A Multilevel Hierarchical Model for Basketball Player Performance

Eric Gerber, Evidence Matangi, Yucong Zhang

Department of Statistics, Purdue University.

250 N. University Street, West Lafayette, IN 47907-2066, USA

December 14, 2016

Abstract

Using NBA play-by-play data for the 2009-2010 season, we plan to construct a multilevel hierarchical model, with players nested within teams in order to predict shooting performance for players post Christmas eve, the 24^{th} of December 2009. The Bayesian framework takes into account the prior information about the players nested within the teams to predict their end of season shooting performance. Multilevel models allow us to analyze both the team and player shooting performances, which is helpful in motivating the players and also in team selection over the course of the seasons. We find that a simple multilevel model does a slightly better job than the sample mean in predicting team field goal percentage.

Keywords: Multilevel Normal Hirerarchical Model, NBA, Shooting, Prediction, Performance.

---

1. Introduction

It is important to evaluate individual player performance in team sports such as basketball, but this is very difficult as there is a need to explicitly consider team effects (Piette *et al*, 2011). The composition of a team depends on team selection, the health of the players, the type of play of their opponents, home or away fixture, and many other decision-making factors surrounding the team. Non-hierarchical statistical models tend to ignore structure and report underestimated standard errors, while multilevel models are capable of building in structural relationships (Roberts, 2009). Basketball players' performance can be assessed through many skills sets, but here we will use their field goal percentage. Since we are aware of the fact that players are nested within teams, we investigate their performance taking into consideration the multilevel nature of their data. Our goal is to predict the performance of NBA teams and players in the last section of the 2009/10 season, based on the prior knowledge of their early season performance.

2. Literature review

Multilevel modeling can be used for prediction, data reduction, causal inference, among a variety of purposes. It is an approach in data analysis that can be applied to handle data that is naturally hierarchical or grouped/nested. They account for association among observations in groups and hence facilitate the making of valid and efficient inferences. Multilevel modeling aids group comparison while assessing individual group members. The multilevel Bayesian hierarchical modeling technique has been discussed widely for many years now, and has been generalized (Lindley and Smith, 1972; Smith, 1973), refined for specific cases (Wong and Mason, 1985; Kass and Steffey, 1989), and further rationalized to the general statistical public and other statistical minded fields (Gelman, 2006; Wright and London, 2009). Use of multilevel hierarchical models, specifically Normal hierarchical models, has been used in sports data analysis before, but basketball provides its own unique challenges.

While all sports have been trending towards a more analytic approach to player performance, basketball data is numerous hence basketball teams could benefit greatly if they elected to be statistically supported (Nikolaidis, 2015). The NBA launched a public database in September 2013 containing players' statistics, through this, fans and analysts are able to explore and do research on it (Bruce, 2016). This NBA player tracking data has the ability to differentiate player performance across more dimensions than before, hence improving on the performance analysis. However, this places a demand for more computational resources and reduces the ability to easily analyze and interpret findings.

On another viewpoint, the ability to accurately predict the performance of NBA players depends upon determining and using features that are strongly correlated with the number of points an NBA player will score against any given opponent (Wheeler, 2015). These can be either team-level or player-specific features, and one such feature that cuts across these levels is field goals. The need to explore both the player and team performances entail that there is a need to implement some dimension reduction techniques on the data priory (Bruce, 2016). Performance analysis in basketball is useful in giving a better insight into ways to optimize players and team resources, understand competition demands and possible adjustments to training processes (Sindik, 2015). Basketball is also largely a strategic team sport which requires teamwork through synchronization of individual player technique and tactics with those of his team mates (Sindik, 2015). Basketball players differ according to five playing positions; small forward, point guard, center, power forward, and shooting guard. While a more comprehensive hierarchical model will take all these important factors into consideration, for the purposes of this project, we focus solely on the hierarchical model's ability to account for the nested relationship of players to teams.

3. Methods

The data used in this data is the NBA play-by-play for the 2009-2010 season. We seek to construct a multilevel hierarchical model, with players nested within teams, in order to predict shooting performance for players post Christmas holiday. The sample under study consisted of 192 NBA players. Only the players who had attempted at least 150 field goals by the 24^{th} of December of the 2009/10 season were included in this investigation, to prevent the influence of the player statistics derived from a few games. The field goals made for individual players follow the Binomial distribution and if the sample size is large enough, by the Central Limit Theorem, it can be approximated by the Normal distribution. However, to simplify adherence to assumptions, a more natural Normal approximation of a transformation of the actual percentage of field goals made can be used as the response for the Bayesian hierarchical model. A Bayesian framework naturally accommodates hierarchical models and ensures that future predictions can be carried out by means of the posterior predictive distribution of the data (Baio and Blangiardo, 2010).

Much research has been done on approximating binary responses within a hierarchical framework (Rodriguez and Goldman, 1995; Goldstein and Rasbash, 1996). Since we are focusing on the percentage itself, we can motivate the Normal approximation with some cut-off for minimum number of attempts, which has been done before with a similar approach involving baseball player batting averages (Efron and Morris, 1975). The Normal approximation becomes even more reasonable when we use the logit transformation on the percentage, since while probabilities follow Beta distributions, the support for the logit transformation exists on an infinite support and should thus be Normal. This line of thinking motivated our choice in response:

$$y_{j(k)} = \log\left(\frac{p_{j(k)}}{1 - p_{j(k)}}\right)$$

Where $p_{j(k)}$ is the observed field goal percentage of player j on team k before Christmas.

Thus the model will seek to recover posterior means for the players logit of field goal percentage ($\theta_{j(k)}$) as the lower level of the multilevel model nested within the team means (ψ_k).

If we consider the variation for the players observed logit as a known quantity ($\sigma^2_{j(k)}$), the priors at the highest level of the model will be placed on only three parameters; the mean for the team means μ, the variance for the team means τ^2, and the variances for the player means ν_k^2. Treating $\sigma^2_{j(k)}$ as unknown is a more intuitive approach, but complicates matters to beyond the scope of this project.

A multilevel Normal hierarchical model can then follow the general three-stage process (Lindley and Smith, 1972) as follows:

$$\begin{aligned}
y_{j(k)} &\sim N(\theta_{j(k)}, \sigma^2_{j(k)}) \\
\theta_{j(k)} &\sim N(\psi_k, \nu_k^2) \\
\psi_k &\sim N(\mu, \tau^2) \\
p(\mu, \tau^2, \nu_k^2) \propto p(\tau^2, \nu_k^2) &= \frac{1}{\sqrt{\tau^2 + \nu_k^2}}
\end{aligned}$$

The three unknown parameters are actually 32 unknown parameters; one μ and τ^2, and thirty ν_k^2 for the thirty NBA teams. The joint reference prior distribution on τ^2 and the vector of ν_k^2's allows us to directly sample all remaining parameters (including μ) given the sample data. However, sampling from this joint distribution, given the data, is not an easy task.

The MCMC method we used in this study is the Metropolis-Hastings algorithm. While this requires some work to determine the posterior distribution for the parameters we are interested, Metropolis-Hastings has the advantage that we do not have to calculate the whole conditional posterior distribution but rather simply evaluate it at two points per iteration.

With some use of the lemmas (Lindley and Smith, 1972), and the framework for the Metropolis-Hastings (Gelman, 2014), we can map out the iterative process for sampling τ^2 and ν_k^2 and then successively sampling the remaining parameters given those variance terms:

Via Metropolis-Hastings, sample:

$$\begin{aligned}\log p(\tau^2, \nu_k^2 | y_{j(k)}, \sigma_{j(k)}^2) \propto \log p(\tau^2, \nu_k^2) \times & \\ & \sum_{k=1}^{30} \Big[(-.5) \sum_{j=1}^{192} \log(\tau^2 + \nu_k^2 + \sigma_{j(k)}^2) \times \\ & (-.5) \log \left(\sum_{j=1}^{192} \frac{1}{\tau^2 + \nu_k^2 + \sigma_{j(k)}^2} \right) \times \\ & (-.5) \sum_{j=1}^{192} \frac{(y_{j(k)} - \hat{\psi}_k)^2}{\tau^2 + \nu_k^2 + \sigma_{j(k)}^2} \Big] \end{aligned} \tag{1}$$

Then, conditional on sampled variance parameters τ^2 and ν_k^2, and observed team logits ($y_{.(k)}$), sample:

$$\mu | \tau^2, \nu_k^2, y_{.(k)} \sim N \left(\frac{\sum_{k=1}^{30} (\tau^2 + \nu_k^2)^{-1} y_{.(k)}}{\sum_{k=1}^{30} (\tau^2 + \nu_k^2)^{-1}}, \left(\sum_{k=1}^{30} \frac{1}{\tau^2 + \nu_k^2} \right)^{-1} \right) \tag{2}$$

And the sampling continues thusly:

$$\psi_k | \mu, \tau^2, \nu_k^2, y_{.(k)} \sim N \left(\frac{\tau^{-2}\mu + \nu_k^{-2} y_{.(k)}}{\tau^{-2} + \nu_k^{-2}}, \left(\tau^{-2} + \nu_k^{-2} \right)^{-1} \right) \tag{3}$$

$$\begin{aligned}\theta_{j(k)} | \psi_k, \tau^2, \nu_k^2, y_{j(k)}, \sigma_{j(k)}^2 \sim N \Bigg(& \frac{(\tau^2 + \nu_k^2)^{-1} \psi_k + \sigma_{j(k)}^{-2} y_{j(k)}}{(\tau^2 + \nu_k^2)^{-1} + \sigma_{j(k)}^{-2}}, \\ & \left((\tau^2 + \nu_k^2)^{-1} + \sigma_{j(k)}^{-2} \right)^{-1} \Bigg) \end{aligned} \tag{4}$$

With the resulting samples, we will be able to compare not only the recovered player means with the true shooting performance after Christmas, but the team means as well. We can then assess the Bayes' estimators performance for both team and player versus simply using the frequentist estimator of the sample mean.

4. Results

Metropolis Hastings Algorithm

We used Metropolis Hastings to sample the unknown variance parameters. For the multilevel model this involves sampling from the joint posterior of τ^2: the overall team variance parameter, and ν_k^2: the player variance parameter for the players on team k, given the data $y_{j(k)}$ and $\sigma_{j(k)}$. The prior placed on the variance parameters will be a simple reference prior, to avoid making the computation any more difficult. Then we sample the successful level parameters given the sampled variance components until the player posterior means are sampled.

Shrinkage Plots

Shrinkage plots are useful tools for examining how the Bayesian hierarchical model influences the estimation of the parameters over that of the traditional estimators based on sample means. A shrinkage plot showing parallel lines indicates the Bayes' estimators performing no differently than the sample means, while observing actual shrinkage indicates a real influence of the model on prediction. For the shrinkage plots displayed on the next two pages, we back convert the posterior logit to be a posterior proportion, and examine the shrinkage plot for both the teams (constructed with the posterior means of ψ) and the players (with the posterior means of θ). Also plotted are what can call a "True" shrinkage plot, which shows what a shrinkage plot that was producing perfect predictors would look like. This is one way of assessing prediction of our Bayes' estimators: if the shrinkage plot of the posterior means looks closer to the "True" shrinkage plot than parallel lines, the Bayes' estimators are likely doing a better job than the sample means would as predictors.

Figures 1 and 2 show the shrinkage plots for the predicted and true Team Field Goal Percentage, respectively.

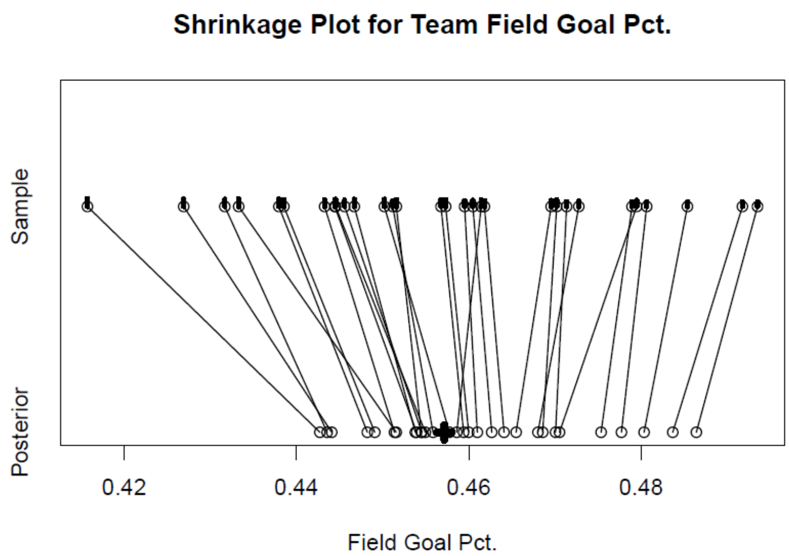

Figure 1: Shrinkage Plot for Team Field Goal Percentage

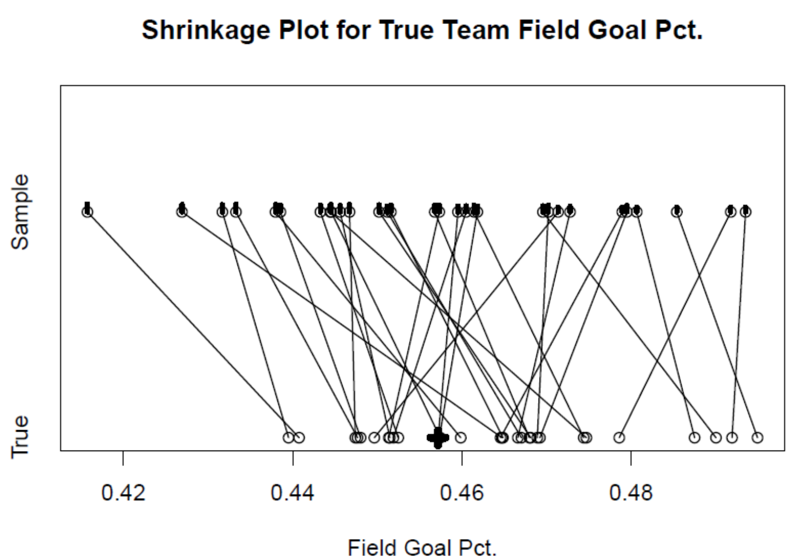

Figure 2: Shrinkage Plot for True Team Field Goal Percentage

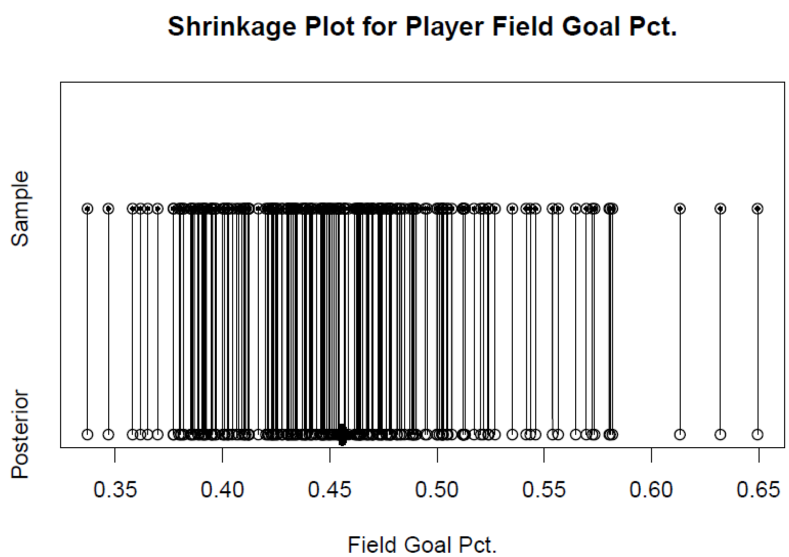

Figure 3: Shrinkage Plot for Player Field Goal Percentage

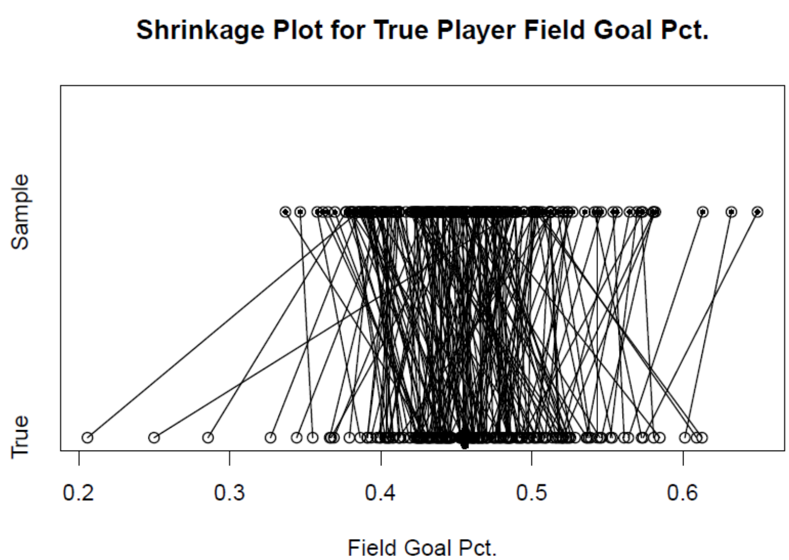

Figure 4: Shrinkage Plot for True Player Field Goal Percentage

Figure 3 and Figure 4 show the shrinkage plots for the predicted and true Player Field Goal Percentage, respectively. However, Figure 3 shows that the hierarchical model does not improve the prediction for players.

Performance

Finally, we look at the average $l2$ distance of the Bayes' estimators of the log-odds to the true post-Christmas shooting log-odds which, if smaller for the Bayes' estimators, indicate a better performance over the sample means. The results for both the team and player means are shown in Table 1.

Table 1: Euclidean distances

	Posterior Preds	Sample Mean Preds
Team Means	0.001735	0.003706
Player Means	0.040872	0.040870

Based on comparing the shrinkage plots and the distances, we can see that in estimating the Team shooting percentages, the Multilevel Hierarchical Normal Model does better in prediction than the sample means, but for the players, the Bayes' estimators are no better than the sample means for prediction. The team shrinkage plot is fairly similar to the "True" shrinkage plot, and based on the Euclidean distance, the Bayes' estimators do a better job of prediction than the sample means.

Conversely, predicting the player's shooting performance post Christmas with the Bayes' estimators seems to be no better than using the sample means. This might be because we treated the player standard errors σ as known (and they were very small using the sample data), so we could likely do much better by treating the player variances as unknown and including them in the initial unknown parameters. It would also be prudent to include some player-specific covariates in the estimation of the player mean, θ, though estimation of those covariate parameters would further complicate the model.

5. Discussion

The simplicity of the multilevel Normal hierarchical model makes it an enticing tool for analyzing sports data, and with reasonable satisfaction of assumptions, the prospect of improved prediction over that of the sample means is a powerful lure. The results indicate that even with a non-informative reference prior and without introducing covariates, team field goal percentage prediction can be improved with Bayes estimation. While the nesting structure of the model made it reasonable to assume that player prediction could also be improved, that turned out not to be the case. Though, as we have noted, we might expect significant improvement in prediction of player means by either treating player variances as unknown, inclusion of player-specific covariates, or both.

Future work
As we saw in the results section, prediction for team means based on the Bayes' estimators from the multilevel model was reasonably better than prediction for the player means. In the future, inclusion of player-specific covariates as well as estimation of player variability might make posterior draws of player field goal percentage more trustworthy as predictors. Field goals are also not the only way to assess a basketball player's contribution to the team, as there are other indicators such as rebounding or steals, which might make a Dirichlet model more appropriate.

6. Conclusions

What we have seen is that while the results for our simple multilevel model did not live up to expectations, there is no question that a Bayesian hierarchical model approach can be beneficial to the analysis of sports data. An increase in complexity, from estimation of player variances to inclusion of additional player- or team-specific parameters, would almost certainly increase posterior predictive ability.

7. Code

All code used in this project available via a GitHub repository:
https://github.com/gerber19/fieldgoalpbayes

8. Acknowledgments

Both Professors Sabbaghi and Rao were incredibly helpful, and Prf. Sabbaghi especially motivated many of the major steps we took. Our classmates proved very useful throughout, occasionally handing out tips on the R coding or writing in LaTeX; perhaps most helpful was Tim Keaton.

9. References

1. Baio, G., and Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics, 37(2), 253-264.
2. Bruce, S. (2016). A scalable framework for NBA player and team comparisons using player tracking data. Journal of Sports Analytics, 2(2), 107-119.
3. Efron, B., and Morris, C. (1975). Data Analysis Using Stein's Estimator and its Generalizations. Journal of the American Statistical Association, 70(350), 311-319.
4. Gelman, A. (2006). Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. Technometrics, 48(3), 432-435.
5. Andrew Gelman, Bayesian Data Analysis, CRC Press, Boca Raton, 3rd Edition, 2014.
6. Goldstein, H., and Rasbash, J. (1996). Improved Approximations for Multilevel Models with Binary Responses. Journal of the Royal Statistical Society. Series A (Statistics in Society), 159(3), 505-513.
7. Kass, R., and Steffey, D. (1989). Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models). Journal of the American Statistical Association, 84(407), 717-726.
8. Lindley, D.V., and Smith, A.F.M. (1972). Bayes Estimates for the Linear Model. Journal of the Royal Statistical Society. Series B (Methodological), 34(1), 1-41.
9. Nikolaidis, Y. (2015). Building a basketball game strategy through statistical analysis of data. Annals of Operations Research, 227(1), 137-159.
10. Piette, J., Pham, L., and Anand, S. (2011, May). Evaluating basketball player performance via statistical network modeling. MIT Sloan Sports Analytics Conference.
11. Roberts, Kyle. "Introduction to Hierarchical Linear Modeling with R", 2009. [Online]. Available: http://faculty.smu.edu/kyler/training/AERA_overheads.pdf
12. Rodriguez, G., and Goldman, N. (1995). An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. Journal of the Royal Statistical Society. Series A (Statistics in Society), 158(1), 73-89.

13. Sindik, J. (2015). Performance Indicators of the Top Basketball Players: Relations with Several Variables. Collegium antropologicum, 39(3), 617-624.
14. Smith, A.F.M. (1973). A General Bayesian Linear Model. Journal of the Royal Statistical Society. Series B (Methodological), 35(1), 67-75.
15. Wheeler, Kevin. Predicting NBA Player Performance", 2012. [Online]. Available: http://cs229.stanford.edu/proj2012/WheelerPredictingNBAPlayerPerformance.pdf
16. Wong, G., and Mason, W. (1985). The Hierarchical Logistic Regression Model for Multilevel Analysis. Journal of the American Statistical Association, 80(391), 513-524.
17. Wright, D., and London, K. (2009). Multilevel modelling: Beyond the basic applications. British Journal of Mathematical and Statistical Psychology, 62, 439-456.