Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers

Joshua B. Miller and Adam Sanjurjo *†

July 6, 2015

Abstract

We find a subtle but substantial bias in a standard measure of the conditional dependence of present outcomes on streaks of past outcomes in sequential data. The mechanism is driven by a form of selection bias, which leads to an underestimate of the true conditional probability of a given outcome when conditioning on prior outcomes of the same kind. The biased measure has been used prominently in the literature that investigates incorrect beliefs in sequential decision making—most notably the Gambler's Fallacy and the Hot Hand Fallacy. Upon correcting for the bias, the conclusions of some prominent studies in the literature are reversed. The bias also provides a structural explanation of why the belief in the law of small numbers persists, as repeated experience with finite sequences can only reinforce these beliefs, on average.

JEL Classification Numbers: C12; C14; C18;C19; C91; D03; G02.

Keywords: Law of Small Numbers; Alternation Bias; Negative Recency Bias; Gambler's Fallacy; Hot Hand Fallacy; Hot Hand Effect; Sequential Decision Making; Sequential Data; Selection Bias; Finite Sample Bias; Small Sample Bias.

*Miller: Department of Decision Sciences and IGIER, Bocconi University, Sanjurjo: Fundamentos del Análisis Económico, Universidad de Alicante. Financial support from the Department of Decision Sciences at Bocconi University, and the Spanish Ministry of Economics and Competitiveness under project ECO2012-34928 is gratefully acknowledged.

†Both authors contributed equally, with names listed in alphabetical order.

We shall encounter theoretical conclusions which not only are unexpected but actually come as a shock to intuition and common sense. They will reveal that commonly accepted notions concerning chance fluctuations are without foundation and that the implications of the law of large numbers are widely misconstrued. (Feller 1968)

1 Introduction

Jack takes a coin from his pocket and decides that he will flip it 4 times in a row, writing down the outcome of each flip on a scrap of paper. After he is done flipping, he will look at the flips that immediately followed an outcome of heads, and compute the relative frequency of heads on those flips. Because the coin is fair, Jack of course expects this *conditional* relative frequency to be equal to the probability of flipping a heads: 0.5. Shockingly, Jack is wrong. If he were to sample 1 million fair coins and flip each coin 4 times, observing the conditional relative frequency for each coin, on average the relative frequency would be approximately 0.4.

We demonstrate that in a finite sequence generated by i.i.d. Bernoulli trials with probability of success p, the relative frequency of success on those trials that immediately follow a streak of one, or more, consecutive successes is expected to be *strictly less than* p, i.e. the conditional relative frequency of success is a biased estimator of the conditional probability of success. While, in general, the bias does decrease as the sequence gets longer, it remains substantial for a range of sequence lengths often used in empirical work.

This result has considerable implications for the study of decision making in any environment which involves sequential data. For one, it provides a structural explanation for the persistence of one of the most well-documented, and robust, systematic errors in beliefs—that people have an *alternation bias* (also known as *negative recency bias*; see Bar-Hillel and Wagenaar [1991]; Nickerson [2002]; Oskarsson, Boven, McClelland, and Hastie [2009]) —by which they believe, for example, that when observing multiple flips of a fair coin, an outcome of heads is more likely to be followed by a tails than by another heads, as well as the closely related *gambler's fallacy* (see Bar-Hillel and Wagenaar (1991)), in which this alternation bias *increases* with the length of the streak of heads. Thus, empirical properties of conditional relative frequencies are found to be consistent with the types of subjective inference that have been conjectured in behavioral models of the law of small numbers, as in Rabin (2002); Rabin and Vayanos (2010). Further, the result shows that data in the *hot hand fallacy* literature (see Gilovich, Vallone, and Tversky [1985] and Miller and Sanjurjo [2014, 2015a,b]) has been systematically misinterpreted by researchers; for those trials that immediately follow a streak of successes, observing that the relative frequency of successes is equal to the base rate of success, is in fact evidence in favor of the hot hand, rather than evidence against it. Tying these two implications together, it becomes clear why the inability of the gambler to detect the fallacy of his belief in alternation has an exact parallel with the researcher's inability to detect his mistake when concluding that experts' belief in the hot hand is a fallacy.

In addition, the result has implications for evaluation and compensation systems, and suggests successful gambling systems for (particular) games of finite length.[1] The fact that a coin is expected to exhibit an alternation "bias" in finite sequences, implies that coin flips can be successfully "predicted" in finite sequences at rates better than that of chance. For example, suppose that each day a stock index goes either up or down, according to a random walk in which the probability of going up is, say, 0.6. A financial analyst who can predict the next day's performance on the days she chooses to, and whose predictions are evaluated in terms of how her success rate on predictions in a given month compares to that of chance, can expect to outperform this benchmark by using any of a number of different decision rules. For instance, she can simply predict "up" immediately following down days, or increase her expected relative performance even further by predicting "up" only immediately following longer streaks of consecutive down days.[2]

On the surface, it appears impossible that the relative frequency of heads, on those flips of a fair coin that immediately follow a heads, is expected to be less than 1/2, and thus, by implication, that the relative frequency of heads on those flips that immediately follow a tails is expected to be greater than 1/2. After all, each coin flip is assigned at random either to the group of flips that immediately follow a heads, or to the group of flips that immediately follow a tails. However, what this simple argument overlooks is that, for any given sequence, when calculating the relative

[1] See Feller (1968) for a treatment of gambling systems for games of infinite length.

[2] Similarly, it is easy to construct betting games that act as money pumps while defying intuition. For example, we can offer the following lottery at a $5 ticket price: a fair coin will be flipped 4 times. if the relative frequency of heads on flips that immediately follow a heads is greater than 0.5 then the ticket pays $10; if the relative frequency is less than 0.5 then the ticket pays $0; if the relative frequency is exactly equal to 0.5, or if no flip is immediately preceded by a heads, then a new sequence of 4 flips is generated. While, intuitively, it seems like the expected payout of this ticket is $0, it is actually $-0.71 (see Table 1). Curiously, this betting game may be more attractive to someone who believes in the independence of coin flips, rather that someone who holds the Gambler's fallacy.

frequency of heads in each group, the assignment of flips to each group occurs *after* the sequence has been generated. Thus, when selecting, from the sequence, the group of flips that immediately follow a heads, the first flip from each run of heads is excluded, and this *selection bias* is more severe the more runs there are in the sequence.[3] This problem can be understood clearly by returning to Jack's 4 coin flips in the opening example, and considering all 16 possible sequences that can be generated. For each of these sequences, Table 1 reports the conditional relative frequency $\hat{p}(H|H)$, which is the empirical rate of heads on those trials that immediately follow a heads (repetition rate for heads); the conditional relative frequency $\hat{p}(H|T)$, which is the empirical rate of heads on those trials that immediately follow a tails (alternation rate for tails), along with the difference between these two quantities $\hat{p}(H|H) - \hat{p}(H|T)$. The expectation of each quantity, as well as the expectation of the difference, is computed in the last row of the table. As can be seen, the expected rate of repetition on trials that immediately follow a heads is not 0.5, but 0.4, while the expected rate of alternation on trials that immediately follow a tails is not 0.5, but 0.6, and the expected difference (which is based only on sequences in which the difference can be computed) is not 0, but -0.33.[4] This example illustrates why the law of large numbers, as the number of observed sequences of fixed size goes to infinity, will not lead the conditional relative frequencies to converge to their respective underlying conditional probabilities; here, instead they converge to the levels presented in the final row of Table 1.

A counterfactual approach to measuring relative frequency, which eliminates the bias and helps in understanding it further, involves computing a single relative frequency, using all flips that immediately follow a heads, regardless of the sequence in which they appear (rather than averaging the relative frequencies across the sequences). With this approach there is no bias, which reveals that the bias arises from the fact that the expected relative frequency is a clustered average, i.e. the relative frequency for each sequence is unweighted with respect to the number of flips for which it is calculated. As can be seen in Table 1, sequences that have more flips of heads tend to have a higher relative frequency of repetition. Thus, these higher relative frequency sequences are "underweighted" when taking the (clustered) average of the relative frequencies across all (here,

[3] For example, in the sequence $0\underline{1}100\underline{1}110\underline{1}0$ in which there are three runs of ones, 1, 11 and 111, and two runs of zeros, 0 and 00, the flips with underlined 1s are excluded, i.e. they cannot immediately follow a 1.

[4] This type of expected difference in conditional relative frequencies is the standard measure of effect size in the most prominent studies of the hot hand fallacy, under the incorrect assumption that the expected difference is zero (see Section 3).

Table 1: *Illustrating the true expected conditional sample relative frequencies of heads for 4 flips of a fair coin.*

# Heads	Sequence	$\hat{p}(H\|H)$	$\hat{p}(H\|T)$	$\hat{p}(H\|H) - \hat{p}(H\|T)$
0	TTTT	-	0	-
1	TTTH	-	.33	-
	TTHT	0	.5	-.5
	THTT	0	.5	-.5
	HTTT	0	0	0
2	TTHH	1	.5	.5
	THTH	0	1	-1
	THHT	.5	1	-.5
	HTTH	0	.5	-.5
	HTHT	0	1	-1
	HHTT	.5	0	.5
3	THHH	1	1	0
	HTHH	.5	1	-.5
	HHTH	.5	1	-.5
	HHHT	.66	-	-
4	HHHH	1	-	-
Average		.40	.60	-.33

equiprobable) sequences, which results in an average that indicates alternation. The implications for learning are stark: so long as decision makers experience finite length sequences, and simply observe the relative frequencies of one outcome when conditioning on previous outcomes in each sequence, they will never unlearn a belief in the gambler's fallacy.[5,6]

That the bias emerges when the conditional relative frequency—the estimator of conditional probability—is clustered at the sequence level suggests a connection to the well-known finite sample bias of the least squares estimator of autocorrelation in time series data (Stambaugh 1986; Yule 1926). Indeed, we find that this connection is more than suggestive; letting $\mathbb{P}(x_i = 1|x_{i-1} = \cdots = x_{i-k} = 1)$ be the probability of a successes in trial i, conditional on a success streak of length k (or more) in immediately preceding k trials, the least squares estimator for the coefficients in the associated linear probability model, $x_i = \beta_0 + \beta_1 \mathbb{1}_{[x_{i-1}=\cdots=x_{i-k}=1]}(i)$, happens to be the conditional relative frequency $\hat{p}(x_i = 1|x_{i-1} = \cdots = x_{i-k} = 1)$. Therefore, the explicit formula for the bias

[5] For a full discussion see Section 3.

[6] When learning is deliberate rather than intuitive, this result holds as well if conditional relative frequencies are instead represented as conditional natural frequencies (e.g. 3 in 4 rather than .75), under the assumption that people neglect sample size (see Kahneman and Tversky (1972) and Benjamin, Rabin, and Raymond (2014)).

that we find below can be applied directly to the coefficients of the associated linear probability model (see Appendix B.3). Furthermore, this bias can be corrected with the appropriate form of re-sampling, related to what has been done in the study of time series data (Nelson and Kim 1993), but in this case using permutation test procedures (see Miller and Sanjurjo (2014) for details).

In Section 2 (and in Appendix B), we derive an explicit formula for the expected conditional relative frequency of successes for any probability of success p, any streak length k, and any sample size n. While this formula does not appear, in general, to admit a closed form representation, for the special case of streaks of length $k = 1$ (as in the example above) one is provided. For k larger than one, we compute the expected conditional relative frequency for sequences with lengths relevant for empirical work, using results on the sampling distribution developed in Appendix B.

Section 3 begins by detailing how the result provides a structural explanation of the persistence of both alternation bias, and gambler's fallacy beliefs. In addition, results are reported from a simple survey that we conducted in order to gauge whether the degree of alternation bias that has typically been observed from experimental subjects in previous studies is roughly consistent with subjects' experience with finite sequences outside of the laboratory. Then we discuss how the result reveals a incorrect assumption made in the analyses of the most prominent hot hand fallacy studies. Once corrected for, the previous findings are reversed.

2 Result

We find an explicit formula for the expected conditional relative frequency of a success in a finite sequence of i.i.d. Bernoulli trials with any probability of success p, any sequence length n, and when conditioning on streaks of successes of any length k. Further, we quantify the bias for a range of n and k relevant for empirical work. In this section, for the simplest case of $k = 1$, we present a closed form representation of this formula. To ease exposition, the statement and proofs of results for k larger than one are treated in Appendix B.

2.1 Expected Bias

Let $X_1, \ldots, X_n$ be a sequence of n i.i.d Bernoulli trials with probability of success $p := \mathbb{P}(X_i = 1)$. Let the sequence $\boldsymbol{x} = (x_1, ..., x_n) \in \{0, 1\}^n$ be the realization of the n trials, $N_1(\boldsymbol{x}) := \sum_{i=1}^{n} x_i$ the number of ones, and $N_0(\boldsymbol{x}) := N - N_1(\boldsymbol{x})$ the number of zeros, with their respective realizations n_1 and n_0. We begin with two definitions.

Definition 1 *For $\boldsymbol{x} \in \{0,1\}^n$, $n_1 \geq 0$, and $k = 1, \ldots, n_1$, the set of* k/1-streak *successors $I_{1k}(\boldsymbol{x})$ is the subset of trials that immediately succeed a streak of ones of length k or more, i.e.*

$$I_{1k}(\boldsymbol{x}) := \{i \in \{k+1, \ldots, n\} : x_{i-1} = \cdots = x_{i-k} = 1\}$$

The *k/1-streak momentum statistic* $\hat{P}_{1k}(\boldsymbol{x})$ is the conditional relative frequency of 1, for the subset of trials which are k/1-streak successors, i.e.

Definition 2 *For $\boldsymbol{x} \in \{0,1\}^n$, if the set of k/1-streak successors satisfies $|I_{1k}(\boldsymbol{x})| > 0$, then the k/1-streak momentum statistic is defined as:*

$$\hat{P}_{1k}(\boldsymbol{x}) := \frac{\sum_{i \in I_{1k}} x_i}{|I_{1k}(\boldsymbol{x})|}$$

otherwise it is not defined

Thus, this statistic measures the success rate on the subset of trials that immediately succeed a streak of success(es). Let P_{1k} be the derived distribution with support $\{p_{1k} \in [0,1] : p_{1k} = \hat{P}_{1k}(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \{0,1\}^n,\ |I_{1k}(\boldsymbol{x})| > 0\}$ and distribution determined by the Bernoulli trials. The expected value of P_{1k} is equal to the expected value of $\hat{P}_{1k}(\boldsymbol{x})$ across all sequences $\boldsymbol{x}$ for which $\hat{P}_{1k}(\boldsymbol{x})$ is defined, and can be represented as follows:

$$\begin{aligned}
E[P_{1k}] &= E\left[\ \hat{P}_{1k}(\boldsymbol{x})\ \middle|\ |I_{1k}(\boldsymbol{x})| > 0\ \right] \\
&= C \sum_{n_1=k}^{n} \sum_{\boldsymbol{x} \in D(n_1,k)} p^{n_1}(1-p)^{n-n_1} \hat{P}_{1k}(\boldsymbol{x}) \\
&= C \sum_{n_1=k}^{n} |D(n_1,k)| p^{n_1}(1-p)^{n-n_1} E\left[\ \hat{P}_{1k}(\boldsymbol{x})\ \middle| |I_{1k}(\boldsymbol{x})| > 0,\ N_1(\boldsymbol{x}) = n_1\ \right] \\
&= C \sum_{n_1=k}^{n} |D(n_1,k)| p^{n_1}(1-p)^{n-n_1} E[P_{1k}|N_1 = n_1] \qquad (1)
\end{aligned}$$

where $D(n_1, k)$ is the set of sequences for which $\hat{P}_{1k}(\boldsymbol{x})$ is defined and C is the constant which normalizes the total probability to 1.[7] The distribution of $P_{1k}|N_1$ derived from $\boldsymbol{x}$ has support $\{p_{1k} \in [0,1] : p_{1k} = \hat{P}_{1k}(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in \{0,1\}^n,\ |I_{1k}(\boldsymbol{x})| > 0, \text{ and } N_1(\boldsymbol{x}) = n_1\}$ for all $k \geq 1$ and $n_1 \geq 1$. In Appendix B we determine this distribution for all $k > 1$. Here we consider the case of $k = 1$, and compute the expected value directly. First we establish the main lemma.

Lemma 3 *For $n > 1$ and $n_1 = 1, \dots, n$*

$$E[P_{11}|N_1 = n_1] = \frac{n_1 - 1}{n - 1} \tag{2}$$

Proof: See Appendix A.

The quantity $E\left[\hat{P}_{11}(\boldsymbol{x}) \,\middle|\, |I_{11}(\boldsymbol{x})| > 0,\ N_1(\boldsymbol{x}) = n_1\right]$ can, in principle, be computed directly by calculating $\hat{P}_{11}(\boldsymbol{x})$ for each sequence of length n_1, and then averaging across these sequences, but the number of sequences in $D(n_1, 1)$ is typically too large to perform the complete enumeration needed.[8] The key to the proof is to reduce the dimensionality of the problem by identifying the set of sequences over which $\hat{P}_{11}(\boldsymbol{x})$ is constant (the same approach is taken when working with the conditional expectation of $\hat{P}_{1k}(\boldsymbol{x})$ for $k > 1$ in Appendix B). In order to do this, we introduce the concept of a run: let $R_1(\boldsymbol{x})$ be the number of runs of ones, i.e. the number of subsequences of consecutive ones in sequence $\boldsymbol{x}$ which are flanked by zeros or an end point.[9] For all sequences with $R_1(\boldsymbol{x}) = r_1$, $\hat{P}_{11}(\boldsymbol{x})$ is (i) constant and equal to $(n_1 - r_1)/n_1$ across all those sequences which terminate with a zero, and (ii) constant and equal to $(n_1 - r_1)/(n_1 - 1)$ across all those sequences which terminate with a one. The distribution of r_1 in each of these cases can be found by way of combinatorial argument, and the expectation computed directly, yielding Equation 2. This result is quantitatively identical to the formula one would get for the probability of success, conditional on n_1 successes and $n_0 = n - n_1$ failures, if one were to first draw one success from an unordered set of

---

[7] More precisely, $D(n_1, k) := \{\boldsymbol{x} \in \{0,1\}^n : \sum_{i=1}^n x_i = n_1 \text{ and } |I_{1k}(\boldsymbol{x})| > 0\}$ and $C := 1/\sum_{n_1=k}^n |D(n_1,k)| p^{n_1}(1-p)^{n-n_1}$. Note that for $n_1 < k$ we have $D(n_1, k) = \emptyset$, and for $n_1 = k$, we have $\hat{P}_{1k}(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in D(n_1, k)$.

[8] For example, with $n = 100$ and $n_1 = 50$, there are $D(50, 1) = 100!/(50!50!) > 10^{29}$ distinguishable sequences, which is greater than the nearly 10^{24} microseconds since the "big bang."

[9] The number of runs of ones can be defined explicitly to be the number of trials in which a one occurs and is immediately followed by a zero on the next trial or has no following trial, i.e. $R_1(\boldsymbol{x}) := |\{i \in \{1, \dots, n\} : x_i = 1 \text{ and, if } i < n, \text{ then } x_{i+1} = 0\}|$

n trials with n_1 successes, and then draw one of the remaining $n-1$ trials without replacement.[10,11]

Given (2), Equation 1 can now be simplified, with $E[P_{11}]$ expressed in terms of only n and p

Theorem 4 *For $p > 0$*

$$E[P_{11}] = \frac{n}{(n-1)[1-(1-p)^n - p(1-p)^{n-1}]}\left(p - \frac{1-(1-p)^n}{n}\right) \tag{3}$$

Proof: See Appendix A.

In the following subsection we plot $E[P_{11}]$ as a function of n, for several values of p, and we also plot $E[P_{1k}]$ for $k > 1$.

In Section 3.2 we compare the relative frequency of successes for k/1-streak successor trials (repetitions) to the relative frequency of successes for k/0-streak successor trials (alternations), which requires explicit consideration of the expected difference between these conditional relative frequencies. We now consider the case in which $k = 1$ (for $k > 1$ see Appendix B), and find the difference to be independent of p. Before stating the theorem, we first define P_{0k} to be the relative frequency of a 0 for k/0-streak successor trials, i.e.

$$\hat{P}_{0k}(\boldsymbol{x}) := \frac{\sum_{i \in I_{0k}} 1 - x_i}{|I_{0k}(\boldsymbol{x})|}$$

where P_{0k} is the derived distribution. We find that for $k = 1$, the expected difference between the relative frequency of successes for k/1-streak successor trials, and the relative frequency of successes for k/0-streak successor trials, $D_1 := P_{11} - (1 - P_{01})$, depends only on n.[12]

Theorem 5 *Letting $D_1 := P_{11} - (1 - P_{01})$, then for any $0 < p < 1$, we have:*

$$E[D_1] = -\frac{1}{n-1}$$

[10]For $k > 1$ this identity does not hold. Instead, it appears to be the case that $E[P_{1k}|N_1 = n_1] < (n_1 - k)/(n - k)$ for all $k > 1$ and $k < n_1 < n$. The intuition why is similar to that given in the discussion of Table 1, but in this case at the level of sequences with fixed $N_1 = n_1$, rather than across n_1. That is, those sequences with $N_1 = n_1$ which have fewer runs of successes, and thus more streak successors and higher repetition rates, are relatively underweighted in the (unweighted) average across all sequences in which $N_1 = n_1$.

[11]Note that for any trial i within the set of sequences satisfying $N_1(\boldsymbol{x}) = n_1$, $E[x_i|x_{i-1} = \cdot x_{i-k} = 1, N_1(\boldsymbol{x}) = n_1] = (n_1 - k)/(n - k)$.

[12]This independence from p does not extend to the case of $k > 1$, see next section and appendix.

Proof: See Appendix A

2.2 The degree of bias as p, n, and k vary

In Section 1, for sequences of trials of length $n = 4$ and trials with success probability $p = .5$, Table 1 reported the expected conditional relative frequency of success (heads) on those trials that immediately follow a success, the expected conditional relative frequency of success (heads) on those trials that immediately follow a failure (tails), and the expected difference between these relative frequencies. This subsection illustrates how these quantities depend on p, n and k, more generally.

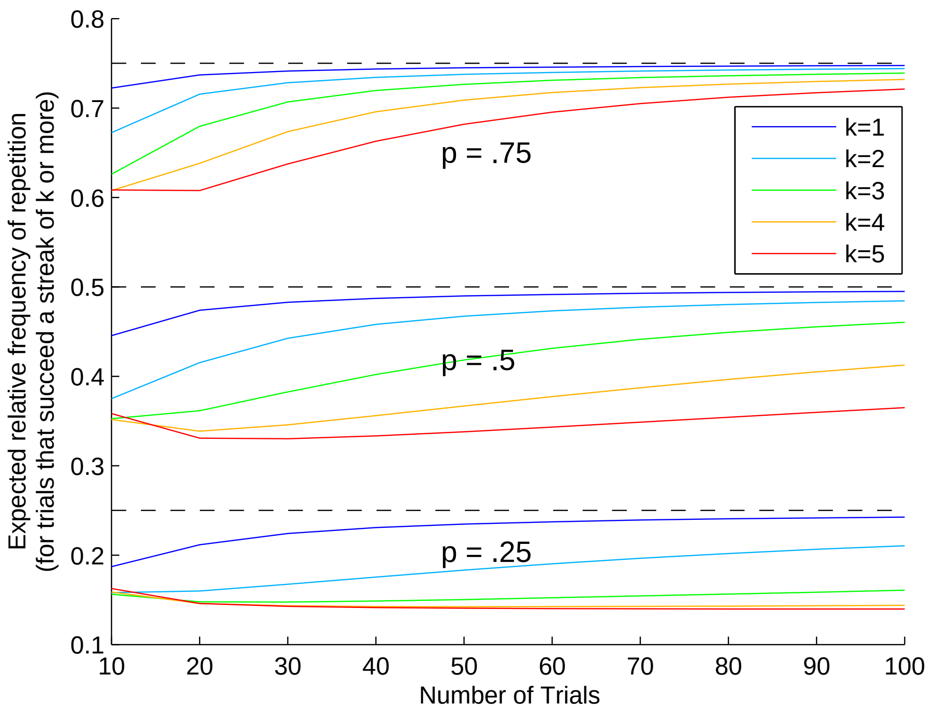


Figure 1: *The expected conditional relative frequency of a 1 on k/1 streak successor trials, as a function of n, for different values of k and p.*

Figure 1, which was produced by combining Equation 1 with the explicit formula provided in Theorem 7 (Appendix B), shows how the expected conditional relative frequency of successes, on those trials that immediately follow k (or more) successes, changes as n, p, and k vary. Each dotted line represents the true conditional (and unconditional) probability of success for $p = 0.25, 0.50,$

and 0.75, with the five solid lines immediately below each dotted line representing the respective expected conditional relative frequencies. Whereas intuition suggests that these expected conditional relative frequencies should be equal to p, one can see that they are strictly less than the true conditional probability p in all cases. One can also see that as n gets larger, the difference between expected conditional relative frequencies and respective probabilities of success generally decrease. Nevertheless, these differences can be substantial, even for long sequences, as can be seen, for example, in the case of $n = 100$, $p = 0.5$, and $k = 5$, in which the absolute difference is $0.50 - 0.35 = 0.15$, or in the case of $n = 100$, $p = 0.25$, and $k = 3$, in which the absolute difference is $0.25 - 0.16 = 0.09$.[13]

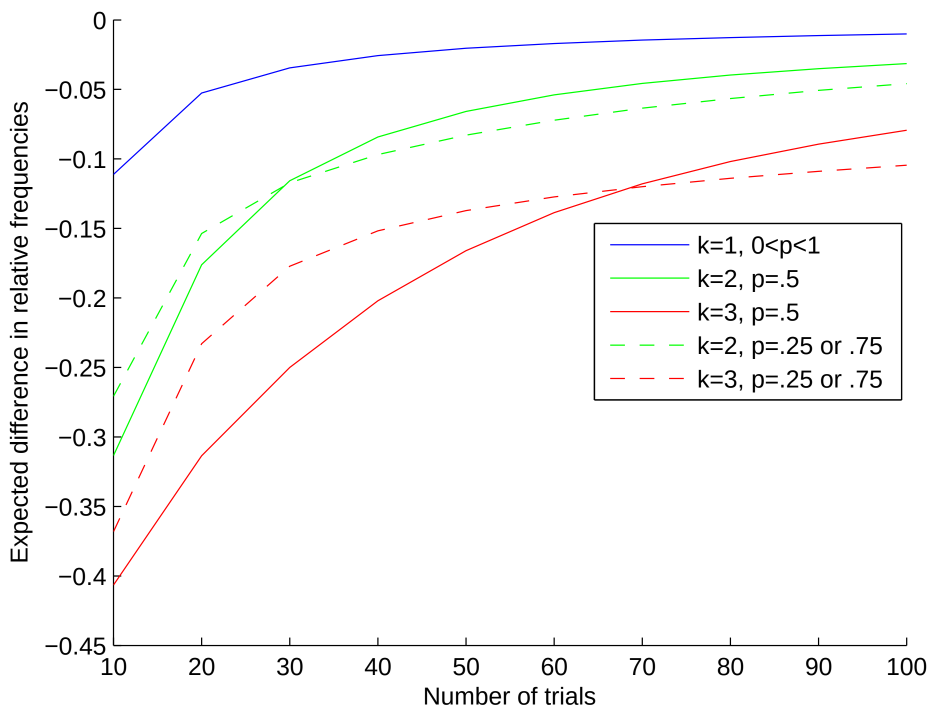


Figure 2: *The expected difference between the conditional relative frequency of a 1 on k/1 streak successor trials and the conditional relative frequency of a 1 on k/0 streak successor trials, as a function of n, for different values of k and p.*

Figure 2 illustrates the expected difference between the conditional relative frequency of a success on those trials that immediately follow a streak of k or more successes (k/1-streak successor

[13] The fact that there is an initial decrease only for the larger values of k is an artifact of the x-axis beginning at 10; though not observable here, there is also an initial decrease for $k = 1$, between $n = 3$ and $n = 4$.

trials), and the conditional relative frequency of a success on those trials that immediately follow a streak of k or more failures (k/0-streak successor trails), as k goes from 1 to 3, and p takes the values of 0.25, 0.50, and 0.75. Whereas intuition suggests that these differences should be zero, one can see that they are strictly negative in all cases, and can be substantial even for long sequences. Further, the bias revealed by comparing the expected difference to zero is more than twice the bias related to the conditional relative frequency reported in Figure 1.[14]

3 Applications to the Gambler's and Hot Hand Fallacies

Inferring serial dependence of any order from sequential data is an important feature of decision making in a variety of important economic domains, and has been studied in financial markets,[15] sports wagering,[16] casino gambling,[17] and lotteries.[18] The most controlled studies of decision making based on sequential data have occurred in a large body of laboratory experiments (for surveys, see Bar-Hillel and Wagenaar [1991] [1991], Nickerson [2002], Rabin [2002], and Oskarsson et al. [2009]).

First, we explain how the result from Section 2 provides a structural explanation for the persistence of alternation bias and gambler's fallacy beliefs. Then we conduct a simple survey to check whether peoples' experience outside of the laboratory is consistent with the degrees of alternation bias and gambler's fallacy that they exhibit within the laboratory. Lastly, we explain how the result reveals a common error in the statistical analyses of the most prominent hot hand fallacy studies (including the original), which when corrected for, reverses the strongest evidence that belief in the hot hand is a "cognitive illusion." In particular, what had previously been considered nearly conclusive evidence of a *lack of* hot hand performance, was instead strong evidence of hot hand performance all along.

[14] Similarly, for $p = 0.25$ and $p = 0.75$ the differences are larger than double the average difference between the expected conditional relative frequencies and p (seen in Figure 1), when averaging across $p = 0.25$ and $p = 0.75$.

[15] Barberis and Thaler (2003); De Bondt (1993); De Long, Shleifer, Summers, and Waldmann (1991); Kahneman and Riepe (1998); Loh and Warachka (2012); Malkiel (2011); Rabin and Vayanos (2010)

[16] Arkes (2011); Avery and Chevalier (1999); Brown and Sauer (1993); Camerer (1989); Durham, Hertzel, and Martin (2005); Lee and Smith (2002); Paul and Weinbach (2005); Sinkey and Logan (2013)

[17] Croson and Sundali (2005); Narayanan and Manchanda (2012); Smith, Levere, and Kurtzman (2009); Sundali and Croson (2006); Xu and Harvey (2014)

[18] Galbo-Jørgensen, Suetens, and Tyran (2013); Guryan and Kearney (2008); Yuan, Sun, and Siu (2014)

3.1 Alternation Bias and Gambler's Fallacy

> *Why, if the gambler's fallacy is truly fallacious, does it persist? Why is it not corrected as a consequence of experience with random events?* (Nickerson 2002)

A classic result in the literature on the human perception of randomly generated sequential data is that people believe outcomes alternate more than they actually do, e.g. for a fair coin, after observing a flip of a tails, people believe that the next flip is more likely to produce a heads than a tails (Bar-Hillel and Wagenaar 1991; Nickerson 2002; Oskarsson et al. 2009; Rabin 2002).[19] Further, as a streak of identical outcomes increases in length, people also tend to think that the alternation rate on the outcome that follows becomes even larger, which is known as the gambler's fallacy (Bar-Hillel and Wagenaar 1991).

The result presented in Section 2 provides a structural explanation for the persistence of both of these systematic errors in beliefs. Independent of how these beliefs arise, it is unlikely that any amount of exposure to finite sequences of i.i.d. outcomes will make them go away. The reason why is that for any sequence length n, even as the number of sequences observed goes to infinity, the expected conditional alternation rate is strictly larger than the (un)conditional probability of the outcome, and this expected alternation rate grows larger when conditioning on streaks of preceding outcomes of increasing length. These effects can be seen clearly in Figure 1 of Section 2. Thus, experience can, in a sense, train people to have an alternation bias, as well as gambler's fallacy beliefs.

A possible solution to the problem is that rather than observing more sequences of size n, one could instead observe *longer* sequences; as n goes to infinity the difference between the conditional relative frequency of, say, a success, and the underlying conditional probability of a success, disappear. Nevertheless, this possibility seems unlikely to fix the problem, for the following reasons: (1) these differences only go away when n is extremely large relative to the lengths of sequences that people are likely to typically observe (as can be seen in Figure 1 and the survey results below), (2) even if one were to observe sufficiently long sequences of outcomes, it seems unlikely that memory and attention limitations would allow them to consider more than relatively short subsequences at a time (Bar-Hillel and Wagenaar 1991; Cowan 2001; Miller 1956; Nickerson 2002), thus effectively

[19] This *alternation bias* is also sometimes referred to as *negative recency bias*.

converting the long sequence into many shorter sub-sequences in which expected differences will again be relatively large.

Another possibility is that people become able to interpret observed relative frequencies in such a way as to make the difference between the underlying probability and expected conditional relative frequencies disappear. The way of doing this, as explained briefly in Section 1, is to count the number of observations that were used in computing the conditional relative frequency for each sequence, and then to use these as weights in a weighted average of the conditional relative frequencies across sequences. Doing so will make the difference disappear entirely. Nevertheless, it seems unlikely that people will do this, as it requires relatively more effort, mental accounting, and memory capacity, and at the same time does not intuitively seem to yield any benefit relative to the simpler and more natural approach of taking the standard (unweighted) average relative frequency across sequences. A simpler, and equivalent, correction, is that a person could instead pool observations from all sequences and compute the conditional relative frequency in the new "composite" sample. While simple in theory, this seems unlikely to occur in practice due to similar arguments, such as the immense demands on memory and attention that it would require, combined with the apparent suitability of the more natural, and simpler, alternative approach. Thus, one might conclude that because people are effectively only exposed to finite sequences of outcomes, the natural learning environment is "wicked," in the sense that it does not allow people to calibrate to the true conditional probabilities with experience alone (Hogarth 2010).

Another example of how experience may not be helpful in ridding of alternation bias and gambler's fallacy beliefs is that gambling, in games such as roulette, places no pressure on these beliefs to go away. In particular, while people can learn via reinforcement that gambling is not profitable, they cannot learn via reinforcement that it is disadvantageous to believe in excessive alternation, or in streak-effects, as the expected return is the same for all choices (Croson and Sundali 2005; Rabin 2002).

Thus, while it seems unlikely that experience alone will make alternation bias and gambler's fallacy beliefs disappear, studies have shown that people can learn to perceive randomness correctly in experimental environments, when given proper feedback and incentives (Budescu and Rapoport 1994; Lopes and Oden 1987; Neuringer 1986; Rapoport and Budescu 1992). However, to the extent that such conditions are not satisfied in real world settings, and people adapt to the natural statistics in their environment (Atick 1992; Simoncelli and Olshausen 2001), we suspect that the structural limitation to learning, which arises from the fact that sequences are of finite length, may ensure that some degree of alternation bias and gambler's fallacy persist, particularly among amateurs with little incentive to eradicate such beliefs.

It is worth noting that in light of the result presented in Section 2, behavioral models of the belief in the law of small numbers (e.g. Rabin [2002]; Rabin and Vayanos [2010]), in which the subjective probability of alternation exceeds the true probability, and grows as streak lengths increase, not only qualitatively describe behavioral patterns, but also happen to be consistent with the properties of a statistic that is natural to use in environments with sequential data—the relative frequency of successes on those outcomes that immediately follow a salient streak of successes.

Survey

A stylized fact about experimental subjects' perceptions of sequential dependence is that, on average, they believe that random processes, such as a fair coin, alternate at a rate of roughly 0.6, rather than 0.5 (Bar-Hillel and Wagenaar 1991; Nickerson 2002; Oskarsson et al. 2009; Rabin 2002). This expected alternation rate, of course, corresponds precisely with the alternation rate reported for coin flip sequences of length four in Table 1. If it were the case that peoples' experience with finite sequences, either by observation or by generation, tended to involve sequences this short, then this could provide a structural explanation of why the alternation bias and gambler's fallacy persist.

In order to get a sense of what people might expect alternation rates to be, based on their experience with binary outcomes outside of the laboratory, we conduct a simple survey, which is designed to elicit the typical number of sequential outcomes that people observe when repeatedly flipping a coin, as well as their perceptions of expected conditional probabilities, based on recent outcomes. We recruited 649 subjects to participate in a survey in which they could be paid up to 25 Euros to answer the following questions as best as possible[20]: (1) what is the largest number of successive coin flips that they have observed in one sitting (2) what is the average number of

[20]The exact email invitation to the survey was as follows. "This is a special message regarding an online survey which pays up to 25 Euros for 3 minutes of your time. If you complete this survey (**link**) by 02:00 on Friday 05-June Milan time and the Thursday June 4th evening drawing of the California Lottery Pick 3 matches the final 3 digits of your student ID, you will be paid 25 dollars. For details on the Pick 3, see: http://www.calottery.com/play/draw-games/daily-3/winning-numbers for detail."

sequential coin flips that they have observed (3) given that they observe a fair coin land heads one (two; three) consecutive times, what do they feel the chances are that the next flip will be heads (tails).[21,22]

The subjects were recruited from Bocconi University in Milan. All subjects responded to each of the three questions. We observe that the median of the maximum number of sequential coin flips that a subject has seen is 6, and the median of the average number of coin flips is 4. As mentioned previously, given the result presented in Section 2, for sequences of 4 flips of a fair coin, the true expected conditional alternation rate is 0.6 (as illustrated in Table 1 of Section 1), precisely in line with the average magnitude of alternation bias observed in laboratory experiments.[23] This result suggests that experience outside of the laboratory may have a meaningful effect on the behavior observed inside the laboratory.

For the third question, which regarded perceptions of sequential dependence, subjects were randomly assigned into one of two treatments. Subjects in the first treatment were asked about the probability of a head immediately following a streak of heads (repetition), while subjects in the second were asked about the probability of a tail immediately following a streak of heads (alternation). Subjects' responses with respect to streaks of one, two, as well as three heads, are shown in Table 2. One can see that the perceived probability that a streak of a single head will be followed by a head $\mathbb{P}(H|\cdot)$ [tails $\mathbb{P}(T|\cdot)$] is 0.49 [0.50], a streak of two heads 0.45 [0.53], and a streak of three heads 0.44 [0.51]. Thus, subjects' responses are generally directionally consistent with

[21] The subjects were provided a visual introduction to heads and tails for a 1 Euro coin. Then, the two questions pertaining to sequence length were asked in random order: (1) Please think of all the times in which you have observed a coin being flipped, whether it was flipped by you, or by somebody else. To the best that you can recall, what is the maximum number of coin flips that you have ever observed in one sitting?, (2) Please think of all the times in which you have observed a coin being flipped, whether it was flipped by you, or by somebody else. Across all of the times you have observed coin flips, to the best that you can recall, how many times was the coin flipped, on average?

[22] The three questions pertaining to perceived probability were always presented at the end, in order, with each subject assigned either to a treatment in which they were asked for the probability of heads (repetition), or the probability of tails (alternation). The precise language used was: (3) (a) Imagine that you flip a coin you know to be a fair coin, that is, for which heads and tails have an equal chance. First, imagine that you flip the coin one time, and observe a heads. On your second flip, according to your intuition, what do you feel is the chance of flipping a heads (T2: tails) (in percentage terms 0-100)? (b) Second, imagine that you flip the coin two times, and observe two heads in a row. On your third flip, according to your intuition, what do you feel is the chance of flipping a heads (T2: tails) (in percentage terms 0-100)? (c) Third, imagine that you flip the coin three times, and observe three heads in a row. On your fourth flip, according to your intuition, what do feel is the chance of flipping a heads (T2: tails) (in percentage terms 0-100)?

[23] In addition, peoples' maximum working memory capacity has been found to be around four (not seven), in a meta-analysis of the literature in Cowan (2001), which could translate into sequences longer than four essentially being treated as multiple sequences of around four.

true sample conditional alternation and repetition rates, given that in Section 2 it is demonstrated that the conditional probability of heads, when conditioning on streaks of heads, is consistently below 0.5, and is consistently above or equal to 0.5 when the outcome in question is instead a tails. Further, between subjects, perceptions on average satisfy

$$\mathbb{P}(T|H) - \mathbb{P}(H|H),\ \mathbb{P}(T|H) - \mathbb{P}(H|H),\ \mathbb{P}(T|HHH) - \mathbb{P}(H|HHH) > 0$$

and with 649 subjects, all three differences are significant. Thus, subjects' perceptions are consistent with the true positive differences between expected alternation and repetition rates (see Table 1 of Section 1 for an example), when these perceptions are based on the finite sequences of outcomes that a subject has observed in the past. Notice also that average perceived expected conditional alternation and repetition rates lie somewhere between those that subjects have observed in the past, and the true (unconditional) probabilities of alternation and repetition (0.5), and more closely resemble the latter than the former.[24,25]

Table 2: *Average perceived repetition rate $P(H|\cdot)$, and alternation rate $P(T|\cdot)$, when conditioning on streak length of H's, for 649 survey participants*

	Streak Length of H's			
	H	HH	HHH	obs.
$P(H\|\cdot)$	.49	.45	.44	304
$P(T\|\cdot)$	.50	.53	.51	345

3.2 The Hot Hand Fallacy

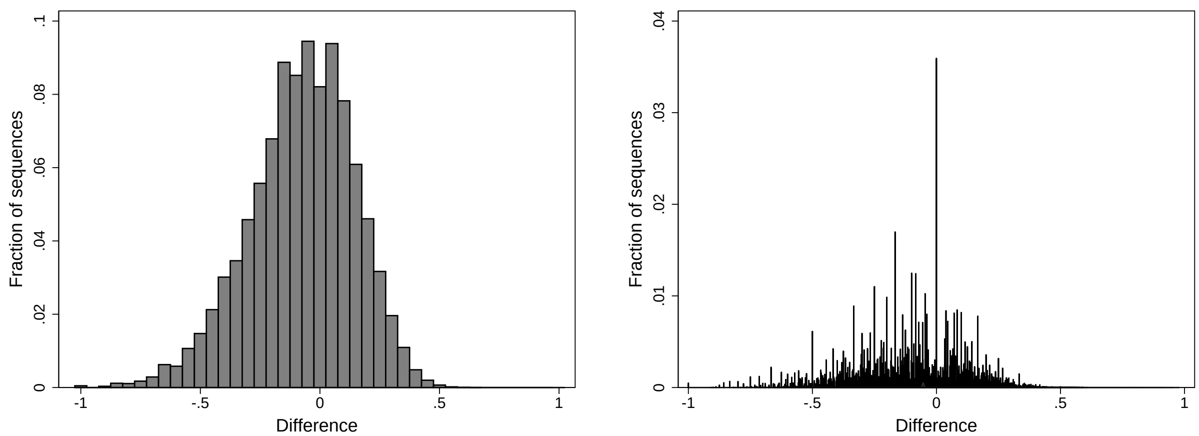
> *This account explains both the formation and maintenance of the erroneous belief in the hot hand: if random sequences are perceived as streak shooting, then no amount of exposure to such sequences will convince the player, the coach, or the fan that the*

[24] Given the fact that subjects were randomly assigned to treatments, with as sample size sufficiently large for significance tests to be robust to adjustments for multiple comparisons, it is a mystery why, for example, on average $\mathbb{P}(H|HHH) < \mathbb{P}(H|HH)$ and $\mathbb{P}(T|HHH) < \mathbb{P}(T|HH)$. A possible explanation is that eliciting the probability of repetition is different than eliciting the probability of alternation.

[25] These findings suggest a possible answer to the following puzzle: "Study of the gambler's fallacy is complicated by the fact that people sometimes make predictions that are consistent with the assumption that they believe that independent events are contingent even when they indicate, when asked, that they believe them to be independent" (Nickerson 2002). Subjects' responses lie between their beliefs based on experience, and their "book knowledge."

Figure 3: *The graph of the (exact) discrete probability distribution of* $E[P_{1k} - (1 - P_{0k})|N_1 = n_1]$, *the difference between the conditional relative frequency of a success on those trials that immediately follow a streak of* 3 *or more successes (3/1-streak successor trials), and the conditional relative frequency of a success on those trials that immediately follow a streak of* 3 *or more failures (3/0-streak successor trails), with* $n = 100$, $n_1 = 50$.

(a) *Bin width = 0.05* **(b)** *Bin width = 0.0001*



sequences are in fact random." (Gilovich et al. 1985)

The hot hand fallacy typically refers to the mistaken belief that success tends to follow success (hot hand), when in fact observed successes are consistent with the typical fluctuations of a chance process. The original evidence of the fallacy was provided by the seminal paper of Gilovich et al. (1985), in which the authors found that basketball players' beliefs that a player has "a better chance of making a shot after having just made his last two or three shots than he does after having just missed his last two or three shots" were not supported by the analysis of shooting data. Because these players were experts who, despite the evidence, continued to make high-stakes decisions based on their mistaken beliefs, the hot hand fallacy came to be characterized as a "massive and widespread cognitive illusion" (Kahneman 2011).[26] The strength of the original results has had a large influence on empirical and theoretical work in all areas related to decision making with sequential data, both in economics and psychology.

In the Gilovich et al. (1985) study, and the other studies in basketball which follow, player performance records – patterns in hits (successes) and misses (failures) – are checked to see if they

[26] The fallacy view has been the consensus in the economics literature on sequential decision making, and the existence of the fallacy itself has been highlighted by prominent main stream economists as important example of how statistical analysis can reveal the flaws of expert intuition (e.g. see (Davidson 2013)).

"differ from sequences of heads and tails produced by [weighted] coin tosses"(Gilovich et al. 1985). The standard measure of hot hand effect size in these studies is to compare the conditional relative frequency of a hit on those shots that immediately follow a streak of hits to the conditional relative frequency of a hit on those shots that immediately follow a streak of misses (Avugos, Bar-Eli, Ritov, and Sher 2013; Gilovich et al. 1985; Koehler and Conley 2003). This comparison appears to be a sound one: if a player is consistent, i.e. always shoots with the same probability of a hit, then the fact that a given shot attempt is taken immediately following a streak of hits, or misses, is determined by chance. Therefore, it appears that these two sets of shot attempts can be regarded as two treatments that are statistically independent. However, what this reasoning ignores is that within any given sequence of finite length, when conditioning on a streak of three (or more) hits, the first (up to) three hits within the subsequences MHM, MHHM, and MHHH cannot immediately follow a streak of hits, and therefore these shot attempts are excluded from the "treatment." Thus, by conditioning on a streak of hits within a sequence of finite length, one creates a selection bias towards observing shots that are misses.[27] An analogous argument applies to conditioning on streaks of misses, but in this case the selection bias is towards observing shots that are hits.

The seemingly correct, but mistaken, assumption that in a sequence of coin tosses, the relative frequency of heads on those flips that immediately follow a streak of heads is expected to be equal to the relative frequency of heads on those flips that immediately follow a streak of tails, has striking effects on the interpretation of performance data. As was shown in Figure 2 of Section 2.2, the bias in this comparison between conditional relative frequencies is more than double that of the comparison of either conditional relative frequency to the true probability (under the Bernoulli assumption). If players were to shoot with a fixed hit rate (the null Bernoulli assumption), then, given the parameters of the original study, one should in fact expect the difference in these relative frequencies to be -0.08, rather than 0. Moreover, the distribution of the differences will have a pronounced negative skew. In Figure 3 we present the *exact* distribution of the difference between these conditional relative frequencies (for two different bin sizes), following Theorem 8 of Appendix B. The distribution is generated subject to the target parameters of the original study: sequences of length $n = 100$, $n_1 = 50$ hits, and streaks of length $k = 3$ or more. As can be seen, the distribution

[27] This selection bias, for streaks of length one, or greater, was mentioned in Section 1 for the special case of streaks of length one.

Table 3: *Columns 4-5 reproduce columns 2 and 8 of Table 4 from Gilovich et al. (1985). Column 6 reports the difference between the report relative frequencies, and column 7 adjusts for the bias (mean correction), based on the player's field goal percentage (probability in this case) and number of shots.*

					$\hat{p}(hit\|3\ hits) - \hat{p}(hit\|3\ misses)$	
Player	# shots	fg%	$\hat{p}(hit\|3\ hits)$	$\hat{p}(hit\|3\ misses)$	GVT est.	bias adj.
Males						
1	100	.54	.50	.44	.06	.14
2	100	.35	.00	.43	−.43	−.33
3	100	.60	.60	.67	−.07	.02
4	90	.40	.33	.47	−.13	−.03
5	100	.42	.33	.75	−.42	−.33
6	100	.57	.65	.25	.40	.48
7	75	.56	.65	.29	.36	.47
8	50	.50	.57	.50	.07	.24
9	100	.54	.83	.35	.48	.56
10	100	.60	.57	.57	.00	.09
11	100	.58	.62	.57	.05	.14
12	100	.44	.43	.41	.02	.10
13	100	.61	.50	.40	.10	.19
14	100	.59	.60	.50	.10	.19
Females						
1	100	.48	.33	.67	−.33	−.25
2	100	.34	.40	.43	−.03	.07
3	100	.39	.50	.36	.14	.23
4	100	.32	.33	.27	.07	.17
5	100	.36	.20	.22	−.02	.08
6	100	.46	.29	.55	−.26	−.18
7	100	.41	.62	.32	.30	.39
8	100	.53	.73	.67	.07	.15
9	100	.45	.50	.46	.04	.12
10	100	.46	.71	.32	.40	.48
11	100	.53	.38	.50	−.12	−.04
12	100	.25	.	.32	.	.
Average		.47	.49	.45	.03	.13

has a pronounced negative skew, with 63 percent of observations less than 0 (median = -.06).[28]

The effects of this bias can be seen in Table 3, which reproduces data from Table 4 of Gilovich et al. (1985). The table presents shooting performance records for each of the 14 male and 12 female Cornell University basketball players who participated in the controlled shooting experiment of the original hot hand study (Gilovich et al. 1985). One can see the number of shots taken, overall field goal percentage, relative frequency of a hit on those shots that immediately follow a streak of three (or more) hits, $\hat{p}(hit|3\ hits)$, relative frequency of a hit on those shots that immediately follow a streak of three (or more) misses, $\hat{p}(hit|3\ misses)$, and the expected difference between these quantities. Under the incorrect assumption that these conditional relative frequencies are expected to be equal to each player's overall field goal percentage, it indeed appears as if there is little to no evidence of hot hand shooting. Likewise, under the incorrect assumption that the difference between these two conditional relative frequencies is expected to be zero, the difference of +0.03 may seem like directional evidence of a slight hot hand, but it is not statistically significant.[29] The last column corrects for the incorrect assumption by subtracting from each player's observed difference the actual expected difference under the null hypothesis that the player shoots with a constant probability of a hit, given his/her overall field goal percentage. Once this correction is made, one can see that 19 of the 25 players directionally exhibit hot hand shooting, and that the average difference in shooting percentage when on a streak of hits vs. misses is +13 percentage points.[30] This is a substantial effect size, given that all shooters are included in the average, and that in the 2013-2014 NBA season, the difference between the very best three point shooter in the league, and the median shooter, was +10 percentage points.

Miller and Sanjurjo (2014) introduce a statistical approach that corrects for this downward bias, and then analyze data from all extant controlled shooting studies. They find substantial evidence of hot hand shooting in all studies. Miller and Sanjurjo (2015a) apply the same corrected statistical approach to 29 years of NBA Three-Point Contest data, and again find substantial evidence of hot

[28] The distribution is displayed using 6 decimal digits of precision. For this precision, the more than 10^{29} distinguishable sequences take on 19,048 distinct values when calculating the difference in relative frequencies. In the computation of the expected value in Figures 1 and 2, each difference is represented with the highest floating point precision possible.

[29] A common criticism of the original study, as well as subsequent studies, is that they are under-powered, thus even substantial differences are not registered as significant (see Miller and Sanjurjo (2014) for a power analysis, and the complete discussion of references).

[30] The effective adjustment after pooling (0.09) is higher than the expected bias for a $p = .5$ shooter due to the different shooting percentages.

hand shooting, in stark contrast with Koehler and Conley (2003), which has been considered a particularly clean replication of the original study, but which happened to use the same incorrect measure of hot hand effect size as Gilovich et al. (1985).[31]

4 Conclusion

We prove that in a finite sequence generated by repeated trials of a Bernoulli random variable the expected conditional relative frequency of successes, on those realizations that immediately follow a streak of successes, is *strictly less than* the fixed probability of success. One direct corollary of this result is a structural explanation for the persistence of alternation bias and gambler's fallacy beliefs. Another is that empirical approaches of the most prominent studies in the hot hand fallacy literature are incorrect. Once corrected for, the data that was previously interpreted as providing substantial evidence that the belief in the hot hand is fallacy, reverses, and becomes substantial evidence that it is not a fallacy to believe in the hot hand. Finally, the respective errors of the gambler and the hot hand fallacy researcher are found to be analogous.

References

Arkes, J. (2011): "Do Gamblers Correctly Price Momentum in NBA Betting Markets?" *Journal of Prediction Markets*, 5, 31–50.

Atick, J. J. (1992): "Could information theory provide an ecological theory of sensory processing?" *Network: Computation in neural systems*, 3, 213–251.

Avery, C. and J. Chevalier (1999): "Identifying Investor Sentiment from Price Paths: The Case of Football Betting," *Journal of Business*, 72, 493–521.

Avugos, S., M. Bar-Eli, I. Ritov, and E. Sher (2013): "The elusive reality of efficacy performance cycles in basketball shooting: analysis of players performance under invariant conditions," *International Journal of Sport and Exercise Psychology*, 11, 184–202.

Bar-Hillel, M. and W. A. Wagenaar (1991): "The perception of randomness," *Advances in Applied Mathematics*, 12, 428–454.

Barberis, N. and R. Thaler (2003): "A survey of behavioral finance," *Handbook of the Economics of Finance*, 1, 1053–1128.

[31] Koehler and Conley (2003) studies 4 years of NBA Three-Point contest data.

Benjamin, D. J., M. Rabin, and C. Raymond (2014): "A Model of Non-Belief in the Law of Large Numbers," Working Paper.

Brown, W. A. and R. D. Sauer (1993): "Does the Basketball Market Believe in the Hot Hand? Comment," *American Economic Review*, 83, 1377–1386.

Budescu, D. V. and A. Rapoport (1994): "Subjective randomization in one-and two-person games," *Journal of Behavioral Decision Making*, 7, 261–278.

Camerer, C. F. (1989): "Does the Basketball Market Believe in the 'Hot Hand,'?" *American Economic Review*, 79, 1257–1261.

Cowan, N. (2001): "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behavioral and Brain Sciences*, 24, 87–114.

Croson, R. and J. Sundali (2005): "The Gamblers Fallacy and the Hot Hand: Empirical Data from Casinos," *Journal of Risk and Uncertainty*, 30, 195–209.

Davidson, A. (2013): "Boom, Bust or What?" New York Times.

De Bondt, W. P. (1993): "Betting on trends: Intuitive forecasts of financial risk and return," *International Journal of Forecasting*, 9, 355–371.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1991): "The Survival of Noise Traders In Financial-markets," *Journal of Business*, 64, 1–19.

Durham, G. R., M. G. Hertzel, and J. S. Martin (2005): "The Market Impact of Trends and Sequences in Performance: New Evidence," *Journal of Finance*, 60, 2551–2569.

Feller, W. (1968): *An Introduction to Probability Theory and Its Applications*, New York: John Wiley & Sons.

Galbo-Jørgensen, C. B., S. Suetens, and J.-R. Tyran (2013): "Predicting Lotto Numbers A natural experiment on the gamblers fallacy and the hot hand fallacy," Working Paper.

Gibbons, J. D. and S. Chakraborti (2010): *Nonparametric Statistical Inference*, New York: CRC Press, Boca Raton, Florida.

Gilovich, T., R. Vallone, and A. Tversky (1985): "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 17, 295–314.

Guryan, J. and M. S. Kearney (2008): "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales," *American Economic Review*, 98, 458–473.

Hogarth, R. M. (2010): "Intuition: A Challenge for Psychological Research on Decision Making," *Psychological Inquiry*, 21, 338–353.

Kahneman, D. (2011): *Thinking, Fast and Slow*, Farrar, Straus and Giroux.

Kahneman, D. and M. W. Riepe (1998): "Aspects of Investor Psychology: Beliefs, preferences, and biases investment advisors should know about," *Journal of Portfolio Management*, 24, 1–21.

Kahneman, D. and A. Tversky (1972): "Subjective Probability: A Judgement of Representativeness," *Cognitive Psychology*, 3, 430–454.

Koehler, J. J. and C. A. Conley (2003): "The "hot hand" myth in professional basketball," *Journal of Sport and Exercise Psychology*, 25, 253–259.

Lee, M. and G. Smith (2002): "Regression to the mean and football wagers," *Journal of Behavioral Decision Making*, 15, 329–342.

Loh, R. K. and M. Warachka (2012): "Streaks in Earnings Surprises and the Cross-Section of Stock Returns," *Management Science*, 58, 1305–1321.

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

Malkiel, B. G. (2011): *A random walk down Wall Street: the time-tested strategy for sucessful investing*, New York: W. W. Norton & Company.

Miller, G. A. (1956): "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review*, 63, 81–97.

Miller, J. B. and A. Sanjurjo (2014): "A Cold Shower for the Hot Hand Fallacy," Working Paper.

——— (2015a): "Is the Belief in the Hot Hand a *Fallacy* in the NBA Three Point Shootout?" Working Paper.

——— (2015b): "Surprised by the Gambler's and Hot Hand Fallacies? A Proof of the Law of Small Numbers with Implications For Inferring Serial Dependence in Finite Sequential Data." Working Paper.

Narayanan, S. and P. Manchanda (2012): "An empirical analysis of individual level casino gambling behavior," 10, 27–62.

Nelson, C. R. and M. J. Kim (1993): "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48, 641–661.

Neuringer, A. (1986): "Can people behave "randomly?": The role of feedback." *Journal of Experimental Psychology: General*, 115, 62.

Nickerson, R. S. (2002): "The production and perception of randomness," *Psychological Review*, 109, 350–357.

Oskarsson, A. T., L. V. Boven, G. H. McClelland, and R. Hastie (2009): "Whats next? Judging sequences of binary events," *Psychological Bulletin*, 135, 262–385.

Paul, R. J. and A. P. Weinbach (2005): "Bettor Misperceptions in the NBA: The Overbetting of Large Favorites and the 'Hot Hand'," *Journal of Sports Economics*, 6, 390–400.

Rabin, M. (2002): "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, 117, 775–816.

Rabin, M. and D. Vayanos (2010): "The Gamblers and Hot-Hand Fallacies: Theory and Applications," *Review of Economic Studies*, 77, 730–778.

Rapoport, A. and D. V. Budescu (1992): "Generation of random series in two-person strictly competitive games." *Journal of Experimental Psychology: General*, 121, 352–.

Riordan, J. (1958): *An Introduction to Combinatorial Analysis*, New York: John Wiley & Sons.

Simoncelli, E. P. and B. A. Olshausen (2001): "Natural Image Statistics And Neural Representation," *Annual Review of Neuroscience*, 24, 1193–1216.

Sinkey, M. and T. Logan (2013): "Does the Hot Hand Drive the Market?" *Eastern Economic Journal*, Advance online publication, doi:10.1057/eej.2013.33.

Smith, G., M. Levere, and R. Kurtzman (2009): "Poker Player Behavior After Big Wins and Big Losses," *Management Science*, 55, 1547–1555.

Stambaugh, R. F. (1986): "Bias in Regressions with lagged Stochastic Regressors," Working Paper No. 156.

——— (1999): "Predictive regressions," *Journal of Financial Economics*, 54, 375–421.

Sundali, J. and R. Croson (2006): "Biases in casino betting: The hot and the gamblers fallacy," *Judgement and Decision Making*, 1, 1–12.

Xu, J. and N. Harvey (2014): "Carry on winning: The gambler's fallacy creates hot hand effects in online gambling," *Cognition*, 131, 173 – 180.

Yuan, J., G.-Z. Sun, and R. Siu (2014): "The Lure of Illusory Luck: How Much Are People Willing to Pay for Random Shocks," *Journal of Economic Behavior & Organization*, forthcoming.

Yule, G. U. (1926): "Why do we Sometimes get Nonsense-Correlations between Time-Series?–A Study in Sampling and the Nature of Time-Series," *Journal of the Royal Statistical Society*, 89, 1–63.

A Appendix: Section 2 Proofs

Proof of Lemma 3

For $n_1 = 1$, clearly $\hat{P}_{11}(\boldsymbol{x}) = 0$ for all $\boldsymbol{x}$, and the identity is satisfied. For $n_1 > 1$ this quantity cannot be computed directly by calculating its value for each sequence because the number of sequences in $D(n_1, 1)$ is typically too large.[32] In order to handle the case of $n_1 > 1$, we first define $R_1(\boldsymbol{x})$ as the number of runs of ones, i.e. the number of subsequences of consecutive ones in sequence $\boldsymbol{x}$ which are flanked by zeros or an end point.[33] The key observation is that for all sequences with $R_1(\boldsymbol{x}) = r_1$, $\hat{P}_{11}(\boldsymbol{x})$ is (i) constant and equal to $(n_1 - r_1)/n_1$ across all those sequences which terminate with a zero, and (ii) constant and equal to $(n_1 - r_1)/(n_1 - 1)$ across all those sequences which terminate with a one. The number of sequences in each of these cases can be counted using a combinatorial argument.

Any sequence with r_1 ones can be constructed, first, by building the runs of ones of fixed length with an ordered partition of the n_1 ones into r_1 cells (runs) which can be performed in $\binom{n_1-1}{r_1-1}$ ways by inserting $r_1 - 1$ dividers into the $n_1 - 1$ available positions between ones, and second, by placing the r_1 ones into the available positions to the left or the right of a zero among the n_0 zeros to form the final sequence. For the case in which $x_n = 0$ there are n_0 available positions to place the runs, and therefore $\binom{n_0}{r_1}$ possible placements, while in the case in which $x_n = 1$ (which must end in a run of ones) there are n_0 available positions to place the $r_1 - 1$ remaining runs, and therefore $\binom{n_0}{r_1-1}$ possible placements. Note that for $n_1 > 1$, we have $D(n_1, 1) = \{\boldsymbol{x} \in \{0,1\}^n : N_1(\boldsymbol{x}) = n_1\}$ and $|D(n_1, 1)| = \binom{n}{n_1}$, and $r_1 \leq n_1$, therefore

$$E[P_{11}|N_1 = n_1] = \frac{1}{\binom{n}{n_1}} \sum_{d=0}^{1} \sum_{r_1=1}^{\min\{n_1, n_0+d\}} \binom{n_1 - 1}{r_1 - 1} \binom{n_0}{r_1 - d} \frac{n_1 - r_1}{n_1 - d}$$

[32] For example, with $n = 100$, $n_1 = 50$ and $k = 1$ there $100!/(50!50!) > 2^{50} > 10^{15}$ such sequences.

[33] The number of runs of ones can be defined explicitly to be the number of trials in which a one occurs and is immediately followed by a zero on the next trial or has no following trial, i.e. $R_1(\boldsymbol{x}) := |\{i \in \{1, \ldots, n\} : x_i = 1 \text{ and, if } i < n \text{ then } x_{i+1} = 0\}|$

For the case in which $d = 0$, the inner sum satisfies:

$$\begin{aligned}
\sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1}\frac{n_1-r_1}{n_1} &= \sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1}\left(1-\frac{r_1}{n_1}\right) \\
&= \binom{n-1}{n_0-1} - \frac{1}{n_1}\sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1} r_1 \\
&= \binom{n-1}{n_0-1} - \frac{n_0}{n_1}\sum_{r_1=1}^{\min\{n_1,n_0\}} \binom{n_1-1}{r_1-1}\binom{n_0-1}{r_1-1} \\
&= \binom{n-1}{n_0-1} - \frac{n_0}{n_1}\sum_{x=0}^{\min\{n_1-1,n_0-1\}} \binom{n_1-1}{x}\binom{n_0-1}{x} \\
&= \binom{n-1}{n_0-1} - \frac{n_0}{n_1}\binom{n-2}{n_1-1}
\end{aligned}$$

The left term of the second line follows because it is the total number of sequences that can be formed in the first $n-1$ positions with $n_0 - 1$ zeros and $n_1 = n - n_0$ ones. The final line follows from an application of Vandermonde's convolution.[34]

For the case in which $d = 1$, the inner sum can be reduced using similar arguments:

[34] Vandermonde's convolution is given as

$$\sum_{k=\max\{-m,n-s\}}^{\min\{r-m,n\}} \binom{r}{m+k}\binom{s}{n-k} = \binom{r+s}{m+n}$$

from which one can derive the following identity, which we apply

$$\sum_{k=\max\{-m,-n\}}^{\min\{\ell-m,s-n\}} \binom{\ell}{m+k}\binom{s}{n+k} = \sum_{k=\max\{-m,-n\}}^{\min\{\ell-m,s-n\}} \binom{s}{n+k}\binom{\ell}{(\ell-m)-k} = \binom{\ell+s}{\ell-m+n}$$

$$
\begin{aligned}
\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1-1}\frac{n_1-r_1}{n_1-1} &= \frac{n_1}{n_1-1}\binom{n-1}{n_0} - \frac{1}{n_1-1}\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1-1} r_1 \\
&= \binom{n-1}{n_0} - \frac{1}{n_1-1}\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0}{r_1-1}(r_1-1) \\
&= \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\sum_{r_1=2}^{\min\{n_1,n_0+1\}} \binom{n_1-1}{r_1-1}\binom{n_0-1}{r_1-2} \\
&= \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\sum_{x=0}^{\min\{n_1-2,n_0-1\}} \binom{n_1-1}{x+1}\binom{n_0-1}{x} \\
&= \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\binom{n-2}{n_1-2}
\end{aligned}
$$

Combining both cases we have:

$$
\begin{aligned}
E[P_{11}|N_1=n_1] &= \frac{1}{\binom{n}{n_1}}\left[\binom{n-1}{n_0-1} - \frac{n_0}{n_1}\binom{n-2}{n_1-1} + \binom{n-1}{n_0} - \frac{n_0}{n_1-1}\binom{n-2}{n_1-2}\right] \\
&= \frac{1}{\binom{n}{n_1}}\left[\binom{n}{n_0} - \frac{n_0}{n-1}\binom{n}{n_1}\right] \\
&= \frac{n_1-1}{n-1}
\end{aligned}
$$

■

Proof of Lemma 4

The full derivation of Equation 3 is as follows.

$$
\begin{aligned}
E[P_{11}] &= C\sum_{n_1=1}^{n} |D(n_1,1)| p^{n_1}(1-p)^{n-n_1} E[P_{11}|N_1=n_1] \\
&= \frac{\sum_{n_1=2}^{n}\binom{n}{n_1}p^{n_1}(1-p)^{n-n_1}\frac{n_1-1}{n-1}}{1-(1-p)^n-p(1-p)^{n-1}} \\
&= \frac{\frac{1}{n-1}\left[\left(np-np(1-p)^{n-1}\right)-\left(1-(1-p)^n-np(1-p)^{n-1}\right)\right]}{1-(1-p)^n-p(1-p)^{n-1}} \\
&= \frac{n}{(n-1)[1-(1-p)^n-p(1-p)^{n-1}]}\left(p-\frac{1-(1-p)^n}{n}\right)
\end{aligned}
$$

where the second line follows because $|D(n_1,k)| = \binom{n}{n_1}$ for $n_1 > 1$ and $C = 1/[1-(1-p)^n - p(1-p)^{n-1}]$. Clearly $E[P_{11}] < p$.

■

Proof of Lemma 5

First we show that for $n_1 = 1, \ldots, n-1$

$$E[D_1|N_1 = n_1] = -\frac{1}{n-1}$$

If $n_1 = 1$, then for each sequence for which D_1 is defined, $D_1 = 0 - 1/(n-1) = -1/(n-1)$, and therefore the average is $-1/(n-1)$. If $n_1 = n-1$ then for each sequence for which D_1 is defined, $D_1 = (n-2)/(n-1) - 1 = -1/(n-1)$, and therefore the average is $-1/(n-1)$. For n_1 with $1 < n_1 < n-1$, $D_1 = P_{11} - (1 - P_{01})$ is defined for all sequences and therefore the linearity of the expectation, and Lemma 3 yields

$$\begin{aligned} E[D_1|N_1 = n_1] =& E[P_{11}|N_1 = n_1] - E[(1 - P_{01}|N_1 = n_1] \\ =& \frac{n_1 - 1}{n-1} - \left(1 - \frac{n_0 - 1}{n-1}\right) \\ =& -\frac{1}{n-1} \end{aligned}$$

The fact that the conditional expectation is independent of N_1 implies that $E[D_1]$ is independent of p, and we have the result.

■

B Appendix: Quantifying the Bias for $k > 1$

In this section, for $k > 1$, we obtain the expected relative frequency of a 1 for k/1-streak successor trials, $E[P_{1k}]$, and the expected difference in the relative frequency of 1, between k/1-streak successor trials and k/0-streak successor trials, $E[P_{1k} - (1 - P_{0k})]$. Similar to what was done in in the proof of the $k = 1$ case, representing the relative frequency in terms of runs allows us to reduce the dimensionality of the problem by identifying the set of sequences over which $\hat{P}_{1k}(\boldsymbol{x})$ is constant. We begin with some basic definitions.

Given the sequence $\boldsymbol{x} = (x_1, \ldots, x_n)$, recall that a run of 1s is a subsequence of consecutive 1s in $\boldsymbol{x}$ that is flanked on each side by a 0 or an endpoint.[35] Define runs of 0s analogously to runs of 1s. Let $R_{1j}(\boldsymbol{x}) = r_{1j}$ be the number of runs of 1s of exactly length j for $j = 1, \ldots, n_1$. Let $R_{0j}(\boldsymbol{x}) = r_{0j}$ be defined similarly. Let $S_{1j}(\boldsymbol{x}) = s_{1k}$ be the number of runs of 1s of length j or more, i.e. $S_{1j}(\boldsymbol{x}) := \sum_{i=j}^{n_1} R_{1j}(\boldsymbol{x})$ for $j = 1, \ldots, n_1$, with $S_{0j}(\boldsymbol{x}) = s_{0j}$ defined similarly. Let $R_1(\boldsymbol{x}) = r_1$, be the number of runs of 1s, i.e. $R_1(\boldsymbol{x}) = S_{11}(\boldsymbol{x})$, and $R_0(\boldsymbol{x}) = r_0$ be the number of runs of 0s. Let $R(\boldsymbol{x}) = r$ be the total number of runs, i.e. $R(\boldsymbol{x}) := R_1(\boldsymbol{x}) + R_0(\boldsymbol{x})$. Further, let the k/1-streak *frequency* statistic $F_{1k}(\boldsymbol{x}) = f_{1k}$ be defined as the number of (overlapping) 1-streaks of length k, i.e. $F_{1k}(\boldsymbol{x}) := \sum_{j=k}^{n_1}(i-k+1)R_{1j}(\boldsymbol{x})$, with $F_{0k}(\boldsymbol{x}) = f_{0k}$ defined analogously. Notice that $f_{1k} = |I_{1k}(\boldsymbol{x})|$ if $\exists i > n-k$ with $x_i = 0$, and $f_{1k} = |I_{1k}| + 1$ otherwise. Also note that $n_1 = f_{11} = \sum_{j=1}^{n_1} j r_{1j}$ and $n_0 = f_{01} = \sum_{j=1}^{n_0} j r_{0j}$.

To illustrate the definitions, consider the sequence of 10 trials 1101100111. The number of 1s is given by $n_1 = 7$. For $j = 1, \ldots, n_1$, the number of runs of 1s of exactly length j for are given by $r_{11} = 0$, $r_{12} = 2$, $r_{13} = 1$ and $r_{1j} = 0$ for $j \geq 4$, the number of runs of 1s of length j or more are given by $s_{11} = 3$, $s_{12} = 3$, $s_{13} = 1$ and $s_{1j} = 0$ for $j \geq 4$. The total number or runs is $r = 5$. The k/1-streak *frequency* statistic satisfies $f_{11} = 7$, $f_{12} = 4$, $f_{13} = 1$, and $f_{1j} = 0$ for $j \geq 4$. Finally, the k/1-streak momentum statistic satisfies $p_{11} = 4/6$, $p_{12} = 1/3$, with p_{1j} undefined for $j \geq 3$.

B.1 Expected Proportion

In this section we obtain the expected value of the k/1-streak momentum statistic $E[P_{1k}]$ for $k > 1$. As in Section 2 we first compute $E[P_{1k}|N_1 = n_1]$. The fact that $E[P_{1k}|N_1 = n_1]$ was shown to be equal to $(n_1 - 1)/(n - 1)$ in Lemma 3 suggests the possibility of a similar formula for $k > 1$, $(n_1 - k)/(n - k)$, in the spirit of sampling without replacement. This formula does not extend to the case of $k > 1$ as can be readily confirmed by setting $k = 2$, $n_1 = 4$, and $n = 5$.[36] As in Section 2, $\hat{P}_{1k}(\boldsymbol{x})$ cannot be determined directly by computing its value for each sequence as the number

[35] More precisely, it is a subsequence with successive indices $j = i_1 + 1, i_1 + 2, \ldots, i_1 + k$, with $i_1 >= 0$ and $i_1 + k <= n$, in which $x_j = 1$ for all j, and (1) either $i_1 = 0$ or if $i_1 > 0$ then $x_{i_1} = 0$, and (2) either $i_1 + k = n$ or if $i_1 + k < n$ then $i_1 + k + 1 = 0$

[36] If $k = 2$ then for $n_1 = 4$ and $n = 5$, then $E[P_{1k}|N_1 = n_1] = (0/1 + 1/1 + 1/2 + 2/2 + 2/3)/5 = 19/30 < 2/3$. It appears to be the case that $E[P_{1k}|N_1 = n_1] < (n_1 - k)/(n - k)$ for all $k > 1$ and $k < n_1 < n$.

of sequences in $D(n_1, k)$ are typically too large.

We observe that the k/1-streak momentum statistic $\hat{P}_{1k}(\boldsymbol{x})$ can be represented as the ratio of the frequency of length $k+1$ 1-streaks in the sequence and the frequency of length k 1-streaks in the sub-sequence which does not include the final term (assuming $F_{1k}(x_1, \ldots, x_{n-1}) > 0$),

$$\hat{P}_{1k}(\boldsymbol{x}) = \frac{F_{1k+1}(x_1, \ldots, x_n)}{F_{1k}(x_1, \ldots, x_{n-1})}$$

and represented in terms of the entire sequence we have two cases:

$$\hat{P}_{1k}(\boldsymbol{x}) = \begin{cases} F_{1k+1}(\boldsymbol{x})/F_{1k}(\boldsymbol{x}) & \text{if } F_{1k}(\boldsymbol{x}) > 0 \text{ and } \exists i > n-k \text{ with } x_i = 0 \\ F_{1k+1}(\boldsymbol{x})/(F_{1k}(\boldsymbol{x}) - 1) & \text{if } F_{1k}(\boldsymbol{x}) > 1 \text{ and } \forall i > n-k, x_i = 1 \end{cases}$$

The first case in the identity above holds because for a sequence that ends in a 0 or a streak of 1s of length $k-1$ or less, the frequency of length k 1-streaks is exactly the number of length k/1-streak successors, i.e. $f_{1k} = |I_{1k}(\boldsymbol{x})|$. The second case holds because for a sequence that ends with a streak of 1s of length k or more, there is no successor for the final term, which is a length k 1-streak. Note that for all other cases $|I_{1k}(\boldsymbol{x})| = 0$ and thus $\hat{P}_{1k}(\boldsymbol{x})$ is undefined.

Taking an approach similar to that in the proof of Lemma 3, we make the key observation that for all sequences with $R_{1j}(\boldsymbol{x}) = r_{1j}$ for $j = 1, \ldots, k-1$ and $S_{1k}(\boldsymbol{x}) = s_{1k}$, the k/1-streak momentum statistic $\hat{P}_{11}(\boldsymbol{x})$ is (i) constant and equal to $(f_{1k} - s_{1k})/f_{1k}$ across all those sequences which have a 0 in the final k positions, and (ii) constant and equal to $(f_{1k} - s_{1k})/(f_{1k} - 1)$ across all those sequences which have a 1 in each of the final k positions. This can be seen because $f_{1k+1} = f_{1k} - s_{1k}$, and $f_{1k} = n_1 - \sum_{j=1}^{k-1} j r_{1j} - (k-1)s_{1k}$. Notice that for each case $\hat{P}_{1k}(\boldsymbol{x}) = G(R_{11}(\boldsymbol{x}), \ldots, R_{1k-1}(\boldsymbol{x}), S_{1k}(\boldsymbol{x}))$ for some G, and therefore finding the joint distribution of $(R_{11}, \ldots, R_{1k-1}, S_{1k})$ conditional on N_1 will allow us to obtain $E[P_{1k}|N_1 = n_1]$. A classic reference for non-parametric statistical theory Gibbons and Chakraborti (2010) contains a theorem (Theorem 3.3.2, p.87) for the joint distribution $(R_{11}, \ldots, R_{1n_1})$ conditional on N_1 and R_1 from which the joint distribution $(R_{11}, \ldots R_{1k-1}, S_{1k})$ can in principle be derived. We note that the distribution in the theorem is in fact not conditional R_1.[37] Instead of employing the distribution in the theorem, for completeness we obtain the joint

[37] For the distribution conditional on r_1 and n_1 it is straightforward to show that

$$\mathbb{P}(R_{11} = r_{11}, \ldots, R_{1n_1} = r_{1n_1} | N_1 = n_1, R_1 = r_1) = \frac{r_1!}{\binom{n_1 - 1}{r_1 - 1} \prod_{j=1}^{n_1} r_{1j}!}$$

distribution $(R_{11}, \dots R_{1k-1}, S_{1k})$ via direct combinatorial proof. With $\binom{n}{n_1}$ sequences $\boldsymbol{x} \in \{0,1\}^n$ that satisfy $N_1(\boldsymbol{x}) = n_1$, the joint distribution of $(R_{11}(\boldsymbol{x}), \dots R_{1k-1}(\boldsymbol{x}), S_{1k}(\boldsymbol{x}))$ is fully characterized by the number of distinguishable sequences $\boldsymbol{x}$ which satisfy $R_{11}(\boldsymbol{x}) = r_{11}, \dots R_{1k-1}(\boldsymbol{x}) = r_{1k-1}$, and $S_{1k}(\boldsymbol{x}) = s_{1k}$, which we obtain in the following lemma

Lemma 6 *The number of distinguishable sequences $\boldsymbol{x} \in \{0,1\}^n$, $n \geq 1$, with $n_1 \leq n$ 1s, $r_{1j} \geq 0$ runs of 1s of exactly length j for $j = 1, \dots k-1$, and $s_{1k} \geq 0$ runs of 1s of length k or more satisfies:*

$$C_{1k} = \frac{r_1!}{s_{1k}! \prod_{j=1}^{k-1} r_{1j}!} \binom{n_0+1}{r_1} \binom{f_{1k}-1}{s_{1k}-1}$$

where $r_1 = \sum_{j=1}^{k-1} r_{1j} + s_{1k}$ and $f_{1k} = n_1 - \sum_{j=1}^{k-1} j r_{1j} - (k-1)s_{1k}$. Further, we define $\binom{n}{k} = n!/k!(n-k)!$ if $n \geq k$ and $n, k \geq 0$ and $\binom{n}{k} = 0$ otherwise, except for the special case $\binom{-1}{-1} = 1$.[38]

Proof:

Any sequence with $r_{11}, \dots, r_{1k-1}$ runs of 1s of fixed length, and s_{1k} runs of 1s of length k or more can be constructed in three steps by (1) selecting a distinguishable permutation of the $r_1 = \sum_{j=1}^{k-1} r_{1j} + s_{1k}$ cells which correspond to the r_1 runs, which can be done in $r_1!/(s_{1k}! \prod_{j=1}^{k-1} r_{1j})$ unique ways, as for each j, the $r_{1j}!$ permutations of the r_{1j} identical cells across their fixed positions do not generate distinguishable sequences (nor for the s_{1k} identical cells), (2) placing the r_1 1s into the available positions to the left or the right of a 0 among the n_0 0s; with n_0+1 available positions, there are $\binom{n_0+1}{r_1}$ ways to do this, (3) filling the "empty" run cells, first by filling the r_{1j} run cells of length j with exactly $j r_{1j}$ 1s for $j < k$, and then the s_{1k} run cells with k or more runs first with $k-1$ 1s each, and then with remaining f_{1k} 1s (the number of 1s which succeed a streak of $k-1$ or more 1s), taking an ordered partition of these 1s into s_{1k} cells (runs) which can be performed in $\binom{f_{1k}-1}{s_{1k}-1}$ ways by inserting $s_{1k}-1$ dividers into the $f_{1k}-1$ available positions between 1s (which guarantees that each s_{1k} cell has at least k 1s).

■

We are now ready for the main theorem which provides the formula for the expected value of the k/1-streak momentum statistic conditional on the number of 1s:

---

[38] Note with this definition of $\binom{n}{k}$, we have $C_{1k} = 0$ if $r_1 > n_0 + 1$, or $\sum_{j=1}^{k-1} j r_{1j} + k s_{1k} > n_1$ (the latter occurs if $s_{1k} > \lfloor \frac{n_1 - \sum_j j r_{1j}}{k} \rfloor$, or $r_{1\ell} > \lfloor \frac{n_1 - \sum_{j \neq \ell} j r_{1j} - k s_{1k}}{\ell} \rfloor$ for some $\ell = 1, \dots, k-1$, where $\lfloor \cdot \rfloor$ is the floor function). Further, because $r_1 > n_1$ implies that latter condition, it also implies $C_{1k} = 0$.

Theorem 7 *For $n > 0$ and n_1, k with $1 < k \leq n_1 \leq n$*

$$E[P_{1k}|N_1 = n_1] = \frac{1}{\binom{n}{n_1} - U_{1k}} \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k \\ s_{1k} \geq 1}} C_{1k} \left[\frac{s_{1k}}{n_0+1} \left(\frac{f_{1k} - s_{1k}}{f_{1k} - 1} \right) + \frac{n_0 + 1 - s_{1k}}{n_0+1} \left(\frac{f_{1k} - s_{1k}}{f_{1k}} \right) \right]$$

where f_{1k} and C_{1k} depend on $n_0, n_1, r_{11}, \ldots, r_{1k-1}$ and s_{1k} and are defined as in Lemma 6.[39] *U_{1k} is defined to be the number of sequences in which the k/1-streak momentum statistic is undefined, and satisfies*

$$\begin{aligned} U_{1k} = & \sum_{r_1=1}^{\min\{n_1, n_0+1\}} \binom{n_0+1}{r_1} \sum_{\ell=0}^{\lfloor \frac{n_1 - r_1}{k-1} \rfloor} (-1)^\ell \binom{r_1}{\ell} \binom{n_1 - 1 - \ell(k-1)}{r_1 - 1} \\ & + \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1 - k + 1, n_0+1\}} \binom{n_0}{r_1 - 1} \sum_{\ell=0}^{\lfloor \frac{n_1 - k - r_1 + 1}{k-1} \rfloor} (-1)^\ell \binom{r_1 - 1}{\ell} \binom{n_1 - k - 1 - \ell(k-1)}{r_1 - 2} \end{aligned}$$

Proof:

For all sequences $\boldsymbol{x} \in \{0,1\}^n$ with n_1 1s, we have three possible cases for how the k/1-streak momentum statistic is determined by r_{1j} $j < k$ and s_{1k}: (1) $\hat{P}_{1k}(\boldsymbol{x})$ is not defined, which arises if (i) $f_{1k} = 0$ or (ii) $f_{1k} = 1$ and $\sum_{i=n-k+1}^n x_i = k$, (2) $\hat{P}_{1k}(\boldsymbol{x})$ is equal to $(f_{1k} - s_{1k})/(f_{1k} - 1)$, which arises if $f_{1k} \geq 2$ and $\sum_{i=n-k+1}^n x_i = k$ or (3) $\hat{P}_{1k}(\boldsymbol{x})$ is equal to $(f_{1k} - s_{1k})/f_{1k}$, which arises if $f_{1k} \geq 1$ and $\sum_{i=n-k+1}^n x_i < k$. In case 1i, with $f_{1k} = 0$, the number of terms, which we denote

[39] Note that $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$ implies $f_{1k} > s_{1k} \geq 1$, which guarantees $f_{1k} \geq 2$.

U^1_{1k}, satisfies:

$$\begin{aligned}
U^1_{1k} &:= \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j}=n_1 \\ s_{1k}=0}} C_{1k} \\
&= \sum_{\substack{r_{11},\ldots,r_{1k-1} \\ \sum_{j=1}^{k-1} jr_{1j}=n_1}} \frac{r_1!}{\prod_{j=1}^{k-1} r_{1j}!} \binom{n_0+1}{r_1} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\substack{r_{11},\ldots,r_{1k-1} \\ \sum_{j=1}^{k-1} jr_{1j}=n_1 \\ \sum_{j=1}^{k-1} r_{1j}=r_1}} \frac{r_1!}{\prod_{j=1}^{k-1} r_{1j}!} \\
&= \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\ell=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^\ell \binom{r_1}{\ell} \binom{n_1-1-\ell(k-1)}{r_1-1}
\end{aligned}$$

where the last line follows by first noting that the inner sum of the third line is the number of compositions (ordered partitions) of $n_1 - k$ into $r_1 - 1$ parts, which has generating function $(x + x^2 + \cdots x^{k-1})^{r_1}$ (Riordan 1958, p. 124). Therefore the inner sum can be generated as the coefficient on x^{n_1} in the multinomial expansion of $(x+x^2+\cdots x^{k-1})^{r_1}$. The inner sum of binomial coefficients in the fourth line corresponds to the coefficient on x^{n_1} in the binomial expansion of of an equivalent representation of the generating function $x^{r_1}(1-x^{k-1})^{r_1}/(1-x)^{r_1} = (x+x^2+\cdots x^{k-1})^{r_1}$. The coefficient in the binomial expansion must agree with the coefficient in the multinomial expansion.[40]

In case 1ii, with $f_{1k} = 1$ and $\sum_{i=n-k+1}^{n} x_i = k$, in which case $\hat{P}_{1k}(\boldsymbol{x})$ is also undefined, all sequences that satisfy this criteria can be constructed by first forming a distinguishable permutation of the $r_1 - 1$ runs of 1s not including the final run of k 1s, which can be done in $r_1!/(\prod_{j=1}^{k-1} r_{1j}!)$ ways, and second placing these runs to the left or the right of the available n_0 0s, not including the right end point, which can be done in $\binom{n_0}{r_1-1}$ ways with the n_0 positions. Summing over all possible

[40] The binomial expansion is given by:

$$x^{r_1}\frac{(1-x^{k-1})^{r_1}}{(1-x)^{r_1}} = x^{r_1}\left[\sum_{t_1=0}^{r_1}\binom{r_1}{t_1}(-1)^{t_1}x^{t_1(k-1)}\right]\cdot\left[\sum_{t_2=0}^{+\infty}\binom{r_1-1+t_2}{r_1-1}x^{t_2}\right]$$

therefore the coefficient on x^{n_1} is $\sum(-1)^{t_1}\binom{r_1}{t_1}\binom{r_1-1+t_2}{r_1-1}$ where the sum is taken over all t_1, t_2 such that $r_1 + t_1(k-1) + t_2 = n_1$.

runs, the number of terms U_{1k}^2 satisfies:

$$
\begin{aligned}
U_{1k}^2 &:= \sum_{\substack{r_{11},\dots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1-k\\ s_{1k}=1}} \frac{s_{1k}}{n_0+1} C_{1k} \\
&= \sum_{\substack{r_{11},\dots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1-k\\ s_{1k}=1}} \frac{(r_1-1)!}{\prod_{j=1}^{k-1} r_{1j}!}\binom{n_0}{r_1-1} \\
&= \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\substack{r_{11},\dots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1-k\\ \sum_{j=1}^{k-1} r_{1j}=r_1-1}} \frac{(r_1-1)!}{\prod_{j=1}^{k-1} r_{1j}!} \\
&= \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\ell=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^\ell \binom{r_1-1}{\ell}\binom{n_1-k-1-\ell(k-1)}{r_1-2}
\end{aligned}
$$

and we assume $\sum_{j=m}^{n} a_j = 0$ if $m > n$. The Kronecker delta in the third line appears because when $s_{1k} = 1$ and $\sum_{j=1}^{k-1} jr_{1j} = n_1 - k$ and there is only one sequence for which the k/1-streak momentum statistic is undefined. The last line follows because the inner sum of the third line can be generated as the coefficient on x^{n_1-k} in the multinomial expansion of $(x+x^2+\cdots x^{k-1})^{r_1-1}$, which, as in determining $U_1(n_0, n_1)$, corresponds to the coefficient on the binomial expansion. Taking case 1i and 2ii together, the total number of sequences in which $\hat{P}_{1k}(\boldsymbol{x})$ is undefined is equal to $U_{1k} = U_{1k}^1 + U_{1k}^2$

In case 2 in which $\hat{P}_{1k}(\boldsymbol{x})$ is defined with $\sum_{i=n-k+1}^{n} x_i = k$ and $f_{1k} \geq 2$, it must be the case that $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$, and all sequences that satisfy this criteria can be constructed in three steps analogously to Lemma 6 by (1) selecting a distinguishable permutation of the $r_1 - 1$ remaining runs, (2) placing the $r_1 - 1$ 1s into the n_0 available positions to the left or the right of a 0, (3) filling the "empty" run cells. For a given $(r_{11}, \dots, r_{1k-1}, s_{1k})$ the total number of sequences satisfying this criteria is:

$$
\frac{(r_1-1)!}{(s_{1k}-1)!\prod_{j=1}^{k-1} r_{1j}!}\binom{n_0}{r_1-1}\binom{f_{1k}-1}{s_{1k}-1} = \frac{s_{1k}}{n_0+1} C_{1k}
$$

In case 3 in which $\hat{P}_{1k}(\boldsymbol{x})$ is defined with $\sum_{i=n-k+1}^{n} x_i < k$ and $f_{1k} \geq 1$, it must be the case that $\sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k$, as before all sequences that satisfy this criteria can be constructed in three steps analogously to Lemma 6, and we consider two subcases, sequences which terminate in a 1 (i.e. a run of 1s of length less than k) and sequences which terminate in a 0 (i.e. a run of 0s). For those sequence which terminate in a 1, for a given $(r_{11}, \ldots, r_{1k-1}, s_{1k})$ the total number of sequences satisfying this criteria is:

$$\left(\frac{r_1!}{s_{1k}! \prod_{j=1}^{k-1} r_{1j}!} - \frac{(r_1 - 1)!}{(s_{1k} - 1)! \prod_{j=1}^{k-1} r_{1j}!}\right) \binom{n_0}{r_1 - 1} \binom{f_{1k} - 1}{s_{1k} - 1} = \frac{r_1 - s_{1k}}{n_0 + 1} C_{1k}$$

with $(r_1 - 1)!/((s_{1k} - 1)! \prod_{j=1}^{k-1} r_{1j}!)$ being the number of sequences which terminate in a run of 1s of length k or more. For those sequences which terminate in a 0, for a given $(r_{11}, \ldots, r_{1k-1}, s_{1k})$ the total number of sequences satisfying this criteria is:

$$\frac{r_1!}{s_{1k}! \prod_{j=1}^{k-1} r_{1j}!} \binom{n_0}{r_1} \binom{f_{1k} - 1}{s_{1k} - 1} = \frac{n_0 + 1 - r_1}{n_0 + 1} C_{1k}$$

therefore, the sum of the k/1-streak momentum statistic across all sequences for which it is defined satisfies:

$$\begin{aligned} E[P_{1k}|N_1 = n_1] \left[\binom{n}{n_1} - U(n_0, n_1)\right] = & \sum_{\substack{r_{11}, \ldots, r_{1k-1}, s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{s_{1k}}{n_0 + 1} \frac{f_{1k} - s_{1k}}{f_{1k} - 1} \\ & + \sum_{\substack{r_{11}, \ldots, r_{1k-1}, s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{r_1 - s_{1k}}{n_0 + 1} \frac{f_{1k} - s_{1k}}{f_{1k}} \\ & + \sum_{\substack{r_{11}, \ldots, r_{1k-1}, s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{n_0 + 1 - r_1}{n_0 + 1} \frac{f_{1k} - s_{1k}}{f_{1k}} \end{aligned}$$

and this reduces to the formula in the theorem because the final two terms can be combined, and then can be summed over only runs which satisfy $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$ and combined with the first term (because $f_{1k} - s_{1k} = 0$ if $\sum_{j=1}^{k-1} jr_{1j} = n_1 - k$).[41]

■

B.2 Expected Difference in Proportions

The exact formula for the expected difference between the relative frequency of 1 for k/1-streak successor trials and the relative frequency of 1 for k/0-streak successor trials can be obtained with an approach similar to the previous section. The difference satisfies $D_k := P_{1k} - (1 - P_{0k})$, and there are three categories of sequences for which D_k is defined: (1) a sequence which ends in a run of 0s of length k or more with $f_{0k} \geq 2$ and $f_{1k} \geq 1$, with the difference equal to $D_k^1 = (f_{1k} - s_{1k})/f_{1k} - (s_{0k} - 1)/(f_{0k} - 1)$, (2) a sequence which ends in a run of 1s of length k or more with $f_{0k} \geq 1$ and $f_{1k} \geq 2$, with the difference equal to $D_k^2 := (f_{1k} - s_{1k})/(f_{1k} - 1) - s_{0k}/f_{0k}$ (3) a sequence which ends in a run of 0s of length $k-1$ or less or a run of 1s of length $k-1$ or less, with $f_{0k} \geq 1$ and $f_{1k} \geq 1$, with the difference equal to $D_k^3 := (f_{1k} - s_{1k})/f_{1k} - s_{0k}/f_{0k}$. For all other sequences the difference is undefined.

Theorem 8 *For $n > 0$ and n_1, n_0, k with $n_0 + n_1 = n$ and $1 < k \leq n_0, n_1 \leq n$, the expected difference in the relative frequency of 1 for k/1-streak successors that are 1 and the relative frequency*

[41] The formula

$$\sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k \\ s_{1k} \geq 1}} C_{1k} \frac{s_{1k}}{n_0 + 1} \frac{f_{1k} - s_{1k}}{f_{1k} - 1}$$

has a closed form solution in terms of binomial coefficients, but the other terms appear to not have one. Because U_{1k} does not have a closed form solution, $E[P_{1k}|N_1 = n_1]$ cannot have a closed form solution in any case..

of 1 for k/0-streak successors, $D_k := P_{1k} - (1 - P_{0k})$*, satisfies*

$$E[D_k \mid N_1 = n_1] = \frac{1}{\binom{n}{n_1} - U_k} \left[\sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} < n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} C_k \left[\frac{s_{0k}}{r_0} D_k^1 + \frac{r_0 - s_{0k}}{r_0} D_k^3 \right] \right.$$

$$\left. + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} C_k \left[\frac{s_{1k}}{r_1} D_k^2 + \frac{r_1 - s_{1k}}{r_1} D_k^3 \right] \right]$$

where $D_k^1 = (f_{1k} - s_{1k})/f_{1k} - (s_{0k} - 1)/(f_{0k} - 1), D_k^2 := (f_{1k} - s_{1k})/(f_{1k} - 1) - s_{0k}/f_{0k}, D_k^3 := (f_{1k} - s_{1k})/f_{1k} - s_{0k}/f_{0k}$*, and*

$$C_k := \frac{r_0!}{s_{0k}! \prod_{i=1}^{k-1} r_{0i}!} \frac{r_1!}{s_{1k}! \prod_{i=1}^{k-1} r_{1i}!} \binom{f_{0k} - 1}{s_{0k} - 1} \binom{f_{1k} - 1}{s_{1k} - 1}$$

and U_k *is the number of sequences in which either there are no k/1-streak successors or no k/0-streak successors.*

Proof:

Note that for the case in which $|r_1 - r_0| = 1$, C_k is the number of sequences with $N_1 = n_1$ in which the number of runs of 0s and runs 1s satisfy run profile $(r_{01}, \ldots, r_{0k-1}, s_{0k}; r_{11}, \ldots, r_{1k-1}, s_{1k})$, and for the cases with $r_1 = r_0$, C_k is equal to half the number of these sequences (because each sequence can end with a run of 1s or a run of 0s). The combinatorial proof of this formula, which we omit, is similar to that of Lemma 6.

The sum total of the differences across all sequences for which the difference is defined and $N_1 = n_1$ is

$$
\begin{aligned}
E[D_k \mid N_1 = n_1] \cdot \left[\binom{n}{n_1} - U_k \right] \\
= \sum_{\substack{r_{01},\dots,r_{0k-1},s_{0k} \\ r_{11},\dots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} < n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} \frac{s_{0k}}{r_0} C_k D_k^1 \quad + \sum_{\substack{r_{01},\dots,r_{0k-1},s_{0k} \\ r_{11},\dots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} \frac{s_{1k}}{r_1} C_k D_k^2 \\
+ \sum_{\substack{r_{01},\dots,r_{0k-1},s_{0k} \\ r_{11},\dots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} \frac{r_0 - s_{0k}}{r_0} C_k D_k^3 \quad + \sum_{\substack{r_{01},\dots,r_{0k-1},s_{0k} \\ r_{11},\dots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} \frac{r_1 - s_{1k}}{r_1} C_k D_k^3
\end{aligned}
$$

where the first sum relates to those sequences which end in a run of 0s of length k or more (whence $r_0 \geq r_1$, the multiplier s_{0k}/r_0 and $\sum_{j=1}^{k-1} jr_{0j} < n_0 - k$),[42] the second sum relates to those sequences which end in a run of 1s of length k or more (whence $r_1 \geq r_0$, the multiplier s_{1k}/r_1 and $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$), the third sum relates to those sequences which end on a run of 0s of length k-1 or less (whence $r_0 \geq r_1$, the multiplier $(r_0 - s_{0k})/r_0$) and $\sum_{j=1}^{k-1} jr_{0j} < n_0 - k$),[43] and the fourth sum relates to those sequences which end on a run of 1s of length $k-1$ or less (whence $r_1 \geq r_0$, the multiplier $(r_1 - s_{1k})/r_1$) and $\sum_{j=1}^{k-1} jr_{1j} < n_1 - k$). The four terms can be combined to the

[42] Note $\sum_{j=1}^{k-1} jr_{0j} < n_0 - k \Longleftrightarrow f_{0k} \geq 2$.

[43] The multiplier $(r_0 - s_{0k})/r_0$ arises because the number of distinguishable permutations of the 0 runs that end with a run of length $k-1$ or less is equal to the total number of distinguishable permutations of the 0 runs minus the number of distinguishable permutations of the 0 runs that end in a run of length k or more, i.e.

$$\frac{r_0!}{s_{0k}! \prod_{i=1}^{k-1} r_{0i}!} - \frac{(r_0 - 1)!}{(s_{0k} - 1)! \prod_{i=1}^{k-1} r_{0i}!} = \frac{r_0 - s_{0k}}{r_0} \frac{r_0!}{s_{0k}! \prod_{i=1}^{k-1} r_{0i}!}$$

following two terms:

$$E[D_k \mid N_1 = n_1] \cdot \left[\binom{n}{n_1} - U_k\right] = \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} < n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} \leq n_1 - k, s_{1k} \geq 1 \\ r_0 \geq r_1}} C_k \left[\frac{s_{0k}}{r_0} D_k^1 + \frac{r_0 - s_{0k}}{r_0} D_k^3\right] + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} \leq n_0 - k, s_{0k} \geq 1 \\ \sum_{j=1}^{k-1} jr_{1j} < n_1 - k, s_{1k} \geq 1 \\ r_1 \geq r_0}} C_k \left[\frac{s_{1k}}{r_1} D_k^2 + \frac{r_1 - s_{1k}}{r_1} D_k^3\right]$$

which can be readily implemented numerically for $n \leq 300$.[44] The total number of sequences for which the difference is undefined, U_k, can be counted analogously to Theorem 7 with an application of the inclusion-exclusion principle:

$$\begin{aligned} U_k := & \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} = n_1 \\ s_{1k} = 0}} C_{1k} + \sum_{\substack{r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{1j} = n_1 - k \\ s_{1k} = 1}} \frac{s_{1k}}{n_0 + 1} C_{1k} + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ \sum_{j=1}^{k-1} jr_{0j} = n_1 \\ s_{0k} = 0}} C_{0k} + \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ \sum_{j=1}^{k-1} jr_{0j} = n_1 - k \\ s_{0k} = 1}} \frac{s_{0k}}{n_1 + 1} C_{0k} \\ & - \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} = n_0, s_{0k} = 0 \\ \sum_{j=1}^{k-1} jr_{1j} = n_1, s_{1k} = 0 \\ |r_0 - r_1| \leq 1}} \left(2 \cdot \mathbb{1}_{\{r_1 = r_0\}} + \mathbb{1}_{\{|r_1 - r_0| = 1\}}\right) C_k \\ & - \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} = n_0 - k, s_{0k} = 1 \\ \sum_{j=1}^{k-1} jr_{1j} = n_1, s_{1k} = 0 \\ r_0 \geq r_1}} \frac{s_{0k}}{r_0} C_k - \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k} \\ r_{11},\ldots,r_{1k-1},s_{1k} \\ \sum_{j=1}^{k-1} jr_{0j} = n_0, s_{0k} = 0 \\ \sum_{j=1}^{k-1} jr_{1j} = n_1 - k, s_{1k} = 1 \\ r_1 \geq r_0}} \frac{s_{1k}}{r_1} C_k \end{aligned}$$

where C_{0k} is a function of $(r_{01}, \ldots, r_{0k-1}, s_{0k}; n_0, n_1)$ and defined analogously to C_{1k}. We can simplify the above expression by first noting that the third term, corresponding to those sequences

[44] In the numerical implementation one can consider three sums $r_0 = r_1 + 1$, $r_1 = r_0 + 1$, and for the case of $r_1 = r_0$ the sums can be combined, with a sum of four terms for each term of the finite series.

in which there are no k/1-streak successors and no k/0-streak successors, can be reduced to a sum of binomial coefficients:

$$\begin{aligned}
&\sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k}\\ r_{11},\ldots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0,s_{0k}=0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1,s_{1k}=0\\ |r_0-r_1|\le 1}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}}+\mathbb{1}_{\{|r_1-r_0|=1\}}\right) C_k \\
&\quad = \sum_{\substack{r_{01},\ldots,r_{0k-1},s_{0k}\\ r_{11},\ldots,r_{1k-1},s_{1k}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0,s_{0k}=0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1,s_{1k}=0\\ |r_0-r_1|\le 1}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}}+\mathbb{1}_{\{|r_1-r_0|=1\}}\right) \frac{r_0!}{s_{0k}!\prod_{i=1}^{k-1} r_{0i}!}\frac{r_1!}{s_{1k}!\prod_{i=1}^{k-1} r_{1i}!} \\
&\quad = \sum_{\substack{r_{01},\ldots,r_{0k-1}\\ r_{11},\ldots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1\\ |r_0-r_1|\le 1}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}}+\mathbb{1}_{\{|r_1-r_0|=1\}}\right) \frac{r_0!}{\prod_{i=1}^{k-1} r_{0i}!}\frac{r_1!}{\prod_{i=1}^{k-1} r_{1i}!} \\
&\quad = \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}}+\mathbb{1}_{\{|r_1-r_0|=1\}}\right) \sum_{\substack{r_{01},\ldots,r_{0k-1}\\ r_{11},\ldots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0\\ \sum_{j=1}^{k-1} jr_{1j}=n_1\\ \sum_{j=1}^{k-1} r_{0j}=r_0\\ \sum_{j=1}^{k-1} r_{1j}=r_1}} \frac{r_0!}{\prod_{i=1}^{k-1} r_{0i}!}\frac{r_1!}{\prod_{i=1}^{k-1} r_{1i}!} \\
&\quad = \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}}+\mathbb{1}_{\{|r_1-r_0|=1\}}\right) \sum_{\substack{r_{01},\ldots,r_{0k-1}\\ \sum_{j=1}^{k-1} jr_{0j}=n_0\\ \sum_{j=1}^{k-1} r_{0j}=r_0}} \frac{r_0!}{\prod_{i=1}^{k-1} r_{0i}!} \sum_{\substack{r_{11},\ldots,r_{1k-1}\\ \sum_{j=1}^{k-1} jr_{1j}=n_1\\ \sum_{j=1}^{k-1} r_{1j}=r_1}} \frac{r_1!}{\prod_{i=1}^{k-1} r_{1i}!} \\
&\quad = \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2\cdot \mathbb{1}_{\{r_1=r_0\}}+\mathbb{1}_{\{|r_1-r_0|=1\}}\right) \sum_{\ell_0=0}^{\lfloor \frac{n_0-r_0}{k-1}\rfloor} (-1)^{\ell_0}\binom{r_0}{\ell_0}\binom{n_0-1-\ell_0(k-1)}{r_0-1} \\
&\qquad \times \sum_{\ell_1=0}^{\lfloor \frac{n_1-r_1}{k-1}\rfloor} (-1)^{\ell_1}\binom{r_1}{\ell_1}\binom{n_1-1-\ell_1(k-1)}{r_1-1}
\end{aligned}$$

For the final two negative terms in the formula for U_k, we may apply a similar argument to represent them as a sum of binomial coefficients. For the first four positive terms we can use the argument provided in Theorem 7 to represent them as sums of binomial coefficients, and therefore U_k reduces to a sum of binomial coefficients:

$$
\begin{aligned}
U_k = & \sum_{r_1=1}^{\min\{n_1,n_0+1\}} \binom{n_0+1}{r_1} \sum_{\ell=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^\ell \binom{r_1}{\ell} \binom{n_1-1-\ell(k-1)}{r_1-1} \\
& + \delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \binom{n_0}{r_1-1} \sum_{\ell=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^\ell \binom{r_1-1}{\ell} \binom{n_1-k-1-\ell(k-1)}{r_1-2} \\
& + \sum_{r_0=1}^{\min\{n_0,n_1+1\}} \binom{n_1+1}{r_0} \sum_{\ell=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^\ell \binom{r_0}{\ell} \binom{n_0-1-\ell(k-1)}{r_0-1} \\
& + \delta_{n_0 k} + \sum_{r_0=2}^{\min\{n_0-k+1,n_1+1\}} \binom{n_1}{r_0-1} \sum_{\ell=0}^{\lfloor \frac{n_0-k-r_0+1}{k-1} \rfloor} (-1)^\ell \binom{r_0-1}{\ell} \binom{n_0-k-1-\ell(k-1)}{r_0-2} \\
& - \left[\sum_{r_1=1}^{\min\{n_1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1+1,n_0\}} \left(2 \cdot \mathbb{1}_{\{r_1=r_0\}} + \mathbb{1}_{\{|r_1-r_0|=1\}}\right) \times \right. \\
& \quad \left. \times \sum_{\ell_0=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^{\ell_0} \binom{r_0}{\ell_0} \binom{n_0-1-\ell_0(k-1)}{r_0-1} \sum_{\ell_1=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^{\ell_1} \binom{r_1}{\ell_1} \binom{n_1-1-\ell_1(k-1)}{r_1-1} \right] \\
& - \left[\delta_{n_0 k} + \sum_{r_0=2}^{\min\{n_0-k+1,n_1+1\}} \sum_{r_1=\max\{r_0-1,1\}}^{\min\{r_0,n_1\}} \sum_{\ell_0=0}^{\lfloor \frac{n_0-k-r_0+1}{k-1} \rfloor} (-1)^{\ell_0} \binom{r_0-1}{\ell_0} \binom{n_0-k-1-\ell_0(k-1)}{r_0-2} \times \right. \\
& \quad \left. \times \sum_{\ell_1=0}^{\lfloor \frac{n_1-r_1}{k-1} \rfloor} (-1)^{\ell_1} \binom{r_1}{\ell_1} \binom{n_1-1-\ell_1(k-1)}{r_1-1} \right] \\
& - \left[\delta_{n_1 k} + \sum_{r_1=2}^{\min\{n_1-k+1,n_0+1\}} \sum_{r_0=\max\{r_1-1,1\}}^{\min\{r_1,n_0\}} \sum_{\ell_1=0}^{\lfloor \frac{n_1-k-r_1+1}{k-1} \rfloor} (-1)^{\ell_1} \binom{r_1-1}{\ell_1} \binom{n_1-k-1-\ell_1(k-1)}{r_1-2} \times \right. \\
& \quad \left. \times \sum_{\ell_0=0}^{\lfloor \frac{n_0-r_0}{k-1} \rfloor} (-1)^{\ell_0} \binom{r_0}{\ell_0} \binom{n_0-1-\ell_0(k-1)}{r_0-1} \right]
\end{aligned}
$$

■

B.3 The relationship to finite sample bias when estimating autocorrelation in time series data

It is well known that standard estimates of autocorrelation are biased in finite samples (Yule 1926), as well as for regression coefficients in time series data (Stambaugh 1986, 1999). Below we show that the bias in the estimate of the probability of a 1 conditional on a k/1-streak has direct relationship with these least squares estimators via a linear probability model with an endogenous dummy covariate that indicates if a trial is a k/1-streak successor.

Theorem 9 *Let* $\boldsymbol{x} \in \{0,1\}^n$ *with* $I_{1k}(\boldsymbol{x}) \neq \emptyset$. *If* $\boldsymbol{\beta}_k(\boldsymbol{x}) = (\beta_{0k}(\boldsymbol{x}), \beta_{1k}(\boldsymbol{x}))$ *is defined to be the solution to the least squares problem,* $\boldsymbol{\beta}_k(\boldsymbol{x}) \in \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^2} \|\boldsymbol{x} - [\mathbf{1}\ \boldsymbol{d}]^\intercal \boldsymbol{\beta}\|^2$ *where* $\boldsymbol{d} \in \{0,1\}^n$ *is defined so that* $d_i := \mathbb{1}_{I_{1k}(\boldsymbol{x})}(i)$ *for* $i = 1, \ldots, n$, *then*[45]

$$\beta_{0k}(\boldsymbol{x}) + \beta_{1k}(\boldsymbol{x}) = \hat{P}_{1k}(\boldsymbol{x})$$

Proof:

If $\boldsymbol{\beta}_k(\boldsymbol{x})$ minimizes that sum of squares then $\beta_{1k}(\boldsymbol{x}) = \sum_{i=1}^n (x_i - \bar{x})(d_i - \bar{d}) / \sum_{i=1}^n (d_i - \bar{d})^2$. First, working with the numerator, letting $I_{1k} \equiv I_{1k}(\boldsymbol{x})$ we have

$$\begin{aligned}
\sum_{i=1}^n (x_i - \bar{x})(d_i - \bar{d}) &= \sum_{i\in I_{1k}} (x_i - \bar{x})(d_i - \bar{d}) + \sum_{i\in I_{1k}^C} (x_i - \bar{x})(d_i - \bar{d}) \\
&= \left(1 - \frac{|I_{1k}|}{n}\right) \sum_{i\in I_{1k}} (x_i - \bar{x}) - \frac{|I_{1k}|}{n} \sum_{i\in I_{1k}^C} (x_i - \bar{x}) \\
&= \left(1 - \frac{|I_{1k}|}{n}\right) \sum_{i\in I_{1k}} x_i - \frac{|I_{1k}|}{n} \sum_{i\in I_{1k}^C} x_i - \left(1 - \frac{|I_{1k}|}{n}\right) |I_{1k}|\bar{x} + \frac{|I_{1k}|}{n}(n - |I_{1k}|)\bar{x} \\
&= |I_{1k}| \left(1 - \frac{|I_{1k}|}{n}\right) \left(\frac{\sum_{i\in I_{1k}} x_i}{|I_{1k}|} - \frac{\sum_{i\in I_{1k}^C} x_i}{n - |I_{1k}|} \right)
\end{aligned}$$

[45] When $I_{1k}(\boldsymbol{x}) = \emptyset$ the solution set of the least squares problem is infinite, i.e. $\operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^2} \|\boldsymbol{x} - \beta_0\|^2 = \{(\beta_0, \beta_1) \in \mathbb{R}^2 : \beta_0 = (1/n)\sum_{i=1}^n x_i\}$. If we treat $\boldsymbol{\beta}_k(\boldsymbol{x})$ as undefined in this case, then the bias from using the conditional relative frequency is equal to the finite sample bias in the coefficients of the associated linear probability model. If instead we define $\beta_{1k}(\boldsymbol{x}) = 0$, then the bias in the coefficients of the associated linear probability model will be less than the bias in the conditional relative frequency.

second, with the denominator of $\beta_{1k}(\boldsymbol{x})$ we have

$$\begin{aligned}\sum_{i=1}^{n}(d_i - \bar{d})^2 &= \sum_{i\in I_{1k}}\left(1-\frac{|I_{1k}|}{n}\right)^2 + \sum_{i\in I_{1k}^C}\left(\frac{|I_{1k}|}{n}\right)^2 \\ &= |I_{1k}|\left(1-\frac{|I_{1k}|}{n}\right)^2 + (n-|I_{1k}|)\left(\frac{|I_{1k}|}{n}\right)^2 \\ &= |I_{1k}|\left(1-\frac{|I_{1k}|}{n}\right)\end{aligned}$$

therefore we have

$$\beta_{1k}(\boldsymbol{x}) = \frac{\sum_{i\in I_{1k}} x_i}{|I_{1k}|} - \frac{\sum_{i\in I_{1k}^C} x_i}{n-|I_{1k}|}$$

now

$$\begin{aligned}n\beta_{0k}(\boldsymbol{x}) &= n(\bar{x} - \beta_1(\boldsymbol{x})\bar{d}) \\ &= \sum_{i=1}^{n} x_i - \left(\frac{\sum_{i\in I_{1k}} x_i}{|I_{1k}|} - \frac{\sum_{i\in I_{1k}^C} x_i}{n-|I_{1k}|}\right)|I_{1k}| \\ &= \sum_{i\in I_{1k}^C} x_i + \frac{|I_{1k}|\sum_{i\in I_{1k}^C} x_i}{n-|I_{1k}|} \\ &= \frac{n\sum_{i\in I_{1k}^C} x_i}{n-|I_{1k}|}\end{aligned}$$

and summing both coefficients, the result follows.

■

Note that the bias in the coefficients follows from the bias in the estimate of the conditional probability, i.e. $E[\hat{P}_{1k}(\boldsymbol{x})] < p$ implies $E[\beta_{1k}(\boldsymbol{x})] < 0$ and $E[\beta_{0k}(\boldsymbol{x})] > p$.[46]

[46]Note that $p = E\left[(1/n)\sum_{i=1}^{n} x_i\right]$, and the sum can be broken up and re-arranged as in the theorem.