Adaptive kernel estimation of the baseline function in the Cox model with high-dimensional covariates

Sarah LEMLER, École CentraleSupélec

Mots-clés : Cox model, Kernel estimator, Goldenschluger and Lepski method, model selection, high-dimension, non-asymptotic oracle inequality

Recurrent event data arise in such fields as medicine, insurance, economics, and reliability. Such events include for example relapse from a disease in biomedical research, monetization in marketing or blogging in social network study. In this context, proportional hazards models have been largely studied in the literature to model the rate functions of recurrent event data, that represents the instantaneous probability of experiencing a recurrent event at a given time. In this work, we propose a novel kernel estimator of the baseline function in a general high-dimensional Cox model, for which we derive non-asymptotic rates of convergence.

To construct our estimator, we first estimate the regression parameter in the Cox model via a Lasso procedure. We then plug this estimator into the classical kernel estimator of the baseline function, obtained by smoothing the so-called Breslow estimator of the cumulative baseline function. We propose and study an adaptive procedure for selecting the bandwidth, in the spirit of Goldenshluger and Lepski (2011). We state non-asymptotic oracle inequalities for the final estimator, which reveal the reduction of the rates of convergence when the dimension of the covariates grows.

Lastly, we conduct a study to measure the practical performances of the resulting estimator on simulated data and we apply the implemented procedure to a real dataset.

Références

[1] O. Bouaziz, F. Comte, A. Guilloux, *Nonparametric estimation of the intensity function of a recurrent event process*, Statist. Sinica 23 (2) (2013) 635-665.

[2] Guilloux, A., Lemler, S., Taupin, M. L. (2016). *Adaptive kernel estimation of the baseline function in the Cox model with high-dimensional covariates.* Journal of Multivariate Analysis, 148, 141-159.

[3] A. Goldenshluger, O. Lepski, *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality*, Ann. Statist. 39 (3) (2011) 1608-1632.

[4] H. Ramlau-Hansen, *The choice of a kernel function in the graduation of counting process intensities*, Scand. Actuar. J. 3 (1983) 165-182.

Sarah LEMLER, École CentraleSupélec, 9 rue Joliot Curie, 91192 Gif-sur-Yvette