relationship between new assumption and rho

2020-01-26

Suppose the random variable for survival time is T, with CDF $P(T < t) = F(t)$ and PDF $f(t)$. Suppose the random variable for censoring time is C, with CDF $P(C < t) = G(t)$ and PDF $g(t)$. The joint distribution for T, C is

$$P(T < u, C < s) = H_{T,C}(u, s), \text{ and its pdf is } h_{T,C}(u, s)$$

Calculate the $\rho(t)$ function when the following assumption is true.

Condition A:

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

Or we may write it as

$$\exists \epsilon > 0, s.t. \text{ for } \forall |dt| < \epsilon, P(T \geq t + dt, C > t) - P(T \geq t + dt)P(C > t) = 0, \text{ for } \forall |dt| < \epsilon$$

We know the $\rho(t)$ function is

Condition B:

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

However, this two conditions are not equivalent.

$$\text{Condition B } \subseteq \text{Condition A}$$

Our new assumption is looser than $\rho = 1$ in terms of independent relationship between death time and censor time.

Direction 1: $\rho = 1 \Rightarrow$ new assumption is true.

Back to our condition A, when $dt = 0$, it is $P(T > t, C > t) = P(T > t)P(C > t)$. To prove $P(T > t, C > t) = P(T > t)P(C > t)$ is equivalent to prove the production of the pdf equals to the joint distribution:

$$h_{T,C}(t, t) = f(t)g(t)$$

Proof

First we show that $\rho(t) = 1 \iff \frac{f(t)}{\psi(t)} = \frac{S(t)}{S_x(t)} = \frac{P(T>t)}{P(T>t,C>t)}$.

If $\rho(t) = 1$, then

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

$\Longrightarrow$

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

$\Longrightarrow$

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

$\Longrightarrow$

$$\lim_{x\to 0}\frac{P(t<T<t+x)}{P(t<T<t+x,C>t)} = \frac{P(T>t)}{P(T>t,C>t)}$$

Then we would like to prove that

$$\lim_{x\to 0}\frac{P(t<T<t+x)}{P(t<T<t+x,C>t)} = \frac{P(T>t)}{P(T>t,C>t)}$$

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

Since as $x \to 0$, both of the nominator and denominator go to 0. Apply L'hopital law, calculate the derivations of nominator and denominator, we get:

$$\lim_{x\to 0}\frac{P(t<T<t+x)}{P(t<T<t+x,C>t)} = \lim_{x\to 0}\frac{f(t+x)}{f(t+x)-P'(T<t+x,C<t)}$$

And

$$\begin{aligned}
P'(T<t+x,C<t) =& \frac{\partial}{\partial x}P(T<t+x,C<t)\\
=& \frac{\partial}{\partial x}\int_0^{t+x}\int_0^t h_{T,C}(u,s)dsdu\\
=& \frac{\partial}{\partial x}\int_0^{x}\int_0^t h_{T,C}(u+t,s)dsdu\\
=& \int_0^t h_{T,C}(x+t,s)ds
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{x\to 0}\frac{f(t+x)}{f(t+x)-P'(T<t+x,C<t)} &= \lim_{x\to 0}\frac{f(t+x)}{f(t+x)-\int_0^t h_{T,C}(x+t,s)ds}\\
&= \frac{f(t)}{f(t)-\int_0^t h_{T,C}(t,s)ds} = \frac{f(t)}{\psi(t)},\ \text{where } \psi(t) = \int_t^\infty h_{T,C}(t,s)ds
\end{aligned}$$

Therefore,

$$\frac{f(t)}{f(t)-\int_0^t h_{T,C}(t,s)ds} = \frac{f(t)}{\psi(t)} = \frac{S(t)}{S_x(t)} = \frac{P(T>t)}{P(T>t,C>t)}$$

Next we prove that $f(t)/\psi(t) = S(t)/S_x(t) \Rightarrow f(t)g(t) = h_{T,C}(t,t)$

If $f(t)/\psi(t) = S(t)/S_x(t)$, we have

$$\begin{aligned}
\frac{f(t)}{f(t)-\int_0^t h_{T,C}(t,s)ds} =& \frac{P(T>t)}{P(T>t,C>t)} \text{ (both denominators are non zero)}\\
=& \frac{1-P(T<t)}{1-P(T<t)-P(C<t)+P(T<t,C<t)}
\end{aligned}$$

$\Longrightarrow$

$$f(t)\big[1-P(T<t)-P(C<t)+P(T<t,C<t)\big] = \big[f(t)-\int_0^t h_{T,C}(t,s)ds\big]\big[1-P(T<t)\big]$$

$\Longrightarrow$

$$
\begin{aligned}
&f(t)-f(t)P(T<t)-f(t)P(C<t)+f(t)P(T<t,C<t)\\
=&f(t)-f(t)P(T<t)-\int_0^t h_{T,C}(t,s)ds+P(T<t)\int_0^t h_{T,C}(t,s)ds
\end{aligned}
$$

$\Longrightarrow$

$$
\begin{aligned}
f(t)P(C<t)-f(t)P(T<t,C<t)&=\int_0^t h_{T,C}(t,s)ds-P(T<t)\int_0^t h_{T,C}(t,s)ds\\
f(t)\int_0^t g(s)ds-f(t)\int_0^t\int_0^t h_{T,C}(u,s)duds&=\int_0^t h_{T,C}(t,s)ds-\int_0^t f(u)du\int_0^t h_{T,C}(t,s)ds\\
\int_0^t f(t)g(s)ds-\int_0^t\int_0^t f(t)h_{T,C}(u,s)duds&=\int_0^t h_{T,C}(t,s)ds-\int_0^t\int_0^t f(u)h_{T,C}(t,s)duds
\end{aligned}
$$

The we just need to show that

$$
\int_0^t f(t)g(s)ds-\int_0^t\int_0^t f(t)h_{T,C}(u,s)duds=\int_0^t h_{T,C}(t,s)ds-\int_0^t\int_0^t f(u)h_{T,C}(t,s)duds
$$

$$\Longrightarrow$$

$$f(t)g(t)=h_{T,C}(t,t)$$

Notice that both of the left and right sides are continuous functions of t, let

$$L(t)=\int_0^t f(t)g(s)ds-\int_0^t\int_0^t f(t)h_{T,C}(u,s)duds$$

$$R(t)=\int_0^t h_{T,C}(t,s)ds-\int_0^t\int_0^t f(u)h_{T,C}(t,s)duds$$

If function $L(t)=R(t),\forall t\in\mathbb{R}$, then

$$\frac{\partial}{\partial t}L(t)=\frac{\partial}{\partial t}R(t)$$

Also we notice that

$$\frac{\partial}{\partial t}\int_0^t\int_0^t f(t)h_{T,C}(u,s)duds=\int_0^t f(t)h_{T,C}(u,t)du,$$

$$\frac{\partial}{\partial t}\int_0^t f(t)h_{T,C}(u,t)du=f(t)h_{T,C}(t,t)$$

And

$$\frac{\partial}{\partial t}\int_0^t\int_0^t f(u)h_{T,C}(t,s)duds=\int_0^t f(u)h_{T,C}(t,t)du,$$

$$\frac{\partial}{\partial t}\int_0^t f(u)h_{T,C}(t,t)du=f(t)h_{T,C}(t,t)$$

Therefore, to make $L(t)=R(t)$, we need,

$$\int_0^t f(t)g(s)ds=\int_0^t h_{T,C}(t,s)ds$$

$\Longrightarrow$

$$\frac{\partial}{\partial t}\int_0^t f(t)g(s)ds=\frac{\partial}{\partial t}\int_0^t h_{T,C}(t,s)ds$$

$$f(t)g(t)=h_{T,C}(t,t)$$

That is, when $\Rightarrow P(T>t,C>t)=P(T>t)P(C>t)$.

Since $P(T > u), P(C > s), P(T > u, C > s)$ are continuous functions,

$$\lim_{dt \to 0} \{P(T > u + dt) - P(T > u)\} = 0$$

$$\lim_{dt \to 0} \{P(C > s + dt) - P(C > s)\} = 0$$

$$\lim_{dt \to 0} \{P(T > u + dt, C > s) - P(T > u, C > s)\} = 0$$

Therefore,

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

which is our new condition A.

Direction 2

When new assumption A is true $\not\to \rho(t) = 1$.

Counter example: suppose we have a joint distribution of T and C,

$$S_{T,C}(x, y) = (1 - x)(1 - y)(1 + \frac{C}{8}xy(x - y)(x + y - 1))$$

$$S_T(x) = 1 - x, S_C(y) = 1 - y$$

where $(x, y) \in [0, 1] \times [0, 1], C \in [-4, 4]$. It satifies the condition A, since:

$$\begin{aligned}
P(T > x + y, C > x) =& (1 - x - y)(1 - x)(1 + \frac{C}{8}xy(x + y)(2x + y - 1)) \\
=& [(1 - x)^2 - (1 - x)y][1 + \frac{C}{8}\{(2x^3 - x^2)y + (3x^2 - x)y^2 + xy^3\}] \\
=& (1 - x)^2 - (1 - x)y \\
&+ \frac{C}{8}\{(1 - x)^2(2x^3 - x^2)y + (1 - x)^2(3x^2 - x)y^2 + x(1 - x)^2y^3\} \\
&- \frac{C}{8}\{(1 - x)(2x^3 - x^2)y^2 + (1 - x)(3x^2 - x)y^3 + x(1 - x)y^4\} \\
=& (1 - x)^2 + \frac{C}{8}[(1 - x)^2(2x^3 - x^2) - (1 - x)]y \\
&+ \frac{C}{8}[(1 - x)^2(3x^2 - x) - (1 - x)(2x^3 - x^2)]y^2 \\
&+ \frac{C}{8}[x(1 - x)^2 - (1 - x)(3x^2 - x)]y^3 - \frac{C}{8}x(1 - x)y^4 \\
=& (1 - x)^2 + A_1y + A_2y^2 + A_3y^3 + A_4y^4
\end{aligned}$$

where

- $A_1 = \frac{C}{8}[(1 - x)^2(2x^3 - x^2) - (1 - x)]$
- $A_2 = \frac{C}{8}[(1 - x)^2(3x^2 - x) - (1 - x)(2x^3 - x^2)]$
- $A_3 = \frac{C}{8}[x(1 - x)^2 - (1 - x)(3x^2 - x)]$
- $A_4 = -\frac{C}{8}[x(1 - x)]$

And when $y \to 0$,

$$\lim_{y \to 0} P(T > x + y, C > x) = \lim_{y \to 0} \{(1 - x)^2 + A_1y + A_2y^2 + A_3y^3 + A_4y^4\} = (1 - x)^2 = P(T > t)P(C > t)$$

For $\rho(t)$ calculation,

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

For $\frac{P(T>t, C>t)}{P(T>t, C \leq t)}$, under our assumption,

$$
\frac{P(T > t, C > t)}{P(T > t, C \leq t)} = \frac{P(T > t)P(C > t)}{P(T > t) - P(T > t, C > t)} = \frac{P(T > t)P(C > t)}{P(T > t) - P(T > t)P(C > t)} = \frac{P(C > t)}{1 - P(C > t)}
$$

when $P(T > t) \neq 0$

And we know that

$$
P(t < T < t + dt) = dt
$$

$$
\begin{aligned}
P(t < T < t + dt, C > t) =& P(T > t, C > t) - P(T > t + dt, C > t) \\
=& (1 - t)^2 - (1 - t)^2 - A_1 dt - A_2 (dt)^2 - A_3 (dt)^3 - A_4 (dt)^4 \\
=& - A_1 dt - A_2 (dt)^2 - A_3 (dt)^3 - A_4 (dt)^4
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{P(t < T < t + dt, C \leq t)}{P(t < T < t + dt, C > t)} =& \frac{P(t < T < t + dt) - P(t < T < t + dt, C > t)}{P(t < T < t + dt, C > t)} \\
=& \frac{dt + A_1 dt + A_2 (dt)^2 + A_3 (dt)^3 + A_4 (dt)^4}{-A_1 dt - A_2 (dt)^2 - A_3 (dt)^3 - A_4 (dt)^4}
\end{aligned}
$$

$$
\begin{aligned}
\lim_{dt \to 0} \frac{P(t < T < t + dt, C \leq t)}{P(t < T < t + dt, C > t)} &= \lim_{dt \to 0} \frac{P(t < T < t + dt) - P(t < T < t + dt, C > t)}{P(t < T < t + dt, C > t)} \\
&= \lim_{dt \to 0} \frac{dt + A_1 dt + A_2 (dt)^2 + A_3 (dt)^3 + A_4 (dt)^4}{-A_1 dt - A_2 (dt)^2 - A_3 (dt)^3 - A_4 (dt)^4} \\
&= \lim_{dt \to 0} \frac{1 + A_1 + 2A_2 (dt) + 3A_3 (dt)^2 + 4A_4 (dt)^3}{-A_1 - 2A_2 (dt) - 3A_3 (dt)^2 - 4A_4 (dt)^3} \\
&= \frac{1 + A_1}{-A_1} = \frac{1 + \frac{C}{8}[(1 - x)^2 (2x^3 - x^2) - (1 - x)]}{-\frac{C}{8}[(1 - x)^2 (2x^3 - x^2) - (1 - x)]}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\rho(t) =& \lim_{dt \to 0} \frac{P(t < T < t + dt, C \leq t)}{P(t < T < t + dt, C > t)} \times \frac{P(C > t)}{1 - P(C > t)} \\
=& \frac{1 + \frac{C}{8}[(1 - x)^2 (2x^3 - x^2) - (1 - x)]}{-\frac{C}{8}[(1 - x)^2 (2x^3 - x^2) - (1 - x)]} [\frac{1 - x}{x}] \\
=& \frac{8 + C(2x - 1)(1 - x)^2 x}{8 + Cx^2 (x - 1)(2x - 1)} \\
\neq& 1
\end{aligned}
$$

Therefore

$$
\text{Condition B } \subseteq \text{Condition A}
$$

Our new assumption is looser than $\rho = 1$ in terms of independent relationship between death time and censor time.