ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

A Bayesian test for the hot hand phenomenon

Ruud Wetzels [a,*], Conor Dolan [b], Darja Tutschkow [c], Sophie van der Sluis [b], Gilles Dutilh [d], Eric-Jan Wagenmakers [e]

[a] *PricewaterhouseCoopers, The Netherlands*
[b] *Free University of Amsterdam, The Netherlands*
[c] *University of Dortmund, Germany*
[d] *University of Basel, Switzerland*
[e] *University of Amsterdam, The Netherlands*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Bayes factor
Streakiness
Sports
Hot hand

ABSTRACT

The hot hand phenomenon refers to the popular notion that the performance of sports players is punctuated by streaks of exceptional performance. During these streaks, the player is said to be 'hot', or even 'on fire'. Unfortunately, when it comes to assessing evidence for the hot hand phenomenon, human intuition is inadequate—people are known to perceive streaks even in sequences that are purely random. Here we develop a new statistical test for the presence of the hot hand phenomenon for binary sequences of successes and failures. The test compares a constant performance model to a hidden Markov model with two states (one representing hot performance, and one representing cold performance) and one probability of switching from one state to the other. We assume appropriately restricted uniform priors on the model parameters and compute the Bayes factor by integrating the likelihood over the prior. The test is assessed in a simulation study and applied to real data sets from basketball and from psychology. Our analysis suggests that it is difficult to find compelling evidence for and against streakiness except for very long data sequences and extreme forms of streakiness.

1. Introduction

For more than 25 years, the existence of the hot hand phenomenon has been the topic of intense debate in the academic literature on sports, statistics, and psychology. A player is called 'hot' or is said to have a hot hand if "(…) the performance of a player during a particular period is significantly better than expected on the basis of the player's overall record" (Gilovich, Vallone, & Tversky, 1985, p. 295–296). Sports fans, players, and coaches often express belief in the hot hand phenomenon; however, several researchers have argued that the hot hand is nothing but a cognitive illusion. For instance, Tversky and Kahneman (1974) claimed that people rely on heuristics when judging the probability of an event and that these heuristics lead to systematic biases in people's perception. Specifically, Gilovich et al. (1985) analyzed shooting records of basketball players, failed to reject the null hypothesis of constant performance, and concluded that the belief in the hot hand rests on "a general misconception of chance according to which even short random sequences are thought to be highly representative of their generating process" (p. 295; but see Wardrop, 1995).

Over time, initial academic skepticism towards the existence of the hot hand phenomenon has given way to a more balanced view. Psychologists Gilden and Wilson (1995b) explained the occurrence of streaks in skilled performance by the concept of flow (Csikszentmihalyi, 1990). Statisticians applied a series of different tests to sports such as baseball (Albert, 2008; Albright, 1993; Barry & Hartigan, 1993), basketball (Albert & Williamson, 2001; Gilovich et al., 1985; Shea, 2014; Wardrop, 1999), golf (Clark, 2005), bowling (Dorsey-Palmateer & Smith, 2004), volleyball (Raab, Gula, & Gigerenzer, 2012), and others, finding mixed support for the hot hand phenomenon. In a review paper, Bar-Eli, Avugos, and Raab (2006) listed 11 studies that found support for the hot hand phenomenon and 13 studies that did not.

The importance of the hot hand phenomenon transcends the domain of sports. As noted by Bar-Eli et al. (2006), "the hot hand debate in sport may well influence other domains and provide boundaries for theories that attempt to explain beliefs and behavior in real environments other than sport" (p. 526). One example of this

* Correspondence to: PricewaterhouseCoopers, Thomas R. Malthusstraat, 1006 GC, Amsterdam, The Netherlands.
E-mail address: wetzels.ruud@gmail.com (R. Wetzels).

http://dx.doi.org/10.1016/j.jmp.2015.12.003

general relevance is the study by Gilden and Wilson (1995a), whose work concerned the occurrence of streaky performance in a simple perceptual task.

The current status of the hot hand phenomenon is not entirely clear. Part of the problem is that different sports and tasks may elicit streakiness more than others; an additional complication is that different researchers use different tests to assess streakiness. Moreover, classical tests for streakiness such as tests of serial correlation and the popular Wald–Wolfowitz runs test generally have low power (Albert & Williamson, 2001; Wardrop, 1999). With low power to detect deviations from the null model of constant performance, the absence of evidence for the hot hand phenomenon does not equal evidence for its absence.

A related issue is that classical tests focus exclusively on the null hypothesis of constant performance, and do not consider the plausibility of the data under a specific alternative hypothesis. Ideally, a test for the hot hand phenomenon compares the null hypothesis against a concrete alternative model for streakiness, as this allows one to compute the extent to which the data support one model over the other (for a brief summary of these and other Bayesian advantages, see Mulder & Wagenmakers, in press). One simple model for streakiness, proposed by Albert (1993) in the context of baseball batting, is a hidden Markov model with two states and one transition probability (for a different model see Albert, 2008). In each baseball game i, the number of successful at-bats follows a binomial distribution with success probability p_i; when the player is in the hot state, $p_i = p_h$, and when the player is in the cold state, $p_i = p_c$, with $p_h > p_c$. Following each game, the player switches states with a fixed probability $\alpha = 0.1$. Similar models have been proposed, applied, and evaluated in other work (Albert & Williamson, 2001; Lopes & Oden, 1987; Sun, 2004; Sun & Wang, 2012).

Inspired by the work of Albert, our test for the hot hand phenomenon uses the Bayes factor to quantify the adequacy of a constant performance model against that of a streaky performance model. The streaky performance model is a hidden Markov model with two states and one transition probability. In contrast to Albert (1993) we do not assign the transition probability α a fixed value, but rather treat it as a free parameter. Furthermore, Albert (1993) assumed that a player is in a particular state during an entire game i (or sometimes an epoch i of arbitrary length), whereas we assume that a player can switch states at any time point t. Hence the binary random variable that indicates success or failure at time t follows a Bernoulli distribution with a success probability that depends on the hidden state at time t. The underlying process is assumed to follow a stationary first-order Markov chain, meaning that the probability of being in a certain state at time t depends only on the state occupied at time $t - 1$.

The outline of this paper is as follows. The first section provides the mathematical details of the hidden Markov model and the proposed Bayesian test. The second section reports a simulation study to assess the performance of the Bayesian test. The third and fourth sections provide application examples with data from basketball free-throw shooting and perceptual identification, respectively. The final section summarizes our findings and discusses their ramifications.

2. A two-state Bernoulli hidden Markov model

Consider a first-order hidden Markov model (HMM) with two possible states at each discrete time point t: $S_t \in \{0, 1\}$, where $S_t = 0$ represents the cold state and $S_t = 1$ represents the hot state. We use upper-case letters to denote random variables and lower-case letters to denote the realization of these random variables. Switches between the states are governed by so-called transition probabilities. The one-step transition probability matrix $\Gamma = (\gamma_{ij})_{i,j\in\{1,2\}}$ contains the probability of switching from the hot to the cold state and vice versa: $\gamma_{ij} = p(S_{t+1} = 0 \mid S_t = 1) = p(S_{t+1} = 1 \mid S_t = 0) = \alpha$ for $i \neq j$ and the probability of staying in a state $\gamma_{ij} = p(S_t = 1 \mid S_{t-1} = 1) = p(S_t = 0 \mid S_{t-1} = 0) = 1 - \alpha$ for $i = j$. Thus, when $\alpha < .5$ the states are "sticky" and when $\alpha > .5$ the states are "repelling". Only sticky states produce performance that is consistent with streakiness and the hot hand phenomenon, and hence the remainder of this paper focuses on switching probabilities lower than .5.

The state-dependent sequence of random variables $\{Y_t : t \in \mathbb{N}\}$ produces the sequence of observations $y_t, t \in \{1, \ldots, T\}$. Since we are concerned only with binary data, Y_t is distributed according to a Bernoulli distribution for all t. Here $Y_t = 0$ indicates failure (e.g., a miss) and $Y_t = 1$ indicates success (e.g., a hit). A player can have success both in the hot and in the cold state. However, the probability of success is by definition higher in the hot than in the cold state. The random variable Y_t therefore has a different Bernoulli distribution $Y_t \sim \text{Bern}(p_{S_t})$ depending on the current state S_t. We denote the probability of success in the hot state by $\theta_h = p(Y_t = 1 \mid S_t = 1)$, and the probability of success in the cold state by $\theta_c = p(Y_t = 1 \mid S_t = 0)$. For compactness we define two diagonal matrices $\mathbf{p}(y_t)$ with $t = 1, \ldots, T$ and $y_t \in \{0, 1\}$ which contain the success and failure probabilities for both states (Zucchini & MacDonald, 2009):

$$\mathbf{p}(y_t = 1) = \begin{pmatrix} \theta_h & 0 \\ 0 & \theta_c \end{pmatrix} \quad \text{and}$$

$$\mathbf{p}(y_t = 0) = \begin{pmatrix} 1 - \theta_h & 0 \\ 0 & 1 - \theta_c \end{pmatrix}.$$

The likelihood L_{HMM} of the two-state Bernoulli hidden Markov model is:

$$L_{\text{HMM}} = \boldsymbol{\delta}\mathbf{p}(y_1)\Gamma p(y_2) \cdots \Gamma p(y_T)\mathbf{1}' \quad (1)$$

(Zucchini & MacDonald, 2009, p. 37), where $\mathbf{1}'$ is a 2-dimensional row vector and $\boldsymbol{\delta}$ is the initial distribution of the Markov chain. Here we assume that a player is equally likely to start in one or the other state which means $\boldsymbol{\delta} = (1/2, 1/2)$. Hence, our two-state HMM has three free parameters: the probability θ_h of success in the hot state, the probability θ_c of success in the cold state, and the probability α of switching between states.

To illustrate the typical shape of the HMM likelihood function we generated a synthetic data set with 1000 observations from a HMM with parameters $\theta_h = .7$, $\theta_c = .4$, and $\alpha = .1$. Fig. 1 shows the corresponding likelihood function as a series of contour plots. These plots reveal two kinds of non-identifiability (Allman, Matias, & Rhodes, 2009; Petrie, 1969). First, for every value of α the likelihood is symmetric around the main diagonal, indicating label-switching between θ_h and θ_c. This problem can be overcome by enforcing the constraint $\theta_h > \theta_c$. Second, when $\alpha = .5$ there are infinitely many combinations of θ_h and θ_c that yield the same likelihood. Although important for parameter point estimation, these HMM concerns about identifiability are irrelevant for Bayesian model selection using the Bayes factor.

3. A bayes factor test for streakiness

In order to assess the evidence for and against streaky performance we compare two models. The first model is the HMM from the previous section, which represents streaky performance. The second model is a baseline model that assumes a single, constant success probability $\theta = p(Y_t = 1)$ for all time points $t \in \mathbb{N}$: the constant performance model (CPM). In the case of the CPM