

Julien DUPIN <julien.dupin@roche.com>

TR: Journées de Biostatistique: soumission de résumé

Julien Dupin <dupin.julien@hotmail.fr>

30 septembre 2022 à 11:17

À : Ahmadou Alioum <amadou.alioum@u-bordeaux.fr>, "bertrand.frederic@gmail.com" <bertrand.frederic@gmail.com>, David Magrez <d_magrez@yahoo.fr>, Hélène SAVEL <helene.savel@ipsen.com>, Louise BASCHET <louise.baschet@horiana.com>, Nicolas Savy <nicolas.savy@math.univ-toulouse.fr>, Pascale Tubert <pascale.tubert@inserm.fr>, Philippe Saint Pierre <Philippe.Saint-Pierre@math.univ-toulouse.fr>, "sandrine.guillemot@servier.com" <sandrine.guillemot@servier.com>, "julien.dupin@roche.com" <julien.dupin@roche.com>

De : Ariane Cwiling <ariane-cwiling@orange.fr>**Envoyé :** vendredi 30 septembre 2022 11:15**À :** biopharma@sdfs.asso.fr <biopharma@sdfs.asso.fr>**Objet :** Journées de Biostatistique: soumission de résumé

Bonjour,

Je suis doctorante au MAP5 et je souhaiterais soumettre mon résumé pour les journées de Biostatistique.

Machine learning for survival data prediction: Prediction of the restricted mean survival time

Because of its simple interpretation, the restricted mean survival time (RMST) is an interesting quantity in survival analysis. Its prediction with regard to the attributes of a patient can be especially relevant in health care. Yet few survival methods exist for this purpose.

We present a new prediction method based on pseudo-observations, which transform incomplete observed times into outcomes that can be handled as uncensored (Andersen and Perme, 2010). It is then possible to apply any standard prediction method to link them with covariates, like generalized linear models (GLM) or deep neural networks (Andersen et al., 2003, Zhao, 2021). We combine pseudo-observations with the super learner, an ensemble algorithm that fits an optimal weighted combination of candidate learners (Van der Laan et al., 2007). Note that assessing the performance of this kind of model on real censored data requires specific error measures, e.g. the C-index (Heagerty and Zheng, 2005). We suggest instead a new estimator of the mean squared error (MSE) that we adapted from the Brier score estimator of Gerds and Schumacher (2006). We also developed an algorithm derived from conformal inference (Lei et al., 2018) to construct prediction intervals for right-censored data.

We showed that fitting a GLM on pseudo-observations or on true event times is asymptotically equivalent in terms of MSE. Simulation results give us good confidence in the extension of this result to the super learner. Moreover we proved that our estimator of the MSE and our conformal prediction intervals are asymptotically valid. These results are illustrated through extensive simulations and applications on real data.

Auteurs:

1- Ariane Cwiling (ariane.cwiling@u-paris.fr)2- Olivier Bouaziz (olivier.bouaziz@parisdescartes.fr)3- Vittorio Perduca (vittorio.perduca@parisdescartes.fr)

1,2,3 - Université Paris Cité, MAP5, UMR 8145, France

Bien à vous,

Ariane Cwiling