Cox ph model and m() function

2020-02-24

Estimator with covariates

When there are covariates that are needed to be considered in the model to estimate the survival time, the most commonly used method is Cox PH model, by

$$S^{cox}(t|X=x) = S_0(t)^{\exp(\beta' x)} \tag{1}$$

where $x = (x_1, .., x_p)', \beta = (\beta_1, ..., \beta_p)'$ are vectors. Also,

$$\lambda_T(t|x) = \lambda_0(t)\exp(\beta' x), \Lambda_T(t|x) = \Lambda_0(t)\exp(\beta' x) \tag{2}$$

Dikta proposed another semi-parametric estimator with consideration of covariates. The estimator is derivatived from the estimator generated by Stute (1993), instead of the product limit estimator from Kaplan Meier.

The Stute estimator (which estimates $E(\phi(x,t))$):

$$F_0(x,t) = P(X \le x, T \le t) = \sum_{i=1}^{n} W_i \phi(x,t) \tag{3}$$

If we set $\phi(x,t) = I_{(X\le x, T\le t)}$, it leads to

$$F_0(x,t) = \sum_{i=1}^{n} W_i I_{(X\le x,T\le t)} = \sum_{i=1}^{n}\Big\{\frac{\delta_i}{n-i+1}\Big[\prod_{j=1}^{i-1}(1-\frac{1}{n-j+1})^{\delta_i}\Big]\Big\} I_{(X\le x,T\le t)} \tag{4}$$

where $W_i = \frac{\delta_i}{n-i+1}\prod_{j=1}^{i-1}(1-\frac{1}{n-j+1})^{\delta_i}$.

Question: when $i=1$, what is the production value?

Dikta replaced δ_i with $m(t,x)$ and got

$$S^{D3} = \sum_{i=1}^{n}\Big\{\frac{m(Z_i,X_i)}{n-i+1}\Big[\prod_{j=1}^{i-1}(1-\frac{m(Z_i,X_i)}{n-j+1})\Big]\Big\} I_{(X\le x,T\le t)} \tag{5}$$

Therefore, the esimators of Cox PH model and of Dikta's model are actually estimating different things, i.e.

Cox

$$S^{cox}(t|X=x) = P(T>t|X=x)$$

Dikta

$$S^{D3} = P(T < t, X < x) = P(T < t, X_1 < x_1, ..., X_p < x_p)$$

Should $S^{cox}(t|X = x)$ and $S^{D3} = P(T < t, X < x)$ close to each other?

We then need some transformations. Since,

$$\begin{aligned}\int_0^t f_{T,X}(t,x)dt =& \int_0^t f_{T|X}(t|x)f_X(x)dt \\ =& f_X(x)\int_0^t f_{T|X}(t|x)dt \\ =& f_X(x)P(T < T|X = x)\end{aligned}$$

then

- $P(T < t|X = x) = \frac{1}{f_X(x)}\int_0^t f_{T,X}(t,x)dt$
- $f_{T,X}(t,x) = f_X(x)\frac{\partial}{\partial t}P(T < t|X = x)$
- $$\begin{aligned}F_{T,X}(t,x) =& \int_0^x\int_0^t f_{T,X}(t,x)dtdx \\ =& \int_0^x\int_0^t f_X(x)\frac{\partial}{\partial t}P(T < t|X = x)dtdx \\ =& \int_0^x f_X(x)P(T < t|X = x)dx\end{aligned}$$

Then for example, if we have one dimension covariate $X \sim N(0,1)$. The conditional distribution

$$\begin{aligned}P(T > t|X = x) =& S(t|x) = S_0(t)^{\exp(x)} = \exp(-t \times \exp(x)) \\ P(T < t|X = x) =& 1 - P(T > t|X = x) = 1 - \exp(-t \times \exp(x)) \\ F_{T,X}(t,x) =& \int_0^x f_X(x)P(T < t|X = x)dx \\ =& \int_0^x \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})(1 - \exp(-t \times \exp(x)))dx \\ =& \Phi(x) - \int_0^x \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2} - t\exp(x))dx\end{aligned}$$

which does not have a close form.

Uniform setting

To make the calculation easy, let's try to have X generated from other distributions. Suppose $X \sim UNI(0,1)$We then have

$$\begin{aligned}f_{T,X}(t,x) =& f_X(x)\frac{\partial}{\partial t}P(T < t|X = x) \\ =& 1 \times \exp(x)\exp(-t \times \exp(x))\end{aligned}$$

$$F_{T,X}(t,x) = \frac{1}{\exp(x)} - \frac{1}{\exp(x)}\exp(-t \times \exp(x)) = \exp(-x) - \exp(-x - t\exp(x))$$