Influence function calculation for Brier score for event time data

Johan Sebastian Ohlendorff & Thomas Alexander Gerds

2022-01-09

Introduction

In this document, we consider the estimation of the Brier score - a discrimination measure. Let $R(D_m)(X)$ be a (risk) prediction for X for a model R trained on a data set D_m of size m. Then, for binary data $Z = (Y, X)$, the Brier score is defined as:

$$\text{Brier}_{R,D_m} = \mathbb{E}[(Y - R(D_m)(X))^2] = \int (y - R(D_m)(x))^2 P(dz)$$

Note that in the above definition that D_m is fixed, so the Brier score above will depend on which data set that the model is trained on. 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 $\{0, 1, 2\}$. Also let G denote the survival function for the censoring distribution. Now for event type data $Z = (T, D, X)$, we consider the above as a time-dependent discrimination measure with $Y = I(T \leqslant \tau)$ for some fixed τ. Thus,

$$\begin{aligned}\text{Brier}_{R,D_m,\tau} &= \mathbb{E}[(I(T \leqslant \tau, D = 1) - R_\tau(D_m)(X))^2] = \int (I(t \leqslant \tau, d = 1) - R_\tau(D_m)(x))^2 Q(dz) \\ &= \int (I(t \leqslant \tau, \delta = 1) - R_\tau(D_m)(x))^2 W_\tau(z; G) P(dz) = \sum_{\delta=0,1,2} \int \left\{1_{\{t \leqslant \tau, \delta=1\}} R_\tau(D_m)(x)\right\}^2 W_\tau(z; G) P(dt, \delta, dx)\end{aligned}$$

Here we used IPCW with $W_\tau(z; G) = \frac{I(t \leqslant \tau, \delta=1)}{G(t-|x)} + \frac{I(t \leqslant \tau, \delta=2)}{G(t-|x)} + \frac{I(t > \tau)}{G(\tau|x)}$.

In the situation with cross-validation, it will be of interest to estimate $\mathbb{E}_{D_m}[\text{Brier}_{R,D_m}]$ or $\mathbb{E}_{D_m}[\text{Brier}_{R,D_m,\tau}]$, i.e. the expected performance of the model over all data sets of size m.

In the below sections, we will suggest some estimators of the Brier score and their asymptotic variances (by using influence functions). Also, we will calculate the efficient influence function.

Estimators of the Brier score

When we are dealing with the Brier score for binary data, we can consider the Brier score as a functional of the probability measure P. Plugging in the empirical measure $\mathbb{P}_n$ instead of P, we get

$$\widehat{\text{Brier}}_{R,D_m} = \int (y - R(D_m)(x))^2 \mathbb{P}_n(dz) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - R(D_m)(X_i))^2$$

By the functional delta method, one can derive the influence function of the above estimator as,

$$\text{IF}_{\widehat{\text{Brier}}_{R,D_m}}(Y_i, X_i) = (Y_i - R(D_m)(X_i))^2 - \text{Brier}_{R,D_m}$$

This is also the efficient influence function (because the estimator is fully nonparametric). One can estimate this consistently by

$$\hat{\mathrm{IF}}_{\widehat{\mathrm{Brier}}_{R,D_m}}(Y_i, X_i) = (Y_i - R(D_m)(X_i))^2 - \widehat{\mathrm{Brier}}_{R,D_m}$$

The situation with type data is more complicated as one needs to estimate the censoring weights in $W_\tau(z; G)$. However, if we assume that the censoring does not depend on the covariates, then it is natural to estimate G with the Kaplan-Meier estimator $\hat{G}_{\mathrm{KM}}$. Then by the plug-in principle, we consider the estimator,

$$\widetilde{\mathrm{Brier}}_{R,D_m,\tau} = \int (I(t \leqslant \tau, d = 1) - R(D_m)(x))^2 W_\tau(z; \hat{G}_{\mathrm{KM}})\mathbb{P}_n(dz) = \frac{1}{n}\sum_{i=1}^{n}(I(\tilde{T}_i \leqslant \tau, \Delta_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i; \hat{G}_{\mathrm{KM}})$$

The influence function of this may again be found by functional delta method as

$$\begin{aligned}\mathrm{IF}_{\widetilde{\mathrm{Brier}}_{R,D_m,\tau}}(\tilde{T}_i, \Delta_i) &= (I(\tilde{T}_i \leqslant \tau, \Delta_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i; G) - \mathrm{Brier}_{R,D_m,\tau} \\ &\quad + \int (I(t \leqslant \tau, \delta = 1) - R(D_m)(x))^2 \mathrm{IF}_{W_\tau, \hat{G}_{\mathrm{KM}}}(z; G)(\tilde{T}_i, \Delta_i) P(dz)\end{aligned}$$

with

$$\begin{aligned}\mathrm{IF}_{W_\tau, \hat{G}_{\mathrm{KM}}}(z; G)(\tilde{T}_i, \Delta_i) &= \frac{I(t \leqslant \tau, \delta = 1)}{G(t-)}\mathrm{IF}_{\hat{\Lambda}_C}(t-)(\tilde{T}_i, \Delta_i) + \frac{I(t \leqslant \tau, \delta = 2)}{G(t-)}\mathrm{IF}_{\hat{\Lambda}_C}(t-)(\tilde{T}_i, \Delta_i) \\ &\quad + \frac{I(t > \tau)}{G(\tau)}\mathrm{IF}_{\hat{\Lambda}_C}(\tau)(\tilde{T}_i, \Delta_i)\end{aligned}$$

where $\mathrm{IF}_{\hat{\Lambda}_C}(t)$ is the influence function of the cumulative hazard at time t of the Kaplan-Meier estimator of the censoring. We can estimate the influence function of the estimator, by plugging in $\hat{G}_{\mathrm{KM}}$ for G, $\hat{\mathrm{IF}}_{\hat{\Lambda}_C}(t)(\tilde{T}_i, \Delta_i)$ for $\mathrm{IF}_{\hat{\Lambda}_C}(t)(\tilde{T}_i, \Delta_i)$ and $\mathbb{P}_n$ for P and of course, the Brier score with this estimator.

In a similar manner, we can also derive the influence function, when we instead try to estimate G by a Cox model. Then,

$$\widetilde{\mathrm{Brier}}_{R,D_m,\tau} = \int (I(t \leqslant \tau, d = 1) - R(D_m)(x))^2 W_\tau(z; \hat{G}_{\mathrm{Cox}})\mathbb{P}_n(dz) = \frac{1}{n}\sum_{i=1}^{n}(I(\tilde{T}_i \leqslant \tau, \Delta_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i; \hat{G}_{\mathrm{Cox}})$$

The influence function of this may again be found by functional delta method as

$$\begin{aligned}\mathrm{IF}_{\widetilde{\mathrm{Brier}}_{R,D_m,\tau}}(\tilde{T}_i, \Delta_i, X_i) &= (I(\tilde{T}_i \leqslant \tau, \Delta_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i; G) - \mathrm{Brier}_{R,D_m,\tau} \\ &\quad + \int (I(t \leqslant \tau, \delta = 1) - R(D_m)(x))^2 \mathrm{IF}_{W_\tau, \hat{G}_{\mathrm{Cox}}}(z; G)(\tilde{T}_i, \Delta_i, X_i) P(dz)\end{aligned}$$

with

$$\begin{aligned}\mathrm{IF}_{W_\tau, \hat{G}_{\mathrm{Cox}}}(z; G)(\tilde{T}_i, \Delta_i, X_i) &= \frac{I(t \leqslant \tau, \delta = 1)}{G(t - |x)}\mathrm{IF}_{\hat{\Lambda}_C}(t-, x)(\tilde{T}_i, \Delta_i, X_i) + \frac{I(t \leqslant \tau, \delta = 2)}{G(t - |x)}\mathrm{IF}_{\hat{\Lambda}_C}(t-, x)(\tilde{T}_i, \Delta_i, X_i) \\ &\quad + \frac{I(t > \tau)}{G(\tau|x)}\mathrm{IF}_{\hat{\Lambda}_C}(\tau, x)(\tilde{T}_i, \Delta_i, X_i)\end{aligned}$$

where $\mathrm{IF}_{\hat{\Lambda}_C}(t, x)$ is the influence function of the cumulative hazard estimated by a Cox regression at time t given covariate x. We can estimate the influence function of the estimator, by plugging in $\hat{G}_{\mathrm{Cox}}$ for G, $\hat{\mathrm{IF}}_{\hat{\Lambda}_C}(t, x)(\tilde{T}_i, \Delta_i)$ for $\mathrm{IF}_{\hat{\Lambda}_C}(t, x)(\tilde{T}_i, \Delta_i)$ and $\mathbb{P}_n$ for P. This yields that

$$\begin{aligned}\hat{\mathrm{IF}}_{\widetilde{\mathrm{Brier}}_{R,D_m,\tau}}(\tilde{T}_i, \Delta_i, X_i) &= (I(\tilde{T}_i \leqslant \tau, \Delta_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i; \hat{G}_{\mathrm{Cox}}) - \widetilde{\mathrm{Brier}}_{R,D_m,\tau} \\ &\quad + \int (I(t \leqslant \tau, \delta = 1) - R(D_m)(x))^2 \hat{\mathrm{IF}}_{W_\tau, \hat{G}_{\mathrm{Cox}}}(z; \hat{G}_{\mathrm{Cox}})(\tilde{T}_i, \Delta_i, X_i)\mathbb{P}_n(dz) \\ &= (I(\tilde{T}_i \leqslant \tau, \Delta_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i; \hat{G}_{\mathrm{Cox}}) - \widetilde{\mathrm{Brier}}_{R,D_m,\tau} \\ &\quad + \frac{1}{n}\sum_{j=1}^{n}(I(\tilde{T}_j \leqslant \tau, \Delta_j = 1) - R(D_m)(X_j))^2 \hat{\mathrm{IF}}_{W_\tau, \hat{G}_{\mathrm{Cox}}}(Z_j; \hat{G}_{\mathrm{Cox}})(\tilde{T}_i, \Delta_i, X_i)\end{aligned}$$

with

$$\hat{\mathrm{IF}}_{W_\tau,\hat{G}_{\mathrm{Cox}}}(z;\hat{G}_{\mathrm{Cox}})(\tilde{T}_i,\Delta_i,X_i) = \frac{I(t \leqslant \tau, \delta = 1)}{\hat{G}_{\mathrm{Cox}}(t-|x)}\hat{\mathrm{IF}}_{\hat{\Lambda}_C}(t-,x)(\tilde{T}_i,\Delta_i,X_i) + \frac{I(t \leqslant \tau, \delta = 2)}{\hat{G}_{\mathrm{Cox}}(t-|x)}\hat{\mathrm{IF}}_{\hat{\Lambda}_C}(t-,x)(\tilde{T}_i,\Delta_i,X_i) + \frac{I(t > \tau)}{\hat{G}_{\mathrm{Cox}}(\tau|x)}\hat{\mathrm{IF}}_{\hat{\Lambda}_C}(\tau,x)(\tilde{T}_i,\Delta_i,X_i)$$

Here is $\hat{\mathrm{IF}}_{\hat{\Lambda}_C}(t,x)$ is the estimated influence function of the cumulative hazard when estimated by a Cox regression at time t given covariate x.

Calculation of the efficient influence function for event type data

Taking the Gateaux derivative of the Brier score written as a functional of P and G yields,

$$\mathrm{IF}_{\mathrm{Brier}_{R,D_m,\tau}}(\tilde{T}_i,\Delta_i,X_i) = (I(\tilde{T}_i \leqslant \tau, D_i = 1) - R_\tau(D_m)(X_i))^2 W_\tau(Z_i;G) - \mathrm{Brier}_{R,D_m,\tau} + \int (I(t \leqslant \tau, \delta = 1) - R(D_m)(x))^2 \mathrm{IF}_{W_\tau,G}(z;G)(\tilde{T}_i,\Delta_i,X_i)P(dz)$$

with

$$\mathrm{IF}_{W_\tau,\hat{G}_{\mathrm{Cox}}}(z;G)(\tilde{T}_i,\Delta_i,X_i) = \frac{I(t \leqslant \tau, \delta = 1)}{G(t-|x)}\mathrm{IF}_{\Lambda_C}(t-,x)(\tilde{T}_i,\Delta_i,X_i) + \frac{I(t \leqslant \tau, \delta = 2)}{G(t-|x)}\mathrm{IF}_{\Lambda_C}(t-,x)(\tilde{T}_i,\Delta_i,X_i) + \frac{I(t > \tau)}{G(\tau|x)}\mathrm{IF}_{\Lambda_C}(\tau,x)(\tilde{T}_i,\Delta_i,X_i)$$

Here we have that

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

which is the efficient influence function of the cumulative hazard of the censoring. This is in general not a proper function of x. However, when both the outcome and the censoring does not depend on the covariates, one can see that this influence function coincides with the one for Kaplan-Meier censoring. In general though, our estimators are not efficient. Using the properties of the dirac measure yield that,

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

where Λ_C is the cumulative hazard of the censoring and F_1 and F_2 are the subdistribution functions (or risk functions) of event 1 and 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 $m \leqslant n$ 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:

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

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:

$$\begin{aligned}
\mathrm{IF}_{\psi}(m; o_k) &= \frac{\partial}{\partial \epsilon} \psi_m(P_{\epsilon,k}) \Big|_{\epsilon=0} && (1)\\
&= \int \frac{\partial}{\partial \epsilon} \left(\int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^{m} P_{\epsilon,k}(do_i) \right) P_{\epsilon,k}(do_0) \Big|_{\epsilon=0} && (2)\\
&= \int \frac{\partial}{\partial \epsilon} \left(\int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^{m} P_{\epsilon,k}(do_i) \right) \Big|_{\epsilon=0} P(do_0) && (3)\\
&+ \int \left(\int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \prod_{i=1}^{m} P(do_i) \right) \frac{\partial}{\partial \epsilon} P_{\epsilon,k}(do_0) \Big|_{\epsilon=0} && (4)\\
&= \int \sum_{j=1}^{m} \left(\int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} P(do_i) \right) P(do_0) - m\psi_m(P) && (5)\\
&+ \omega_m(Y_k, X_k) - \psi_m(P) && (6)\\
&= \omega_m(Y_k, X_k) - (m+1)\psi_m(P) && (7)\\
&+ \int \sum_{j=1}^{m} \left(\int \cdots \int \{y_0 - R(D_m)(x_0)\}^2 \delta_{o_k}(o_j) \prod_{i \neq j} P(do_i) \right) P(do_0) && (8)
\end{aligned}$$

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

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

Then

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

For estimating the influence function, we suggest the estimator:

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

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 $\bar{O} = (\tilde{T}, X)$ 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 $\mu_{\tau,m}$

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

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

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

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

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

Here, we redefine that

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

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.

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

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

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

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

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

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

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

as an estimator.