Journal of Mathematical Psychology 72 (2016) 200–209

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,*], Darja Tutschkow [c], Conor Dolan [b], 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 2 February 2016

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
0022-2496/

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, 2016). 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:

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

(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

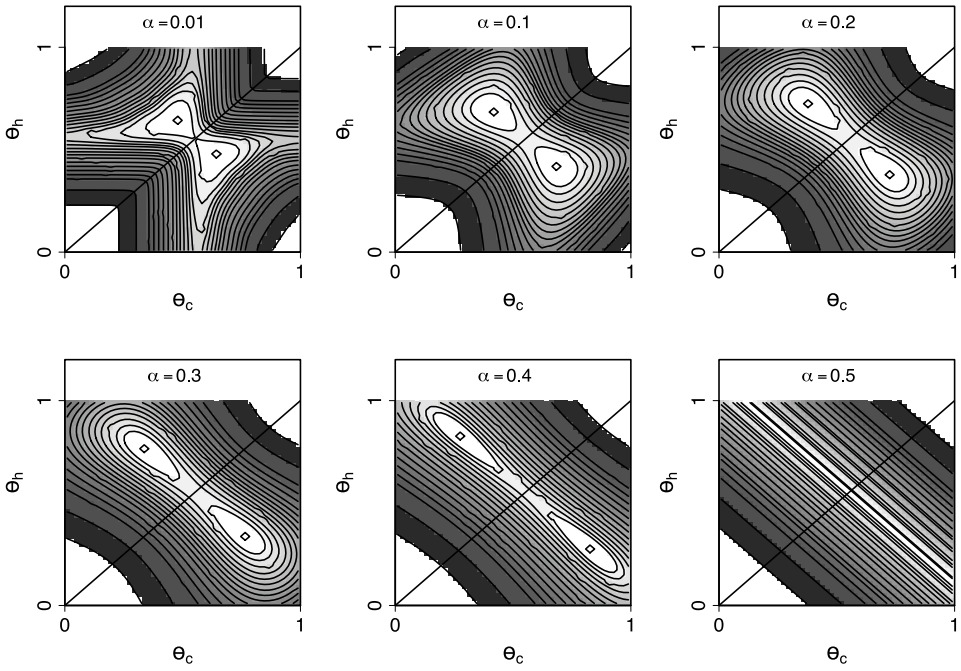

Fig. 1. Contour plots of the likelihood function for a HMM. Each panel represents a different value of the switching parameter α. The associated synthetic data set featured 1000 observations and was generated under a HMM with parameters $\theta_h = .7$, $\theta_c = .4$, and $\alpha = .1$.

the observed data y_t are the outcome of a sequence of independent random variables $Y_1, Y_2, Y_t, \ldots, Y_T$ with $Y_t \sim \text{Bern}(\theta)$ for all $t \in \mathbb{N}$. It follows that the likelihood L_{CPM} of the CPM is:

$$L_{\text{CPM}} = \theta^k (1-\theta)^{T-k}, \tag{2}$$

with k the number of successes in the observed sequence of length T. When $\alpha = .5$ the HMM reduces to the CPM with parameter $\theta = (\theta_h + \theta_c)/2$ (see Appendix A for a proof).

In order to assess the evidence that the data provide for the HMM versus the CPM we compute the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995), that is, the ratio of marginal likelihoods:

$$\text{BF}_{\text{HC}} = \frac{p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \text{HMM})}{p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \text{CPM})}, \tag{3}$$

where $\mathbf{Y}^{(T)}$ and $\mathbf{y}^{(T)}$ denote the time-ordered vector of random variables $Y_1, Y_2, \ldots, Y_T$ and observations $y_1, y_2, \ldots, y_T$, respectively. The Bayes factor indicates the change from prior to posterior odds brought about by the data. When $\text{BF}_{\text{HC}} = .20$, for instance, this indicates that the observed data are 5 times more likely to occur under the CPM than under the HMM.

The marginal likelihoods are obtained by integrating out the model parameters over the prior distribution:

$$\begin{aligned}
&\frac{p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \text{HMM})}{p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \text{CPM})} \\
&\quad = \frac{\int_0^1 \int_0^1 \int_0^1 p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \theta_c, \theta_h, \alpha) p(\theta_c) p(\theta_h) p(\alpha)\, \mathrm{d}\theta_c \mathrm{d}\theta_h \mathrm{d}\alpha}{\int_0^1 p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \theta) p(\theta)\, \mathrm{d}\theta}.
\end{aligned} \tag{4}$$

Here we pursue a default, reference-style analysis with independent uniform prior distributions for all parameters in both models. For the HMM, the resulting joint prior specification assigns equal mass to all parameter values in the unit cube. However, as priors in Bayesian theory can be used to represent theory (Vanpaemel, 2010; Vanpaemel & Lee, 2012), we restrict the uniform integration space in two ways. First, sticky-state values for α range from 0 to .5, halving the integration space. Second, we eliminate label-switching and consider only those parameter values where $\theta_h > \theta_c$, halving the integration space once more—hence, the final integration space covers only a quarter of the unit cube. Hence we obtain:

$$\begin{aligned}
&\frac{p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \text{HMM})}{p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \text{CPM})} \\
&\quad = \frac{4 \int_0^{.5} \int_0^1 \int_0^{\theta_h} p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \theta_c, \theta_h, \alpha)\, \mathrm{d}\theta_c \mathrm{d}\theta_h \mathrm{d}\alpha}{\int_0^1 p(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid \theta)\, \mathrm{d}\theta} \\
&\quad = \frac{4 \int_0^{.5} \int_0^1 \int_0^{\theta_h} \boldsymbol{\delta}\mathbf{p}(y_1) \Gamma p(y_2) \cdots \Gamma p(y_T) \mathbf{1}'\, \mathrm{d}\theta_c \mathrm{d}\theta_h \mathrm{d}\alpha}{\int_0^1 \theta^k (1-\theta)^{T-k}\, \mathrm{d}\theta} \\
&\quad = \frac{4(T+1)!}{k!(T-k)!} \int_0^{.5} \int_0^1 \int_0^{\theta_h} \boldsymbol{\delta}\mathbf{p}(y_1) \Gamma p(y_2) \cdots \\
&\qquad \times \Gamma p(y_T) \mathbf{1}'\, \mathrm{d}\theta_c \mathrm{d}\theta_h \mathrm{d}\alpha,
\end{aligned} \tag{5}$$

where the last step follows from Euler's beta integral.

Even though the Bayes factor has an unambiguous and continuous scale of evidential strength, the upcoming presentation of results is made easier by using Jeffreys (1961, Appendix B) discrete classification scheme shown in Table 1. Jeffreys' labels facilitate communication but should be considered only as an approximate descriptive articulation of different standards of evidence. Note that, in contrast to classical tests, the Bayes factor allows one to quantify evidence in favor of the CPM. In addition, the Bayes factor BF_{HC} can also indicate that the data provide only anecdotal evidence that is "not worth more than a bare mention".

4. Implementation

In order to approximate the integral for the HMM in Eq. (5), we used a simple midpoint rule with a subdivision into 100 intervals. If overflow occurred, fewer subdivisions were used. The numerical integration routine was programmed using the R system for statistical computing (R Development Core Team, 2012). In addition, since the likelihood is a product of matrices with

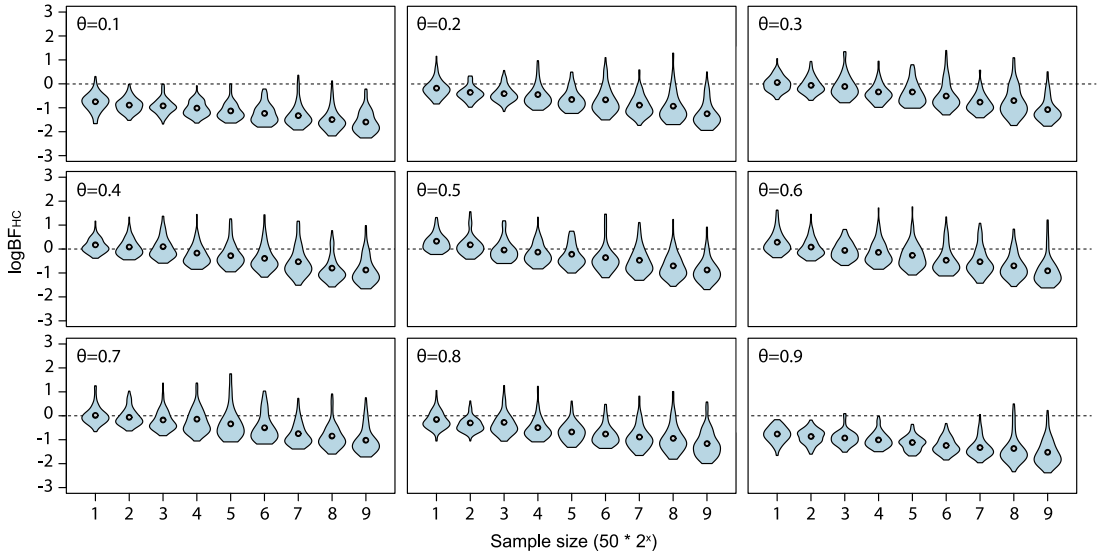

Fig. 2. Distribution of log BF_{HC} for data generated under the CPM as a function of sample size. Each panel is associated with a different value of the success probability θ. The Bayes factors generally indicate support for the correct data generating model, and this support increases with sample size.

Table 1
Evidence categories for the Bayes factor BF_{HC} (Jeffreys, 1961).

BF_{HC}	Strength of evidence
0–1	Negative (supports CPM)
1–3	Not worth more than a bare mention
3–10	Substantial
10–30	Strong
30–100	Very strong
>100	Decisive

elements ranging between 0 and 1 (see Eq. (1)) a scaling algorithm had to be applied to prevent underflow before sampling from the likelihood (Zucchini & MacDonald, 2009). The likelihood – or more precisely the logarithm of the likelihood – was evaluated using the **HiddenMarkov** package (Harte, 2011). The R code used to calculate the Bayes factor BF_{HC} can be found in Appendix B.

5. Simulation study

5.1. Methods

In order to confirm the correctness of the algorithm and evaluate the informativeness of data for discriminating between the HMM and the CPM we conducted a model recovery simulation study. In this study we generated synthetic data from both models and then applied our Bayes factor test procedure. We considered nine different sample sizes, $T = (50, 100, 200, 400, 800, 1600, 3200, 6400, 12800)$. For each sample size category, we simulated data from the CPM for nine different values of the success rate parameter, $\theta = (.1, .2, .3, .4, .5, .6, .7, .8, .9)$. We also simulated data from the HMM; here we first fixed θ_h to .7, as earlier simulation studies showed that BF_{HC} is mostly affected by the difference $d = \theta_h - \theta_c$ and is relatively insensitive to the absolute values of θ_h and θ_c. We then examined the factorial parameter combinations involving three levels of the switching parameter, $\alpha = (.1, .25, .40)$, and three levels of the parameter that quantifies the success probability in the cold state, $p_c = (.4, .5, .6)$ (i.e., $d = (.1, .2, .3)$), for a total of 9 cells in the design. For both models, each of the 9×9 cells in the design contained 100 repetitions.

5.2. Results

We use violin plots (Hintze & Nelson, 1998) to represent the distribution of log BF_{HC} for each cell in the design. For ease of interpretation we employ Jeffreys' classification scheme (Table 1).

5.2.1. Data generated under the CPM

Fig. 2 shows the distribution of log BF_{HC} for nine different data-generating values of the CPM success probability θ. Because the data were generated with the CPM, we expect to see support in favor of the CPM over the HMM (i.e., negative values of log BF_{HC}), and we expect this support to increase with sample size. Both expectations are confirmed. Overall, 6690 out of 8100 data sets (83%) show support for the CPM, and 1922 out of 6690 data sets (29%) show support for the CPM that is worth more than a bare mention. Also, evidence in favor of the correct data-generating CPM grows with the available information: the log BF_{HC} values in each panel decrease with sample size. Finally, the evidence in favor of the CPM is larger when the success probability θ takes on more extreme values (e.g., $\theta = (.1, .9)$).

5.2.2. Data generated under the HMM

Fig. 3 shows the distribution of log BF_{HC} for different data-generating values of the HMM parameters. Each row corresponds to a particular value of the switching probability α, and each column corresponds to a particular value of the success differential d (i.e., $\theta_h - \theta_c$ with $\theta_h = .7$). Because the data were generated with the HMM, we expect to see support in favor of the HMM over the CPM (i.e., positive values of log BF_{HC}), and we expect this support to increase with sample size. Both expectations are confirmed, albeit with important qualifications. Specifically, the expected pattern occurs when the states are sufficiently sticky (i.e., α is relatively low) and the success differential is sufficiently large (i.e., d is relatively high).

In general, it appears that even for medium sample sizes the support in favor of the (correct, data-generating) HMM often fails to exceed the threshold for anecdotal evidence. This underscores the fact that Bernoulli time series contains relatively little information, and this tentatively suggests an explanation as to why the evidence for streakiness in sports is so mixed: in real-life applications, the data may not be informative enough to distinguish between a streaky and a non-streaky model.

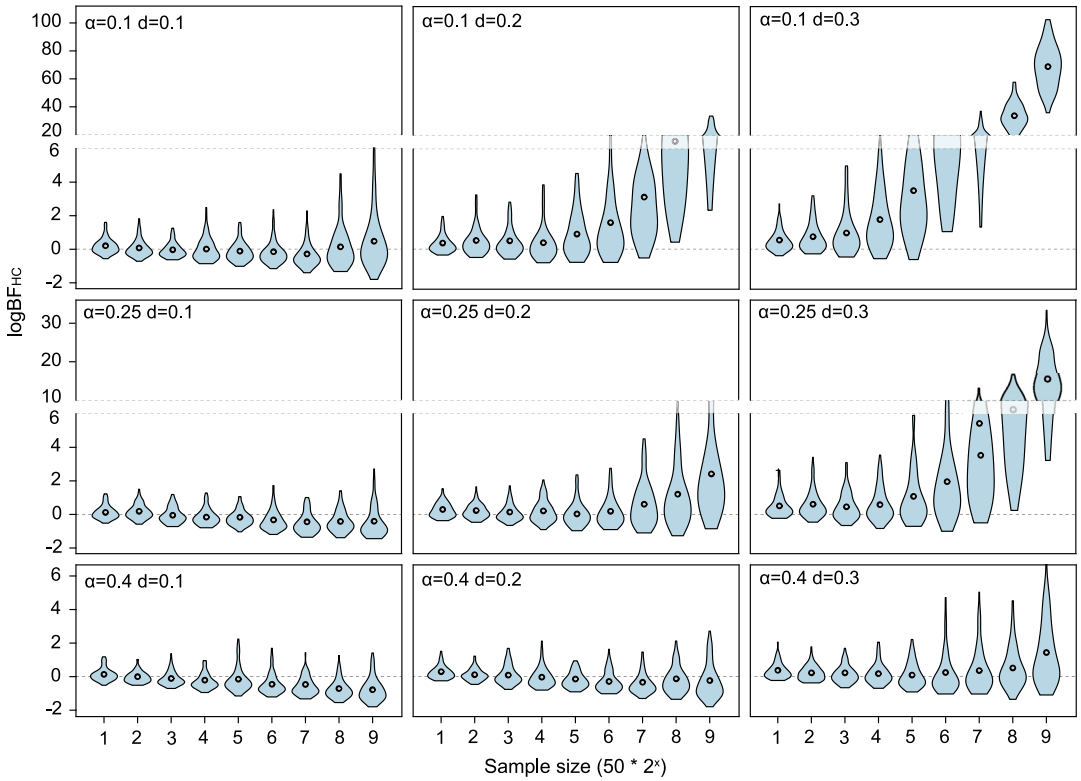

Fig. 3. Distribution of log BF_{HC} for data generated under the HMM as a function of sample size. Each panel row is associated with a different switching probability α, and each column is associated with a different success differential d.

6. Example application 1: basketball free-throw shooting

Here we analyze binary basketball free-throw shooting data from six consecutive NBA seasons (Yaari & Eisenmann, 2011). In particular, we illustrate the behavior of our test for free-throw shooting performance of two iconic basketball players: Kobe Bryant and Shaquille O'Neal. Kobe Bryant is among the best free-throw shooters in the NBA.[1] Each panel of Fig. 4 shows Bryant's performance for a specific season; in addition, each panel provides the log Bayes factor in favor of the HMM over the CPM. For every season, Bryant's data are more likely under the CPM than under the HMM. The extent of the support in favor of the CPM ranges from $BF_{CH} = 1/\exp(-0.26) = 1.3$ in 2008 to $BF_{CH} = 1/\exp(-0.86) = 2.6$ in 2006.

In contrast to Kobe Bryant, Shaquille O'Neal is known as an erratic free-throw shooter at best.[2] Each panel of Fig. 5 shows O'Neal's performance for a specific season. For every season, O'Neal's data are more likely under the HMM than under the CPM. The extent of the support ranges from $BF_{HC} = \exp(0.20) = 1.2$ in 2010 to $BF_{HC} = \exp(6.99) = 1085.7$ in 2006.

In sum, our Bayes factor test reveals that the data consistently suggest Kobe Bryant to be a non-streaky free-throw shooter and Shaquille O'Neal to be a streaky free-throw shooter.

7. Example application 2: perceptual identification

As a second, more elaborate example we analyze data from a visual discrimination task (Gilden & Wilson, 1995a). On each trial, a computer monitor showed two gray squares, one of which brightened for 16 ms; this was the target square that participants were instructed to identify. Brightness was adjusted individually to achieve three different difficulty levels (60, 70, and 90% correct). Each participant completed three blocks in each difficulty condition for a total of nine blocks. The experiment featured four participants, yielding $4 \times 3 \times 3 = 36$ time series overall. Each time series contained 500 trials.

7.1. Methods

We will analyze the Gilden and Wilson time series using our Bayes factor test and compare the results to those of the popular Wald–Wolfowitz runs test (Bradley, 1968). Under the null hypothesis of a constant hitting probability, the number of runs R is asymptotically normally distributed, $R \sim \mathcal{N}(\mathrm{E}(R), \mathrm{Var}(R))$, with $\mathrm{E}(R) = 1 + (2n_1n_2)/n$ and $\mathrm{Var}(R) = (2n_1n_2(2n_1n_2 - n))/(n^2(n-1))$, where n_1 represents the number of runs of successes, n_2 represents the number of runs of failures, and n represents the sequence length. The standardized test statistic R' is the runs z score which is asymptotically distributed as a standard normal. Here we reject the null hypothesis of constant performance (and prefer the alternative hypothesis of streaky performance) whenever $R' < -1.65$ (i.e., a one-side test with significance level .05).

7.2. Results

Fig. 6 plots log BF_{HC} values against runs z scores for each of the 36 time series from Gilden and Wilson. According to the Bayes factor test, 15 time series (42%) show evidence for streakiness, 13 time series (36%) show evidence that is not worth more than a bare mention for streakiness, and 8 time series (22%) show

[1] Among all 197 players active during the 2005–2010 seasons, only 25 had a higher success rate than Bryant.

[2] Among all 197 players active during the 2005–2010 seasons, only 22 had a lower success rate than O'Neal.

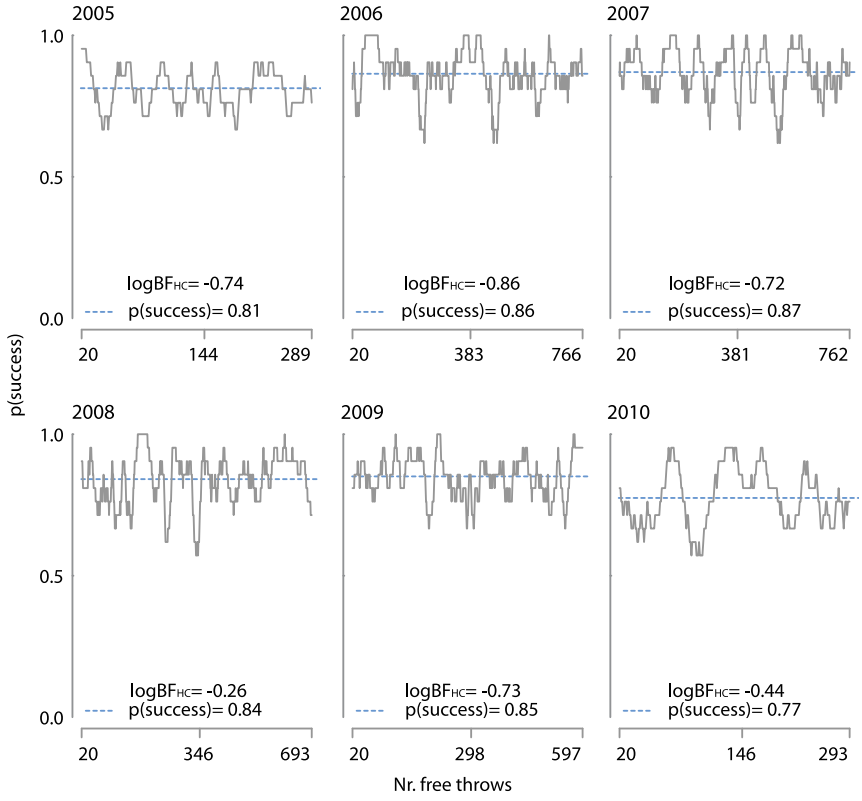

Fig. 4. Kobe Bryant's free-throw shooting across six consecutive NBA seasons. For every season, the Bayes factor favors the constant performance model over the hidden Markov model. For better visualization, all time series are smoothed using a moving window of width 20. The analysis is based on the raw binary results.

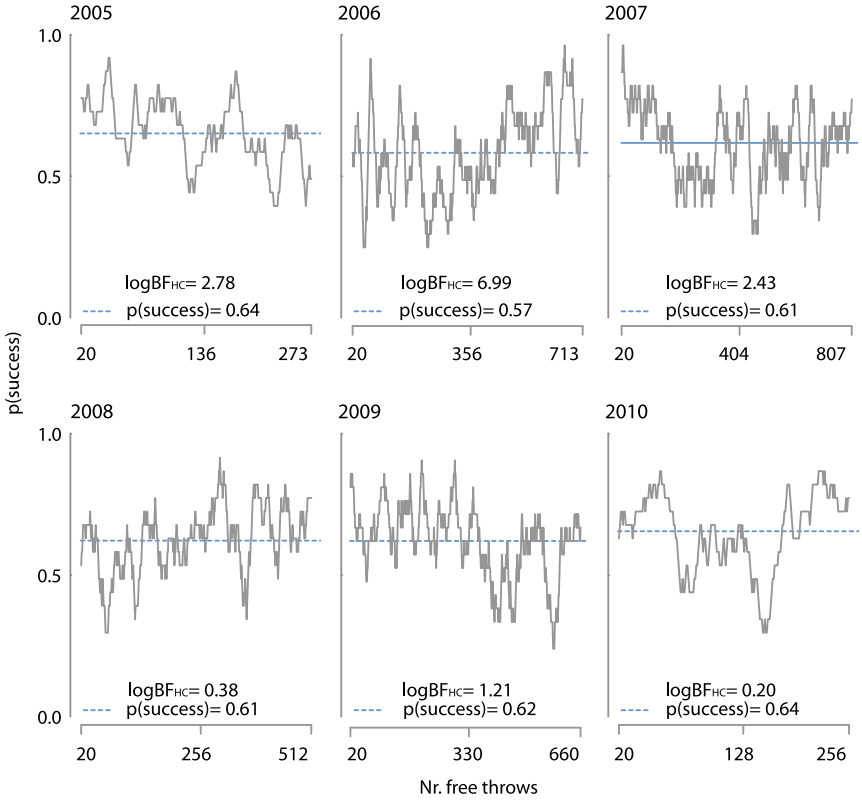

Fig. 5. Shaquille O'Neal's free-throw shooting across six consecutive NBA seasons. For every season, the Bayes factor favors the HMM over the CPM. For better visualization, all time series are smoothed using a moving window of width 20. The analysis is based on the raw binary results.

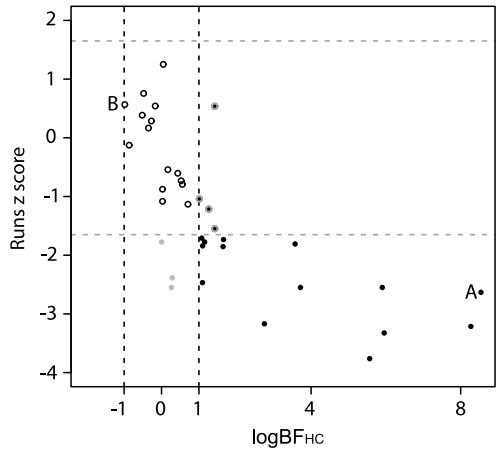

Fig. 6. Results of the Bayes factor test against those of the runs test for 36 visual discrimination time series from Gilden and Wilson. Values of $\log \mathrm{BF}_{\mathrm{HC}} > \log(3)$ (right of the vertical dashed black line) indicate evidence in favor of the HMM, whereas $\log \mathrm{BF}_{\mathrm{HC}} < -\log(3)$ (left of the vertical dashed black line) indicate evidence in favor of the CPM. For the runs test, the null hypothesis of constant performance is rejected when $R' < -1.65$ (below the horizontal dashed gray line).

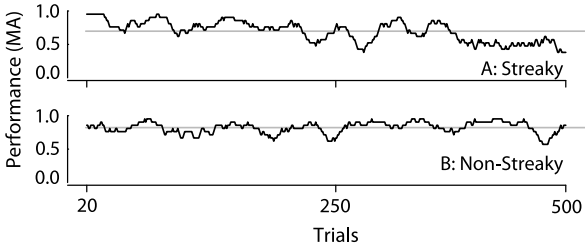

Fig. 7. The upper panel (A: Streaky) shows the Gilden and Wilson time series with the most evidence in favor of the HMM, whereas the lower panel (B: Non-Streaky) shows the Gilden and Wilson time series with the most evidence in favor of the CPM. The gray lines show the average performance. For better visualization, both time series are smoothed using a moving window of width 20.

evidence (not worth more than a bare mention) for the constant performance model. For seven data sets, the evidence in favor of streakiness is either very strong or decisive. According to the runs test, 17 out of 36 time series (47%) are significantly streaky. The solid black dots in Fig. 6 represent those time series for which both tests suggest streakiness. The four black dots with gray border represent the time series for which the Bayes factor indicates evidence for streakiness, while the runs test is undecided. The three solid gray dots represent the time series for which the runs test indicates streakiness, while the Bayes factor approach is undecided. The other dots represent time series for which both methods are undecided.

To provide a visual impression of the time series, Fig. 7 shows the time series marked "A" and "B" in Fig. 6. The time series from panel A yields a Bayes factor of approximately 3900 in favor of the HMM over the CPM. This time series shows a pattern that appears to be streaky with sequences of high performance, and sequences of low performance. In contrast, the non-streaky time series (panel B) does not show periods of pronounced deviation from average performance.

8. Conclusions

Here we outlined a Bayes factor test for the hot hand phenomenon. Inspired by the work of Albert, the Bayes factor test compares the evidence for two models: a hidden Markov model (HMM) that represents streaky performance and a constant performance model (CPM) that represents non-streaky performance. Our implementation of the HMM used appropriately restricted uniform priors on the model parameters: first, the success probability is higher in the hot state than in the cold state, preventing label-switching; second, the switching probability α is only allowed to take on values lower than .5, ensuring states are sticky and not repelling. This second restriction resembles that used in a one-sided hypothesis test where researchers have strong prior expectations about the direction of an effect. For instance, a replication effort by Donnellan, Lucas, and Cesario (2015) concerned the hypothesis that lonely people take hotter showers than people who are not lonely (because lonely people seek to nullify the lack of social warmth with physical warmth from the shower, Bargh & Shalev, 2012). Although prior mass can be assigned to both negative and positive values of effect size, the hypothesis under test emphatically predicts that lonely people take hotter showers, not colder showers (for a Bayesian reanalysis of the bathing data, see Wagenmakers, Verhagen, & Ly, in press). In the same manner, the hot hand phenomenon refers to sticky states, not to repelling states. By incorporating order-restrictions in the specification of the priors, the statistical model becomes a more veridical reflection of the substantive hypothesis at hand, allowing a more informative assessment (e.g., Hoijtink, 2011; Hoijtink, Klugkist, & Boelen, 2008; Mulder et al., 2009; Vanpaemel, 2010; Vanpaemel & Lee, 2012).

Simulation studies showed that when the data are generated by the HMM, the evidence in its favor is often unimpressive, unless the time series is very large (i.e., >2000 observations) or the HMM parameters are extreme (i.e., very sticky states and large discrepancy between the success probabilities in the hot and the cold state). Nevertheless, the time series of Shaquille O'Neal's free-throw performance in the 2006 NBA season produced a Bayes factor of 1085.7 in favor of HMM over CPM; in addition, in the example application of 36 perceptual identification time series each 500 trials long, 7 time series yielded very strong or decisive Bayes factors in favor of the HMM. Such surprisingly high Bayes factors may arise from model misspecification. For instance, the presence of a trend (e.g., a gradual loss of concentration) harms the CPM much more than it harms the HMM. In general, Bayes factors are a measure of relative support, not absolute support. In other words, Bayes factors depend not just on the specification of the streaky model as a two-state HMM, but also on the specification of the non-streaky model as a CPM. Alternative implementations of the non-streaky model are possible and worth considering—for instance, the O'Neal data suggest a non-streaky model that features a linear trend.

Our analysis demonstrates that for Bernoulli time series it may be difficult in practice to discriminate an HMM from a CPM. Perhaps this inherent ambiguity is to blame for the fact that the support in favor of the hot hand in sports is as mixed as it is (Avugos, Köppen, Czienskowski, Raab, & Bar-Eli, 2013; Bar-Eli et al., 2006). Future work can explore several options to create a more powerful test. First, subject-specific knowledge can be included in the prior distributions for the parameters; here we pursued a reference test and assumed independent uniform distributions, but informed priors are likely to result in less ambiguous results (Albert, 1993). Second, one can explore the possibility of collecting and analyzing continuous variables instead of binary variables. Finally, one can extend each of the two competing models by allowing the switching probability to depend on the state, or by adding a hierarchical structure that simultaneously takes into account multiple seasons and multiple players.

Acknowledgments

We thank Dr. David Gilden for kindly providing the data from Gilden and Wilson (1995a). We also thank Dr. Florian Wickelmaier for helpful comments on the code (R Development Core Team, 2012).

Appendix A

Assumption. If $\alpha = .5$ and $\frac{1}{2}(\theta_h + \theta_c) = \theta$ it follows that $L^{(T)}_{\text{HMM}} = L^{(T)}_{\text{CPM}}$.

Proof. The likelihood of the CPM is defined as

$$L^{(T)}_{\text{CMP}} = \theta^k(1-\theta)^{T-k},$$

where k is the number of 1's and T is the length of the data set. The likelihood of the HMM can be written as

$$\begin{aligned} L^{(T)}_{HMM} &= \boldsymbol{\delta}\mathbf{p}(y_1)\Gamma\mathbf{p}(y_2)\ldots\mathbf{p}(y_T)\mathbf{1}' \\ &= \sum_{s_1,\ldots,s_T=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T}p_{s_T}(y_T), \end{aligned}$$

where s_t is the hidden state in time point t, $\gamma_{s_t,s_{t+1}}$ is the probability of switching from state s_t in time t to state s_{t+1} in time $t+1$, and $p_{s_t}(y_t)$ is the probability of the observation y_t given state s_t in time t with $t \in \{1,\ldots,T\}$.

We prove this assumption by induction over the length of the data set T. We begin with the base case for $T = 1$ and differentiate between two cases. In the case of observing a hit, coded as $y_t = 1$ the likelihood of the HMM is:

$$\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}\theta_h & 0\\ 0 & 2\theta-\theta_h\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} = \tfrac{1}{2}\theta_h + \tfrac{1}{2}(2\theta-\theta_h) = \theta.$$

If we observe a miss which is coded as $y_t = 0$ the likelihood of the HMM is:

$$\begin{aligned}&\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}1-\theta_h & 0\\ 0 & 1-(2\theta-\theta_h)\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix} \\ &= \tfrac{1}{2}(1-\theta_h+1-2\theta+\theta_h) = 1-\theta.\end{aligned}$$

Therefore, the assumption holds for $T = 1$. We now show that if the induction hypothesis (IH) $L^{(T-1)}_{\text{HMM}} = \theta^k(1-\theta)^{(T-1)-k}$ holds for $T-1$, it also holds for T. Again we distinguish two cases. The case where the last observation is a hit, $y_T = 1$ and the case were the last observation is a miss $y_T = 0$. If we observe a hit in time T, the likelihood of the HMM can be written as:

$$\begin{aligned} L^{(T)}_{\text{HMM}} &= \sum_{s_1,\ldots,s_T=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T}p_{s_T}(y_T=0) \\ &= \sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T=1}p_{s_T=1}(y_T=0) \\ &\quad + \sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T=1}p_{s_T=2}(y_T=0). \end{aligned}$$

Since $\Gamma = \begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$ if $\alpha = \frac{1}{2}$ it follows that $\gamma_{s_{t-1},s_t} = \frac{1}{2}$ for all t. Additionally, if $\frac{1}{2}(\theta_h + \theta_c) = \theta$ the probability of a miss in the "hot" state, $s_T = 1$, equals $p_{s_T=1}(y_T = 0) = 1-\theta_h$ and the probability of a miss in the "cold" state, $s_T = 0$, equals $p_{s_T=2}(y_T = 0) = 1-(2\theta-\theta_h)$. Using the induction hypothesis for $T-1$, we get:

$$\begin{aligned} L^{(T)}_{\text{HMM}} &= \frac{1}{2}(1-\theta_h)\sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}} \\ &\quad + \frac{1}{2}(1-(2\theta-\theta_h))\sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2} \\ &\quad \times p_{s_2}(y_2)\cdots\gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}} \\ &\overset{IH}{=} \frac{1}{2}(1-\theta_h)\theta^k(1-\theta)^{(T-1)-k} \\ &\quad + \frac{1}{2}(1-(2\theta-\theta_h))\theta^k(1-\theta)^{(T-1)-k} \\ &= \theta^k(1-\theta)^{(T-1)-k}\frac{1}{2}(1-\theta_h+1-2\theta+\theta_h) \\ &= \theta^k(1-\theta)^{T-k}. \end{aligned}$$

If we observe a miss in time T the likelihood of the HMM can be written as:

$$\begin{aligned} L^{(T)}_{\text{HMM}} &= \sum_{s_1,\ldots,s_T=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T}p_{s_T}(y_T=1) \\ &= \sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T=1}p_{s_T=1}(y_T=1) \\ &\quad + \sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}}\gamma_{s_{T-1},s_T=1}p_{s_T=2}(y_T=1). \end{aligned}$$

Again, $\gamma_{s_{t-1},s_t=1} = \frac{1}{2}$ for all t and the probability of a hit in the "hot" state equals $p_{s_T=1}(y_T = 1) = \theta_h$ and the probability of a hit in the "cold" state $p_{s_T=2}(y_T = 1) = 2\theta-\theta_h$. Again, we use the induction hypothesis for $T-1$ and get:

$$\begin{aligned} L^{(T)}_{\text{HMM}} &= \frac{1}{2}\theta_h\sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots\gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}} \\ &\quad + \frac{1}{2}(2\theta-\theta_h)\sum_{s_1,\ldots,s_{T-1}=1}^{2} \delta_{s_1} p_{s_1}(y_1)\gamma_{s_1,s_2}p_{s_2}(y_2)\cdots \\ &\quad \times \gamma_{s_{T-1},s_{T-2}}p_{s_{T-1}} \\ &\overset{IH}{=} \frac{1}{2}\theta_h\theta^k(1-\theta)^{(T-1)-k} + \frac{1}{2}(2\theta-\theta_h)\theta^k(1-\theta)^{(T-1)-k} \\ &= \theta^k(1-\theta)^{(T-1)-k}\frac{1}{2}(\theta_h+2\theta-\theta_h) \\ &= \theta^{k+1}(1-\theta)^{(T-1)-k} = \theta^{k+1}(1-\theta)^{T-(k+1)}. \end{aligned}$$

Appendix B

```
## requires lattice and HiddenMarkov to run
library(lattice)
library(HiddenMarkov)

# Example: Sequence of Carlos Guillen's batting outcomes
# for the 2005 season (Albert, 2008), 1 is a hit and 0 is
# out.

Guillen.data <-c(
0,1,0,1,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,1,1,0,0,0,0,0,0,
1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,1,0,1,0,0,1,1,0,
1,0,1,1,0,1,0,1,1,0,0,0,0,0,1,1,1,1,0,0,1,0,1,0,0,1,1,0,0,
0,1,0,1,0,0,0,1,1,1,0,1,1,1,1,0,0,1,1,1,1,0,0,1,0,0,1,0,1,
0,0,0,1,0,1,0,0,0,0,0,0,1,1,1,0,0,1,0,0,0,0,0,0,1,1,1,0,1,0,
0,0,0,1,1,1,1,0,1,0,0,0,1,0,0,0,0,0,0,1,0,0,0,0,1,1,0,1,0,
0,0,0,0,0,0,0,0,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,
0,1,0,0,0,0,1,1,0,1,0,0,1,0,0,0,1,0,1,0,0,0,0,0,0,0,0,1,0,
1,0,0,1,1,1,1,0,0,0,0,0,1,1,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,1,0,0,
0,1,1,0,0,1,0,0,0,1,0,0,0,0,1,0,0,0,0,1,1,0,1,1,1,0,1,0,0,
0,0,0,1,0,1,1,0,0,0,1,0,0,0,1)

## the Bayes factor function
## use LogBF = LogBayesfactorHMM(data, gridprecision)
## input parameters are:
## dat: the data in vector form
## intprec: the precision of the integration process, higher is more precise
LogBayesfactorHMM <- function(dat,intprec=50){
 grid1 <- 1 / (2 * (intprec - 1)) + seq(0, 1, length = intprec)[-intprec]
 grid2 <- 1 / (2 * intprec) + seq(0, 0.5, length = intprec)[-intprec]
 len1 <- length(grid1)
 len2 <- length(grid2)
 indices <- which(upper.tri(matrix(TRUE, len1, len1)), arr.ind = TRUE)
 zzz <- matrix(0, nrow(indices), len1)
 n <- length(dat)
 m <- 2L
 phi <- as.double(c(0.5, 0.5))
 logalpha <- matrix(as.double(rep(0, m * n)), nrow = n)
 lscale <- as.double(0)
 memory0 <- rep(as.double(0), m)
 for(k in seq_len(len2)){
  xPi  <- matrix(c(1 - grid2[k], grid2[k],
                   grid2[k], 1 - grid2[k]), 2, 2)
   for(i in seq_len(nrow(zzz))){
    prob <-
    cbind(grid1[indices[i, 1]]^dat * (1 - grid1[indices[i, 1]])^(1 - dat),
          grid1[indices[i, 2]]^dat * (1 - grid1[indices[i, 2]])^(1 - dat))
    zzz[i, k] <- .Fortran("loop1", m, n, phi, prob, xPi, logalpha,
                          lscale, memory0, PACKAGE = "HiddenMarkov")[[7]]
   }
 }
 f <- 699 - max(zzz, na.rm = TRUE)
 zzz <- exp(zzz + f)
 kMarginalHMM <- log(mean(zzz,na.rm=T)) - f
 kMarginalCPM  <- lgamma(length(dat[dat == 1]) + 1) + lgamma(length(dat)
                 - length(dat[dat == 1]) + 1) - lgamma(length(dat) + 2)
 BF <- kMarginalHMM - kMarginalCPM
 return(BF)
}

# compute log Bayes factor HMM/CPM
# positive log(BF) indicates evidence in favor of HMM
LogBayesfactorHMM(Guillen.data) ## should be approx 0.45
```

References

Albert, J. (1993). A statistical analysis of hitting streaks in baseball: Comment. *Journal of the American Statistical Association*, *424*, 1184–1188.

Albert, J. (2008). Streaky hitting in baseball. *Journal of Quantitative Analysis in Sports*, *4*, 1–32.

Albert, J., & Williamson, P. (2001). Using model/data simulations to detect streakiness. *The American Statistician*, *55*, 41–50.

Albright, S. C. (1993). A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association*, *88*, 1175–1183.

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, *37*, 3099–3132. http://dx.doi.org/10.1214/09-AOS689.

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

Bar-Eli, M., Avugos, S., & Raab, M. (2006). Twenty years of "hot hand" research: Review and critique. *Psychology of Sports and Exercise*, *7*, 525–553.

Bargh, J. A., & Shalev, I. (2012). The substitutability of physical and social warmth in daily life. *Emotion*, *12*, 154–162.

Barry, D., & Hartigan, J. A. (1993). Choice models for predicting divisional winners in major league baseball. *Journal of the American Statistical Association*, *88*, 766–774.

Bradley, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice-Hall.

Clark, R. D. (2005). Examination of hole-to-hole streakiness on the PGA tour. *Perceptual and Motor Skills*, *100*, 806–814.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Donnellan, M. B., Lucas, R. E., & Cesario, J. (2015). On the association between loneliness and bathing habits: Nine replications of Bargh and Shalev (2012) Study 1. *Emotion*, *15*, 109–119.

Dorsey-Palmateer, R., & Smith, G. (2004). Bowlers' hot hands. *The American Statistician*, *58*, 38–45.

Gilden, D. L., & Wilson, S. G. (1995a). On the nature of streaks in signal detection. *Cognitive Psychology*, *28*, 17–64.

Gilden, D. L., & Wilson, S. G. (1995b). Streaks in skilled performance. *Psychonomic Bulletin & Review*, *2*, 260–265.

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

Harte, D. (2011). HiddenMarkov: Hidden Markov Models [Computer software manual]. Wellington Retrieved from http://cran.at.r-project.org/web/packages/HiddenMarkov (R package version 1.5–0).

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot–density trace synergism. *The American Statistician*, *52*, 181–184.

Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: Chapman & Hall/CRC.

Hoijtink, H., Klugkist, I., & Boelen, P. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.

Jeffreys, H. (1961). *Theory of probability* (3rd ed). Oxford, UK: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.

Lopes, L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 392–400.

Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, *53*, 530–546.

Mulder, J., & Wagenmakers, E.-J. (2016). Editor's introduction to the special issue on bayes factors for testing hypotheses in psychological research: practical relevance and new developments. *Journal of Mathematical Psychology*, *72*, 1–5.

Petrie, T. (1969). Probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, *40*, 97–115.

R Development Core Team, (2012). R: a language and environment for statistical computing [omputer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/ (ISBN 3-900051-07-0).

Raab, M., Gula, B., & Gigerenzer, G. (2012). The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimantal Psychology*, *18*, 81–94.

Shea, S. (2014). In support of a hot hand in professional basketball and baseball. *PsyCh Journal*, *3*, 159–164.

Sun, Y. (2004). Detecting the hot hand: An alternative model. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual conference of the cognitive science society* (pp. 1279–1284). Chicago, IL: Lawrence Erlbaum.

Sun, Y., & Wang, H. (2012). The "hot hand" revisited: A nonstationarity argument. *PsyCh Journal*, *1*, 28–39.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Vanpaemel, W. (2010). Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, *54*, 491–498.

Vanpaemel, W., & Lee, M. D. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, *19*, 1047–1056.

Wagenmakers, E.-J., Verhagen, A. J., & Ly, A. How to quantify the evidence for the absence of a correlation, Behavior Research Methods (in press).

Wardrop, R. L. (1995). Simpson's paradox and the hot hand in basketball. *The American Statistician*, *49*, 24–28.

Wardrop, R. L. (1999). Statistical tests for the hot-hand in basketball in a controlled setting. http://www.stat.wisc.edu/~wardrop/papers/tr1007.pdf.

Yaari, G., & Eisenmann, S. (2011). The hot (invisible?) hand: Can time sequence patterns of success/failure in sports be modeled as repeated random independent trials? *PLoS ONE*, *6*, e24532.

Zucchini, W., & MacDonald, I. L. (2009). *Hidden Markov models for time series: An introduction using R*. Boca Raton, FL: Chapman and Hall/CRC.