Bayesian Analysis of Player Performance over Time

Nathaniel Brown

April 10, 2018

Objective

- Investigate Bayesian modelling techniques on shooting data, and to learn more about time-dependency and the “hot hand” concept in sports.

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.

Data Cleaning (cont.)

Table 1: Summary of Dataset

Name	Type	Values	Extra Details
season	categorical	$\{2014, \ldots, 2017\}$	
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 occured 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.)

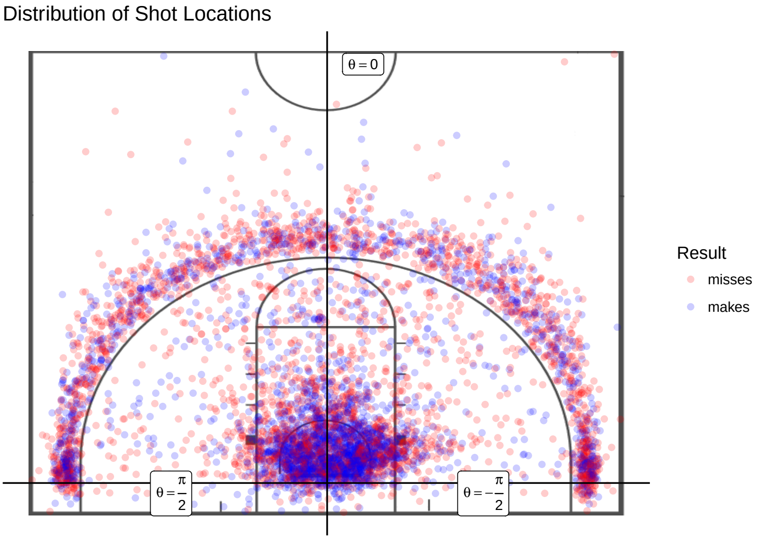
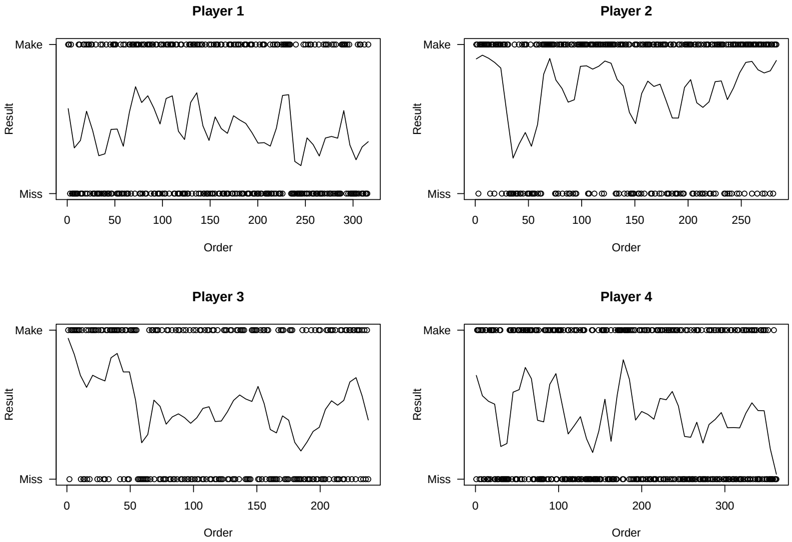


Figure 1: Locations and Results of All Shots

Exploratory Data Analysis



Exploratory Data Analysis (cont.)

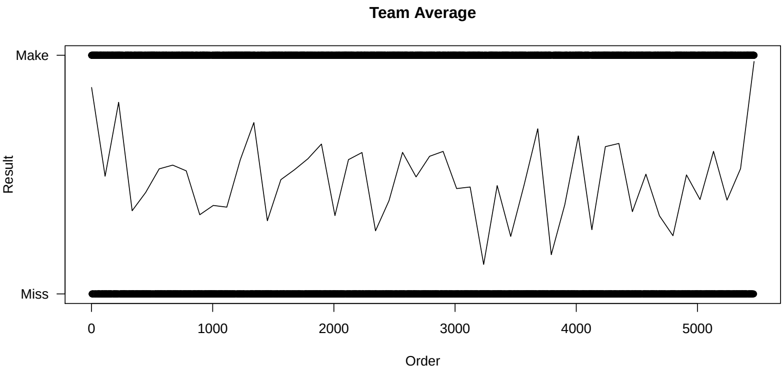


Figure 2: Loess Smoothing Curve on Shot Success Rate

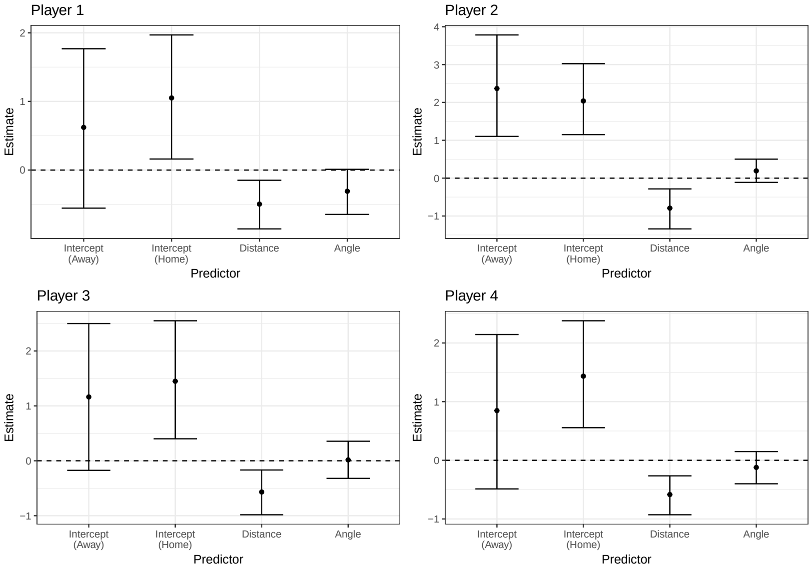
Model-Building

- 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

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

Generalized Linear Model: Results



Generalized Linear Model: Results (cont.)

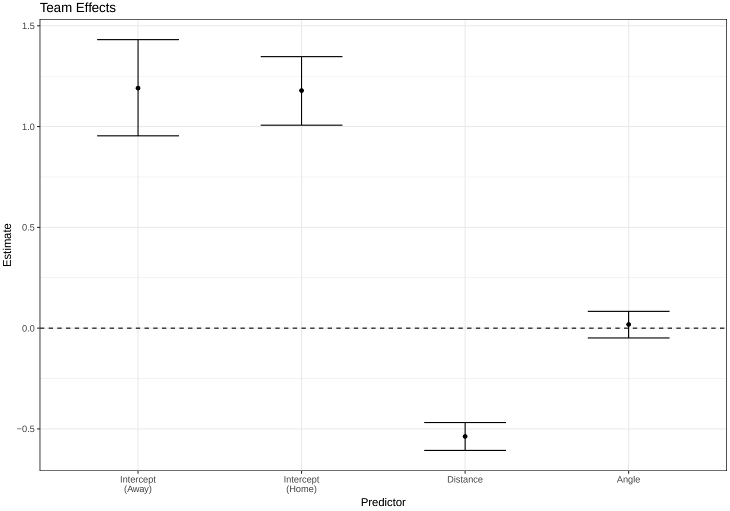


Figure 3: Posterior Distributionsof GLM Parameters

Hierarchical Model: Notation

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

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

$$\beta_{\text{r, j}} \sim N(\beta_{\text{r}}, \tau^2_{\text{r}}),$$

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

$$\beta_{\text{H, j}} \sim N(\beta_{\text{H}}, \tau^2_{\text{H}}).$$

Hierarchical Model: Results

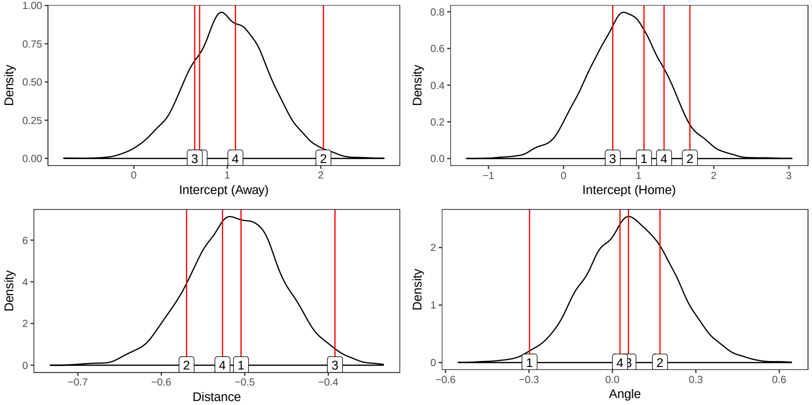
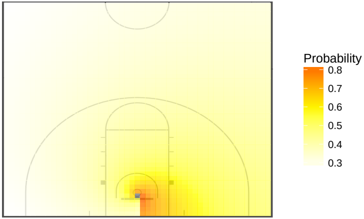


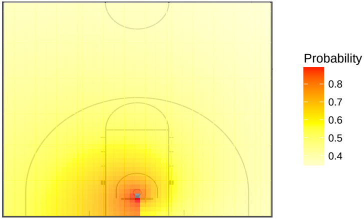
Figure 4: Median Player Effects over Population Distribution

Hierarchical Model: Contour Plots

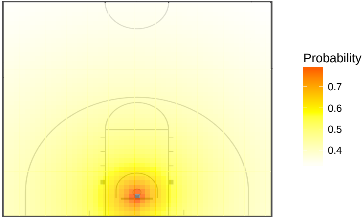
Player 1



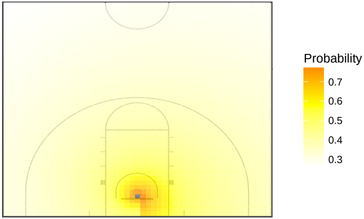
Player 2



Player 3



Player 4



Hierarchical Model: Contour Plots (cont.)

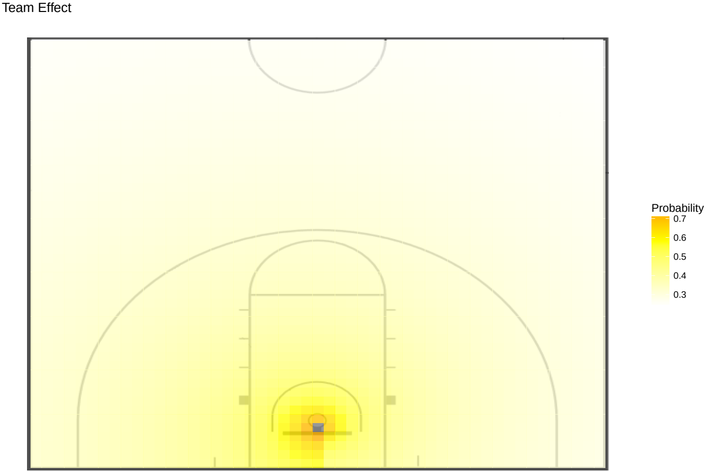


Figure 5: Contour Plots

Discounted Model: Notation

Hierarchical model equation:

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

Binomial likelihood term:

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

Exponential discounting on outcomes:

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

Discounted likelihood:

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

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

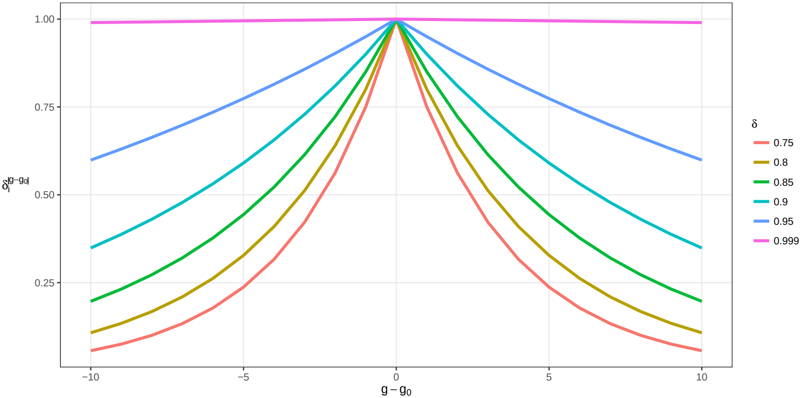


Figure 6: Discount Weights by δ and Time Difference

Discounted Model: Results

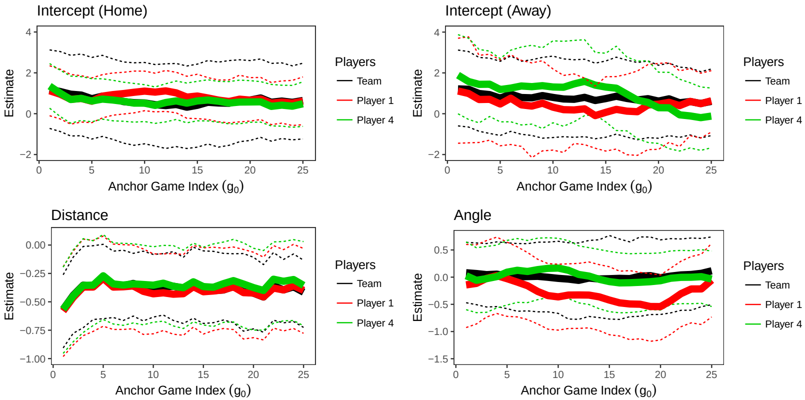


Figure 7: Parameters for Two Players and Population over Time, $\delta = 0.750$

Discounted Model: Results (cont.)

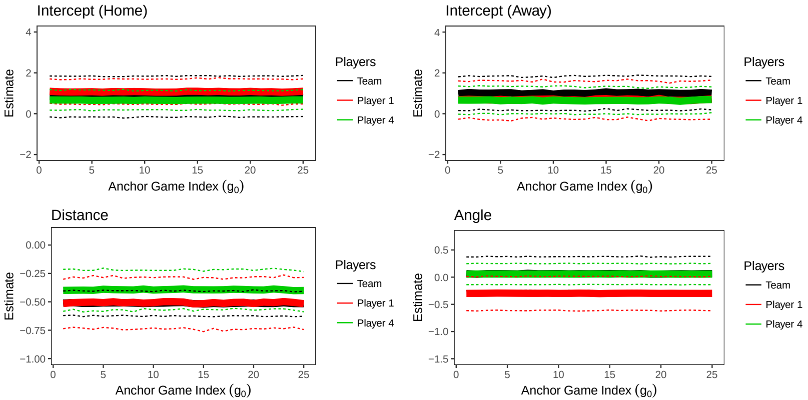


Figure 8: Parameters for Two Players and Population over Time, $\delta = 0.999$

Evaluation of Models

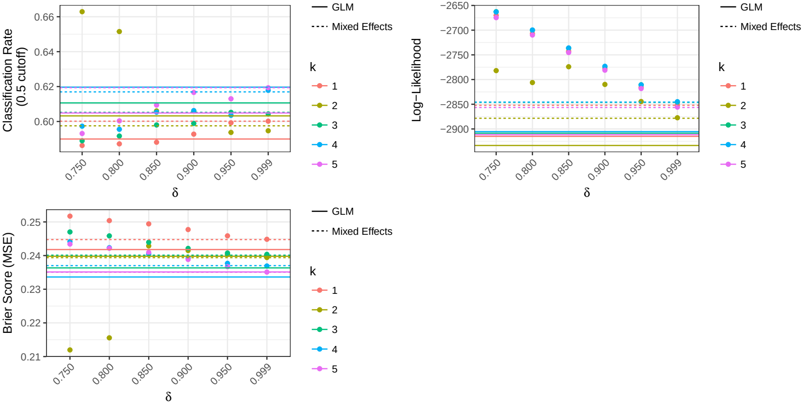


Figure 9: Model Evaluation

Calibration Plots

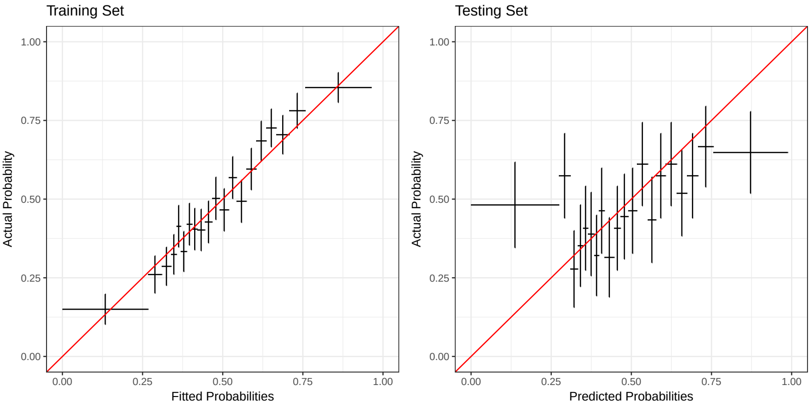


Figure 10: Calibration Plots for Discounted Likelihood Model, $\delta = 0.850$

Results from Model with $\delta = 0.850$

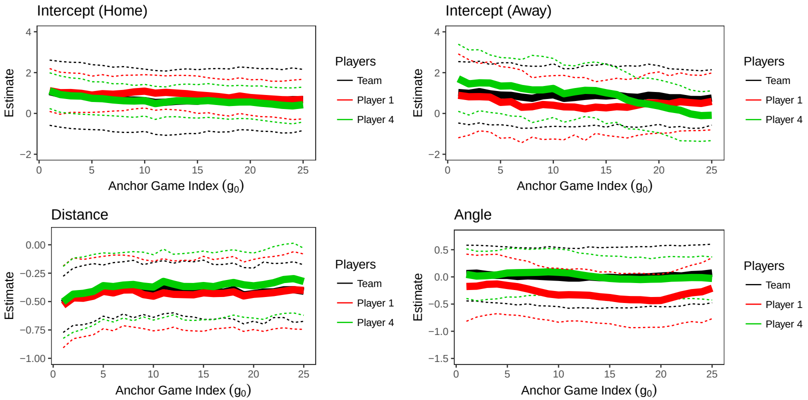


Figure 11: Parameters for Two Players and Population over Time, $\delta = 0.850$

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.

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.