Proportional Hazard Models

CHRIS SUTTON

OCTOBER 2020
Topic outline

1. Covariate data
2. Proportional Hazard (PH) models
3. The Cox PH model
4. Model fitting criteria
If you were asked to carry out an investigation into mortality in Perth over 5 years, what data (in addition to observed deaths) would you collect on the population of the town?
Covariate data
Covariates

so far the models and estimators we have looked at in this course have used only age and duration data \((x, t)\)

in practice more data would usually be recorded for each life in a study which might be valuable for modelling purposes

<table>
<thead>
<tr>
<th>Age</th>
<th>Male / Female</th>
<th>Smoker / non-smoker</th>
<th>Type of treatment</th>
<th>Symptom severity</th>
<th>Postcode</th>
<th>Time since last medical</th>
</tr>
</thead>
</table>

this data is called **covariate data**
2 ways to deal with covariate data

- sub-divide the population into smaller groups and model separately
- model the effects of covariates directly using some ‘regression model’
covariate notation

there are p covariate data measures obtained for each life

\( z_i \) is a 1xp vector of covariates for the \( i^{th} \) life

\( z_i = (X_{i1}, X_{i2}, \ldots, X_{ip}) \)

the covariates can be collected in one of three ways

<table>
<thead>
<tr>
<th>Covariate measure</th>
<th>Example time since last medical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw numerical value</td>
<td>Actual time in months</td>
</tr>
<tr>
<td>0 or 1 value assigned</td>
<td>1 if within last year, 0 otherwise</td>
</tr>
<tr>
<td>Score on some other scale e.g. 1 to 5 (qualitative)</td>
<td>1 if 0-3 months; 2 if 3-6 months; 3 if 6-12 months; 4 if 12-24 months; 5 if &gt; 2 years</td>
</tr>
</tbody>
</table>
example covariates

<table>
<thead>
<tr>
<th>Life</th>
<th>Male or Female (F=1; M=0)</th>
<th>Weekly alcohol consumption (units)</th>
<th>Time since last medical (scored 1-5)</th>
<th>Prior history of heart disease (yes=1; no=0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>26</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>12</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>9</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
Proportional Hazard models
PH models

The most commonly used regression models in survival analysis
- can be built using non-parametric or parametric approaches

Instead of the force of mortality $\mu_{x+t}$ used in our simpler models, we introduce the **hazard function** $\lambda_i$ for the $i^{th}$ life where the hazard is now a function of both duration $t$ and the covariate data vector $z_i$

In a **Proportional Hazard** model,

$$\lambda_i(t, z_i) = \lambda_0(t) \cdot g(z_i)$$

where $\lambda_0(t)$ is a function of duration only
and $g(z_i)$ is a function of the covariate vector only
PH model simplifying assumption

\[ \lambda_i(t, z_i) = \lambda_0(t) \cdot g(z_i) \]

\( \lambda_0(t) \) a function of \( t \) only is the “baseline hazard”, it is the hazard for an individual with covariate vector of zero

\( g(z_i) \) is a function of covariate data only

This is much simpler to model than a single function that varies with both covariate data and duration
Parametric Proportional Hazard model
parametric PH model

we can construct PH models on a non-parametric or parametric basis

in general in a parametric model we assume the lifetime distribution follows a certain functional form

in a parametric PH model we assume the hazard function follows one of the types of parametric survival models e.g. exponential; Gompertz; Makeham (or others) and we bring the covariate vector into the model parameters
parametric PH using Gompertz

Example:

we know Gompertz as $\mu_x = Bc^x$ for some parameters $B$ and $c$
in terms of a hazard function rather than a force of mortality this translates to
$\lambda(t) = Bc^t$

Now in a PH model we can let parameter $B$ depend on the covariates
if $z_i$ is our 1xp covariate vector and $z_i^T$ is the transpose of that vector (so px1)
then we can set Gompertz parameter $B$ to be $B = \exp(\beta.z_i^T)$
where $\beta$ is a 1xp vector of regression coefficients ($\beta_1, \beta_2, .., \beta_p$)
PH with Gompertz e.g. cont’d

\[ \lambda_i(t, z_i) = c^t \exp(\beta.z_i^T) \]

where
- the baseline hazard is \( c^t \)
- \( \exp(\beta.z_i^T) \) is the [assumed] effect of the covariates

the log-hazard is linear and separates the baseline hazard term

\[ \log[\lambda_i(t, z_i)] = t \log(c) + \beta.z_i^T \]

which can be very convenient to work with

however the usefulness of this model will depend entirely on how effectively we can estimate the regression coefficients \( \beta \) [the 1xp vector \((\beta_1, \beta_2, .., \beta_p)\)]
The Cox PH model
introduction to Cox model


- Imperial College London
- won the first International Statistics Prize in 2016 (awarded by the American Statistical Association, see statprize.org) for this survival models work

Professor Sir David Cox, 1980. Source: General Motors Cancer Research Foundation, National Cancer Institute.
what and why

situations where we do not need to know the precise rate of mortality but instead are interested in relative levels of mortality between different individuals

we assume each individual’s mortality is proportional to some general function (the baseline hazard)

- we do not worry about the shape of this baseline hazard
- instead we focus on the constant of proportionality for each individual which will depend on the covariates
- this is a widely used survival model
Cox PH model

Hazard is in the form

\[ \lambda_i(t, z_i) = \lambda_0(t) \exp(\beta.z_i^T) \]

- so the general shape of the hazard function depends on the baseline hazard \( \lambda_0(t) \)
- the differences between individuals are given by \( \beta.z_i^T \)
- if we are conducting a study (e.g. a medical trial) where we are more interested in the effect of covariates than in the shape of the hazard, this means we can ignore the baseline hazard \( \lambda_0(t) \) and look to estimate the regression coefficients \( \beta \) irrespective of \( \lambda_0(t) \)
- this is called a “semi-parametric” approach which is widely used in statistical survival models
- the Cox model uses the method of “partial likelihood” to estimate the coefficients \( \beta \) but not the baseline hazard
- partial likelihood statistics behave in similar way to the more usual maximum likelihood
Partial likelihood in Cox

assume deaths are observed at times $t_1$, $t_2$, ..., $t_k$ with just one death at each $t_j$

let $R(t_j)$ be the set of lives at risk of death at time $t_j$ (just prior to the $j^{th}$ death)

the partial likelihood calculation depends only on the order in which deaths are observed

the probability that life 1 [out of the set $R(t_1)$ ] is the life that dies at $t_1$ (conditional on one death being observed at that time) is

$$\frac{\lambda_0(t) \exp(\beta . z_1^T)}{\sum_{i \in R(t_1)} \lambda_0(t) \exp(\beta . z_i^T)}$$

the baseline hazard $\lambda_0(t)$ will cancel top and bottom here
Partial likelihood in Cox (cont’d)

repeating this calculation for all observed deaths at $t_1, t_2, \ldots, t_k$ the probability that life 1 out of set $R(t_1)$ dies at time $t_1$ and life 2 out of set $R(t_2)$ dies at time $t_2$ and … life k out of set $R(t_k)$ dies at time $t_k$ is given by the product of these probabilities.

This is the **partial likelihood function** which here is a function of parameters $\beta$ (the regression coefficients)

$$L(\beta) = \prod_{j=1}^{k} \frac{\exp(\beta . z_j^T)}{\sum_{i \in R(t_j)} \exp(\beta . z_i^T)}$$
Partial likelihood in Cox

this $L(\beta)$ is a partial likelihood function because it only uses the order in which deaths are observed and the rest of the observed data is discarded.

To find the likelihood estimates for the regression coefficients $\beta$ we would need to differentiate with respect to each of the $p$ coefficients that make up $\beta$, set to zero and solve for vector $\hat{\beta}$ our likelihood estimate of vector $\beta$.

This vector of derivatives which we set to zero is the efficient score function $u(\beta)$

$$u(\beta) = \begin{bmatrix} \frac{d}{d\beta_1} \log L(\beta), & \ldots, & \frac{d}{d\beta_p} \log L(\beta) \end{bmatrix}$$

then $\hat{\beta}$ found by solving $u(\hat{\beta}) = 0$

in practice we will not be able to do this algebraically but will need a computer package.
Breslow’s approximation

The Cox PH model assumes there is one death at time $t_j$. If instead there are $d_j > 1$ deaths at time $t_j$ the modelling becomes much more complex because all of the possible combinations of the $d_j$ deaths out of $R(t_j)$ need to be included in the likelihood function.

In this scenario, “Breslow’s approximation” is sometimes used:

$$L(\beta) \approx \prod_{j=1}^{k} \frac{\exp(\beta.s_j^T)}{\sum_{i \in R(t_j)} \exp(\beta.z_i^T)^{d_i}}$$

where $s_j$ is the sum of the $z$ covariate vectors for the $d_j$ lives observed to die at time $t_j$. 
Model fitting criteria
assessing covariates

In PH models (including Cox) we need criteria for assessing the effects of the different covariates

the **likelihood ratio statistic** gives one method for doing this

- model 1 has \( p \) covariates
- model 2 has additional \( q \) covariates (so \( p+q \) in total)

\[
\log L_p = \text{the maximised log-likelihood of model 1}
\]
\[
\log L_{p+q} = \text{the maximised log-likelihood of model 2}
\]

then

likelihood ratio statistic = \(-2(\log L_p - \log L_{p+q})\)
likelihood ratio statistic comments

generally the likelihood ratio statistic will use the full likelihood function (the one used to derive MLEs) but for the Cox model it is okay to use it with partial likelihoods

this likelihood ratio statistic has an asymptotic [that is it approaches as a limit] $\chi^2$ distribution on q degrees of freedom under the hypothesis that the additional q covariates have no effect when the first p covariates are there

further tests of interactions between different covariates will not be covered in this module
2 model building strategies

Start with null model that has no covariates then add new covariates one at a time and evaluate with likelihood ratio statistic.

Begin with full model that has all the possible covariates and use likelihood ratio statistics to eliminate covariates that have no statistically significant affect.