New estimator with larger sample size

2019-08-20

Setting

Functions

Let $\rho(t) = 0.1$, the piecewise example:

$$f(t, s) = \begin{cases} exp(-t - s) & (t \leq s) \\ \rho exp(-0.1t - 1.9s) & (t > s) \end{cases}$$

And

$$f(t) = \frac{18}{19}exp(-2t) + \frac{10}{19}exp(-10t)$$

$$S(t) = \frac{9}{19}exp(-2t) - \frac{10}{19}exp(-\rho t)$$

$$\psi(t) = exp(-2t)$$

$$S_x(t) = exp(-2t)$$

The simulation:

- Generated dataset size: 1000
- Notation

True $S(t)$	Slud's estimator	Corrected Slud's estimator	$S(t)$ integral estimated by Riemann	$S(t)$ integral estimated by expectation
$S(t)$	$S_{p1}(t)$	$S_{p2}(t)$	$S_r(t)$	$S_e(t)$

All the estimators except KM were using the true value of $\rho(t)$, also $S_x(t)$.

Mean absolute difference

KM estimator

```
round(mean(abs(s_est - S(t_est))),3)
```

```
## [1] 0.167
```

Slud's estimator

```
round(mean(abs(result2$s1 - S(t_est))),3)
```

```
## [1] 0.119
```

Corrected Slud's estimator

```
round(mean(abs(result2$s2 - S(t_est))),3)
```

```
## [1] 0.119
```

Integral estimated by Riemann

```r
round(mean(abs(s_r - S(t_est))),3)
```

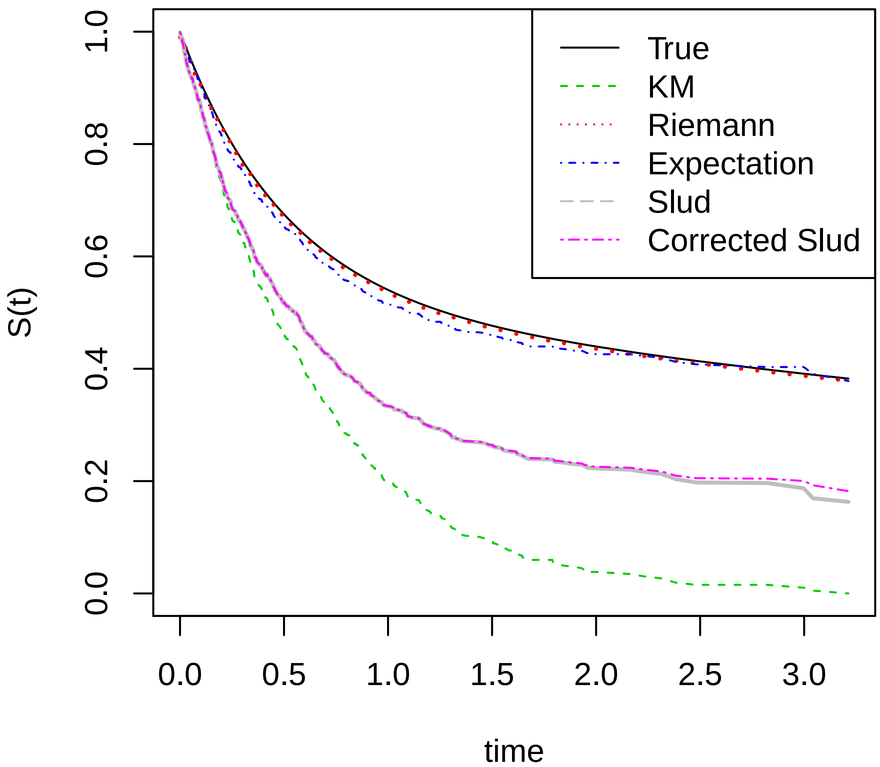
```
## [1] 0.005
```

Integral estimated by the expectation idea

```r
round(mean(abs(s_e - S(t_est))),3)
```

```
## [1] 0.016
```

The plot



Why biased?

For KM estimator and Slud's estimator:

When $\rho_i = 0$, which means that $f(t) = \psi(t) \to \int_0^\infty (t,s)ds = \int_t^\infty f(t,s)ds \to \int_0^t f(t,s)ds = 0$. That is, when $s < t$, $f(t,s) = 0$, there is no censoring.

$S_{p1}(t) = \prod_{i=1}^{d(t)} \frac{n_i - 1}{n_i + \rho_i - 1} + \frac{1}{N} \sum_{k=1}^{d(t)} (\rho_k - 1) \prod_{i=k}^{d(t)} \frac{n_i - 1}{n_i + \rho_i - 1} = 1 - \frac{d(t)}{N}$

$S_{km}(t) = \prod_{i=1}^{d(t)} \frac{n_i - 1}{n_i}$.

And if there is no censor, $n_i = N - i$, $\prod_{i=1}^{d(t)} \frac{n_i - 1}{n_i} = \prod_{i=1}^{d(t)} \frac{N-i-1}{N-i} = \frac{N-1-d(t)}{N-1} = 1 - \frac{d(t)}{N-1}$.

Therefore, $S_{p1}(t) \approx S_{km}(t)$. Since $S_{km}(t)$ supposes independent, it is biased, and $S_{p1}(t)$, $S_{p2}(t)$ are also biased.

How could we proof that the integral estimation method is unbiased?

I feel a little bit confused about how to proof the unbias.

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 intergal term can be estimated as:

$$\begin{aligned} g(t) \equiv \int_0^t \frac{\psi(s)\rho(s)}{Sx(s)}ds &= \int_0^t \frac{\rho(s)}{Sx(s)}d\Psi(s) \\ &= \int_0^t \frac{\rho(s)}{Sx(s)}P(I=1)d\frac{\Psi(s)}{P(I=1)} \\ &= \int_0^t \frac{\rho(s)}{Sx(s)}P(I=1)d\Psi_c(s) \\ &\approx \frac{1}{N}\sum_{0\leq s\leq t}\frac{\rho(s)}{Sx(s)}P(I=1) \equiv \hat{g}(t) \end{aligned}$$

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

Empirically, the $\hat{g}(t)$ is the unbiased estimator of $g(t)$, i.e. $E(\hat{g}(t)) = g(t)$

However, how could we say that $E(exp(-\hat{g}(t))) = exp(-g(t))$? Based on Jensen's inequality, $f(x) = exp(-x)$ is a concave function. Then

$$E(exp(-\hat{g}(t))) < exp(-E(\hat{g}(t))) = exp(-g(t))$$

Then how could we show it?