Wharton SPORTS ANALYTICS Student Research Journal

2022, Vol 4

wsb.wharton.upenn.edu/student-research-journal

Hot Hand Fallacy, Fallacy, Fallacy?

Kevin Babitz*, W'21*

The Wharton School of the University of Pennsylvania, Philadelphia PA, USA

Advisor: Abraham Wyner, PhD

The Wharton School of the University of Pennsylvania, Philadelphia PA, USA

Abstract

Common across many domains, and especially prevalent in basketball, the 'hot hand' suggests that a person who has experienced a recent period of success has a greater likelihood of future success. In basketball, the player with the hot hand who has recently made a series of baskets is more likely to continue their recent success. The seminal research on the hot hand, identifies it as a fallacy after observing 26 Cornell basketball players' shooting sequences (Gilovich et al., 1985). The shooting percentages observed after a hot streak should be less than those after a cold streak when holding the probability of success and length of the sequence constant. This work instead looks at how much different what we observe is from what we would expect. We call this expected negative difference between the hot and marginal or cold shooting probability a bias adjustment (Miller and Sanjuro, 2018). We extend the hot hand literature in three ways: we add permutation tests and apply more robust hypothesis testing to the resulting p-values to the Cornell dataset, we demonstrate the effectiveness of a new formula to approximate the bias term in the literature, and we apply the hot hand analysis framework to PGA putting data from the 2019 season. The results suggest that the Gilovich et al. paper was close to accurate in determining if the hot hand was true, even with their failure to recognize the bias adjustment needed in the analysis. The hot hand appears to be hard to detect statistically in the Cornell basketball player dataset and PGA putting data using the simulation framework discussed in these analyses. Additionally, we have found a relatively accurate approximation of the bias term used in correctly assessing if a player is significantly hot.

key words: basketball, golf, hot hand, shot

Introduction

The hot hand literature's seminal paper calls the hot hand a fallacy after observing 26 Cornell basketball players' shooting sequences. In this paper, they observed that many of the player's percentage of made shots on a hot streak in a sequence of 100 shots were less than their percentage on a cold streak. They claim only one of the players is significantly hot (Gilovich et al., 1985). Later research showed that in a randomly generated sequence, we would expect see the results seen in the original paper. The shooting percentages observed after a hot streak should be less than those after a cold streak when holding the probability of success and length of the sequence constant. For that reason, we must instead look at how much different what we observe is from what we would expect. We call this expected negative difference between the hot and marginal or cold shooting probability a bias adjustment (Miller and Sanjuro, 2018).

We explore and extend the hot hand literature in three ways: we add permutation tests and apply more robust hypothesis testing to the resulting p-values to the Cornell dataset, we demonstrate the effectiveness of a new formula to approximate the bias term in the literature, and we apply the hot hand analysis framework to PGA putting data from the 2019 season.

Methods

Study 1: Permutation Tests on Cornell Basketball Data

We use the same data that both the Gilovich, Vallone, and Tversky (GVT) and Miller and Sanjuro (MS) papers relied on for our analysis. The data contain shot data from 26 Cornell basketball players, 14 males and 12 females. Each player shot 100 times, from a position on the court they believed to be a 50% probability of success for themselves. From the data, marginal shooting percentages and shooting percentages on hot and cold streaks of length one to three were calculated. The data are seen in Figure 1.

Player	P(hit| 3 misses)	P(hit| 2 misses)	P(hit| 1 miss)	P(hit)	P(hit| 1 hit)	P(hit| 2 hit)	P(hit| 3 hit)
Male							
M1	0.44	0.5	0.61	0.54	0.49	0.48	0.5
M2	0.43	0.33	0.35	0.35	0.35	0.25	0
M3	0.67	0.68	0.49	0.6	0.67	0.62	0.6
M4	0.47	0.45	0.43	0.4	0.36	0.23	0.33
M5	0.75	0.6	0.47	0.42	0.36	0.4	0.33
M6	0.25	0.38	0.48	0.57	0.65	0.62	0.65
M7	0.29	0.5	0.47	0.56	0.64	0.63	0.65
M8	0.5	0.5	0.52	0.5	0.46	0.64	0.57
M9	0.35	0.33	0.35	0.54	0.72	0.79	0.83
M10	0.57	0.5	0.64	0.59	0.79	0.6	0.57
M11	0.57	0.61	0.56	0.58	0.59	0.62	0.62
M12	0.41	0.43	0.46	0.44	0.42	0.39	0.43
M13	0.4	0.62	0.67	0.61	0.58	0.56	0.5
M14	0.5	0.62	0.6	0.59	0.58	0.59	0.6
Female							
F1	0.67	0.61	0.55	0.48	0.42	0.45	0.33
F2	0.43	0.36	0.31	0.34	0.41	0.36	0.4
F3	0.36	0.38	0.33	0.39	0.49	0.42	0.5
F4	0.27	0.33	0.34	0.33	0.29	0.33	0.33
F5	0.22	0.36	0.34	0.35	0.37	0.5	0.2
F6	0.54	0.58	0.52	0.46	0.38	0.41	0.29
F7	0.32	0.28	0.36	0.41	0.49	0.65	0.62
F8	0.67	0.55	0.57	0.53	0.5	0.58	0.73
F9	0.46	0.55	0.47	0.45	0.41	0.47	0.5
F10	0.32	0.34	0.46	0.47	0.47	0.67	0.71
F11	0.5	0.56	0.51	0.53	0.56	0.5	0.39
F12	0.32	0.32	0.27	0.25	0.2	0	-
Average	0.45	0.47	0.47	0.47	0.48	0.49	0.49

Figure 1: Cornell Basketball Player Data from GVT Paper (1985)

To determine significance of hotness in the data, we formally state the null hypothesis saying that all shots are independent from each other. To conduct significance testing in the Cornell data, we need independently generated sequences to compare to the observed data. We use permutations of the observed Cornell data to do this.

Specifically, we create a sequence of 100 shots (represented as 0's and 1's for failure and successrespectively) with an underlying probability of success of the observed probability of success, named P(hit) in Figure 1. We then permute each of these sequences 10,000 times. Permutations were used as they allowed us to assume that the player's "true" shooting percentage is the one we observe in the study. We can then observe sequences with independence between each shot while making minimal changes to the underlying data.

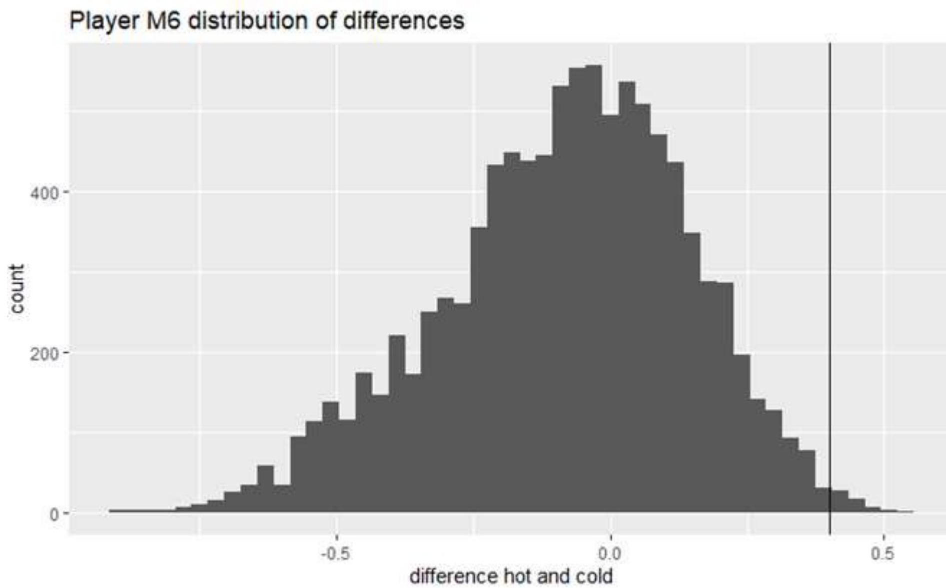

Figure 2: Male 6's distribution of differences between hot and cold success rates in 10,000 permutations

From these permuted sequences, we can create a distribution of expected differences between hotand cold and hot and marginal shooting probabilities under the null hypothesis. We then comparethese distributions to the observed values to calculate p-values. An example of one of these distributions and comparison of the observed difference can be seen in Figure 2.

From these distributions, we calculate p-values to test for significance in hot handedness when comparing hot and cold shooting success rates (seen in Figure 3) and hot and marginal shooting success rates (seen in Figure 4).

We observe, when using a significance level of 0.05, that five of the players exhibit a significantly hot hand when comparing the hot and cold shooting success rates and just three of the players is significantly hot when comparing hot and marginal shooting success rates. If we use a Bonferroni correction to adjust the false discovery rate for multiple hypotheses, we observe that just one of the players is significantly hot at a significance level of 0.002 (Male number 9).

To calculate significance in the overall data, we use a permuted t-distribution and two methods for combining multiple p-values. To generate the t-distribution, we permute each player's sequence once, sum the bias adjusted differences between hot and cold or hot and marginal success rates, and do this 10,000 times. We then compare the observed sum of bias adjusted differences to this simulated distribution.

Player	P (hit)	P(hit| 3 hits)	P(hit| 3 misses)	GVT est.	Permutation Bias Adj.	p_value
Males						
M1	0.54	0.5	0.44	0.06	0.14	0.28
M2	0.35	0	0.43	-0.43	-0.33	0.9
M3	0.6	0.6	0.67	-0.07	0.02	0.5
M4	0.4	0.33	0.47	-0.14	-0.05	0.6
M5	0.42	0.33	0.75	-0.42	-0.33	0.9
M6	0.57	0.65	0.25	0.4	0.48	0.01*
M7	0.56	0.65	0.29	0.36	0.44	0.02*
M8	0.5	0.57	0.5	0.07	0.14	0.26
M9	0.54	0.83	0.35	0.48	0.56	0*
M10	0.59	0.57	0.57	0	0.09	0.38
M11	0.58	0.62	0.57	0.05	0.13	0.3
M12	0.44	0.43	0.41	0.02	0.1	0.36
M13	0.61	0.5	0.4	0.1	0.19	0.23
M14	0.59	0.6	0.5	0.1	0.18	0.21
Females						
F1	0.48	0.33	0.67	-0.34	-0.26	0.88
F2	0.34	0.4	0.43	-0.03	0.07	0.42
F3	0.39	0.5	0.36	0.14	0.23	0.18
F4	0.33	0.33	0.27	0.06	0.16	0.29
F5	0.35	0.2	0.22	-0.02	0.08	0.41
F6	0.46	0.29	0.54	-0.25	-0.17	0.79
F7	0.41	0.62	0.32	0.3	0.39	0.04*
F8	0.53	0.73	0.67	0.06	0.14	0.29
F9	0.45	0.5	0.46	0.04	0.12	0.31
F10	0.47	0.71	0.32	0.39	0.47	0.01*
F11	0.53	0.39	0.5	-0.11	-0.03	0.59
F12	0.25	-	0.32	-	-	
Average	0.47	0.49	0.45	0.04	0.12	0.31

Player	P(hit)	P(hit| 3 hits)	Marginal Difference	Marginal Bias Adj.	P-Value
Male					
M1	0.54	0.5	-0.04	-0.01	0.50
M2	0.35	0	-0.35	-0.27	0.66
M3	0.6	0.6	0	0.02	0.42
M4	0.4	0.33	-0.07	0.00	0.59
M5	0.42	0.33	-0.09	-0.03	0.64
M6	0.57	0.65	0.08	0.11	0.16
M7	0.56	0.65	0.09	0.11	0.16
M8	0.5	0.57	0.07	0.11	0.25
M9	0.54	0.83	0.29	0.32	0.00*
M10	0.59	0.57	-0.02	0.01	0.52
M11	0.58	0.62	0.04	0.06	0.29
M12	0.44	0.43	-0.01	0.04	0.42
M13	0.61	0.5	-0.11	-0.09	0.79
M14	0.59	0.6	0.01	0.03	0.38
Female					
F1	0.48	0.33	-0.15	-0.11	0.80
F2	0.34	0.4	0.06	0.14	0.26
F3	0.39	0.5	0.11	0.18	0.15
F4	0.33	0.33	0	0.09	0.45
F5	0.35	0.2	-0.15	-0.07	0.60
F6	0.46	0.29	-0.17	-0.12	0.77
F7	0.41	0.62	0.21	0.27	0.07
F8	0.53	0.73	0.2	0.23	0.02*
F9	0.45	0.5	0.05	0.10	0.24
F10	0.47	0.71	0.24	0.28	0.02*
F11	0.53	0.39	-0.14	-0.11	0.80
F12	0.25	-	-	-	-
Average	0.47	0.49	0.02	0.06	0.41

Figure 3 (left) and 4 (right): p-values for each player comparing hot and cold success rates(left) and hot and marginal success rates (right)

When comparing hot and cold success rates, we observe p-values of 0.0115, 0.002, and 0.0008 using our t-test, Stouffer’s Method, and Fisher’s Method, respectively. When comparing hot shooting success rates to marginal shooting success rates, we observe p-values of 0.0377, 0.0165, and 0.0209 using the same methods above.

Finally, we observe that removing the extremely significant player from the data gives us a large change in significance of the hot hand in the overall data. Using hot and cold probabilities, we observe p-values of 0.0101, 0.0115, and 0.0109 using our t-test, Stouffer’s Method, and Fisher’s Method, respectively. When using marginal and hot shooting percentages, we observe p-values of 0.0329, 0.0619, and 0.1456 for the same methods above.

Overall, we observe that only a few players potentially demonstrate a hot hand in the Cornell data. This is similar to the finding in the original GVT paper in 1985 and conflicts with findings from the MS paper from 2018. We observe much of the overall significance is driven by the 9th male player and the significance is magnified by comparing hot and cold shooting percentages rather than hot and marginal shooting percentages.

Study 2: Bias Adjustment Value Formula

We use a similar simulation approach to demonstrate the strength of a formula that does a relatively good job of approximating the bias adjustment value that is necessary for understanding if a player is hot.

We claim the following formula,

(1) $$p \approx \left(1 + \mathrm{E}\left[\frac{1}{K} \mid K > 0\right]\right) * \mathrm{Cond} - \hat{\mathrm{p}}(\mathrm{L}),$$

where $\mathrm{E}\left[\frac{1}{K} \mid K > 0\right]$ is the expected value of 1 over the number of hot streaks of length L observed given that a hot streak occurs in the sequence,

$\mathrm{Cond} - \hat{\mathrm{p}}(\mathrm{L})$ is the probability of a successful shot being observedafter a streak of length L given a hot streak occurs in the sequence, and

p is the underlying probability of a success for the sequence.

To test the success of this formula, we generate permuted sequences of lengths 50, 100, and 200 with probabilities ranging from 0.3 to 0.7 in increments of 0.05. We generate 10,000 permuted sequences for each length and probability combination and calculate the values in the formula above. We observe the results in figures 5, 6, and 7 below.

Permutatation Bias Calculations for n = 50

pi	E(K)	E(K | K > 0)	E(1 / K)	E(1 / K | k > 0)	P-Hat(L)	Cond-P-hat(L)	(1 + E[1/K| K> 0])* Cond-P-Hat(L)
0.3	1.165	1.7511	0.4775	0.7177	0.1084	0.1629	0.2799
0.36	2.0746	2.4042	0.4838	0.5607	0.1971	0.2284	0.3565
0.4	2.916	3.0815	0.4179	0.4416	0.2666	0.2817	0.4061
0.46	4.5662	4.5993	0.2801	0.2822	0.3631	0.3657	0.4689
0.5	5.8959	5.9036	0.2053	0.2056	0.422	0.4226	0.5095
0.56	9.2542	9.2542	0.1178	0.1178	0.5246	0.5246	0.5864
0.6	10.418	10.418	0.1029	0.1029	0.5559	0.5559	0.6131
0.66	13.9135	13.9135	0.0749	0.0749	0.6227	0.6227	0.6693
0.7	16.7271	16.7271	0.0614	0.0614	0.6719	0.6719	0.7131

Figure 5: Bias Adjustment formula calculations for sequences of length 50

Permutatation Bias Calculations for n = 100

pi	E(K)	E(K | K > 0)	E(1 / K)	E(1 / K | k > 0)	P-Hat(L)	Cond-P-hat(L)	(1 + E[1/K| K> 0])* Cond-P-Hat(L)
0.3	2.5294	2.7943	0.4471	0.494	0.1826	0.2017	0.3014
0.35	4.0395	4.1304	0.3215	0.3287	0.261	0.2669	0.3546
0.4	6.1775	6.1899	0.1998	0.2002	0.3399	0.3406	0.4088
0.45	8.8032	8.8032	0.1291	0.1291	0.4034	0.4034	0.4555
0.5	12.1302	12.1302	0.089	0.089	0.4634	0.4634	0.5047
0.55	17.1549	17.1549	0.0608	0.0608	0.5311	0.5311	0.5634
0.6	22.2151	22.2151	0.0462	0.0462	0.5884	0.5884	0.6156
0.65	26.9376	26.9376	0.0378	0.0378	0.6312	0.6312	0.655
0.7	33.8451	33.8451	0.0299	0.0299	0.6863	0.6863	0.7068

Figure 6: Bias Adjustment formula calculations for sequences of length 100

Permutatation Bias Calculations for n = 200

pi	E(K)	E(K | K > 0)	E(1 / K)	E(1 / K | k > 0)	P-Hat(L)	Cond-P-hat(L)	(1 + E[1/K| K> 0])* Cond-P-Hat(L)
0.3	5.1953	5.2383	0.2512	0.2532	0.2405	0.2425	0.3039
0.35	8.3652	8.3669	0.1406	0.1406	0.312	0.312	0.3559
0.4	12.5351	12.5351	0.0871	0.0871	0.3722	0.3722	0.4047
0.45	17.8572	17.8572	0.0591	0.0591	0.4256	0.4256	0.4507
0.5	24.6047	24.6047	0.042	0.042	0.482	0.482	0.5022
0.55	33.7724	33.7724	0.0302	0.0302	0.541	0.541	0.5573
0.6	43.8265	43.8265	0.0231	0.0231	0.5945	0.5945	0.6083
0.65	54.4549	54.4549	0.0185	0.0185	0.6409	0.6409	0.6528
0.7	68.2117	68.2117	0.0147	0.0147	0.6931	0.6931	0.7033

Figure 7: Bias Adjustment formula calculations for sequences of length 200

In comparing the right-most column to the left-most column, we observe that the hypothesized formula does very well in approximating the underlying probability of the sequences for longer sequences and does relatively well with shorter sequences.

The success of this formula suggests that if we are confident about a player's true shotting percentage and have a good approximation for one over the expected value of the number of

shots taken on a hot streak, we can accurately estimate the player's expected probability of success on a hot streak and the bias adjustment value.

Study 3: Applying Hot Hand Analysis to PGA Putting Data

We conclude this analysis with an application to professional sports data. It is interesting to apply the framework used to analyze the Cornell basketball shooting dataset to a professional sports context. For this analysis, we use putting data from all PGA tournaments in 2019 tracked on PGA's ShotLink tracker. We rely on the Strokes Gained statistic, a measure of how good a shot is compared to what the average golfer's expected outcome would be. Strokes gained is calculated by comparing the number of strokes expected to get the ball in the hole from an old position to the position after the ball is hit and then by subtracting 1 from that difference. These expectations are computed separately for each location on the golf course (fairway, rough, green, etc.) in binned distances of either yards for non-green shots or inches for shots on the green.

For the hot hand analysis, we make three assumptions: a successful putt has a positive or 0 value for strokes gained, all putts in a single round make up a sequence of putts that can be used to consider if a player is “hot” and all rounds are separate from each other, and all players have the same expected strokes from each distance bin and location on a green regardless of the player. Strokes gained is a widely used metric for measuring success of a stroke in golf and these are reasonable assumptions when using this metric.

To calculate hotness in the data, we begin by creating success and failure sequences for each player and for each round using strokes gained (the putt is successful if it has a zero or larger strokes gained value for a particular putt). We then calculate hot and marginal probability of success in each of these sequences where a hot streak is defined as three or more successes in a row. 20,000 permutations are then created for each unique combination of probability and sequence length. Again, hot probabilities and marginal probabilities are calculated for each permuted sequence and sequences with no hot streaks are removed. Finally, we compare the permuted distributions to the observed values for each round to determine statistically significant hot rounds in the data.

From this analysis, we end up with 4425 total rounds from the 2019 season that had at least one hot streak in them. From these rounds, only 72 of them (1.63%) are determined to be significantly hot using the above procedure. No individual player had more than 3 hot rounds in the 2019 season and a majority did not have any hot rounds.

Conclusion

The results in studies 1 and 2 suggest that the GVT paper was very close to accurate in determining if the hot hand was true, even with their failure to recognize the bias adjustment needed in the analysis. The hot hand appears to be hard to detect statistically in the Cornell basketball player dataset and PGA putting data using the simulation framework discussed in these analyses.

Additionally, we have found a relatively accurate approximation of the bias term used in correctly assessing if a player is significantly hot. We note that the approximation is significantly more accurate as the length of the sequence being analyzed increases but performs relatively well on shorter sequences.