



Large sample results for frequentist multiple imputation for Cox regression with missing covariate data

Frank Eriksson¹ · Torben Martinussen¹ · Søren Feodor Nielsen²

Received: 31 July 2018 / Revised: 7 March 2019 / Published online: 4 April 2019
© The Institute of Statistical Mathematics, Tokyo 2019

Abstract

Incomplete information on explanatory variables is commonly encountered in studies of possibly censored event times. A popular approach to deal with partially observed covariates is multiple imputation, where a number of completed data sets, that can be analyzed by standard complete data methods, are obtained by imputing missing values from an appropriate distribution. We show how the combination of multiple imputations from a compatible model with suitably estimated parameters and the usual Cox regression estimators leads to consistent and asymptotically Gaussian estimators of both the finite-dimensional regression parameter and the infinite-dimensional cumulative baseline hazard parameter. We also derive a consistent estimator of the covariance operator. Simulation studies and an application to a study on survival after treatment for liver cirrhosis show that the estimators perform well with moderate sample sizes and indicate that iterating the multiple-imputation estimator increases the precision.

Keywords Asymptotic distribution · Coarsened data · Semiparametric · Survival · Variance estimator

✉ Frank Eriksson
eriksson@sund.ku.dk

¹ Section of Biostatistics, Department of Public Health, University of Copenhagen, 1014 Copenhagen, Denmark

² Center for Statistics, Department of Finance, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark