Applied Survival Analysis Using R
Additional

Qi Guo

Department of Mathematical Sciences
The University of Texas at Dallas

May, 4 2019

Additional

- The *additional* part is to replenish the stuff after reading the notes.
- If you have questions about the replenishment, send me an email($qxg160330@utdallas.edu$), and I will update it.
- Thanks for the guidance from Dr.Steven, and it's our pleasure if everyone in our department can really learn something or jog your memory from notes.

Kernel Smoother

- Question: What are your thoughts on kernel smoother used in Chapter 3, e.g., Figure 3.5?
- Answer: Now we know, for *Nelson-Aalen estimator*, the cumulative hazard function is the sum of the estimated hazards up to a time t_i

$$H(t) = \sum_{t_i \le t} \frac{d_i}{n_i} \tag{1}$$

  But, this estimate of the hazard function is quite unstable from one time to the next, and thus is of limited value in illustrating the true shape of the hazard function.
- A better way to *visualize* the hazard function estimate is by using a "`kernel`" smoother, $K(\mu)$.

Kernel Smoother

- The amount of smoothing controlled by a parameter b. The estimate of the hazard function is given by:

$$\hat{h}(t) = \frac{1}{b}\sum_{i=1}^{D} K\left(\frac{t - t_{(i)}}{b}\right)\frac{d_i}{n_i} \tag{2}$$

  where $t_{(1)} < t_{(2)} < \cdots < t_{(D)}$ are distinct ordered failure times, the subscript "(i)" in $t_{(i)}$ indicates that this is the ith ordered failure time, d_i is the number of deaths at time $t_{(i)}$, and n_i is the number at risk at that time.
- Larger values of b result in wider kernel functions, and hence more smoothing

Kernel Smoother

- This is illustrated in Fig. 3.5.Here the three failure times $t = 2, 4, 6$ are indicated by gray triangles, and the kernels, are dashed gray.The sum, the smoothed estimate of the hazard, is given by the blue curve.

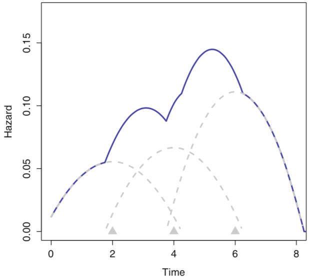

Fig. 3.5 Illustration of the hazard kernel smoother using the example data from Table 1.1 and the Kaplan-Meier estimate in Table 3.1

Kernel Smoother

- The first two arguments are the failure times and censoring indicators, the maximum time is set at 8; the smoothing parameter *b* is specified by "`bw.grid=2.25`"; the "`global`" option means that a constant smoothing parameter is use for all times; and the "`b.cor`" option is set to "`none`" indicating that no boundary correction is to be done.

```
> library(muhaz)
> t.vec <- c(7,6,6,5,2,4)
> cens.vec <- c(0,1,0,0,1,1)
> result.simple <- muhaz(t.vec, cens.vec, max.time=8,
bw.grid=2.25, bw.method="global", b.cor="none")
> plot(result.simple)
```

Kernel Smoother

- One use of smoothing the hazard function is to obtain a smooth estimate of the survival function by:

$$S(t) = e^{-\int_{\mu=0}^{t} h(\mu)d\mu} \tag{3}$$

```
surv<-exp(-cumsum(haz[1:(length(haz)-1)]*diff(times)))
```

The survival curve estimation uses the "`cumsum`" function. Since the length of "`diff(times)`" is one less than the length of "`times`" and "`haz`", we need to drop the last element of "`haz`".

Weight

- Question: When to use weights in the tests considered in Ch4? Is it always a good idea to stratify?
- Answer: [1] *Weight* is a value assigned to each case in the data file. Normally used to make statistics computed from the data more representative of the population.
  Examples: A weight of 2 means that the case counts in the dataset as two identical cases.
  Two most common types:
  - Design Weights(DW)
  - Post-Stratification(PS)

Weight

- DW is normally used to compensate for over- or under-sampling of specific cases or for disproportionate stratification.
- Example: It is a common practice to over-sample minority group members or persons living in areas with larger percentage minority. If we doubled the size of our sample from minority areas, then each case in that area would get a design weight of $\frac{1}{2}$.
- PS is used to compensate for that fact that persons with certain characteristics are not as likely to respond to the survey.

Weight

- In survival analysis, the weight w_i is compensate for different survival function in different t_i.

$$w_i = \left\{\hat{S}(t_i)\right\}^{\rho} \tag{4}$$

- So we can control higher or lower weight on anytime we want during survival analysis.
- Weights almost always increase the standard errors of estimates. Introduce instability into data.
- Very large weights (or very small ones) can also introduce instabilities.

Partial Likelihood for Cox model

- Question: The concept of the partial likelihood for Cox model and why it is preferred over full likelihood?
- Answer: Some parametric distributions require strong assumptions about the form of the underlying survival distribution, and *the partial likelihood* use an unspecified *baseline survival distribution* to define the survival distributions of subjects based on their covariates.
  Why partial likelihood?
  - It is a product of expressions, one for each failure time, while censoring times do not contribute any factors.
  - The factors of a partial likelihood are *conditional probabilities*.

Score function for Cox model

- Question: How the score function derive from Cox model?
- Answer: The Score function is the first derivative of the partial likelihood function(after take log)

$$S(\beta) = \ell^{'}(\beta) \tag{5}$$

- We use it to check the β is significant or not by different test in cox model.

Outlier and influential point

- Question: What's the difference between an outlier and influential point?
- Answer: An *outlier* is a data point that *diverges* from an overall pattern in a sample. An outlier has a large residual (the distance between the predicted value ($\hat{y}$)) and the observed value (y)). Outliers lower the significance of the fit of a statistical model because they do not coincide with the model's prediction.
  An *influential* point is any point that has a large effect on the slope of a regression line fitting the data. They are generally extreme values
- Check outliers: plot with residuals
- Check influential: Just use the *Case Deletion Residuals* in note7. By removing the suspected influential point from the data set. If this removal significantly changes the slope of the regression line, then the point is considered an influential point.

Time dependent Covariates

- Question: Time-dependent variables that increase with time has no effects in the Cox model, why?
- Answer: Denote age at entry into the trial by $z(0)$and current age by $z(t) = z(0) + t$. Then the hazard function is given by:

$$h(t) = h_0(t)e^{\beta z(t)} = \{h_0(t)e^{\beta t}\} \cdot e^{\beta z(0)} \tag{6}$$

When we plug in the formula for the partial likelihood(Chapter 5), the time dependent part, $e^{\beta t}$, appears in both the numerator and the denominator of each factor, as does the baseline hazard. Both cancel, leaving only the age at entry variable $z(0)$ Thus, the coefficient β for the time dependent model is *identical* to that from the non-time dependent model.

Marginal model

- Question: When do we call an approach "marginal"?
- Answer: From the note 9, we know the marginal consider each event to be a separate process, so subjects are at risk for all events from the start of follow-up, regardless of whether they experienced a prior event. We take it without clustered (independent), and the standard error of the estimate $\hat{V}^{1/2}$ times a quantity C for a clustered-adjusted standard error.
- This model is appropriate when the events are thought to result from *different underlying processes*, so that a subject could experience a 3rd event, for example, without experiencing the 1st. Although this assumption seems implausible with some types of data, like cancer recurrences, it could be used to model injury recurrences over a period of time, when subjects could experience different types of injuries over the time period that have no natural order.

Frailties

- Question: Why would the full likelihood(Chapter 9,eq8) can not be solved directly?
- Answer: Because the frailties(ω_i) are *latent* variables, we can't directly observe, so that's why we use *EM* algorithm to estimate β and θ.

$$L(\beta,\theta) = \prod_{i=1}^{G} \prod_{j=1}^{n_i} L_{ij}(\beta,\theta;\omega_i,t_{ij},\delta_{ij},z_{ij}) \tag{7}$$

+cluser and +frailty

- Question: What's the difference between "+cluster()" and "+frailty()" in coxph? Can we use both in the same model?
- Answer: For "+cluster()", The parameter estimate and it?s standard error are the same as in the ordinary Cox model, but we have an additional estimate of the standard error, the "robust se", this is only slightly *higher* than the one from the standard Cox model, indicating that the effect of clustering within first-degree relatives is small.
  "+frailty" account for the clustering using a *gamma frailty model* for default, and it usually more than one estimates of standard error. We *can't* use "+frailty()" and "+cluster()" both in the same model.

Reference

[1] Using Weights in the Analysis of Survey Data, David R. Johnson, Department of Sociology Population Research Institute, The Pennsylvania State University ,November 2008