All-in

elaidlaw, lbai5, jgood1, awey, ssyed7

Hypothesis

As avid poker players, we sought out to investigate at which moment during a poker session does a player's performance change. More specifically, we are seeking to understand whether a player is more likely to win a hand in a winning session (cumulative net gain is positive) or losing session (cumulative net gain is negative) on a given day. Thus, our hypothesis tests whether the win rate of players who had a negative cumulative net gain for the session and that of players who had a positive cumulative net gain for the session are equivalent.

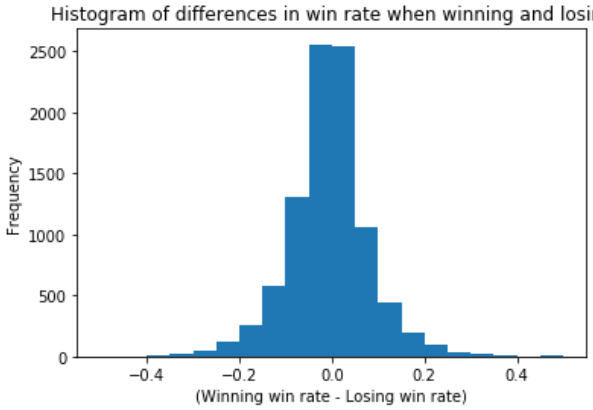
Data

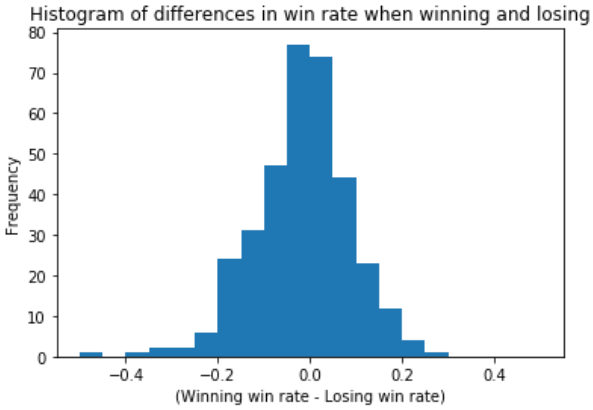
We collected data from Kaggle and Michael Maurer's IRC Poker Database, both of which are data sets of players' hand history of online texas holdem games. Both sources provide thorough logs of online games, though we cannot be sure that all players are playing seriously or even that all players are actual players and not poker bots.

The raw data is in text format and we parsed the text files to create three data tables: Games, Hands, and Actions. 'Games' contains fields: game ID, data source, big blind, community cards, and pot sizes pre-flop, post-flop, post-turn, and post-river. 'Hands' contains fields: game ID, data source, player ID, card_1, card_2, bets, net gain, and chips at the beginning of hand. 'Actions' contains fields: game ID, data source, player ID, position in rounds, and an action ID.

Findings

Claim: Players tend to improve their performance when in a losing session.
Support for Claim: For each player in our database, we calculated the difference between two averaged win rates, their win rate within positive sessions and their win rate within negative sessions. With the null hypothesis that this difference is zero, we ran z-tests on the data from the IRC dataset and kaggle dataset. The samples can be seen in the histograms below (IRC left, kaggle right). When running z-tests against a zero mean, the p-value for the IRC dataset was 1×10^{-8}, while the kaggle dataset resulted in a p-value of 1.1×10^{-3}. The respective means were -0.50% and -1.7%. Therefore, the results are significant and we reject the null hypothesis and find that players are more likely to win in a losing session than in a winning session on average.


Histogram of differences in win rate when winning and losing
Frequency
2500
2000
1500
1000
500
0
−0.4
−0.2
0.0
0.2
0.4
(Winning win rate - Losing win rate)


Histogram of differences in win rate when winning and losing
Frequency
80
70
60
50
40
30
20
10
0
−0.4
−0.2
0.0
0.2
0.4
(Winning win rate - Losing win rate)