**Maximum Likelihood Estimation of Home Court Advantage**

One of the most important factors that go into which team will win a game is the home court advantage of that team. In both the NBA and the NFL, it is often approximated that home court or home field is worth about 3 points to the game score. That is, if the teams are otherwise evenly matched, the home team should be expected to win by about 3 points on average.

However, it is certainly not the case that every team’s home environment is worth the same amount of points. For instance, the Clippers playing a ‘home’ game against the Lakers might see the crowd equally split. Moreover, neither team has to travel so the Clippers might not benefit at all from this being a home game.

In this article, we’ll use a statistical estimation technique to compute home court advantages in the NBA. In particular, we’ll use maximum likelihood estimation to compute the optimal home court advantage for each team in terms of points.

**Past Methods for Home Court Advantage**

I am not the first person to want to compute home court advantage. In fact this is quite a well-studied topic. Different methods come up with different numbers for how much playing at home is worth. These methods are all valuable, I am simply going to introduce a new method that addresses some mild concerns I have with the statistical robustness presented here.

The first, simplest method that is used is this. Over the entire course of the season, compute the average margin of victory for the home team. If you do this in the NBA. This is reasonable because, all things held equally, the good teams play at home just as often as the bad teams so the only thing to explain the difference is the presence of the home court advantage. In the NBA, home-court advantage by this method is worth 2.6 points.

Another way home court advantage is computed on a team-by-team basis is the [method described by Ken Pomeroy](https://kenpom.com/blog/how-to-measure-home-court-advantage/). More-or-less, you take a team’s average margin of victory at home and compare it to the margin of victory on the road. For example, if my average home margin is 3 points and my road margin is -4 points, then the difference is 7 points per game. Then, we declare the home court advantage to be 3.5 points when measured relative to a neutral court (because not only do I lose my 3.5 point advantage playing at home, the team gains a 3.5 point advantage by playing at their court for a total of 7 points).

**The Ideal Home Court Advantage Measure**

For us, we want to define home court advantage:

* Individually for each team in a league
* Taking into account the relative difficulty of home/away schedules for a particular team

I suggest that doing this requires more advanced mathematics than what we’ve seen above already. Each of the above methods, while undeniably useful, suffer from the same flaw: they don’t take into account the quality of teams playing at home and on the road.

For the first method, let’s take College Basketball as an example to understand why good teams play at home more often than bad teams. At the beginning of the 2019-2020 season, Kentucky played Georgetown College, Kentucky State, Eastern Kentucky University, and University of Evansville all at home. (Sorry big blue fans, I will never let you forget the University of Evansville loss). Good college basketball teams start the season playing many inferior opponents at home. Bad college basketball teams start the season playing superior opponents on the road. Thus, a higher percentage of the home games in the NCAAM season are played by quality teams, inflating the average home margin of victory.

For the second method, the issue remains but I want to explain that what Ken Pomeroy did is still not optimal. That technique used only conference games over the course of seven years with the idea that on average the home strength of schedule and away strength of schedule will balance out. This is mostly true though there will always be slight inaccuracies that lead to unwarranted conclusions.

The home/away strength of schedule over the course of seven years will mostly balance out, but they will never exactly balance out. If the teams played at home were on average one point per game better than the teams played on the road, that would be reflected in a half a point of home court advantage which is huge. This idea is very close to what we want to do, but we can do just a little bit better using some statistics.

**Parameter Estimation**

In statistics, if there is an unknown quantity (parameter) that you want to know the value of, what do you do? Run experiments and test it. Let’s show a quick example.

Suppose I have a weighted coin that I know is weighted some way but I don’t know the heads probability. Suppose I flip the coin 20 times and use this to guess the heads probability**.** If we observe

* 10 heads & 10 tails then our best guess should be that the coin is heads 50% of the time
* 15 heads & 5 tails then our best guess should be that the coin is heads 75% of the time

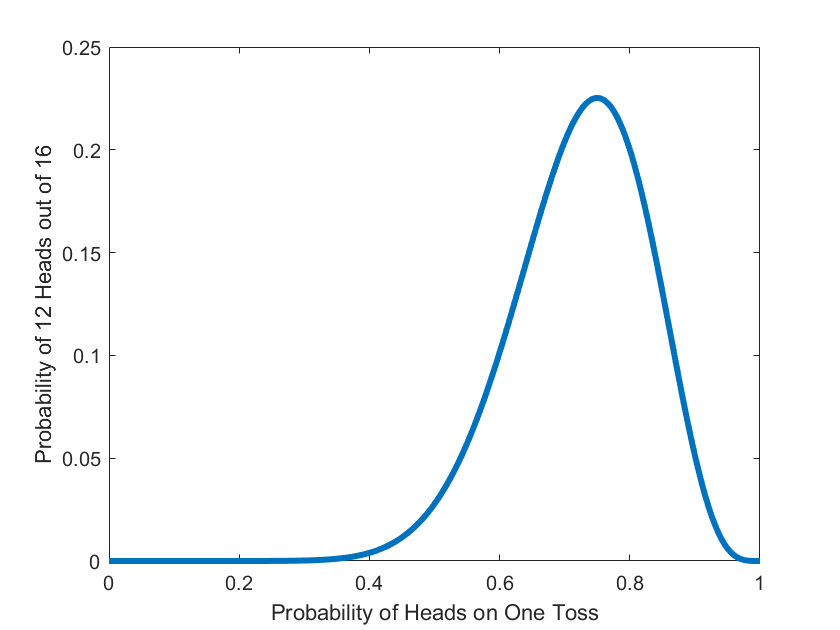
This seems like common sense. Use what the coin says to inform your guess, or estimate, of the unknown parameter. In mathematics, though, we formalize everything. We start with a simple idea and distill the important characteristics in order to be able to extend ideas into murkier situations.

**Maximum Likelihood Estimation**

The technique of maximum likelihood estimation is a generic method to estimate unknown parameters from observed data. The idea is simple: if we *knew* the value of the parameter, we could compute the probability of observing our sample:

* If a coin is 50% heads, then observing two heads out of four happens 37.5% of the time
* If a coin is 70% heads, then observing two heads out of four happens 26.5% of the time
* If a coin is 45% heads, then observing two heads out of four happens 36.8% of the time.

In fact, observing two heads out of four flips is most likely provided the coin is fair. To illustrate this point further, suppose we flip 12 heads out of 16. This would suggest a heads rate of 75%. I plotted below the probability of flipping 12 heads out of 16 versus the probability of a head on a single throw. Notice that the peak is precisely at 75%. The 75% heads parameter is the choice that best explains the data we’ve seen.



Again, predicting that my coin comes up heads 75% of the time is equivalent to picking the parameter that best explains us having seen 12 heads out of 16 tosses. We pick the parameter value that maximizes the probability of what we have seen. We pick the parameter value that minimizes how much of an ‘outlier’ our observed data is. One final way to view it: we are picking the parameter value that needs the least explanation. It describes exactly what we have seen in practice without any alteration or imparted bias.

**Home Court Advantage with Maximum Likelihood Estimation.**

In the case of home court advantage, maximum likelihood estimation is used to assign a point value for each team’s home court advantage that best explains their margin of victory at home and away while also taking into account the quality of teams we played.

Luckily, our [ensemble ratings](https://thedatajocks.com/introducing-ensemble-ratings/) give us a way to predict margin of victory on a neutral court in the NBA. So, for each individual game we can see how much better or worse we did compared to the neutral court predicted margin. This gives us a proxy to measure the impact of home court advantage that takes into account the quality of teams that played.

A little more specifically, if we out perform our neutral court predicted margins by an average of 3.2 points per game when we play at home, then this is evidence suggests that our home court advantage is about 3.2 points.

Where the technique of maximum likelihood estimation comes into play is

* Estimating the neutral court margins-of-victory given data from games that are not necessarily played on a neutral court.
* Picking the home-court advantage parameters that best explain the difference between predicted margins and actual margins.

**Home Court Advantage in the NBA**

It turns out that our technique, when used to assign a home court advantage number to the NBA as a whole, we get the same number as the average home margin of victory described far up above. This isn’t a coincidence; any time every team plays the same number of home/away games (like in the NBA), our technique will recover this simplified method. But remember, our technique lets us estimate for individual teams. See the sections below, but first….

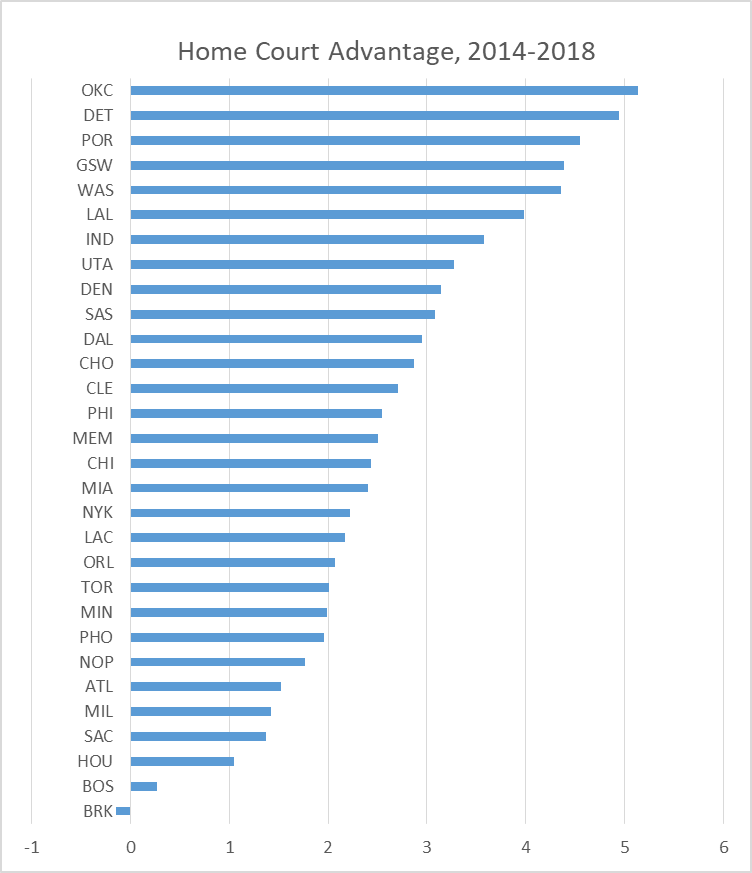
Let me comment on the 2.6 home court advantage I computed. This is probably quite a bit lower than you might expect. It might even seem insignificant. But, it is close to the difference between the Nuggets and the Trailblazers in 2020. It is roughly comparable to the point differential we assigned to explain how much harder the [West was than the East](https://thedatajocks.com/how-imbalanced-are-the-nba-conferences/).

One last way to view this 2.6 number. If two evenly matched teams play on a neutral court, they each win about 50% of the time. A home court advantage of 2.6 points gives the home team roughly a 60% chance of winning against an evenly matched opponent.

**Long-Term Average Home Court Advantage**

Perhaps more interesting is the different home court advantages for different teams across the NBA. The 2.6 number above was the league average. Individual teams’ home court advantages vary quite a bit from this number. If we want to predict who will win in any individual game, you need to know the specific HCA for the given team.

We will study two different types of individualized home court advantage. First, we’ll look at long term numbers over the course of a few years. The point of studying average HCA over the course of a few years is to try to capture actual advantages that correspond to playing in a specific city, with your specific fans, etc. To compute these, I looked at all NBA data between 2014 and 2018 and here is what we found.



Some of these advantages match our intuition. Oklahoma City and Golden State have been well-known to have good fan bases and are a tough road environment (at least over the period discussed). Smaller teams in big markets (Brooklyn, LAC) typically have worse than league average fan support.

This brings out some surprising facts, though. I doubt anybody would have thought that Detroit was a tough road environment to play in, but the numbers bear that out. Most people probably would have thought that the Boston Garden was a tough road environment, but the data doesn’t support that. My main conclusion here is that it is much harder than expected to guess which road environments are the hardest to play in.

**Single Season Home Court Advantage**

The second type of individualized home court advantage I’ll present here is for a single season. While a smaller number of games induce a higher amount of variance (or, noise) in the data, we can still make conclusions. Over the course of one season, if we measure home court advantage for a given team we are actually measuring more of a difference in an individual team’s performance playing at home versus on the road.

When looking at one single season, you can’t attribute all of the home court advantage to environmental factors (fans, distance travelled, etc.). Rather, this number is better at explaining the difference between home/away performance for a single team over the course of the season. For the shortened 2019-2020 season, here is the computed home court advantages.

Is Philadelphia really 12 points better at home than on a neutral court? Well, this year YES. They are 29-2 at home and 10-24 on the road. Minnesota, on the other hand, is 8-24 at home and 11-21 on the road. They are playing better on the road this year.

Curiously, Sacramento has supposedly been better on the road than at home this year too. Their home record is 14-17 while their road record is 14-19. What gives? Well, remember our ratings take into account quality of opponents at home and away in determine the numbers based on how a team has done relative to expectation. Sacramento has two road wins this year against the Clippers by an average margin of 15 points.

And, one final note. Even though these numbers swing wildly from -4 to +12, if you compute the average, it still suggests about a 2-2.5 point home court advantage league-wide this year.

**A Note for the Detail Oriented or Mathematically Inclined**

That is the end of the article, but I wanted to add a little extra detail if you’re like me and the specific mathematical details are the interesting part, then this section is written for you. Otherwise, my conclusions and findings are entirely contained in the following sections.

Above, I spoke about ensemble ratings predicting the margin of victory on a neutral court. The way we predict margins of victory are also using maximum likelihood to assign each team a rating and the difference between two team’s ratings is the predicted margin of victory on a neutral court.

However, I don’t have any data about games played on a neutral court [yet, the NBA restart in Orlando is the only time I can remember where this is going to happen]. So, I have to be able to estimate neutral court margins given games where a team has a home-court advantage. It seems as if I need to know HCA before computing a team’s rating in the first place.

It turns out that in the model I use, I have to simultaneously estimate home court advantage and team ratings. Essentially, I am solving a system of 60 equations in 60 variables (2 variables per team). The bad part though: the resulting matrix is singular, it isn’t invertible. You can’t uniquely solve a system of equations that aren’t invertible. This is because we only ever observe differences between team’s ratings and shifting every team’s rating by the same amount results in the same predicted margin.

There are two ways to overcome this. The first is to add an extra row to your matrix forcing the average rating to be whatever value you like. The second is to just use the Moore-Penrose (pseudo)inverse to compute *a* solution to the system.