Some results based on the m(t) function

2019-10-14

Contents

Introduction: the new assumption for a semi-parameteric model

We denote $Y_i, i = 1, ..., N$ are the independent, identically, distributed (iid) lifetimes, whose corresponding cumulative distriubtion function (CDF) is F, probability distribution function (PDF) is f; the censoring time is defined as $C_i, i = 1, ..., N$. C_is are also iid, with CDF denoted as G and PDF denoted as g. We set the censors happen on the right and the ovserved time is $Z_i = Y_i \wedge C_i$, whose CDF is H and PDF is h. The $\delta_i = I_{[T_i \leq C_i]}$ is the status indicator, which shows whether subject i is censored ($\delta_i = 0$) or not ($\delta_i = 0$). The corresponding hazard function of lifetime is λ_F and cumulative hazard function is Λ_F.

Instead of the strong assumption of independent between Y_i and C_i, we proposed that $T \perp\!\!\!\perp C$ at a small neighborhood, where $T = C$. That is, we have

$$\lim_{dt \to 0} P(C > t, T \geq t + dt) = P(C > t)P(T \geq t + dt) \tag{1}$$

As well as

$$P(C > t, T \geq t) = P(C > t)P(T \geq t) \tag{2}$$

With this assumption, we can show:

$$
\begin{aligned}
P(C>t|T=t) &= \lim_{dt\to 0} P(C>t|t\le T<t+dt) \\
&= \lim_{dt\to 0} \frac{P(C>t, t\le T<t+dt)}{P(t\le T<t+dt)} \\
&= \lim_{dt\to 0} \frac{P(C>t, T\ge t) - P(C>t, T>t+dt)}{P(T\ge t) - P(T>t+dt)} \\
&= \lim_{dt\to 0} \frac{P(C>t)\Big(P(T\ge t) - P(T>t+dt)\Big)}{P(T\ge t) - P(T>t+dt)} \\
&= P(C>t)
\end{aligned} \tag{3}
$$

And since indpendent,

$$
P(C>t|T>t) = \frac{P(C>t, T>t)}{P(T>t)} = \frac{P(C>t)P(T>t)}{P(T>t)} = P(C>t)
$$

Therefore,

$$
P(C>t|T>t) = P(C>t|T=t) \tag{4}
$$

Given (Eq 1), we could derive that

$$
P(\delta=1|X=t) = \frac{P(C>t, T=t)}{P(X=t)} = \frac{P(T=t)}{P(X=t)} \frac{P(C>t, T>t)}{P(T>t)} = \frac{f(t)S_x(t)}{h(t)S(t)} = \frac{\lambda_F(t)}{\lambda_H(t)}
$$

where $\lambda_H(t)$ is the hazard function corresponding to Z, which is known as crude hazard rate as well.

We may define $m(t) = P(\delta=1|X=t) = E(\delta|X=t)$. Then

$$
m(t) = \frac{\lambda_F(t)}{\lambda_H(t)} \tag{5}
$$

which is the same parameter defined in Dikta's papers. Therefore, the independence between Y and C is not the necessory condition for equation (2).

The "if and only if" relationship between $\rho(t)$ and diagonial independence

Recall the definition of $\rho(t)$ in Slud's paper:

$$
\rho(t) = \lim_{\delta\to 0} \frac{P(t<T<t+\delta|T>t, C\le t)}{P(t<T<t+\delta|T>t, C>t)} \tag{6}
$$

The $\rho(t) = 1$ is equivalent to the independnet condition.

Proof

- If $\lim_{dt\to 0} P(C > t, T \geq t + dt) = P(C > t)P(T \geq t + dt)$

Then

$$\begin{aligned}
&\lim_{\delta\to 0} P(t < T < t + \delta | T > t, C \leq t) = \lim_{\delta\to 0} \frac{P(t < T < t + \delta, C \leq t)}{P(T > t, C \leq t)} \\
&= \lim_{\delta\to 0} \frac{P(t < T < t + \delta) - P(t < T < t + \delta, C > t)}{P(T > t) - P(T > t, C > t)} \\
&= \lim_{\delta\to 0} \frac{P(T > t) - P(T > t + \delta) - P(T > t, C > t) + P(T > t + \delta, C > t)}{P(T > t) - P(T > t, C > t)} \\
&= \lim_{\delta\to 0} \frac{P(T > t) - P(T > t + \delta) - P(T > t)P(C > t) + P(T > t + \delta)P(C > t)}{P(T > t) - P(T > t)P(C > t)} \\
&= \lim_{\delta\to 0} \frac{(P(T > t) - P(T > t + \delta))(1 - P(C > t))}{P(T > t)(1 - P(C > t))} \\
&= \lim_{\delta\to 0} \frac{P(T > t) - P(T > t + \delta)}{P(T > t)}
\end{aligned}$$

On the other hand

$$\begin{aligned}
\lim_{\delta\to 0} P(t < T < t + \delta | T > t, C > t) &= \lim_{\delta\to 0} \frac{P(t < T < t + \delta, C > t)}{P(T > t, C > t)} \\
&= \lim_{\delta\to 0} \frac{P(T > t, C > t) - P(T > t + \delta, C > t)}{P(T > t)P(C > t)} \\
&= \lim_{\delta\to 0} \frac{P(T > t)P(C > t) - P(T > t + \delta)P(C > t)}{P(T > t)P(C > t)} \\
&= \lim_{\delta\to 0} \frac{P(T > t) - P(T > t + \delta)}{P(T > t)}
\end{aligned}$$

Therefore, under the condition $\lim_{dt\to 0} P(C > t, T \geq t + dt) = P(C > t)P(T \geq t + dt)$,

$$\lim_{\delta\to 0} P(t < T < t+\delta | T > t, C \leq t) = \lim_{\delta\to 0} \frac{P(T > t) - P(T > t + \delta)}{P(T > t)} = \lim_{\delta\to 0} P(t < T < t+\delta | T > t, C > t)$$

$$\to \rho(t) = \lim_{\delta\to 0} \frac{P(t < T < t + \delta | T > t, C \leq t)}{P(t < T < t + \delta | T > t, C > t)} = 1$$

- If $\rho(t) = \lim_{\delta\to 0} \frac{P(t<T<t+\delta|T>t,C\leq t)}{P(t<T<t+\delta|T>t,C>t)} = 1$

$$
\begin{aligned}
&\lim_{\delta\to 0}\frac{P(t<T<t+\delta|T>t,C\le t)}{P(t<T<t+\delta|T>t,C>t)}=1\\
\to&\lim_{\delta\to 0}\frac{P(P(t<T<t+\delta,C\le t)}{P(T>t,C\le t)}=\lim_{\delta\to 0}\frac{P(t<T<t+\delta,C>t)}{P(T>t,C>t)}\\
\to&\\
&\lim_{\delta\to 0}\frac{P(t<T<t+\delta)-(P(T>t,C>t)-P(T>t+dt,C>t))}{P(T>t)-P(T>t,C>t)}=\\
&\lim_{\delta\to 0}\frac{P(T>t,C>t)-P(T>t+\delta,C>t)}{P(T>t,C>t)}
\end{aligned}
$$

That is,

$$
\begin{aligned}
&P(T>t,C>t)\Big[P(t<T<t+\delta)-(P(T>t,C>t)+P(T>t+dt,C>t))\Big]\\
&=\Big[P(T>t)-P(T>t,C>t)\Big]\Big[P(T>t,C>t)-P(T>t+\delta,C>t)\Big]\\
&\to\\
&P(T>t,C>t)P(t<T<t+\delta)-P(T>t,C>t)^2+P(T>t,C>t)P(T>t+dt,C>t)\\
&=P(T>t)P(T>t,C>t)-P(T>t,C>t)^2-\\
&P(T>t)P(T>t+\delta,C>t)+P(T>t,C>t)P(T>t+\delta,C>t)\\
&\to\\
&P(T>t,C>t)P(t<T<t+\delta)=P(T>t,C>t)P(T>t)-P(T>t,C>t)P(T>t+\delta)\\
&=P(T>t)P(T>t,C>t)+P(T>t)P(T>t+\delta,C>t)
\end{aligned}
$$

That is

$$
P(T>t,C>t)P(T>t+\delta)=P(T>t)P(T>t+\delta,C>t)
$$
$$
P(C>t|T>t)=P(C>t|T>t+\delta)
$$

Therefore, if $\rho(t)=1$, we should have that $T\perp\!\!\!\perp C$ at a small neighborhood, where $T=C$, which is $P(C>t|T>t)=P(C>t|T>t+\delta)$.

The relationship between $\rho(t)$ and $m(t)$

$\rho(t)=\frac{f(t)/\psi(t)-1}{S(t)/S_x(t)-1}$

$$
\begin{aligned}
\psi(t)=&\int_t^\infty f(t,s)ds\\
=&\int_t^\infty f(s|t)f(t)ds=f(t)P(C>t|T=t)\\
=&f(t)\frac{P(C>t,T=t)}{P(T=t)}\\
=&m(t)\frac{P(X=t)}{P(T=t)}
\end{aligned}
$$

Therefore,

$$\rho(t) = \frac{f(t)/\Big(m(t)\frac{P(X=t)}{P(T=t)}\Big) - 1}{S(t)/S_x(t) - 1}$$

(Question: is there a way to write it into simpler version or odds ratio version?)

Maximum likelihood

Under our new assumption,

$$m_\theta(t) = P(\delta = 1|Z = z) = \lambda_F(t)/\lambda_H(t)$$

And the Z, which is the observated time, has pdf $f_H(z) = \lambda_H(z)S_H(z)$. The likelihoood function can be written as:

$$L_\theta = \prod_{i=1}^{n} m_\theta(z_i)^{\delta_i}(1 - m_\theta(z_i))^{1-\delta_i}\lambda_H(z_i)S_H(z_i)$$

where $f_\theta(\delta_i, z_i) = \Big[m_\theta(z_i)\lambda_H(z_i)S_H(z_i)\Big]^{\delta_i}\Big[(1 - m_\theta(z_i))\lambda_H(z_i)S_H(z_i)\Big]^{1-\delta_i}$ And

$$l_\theta = \log(L_\theta) = \sum_{i=1}^{n}\Big[\delta_i \log(m_\theta(z_i)\lambda_H(z_i)S_H(z_i)) + (1 - \delta_i)\log((1 - m_\theta(z_i))\lambda_H(z_i)S_H(z_i))\Big]$$

We may show that the true θ_0^* is the one that maximize the likelihood function.

Proof:

Suppose θ_0^* is the true vaue of θ. Suppose $f_H^*(z)$ is the true density. We would like to prove that

$$l_{\theta_0^*} = sup l_\theta$$

Which equivalent to

$$\begin{aligned}
&\sum_{i=1}^{n}\Big[\delta_i \log(m_{\theta_0^*}(z_i)f_H(z_i)) + (1 - \delta_i)\log((1 - m_{\theta_0^*}(z_i))f_H(z_i))\Big] \\
&- \sum_{i=1}^{n}\Big[\delta_i \log(m_\theta(z_i)f_H(z_i)) + (1 - \delta_i)\log((1 - m_\theta(z_i))f_H(z_i))\Big] \geq 0 \\
\rightarrow &\frac{1}{n}\sum_{i=1}^{n}\delta_i \log\Big(\frac{m_{\theta_0^*}(z_i)f_H(z_i)}{m_\theta(z_i)f_H(z_i)}\Big) + \frac{1}{n}\sum_{i=1}^{n}(1 - \delta_i)\log\Big(\frac{(1 - m_{\theta_0^*}(z_i))f_H(z_i)}{(1 - m_\theta(z_i))f_H(z_i)}\Big) \geq 0
\end{aligned}$$

Based on Law of Large Number (LLN),

$$\frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big) \rightarrow E(\log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big))$$

Since

$$E(\log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)) = \int_0^\infty \log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)[m_{\theta_0^*}(z_i)f_H^*(z_i)]dz_i$$

According to Kullback–Leibler divergence,

$$D_{KL}(F||G) = \int f\log(\frac{f}{g}) \geq 0$$

Therefore,

$$E(\log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)) \geq 0$$

Similiarly,

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\big) \to (1-\delta_i)E(\log\big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\big)) \geq 0$$

Therefore, $l_{\theta_0^*} \geq l_\theta$ for any other θ that is not the true θ_0^*.

The true θ_0^* maximizes the likelihood function.

Simulation

Example 1

For a joint pdf function $f_{T1,T2}(t_1,t_2)$, if it equals to

$$f_{T1,T2}(x,y) = 16(x-\frac{1}{2})(y-\frac{1}{2})(x-y)(x-y+1)+1$$

Actually, the $f_{T1,T2}(x,y) = C_0(x-\frac{1}{2})(y-\frac{1}{2})(x-y)(x-y+1)+1$, the C_0 can be any positive number to make it work

Then we have survival function $S_{T_1,T_2} = P(T_1 > t_1, T_2 > t_2)$ as:

$$\begin{aligned}
S_{T_1,T_2} =& P(T_1 > t_1, T_2 > t_2) = \int_{t_2}^1\int_{t_1}^1 f_{T_1,T_2}(x,y)dxdy \\
=& \int_{t_2}^1\int_{t_1}^1 f_{T_1,T_2}(x,y)dxdy \\
=& \int_{t_2}^1\int_{t_1}^1 \Big[16(x-\frac{1}{2})(y-\frac{1}{2})(x-y)(x-y+1)+1\Big]dxdy \\
=& \int_{t_2}^1 \Big\{4(y-\frac{1}{2})\Big[x^4-2x^3+(-2y^2+2y+1)x^2+(2y^2-2y)x\Big]+x\Big\}|_{t_1}^1 dy \\
=& \int_{t_2}^1 \Big\{(2-4y)t_1^4+(8y-4)t_1^3+(8y^3-12y^2+2)t_1^2+(-8y^3+12y^2-4y-1)t_1+1\Big\}dy \\
=& (t_1-1)y(2t_1y^3-4t_1y^2+(-2t_1^3+2t_1^2+2t_1)y+2t_1^3-2t_1^2-1)|_{t_2}^1 \\
=& (1-t_1)(1-t_2)(1-2t_1t_2(t_2-t_1)(t_1+t_2-1))
\end{aligned}$$

The marginal function for the survival time and censoring time are all uniform distributions:

$$
\begin{aligned}
f_{t_1}(x) &= \int_0^1 f_{t_1,t_2}(x,y)dy \\
&= \Big\{ y - 4(x - \frac{1}{2})(y^4 - 2y^3 + (-2x^2 + 2x + 1)y^2 + (2x^2 - 2x)y) \Big\}|_0^1 \\
&= 1
\end{aligned}
$$

$$
\begin{aligned}
f_{t_2}(y) &= \int_0^1 f_{t_1,t_2}(x,y)dx \\
&= \Big\{ 4(y - \frac{1}{2})\Big[x^4 - 2x^3 + (-2y^2 + 2y + 1)x^2 + (2y^2 - 2y)x\Big] + x \Big\}|_0^1 \\
&= 1
\end{aligned}
$$

That is,

$$
f_{T_1}(t_1) = I_{[0,1]}(t_1),\ f_{T_2}(t_2) = I_{[0,1]}(t_2)
$$
$$
P(T_1 > t_1) = 1 - t_1,\ P(T_2 > t_2) = 1 - t_2
$$

Therefore, the hazard rate function λ_F for the survival time is:

- $S_F(t) = 1 - t$, $\Lambda_F(t) = -log(1-t)$, $\lambda_F(t) = \frac{1}{1-t}$

The hazard rate function λ_H for the observed time is:

- $S_H(t) = P(Z > t) = (1-t)^2$, $\Lambda_H(t) = -2log(1-t), \lambda_H(t) = \frac{2}{1-t}$

Then

$$
m(t) = \frac{\lambda_F(t)}{\lambda_H(t)} = 0.5
$$

Let's make a simulation to show it works.

Data generation

T_2 is generated from the UNI(0,1).

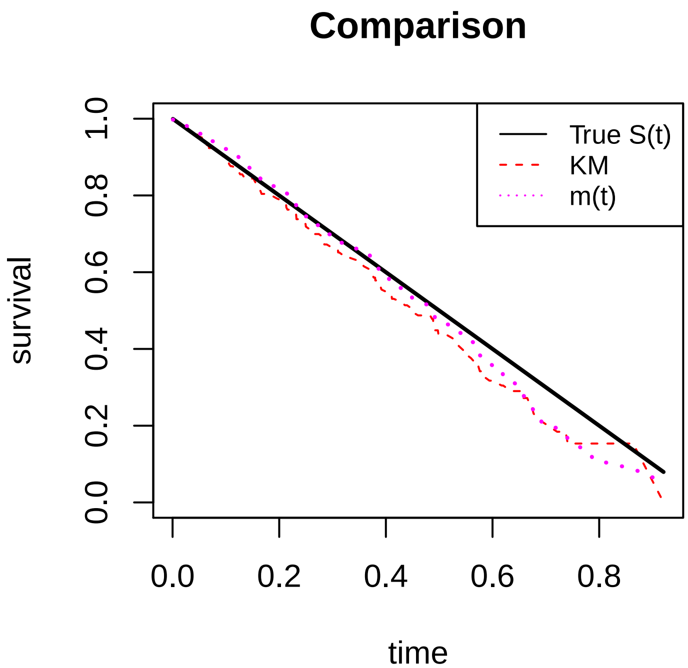
Given T_2, T_1 is generated from $f_{T_1|T_2}(x|y) = \frac{f_{T_1,T_2}(x,y)}{f_{T_2}(y)} = f_{T_1,T_2}(x,y)$, since $f_{T_2}(y) = 1$.

Then $F_{T_1|T_2}(x|y) = x((4y-2)x^3 + (4-8y)x^2 + (-8y^3 + 12y^2 - 2)x + 8y^3 - 12y^2 + 4y + 1)$. Then sample x by inverse probability sampling.

Results:

Censoring percentage: 52.5%

The KM estimator:



Bias:

Kaplan Meier:

```
mean(abs(fit_km$surv - Sx(fit_km$time)))
```

```
## [1] 0.03419431
```

Semi parametric model: $m(t) = \frac{\lambda_F(t)}{\lambda_H(t)}$

```
mean(abs(sest - Sx(fit_km$time)))
```

```
## [1] 0.02045551
```

If we do not know the $m(t)$ function, but know that it is a constant, i.e. $m(t;\theta) = \theta$,we many estimate the parameter by using the MLE:

$$L_n(\theta) = \prod_{i=1}^{n} m(\theta)^{\delta_i}(1 - m(\theta))^{\delta_i}$$

The estimated value is $m(t) = 0.525$. The bias is

```
## [1] 0.0263961
```

Example 2: Zhiliang Ying's paper

In Zhiliang Ying's paper, the Joint CDF is:

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

The corresponding marginal distributions:

- $P(T > x) = P(T > x, U > 0) = e^{-x}e^{-(e^0-1)\left((x-0)^2+1\right)} = e^{-x}$
- $F_T(x) = 1 - e^{-x}, f_T(x) = e^{-x}$
- $P(U > x) = P(U > x, T > 0) = e^{-0}e^{-(e^y-1)} = e^{-(e^y-1)}$
- $F_U(x) = 1 - e^{-(e^y-1)}, f_U(x) = e^{1+y-e^y}$

And the distribution of $X = T \wedge U$ is

$$P(X > x) = P(T > x, U > x) = e^{-x}e^{-(e^x-1)}$$

Therefore,

$$F_X(x) = 1 - e^{1-x-e^x}, f_X(x) = (1 + e^x)e^{1-x-e^x}$$

The $m()$ function is:

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

The censoring percentage Since

$$\begin{aligned} P(T < x < U) =& P(T < x, U > x) = P(U > x) - P(T > x, U > x) \\ =& \exp(-(\exp(x) - 1)) - \exp(-x)\exp(-\exp(x) + 1) \\ =& (1 - \exp(-x))\exp(-(\exp(x) - 1)) \end{aligned}$$

Then we can calculate $P(T < U)$ as:

$$\begin{aligned} P(T < U) =& \int_0^\infty P(T < x < U)dx \\ =& \int_0^\infty (1 - \exp(-x))\exp(-(\exp(x) - 1))dx \\ =& [-e(\Gamma(0, e^x)) - \Gamma(-1, e^x)]|_0^\infty \\ \approx& 0.2 \end{aligned}$$

The censoring percentage is 1 - 0.2 = 0.8.

There was some bug in my simulation for this example. I haven't finished it yet.