

3 BATTED BALL OUTCOME MODEL

In the first chapter, we discussed where wins come from. Wins come from runs. In the second chapter, we discussed where runs come from. Runs come from plate appearance outcomes (walks, singles, etc.). In this chapter (you guessed it!) we will discuss where plate appearance outcomes come from.

3.1 TRACKING DATA

In 2009, a company called Sportvision introduced a product to MLB called HITf/x. This product was a camera system that used computer vision to track the baseball after it was hit by the batter. The implications were huge. So often, a batter will hit a weak ground ball that luckily finds space between two fielders (reflecting luck, not a repeatable skill). “That’s a hit in the scorebook,” announcers would say. Or a batter will hit a line drive right at a fielder. The announcers are right—in the scorebook, we have no way of distinguishing between different types of singles or different types of flyouts. With batted ball tracking, now we do more to distinguish between skill and luck on batted balls.

The California-based Sportvision was usurped ca. 2015 by a Danish company called TrackMan¹ whose technology was based on Doppler radar. TrackMan was usurped ca. 2020 by a London-based company called Hawk-Eye,² who reverted us to computer vision. For the most part, the data we observe for each batted ball are the exit speed, the launch angle and the spray angle. The exit speed is the speed of the ball as it leaves the bat. The launch angle is the initial vertical angle of the ball trajectory off the bat. The spray angle is the horizontal angle of the initial trajectory. MLB teams have access to additional data (ball spin, hang time, distance, etc.), but none of those data are publicly available.³

3.2 OUTCOME MODEL

Suppose we observe a dataset of n batted balls indexed by $i \in \{1, \dots, n\}$. For each batted ball, we observe the batter b_i , the exit speed x_{i1} , the launch angle x_{i2} , the spray angle x_{i3} , and the outcome o_i . We use y_i to denote the linear weight of the outcome o_i ; in the notation of the previous chapter, $y_i = \ell(o_i)$. We use Y_i to denote the random variable of which y_i is an observation. We will present three different probability models for Y_i , starting with Model #3 (for notational reasons that will become clear).

$$\text{Model \#3: } Y_i \sim \text{Normal}(f_3(x_{i1}, x_{i2}, x_{i3}), \sigma^2)$$

Model #3 is a regression model. To fit the model, the task is estimating the function f_3 . There are many methods for estimating f_3 , one of which is linear regression, which finds the best function of the form:

$$f_3(x_{i1}, x_{i2}, x_{i3}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3}.$$

It turns out that strict linear regression is not well suited for this problem because (a) there are strongly nonlinear effects, especially for launch angle and (b) there are strong interactions between the covariates (e.g. exit speed and launch angle). We want a more flexible model, and there are many ways to achieve this. There is a whole subset of machine learning called *supervised regression* that is well suited for this problem, but it is outside the scope of this course.⁴

¹TrackMan got their start in golf. Many sports tech companies start in golf because enthusiasts often have disposable income.

²Hawk-Eye got their start in tennis. They were the ones responsible for the in/out replay challenges you’ve seen on TV.

³Incidentally, the fact that even these data are publicly available is a happy accident. MLB left an API intended for media companies exposed to the public, and they decided to leave it as is after it was discovered by the public.

⁴You can learn about this in STAT 413.

3.3 PLAYER EVALUATION

Jumping ahead, suppose we have (one way or another) come up with an estimate \hat{f}_3 of f_3 . Let's now introduce two competing metrics for batter evaluation.

$$\begin{aligned} \text{LW}(b) &= \sum_{i=1}^n \mathbb{I}(b_i = b) \ell(o_i) \\ \text{xLW}_3(b) &= \sum_{i=1}^n \mathbb{I}(b_i = b) \hat{f}_3(x_{i1}, x_{i2}, x_{i3}) \end{aligned}$$

LW we have previously covered. xLW stands for expected linear weight. The x* prefix is a convention in sabermetrics for indicating that we are using batted ball trajectories rather than actual outcomes. The subscript 3 indicates that we are using Model #3.

Which do we prefer between LW and xLW₃? This question is analogous to the one we faced when choosing between RE24 and LW. The metric xLW₃ introduces a restrictive assumption, that the batter only controls their LW through their batted ball trajectories. This assumption introduces bias into our batter evaluation, but the upshot is that it reduces variance. This is an example of the famous bias-variance tradeoff from statistical machine learning.⁵ The short of it is that the optimal estimate of batter skill will lie somewhere between the extremes of high variance and high bias, and it depends on your sample size. The smaller the sample, the more bias you want. The larger your sample, the less bias you want. For tiny samples, we prefer xLW₃. For infinite samples, we prefer LW. Note that the decision between RE24 and LW was also an example of the bias-variance tradeoff (LW is higher bias, lower variance relative to RE24).

Because we are facing the same problem as the previous two chapters, we use the same tools to solve it. Suppose we have batters $j \in \{1, \dots, p\}$, each with n_j plate appearances. Let the random variable Z_j denote the per-PA residual $(\text{LW}(j) - \text{xLW}_3(j))/n_j$ for batter j . Our model for Z_j is:

$$\begin{aligned} Z_j &\sim \text{Normal}(\eta_j, \sigma_Z^2/n_j) \\ \eta_j &\sim \text{Normal}(0, \sigma_\eta^2). \end{aligned}$$

As the sample size increases, when do we switch from preferring xLW₃ to preferring LW? When $n > \sigma_Z^2/\sigma_\eta^2$ (see Chapter 1). What is our mean-regressed estimate of the “true talent” residual η_j ? From Chapter 2:

$$\frac{n_j/\sigma_Z^2 \cdot z_j}{n_j/\sigma_Z^2 + 1/\sigma_\eta^2}.$$

3.4 ALTERNATIVE OUTCOME MODELS

Let's keep going down this path of reducing variance. First, we introduce two alternatives to Model #3:

$$\begin{aligned} \text{Model \#2: } Y_i &\sim \text{Normal}(f_2(x_{i1}, x_{i2}), \sigma^2) \\ \text{Model \#1: } Y_i &\sim \text{Normal}(f_1(x_{i1}), \sigma^2) \end{aligned}$$

Model #2 drops spray angle and only considers exit speed and launch angle. Model #1 drops launch angle and only considers exit speed. Why would we do this? When we remove a covariate from the model, we increase the bias of the model and reduce its variance. Now we have a sequence of batter evaluation metrics with decreasing levels of bias and increasing levels of variance:

$$\text{xLW}_1 \rightarrow \text{xLW}_2 \rightarrow \text{xLW}_3 \rightarrow \text{LW} \rightarrow \text{RE24}.$$

How do we choose which metric to use?

⁵See STAT 413.

4 DISCUSSION QUESTIONS

1. Find the Baseball Savant player batting leaderboard. Who is leading the league in xwOBA? Who has the biggest gap between their wOBA and their xwOBA?
2. What is the relationship between xwOBA (from Baseball Savant) and xLW3 (from the lecture notes)?
3. Why might a batter's true-talent LW on batted balls be higher than their true-talent xLW3?
4. Suppose that the true talent of the residual $LW - xLW3$ for MLB batters is normally distributed with mean zero and standard deviation 0.02. Suppose further that the noise variance of observed performance is $0.2 / n$, where n is the number of plate appearances observed. If a batter accumulates 20 xLW3 and 60 LW over 500 plate appearances, what is the posterior mean of their "true talent" residual per plate appearance $(LW - xLW3) / n$?
5. Which do you think is the best metric to evaluate batters: xLW3, xLW2 or xLW1?