

# Biostats: The Second Half

Rui Qiu

June 8, 2018

## 1 Cox Regression (week 5 continued and week 6)

**When?** When we are interested in the impact of a number of (usually continuous) covariates on the survival distribution.

**Basic Cox Regression:** the hazard at time  $t$  for an individual with covariates given by  $p$ -length vector  $\mathbf{x}$ :

$$\lambda(t; x) = \lambda_0(t) \exp(\beta^T x)$$

$\beta$  is a  $p$ -vector of unknown parameters.  $\lambda_0(t)$  is the **baseline hazard**: the hazard for a life at  $t$  when all covariates are 0, also it is arbitrary. So the whole  $\lambda(t; x)$  is a semi-parametric model.

### Cox Proportional Hazards Regression

Recall  $S_0(t) = \exp\left(-\int_0^t \lambda_0(s) ds\right)$  is the survival function for an individual whose covariates are all zero.

$$\begin{aligned} S(t) &= \exp\left(\int_0^t \lambda(s, x) \cdot ds\right) \\ &= \exp\left(-\int_0^t \lambda_0(s) \exp(\beta^T x) ds\right) \\ &= \exp\left(-\int_0^t \lambda_0(s) ds \cdot \exp(\beta^T x)\right) \\ &= \left[\exp\left(-\int_0^t \lambda_0(s) ds\right)\right]^{\exp(\beta^T x)} \\ &= S_0^{\exp(\beta^T x)} \end{aligned} \tag{1}$$

Hence, **the survival functions for different covariate values cannot cross.**

Cox regression estimations: maximizing the **partial likelihood (PL)**. Define

- $x_{(i)}$  the covariates for the person that was observed to die at time  $t_i$ .
- $R(t_i)$  the set of individuals still under study at  $t_i$ .

The calculation won't be tested. (Treat as a typical likelihood).

**PL idea:** what is the probability that a person with covariate given by  $x_{(s)}$  dies at  $t_s$  given that one of the persons in  $R(t_s)$  dies at this time?

$R(t_i)$  including censored

$$L(\beta) = \prod_{i=1}^m \frac{\exp(\beta^T x_{(i)})}{\sum_{j \in R(t_i)} \exp(\beta^T x_j)}$$

$m$  is the number of deaths. When there are tied deaths,

$$L(\beta) = \prod_{i=1}^m \frac{\exp(\beta^T s_{(i)})}{[\sum_{j \in R(t_i)} \exp(\beta^T x_j)]^{d_i}}$$

Find one example in slides

where  $s_{(i)} = x_{(i),1} + \dots + x_{(i),d}$  the sum of the covariates for the persons observed to die at  $t$ .

**Significance of one parameter in Cox model:** test statistic  $\frac{\hat{\beta}}{SE(\hat{\beta})}$  vs  $N(0, 1)$ .

**Overall significance of Cox model:** testing  $H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  vs.  $H_1 : \text{at least one of } \beta_i \neq 0, i = 1, \dots, p$ . Test statistic  $-2(LL(H_0) - LL(H_1)) \xrightarrow{D} \chi_k^2$  where  $k$  is the number of different parameters in the model specified under the null and alternative. (Recall inference.)

**Estimated survival curves** often plotted based on mean or median.

## 2 Two-state Markov model (week 7)

prob of trans between states

**Two-state Markov model:**  $\mu(x) = \mu_x$  is the hazard (aka **transition intensity**) at  $x$ . Alive to dead state is one-direction. Prob a life alive at a given age will be dead at any subsequent age is determined only by the transition intensity  $\mu_{x+t}$  for  $t \geq 0$ .

**assumption 1: Markov:** the prob of moving from one state to another depends **only on the current state**. Past events have no influence on the future.

**assumption 2:** For a short time interval  $dt$ ,

$$dt q_{x+t} \approx \mu_{x+t} dt + o(dt),$$

$o(dt)$  goes to zero faster than  $dt$ . Meaning: the probability of dying in a short time is roughly = the transition intensity times the length time interval.

**assumption 3: Transition intensity**  $\mu_{x+t}$  is a constant between 0 and 1.

**A model example:**

- $x + a_i$  is the age that individual  $i$  comes under observation.
- $x + b_i$  is the age that individual  $i$  will cease being observed if life survives to that age  $x + b_i$ .
- $x + T_i$  is the age at which observation of life  $i$  ends.
- $V_i$  is a rv: the length of time that individual  $i$  is under observation, called **waiting time**.

- $\delta_i$  is an indicator variable that equals to 1 if  $i$  dies and 0 otherwise. (censored = 0) If  $\delta_i = 0 \implies V_i = b_i - a_i, T_i = b_i$ . If  $\delta_i = 1 \implies V_i = T_i - a_i$ .

**From model to likelihood:** For data  $(V_i, \delta_i), i = 1, \dots, N$ .

- For  $\delta_i = 0$ , lives from  $x + a_i$  to  $x + a_i + v_i$ , contribution of this event to the likelihood is  $v_i p_{x+a_i}$ .
- For  $\delta_i = 1$ , lives from  $x + a_i$  to  $x + a_i + v_i$  and dies at  $x + a_i + v_i$ , the contribution to the likelihood is  $v_i p_{x+a_i} \mu(x + a_i + v_i)$
- Thus the individual contribution to the likelihood is  $v_i p_{x+a_i} \mu(x + a_i + v_i)^{\delta_i}$ .
- Recall  $v_i p_{x+a_i} = \exp(-\int_{x+a_i}^{x+a_i+v_i} \mu(s) ds)$ .
- Recall assumption  $\mu(s) = \mu$ .
- 

$$\begin{aligned}
 L &= \prod_{i=1}^N \exp(-v_i \mu) \mu^{\delta_i} \\
 &= \exp(-\mu \sum_{i=1}^N v_i) \mu^{\sum_{i=1}^N \delta_i} \\
 \implies l &= -\mu \sum_{i=1}^N v_i + \log(\mu) \sum_{i=1}^N \delta_i \\
 &= -\mu v + \log(\mu) \delta
 \end{aligned} \tag{2}$$

$v = \sum_{i=1}^N v_i$  is the total time under observation of all  $N$  individuals.

$\delta = \sum_{i=1}^N \delta_i$  is the total number of deaths.

**MLE of  $\hat{\mu}$**  is

$$\hat{\mu} = \frac{\delta}{v} \quad \text{SE}(\hat{\mu}) = \frac{\sqrt{\delta}}{v}$$

- asymptotically,  $\hat{y} \approx N(\mu, \frac{\mu^2}{E(\delta)})$ , use  $\hat{\mu} = \mu$  and  $\delta = E(\delta)$ .

### 3 Multi-state Markov model (week 7 and week 8)

**Notations:**

- ${}_t p_x^{gh}$  is the prob a person in state  $g$  at age  $x$  is in state  $h$  at age  $x + t$ .
- ${}_t p_x^{\overline{gg}}$  is the prob a person in state  $g$  at age  $x$  remains in that state continuously until age  $x + t$ .
- $\mu^{gh}$  is the transition intensity (or hazard) between state  $g$  and  $h$ . Assuming hazard is constant over each year of age.
- $v_i^g$  is the amount of time spent by individual  $i$  in state  $g$ .

- $\delta_i^{gh}$  is the number of transitions by individual  $i$  from state  $g$  into state  $h$ .

**Kolmogorov Forward Equations:** compute probabilities of interest from our estimated hazards or transition intensities:

$$\begin{aligned} \frac{d}{dt} {}_t p_x^{gh} &= \sum_{r \neq h} ({}_t p_x^{gr} \mu(x+t)^{rh} - {}_t p_x^{gh} \mu(x+t)^{hr}) \\ \frac{d}{dt} {}_t p_x^{\bar{g}\bar{g}} &= -{}_t p_x^{\bar{g}\bar{g}} \sum_{r \neq g} \mu(x+t)^{gr} \end{aligned} \quad (3)$$

NB:

$${}_t p_x^{\bar{g}\bar{g}} = \exp\left(-t \sum_{r \neq g} \mu^{gr}\right)$$

Further derivation leads to the MLEs:

$$\begin{aligned} \hat{\mu}^{gh} &= \frac{\delta^{gh}}{v^g} \\ \hat{\mu}^{gh} &\approx N(\mu^{gh}, \frac{\delta^{gh}}{(v^g)^2}). \end{aligned} \quad (4)$$

$\delta^{gh}$  is the total number of transitions from state  $g$  to  $h$  and  $v^g$  is the total time spent in state  $g$ .

Then

$$\hat{p}^{\bar{g}\bar{g}} = \exp\left(-t \sum_{r \neq g} \hat{\mu}^{gr}\right)$$

$$V(\hat{p}^{\bar{N}\bar{N}}) = V(\hat{\mu}^{NC} + \hat{\mu}^{NT}) \cdot (\hat{p}^{\bar{N}\bar{N}})^2$$

Find SE, find CI.

## 4 Binomial Markov Model (week 8)

Instead of estimating hazard  $\mu_x$ , we estimate  $q_x$  the **probability an individual aged  $x$  dies before reaching age  $x+1$** .

Setup:  $x+a_i$  enter,  $x+b_i$  stop,  $\delta_i$  indicating death. To estimate  $q_x$  we need to write the likelihood for  $\delta_i$ :

$$L = \prod_{i=1}^N b_i - a_i q_{x+a_i}^{\delta_i} (1 - b_i - a_i q_{x+a_i})^{(1-\delta_i)}$$

**assumption 1:** UDD  ${}_t q_x = tq_x$

**assumption 2:** constant hazard  ${}_t q_x = 1 - \exp(-\mu t)$

**assumption 3:** behaviour between 0, 1  $\mu_x(t) = \frac{q_x}{1-tq_x}$

**Example for calculating MLE:** let  $a_i = b_i = 0.5$ ,  $\omega$  = the number of censored individuals,  $\delta$  = total number of observed deaths, then likelihood:

$$q_x^\delta (1 - q_x)^{N-\omega-\delta} (1 - q_x)^{0.5\omega}$$

MLE:

$$\hat{q}_x = \frac{\delta}{N - 0.5\omega}$$

## 5 Statistical Tests (week 9)

**graduation: smoothing a set of crude mortality rates. Why?** 1. pooling of information, 2. belief is that mortality rates are a smooth function of age, 3. practical reasons.

tradeoff: smoothness vs adherence.

**focus:** statistical tests of whether a set of crude mortality rates are consistent with a **standard table** or a set of smoothed rates.

### 5.1 Chi-square Test

overall assessment of the closeness of the observed mortality and that expected given by the standard table.

Under null hypo, the obs deaths is binomial  $Bino(E_x, q_x)$ .

TS:

$$\chi^2 = \sum_{\text{age groups}} \frac{(\theta_x - E_x q_x)^2}{E_x q_x}$$

$\theta_x$  is obs. death, DegOfFreedom is the number of age groups (constraints).

- Within each group should be homogeneous.
- Rule of thumb: each group should contain about 5 deaths.
- advantage: can detect a single difference between observed and expected
- disadvantage: fail to detect certain issues like all observed deaths slightly greater than the expected deaths; may not detect large groups of positive or negative deviations

### 5.2 Standard Deviation Test

TS:

$$Z_x = \frac{\theta_x - E_x q_x}{\sqrt{E_x q_x (1 - q_x)}} \sim N(0, 1)$$

1. Decide on a set of intervals (equal lengths)
2. compute the observed number of  $Z_x$  falling in each interval
3. compute the expected number of  $Z_x$  falling in each interval, under null.  
The expected number is the number of age ranges times the appropriate area from the standard normal.
4. use a  $\chi^2$  test to compare the observed and expected value.

DegOfFreedom is the number of intervals -1

**can detect situation of a few large deviations being masked by an excessive number of smaller deviations.**

### 5.3 Sign Test

**Compare the observed number of positive deviations to the Binomial( $m, 1/2$ ) dist.**  $m$  is the number of age groups. Approx to  $N(0.5m, 0.25m)$ .

**Two sided test: both too many or too few positive deviations are evidence against  $H_0$ .** Can detect an excessive number of positive or negative deviations.

Generally, same result from sign test and standard deviation test.

### 5.4 Cumulative Deviation Test

TS:

$$\frac{\sum_a^b (\theta_x - E_x q_x)}{\sqrt{\sum_a^b (E_x q_x p_x)}} \sim N(0, 1)$$

can be applied to the whole age range or over particular ranges. **two-sided**

### 5.5 Runs Test

The **runs test** is based on the **observed number of runs of positive deviations**. A run is a groupings of positive (or negative deviations). E.g. 1, 2, -1, -2, 3, 1 have 2 runs of positive deviations.

The possibility of seeing  $t$  or less positive runs:

$$\sum_{j=1}^t \frac{\binom{n_1-1}{j-1} \binom{n_0+1}{j}}{\binom{m}{n_1}}$$

where  $n_1$  and  $n_0$  are the number of positive/negative deviations.  $m$  is the number of age groups. For large values of  $m$  a normal approximation can be used,

$$t \sim N\left(\frac{n_1(n_0+1)}{n_0+n_1}, \frac{(n_0 n_1)^2}{(n_0+n_1)^3}\right)$$

**One-sided**, a small number of runs is evidence against the null hypothesis. (too stacked! too grouped!)

test name	TS approx	one or two sided
$\chi^2$ test	$\chi_{(m)}^2$	one-sided
Std-dev test (under Normal approx)	$\chi_{(-1)}^2$	one-sided
cum-dev test	Z	separate groups (two-sided)
sign test	Z	two-sided
runs test	Z	one-sided

## 6 Methods of Graduations/Smoothing (week 10)

### 6.1 Kernel Smoothing

**idea:** estimate  $f(x^*)$  using information contained in observations that are close to the point of interest  $x^*$ . The weight assigned to each obs is determined by

Common kernel  
formula

the form of the **kernel**.

**Kernel properties:** 1. obs further away from  $x^*$  receive less weight. 2. symmetric kernel

kernel smoothed value at  $x^*$ :

$$\hat{f}(x^*) = \hat{y}(x^*) = \sum_{j=1}^n w_j y_j$$
$$w_j = \frac{K\left(\frac{|x^* - x_j|}{b}\right)}{\sum_{m=1}^n K\left(\frac{|x^* - x_m|}{b}\right)}$$

where  $K(t)$  is the kernel and  $b$  is the bandwidth.  $b$  decreases,  $t$  increases, the smaller  $b$  the quicker the weights assigned to observations decline. (in fact, observations very close to  $x^*$  receive larger weights. So the weights decline for majority of observations but increase for the observations close to  $x^*$ . So we get a least smooth curve. So vice versa, greater  $b$ , smoother.

**bandwidth** ( $b$ ) is the tuning parameter.

## 6.2 Spline Smoothing

**natural cubic splines** estimate  $f(x)$  by using piece-wise cubic polynomials. Fitting them with **constraints**:

- pieces are connected
- smooth = derivatives of endpoints should be the same
- the 2nd derivative at the connection points should be the same

We customize the number of **knots**. Knots = connection points. The num of knots is the tuning parameter for splines. More knots, less smooth.

**Knots selection:** minimize AIC or BIC (measure goodness of fit, also penalize for the number of parameters used)

- $AIC = -2 \ln(L) + 2 \times p$
- $BIC = -2 \ln(L) + \log(n) \times p$

**Other smoothers** loess = locally weighted scatterplot smoothing...

## 7 Graduation: ALT 2005-07 (week 11)

Manual graduation will be used at the extremes of age. Cubic splines used to graduate (the crude rates) as well. For cubic splines, the number of knots is undetermined yet. Tests will be conducted to assess the appropriateness of graduated rates.

Knots selected by minimizing

$$\frac{(A_x - E_x^c m_x)^2}{E_x^c m'_x (1 - m'_x)}$$

where  $A_x$  is the observed number of deaths aged  $x$ ,  $E_x^c$  is the central exposed to risk aged  $x$ ,  $m_x$  is the graduated value of the central mortality rate at age  $x$ .  $m'_x$  is a preliminary value of  $m_x$ .

## 8 Exposed to Risk: how to obtain crude mortality (week 12)

**what is involved?** Requires a measure of the number of deaths and the population which was at risk of dying, the exposed-to-risk- over the same period.

$$\begin{aligned}\hat{q}_x &= \frac{d_x}{E_x} \\ \hat{\mu}_x &= \frac{d_x}{E_x^c}\end{aligned}\tag{5}$$

$d_x$  is the number of deaths and  $E_x$  and  $E_x^c$  are initial and central exposed to risk.

Under constant force of mortality, central mortality=force of mortality:

$$\begin{aligned}m_x &= \frac{d_x}{\int_0^1 l_{x+t} dt} = \frac{q_x}{\int_0^1 {}_t p_x dt} \\ &= \frac{1 - e^{-\mu}}{\int_0^1 e^{-\mu t} dt} = \frac{1 - e^{-\mu}}{\frac{1}{\mu} [-e^{-\mu t}]_0^1} \\ &= \frac{1 - e^{-\mu}}{\frac{1}{\mu} (1 - e^{-\mu})} \\ &= \mu\end{aligned}\tag{6}$$

$E_x^c$ : is the total time that all individuals under study are observed. This is the same as the waiting time.

$E_x$ : a less natural quantity than  $E_x^c$ , requires adjustment for individuals who die while under observation.

Under the assumption that on average deaths occur half way through the year, the following approx can be used to estimate  $E_x$ :

$$E_x \approx E_x^c + 0.5d_x$$

But computation  $E_x^c$  requires lots of detailed information. Sometimes we only have information at one or more times, we use  $E_x^c = \int P(x, t) dt$ .

Use the **Trapezium approx** to estimate the area of trapezium: **(sum of sides / 2) \* height**.

**Rate intervals:** intervals of time over which an individual's age remains the same. if die in this interval, contribute to  $d_x$ . age definition is important.