Covariates example

2019-12-17

Suppose the survival time T and the censoring time C are independent given covariates.

$$S_T(t|x) = \exp(-(\beta^T x)t)$$

where that, given the covariates, the survival time follows an exponential distribution. Also can be fitted by a cox model. For the censoring time,

$$S_c(t|x) = e^{-(e^{(\beta^T x)t}-1)}$$

However, this is not a cox ph model.

We can replace $\beta^T x$ with θ, then

$$S_T(t|\theta) = \exp(-\theta t), S_c(t|\theta) = e^{-(e^{\theta t}-1)}$$

$$P(T > t, C > t|x) = P(T > t|x)P(C > t|x) = e^{-(\beta^T x)t} e^{-(e^{(\beta^T x)t}-1)}$$

The $m()$ function is:

$$m(t|x) = \frac{\lambda_F(t|x)}{\lambda_H(t|x)} = \frac{f_T(t|x)}{S_T(t|x)} / \frac{f_Z(t|x)}{S_Z(t|x)} = \frac{\theta e^{-\theta t}}{e^{-\theta t}} / \frac{\theta(1+e^{\theta t})e^{-e^{\theta t}-\theta t+1}}{e^{-e^{\theta t}-\theta t+1}} = \frac{1}{1+e^{\theta t}}$$

$$\theta = \beta^T x$$

Therefore, the β can be estimated by fit logistic regression model with

$$logit(\delta) \sim (\beta^T x)t$$

For the censoring percentage, it is not affected by the covariates, since the $P(T < C)$ does not contains θ.

$$\begin{aligned}
P(T < C) =& \int_0^\infty \int_0^y f_{t,c}(x,y)dxdy \\
=& \int_0^\infty \int_0^y \theta^2 e^{-e^{\theta y}+\theta y-\theta x+1} dxdy \\
=& \int_0^\infty \theta(e^{\theta y}-1)e^{1-e^{\theta y}} \\
=& 1 - e\Gamma(0,1) \\
\approx& 0.4
\end{aligned}$$

For the true survival time $S_T(t)$ and $f_T(t)$,

$$f_T(t) = \int_{-\infty}^{\infty} f_T(t|x)g(x)dx,$$

where $g(x)$ is the distribution of $\beta^T x$.

Simulation

Suppose $\beta^T x = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, where $\beta_0 = 2, \beta_1 = 0.5, \beta_2 = 0.5$. $x_1 \sim N(1, 0.5^2), x_2 \sim N(2, 0.5^2)$.

Suppose x_1 and x_2 are independent, therefore, $\beta^T x \sim N(3.5, 0.5^3)$, i.e. $g(x)$ is known.

Sample size = 500, 500 iterations were conducted.

For data generation,

- Simulate $x_1 \sim N(1, 0.5^2), x_2 \sim N(2, 0.5^2)$ and then calculate the $\theta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
- For each *theta* above, simulate one survival time from $Exp(\theta)$, simulate one cenorsing time from extreme distribution, with cdf $S_c(t|\theta) = e^{-(e^{\theta t}-1)}$
- Combine the survival time and censoring time, calculate the observed time $Z = T \wedge C$ and status indicator $\delta = I(T < C)$
- Save the dataset with $T, C, Z, x_1, x_2, \theta$

The $\hat{S}(t)$ was estimated by Kaplan Meier, $exp(-m(t)\lambda_H(t))$, Dikta method 1, Dikta method 2.

The true $S(t)$

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_T(t|x)g(x)dx \\ &= \int_{-\infty}^{\infty} \beta^T x \exp(-\beta^T xt) \frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2)dx \end{aligned}$$

However, it is hard to calculate unless $\mu = 0$. But if $\mu = 0$, the $\beta^T x$ can be easily less then 0, which then cannot simulate $\exp(-\beta^T x)$.

Howver, the $\exp(-\bar{\beta^T} x)$ looks similar with the true $S(t)$

Results

The mean value of the estimations of the quantiles of the KM, exponential m(), Dikta method 1 and Dikta method 2 were calculated.

The cells in the table are the mean value of the estimations over 500 iterations minus the true quantiles value i.e (0.9, 0.75, 0.5, 0.25, 0.1)

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

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	0.01084	0.00756	0.00757	0.00757	0.01084	0.00754	0.00755	0.00755
t0.25	0.01695	0.01158	0.01159	0.01161	0.01695	0.01179	0.01180	0.01180
t0.5	0.02345	0.01418	0.01411	0.01408	0.02345	0.01905	0.01892	0.01884
t0.75	0.03964	0.02121	0.02101	0.02131	0.03964	0.03448	0.03430	0.03446
t0.9	0.10000	0.06567	0.09902	0.09835	0.10000	0.07027	0.09907	0.09846

To make the table easy to look at, I used the column 2,3,4,5 to divide the column 1, column 7,8,9,10 to divide column 6.

The values that are less than 1 are showing that the methods have less bias than the KM.

Table 2: Mean absolute difference between estimated and true S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	1	0.69766	0.69822	0.69859	1	0.69589	0.69642	0.69678
t0.25	1	0.68310	0.68415	0.68487	1	0.69597	0.69607	0.69624
t0.5	1	0.60481	0.60187	0.60067	1	0.81253	0.80693	0.80349
t0.75	1	0.53506	0.53011	0.53763	1	0.86987	0.86517	0.86937
t0.9	1	0.65675	0.99028	0.98356	1	0.70273	0.99070	0.98468

The standard deviation of the estimations of each quantiles are reported in the following table.

Table 3: Standard deviations of the estimated S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	0.01362	0.00947	0.00948	0.00949	0.01362	0.00945	0.00946	0.00946
t0.25	0.02113	0.01443	0.01444	0.01445	0.02113	0.01468	0.01469	0.01471
t0.5	0.02928	0.01748	0.01754	0.01757	0.02928	0.02310	0.02315	0.02317
t0.75	0.04859	0.02441	0.02567	0.02688	0.04859	0.04168	0.04252	0.04320
t0.9	0.09232	0.01099	0.03937	0.03645	0.09232	0.04420	0.04060	0.03785

To make the table easy to look at, I used the column 2,3,4,5 to divide the column 1, column 7,8,9,10 to divide column 6.

The values that are less than 1 are showing that the methods have less standard deviation than the KM.

Table 4: Standard deviations of the estimated S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	1	0.69572	0.69641	0.69682	1	0.69398	0.69468	0.69508
t0.25	1	0.68257	0.68325	0.68379	1	0.69469	0.69534	0.69586
t0.5	1	0.59724	0.59901	0.60031	1	0.78910	0.79062	0.79156
t0.75	1	0.50235	0.52827	0.55324	1	0.85774	0.87507	0.88919
t0.9	1	0.11900	0.42647	0.39483	1	0.47880	0.43974	0.40998

The MSE of each estimation

Table 5: MSE

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	0.00019	0.00009	0.00009	0.00009	0.00019	0.00009	0.00009	0.00009
t0.25	0.00045	0.00021	0.00021	0.00021	0.00045	0.00022	0.00022	0.00022
t0.5	0.00086	0.00032	0.00031	0.00031	0.00086	0.00056	0.00056	0.00055
t0.75	0.00246	0.00069	0.00068	0.00072	0.00246	0.00189	0.00187	0.00189
t0.9	0.01326	0.00443	0.00983	0.00973	0.01326	0.00676	0.00992	0.00982

To make the table easy to look at, I used the column 2,3,4,5 to divide the column 1, column 7,8,9,10 to divide column 6.

The values that are less than 1 are showing that the methods have less MSE than the KM.

Table 6: MSE

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	1	0.48399	0.48488	0.48544	1	0.48172	0.48252	0.48303
t0.25	1	0.46692	0.46714	0.46765	1	0.48563	0.48523	0.48537

t0.5	1	0.36827	0.36441	0.36299	1	0.65033	0.64338	0.63977
t0.75	1	0.27853	0.27749	0.29388	1	0.76873	0.75991	0.76577
t0.9	1	0.33440	0.74122	0.73379	1	0.50951	0.74810	0.74079

Estimate of β

Table 7: MSE

True b0 = 2	True b1= 0.5	True b2 = 0.5
2.137	0.518	0.474

Estimation of $m()$

```r
mean(result$m_diff)
```

```
## [1] 0.02232238
```

Discrete setting

Since the pdf of the $S_T(t)$ is not easy to calculate, in previous results with continuous $\beta^T x$ distribution. Let's then make it simple.

Suppose $\beta^T x = b_0 + b_1 x$, where $b_0 = 1, b_1 = 0.3$, $x \sim$ discrete distribution:

Table 8: MSE

x	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
beta x	1.3	1.6	1.9	2.2	2.5	2.8	3.1	3.4	3.7	4.0
probability	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Then the marginal survival time $f_T(t)$ is,

$$\begin{aligned} f_T(t) =& \sum_{x=1}^{10} f_T(t|x)g(x) \\ =& \frac{1}{10}[1.3\exp(-1.3t) + 1.6\exp(-1.6t) + 1.9\exp(-1.9t) + ... + 4\exp(-4t)] \end{aligned}$$

$$\begin{aligned} F_T(t) =& \int_{-\infty}^{+\infty} f(t)dt \\ =& \frac{1}{10}[\exp(-1.3t) + \exp(-1.6t) + \exp(-1.9t) + ... + \exp(-4t)] \end{aligned}$$

The other settings are the same as the previous example.

Results

The mean value of the estimations of the quantiles of the KM, exponential m(), Dikta method 1 and Dikta method 2 were calculated.

The cells in the table are the mean value of the estimations over 500 iterations minus the true quantiles value i.e (0.9, 0.75, 0.5, 0.25, 0.1)

Table 9: Mean absolute difference between estimated and true S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	0.01172	0.00773	0.00774	0.00775	0.01172	0.00775	0.00776	0.00777
t0.25	0.01807	0.01119	0.01115	0.01113	0.01807	0.01177	0.01175	0.01173
t0.5	0.02787	0.02220	0.02167	0.02136	0.02787	0.02348	0.02305	0.02279
t0.75	0.05478	0.05218	0.04971	0.04806	0.05478	0.05347	0.05138	0.04996
t0.9	0.06658	0.05197	0.07234	0.06487	0.06658	0.05889	0.07462	0.07011

To make the table easy to look at, I used the column 2,3,4,5 to divide the column 1, column 7,8,9,10 to divide column 6.

The values that are less than 1 are showing that the methods have less bias than the KM.

Table 10: Mean absolute difference between estimated and true S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	1	0.65960	0.66057	0.66108	1	0.66101	0.66195	0.66245
t0.25	1	0.61900	0.61702	0.61611	1	0.65115	0.64989	0.64921
t0.5	1	0.79684	0.77778	0.76653	1	0.84273	0.82706	0.81800
t0.75	1	0.95256	0.90760	0.87731	1	0.97616	0.93803	0.91218
t0.9	1	0.78057	1.08661	0.97440	1	0.88446	1.12078	1.05305

The standard deviation of the estimations of each quantiles are reported in the following table.

Table 11: Standard deviations of the estimated S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	0.01454	0.00984	0.00985	0.00986	0.01454	0.00987	0.00988	0.00989
t0.25	0.02203	0.01338	0.01340	0.01341	0.02203	0.01398	0.01400	0.01401
t0.5	0.02994	0.01786	0.01792	0.01796	0.02994	0.02195	0.02201	0.02204
t0.75	0.04266	0.02389	0.02442	0.02480	0.04266	0.03600	0.03641	0.03664
t0.9	0.06845	0.02769	0.08034	0.07647	0.06845	0.04794	0.08708	0.08357

To make the table easy to look at, I used the column 2,3,4,5 to divide the column 1, column 7,8,9,10 to divide column 6.

The values that are less than 1 are showing that the methods have less standard deviation than the KM.

Table 12: Standard deviations of the estimated S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	1	0.67669	0.67745	0.67784	1	0.67884	0.67960	0.67999
t0.25	1	0.60731	0.60819	0.60866	1	0.63475	0.63564	0.63610
t0.5	1	0.59670	0.59862	0.59982	1	0.73339	0.73524	0.73612

t0.75	1	0.55987	0.57230	0.58132	1	0.84370	0.85346	0.85878
t0.9	1	0.40458	1.17371	1.11719	1	0.70035	1.27214	1.22082

The MSE of each estimation

Table 13: MSE

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	0.00021	0.00010	0.00010	0.00010	0.00021	0.00010	0.00010	0.00010
t0.25	0.00049	0.00019	0.00019	0.00019	0.00049	0.00020	0.00020	0.00020
t0.5	0.00115	0.00067	0.00064	0.00063	0.00115	0.00080	0.00078	0.00076
t0.75	0.00421	0.00321	0.00297	0.00281	0.00421	0.00384	0.00361	0.00345
t0.9	0.00678	0.00346	0.00646	0.00584	0.00678	0.00512	0.00758	0.00697

To make the table easy to look at, I used the column 2,3,4,5 to divide the column 1, column 7,8,9,10 to divide column 6.

The values that are less than 1 are showing that the methods have less MSE than the KM.

Table 14: MSE

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
t0.1	1	0.45839	0.45981	0.46059	1	0.46143	0.46290	0.46369
t0.25	1	0.38768	0.38496	0.38364	1	0.41815	0.41581	0.41465
t0.5	1	0.58173	0.55801	0.54428	1	0.69680	0.67453	0.66141
t0.75	1	0.76209	0.70416	0.66669	1	0.91101	0.85567	0.81904
t0.9	1	0.51063	0.95334	0.86122	1	0.75528	1.11830	1.02844

Estimate of β

```
temp = c(mean(-result$b0_est), mean(-result$b1_est))
temp = round(temp,2)
temp = matrix(temp,1,2)
colnames(temp) = c('True b0 = 1','True b1= 0.3')
kable(temp, "latex", longtable = T, align="c",
      caption = 'MSE',
      booktabs=TRUE, escape = F)
```

Table 15: MSE

True b0 = 1	True b1= 0.3
1.03	0.29

Estimation of $m()$

```
## [1] 0.0181984
```