Fundamentals of Statistical Modeling (VT20)

Andrea Discacciati
Karolinska Institutet
Stockholm, Sweden

Lab 4

Load the dataset and the `mlci` command

{{1}}

Exercise 1

We use data from 267 patients diagnosed with oral cancer. We measured time to death (y) in subjects with low-grade cancer ($x = 0$) and high-grade cancer ($x = 1$). Some survival times are censored ($d = 0$). First, we plot Kaplan-Meier estimates of the survival functions.

{{2}}

We consider a log-logistic model (AFT) for $f(y|x)$ (see slides 112). Estimate the model's parameters. Constrain the parameter λ to be positive. Take into account right censoring (see slide 76).

The PDF of a (standard) logistic distribution is:

$$f(y) = \frac{\exp(-y)}{(1 + \exp(-y))^2}$$

while the Survival function is:

$$S(y) = 1 - \frac{1}{1 + \exp(-y)}$$

{{3}}

Plot the estimated densities $\hat{f}(y|x)$.

{{4}}

Plot the estimated survival functions $\hat{S}(y|x)$ together with the Kaplan-Meier estimates.

{{5}}

Exercise 2

Use a RCS transformation of time to death to make the log-logistic model more flexible (see slide 118).

Generate a RCS transform of y (V2) and its derivative (v2) using `rcsgen`. Estimate the model's parameters. Constrain the parameter λ to be positive

{{6}}

Plot the estimated densities $\hat{f}(y|x)$

{{7}}

Plot the estimated survival functions $\hat{S}(y|x)$ over the Kaplan-Meier estimates.

{{8}}

Exercise 3

We consider a Weibull model (PH) (see slide 124). Estimate the model's parameters. Constrain the parameter k to be positive.

But first: we need to extend the likelihood on slide 123 in order to accomodate the censored observations.

The log-likelihood, in the presence of right censoring, is (see slides 76):

$$\log[L(\theta)] = \sum_{i=1}^{n} I(d_i = 1)\log[f(z_i)] + I(d_i = 0)\log[Sz_i].$$

Knowing that $f(y) = S(y)h(y)$ and $S(y) = \exp(-H(y))$ (see slide 121), we can rewrite it as

$$\log[L(\theta)] \doteq \sum_{i=1}^{n} I(d_i = 1)(\log[\exp(-H(z_i))] + \log[h(z_i)]) + I(d_i = 0)\log[\exp(-H(z_i))] = \sum_{i=1}^{n} I(d_i = 1)\log[h(z_i)] - H(z_i).$$

This equation justifies the form of the log-likelihood passed to `mlexp` in Exercise 4 and 5.

What's the interpretation of β_1?

{{9}}

Plot the estimated densities $\hat{f}(y|x)$

{{10}}

Exercise 4

We now use a RCS transformation of time to death to make the Weibull model more flexible (see slide 128).

{{11}}

Plot the estimated densities $\hat{f}(y|x)$

{{12}}

Plot the estimated survival functions $\hat{S}(y|x) = \exp(-H(y|x))$ over the Kaplan-Meier estimates.

{{13}}

Plot the model-based estimated hazards functions $\hat{h}(y|x)$. Use a log scale for the vertical axis to visually check that the model-based hazard functions are actually proportional. **Important**: the hazard functions are proportional because we forced them to be so (using a Weibull PH model)!

Now interpret them, if you can :-)

{{14}}

Extra

Can we fit a so-called "flexible parametric survival model"[1] using the tools we've learned so far? Of course. To us, it's just one possible way of modeling y.

Note that here we apply RCS transforms to $z = \log(y)$ instead of y.

(You'll need to install the command `stpm2`, first)

{{15}}

[1] Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in medicine, 21(15), 2175-2197.