Class 13: Respondent-driven sampling

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University

3/3 New approaches to respondent-driven sampling

In the previous video we saw that even if assumptions are met, respondent-driven sampling seems to have high sample-to-sample variability in real social networks. Could we improve respondent-driven sampling to address these problems?
3 approaches:

- change offspring distribution
- change data collection to avoid bottlenecks
- diagnostics to detect problems

How can we improve RDS?

Change offspring distribution by changing from 3 to 2 coupons. Switching from:

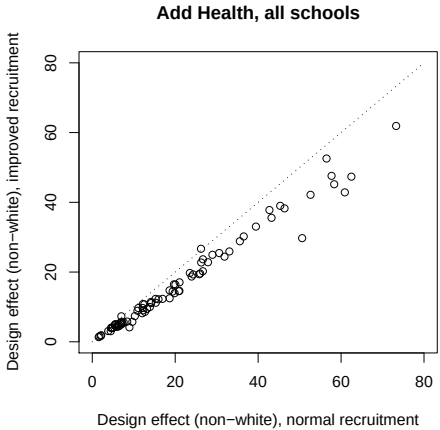
	Number of recruits			
	0	1	2	3
Prob.	1/3	1/6	1/6	1/3

to

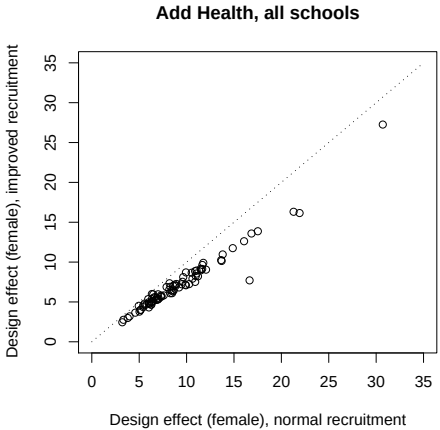
	Number of recruits		
	0	1	2
Prob.	1/4	1/4	1/2

reduces design effects for prop. female and prop. non-white by about 20% in Add Health data.

Reducing multiple recruitment improves estimates



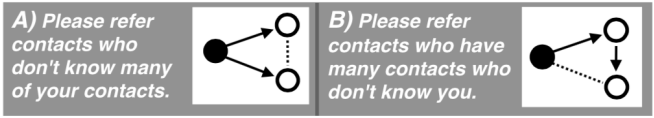
(a) Prop. non-White

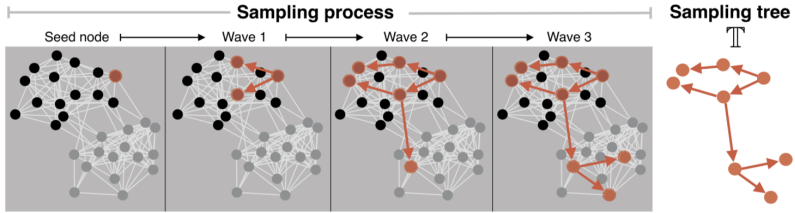


(b) Prop. Female

New approaches to sample that avoid bottlenecks.

Anti-cluster referral requests




Sampling process
Seed node
Wave 1
Wave 2
Wave 3
Sampling tree
$\mathbb{T}$

Electronic Journal of Statistics
Vol. 11 (2017) 4769–4812
ISSN: 1935-7524
DOI: 10.1214/17-EJS1358

Novel sampling design for respondent-driven sampling

Mohammad Khabbazian*

Department of Electrical Engineering
University of Wisconsin-Madison
1415 Engineering Dr, Madison, WI 53706
e-mail: khabbazian@wisc.edu

Bret Hanlon

Department of Surgery
University of Wisconsin School of Medicine and Public Health
600 Highland Avenue Madison, WI 53792
e-mail: hanlon@surgery.wisc.edu

and

Zoe Russek, Karl Rohe

Department of Statistics
University of Wisconsin-Madison
1300 University Ave, Madison, WI 53706
e-mail: zrussek@wisc.edu; karlrohe@stat.wisc.edu

https://doi.org/10.1214/17-EJS1358

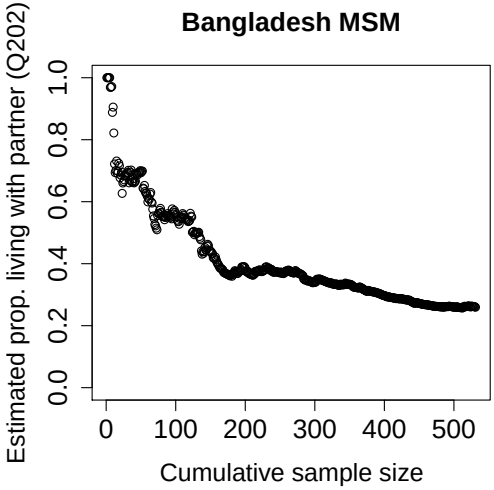
How can we improve RDS? Diagnostics

Diagnostics that can be run as the sample progress. In this case, we might actually be able to correct problems during data collection, but this is not without difficulty.

Examples from a study of men who have sex with men in Dhaka, Bangladesh (Johnston et al. 2007). 7 seeds grew to a sample size of 531 in 7 weeks. Thanks to Lisa Johnston for providing the data.

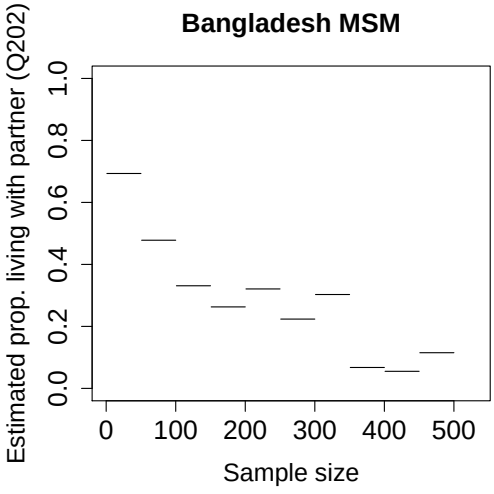
How can we improve RDS? Diagnostics

Plot the cumulative estimate over time. In this case, the estimate doesn't really stabilize.



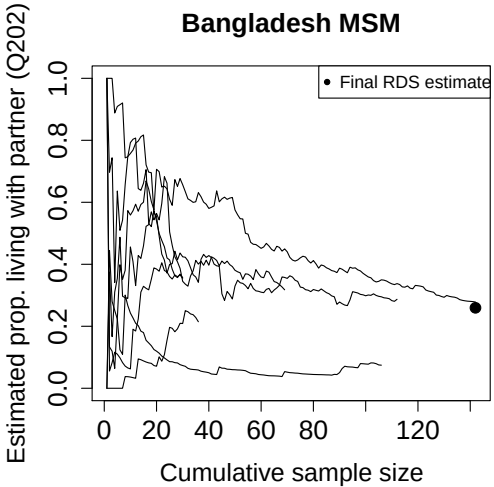
How can we improve RDS? Diagnostics

Batched estimates look even worse.



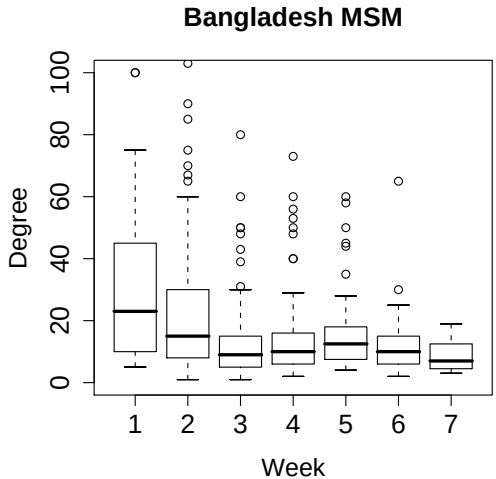
How can we improve RDS? Diagnostics

Plot the estimates for each seed separately; inspired by (but different from) $\hat{R}$ method of Gelman and Rubin (1992).



How can we improve RDS?

Look for burn-out as sample progresses. The degree of respondents decreases over time as the sample progresses.



J. R. Statist. Soc. A (2015)
178, *Part* 1, *pp.* 241–269

Diagnostics for respondent-driven sampling

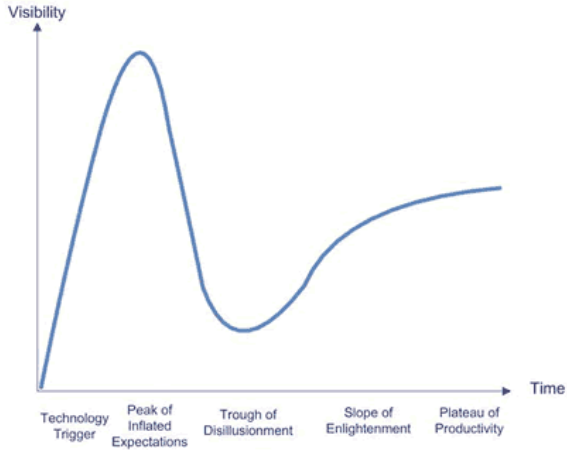
Krista J. Gile,
University of Massachusetts, Amherst, USA

Lisa G. Johnston
Tulane University, New Orleans, and University of California, San Francisco, USA

and Matthew J. Salganik
Microsoft Research, New York, and Princeton University, USA

https://doi.org/10.1111/rssa.12059

Where are we and where are we going?



Gartner consulting.

Conclusion

- Network-based sampling methods can help us collect data from hidden populations, such as the groups most at-risk for HIV.

Conclusion

- Network-based sampling methods can help us collect data from hidden populations, such as the groups most at-risk for HIV.
- Under certain conditions, people are selected with probability proportional to degree.

Conclusion

- Network-based sampling methods can help us collect data from hidden populations, such as the groups most at-risk for HIV.
- Under certain conditions, people are selected with probability proportional to degree.
- If people are selected with probability proportional to degree, then it is possible to produce estimates that are unbiased (correct on average).

Conclusion

- Network-based sampling methods can help us collect data from hidden populations, such as the groups most at-risk for HIV.
- Under certain conditions, people are selected with probability proportional to degree.
- If people are selected with probability proportional to degree, then it is possible to produce estimates that are unbiased (correct on average).
- Even if all the conditions are met, respondent-driven sampling seems to have high sample-to-sample variability because of the properties of real social networks (specifically bottlenecks between groups that have differnt characteristics).

Next class:

- Granovetter, M. (1973). "The strength of weak ties." *American Journal of Sociology*.
- Smith, S. (2005). ""Don't put my name on it": Social capital activation and job-finding assistance among the black urban poor." *American Journal of Sociology*.