

# 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  $\Delta = D1_{\{T \leq C\}}$  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:

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

with  $W_\tau(z; G) = \frac{I(t \leq \tau, \delta=1)}{G(t-|x)} + \frac{I(t \leq \tau, \delta=2)}{G(t-|x)} + \frac{I(t > \tau)}{G(\tau|x)}$ . Taking the Gateaux derivative yields,

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

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 (see also

there to find the definition of  $f_i(t, x)$  to get that,

$$\begin{aligned}
& \int I(t > \tau) R(\tau|x)^2 \frac{f_i(\tau, x)}{G(\tau|x)} P(dt, dx) + \int I(t \leq \tau) (1 - R(\tau|x))^2 \frac{f_i(t^-, x)}{G(t - |x)} P(dt, 1, dx) + \int I(t \leq \tau) R(\tau|x)^2 \frac{f_i(t^-, x)}{G(t - |x)} P(dt, 2, dx) \\
&= R(\tau|X_i)^2 S(\tau|X_i) \left( \frac{I(\tilde{T}_i \leq \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge \tau} \frac{1}{G(s|X_i)^2 S(s|X_i)^2} P(ds, 0|X_i) \right) \\
&+ (1 - R(\tau|X_i))^2 \left( \frac{I(\tilde{T}_i \leq \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} (F_1(\tau|X_i) - F_1(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_1(\tau|X_i) - F_1(s|X_i))}{G(s|X_i)^2 S(s|X_i)^2} P(ds, 0|X_i) \right) \\
&+ R(\tau|X_i)^2 \left( \frac{I(\tilde{T}_i \leq \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} (F_2(\tau|X_i) - F_2(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_2(\tau|X_i) - F_2(s|X_i))}{G(s|X_i)^2 S(s|X_i)^2} P(ds, 0|X_i) \right) \\
&= R(\tau|X_i)^2 S(\tau|X_i) \left( \frac{I(\tilde{T}_i \leq \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i \wedge \tau} \frac{1}{G(s|X_i) S(s|X_i)} \Lambda_C(ds|X_i) \right) \\
&+ (1 - R(\tau|X_i))^2 \left( \frac{I(\tilde{T}_i \leq \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} (F_1(\tau|X_i) - F_1(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_1(\tau|X_i) - F_1(s|X_i))}{G(s|X_i) S(s|X_i)} \Lambda_C(ds|X_i) \right) \\
&+ R(\tau|X_i)^2 \left( \frac{I(\tilde{T}_i \leq \tau, \Delta_i = 0)}{G(\tilde{T}_i|X_i) S(\tilde{T}_i|X_i)} (F_2(\tau|X_i) - F_2(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i \wedge \tau} \frac{(F_2(\tau|X_i) - F_2(s|X_i))}{G(s|X_i) S(s|X_i)} \Lambda_C(ds|X_i) \right)
\end{aligned}$$

Here we used that

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

## 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  $m \leq n$  with or without replacement from the data  $D_n$ . For  $i = 1, \dots, n$  let  $N_i^b$  be the number of times subject  $i$  is included in  $D_{m,b}^*$ . For subsampling bootstrap (without replacement),  $N_i^b$  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  $N_i^b = 0$  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 0:

$$\psi_m(P) = \int \left( \int \dots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^m P(do_i) \right) P(do_0)$$

for some sample size  $m < n$ . Let

$$\omega_m(Y_0, X_0) = \mathbb{E}_{D_m}[(Y_0 - R(D_m)(X_0))^2 | Y_0, X_0]$$

because then  $\psi_m(P) = \mathbb{E}_{O_0}[\omega_m(Y_0, X_0)]$ . For estimating  $\omega_{\tau,m}$ , we propose to use leave one-out bootstrap estimation:

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

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:

$$\text{IF}_\psi(m; o_k) = \frac{\partial}{\partial \epsilon} \psi_m(P_{\epsilon, k}) \Big|_{\epsilon=0} \quad (1)$$

$$= \int \frac{\partial}{\partial \epsilon} \left( \int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^m P_{\epsilon, k}(do_i) \right) P_{\epsilon, k}(do_0) \Big|_{\epsilon=0} \quad (2)$$

$$= \int \frac{\partial}{\partial \epsilon} \left( \int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^m P_{\epsilon, k}(do_i) \right) \Big|_{\epsilon=0} P(do_0) \quad (3)$$

$$+ \int \left( \int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^m P(do_i) \right) \frac{\partial}{\partial \epsilon} P_{\epsilon, k}(do_0) \Big|_{\epsilon=0} \quad (4)$$

$$= \int \sum_{j=1}^m \left( \int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} P(do_i) \right) P(do_0) - m \psi_m(P) \quad (5)$$

$$+ \omega_m(Y_k, X_k) - \psi_m(P) \quad (6)$$

$$= \omega_m(Y_k, X_k) - (m+1) \psi_m(P) \quad (7)$$

$$+ \int \sum_{j=1}^m \left( \int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} P(do_i) \right) P(do_0) \quad (8)$$

wherein we used the product rule of differentiation. Note we use the approximation that

$$\int \sum_{j=1}^m \left( \int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} P(do_i) \right) P(do_0) \approx m \psi_m(P)$$

Then

$$\text{IF}_\psi(m; o_k) = \omega_m(Y_k, X_k) - \psi_m(P)$$

For estimating the influence function, we suggest the estimator:

$$\widehat{\text{IF}}_\psi(m; O_k) = \hat{\omega}_m^{(1)}(Y_k, X_k) - \hat{\psi}_m^{(1)}$$

### 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  $\bar{O} = (\bar{T}, X)$  denote the true event time, i.e.  $T = \min\{\bar{T}, C\}$  and  $\Delta = I(T \leq C)$ , where  $C$  is the censoring time. Also let  $Y_i = I(\bar{T} \leq \tau)$  (or  $T$ , depending on whichever is the most appropriate), where  $\tau$  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  $\mu_{\tau, m}$

$$\mu_{\tau, m} = \mathbb{E}_{O_0} [\mathbb{E}_{D_m} [\{\mathcal{I}_{\{\bar{T}_0 \leq \tau\}} - R_\tau(D_m)(X_0)\}^2 | \bar{T}_0, X_0]]$$

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  $\psi_{\tau, m} : \mathcal{P} \rightarrow [0, 1]$

$$\begin{aligned} \psi_{\tau, m}(P) &= \int \left( \int \cdots \int \{\mathcal{I}_{\{u_0 \leq \tau\}} - R_\tau(D_m)(x_0)\}^2 \prod_{i=1}^m P(do_i) \right) \\ &\quad \times W_\tau(o_0, \kappa_{\tau, x_0}(P)) P(do_0) \\ &= \mu_{\tau, m}. \end{aligned}$$

with

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

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

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

Here, we redefine that

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

In much the same way as before, we may find the the influence function of  $\psi_{\tau,m}(\mathbf{P})$  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.

$$\begin{aligned} \text{IF}_{\psi}(\tau, m; o_k) &= \omega_{\tau,m}(t_k, x_k) W_{\tau}(o_k, \kappa_{\tau, x_0}(\mathbf{P})) - (m+1) \mu_{\tau,m} \\ &+ \int \left[ \sum_{j=1}^m \int \cdots \int \{\mathcal{I}_{\{t_0 \leq \tau\}} - R_{\tau}(\{o_i\}_{i=1}^n)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} \mathbf{P}(\text{d}o_i) \right] W_{\tau}(o_0, \kappa_{\tau, x_0}(\mathbf{P})) \mathbf{P}(\text{d}o_0) \\ &+ \int \omega_{\tau,m}(t_0, x_0) \left[ \frac{\mathcal{I}_{\{t_0 \leq \tau\}} \delta_0}{G(t_0 - |x_0|)} f_k(t_0, x_0) + \frac{\mathcal{I}_{\{t_0 > \tau\}}}{G(\tau | x_0)} f_k(\tau, x_0) \right] \mathbf{P}(\text{d}o_0) \end{aligned}$$

We can use the approximation that  $\int \left[ \sum_{j=1}^m \int \cdots \int \{\mathcal{I}_{\{t_0 \leq \tau\}} - R_{\tau}(\{o_i\}_{i=1}^n)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} \mathbf{P}(\text{d}o_i) \right] W_{\tau}(o_0, \kappa_{\tau, x_0}(\mathbf{P})) \mathbf{P}(\text{d}o_0) \approx m \mu_{\tau,m}$ . Then approximately,

$$\begin{aligned} \text{IF}_{\psi}(\tau, m; o_k) &\approx \omega_{\tau,m}(t_k, x_k) W_{\tau}(o_k, \kappa_{\tau, x_0}(\mathbf{P})) - \mu_{\tau,m} \\ &+ \int \omega_{\tau,m}(t_0, x_0) \left[ \frac{\mathcal{I}_{\{t_0 \leq \tau\}} \delta_0}{G(t_0 - |x_0|)} f_k(t_0, x_0) + \frac{\mathcal{I}_{\{t_0 > \tau\}}}{G(\tau | x_0)} f_k(\tau, x_0) \right] \mathbf{P}(\text{d}o_0) \end{aligned}$$

This approximation can be justified in the sense that influence functions should have mean zero. This can then be estimated in much the same way as before

$$\begin{aligned} \widehat{\text{IF}}_{\psi}(\tau, m; O_k) &= \hat{\omega}_{\tau,m}^{(1)}(T_k, X_k) W_{\tau}(O_k, \hat{G}_n) - \hat{\mu}_{\tau,m}^{(1)} \\ &+ \frac{1}{n} \sum_{i=1}^n \hat{\omega}_{\tau,m}^{(1)}(T_i, X_i) \left[ \frac{\mathcal{I}_{\{T_i \leq \tau\}} \Delta_i}{\hat{G}_n(T_i - |X_i|)} \hat{f}_k(T_i -, X_i) + \frac{\mathcal{I}_{\{T_i > \tau\}}}{\hat{G}_n(\tau | X_i)} \hat{f}_k(\tau, X_i) \right] \end{aligned}$$

Here the last term corresponds to the censoring being unknown. Note that this corresponds exactly to estimating as the train-validation case situation with the residuals now being the cross-validated residuals  $\hat{\omega}_{\tau,m}^{(1)}(T_k, X_k) W_{\tau}(O_k, \hat{G}_n)$ . This is for the survival case; for the competing risk case, we would use

$$\begin{aligned} \widehat{\text{IF}}_{\psi}(\tau, m; O_k) &= \hat{\omega}_{\tau,m}^{(1)}(T_k, X_k) W_{\tau}(O_k, \hat{G}_n) - \hat{\mu}_{\tau,m}^{(1)} \\ &+ \frac{1}{n} \sum_{i=1}^n \hat{\omega}_{\tau,m}^{(1)}(T_i, X_i) \left[ \frac{\mathcal{I}_{\{T_i \leq \tau, \Delta_i=1\}}}{\hat{G}_n(T_i - |X_i|)} \hat{f}_k(T_i -, X_i) + \frac{\mathcal{I}_{\{T_i \leq \tau, \Delta_i=2\}}}{\hat{G}_n(T_i - |X_i|)} \hat{f}_k(T_i -, X_i) + \frac{\mathcal{I}_{\{T_i > \tau\}}}{\hat{G}_n(\tau | X_i)} \hat{f}_k(\tau, X_i) \right] \end{aligned}$$

as an estimator.