

Inferens på sociala nätverk i "gömda" populationer

Anni Pilbacka*

April 2010

Sammanfattning

Populationer där man inte känner till sannolikheten för att välja en viss individ till stickprovet, och man därmed inte kan tillämpa vanlig stickprovs- och skattningsteknik, kallas för gömda populationer. I den här uppsatsen ska vi simulera ett socialt nätverk av individer och göra ett stickprov på nätverket som vi ska tillämpa en relativt ny skattningsteknik på - Respondent Driven Sampling (RDS). Metoden går ut på att individerna, som ingår i den population man är intresserad av att studera, väljs genom vänskapsbanden mellan de redan existerande medlemmarna i stickprovet. Vi skattar populationsandelar med RDS för ett antal olika modeller. Syftet är att se om skattningarna närmar sig den sanna andelen i populationen samt att se hur skattningens precision förbättras i takt med att urvalet växer. Vi kommer att visa att RDS fungerar bra när vänskapsbanden är ömsesidiga samt när endast enstaka vänskapsband som inte är ömsesidiga finns med. Vi kommer fram till att RDS fungerar mindre bra när merparten av vänskapsbanden inte är ömsesidiga.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: anni_pilbacka@hotmail.com. Handledare: Tom Britton.

Inference on social networks in hidden populations

Abstract

Populations where the probability of choosing a certain individual to a sample is unknown and where you thus cannot apply regular sampling and estimation technique are called hidden populations. In this paper we will simulate a social network of individuals and make a sample of the network that we will apply a relatively new estimation methodology on – Respondent-Driven Sampling (RDS). The method is based on that the individuals in the population that we are interested in studying recruit their friends to be included in the study. We will estimate the population proportions with RDS for a number of different models. The aim is to see if the estimates are approaching the true proportion of the population, and to see how the precision is improved as the sample is growing. We will show that RDS works well when the friendships are reciprocal and in the case when there are only a few friendships that are not reciprocal. We conclude that RDS works less good when the majority of friendships are not reciprocal.