Do Steph Curry and Klay Thompson Have Hot Hands?

Alon Daks Nishant Desai Lisa Goldberg

February 20, 2019

Berkeley Math Circle

The hot hand in basketball

The hot hand in basketball

The hot hand is a widely accepted phenomenon in basketball.

Pass to a player on a scoring streak to benefit from an elevated probability of a hit.

The hot hand in basketball

photos by Ken Ribet

The hot hand in basketball

The original hot hand study

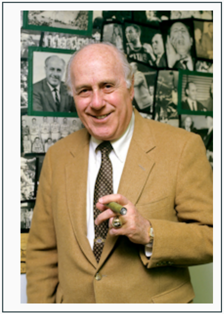
In 1985, three researchers published a study showing that the hot hand in basketball was a cognitive illusion.

Using data on the Philadelphia 76ers, the Boston Celtics, and men's and women's teams at Cornell University, Thomas Gilovich, Robert Vallone and Amos Tversky found no statistical evidence that a player was more likely to score after a string of successful shots than at any other time.

The result was unexpected

and while scholars found the result interesting, its reception in the sports community was lukewarm.

The reception



When told about the finding, NBA President Red Auerbach, who had coached the Boston Celtics during their 1956–1966 dynasty, said (of Amos Tversky):

> *So he makes a study. I couldn't care less.*

Framework of the original hot hand study

The null hypothesis in [Gilovich et al., 1985] was that a hit was just as likely to follow a string of hits as a string of misses.

But they did not find a statistically significant difference in the data

and failed to reject their null hypothesis.

The small sample bias

Realization of the null hypothesis in a finite data set

Thirty years later, two statisticians, Josh Miller and Adam Sanjurjo, noticed that Gilovitch, Vallone and Tversky had not properly set up their own null hypothesis.

A coin flipping experiment illustrates the point

Flip a fair coin three times.

Record the outcome of a flip following an observation of heads.

What is the expected fraction of heads in the recorded string?

The bias

Realization	Record	Fraction
TTT		
TTH		
THT	T	0.0
HTT	T	0.0
HHT	HT	0.5
HTH	T	0.0
THH	H	1.0
HHH	HH	1.0
		2.5

On average, the likelihood that a head follows a head is:

$$\frac{2.5}{6} < 0.5$$

The importance of the bias

Under the null hypothesis of identical independent coin flips,

a reversal is more likely than a continuation in a finite data set.

[Gilovich et al., 1985] neglected this small sample effect.

The magnitude of the bias

It depends on the length of the conditioning set and the length of the sequence.

It can be substantial, even for a sequence of length $N = 100$, and certainly for sequences of whose length ranges between $N = 20$ and $N = 30$, roughly the number of shots a star shooter takes in a game, or a team takes in a quarter.

The revised finding

Miller and Sanjurjo revisited the Cornell study in [Gilovich et al., 1985], where each subject made $N = 100$ shots.

When the small sample correction was applied to generate a correct null hypothesis, the original finding was reversed in some examples.

Measuring hot-handedness in a finite setting with modern data science

Thompson versus Pistons

Consider a binary string representing hits (1's) and missed shots (0's), like this one:

$$111010011000011$$

That string describes Klay Thompson's shooting record, including field goals and free throws, against the Detroit Pistons on December 13, 2016.

Thompson made half his shots and missed half his shots.

The conditional probability of a hit, given two hits

111010011000011

There are four instances of two hits in a row, indicated by the string "11."

In one case, Thompson made the next shot. In two cases, he missed the next shot. In one case, nothing happened.

The conditional probability of a hit, given two hits

$$11\textcolor{orange}{10}10011\textcolor{orange}{0}000011$$

There are four instances of two hits in a row, indicated by the string “11.”

In one case, Thompson made the next shot. In two cases, he missed the next shot. In one case, nothing happened.

$$\hat{P}(\text{hit} \mid 2 \text{ hits}) = 1/3$$

The conditional probability of hit, given two misses

$$111010011000000 11$$

There are five instances of two misses in a row, indicated by the string “00.”

In two cases, Thompson made the next shot and in three cases, he missed the next shot.

The conditional probability of hit, given two misses

$$1110100\textcolor{orange}{1}100\textcolor{orange}{0001}1$$

There are five instances of two misses in a row, indicated by the string "00."

In two cases,Thompson made the next shot and in three cases, he missed the next shot.

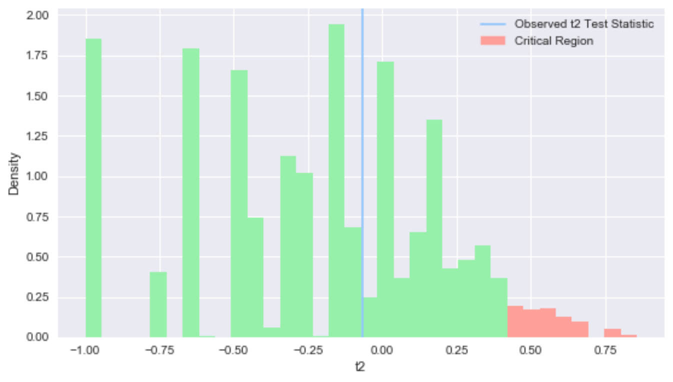
$$\hat{P}(\text{hit} \,|\, 2 \text{ misses}) = 2/5$$

Is a hit more likely to follow two hits than two misses?

$$\begin{aligned} t_2 &= \hat{P}(\text{hit} \mid 2 \text{ hits}) - \hat{P}(\text{hit} \mid 2 \text{ misses}) \\ &= 1/3 - 2/5 \\ &= -1/15. \end{aligned}$$

Is this exceptional?

The null hypothesis: the t_2 statistic for permutations of the observed string



What happened to the small sample bias?

Modern data science takes care of it.

Since the null hypothesis was generated by a finite data set

and not inferred from an asymptotic approximation to an infinite data set,

the size of the sample is automatically accounted for.

Do Steph Curry and Klay Thompson have hot hands?

The Warriors: 2016–2017

The team played 82 games in the regular season and 17 in the playoffs.

Curry played a total of 96 games, and Thompson played 95.

We calculated the t_2 statistic for Curry and Thompson in each game played

and we also calculated the t_2 statistic for the Warriors, quarter by quarter.

Warriors' shooting records (including field goals and free throws) 2016–2017

	Curry	Thompson	Warriors
Games or Quarters	96	95	396
Season Percentage	56%	51%	56%
Average Number of Shots	24	20	28

Design of the experiment

For each Splash Brother, game by game, and for the Warriors, quarter by quarter:

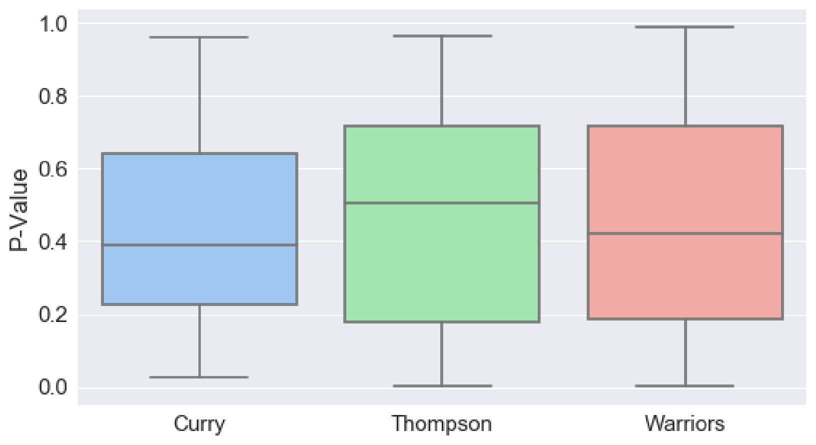
- Calculate t_2 for the string of observed hits and misses.
- Then compute the histogram of t_2 values for the null hypothesis, no hot hand, using either an enumeration of strings or 10,000 random permutations.
- Finally calculate the p-value of the observed string.

How will we interpret the results?

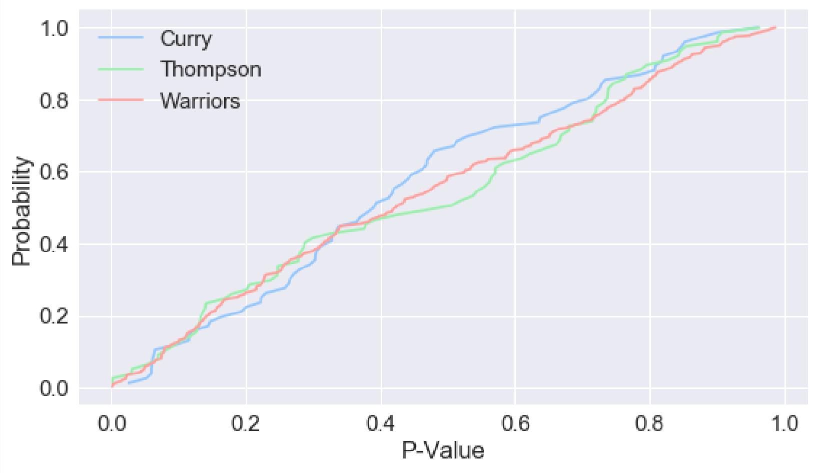
Under the null hypothesis:

- The p-values for Curry, Thompson and the Warriors are uniformly distributed on the unit interval
- and the CDF associated with the p-values is a 45-degree line.

Results: boxplots



Results: CDFs



Results: 5% significance of t_1,, t_2 and t_3

	Curry	Thompson	Warriors
Games	96	95	99
Observations	96	95	396
Conditioning set depth			
1	12	6	35
2	2	4	21
3	3	3	21

Thompson's 60-point game

The record

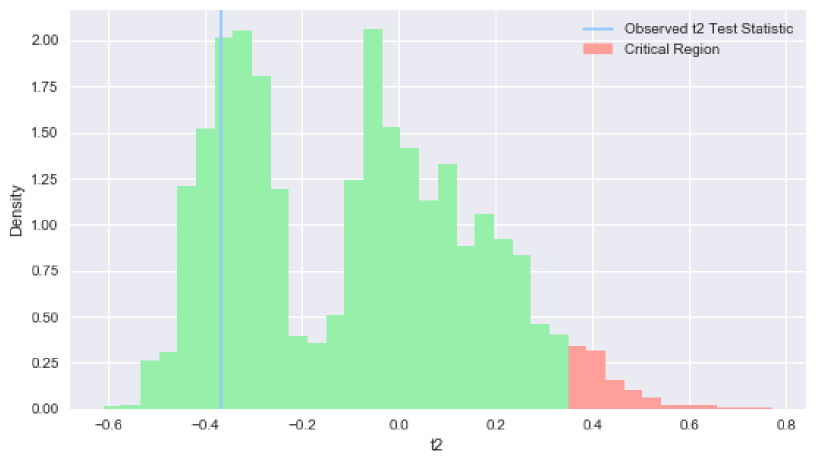
On December 6, 2016, Klay Thompson scored 60 points in a game against the Indiana Pacers

He hit 31 of 44 shots:

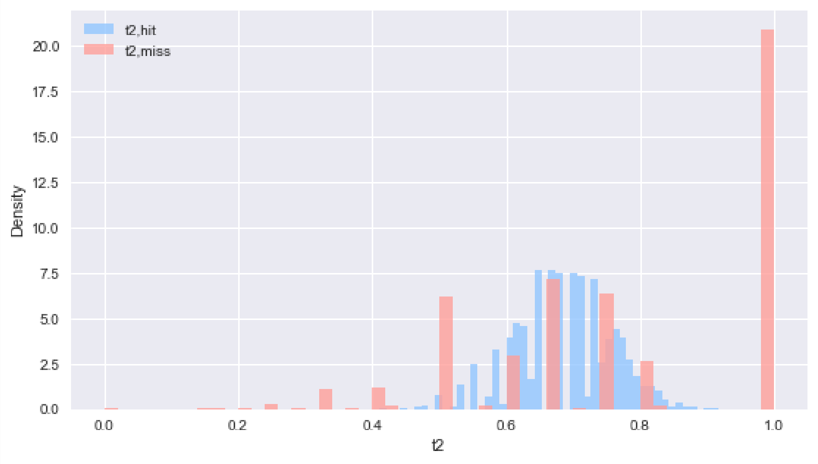
$$11011110010111111001110111101110111101010101$$

But did he have a hot hand?

The statistical analysis



A deeper statistical analysis: deconstructing t_2



Summary

Summary

We ran a version of the hot hand experiment pioneered in [Gilovich et al., 1985] on Golden State Warriors shooting data from the 2016–2017 regular season and the NBA playoffs.

We examined shooting records of Steph Curry and Klay Thompson game by game, and of the Golden State Warriors quarter by quarter.

Our version of the experiment relied on modern data science methods, which automatically take account of the small sample correction from [Miller and Sanjurjo, 2016].

Summary

The evidence from this experiment supports Gilovitch, Vallone and Tversky's original finding: no hot hand.

This is true even though the bar for rejecting the hot hand was lowered, relative to the original experiment, by the small sample correction.

While this is not the end of the story: there are plenty of unexplored data sets and alternative statistical formulations of hot hand,

the growing body of evidence that humans misconstrue randomness is awfully convincing.

Acknowledgement

Dedicated to Amos Tversky and Daniel Kahneman

The true believer in the law of small numbers commits his multitude of sins against the logic of statistical inference in good faith.

Amos Tversky and Daniel Kahneman, 1971, "Belief in the Law of Small Numbers"

to the Splash Brothers

and, of course, to KD

He doesn't have a hot hand either

	Curry	Thompson	Durant	Warriors
Games	96	95	77	99
Observations	96	95	77	396
Conditioning set depth				
1	12	6	8	35
2	2	4	4	21
3	3	3	3	21

Thank You

The formula for t_k

$$
\begin{aligned}
t_k(X) &= \frac{1}{H_k}\sum_{\tau=k+1}^{L} I\left(X_\tau = 1 \middle| \prod_{u=\tau-k}^{\tau-1} X_u = 1\right) \\
&\qquad - \frac{1}{M_k}\sum_{\tau=k+1}^{L} I\left(X_\tau = 1 \middle| \prod_{u=\tau-k}^{\tau-1} (1-X_u) = 1\right) \\
&= \hat{P}\left(X_\tau = 1 \middle| \prod_{u=\tau-k}^{\tau-1} X_u = 1\right) \\
&\qquad - P\left(X_\tau = 1 \middle| \prod_{u=\tau-k}^{\tau-1} (1-X_u) = 1\right) \\
&= t_{k,hit}(X) - t_{k,miss}(X)
\end{aligned}
$$

The Warriors: 2016–2017

	Curry	Thompson	Durant	Warriors
Games	96	95	77	99
Observations	96	95	77	396
Season Percentage	56%	51%	63%	56%
Average Game Percentage	56%	50%	62%	56%
StDev Game Percentage	11%	12%	12%	10%
Average Number of Shots	24	20	23	28
StDev Number of Shots	5	5	7	4

References

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

[Miller and Sanjurjo, 2016] Miller, J. B. and Sanjurjo, A. (2016).
Surprised by the gambler's and hot hand fallacies? a truth in the law of small numbers.
IGIER Working Paper No. 552.