האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

COMMENTS ON A "HOT HAND" PAPER BY MILLER AND SANJURJO (2015)

By

YOSEF RINOTT and MAYA BAR-HILLEL

Discussion Paper # 688   August 2015

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Edmond J. Safra Campus,
Jerusalem 91904, Israel
PHONE: [972]-2-6584135   FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Comments on a "Hot Hand" Paper by Miller and Sanjurjo (2015)

Yosef Rinott and Maya Bar-Hillel

The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem and LUISS, Rome
and
The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem

September 20, 2015

Abstract

Miller and Sanjurjo (2015) suggest that many analyses of the hot hand and the gambler's fallacies are subject to a bias. The purpose of this note is to describe our understanding of their main point in terms we hope are simpler and more accessible to non-mathematicians than is the original.

1 Introduction

Miller and Sanjurjo (2015), henceforth MS, make a startling and far-reaching claim about the hot hand in basketball, and about the analyses thereof, as presented in the seminal 1985 paper by Gilovich, Vallone and Tversky, henceforth GVT. If true, their claim is surprising, because it challenges a result that has withstood three decades of often skeptical research; see, e.g., Koehler and Conley (2003), Avugos et al. (2013). It is far-reaching, because the same allegedly faulty methods used by GVT underlie other lines of research into possibly-streaky phenomena; see, e.g., Rabin and Vayanos (2010). Early reactions to MS's paper suggest that it might have the potential to divide knowledgeable professionals, at least at first glance. We found MS intriguing but hard to follow. We therefore present the following brief note in the hope that it will make their important and valid critique, based on a hitherto unnoticed point, easier to understand, and to accept. We do not presume to address MS's body of work comprehensively.

Consider a basketball player shooting for the basket n times. Assuming that his probability of a Hit in each shot remains fixed throughout the sequence, denote it by $P(H)$. As is customary, in the present note as well as throughout all studies we shall discuss, we assume that all types of conditional probabilities also do not depend upon where in the sequence the conditioning event occurs [1]. The assumption of *independence* means that following any sequence of shots, the conditional probability that the next shot will be a Hit equals the unconditional probability of a Hit, $P(H)$. If, however, the probability that a streak of (one or more) Hits will be followed in the next shot by another Hit is larger than $P(H)$, then independence does not hold. Such a pattern of positive dependence is commonly referred to as a hot hand, a pattern whereby "success breeds success", if you will.

GVT analyzed data from actual games of basketball and from designed basketball shooting experiments, in order to study the question of whether or not they exhibit a hot hand. In GVT's Study 2, records of consecutive shots for individual Philadelphia 76ers were obtained, and subjected to the following analyses, which seem natural: "Players' shooting percentages conditioned on having hit their last shot, their last two shots, and their last three shots . . . [were compared with] their shooting percentages conditioned on having missed their last

[1] This property follows from *stationarity*, a technical form of stability that need not concern us here. It is always assumed implicitly, since without it statistical analysis is hardly possible. We shall mention it no further.

shot, their last two shots, and their last three shots, respectively" (p. 298). Subjecting these numbers to statistical significance testing, which we shall not describe here, led GVT to conclude that the field goal data showed no evidence of streak shooting, and indeed seemed quite consistent with independence. Similar statistical procedures were followed in GVTs Studies 3 and 4, and similar conclusions were reached.

Obviously, these percentages (or relative frequencies) are not the true conditional probabilities $P(H|H)$, $P(H|2H)$ or $P(H|3H)$, where $P(H|kH)$ stands for the conditional probability that a streak of k Hits is followed by another Hit; they are only estimators of the unknown conditional probabilities. This distinction between estimators and true probabilities is not made in GVT's text and Table headings. In our understanding, the main point in MS is that these estimators of the conditional probabilities that streaks of any length would continue, albeit standard, are nonetheless, on average, biased downwards, under the assumption of independence. Furthermore, since the (conditional) probabilities of H or T (H's complement, or a missed shot) add up to one, the estimators of the conditional probability that a streak will end on the next shot are biased upwards. The fact that such estimators are biased is not new. See for example Bai (1975) and references therein. However, the direction of the bias was discovered by MS, and is essential for the arguments in their paper.

In Section 3 we describe some implications of the bias. In particular we explain, following MS, that the above bias may mask a hot hand; that is, data which in fact exhibit a hot hand may seem to have been generated by independent basketball shooting.

2 A mathematical presentation of the bias

Given a sequence of n shots, the standard estimator of $P(H)$, denoted by $\widehat{P}(H)$, is the number of Hits in the sequence divided by its length n. This is an unbiased estimator. Next consider $P(HH)$, the probability of observing two consecutive Hits at any eligible shot in the sequence, where an eligible shot is one which has a successor, so the first n-1 shots are eligible. The standard estimator of $P(HH)$, denoted by $\widehat{P}(HH)$, is the number of eligible Hits that are followed by a Hit, divided by the total number of eligible shots, n-1. This estimator is also unbiased. The standard estimator of $P(H|H)$, $\widehat{P}(H|H)$, is the number of eligible Hits whose successor is another Hit, divided by the total number of eligible Hits. If there are no eligible Hits, there is no estimator. This estimator is biased [2].

What makes it biased? Since the estimator $\widehat{P}(H|H)$ is itself a ratio of two estimators, then although these estimators are themselves unbiased, their ratio is generally prone to bias, for various reasons. In particular, division by an estimator is not a linear function. And here, moreover, the two estimators are obviously not independent. In fact, they are positively and strongly dependent, and this determines the direction of the bias. MS computed the bias of $\widehat{P}(H|H)$ as an estimator of $P(H|H)$, as well as the bias conditioned on streaks of length k for k up to 5. They showed that the bias is downward, which is the crux of the matter.

Below we present our derivation for the case $k = 1$, that is, we compute the bias of $\widehat{P}(H|H)$. It is shorter and simpler than that of MS. Formula (1) below is equivalent to MS's somewhat more cumbersome formula (3). Part *(b)* follows from (1), and Part *(c)* of the proposition, which follows from Parts *(a)* and *(b)*, says that under independence, $\widehat{P}(H|H)$ is biased downward: its expectation is smaller than the estimated $P(H|H)$, which in turn equals $P(H)$ under independence. We remark that the formulas for the bias below show that it is unimodal both as a function of n and of p, first increasing and then decreasing [3]. MS's calculations, which we verified by simulations, show that the bias grows as the length of the conditioning streak increases. When conditioning on 3H, even for p=0.5 and n=100 (the parameters of GVT's study 4), the bias is not negligible.

[2] In technical terms, not essential for this note, these estimators are the Maximum Likelihood Estimators if one assumes that the sequence of shots or coin tosses forms a suitable homogeneous Markov Chain, and in particular, when they are independent. Moreover, such estimators are asymptotically normal and asymptotically optimal in a suitable sense. They are consistent, and they are known to be generally biased; see Bai (1975) and references therein for all these facts. These facts, including the bias, apply also to k-step Markov Chains, since it is well known that they can be regarded as Markov Chains by enlarging the state space; this is relevant when conditioning on the last k shots for $k > 1$.

[3] As a function of n the turning point for $p = 0.15$, for example, is $n = 6$ and for $p = 0.5$ it is $n = 3$, but it is arbitrarily large for small p's. For values of p of interest in basketball this curious non-monotonicity in n occurs only for small n's.

Proposition. [4] *Set $p = P(H)$, $q = P(T) = 1 - p$, $m = n - 1$; let F be the event that the number of eligible H's, that is, H's among the first n-1 shots, is non zero, so that the estimator $\widehat{P}(H|H)$ is well defined, and assume the tosses are independent. Then*

(a) The conditional expectation of the estimator $\widehat{P}(H|H)$ given the event F, satisfies for any $0 < p < 1$,

$$E(\widehat{P}(H|H) \mid F) = \frac{1-q}{1-q^m} - \frac{q}{m}. \tag{1}$$

(b) The bias can be expressed as

$$P(H) - E(\widehat{P}(H|H) \mid F) = q(1-q)[1 + q + \cdots + q^{m-1} - mq^{m-1}]/[m(1-q^m)]. \tag{2}$$

(c)

$$E(\widehat{P}(H|H) \mid F) < P(H).$$

Proof. Note that given F, one can interpret $\widehat{P}(H|H)$ as the probability of the event A that a randomly chosen H among the first $n-1$ tosses in the observed sequence, if one exists, is followed by another H. Consider now the sequence of the first $m = n-1$ tosses. If it is all T's then F does not hold. So when conditioning on F, the first m tosses can be one of $2^m - 1$ sequences, and $P(F) = 1 - q^m$. The probability that a random sequence has in the first m tosses i H's and $m - i$ T's is $\binom{m}{i} p^i q^{m-i}$. Choose one of the i H's and place it in one of the first m positions at random. It will be last (that is, mth) with probability $\frac{1}{m}$, and then it will be followed by H with probability p. With probability $\frac{m-1}{m}$ the chosen H is not last. Conditioned on this case, it is followed by H with probability $\frac{i-1}{m-1}$. Combining these facts we obtain by the formula of total probability that the probability of the event A conditioned on F is

$$P(A \mid F) = P(A \cap F)/P(F) = \sum_{i=1}^{m} \binom{m}{i} p^i q^{m-i} [\frac{1}{m} p + \frac{m-1}{m} \cdot \frac{i-1}{m-1}] \Big/ (1 - q^m).$$

Using $\sum_{i=1}^{m} \binom{m}{i} p^i q^{m-i} = 1 - q^m$ and $\sum_{i=1}^{m} i \binom{m}{i} p^i q^{m-i} = mp$, that is, the binomial sum and expectation formulas, respectively, we obtain (1) as follows:

$$P(A \mid F) = \frac{p}{m} + \frac{p}{1-q^m} - \frac{1}{m} = \frac{p}{1-q^m} - \frac{1-p}{m} = \frac{1-q}{1-q^m} - \frac{q}{m}.$$

Somewhat tedious but elementary algebra yields (2), which implies Part *(c)* by the fact that for $0 < q < 1$, $1 + q + \cdots + q^{m-1} > mq^{m-1}$. ∎

3 Implications for the hot hand

We shall not take a stand regarding the question of whether or not the hot hand phenomenon is genuine. For one, we didn't read MS's other papers on the subject, nor did we perform any analyses on any data ourselves. We endorse their mathematical point, but leave it for others to determine whether or not a hot hand exists regardless. On the other hand, the rationale whereby the bias may cause a genuine hot hand to appear erroneously as independence is important to understand. A similar rationale can disguise the fact that the so-called Gambler's Fallacy may sometimes not be a fallacy at all.

Consider the implications of this bias. Under independence $P(H|H) = P(H|T)$, while hot hand suggests $P(H|H) > P(H|T)$. The above biases imply that under independence, which implies no hot hand, we have in expectations $E\widehat{P}(H|H) < E\widehat{P}(H|T)$[5], with a difference that is not negligible in many relevant data set sizes. Roughly speaking this implies that more often than not, under independence, one will observe $\widehat{P}(H|H) < \widehat{P}(H|T)$. Now suppose you observe small positive and negative values of $\widehat{P}(H|H) - \widehat{P}(H|T)$ as did GVT, rather than mostly negative values, as expected under independence. This suggests that independence may be

[4] A reader who wishes to avoid technical details can skip the proposition below and its proof with no loss of continuity.

[5] These expectations should be conditioned on the existence of the estimators as in Section 2. Here we suppress this conditioning.

violated, and since observed values of $\widehat{P}(H|H) - \widehat{P}(H|T)$ are larger than expected under independence, a hot hand is suggested. Clearly, if one is not aware of the bias, observing small values of $|\widehat{P}(H|H) - \widehat{P}(H|T)|$, as did GVT, would suggest independence, possibly wrongly. The same arguments apply to conditioning on longer streaks of Hits, in which case it turns out that the biases are larger.

MS's toy-example of a game where a player whose probability of a Hit is known to be 0.5 and who takes only 4 shots is a useful platform. As MS show in their Table 1, even though this player's true $P(H|H)$ is 0.5, and so he does not have a hot hand, the expected value of the observed conditional probability, $\widehat{P}(H|H)$, is not - surprisingly enough - 0.5, as one might guess intuitively, but rather only 0.405 ! Thus, if the observed rate at which a streak continues is 0.5, the data are actually indicative of positive dependence, because under independence that rate would only be expected to be 0.405. A comparison of the observed rate to 0.5 is compelling, but nonetheless misleading. Similarly, if $P(H)$ stands for the probability of a coin toss coming up Heads, then, when estimating alternation rates as above, only observed alternation rates higher than 0.595, not merely higher than 0.5, can be regarded as evidence of the Gambler's Fallacy.

Clearly, in real data, any observed results could also happen by random fluctuations, so that a test of statistical significance is required in order to assess their actual significance. MS report no statistical significance tests[6], however, we performed one ourselves. In their Table 3, MS took data from GVTs Table 4 and corrected the latter's reported differences between $\widehat{P}(H|3H)$ and $\widehat{P}(H|3T)$ by the calculated bias, thus obtaining unbiased estimates of the difference $P(H|3H) - P(H|3T)$. Based on a simulation we ran, their corrections are accurate (they themselves derived them analytically, not by simulation). Following this correction, 19 of 25 players exhibit a positive difference. By a sign test, this outcome is significant ($p < .01$), indicating that at a group level, players tended to score better following a streak of 3 Hits than following a streak of 3 misses [7]. Is this evidence of a hot hand? Clearly, more careful testing is needed. Since hot hands are attributed to individual players, the group results hardly suffice for a responsible or definitive reply, but is nonetheless intriguing. The customary comparisons underestimate the true probability of hot hand seeming streaks, and recalculations are called for throughout the literature to see if the bias in question changes the conclusions of hot hand and related studies significantly.

In conclusion it is our opinion that 30 years after GVT, MS raised a valid and overlooked criticism of their and subsequent studies. Comments on this note by Gilovich as well as from Miller and Sanjurjo promise that the hot hand debate will continue.

References

Avugos, S., Köppen, J., Czienskowski, U., Raab, M., and Bar-Eli, M. (2013). The hot-hand reconsidered: A meta-analytic approach. *Psychology of Sport and Exercise* 14, 21-27.

Bai, D. S. (1975) Efficient estimation of transition probabilities in a Markov Chain. *The Annals of Statistics* 3, 1305-1317.

Gilovich, T., Vallone, R. and Tversky, A. (1985). On the misperception of random sequences. *Cognitive Psychology* 17, 295-314.

Koehler, J. J., and Conley, C. (2003). The 'Hot-hand' myth in professional basketball. *Journal of Sport and Exercise Psychology* 25, 253-259.

Miller, J. B., and Sanjurjo, A. (2014). A cold shower for the hot hand fallacy *http://ssrn.com/abstract=2450479*.

Miller, J. B., and Sanjurjo, A. (2015). Surprised by the gambler's and hot hand fallacies? A truth in the Law of Large Numbers. *http://ssrn.com/abstract=2627354*.

Rabin, M., and Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies* 77, 730-778.

[6] But see, e.g., Miller and Sanjurjo (2014)

[7] For the sign test one should correct the median rather than the mean. Since the estimated probabilities are close to 1/2, and the sample size of 100 is not small, the distribution in question is close to being symmetric, so the mean and the median are close.