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 derive the influence function of the inverse probability of censoring weighted (IPCW) estimator of the time-dependent AUC in the cases with and without competing risks. We denote the uncensored time to event by T, the right censoring time by C, and by X a continuous marker. Let us first look at the case, where T is not censored (and there is only one event).

Influence function uncensored case

In the uncensored case, we observe (T, X) and denote Q for the joint probability measure Q, i.e., $(T, X) \sim Q$. The time-dependent AUC at time τ for marker X is defined as:

$$\text{AUC}(\tau) = Q(X > X' | T \leqslant \tau, T' > \tau) = \frac{Q(T \leqslant \tau, T' > \tau, X > X')}{Q(T \leqslant \tau, T' > \tau)}$$

Let $\mathcal{Q}$ be the set of all probability measures of T, X and consider the functional $\nu_\tau : \mathcal{Q} \to [0, 1]$

$$\nu_\tau(Q) = \int\int 1_{\{t \leqslant \tau, t' > \tau, x > x'\}} dQ(t, x) dQ(t', x')$$

Note that the value of the functional is the numerator of the AUC:

$$\nu_\tau(Q) = Q(T \leqslant \tau, T' > \tau, X > X')$$

To derive the Gateaux derivative of the functional at Q in the direction of an observation T_i, X_i we introduce the corresponding path,

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

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

$$\partial_\varepsilon Q^i_\varepsilon = \delta_{\{T_i, X_i\}} - Q.$$

The Gateaux derivative of the functional ν_τ is obtained using straight-forward calculus:

$$\begin{aligned}
\partial_\varepsilon \nu_\tau(Q_\varepsilon) &= \int\int 1_{\{t \leqslant \tau, t' > \tau, x > x'\}} \partial_\varepsilon Q^i_\varepsilon(t,x) dQ(t',x') \\
&\quad + \int\int 1_{\{t \leqslant \tau, t' > \tau, x > x'\}} dQ(t,x) \partial_\varepsilon Q^i_\varepsilon(t',x') \\
&= \int\int 1_{\{t \leqslant \tau, t' > \tau, x > x'\}} d[\delta_{\{T_i,X_i\}} - Q](t,x) dQ(t',x') \\
&\quad + \int\int 1_{\{t \leqslant \tau, t' > \tau, x > x'\}} dQ(t,x) d[\delta_{\{T_i,X_i\}} - Q](t',x') \\
&= \int_{-\infty}^{\infty}\int_0^\tau Q(T > \tau, X < x) d\left[\delta_{\{T_i,X_i\}} - Q\right](t,x) \\
&\quad + \int_{-\infty}^{\infty}\int_\tau^\infty Q\left(T \leqslant \tau, X > x'\right) d\left[\delta_{\{T_i,X_i\}} - Q\right](t',x') \\
&= 1_{\{T_i \leqslant \tau\}} Q\left(T > \tau, X < X_i\right) + 1_{\{T_i > \tau\}} Q\left(T \leqslant \tau, X > X_i\right) - 2\nu_\tau(Q).
\end{aligned}$$

Now we consider the functional $\mu_\tau : \mathcal{Q} \to [0,1]$ corresponding to the denominator of the AUC:

$$\mu_\tau(Q) = \int\int 1_{\{t \leqslant \tau, t' > \tau\}} dQ(t,x) dQ(t',x') = Q(T \leqslant \tau, T' > \tau).$$

An analogous calculation yields that

$$\partial_\varepsilon \mu_\tau(Q_\varepsilon) = 1_{\{T_i \leqslant \tau\}} Q\left(T > \tau\right) + 1_{\{T_i > \tau\}} Q\left(T \leqslant \tau\right) - 2\mu_\tau(Q)$$

The quotient rule now yields that the Gateaux derivative for the functional $\mathrm{AUC}_\tau(Q) = \nu_\tau(Q)/\mu_\tau(Q)$ is given by

$$\begin{aligned}
&IF_{\mathrm{AUC}}(T_i, X_i; \tau) = (\mu_\tau(Q))^{-2} \\
&\quad \left[1_{\{T_i \leqslant \tau\}}(Q\left(T > \tau, X < X_i\right)\mu_\tau(Q) - Q(T > \tau)\nu_\tau(Q)) \right. \\
&\quad \left. + 1_{\{T_i > \tau\}}(Q\left(T \leqslant \tau, X > X_i\right)\mu_\tau(Q) - \nu_\tau(Q) Q(T \leqslant \tau))\right].
\end{aligned}$$

Influence function with censoring and multiple events

In this section, we derive the Influence function for the AUC with more than event and censoring (where the censoring is allowed to depend on the covariates). We now suppose that the censoring distribution depends on the baseline covariates but continue to assume that the censoring time is conditionally independent of the event time and event type given the covariates. To describe the situation with competing risks we introduce a new 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 $\Delta = D 1_{\{T \leqslant C\}}$ taking values in the set $\{0,1,2\}$. We are interested in the following definition of the time-dependent discrimination measure for cause 1:

$$\begin{aligned}
\mathrm{AUC}^1(\tau) &= Q(X_i > X_j \mid T_i \leqslant \tau, D_i = 1, (T_j > \tau \cup D_j = 2)) \\
&= \frac{Q(X_i > X_j, T_i \leqslant \tau, D_i = 1, (T_j > \tau \cup D_j = 2))}{Q(T_i \leqslant \tau, D_i = 1, (T_j > \tau \cup D_j = 2))}
\end{aligned}$$

We define a functional $\nu_\tau^1(Q)$ to represent the numerator:

$$\begin{aligned}
\nu_\tau^1(Q) &= \iint 1_{\{x>x', t\leqslant\tau, d=1, (t'>\tau\cup d'=2)\}} dQ(t,d,x) dQ\left(t',d',x'\right) \\
&= \iint 1_{\{x>x', t\leqslant\tau, (t'>\tau\cup d'=2)\}} dQ(t,1,x) dQ\left(t',d',x'\right) \psi_\tau^1(P) \\
&= \iint 1_{\{x>x', t\leqslant\tau, t'>\tau\}} + 1_{\{x>x', t\leqslant\tau, t'\leqslant\tau, d'=2)\}} dQ(t,1,x) dQ\left(t',d',x'\right) \\
&= \iint 1_{\{x>x', t\leqslant\tau, t'>\tau\}} dQ(t,1,x) dQ\left(t',d',x'\right) \\
&\quad + \iint 1_{\{x>x', t\leqslant\tau, t'\leqslant\tau\}} dQ(t,1,x) dQ\left(t',2,x'\right) \\
&= \sum_{\delta'=0,1,2} \iint 1_{\{x>x', t\leqslant\tau, t'>\tau\}} \frac{dP(t,1,x)}{G(t-|x)} \frac{dP\left(t',\delta',x'\right)}{G\left(\tau|x'\right)} \\
&\quad + \iint 1_{\{x>x', t\leqslant\tau, t'\leqslant\tau\}} \frac{dP(t,1,x)}{G(t-|x)} \frac{dP(t',2,x')}{G(t'-|x')}
\end{aligned}$$

We also need the influence function for the censoring as part of our calculations:

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

We have when Z is discrete,

$$\begin{aligned}
&\partial_\varepsilon \left[\int_0^t \frac{dP_\varepsilon(s,0|z)}{P_\varepsilon(\tilde{T}>s|Z=z)}\right] \\
&= \int_0^t \frac{\partial_\varepsilon dP_\varepsilon(s,0,z) P(\tilde{T}_i>s, Z_i=z)}{P(\tilde{T}>s, Z_i=z)^2} - \frac{dP(s,0,z)\partial_\varepsilon P_\varepsilon(\tilde{T}>s, Z_i=z)}{P(\tilde{T}>s, Z_i=z)^2} \\
&= \int_0^t \frac{(d\mathbb{1}_{\{\tilde{T}_i\leqslant s, \Delta_i=0, Z_i=z\}} - dP(s,0,z))P(\tilde{T}>s, Z_i=z)}{P(\tilde{T}>s, Z=z)^2} \\
&\quad - \frac{dP(s,0,z)(\mathbb{1}_{\{\tilde{T}_i>s, Z_i=z\}} - P(\tilde{T}>s, Z=z))}{P(\tilde{T}>s, Z=z)^2} \\
&= \int_0^t \frac{d\mathbb{1}_{\{\tilde{T}_i\leqslant s, \Delta_i=0, Z_i=z\}} P(\tilde{T}>s, Z=z) - dP(s,0,z)\mathbb{1}_{\{\tilde{T}_i>s, Z_i=z\}}}{P(\tilde{T}>s, Z=z)^2} \\
&= \frac{\mathbb{1}_{\{\tilde{T}_i\leqslant t, \Delta_i=0, Z_i=z\}}}{P(\tilde{T}>\tilde{T}_i, Z=Z_i)} - \int_0^t \frac{\mathbb{1}_{\{\tilde{T}_i>s, Z_i=z\}} dP(s,0,z)}{P(\tilde{T}>s, Z=z)^2} \\
&= \frac{\mathbb{1}_{\{\tilde{T}_i\leqslant t, \Delta_i=0, Z_i=z\}}}{G(\tilde{T}_i|Z_i)S(\tilde{T}_i|Z_i)P(Z=Z_i)} - \int_0^{\tilde{T}_i\wedge t} \frac{\mathbb{1}_{\{Z_i=z\}} dP(s,0,z)}{G(s|Z_i)^2 S(s|Z_i)^2 P(Z=Z_i)^2} \\
&:= f_i(t,z)
\end{aligned}$$

When Z is continuous or mixed, the above has to be modified. Luckily, if we use the trick from before, we can do the calculation again which will mostly be the same, i.e.

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

This just corresponds to setting $\delta_{Z_i}(z) = \frac{1_{\{Z_i=z\}}}{P(Z=Z_i)}$ in the previous calculation.

We have

$$\partial_\varepsilon \left[\frac{dP_\varepsilon\,(t,\delta,x)}{\kappa(P_\varepsilon)\,(s)}\right] = \frac{d(\delta_{\{\tilde{T}_i,\Delta_i,X_i\}})(t,\delta,x) + dP(t,\delta,x)\,[f_i(s,x)-1]}{G(s|x)}$$

So calculations for $\nu_\tau^1(Q)$ can now be done using the above. This yields that the influence function for the numerator can be written as:

$$\begin{aligned}
\mathrm{IF}_\nu(Z_i;\tau) &= \frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=1\}}}{G(\tilde{T}_i-|X_i)}\int 1_{\{X_i>x',t'>\tau\}}\frac{dP(t',x')}{G(\tau|x')} && (1)\\
&+\int 1_{\{t\leqslant\tau\}}\int 1_{\{x>x',t'>\tau\}}\frac{dP(t',x')}{G(\tau|x')}\frac{[f_i(t-,x)-1]\,dP(t,1,x)}{G(t-|x)} && (2)\\
&+\iint 1_{\{x'<x,t\leqslant\tau\}}\frac{dP(t,1,x)}{G(t-|x)}\frac{1_{\{t'>\tau\}}\,[f_i(\tau,x')-1]\,dP(t',x')}{G(\tau|x')} && (3)\\
&+\left(\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=2\}}}{G(\tilde{T}_i-|X_i)}+\frac{1_{\{\tilde{T}_i>\tau\}}}{G(\tau|X_i)}\right)\int 1_{\{X_i<x,t\leqslant\tau\}}\frac{dP(t,1,x)}{G(t-|x)} && (4)\\
&+\iint 1_{\{x'<x,t\leqslant\tau\}}\frac{dP(t,1,x)}{G(t-|x)}1_{\{t'\leqslant\tau\}}(f_i(t'-,x')-1)\frac{dP(t',2,x')}{G(t'-|x')} && (5)\\
&+\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=1\}}}{G(\tilde{T}_i-|X_i)}\int 1_{\{X_i>x',t'\leqslant\tau\}}\frac{dP\,(t',2,x')}{G(t'-|x')} && (6)\\
&+\iint 1_{\{x>x',t'\leqslant\tau\}}\frac{dP(dx')\mathcal{I}_{\{\delta'=2\}}}{G(t'-|x')}1_{\{t\leqslant\tau\}}\frac{[f_i(t-,x)-1]\,dP(t,1,x)}{G(t-|x)} && (7)
\end{aligned}$$

Next, the denominator is

$$\begin{aligned}
\mathrm{IF}_\mu(Z_i;\tau) &= \frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=1\}}}{G(\tilde{T}_i-|X_i)}\int 1_{\{t'>\tau\}}\frac{dP(t',x')}{G(\tau|x')} && (8)\\
&+\int 1_{\{t'>\tau\}}\frac{dP(t',x')}{G(\tau|x')}\int 1_{\{t\leqslant\tau\}}\frac{[f_i(t-,x)-1]\,dP(t,1,x)}{G(t-|x)} && (9)\\
&+\int 1_{\{t\leqslant\tau\}}\frac{dP(t,1,x)}{G(t-|x)}\int\frac{1_{\{t'>\tau\}}\,[f_i(\tau,x')-1]\,dP(t',x')}{G(\tau|x')} && (10)\\
&+\left(\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=2\}}}{G(\tilde{T}_i-|X_i)}+\frac{1_{\{\tilde{T}_i>\tau\}}}{G(\tau|X_i)}\right)\int 1_{\{t\leqslant\tau\}}\frac{dP(t,1,x)}{G(t-|x)} && (11)\\
&+\int 1_{\{t\leqslant\tau\}}\frac{dP(t,1,x)}{G(t-|x)}\int 1_{\{t'\leqslant\tau\}}(f_i(t'-,x')-1)\frac{dP(t',2,x')}{G(t'-|x')} && (12)\\
&+\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=1\}}}{G(\tilde{T}_i-|X_i)}\int 1_{\{t'\leqslant\tau\}}\frac{dP\,(t',2,x')}{G(t'-|x')} && (13)\\
&+\int 1_{\{t\leqslant\tau\}}(f_i(t-,x)-1)\frac{dP(t,1,x)}{G(t-|x)}\int 1_{\{t'\leqslant\tau\}}\frac{dP(t',2,x')}{G(t'-|x')} && (14)
\end{aligned}$$

The quotient rule now yields that the Gateaux derivative for the functional $\mathrm{AUC}_\tau(P) = \nu_\tau(Q)/\mu_\tau(Q)$ is given by

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

Note that the terms involving 1 actually cancel each other out in the final calculation.

Some simpler cases

- Whenever we have survival data instead of competing risk data, the terms corresponding to $\Delta = 2$ can be set to zero, i.e. term 5-7 and 12-14. Also, the first part of 4 and 11 is set to zero.
- Whenever we have censoring that does not depend on the covariates, terns 2-3 simplify a bit and are instead written as:

$$\frac{f_i(\tau)-2}{G(\tau)}\int 1_{\{t\leqslant\tau\}}P(X<x,\tilde{T}>\tau)\frac{1}{G(t-)}dP(t,1,x)$$
$$+\frac{1}{G(\tau)}\int 1_{\{t\leqslant\tau\}}P(X<x,\tilde{T}>\tau)\frac{f_i(t-)}{G(t-)}dP(t,1,x)$$

Similar corrections hold for the denominator. In the code, terms (1-7) are now termed (8-14) and (8-14) are now termed (15-21). Note that $P(X<X_i,\tilde{T}>\tau)=\int 1_{\{t>\tau,x<X_i\}}dP(t,x)$.

Notes on the case with covariates

- Term (2), (5) and (7) (and also (9), (12), and (14)) may be evaluated by noting the following

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

  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

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

  where H is again an arbitrary function.

Estimation

We estimate as follows:

$$\mathrm{IF}_\nu(Z_i;\tau)=\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=1\}}}{\hat{G}(\tilde{T}_i-|X_i)}\int 1_{\{X_i>x',t'>\tau\}}\frac{d\mathbb{P}_n(t',x')}{\hat{G}(\tau|x')} \tag{15}$$
$$+\int 1_{\{t\leqslant\tau\}}\int 1_{\{x>x',t'>\tau\}}\frac{d\mathbb{P}_n(t',x')}{\hat{G}(\tau|x')}\frac{\left[\hat{f}_i(t-,x)-1\right]d\mathbb{P}_n(t,1,x)}{\hat{G}(t-|x)} \tag{16}$$
$$+\iint 1_{\{x'<x,t\leqslant\tau\}}\frac{d\mathbb{P}_n(t,1,x)}{\hat{G}(t-|x)}\frac{1_{\{t'>\tau\}}\left[\hat{f}_i(\tau,x')-1\right]d\mathbb{P}_n(t',x')}{\hat{G}(\tau|x')} \tag{17}$$
$$+\left(\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=2\}}}{\hat{G}(\tilde{T}_i-|X_i)}+\frac{1_{\{\tilde{T}_i>\tau\}}}{\hat{G}(\tau|X_i)}\right)\int 1_{\{X_i<x,t\leqslant\tau\}}\frac{d\mathbb{P}_n(t,1,x)}{\hat{G}(t-|x)} \tag{18}$$
$$+\iint 1_{\{x'<x,t\leqslant\tau\}}\frac{d\mathbb{P}_n(t,1,x)}{\hat{G}(t-|x)}1_{\{t'\leqslant\tau\}}(f_i(t'-,x')-1)\frac{d\mathbb{P}_n(t',2,x')}{\hat{G}(t'-|x')} \tag{19}$$
$$+\frac{1_{\{\tilde{T}_i\leqslant\tau,\Delta_i=1\}}}{\hat{G}(\tilde{T}_i-|X_i)}\int 1_{\{X_i>x',t'\leqslant\tau\}}\frac{d\mathbb{P}_n(t',2,x')}{\hat{G}(t'-|x')} \tag{20}$$
$$+\iint 1_{\{x>x',t'\leqslant\tau\}}\frac{d\mathbb{P}_n(t',2,x')}{\hat{G}(t'-|x')}1_{\{t\leqslant\tau\}}\frac{\left[\hat{f}_i(t-,x)-1\right]d\mathbb{P}_n(t,1,x)}{\hat{G}(t-|x)} \tag{21}$$

Next, the denominator is

$$\mathrm{IF}_\mu(Z_i;\tau) = \frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 1\}}}{\hat{G}(\tilde{T}_i - |X_i)} \int 1_{\{t' > \tau\}} \frac{d\mathbb{P}_n(t', x')}{\hat{G}(\tau|x')} \tag{22}$$

$$+ \int 1_{\{t' > \tau\}} \frac{d\mathbb{P}_n(t', x')}{\hat{G}(\tau|x')} \int 1_{\{t \leqslant \tau\}} \frac{\left[\hat{f}_i(t-, x) - 1\right] d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} \tag{23}$$

$$+ \int 1_{\{t \leqslant \tau\}} \frac{d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} \int \frac{1_{\{t' > \tau\}} \left[\hat{f}_i(\tau, x') - 1\right] d\mathbb{P}_n(t', x')}{\hat{G}(\tau|x')} \tag{24}$$

$$+ \left(\frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 2\}}}{\hat{G}(\tilde{T}_i - |X_i)} + \frac{1_{\{\tilde{T}_i > \tau\}}}{\hat{G}(\tau|X_i)} \right) \int 1_{\{t \leqslant \tau\}} \frac{d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} \tag{25}$$

$$+ \int 1_{\{t \leqslant \tau\}} \frac{d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} \int 1_{\{t' \leqslant \tau\}} (\hat{f}_i(t'-, x') - 1) \frac{d\mathbb{P}_n(t', 2, x')}{\hat{G}(t' - |x')} \tag{26}$$

$$+ \frac{1_{\{\tilde{T}_i \leqslant \tau, \Delta_i = 1\}}}{\hat{G}(\tilde{T}_i - |X_i)} \int 1_{\{t' \leqslant \tau\}} \frac{d\mathbb{P}_n(t', 2, x')}{\hat{G}(t' - |x')} \tag{27}$$

$$+ \int 1_{\{t \leqslant \tau\}} (\hat{f}_i(t-, x) - 1) \frac{d\mathbb{P}_n(t, 1, x)}{\hat{G}(t - |x)} \int 1_{\{t' \leqslant \tau\}} \frac{d\mathbb{P}_n(t', 2, x')}{\hat{G}(t' - |x')} \tag{28}$$

We suggest for estimating this, that the inner part of the integrals are estimated first, because (1) some of the inner integrals appear more than once and (2) for efficiency reasons. Note that, when we have that the censoring is independent of the covariates, we might do some other optimizations as well: We assume that the survival times are sorted and possibly with ties such that $\tilde{T}_1 < \ldots < \tilde{T}_i = \ldots = \tilde{T}_{i+k} < \tilde{T}_{i+k+1} < \ldots < \tilde{T}_n$. We use the following algorithm to preserve memory and the number of iterations for say $\mu^{(i)} = \int 1_{\{t \leqslant \tau\}} \frac{f_i(t-)}{G(t-)} dP(t, 1, x)$. The idea is to split the sum into two terms:

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\hat{f}_i(\tilde{T}_j -) 1_{\{\tilde{T}_j \leqslant \tau, \Delta_j = 1\}}}{\hat{G}(\tilde{T}_j -)} =$$
$$\frac{1}{n} \left(\sum_{j=2}^{i+k} \frac{g(j) 1_{\{\tilde{T}_j \leqslant \tau, \Delta_j = 1\}}}{\hat{G}(\tilde{T}_j -)} + h(i) \sum_{j=i+k+1}^{n} \frac{1_{\{\tilde{T}_j \leqslant \tau, \Delta_j = 1\}}}{\hat{G}(\tilde{T}_j -)} \right)$$

since $\hat{f}_i(\tilde{T}_j -)$ only depends on i for $i + k > j$ and only depends on j for $i + k \leqslant j$, so these values are calculated a priori. Also the first term will always be zero, since we are looking at the value of the integral before any observed event (hence the sum starts at $j = 2$). One can check in the estimation of the Influence Curve for the censoring, which does not depend on the covariates that we need to calculate $2n$ values (i.e. n values for $g(i)$ and n for $h(j)$). This is how we can avoid memory issues. The algorithm is:

$t := 1$
$\hat{\mu}_2 := \sum_{j=1}^{n} \frac{1_{\{\tilde{T}_j \leqslant \tau, \Delta_j = 1\}}}{\hat{G}(\tilde{T}_j -)}$
while $\tilde{T}_1 = \tilde{T}_t$ *and* $t \leqslant n$ **do**
  if $\tilde{T}_t \leqslant \tau$ *and* $\Delta_t = 1$ **then**
    $\hat{\mu}_2 = \hat{\mu}_2 - \frac{1}{G(\tilde{T}_t -)}$
  end
  $t = t + 1$
end
$tieEnd := t - 1$
$\hat{\mu}_1 := 0$
for $i = 1$ *to* n **do**
  $\hat{\mu}^{(i)} = \frac{1}{n} (\hat{\mu}_1 + h(i)\hat{\mu}_2)$
  if $tieEnd \leqslant i$ **then**
    $t = i + 1$
    while $\tilde{T}_1 = \tilde{T}_t$ *and* $t \leqslant n$ **do**
      if $\tilde{T}_t \leqslant \tau$ *and* $\Delta_t = 1$ **then**
        $\hat{\mu}_2 = \hat{\mu}_2 - \frac{1}{G(\tilde{T}_t -)}$
        $\hat{\mu}_1 = \hat{\mu}_1 + \frac{g(t) 1_{\{\tilde{T}_t \leqslant \tau, \Delta_t = 1\}}}{\hat{G}(\tilde{T}_t -)}$
      end
      Let $t = t + 1$
    end
  end
  Let $tieEnd = t - 1$
end
return $\hat{\mu}^{(i)}$ for each $i = 1, \ldots, n$

The idea is that but we keep on adding and subtracting the terms with tied values in the event times. Then we do not need to calculate a sum for each i.

On the other hand, efficient calculation of some of the estimates can be done for example by sorting the risks X in (1) for the integral there.

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 $b = 1, \ldots, B$ we draw a bootstrap dataset $D^*_{m,b} = \{X^*_{b,1}, \ldots, X^*_{b,m}\}$ of size m with replacement from the data D_n. For $i = 1, \ldots, n$ 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

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

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

$$\begin{aligned}\theta_m &= \mathbb{E}_{D_m}\left[\frac{\mathbb{E}_{X_{m+1},X_{m+2}}\Big[\mathcal{I}_{\{R_\tau(D_m)(Z_{m+1})>R_\tau(D_m)(Z_{m+2})\}}\mathcal{I}_{\{Y_{m+1}=1\}}\mathcal{I}_{\{Y_{m+2}=0\}}\Big|D_m\Big]}{\mathbb{E}_{X_{m+1}}\Big[\mathcal{I}_{\{Y_{m+1}=1\}}\Big]\mathbb{E}_{X_{m+2}}\Big[\mathcal{I}_{\{Y_{m+2}=0\}}\Big]}\right]\\ &= \frac{\mathbb{E}_{X_{m+1},X_{m+2}}\big[\Theta_m(Z_{m+1}, Z_{m+2})\mathcal{I}_{\{Y_{m+1}=1\}}\mathcal{I}_{\{Y_{m+2}=0\}}\big]}{\mathbb{E}_{X_{m+1}}\Big[\mathcal{I}_{\{Y_{m+1}=1\}}\Big]\mathbb{E}_{X_{m+2}}\Big[\mathcal{I}_{\{Y_{m+2}=0\}}\Big]} = \frac{\mathcal{A}_{\tau,m}}{\mathcal{B}_\tau}\end{aligned}\tag{29}$$

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

$$\hat{\Theta}^{(1,1)}_m(Z_i, Z_j) = \frac{\sum_{b=1}^B \mathcal{I}_{\{R(D^*_{m,b})(Z_i)>R(D^*_{m,b})(Z_j)\}}\mathcal{I}_{\{N_i^b=0\}}\mathcal{I}_{\{N_j^b=0\}}}{\sum_{b=1}^B \mathcal{I}_{\{N_i^b=0\}}\mathcal{I}_{\{N_j^b=0\}}}\tag{30}$$

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

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

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

Influence function

The influence function of the functional $\nu_m(\mathrm{P}) = \mu_m$ in the direction $x_k = (t_k, \delta_k, z_k)$ is given by

$$\begin{aligned}\mathrm{IF}_\nu(m; x_k) &= \frac{\partial}{\partial\epsilon}\nu_m(\mathrm{P}_{\epsilon,k})\Big|_{\epsilon=0}\\ &= \frac{1}{\eta(\mathrm{P})}\frac{\partial}{\partial\epsilon}\Bigg[\int\int\Bigg(\int\cdots\int\mathcal{I}_{\{R_\tau(\{x_i\}_{i=1}^m)(z_{m+1})>R_\tau(\{x_i\}_{i=1}^m)(z_{m+2})\}}\prod_{i=1}^m \mathrm{P}_{\epsilon,k}(\mathrm{d}x_i)\Bigg)\\ &\qquad \mathcal{I}_{\{y_{m+1}=1,y_{m+2}=0\}}\mathrm{P}_{\epsilon,k}(\mathrm{d}x_{m+1})\mathrm{P}_{\epsilon,k}(\mathrm{d}x_{m+2})\Bigg]\Bigg|_{\epsilon=0} - \frac{\partial}{\partial\epsilon}\frac{\eta(\mathrm{P}_{\epsilon,k})}{\eta(\mathrm{P})}\nu_m(\mathrm{P})\Big|_{\epsilon=0}\end{aligned}$$

It follows that

$$\begin{aligned}
\mathrm{IF}_\nu(m;x_k) &= \frac{1}{\eta(\mathrm{P})}\int \Theta_m(z_k,z_{m+2})\mathcal{I}_{\{y_k=1,y_{m+2}=0\}}\mathrm{P}(\mathrm{d}x_{m+2})-\nu_m(\mathrm{P}) \\
&+\frac{1}{\eta(\mathrm{P})}\int \Theta_m(z_{m+1},z_k)\mathcal{I}_{\{y_{m+1}=1,y_k=0\}}\mathrm{P}(\mathrm{d}x_{m+1})-\nu_m(\mathrm{P}) \\
&+\frac{1}{\eta(\mathrm{P})}\int\int\frac{\partial}{\partial\epsilon}\int\cdots\int \mathcal{I}_{\{R_\tau(\{x_i\}_{i=1}^m)(z_{m+1})>R_\tau(\{x_i\}_{i=1}^m)(z_{m+2})\}}\prod_{i=1}^m \mathrm{P}_{\epsilon,k}(\mathrm{d}x_i) \\
&\mathcal{I}_{\{y_{m+1}=1,y_{m+2}=0\}}\mathrm{P}(\mathrm{d}x_{m+1})\mathrm{P}(\mathrm{d}x_{m+2})\Big|_{\epsilon=0} \\
&-\frac{\partial}{\partial\epsilon}\frac{\nu_m(\mathrm{P})}{\eta(\mathrm{P})}\eta(\mathrm{P}_{\epsilon,k})\Big|_{\epsilon=0}
\end{aligned}$$

Hence, after some calculations,

$$\begin{aligned}
\mathrm{IF}_\nu(m;x_k) = \frac{1}{\eta(\mathrm{P})}\Big[&\int \Theta_m(z_{m+1},z_k)\mathcal{I}_{\{y_k=1,y_{m+2}=0\}}\mathrm{P}(\mathrm{d}x_{m+1}) \\
&+\int \Theta_m(z_k,z_{m+2})\mathcal{I}_{\{y_{m+1}=1,y_k=0\}}\mathrm{P}(\mathrm{d}x_{m+2})\Big]-m\,\nu_m(\mathrm{P}) \\
+\frac{1}{\eta(P)}&\int\int\sum_{j=1}^m\int\cdots\int \mathcal{I}_{\{R_\tau(\{x_i\}_{i=1}^m)(z_{m+1})>R_\tau(\{x_i\}_{i=1}^m)(z_{m+2})\}}\delta_{x_k}(x_j)\prod_{i\neq j}\mathrm{P}(\mathrm{d}x_i) \\
\mathcal{I}_{\{y_{m+1}=1,y_{m+2}=0\}}&\mathrm{P}(\mathrm{d}x_{m+1})\mathrm{P}(\mathrm{d}x_{m+2}) \\
-\frac{\nu_m(\mathrm{P})}{\eta(\mathrm{P})}&\Big[\mathcal{I}_{\{y_k=1\}}\Big]\Big[\int \mathcal{I}_{\{y_{m+2}=0\}}\mathrm{P}(\mathrm{d}x_{m+2})\Big] \\
-\frac{\nu_n(\mathrm{P})}{\eta(\mathrm{P})}&\Big[\int \mathcal{I}_{\{y_{m+1}=1\}}\mathrm{P}(\mathrm{d}x_{m+1})\Big]\Big[\mathcal{I}_{\{y_k=0\}}\Big]
\end{aligned}$$

Survival and Competing risk cases (and more 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 $\Phi_\tau := \mathbb{E}_{D_m}[\phi(D_m)_\tau]$ from the test-train situation as follows by finding the Gateaux derivative for observation i

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

The idea is now to interchange the order of integration (both terms), so that we may use leave-one-out bootstrap estimates. Indeed, for the most general case (competing risk case), we get that

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

where

$$\begin{aligned}
\nu^1_{\tau,m}(P) &= \iint \Theta_m(x,x')1_{\{t\leqslant\tau,t'>\tau\}}\frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)}\frac{P(dx')}{G(\tau|z')} \\
&+\iint \Theta_m(x,x')1_{\{t\leqslant\tau,t'\leqslant\tau\}}\frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)}\frac{P(dx')\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}
\end{aligned}$$

and similarly defined

$$\nu^2_{\tau,m} = \iint \sum_{j=1}^{m} \Theta_m(x,x')^{(j,k)} 1_{\{t\leqslant\tau, t'>\tau\}} \frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)} \frac{P(dx')}{G\left(\tau|z'\right)}$$
$$+ \iint \sum_{j=1}^{m} \Theta_m(x,x')^{(j,k)} 1_{\{t\leqslant\tau, t'\leqslant\tau\}} \frac{P(dx)\mathcal{I}_{\{\delta=1\}}}{G(t-|z)} \frac{P(dx')\mathcal{I}_{\{\delta'=2\}}}{G(t'-|z')}$$

where

$$\Theta_m(x,x')^{(j,k)} = \int \dots \int \mathcal{I}_{\{R(D_m)(z)>R(D_m)(z')\}} \delta_{x_j}(x_k) \prod_{i\neq j} P(dx_i)$$

Note that $\mu_\tau(P)$ is defined precisely as before (and hence also IF_μ is the same), but we must redefine, IF_ν

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

The functional can also be found as $\frac{\nu^1_{\tau,m}(P)}{\mu_\tau(P)}$, and the functional and the terms in its influence function can be estimated in mostly the same way as Brier. Most likely, we don't need the ones here as well.