Respondent driven sampling and random walks on directed networks

Jens Malmros[1], Naoki Masuda[2], and Tom Britton[1]

[1] Department of Mathematics, Stockholm University, Sweden
[2] Department of Mathematical Informatics, University of Tokyo, Japan

Respondent driven sampling (RDS) is often used to estimate population properties (e.g. sexual risk behavior) in hard-to-reach populations. It combines a modified snowball sampling methodology with an estimation procedure that yields unbiased population estimates under the assumption that the sampling process behaves like a random walk on the social network of the population. Current RDS estimation methodology assumes that the social network is undirected, i.e. all edges are reciprocal. However, empirical social networks in general also have non-reciprocated edges. We develop an estimation method for RDS on the basis of random walks on directed networks, where we derive estimators of the selection probabilities of individuals as a function of the number of outgoing edges of sampled individuals. The proposed estimators are evaluated on artificial and empirical networks and are shown to generally perform better than existing estimators. This is in particular the case when the fraction of directed edges in the network is large.