To skip or not to skip

The quality of estimation depends on the length of the chain and on the number of participants in the estimation. In order to make the chain longer we can separate the payment for coming and taking part in the testing and the payment only for providing the list of friends. This can be especially suitable for people from hard-to-reach populations, where one can obtain some amount of money without revealing needed information about him but only by pointing (or recruiting) his friends. There is a trade-off: on one hand we make the chain longer and reduce dependency between participants. On other hand we spend money on people who do not bring any information needed for research and finally there will be less participants. For now we will assume that it is not people who decide to participate or just provide the list of their friends, but researches. Thus, having fixed budget to conduct the RDS there is need to answer following questions: how much to pay for the participation in their research, how much to pay for simply providing the names of the friends and how many people invite for participation. The goal is to minimize error of the estimated parameter.

Let's say that the payment or cost of providing list of the friends is c_f and cost of participation is c_p. In this way each of n person that does one of actions gets the payment for providing friends c_f and part p of n people get additional payment c_p for also participation in test. Thus having fixed budget B:

$$B = n \cdot c_f + np \cdot c_p$$

Let's say that error of the estimated parameter is the function of n and p, $f(n, p)$ that decreases with increasing n and with decreasing p.

Insert here graph with the same p and increasing n and with the same n but increasing p.

I simulated values on the nodes according to the Gibbs distribution. Now, I want to estimated the average of these values with the help of the random walk. I use Metropolis-Hasting method to take the samples uniformly(maybe try simple random walk). The question is whether it is better to estimate the average value taking each sample or to skip some samples.

Assigning values to the nodes with the Gibbs distribution brings dependency of values between neighbors (the value of the property on one node depends on the values on its neighbors).

In this way the values of the nodes that we see on step k and on step

$k + 1$ are dependent. So by skipping some nodes can decreases dependency.

The quality of estimation depends on the number of participants in the estimation and on the dependency between them. More participants there are the better is estimation. The less correlation between participants the better. In order to make the chain longer we can separate the payment for coming and taking part in the testing and the payment only for providing the list of friends. This can be especially suitable for people from hard-to-reach populations, where one can obtain some amount of money without revealing needed information about him but only by pointing (or recruiting) his friends. There is a trade-off: on one hand we make the chain longer and reduce dependency between participants. On other hand we spend money on people who do not bring any information needed for research and finally there will be less participants. For now we will assume that it is not people who decide to participate or just provide the list of their friends, but researches. Thus, having fixed budget to conduct the RDS there is need to answer following questions: how much to pay for the participation in their research, how much to pay for simply providing the names of the friends and how many people invite for participation. The goal is to minimize error of the estimated parameter.

Let's say that the payment or cost of providing list of the friends is c_f and cost of participation is c_p. In this way each of n person that does one of actions gets the payment for providing friends c_f and part p of n people get additional payment c_p for also participation in test. Thus having fixed budget B:

$$B = n \cdot c_f + np \cdot c_p$$

Let's say that error of the estimated parameter is the function of n and p, $f(n, p)$ that decreases with increasing n and with decreasing p.