Influence function calculation for Brier score for event time data

Johan Sebastian Ohlendorff & Thomas Alexander Gerds

2022-01-09

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 $O = (\tilde{T}, \Delta, X) \sim P$ 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:

$$
\begin{aligned}
&\int \left\{1_{\{t \leqslant \tau\}} - R(\tau \mid x)\right\}^2 Q(dt, dx) \\
&= \int 1_{\{t \leqslant \tau\}} - 21_{\{t \leqslant \tau\}} R(\tau \mid x) + R(\tau \mid x)^2 Q(dt, dx) \\
&= \int 1_{\{t \leqslant \tau\}} (1 - 2R(\tau \mid x)) Q(dt, dx) + \int R(\tau \mid x)^2 Q(dx) \\
&= \int \frac{1_{\{t \leqslant \tau\}}}{G(t - |x)} (1 - 2R(\tau \mid x)) P(dt, 1, dx) + \int R(\tau|x)^2 P(dx)
\end{aligned}
$$

We find

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

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 (plugging in $f(t,x)$ instead of $f(t-,x)$!)

$$\begin{aligned}\int_X \int_0^\tau (1-2R(\tau|x)) f_i(t-,x) \frac{P(dt,1,x)}{G(t-|x)} &= \int_X \int_0^\tau (1-2R(\tau|x)) \left[\frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}} \delta_{X_i}(x)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge t} \frac{\delta_{X_i}(x) P(ds,0|x)}{G(s|X_i)^2 S(s|X_i)^2} \right] \frac{P(dt,1,x)}{G(t-|x)} \\ &= \int_0^\tau (1-2R(\tau|X_i)) \left[\frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge t} \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2} \right] \frac{P(dt,1|X_i)}{G(t-|X_i)} \\ &= (i) - (ii)\end{aligned}$$

where

$$\begin{aligned}(i) &= \int_0^\tau (1-2R(\tau|X_i)) \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant t, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} \frac{P(dt,1|X_i)}{G(t-|X_i)} \\ &= (1-2R(\tau|X_i)) \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} \int_{\tilde{T}_i}^\tau \frac{P(dt,1|X_i)}{G(t-|X_i)} \\ &= (1-2R(\tau|X_i)) \frac{\mathbb{1}_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 0\}}}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} (F_1(\tau|X_i) - F_1(\tilde{T}_i|X_i))\end{aligned}$$

Similarly, we have

$$(ii) = (1-2R(\tau|X_i)) \int_0^\tau \int_0^{\tilde{T}_i \wedge t} \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2} \frac{P(dt,1|X_i)}{G(t-|X_i)}$$

If $\tilde{T}_i > \tau$, then this can be written as

$$\begin{aligned}(ii) &= (1-2R(\tau|X_i)) \int_0^\tau \int_s^\tau \frac{P(dt,1|X_i)}{G(t-|X_i)} \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2} \\ &= (1-2R(\tau|X_i)) \int_0^\tau (F_1(\tau|X_i) - F_1(s|X_i)) \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2}\end{aligned}$$

On the other hand, if $\tilde{T}_i \leqslant \tau$, then

$$\begin{aligned}(ii) &= (1-2R(\tau|X_i)) \int_0^{\tilde{T}_i} \int_s^\tau \frac{P(dt,1|X_i)}{G(t-|X_i)} \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2} \\ &= (1-2R(\tau|X_i)) \int_0^{\tilde{T}_i} (F_1(\tau|X_i) - F_1(s|X_i)) \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2}\end{aligned}$$

Thus

$$(ii) = (1-2R(\tau|X_i)) \int_0^{\tilde{T}_i \wedge \tau} (F_1(\tau|X_i) - F_1(s|X_i)) \frac{P(ds,0|X_i)}{G(s|X_i)^2 S(s|X_i)^2}$$

Hence,

$$(i)-(ii) = (1-2R(\tau|X_i)) \left(\frac{I(\tilde{T}_i \leqslant \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} (F_1(\tau|X_i) - F_1(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_1(\tau|X_i) - F_1(s|X_i))}{G(s|X_i)^2 S(s|X_i)^2} P(ds,0|X_i) \right)$$

Not expanding the paranthesis

We can also derive the Brier score as:

$$\sum_{\delta=0,1} \int \left\{ 1_{\{t \leqslant \tau\}} - R(\tau \mid x) \right\}^2 W_\tau(z;G) P(dt,\delta,dx)$$

with $W_\tau(z;G) = \frac{\delta I(t\leqslant\tau)}{G(t-|x)} + \frac{I(t>\tau)}{G(\tau|x)}$. Taking the Gateaux derivative yields,

$$\begin{aligned}
&I(\tilde{T}_i \leqslant \tau, \Delta_i = 1)\frac{(1-R(\tau \mid X_i))^2}{G(\tilde{T}_i - |X_i)} + I(\tilde{T}_i > \tau)\frac{R(\tau \mid X_i)^2}{G(\tau|X_i)} \\
&- Brier \\
&+ \int I(t \leqslant \tau)(1-R(\tau|x))^2 \frac{f_i(t-,x)}{G(t-|x)} P(dt,1,dx) \\
&+ \int I(t > \tau) R(\tau|x)^2 \frac{f_i(\tau,x)}{G(\tau|x)} P(dt,dx)
\end{aligned}$$

Whenever the censoring does not depend on covariates, the two last integrals are easy to estimate. However, when the censoring does depend on the covariates, we can use the tricks mentioned in IFAUC.pdf to get that,

$$\begin{aligned}
&\int I(t > \tau) R(\tau|x)^2 \frac{f_i(\tau,x)}{G(\tau|x)} P(dt,dx) + \int I(t \leqslant \tau)(1-R(\tau|x))^2 \frac{f_i(t-,x)}{G(t-|x)} P(dt,1,dx) \\
&= R(\tau|X_i)^2 S(\tau|X_i)\left(\frac{I(\tilde{T}_i \leqslant \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge \tau} \frac{1}{G(s|X_i)^2 S(s|X_i)^2} P(ds,0|X_i)\right) \\
&+ (1-R(\tau|X_i))^2\left(\frac{I(\tilde{T}_i \leqslant \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)}(F_1(\tau|X_i) - F_1(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_1(\tau|X_i) - F_1(s|X_i))}{G(s|X_i)^2 S(s|X_i)^2} P(ds,0|X_i)\right) \\
&= R(\tau|X_i)^2 S(\tau|X_i)\left(\frac{I(\tilde{T}_i \leqslant \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge \tau} \frac{1}{G(s|X_i) S(s|X_i)} \Lambda_C(ds|X_i)\right) \\
&+ (1-R(\tau|X_i))^2\left(\frac{I(\tilde{T}_i \leqslant \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)}(F_1(\tau|X_i) - F_1(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_1(\tau|X_i) - F_1(s|X_i))}{G(s|X_i) S(s|X_i)} \Lambda_C(ds|X_i)\right)
\end{aligned}$$

IF with cross-validation

Binary case

Our cross-validation algorithm repeatedly splits the dataset D_n of size n into training and validation datasets as follows. Let B be a large integer. For each $b = 1, \dots, B$ we draw a bootstrap dataset $D^*_{m,b} = \{O^*_{b,1}, \dots, O^*_{b,m}\}$ of size $m \leqslant n$ with or without replacement from the data D_n. For $i = 1, \dots, n$ let N^b_i be the number of times subject i is included in $D^*_{m,b}$. For subsampling bootstrap (without replacement), N^b_i is either 0 or 1. In step b of the cross-validation algorithm the bootstrap dataset $D^*_{m,b}$ is used for training. We apply R to $D^*_{m,b}$ to obtain the prediction model $R(D^*_{m,b})$. All subjects i for which $N^b_i = 0$ are out-of-bag and we let these subjects form the validation dataset of step b.

We now calculate the influence function in the case with binary outcome data ($Y \in \{0,1\}$. In this case, we consider the following functional which describes the expected Brier score of the model R on average across all possible training datasets D_m of size m. The expectation is taken with respect to the data of subject $m+1$:

$$\psi_m(P) = \int \left(\int \cdots \int \{y_{m+1} - R(D_m)(x_{m+1})\}^2 \prod_{i=1}^m P(do_i) \right) P(do_{m+1})$$

for some sample size $m < n$. Let

$$\omega_m(Y_{m+1}, X_{m+1}) = \mathbb{E}_{D_m}[(Y_{m+1} - R(D_m)(X_{m+1}))^2 | Y_{m+1}, X_{m+1}]$$

because then $\psi_m(P) = \mathbb{E}_{O_{m+1}}[\omega_m(Y_{m+1}, X_{m+1})]$. For estimating $\omega_{\tau,m}$, we propose to use leave one-out bootstrap estimation:

$$\hat{\omega}_m(Y_i, X_i) = \frac{\sum_{b=1}^B (Y_i - R(D_m)(X_i))^2 I(N^b_i = 0)}{\sum_{b=1}^B I(N^b_i = 0)}$$

Finally, the Brier score may then be estimated by leave one-out bootstrap estimation as ´

$$\hat{\psi}_m^{(1)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\omega}_m(Y_i, X_i)$$

Finally, we also want standard errors of the estimates. For this, we consider the influence function by taking the Gateaux derivative and get:

$$
\begin{aligned}
\mathrm{IF}_\psi(m; o_k) &= \frac{\partial}{\partial \epsilon}\psi_m(P_{\epsilon,k})\Big|_{\epsilon=0} && (1)\\
&= \int \frac{\partial}{\partial \epsilon}\left(\int\cdots\int \{y_{m+1} - R(D_m)(x_{m+1})\}^2 \prod_{i=1}^{m} P_{\epsilon,k}(do_i)\right) P_{\epsilon,k}(do_{m+1})\Big|_{\epsilon=0} && (2)\\
&= \int \frac{\partial}{\partial \epsilon}\left(\int\cdots\int \{y_{m+1} - R(D_m)(x_{m+1})\}^2 \prod_{i=1}^{m} P_{\epsilon,k}(do_i)\right)\Big|_{\epsilon=0} P(do_{m+1}) && (3)\\
&+ \int \left(\int\cdots\int \{y_{m+1} - R(D_m)(x_{m+1})\}^2 \prod_{i=1}^{m} P(do_i)\right) \frac{\partial}{\partial \epsilon} P_{\epsilon,k}(do_{m+1})\Big|_{\epsilon=0} && (4)\\
&= \int \sum_{j=1}^{m}\left(\int\cdots\int \{y_{m+1} - R(D_m)(x_{m+1})\}^2 \delta_{o_k}(o_j)\prod_{i\neq j}(do_i)\right) P(do_{m+1}) - m\psi_m(P) && (5)\\
&+ \omega_m(Y_k, X_k) - \psi_m(P) && (6)\\
&= \omega_m(Y_k, X_k) - (m+1)\psi_m(P) && (7)\\
&+ \int \sum_{j=1}^{m}\left(\int\cdots\int \{y_{m+1} - R(D_m)(x_{m+1})\}^2 \delta_{o_k}(o_j)\prod_{i\neq j}(do_i)\right) P(do_{m+1}) && (8)
\end{aligned}
$$

wherein we used the product rule of differentiation. For estimating the influence function, we suggest the estimator:

$$\widehat{\mathrm{IF}}_\psi(m; O_k) = \hat{\omega}_m^{(1)}(Y_k, X_k) - (m+1)\,\hat{\psi}_m^{(1)} + \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_b \{Y_i - R(D^*_{m,b})(X_i)\}^2 \mathcal{I}_{\{N_i^b=0\}} N_k^b}{\sum_b \mathcal{I}_{\{N_i^b=0\}}}$$

Survival and competing risk case

Let us now try to expand this to the case with (right-censored) survival data, i.e. $O = (T, \Delta, X)$ and let $\bar{O} = (\tilde{T}, X)$ denote the true event time, i.e. $T = \min\{\tilde{T}, C\}$ and $\Delta = I(T \leqslant C)$, where C is the censoring time. Also let $Y_i = I(\tilde{T} \leqslant \tau)$ (or T, depending on whichever is the most appropriate), where τ be some prespecified time point and $\hat{G}_n$ be an estimate of the censoring distribution G based on D_n. Then we are concerned with the functional $\mu_{\tau,m}$

$$\mu_{\tau,m} = \mathbb{E}_{O_{m+1}}\left[\mathbb{E}_{D_m}\left[\{\mathcal{I}_{\{\tilde{T}_{m+1}\leqslant\tau\}} - R_\tau(D_m)(X_{m+1})\}^2 \middle| \tilde{T}_{m+1}, X_{m+1}\right]\right]$$

By rewriting the above above a bit (i.e. by using standard tricks when rewriting in terms of the observed data), it can be shown that this quantity can be defined in terms of the observed data, i.e. it can be expressed as the value of a statistical functional $\psi_{\tau,m} : \mathcal{P} \to [0,1]$

$$
\begin{aligned}
\psi_{\tau,m}(\mathrm{P}) = \int \Bigg(\int\cdots\int &\left\{\mathcal{I}_{\{u_{m+1}\leqslant\tau\}} - R_\tau(D_m)(x_{m+1})\right\}^2 \prod_{i=1}^{m}\mathrm{P}(\mathrm{d}o_i)\Bigg)\\
&\times W_\tau(o_{m+1}, \kappa_{\tau,x_{m+1}}(\mathrm{P}))\mathrm{P}(\mathrm{d}o_{m+1})\\
= \mu_{\tau,m}.
\end{aligned}
$$

with

$$W_\tau(X_i, \hat{G}_n) = \frac{\mathcal{I}_{\{T_i\leqslant\tau\}}\Delta_i}{\hat{G}_n(T_i|X_i)} + \frac{\mathcal{I}_{\{T_i>\tau\}}}{\hat{G}_n(\tau|X_i)}$$

then this can be estimated in much the same way as before, i.e. as

$$\hat{\mu}^{(1)}_{\tau,m} = \frac{1}{n}\sum_{i=1}^{n} \hat{\omega}^{(1)}_{\tau,m}(T_i, X_i) W_\tau(X_i, \hat{G}_n)$$

Here, we redefine that

$$\hat{\omega}^{(1)}_{\tau,m}(\tilde{T}_i, X_i) = \frac{\sum_{b=1}^{B}\{\mathcal{I}_{\{\tilde{T}_i \leqslant \tau\}} - R_\tau(D^*_{m,b})(X_i)\}^2 \mathcal{I}_{\{N_i^b=0\}}}{\sum_{b=1}^{B} \mathcal{I}_{\{N_i^b=0\}}}$$

In much the same way as before, we may find the the influence function of $\psi_{\tau,m}(\mathrm{P})$ to be almost the same as before with an additional term corresponding to the fact that the censoring distribution has to be estimated, i.e.

$$\begin{aligned}
\mathrm{IF}_\psi(\tau, m; o_k) &= \omega_{\tau,m}(t_k, x_k) W_\tau(o_k, \kappa_{\tau, x_{m+1}}(\mathrm{P})) - (m+1)\,\mu_{\tau,m} \\
&\quad + \int \Big[\sum_{j=1}^{m} \int \cdots \int \{\mathcal{I}_{\{t_{m+1} \leqslant \tau\}} - R_\tau(\{o_i\}_{i=1}^n)(x_{m+1})\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} \mathrm{P}(\mathrm{d}o_i) \Big] W_\tau(o_{m+1}, \kappa_{\tau, x_{m+1}}(\mathrm{P})) \mathrm{P}(\mathrm{d}o_{m+1}) \\
&\quad + \int \omega_{\tau,m}(t_{m+1}, x_{m+1}) \left[\frac{\mathcal{I}_{\{t_{m+1} \leqslant \tau\}} \delta_{m+1}}{G(t_{m+1} - | x_{m+1})} f_k(t_{m+1}, x_{m+1}) + \frac{\mathcal{I}_{\{t_{m+1} > \tau\}}}{G(\tau | x_{m+1})} f_k(\tau, x_{m+1}) \right] \mathrm{P}(\mathrm{d}o_{m+1})
\end{aligned}$$

This can then be estimated in much the same way as before

$$\begin{aligned}
\widehat{\mathrm{IF}}_\psi(\tau, m; O_k) &= \hat{\omega}^{(1)}_{\tau,m}(T_k, X_k) W_\tau(O_k, \hat{G}_n) - (m+1)\,\hat{\mu}^{(1)}_{\tau,m} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_b \{\mathcal{I}_{\{T_i \leqslant \tau\}} - R_\tau(D^*_{m,b})(X_i)\}^2 \mathcal{I}_{\{N_i^b=0\}} N_k^b}{\sum_b \mathcal{I}_{\{N_i^b=0\}}} W_\tau(O_i, \hat{G}_n) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \hat{\omega}^{(1)}_{\tau,m}(T_i, X_i) \left[\frac{\mathcal{I}_{\{T_i \leqslant \tau\}} \Delta_i}{\hat{G}_n(T_i - | X_i)} \hat{f}_k(T_i -, X_i) + \frac{\mathcal{I}_{\{T_i > \tau\}}}{\hat{G}_n(\tau | X_i)} \hat{f}_k(\tau, X_i) \right]
\end{aligned}$$

Here the last term corresponds to the censoring being unknown.