

The amount of luck in competitive bridge

Dah Ming Chiu, November 22, 2020.

1. Introduction

Recently, a report was published on the Bridge Winners website alleging cheating by a world-class bridge player Giorgio Duboin [1] that brought heated discussions. The report was authored by a team of top-notch bridge players with detailed analysis of many boards played by Duboin, the statistics of which was also compared with that of some other world-class players for benchmarking. The methodology, though relying to a large extent on manual analysis, suggests a nice framework for understanding evaluating skill level in competitive bridge, with and without unauthorized information (UI). This also sheds light on the question of the amount of luck in bridge, which is the main topic of this article. The determination of cheating in bridge based on play records is a very hard problem, conceptually similar to the problem of determining insider trading based on trading records. Without having carefully read through the 189 pages of [1], we will not discuss here the conclusions of reached in that report, but that report is definitely interesting reading for bridge lovers and they can try to reach their own conclusions.

The question of luck versus skill has been widely discussed before, in the context of different sport games, and evaluation of professional skills that depend on random events, for example in the case of investment advisors. A very good survey and discussion can be found in [2].

For the rest of the article, we assume you have some rough knowledge about the game of bridge and how the format of bridge competition. But the question we discuss does not rely on detailed knowledge of bridge to appreciate.

2. Definition of score card

Bridge is a game that requires a player to make bidding, defending and declaring decisions on each board played. The bridge competition format that arguably requires the most skill is played between two competing teams, each with four players, playing a fixed number of boards. Team 1's North and South play against team 2's East and West, and vice versa; so the same challenges for each board will be faced by players from both teams, with the team making the decision that yielded better results winning the board.

Each playing decision can be classified as being one of two types: (a) decision with unpredictable outcome, and (b) decision with predictable outcome. For case (a), if the decision is to choose between two (or more) options, then the option has equal chance of being the winning option. In other words, the player's decision is like flipping a coin and calling it heads. For case (b), some options are better than others, statistically speaking; for example, option 1 has 70% of winning whereas option 2 has

30% of winning. A skillful player can quickly evaluate the available options and pick the winning option to play. It should be noted that in case (b), the right decision may not win. In the above example scenario, 30% of time the right decision does not win. Accordingly, report [1] classifies each board based on the most crucial decision that determines the outcome of the board as follows (using our own notations):

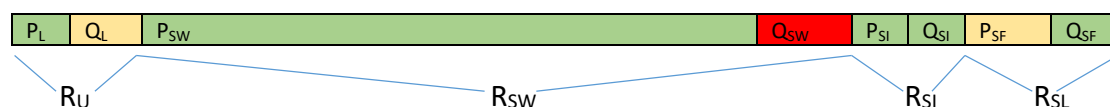
1. **Luck Boards (denoted L)**: When play options have equal chance to win the board.
2. **Skill Boards (denoted S)**: When the better play option has higher chance to win the board than other options. But there are three subcategories:
 - a) **Win (SW)** by playing best option, as predicted.
 - b) **Inconsequential (SI)**, whether you play the best option or not, the result is the same
 - c) **Failed (SF)** to win by playing the best option, the less likely event occurred.

Let the frequency of these four types of boards be R_L , R_{SW} , R_{SI} and R_{SF} respectively; they sum up to 1 by definition. In [1], R_L was found to be around 15%.

We define the following notations to refer to the frequency (out of all boards played) of different kinds of decisions (and outcomes) made by a player:

Frequency	From decisions	Symbols used in [1]
P_L	Frequency L boards won	G (Good)
$Q_L=R_L-P_L$	Frequency L boards lost	B (Bad)
P_{SW}	Frequency best option taken for SW boards	Part of N (Normal)
$Q_{SW}=R_{SW}-P_{SW}$	Frequency inferior option taken for SW boards	A (Anti-suspicious)
P_{SI}	Frequency best option taken for SI boards	Part of N (Normal)
$Q_{SI}=R_{SI}-P_{SI}$	Frequency inferior option taken for SI boards	L (Lazy)
P_{SF}	Frequency best option taken for PL boards	Part of N (Normal)
$Q_{SF}=R_{SF}-P_{SF}$	Frequency inferior option taken for PL boards	S (Suspicious)

This classification of the boards and definition of result for each type of board, essentially defines a **score card** for each player after a match, illustrated pictorially:



All the green parts are winning results. The red part signifies the losses due to the player's mistake of not taking the best option. The first yellow part signifies the losses due to luck. The second yellow part signifies the losses due to bad luck despite the player make the best play. Assuming the player played skillfully, the quantity Q_{SI} and Q_{SF} should be very small (relative to R_{SI} and R_{SF}), since those are frequency when the best plays are not made. On the other hand, P_L and Q_L should be roughly equal since they are results of tossing a coin.

A cheater (using unauthorized information) can see all the cards, so the boards in L do not need to depend on luck – it is possible to make $P_L \gg Q_L$ (or Good \gg Bad in [1]), to the cheater's advantage. Secondly, the cheater can make $Q_{SF} \gg P_{SF}$, by choosing

the inferior play when that yields better result for boards in SF (Q_{SF} is known as Suspicious in [1]). Thirdly, it is noticed that Q_{SI} (called Lazy in [1]) also tends to increase, since keeping Q_{SI} small requires more effort. This is essentially the abnormal behavior that led [1] to detect cheating. The table below from [1] summarizes the result of that report in terms of score cards:

Table 8: de Wijs and Muller compared to Mr. Duboin kibitzers/no kibitzers

Player	Simon de Wijs		Bauke Muller		Giorgio Duboin Kib. Disallowed		Giorgio Duboin Kib. Allowed	
	#	%	#	%	#	%	#	%
N	118	79.73%	127	85.81%	122	76.25%	406	74.09%
G	11	7.43%	7	4.73%	12	7.50%	53	9.67%
B	8	5.41%	4	2.70%	12	7.50%	14	2.55%
S	1	0.68%	0	0.00%	0	0.00%	47	8.58%
A	8	5.41%	9	6.08%	9	5.63%	3	0.55%
L	2	1.35%	1	0.68%	2	1.25%	15	2.74%
W	0	0.00%	0	0.00%	3	1.88%	9	1.64%
Excl.	0	0.00%	0	0.00%	0	0.00%	1	0.18%
Total	148	100.00%	148	100.00%	160	100.00%	548	100.00%

From the statistics of G, B, S, A and L, we can see how [1] build up its case against the suspected player, using some other players' score statistics for benchmarking.

3. The luck curve

We have seen how the *score card* of a player can be used to analyze if he/she is using unauthorized information. It can also be used to tell the amount of luck in bridge. For that purpose, we need to compare the score card of two players, competing in the same position representing the competing teams. The difference between the score cards give the margin of win (or loss).

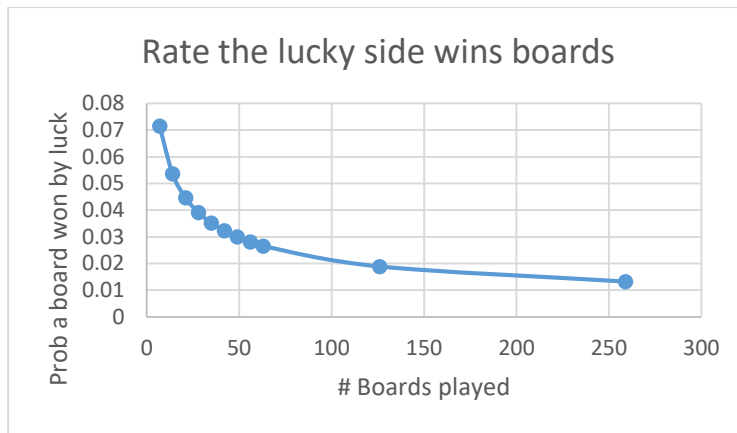
The first step is to find out the difference in the outcome of the Luck boards. For each board in L, whether a given player will get it "right" is just like the result of tossing a coin for heads. For n such boards, the number of heads is a random variable with Binomial distribution, denoted $B(n, 1/2)$. Let us denote the random variable for the first team as X , and that from the second team as Y . Then the difference $|X-Y|$ is also a random variable with distribution:

$$\text{Prob}(|X-Y|=0) = \binom{2n}{n} \frac{1}{2^{2n}}$$

$$\text{Prob}(|X-Y|=k) = \binom{2n}{n+k} \frac{1}{2^{2n-1}} \quad k = 1, 2, \dots, n$$

This is the distribution either team wins by a margin of k boards, out of a total of n Luck boards played.

Assuming we have on average 1 Luck board for every 7 boards played, which is roughly the 15% rate observed in [1], then the rate the lucky side will win by luck can be plotted against the number of boards played, as shown.



If we play only 7 boards in a match, then luck factor is as high as around 7% in playing each board. As we play more boards for a match, the luck factor reduces steadily, to only slightly over 1% for a 256 board match. This phenomenon, the reduction of the luck factor as we play more times, is mentioned many times in [2], referred to as “reversion to the mean”. We refer to this as the luck curve (for bridge).

4. Evaluation of relative skill

If the two competing players representing two teams have similar skill and make all the right decisions (picking the better option) for each board, then the result of the match will be determined by luck, and the luck curve will give the winning margin. Out of the three type of Skill boards, since most are expected to be in the SW set, let us focus on that first, and assume SI and SF are negligibly small. If team 1 makes mistake on $X=P_{SW}(1)$ percent of the boards in SW, and team 2 makes $Y=P_{SW}(2)$ percent, and assume $X>Y$ without loss of generality, then we can compute the skill difference $SD=(X-Y)/R_{SW}$ as the rate team 1 is out-winning team 2 by skill on each board. If we play enough boards so that SD is larger than the corresponding lucky curve value, then we can claim skill will determine the winner.

According to [1], top players are making errors at the rate of 5-6% (Q_{SW} , the red colored part of the score care). The difference between competing top players may be small (1% or less). Even if they play a 256 board match, the result is usually not very predictable – since the luck factor is greater than the skill difference. On the other hand, the error rate for active but average players may be considerably higher; it would not be surprising Q_{SW} is on the order of 20%. In that case, the skill difference when playing against top players is quite large, and such players will be consistently out-played with a 32 board or even 16 board match. These observations are consistent with experiences by tournament bridge players.

The above analysis is based on comparing the score card of two players playing in the same position from two competing teams. How to apply the methodology to the teams? Strictly speaking, we need to merge the players score cards and arrive at a team score card, and compare at the team level. What is the statistics of the score card at the team level look like? Is it much different than that at the individual level? Since we do not have the benefit of example statistics as that provided by [1], it is

hard to say.

The above analysis also ignored the effect of Q_{SF} , unexpected good result from bad play. This is really a second order effect, since this should happen at the rate equal to the product of R_{SF} and play error rate, meaning $Q_{SF} \sim R_{SF} * P_{SW}$ for non-cheating players. The way to incorporate this factor is to subtract this rate from the error rate, so the effective error rate becomes $P'_{SW} = (1 - R_{SF}) * P_{SW}$. Unfortunately, [1] did not provide statistics for R_{SF} .

5. Discussion and Conclusion

Inspired by the bridge analysis on cheating in [1], we look at a related question of how much luck versus skill is in the game of bridge. We first review the method used by [1], and extract from it the concept of a score card. Then we point out how the result of a duplicate bridge match is determined by comparing two scoring cards. Determining the luck factor is particularly interesting as it depends on the difference between two random variables, which itself is a random variable with known distribution. The expected value of this difference is reduced as we play more boards, which is plotted as the “luck curve”. We then explain how the relative level of the luck and skill factor can be compared.

It goes without saying that our analysis is more at the qualitative and conceptual level, rather than at a level that can be readily applied to evaluate real-world bridge matches. Bridge scoring is complicated. Making a suboptimal play on different boards can have very different effects towards the match, rather than simply winning or losing a board as we assumed. Also, we have implicitly assumed that plays are mostly at expert level with few errors, so a certain play can be assumed to have some skill level with some consistency. At an average player level, error rate cannot be assumed to be consistent from match to match, then the concept of skill may need to be model as a random variable itself, and becomes more complicated. But hopefully our analysis shed some light on the question of skill versus luck for competitive bridge.

References

[1] Sjoert Brink et al, “Report on Giorgio Duboin”, Nov 5, 2020,
<https://bridgewinners.com/article/view/the-hand-records-speak-giorgio-duboin/>

[2] Michael Mauboussin, “Untangling Skill and Luck, how to think about outcomes – past, present, and future”, Legg Mason Capital Management, July 15, 2010,
<https://www.trendfollowing.com/pdfs/UntanglingSkillandLuck.pdf>