Prediction?

2020-02-02

Mimic the simulation in Dikta's paper.

Assume the joint distribution

$$S_{T,C}(t, s|x) = \begin{cases} \exp\big(-(\exp(-\beta^T x))t\big)\exp\big(-s(t-s+1)\big) & t \geq s \\ \exp\big(-(\exp(-\beta^T x))t\big)\exp\big(-s\big) & t < s \end{cases}$$

The censoring percentage is When $t < s$, the pdf is

$$\frac{\partial}{\partial s \partial t}\exp\big(-(\exp(-\beta^T x))t\big)\exp\big(-s\big) = \exp(-\beta^T x) \times \exp\big(-(\exp(-\beta^T x))t\big)\exp\big(-s\big)$$

$$\begin{aligned} P(T < C) &= \int_0^\infty \int_0^s f_{T,C}(t,s)dtds \\ &= \int_0^\infty \int_0^s \exp(-\beta^T x) \times \exp\big(-(\exp(-\beta^T x))t\big)\exp\big(-s\big)dtds \\ &= \int_0^\infty \big\{-\exp\big(-(\exp(-\beta^T x))t\big)\exp\big(-s\big)|_0^s\big\}ds \\ &= \frac{\exp(-\beta^T x)}{1+\exp(-\beta^T x)} \end{aligned}$$

The survival functions are:

$$S_T(t) = \exp\big(-\exp(-\beta^T x)t\big), S_C(t) = \exp\big(-s\big)$$

The $m()$ function is

$$m(t, x) = \frac{1}{1+\exp(-\beta^T x)}$$

which can be estimated by fitting a logistic regression.

Simulation

Consider $n = 500$, $x = 0.1, \beta = 2$. Generate datasets.

Table 1: Mean absolute difference between estimated and true S()

	KM	Exp m()	Dikta 1	Dikta 2
q0.1	0.036	0.143	0.144	0.144
q0.25	0.047	0.135	0.135	0.136
q0.5	0.037	0.016	0.016	0.015
q0.75	0.028	0.127	0.126	0.125
q0.9	0.018	0.150	0.148	0.147

Table 2: Standard deviations of the estimated S()

	KM	Exp m()	Dikta 1	Dikta 2
q0.1	0.012	0.011	0.011	0.011
q0.25	0.025	0.010	0.010	0.010
q0.5	0.032	0.015	0.015	0.015
q0.75	0.029	0.018	0.019	0.019
q0.9	0.022	0.023	0.023	0.023

Table 3: MSE

	KM	Exp m()	Dikta 1	Dikta 2
q0.1	0.001	0.021	0.021	0.021
q0.25	0.003	0.018	0.018	0.019
q0.5	0.002	0.000	0.000	0.000
q0.75	0.001	0.016	0.016	0.016
q0.9	0.001	0.023	0.022	0.022