Pitch Prediction: Statistical Learning

Christian Stratton, Andrew Hoegh, Jennifer Green
Montana State University
Department of Mathematical Sciences

January 13, 2020

Abstract

Keywords:

1 Introduction

Analytics have long played a role in Major League Baseball (MLB) and knowing the type of pitch that will be thrown can give a hitter a major advantage; however, in recent years, teams have used illegal methods to video record, decipher pitch signals, and relay them to hitters. This manuscript explores the use of statistical modeling to predict the type of pitch thrown in various scenarios.

On January 13, 2020 MLB imposed one of the largest penalties in history on the Houston Astros for a scandal that involved video recording hitters and banging on trash cans to relay signals. The Astros were fined 5 million dollars, the maximum fine allowed in MLB; stripped of first and second round draft picks for multiple years; and the manager and general manager (GM) were suspended for one year. The manager and GM were ultimately fired by the organization.

The Boston Red Sox and ... have also been fined for using methods to video tape and relay signals to hitters.

In recent years, MLB changed the rules to permit the use of technology in the dugout as teams are now allowed to use league-provided iPad/laptops.

In line with historical advances in analytics in baseball and the sabremetrics movement, we seek to explore using statistical models to predict the next pitch thrown depending on the count and other scenarios. Section 2 describes the data used in this analysis. Section 3 highlights the statistical models used for prediction as well as the loss functions used to evaluate different models. Section 4 describes the model results and Section 5 concludes with a discussion.

2 Data

The data used for this

3 Statistical Framework

3.1 Loss Functions

3.2 Model Specification

4 Results

5 Discussion