Predicting NBA All-Stars Using Bayesian Logistic Regression

Timo Pew

December 10, 2019

NBA All-Stars

https://en.wikipedia.org/wiki/2023_NBA_All-Star_Game

- Halfway through each season, 12 best players from both conferences are selected to play in the All-Star game
- All-Star selection is one of the greatest honors NBA players can receive. It impacts how the general public perceive the legacy of players. It also has financial consequences for some players.
- Though the intent of the game is to recognize and showcase the 24 best players in the league of the given season, every year there appears to be players deserving of the recognition that do not receive it

Data and Objective

- We have data from every player that played more than 300 minutes from the 2008-2009 season through the 2018-2019 season
- For each player in each season, we have what we consider two types of variables:
  - (i) Performance: typical game averages (including points, rebounds, and assists per game), and metrics that attempt to quantify the overall value of a player (such as win shares per 48 minutes)
  - (ii) Non-performance: whether or not the player has previously been selected to an All-Star game, the proportion of games the player's team has won, and the number of TV households in the local market
- Our primary objective is to determine if, or to what extent, the non-performance variables have an effect on All-Star selection

Model Specification

Let Y_i denote the dichotomous All-Star selection outcome, and π_i denote the probability of being selected for a given player and season. We assume:

$$\begin{aligned} Y_i|\pi_i &\overset{ind}{\sim} Bernoulli(\pi_i) \quad \text{for } i = 1, \ldots, n \\ log\left(\frac{\pi_i}{1-\pi_i}\right) &= \beta_0 + \boldsymbol{x}_i'\boldsymbol{\beta} \quad \text{for } i = 1, \ldots, n \end{aligned} \tag{1}$$

where $n = 4206$ and $\boldsymbol{x}_i'$ contains p explanatory variables, such as those previously mentioned

Model Specification

We have little intuition about the location or scale of the regression coefficients, so we chose the following hierarchical specification:

$$
\begin{aligned}
\beta_j | \mu_j, \tau_j &\overset{ind}{\sim} \mathcal{N}(\mu_j, \tau_j) \quad j = 0, 1, \ldots, p \\
\mu_j &\overset{iid}{\sim} \mathcal{N}(0, b) \quad j = 0, 1, \ldots, p \\
\tau_j &\overset{iid}{\sim} Gamma(c, d) \quad j = 0, 1, \ldots, p
\end{aligned} \tag{2}
$$

- We chose Normal priors for the regression coefficients because we want to allow them to be positive or negative
- We chose Normal and Gamma hyperpriors on the mean and precision of the regression coefficients to take advantage of conditional conjugacy
- We set the mean of each μ_j to be zero as something of a null hypothesis that the variables have no effect on making the All-Star roster
- For our primary model we use diffuse hyperparameters $b = 0.1$, $c = 1$, and $d = 1$. We also use $b = 64$, $c = 40$, and $d = 20$ to investigate the sensitivity of our chosen specification

Joint Posterior Density

Under the assumed sampling model and prior specification, the joint posterior density of the parameters $\beta_0, \ldots, \beta_p$; $\mu_0, \ldots, \mu_p$; and $\tau_0, \ldots, \tau_p$ is proportional to:

$$\begin{aligned}
&\prod_{i=1}^{n} \left(\frac{exp(\beta_0 + \mathbf{x}_i' \boldsymbol{\beta})}{1 + exp(\beta_0 + \mathbf{x}_i' \boldsymbol{\beta})} \right)^{y_i} \left(\frac{1}{1 + exp(\beta_0 + \mathbf{x}_i' \boldsymbol{\beta})} \right)^{1-y_i} \times \\
&\prod_{j=0}^{p} \sqrt{\tau_j}\, exp\left(\frac{-\tau_j}{2} (\beta_j - \mu_j)^2 \right) \times \\
&\prod_{j=0}^{p} exp\left(\frac{-b}{2} \mu_j^2 \right) \times \prod_{j=0}^{p} \tau_j^{c-1} exp(-\tau_j d)
\end{aligned} \tag{3}$$

Because the joint posterior density has an unknown form, we must use a simulation based approach to approximate the posterior density

Sampling from the Posterior Density

Our particular selection of priors and hyperpriors allows us to take advantage of conditional conjugacy to sample more easily from the posterior distribution. The full conditional density of an individual β_j, μ_j, and τ_j for $j = 0, 1, \ldots, p$ are shown below.

$$p(\beta_j|\cdot) \propto \left(\prod_{i=1}^{n} \frac{exp(x_{ij}\beta_j y_i)}{1 + exp(\beta_0 + \mathbf{x}_i' \boldsymbol{\beta})} \right) exp\left(\frac{-\tau_j}{2}(\beta_j - \mu_j)^2 \right)$$

$$p(\mu_j|\cdot) \propto exp\left(\frac{\tau_j + b}{2} \left(\mu_j - \frac{\beta_j}{1 + b/\tau_j} \right)^2 \right) \implies \mu_j|\cdot \sim N\left(\frac{\beta_j}{1 + b/\tau_j}, \tau_j + b \right)$$

$$p(\tau_j|\cdot) \propto \tau_j^{c-1/2} exp\left(-\tau_j \left(\frac{(\beta_j - \mu_j)^2}{2} + d \right) \right) \implies \tau_j|\cdot \sim Gamma\left(c + \frac{1}{2}, \frac{(\beta_j - \mu_j)^2}{2} + d \right)$$

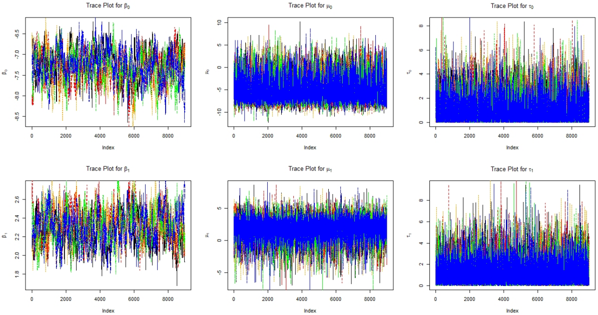
Sampling from the Posterior Density

The algorithm to obtain a single chain of draws from the posterior distribution is given in Algorithm 1. We ran 5 chains of 10,000 iterations with 1,000 burn-in iterations.

Algorithm 1: Metropolis-within-Gibbs

1. Choose starting points for β_j, μ_j, and τ_j ($\beta_j^{(1)}$, $\mu_j^{(1)}$, $\tau_j^{(1)}$) for $j = 0, 1, \ldots, p$
2. For i $= 2, \ldots, B$:
   (a) For j $= 0, 1, \ldots, p$:
      (i) Draw a new proposed value, $\beta_j^{(i)*}$ from $\mathcal{N}(\beta_j^{(i-1)}, \sigma)$, where σ is a tuning parameter
      (ii) Compute the Metropolis ratio, $M_r = p(\beta_j^{(i)*}|\cdot) \,/\, p(\beta_j^{(i-1)}|\cdot)$
      (iii) With probability $\min(1, M_r)$, set $\beta_j^{(i)} = \beta_j^{(i)*}$, otherwise set $\beta_j^{(i)} = \beta_j^{(i-1)}$
   (b) Draw $\mu_j^{(i)}$ from its full conditional distribution, $\mu_j|\cdot \sim N\left(\frac{\beta_j^{(i)}}{1+b/\tau_j^{(i-1)}}, \tau_j^{(i-1)} + b\right)$ for $j = 0, 1, \ldots, p$
   (c) Draw $\tau_j^{(i)}$ from its full conditional distribution, $\tau_j|\cdot \sim Gamma\left(c + \frac{1}{2}, \frac{(\beta_j^{(i)} - \mu_j^{(i)})^2}{2} + d\right)$ for $j = 0, 1, \ldots, p$

MCMC Diagnostics



Trace plots for β_0, β_1, μ_0, μ_1, τ_0, and τ_1

Effective samples, $\hat{R}$, and acceptance rate for each β_j

Diagnostic	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Effective Samples	451	505	1142	1548	1230	2795	1175	6314
$\hat{R}$	1.019	1.020	1.005	1.004	1.002	1.000	1.007	1.001
Acceptance Rate	0.255	0.168	0.162	0.167	0.215	0.333	0.256	0.247

Posterior Predictive Test

- We simulated All-Star selections using our posterior draws from β_0, β_1, ..., β_p and the observed explanatory variables
- For each simulation (45,000 total) we calculated the number of All-Stars selected from each season
- The posterior predictive p-values are the proportion of simulated number of All-Stars in each season that surpassed the actual number of All-Stars selected that season.
- This test helps us see if the data we observed is a reasonable realization under our posterior distribution and if there is a year-to-year trend that we are not capturing

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
0.570	0.080	0.595	0.269	0.151	0.499	0.054	0.470	0.911	0.395	0.907

Posterior predictive test for primary prior parameterization

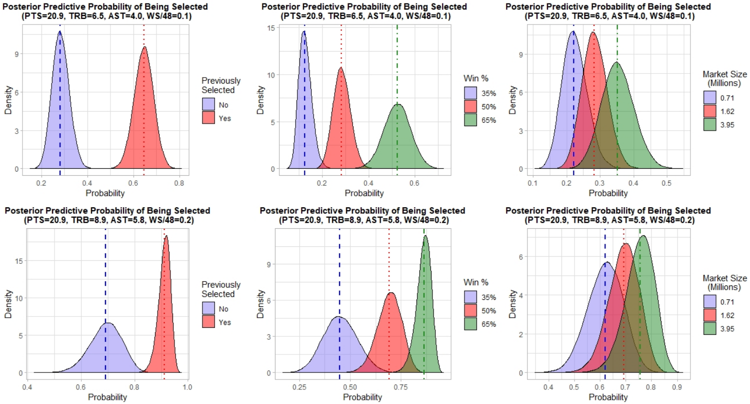
Posterior Inference on $\beta_0, \ldots, \beta_7$

Method	Type	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Bayesian (1)	Point Estimate	-7.301	2.306	0.679	0.684	0.393	1.535	1.025	0.320
	Lower 95% CI	-8.068	1.970	0.467	0.485	0.049	1.109	0.715	0.106
	Upper 95% CI	-6.610	2.658	0.896	0.874	0.728	1.970	1.332	0.526
Bayesian (2)	Point Estimate	-6.236	1.934	0.570	0.556	0.417	1.341	0.807	0.223
	Lower 95% CI	-6.848	1.621	0.364	0.363	0.117	0.946	0.537	0.027
	Upper 95% CI	-5.628	2.241	0.779	0.749	0.726	1.739	1.083	0.415
Frequentist	Point Estimate	-7.267	2.297	0.677	0.680	0.385	1.536	1.024	0.321
	Lower 95% CI	-8.075	1.954	0.459	0.485	0.049	1.108	0.720	0.110
	Upper 95% CI	-6.556	2.668	0.903	0.880	0.729	1.973	1.339	0.531

Point Estimates and Confidence Intervals for Regression Coefficients Using Bayesian and Frequentist Methods

- Bayesian and frequentist approaches produce similar results
- From either perspective, we have strong evidence that having previously being selected, being on a better team, and playing in a well-populated city have an effect on the probability of being selected to the All-Star roster increase
- Coefficients are shrunk towards zero using alternative priors

Visualizing Effect of Non-Performance Variables



Posterior predictive probability of being selected under various conditions. The dotted lines represents the frequentist point estimates

Predicting 2018-2019 All-Stars

- For the Bayesian predictions, players with greater than half of their posterior predictive mass above 0.5 were considered predicted All-Stars

		Predicted Positive	Predicted Negative
Truth	Positive	20	7
	Negative	10	366

		Predicted Positive	Predicted Negative
Truth	Positive	23	4
	Negative	5	375

Confusion matrix for model built with only performance variables on the left and for model built with both types of variables on the right using Bayesian approach

- For the frequentist predictions, players with predicted probabilities of being selected greater than 0.5 were considered predicted All-Stars

		Predicted Positive	Predicted Negative
Truth	Positive	20	7
	Negative	10	366

		Predicted Positive	Predicted Negative
Truth	Positive	23	4
	Negative	5	375

Confusion matrix for model built with only performance variables on the left and for model built with both types of variables on the right using Bayesian approach

Conclusion

- We found strong evidence that having previously been selected to an All-Star roster, team success, and the number of TV households in the local market effect the probability of being selected to play in the NBA All-Star game
- This was manifest in two ways: (1) the posterior densities for the effects of non-performance variables were largely above zero and (2) a model including non-performance variables predicted 2018-2019 All-Stars better than one including only performance variables
- Using a hierarchical structure with diffuse hyperiors resulted in very similar results to a frequentist analysis