Derivation of the Influence Function for AUC for competing risk data and survival data

Johan Sebastian Ohlendorff & Thomas Alexander Gerds

2022-01-09

Introduction

In this document, we consider the estimation of the (possibly time-dependent) AUC. 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 $Z = (Y, X)$ and $Z' = (Y', X')$ with $Z \perp\!\!\!\perp Z'$, the AUC score is defined as:

$$\text{AUC}_{R,D_m} = P(R(D_m)(X) > R(D_m)(X')|Y = 1, Y' = 0) = \frac{\int\int I(R(D_m)(x) > R(D_m)(x'), y = 1, y' = 0)P(dz')P(dz)}{P(Y = 1)P(Y' = 0)}$$

Note that in the above definition that D_m is fixed, so the AUC 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. 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 $Z = (T, D, X)$ and $Z' = (T', D', X')$ with $Z \perp\!\!\!\perp Z'$, we consider the above as a time-dependent discrimination measure for some fixed τ:

$$\begin{aligned}
\text{AUC}_{R,D_m,\tau} &= Q(R(D_m)(X) > R(D_m)(X')|T \leqslant \tau, D = 1, (T' > \tau \cup D' = 2)) \\
&= \frac{Q(R(D_m)(X) > R(D_m)(X'), T \leqslant \tau, D = 1, (T' > \tau \cup D' = 2))}{Q(T \leqslant \tau, D = 1, (T' > \tau \cup D' = 2))} \\
&= \frac{\nu_\tau(Q)}{\mu_\tau(Q)}
\end{aligned}$$

with

$$\begin{aligned}
\nu_\tau(Q) &= \iint 1_{\{R(D_m)(x) > R(D_m)(x'), t \leqslant \tau, t' > \tau, d = 1\}} \frac{P(dz)}{G(t-|x)} \frac{P(dz')}{G(\tau|x')} \\
&+ \iint 1_{\{R(D_m)(x) > R(D_m)(x'), t \leqslant \tau, t' \leqslant \tau, d = 1, d' = 2\}} \frac{P(dz)}{G(t-|x)} \frac{P(dz')}{G(t'-|x')}
\end{aligned}$$

and

$$\mu_\tau(Q) = \iint 1_{\{t \leqslant \tau, t' > \tau, d = 1\}} \frac{P(dz)}{G(t-|x)} \frac{P(dz')}{G(\tau|x')} + \iint 1_{\{t \leqslant \tau, t' \leqslant \tau, d = 1, d' = 2\}} \frac{P(dz)}{G(t-|x)} \frac{P(dz')}{G(t'-|x')}$$

When there are no competing risks, the definition of the above is the same as for binary data with $Y = I(T \leqslant \tau)$ and $Y' = I(T' > \tau)$. Here we used IPCW to rewrite the above, such that we can actually estimate it from observed data. It is straightforward to write down these plugin-in estimators.

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

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

Efficient influence function with known censoring

We will only cover the competing risk case with uncensored data (or more generally with known censoring distribution), as the efficient influence function for AUC for binary data and survival data follow from this. To derive the Gateaux derivative of the functional at $\mathrm{AUC}_{R,D_m,\tau}$ in the direction of an observation (T_i, D_i, X_i) we introduce the corresponding path,

$$Q_\varepsilon^i = Q + \varepsilon \left\{ \delta_{\{T_i,D_i,X_i\}} - Q \right\}.$$

and the following short notation $\partial_\varepsilon = \frac{\partial}{\partial\varepsilon}\Big|_{\varepsilon=0}$ to obtain the directional derivative:

$$\partial_\varepsilon Q_\varepsilon^i = \delta_{\{T_i,D_i,X_i\}} - Q.$$

The Gateaux derivative of the functional $\mathrm{AUC}_{R,D_m,\tau}$ is obtained using straight-forward calculus:

$$\mathrm{IF}_{G \text{ known},\mathrm{AUC}}(Z_i,\tau) = \frac{\mathrm{IF}_\nu(Z_i;\tau)\mu_\tau(Q) - \nu_\tau(Q)\mathrm{IF}_\mu(Z_i;\tau)}{\mu_\tau(Q)^2}$$

where

$$\begin{aligned}\mathrm{IF}_\nu(Z_i;\tau) = &\frac{1_{\{\tilde{T}_i \leqslant \tau, D_i=1\}}}{G(\tilde{T}_i - |X_i)} \int 1_{\{R(D_m)(X_i) > R(D_m)(x')\}} \left(1_{\{t' > \tau\}} \frac{1}{G(\tau|x')} + 1_{\{t' \leqslant \tau, d'=2\}} \frac{1}{G(t' - |x')} \right) P(dz') \\ &+ \left(\frac{1_{\{\tilde{T}_i \leqslant \tau, D_i=2\}}}{G(\tilde{T}_i - |X_i)} + \frac{1_{\{\tilde{T}_i > \tau\}}}{G(\tau|X_i)} \right) \int 1_{\{R(D_m)(x) > R(D_m)(X_i), t \leqslant \tau, d=1\}} \frac{P(dz)}{G(t - |x)}\end{aligned}$$

and

$$\begin{aligned}\mathrm{IF}_\mu(Z_i;\tau) = &\frac{1_{\{\tilde{T}_i \leqslant \tau, D_i=1\}}}{G(\tilde{T}_i - |X_i)} \int \left(1_{\{t' > \tau\}} \frac{1}{G(\tau|x')} + 1_{\{t' \leqslant \tau, d'=2\}} \frac{1}{G(t' - |x')} \right) P(dz') \\ &+ \left(\frac{1_{\{\tilde{T}_i \leqslant \tau, D_i=2\}}}{G(\tilde{T}_i - |X_i)} + \frac{1_{\{\tilde{T}_i > \tau\}}}{G(\tau|X_i)} \right) \int 1_{\{t \leqslant \tau, d=1\}} \frac{P(dz)}{G(t - |x)}\end{aligned}$$

Efficient influence function

Similar to the case from before, we can obtain the efficient influence function when the censoring is unknown and has to be estimated. We need the influence function for the censoring as part of these calculations:

$$\begin{aligned}\kappa_{t,x}(P_\varepsilon^i) &= \exp\left(-\int_0^t \frac{dP_\varepsilon(s,0|x)}{P_\varepsilon(\tilde{T} > s|X = x)}\right) \\ \partial_\varepsilon P_\varepsilon(\tilde{T} > s, X = x) &= \mathbb{1}_{\{\tilde{T}_i > s, X_i = x\}} - P(\tilde{T} > s, X = x) \\ \partial_\varepsilon P_\varepsilon(\tilde{T} \leqslant s, \Delta = 0, X = x) &= \mathbb{1}_{\{\tilde{T}_i \leqslant s, \Delta_i = 0, X_i = x\}} - P(\tilde{T} \leqslant s, \Delta = 0, X = x)\end{aligned}$$

Let $f_i(t,x) = \partial_\varepsilon \left[\int_0^t \frac{dP_\varepsilon(s,0|x)}{P_\varepsilon(\tilde{T}>s|X=x)} \right]$, so that $f_i(t,x)$ is the influence function of the cumulative hazard (conditioned on x) of the censoring. We observe that

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

Here δ denotes the Dirac measure. When X is discrete, $\delta_{X_i}(x)$ is to be understood as $\frac{I(X_i=x)}{P(X_i=x)}$. With this in mind, we find that

$$\mathrm{IF}_{\mathrm{AUC}}(Z_i,\tau) = \mathrm{IF}_{G \text{ known},\mathrm{AUC}}(Z_i,\tau) + \mathrm{IF}_{\mathrm{AUC}}(Z_i,\tau)_{\mathrm{cens}}$$

where

$$\mathrm{IF}_{\mathrm{AUC}}(Z_i,\tau)_{\mathrm{cens}} = \frac{\mathrm{IF}_\nu(Z_i;\tau)_{\mathrm{cens}}\mu_\tau(Q) - \nu_\tau(Q)\mathrm{IF}_\mu(Z_i;\tau)_{\mathrm{cens}}}{\mu_\tau(Q)^2}$$

and

$$\mathrm{IF}_\nu(Z_i;\tau)_{\mathrm{cens}} = \iint 1_{\{R(D_m)(x)>R(D_m)(x')\}} \left(1_{\{t'>\tau\}} \frac{1}{G(\tau|x')} + 1_{\{t'\leqslant\tau,d'=2\}} \frac{1}{G(t'-|x')}\right) P(dz') 1_{\{t\leqslant\tau,d=1\}} \frac{f_i(t-,x)P(dz)}{G(t-|x)}$$
$$+ \iint 1_{\{R(D_m)(x)>R(D_m)(x'),t\leqslant\tau,d=1\}} \frac{P(dz)}{G(t-|x)} \left(\frac{1_{\{t'>\tau\}} f_i(\tau,x')}{G(\tau|x')} + 1_{\{t'\leqslant\tau,d'=2\}} f_i(t'-,x') \frac{1}{G(t'-|x')}\right) P(dz')$$

and

$$\mathrm{IF}_\mu(Z_i;\tau)_{\mathrm{cens}} = \int \left(1_{\{t'>\tau\}} \frac{1}{G(\tau|x')} + 1_{\{t'\leqslant\tau,d'=2\}} \frac{1}{G(t'-|x')}\right) P(dz') \int 1_{\{t\leqslant\tau,d=1\}} \frac{f_i(t-,x)P(dz)}{G(t-|x)}$$
$$+ \int 1_{\{t\leqslant\tau,d=1\}} \frac{P(dz)}{G(t-|x)} \int \left(\frac{1_{\{t'>\tau\}} f_i(\tau,x')}{G(\tau|x')} + 1_{\{t'\leqslant\tau,d'=2\}} f_i(t'-,x') \frac{1}{G(t'-|x')}\right) P(dz')$$

Notes on the case with covariates

- Some of the censoring may be evaluated by noting the following

  $$\int_X \int_0^\tau 1_{\{d=k\}} H(x) f_i(t-,x) \frac{P(dz)}{G(t-|x)} = H(X_i)\left(\frac{I(\tilde{T}_i\leqslant\tau,\Delta_i=0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)}(F_k(\tau|X_i)-F_k(\tilde{T}_i|X_i)) - \int_0^{\tilde{T}_i\wedge\tau} \frac{(F_k(\tau|X_i)-F_k(s|X_i))}{G(s|X_i)^2 S(s|X_i)^2} P(ds,0|X_i)\right)$$

  where H is an arbitrary function of x (but not t!), $k \in \{1,2\}$ and F_k is the subdistribution function for cause k.
- For Term (3) (and similarly (10)), we can use that

  $$\int_X \int H(x') 1_{\{t'>\tau\}} f_i(\tau,x') \frac{P(dz')}{G(\tau|x')} = \frac{H(X_i)}{G(\tau|X_i)} P(\tilde{T}>\tau|X_i) \left(\frac{I(\tilde{T}_i\leqslant\tau,\Delta_i=0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i\wedge\tau} \frac{1}{G(s|X_i)^2 S(s|X_i)^2} P(ds,0|X_i)\right)$$
  $$= H(X_i)S(\tau|X_i)\left(\frac{I(\tilde{T}_i\leqslant\tau,\Delta_i=0)}{G(\tilde{T}_i|X_i)S(\tilde{T}_i|X_i)} - \int_0^{\tilde{T}_i\wedge\tau} \frac{1}{G(s|X_i)^2 S(s|X_i)^2} P(ds,0|X_i)\right)$$

  where H is again an arbitrary function.

Estimation

We can estimate the AUC by using the plugin-estimator $\mathbb{P}_n$ instead of P. We also need to estimate the censoring distribution, typically either by a Cox model $\hat{G}_{\mathrm{cox}}$ or a Kaplan-Meier model $\hat{G}_{\mathrm{KM}}$, or maybe some other model. The influence function for the first two estimators is fairly easy to estimate (the functional delta theorem will yield that the influence function for these estimators is pretty much the same as the efficient influence function, except $f_i(t,z)$ being given in a slightly different way). In fact, one can fairly easily implement $\hat{f}_i(t,z)$ for these models, so that the influence function of these estimators can be estimated. However, this is generally not so for other types of estimators, and for these estimators, we will drop the censoring term, leading to conservative confidence interval, i.e. assume that $\mathrm{IF}_{\mathrm{AUC}}(Z_i,\tau)_{\mathrm{cens}} = 0$.

In any case, we can actually estimate the terms in the influence function very (computationally) efficiently that do not involve the censoring by sorting the risk predictions.

For some ideas of how to implement the estimation of the censoring terms when we have Kaplan-Meier censoring, see **IFKMCens.pdf**.

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 $D^*_{m,b} = \{X^*_{b,1},\dots,X^*_{b,m}\}$ of size m with 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 $D^*_{m,b}$. 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. First let us construct the influence function in this case with binary data, $X = (Y, Z)$. In this case, we consider estimation of the functional (i.e. the definition of the AUC in cross-validation)

In this section we extend the ideas from the previous section and define a leave-pair-out bootstrap IPCW estimator of AUC. The AUC can be written as

$$\begin{aligned}\mathrm{AUC}(D_m) &= \mathbb{E}_{X_{m+1},X_{m+2}}\Big[\mathcal{I}_{\{R_\tau(D_m)(Z_{m+1})>R_\tau(D_m)(Z_{m+2})\}}\big|Y_{m+1}=1, Y_{m+2}=0, D_m\Big] \\ &= \frac{\mathbb{E}_{X_{m+1},X_{m+2}}\Big[\mathcal{I}_{\{R_\tau(D_m)(Z_{m+1})>R_\tau(D_m)(Z_{m+2})\}}\mathcal{I}_{\{Y_{m+1}=1\}}\mathcal{I}_{\{Y_{m+2}=0\}}\big|D_m\Big]}{\mathbb{E}_{X_{m+1}}\Big[\mathcal{I}_{\{Y_{m+1}=1\}}\Big]\mathbb{E}_{X_{m+2}}\Big[\mathcal{I}_{\{Y_{m+2}=0\}}\Big]}\end{aligned}$$

With the notation

$$\Theta_m(Z_{m+1}, Z_{m+2}) = \mathbb{E}_{D_m}\big[\mathcal{I}_{\{R_\tau(D_m)(Z_{m+1})>R_\tau(D_m)(Z_{m+2})\}}\big|Z_{m+1}, Z_{m+2}\big],$$

the expected parameter $\theta_m = \mathbb{E}_{D_n}\big[\mathrm{AUC}(D_n)\big]$ can then be written as

$$\begin{aligned}\theta_m &= \mathbb{E}_{D_m}\left[\frac{\mathbb{E}_{X_{m+1},X_{m+2}}\Big[\mathcal{I}_{\{R_\tau(D_m)(Z_{m+1})>R_\tau(D_m)(Z_{m+2})\}}\mathcal{I}_{\{Y_{m+1}=1\}}\mathcal{I}_{\{Y_{m+2}=0\}}\big|D_m\Big]}{\mathbb{E}_{X_{m+1}}\Big[\mathcal{I}_{\{Y_{m+1}=1\}}\Big]\mathbb{E}_{X_{m+2}}\Big[\mathcal{I}_{\{Y_{m+2}=0\}}\Big]}\right] \\ &= \frac{\mathbb{E}_{X_{m+1},X_{m+2}}\left[\Theta_m(Z_{m+1}, Z_{m+2})\mathcal{I}_{\{Y_{m+1}=1\}}\mathcal{I}_{\{Y_{m+2}=0\}}\right]}{\mathbb{E}_{X_{m+1}}\Big[\mathcal{I}_{\{Y_{m+1}=1\}}\Big]\mathbb{E}_{X_{m+2}}\Big[\mathcal{I}_{\{Y_{m+2}=0\}}\Big]} = \frac{\mathcal{A}_{\tau,m}}{\mathcal{B}_\tau}\end{aligned} \tag{1}$$

As before, let $D^*_{m,b}$ be a bootstrap sample drawn with replacement from D_m, and let $R(D^*_{m,b})$ be a prediction modeling algorithm trained in $D^*_{m,b}$. The idea of the leave-pair-out bootstrap estimator is to evaluate the concordance of predicted risks obtained from $R(D^*_{m,b})$ for all pairs of subjects (i, j) for which both subject i and subject j is out-of-bag

$$\mathcal{I}_{\{R(D^*_{m,b})(Z_i)>R(D^*_{m,b})(Z_j)\}}\mathcal{I}_{\{N_i^b=0\}}\mathcal{I}_{\{N_j^b=0\}}$$

The contribution of pair (i, j) to the leave-pair-out bootstrap estimator is the average concordance over all bootstrap samples for which both i and j are out-of-bag

$$\hat{\Theta}_m^{(1,1)}(Z_i, Z_j) = \frac{\sum_{b=1}^B \mathcal{I}_{\{R(D^*_{m,b})(Z_i)>R(D^*_{m,b})(Z_j)\}}\mathcal{I}_{\{N_i^b=0\}}\mathcal{I}_{\{N_j^b=0\}}}{\sum_{b=1}^B \mathcal{I}_{\{N_i^b=0\}}\mathcal{I}_{\{N_j^b=0\}}} \tag{2}$$

In the case of no censoring, the leave-pair-out bootstrap estimate of $\theta_{\tau,m}$ is obtained by inserting $\hat{\Theta}_m^{(1,1)}$ in θ_m and replacing the expectations by empirical means

$$\frac{1}{n^2}\frac{1}{\hat{\mathcal{B}}}\sum_{i=1}^n\sum_{j=1}^n \hat{\Theta}_m^{(1,1)}(Z_i, Z_j)\mathcal{I}_{\{Y_i=1\}}\mathcal{I}_{\{Y_j=0\}},$$

with $\hat{\mathcal{B}} = n^{-2}\big[\sum_{i=1}^n \mathcal{I}_{\{Y_i=1\}}\big]\big[\sum_{j=1}^n \mathcal{I}_{\{Y_j=0\}}\big]$.

Influence function (generally)

We note that we may actually use our results from the test-train section on any functional, $\phi(D_m)_\tau$ defined in the test-train situation. Then the influence function of the function $\Phi_\tau := \mathbb{E}_{D_m}[\phi(D_m)_\tau]$ (with τ removed from the notation if we are in the binary sitaution) from the test-train situation as follows by finding the Gateaux derivative for observation i

$$\int \cdots \int \mathrm{IC}^i_{\phi(D_m)_\tau} P(dx_1)\cdots P(dx_m) + \sum_{j=1}^{m} \int \cdots \int \phi(D_m)_\tau \delta_{x_j}(x_k) \prod_{i\neq j} P(dx_i) - m\Phi_\tau$$

Then using that approximately, we have that

$$\int \cdots \int \phi(D_m)_\tau \delta_{x_j}(x_k) \prod_{i\neq j} P(dx_i) \approx \int \cdots \int \phi(D_m)_\tau P(dx_j) \prod_{i\neq j} P(dx_i) := \Phi_\tau$$

i.e. we use the expected value as an approximation of the actual value (this is ok assuming that the variance of the LHS is not too large), so we have the Gateux derivative

$$\int \cdots \int \mathrm{IC}^i_{\phi(D_m)_\tau} P(dx_1)\cdots P(dx_m)$$

and we know

$$\mathrm{IC}^i_{\phi(D_m)_\tau}$$

from the train-test situation. Thus, by interchanging the integrals, we get after some calculations,

$$\begin{aligned}
\mathrm{IF}_\nu(m; x_k) = \frac{1}{\eta(\mathrm{P})}\Big[&\int \Theta_m(z_{m+1}, z_k)\mathcal{I}_{\{y_k=1, y_{m+2}=0\}}\mathrm{P}(\mathrm{d}x_{m+1}) \\
&+ \int \Theta_m(z_k, z_{m+2})\mathcal{I}_{\{y_{m+1}=1, y_k=0\}}\mathrm{P}(\mathrm{d}x_{m+2})\Big] \\
- \frac{\nu_m(\mathrm{P})}{\eta(\mathrm{P})}&\Big[\mathcal{I}_{\{y_k=1\}}\Big]\Big[\int \mathcal{I}_{\{y_{m+2}=0\}}\mathrm{P}(\mathrm{d}x_{m+2})\Big] \\
- \frac{\nu_n(\mathrm{P})}{\eta(\mathrm{P})}&\Big[\int \mathcal{I}_{\{y_{m+1}=1\}}\mathrm{P}(\mathrm{d}x_{m+1})\Big]\Big[\mathcal{I}_{\{y_k=0\}}\Big]
\end{aligned}$$

Estimating the Influence function

We suggest to estimate the influence function by

$$\begin{aligned}
\hat{\mathrm{IF}}_\nu(\tau, m; X_k) = \frac{1}{n}\frac{1}{\hat{\mathcal{B}}}&\left[\sum_{i=1}^{n} \hat{\Theta}^{(1,1)}_m(Z_i, Z_k)\mathcal{I}_{\{Y_i=1, Y_k=0\}} + \sum_{j=1}^{n} \hat{\Theta}^{(1,1)}_m(Z_k, Z_j)\mathcal{I}_{\{Y_k=1, Y_j=0\}}\right] \\
- \frac{\hat{\theta}^{(1,1)}_m}{\hat{\mathcal{B}}}&\left[\mathcal{I}_{\{Y_k=1\}}\right]\left[\frac{1}{n}\sum_{j=1}^{n}\mathcal{I}_{\{Y_j=0\}}\right] \\
- \frac{\hat{\theta}^{(1,1)}_m}{\hat{\mathcal{B}}_\tau}&\left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_{\{Y_i=1\}}\right]\left[\mathcal{I}_{\{Y_k=0\}}\right]
\end{aligned}$$

Survival and Competing risk cases

The idea is the same as before. Let $\nu^1_{\tau,m}(P)$ be the numerator of the cross-validated AUC. Then, for the most general case (competing risk case), we get that

$$\frac{\mathrm{IF}_\nu(Z_i; \tau, m)\mu_\tau(P) - \nu^1_{\tau,m}(P)\mathrm{IF}_\mu(Z_i; \tau)}{\mu_\tau(P)^2}$$

Note that $\mu_\tau(P)$ is defined precisely as before (and also IF_μ is the same, because they are constants in terms of $(x_1, ..., x_m)$), but we must redefine IF_ν, so

$$\text{IF}_\nu(Z_i;\tau,m) = \frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i=1\}}}{G(\tilde{T}_i - |Z_i)} \int \Theta_m(X_i,x') \left(1_{\{t'>\tau\}} \frac{1}{G(\tau|z')} + 1_{\{t' \leqslant \tau\}} \frac{\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}\right) P(dx') \tag{3}$$

$$+ \iint \Theta_m(x,x') \left(1_{\{t'>\tau\}} \frac{1}{G(\tau|z')} + 1_{\{t' \leqslant \tau\}} \frac{\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}\right) P(dx') 1_{\{t \leqslant \tau\}} \frac{f_i(t-,z)P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)} \tag{4}$$

$$+ \iint \Theta_m(x,x') 1_{\{t \leqslant \tau\}} \frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)} \left(\frac{1_{\{t'>\tau\}} f_i(\tau,z')}{G(\tau|z')} + 1_{\{t' \leqslant \tau\}} f_i(t'-,z') \frac{\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}\right) P(dx') \tag{5}$$

$$+ \left(\frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i=2\}}}{G(\tilde{T}_i - |Z_i)} + \frac{1_{\{\tilde{T}_i > \tau\}}}{G(\tau|Z_i)}\right) \int \Theta_m(x,X_i) 1_{\{t \leqslant \tau\}} \frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|x)} \tag{6}$$

Similarly for the denominator, we need to do this, so

$$\text{IF}_\mu(Z_i;\tau) = \frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i=1\}}}{G(\tilde{T}_i - |Z_i)} \int \left(1_{\{t'>\tau\}} \frac{1}{G(\tau|z')} + 1_{\{t' \leqslant \tau\}} \frac{\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}\right) P(dx') \tag{7}$$

$$+ \int \left(1_{\{t'>\tau\}} \frac{1}{G(\tau|z')} + 1_{\{t' \leqslant \tau\}} \frac{\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}\right) P(dx') \int 1_{\{t \leqslant \tau\}} \frac{f_i(t-,z)P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)} \tag{8}$$

$$+ \int 1_{\{t \leqslant \tau\}} \frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)} \int \left(\frac{1_{\{t'>\tau\}} f_i(\tau,z')}{G(\tau|z')} + 1_{\{t' \leqslant \tau\}} f_i(t'-,z') \frac{\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}\right) P(dx') \tag{9}$$

$$+ \left(\frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i=2\}}}{G(\tilde{T}_i - |Z_i)} + \frac{1_{\{\tilde{T}_i > \tau\}}}{G(\tau|Z_i)}\right) \int 1_{\{t \leqslant \tau\}} \frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|x)} \tag{10}$$