Respondent driven-sampling

Procedure to sample from hidden or hard-to-reach populations

Lucas Moschen

School of Applied Mathematics
Fundação Getulio Vargas

May 23, 2021

FGV EMAp
ESCOLA DE MATEMÁTICA APLICADA

Table of Contents

1. Introduction
2. Mathematical formulation
3. Examples and usages
4. Applications with real data

Table of Contents

1. Introduction
2. Mathematical formulation
3. Examples and usages
4. Applications with real data

Hidden and hard-to-reach populations

- No sampling frame exists: size and boundaries of the population are unknown.
- Privacy concerns: stigmatized or illegal behavior.
- Fear of exposition or prosecution complicates the enumeration and learning about these populations.
- High logistic cost when the occurrence frequency is low.
- Examples: Heavy drug users, sex workers, homeless people, and men who have sex with men.

Existing sampling methods

- Snowball (Goodman (1961))

From starting individuals, each subject provides a list of names of known individuals from the target population. The researcher invites this person to participate, who can agree or deny it.

- Key informant (Deaux and Callaghan (1985))

Expert respondents are selected to answer about others' behavior. For instance, social workers, drug abuse counselors, official, etc.

- Targeted (Watters and Biernacki (1989))

Field researchers build an ethnographic mapping of a target population, and recruit a number of individuals at sites identified by this map.

Problems with snowball sampling

- Inferences about the individuals depend on the initial sample.
- Bias towards individuals who are more cooperative and agree to participate.
- Bias because of masking, that is, protecting friends by not referring them.
- Individuals with more links may be oversampled.

Respondent-driven sampling

Table of Contents

1. Introduction
2. Mathematical formulation
3. Examples and usages
4. Applications with real data

Markov chain

Network model

Table of Contents

1. Introduction
2. Mathematical formulation
3. Examples and usages
4. Applications with real data

Table of Contents

1. Introduction
2. Mathematical formulation
3. Examples and usages
4. Applications with real data

References I

Deaux, E. and Callaghan, J. W. (1985). Key informant versus self-report estimates of health-risk behavior. *Evaluation Review*, 9(3):365–368.

Goodman, L. A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics*, 32(1):148–170.

Watters, J. K. and Biernacki, P. (1989). Targeted sampling: Options for the study of hidden populations. *Social Problems*, 36(4):416–430.