Bootstrapping of variance

2019-11-18

Data generation

CDF of the death time and censor time

$$S(T \geq x, C \geq y) = \begin{cases} e^{-\theta x} e^{-(e^{\theta y}-1)\left((\theta x-\theta y)^2+1\right)} & x \geq y \\ e^{-\theta x} e^{-(e^{\theta y}-1)} & x < y \end{cases}$$

And

$$S_T(t) = P(T > t) = P(T > t, C > 0) = e^{-\theta t} e^{-(e^{\theta 0}-1)\left((t-0)^2+1\right)} = e^{-\theta t}$$

$$f_T(t) = \frac{\partial}{\partial t}(1 - S_T(t)) = \frac{\partial}{\partial t}(1 - e^{-\theta t}) = \theta e^{-\theta t}$$

$$S_x(t) = P(T > t, C > t) = e^{-\theta t} e^{-(e^{\theta t}-1)} = e^{-e^{\theta t}-\theta t+1}$$

$$f_x(t) = \frac{\partial}{\partial t}(1 - S_x(t)) = 1 - e^{-e^{\theta t}-\theta t+1} = \theta(1 + e^{\theta t}) e^{-e^{\theta t}-\theta t+1}$$

$$\psi(t) = \int_t^\infty f(t, c) dc = \int_t^\infty \theta^2 e^{-e^{\theta c}+\theta c-\theta t+1} dc = \theta e^{-e^{\theta t}-\theta t+1}$$

Therefore, the $m()$ function is:

$$m(t) = \frac{\lambda_F(t)}{\lambda_H(t)} = \frac{f_T(t)}{S_T(t)} / \frac{f_X(t)}{S_X(t)} = \frac{\theta e^{-\theta t}}{e^{-\theta t}} / \frac{\theta(1 + e^{\theta t}) e^{-e^{\theta t}-\theta t+1}}{e^{-e^{\theta t}-\theta t+1}} = \frac{1}{1 + e^{\theta t}}$$

And for the $\rho()$ function,

$$\begin{aligned} \rho =& \frac{f(t)/\psi(t) - 1}{S(t)/S_x(t) - 1} \\ =& \frac{\theta e^{-\theta t}/(\theta e^{-e^{\theta t}-\theta t+1}) - 1}{e^{-\theta t}/e^{-e^{\theta t}-\theta t+1} - 1} \\ =& 1 \end{aligned}$$

PDF of the death time and censor time

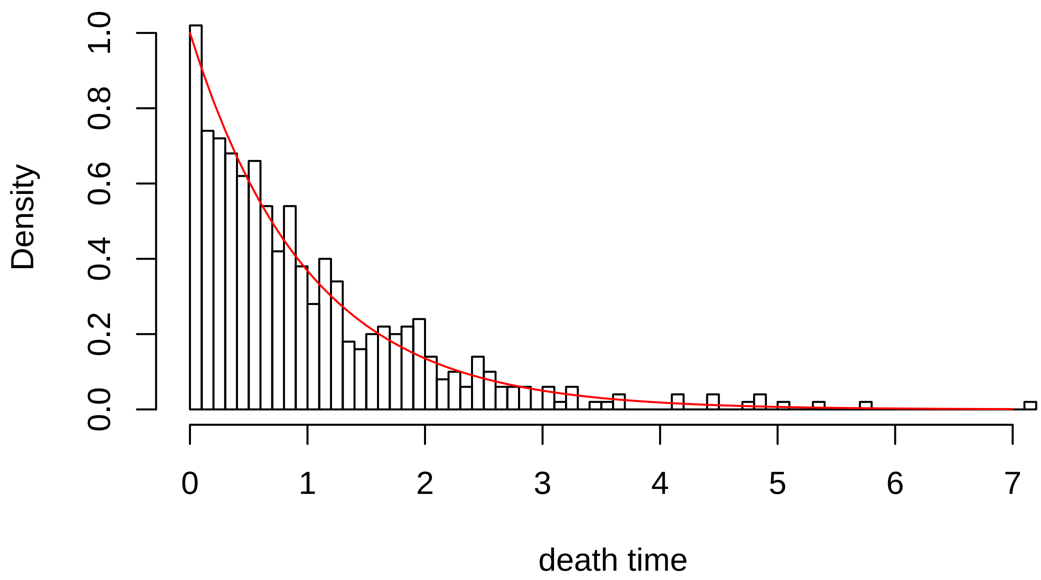
$$f_{T,C}(x, y) = \begin{cases} ((2\theta^2 x - 2\theta^2 y)(1 - e^{\theta y}) - \theta)((2\theta^2 y - 2\theta^2 x)(1 - e^{\theta y}) - \theta(\theta^2 y^2 - 2\theta^2 xy + \theta^2 x^2 + 1) e^{\theta y}) \times \\ e^{(\theta^2 y^2 - 2\theta^2 xy + \theta^2 x^2 + 1)(1 - e^{\theta y}) - \theta x} \\ + (\theta(2\theta^2 y - 2\theta^2 x) e^{\theta y} - 2\theta^2(1 - e^{\theta y})) e^{(\theta^2 y^2 - 2\theta^2 xy + \theta^2 x^2 + 1)(1 - e^{\theta y}) - \theta x} & x \geq y \\ \\ \theta^2 e^{-e^{\theta y}+\theta y-\theta x+1} & x < y \end{cases}$$

Data generation, check margnial

For example, when theta = 1,

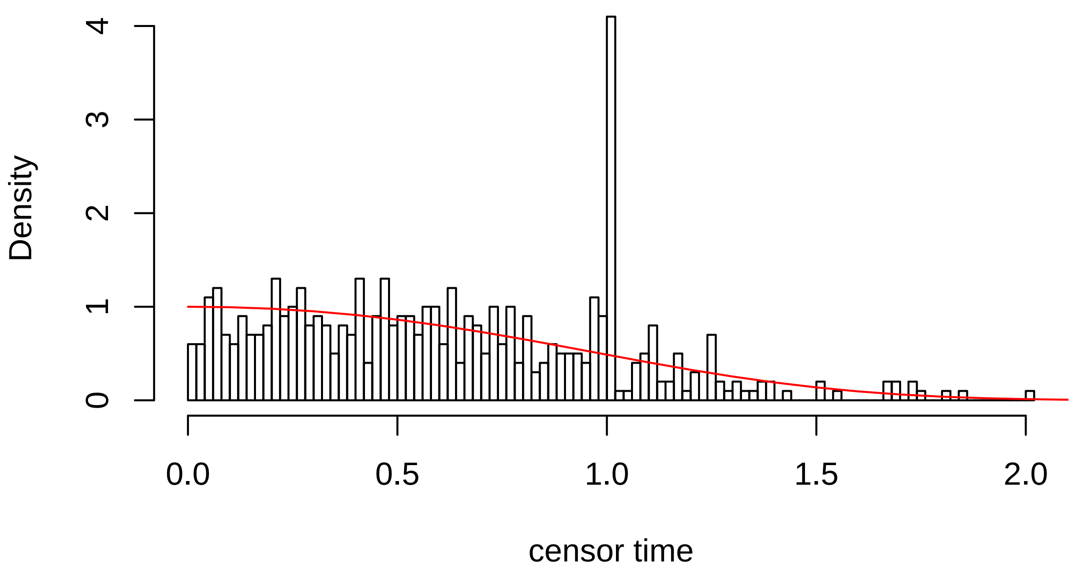
- Histogram of T estimation

The red line is the true density of T



- Histogram of C estimation

The red line is the true density of C



Settings:

$\theta = 0.8, 1, 1.5, 2, 5$

Results

Estimation of θ by logistic regression

	0.8	1	1.5	2	5
Estimated theta	0.2109	0.6887	1.5115	2.0562	5.1313

Estimation of $m()$ by logistic regression

Plug in the estimated $\hat{\theta}$ to get the estimated $\hat{m}(t)$. Calculate the mean absolute difference between true $m(t)$ and $\hat{m}(t)$.

	0.8	1	1.5	2	5
|hat m(t) - m(t)|	0.0598	0.0268	0.0088	0.0086	0.0086

Estimation of $S(t)$ with true $m(t)$

	KM	new m()	Dikta1	Dikta2
0.8	0.0532	0.0635	0.0640	0.0643
1	0.0308	0.0342	0.0346	0.0350
1.5	0.0148	0.0118	0.0120	0.0120
2	0.0139	0.0122	0.0124	0.0123
5	0.0143	0.0155	0.0170	0.0168

Estimation of $S(t)$ with logistic regression estimated $m(t)$

	KM	new m()	Dikta1	Dikta2
0.8	0.0532	0.0648	0.0654	0.0658
1	0.0308	0.0351	0.0357	0.0361
1.5	0.0148	0.0123	0.0126	0.0127
2	0.0139	0.0130	0.0133	0.0132
5	0.0143	0.0159	0.0173	0.0172

Variances of the quantiles of $S(t)$

Let's look at the variance of the estimation: S(t) at the '10th', '20th', '50th', '125th', '250th', '325th', '400th' out of 500 subjects.

The following table shows the fraction of standard deviation of new methods over Kaplan Meier estiamter: $\frac{v}{v_{km}}$.

1. Estimated by $\lambda_F(t) = m(t)\lambda_H(t)$ $(\frac{v}{v_{km}})$

```
## Warning in rep(linesep, length.out = nrow(x) - 2): 'x' is NULL so the
## result will be NULL
```

	With true m()					Logistic estimated m()				
	0.8	1	1.5	2	5	0.8	1	1.5	2	5
10th	0.0499	0.0436	0.0332	0.0421	0.0624	0.0503	0.0437	0.0334	0.0422	0.0623
20th	0.0749	0.0595	0.0402	0.0489	0.0883	0.0760	0.0602	0.0406	0.0494	0.0883
50th	0.1301	0.1103	0.0693	0.0759	0.1304	0.1350	0.1130	0.0710	0.0777	0.1312
125th	0.2358	0.2032	0.1282	0.1337	0.2026	0.2576	0.2155	0.1384	0.1512	0.2138
250th	0.4663	0.4262	0.2936	0.2994	0.4047	0.5331	0.4645	0.3183	0.3429	0.4381
325th	0.6493	0.5570	0.4527	0.4816	0.5953	0.7473	0.6055	0.4885	0.5407	0.6401
400th	0.8087	0.7945	0.6155	0.7220	0.7499	0.9142	0.8580	0.6595	0.7944	0.7922

2. Estimated by Dikta's 1st formula ($\frac{v}{v_{km}}$)

```
## Warning in rep(linesep, length.out = nrow(x) - 2): 'x' is NULL so the
## result will be NULL
```

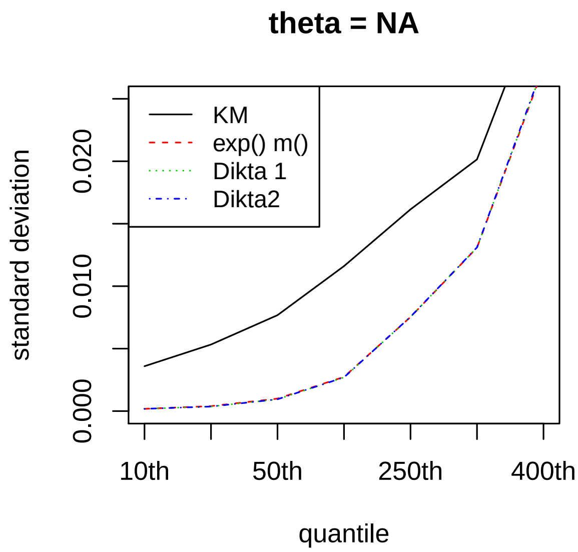
	With true m()					Logistic estimated m()				
	0.8	1	1.5	2	5	0.8	1	1.5	2	5
10th	0.0500	0.0385	0.0194	0.0198	0.0285	0.0505	0.0387	0.0197	0.0201	0.0289
20th	0.0683	0.0550	0.0276	0.0278	0.0395	0.0694	0.0556	0.0283	0.0288	0.0407
50th	0.1232	0.1013	0.0513	0.0514	0.0688	0.1284	0.1043	0.0534	0.0549	0.0720
125th	0.2324	0.2000	0.1135	0.1064	0.1479	0.2554	0.2130	0.1211	0.1193	0.1573
250th	0.4668	0.4237	0.2868	0.2851	0.3517	0.5355	0.4621	0.3089	0.3161	0.3673
325th	0.6496	0.5555	0.4499	0.4720	0.5493	0.7503	0.6046	0.4800	0.5159	0.5701
400th	0.8116	0.7937	0.6137	0.7156	0.7088	0.9198	0.8584	0.6518	0.7692	0.7266

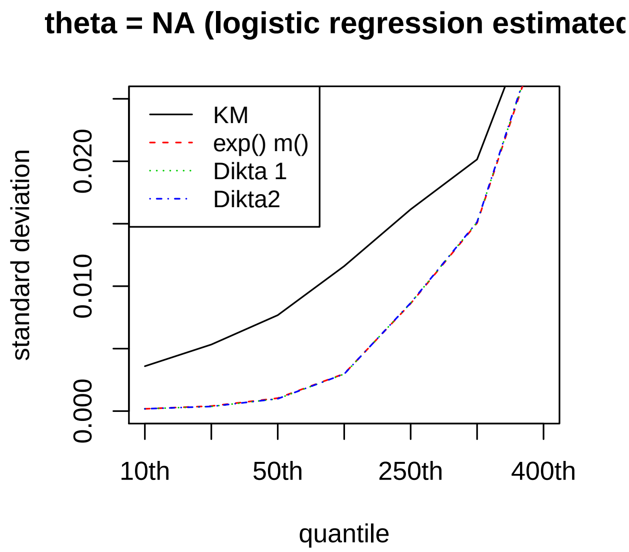
3. Estimated by Dikta's 2nd formula ($\frac{v}{v_{km}}$)

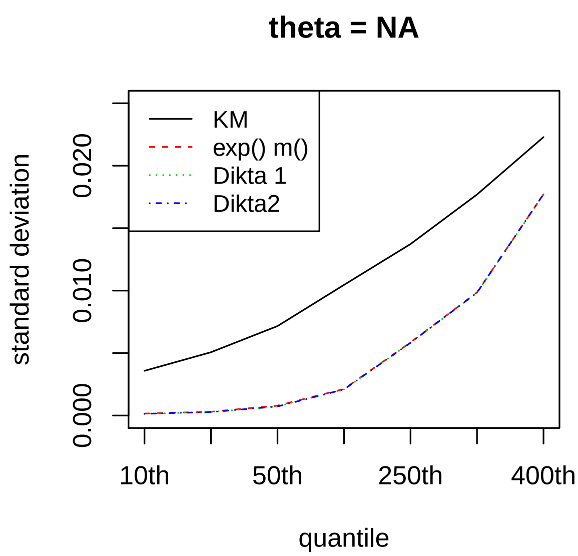
```
## Warning in rep(linesep, length.out = nrow(x) - 2): 'x' is NULL so the
## result will be NULL
```

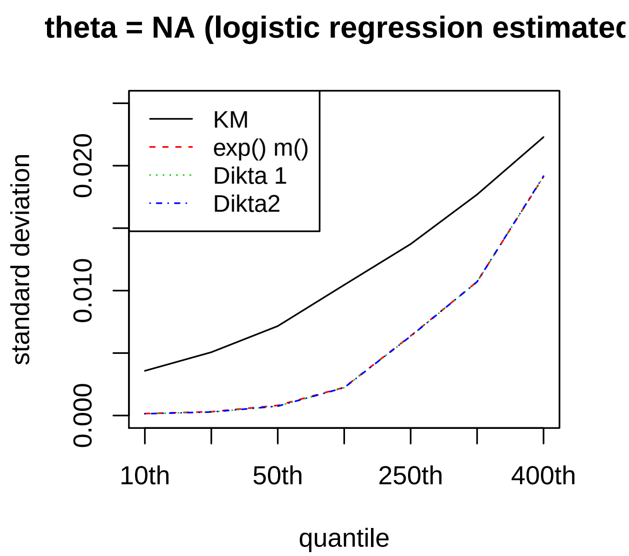
	With true m()					Logistic estimated m()				
	0.8	1	1.5	2	5	0.8	1	1.5	2	5
10th	0.0501	0.0385	0.0194	0.0198	0.0285	0.0505	0.0387	0.0197	0.0201	0.0289
20th	0.0683	0.0550	0.0276	0.0278	0.0395	0.0695	0.0556	0.0283	0.0288	0.0407
50th	0.1233	0.1014	0.0514	0.0515	0.0688	0.1286	0.1044	0.0535	0.0549	0.0721
125th	0.2326	0.2002	0.1136	0.1065	0.1480	0.2557	0.2132	0.1212	0.1194	0.1575
250th	0.4674	0.4241	0.2870	0.2852	0.3519	0.5360	0.4625	0.3090	0.3163	0.3676
325th	0.6504	0.5560	0.4501	0.4722	0.5497	0.7512	0.6052	0.4802	0.5161	0.5705
400th	0.8147	0.7951	0.6140	0.7160	0.7096	0.9227	0.8599	0.6521	0.7696	0.7274

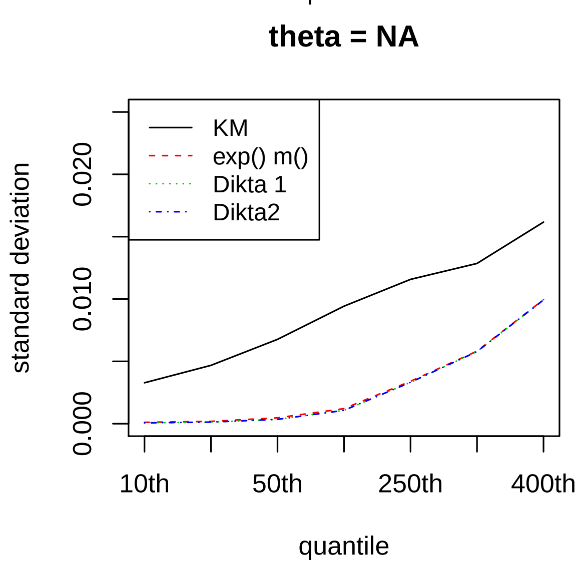
Plots of the variances

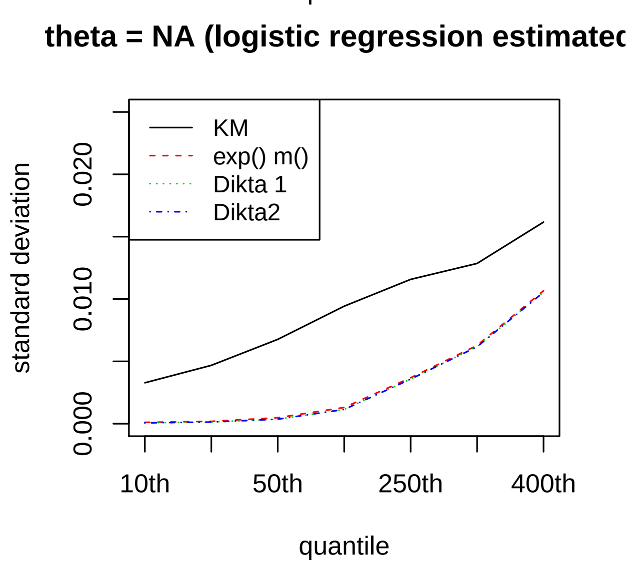


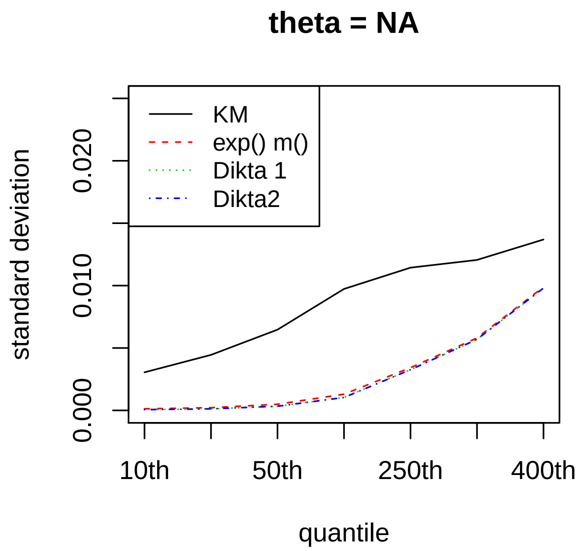


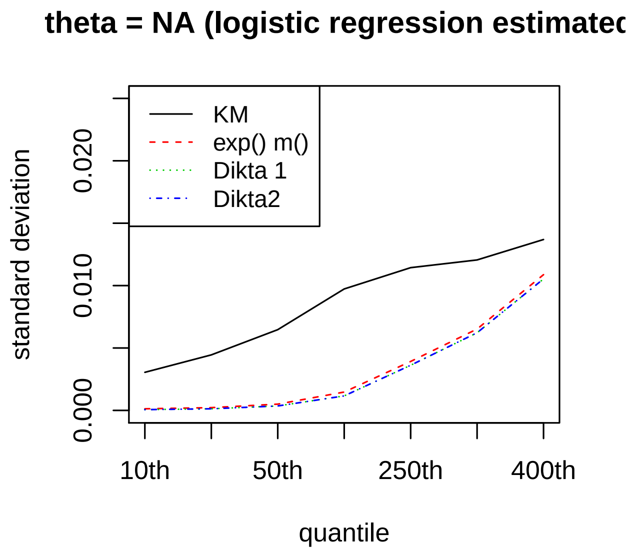


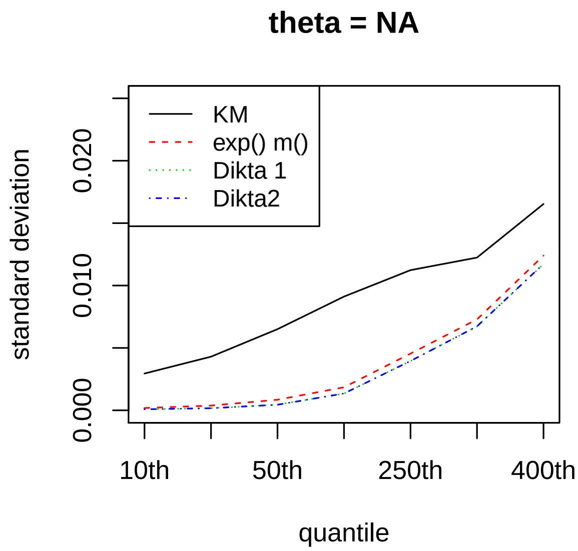


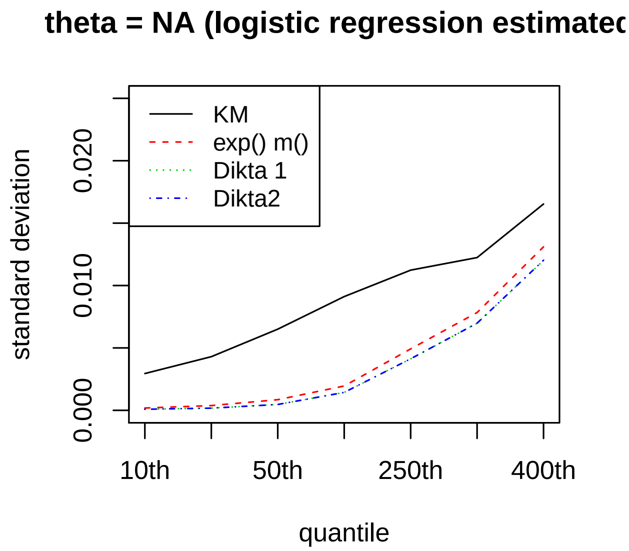





theta = NA
KM
exp() m()
Dikta 1
Dikta2
standard deviation
0.000
0.010
0.020
10th
50th
250th
400th
quantile


theta = NA (logistic regression estimatec
KM
exp() m()
Dikta 1
Dikta2
standard deviation
0.000
0.010
0.020
10th
50th
250th
400th
quantile


theta = NA
KM
exp() m()
Dikta 1
Dikta2
standard deviation
0.000
0.010
0.020
10th
50th
250th
400th
quantile


theta = NA (logistic regression estimatec
KM
exp() m()
Dikta 1
Dikta2
standard deviation
0.000
0.010
0.020
10th
50th
250th
400th
quantile