MATH 60604A
Statistical modelling
§ 7d - Cox proportional hazard model

HEC Montréal
Department of Decision Sciences

Motivation

- What could we do if we wanted to measure the effect of explanatory variables $X_1, \ldots, X_p$ on the survival?
  - with categorical variables (and a lot of observations), we could estimate the survival function in each sub-group using Kaplan–Meier's estimator.
  - this approach does not work if X_j is continuous or the number of observations per group is small.

Cumulative hazard function

For T continuous*, we define the cumulative hazard function as

$$H(t) = \int_0^t h(u)\mathrm{d}u = \int_0^t \frac{f(u)}{S(u)}\mathrm{d}u = -\ln\{S(t)\}$$

and we can write the survival function

$$S(t) = \exp\{-H(t)\}.$$

We can write the log likelihood in terms of the (cumulative) hazard function

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\{\delta_i \ln h(t_i;\boldsymbol{\theta}) - H(t_i;\boldsymbol{\theta})\}$$

Proportional hazard assumptions

In the proportional hazard model, the hazard is parametrized as

$$h(t; \mathbf{x}_i) = h_0(t) \exp(\mathbf{x}_i \boldsymbol{\beta})$$

where

- the baseline hazard $h_0(t)$ is the only term that varies through time.
- the proportional hazard assumption implies that the ratio $h(t; \mathbf{x}_i)/h(t; \mathbf{x}_j)$ is constant regardless of time t.
- the interpretation of the effect of explanatory variables is simpler because these effects don't vary over time.
- this assumption is restrictive and must be validated in practice, but it is particularly convenient.

Note: there is no intercept in the Cox proportional hazard model: the latter is included in $h_0(t)$.

Derivation of the proportional hazards

We consider observed failure times $0 \leq t_1 < \cdots < t_D$, assuming no ties to simplify the derivation.
The baseline cumulative hazard

$$H_0(t) = \sum_{j:t_j \leq t} h_0(t_j),$$

is a step fonction with jumps only at the observed failure times.
We consider

- $\mathcal{R}_j$, the set of individuals at risk t_j
- δ_i, a binary indicator worth 1 for the observed failure and 0 if the observation is right-censored.

Likelihood of the Cox proportional hazard model

Let $h_j = h_0(t_j)$. The log likelihood is

$$\begin{aligned}\ell(h, \boldsymbol{\beta}) &= \sum_{i=1}^{n} \{\delta_i \ln\{\exp(\mathbf{x}_i\boldsymbol{\beta})h_i\} - \exp(\mathbf{x}_i\boldsymbol{\beta})H_0(t_j)\} \\ &= \sum_{i=1}^{n} \left\{ \delta_i \mathbf{x}_i\boldsymbol{\beta} + \delta_i \ln h_i - h_i \sum_{j\in\mathcal{R}_i} \exp(\mathbf{x}_j\boldsymbol{\beta}) \right\}\end{aligned}$$

- Since we are primarily interested in the effect of explanatories $\mathbf{X}$, we treat $h_1, \ldots, h_D$ as nuisance parameters.
- If $\boldsymbol{\beta}$ are fixed, the maximum likelihood estimator of h_i is $\widehat{h}_i = \delta_i / \sum_{j\in\mathcal{R}_i} \exp(\mathbf{x}_j\boldsymbol{\beta})$.
- This estimate is nonzero only if $\delta_i = 1$ (observed failure time).

Profile likelihood

The profile log likelihood for $\boldsymbol{\beta}$ is

$$\ell_{\mathsf{p}}(\boldsymbol{\beta}) = \max_{h} \ell(h, \boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \ln \left(\frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{\sum_{j \in \mathcal{R}_i} \exp(\mathbf{x}_j \boldsymbol{\beta})} \right)$$

It remains to maximimize $\ell_{\mathsf{p}}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$.
Even if the number of parameters of this model exceeds the number of observations (!), $\ell_{\mathsf{p}}(\boldsymbol{\beta})$ behaves like a regular likelihood.

- Standard errors are obtained from the observed information.
- We can perform likelihood ratio, score or Wald tests for $\boldsymbol{\beta}$.

The derivation is more complex with ties, but automatic adjustments are made by software (various alternatives, some are higher quality but more costly).

Once we recover the maximum likelihood estimators of $\widehat{\boldsymbol{\beta}}$, we can recover the cumulative hazard

$$\widehat{H}_0(t) = \sum_{i:t_i \leq t} \frac{\delta_i}{\sum_{j \in \mathcal{R}_i} \exp(\mathbf{x}_j \widehat{\boldsymbol{\beta}})},$$

from which the estimated survival function for an individual with covariates $\mathbf{x}$ follows

$$\widehat{S}(t; \mathbf{x}) = \exp\left\{ - \exp(\mathbf{x}\widehat{\boldsymbol{\beta}}) \widehat{H}_0(t) \right\}$$

Parameter interpretation

- In order to interpret the parameters in the Cox proportional hazards model, we can compare the hazard rates (multiplicative model).
- Consider two individuals who are similar in all ways, except that their X_j values differ by one unit.
  - For individual i with $X_{ij} = x_j$, the hazard rate is

$$h(t; \mathbf{x}_i) = h_0(t) \exp(\beta_1 x_1 + \cdots + \beta_j x_j + \cdots + \beta_p x_p)$$

  - For individual k with $X_{kj} = x_j + 1$, the hazard rate is

$$h(t; \mathbf{x}_k) = h_0(t) \exp(\beta_1 x_1 + \cdots + \beta_j (x_j + 1) + \cdots + \beta_p x_p)$$

- The hazard ratio is

$$\frac{h(t; \mathbf{x}_k)}{h(t; \mathbf{x}_i)} = \exp(\beta_j)$$

Hazard ratio

- For each increase of X_j by one unit, the hazard rate is multiplied by $\exp(\beta_j)$, *ceteris paribus*.
  - If $\exp(\beta_j) = 1$, the hazard rate is not affected by X_j.
  - If $\exp(\beta_j) > 1$, the hazard rate **increases** when X_j increases.
    - Higher values of X_j correspond to a higher risk of an event occurring, that is, shorter expected survival times.
  - If $\exp(\beta_j) < 1$, the hazard rate **decreases** when X_j increases.
    - Higher values of X_j correspond to a lower risk of an event occurring, that is, longer expected survival times.

Example with `melanoma` data

The `melanoma` data contains survival time of patients with malignant melanoma who had their tumour surgically removed, along with the following variables

- `time`: the survival time (in days) since the operation
- `status`: 1 if the patient died, 0 if censored
- `sex`: patient's sex, 1 for male, 0 for female
- `age`: patient's age (in years) at the time of the operation
- `thickness`: thickness of the tumour (in mm)
- `ulcer`: indicator variable, 1 if ulceration present and 0 otherwise

Descriptive statistics for `melanoma` data

Variable	Mean	Maximum	Minimum	Std Dev
time	2152.80	5565.00	10.00	1122.06
age	52.46	95.00	4.00	16.67
year	1969.91	1977.00	1962.00	2.58
thickness	2.92	17.42	0.10	2.96

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
205	57	148	72.20

Cox proportional hazard model for `melanoma`

The Cox proportional hazards model is

$$h(t) = h_0(t)\exp(\beta_1 \texttt{sex} + \beta_2 \texttt{age} + \beta_3 \texttt{thickness} + \beta_4 \texttt{ulcer})$$

We can fit this model in SAS using the `phreg` procedure:

SAS code for fitting a proportional hazard model

```
proc phreg data=statmod.melanoma;
model time*status(0) = sex age thickness ulcer / ties=exact;
run;
```

Likelihood-based tests

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	566.398	524.779
AIC	566.398	532.779
SBC	566.398	540.951

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	41.6195	4	<.0001
Score	46.6689	4	<.0001
Wald	39.4154	4	<.0001

The output includes the log likelihood value with and without explanatory variables, along with the usual global significance tests for $\mathscr{H}_0 : \boldsymbol{\beta} = \mathbf{0}_p$ versus $\mathscr{H}_a : \boldsymbol{\beta} \neq \mathbf{0}_p$.

Estimated coefficients of the Cox model

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
sex	1	0.43282	0.26741	2.6197	0.1055	1.542
age	1	0.01220	0.00830	2.1616	0.1415	1.012
thickness	1	0.10895	0.03773	8.3362	0.0039	1.115
ulcer	1	1.16448	0.30975	14.1330	0.0002	3.204

Interpretation

- For `sex`, $\exp(\widehat{\beta}_1) = 1.542$ represents the hazard ratio between a man versus a woman of the same age, with the same thickness of tumour and the same ulceration status. Thus, the hazard rate for males is 1.542 times that for females, *ceteris paribus*.
- For the variable `thickness`, $\exp(\widehat{\beta}_3) = 1.115$. For every 1mm increase in the tumour thickness, the hazard rate increases by a factor of 1.115 (or 11.5%), when all other variable are held constant.