Bayesian Inference from Non-Ignorable Network Sampling Designs

Simon Lunagomez*

Harvard University, Cambridge, USA simon.lgz@gmail.com

Edoardo M. Airoldi

Harvard University, Cambridge, USA airoldi@stat.harvard.edu

Consider a population where subjects are susceptible to a disease (e.g. AIDS). The objective is to perform inferences on a population quantity (like the incidence of HIV on a high-risk subpopulation, e.g. intra-venous drug abusers) via sampling mechanisms based on a social network (link-tracing designs, respondent-driven sampling). We phrase this problem in terms of the framework proposed by Rubin for making inferences on a population quantity and, within this context, prove that respondent-driven sampling is non-ignorable. By non-ignorable it is meant that the uncertainty of the sampling mechanism must be modeled in the likelihood in order to get valid inferences. We develop a general framework for making Bayesian inference on the population quantity that: models the uncertainty in the underlying social network, incorporates dependence among the individual responses according to the network, and deals with the non-ignorability of the sampling design. The proposed framework is general in the sense that it allows a wide range of different specifications for the components of the model we just mentioned. Our model is compared with state of the art methods in simulation studies and it is applied to real data.

Key Words: Social networks, Bayesian inference, hard to reach population, random graph