The Cox model

TS

Outline

- Cox's proportional hazards model.
  - Partial likelihood principle
- Goodness-Of-Fit for Cox's Regression Model.
  - practical implementation, different approaches
- Extensions and variations of Cox's Regression Model.
  - stratified Cox model
  - nested case control sampling
  - robust standard errors
  - delayed entry, left-truncation
- Important tools
  - Resampling

Survival Setting

Standard setup for right-censored survival data. IID copies of

$$(T, D)$$

where

$$T = T^* \wedge C \quad D = I(T^* \leq C)$$

with T^* being the true survival time and C the (potential) censoring time and possibly covariates $X_i(t)$.
Hazard-function:

$$\alpha(t) = \lim_{h \downarrow 0} \frac{1}{h} P(t \leq T^* < t + h | T^* \geq t).$$

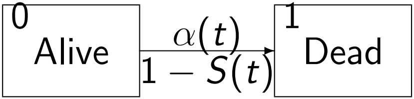
And given covariates

$$\alpha(t, X) = \lim_{h \downarrow 0} \frac{1}{h} P(t \leq T^* < t + h | T^* \geq t, X).$$

Importantly : we have independent right-censoring $C \perp T$ given X.

Modelling Event Histories

Simple life-death model, a so called 2-state model



Here, the hazard rate $\alpha(t)$ is the conditional probability of being in state 1 at time $t + \Delta t$ given state 0 just before time t is approximately $\alpha(t)\Delta t$ when $\Delta t > 0$ is small.

Survival model

For simple survival model we know how to do regression modelling for

$$\lambda(t) = \lambda_0(t) \exp(Z_1\beta_1 + \cdots + Z_p\beta_p) \tag{1}$$

when there is right censoring. Further and very importantly we can also estimate the survival probabilities in this model

$$S(t) = \exp(-\Lambda_0(t) \exp(Z_1\beta_1 + \cdots + Z_p\beta_p)) \tag{2}$$

that gives the relevant probabilities in this framework.

Counting process notation $N(t)$

- Useful notation and link to martingale theory
- Formulations often in terms of Counting processes

Counting process: $N_i(t) = I(T_i \leq t, D_i = 1)$

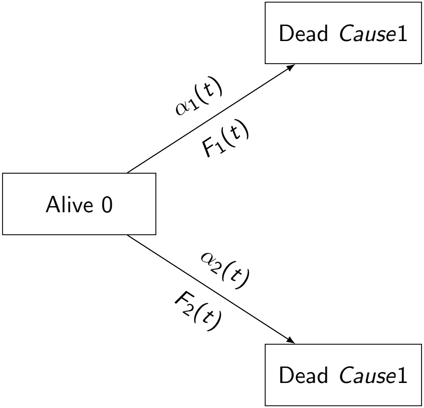
Martingale (mean-0 process iid): $\quad M_i(t) = N_i(t) - \Lambda_i(t)$

where $\Lambda_i(t) = \int_0^t Y_i(s)\alpha(s)\, ds$ (compensator),
$Y_i(t) = I(t \leq T_i)$ (at risk process).

- Intensity: $Y(t)\alpha(t)$.
- Intensity is rate given history

Competing Risk Models

With several causes of death forced to use `competing risks` model:



- cause specific hazards $\alpha_1(t)$ $\alpha_1(t)$
- cumulative incidence $F_1(t)$ $F_2(t)$

Malignant melanoma

In the period 1962-77 205 patients had their tumour removed and were followed until 1977.

- 57 died of mgl. mel.
- 14 died of non-related mgl. mel.
- 134 were still alive.

Purpose: Study effect on survival of sex, age, thickness of tumour, ulceration, ..

```
head(melanoma)
```

```
   no status days ulc thick sex
1 789      3   10   1   676   1
2  13      3   30   0    65   1
3  97      2   35   0   134   1
4  16      3   99   0   290   0
5  21      1  185   1  1208   1
6 469      1  204   1   484   1
```

The Cox model

- The regression coefficients $\beta_1, ..., \beta_p$ represent the effects of the covariates.
  - β_1 is the effect of X_{i1} when we have corrected for the other covariates.
  - β_1 may be interpreted in terms of the relative risk when the covariate X_{i1} is increased 1:

$$\frac{\lambda_0(t)\exp(\beta_1(X_{i1}+1)+...+\beta_p X_{ip})}{\lambda_0(t)\exp(\beta_1 X_{i1}+...+\beta_p X_{ip})} = \exp(\beta_1)$$

- If $\beta_1 > 0$ the risk of dying increases as X_{i1} increases, and if $\beta_1 < 0$ the risk of dying decreases as X_{i1} increases.
- The quantity $\hat{\beta}_1 X_{i1} + ... + \hat{\beta}_p X_{ip}$ is called the prognostic index, linear predictor, for the ith subject.
- For survival modelling β_1 is also survival difference on cloglog survival scale.

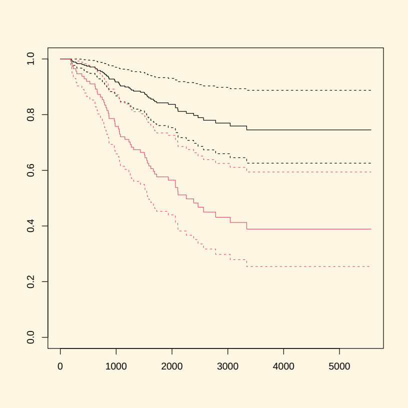
Melanoma Data

```r
out <- coxph(Surv(days,status==1)~thick+sex+ulc,melanoma)
summary(out)
```

```
Call:
coxph(formula = Surv(days, status == 1) ~ thick + sex + ulc,
    data = melanoma)

  n= 205, number of events= 57

           coef exp(coef)  se(coef)     z Pr(>|z|)
thick 0.0011345 1.0011351 0.0003794 2.990  0.00279 **
sex   0.4594907 1.5832675 0.2667580 1.723  0.08498 .
ulc   1.1668079 3.2117240 0.3114615 3.746  0.00018 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

      exp(coef) exp(-coef) lower .95 upper .95
thick     1.001     0.9989    1.0004     1.002
sex       1.583     0.6316    0.9386     2.671
ulc       3.212     0.3114    1.7443     5.914

Concordance= 0.76  (se = 0.034 )
Likelihood ratio test= 39.39  on 3 df,   p=1e-08
Wald test            = 37.75  on 3 df,   p=3e-08
Score (logrank) test = 44.96  on 3 df,   p=9e-10
```

Survival predictions pay attention

```
par(bg="#fdf6e3")
out=coxph(Surv(days,status==1)~thick+sex+ulc,data=melanoma
    )
newdata=data.frame(ulc=c(0,1),thick=300,sex=1);
pred=survfit(out,newdata);
plot(pred,conf.int=TRUE,col=1:2);
```



Likelihood construction

- Likelihood function based on
  $T_i = T_i^* \wedge C_i \quad D_i = I(T_i^* \leq C_i), X_i$ is:

$$\prod_i \alpha(T_i)^{D_i} e^{-\int_0^{T_i} \alpha(t)\,dt} = \prod_i \left\{ \prod_t \alpha(t)^{\Delta N_i(t)} \right\} e^{-\int_0^{\tau} Y_i(t)\alpha(t)\,dt}$$

- τ is end of observation period
- $Y_i(t) = I(t \leq T_i)$ is the at risk indicator.
- Can also write it as

$\prod_i \left\{ \prod_t dA(t)^{\Delta N_i(t)} \right\} e^{-\int_0^{\tau} Y_i(t) dA(t)}$ where $A(t) = \int_0^t \alpha(s)\,ds$.

Cox-model: $A(t) = A_0(t) e^{X^T \beta}$ with $A_0(t) = \int_0^t \alpha_0(s)\,ds$.

Likelihood construction

The Breslow-estimator: $\hat{A}_0(t,\beta) = \int_0^t \frac{1}{S_0(t,\beta)} dN.(s)$

- $S_0(t,\beta) = \sum_{i=1}^n Y_i(t)\exp(X_i^T\beta)$.
- Solving for $dA(t)$ where the jumps are that gives $dN.(t)/S_0(t,\beta)$
- Now plug this estimator into the likelihood function, and arrive at a function only depending on β !
- This function, $L(\beta)$ is called Cox's partial likelihood function.

Cox's proportional hazards model

- The Cox model is
$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp(X_i^T\beta), \tag{3}$$
where $X = (X_1, ..., X_p)$ is a p-dimensional locally bounded predictable covariate and $Y_i(t)$ is an at risk indicator.
- The regression parameter β is estimated as the maximizer to Cox's partial likelihood function
$$L(\beta) = \prod_t \prod_i \left(\frac{\exp(X_i^T\beta)}{S_0(t,\beta)} \right)^{\Delta N_i(t)}, \tag{4}$$
where $S_0(t,\beta) = \sum_{i=1}^n Y_i(t)\exp(X_i^T\beta)$.
- $\Lambda_0(t)$ is estimated by the Breslow estimator
$$\hat{\Lambda}_0(t) = \hat{\Lambda}_0(t,\hat{\beta}) = \int_0^t \frac{1}{S_0(t,\hat{\beta})} dN_\cdot(t). \tag{5}$$

Cox's proportional hazards model

- We have presented theory with time-const. X's, but can easily handle time-var. covariates.
- Now let's look into the asymptotical properties of Cox's estimators.
- The first partial derivative of $S_0(t,\beta)$ with respect to β is denoted : $S_1(t,\beta) = \sum_{i=1}^{n} Y_i(t)\exp(X_i^T\beta)X_i$
- Show that $\hat{\beta}$ is found as the solution to the score equation $U(\hat{\beta}) = 0$, where

$$U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} (X_i(t) - E(t,\beta))dN_i(t) \quad E(t,\beta) = \frac{S_1(t,\beta)}{S_0(t,\beta)}. \tag{6}$$

- Show also that $U(\beta_0)$ can be written as a martingale (mean-0).
- Partial likelihood concave (almost always).

Cox's proportional hazards model

- Cox (1975, 1972)
- Asymptotics for Cox model
  - Andersen and Gill (1982)
  - Tsiatis (1981)
- Existence, Jacobsen (1984)

Cox's proportional hazards model

Estimating equations provide anonther general approach for hazard models.
Let $N(t) = (N_1(t), .., N_n(t))^T$ $\lambda(t) = (\lambda_1(t), ..., \lambda_n(t))^T$, $X(t) = (Y_1(t)X_1(t), ..., Y_n(t)X_n(t))^T$. The martingale decomposition of $dN(t)$ then reads

$$dN(t) = \lambda(t)dt + dM(t) = \text{diag}(\exp(X_i^T(t)\beta))Y(t)d\Lambda_0(t) + dM(t), \tag{7}$$

Least squares score equations

$$\int X^T \text{diag}(\lambda_i) W_1 \{dN - \text{diag}(\exp(X_i^T\beta))Yd\Lambda_0\} = 0, \tag{8}$$

$$\lambda_0^{-1}(t) Y^T \text{diag}(\lambda_i) W_2 \{dN - \text{diag}(\exp(X_i^T\beta))Yd\Lambda_0\} = 0, \tag{9}$$

Optimal choice of W_1 and W_2 (\$var^{-1}(dN$_i$(t))=1/λ_i(t)\$).

Cox's proportional hazards model

Solving (9) for fixed β gives

$$\tilde{\Lambda}_0(t) = \int_0^t Y^-(t) dN(t), \tag{10}$$

where $Y^-(t)$ is the generalized inverse

$$Y^-(t) = (Y^T(t)\mathrm{diag}(\exp(X_i^T(t)\beta))Y(t))^{-1} Y^T(t)$$

of $Y(t)$.
Inserting this solution into (8) and solving for β gives

$$\int X^T(t)(dN(t) - \mathrm{diag}(\exp(X_i^T(t)\beta))Y(t)Y^-(t)dN(t)) = 0 = U(\beta) \tag{11}$$

Asymptotic properties

- Under the standard conditions, then, as $n \to \infty$,

$$n^{-1/2} U(\beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma),$$
$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1}),$$

and Σ is estimated consistently by $n^{-1} I(\hat{\beta})$, with I minus the derivative of U.

- Under the standard conditions, then, as $n \to \infty$,

$$n^{1/2}(\hat{\Lambda}_0(t, \hat{\beta}) - \Lambda_0(t)) \xrightarrow{\mathcal{D}} U(t)$$

where $U(t)$ is a zero-mean Gaussian process with covariance function $\Phi(t)$. Not a martingale though, why?
Get back to how $\Phi(t)$ can be estimated consistently

Asymptotic properties

- Key to the proof is that the score evaluated in the true point β_0 is a martingale (evaluated at τ):

$$U(\beta_0) = \sum_{i=1}^{n} \int_0^{\tau} (X_i - E(t, \beta_0)) dM_i(t) \tag{12}$$

- The predictable variation process of $n^{-1/2} U(\beta_0)$ is

$$\begin{aligned} \langle n^{-1/2} U(\beta_0) \rangle &= n^{-1} \sum_{i=1}^{n} \int_0^{\tau} (X_i(t) - E(t, \beta_0))^{\otimes 2} Y_i(t) \exp(X_i^T(t) \beta_0) d\Lambda_0(\\ &= n^{-1} \int_0^{\tau} V(t, \beta_0) S_0(t, \beta_0) d\Lambda_0(t) \to \Sigma. \end{aligned}$$

where

$$V(t, \beta) = \frac{S_2(t, \beta)}{S_0(t, \beta)} - E(t, \beta)^{\otimes 2}. \tag{13}$$

Asymptotic properties

- It follows that $U(\beta_0)$ converges in distribution to a normal variate with zero-mean and variance Σ.
- A Taylor series expansion of the score gives $n^{1/2}(\hat{\beta} - \beta_0) = (n^{-1}I(\beta^*))^{-1}n^{-1/2}U(\beta_0)$, where β^* is on the line segment between β_0 and $\hat{\beta}$.
- Consistency of $\hat{\beta}$ and the results above give that $n^{1/2}(\hat{\beta} - \beta_0)$ converges to the postulated normal distribution.

Cox's proportional hazards model

The asymptotics for the baseline and as a consequence also the survival function makes it a little bit tricky to construct confidence bands.
One practical way to go about this is to an IID decomposition.

- write estimators as sums of iid terms
- based on this we can resample to get asymptotics for processes of interest (baseline, score for testing).
- also shows how results are extended to rate situation not relying on martingales.

IID decomposition

- The score process evaluated at β_0

$$n^{-1/2}U(\beta_0, t) = n^{-1/2}\sum_{i=1}^{n}\epsilon_{1i}(t) + o_p(1)$$

where

$$\epsilon_{1i}(t) = \int_0^t \left(X_i - \frac{s_1(\beta_0, t)}{s_0(\beta_0, t)}\right) dM_i(s)$$

with $s_j(\beta_0, t)$ the limit in prob. of $n^{-1}S_j(\beta_0, t)$.

- That is a sum of zero-mean iid terms!

The variance of $n^{1/2}(\hat{\beta} - \beta_0)$ may be estimated consistently by

$$\tilde{\Sigma} = nI^{-1}(\hat{\beta}, \tau)\left\{\sum_{i=1}^{n}\hat{\epsilon}_{1i}^{\otimes 2}(\tau)\right\} I^{-1}(\hat{\beta}, \tau).$$

where $\hat{\epsilon}_{1i}(\tau)$ is given by $\epsilon_{1i}(\tau)$ replacing unknowns with their empirical counterparts.

IID decomposition

- Try the same for the Breslow-estimator.

$$n^{1/2}(\hat{\Lambda}_0(t,\hat{\beta}) - \Lambda_0(t)) \approx n^{1/2}(\hat{\Lambda}_0(t,\beta_0) - \Lambda_0(t)) + D_\beta \hat{\Lambda}_0(t,\beta_0) n^{1/2}(\hat{\beta} - \beta_0$$

- $D_\beta \hat{\Lambda}_0(t,\hat{\beta})$ conv. in probability so last term is essentially a sum of zero-mean iid's.

But also

$$\begin{aligned} n^{1/2}(\hat{\Lambda}_0(t,\beta_0) - \Lambda_0(t)) &= \int_0^t \frac{1}{S_0(s,\beta)} dM.(s) \\ &\approx n^{-1/2} \int_0^t \frac{1}{s_0(s,\beta)} dM_i(s), \end{aligned}$$

since $n^{-1} S_0(s,\beta) \to s_0(s,\beta)$.

IID decomposition

- It thus follows that \$ n$^{1/2}$(Λ_0(t,β) - Λ_0(t)) \$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\epsilon_{2i}(t), \tag{14}$$

where

$$\begin{aligned}\epsilon_{2i}(t) =& \epsilon_{3i}(t) + H^T(\beta_0, t) I(\beta_0, \tau)^{-1}\epsilon_{1i}(\tau),\\ \epsilon_{3i}(t) =& \int_0^t S_0^{-1}(\beta_0, s) dM_i(s),\end{aligned}$$

and where $H(\beta, t)$ is the derivative of $\int_0^t S_0^{-1}(\beta, s)dN.(s)$ wrt β.

- The variance of $n^{1/2}(\hat{\Lambda}_0(t, \hat{\beta}) - \Lambda_0(t))$ thus may be estimated by

$n^{-1}\sum_{i=1}^{n}\epsilon_{2i}^{\otimes 2}(t)$

Resampling

- Now, $n^{1/2}(\hat{\Lambda}_0(t,\hat{\beta}) - \Lambda_0(t))$ is asymptotically equivalent to

$$n^{-1/2}\sum_{i=1}^{n}\hat{\epsilon}_{2i}(t)G_i \tag{15}$$

where $G_1, ..., G_n$ are iid $N(0,1)$.

- This is very useful for constructing confidence bands for the survival predictions and to make tests about the baseline and survival function predictions.
- Taylor expansion:

$$n^{1/2}(S_0(\hat{\Lambda}_0,\hat{\beta},t) - S_0(\Lambda_0,\beta,t)) = -S_0(\Lambda_0,\beta,t) \\ n^{1/2}\left\{\exp(X_0^T\beta)(\hat{\Lambda}_0(t) - \Lambda_0(t)) + \Lambda_0(t)\exp(X_0^T\beta)X_0^T(\hat{\beta}-\beta)\right\}. \tag{16}$$

Resampling

- Just saw that : $n^{1/2}(\hat{\beta} - \beta, \hat{\Lambda}_0 - \Lambda_0)$ was asymptotically equivalent to the processes

$$\Delta_1 = n^{-1/2} \sum_{i=1}^{n} \hat{\epsilon}_{1i} G_i,$$

$$\Delta_2(t) = n^{-1/2} \sum_{i=1}^{n} \hat{\epsilon}_{2i}(t) G_i,$$

where $G_1, ..., G_n$ are independent standard normals.

- It follows that $n^{1/2}(\hat{S}_0 - S_0)$ has the same asymptotic distribution as

$$\Delta_S(t) = S_0(t) n^{-1/2} \sum_{i=1}^{n} \hat{\epsilon}_{4i}(t) G_i, \tag{17}$$

where

$$\epsilon_{4i}(t) = \exp(Z^T \beta)\epsilon_{2i}(t) + \Lambda(t) \exp(X^T \beta) X^T \epsilon_{1i}$$

Cox-Regression: Partial likelihood principle

The regression parameter β is estimated as the maximizer to Cox's partial likelihood function

$$L(\beta) = \prod_t \prod_i \left(\frac{\exp\left(X_i^T \beta\right)}{S_0(t, \beta)} \right)^{\Delta N_i(t)}, \tag{18}$$

where $S_0(t, \beta) = \sum_{i=1}^n Y_i(t) \exp(X_i^T \beta)$.
This product of probabilities can be used in many other settings. It links directly to structure of model

- nested case control
- twin-settings
- conditional logistic regression

Cox-Regression: checking assumptions

Wish to study effect of covariates using Cox-model:
$\lambda_i(t) = \lambda_0(t)\exp(\beta_1 X_{i1} + ... + \beta_p X_{ip})$ $\lambda_0(t)$ is the baseline hazard for a subject with covariates 0.
The Cox models ability to deal with many covariates comes from the regression structure. Some assumptions have been made

- The effects of covariates are additive and linear on the log risk scale.
- If covariates interact with each other the regression model should include interaction terms.
- The relative risk between the hazard rate for two subjects is constant over time $c(\beta_1, ..., \beta_p) = \frac{\lambda_i'(t)}{\lambda_i(t)}$
- We will try to check all these assumptions with the latter being the key assumption.

Cox's proportional hazards model, Graphical approach

Traditional goodness-of-fit tools. Model:
$\alpha_i(t) = \alpha_0(t)\exp(\beta_1 X_{i1} + ... + \beta_p X_{ip})$ Investigate if each of the covariates are consistent with the proportional hazards assumption. Stratify based on a grouping (k=1,...,K) based on X_{i1}'s values:
$\alpha_i(t) = \alpha_{0k}(t)\exp(\beta_2 X_{i2} + ... + \beta_p X_{ip});\quad \text{if} X_{i1} \in A_k$ Now, if the underlying full Cox-model is true the baseline estimates $\alpha_{0k}(t)$ should satisfy

$$\alpha_{0k}(t) = \alpha_0(t)\exp(\sum_{k=1}^{K} \beta_{1k} I(X_{i1} \in A_k)).$$

Graphical model-check of proportionality by making graphs of estimates of $\log(\int_0^t \alpha_{0k}(s)ds)$. Plotted against t they should be parallel.

Checking proportionality

- In the 2-sample case the assumption is equivalent to proportional intensities for the two groups, i.e.,

$$\exp(\beta_1)\lambda_1(t) = \lambda_2(t).$$

- Implies that the cumulative intensities are proportional

$$\Lambda_2(t) = \int_0^t \lambda_2(s)ds = \exp(\beta_1)\Lambda_1(t)$$

- A plot of $\hat{\Lambda}_2(t)$ versus $\hat{\Lambda}_1(t)$ should give a straight line through (0,0) with slope $\exp(\beta_1)$.
- Similarly

$$\log(\Lambda_2(t)) - \log(\Lambda_1(t)) = \beta_1$$

so plotting $\log(\hat{\Lambda}_k(t))$, $k = 1, 2$ versus t should give parallel curves.

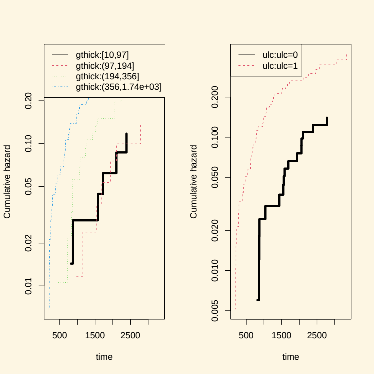
The Stratified Cox model

The stratified Cox model contains different baselines for different strata : $\lambda_{ik}(t) = \lambda_k(t)\exp(\beta_1 X_{i1} + ... + \beta_p X_{ip})$ $k = 1, ..., K$ $\lambda_k(t)$ is the baseline hazard for a subject in strata k.

- The regression coefficients $\beta_1, ..., \beta_p$ represent the effects of the covariates as in the simple Cox regression model.
- Melanoma-data: we stratify according to all covariates and look at graphs.
- The baselines give the mortality of the two defined by ulceration (yes/no) when all other covariates are 0.

Graphical procedure (mets)

```
par(bg="#fdf6e3")
dcut(melanoma) <- gthick~thick
out1 <- phreg(Surv(days,status==1)~strata(gthick)+sex+ulc,
    melanoma)
out2 <- phreg(Surv(days,status==1)~thick+sex+strata(ulc),
    melanoma)
par(mfrow=c(1,2))
plot(out1,log="y",lwd=4)
plot(out2,log="y",lwd=4)
```



Time-dependent covariates

- An important and useful extension of the Cox model is that we can allow time-dependent covariates
  - that are predictable (thus observed just before time t)

$$\lambda_0(t)\exp(Z^T(t)\beta) = \lim_{h\downarrow 0}\frac{1}{h}P(t \le T^* < t+h | T^* \ge t, X(s) : s \le t).$$

this model can be fitted and analysed as before. Note however, that survival is more complicated to compute unless $X(t)$ is deterministic in time (given $X(0)$) such that

$$S(t|X(0)) = \exp(-\int_0^t \exp(Z^T(t)\beta)d\Lambda_0(t))$$

- this can be used to make time-interactions between $X(0)$ and time.
- or to deal with real stochastic changes of $X(t)$.

Checking Proportionality: tests

- Graphical test may reveal serious violations, but they may be difficult to use in general.
- If effect is expected to wear off with time or to increase with time, one may try to describe this effect and then fit a Cox model with a time-varying explanatory variable $f(t)$ $\exp(\beta + \theta \cdot f(t)) = \frac{\lambda_1(t)}{\lambda_2(t)}$
- For a given choice of $f(t)$ we can then test if θ is 0, i.e, if the time-varying effect can be left out of the model. A common choice of $f(t)$ is $ln(t)$. More specialized methods and programs are available.

Checking Proportionality: tests

```
> fit=coxph(Surv(days/365,status==1)~ factor(sex)+ulc+lthick,
> data=melanoma)
> time.test=cox.zph(fit,transform="log")
> time.test
                  rho chisq      p
factor(sex)1 -0.0627 0.230 0.6312
ulc          -0.1335 0.956 0.3283
lthick       -0.2942 4.022 0.0449
GLOBAL            NA 8.773 0.0325
```

Based on this, the Cox model is rejected.

Checking Proportionality: tests

- Another option is to have a time-changing effect of ulc.
- Imagine for instance that effect of ulc is most important in an initial phase. We try the first 4 years, approx 1400 days.
- We hence replace $\beta_{ulc} \cdot ulc$ by $\beta_{ulc}(t) \cdot ulc$ with

$\beta_{ulc}(t) = \beta_{ulc,1} + \beta_{ulc,2} I(t > 1400)$

- Note that

$$\begin{aligned}\beta_{ulc}(t) \cdot ulc &= (\beta_{ulc,1} + \beta_{ulc,2} I(t > 1400)) \cdot ulc \\ &= \beta_{ulc,1} \cdot ulc + \beta_{ulc,2} \cdot X(t)\end{aligned}$$

with $X(t) = I(t > 1400)) \cdot ulc$ being a time-varying covariate.

- So this is just a Cox-model again, but now with a time-varying covariate.

Time-changing effect of covariates

- Such a model can easily be fitted in R using the survSplit-function.

```
> melanoma1=survSplit(melanoma,cut=c(1400),end="days",start="start",event="status
> melanoma1$ulcnew=melanoma1$ulc*as.numeric(melanoma1$days>1400)
>
> fit1=coxph(Surv(start,days,status==1)~ factor(sex)+lthick+factor(ulc)+ factor(u
> summary(fit1)
Call:
coxph(formula = Surv(start, days, status == 1) ~ factor(sex) +
    lthick + factor(ulc) + factor(ulcnew), data = melanoma1)

  n= 367

   coef exp(coef) se(coef)      z Pr(>|z|)
factor(sex)1     0.3744    1.4541   0.2701  1.386 0.165759
lthick           0.5741    1.7756   0.1801  3.187 0.001436 **
factor(ulc)1     1.6967    5.4561   0.5024  3.377 0.000732 ***
factor(ulcnew)1 -1.5515    0.2119   0.6451 -2.405 0.016170 *
---
```

- Relative risk Ulc/no Ulc in the first 4 years, and after the first 4 years? Conclusion?

Cox's proportional hazards model

- The procedures described above are the traditional goodness-of-fit tools.
- Make tests against specific deviations: Replace X_1 with $(X_1, X_1(\log(t)))$, say ($\beta_1 \to \beta_1 + \beta_{p+1} \cdot \log(t)$). Test the null $\beta_{p+1} = 0$.
- Graphical method
- Time-changing effect.

These methods are quite useful but also have some limitations:

- Graphical method:
  - Not parallel. What is acceptable?
  - What if a given covariate is continuous?
- Test: Ad hoc method. Which transformation to use?
- Time-changing effect. Also ad-hoc in that the time where effects changes are rarely known in practice.
- All methods: They assume that model is ok for all the other covariates.

Cumulative martingale residuals

Alternative: Cumulative martingale residuals, Lin, Wein and Ying (1993). The martingales under the Cox regression model can be written as

$$\begin{aligned} M_i(t) &= N_i(t) - \int_0^t Y_i(s)\exp(X^T\beta)d\Lambda_0(s) \\ \hat{M}_i(t) &= N_i(t) - \int_0^t Y_i(s)\exp(X^T\hat{\beta})d\hat{\Lambda}_0(s) \\ &= N_i(t) - \int_0^t Y_i(s)\exp(X_i^T(s)\hat{\beta})\frac{1}{S_0(s,\hat{\beta})}dN_\cdot(s). \end{aligned}$$

One idea is now to look at different groupings of the these residuals and see if they behave as they should under the model.

Cumulative martingale residuals

The score function, evaluated in the estimate $\hat{\beta}$, and seen as a function of time, can for example be written as $U(\hat{\beta}, t) = \sum_{i=1}^{n} \int_0^t (X_i - \frac{S_1(s,\hat{\beta})}{S_0(s,\hat{\beta})}) dN_i(s) = \sum_{i=1}^{n} \int_0^t X_i d\hat{M}_i(s)$ so it is equivalent to cumulating the martingale residuals (in certain way). Also, by Taylor series expansion,

$$n^{-1/2} U(\hat{\beta}, t) \approx n^{-1/2} U(\beta_0, t) - (n^{-1} I(\hat{\beta}, t)) n^{1/2} (\hat{\beta} - \beta)$$
$$n^{1/2} (\hat{\beta} - \beta) \approx (n^{-1} I(\hat{\beta}, \tau))^{-1} n^{-1/2} U(\beta_0, \tau)$$

Hence

$$n^{-1/2} U(\hat{\beta}, t) \approx n^{-1/2} \left\{ U(\beta_0, t) - I(\hat{\beta}, t)(I(\hat{\beta}, \tau))^{-1} U(\beta_0, \tau) \right\}$$

Cumulative martingale residuals

$n^{-1/2}U(\hat{\beta}, t)$ is thus asymptotically equivalent to the process

$$n^{-1/2}\left(M_1(t) - I(t,\hat{\beta})I^{-1}(\tau,\hat{\beta})M_1(\tau)\right), \qquad (19)$$

where

$$M_1(t) = \sum_{i=1}^{n} M_{1i}(t) = \sum_{i=1}^{n}\int_0^t (X_i(s) - e(s,\beta_0))dM_i(s)$$

with $e(t,\beta_0) = \lim_p E(t,\beta_0)$.

Note: $n^{-1/2}U(\hat{\beta},\tau) = 0$

Cumulative martingale residuals, resampling

The distribution of the process $n^{-1/2}M_1(t)$ ($t \in [0, \tau]$) is asymptotically equivalent to

$$n^{-1/2} \sum_{i=1}^{n} \int_0^t (X_i(s) - E(s, \hat{\beta})) dN_i(s) G_i$$

where $G_1, ..., G_n$ are independent standard normals.
Therefore it can be shown that $n^{1/2}U(\hat{\beta}, t)$ is asymptotically equivalent to

$$n^{-1/2} \sum_{i=1}^{n} \left(\tilde{N}_i(t) - I(t, \hat{\beta}) I^{-1}(\tau, \hat{\beta}) \tilde{N}_i(\tau) \right),$$

with $\tilde{N}_i(t) = \int_0^t (X_i(s) - E(s, \hat{\beta})) dN_i(s) G_i$ and for almost any sequence of the counting processes and iid G_i.

Cumulative martingale residuals, resampling

Can also use "martingale" resampling that holds more generally (clusters, rate models)
First $n^{-1/2} M_1(t)$ is also approximated asymptotically by

$$n^{-1/2} \sum_{i=1}^{n} \hat{M}_{1i}(t) G_i, \text{ with } \hat{M}_{1i}(t) = \int_0^t (X_i(s) - E(s, \hat{\beta})) d\hat{M}_i(s).$$

and therefore

$$n^{-1/2} \sum_{i=1}^{n} \left(\hat{M}_{1i}(t) - I(t, \hat{\beta}) I^{-1}(\tau, \hat{\beta}) \hat{M}_{1i}(\tau) \right) G_i,$$

for almost any sequence of the counting processes

- simple dN resamping easier on the computer
- G_i often standard normal
- also may other types of resampling, bootstrap, nice way of getting asymptotics of complicated objects

Cumulative martingale residuals

To summarize the cumulative score process plots one may look at test statistics like

$$\sup_{t \in [0,\tau]} |U_j(\hat{\beta}, t)| \quad |, \quad j = 1 \ldots, p,$$

Same as cumulative sum of Schoenfeld residuals, $(Z_i - E(Z))$ for event times, and a cumulative martingale residual sum

$$\sum_i Z_i \hat{M}_i(t)$$

These are called cumulative martingale residuals, or Lin, Wei, Ying score process test.

PBC-data

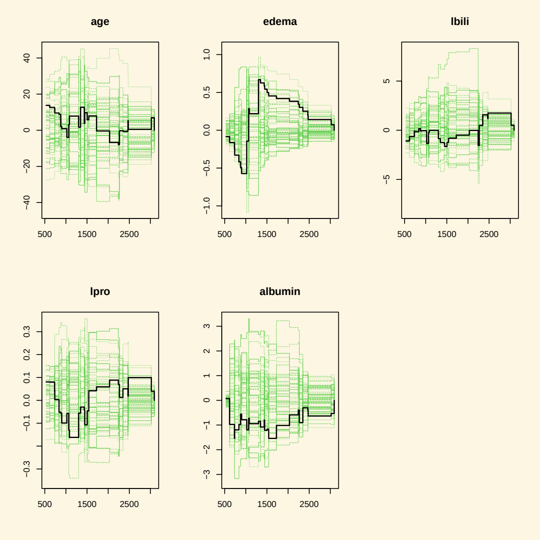
```
> library(timereg)
> fit=cox.aalen(Surv(days,status==1)~ prop(factor(sex))+prop(lthick)+
      prop(ulc), data=melanoma)
> summary(fit)

Proportional Cox terms :
                   Coef.    SE Robust SE D2log(L)^-1    z    P-val
prop(factor(sex))1 0.381 0.274     0.281       0.271 1.36 0.174000
prop(lthick)       0.576 0.162     0.172       0.179 3.34 0.000847
prop(ulc)          0.939 0.315     0.307       0.324 3.06 0.002230
Test for Proportionality
                   sup|  hat U(t) | p-value H_0
prop(factor(sex))1           3.27       0.282
prop(lthick)                 8.17       0.012
prop(ulc)                    4.31       0.054
> plot(fit,score=T)
```

PBC-data, mets

```
library(mets)
data(pbc); pbc=transform(pbc,lbili=log(bili),lpro=log(
    protime))
cox1 <- phreg(Surv(time,status==1)~age+edema+lbili+lpro+
    albumin,data=pbc)
gof(cox1)
```

```
Cumulative score process test for Proportionality:
         Sup|U(t)|  pval
age     13.7758260 0.881
edema    0.6696012 0.291
lbili    1.7491295 0.904
lpro     0.1619354 0.666
albumin  1.5415120 0.435
```

PBC

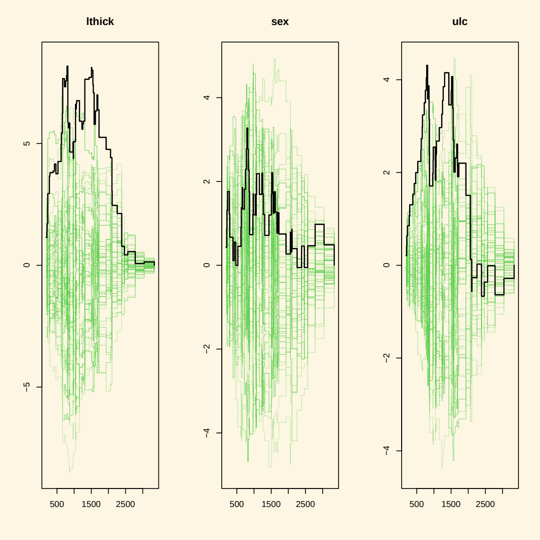


Melanoma

```r
melanoma=transform(melanoma,lthick=log(thick))
cox1 <- phreg(Surv(days,status==1)~lthick+sex+ulc,
    melanoma)
gof(cox1)
```

```
Cumulative score process test for Proportionality:
       Sup|U(t)|  pval
lthick  8.174262 0.017
sex     3.274663 0.312
ulc     4.310847 0.035
```

Melanoma


lthick
sex
ulc
5
0
−5
4
2
0
−2
−4
4
2
0
−2
−4
500
1500
2500
500
1500
2500
500
1500
2500

Important extension Stratified Cox

Now

$$\lambda(t) = Y(t)\lambda_s(t)\exp(X^T(t)\beta), \tag{20}$$

- Likelihood
  - baseline increments within strata
- Partial likehood for each strata
  - sums over strata to get $S_j^s(t, \beta)$

Robust standard errors

When model is not correct then robust standard errors will give a better estimate of standard errors

- time-dependent covariates (intensity model) $\alpha(t, X) = \lim_{h \downarrow 0} \frac{1}{h} P(t \leq T^* < t + h | T^* \geq t, X(s) : s \leq t)$.
- misspecified in terms of covariates (RCT's).

```
melanoma=transform(melanoma,lthick=log(thick))
cox1 <- phreg(Surv(days,status==1)~lthick+sex+ulc,
    melanoma)
summary(cox1)$coef
cox1 <- coxph(Surv(days,status==1)~lthick+sex+ulc,
    melanoma,robust=TRUE)
coefcox(cox1)
cox1 <- coxph(Surv(days,status==1)~lthick+sex+ulc,
    melanoma)
coefcox(cox1)
```

```
        Estimate      S.E.   dU^-1/2      P-value
lthick 0.5755837 0.1724857 0.1793779 0.0008469015
sex    0.3812724 0.2806747 0.2705711 0.1743323319
ulc    0.9388685 0.3071085 0.3243257 0.0022347260
        Coef.    SE    z    P-val lower2.5% upper97.5%
lthick 0.576 0.172 3.34 0.000847     0.239      0.913
sex    0.381 0.281 1.36 0.174000    -0.170      0.932
ulc    0.939 0.307 3.06 0.002230     0.337      1.540
        Coef.    SE    z   P-val lower2.5% upper97.5%
lthick 0.576 0.179 3.21 0.00133     0.225      0.927
sex    0.381 0.271 1.41 0.15900    -0.150      0.912
ulc    0.939 0.324 2.89 0.00379     0.304      1.570
```

Robust standard errors

Struthers and Kalbfleisch (1986),Lin and Wei (1989)
Fit Cox model $\lambda_0(t)\exp(X^T\beta)$ when true model is $\lambda(t, X)$.
It follows that $\hat{\beta}$ that converge to the solution of

$$U(\beta) = \int_0^\tau \left[s_1(s) - \frac{s_1(s,\gamma,\beta)}{s_0(s,\gamma,\beta)} s_0(s) \right] ds$$

$$s_j(t) = \lim_p n^{-1} \sum Y_i(s) X^j \lambda(t, X)$$

$$s_j(t,\beta) = \lim_p n^{-1} \sum Y_i(s) X^j \exp(X\beta)$$

and is asymptotically normal with standard errors estimated by the robust standard errors via sandwhich formula.

- This would suggest that we should use robust standard errors as the default.
  - luckily, however, often it does not matter very much.
  - important consequences for RCT's (A, Z)

Summary

- Cox's proportional hazards model.
  - Is is used heavily in Biostatistcs.
- Nice model with easily interpretable parameters - relative risks! Useful to also compute absolute risk, for example making survival predictions.
- Proportionality test important.
- Resampling techniques useful for evaluating asymptotics distributions.
- Useful goodness-of-fit based on cumulative residuals with p-values.

Some selected references

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10**, 1100–1120.

Cox, D.R. (1972). Regression models and life tables. *J. Roy. Stat. Soc. Ser. B* **34**, 406–424.

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Jacobsen, M. (1984). Maximum likelihood estimation in the multiplicative intensity model. *Internat. Statist. Review* **52**, 193–207.

Lin, D.Y. and Wei, L.J. (1989). The robust inference for the Cox proportional hazards model. *J. Amer. Statist. Assoc.* **84**(408), 1074–1078.

Struthers, C.A. and Kalbfleisch, J.D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363–369.

Tsiatis, A.A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.