Jordan Fresque

“Hot hand” Analysis and Predicting Makes Using NBA Shot Charts

Introduction

The concept of a “hot hand” is often applied to sports or other skill-based tasks and originates from basketball. It’s the idea that a player is allegedly more likely to score if their previous attempts were successful, as in when they are "hot” or “on fire”. While previous success at a task can indeed change the psychological attitude and maybe even the subsequent success rate of a player, researchers for many years did not find evidence of a "hot hand" in practice and dismissed this notion as a fallacy. I was motivated to recreate this type of analysis on my own because of a publication I came across recently, titled “Surprised by the hot hand fallacy?”. In this study, Miller and Sanjuro (2018) demonstrated that a subtle but substantial bias exists in this type of analysis. One that’s been inadvertently overlooked by researchers, including Gilovich, Vallone & Tversky (1985) who first described this phenomenon as a fallacy.

The bias can be explained with a simple example. Imagine you were to flip a coin three consecutive times, and you are only interested in the times you flip heads directly following a flip that was heads. Intuitively you might be thinking that each flip is independent from the next so the probability of flipping heads should be 1/2 (50%) but because we are dealing with a finite number of flips, not all of those conditions will be realized and flipping tails directly after flipping heads is actually slightly more likely. To demonstrate, let’s look at all the possible ways our three flips might occur (Figure 1).

Figure 1. The bias in the case of three coin flips (Taken from Miller and Sanjurjo 2018)

Three-flip sequence	Proportion of Hs on recorded flips
TTT	–
TTH	–
THT	0
HTT	0
THH	1
HTH	0
HHT	$\frac{1}{2}$
HHH	1
Expectation:	$\frac{5}{12}$

The expected probability of getting heads in this situation is actually slightly less than 50%. This bias translates to the finite number of shots in a basketball game as well. In the original “hot hand” study, Gilovich, Vallone & Tversky (1985) relied on a difference of conditional probabilities: the probability of making a shot, given a string of made shots minus the probability of making a shot, given an equally long string of missed shots and compared this to a benchmark of 50%, a null hypothesis that assumes makes given makes were equally as likely as makes given misses. As analogously demonstrated in the coin example, this is a biased benchmark and led to several type 1 errors in their analysis (Miller and Sanjurjo 2018).

Data Collection

To collect the data for my analysis, I used Beautiful Soup to scrape the shot-charts found at basketball-reference.com for every game in Steph Curry’s 2015-2016 and 2016-2017 seasons (including play-offs). The points on these charts contain details about each shot a player took in a particular game. From the tooltip of each point on the chart, I was able to scrape the data for each shot and organize each game into a sequential string of makes (1) or misses (0).

Figure 2. Example string for one of Curry’s games

0 0 0 0 0 1 1 1 1 1 0 0 0 0 1 0

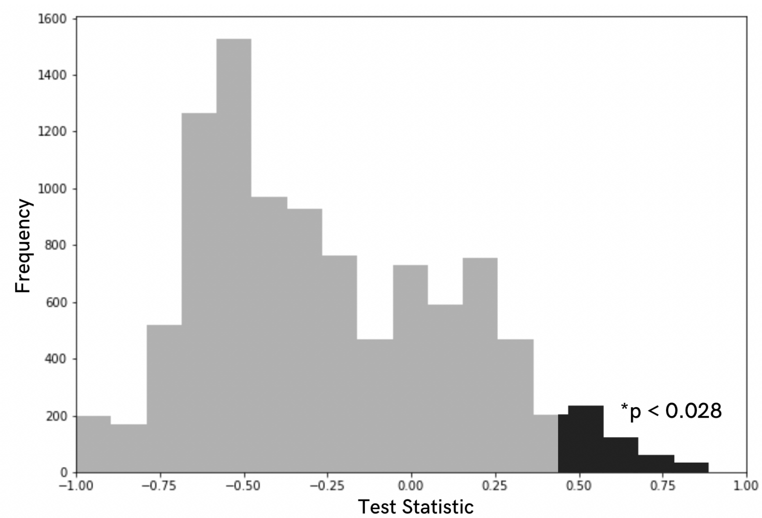
Experimental Design

With this string of 1s and 0s (Figure 2, for example), I identified all the instances of a shot taken following *n* made shots, and calculated the conditional fraction of those attempts (3/4 = 75%, *n*=2). I then did the same thing for all the instances of a shot taken following *n* missed shots (2/7 = 28.6%, *n*=2). The difference between those conditional fractions gives us a test statistic (t_2 = 44.5%), which will act as our gauge of “hot handedness”. If this number is abnormally high, the player will have made a much higher percentage of shots following *n* makes compared to shots following *n* misses, something that we might expect from a player with a “hot hand”.

To establish a null hypothesis that accounts for the small sample bias, that original string of 1s and 0s is shuffled so the order doesn't matter anymore (i.e. the previous attempt will have had no influence on the following attempt). After shuffling the original string 10,000 times and calculating the test statistic for each random permutation, we can compare the test statistic of the original string to the distribution of test statistics for its 10,000 random permutations.

Using our example string from Figure 2 to illustrate, we see that the particular order of 1s and 0s is actually on the "hotter" end of the many possible ways we can arrange this many made shots given a string of this length (Figure 3).

Figure 3. Comparison of example string to the distribution 10,000 random permutations

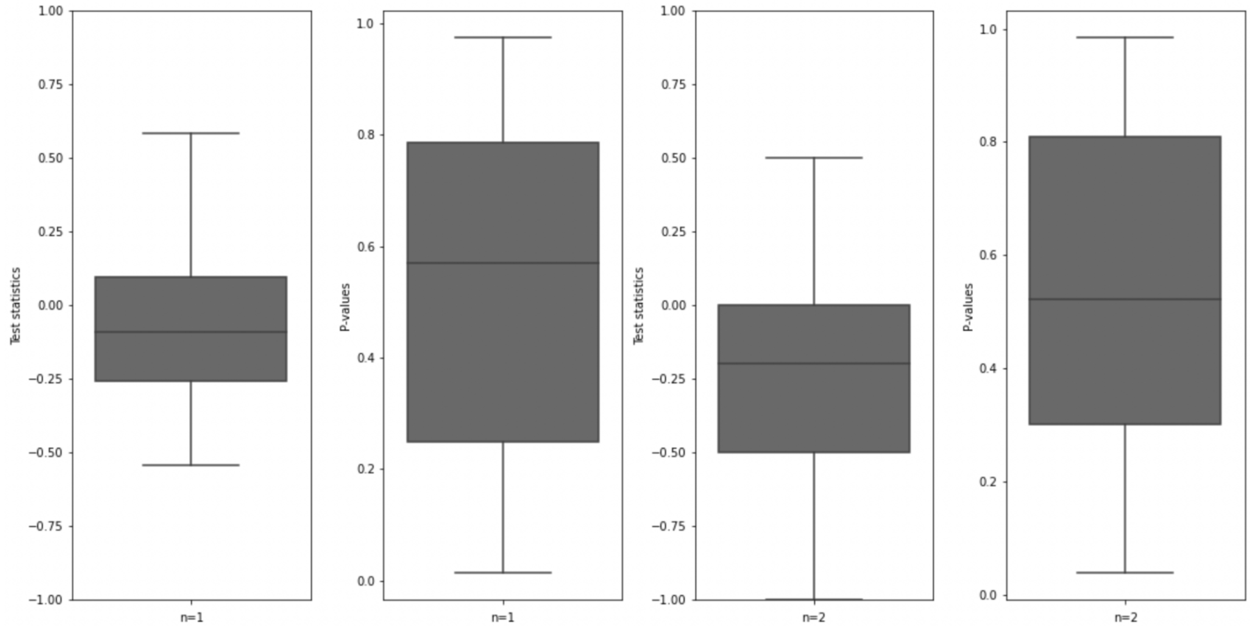


Results

Even though the example string used to illustrate the methodology of this analysis provides evidence of a "hot hand", at least for how we've defined it, this wasn't usually the case for Steph Curry. Generally his shooting percentage following makes was slightly lower than his shooting percentage following misses and very few* observations demonstrated evidence of a "hot hand" (Figure 4).

*Only 3/97 (n=1) and 2/97 (n=2) observations significant at the 5% level

Figure 4. Test statistics and p-values for all games (Steph Curry, 2016-2017)



Predicting makes

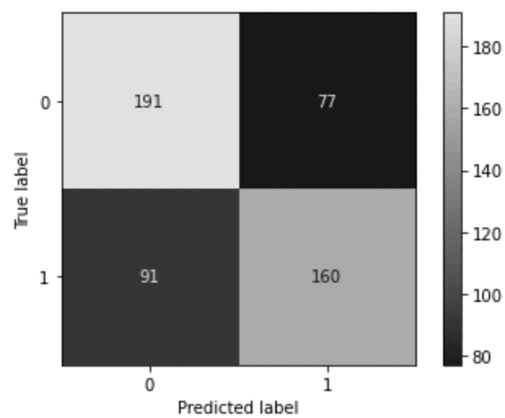
I extended this project by training a model with the information I collected about each shot to predict makes or misses. To establish a baseline, I fit a vanilla logistic regression to the data I had already scraped from those shot charts. There wasn't much detail here, mostly when and where the shot was taken, and whether it was a jump shot or not. The model trained on this data wasn't very effective (55% on test data), it only performs slightly better than just predicting the majority class every time (52.3%).

I expanded my data wrangling efforts in order to find as much information about each shot as possible. The NBA stats API provided some additional categorical features like what kind of shot it was or what region of the court it was taken from. Some more advanced features were available but only for a few years (between 2013-2016). Thankfully someone scraped the shot charts with these advanced features while they were still available and put them up on NBAsavant.com. This dataset provides some really great details like how many dribbles the player took before they shot, how long they had the ball for, who the nearest defender was, how closely they were guarded and time remaining on the shot clock.

Model Evaluation

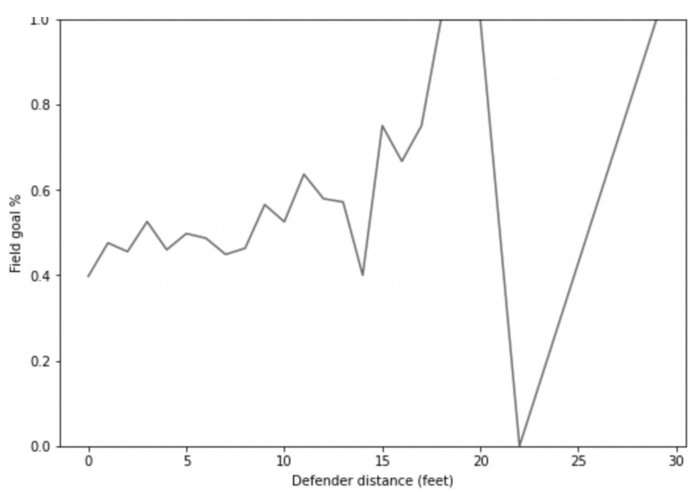
With a lot more features to train the model on, the accuracy improved substantially (67% on test data) compared to our baseline (55% on test data). The optimized model managed to accurately predict a few more misses (71% recall) than makes (64% recall) but had nearly identical precision for both classes (68%) (Figure 5).

Figure 5. Confusion matrix for optimized logistic regression



The model was able to learn some pretty valuable insights that agree with exploratory data analysis. For example, being separated from a defender (Figure 6) and having ample time left on the shot clock are predictive of a make. Conversely, the longer Steph holds the ball, the more dribbles he takes, and the further away from the basket he is, the less likely he is to score.

Figure 6. Field goal % by defender distance (feet)



Conclusion

After becoming more familiar with this process, I still think the biggest issue with this type of analysis is how to quantitatively define a “hot hand”. I see a major issue with the way it has been defined by the research so far. This standard definition of a “hot hand” allows incredibly high scores for games with incredibly low field goal percentages. Take the string [0 0 0 0 0 0 0 0 0 0 1 1 1] for example, this string would have a test statistic of 90% (*n*=1) but would only translate to 3 for 13 (23%) from the field. I can hardly consider that a “hot” night. Conversely, a player could make every shot except for one [1 1 1 1 1 1 1 1 1 1 0 1], shoot 11 for 12 (92%) from the field, but only land a measly test statistic of - 8% because they were 1 for 1 (100%) following their single miss. With this issue in mind, I’d like to explore more ways of framing this problem in the future.

As a final note, when we don't find evidence of an alternative hypothesis, we cannot reject it. We can only fail to reject the null. A lack of evidence only means that you haven’t proven that something exists. It does not prove that something doesn’t exist. So is it fair to dismiss the "hot hand" as a fallacy?

References

Gilovich, Thomas, Robert Vallone & Amos Tversky (1985), ‘The hot hand in basketball: on the misperception of random sequences’, Cognitive Psychology 17, 295–314.

Miller, J. B., & Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. Econometrica, Vol. 86, No.6, pp. 2019–2047, Available at SSRN: https://ssrn.com/abstract=2627354 or http://dx.doi.org/10.2139/ssrn.2627354