October 2012

Inference with Missing data

Christiana Drake

Department of Statistics
University of California, Davis
One Shields Avenue
Davis, CA 95616, USA

Observational studies and to a lesser degree randomized experiments often are confronted with missing data. The validity of inference depends on the missingness mechanism [Little J.A, and Rubin, D.B., 2002]. When the missing data mechanism depends on observed data only, estimation of means and/or regression coefficients requires adjustment but no further information. If the missingness mechanism depends on unobserved data, unbiased estimation requires further information. Multiple imputation [Rubin, D.B. 1987] complete data sets by imputing the missing values and leads to standard statistical analysis. Likelihood based inference distinguishes between data missing at random (MAR) where inference is based on the observed data likelihood and data not missing at random (NMAR) where the joint distribution of the data and missingness mechanism is modelled. Estimation in this case cannot be based on the observed data alone. The concepts will be illustrated with examples from survey sampling and time series analysis. The time series application is joint work with Oskar Knapik and Jacek Leśkow.