Semi-parametric Bayesian analysis of binary responses with a continuous covariate subject to non-random missingness

Frederico Z. Poleto
IME, University of São Paulo, Brazil frederico@poleto.com

Carlos Daniel Paulino*
IST, Technical University of Lisbon, Portugal dpaulino@math.ist.utl.pt

Julio M. Singer
IME, University of São Paulo, Brazil jmsinger@ime.usp.br

Geert Molenberghs
Hasselt University, Belgium geert.molenberghs@uhasselt.be

Missingness in explanatory variables requires a model for the covariates even if the interest lies only in a conditional model for the outcomes given the covariates. An incorrect specification of the models for the covariates or for the missingness mechanism may lead to biased inferences for the parameters of interest. Previously published articles either use semi-/non-parametric flexible distributions for the covariates and identify the model via a missing at random assumption, or employ parametric distributions for the covariates and allow a more general non-random missingness mechanism. We consider the analysis of binary responses, combining a MNAR mechanism with a non-parametric model based on a Dirichlet process mixture for the continuous covariates. We illustrate the proposal with simulations and by analyzing a real dataset.