

Predicting Cricket Outcomes using Bayesian Priors

Mohammed Quazi

PhD Candidate – Department of Mathematics and Statistics

Statistician – Translational Informatics Division

University of New Mexico

<https://math.unm.edu/~mquazi/>

https://github.com/mquazi/cricket_2023

Outline

1 Background

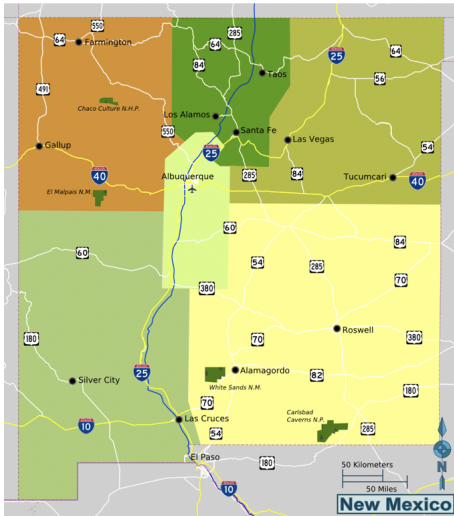
2 Methodology

3 Conclusion

Background

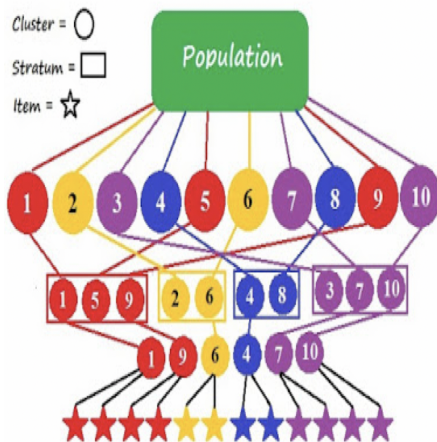
- ▶ Cricket is a game played between 2 teams with eleven players on each side
- ▶ It is the second most popular sport in the world, with 2.5 billion fans, second only to soccer's 4 billion fans (worldatlas.com, 2020)
- ▶ This research paper investigates the impact of incorporating survey sampling theory into team selection patterns to predict outcomes of cricket matches using Bayesian priors
- ▶ Web scraped data using R for over 350 players spanning every One-Day-International game played from 1999 to 2020 and every Indian Premier League game played 2008–2020

Example of a survey design



- ▶ 6 regions form 6 strata (singular stratum)
- ▶ Cities within regions form clusters
- ▶ Neighborhoods within cities form another level of clusters
- ▶ Households within neighborhoods form another level of clusters
- ▶ All individuals from randomly selected households are sampled

Example of a survey design



- ▶ Survey designs could get really complicated
- ▶ Estimating the parameters could get extremely difficult

Figure: 2-stage stratified sampling

Example of another survey design

Population



1 Ultimate Sampling Unit with four clustered households



Four individuals sampled from each USU



Figure: USUs are black rectangles within the population. Oval shows individuals sampled using probability proportional to size method.

Methodology – Survey design for this study

Player roles
form 6 strata
which form the
population



11 individuals
sampled from
the population
to play in a
match



Figure: Stratified random sampling – sampling technique employed for this study

Methodology

Gamma(α , β)

$$f_X(x) = f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 \leq x < \infty, \quad \alpha, \beta > 0$$

$$E(X) = \alpha\beta \quad \text{and} \quad Var(X) = \alpha\beta^2$$

α is called the shape parameter because it defines the peakedness of the distribution and β is called the scale parameter because it shows the spread

- ▶ Model constraints

$$\alpha = \frac{\text{Average score}}{\beta}$$

$$P(X > \text{Highest score}) \leq 0.05$$

- ▶ Used Gamma distribution because of its asymptotic properties

Results – Cricket World Cup 2023 predictions

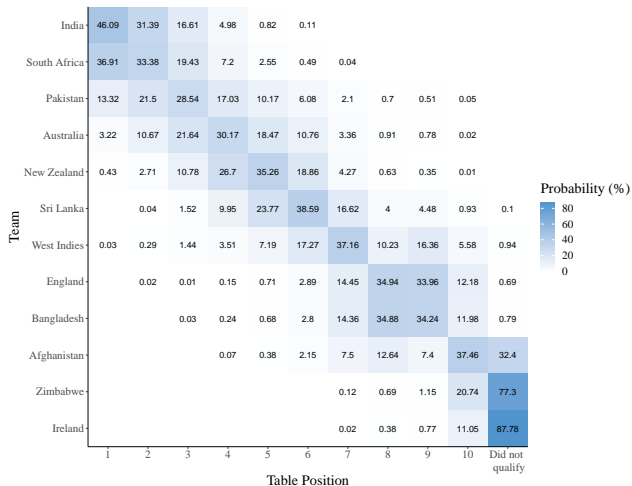


Figure: ICC 2023 CWC predicted team standings probability distribution (in %).

Model Validation using Indian Premier League 2020

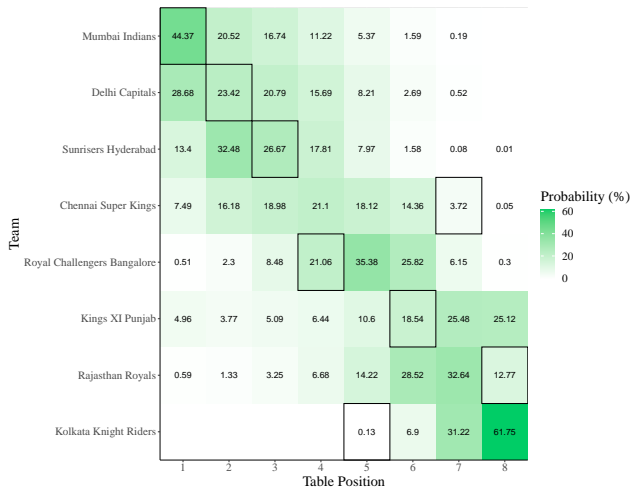


Figure: IPL 2020 predicted team standings probability distribution (in %). Actual IPL 2020 standings are shown for each team with a black rectangle.

AB de Villiers

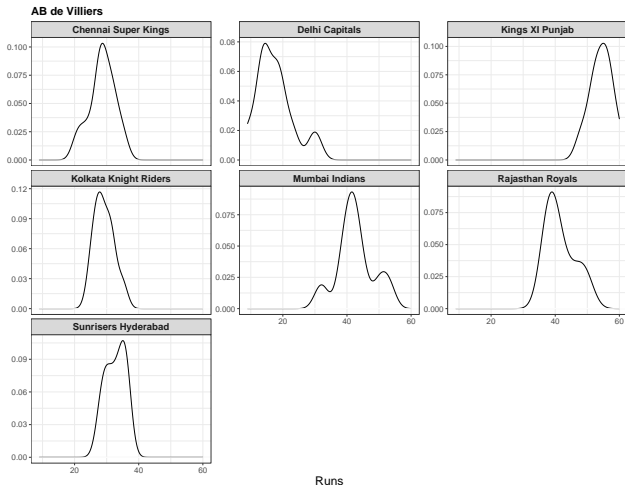
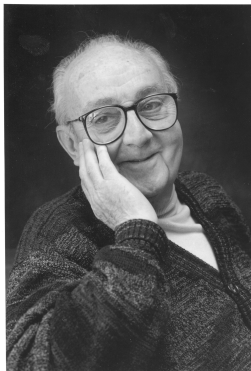


Figure: Distribution of predicted scores for IPL 2020 for RC Bangalore's batsman AB de Villiers. He was expected to score the least against the Delhi Capitals. He actually averaged only 22 against Delhi Capitals in IPL 2020 season.

Conclusion and Impact on the Game

- ▶ Predict probabilities of winning which could inform setting gambling odds
- ▶ This study could be applied to all league-format cricket tournaments
- ▶ The model could be implemented for other sports as well, such as to predict soccer outcomes
- ▶ Set/predict pregame strategy, in-game strategy, postgame strategy, and debutants' performances

Is this the *best* model?



- ▶ All models are wrong, but some are useful – George Box

Selected References

- ▶ Allsopp, P. E., & Clarke, S. R. (2004). Rating teams and analysing outcomes in one-day and test cricket. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167(4), 657-667.
- ▶ Baker, R., & Scarf, P. (2006). Predicting the outcomes of annual sporting contests. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(2), 225-239.
- ▶ Lohr, S. L. (2010). *Sampling design and analysis*, 2nd edition. Chapman & Hall/CRC Press.
- ▶ Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*, 82(398), 499-508.

Thank You

https://github.com/mquazi/cricket_2023

Predicting Cricket Outcomes using Bayesian Priors – Mohammed Quazi, Joshua Clifford, and Pavan Datta