Evaluating the Proportional Hazards Assumption (Chapter 4)

Thomas Cayé, Oscar Perez, Yin Zhang

March 20, 2011

1 Cox Proportional Hazards hypothesis

The Cox Proportional Hazard model gives an expression for the hazard at time t, as the product of a baseline hazard function (intuitively, what we have without explaining variables) and the exponential of a term linear in the predictors X_i's. The expression is divided in two terms : one depending on time only, and the other depending only on the X_i's. These X_i's are said to be time-independent. Here is the formula for p predictors X_i's. The β's are the coefficient of the parametrized part of the model.

$$h(t, \mathbf{X}) = h_0(t) \cdot \exp(\sum_{i=1}^{p} \beta_i \cdot X_i)$$

The Cox model can be extended to allows time dependent predictors.

This model provides simple formulas for different quantities :

- Hazard ratio : $\hat{HR} = \exp(\sum_{i=1}^{p} \beta_i \cdot (X_i^* - X_i))$ (for two series of observations for p predictors : X_i^*'s and X_i's)
- Survival function : $S(t, \mathbf{X}) = [S_0(t)]^{\exp(\sum_{i=1}^{p} \beta_i X_i)}$
- Adjusted survival curves : $\hat{S}(t, \mathbf{X}) = [\hat{S}_0(t)]^{\exp(\hat{\beta}_1 X_1 + \sum_{i=2}^{p} \hat{\beta}_i \bar{X}_i)}$

The PH assumption means that the hazard for an individual is proportional to the hazard of an other one. The proportionality constant depends on the predictors' values.

This assumption needs to be checked. Here are a few methods.

2 Graphical analysis

Graphical technics allow to see on a graph if the PH assumption looks valid, or if there are obvious reasons why it is not. A rather conservative strategy is necessary : reject only the extreme cases. Furthermore reading a graph, one can detect local criteria that contradict Cox hypothesis, while it could be missed by a global test.

2.1 Log-log Kaplan-Meier curves

$$-ln(-ln(S(t, \mathbf{X})) = -\sum_{i=1}^{p} \beta_i \cdot X_i - ln(-ln(S_0(t))$$

For two data sets, the distance between the curves must be constant over time. We can use the KM curves, obtained from the observations, or adjusted curves for predictors satisfying the assumption

2.2 Observed versus Expected plots

Plotting the observed curves (obtained with KM formula, or adjusted for other predictors), and the expected curve (the curve obtained when fitting a Cox model with a regression software) one can assess if the best curve we could get with a Cox model, looks like what we actually observed.

For continuous variables, one has two ways to obtain expected curves. The first is to categorize the observations and fit the PH model with $k-1$ dummy variables (for k categories) and obtain a survival curve for each of these categories. The second is to fit the model for the continuous variable and then use the mean value of the predictor to plot the curve.

3 Use of time-dependent covariates

Using the following model : $h(t, \mathbf{X}) = h_0(t)exp(\beta \cdot X + \gamma \cdot X \cdot g(t))$, for some time function g, $(t$, $log(t)$, Heaviside function), we test the significance of the product term. The null hypothesis is "$\gamma = 0$". One can use Wald, or likelihood ratio statistics : its distribution is then a χ^2 under the null hypothesis. It is possible to assess simultaneously several predictors, with different time function. The distribution of the likelihood ratio statistic is a χ^2 with p degree of freedom (number of predictors being assessed) under the null hypothesis. If the test is significant we reason by backward induction.

4 Goodness of fit : a statistical test

Residual : Observable estimate of the unobservable statistical error.

Here censorship of data, and different relation to time make it necessary to define new residuals. Scaled Schoenfeld residual are used :

$$\hat{\mathbf{r_i}}^* = [\hat{Var}(\hat{\mathbf{r_i}})]^{-1}\hat{\mathbf{r_i}}$$

where $\hat{r_{i,k}} = c_i(x_{i,k} - \hat{\hat{x_{i,k}}})$, the $\hat{\hat{x_{i,k}}}$ are weighted means :

$$\hat{\hat{x_{i,k}}} = \frac{\sum_{j \in R(t_i)} x_{j,k} e^{x'_j \beta}}{\sum_{j \in R(t_i)} e^{x'_j \beta}}$$

and $[\hat{Var}(\hat{\mathbf{r_i}})]^{-1}$is estimated by $m\hat{Var}(\hat{\beta})$ (m is the number of non-censored observations) and c_i is the indicator variable which equals 1 if the individual fails, or 0 if he is censored.

These residuals are tested for significance : they must not be correlated to ranked failure time. We can only reject H_0, and not prove the PH hypothesis.

A parallel graphical test can be conducted : indeed if one take time-varying coefficients β_i's : $\beta_j(t) = \beta_j + \gamma_j g_j(t)$ for some functions of time g_j's, it is possible to prove that $E[r_j^*(t)] \cong \gamma_j g_j(t)$. Therefore, plotting the curve of the scaled residuals against time allows to assess if the time independence of the coefficients seems reasonable. Experience shows that $g(t) = \ln(t)$ is a very good choice.

References

[1] Survival statistics : a self-learning text, Kleinbaum, D.G. and Klein, M.(2005)

[2] Applied Survival Analysis, Regression Modelling of time to event data, David W. Hosmer Jr., Stanley Lemeshow.