

STATS 601 Project Proposal:

Caught Looking: Analyzing Variation in Umpire Strike Zones

SIMON FONTAINE ^{1,*}, MORITZ KORTE-STAPFF ^{1,**} and BRIAN MANZO ^{1,†}

¹University of Michigan, Department of Statistics. West Hall, 1085 South University, Ann Arbor, MI, U.S.A., 48109.

E-mail: ^{*}simfont@umich.edu; ^{**}kortest@umich.edu; [†]bmanzo@umich.edu

[Brian] other title option: From Blue to Deep Blue? Assessing the case for robot umpires

1. Introduction

As technology advances, Major League Baseball (MLB) has faced increased pressure from fans, coaches, and players to use video technologies to aid umpires in making calls on the field. In particular, ball and strike calls are notoriously subjective and challenging for a human to make.¹ With this project, we will assess the ability of umpires to make ball and strike calls that match the rulebook and that are consistent across different game situations. The team members for this project are Simon Fontaine, Moritz Korte-Stapff, and Brian Manzo, all team members are first year Ph.D. students in the Department of Statistics.

2. The dataset

PITCHf/x is a system that uses cameras mounted in every MLB stadium to track the speed and location of every pitched baseball. We will access csv files of the PITCHf/x data for the entire 2018 season from Kaggle ([Schale, 2019](#)) and supplement those files with other data from MLB such as player heights. The most important part of the PITCHf/x data for our purposes is the set of coordinates that gives the location of every ball that crosses the plate, but it has many other details about the game and players involved that will enable us to assess the ability of umpires to stay consistent across a variety of situations.

[Simon] maybe put this into the abstract part since it seems unrelated to the rest of the paragraph

[Simon] don't need to be *that* specific

3. Research questions

The first question we are looking to answer is whether umpires' realized strike zones match the requirements prescribed by the official rules of baseball. Once we have strike

[Simon] maybe say pre-processed or something like this since the original data is not from there

[Simon] that data?

[Simon] second clause basically repeats what

¹<https://www.youtube.com/watch?v=Akx2kBavZ9Y>

zone densities for each umpire, we can use simulation to assess their “error rate”—the percentage of pitches they would classify as a ball or strike that, according to the rules of baseball, should have the opposite classification. As time and space permit, we also are looking to answer the following:

- Do umpires’ strike zones change when there are 2 strikes in the count? What about 3 balls?
- Do umpires have different strike zones for star pitchers and batters than they do for mediocre players?
- Do umpires change their strike zone when a game goes into extra innings?

The general theme of each all of these questions is: how good are umpires at calling balls and strikes? The worse the umpires perform in game situations, the louder will be the calls for technology-aided officiating.

4. Methodology

Estimating strike zones from pitch location data is fundamentally a nonlinear classification problem. Given pitches that have an x and y coordinate, we want to classify the pitches in a rectangle in the middle of the grid as strikes, and anything too far away from the middle (outside the rectangle) as balls. Some methods that we think would succeed in drawing a realistic decision boundary are kernel logistic regression, additive logistic regression, support vector machines, and tree-based methods. Once we have found a technique that we feel is giving a fair representation of the umpires’ strike zones based solely on the pitch location, we can train classifiers that take into account game-level and player-level information (i.e., game situation or player skill). The density estimates can be discretized a 2d grid (say 100x100 so we get a vector of size 10000) and then we can run some sort of KernelPCA on it to reduce the dimension. Then finally compare if the reduced strike zone varies between umpires and according to different game situations with some (multivariate) mixed effects model. These methods, which we have studied in 600 and 601, will allow us to construct the strike zone densities, reduce dimension, and then do inference to determine whether strike zones are inconsistent across game situations and different umpires.

References

SCHALE, P. (2019). MLB Pitch Data 2015-2018. Kaggle Dataset (<https://www.kaggle.com/pschale/mlb-pitch-data-20152018>).

[Simon] I think we need a better word than density. Maybe estimated strike zones.

[Simon] I changed - to —, but I don’t if that’s the English way, that might be the French way.

[Simon] maybe be more vague like “...to answer secondary questions such as”

[Simon] ...and what affects this assessment?

[Simon] +N-N

[Simon] seems a weak statement: maybe say Upon selecting the best method modeling the umpires’ SZ

[Simon] I don’t think they will be classifiers.

[Simon] again: estimated SZ?

[Simon] I think we should