Simulation attempt

2019-05-22

The two equations

The two main equations in the paper:

$$\rho(t) = \Big[\Big\{f(t)/\phi(t)\Big\} - 1\Big]\Big[\Big\{S(t)/S_X(t)\Big\}\Big]^{-1}$$

$$\hat{S}_p(t) = \frac{1}{N}\Big\{n(t) + \sum_{k=0}^{d(t)-1} c_k \prod_{i=k+1}^{d(t)} \frac{n_i - 1}{n_i + \rho_i - 1}\Big\}$$

Let's look at the $\rho(t)$ first.

$\rho(t)$ shows the proportion of death hazard at time t, conditioning on censored before t or after t

$$\begin{aligned}
\rho(t) =& lim_{\delta\to 0}\frac{P(t<T<t+\delta|T>t, C\leq t)}{P(t<T<t+\delta|T>t, C<t)} \\
=&\frac{lim_{\delta\to 0}P(t<T<t+\delta|T>t, C\leq t)}{lim_{\delta\to 0}P(t<T<t+\delta|T>t, C<t)} \\
=&\frac{lim_{\delta\to 0}\frac{P(t<T<t+\delta, C\leq t)}{P(T>t, C\leq t)}}{lim_{\delta\to 0}\frac{P(t<T<t+\delta, C>t)}{P(T>t, C>t)}} \\
=&\frac{\frac{P(T=t, C\leq t)}{P(T>t, C\leq t)}}{\frac{P(T=t, C>t)}{P(T>t, C>t)}} = \frac{\frac{P(T=t}{P(T>t, C\leq t)} - 1}{\frac{P(T=t)}{P(T>t, C>t)} - 1} \\
=&\frac{1/P(C>t|T=t)}{S(t)/S_x(t) - 1} \\
=&\Big[\Big\{f(t)/\phi(t)\Big\} - 1\Big]\Big[\Big\{S(t)/S_X(t)\Big\}\Big]^{-1}
\end{aligned}$$

where,

- $\phi(t) = \int_t^\infty f(t,s)ds = \int_t^\infty f(s|t)f(t)ds = f(t)\int_t^\infty f(s|t)ds = f(t)P(C>t|T=t)$
- $S_X(t) = P(T>t, C>t)$

For the $\hat{S}_p(t)$, it contains two parts: 1) the people whose deaths are observed after time t; 2) the people who died after time t but are censored before time t

The empirical odds of an uncensored suriviving individual's dying at X_j is:

$$1/(n_j - 1)$$

The probability of an uncensored suriviving individual's dying at X_j is is that correct?:

$$\frac{1}{n_j}$$

Since the $\rho(t)$ means that

$$\rho(t) = \frac{P(\text{death at time t}|\text{ censor before time t})}{P(\text{death at time t}|\text{ censor after time t})}$$

Then the empirical odds of an individual who was censored before time X_j but acutally died at X_j is:

$$\rho(X_j) \times 1/(n_j - 1) = \rho(X_j)/(n_j - 1),$$

then the probabilty of an individual who was censored before time X_j but acutally died at X_j is:

$$\rho(X_j)/(n_j - 1)/(1 + \rho(X_j)/(n_j - 1)) = \frac{\rho(X_j)}{\rho(X_j) + n_j - 1}$$

Then the product-limit estimator of the probability of being cnesored before X_j but actually survived through X_j is

$$1/N^{-1} \sum_{k=0}^{d(t)-1} c_k \prod_{i=k+1}^{d(t)} (1 - \frac{\rho(X_j)}{\rho(X_j) + n_j - 1}) = 1/N^{-1} \sum_{k=0}^{d(t)-1} c_k \prod_{i=k+1}^{d(t)} \frac{n_i - 1}{n_i + \rho_i - 1}$$

where,

- N is the total subjects in the trial
- X_i is the time, $X_i = min(T_i, C_i)$, and X_i is ordered from 1 to N: $X_1 \leq X_2 \leq ... \leq X_N$
- d the total number of death
- $X_{(i)}$ is the death time, and ordered as $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(d)}$
- $n(t)$ is the number of subjects who are still alive at time t
- $d(t)$ is the number of total death people at time t
- n_i is the number of people who survived after the ith death time $(X_j \geq X_{(i)})$
- c_i is the number of censer between the ith death time $X_{(i)}$ and the $(i+1)$th death time $X_{(i+1)}$

The property of $\rho(t)$

$$\rho(t) = \Big[\Big\{f(t)/\phi(t)\Big\} - 1\Big]\Big[\Big\{S(t)/S_X(t)\Big\}\Big]^{-1}$$

Since

- $f(t)/\phi(t) = \frac{f(t)}{f(t)P(C>t|T=t)} = \frac{1}{P(C>t|T=t)} \geq 1$
- $S(t)/S_X(t) = \frac{P(T>t)}{P(T>t,C>t)} \geq 1$

Therefore, $\rho(t) \in [0, \infty]$

- When $\rho(t) = 0$, $f(t)/\phi(t) = 1$, that is $P(C > t|T = t) = 1$, which means that there is no censoring.
- When $\rho(t) = 1$, $1/P(C > t|T = t) = P(T > t)/P(T > t, C > t)$, which is $P(C > t|T = t) = P(C > t|T > t)$. That is, the C and T are independent. When $\rho(t) = 1$ the survival time and the censor time are independent.
- When $\rho(t) > 1$, we have a positive dependence between censor and death. The larger the ρ is, the larger the dependence is.

Simulation setting

Suppose we simulate the data from the following joint distribution:

$$f(s,t) = \frac{1}{1000}(s+t)$$

where $s \in (0, 10)$ and $t \in (0, 10)$.

$f(s,t)$ is a pdf since

$$\begin{aligned}
\int_0^{10}\int_0^{10} f(s,t)dsdt =& \int_0^{10}\int_0^{10} \frac{1}{1000}(s+t)dsdt \\
=&\frac{1}{1000}\int_0^{10}(10s+50)dt \\
=&\frac{1}{1000}5s^2 + 50s|_0^{10} \\
=&1
\end{aligned}$$

The marginal pdf is:

$$\begin{aligned}
f(s) &= \frac{1}{1000}(10s+50) \\
f(t) &= \frac{1}{1000}(10t+50)
\end{aligned}$$

Therefore,

- $f(t) = \frac{1}{1000}(10t + 50)$
- $S(t) = 1 - \frac{1}{1000}(5t^2 + 50t)$
- $S_X(t) = \frac{1}{1000}(1000 - 100t - 10t^2 + t^3)$
- $\phi(t) = \frac{1}{1000}(50 + 10t - \frac{3}{2}t^2)$

The data is simulated from the joint distribution through grid method.