Best of Three: Mathematical Model

Oats

Smogon University

Motivation

A central problem in Pokemon tournaments is variance between games due to probabilistic factors, hence recently, there has been much interest in running tournaments in a best of three format under the belief that it reduces variance.

On the other hand, a best of three matches in a probabilistic setting may introduce significant bias on the outcome of the tournament. Thus, a mathematical model of the best of three situation could shed important light on whether this new tournament format favours different people.

Toy Model

Consider a scenario where there are four players. Assign to these players score cards which represent how well they play (figure 1).

Deciding who wins or loses a match is done by comparing the score cards. A player with a higher score wins a game, and the score they get from their card is decided randomly. For instance, if Player A with all fives battles Player C with a three, one, and a ten, then Player A wins two-thirds of the time. Tables of probabilities (1) are constructed from these comparisons to find win loss ratios in singles and best of three settings.

Contestant A Score: $\{5, 5, 5\}$ Consistent player	Player A has above-average skill, and consistently plays at the same level of five each game.
Contestant B Score: $\{6, 6, 2\}$ Matchup fisher	Player B has above-average skill, and likes to matchup fish. They usually get it right and score a six, but sometimes they run into a matchup they didn't prepare for, and play like a two.
Contestant C Score: $\{3, 1, 10\}$ Gimmick Abuser	Player C is a below-average player who takes gimmicks, luck, and matchup fishing too far. They attempt to get unreasonably lucky in the hopes to occasionally win without much counter play from the opponent. They usually get it wrong and perform poorly with a score of three, or lose outright with a score of one, but sometimes it works out for them and they are unbeatable with a ten.
Contestant D Score: $\{4, 4, 4\}$ New Consistent Player	Player D is a below-average player who is new to tournaments. They consistently play at the same level of four each game.

Singles

	A	B	C	D
A	-	33.3%	66.6%	100%
B	66.6%	-	55.5%	66.6%
C	33.3%	44.4%	-	33.3%
D	0.00%	33.3%	66.6%	-

Best of Three

	A	B	C	D
A	-	25.9%	74.1%	100%
B	74.1%	-	58.3%	74.1%
C	25.9%	41.7%	-	25.9%
D	0.00%	25.9%	74.1%	-

Table 1. Singles and best of three match probabilities for each contestant winning against another. The table is read as the row player winning against column player by the corresponding percentage.

Players are then ranked by how likely they are to win any given match. The player who wins the most on average is deemed the best (table 2). The probability that someone wins a best of three match is $p^2 + 2p^2(1-p)$ where p is the probability of winning a single. This explains the difference between the rankings in table (2).

	Singles		Best of Three	
Rank	Contestant	Average Win Rate	Contestant	Average Win Rate
1^{st}	A	66.6%	B	68.8%
2^{nd}	B	63.0%	A	66.6%
3^{rd}	C	37.0%	D	33.3%
4^{th}	D	33.3%	C	31.2%

Table 2. Average win rates of contestants in singles and best of three settings, arranged from highest to lowest, and given a rank to emphasise how well they perform. Notice how the contestants ranks changed between the two settings.

Partial Tournament Data

While the model demonstrates that tournaments of singles and tournaments of best of three can favour different players in principal, do real tournaments also exhibit these conditions? To find out, we take a sample of real tournament data (with pseudonyms, table 1) to generate player cards; use these player cards (table 4) to fill in the rest of the table, and then attempt to determine if any players are favoured in one tournament style over the other (table 5).

	Pious	Knight	Throw dust in eyes	Drive on	Ken	Lattice	Tutee	Wreath	Zirconia	Comet	Wind of the spirit	Friction	Hatchet	Christmas	Moloch	Sakedon
Pious	–	0.67	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Knight	0.33	–	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
Throw dust in eyes	0.0	0.67	–	0.67	0.0	0.0	0.67	0.67	0.67	0.0	0.0	1.0	0.0	0.67	1.0	1.0
Drive on	0.0	0.67	0.33	–	0.0	0.33	0.33	0.67	0.67	0.33	0.33	0.33	0.33	0.67	1.0	1.0
Ken	0.5	0.67	1.0	1.0	–	1.0	1.0	1.0	1.0	0.67	1.0	1.0	1.0	0.83	1.0	1.0
Lattice	0.0	0.67	1.0	0.67	0.0	–	1.0	0.67	0.67	1.0	1.0	1.0	1.0	0.67	1.0	1.0
Tutee	0.0	0.67	0.33	0.67	0.0	0.0	–	0.67	0.67	0.0	0.0	1.0	0.0	0.67	1.0	1.0
Zirconia	0.0	0.67	0.33	0.33	0.0	0.33	0.33	0.67	–	0.33	0.33	0.33	0.33	0.67	1.0	1.0
Comet	0.0	0.67	1.0	0.67	0.33	0.0	1.0	0.67	0.67	–	1.0	1.0	1.0	0.67	1.0	1.0
Wind of the spirit	0.0	0.67	1.0	0.67	0.0	0.0	1.0	0.67	0.67	0.0	–	1.0	1.0	0.67	1.0	1.0
Friction	0.0	0.67	0.0	0.67	0.0	0.0	0.0	0.67	0.67	0.0	0.0	–	0.0	0.67	1.0	1.0
Hatchet	0.0	0.67	1.0	0.67	0.0	0.0	1.0	0.67	0.67	0.0	0.0	1.0	–	0.67	1.0	1.0
Christmas	0.0	0.67	0.33	0.33	0.17	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	–	0.33	0.78
Moloch	0.0	0.67	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.67	–	0.67
Sakedon	0.0	0.67	0.0	0.0	0.0	0.0	0.0	0.22	0.0	0.0	0.0	0.0	0.0	0.22	0.33	–

Table 3. Table of Best of Three Win Probabilities. Each entry in the table is read as the probability that the row player defeats the column player. The data was taken from Jirachee's invitational tournament rounds four through to ten. Highlighted in blue is the original data from these rounds, which was used to generate the player cards in figure (4). These player cards were then used to generate the remaining uncoloured probabilities of the table.

Pious	22, 22, 22, 22, 22, 22
Knight	0, 0, 24, 0, 0, 24
Throw dust in eyes	10, 10, 10, 10, 10, 10
Drive on	7, 7, 17, 7, 7, 17
Ken	20a, 20a, 20a, 23a, 23a, 23a
Lattice	15, 15, 15, 15, 15, 15
Tutee	9, 9, 11, 9, 9, 11
Wreath	4, 4, 19, 4, 4, 19
Zirconia	6, 6, 18, 6, 6, 18
Comet	14, 14, 14, 14, 14a, 14a
Wind of the spirit	13, 13, 13, 13, 13, 13
Friction	8, 8, 8, 8, 8, 8
Hatchet	12, 12, 12, 12, 12, 12
Christmas	2, 2, 21, 2, 2, 21
Moloch	3, 3, 3, 3, 3, 3
Sakedon	1, 1, 5, 1, 1, 5

Table 4. The numbers on the player cards which were generated from the tournament data in figure 1. Only six entries were required for each card because players either win half the time against another, or two thirds of the time against another, or never, or all the time. Some player cards contain special characters which allow for situations where three or more players always win against each other in a cyclical way.

Singles Rankings:

Rank:	Name:	Win Ratio:
1	Pious	0.94
2	Ken	0.91
3	Lattice	0.76
4	Comet	0.71
5	Wind of the spirit	0.62
6	Hatchet	0.56
7	Throw dust in eyes	0.47
8	Drive on	0.47
9	Zirconia	0.44
10	Tutee	0.44
11	Wreath	0.41
12	Friction	0.36
13	Christmas	0.35
14	Knight	0.33
15	Moloch	0.13
16	Sakedon	0.10

(a) Singles Rankings

Best of Three Rankings:

Rank:	Name:	Win Ratio:
1	Pious	0.95
2	Ken	0.93
3	Lattice	0.78
4	Comet	0.73
5	Wind of the spirit	0.65
6	Hatchet	0.58
7	Throw dust in eyes	0.50
8	Tutee	0.46
9	Drive on	0.45
10	Zirconia	0.42
11	Friction	0.38
12	Wreath	0.38
13	Christmas	0.30
14	Knight	0.26
15	Moloch	0.15
16	Sakedon	0.08

(b) Best of Three Rankings

Table 5. A comparison of the most likely player rankings in a singles setting (a), and a best of three setting (b). The changes in the placement when going from singles to best of three; were that Drive on went from 8^{th} to 9^{th} place, Zirconia went from 9^{th} to 10^{th} place, Tutee went from 10^{th} to 8^{th} place, Wreath went from 11^{th} to 12^{th} place, and Friction went from 12^{th} to 11^{th} place.

The results from the rankings table (5) using partial tournament data suggest that there could be some rearrangement of the standings in a real tournament, and an early trend has appeared where players near the top of the rankings win more often, while those near the bottom win less. However the incompleteness of the data and its small sample size make it hard to determine how present these effects would be in a real tournament, if at all.

Conclusion

Mathematically speaking, there exist scenarios where tournaments of singles and tournaments of best of three can favour different types of players.

The partial tournament data used that was used to gauge whether these effects translate to real tournaments was not sufficient to determine if they do, or by how much.

Future work

An experiment should be run in the form of a tournament, to determine if the winner of the tournament is the same in a singles setting compared to the best of three.

The format of this tournament should be a 3-of round robin, where the winner is determined by whoever has the highest win loss ratio. This will ensure that there is sufficient data to determine winners for best of three, as well as singles settings.

References

[1] Hiro'.
Jirachee's dpp invitational - teams, replays, and usage statistics, 2023.