Survey Sampling and Cricket: Predicting the 'Gentleman's Game'

Mohammed Quazi*, Pavan Datta†, Joshua Clifford‡

Abstract

Introduction

The 2019 Mens International Cricket Council (ICC) Cricket World Cup (CWC) was the most watched ICC event of all time, with a 1.6 billion cumulative average audience for live coverage, a 38% increase over the 2015 edition. This research develops a statistical modeling procedure to predict outcomes of future ICC cricket events. The proposed model provides an insight into the application of survey sampling to the team selection pattern by incorporating individual players performance history not only against a particular opposition but also against any cricket-playing nation (full member of ICC). A case study for the next ICC CWC 2023 in India is provided and simulations are discussed.

Methods

This study employs stratified random sampling (SRS) technique shown in figure 1 to select team lineups based on player roles. *S1-S6* denote different types of players based on their role in the team - fast bowler, spinner, all-rounder-fast bowler, all-rounder-spinner, batsman, wicket-keeper. Runs scored by every player on the team is derived using an algorithm which is based off of estimated parameters of a gamma distribution against a particular opposition, which is then used to simulate match and tournament results. Data collected spans over 11000 individual one-day international (ODI) innings which cover every competitive game played among the full members of ICC since 1999. 195 international players are considered for the case study. The proposed model accounts for the Indian subcontinent playing conditions and debutants' performances as well.

Results

Figure 2 shows the probability distribution for standings based on points accumulated at the end of group stage of ICC CWC 2023, after which the top 4 teams qualify for the semi finals. Figure 1 shows the predicted probability of winning for teams against each other. Results indicate that India has the highest chance of qualifying for the semis with a 24% chance of winning the cup. Nevertheless, surprisingly, Pakistan are the favorites, with a 52% chance of winning the trophy if they make it to the semifinals living up to their 'unpredictable' tag, setting up a replay of the ICC Champions Trophy Final in 2017 which they won being

*Mohammed Quazi, Department of Mathematics and Statistics, University of New Mexico
†Pavan Datta, Sandia National Laboratories
‡Joshua Clifford, Sandia National Laboratories

underdogs against India, South Africa (another heavy favorite with) are likely to shed their reputation for choking tag and qualify for the semis and have a 9% chance of winning the tournament.

Conclusion

The method can predict probabilities of winning, which is of interest to fans and could even help set gambling odds. The method can also be adapted to recommend team selection strategies to inform team strategy to increase chances of success in a game. This study could be applied to all league-format cricket tournaments including the Indian Premier League (IPL) – one of the top 10 most watched sports leagues in the world (average attendance), and the most attended and watched cricket league in the world. The proposed model could be implemented in other sports as well.
(Word count 489/500)

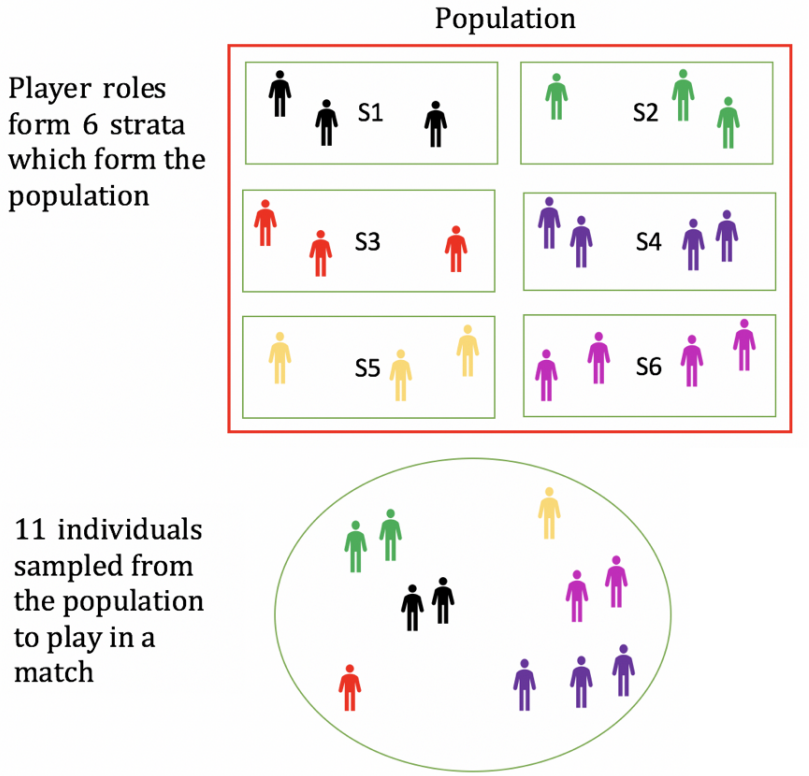

Figure 1: Sampling scheme employed in this study. S1-S6 are strata. Oval shows individuals sampled using stratified random sampling method.

Table 1: Predicted win/loss ratio (in %) at the ICC 2023 CWC

	Losing team											
Winning team	Afghanistan	Australia	Bangladesh	England	India	Ireland	New Zealand	Pakistan	South Africa	Sri Lanka	West Indies	Zimbabwe
Afghanistan		0.20	41.1	4.20	0.00	95.7	0.70	30.1	0.00	8.70	20.3	47.9
Australia	99.8		86.2	61.0	63.8	100	99.5	25.6	0.90	7.30	86.3	100
Bangladesh	58.5	13.6		2.10	2.60	100	0.80	13.4	0.20	2.30	94.9	100
England	95.8	37.7	97.9		10.8	93.8	66.7	37.2	12.5	33.1	37.1	90.2
India	100	35.6	97.4	88.9		100	100	67.7	39.4	99.9	99.3	100
Ireland	3.90	0.00	0.00	5.80	0.00		0.00	0.00	0.00	32.2	0.00	39.5
New Zealand	99.3	0.50	99.0	32.2	0.00	100		85.3	19.0	100	56.8	89.3
Pakistan	68.8	73.0	86.4	61.9	31.7	100	14.3		88.7	100	79.4	100
South Africa	100	99.0	99.8	87.2	59.4	100	80.4	11.0		100	72.8	100
Sri Lanka	90.1	92.6	97.5	66.3	0.10	67.2	0.00	0.00	0.00		71.1	99.6
West Indies	79.2	13.3	4.80	61.9	0.70	100	42.4	19.9	26.8	27.5		97.1
Zimbabwe	49.9	0.00	0.00	9.50	0.00	59.6	9.80	0.00	0.00	0.40	2.80	

Table 2: All time win/loss ratio (in %)

	Losing team											
Winning team	Afghanistan	Australia	Bangladesh	England	India	Ireland	New Zealand	Pakistan	South Africa	Sri Lanka	West Indies	Zimbabwe
Afghanistan		0.00	37.5	0.00	16.7	50.0	0.00	0.00	0.00	25.0	37.5	60.0
Australia	100		95.0	56.8	60.0	100	70.2	67.8	48.5	65.6	55.1	93.1
Bangladesh	62.5	5.00		19.0	14.3	77.8	28.6	13.5	19.0	15.2	41.7	62.7
England	100	43.2	81.0		44.3	83.3	48.9	62.4	48.3	50.0	54.2	72.4
India	83.3	40.0	85.7	55.7		100	52.9	43.0	43.2	61.8	50.4	82.5
Ireland	50.0	0.00	22.2	16.7	0.0		0.00	21.4	0.00	0.00	9.09	50.0
New Zealand	100	29.8	71.4	51.1	47.1	100		46.6	37.9	54.4	48.3	74.3
Pakistan	100	32.2	86.5	37.6	57.0	78.6	53.4		35.9	61.3	45.9	92.1
South Africa	100	51.5	81.0	51.7	56.8	100	62.1	64.1		58.6	74.2	95.0
Sri Lanka	75.0	34.4	84.8	50.0	38.2	100	45.6	38.7	41.4		50.9	80.0
West Indies	62.5	44.9	58.3	45.8	49.6	90.9	51.7	54.1	25.8	49.1		77.7
Zimbabwe	40.0	6.89	37.3	27.6	17.5	50.0	25.7	7.89	5.00	20.0	22.3	

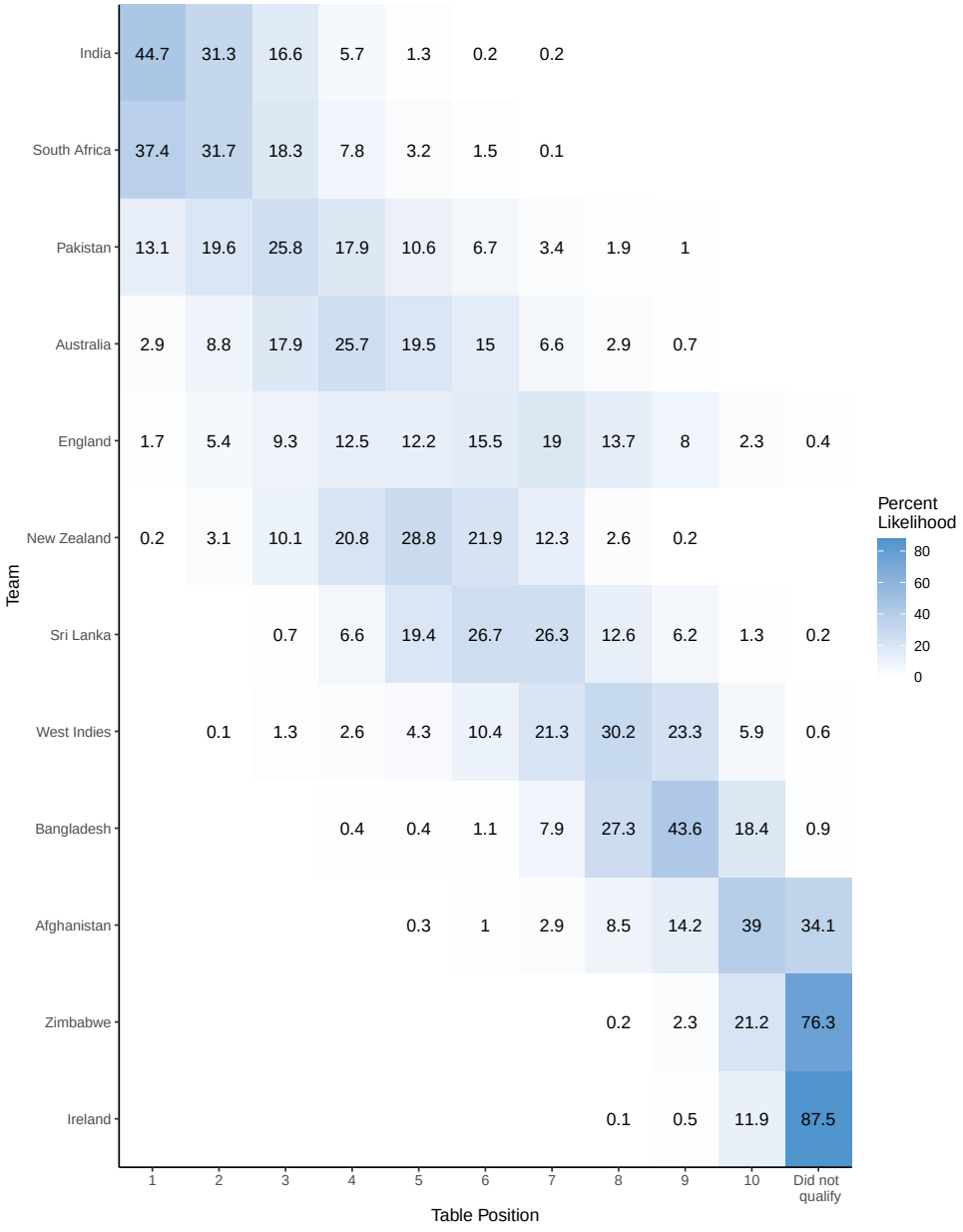

Figure 2: ICC 2023 World Cup Team Table Probability Distribution