Monte Carlo simulations to estimate power

Anna Fedor

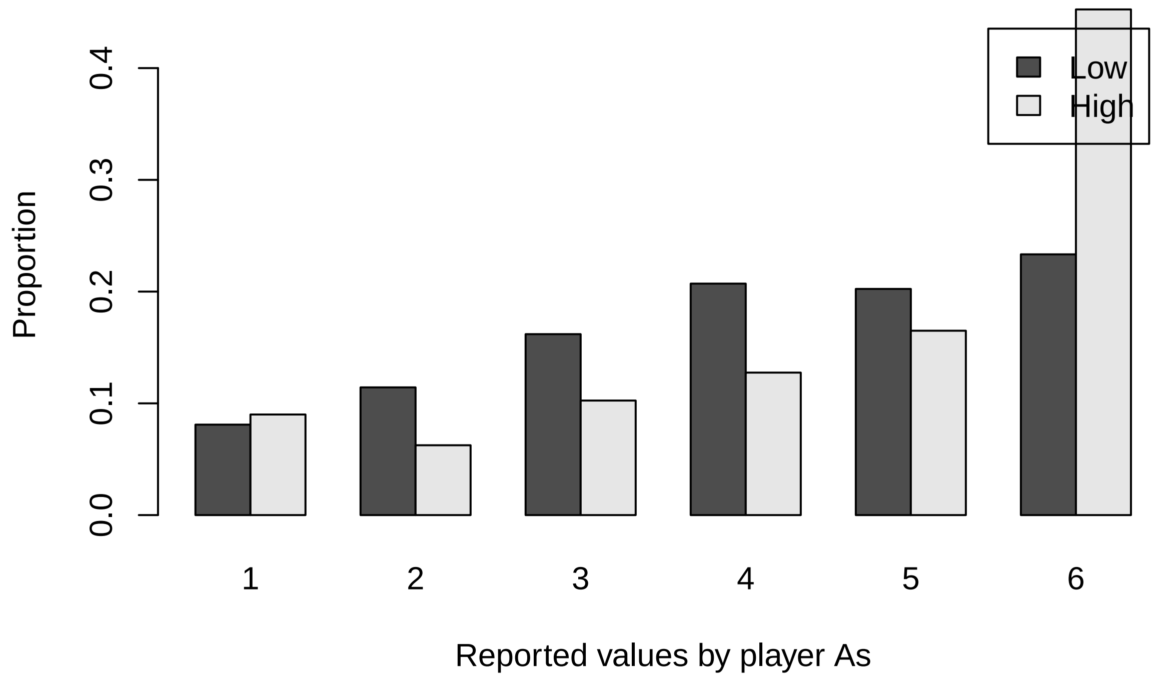
Description

We used this script to estimate the necessary sample size for our experiments to reach enough power. Our goal was to achieve at least 95% power on all of our Wilcoxon signed rank U tests. We simulated participant behaviour based on our assumptions and tested the outcome with a series of sample sizes. We ran **10 simulations with each sample size** and calculated power for a given sample size and a given comparison as the percentage of simulations where the Wilcoxon signed rank U test was significant, i.e., detected a true difference. We decided to choose the smallest sample size with wich we could achieve the desired power on all of our tests.

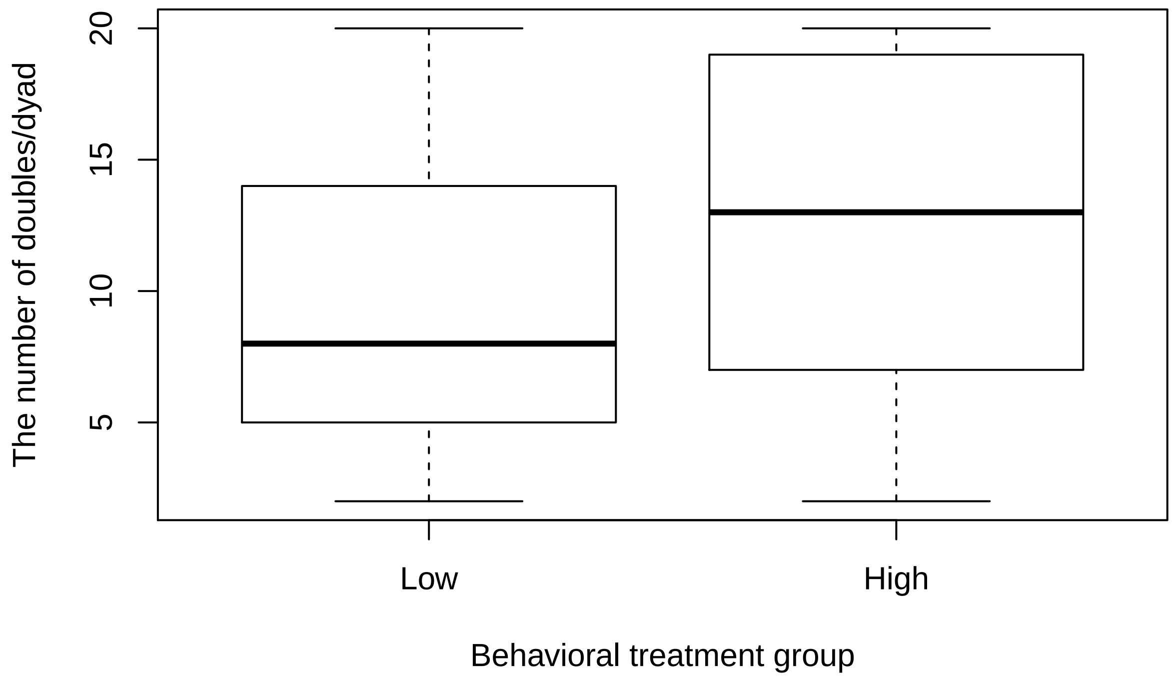
Simulating player A behavior

We planned to test our participants in four conditions with the sequential dyadic die-rolling task used by Weisel & Shalvi, 2015. The main difference is that in our experiment, player A will be simulated by the computer, although player B will be lead to believe that he is playing with a real person. Player A will be either "honest" or "dishonest". The reported values of honest player As will be sampled from a uniform distribution between 1 and 6. The reported values of dishonest player As will be sampled from the values reported by participants who were in the role of player A in Wouda et al., 2017, Study 2, High behavioral norm group.

The original authors sent us their data and script with which they calculated their statistics. In this experiment, participants's norm was manipulated by showing them results from previous experiments: either from an experiment where participants lied very often (High behavioral norm group) or from an experiment where participants lied less often (Low behavioral norm group). The manipulation affected participant behavior: Participants in the high behavioral norm group reported higher values more often and reported double rolls more often than participants in their low behavioral norm group, see the figures below.



Low
High
Proportion
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
Reported values by player As



The distribution of values reported by player As in the high behavioral norm group was the following (we used this distribution to sample values for our simulated dishonest player As):

Var1	Freq
1	36
2	25
3	41
4	51
5	66
6	181

Simulating player B behavior

For the estimation of power we simulated the behavior of player Bs too. For this, we had to guess what values would player Bs report in our control group and three experimental manipulation groups. Player Bs can only increase their payoff by reporting doubles: if they cheated, they would report more doubles than expected by chance. There is no point for player Bs to report higher values than what they actually rolled, except, if they intend to signal to player As, trying to convince them to cheat. We decided not to model this behavior for the purposes of power calculations.

We simulated player B behavior in the following way. For each of our participant groups, we estimated the probability with which player Bs will report doubles. Each roll was a double with this certain probability. If a roll was a double, player B's reported value was the same as player A's value; if a roll was not a double, player B's reported value was sampled from 1 to 6 with a uniform distribution, excluding player A's value. After simulating the reported values of all participants in this way, we calculated the statistical tests and power.

For estimating the probability of reporting a double, we used data from Wouda et al., 2017, Study 2, again. We chose this experiment because this had a manipulation that we assumed that affected participant behavior similarly as our intended manipulations would do.

Simple game with honest partner (control group)

For our simple game with honest partner the probability of reporting a double was taken from the Low behavioral norm group: it was 0.4666667 (see their Table 1). This could potentially overestimate the probability of reporting doubles in our experiment. Differences that might lead to overestimation:

- Their participants were used to participating in economic studies, which supposedly increases the tendency to cheat. Most of our participants probably did not participate in any human experiments before.
- Despite the norm manipulation, their player As still cheated by reporting higher values more often than expected by chance, which in turn might have influenced player B behavior to cheat too. Our simulated player As will be perfectly "honest", which might discourage chating in player Bs too.
- The norm manipulation probably affected player B behavior by its own right. Although it decreased cheating compared to the high behavioral norm group, it might have increased cheating compared to no manipulation at all (which was not tested). The figures that participants studied before the experiment showed data from Study 1, where particpants reported 30% of doubles, which is still higher than expected by chance.

Overestimating the probability of reporting doubles might lead to overestimating power for a given sample size with the one sample Wilcoxon signed rank U test - however, we did not worry about this, since the sample size necessary for two-sample Wilcoxon tests will be higher, i.e., the one sample test won't be the one to define our sample size anyway.

Charity game with dishonest partner

In our view, player B behavior in Wouda et al.'s Study 2 was afftected in two ways:

1. By the experimental manipulation of showing them figures about the results of previous experiments
2. By the behavior (reported values) of player A, which was also affected by the experimental manipulation

In other words, player B in the high behavioral norm group was "encouraged" to cheat not only by the experimental manipulation but also by the behavior of player A.

In our charity game with dishonest partner the case is similar: player A cheats more than in the simple game with honest partner and also, there is the effect of the experimental manipulation of donating to a charity. The difference between player A behavior in our case is higher than in the Wouda et al. exeriment, which could lead to understimation of the effect, what is a safe thing to do when estimating power. As for the experimental manipulation, we have no way to guess how much charity would affect behavior in this experiment, so we just suppose that the effect would be similar to that of the norm manipulation in the Wouda et al. experiment.

For these reasons we took the probability of reporting doubles from the High behavioral norm group: 0.6275. (The value 67% reported in Wouda et al., 2017, Table 2 is a typo.)

Single manipulation groups

If we assume that our manipulations of partner honesty and game type have an added effect, then the probability of reporting doubles in our single manipulation groups (simple game with dishonest partner and charity game with honest partner) should be between that of the control group and the double manipulation group. The question is that how much of the effect can be attributed to either manipulation.

The highest power and lowest sample size would be estimated for the comparison between groups if the effect of the two manipulations are equal. Since we have no reason to assume that te effect of the two manipulations are the same, we assumed that the effect of partner manipulation would have a higher effect: it would be responsible for 50% of the difference between the control group and the double manipulation group.

Summary

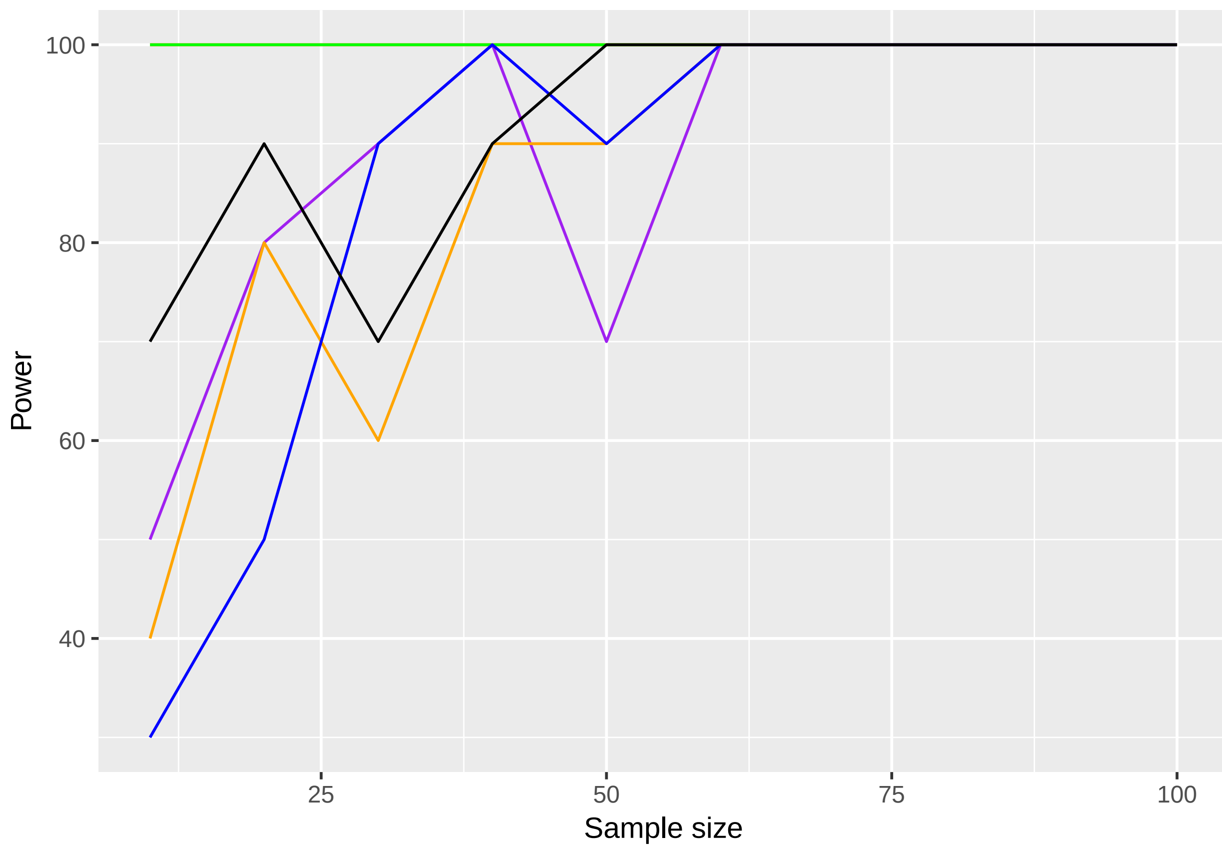
A completely honest player B would report doubles with the probability of 0.1666667. In our power simulations we estimated that player Bs would report doubles in the different experimental groups with the following probabilities:

- Simple game with honest partner: 0.4666667
- Simple game with dishonest partner: 0.5470833
- Charity game with honest partner: 0.5470833
- Charity game with dishonest partner: 0.6275

Results

The results of our simulations are shown in the table and figure below. The table shows the power of all our tests by sample size.

sample_size	SH	SD	CH	CD	SH_SD	CH_CD	SH_CH	SD_CD
10	100	100	100	100	50	40	30	70
20	100	100	100	100	80	80	50	90
30	100	100	100	100	90	60	90	70
40	100	100	100	100	100	90	100	90
50	100	100	100	100	70	90	90	100
60	100	100	100	100	100	100	100	100
70	100	100	100	100	100	100	100	100
80	100	100	100	100	100	100	100	100
90	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100


100
80
60
40
Power
25
50
75
100
Sample size