Survival Analysis I (CHL5209H)

Olli Saarela

Dalla Lana School of Public Health
University of Toronto

olli.saarela@utoronto.ca

February 8, 2017

More on proportional hazards models

A Cox model for MI incidence

```
Call:
coxph(formula = Surv(ageb, ageb + t, e) ~ smoker +
    systbp + bpdrugs + nonhdl + hdl + bmi + hisdiab)

  n= 5934, number of events= 676

             coef  exp(coef)  se(coef)      z Pr(>|z|)
smoker   0.439647  1.552160  0.081829  5.373 7.75e-08 ***
systbp   0.009035  1.009075  0.001936  4.666 3.08e-06 ***
bpdrugs  0.333451  1.395776  0.094090  3.544 0.000394 ***
nonhdl   0.296432  1.345051  0.029253 10.133  < 2e-16 ***
hdl     -0.730499  0.481668  0.138603 -5.270 1.36e-07 ***
bmi     -0.001734  0.998268  0.010415 -0.166 0.867784
hisdiab  0.739579  2.095053  0.118605  6.236 4.50e-10 ***
---
```

Proportional hazards model

- In the model fitted here, the MI hazard for individual i at time t is given by

$$\begin{aligned}\lambda_i(t) = \lambda_0(t)\exp\{&\beta_1 \times \text{smoker}_i \\ &+ \beta_2 \times \text{systbp}_i \\ &+ \beta_3 \times \text{bpdrugs}_i \\ &+ \beta_4 \times \text{nonhdl}_i \\ &+ \beta_5 \times \text{hdl}_i \\ &+ \beta_6 \times \text{bmi}_i \\ &+ \beta_7 \times \text{hisdiab}_i\}.\end{aligned}$$

- $\lambda_0(t)$ is a baseline hazard function which may depend on time, but not on any individual-level characteristics.
- In turn, the regression coefficients $\beta = (\beta_1, \ldots, \beta_7)$ may not depend on time.

Proportional hazards are proportional

- For example, if we compare two hypothetical individuals i and l with $\text{smoker}_i = 1$ and $\text{smoker}_l = 0$, but otherwise same covariate values, we have that

$$\frac{\lambda_i(t)}{\lambda_l(t)} = \exp\{\beta_1\}.$$

- This log-hazard ratio interpretation applies to every regression coefficient, keeping the other covariates constants, although for continuous covariates the interpretation corresponds to a one unit increase in the covariate level.
- Note: such proportionality of hazards is a modeling assumption and is not always appropriate.
- However, when appropriate, it very much simplifies the model, as the covariate effects can be characterized with a single parameter.

Nuisance parameters and parameters of interest

- If we are mainly interested in the proportional covariate effects, we probably do not wish to specify a parametric form for $\lambda_0(t)$. (Why?)
- This is now a nuisance parameter, while the log-hazard ratios β are parameters of interest.
- However, the general likelihood function for a parametric survival model is a function of both, namely

$$\prod_{i=1}^{n} \left[(\lambda_0(t_i) \exp(\beta' x_i))^{e_i} \exp\left\{ -\int_0^{t_i} \lambda_0(u) \exp(\beta' x_i) \,\mathrm{d}u \right\} \right],$$

  where x_i is the covariate vector for individual i.
- How to avoid specification and estimation of $\lambda_0(t)$?

Elimination of nuisance parameters

Alternative estimating functions

- Instead of the likelihood function for both $\lambda_0(t)$ and β, we have to obtain an estimation function that depends on β alone.
- Two possible means to eliminate nuisance parameters are *conditional likelihood* and *profile likelihood*.
- Neither is generally applicable; closed form profile and conditional likelihoods exist only in special cases.
- Let's first recall the general definitions.

Conditional likelihood

- Let the parameter vector of interest be θ, while the nuisance parameters are denoted ψ.
- Suppose that the data vector can be partitioned as $y = (v, w)$.
- If there exist a partition such that

$$p(v, w \mid \theta, \psi) = p(w \mid v, \theta) p(v \mid \theta, \psi),$$

  where the conditional distribution $p(w \mid v, \theta)$ does not depend on the nuisance parameters, $p(w \mid v, \theta)$ w.r.t. θ is a conditional likelihood function.
- If it is also true that

$$p(v, w \mid \theta, \psi) = p(w \mid v, \theta) p(v \mid \psi),$$

  the conditioning statistic v is ancillary, and conditioning does not lose information on the parameters of interest.
- Example: conditioning on the covariates in a regression model.

How to condition?

- If the conditioning statistic is not ancillary, we may lose information, but can still use the conditional likelihood for the estimation of θ.
- The benefit of this is that ψ need not be estimated, and the corresponding model components need not be specified.
- How to choose the conditioning statistic v?
- There are no general rules for this; C&H (1993, p. 129) say

> *However, the conditional approach is not an automatic method, but relies on our ingenuity in recognizing a suitable conditional argument. Such arguments are not always possible. For example, it has not proved possible to find an argument which leads to a conditional likelihood for the rate difference.*

Profile likelihood

- In the profile likelihood approach, we first try to maximize the likelihood function w.r.t. to the nuisance parameters ψ, keeping θ fixed, to get

$$\hat{\psi}(\theta) \equiv \arg\max_{\psi} p(y \mid \theta, \psi).$$

- If this has a closed form solution, $\hat{\psi}$ is a function of the parameters of interest θ and the data y.
- We can now substitute this expression back to the original likelihood function, to the get the profile likelihood expression

$$p(y \mid \theta, \hat{\psi}(\theta)).$$

- This can in turn be maximized w.r.t. θ to obtain the profile likelihood estimate.

Cox partial likelihood as a profile likelihood

Application to the proportional hazards model

- Previously the baseline hazard function $\lambda_0(t)$ was left unspecified.
- To apply the profile likelihood argument, we need to specify this.
- However, using the piecewise constant model, we can do this in a flexible way, specifying a separate baseline rate parameter λ_{0k} for pre-specified time intervals $(s_{k-1}, s_k]$, where $k = 1, \ldots, K$.
- Following the earlier notation for the piecewise constant model, let d_{ik} indicate whether an individual i experienced an event in the interval k, and y_{ik} the follow-up time contributed by individual i in interval k.

Interpretation

- We have now specified a piecewise constant hazard model

$$\lambda_{ik} \equiv \lambda_{0k} \exp(\beta' x_i).$$

- For example, if $n = 3$ and $K = 3$, the observed outcome data are

time interval:	$(0, s_1]$	$(s_1, s_2]$	$(s_2, s_3]$
interval number:	$k = 1$	$k = 2$	$k = 3$
$i = 1$	(y_{11}, d_{11})	(y_{12}, d_{12})	(y_{13}, d_{13})
$i = 2$	(y_{21}, d_{21})	(y_{22}, d_{22})	(y_{23}, d_{23})
$i = 3$	(y_{31}, d_{31})	(y_{32}, d_{32})	(y_{33}, d_{33})

- If the observed event times and types are $(t_1, e_1) = (3, 1)$, $(t_2, e_2) = (5, 1)$, and $(t_3, e_3) = (6, 0)$, and the intervals are specified through $(s_1, s_2, s_3) = (2, 4, 6)$, how does the above table look like?

Long format data

In statistical software, such split follow-up data could be represented as multiple rows per individual:

individual	interval	lower	upper	length	event	covariate
1	1	0	s_1	y_{11}	d_{11}	x_1
1	2	s_1	s_2	y_{12}	d_{12}	x_1
1	3	s_2	s_3	y_{13}	d_{13}	x_1
2	1	0	s_1	y_{21}	d_{21}	x_2
2	2	s_1	s_2	y_{22}	d_{22}	x_2
2	3	s_2	s_3	y_{23}	d_{23}	x_2
3	1	0	s_1	y_{31}	d_{31}	x_3
3	2	s_1	s_2	y_{32}	d_{32}	x_3
3	3	s_2	s_3	y_{33}	d_{33}	x_3

Long format data (cont.)

In the example, this would become:

individual	interval	lower	upper	length	event	covariate
1	1	0	2	2	0	x_1
1	2	2	4	1	1	x_1
1	3	4	6	0	0	x_1
2	1	0	2	2	0	x_2
2	2	2	4	2	0	x_2
2	3	4	6	1	1	x_2
3	1	0	2	2	0	x_3
3	2	2	4	2	0	x_3
3	3	4	6	2	0	x_3

The third row for individual 1 could be omitted as there is no likelihood contribution.

Fitting the model

- The piecewise constant model could be fitted as

```
glm(event ~ as.factor(interval) + covariate,
    offset=log(length),
    family=poisson(link='log'))
```

- However, now we want to avoid estimation of the interval-specific baseline log-rates.

Likelihood construction

- Here each $d_{ik} \in \{0, 1\}$, so they are not really Poisson counts, but the resulting likelihood function is of the familiar Poisson form.
- The rows in the previous long format data have a separate likelihood contribution, and the likelihood expression becomes

$$\prod_{i=1}^{n} \prod_{k=1}^{K} \left[(\lambda_{0k} \exp(\beta' x_i))^{d_{ik}} \exp\left\{ -y_{ik} \lambda_{0k} \exp(\beta' x_i) \right\} \right]. \quad (1)$$

- There are now as many nuisance parameters as time intervals.
- How to eliminate $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0K})$ using the profile likelihood approach?

Profiling

- The corresponding log-likelihood is

$$
\begin{aligned}
& l(\beta, \lambda_0) \\
& \equiv \sum_{i=1}^{n} \sum_{k=1}^{K} \left[d_{ik} \log(\lambda_{0k} \exp(\beta' x_i)) - y_{ik} \lambda_{0k} \exp(\beta' x_i) \right].
\end{aligned}
$$

- Differentiating w.r.t. each λ_{0k} separately gives

$$
\begin{aligned}
\frac{\partial l(\beta, \lambda_0)}{\partial \lambda_{0k}} &= \sum_{i=1}^{n} \frac{d_{ik} \exp(\beta' x_i)}{\lambda_{0k} \exp(\beta' x_i)} - \sum_{i=1}^{n} y_{ik} \exp(\beta' x_i) \\
&= \frac{d_k}{\lambda_{0k}} - \sum_{i=1}^{n} y_{ik} \exp(\beta' x_i),
\end{aligned}
$$

  where we denoted $d_k \equiv \sum_{i=1}^{n} d_{ik}$.
- Setting $\partial l(\beta, \lambda_0)/\partial \lambda_{0k} = 0$ and solving w.r.t. λ_{0k} gives

$$
\hat{\lambda}_{0k}(\beta) = \frac{d_k}{\sum_{i=1}^{n} y_{ik} \exp(\beta' x_i)}. \tag{2}
$$

Profile likelihood

Finally, substituting (2) back into (1) gives the profile likelihood

$$\prod_{i=1}^{n}\prod_{k=1}^{K}\left[(\hat{\lambda}_{0k}(\beta)\exp(\beta' x_i))^{d_{ik}}\exp\left\{-y_{ik}\hat{\lambda}_{0k}(\beta)\exp(\beta' x_i)\right\}\right]$$

$$=\prod_{i=1}^{n}\prod_{k=1}^{K}\left[\left(\frac{d_k\exp(\beta' x_i)}{\sum_{l=1}^{n} y_{lk}\exp(\beta' x_l)}\right)^{d_{ik}}\exp\left\{-\frac{y_{ik}d_k\exp(\beta' x_i)}{\sum_{l=1}^{n} y_{lk}\exp(\beta' x_l)}\right\}\right]$$

$$=\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{d_k\exp(\beta' x_i)}{\sum_{l=1}^{n} y_{lk}\exp(\beta' x_l)}\right)^{d_{ik}}\prod_{k=1}^{K}\exp\left\{-d_k\right\}$$

$$\overset{\beta}{\propto}\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{\exp(\beta' x_i)}{\sum_{l=1}^{n} y_{lk}\exp(\beta' x_l)}\right)^{d_{ik}}.$$

The limiting case

- Note that the last form depends on β only and can be maximized to obtain the profile likelihood estimates $\hat{\beta}$.
- We could imagine repeating the same profiling argument for infinitely many nuisance parameters corresponding to infinitely many time intervals of infinitesimal length $\mathrm{d}t$, so that each time interval will have at most one event.
- Because based on the previous expression, only the intervals with an observed outcome event have a profile likelihood contribution, the resulting expression is of the form

$$\prod_{i=1}^{n}\left(\frac{\exp(\beta' x_i)}{\sum_{l=1}^{n} Y_l(t_i)\,\mathrm{d}t\exp(\beta' x_l)}\right)^{e_i},$$

  where $Y_l(t) \equiv \mathbf{1}_{\{T_l \geq t\}}$ is an indicator for individual l being at risk (that is, without event and uncensored) at t.
- In the denominator, $\mathrm{d}t$ can be omitted as it does not depend on the parameters (cf. C&H 1993, p. 300).

Cox partial likelihood

- The resulting expression is known as the Cox partial likelihood (Cox, 1975).
- It can also be obtained as a partial likelihood, a generalization of conditional likelihood, hence the name.
- It avoids the piecewise constant hazard assumption by letting the length of the time bins go towards zero.
- We note that the Cox partial likelihood contributions can be interpreted as a conditional probabilities; they are the probabilities of event occurring to individual i, given that we know that one event occurred among those at risk at time t_i.
- Check: what is this probability if the covariates have no effect on the hazard?

Fitting Cox models in R

- coxph function in R survival package:

```
coxph(formula, data=, weights, subset,
      na.action, init, control,
      ties=c('efron','breslow','exact'),
      singular.ok=TRUE, robust=FALSE,
      model=FALSE, x=FALSE, y=TRUE, tt, method, ...)
```

- In the formula the response is a survival object returned by

```
Surv(time, time2, event,
     type=c('right', 'left', 'interval', 'counting',
            'interval2', 'mstate'),
     origin=0)
```

References

- Clayton, D. and Hills, M. (1993). Statistical models in epidemiology. Oxford University Press, Oxford.
- Cox, D. R. (1975). Partial likelihood. Biometrika, 62:269–276.