# Probabilistic Modeling of the Game of Cricket

Varun Kishore Kumar, Anand Madhavan, David Johnson

### The Problem

This project analyzes the game of cricket using a probabilistic model. Cricket is a game, similar to baseball, played between two teams of 11 players each. Each team bats once and has 10 outs (called "wickets"), one out per batsman. The pitcher (called "bowler") throws a series of 6 balls (called an "over") and can bowl a maximum of 10 overs. The batting team tries to score as many runs as possible in 50 overs. The score of the first team that bats becomes the target score for the second team. If the second team beats this score, they win.

### Questions

We seek to answer some interesting and important questions through our probabilistic model and through sensitivity analysis.

- At any point in the second half of the game, what is the probability that the team that is batting loses the game and how sensitive is this to the number of wickets in hand? How sensitive is this to the number of balls remaining?
- Often rain and weather disrupt games and officials are forced to shorten the game in order to ensure an outcome. A target score for the shortened number of overs is determined by the officials. The prevalent D/L method used to do this, does not use probability and is a point of some contention. We use a probabilistic approach and seek to answer, how much should the target score be lowered so the second team's probability of winning does not change?
- At any point in the second half of a game, how does the quality of batsmen and bowlers remaining affect the probability of losing?

### Formulation

Our model uses the idea that a team has various resources available to them and as the game progresses those resources get used up. The quality and amount of the resources remaining affect a team's probability distribution of the number of runs the second team is likely to score in the rest of the game.

At any given ball and with a certain number of runs remaining, we compute the probability of the second team losing, using the probabilities of various outcomes during that ball. Possible outcomes include a wicket falling, no runs being scored, 1 run being scored and so on. These in turn are conditional probabilities which we refer prior work in this topic and use a parameterized model that uses a standard normal distribution with coefficients computed from historic data to model this.

We then use these in a set of dynamic equations to find the probability of the second team losing. We implement the series of dynamic equations as a three dimensional table with the number of balls remaining on one axis, the number of runs required on another axis, and the number of wickets remaining on the third axis.

We create another model to take into account the effect of the quality of batsmen and bowlers remaining. This takes as input the batsmen and bowlers that remain and using the quality of these players we adjust the number of balls remaining and the number wickets remaining to compute an adjusted probability of the second team losing.

For example, if there were 117 balls and 3 wickets remaining. A team with their top 3 batsmen remaining would have the same probability of winning as a team where quality of batsmen is not considered and 159 balls remain. Since the top batsmen are likely to score more on average, they would need less balls in order to score the same number of runs. Using an adjusted balls and wickets, we employ our three dimensional table, to determine the probability of the second team winning.

### Analysis

We show through charts and tables, how our probabilistic model answers some of the questions we set out to answer. We show how it is a viable and fairer way of determining target scores during rain situations. We show the sensitivity of quality of players remaining, to the probability of winning or losing.