

Bayesian Analysis of Player Performance over Time

Nathaniel Brown

April 10, 2018

Introduction

Objective

- ▶ Investigate Bayesian modelling techniques on shooting data, and to learn more about time-dependency in the data.

Previous Studies

- ▶ Gilovich, Vallone, & Tversky (1985)
- ▶ Wetzels (2016)
- ▶ Bar-Eli, Avugos, & Raab (2006)

Data

Description of Dataset

- ▶ Player-tracking data provided by the Duke Men's Basketball team.
- ▶ Recorded using SportVU, a player-tracking system from STATS, LLC.
- ▶ **Final Sequence Play-by-Play Optical:** contains information of players' basketball actions (e.g., shot makes and attempts, dribbles, passes, fouls) and their time stamps for every game.
- ▶ **Final Sequence Optical:** contains precise locations for all 10 players and the ball at a rate of 25 times per second for every game.

Missing Data

- ▶ The ability to record this data depends on specialized tracking cameras, and not every arena has the technology installed.
- ▶ Has data for 94 out of the 147 games played between the 2013-2014 and 2016-2017 seasons (82 at Home and 12 Away).

Data Cleaning

- ▶ Translate the locations to a half-court setting.
- ▶ Convert the x-y coordinates (feet) to polar coordinates (feet and radians).
- ▶ Add an indicator for home games.
- ▶ Only use the shots that Duke players attempt (no opposing players appear in more than 5 games).

Data Cleaning (cont.)

Table 1: Summary of Dataset

Name	Type	Values	Extra Details
season	categorical	{2014, ..., 2017}	
gameid	categorical	NA	94 unique values
time	continuous	NA	13-digit timestamp in milliseconds
globalplayerid	categorical	NA	31 unique values
r	continuous	$[0, \infty)$	Distance of shot from hoop (feet)
theta	continuous	$[-\pi, \pi]$	Angle of shot (radians)
home	categorical	{0,1}	1 if shot occurred during a home game
result	categorical	{0,1}	1 if shot was made(response)

Table 2: Sample of Dataset

season	gameid	time	globalplayerid	r	theta	home	result
2014	201401070173	1389141733839	603106	4.2076	1.0746	1	1
2014	201401070173	1389141844712	601140	16.6537	1.2973	1	0
2014	201401070173	1389143172185	696289	18.7901	-0.0581	1	1
2014	201401070173	1389143196303	601140	23.4629	0.9539	1	1
2014	201401070173	1389143220261	756880	6.5365	0.0696	1	0

Data Cleaning (cont.)

Distribution of Shot Locations

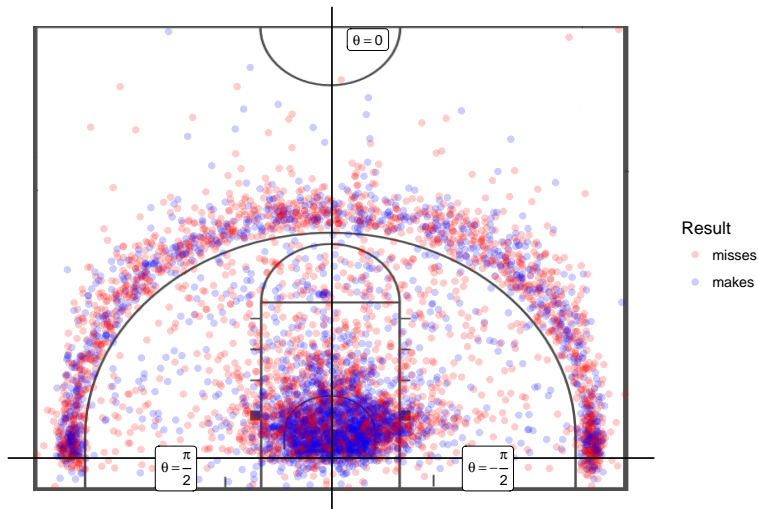


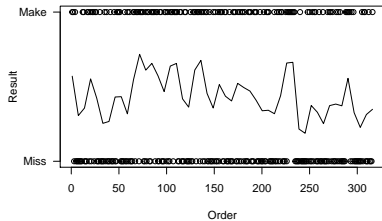
Figure 1: Locations and Results of All Shots

Exploratory Data Analysis

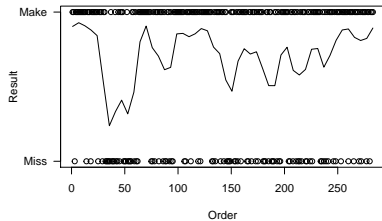
- ▶ Visualize changes in success rate over time in individuals and in population.
- ▶ Moving average only accounts for shot order and outcome.

Exploratory Data Analysis (cont.)

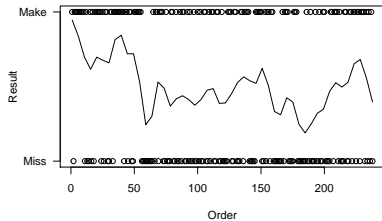
Player 1



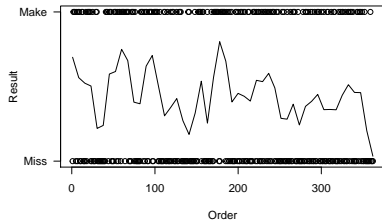
Player 2



Player 3



Player 4



Exploratory Data Analysis (cont.)

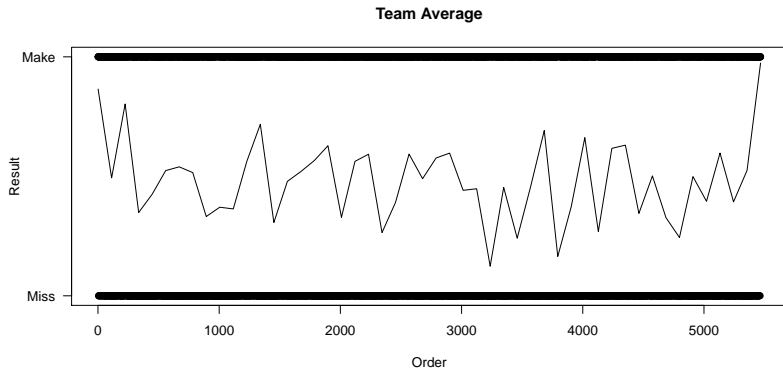


Figure 2: Moving Average of Shot Success Rate

Models & Analysis

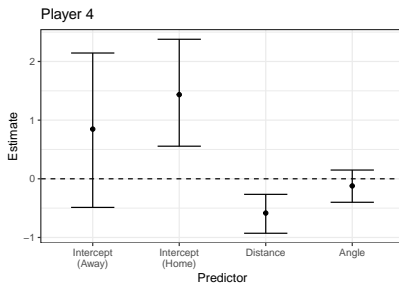
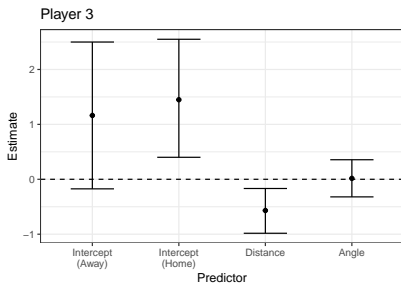
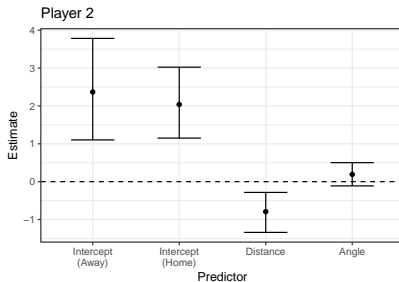
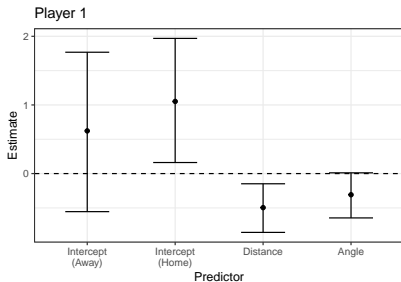
Model-Building

- ▶ Types of Models
- ▶ Generalized Linear Model
- ▶ Hierarchical Generalized Linear Model
- ▶ Discounted Likelihood Hierarchical Model
- ▶ All models based off a logistic regression model
- ▶ Built using JAGS library in R (R2jags)

Generalized Linear Model: Notation

$$\text{logit}(p_i) = \beta_{\text{int}} + x_{r,i}\beta_r + x_{\theta,i}\beta_{\theta} + x_{H,i}\beta_H$$

Generalized Linear Model: Results



Generalized Linear Model: Results (cont.)

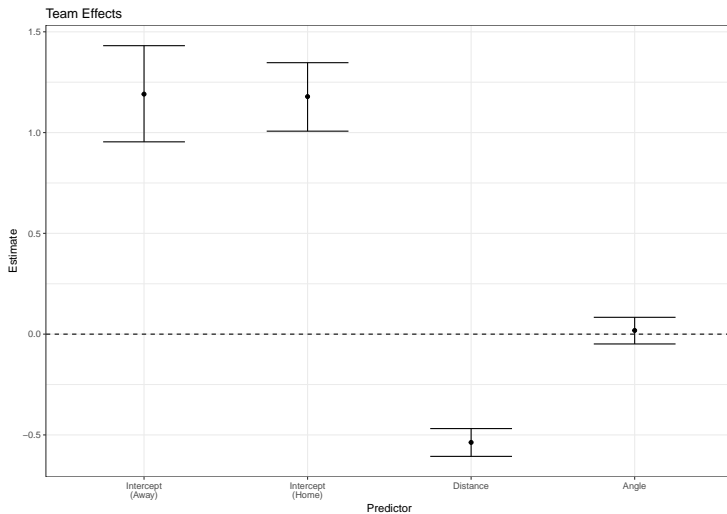


Figure 3: GLM Posterior Distributions for Four Players

Hierarchical Model: Notation

$$\text{logit}(p_{ji}) = \beta_{\text{int}, j} + x_{r,ji}\beta_{r, j} + x_{\theta,ji}\beta_{\theta,j} + x_{H, ji}\beta_{H, j},$$

$$\beta_{\text{int}, j} \sim N(\beta_{\text{int}}, \tau_{\text{int}}^2),$$

$$\beta_{r, j} \sim N(\beta_r, \tau_r^2),$$

$$\beta_{\theta,j} \sim N(\beta_{\theta}, \tau_{\theta}^2),$$

$$\beta_{H, j} \sim N(\beta_H, \tau_H^2).$$

Hierarchical Model: Results

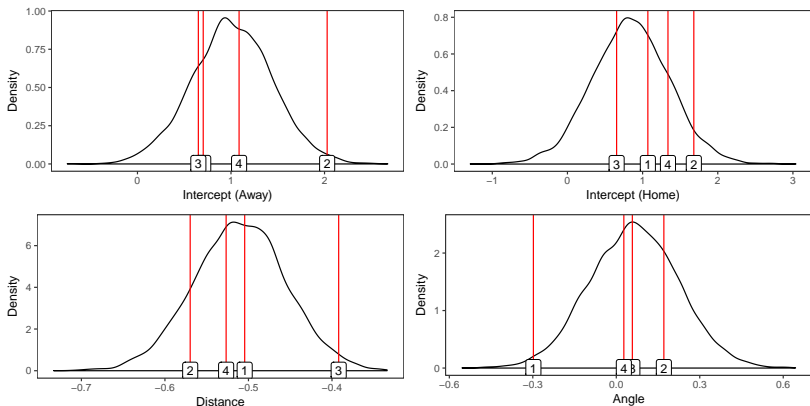
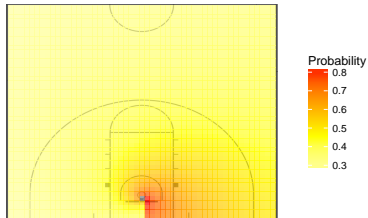


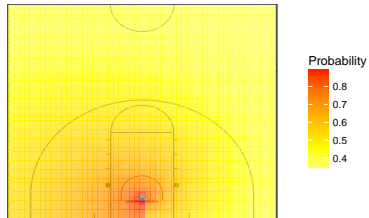
Figure 4: Population Distribution with Four Player Effects

Hierarchical Model: Contour Plots

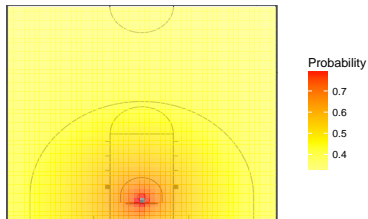
Player 1



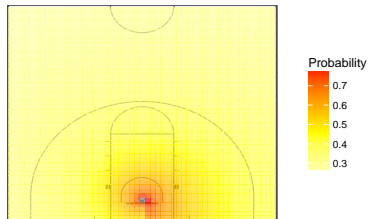
Player 2



Player 3



Player 4



Hierarchical Model: Contour Plots (cont.)

Team Effect

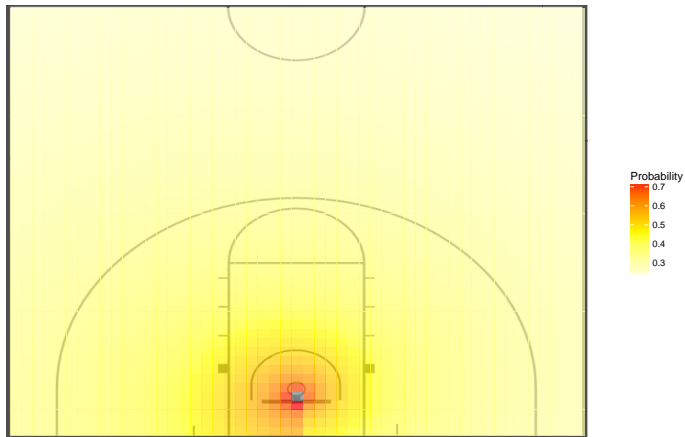


Figure 5: Contour Plots for Four Players and Population of Players

Discounted Model: Notation

Hierarchical model equation:

$$\text{logit}(p_{ji}) = \beta_{\text{int}, j} + x_{r, ji} \beta_{r, j} + x_{\theta, ji} \beta_{\theta, j} + x_{H, ji} \beta_{H, j}.$$

Binomial likelihood term without discounting:

$$L_{gj}(\Theta) = \prod_{i=1}^{n_{gj}} p(y_{gji} | \Theta) \propto \prod_{i=1}^{n_{gj}} p_{gji}^{y_{gji}} (1 - p_{gji})^{1-y_{gji}}.$$

Exponential discounting to the outcomes for $0 < \delta < 1$:

$$\pi_{gji} = \left(p_{gji}^{y_{gji}} (1 - p_{gji})^{1-y_{gji}} \right)^{\delta^{|g-g_0|}},$$

Discounted Likelihood:

$$\Lambda_{gj}(\Theta) = \prod_{i=1}^{n_{gj}} \pi_{gji},$$

Discounted Model: Ones Trick

```
for(i in 1:N){  
  
  # delta = discount rate  
  wt[i] <- delta^abs(games[i]-g0)  
  
  # model equation with random effects by player  
  logit(prob[i]) <-  
    beta_int[player[i]]*int[i] +  
    beta_home[player[i]]*home[i] +  
    beta_r[player[i]]*logr[i] +  
    beta_theta[player[i]]*theta[i]  
  
  # likelihood function  
  p1[i] <- prob[i]^result[i]  
  p2[i] <- (1-prob[i])^(1-result[i])  
  
  # discounted likelihood function  
  pi[i] <- (p1[i] * p2[i])^wt[i]  
  
  # defines correct discounted likelihood function  
  y[i] ~ dbern(pi[i])  
}
```


Discounted Model: Weighting

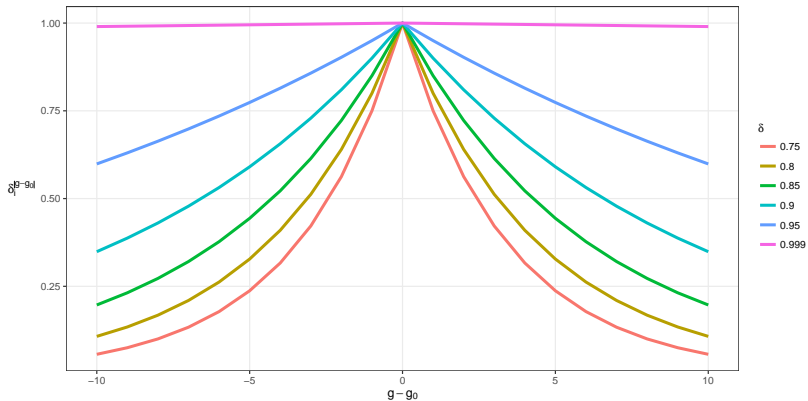


Figure 6: Illustration of Discounted Weighting

Discounted Model: Results

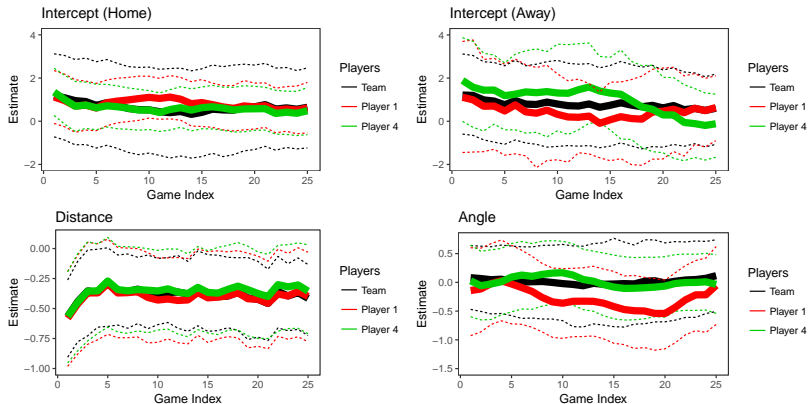


Figure 7: Parameters for Two Players and Population over Time, $\delta = 0.750$

Discounted Model: Results (cont.)

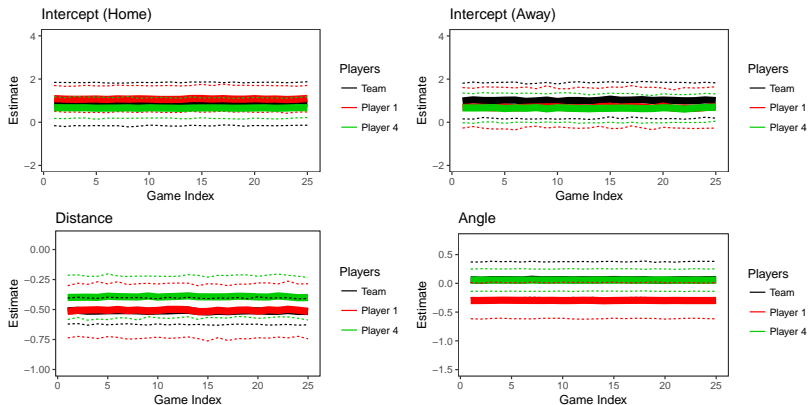


Figure 8: Parameters for Two Players and Population over Time, $\delta = 0.999$

Discussion

Evaluation of Models

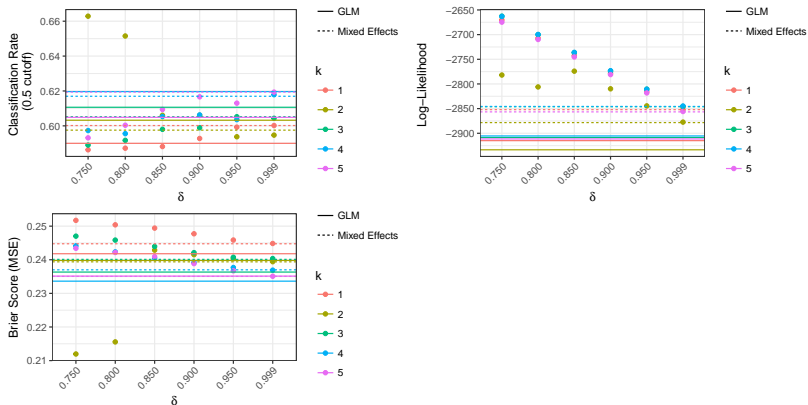


Figure 9: Model Evaluation

Calibration Plots

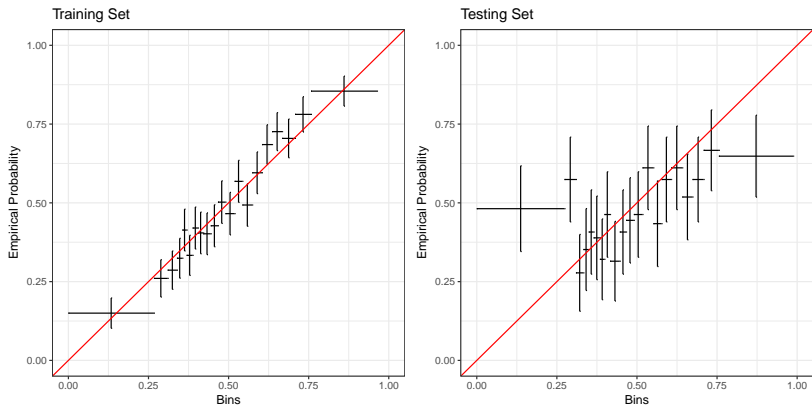


Figure 10: Calibration Plots for Discounted Likelihood Model, $\delta = 0.850$

Results from Model with $\delta = 0.850$

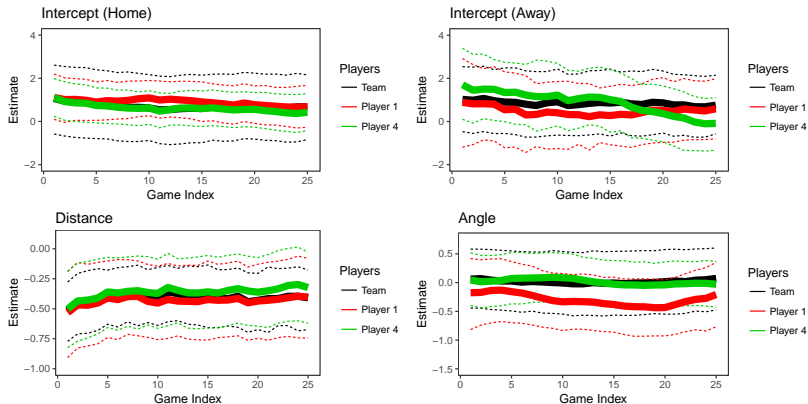


Figure 11: Parameters for Two Players and Population over Time, $\delta = 0.850$

Conclusion

- ▶ Some weak evidence for time-dependency in shooting success rate.
- ▶ Angle only matters for certain players.
- ▶ Effects of Home-court advantage are not strong in this dataset.

Future Goals

- ▶ Incorporate more advanced predictors.
- ▶ Add random effects on discount factor.

Acknowledgments

- ▶ Mike West, for advising me through this project.
- ▶ Merlise Clyde, for introducing me to the ones trick in JAGS.
- ▶ My thesis committee of Cliburn Chan, Merlise Clyde, and Mike West.
- ▶ Kevin Cullen, Basketball Director of Information Technology for the Duke Men's Basketball team, for providing data for this research.