Quantile regression for censored data
Application to a meta-analysis of ovarian cancer treatments
Master 2 internship in Statistics

Olivier Bouaziz
email: olivier.bouaziz@parisdescartes.fr

Aurélien Latouche
email: aurelien.latouche@curie.fr

Context

In oncology, the effect of a treatment may vary in certain subgroups of patients, including during a randomized clinical trial. Classical models usually attempt to estimate the effect of the treatment on an average response - in other words, they model the expectation of the response conditionally on the explanatory factors. An alternative to such approaches is to use quantile regression which allows the effect of variables to be different from one quantile to another. This is also particularly attractive in practice as on the one hand, quantiles are more flexible and robust quantitative tools for characterizing event times than mean-based models and on the other hand it might allow to detect late treatment effects, a phenomenon that often occurs in oncology.

In clinical studies another problem arises from the right-censoring mechanism due to dropout (some patients might leave the study before its end) or end of follow-up (some patients will never develop the outcome during the follow-up). Taking into account right-censoring is a statistical challenge and the analysis of such data needs dedicated tools. Classical methods include the Kaplan-Meier estimator to estimate the survival function in a non-parametric way and the Cox model to model the hazard function given a set covariates. See for instance Kalbfleisch & Prentice (2002) for a thorough presentation of the subject. As an alternative to those methods, quantile regression in the context of right-censored data have also been developed. See for instance Peng & Huang (2008), Peng (2021) for a review of such methods.

The aim of the internship is to apply censored quantile regression methods on ovarian cancer data from several clinical studies in order to assess the effect of age on the treatment response. In order to take into account heterogeneity among clinical trials, a recent proposal from Ahn & Logan (2016) expands the idea of quantile regression to right-censored clustered data.

During his/her internship, the student will review statistical methods for meta-analysis studies and quantile regression for right-censored data. The meta-analysis data contain $11,029$ patients from 17 clinical trials and were collected in a previous study (see Paoletti et al. (2020)). Statistical analyses will be implemented using the R software. Moreover, the work performed during the internship is aimed to be continued during a PhD thesis. In particular, the method from Ahn & Logan (2016) is based on the pseudo-regression approach (see Andersen et al. (2003), Andersen & Pohar Perme (2010)), a technique that will be studied during the PhD and compared to other possible approaches.

Required profile

We are seeking candidates with a strong mathematical background and a keen interest in medical applications. Skills in the R software are also required.

Practical details

The internship will take place within the 2 partner teams of MAP5 (O. Bouaziz, Université de Paris) and the Statistics team for precision medicine of Inserm U900 unit (A. Latouche, Cnam and Institut Curie) in collaboration with the principal investigator of the meta-analysis (X. Paoletti, Institut Curie).

References

AHN, K. W. & LOGAN, B. R. (2016). Pseudo-value approach for conditional quantile residual lifetime analysis for clustered survival and competing risks data with applications to bone marrow transplant data. *The annals of applied statistics* **10**, 618.

ANDERSEN, P. K., KLEIN, J. P. & ROSTHØJ, S. (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika* **90**, 15–27.

ANDERSEN, P. K. & POHAR PERME, M. (2010). Pseudo-observations in survival analysis. *Statistical methods in medical research* **19**, 71–99.

KALBFLEISCH, J. D. & PRENTICE, R. L. (2002). *The statistical analysis of failure time data*. Wiley Series in Probability and Statistics. Wiley-Interscience (John Wiley & Sons), Hoboken, NJ, 2nd ed.

PAOLETTI, X., LEWSLEY, L.-A., DANIELE, G., COOK, A., YANAIHARA, N., TINKER, A., KRISTENSEN, G., OTTEVANGER, P. B., ARAVANTINOS, G., MILLER, A. et al. (2020). Assessment of progression-free survival as a surrogate end point of overall survival in first-line treatment of ovarian cancer: a systematic review and meta-analysis. *JAMA network open* **3**, e1918939–e1918939.

PENG, L. (2021). Quantile regression for survival data. *Annual review of statistics and its application* **8**, 413–437.

PENG, L. & HUANG, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103**, 637–649.