S(t) estimated by the accurate ODE solution

2019-08-14

Estimation of $S(t)$

In previous examples, when the $\rho(t)$ is small, the Slud's equation does not work very well even when the true $\rho(t)$ is known. This may be becasue they estimate the exponential function as $exp(x) \approx 1 + x$, and when $\rho(t)$ is small, the term inside the exponential function becomes large and the estimate is not accurate. Therefore, we may apply the $S(t)$ with the exponential function without linearization.

From Slud's paper, the $S(t)$ has the unique expression:

$$S(t) = exp\big[-\int_0^t \frac{\psi(s)\rho(s)}{Sx(s)}ds\big]\big(1 + \int_0^t \psi(s)\{\rho(s)-1\}exp\big[\int_0^s \frac{\psi(u)\rho(u)}{Sx(u)}du\big]ds\big) \tag{1}$$

And the term can be estimated as:

$$\begin{aligned}
exp\big[-\int_0^t \frac{\psi(s)\rho(s)}{Sx(s)}ds\big] =& exp\big[-\int_0^t \frac{\rho(s)}{Sx(s)}d\Psi(s)\big] \\
=& exp\big[-\int_0^t \frac{\rho(s)}{Sx(s)}P(I=1)d\frac{\Psi(s)}{P(I=1)}\big] \\
=& exp\big[-\int_0^t \frac{\rho(s)}{Sx(s)}P(I=1)d\Psi_c(s)\big] \\
\approx& exp\big[-\frac{1}{n_s}\sum_{0\le s\le t}\frac{\rho(s)}{Sx(s)}P(I=1)\big]
\end{aligned}$$

where $\Psi(t) = P(X < t, I = 1) = P(X < t|I = 1)P(I = 1) = \Psi_c(t)P(I = 1)$

Similiarly,

$$exp\big[\int_0^s \frac{\psi(u)\rho(u)}{Sx(u)}du\big] \approx exp\big[\frac{1}{n_u}\sum_{0\le u\le s}\frac{\rho(u)}{Sx(u)}P(I=1)\big]$$

.

Let $g(s) = exp\big[\frac{1}{n_s}\sum_{0\le u\le s}\frac{\rho(u)}{Sx(u)}P(I=1)\big]$

Then equation (1) can be

$$\begin{aligned}
S(t) =& g(s)^{-1} \times \big(1 + \int_0^t \{\rho(s)-1\}g(s)d\Psi(s)\big) \\
=& g(s)^{-1} \times \big(1 + \int_0^t \{\rho(s)-1\}g(s)P(I=1)d\Psi_c(s)\big) \\
=& g(s)^{-1} \times \big(1 + \sum_{0\le s\le t}\{\rho(s)-1\}g(s)P(I=1)\big)
\end{aligned} \tag{2}$$

A question here:

By using this method, the results do not get estimated very well. I think when we re-write $\int_0^t g(x)f(x)dx$ to $\int_0^t g(x)dF(x) = E(g(x)) \approx \frac{1}{n_x}\sum_{x\in(0,t)} g(x)$, the x need to come from the F(x) distribution. However, we only have the observed time and do not known its distribution. It may be not convenient to estimate it by using the expectation form.

I think it may be more convenient to use the Riemann integral idea to estiamte it. For example,

$$\int_0^t \frac{\psi(s)\rho(s)}{Sx(s)} ds \approx \frac{1}{n_{\Delta_h}} \sum_{i=1}^{n_{\Delta_h}} \frac{\psi(\Delta_h i)\rho(\Delta_h i)}{Sx(\Delta_h i)}$$

Just separate the $(0, t)$, say 100 equal pieces and estiamte. It can return accurate results.

Result

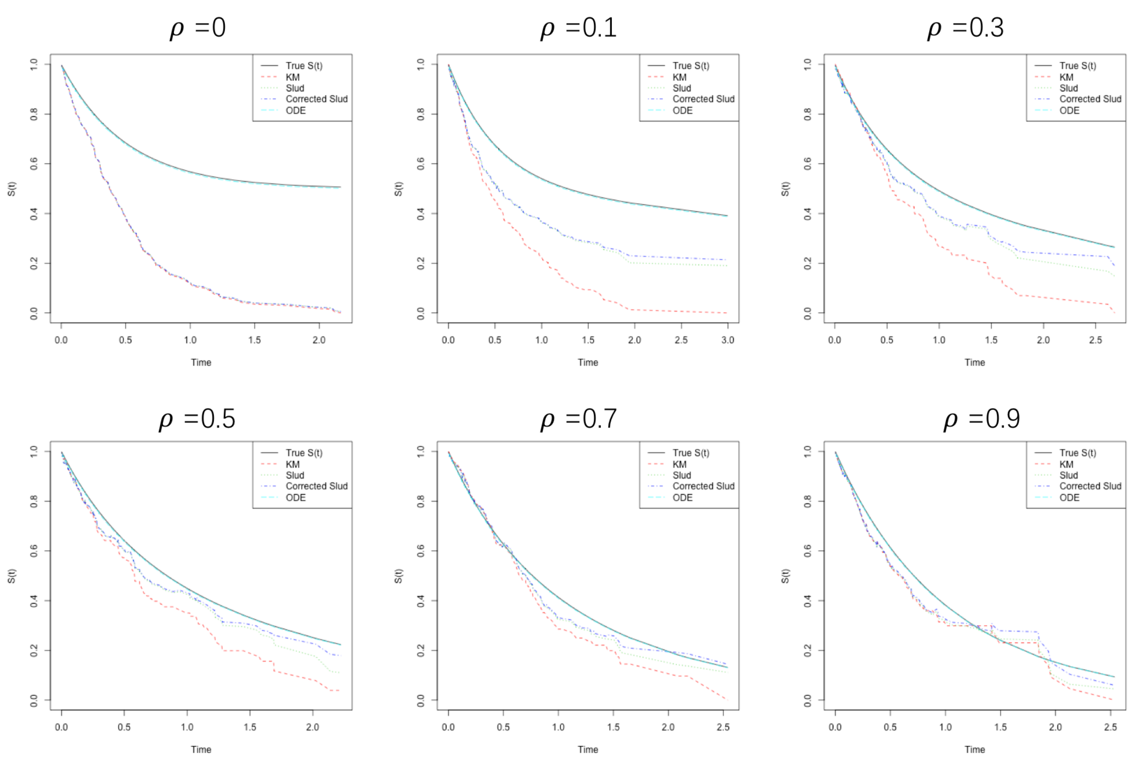
The mean absolute differences table

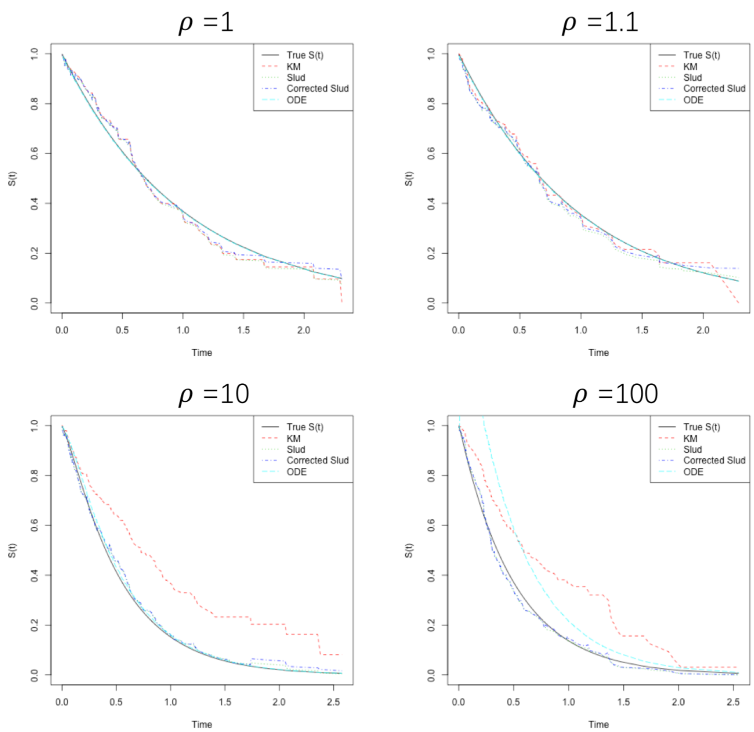
	0	0.1	0.3	0.5	0.7	0.9	1	10	100
KM	0.245	0.162	0.081	0.068	0.040	0.046	0.033	0.133	0.154
Slud	0.243	0.110	0.045	0.040	0.026	0.046	0.031	0.019	0.021
Corrected Slud	0.243	0.109	0.043	0.036	0.024	0.044	0.030	0.019	0.021
No_linear	0.005	0.005	0.005	0.005	0.004	0.004	0.004	0.018	0.281

The new estiamtor has the smallest differences with the true value in most of the scenarios. However, when $\rho(t)$ becomes very large, e.g. 100, the new estimator doesn't work well.

From the following plots, we may see that the line of the new estimator and the line of the true $S(t)$ are almost the same, except when $\rho = 100$

The plots





Iteration

I also tried the iteration method to estimate the ρ and $S(t)$. I simulated 200 subjects. The results depend on how may subjects I used to calculate the mean value.

Simulation setting

To make it easy to control, we would like to try Slud's piecewise example, which is:

$$f(t,s) = \begin{cases} f_1(t)f_C(s) & (t \leq s) \\ f_C(s)\dfrac{S_1(s)}{S_2(s)}f_2(t) & (t > s) \end{cases}$$

We could generate constant $\rho(t)$. Let

- $f_1(t) = exp(-t)$, $S_1(s) = exp(-x)$
- $f_C(s) = exp(-s)$, $S_C(s) = exp(-s)$
- $f_2(t) = \rho exp(-\rho t)$, $S_2(s) = exp(-\rho t)$
- $\rho(t) = \frac{h_2(t)}{h_1(t)} = 10$, which is a constant.

Then

$$f(t,s) = \begin{cases} exp(-t-s) & (t \leq s) \\ \rho exp(-10t+8s) & (t > s) \end{cases}$$

And

$$f(t) = \frac{9}{4}exp(-2t) - \frac{5}{4}exp(-10t)$$

$$S(t) = \frac{9}{8}exp(-2t) - \frac{1}{8}exp(-10t)$$

$$\psi(t) = exp(-2t), \; S_x(t) = exp(-2t)$$

Scenarios

- Scenario 1: I tried to calculate the mean by the 1st to 75% quantile subjects' values
- Scenario 2: I tried to calculate the mean by the 20th to 180th subjects' values
- Scenario 3: I tried to calculate the mean by the 25% quantile to 75% quantile subjects' values

Within those three scenarios, the first scenario's ρ converges the fastest. It can reach to 10. However, after that, it has some turbulence.

The second and the thrid scenario move very slowly and it seems stop at a ρ around 2.

However, if we give them a larger initial ρ, e.g. 4, it can still move towards to 10. (our initial ρ right now is from the KM and is about 0.8).

The ρ values for iterations

Scenario 1 (89 times iteration):

```
##  [1]  0.859  0.883  0.908  0.931  0.955  0.979  1.001  1.022  1.041  1.058
## [11]  1.073  1.088  1.100  1.113  1.126  1.140  1.154  1.167  1.178  1.188
## [21]  1.197  1.204  1.211  1.218  1.224  1.232  1.238  1.243  1.247  1.253
## [31]  1.257  1.261  1.265  1.268  1.269  1.271  1.272  1.272  1.273  1.273
## [41]  1.273  1.273  1.273  1.273  1.273  1.273  1.273  1.273  1.273  1.273
## [51]  4.429  4.181  4.035  3.896  7.774  5.518  4.798  4.421  4.176  4.030
## [61]  3.890  7.766  5.512  4.795  4.419  4.173  4.027  7.967  5.566 10.003
## [71]  5.905  4.999  9.200  5.828  4.955  4.484  4.211  4.014  7.945  5.571
## [81] 10.016  5.906  4.999  9.202  5.829  4.955  4.484  4.211  4.014
```

Scenario 2 (100 times iteration)

```
##   [1] 0.877 0.927 0.977 1.028 1.084 1.136 1.191 1.249 1.304 1.357 1.413
##  [12] 1.470 1.521 1.573 1.625 1.676 1.724 1.767 1.808 1.848 1.887 1.924
##  [23] 1.962 1.999 2.030 2.056 2.081 2.104 2.128 2.151 2.174 2.196 2.212
##  [34] 2.229 2.244 2.258 2.269 2.279 2.289 2.294 2.298 2.302 2.305 2.307
##  [45] 2.310 2.313 2.316 2.320 2.323 2.327 2.329 2.330 2.331 2.332 2.333
##  [56] 2.333 2.334 2.334 2.334 2.335 2.335 2.335 2.335 2.336 2.336 2.336
##  [67] 2.336 2.336 2.336 2.336 2.336 2.336 2.336 2.336 2.337 2.337 2.337
##  [78] 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337
##  [89] 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337 2.337
## [100] 2.337
```

Scenario 3 (50 times iteration)

```
##  [1] 0.859 0.883 0.908 0.931 0.955 0.979 1.001 1.022 1.041 1.058 1.073
## [12] 1.088 1.100 1.113 1.126 1.140 1.154 1.167 1.178 1.188 1.197 1.204
## [23] 1.211 1.218 1.224 1.232 1.238 1.243 1.247 1.253 1.257 1.261 1.265
## [34] 1.268 1.269 1.271 1.272 1.272 1.273 1.273 1.273 1.273 1.273 1.273
## [45] 1.273 1.273 1.273 1.273 1.273 1.273
```