Validating survival models

Young Researcher Day

Anders Munch

November 25, 2021

Risk prediction model and full data

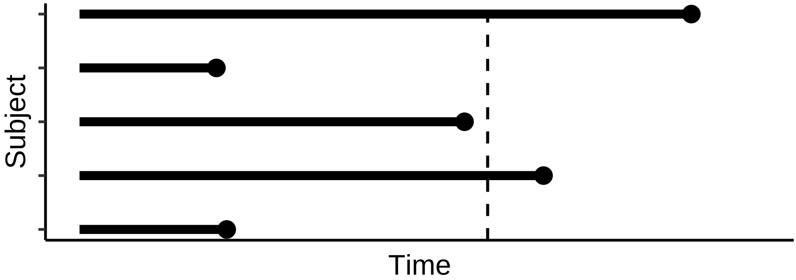
For a fixed time horizon $t \in \mathbb{R}_+$ we have

$X \in \mathbb{R}^p$ Static covariates measured at baseline ($t = 0$)

$T \in \mathbb{R}_+$ Time of event

$r(t \mid X) \in [0, 1]$ Risk prediction at time t given baseline covariates

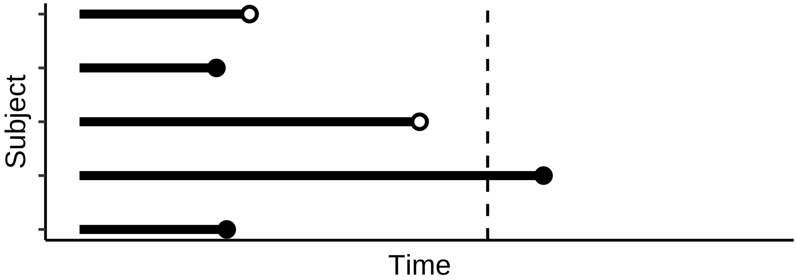
$Y(t) \in \{0, 1\}$ Event status at time t, $Y(t) := \mathbb{1}\{T \leq t\}$



Risk prediction model and censored data

For a fixed time horizon $t \in \mathbb{R}_+$ we have

$X \in \mathbb{R}^p$	Static covariates measured at baseline ($t = 0$)
$\tilde{T} \in \mathbb{R}_+$	Observation time ($\tilde{T} := T \wedge C$)
$\Delta \in \{0, 1\}$	Event indicator ($\Delta := \mathbb{1}\{\tilde{T} = T\}$)
$r(t \mid X) \in [0, 1]$	Risk prediction at time t given baseline covariates
$Y(t) \in \{0, 1\}$	Is unobserved for som subjects



The Brier score

Given a (non-random) risk-prediction model $r \colon \mathbb{R}_+ \times \mathbb{R}^p \to [0,1]$ we want to evaluate the performance of r at a fixed time horizon $t \in \mathbb{R}_+$. For this we use the Brier score

$$\mathbb{E}\left[\{Y(t) - r(t \mid X)\}^2\right], \quad \text{with} \quad Y(t) := \mathbb{1}\{T \leq t\}.$$

Inverse probability of censoring weights (IPCW)

Let $(X, T) \sim Q$ and $(X, \tilde{T}, \Delta) \sim P$. When $T \perp\!\!\!\perp C \mid X$ the Brier score is identifiable from the observed data[1]:

$$\mathbb{E}_Q\left[\{Y(t) - r(t \mid X)\}^2\right] = \mathbb{E}_P\left[W(t)\{Y(t) - r(t \mid X)\}^2\right],$$

with

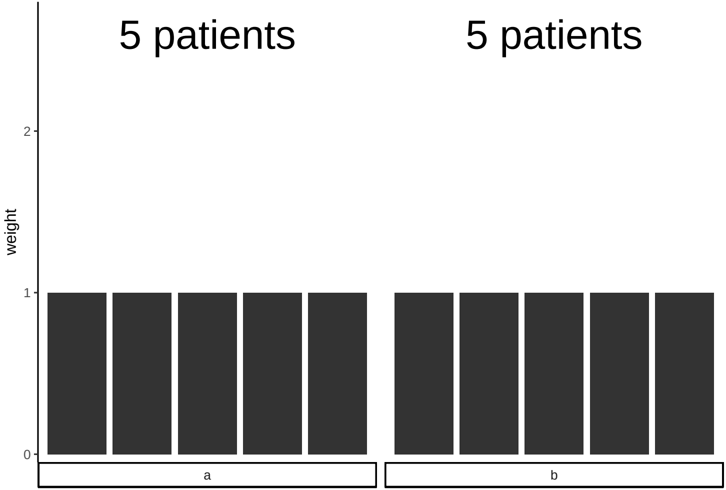
$$W(t) = \frac{\mathbb{1}(\tilde{T} > t)}{G(t \mid X)} + \frac{\mathbb{1}(\tilde{T} \leq t)\Delta}{G(\tilde{T} \mid X)},$$

and where $G(s \mid x) = P(C > s \mid X = x)$.

[1] Note that $W(t)\{Y(t) - r(t \mid X)\}^2$ is a function of the observed data, as $Y(t)$ is observed whenever $W(t)$ is non-zero.

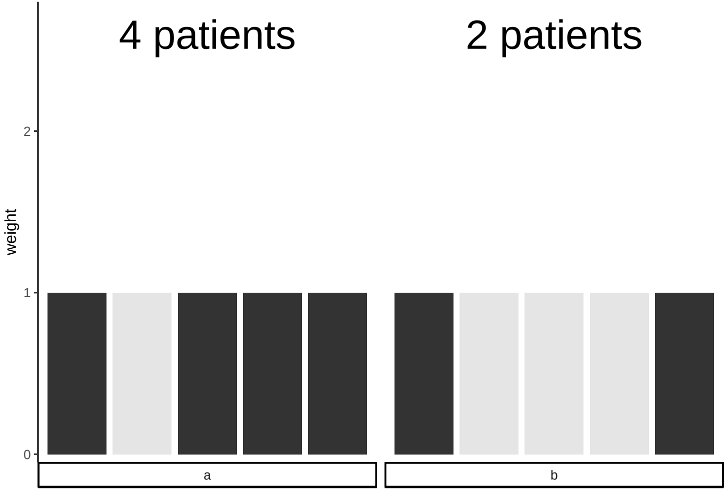
Visualizing the re-weighting

$$\mathbb{E}_Q\left[\{Y(t) - r(t \mid X)\}^2\right] = \mathbb{E}_P\left[W(t)\{Y(t) - r(t \mid X)\}^2\right]$$



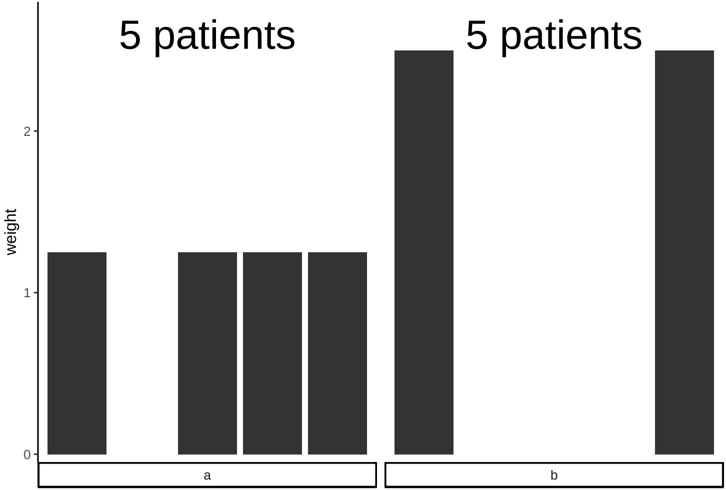
Visualizing the re-weighting

$$\mathbb{E}_Q\left[\{Y(t) - r(t \mid X)\}^2\right] = \mathbb{E}_P\left[W(t)\{Y(t) - r(t \mid X)\}^2\right]$$



Visualizing the re-weighting

$$\mathbb{E}_Q\left[\{Y(t) - r(t \mid X)\}^2\right] = \mathbb{E}_P\left[W(t)\{Y(t) - r(t \mid X)\}^2\right]$$



Non-parametric estimation

By estimating the censoring distribution G we obtain the IPCW estimator

$$\widehat{W}_i(t) = \frac{\mathbb{1}(\tilde{T}_i > t)}{\hat{G}(t \mid X_i)} + \frac{\mathbb{1}(\tilde{T}_i \leq t)\Delta_i}{\hat{G}(\tilde{T}_i \mid X_i)}, \quad \hat{\theta}_n^t = \hat{\mathbb{P}}_n[\widehat{W}_i(t)\{Y_i(t) - r(t \mid X_i)\}^2]$$

Inference and efficiency under non-parametric assumptions

- Existing methods use Kaplan-Meier or Cox models to estimate the censoring distribution.
- Non-parametric, data-adaptive estimation of the censoring distribution + inference → one-step estimators / DML / TMLE.

Data-adaptive selection of the estimator $\hat{G}$

Use some kind of cross-validation to select the best $\hat{G}$ from a collection of candidates.

Non-parametric estimation

By estimating the censoring distribution G we obtain the IPCW estimator

$$\widehat{W}_i(t) = \frac{\mathbb{1}(\tilde{T}_i > t)}{\hat{G}(t \mid X_i)} + \frac{\mathbb{1}(\tilde{T}_i \leq t)\Delta_i}{\hat{G}(\tilde{T}_i \mid X_i)}, \quad \hat{\theta}_n^t = \hat{\mathbb{P}}_n[\widehat{W}_i(t)\{Y_i(t) - r(t \mid X_i)\}^2]$$

Inference and efficiency under non-parametric assumptions

- Existing methods use Kaplan-Meier or Cox models to estimate the censoring distribution.
- Non-parametric, data-adaptive estimation of the censoring distribution + inference $\rightarrow$ one-step estimators / DML / TMLE.

Data-adaptive selection of the estimator $\hat{G}$

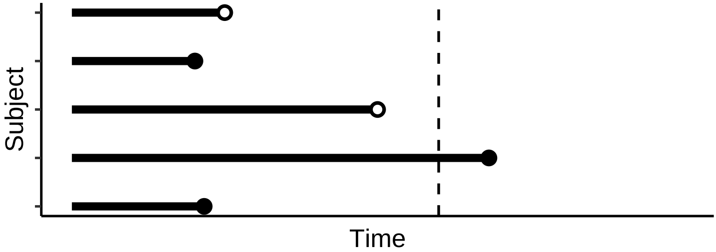
Use some kind of cross-validation to select the best $\hat{G}$ from a collection of candidates. Is it obvious how to do this?

Cross validation for survival models – the Brier score?

Brier score!

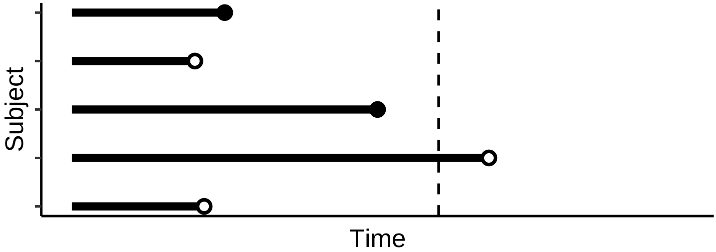
Cross validation for survival models – the Brier score?

Brier score!



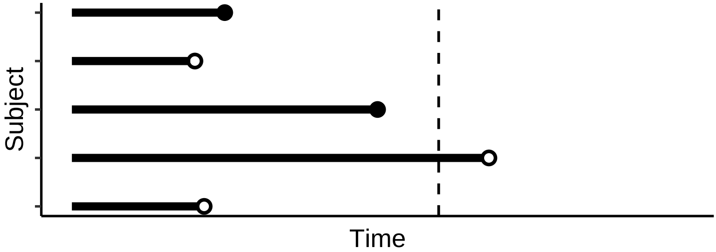
Cross validation for survival models – the Brier score?

Brier score!



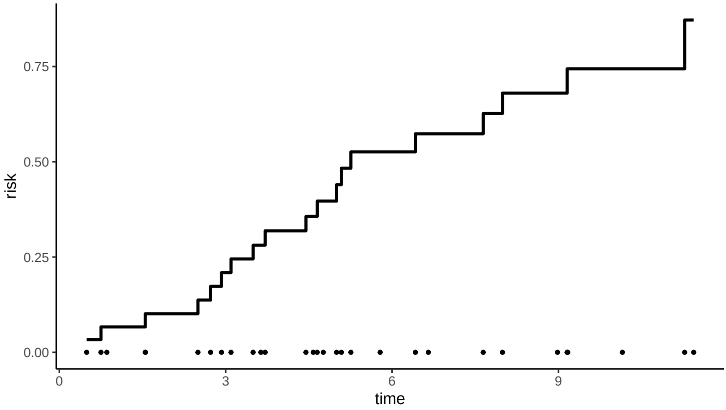
Cross validation for survival models – the Brier score?

Brier score! ... infinite regress ...



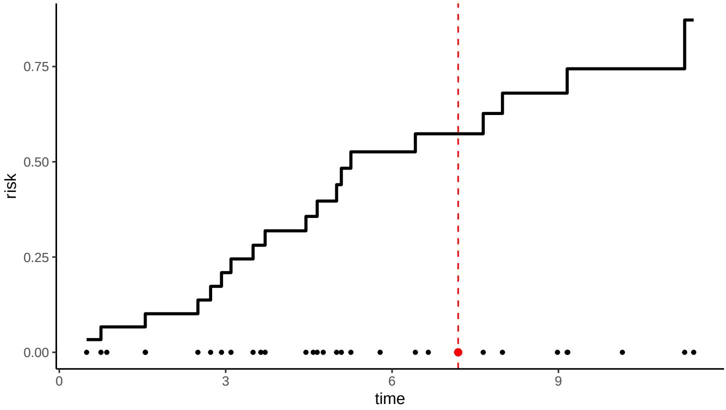
Cross validation for survival models – the likelihood?

Many common estimators of a survival function will be a discrete measure $\rightarrow$ the likelihood will (a.s.) be 0 on any hold-out data sample.



Cross validation for survival models – the likelihood?

Many common estimators of a survival function will be a discrete measure $\rightarrow$ the likelihood will (a.s.) be 0 on any hold-out data sample.



Summary

How should we do cross validation for general survival models when the test and train data are censored?

- Which other loss functions are sensible to use instead?
- How can these (and the risk of an estimator) be approximated with observed data?

Questions, comments, suggestions?

Thank you for listening!