

Bayesian Modeling of Hitting

Daveed Goldenberg, Cristien Wright

December 11, 2019

Modeling Home Run Hitting In Baseball

Using home run rates from 2008 to 2017 we aim to predict the course of a hitter's career in 2018, and answer:

- ▶ How do we balance aging with projecting future hitting?
- ▶ How many seasons of above average hitting do young players need to be considered elite?
- ▶ What is the affect of a bad year on a previously consistent hitting career?

Why This Matters

- ▶ Teams are willing to spend serious money in order to get the best players:
- ▶ Mike Trout 12 years \$430 million
- ▶ Miguel Cabrera 8 years \$248 million
- ▶ Yoenis Cespedes 4 years \$110 million
- ▶ Giancarlo Stanton 13 years \$325 million

Predictors

- ▶ Y_{ij} : home run total for player i in year j
- ▶ M_{ij} : number of at bats for player i in year j
- ▶ A_{ij} : age
- ▶ B_{ij} : home ballpark
- ▶ R_{ij} : position

Logistic Regression Model

$$Y_{ij} \overset{iid}{\sim} \text{Binomial}(M_{ij}, \theta_{ij})$$

θ_{ij} is the home run rate for player i in year j

$$\log\left(\frac{\theta_{ij}}{1 - \theta_{ij}}\right) = \alpha R_{ij} + \beta B_{ij} + f(A_{ij}, R_{ij})$$

$f(A_{ij}, R_{ij})$ is a smoothing function for age based on position, we used cubic B-splines with coefficients γ for each position.

$$\alpha = \begin{cases} \alpha_o & E_{ij} = 0 \\ \alpha_1 & E_{ij} = 1 \end{cases}$$

where $\alpha_{ko} < \alpha_{k1}$, $k = 1, \dots, 9$

E_{ij} is an indicator variable for whether player i is determined to be an elite in year j , this is redetermined each year.

Elite Indicator

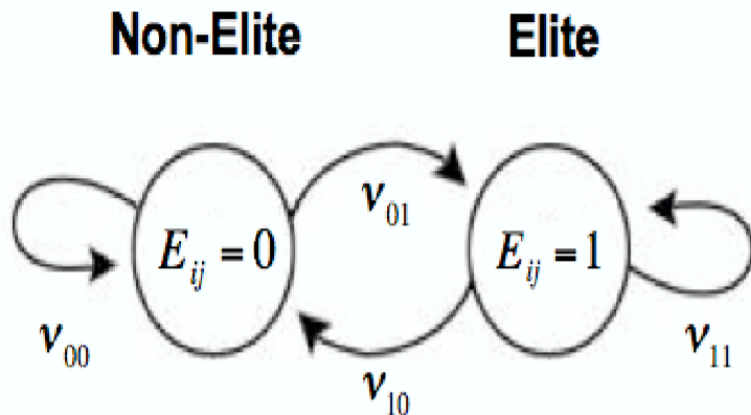


Figure 1: Elite Status

Elite Indicator

$$Pr(E_{i,j+1} = b | E_{ij} = a, R_{ij} = k) = \nu_{abk}$$

► ν_{oo}

Priors

Position Intercepts:

$$\alpha_k \sim MVN(\mathbf{0}, \tau^2 \mathbf{I}_2) * I(\alpha_{k0} < \alpha_{k1}), k = 1, \dots, 9$$

Home Ball Park / Team Intercepts:

$$\beta \sim MVN(0, \tau^2 \mathbf{I})$$

Spline Coefficients:

$$\gamma_{kp} \sim Normal(0, \tau^2), k = 1, \dots, 9; p = 1, 2, 3, 4$$

MCMC Implementation

Future Work

- ▶ Add age as a factor for E_{ij} , as players age they are less likely to maintain elite status
- ▶ Model with a multinomial response to include more than just home runs. Include things like on base percentage, doubles, singles