Influence function calculation for Brier score for event time data

Johan Sebastian Ohlendorff & Thomas Alexander Gerds

2022-06-07

IF Calculation

To describe the situation with competing risks (and also survival) we introduce a random variable $D \in \{1,2\}$ which indicates the cause (i.e., type of the event) observed at time T such that $D=1$ means that the event of interest occurred, and $D=2$ that a competing risk occurred. As in the survival setting we let Q denote the joint probability measure of the uncensored data, $(T,D,X) \sim Q$, and P the joint probability measure of the right censored data $Z=(\tilde{T},\Delta,X) \sim P$ now with $\Delta = D1_{\{T \leqslant C\}}$ taking values in the set $\{0,1,2\}$. We are interested in the following definition of the time-dependent discrimination measure for cause 1. We can easily calculate the influence function for the Brier score, which can be written as:

$$
\begin{aligned}
&\int \left\{1_{\{t \leqslant \tau\}} - R(\tau \mid x)\right\}^2 Q(dt,dx) \\
&= \int 1_{\{t \leqslant \tau\}} - 2 1_{\{t \leqslant \tau\}} R(\tau \mid x) + R(\tau \mid x)^2 Q(dt,dx) \\
&= \int 1_{\{t \leqslant \tau\}} (1-2R(\tau \mid x)) Q(dt,dx) + \int R(\tau \mid x)^2 Q(dx) \\
&= \int \frac{1_{\{t \leqslant \tau\}}}{G(t-|x)} (1-2R(\tau \mid x)) P(dt,1,dx) + \int R\left(\tau|x\right)^2 P(dx)
\end{aligned}
$$

We find

$$
\begin{aligned}
IC_{\text{Brier}}(\tilde{T}_i,\Delta_i,X_i;\tau) &= \partial_\varepsilon \int \frac{1_{\{t \leqslant \tau\}}}{G(t-|x)} (1-2R(\tau \mid x)) P_\varepsilon(dt,1,dx) + \partial_\varepsilon \int R\left(\tau|x\right)^2 P_\varepsilon(dx) \\
&= \int 1_{\{t \leqslant \tau\}} (1-2R(\tau \mid x)) \frac{d(\delta_{\{\tilde{T}_i,\Delta_i,X_i\}})(t,1,x) + dP(t,1,x)\left[f_i(t-,x)-1\right]}{G(t-|x)} \\
&+ \int R\left(\tau|x\right)^2 (\delta_{\{\tilde{T}_i,\Delta_i,X_i\}} - P) \\
&= 1_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 1\}} (1-2R(\tau \mid X_i)) \frac{1}{G(\tilde{T}_i - |X_i)} \\
&+ \int 1_{\{t \leqslant \tau\}} (1-2R(\tau \mid x)) f_i(t-,x) \frac{dP(t,1,x)}{G(t-|x)} + R\left(\tau|X_i\right)^2 \\
&- \int R\left(\tau|x\right)^2 dP(x) - \int 1_{\{t \leqslant \tau\}} (1-2R(\tau \mid x)) \frac{dP(t,1,x)}{G(t-|x)}
\end{aligned}
$$

The last term that is subtracted is the Brier score. Then using that

$$
f_i(t,x) = \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}} \delta_{X_i}(x)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge t} \frac{\delta_{X_i}(x) dP(s,0|x)}{G(s|X_i)^2 S(s|X_i)^2}
$$

we see that by splitting the double integral into two parts that and that $\int f(x)\delta_{X_i}(x)dx = f(X_i)$

$$
\int 1_{\{t \leqslant \tau\}} (1-2R(\tau \mid x)) f_i(t-,x) \frac{dP(t,1,x)}{G(t-|x)} = (1-2R(\tau \mid X_i)) \int 1_{\{t \leqslant \tau\}} \tilde{f}_i(t-) \frac{dP(t,1|X_i)}{G(t-|X_i)}
$$

where $\tilde{f}_i(t) = \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge t} \frac{dP(s,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2}$. Rewriting $\tilde{f}_i(t)$ a bit, we get

$$\tilde{f}_i(t) = \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} + \int_0^{\tilde{T}_i \wedge t} \frac{S(s|X_i)G(ds|X_i)}{G(s|X_i)^2 S(s|X_i)^2} = \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} + \int_0^{\tilde{T}_i \wedge t} \frac{1}{G(s|X_i)S(s|X_i)} \frac{G(ds|X_i)}{G(s|X_i)}$$

As for the estimation, we have

$$\begin{aligned} IC_{\text{Brier}}(\tilde{T}_i, \Delta_i, X_i; \tau) = {} & 1_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 1\}}(1 - 2R(\tau \mid X_i)) \frac{1}{\hat{G}(\tilde{T}_i - |X_i)} \\ & + \int 1_{\{t \leqslant \tau\}}(1 - 2R(\tau \mid x)) \hat{f}_i(t-, x) \frac{d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} + R\left(\tau|X_i\right)^2 \\ & - \int R\left(\tau|x\right)^2 d\mathbb{P}_n(x) - \int 1_{\{t \leqslant \tau\}}(1 - 2R(\tau \mid x)) \frac{d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} \end{aligned}$$

Then using that