

Sensitivity analysis for high-throughput biological data subject to MNAR

Chiu-Hsieh Hsu, Yulei He, Chengcheng Hu, Wei Zhou

Abstract

Missingness mechanism is in theory unverifiable based only on the observed data. If there is a suspicion of missing not at random (MNAR), researchers often perform a sensitivity analysis to evaluate the impact of various missingness mechanisms. In general, sensitivity analysis approaches require a full specification of the relationship between missing values and missingness probabilities. Such relationship can be specified based on a selection model, a pattern-mixture model or a shared parameter model. However, the relationship cannot be specified without some specific parametric assumption. Also, for high-throughput data often more than one variable are subject to missingness, which makes it even more challenging to specify the relationship. Under the selection modeling framework, we propose a sensitivity analysis approach using a nonparametric multiple imputation strategy. The proposed approach does not need to fully specify the relationship and only requires specifying the correlation coefficient between missing values and selection (response) probabilities under a selection model. The correlation coefficient is a standardized measure and can be used as a natural sensitivity analysis parameter. The sensitivity analysis involves multiple imputations of missing values, yet the sensitivity parameter is only used to select imputing/donor sets. Hence, the proposed approach might be more robust against misspecifications of the sensitivity parameter. For illustration, the proposed approach is applied to incomplete measurements of level of pre-operative Hemoglobin A1c, for patients who had high grade carotid artery stenosis and were scheduled for surgery. A simulation study is conducted to evaluate the performance of the proposed approach. Numerical analysis results indicate the proposed approach can be easily applied to perform sensitivity analysis for high-throughput biological data subject to MNAR.

Keywords: Missing data, Sensitivity analysis, Missing not at random, Multiple imputation, Selection model