is associated the history (natural filtration) $\mathcal{F}^\star = \Big\{\mathcal{F}^\star_t = \sigma\Big(N^\star(s), s \leq t\Big), t \geq 0\Big\}$.

Its **intensity** $\lambda^\star$ is the function defined via

$$\lambda^\star(t)dt = \mathbb{P}(dN^\star(t) = 1|\mathcal{F}^\star_{t-})$$

where $dN^\star(t) = N^\star([t, t + dt[) = N^\star(t + dt-) - N^\star(t-)$.

The Poisson process

The counting process $N^\star$ with intensity $\lambda^\star$ is a Poisson process if $N^\star(]a,b])$ is independent of $N^\star(]c,d])$ for all $a \leq b \leq c \leq d$.

Properties

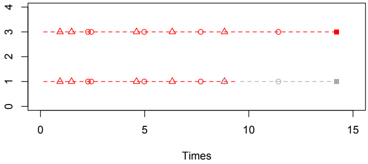
In this case,

- $N^*(t) \sim \mathcal{P}\left(\int_0^t \lambda^\star(s) ds \right)$
- Conditionally to $N^*(t) = n$, the arrival times $T_{(1)} < T_{(2)} < \ldots < T_{(n)} \leq t$ are distributed as the ordered statistics of $T_1, T_2, \ldots, T_n$ i.i.d. with density $\lambda^* / \int_0^t \lambda^\star(s)$ on $[0,t]$.

Censoring

Example: censoring

We observe $N^\star$ only until a censoring C occurs.



Marketing: Monetization for free-to-play games

Times of monetization for players until their **giving-ups**

Filtration/censoring

Definition

Consider an adapted and left continuous process Y (with values in $\{0, 1\}$) and construct the **filtered process**

$$N(t) = \int_0^t Y(s) dN^\star(s)$$

and the new history $\mathcal{F} = \Big\{\mathcal{F}_t = \sigma\big(N(s), Y(s), s \leq t\big), t \geq 0\Big\}$. With the assumptions:

- N is a counting process,
- as $Y(t-) = Y(t))$, $Y(t)$ is $\mathcal{F}_{t-}$-measurable (it is said to be predictable).

Intensity of a filtered counting process

Intensity of a filtered counting process

More generally the intensity of the process N defined as

$$N(t) = \int_0^t Y(s) dN^\star(s)$$

is

$$Y(s)\lambda^\star(s)$$

where $\lambda^\star$ is the intensity of $N^\star$.

Definition of the cumulative intensity

The cumulative intensity is defined as the function Λ

$$\Lambda(t) = \int_0^t \lambda(s) ds.$$

One special case of filtering: at most one event

Let $N^\star$ be a Poisson process with intensity $\lambda^\star$ and construct

$$\tilde{N}_1(t) = \int_0^t \mathbb{1}_{N^\star(t-)<1} dN^\star(t).$$

for all $t \geq 0$ and define

$$\tilde{N}_2(t) = \mathbb{1}_{T \leq t}$$

where T has the hazard rate $\lambda^\star$.

We have:

$$\mathbb{P}(d\tilde{N}_1(t) = 1 | \mathcal{F}^\star_{t-}) = \mathbb{1}_{N^\star(t-)<1} \lambda^\star(t) dt = \mathbb{1}_{T_{(1)} \geq t} \lambda^\star(t) dt \text{ and}$$

$$\mathbb{P}(d\tilde{N}_2(t) = 1 | \mathcal{F}^\star_{t-}) = \mathbb{1}_{T \geq t} \frac{f^\star(t)}{\bar{F}^\star(t)} dt = \mathbb{1}_{T \geq t} \lambda^\star(t) dt$$

so $\tilde{N}_1 \overset{\mathcal{D}}{=} \tilde{N}_2$.

In particular:

$$\mathbb{P}(T \geq t) = \mathbb{P}(\tilde{N}_2([0,t]) = 0) = \mathbb{P}(N^\star([0,t]) = 0) = \exp\Big(- \int_0^t \lambda^*(s) ds \Big).$$

An other special case: at most one event and censoring

Let

- T be a time of interest
- C a censoring time independent of T

We observe

$$T^C = T \wedge C \text{ and } \delta = \mathbb{1}_{T \leq C}.$$

In terms of counting processes, this is equivalent to observing

$$N(t) = \mathbb{1}_{T^C \leq t, \delta=1} \text{ and } Y(t) = \mathbb{1}_{T^C \geq t}.$$

We can write

$$N(t) = \int_0^t Y(s) dN^\star(s)$$

so it is a filtered process and its intensity is given by

$$\mathbb{1}_{T \geq t} \mathbb{1}_{C \geq t} \frac{f^\star(t)}{\bar{F}^\star(t)}$$

Covariates

Stanford Heart Transplant data (**kalbfleisch2011statistical**)

Survival of patients on the waiting list for the Stanford heart transplant program.

- fustat: dead or alive
- surgery: prior bypass surgery
- age: age (in years)
- futime: follow-up time
- wait.time: time before transplant
- transplant: transplant indicator
- accept.yr: acceptance into program

```
##   fustat surgery      age futime wait.time transplant accept.yr
## 1      1       0 30.84463     49        NA          0      1967
## 2      1       0 51.83573      5        NA          0      1968
## 3      1       0 54.29706     15         0          1      1968
## 4      1       0 40.26283     38        35          1      1968
## 5      1       0 20.78576     17        NA          0      1968
## 6      1       0 54.59548      2        NA          0      1968
```

Stanford Heart Transplant data (**kalbfleisch2011statistical)**

Survival of patients on the waiting list for the Stanford heart transplant program.

- fustat: dead or alive $\rightarrow \delta$
- surgery: prior bypass surgery
- age: age (in years)
- futime: follow-up time $\rightarrow T^C$
- wait.time: time before transplant
- transplant: transplant indicator
- accept.yr: acceptance into program

```
##   fustat surgery      age futime wait.time transplant accept.yr
## 1      1       0 30.84463     49        NA          0      1967
## 2      1       0 51.83573      5        NA          0      1968
## 3      1       0 54.29706     15         0          1      1968
## 4      1       0 40.26283     38        35          1      1968
## 5      1       0 20.78576     17        NA          0      1968
## 6      1       0 54.59548      2        NA          0      1968
```

Stanford Heart Transplant data (**kalbfleisch2011statistical)**

Survival of patients on the waiting list for the Stanford heart transplant program.

- fustat: dead or alive $\rightarrow \delta$
- surgery: prior bypass surgery $\rightarrow$ **time independent covariate**
- age: age (in years) $\rightarrow$ **time independent covariate**
- futime: follow-up time $\rightarrow T^C$
- wait.time: time before transplant
- transplant: transplant indicator
- accept.yr: acceptance into program $\rightarrow$ **time independent covariate**

```
##   fustat surgery      age futime wait.time transplant accept.yr
## 1      1       0 30.84463     49        NA          0      1967
## 2      1       0 51.83573      5        NA          0      1968
## 3      1       0 54.29706     15         0          1      1968
## 4      1       0 40.26283     38        35          1      1968
## 5      1       0 20.78576     17        NA          0      1968
## 6      1       0 54.59548      2        NA          0      1968
```

Stanford Heart Transplant data (**kalbfleisch2011statistical)**

Survival of patients on the waiting list for the Stanford heart transplant program.

- fustat: dead or alive $\rightarrow \delta$
- surgery: prior bypass surgery $\rightarrow$ **time independent covariate**
- age: age (in years) $\rightarrow$ **time independent covariate**
- futime: follow-up time $\rightarrow T^C$
- wait.time: time before transplant $\rightarrow$ **time dependent covariate**
- transplant: transplant indicator $\rightarrow$ **time dependent covariate**
- accept.yr: acceptance into program $\rightarrow$ **time independent covariate**

```
##   fustat surgery      age futime wait.time transplant accept.yr
## 1      1       0 30.84463     49        NA          0      1967
## 2      1       0 51.83573      5        NA          0      1968
## 3      1       0 54.29706     15         0          1      1968
## 4      1       0 40.26283     38        35          1      1968
## 5      1       0 20.78576     17        NA          0      1968
## 6      1       0 54.59548      2        NA          0      1968
```

Cox model for the intensity with time-varying covariates

When the covariates are not constant over time, we want the intensity to depend on the covariates at time t

$$\lambda^\star(t) \to \lambda^\star(t, X(t)).$$

The Cox model

The Cox 1972 model for the intensity of a counting process assumes that its intensity has the form

$$\lambda^*(t) = \lambda_0^\star(t) \exp(X(t)\beta^\star).$$

Example with time independent covariates

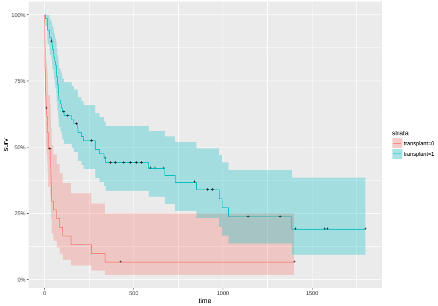
```
coxph(Surv(futime,fustat) ~ accept.yr + surgery + age, data = jasa)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ accept.yr + surgery +
##     age, data = jasa)
##
##              coef exp(coef) se(coef)     z     p
## accept.yr -0.1320    0.8764   0.0681 -1.94 0.053
## surgery   -0.6427    0.5259   0.3673 -1.75 0.080
## age        0.0276    1.0280   0.0134  2.06 0.039
##
## Likelihood ratio test=14.5  on 3 df, p=0.00226
## n= 103, number of events= 75
```

Example with time dependent covariates: false model

- transplant: transplant indicator → **time dependent covariate**

```
autoplot(survfit(Surv(futime,fustat) ~transplant , data = jasa))
```



"The key rule for time dependent covariates in a Cox model is simple and essentially the same as that for gambling: *you cannot look into the future*."
Therneau, Crowson, and Atkinson 2017

Example with time dependent covariates: false model (2)

```
coxph(Surv(futime,fustat) ~ surgery + transplant + age , data = jasa)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ surgery + transplant +
##     age, data = jasa)
##
##               coef exp(coef) se(coef)     z       p
## surgery    -0.4190    0.6577   0.3712 -1.13    0.26
## transplant -1.7171    0.1796   0.2785 -6.16 7.1e-10
## age         0.0589    1.0607   0.0150  3.91 9.1e-05
##
## Likelihood ratio test=45.9  on 3 df, p=6.11e-10
## n= 103, number of events= 75
```

A new format for time dependent covariates: start-stop

```
##     id start stop event transplant        age      year surgery
##     1     0   49     1          0 -17.155373 0.1232033       0
##     2     0    5     1          0   3.835729 0.2546201       0
##     3     0   15     1          1   6.297057 0.2655715       0
##     4     0   35     0          0  -7.737166 0.4900753       0
##     4    35   38     1          1  -7.737166 0.4900753       0
##     5     0   17     1          0 -27.214237 0.6078029       0
```

Notice that for individual 4, we have

- with the old format

```
##   fustat        age futime wait.time transplant
## 4      1   40.26283     38        35          1
```

- with the new format

```
##     id start stop event transplant
##     4     0   35     0          0
##     4    35   38     1          1
```

A new format for time dependent covariates: start-stop (2)

- False model

```
## coxph(formula = Surv(futime, fustat) ~ surgery + transplant +
##     age, data = jasa)
##
##                coef exp(coef) se(coef)     z       p
## surgery    -0.4190    0.6577   0.3712 -1.13    0.26
## transplant -1.7171    0.1796   0.2785 -6.16 7.1e-10
## age         0.0589    1.0607   0.0150  3.91 9.1e-05
```

- Start-stop model

```
## coxph(formula = Surv(start, stop, event) ~ age + surgery +
##      transplant, data = jasa1)
##
##                coef exp(coef) se(coef)     z     p
## age         0.0306    1.0310   0.0139  2.20 0.028
## surgery    -0.7733    0.4615   0.3597 -2.15 0.032
## transplant  0.0141    1.0142   0.3082  0.05 0.964
```

Estimation

The data

We observe for $i = 1, \ldots, n$ i.i.d.

$$\Big(X_i(s)Y_i(s), N_i(s), Y_i(s), s \leq \tau\Big)$$

and we want to learn the influence of X on $t \mapsto \lambda^*(t, X(t))$.

The log-likelihood

In the counting processes setting, the log-likelihood (times $1/n$) is defined as

$$\frac{1}{n}\sum_{i=1}^{n}\{\sum_{T_{i,k}}\delta_{i,k}\log(\lambda(t, X_i(T_{i,k}))) - \int_{[0,\tau]} Y_i(t)\lambda(t, X_i(t))dt\}$$

To ease the notation, I'll consider that each individual has a most one event

$$\frac{1}{n}\sum_{i=1}^{n}\{\delta_i\log(\lambda(t, X_i(T_i^C))) - \int_{[0,\tau]} Y_i(t)\lambda(t, X_i(t))dt\}$$

Partial log-likelihood

In the Cox model,

$$\lambda^*(t) = \lambda_0^\star(t)\exp(X(t)\beta^\star),$$

we can estimate $\beta^\star$ only with the partial likelihood (that's what `coxph` does). In the case where the individuals experience (at most) one event, it writes:

$$\begin{aligned}\ell_n^P(\beta) &= \frac{1}{n}\sum_{i=1}^{n}\delta_i \log \frac{\exp(X_i(T_i^C)\beta)}{\frac{1}{n}\sum_{j:T_j^C \geq T_i^C}\exp(X_j(T_i^C)\beta)} \\ &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\Big\{X_i(T_i^C)\beta - \log\Big(\sum_{j:T_j^C\geq T_i^C}\exp(X_j(T_i^C)\beta)\Big)\Big\}.\end{aligned}$$

Model selection

Moderate p

AIC/BIC criteria

For the Cox model, the AIC and BIC criteria are defined as

$$AIC(\beta) = -2\ell_n^P(\beta) + 2\frac{|\beta|_0}{n}$$
$$BIC(\beta) = -2\ell_n^P(\beta) + \log(n)\frac{|\beta|_0}{n}$$

and choose the model which meets the minimum of the AIC (or BIC) criterion.

Large p

When p grows, one can consider to add a lasso penalty:

$$\ell_n^P(\beta) + \gamma \sum_{j=1}^{p} |\beta_j|$$

or an elastic-net penalty

$$\ell_n^P(\beta) + \gamma\Big(\alpha \sum_{j=1}^{p} |\beta_j| + \frac{1-\alpha}{2} \sum_{j=1}^{p} |\beta_j|^2\Big).$$

```
data("nki70")
model_matrix = model.matrix( ~ as.factor(Grade)  + . - Grade - 1
                            , data = nki70[3:77])

X = model_matrix[,-1]

elasticnet_solution = cv.glmnet(X,Surv(nki70$time, nki70$event),
        family = "cox" , alpha = 0.5,
        penalty.factor = c(rep(0,6),rep(1,70)))

coef(elasticnet_solution)
```

Diagnosis in the Cox model

Beyond linearity

The key assumptions in the Cox model

$$\lambda^*(t) = \lambda_0^\star(t) \exp\left(X(t)\beta^\star\right) = \lambda_0^\star(t) \exp\Big(\sum_{j=1}^{p} X^j(t)\beta_j^\star\Big),$$

are

- $\beta^\star$ is time-independent
- each covariate has a linear effect (in the exponential).

they might be too strong. We need to test them (a least graphically).

The possible extensions are

- to introduce time-dependent coefficients $\beta^\star(t)$
- or to consider a non-parametric effect of the jth covariate, i.e. to replace the term $X^j\beta_j^\star$ by $f_j(X^j)$ (where f_j is a smooth function).

Check for linearity with martingales residuals

Martingale residuals

We know that

$$\mathbb{E}\Big(N_i(\infty)\Big) = \mathbb{E}\Big(\int_0^\infty Y_i(t)\lambda^\star(t)\exp(X_i(t)\beta^\star)dt\Big)$$

so we define the martingale residuals as

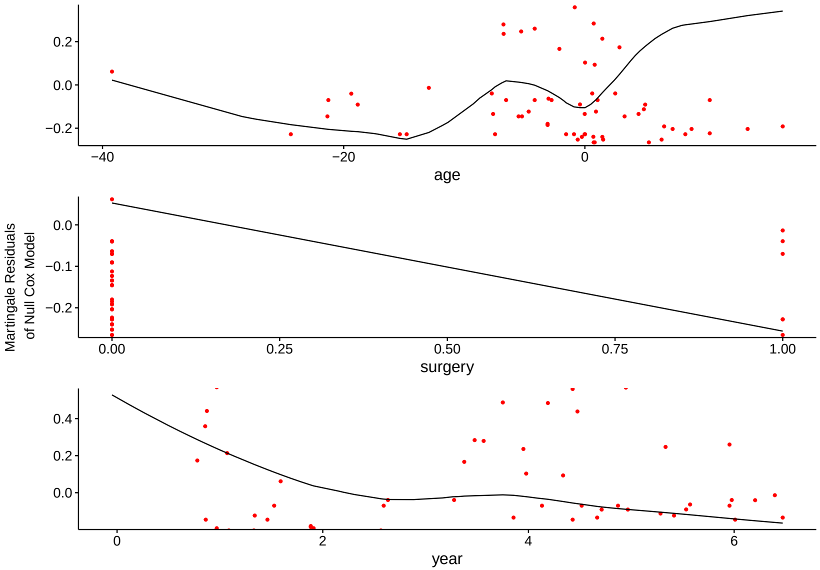
$$N_i(\infty) - \int_0^\infty Y_i(t)\exp(X_i(t)\hat{\beta})\hat{\lambda}_0(t,\hat{\beta})dt$$

To check if the hypothesis that a covariate has a linear effect, plot the martingale residuals against the values of the covariates.

Be careful: this has a sense only for continuous covariates !

Graphical test for $f_j(x) = X^j \beta_j^\star$

```r
library(survminer)
ggcoxfunctional(aic_model  ,data =jasa1)
```



A solution is to consider simple functions f_j (for example splines)

```
coxph(Surv(start, stop, event) ~ pspline(age) + surgery  ,data =jasa1)
```

```
## Call:
## coxph(formula = Surv(start, stop, event) ~ pspline(age) + surgery +
##     pspline(year), data = jasa1)
##
##                              coef se(coef)     se2   Chisq   DF       p
## pspline(age), linear   0.0270   0.0125  0.0123  4.6562 1.00 0.0309
## pspline(age), nonlin                             5.9196 3.00 0.1158
## surgery               -0.8293   0.4041  0.3970  4.2125 1.00 0.0401
## pspline(year), linear -0.1621   0.0700  0.0697  5.3677 1.00 0.0205
## pspline(year), nonlin                           12.2151 2.99 0.0066
##
## Iterations: 5 outer, 15 Newton-Raphson
##      Theta= 0.621
##      Theta= 0.661
## Degrees of freedom for terms= 4 1 4
## Likelihood ratio test=34.6  on 8.96 df, p=6.67e-05  n= 170
```

Check for time invariance via Schoenfeld residuals

From the gradient of the log-likelihood, we can define covariates specific residuals

Schoenfeld residuals (score residuals)

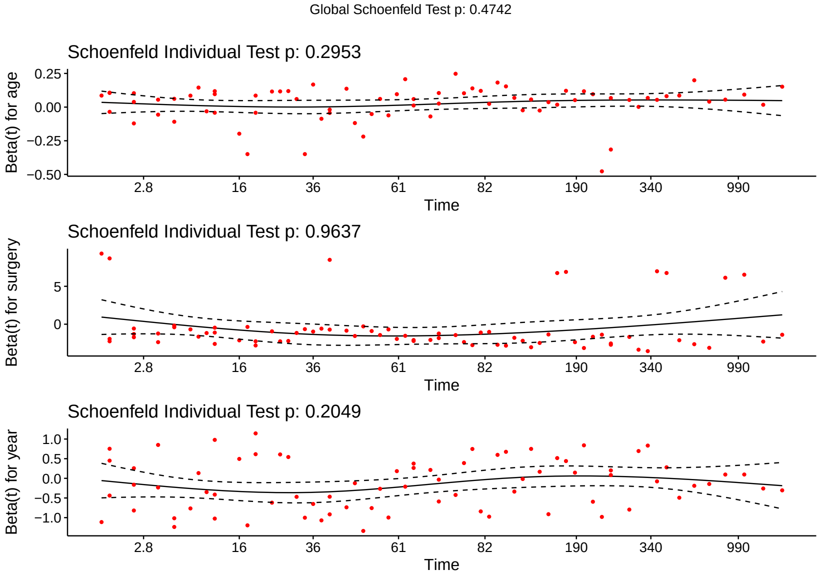
We define the Schoenfled residuals as

$$X_i^j(T_i^C) - \bar{X}^j(T_i^C) = X_i^j(T_i^C) - \frac{\sum_{k=1}^n Y_k(T_i^C) X_k(T_i^C) \exp\left(X_k \hat{\beta}\right)}{\sum_{k=1}^n Y_k(T_i^C) \exp\left(X_k \hat{\beta}\right)}.$$

To check if the hypothesis that a covariate has a constant coefficient, plot the (weighted) Schoenfeld residuals against time.

Test for $\beta_j^\star(t) = \beta_j^\star$

```r
library(survminer)
ggcoxzph(cox.zph(aic_model))
```



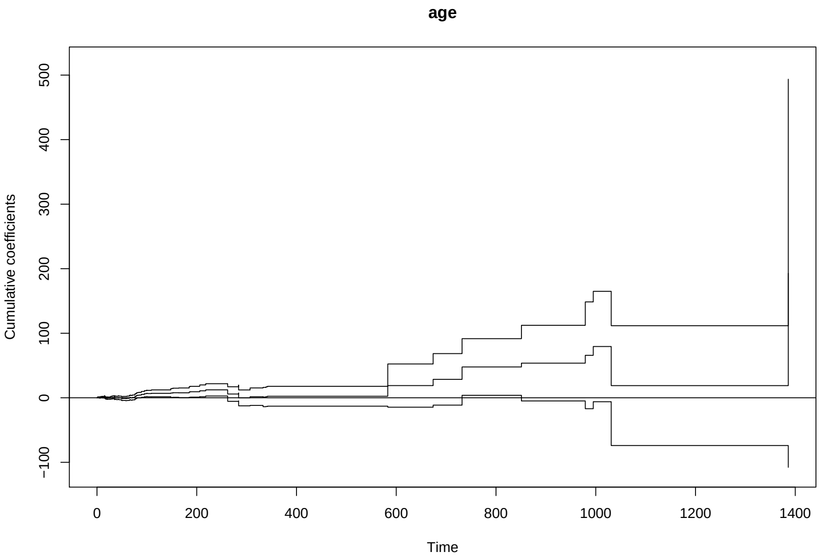
One solution with the timereg package

```
library(timereg)
model_timevarying  = timecox(Surv(start, stop, event) ~ age + surgery ,
    data =jasa1)
summary(model_timevarying)
```

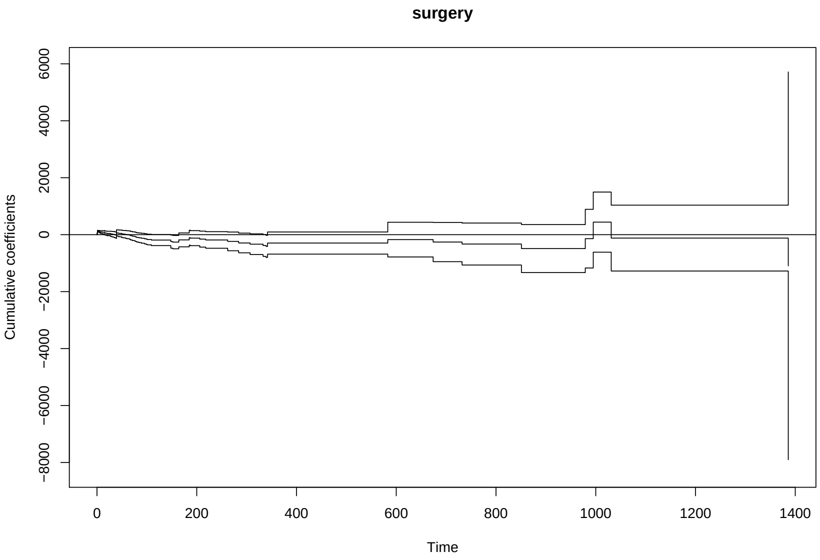
```
## Multiplicative Hazard Model
##
## Test for time invariant effects
##                   Kolmogorov-Smirnov test p-value H_0:constant effec
## (Intercept)                           665                       0.08
## age                                   125                       0.02
## surgery                              1230                       0.10
##                     Cramer von Mises test p-value H_0:constant effec
## (Intercept)                      1.28e+08                       0.15
## age                              3.45e+06                       0.10
## surgery                          1.29e+08                       0.48
##
```

One solution with the timereg package I

```
plot(model_timevarying)
```



One solution with the `timereg` package II



Predictions

Predictions from an adjusted Cox model

Once the regression parameters $\beta^\star$ of the Cox model have been estimated by $\hat{\beta}$, one can compute the Breslow estimator $\hat{\Lambda}_0$.

We get an estimator of the cumulated hazard/intensity function for a value X_+ of the covariates

$$\hat{\Lambda}(t|X_+) = \hat{\Lambda}_0(t) \exp(X_+\hat{\beta}), \text{ for all } t \geq 0.$$

In the case, where only (at most) one event is observed by individual, we derive for that an estimator of the survival function

$$\widehat{\bar{F}}(T|X_+) = \exp\Big(-\hat{\Lambda}(t|X_+)\Big) = \exp\Big(-\hat{\Lambda}_0(t) \exp(X_+\hat{\beta})\Big), \text{ for all } t \geq 0.$$

References I

David R. Cox. "Regression models and life tables (with discussion)". In: *Journal of the Royal Statistical Society* 34 (1972), pp. 187–220.

Terry Therneau, Cindy Crowson, and Elizabeth Atkinson. "Using time dependent covariates and time dependent coefficients in the cox model". In: *Survival Vignettes* (2017).