3-17-2021

A probabilistic approach to identifying run scoring advantage in the order of playing cricket

Manar D. Samad
Tennessee State University

Sumen Sen
Austin Peay State University

Recommended Citation

Samad MD, Sen S. A probabilistic approach to identifying run scoring advantage in the order of playing cricket. International Journal of Sports Science & Coaching. 2021;16(4):1011-1020. doi:10.1177/17479541211000333

International Journal of
Sports Science
& Coaching

Original research

International Journal of Sports Science
& Coaching
2021, Vol. 16(4) 1011–1020

DOI: 10.1177/17479541211000333
journals.sagepub.com/home/spo
SAGE

A probabilistic approach to identifying run scoring advantage in the order of playing cricket

Manar D Samad[1] and Sumen Sen[2]

Abstract

In the game of cricket, the decision to bat first after winning the toss is often taken to make the best use of superior pitch conditions and set a big target for the opponent. However, the opponent may fail to show their natural batting performance in the second innings due to several factors, including deteriorated pitch conditions and excessive pressure of chasing a high target score. The advantage of batting first has been highlighted in the literature and expert opinions. However, the effect of batting and bowling order on match outcome has not been investigated well enough to recommend an adjustment of potential bias. This study proposes a probability-based model to study venue-specific scoring and chasing characteristics of teams with different match outcomes. A total of 1117 one-day international cricket matches held in ten popular venues are analyzed to show substantially high scoring likelihood when the winning team bat in the first innings. In a high scoring match, results suggest that the same bat-first winning team is very unlikely to score or chase the same high score if they bat in the second innings. We use the Bayesian rule to identify the bias in the scoring likelihood due to the playing order (bat-first versus bat-second). The bias is adjusted by revising the second innings target in a way that equalizes winning and run scoring likelihoods of both teams. The data and source code have been shared publicly for future research in creating competitive match outcomes by eliminating the advantage of batting order in run scoring.

Keywords

Choking, one-day international (ODI) match, performance analysis, sport analytics

Introduction

The game of cricket has more than two billion fans and followers over the world with 104 cricket playing nations. One-day international (ODI) and T-20 formats are the most popular versions of cricket that are played between teams in home-away series, world-cup tournaments, and domestically at first class tournaments and premium leagues. In the game of cricket, one team (team A) bat first (first innings) to score runs by competing against the bowling and fielding performance of the opponent team (team B). In ODI, eleven players (ten wickets) of team A bat in the first innings to score a total run facing 50 overs (300 balls, 6 balls per over) of bowling delivery of the opponent team B. Following a half-time break, team B bat in the second innings to chase the target score playing against 50 overs of bowling delivery of team A. Team A win, tie, or lose the match if team B score less, equal, or more than the first innings target score set by team A, respectively. The same strategy is followed in the T-20 version of the game, but each team is given only 20 overs (120 balls) to score or chase instead of 50 overs.

Unlike other sports such as soccer, hockey, and basketball, the game of cricket uniquely involves heavy accounting and diverse statistics to evaluate or predict

Reviewer: Oliver Stevenson (University of Auckland, New Zealand)

[1]Department of Computer Science, Tennessee State University, Nashville, TN, USA
[2]Department of Mathematics and Statistics, Austin Peay State University, Clarksville, TN, USA

Corresponding author:
Manar D Samad, Department of Computer Science, Tennessee State University, Nashville, TN, USA.
Email: msama005@odu.edu

the performance of individual teams or players. Statistical analyses may play a valuable role in determining an effective game strategy and team selection,[1] analyzing outcome in the second innings based on the target set in the first innings,[2] assessing the performance of players,[3] studying the order of eleven players during the batting session,[4] and predicting the outcome of a match.[5,6]

Therefore, historical data obtained from thousands of ODI cricket matches along with data-driven statistical methods can be valuable resources for coaching and operational management, performance evaluation, updating game rules, and forecasting results.[7] One of the most popular statistical models adopted in cricket is the Duckworth-Lewis (DL) method that determines a revised and fair target score when the game is interrupted, and the match duration is shortened due to the rain.[8] In the past, the event of rain not only postponed the match but also unfairly penalized one of the teams by cutting their allotted overs for batting. The DL method has been used for over two decades to set a reasonable target by statistically considering partial match results and resources (wickets and overs) available prior to the rain. This method has been an active subject of research and modification over the last two decades.[9,10]

Apart from rain interrupted events, there are numerous implicit sources of bias that may unfairly favor one team over the other.[11] For example, it is inevitable that home ground advantage is likely to favor the hosting team in a cricket match.[12] The innings played at night time in a day-night match has shown a considerable difference in outcome when compared to its counterpart innings played in the daytime.[13] One of the debated issues in cricket is the decision whether a team should bat or bowl first following the toss of a fair coin. This decision has been argued to give an advantage to the toss winning team. Sood and Willis have shown in a recent study that the winning of coin toss has a significant effect on winning the game, especially when the contesting teams have matching performance, and the match is played in certain conditions such as in the day-night format.[14] In general, a team choose to bat first expecting to set a high-scoring target using superior field conditions in the first half of the match. The pitch condition is expected to deteriorate over time, which may eventually turn less favorable for batting in the second innings. In contrast to this notion, the decision to bat first is also perturbed by a concern that a 'safe' score for confirming victory is unknown while batting in the first innings. Therefore, a team may choose to bowl first when they prefer to have a 'known' target score to chase expecting that field conditions or weather may rather turn unfavorable to bowlers in the second innings. Eventually, the playing order (bat-first versus bat-second) tends to favor one team over the other, which may not always give both teams a fair chance to score runs or win the match. Furthermore, data from the last four world cup cricket tournaments suggest that the teams winning the toss and deciding to bowl first are victorious in less than 50% of similar cases.[15] This phenomenon reveals certain advantages in batting first over bowling apparently due to several confounding factors. In support of this observation, Dawson et al. have concluded in their study that winning the toss and batting first significantly increases the chance of winning the match compared to the decision of bowling first after winning the toss.[16]

Proposed research

In line with the above observations, we identify two cases that may significantly bias the outcome of the game due to the batting or bowling order. First, it is widely known that chasing a target score of 300 and above runs in the second innings is often more challenging than scoring such a high total in the first innings. For example, it becomes even more challenging to chase such a target in the fourth innings of a test cricket match. There are only nine matches in over forty years of world cup cricket history where the bat-second teams have been successful in chasing a target score of 300 and above runs. In contrast, there are at least 19 matches in the 2019 world cup alone where the first innings scores are more than 300 runs. Therefore, the decision to bat first and consequently setting a big target can lower the probability of successful chasing in the second innings even with two equally competitive teams. Second, the first innings batting experience is free of pressure from chasing a target score, which is a psychological advantage. The team batting in the second innings incur this additional pressure, which is considered proportional to the target score. The pressure of large target tempts the bat-second team to take additional risks that may negatively affect their natural batting performance. Even highly competitive teams have been found vulnerable to such high target pressure as their batting performance often collapses after scoring unusually low total while chasing a high target,[17] which is often termed as 'choking' or 'strangling' and has been studied by Lemmer.[18] All these phenomena do not guarantee a level playing field for both teams since the playing order may ultimately influence the outcome of the game. These issues can negatively affect the excitement, competitiveness, and spirit of the game in general.

All these statistics raise several research questions that demand data-driven solutions.

- First, is there any run scoring advantage in the first innings batting that might affect the outcome of the game?

- Second, what would be the team's score if they bat in the second innings given the fact that they have scored a" hard-to-chase" total in the first innings?
- Third, what would be a reasonable and competitive target that accounts for the advantage of setting a hard-to-chase score in the first innings?

Since there is no alternative to coin tossing, we have used probability theory on venue-specific ODI match results to investigate bias in run scoring distributions at different playing conditions and match outcomes. A statistical model has been proposed to capture venue-specific scoring bias for two purposes: 1) identifying the magnitude of bias in run scoring likelihood due to the playing order and 2) recommending a revised second-innings target that will equalize the winning and scoring probability for both teams regardless of the playing order.

Methodology

This study proposes a probability-based approach for investigating potential bias in run scoring likelihood due to the playing order in cricket. First, the run scoring distributions of four-match cases are obtained and then compared within each of ten ODI venues. The four match cases are: 1) bat-first-win, 2) bat-first-lose, 3) bat-second-win, 4) bat-second-lose. The run distributions are modeled using the negative binomial (NB) distribution since it effectively captures run distributions in Scarf et al.[19] Additionally, we compare the results of NB distribution with those of logistic and normal distributions. The NB distribution function modeling a discrete random variable x is shown below

$$P(x, n, p) = \frac{\Gamma(x+n)}{\Gamma(n)x!} p^n (1-p)^x \tag{1}$$

The parameters of NB distribution (n and p) are obtained using maximum likelihood estimates to yield the probability mass function (PMF). For comparison, probability density functions (PDF) of contribution variable distributions are developed using the mean and variance of the scores. The PMF or PDF represents the scoring probability distribution using random variable X, P(X). Figure 1(a) shows the PDF of a normal distribution. Considering a discrete random variable X, the probability of scoring exactly Xt runs P(X = Xt) (e.g. 237 runs) is low when the discrete sample space is large typically ranging from 100 runs to 350 runs. Intuitively, the likelihood of scoring 200 and more runs in an innings is much higher than that of scoring 300 and more runs. This leads to the definition of cumulative PDF that represents the probability of scoring up to Xt runs, P (X ≤Xt) by integrating or summing the PDF or PMF from zero to Xt, respectively as shown in Figure 1(b). We take the complement of the CDF in equation (1) to model the run scoring likelihood such that the likelihood of scoring at least 200 runs P(X > 200) is higher than that of scoring at least 300 runs, P(X > 300), as shown in Figure 1(c)

$$P(X > Xt) = 1 - P(X \leq Xt) = 1 - \sum_{x=0}^{x=xt} P(x, n, p) \tag{2}$$

where, P(X,n,p) is the best fitted PMF on the data. We use the complement of CDF in subsequent comparisons among four match cases to investigate bias in the likelihood of scoring runs.

Analysis of run scoring distribution

To adjust the bias in scoring likelihood, we use the Bayesian rule to recommend a revised target score that will equalize the likelihood of scoring and winning for both teams irrespective of the playing order. First, the variables and outcomes are identified for one-day international (ODI) cricket matches. The match outcome can be either win (W) or lose (L). Runs scored is represented by the random variable S. First and second innings batting conditions are represented by

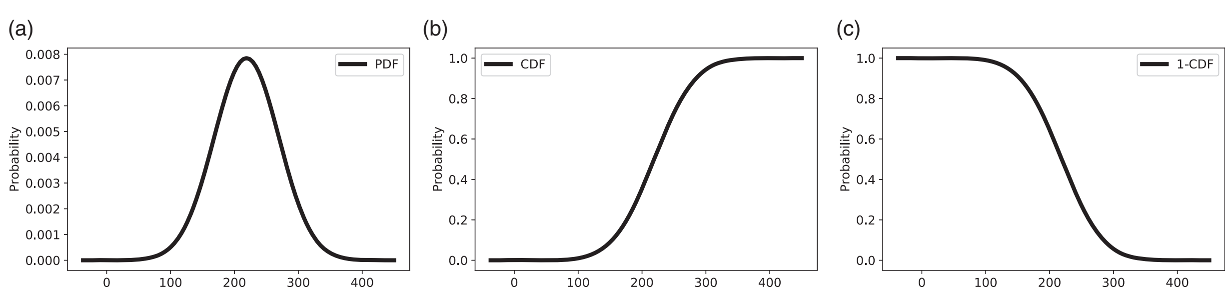

Figure 1. (a) Probability density function, *P*(*x*), (b) cumulative density function (CDF) of *P*(*X*), and (c) complement of the CDF of runs scored to represent the likelihood of scoring above *X* runs.

BF and BS, respectively. We define the posterior probability of winning a match given that the team bat first and score at least Xf runs in the first innings as P (W|S >Xf, BF). Using the Bayes' rule, this posterior winning probability can be calculated as below

$$P(W|S > Xf, BF) = \frac{P(S > Xf, BF|W)\ P(W)}{P(S > Xf, BF)} \tag{3}$$

Similarly, the posterior probability of winning the match given that the winning team bat in the second innings and score Xs runs is as follows

$$P(W|S > Xs, BS) = \frac{P(S > Xs, BS|W)\ P(W)}{P(S > Xs, BS)} \tag{4}$$

Given a first innings score of at least Xf runs, the minimum second innings target score Xs that will equalize the winning probability of any team in a particular venue is obtained by equalizing equations (3) and (4) as below

$$\frac{P(S > Xs, BS|W)}{P(S > Xs, BS)} = \frac{P(S > Xf, BF|W)}{P(S > Xf, BF)} \tag{5}$$

Applying the chain rule of conditional probability

$$\frac{P(S > Xs|BS, W)\ P(BS|W)}{P(S > Xs, BS)} = \frac{P(S > Xf|BF, W)\ P(BF|W)}{P(S > Xf, BF)} \tag{6}$$

Joint probability distributions in the denominator can be expressed in terms of conditional probability distributions as follows

$$\frac{P(S > Xs|BS, W)\ P(BS|W)}{P(S > Xs|BS)\ P(BS)} = \frac{P(S > Xf|BF, W)\ P(BF|W)}{P(S > Xf|BF)\ P(BF)} \tag{7}$$

$$P(S > Xs|BS, W) = \frac{P(S > Xs|BS)}{P(S > Xf|BF)} \frac{P(BF|W)}{P(BS|W)}\ P(S > Xf|BF, W) \tag{8}$$

Here, the probability of batting first or second is equal, P(BF) = P(BS), considering the coin is fair. Given the first innings score Xf, we assume that the revised second innings score Xs will additionally equalize the run scoring likelihood (in addition to the winning probability) of both innings such that P (S >Xf | BF)= P (S >Xs | BS). The revised equation is as follows

$$P(S > Xs|BS, W) = C * P(S > Xf|BF, W) \tag{9}$$

Here, $C = \frac{P(Bf|W)}{P(BS|W)}$ is a constant ratio for a venue, which is the ratio of probabilities of batting first and second given that the team is victorious. When the likelihood of winning team batting in the first innings is higher, the ratio will be greater than 1. Given the first innings score Xf, the second innings score Xs that satisfies the two conditions can be obtained by taking the inverse of the CDF, P(S ≤ Xs|BS,W) as below

$$\begin{aligned} Xs &= \text{Inv}\ (P(S \le Xs|BS, W)) \\ &= \text{Inv}\ (1 - C * P(S > Xf|BF, W)) \end{aligned} \tag{10}$$

Results and discussion

We have performed the development and analysis of the proposed model in Python programming environment using the scipy, NumPy, and pandas packages[20,21] and shared the source code, data, and notebook in a GitHub repository (https://github.com/mdsamad001/CricketStudy). The data set includes scores and outcomes of 1117 ODI matches and is obtained from the webpage of cricket-stats (http://cricket-stats.net/genp/grounds.shtml) for the ten most popular international venues.

Table 1 presents a summary of the data set, including mean scores and the percentage of matches won in different playing orders and outcomes. The mean score is not informative for comparison being a single point in the probability distribution. We hypothesize that an informative scoring bias can be obtained by comparing the run distributions of two match cases instead of their mean values. Furthermore, the winning percentage of a match case (e.g., bat-first) alone may not reveal the complex conditioning among the scoring likelihood, playing order, and winning probability. We propose that run distributions conditioned on winning probability will yield better insights into the scoring advantage of the playing order.

Analysis of run scoring distributions

The probability distribution of runs scored at each venue is studied by fitting the negative binomial distribution. The NB distribution is one of the popular choices for modeling count data like runs in cricket. The overall distribution of runs scored across all

Table 1. Summary of the one-day international cricket match data set used in this study.[a]

Venue	Year	Total matches	Bat-first-win %	Avg. score	Bat-second-lose Avg. score	Bat-second-win %	Avg. score	Bat-first-lose Avg. score
Auckland	1976–2020	71	42.3	240	185	57.7	200	199
Bangalore	1974–2018	22	50.0	294	248	50.0	236	234
Harare	1992–2020	149	49.7	255	183	50.3	205	203
Lahore	1978–2015	58	56.9	266	205	43.1	233	231
Lords	1972–2019	62	48.4	268	215	51.6	218	216
Melbourne	1971–2019	145	49.7	245	191	50.3	202	201
Mirpur	2006–2018	107	46.7	261	194	53.3	203	201
Premadasa	1986–2019	118	58.5	266	196	41.5	204	202
Sharjah	1984–2019	236	53.8	252	189	46.2	195	192
Sydney	1979–2020	149	59.1	248	189	40.9	195	194
Overall		1117	51.5	260	200	48.5	203	201

[a]Average runs are rounded off to the next integer value. Scores of bat-second-win teams and their opponent bat-first-lose teams are adjusted for revised targets set by the Duckworth–Lewis method.

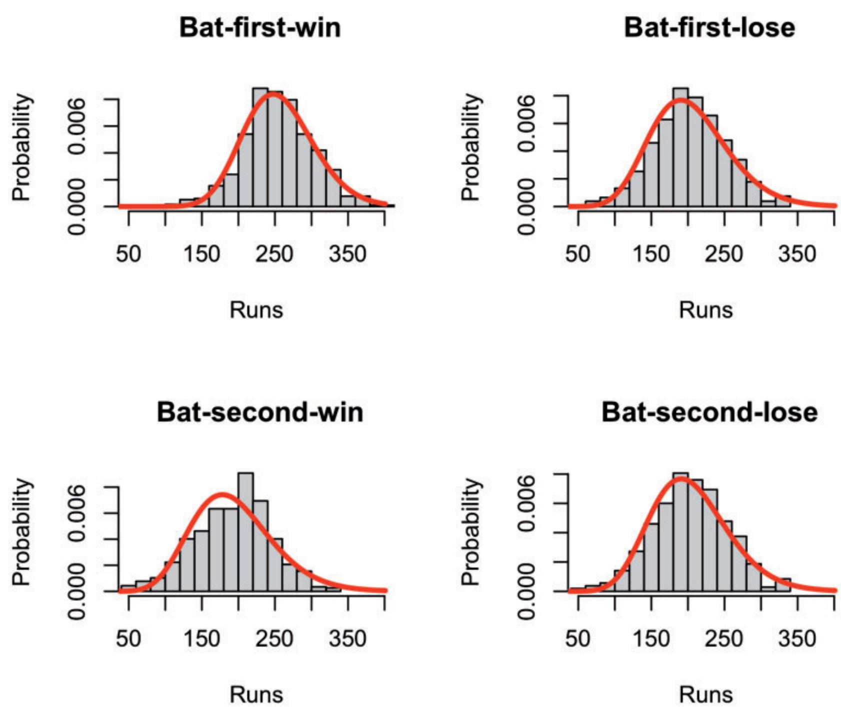

Figure 2. Fitting runs of four experimental cases using the negative binomial distribution.

venues is also modeled by the NB distribution. The effect of playing order (batting and bowling order) on scoring distribution is analyzed by categorizing the distributions into four match cases: 1) bat-first-win, 2) bat-second-lose, 3) bat-first-lose, and 4) bat-second-win. Figure 2 shows scoring distributions of four match cases after fitting the NB distribution using all venue data. This distribution fit shows that venue-specific scores can follow the NB distribution. We compare run scoring likelihoods between relevant match cases as follows.

Bat-first-lose versus bat-second-win

Figure 3 shows that the run distributions of bat-first-lose and bat-second-win are similar and overlapping for all venues. This is intuitive since the bat-second-win team will always score one or several runs more than the bat-first-lose team in a successful run chase. Table 1 shows this trend in the last two columns. Including match outcomes from all ten venues, average scores of bat-second-win and bat-first-lose are 203 and 201 runs, respectively (Table 1). In line with this

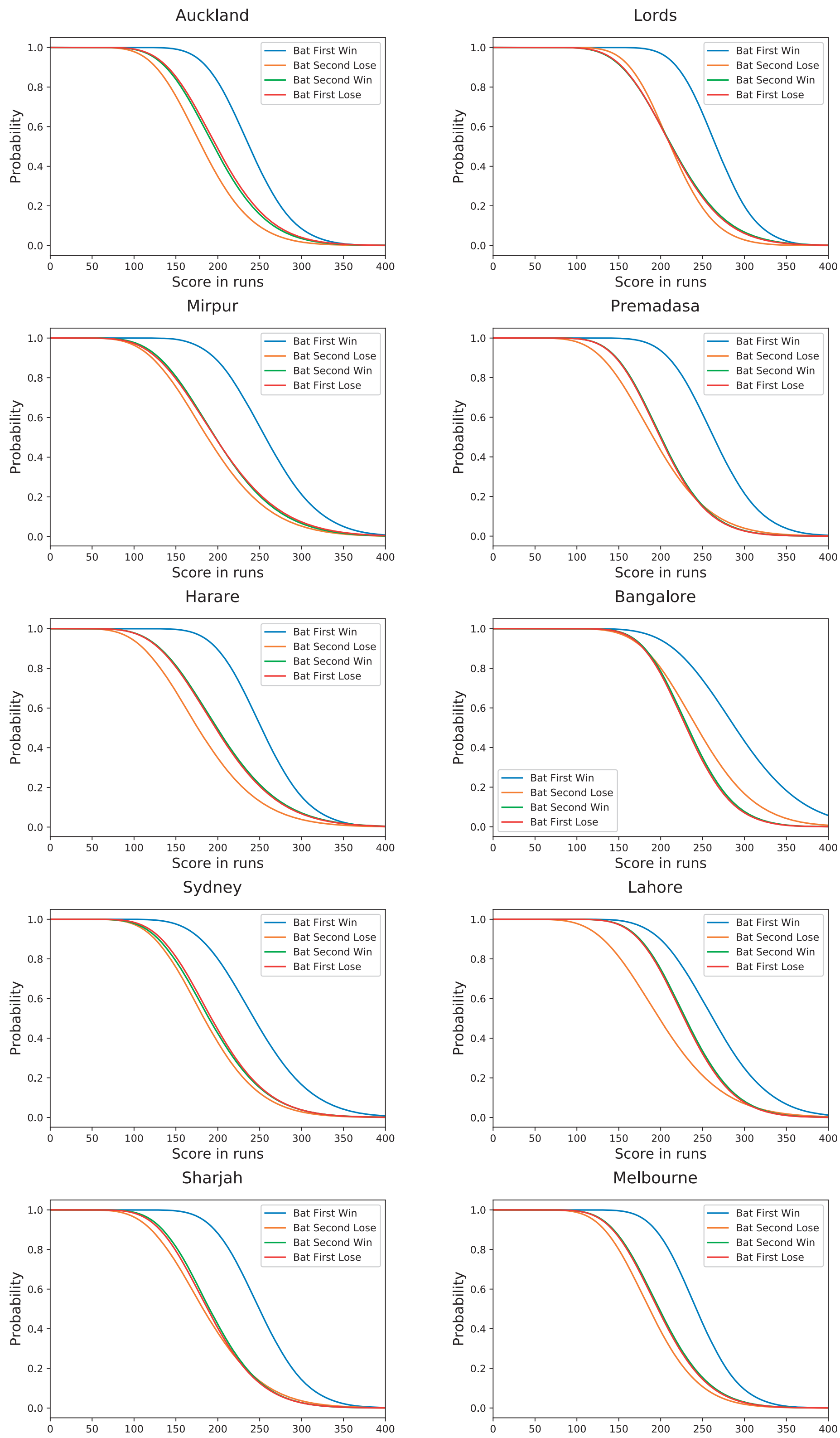

Figure 3. Run scoring distribution of four cases related to two playing orders (Bat first, Bat Second) and two outcomes (win, lose) at 10 specific ODI venues. The complement of cumulative negative binomial distribution is fitted using actual score data. All venues show a large discrepancy between the scoring likelihoods of 'bat-first-win' teams and their opponent 'bat-second-lose' teams. The discrepancy in scoring likelihood is most prominent at two venues: Harare and Lahore.

observation, the difference in their scoring likelihoods appears justifiably insignificant in Figure 3.

Bat-first-lose versus bat-second-lose

In contrast, the scoring likelihood of bat-second-lose teams falls behind that of bat-first-lose teams at almost all venues (Figure 3). That is, the first innings batting yields a higher scoring likelihood than that of the second innings batting even when both cases lose the match. This discrepancy is indicative of a scoring bias or advantage possibly due to the playing order. This advantage in run scoring is more evident at venues like Premadasa, Harare, Lahore and appears lowest at Sharjah, Sydney, and Lords. The only exception is the venue at Bangalore where the bat-second-lose scoring likelihood appears better than that of the bat-first-lose case. This observation may be attributed to the lowest sample size for Bangalore (only 22 ODI matches have to be grouped into four match cases) compared to other venues (Table 1).

Bat-first-win versus bat-second-lose

The discrepancy in scoring likelihoods appears to be the largest when the bat-first team win the match against the bat-second-lose team. Among all four match cases, Figure 3 reveals that the bat-second-lose and their opponent bat-first-win teams have the worst and the best scoring performance, respectively. The scoring likelihood of bat-first-win teams is far ahead of those of the other three cases for all venues (Figure 3). This result infers that the same high-scoring team in the first innings is unlikely to score so high in the second innings. This inference suggests a scoring disadvantage for the bat-second team while chasing a large target set by the bat-first team.

It is noteworthy that these discrepancies are venue-specific under the assumption that stronger and weaker teams have a similar likelihood of winning the toss or batting in the first innings. Figure 4 shows the overall scoring likelihood of four playing order-outcomes obtained using all ten-venue data (1117 ODI match outcomes). This overall scoring likelihood reveals discrepancies similar to those found in most of the venue-specific trends. Therefore, Figure 4 may serve as a general scoring likelihood trend of ODI cricket match outcomes after accounting for all venues, game strategies, and contexts. Furthermore, the discrepancy in run scoring likelihood can be even more striking in specific contexts, especially when the stronger side gets the chance to bat first and sets a large and unattainable target for the opponent. Therefore, it is worth asking the question if the same bat-first team could score the same unattainable score in the second innings. Scoring likelihoods of different match cases may inform us of the total runs that the bat-first-win team would have scored if they bat in the second innings. The difference in scoring likelihood due to the batting order is expected to reveal a bias and subsequently recommends an adjusted target for the second innings.

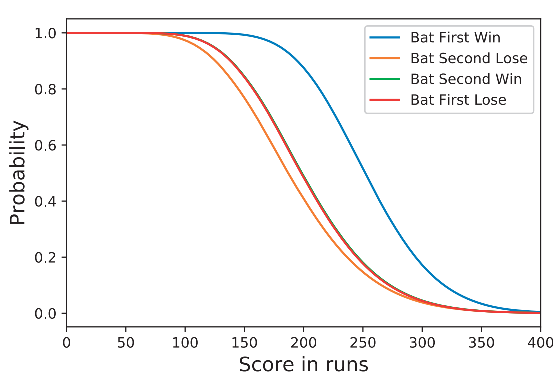

Figure 4. Overall scoring likelihoods for four playing order and outcome cases obtained using all ten-venue data (1117 ODI match outcomes). This scoring distribution accounts for all venues and game strategies.

Probabilistic model for revising target

The previous section reveals a higher scoring advantage of bat-first-win teams compared to that of any other three match cases. The high-scoring advantage can yield a target score that becomes challenging to chase while batting in the second innings. For example, the world cup record of chasing the highest score is 329 runs whereas the first innings score has been as high as 397 runs. There are at least eight matches in the 2019 cricket world cup alone with over 330 runs scored in the first innings, which would require the opponents to break the world cup record to win those matches. Data suggest that there are only a handful of cases in world cup cricket when a score of 300 runs and above has been successfully chased in the second innings. A simple breakdown of 1117 ODI match outcomes in Table 2 reveals that bat-second teams are only 10%

Table 2. First innings scores and percentage of matches successful in chasing the first innings score (bat-second-win cases).

First innings score	≥150	≥180	≥200	≥220	≥240	≥260	≥280	≥300	≥320	≥340
Matches successfully chased	44%	40%	35%	30%	25%	21%	17%	10%	13%	0.0%

successful in chasing 300 and above runs. Notably, our 1117 ODI data include match results from 1971 to 2020, and matches played several decades ago used to yield lower scores than what we see today.

Therefore, after observing a very high score in the first innings, the game is often assumed to be over even before playing the second innings of the match. Commentary is often made highlighting that there has been no record of chasing such a high score at the particular venue. Therefore, the likelihood of chasing such a high score in the second innings remains extremely low. These observations go against 'the game of glorious uncertainties' entitlement of cricket.

One naive approach to tackle this bias is to equalize the scoring likelihood for both teams, especially in a big scoring match. This equalization may be done by adjusting an unusually big target score set in the first innings considering a venue-specific or overall probabilistic model. However, finding the cut-off score for the adjustment is not a trivial task since such a cut-off may change from decade to decade and from venue to venue. Instead, our proposed model provides adjustment of any second innings targets using equation (10) that equalizes both the winning and scoring likelihoods of the competing teams. The extent of adjustment varies across different first innings scores after considering both winning and scoring likelihoods in our model. In Table 3, we demonstrate the estimated bias in the first innings scores for individual venues, which is then used to revise the target scores. The selection of 300 and above runs is only to demonstrate the bias and score adjustment for high scoring matches. Similar bias and adjustment can be obtained for low scoring matches since the bat-second scoring distribution typically falls behind that of bat-first scoring for a wide range of scores. For example, the model will revise first innings scores of 200, 250, and 280 runs in Auckland to 183, 226, and 259 runs, respectively. Additionally, revised scores obtained from the normal and logistic probability distributions are not too different from those obtained using the negative binomial distribution.

The mean run difference between the actual target (first innings score) and the model revised target score (see Table 3) is also a measure of bias for each venue. Results suggest that the venue in Auckland has the lowest mean difference and the one in Premadasa suffers the highest mean difference for all three distribution models. The negative binomial distribution has yielded the lowest mean differences in scoring among the three distributions because the NB distribution is known to best-fit run scoring data. The last row of Table 3 shows the overall model results obtained after fitting all 1117 match outcomes from all ten venues. In all cases, the revised target score for the second innings team is lower than that of the first innings score to compensate for the high-scoring advantage in the first innings.

Effects of playing order in cricket

The findings presented in previous sections require cautious interpretation. First, the findings do not conclude that batting in the first innings is generally preferable or advantageous over batting in the second innings. A team can score a decent total in the first innings and can still lose the match. Therefore, the target score adjustment may be recommended in proportion to the first innings score to alleviate the scoring advantage

Table 3. Revised second innings target scores against first-innings scores (actual score).[a]

Actual target	300	315	330	340	350	Mean difference in runs		
Venue	Negative binomial distribution					Negative binomial	Normal	Logistic
Auckland	283	301	320	332	345	10.8	12.6	1.2
Bangalore	241	251	261	268	275	67.0	65.8	65.4
Harare	276	289	312	327	342	19.0	31.0	30.4
Lahore	251	266	280	289	298	50.0	46.6	56.6
Lords	263	284	306	321	335	24.0	29.0	26.8
Melbourne	267	286	304	317	329	26.2	32.8	32.6
Mirpur	255	273	291	303	316	39.4	42.2	36.8
Premadasa	225	243	260	271	282	71.0	72.6	84.0
Sharjah	242	259	277	289	301	53.4	56.4	61.6
Sydney	230	247	263	274	285	67.2	68.6	82.8
Overall model	249	266	284	295	307	46.8	49.0	52.2

[a]The revised scores are shown using the negative binomial distribution. The mean difference between actual and revised scores is shown for three distributions: negative binomial, normal, and logistic. The overall model (probabilistic) includes data from all 1117 ODI matches and is not a mere aggregation of 10 venue results. The 300+ runs cut-off is chosen for the demonstration of adjustment of high scores.

of the bat-first teams. Second, a common strategy of the team bowling in the first innings is to restrict the bat-first innings score to a chasable target. This strategy implies that above a certain score (often venue-specific), the target becomes less and less chasable for any team as reflected in our scoring likelihood data. However, such a high score is relatively more achievable in the first innings as we know from historical data, which is also supported by our scoring likelihood analysis. Third, the pressure from high target scores tempts the bat-second teams to take additional stress and risks, which the bat-first teams do not incur during their batting performance. As a result, the second innings batting can be vulnerable to collapse, which has been studied as the choking effect in the literature.[18] Our analysis reveals the worst run scoring likelihood of the bat-second-lose teams, which is possibly due to the choking effect and underperformance of the team under pressure. This particular scenario of the game (not all 1117 matches satisfy this scenario) may not give a level playing field for both contesting teams. Therefore, an adjusted target may account for the additional pressure of batting in the second innings of high-scoring matches.

Limitations

The goal of this study is to investigate run scoring advantages in the playing order of cricket. The proposed probabilistic model makes two assumptions to determine the revised score that mitigates the bias present in the first innings score. The model equalizes the winning and scoring likelihood of both contesting teams. The assumption of equal winning probability alone, when the scoring probability is considered unequal ($P(S > Xf \mid BF) \neq P(S > Xs \mid BS)$), complicates the solution and does not yield a robust numerical solution for the revised score. A more conservative estimate of the revised target (closer to the first innings actual target score) may be obtained by relaxing one of the two assumptions in the proposed model. While our probabilistic model reveals bias in run scoring likelihoods, the method for adjusting targets has limitations. It will require further research to measure and consider the choking effect in adjusting the second innings target. Notably, not all target scores will onset choking, which would require an estimation of an adjustment factor that is proportional to the first innings score. Several venues have a comparatively low sample size, which is further reduced due to the grouping of samples into four match cases. Therefore, venue-specific models may not be reliable when the sample size is low. The proposed model is not directly applicable to new venues unless past data are considered from all other existing venues in the model development. To ensure a big sample size, match scores and outcomes over the past fifty years of ODI history have been included in the proposed models. However, the scoring trends and likelihoods have changed significantly over several decades as they change from venue to venue. Therefore, the proposed model will require modification focusing on the current scoring trends in cricket. Furthermore, there are no gold standards and benchmark data to evaluate the performance of our proposed model because of its empirical nature.

Conclusions

This study has investigated run scoring likelihood of teams playing ODI cricket matches under varying playing orders and at different venues. Our result from fifty years of match data reveals a scoring advantage of bat-first teams regardless of venues and strength of the playing teams. The high scoring likelihood of all bat-first-win cases infers that the bat-first-win team is very unlikely to yield the same high score if they are sent to bat in the second innings. Our proposed model captures gaps in venue-specific run scoring likelihood for different match scenarios and recommends revised target scores for the second innings. The revised target score is obtained to ensure equal winning and scoring probability at a particular venue. Despite limitations in numerical computations, we believe that this is one of the first studies to investigate run scoring advantage in the order of playing cricket with a preliminary solution. The proposed model can be used to investigate ordering bias in other game design and operations to subsequently recommend an adjustment for ensuring a fair outcome.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Manar D Samad https://orcid.org/0000-0002-6263-6261

References

1. Amin GR and Kumar Sharma S. Cricket team selection using data envelopment analysis. *Eur J Sport Sci* 2014; 14: S369–S376.
2. Bose D and Chakraborty S. Managing in-play run chases in limited overs cricket using optimized CUSUM charts. *JSA* 2019; 5: 335–346.

3. Akhtar S, Scarf P and Rasool Z. Rating players in test match cricket. *J Oper Res Soc* 2015; 66: 684–695.
4. Swartz TB, Gill PS, Beaudoin D, et al. Optimal batting orders in one-day cricket. *Comput Oper Res* 2006; 33: 1939–1950.
5. Akhtar S and Scarf P. Forecasting test cricket match outcomes in play. *Int J Forecasting* 2012; 28: 632–643.
6. Asif M and McHale I. A generalized non-linear forecasting model for limited overs international cricket. *Int J Forecasting* 2019; 35: 634–640.
7. Mondin L, Weber C, Clark S, et al. Statistical analysis of diagnostic accuracy with applications to cricket. *Involve* 2012; 5: 349–359.
8. Duckworth FC and Lewis AJ. A fair method for resetting the target in interrupted one-day cricket matches. *J Oper Res Soc* 1998; 49: 220–227.
9. McHale IG and Asif M. A modified Duckworth–Lewis method for adjusting targets in interrupted limited overs cricket. *Eur J Oper Res* 2013; 225: 353–362.
10. Preston Thomas J. Rain rules for limited overs cricket and probabilities of victory. *J R Stat Soc Ser D* 2002; 51: 189–202.
11. Forrest D and Dorsey R. Effect of toss and weather on County Cricket Championship outcomes. *J Sports Sci* 2008; 26: 3–13.
12. Morley B and Thomas D. An investigation of home advantage and other factors affecting outcomes in English one-day cricket matches. *J Sports Sci* 2005; 23: 261–268.
13. McGinn E. The effect of batting during the evening in cricket. *J Quant Anal Sports* 2013; 9: 141–150.
14. Sood G and Willis D. Fairly Random: The Impact of Winning the Toss on the Probability of Winning. Retrieved from http://arxiv.org/abs/1605.08753v1 (2016, accessed 4 July 2020).
15. Dalal V. The first innings conundrum in world cup, 27 June 2019, https://www.livemint.com/sports/cricket-news/.the-first-innings-conundrum-in-world-cup/2019-1561614800105.html (accessed 7 May 2020).
16. Dawson P, Morley B, Paton D, et al. To bat or not to bat: an examination of match outcomes in day-night limited overs cricket. *J Oper Res Soc* 2009; 60: 1786–1793.
17. Bhattacharjee D and Lemmer HH. Quantifying the pressure on the teams batting or bowling in the second innings of limited overs cricket matches. *Int J Sports Sci Coaching* 2016; 11: 683–692.
18. Lemmer HH. A method to measure strangling, a dramatic form of choking in cricket. *Int J Sports Sci Coaching* 2015; 10: 717–728.
19. Scarf P, Shi X and Akhtar S. On the distribution of runs scored and batting strategy in test cricket. *J R Stat Soc Ser A* 2011; 174: 471–497.
20. McKinney W. Pandas: a foundational Python library for data analysis and statistics. In: PyHPC 2011:Python for high performance and scientific computing; 2011 Nov 18; Seattle, WA USA. pp.1–9.
21. Oliphant TE, Oliphant TE, Oliphant T, et al. Python for scientific computing. *Comput Sci Eng* 2007; 9: 10–20.