Respondent-driven sampling and random walks on directed networks

Jens Malmros*,
Dept. of mathematics, Stockholm university,
Stockholm, Sweden, jensm@math.su.se

Tom Britton,
Dept. of mathematics, Stockholm university,
Stockholm, Sweden

Respondent-Driven Sampling (RDS) is a commonly used substitute for random sampling, and has proven to be especially useful when studying hidden populations, such as injective drug users or men who have sex with men, for which no sampling frame is known. By utilising the underlying structure of the social network of the population, and approximating the actual RDS process with a random walk, the RDS methodology yields unbiased population estimates using an unequal-probability sampling framework. However, the fact that social networks generally can be considered directed, i.e., edges in the network may be irreciprocal, is neglected in current RDS estimation. When sampling from directed networks using random walks, there are no closed expressions for the sampling probabilities, i.e., the probabilities that individuals are included in the sample, as for the undirected case. As well-estimated sampling probabilities are a key component in RDS estimation, assuming a directed social network thus provides an additional complexity to the estimation procedure. In this work, we study the sampling probabilities from random walks on directed networks using both simulations as well as analytical derivations, and the implications of directedness for RDS estimation. We suggest how to handle RDS estimation on directed networks both for the theoretical case where the full degree is observed as well as for the (practical) situation when only outdegree is observed.

Key Words: Link-tracing sampling, sampling probabilities, irreciprocal relationships