A Bayesian Test for the "Hot Hand"

Darja Tutschkow[1], Conor Dolan[2], Gilles Dutilh[2], Ruud Wetzels[2], Sophie von der Sluis[3], Eric-Jan Wagenmakers[2]

Darja Tutschkow[1]
darja.tutschkow@student.uni-tuebingen.de

[1]Eberhard Karls Universität Tübingen

[2]University of Amsterdam

[3]Free University, Amsterdam

2/9/2012

What is the "Hot Hand"?

- „The ‚hot hand' and ‚streak shooting'-terms refer to the belief that the performance of a player during a particular period is significantly better than could be expected on the basis of the player's overall record." (Gilovich, Vallone and Tversky, 1985)
- The same claim was made by Gilden & Wilson (1995, Cognitive Psychology) about people's performance in simple perceptual tasks.

2005 Batting Outcomes from Carlos Guillen

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, 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

Is Carlos Guillen a streaky player?

Outline

- Current tests for streakiness
- A Bayesian test for streakiness
- Application to real data
- Easy to detect streakiness?

Outline

- Current tests for streakiness
- A Bayesian test for streakiness
- Application to real data
- Easy to detect streakiness?

Current Tests for Streakiness
Examples

- The longest run of hits (Albert, 2008)
- Runs test (Gillovich, Valone & Tversky, 1985)
- Test of stationarity (Gillovich, Valone & Tversky, 1985)
- Black statistic (Albert, 2008)

Current Tests for Streakiness
Problems

- Existing tests are mostly classical or frequentist, and only consider the null hypothesis.
- The tests have very low power.
- This means that it is not very informative when one “fails to reject the null hypothesis”.
- The tests sometimes use ad-hoc division of the data in epochs. But the size of the epoch affects the result (black statistic).

Outline

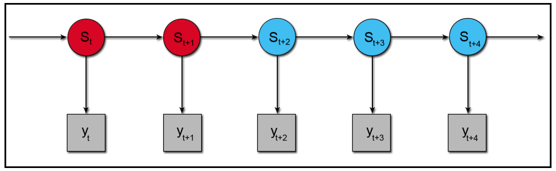
- Current tests for streakiness
- A Bayesian test for streakiness
- Application to real data
- Easy to detect streakiness?

A Bayesian Test for Streakiness

- We want to assess the evidence for and against the hypothesis of streaky performance.
- We contrast two models:
  - The constant-performance model(CpM)
  - A three-parameter hidden Markov model(HMM) – this model is in line with one's intuition of streakiness.

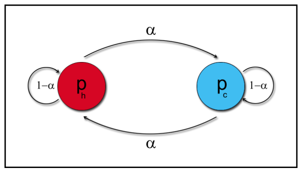
A Bayesian Test for Streakiness
The HMM

- Assume a state space $\{S_t : t \in \mathbb{N}\}$ with two possible states $S_t \in \{0, 1\}$:



- The state space $\{S_t : t \in \mathbb{N}\}$ satisfies the Markov property: $Pr(S_t = s_t | S_{(t-1)} = s_{(t-1)}, \ldots, S_1 = s_1) =$ $Pr(S_t = s_t | S_{(t-1)} = s_{(t-1)})$

A Bayesian Test for Streakiness
The HMM



where

- $p_h = Pr(1|S_t = 1)$ is the probability of a hit in the hot state
- $p_c = Pr(1|S_t = 0)$ is the probability of a hit in the cold state
- $\alpha = Pr(S_t = 1|S_{(t-1)} = 0) = Pr(S_t = 0|S_{(t-1)} = 1)$ is the probability of switchig between states

A Bayesian Test for Streakiness
Bayes Factor

- After seeing the data, which model is preferable?
- The one with the higher posterior probability!

$$\frac{Pr(HMM|Data)}{Pr(CpM|Data)} = \frac{Pr(Data|HMM)}{Pr(Data|CpM)}$$

- Before seeing the data both models are assumed to be equally likely $\Rightarrow \frac{Pr(HMM)}{Pr(CpM)} = 1$
- To choose a model we compute the Bayes factor (BF)

$$\frac{Pr(Data|HMM)}{Pr(Data|CpM)}$$

- The Bayes factor is the change from prior to posterior odds brought about by the data.
- Quantifies the evidence for one versus the other model provided by the data.

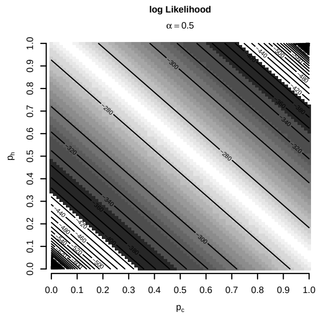
A Bayesian Test for Streakiness
Bayes Factor

$$\frac{Pr(Data|HMM)}{Pr(Data|CpM)} = \frac{\int_0^1 \int_0^1 \int_0^{p_h} Pr(Data|(p_c,p_h,\alpha))Pr(p_c)Pr(p_h)Pr(\alpha)dp_c dp_h d\alpha}{\int_0^1 Pr(Data|p)Pr(p)dp}$$

- with $\alpha, p_h, p_c \in (0,1)$ and $p_h > p_c$.
- we assume independent uniform priors for p_h, p_c and α
- By averaging over the likelihood we discount for model complexity (Myung and Pitt, 1997)

A Bayesian Test for Streakiness
Complications

- Parameter point estimation is useless in many situations because the parameters are highly correlated.
- For example or a paramter value of $\alpha = .5$, p_h and p_c are in perfect tradeoff.



- But this is irrelevant for our test.

Outline

- Current tests for streakiness
- A Bayesian test for streakiness
- Application to real data
- How easy is it to detect streakiness?

Application to Real Data
Flash Data (Gilden & Wilson, 1995)

- 36 time series, each with 500 trials
- each trail involves a brightness discrimination judgment, and is scored as „correct" or "incorrect"
- we compared the results of an often used test for streakiness - runs test - with the Bayes factor

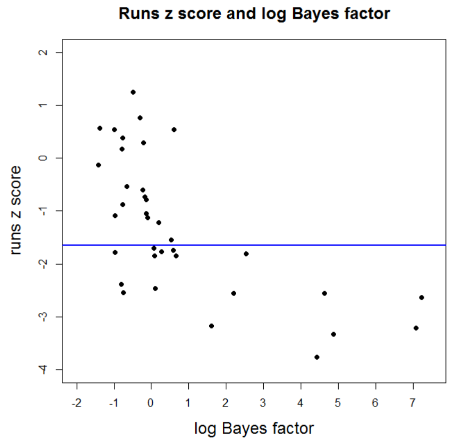
Application to Real Data
Runs z Score

1. The idea: What is the distribution of runs (clusters of 1's and 0's) under a constant hitting probability?
2. The amount of runs R is normally distributed,

   $R \sim N(\frac{2n_1n_2}{n}+1, \frac{2n_1n_2(2n_1n_2-n)}{n^2(n-1)})$

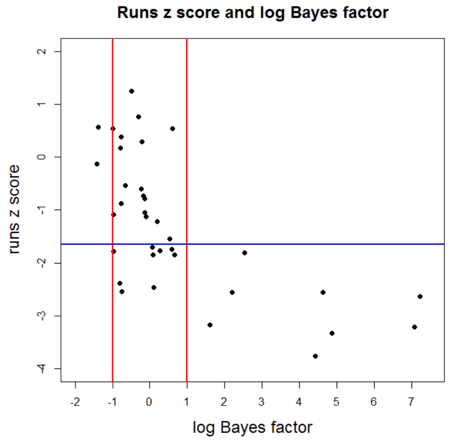
   with n_1 =“runs of hits“, n_2 =“runs of misses“, n=“sequence length“
3. If runs z score < -1.65 there are significantely fewer runs than would be expected under a constant hitting probability.

Application to Real Data
Flash Data (Gilden & Wilson, 1995)



- Runs z score: 48% of the time series are streaky

Application to Real Data
Flash Data (Gilden & Wilson, 1995)



- Runs z score: 48% of the time series are streaky
- Log Bayes factor: 22% of the time series are streaky

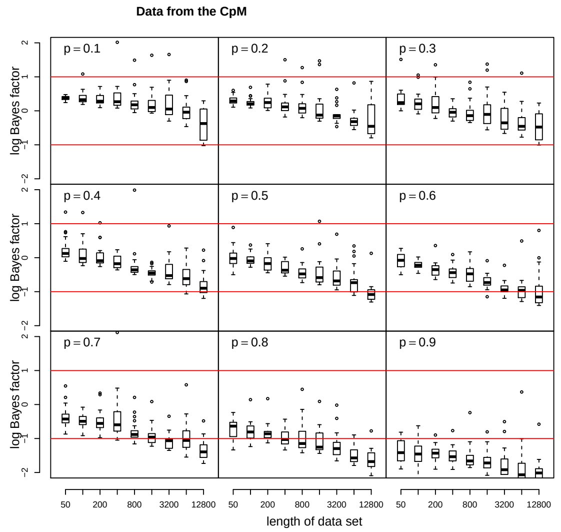
Outline

- Current tests for streakiness
- A Bayesian test for streakiness
- Application to real data
- How easy is it to detect streakiness?

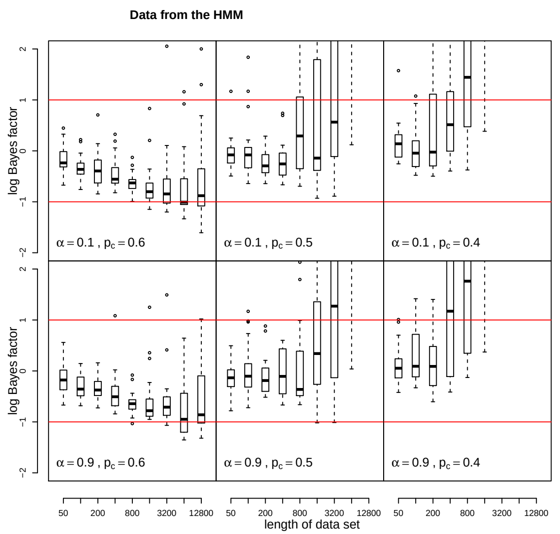
How Easy is it to Detect Streakiness?

1. Simulated data from the HMM
   - for different parameter values of α and p_c, keeping $p_h = .7$ constant. (Former simulation studies showed the difference between p_c and p_h is more important than their absolute values)
   - for different lengths of data sets
2. Simulated data from the CpM
   - for different values of p
   - for different lengths of data sets
3. Calculated the log Bayes factor
4. Calculated the runs z score to compare the results

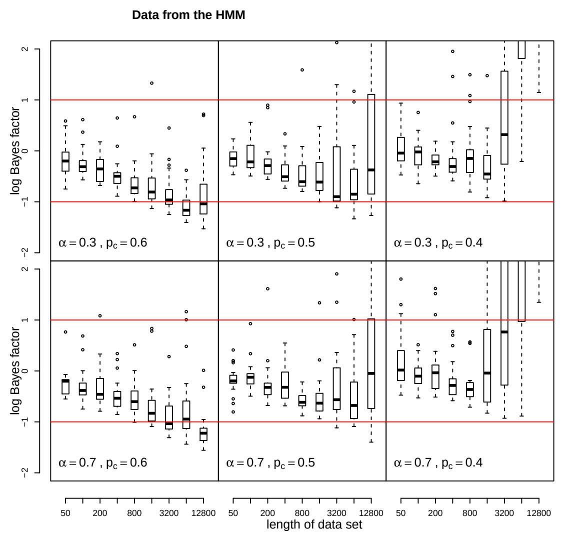
How Easy is it to Detect Streakiness? The Bayes factor



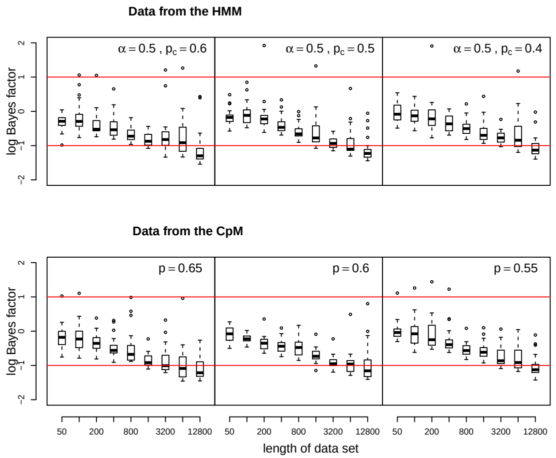
How Easy is it to Detect Streakiness? The Bayes factor



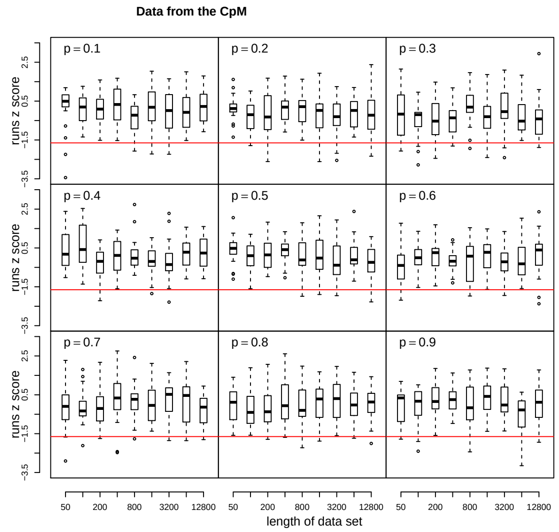
How Easy is it to Detect Streakiness? The Bayes factor



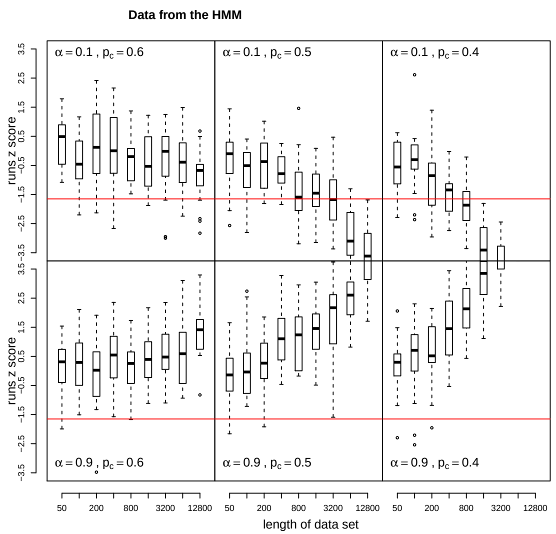
How Easy is it to Detect Streakiness?
The Bayes factor



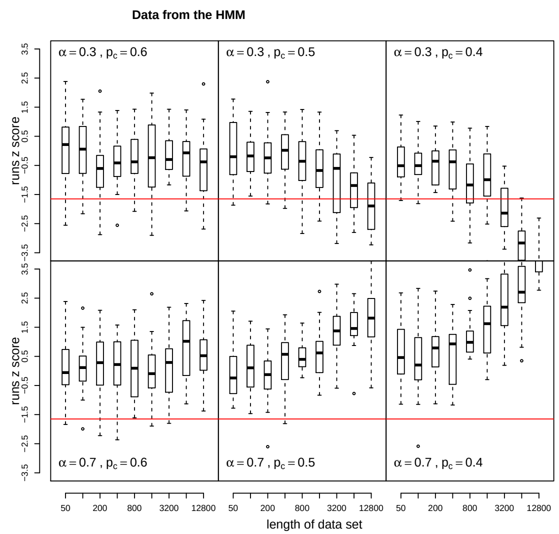
How Easy is it to Detect Streakiness?
Runs z Score



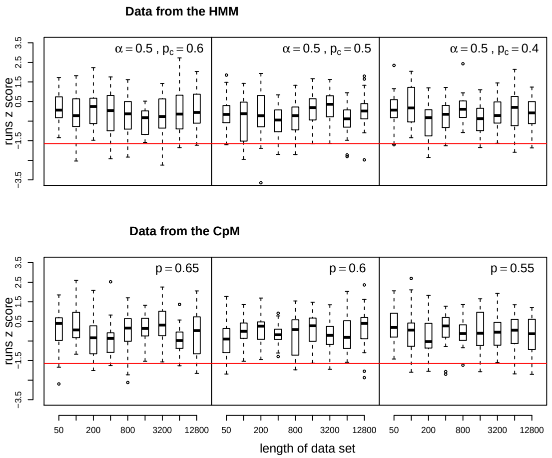
How Easy is it to Detect Streakiness?
Runs z Score



How Easy is it to Detect Streakiness?
Runs z Score



How Easy is it to Detect Streakiness? Runs z Score



Summary

- The higher the probability of a hit under the CpM, the bigger the evidence in favour of the CpM when taking the Bayes factor.
- The runs test is not sensitive to that because it considers just the null hypothesis.
- Given a small value of α - sticky states - the evidence in favour of the HMM gets bigger with a bigger difference between p_h and p_c.
- The difference in the hitting probability between „hot" and „cold" states has to be quite big and the states have to be sticky to have a chance of detecting streakiness.

Summary

- For differences in the hitting probabilities smaller than .3 it is unlikely to detect streakiness even for large datasets.
- The Bayes factor is not able to discriminate between small and high values of α.
- But maybe it would be more in line with the definition of streakiness to assume $\alpha < .5$
- In general the results of the Bayesian test and the runs test show the same pattern.
- But with the Bayes factor you can also get evidence in favour of the CpM.

2005 Batting Outcomes from Carlos Guillen

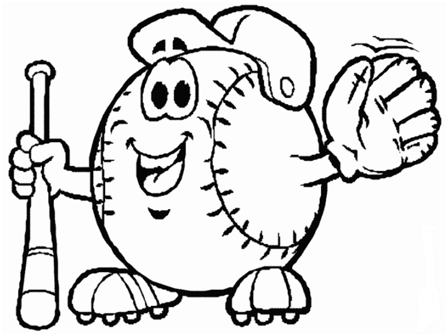
0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0,
1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0,
1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0,
0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 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, 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

Is Carlos Guillen a streaky player?

2005 Batting Outcomes from Carlos Guillen

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, 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

- log Bayes factor = .275 $\Rightarrow$ both models are alomst equally likely given this data
- runs z score = $-.811$ $\Rightarrow$ no significant evidence for streakiness



Thanks for your attention!