The Hot Hand Phenomenon

Leon Wang LJW7736@NYU.EDU
Peter Hu RH3659@NYU.EDU
Yuan Li YL10134@NYU.EDU

Milestone 2 – Midway Checkpoint

1. Background and Methodology

1.1 Two approaches

There are two approaches in the study of the hot hand phenomenon. The first one, developed in GVT (Gilovich and Tversky, 1985) and analyzed by Miller and Sanjurjo (Miller and Sanjurjo, 2018), is to study the phenomenon in controlled environments. For example, in basketball, GVT studied free throw attempts and three-point contests. In another GVT's controlled shooting experiment, each of the 26 players from the Cornell University Men's (14) and Women's (12) basketball teams shot 100 times at a distance from which the experimenters determined he/she would make around 50% of the shots. The reason for using controlled shooting data is that, when there is no hot hand, the sequence of make-or-miss shooting results for each player is independent, identically distributed Bernoulli trials with a fixed probability of success.

The second type of approach, by Pelechrinis and Winston (Pelechrinis and Winston, 2022), uses shooting data from actual NBA games. The key difference between this approach and the first type of controlled shooting approach is that ”in-game situations, where different shots are taken under different conditions and hence, have different probability of success.” (Pelechrinis and Winston, 2022)

1.2 What is the hot hand phenomenon really about?

In the first approach of controlled shootings, each player's shooting results can be treated as a sequence of independent, identically distributed Bernoulli random variables. In real game situations, as in the second approach, each shot has a different probability of success. Thus, the Bernoulli random variables are not identically distributed. However, both approaches share one thing in common—the core of the hot hand phenomenon: under the null hypothesis of no hot hand, the Bernoulli random variables of the outcomes of each shot are independent. Thus, the hot hand phenomenon is not about whether there are variations in the probability of making each shot; rather, it is about whether there are correlations among the outcomes of the shots by a player.

1.3 Streak selection bias

Under the null hypothesis of no hot hand, whether the current shot is a ”make” (M) or ”miss” (X) does not depend on the outcomes of previous shots.

Using the notations of Pelechrinis and Winston, let $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$ be the conditional probability of observing an M following k consecutive M, and let $\Pr[\mathrm{M}|\underbrace{\mathrm{X}\ldots\mathrm{X}}_{\text{k times}}]$ be the conditional probability of observing an M following k consecutive X. In the absence of the hot hand, $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}] = \Pr[\mathrm{M}|\underbrace{\mathrm{X}\ldots\mathrm{X}}_{\text{k times}}] = \Pr[\mathrm{M}]$.

GVT performed a paired t-test on $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$ - $\Pr[\mathrm{M}|\underbrace{\mathrm{X}\ldots\mathrm{X}}_{\text{k times}}] = 0$ for k = 1, 2, 3. And they found no evidence to reject the null hypothesis that there is no hot hand.

A natural way of estimating $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$, which is used by GVT, is to count the proportion of successes among the outcomes that immediately follow a streak of k consecutive successes. However, as Miller and Sanjurjo (Miller and Sanjurjo, 2018) point out, such an estimator is biased, since it is strictly less than the true $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$. This is so-called *Streak Selection Bias*. Furthermore, the bias is more than doubled when using the same method to estimate $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$ - $\Pr[\mathrm{M}|\underbrace{\mathrm{X}\ldots\mathrm{X}}_{\text{k times}}]$. After adjusting the streak selection bias, Miller and Sanjurjo find that the null hypothesis of no hot hand is rejected using the same data from GVT.

1.4 Permutation test

After pointing out the streak selection bias, Miller and Sanjurjo suggest the permutation test. If the outcomes of a player's shots are i.i.d. Bernoulli random variables with a fixed success probability, then any random permutation of the observed sequence has the same chance of occurring. Therefore, we can obtain the value of $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$ by going through all the permutations, via calculation or simulation. With the existence of the streak selection bias, we can reject the null hypothesis of no hot hand if $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{obs}$ - $\Pr[\mathrm{M}|\underbrace{\mathrm{X}\ldots\mathrm{X}}_{\text{k times}}]_{obs}$ exceeds $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{perm}$ - $\Pr[\mathrm{M}|\underbrace{\mathrm{X}\ldots\mathrm{X}}_{\text{k times}}]_{perm}$ by the significance level.

Pelechrinis and Winston extend the permutation test to the case where the probability of success is different for each shot. If a player makes m shots with the known probability of success of $p_1, p_2, \ldots, p_m$, they first randomly generate m independent Bernoulli trials based on $p_1, p_2, \ldots, p_m$ to form the simulated sequence. Using a number of simulated sequences, we can obtain $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{sim}$. If $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{obs}$ exceeds $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{sim}$ by the significance level, we can reject the null hypothesis of no hot hand.

1.5 Our works

1.5.1 Replicate Pelechrinis and Winston model

Pelechrinis and Winston's paper consists of two parts, as shown in the figure below.

Shot-Make Probability Model $\longrightarrow$ Streak Conditional Probability Test

In the first part, a neural network classifier with 3 hidden layers was trained to predict shot-make probabilities based on various factors, such as distance to the basket, distance of the closest defender, etc. In the paper, PW states that their model achieves an out-of-sample accuracy of 66%. With the shot-make probability model, Pelechrinis and Winston perform the streak conditional probability tests by estimating $\mathrm{Pr}[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{obs}$ - $\mathrm{Pr}[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]_{sim}$ on NBA players shooting sequences of each game. Their study showed that while individual players can exhibit the hot hand, the average player tends to regress to their mean performance after consecutive successful shots.

Based on the code supplemented to the paper by Pelechrinis and Winston, we are able to replicate their results. We run their neural network model on the same two-season data (2013, 2014) and achieve around 65% accuracy, depending on runs.

1.5.2 New shot-make probability prediction model using LightGBM

The model of Pelechrinis and Winston selects some features that seems having high power in predicting the probability of making a shot. Those features include Distance to the basket, Distance of the closest defender, Player IDs for the shooter and the closest defender, Touch time prior to shooting, and Number of dribbles before shooting. They claim that their model "is on par with the state-of-the-art shot make probability models." (Pelechrinis and Winston, 2022).

We build a classification model based on the same data used by Pelechrinis and Winston. Our model use LightGBM as the classifier with the following parameter:

```
params = {
    'objective': 'binary',
    'metric': 'binary_logloss',
    'boosting_type': 'gbdt',
    'num_leaves': 31,
    'learning_rate': 0.05,
    'feature_fraction': 0.9,
    'verbose': -1, # Suppress verbose output
    'n_estimators': 100 # Number of boosting rounds
}
```

On the same test data, our model performs slightly better than the model of Pelechrinis and Winston.

1.5.3 New tests for hot hand

The streak conditional probability test based on $\Pr[\mathrm{M}|\underbrace{\mathrm{M}\ldots\mathrm{M}}_{\text{k times}}]$ was introduced by GVT and has been the dominate concept of testing hot hand. Miller and Sanjurjo point out the problem of streak selection bias and suggest permutation test to fix it. Pelechrinis and Winston externd the permutation test into simulation. They still perform the streak conditional probability test, but with unequal probability of success.

The streak conditional probability test for the hot hand is intuitive. It tests the hypothesis that the chance of making the next shot is independent of just making a streak of several shots. However, it suffers the streak selection bias, and it is awkward in estimation.

We propose a new test for the hot hand based on the number of streaks. When a player has a hot hand during the game, the number of k-streaks (k=2, 3, 4, etc) is higher than the case of independent outcomes, even with his probability of making each shot built into the consideration. For example, if a player makes 20 shots and each shot has a 0.45 probability of making. Then the expected number of 3-streaks is $(20 - 3 + 1) * 0.45^3 = 1.64$. If he makes 5 3-streaks, for example, he has the shooting record [X,M,M,M,M,X,X,X,X,M,M,M,M,M,X,X,X,X,X,X,X], then it seems he has unusually higher chance of making streak shots, thus, a hot hand during the game.

In our streak number test, we first use a shot-make probability model to predict the probability of success for each shot a player makes during a game. Then we use simulation to derive the distribution of the number of k-streaks with the predicted shot-make probability and assuming each shot is independent. We then determine the critical number of shots, c_{99}, for the best 99% level. That is, the probability that the number of k-streaks is greater than c_{99} is less than 1%. We can compare the actual number of k-streaks with c_{99} to see if the player has achieved this extraordinary performance during the game. If a player plays 100 games, we should expect him to have an average of 1 game to have an extraordinary performance. If he has 10 such games, then we can say he has a hot hand during the 100 games.

2. Results

2.1 Shot-make probability prediction models

In the paper of Pelechrinis and Winston, they have two shot-make probability models trained on the data of two seasons (2013 and 2014). Then they test the 2013 trained model on 2014 data and the 2014 trained model on 2013 data. In this way, they avoid data leakage while at the same time maximizing the use of data.

We run the neural network model of Pelechrinis and Winston and find the accuracy at 64.87% and AUROC = 69%. Below is the ROC curve of their model.

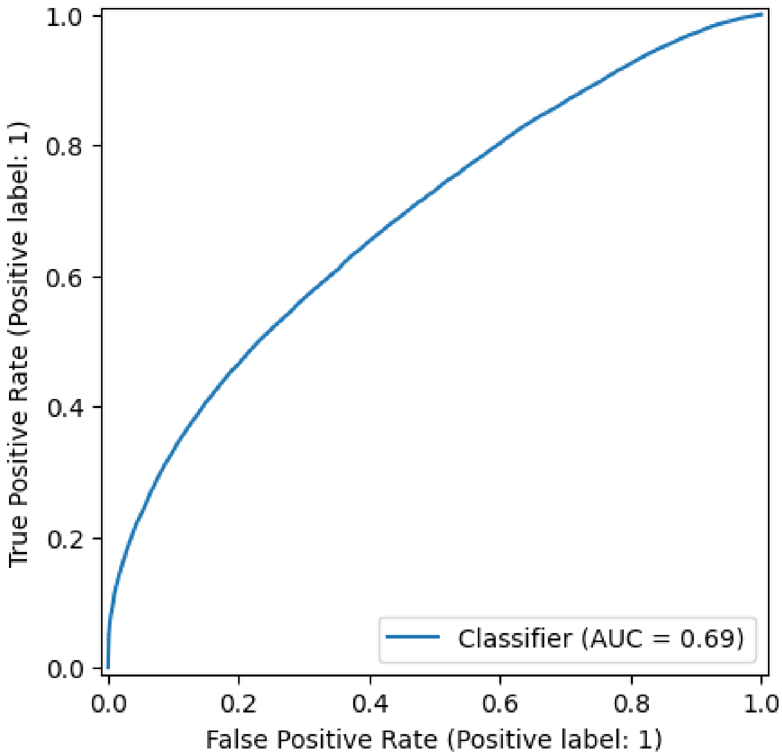

Figure 1: PW Neural Network Model, Accuracy 64.87%

We train and test our LightGBM model based on the same procedure and data as in Pelechrinis and Winston and find the accuracy at 65.46% AUROC = 70%. Below is the ROC curve of our model.

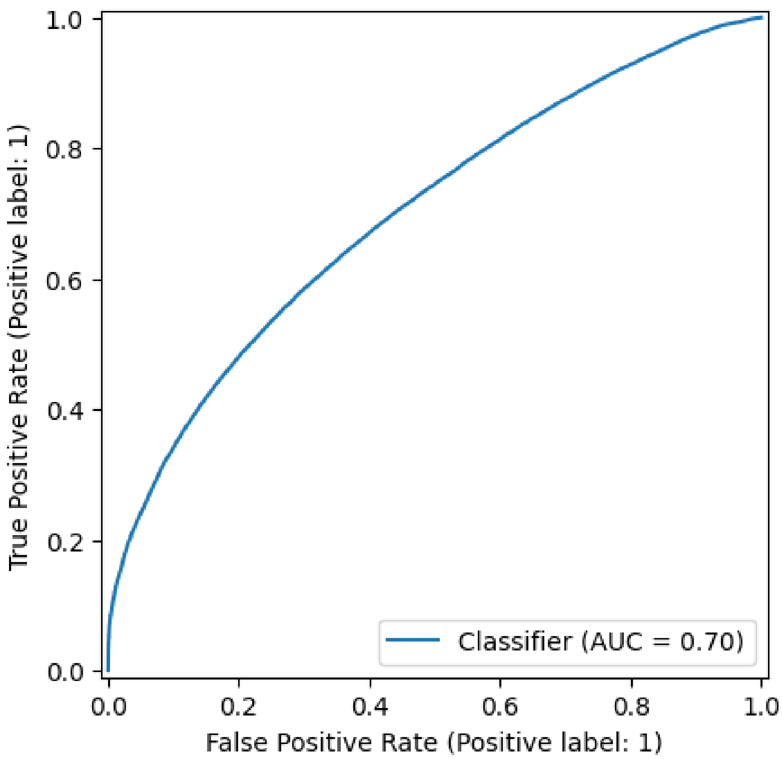

Figure 2: LightGBM Model, Accuracy 65.46%

2.2 Streak Number Tests Confirm Hot Hand

We apply our streak number test on the same 153 players who took at least 1,000 shots over the two seasons with the same data as Pelechrinis and Winston use. For k = 2, with a significance level of 0.1%, we find that 43 out of 153 players had that many high-streak games. Thus, for k = 2, we can reject the null hypothesis of no hot hand.

Here are the best 20 players in terms of 2-streak number performance.

...

Table 1: Top 20 Players in Streak Number Test, k = 2

Player	**N. of Games**	**High-Streak games**	**% of HS games**	**p-value**
Tristan Thompson	140	15	10.71%	1.33e-12
Marreese Speights	117	12	10.26%	2.35e-10
Evan Fournier	104	11	10.58%	7.38e-10
Jordan Hill	106	11	10.38%	9.23e-10
Amir Johnson	132	11	8.33%	1.15e-8
Timofey Mozgov	143	10	6.99%	2.57e-7
Kentavious Caldwell-Pope	125	9	7.20%	6.24e-7
Gerald Henderson	133	9	6.77%	1.10e-6
Tim Hardaway Jr.	114	8	7.02%	2.52e-6
Spencer Hawes	126	8	6.35%	5.75e-6
Darren Collison	80	6	7.50%	1.68e-5
Ersan Ilyasova	88	6	6.82%	3.13e-5
Robin Lopez	122	7	5.74%	3.50e-5
David Lee	95	6	6.32%	5.12e-5
PJ Tucker	130	7	5.38%	5.51e-5
Marco Belinelli	107	6	5.61%	1.09e-4
D.J. Augustin	107	6	5.61%	1.09e-4
Danny Green	111	6	5.41%	1.37e-4
Jodie Meeks	112	6	5.36%	1.45e-4
Mike Conley	85	5	5.88%	2.23e-4

3. Plan for Additional Analysis

3.1 Refining Shot-Make Models and Testing Procedure

Both the neural network model by Pelechrinis and Winston and our LightGBM model have a potential to be improved. We will use some tools learned from the class, such as the Hyperparameter sweeps with Weights and Biases Framework, to search for good hyperparameters for LightGBM.

Also, in the testing of model accuracy and testing for a hot hand, players with combined shots of two seasons fewer than 1000 shots are excluded. This may exclude some centers with fewer shots but high likelihood of streak shots. It would be interesting to explore the impact of relaxing the 1000-shot restriction.

3.2 Incorporating Newer Data

Pelechrinis and Winston use the 2013 and 2014 two-season data. Around 2016, the download of the shot-by-shot data from NBA does not include some detailed tracking data like "touch time" or "closest defender ID" anymore. The neural network model by Pelechrinis and Winston and our LightGBM model use several of those detailed tracking data, and hence, no longer work for newer data. We will explore if it is possible to build a shot-make probability model with the current available information.

3.3 Comparing streak number test with streak conditional probability test

Pelechrinis and Winston give a list of 24 players with a hot hand for the conditional 1-streak test. This is equivalent to our k=2 streak number test. However, only 6 of our top 20 hot hand players are in the list of Pelechrinis and Winston. We will investigate the difference between the tests and how they select different players of a hot hand.

3.4 LLMs' view of hot hand

Hot hand phenomenon is a relatively popular topic in sports. It would be interesting to see how LLMs view this issue. We will explore how ChatGPT responds to the prompts about a hot hand. For example, what are its criteria for hot hand, does the LLM do its statistical calculations?.

Prompt 1:

An NBA player can have a hot hand during a game. We use 1 for a successful shot attempt and 0 for an unsuccessful one. Below are the shooting records of several games of Chris Bosh. Which game do you think the player has a hot hand?

Game, Shooting Record

NOV 23 2014 - MIA vs. CHA , 1000011110100010000101

DEC 10 2014 - MIA @ DEN , 001101001001

FEB 01 2015 - MIA @ BOS , 0101100000010100010

4. Work Plan

We will all be fully involved in the project.

Tentative weekly plan:

1. Week of 11/16: Work on Additional Analysis, 3.1, 3.2
2. Week of 11/23: Work on Additional Analysis, 3.3, 3.4 and other ideas
3. Week of 11/30: Finish the additional work
4. Week of 12/07: Wrap up, write the final report

References

Vallone R. Gilovich, T. and A. Tversky. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3):295–314, 1985.

J. B. Miller and A. Sanjurjo. Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica*, 86(6):2019–2047, 2018.

K. Pelechrinis and W. Winston. The hot hand in the wild. *PLoS ONE*, 17(1):e0261890, 2022.