On a common perception of a random sequence in cricket

Ian N. Durbach and Jani Thiart
Department of Statistical Sciences, University of Cape Town

Abstract

This paper considers a commonly-held belief about sequential cricketing data, which states that batsmen go through periods of good and bad batting 'form' that are longer than one could expect from the outcome of a random process. We test this belief against empirical data, find a distinct lack of support for it, and give some simple psychological explanations for how the erroneous judgement might arise. The paper can be read as an analysis of some previously unexplored aspects of the game of cricket, as an extension of the hot-hand literature to the game of cricket, or as an exploration of some opportunities to increase statistical literacy.

keywords: random sequences, cricket, hot-hand, heuristics and biases

1 Introduction

For most of us, the ability to think clearly and correctly about problems of a statistical nature does not come easily. Evidence abounds that we see patterns in random sequences when there are none, believe that chance processes are self-correcting, are insensitive to prior probabilities or base rates in constructing our beliefs, do not take sample sizes into account when evaluating hypotheses, and are generally poor when it comes to picking up regression toward the mean, even in circumstances like the relationship between heights of fathers and sons in which it is bound to occur (for an overview of judgemental biases see [8] or [18]). Although there are measures, including training in statistical methods, to counter the extent of these biases (e.g. [7]), it would be fair to summarise the enormous literature on so-called heuristics and biases by saying that people are poor intuitive statisticians, and that many of our intuitions and much of our folk psychology should be subjected to analysis using formal statistical methods before being accepted.

In this paper we look at a perceptions commonly held by cricket fans about their sport and hold it up to the scrutiny of formal statistical testing. The perceptions is the widely held belief that during the course of their careers players (batsmen in this instance) go through periods when they are 'in form' and score particularly highly game after game, and periods when they are 'out of form' and struggle to score any runs at all. That a batsman who plays for any reasonable length of time will experience some subsequences of consecutive high-scoring innings and some subsequences of consecutive low-scoring innings is clearly true and not at issue here. What is important is the belief that these streaks are longer than would be expected if the sequence of scores were generated by a random process alone. This belief is suggested by the very use of the term 'form', which has all sorts of connotations that the player is feeling mentally and physically strong. The existence of such sporting form would be supported if the observed streaks are longer than those generated by a random process; otherwise, the observed streaks are merely the wanderings of a random sequence, and cricket fans (and selectors and players) should not read too much into them. The evidence one way or another has clear and important implications for team selection. Simply put, if the process is random, selectors should not drop a batsman because he is thought to be 'out-of-form' – other considerations (e.g. technical deficiencies, batting out of position) are necessary.

Research into the validity of perceptions of random sequences have a fairly long and distinguished history in the American sports of basketball and baseball, some of which is reviewed in the next section. Baseball in particular is a sport rich in official statistics, with several annual volumes devoted to the publication of various quantitative descriptions of the sport. It is therefore an especially fruitful field for finding large volumes of high quality data. Cricket is a similarly data-rich sport, but has not enjoyed the same amount of attention from statisticians. Most research has investigated the distribution of batting scores, beginning with independent research by Wood [21] and Elderton [6] that both found batting scores to be well approximated by a geometric distribution. The implication of the memoryless property possessed by the geometric distribution (that a batsman's chance of getting out is independent of how many runs he has scored) was held by discussants of the paper to be unsupportable by cricketing lore (see [21] and [6]). Much later research [10] showed that properly accounting for incomplete or not-out innings leads to a distinct lack of fit with the geometric distribution, particularly for very low scores (the previous studies had treated not-out scores as complete innings). The measurement of batting ability has also received some related attention. The traditional batting 'average' was augmented in [10] to better account for not-out totals, and Barr and Kantor [3] proposed a risk-return framework for evaluating batsmen, including both a batsman's strike rate (runs per 100 balls faced) and his (constant) probability of going out. They also returned to the debate on the distribution of scores, suggesting that a memoryless geometric approximation may be appropriate where batsmen begin their innings cautiously and adopt increasingly risky strategies as they progress further into their innings, thus maintaining a constant probability of losing their wicket.

Today websites devoted to cricket freely offer enormous amounts of comprehensive and high-quality data on every aspect of the sport: batting, bowling, fielding, even umpiring. What is perhaps more interesting is that there are genuinely interesting questions regarding statistical thinking and statistical education that can be answered using this data, paradoxically precisely because most people have far more carefully thought out views on cricket (e.g. 'form') than they do on statistical issues (e.g. random processes). This paper is an attempt to illustrate that point. In the following section we review previous research into the randomness of sequences and the perception of that randomness. Then, in section 3 we examine the evidence for the existence of batting 'form' that would lead to streaks longer than those expected of a random process, find the evidence for the belief to be lacking, and provide some thoughts on possible explanations of the beliefs. A final discussion in section 4 concludes the paper.

2 Chance processes in sports

The departure of people's conceptions of randomness from the laws of chance is captured in Kahneman and Tversky's 'law of small numbers' [17], by which people expect the characteristics of a global sequence to apply locally in even very short sequences i.e. believe that the law of large numbers applies to small numbers too. This expectation gives rise to two biases: firstly, the belief that chance processes are self-correcting, also known as the gambler's fallacy – to believe, for example, that a coin is more likely to land tails-up on the next toss following a long sequence of heads. The second bias is the tendency to describe sequences that contain the expected number of runs (for a random process) as 'non-random', and those that contain more than the expected number of runs i.e. those that alternate rapidly, as 'random'.

The classic investigation of the second of these biases in sport appeared in a 1985 paper on the so-called 'hot-hand' phenomenon in basketball [9]. Belief in the hot-hand refers to the belief that a player is more likely to be successful with a shot if he has been successful with the previous shot(s). Thus success breeds success, and failure breeds failure. Essentially, the results showed that the probability of a successful shot did not change for 8 of the 9 players analysed when those probabilities were conditioned on a successful previous shot, or any number of successful previous shots for that matter. The ninth player exhibited negative serial autocorrelation, precisely the opposite of the hot-hand hypothesis. This result persisted whether data was taken from general play (which are subject to many other uncontrolled factors) or from free-throw data (where the player attempts a penalty shot from a fixed point with no interference from the opposition). Moreover, the number of runs of successful and unsuccessful shots were exactly what one would expect from a random process for 8 out of the 9 players analysed; again the ninth player had significantly more runs than would be expected, contrary to the hot-hand hypothesis.

Criticism of the original investigation and a follow-up in [16] centered on two issues: firstly, that the hot-hand might be observed in smaller "cognitively manageable chunks" of data, expressed either as a small subset of players or shorter periods of play [11]; secondly, that the statistical tests used (a runs test, a test of serial autocorrelation, and a non-standard and fairly *ad hoc* test of non-stationarity) were insufficiently powerful to detect a departure from Bernoulli trials unless that departure was substantial, and did not distinguish between autocorrelation, in which the probability of success is dependent on history, and non-stationarity, in which the probability of success is subject to occasional elevations [20]. The widespread belief that the hot-hand applies to almost any player reduces the first of these concerns somewhat [15], and for our purposes would apply even more strongly: conventional beliefs suggest that there is surely no player who is entirely exempt from 'form'. The second concern regarding statistical methods is more relevant to the aims of the current paper, and we pay particular attention to the need to distinguish between autocorrelation and non-stationarity in the time series.

The investigation of random sequences in sports was extended to baseball in 1993 [2]. This investigation devotes far more time to methodology and far less time to an explanation of the bias itself, but nevertheless the methodology remains fairly straightforward, employing a logistic regression framework that incorporates a flexible 'history' covariate defined as 'the number of successes in the last k innings', where various values of k are used. This frees the analyst from a specific definition of a 'hot streak'. The analysis makes two further advances over the earlier studies into basketball: firstly, data is available for 501 player-seasons, so that one may plausibly ask if many of these exhibit non-random hitting sequences – this is simply not possible for the previous sample of 9 basketball shooters. Secondly, data on several other covariates are included – these so-called 'situational factors' relate to the external factors influencing the likelihood of success that were not controlled for in the basketball studies (e.g. home/away, economy rate of pitcher). Nevertheless, the results of the study indicate that while some players exhibit clear streakiness in certain years, there are (a) no more of these players than one would expect under a model of randomness, and (b) no players who are streaky over the entire period of study (four years). The problem of confounding covariates, however, is an important point, and some subsequent research [5] has found positive evidence in favour of the hot-hand in professional bowling, a sport in which conditions are highly controlled and thus results are less susceptible to situational factors.

The criticisms of the analysis in [2] focus on the two related issues of statistical power and the method employed. Simulation results in [14] show that logistic regression leads to results biased against finding significant evidence of hitting streaks, and is not powerful enough to detect the likely effect size of previous success (speculated to increase the probability of success by no more than 5%). Combining data across players and seasons (as we do) is advocated to combat the limited power of the individual player-seasons, and using this approach and adjusting for bias provides some weak evidence of streaky hitting, but this appears to be largely due to a single exceptionally streaky player-season. A two-state Markov switching model (where the states are 'hot' and 'cold') has also been used [1], identifying the same exceptionally streaky player-season as the combined logistic regression results in [14].

To summarise, previous research can be categorised into two distinct views. On the one hand, there are those researchers who believe that the widespread belief and regular reference to non-randomness in sequences of successes (hot-hand or streaky hitting sequences) means that the effect is not expected to be particularly subtle, that one can use conventional statistical methods to gain a reasonably reliable insight into the nature of the sequences, and that a failure to reject the null hypothesis of Bernoulli trials constitutes confirmation of the non-existence of streakiness. Proponents of this view would include the authors of both the original papers in basketball [9] and baseball [2]. The other camp consists of those researchers who are more conservative about the size and the prevalence of the possible effect, caution against using the simpler statistical methods that have limited power or ability to detect the different possible interpretations of streaks, and who view the failure to reject the null hypothesis, in light of a lack of a suitable alternative hypothesis, as not necessarily fatal for the ideas of streakiness or the hot-hand. The view we take is that while a lack of statistical evidence may not necessarily conclusively prove the non-existence of the phenomenon (e.g. batting form), it does strongly suggest that its effect is both weaker and less prevalent than is currently thought.

3 Evidence for the presence of 'form' in batting sequences

Batting score data was obtained for the entire test careers (or career-to-date for those still playing) of 16 batsmen from `http://www.cricinfo.com`. Most of the selected batsmen are extremely experienced and have in some cases played in excess of 200 innings, with the sample size i.e. total number of innings, amounting to $n = 2770$. Popular references to batting 'form' are so ubiquitous that one might expect that any batsmen to have played the game for a reasonable period of time would have experienced its associated ups and downs, and the selection of a small number of experienced batsmen does not seem problematic. The selection process therefore involved selecting five premier South African batsmen (D Cullinan, H Gibbs, J Kallis, G Kirsten, G Smith), together with 10 other batsmen who could each claim to be among the 'giants' of the modern batting game (A Border, R Dravid, M Hayden, B Lara, V Sehwag, M Taylor, S Tendulkar, I Ul-Haq, M Waugh, S Waugh). A final batsman (B Dippenaar) was added as a batsmen who has experienced more mixed fortunes over the course of his career. A brief summary of the test careers of the 16 batsmen is provided in Table 3 in the appendix.

The case for sporting form is well-highlighted by the case of one of our selected batsmen, Graeme Smith, whose scores in each of his 96 innings are showed in Figure 1 together with his cumulative career average. There are two clear regions, one starting just before his 20th innings and the other largely occurring between innings 70 to 80, in which scores are consecutively below average and would thus appear to be indicative of a run of 'bad form'. Innings highlighted with large solid circles are those that would be highlighted as 'runs violations' by a heuristic popular in statistical quality control – that any pattern of more than 8 observations either side of the mean is said to indicate some prolonged bias in the underlying process [13]. There are 9 such observations for Smith, 1 in the first region and 8 in the second, more than any other selected batsmen. Yet the term 'violation of runs' is misleading, and does not refer directly to any formal statistical test: that is, over the course of millions of coin flips, we would naturally expect sequences of nine or more heads.

There is no question that such sequences do occur in batting scores. The main question of interest is thus whether the kinds of sequences exhibited by Smith could be expected of a random process such as a coin flip.

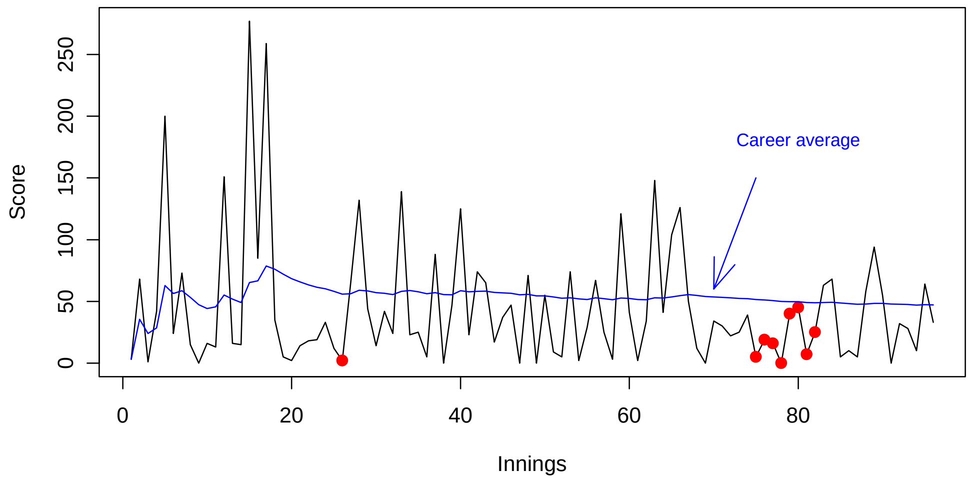


Figure 1: Runs scored by Graeme Smith over his test career-to-date (96 innings)

Previous research on the randomness of sequences of sporting events has largely focused on two areas that will also form the framework for the investigations in this paper: firstly, whether an outcome in the actual sporting sequence (e.g. a basketball shot, baseball hit, batting score) can be predicted using the outcomes of previous attempts in the sequence; secondly, whether there are fewer alternating runs of successes and failures in the observed sequence than could be expected from a sequence generated by a random process. As pointed out in [20], the first of these questions essentially interprets form as a type of autocorrelation; the second as a type of non-stationarity. Turning to the 'form as autocorrelation' interpretation, we may attempt to model the number of runs scored in innings i by some function of the number of runs scored in the period of the previous ϕ completed innings, possibly including some other covariates thought to influence performance. Previous studies have typically carried out this task using either simple correlations [16] or regressing a binary success variable on a function of previous successes [2]. Cricket data is quite different from data obtained from basketball and baseball studies in that a large proportion of innings are censored (i.e. not-out), and there is no objective way to delineate what constitutes a ‘success’ or ‘failure’. This has two implications: firstly, any runs scored in intervening not-out innings (i.e. those up to ϕ completed innings ago) should be added to the total number of runs scored in those last ϕ completed innings; secondly, it makes sense for the time being to directly use the score s_i obtained in innings i as a dependent variable, and model this score using the Cox proportional hazards model

$$h_i(s_i) = h_0(s_i)\exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}), \tag{1}$$

where $h_i(s_i)$ is the usual hazard function giving the instantaneous risk of demise at a score of s_i runs and we make use of the following independent variables:

- x_{i1}: An exponentially-weighted sum of the runs scored in the last $\phi+\theta$ innings, where θ is the number of not-out innings in the period over which ϕ completed innings have occured, and $\phi \in \{2, 4, 6, 10\}$ i.e. $x_{i1} = \sum_{j=1}^{\phi+\theta} 0.85^{j-1} s_{i-j}$,
- x_{i2}: A binary variable indicating whether the innings is the team’s first batting innings in the test match
- x_{i3}: A binary variable indicating whether the match is played against one of the generally weaker test-playing nations (Bangladesh, Zimbabwe, New Zealand, Sri Lanka)
- x_{i4}: A binary variable indicating whether the match is being played at home

The exponentially-weighted sum representing batting ‘form’ is thus allowed to stretch back as far as the previous ten innings, but in all cases gives preferential weighting to the scores obtained in the most recent innings. A similar mechanism for capturing form is provided in [2], which also used a value of 0.85 to weight the lagged observations. Other values between 0.7 and 1 were also tried, with no material influence on the results shown here. Different models are run for each definition of x_{i1} i.e. each value of ϕ, from which one might observe whether any particular length of time is conducive to the emergence of form as a significant predictor. Table 1 gives the regression coefficients for each of the lagged form variables together with their standard errors and significance levels. The results for the other three covariates that are shown were obtained under the most significant time lag $\phi = 2$, but these results do not change materially if other values of ϕ are used. Note that a negative coefficient indicates that an independent variable reduces the hazard of losing one's wicket, and thus can be said to positively influence a batsman's score in the current innings.

	Form $\phi = 2$	Form $\phi = 4$	Form $\phi = 6$	Form $\phi = 10$	First Innings	Weaker Opponent	Home Ground
β	$-5.6E^{-4}$	$-2.8E^{-4}$	$-2E^{-4}$	$-1E^{-4}$	-0.16	-0.12	$8.8E^{-3}$
s.e.	$3.3E^{-4}$	$2.8E^{-4}$	$2.6E^{-4}$	$2.6E^{-4}$	0.04	0.05	0.04
t	-1.67	-0.98	-0.75	-0.38	-3.82	-2.56	0.22
p	0.10	0.33	0.45	0.70	< 0.001	0.01	0.83

Table 1: Results for Cox proportional hazards model predicting current score

Importantly, there is some weak evidence for the presence of short-term batting form using $\phi = 2$. That is, the number of runs scored in the last two innings may exercise a marginal effect on the current score. It is also worth noting that the sign of the coefficients for batting form have the hypothesised negative sign for all values of ϕ i.e. higher previous scores are associated with higher current scores. However, there is no evidence of any form beyond two innings, and the trend toward smaller and less significant coefficients as more distant innings are included in the definitions of batting form suggests that longer-term form of the sort usually alluded to by the sporting press and fans is not a general phenomenon. Although a value of $\phi = 1$ is too small to denote any kind of real 'form', a similar Cox model run with $\phi = 1$ also returned a non-significant regression coefficient for form ($\beta = 3.7E^{-4}, p = 0.37$). In contrast, there is substantial evidence that scores tend to be higher in the first (batting) innings, and that scores also tend to be higher when playing against the so-called weaker opposition. While these effects are hardly surprising, they are encouraging indications that where effects are truly general they are detected as such. Perhaps surprisingly, there is no indication that batsmen score better on home pitches than they do when they away from home.

We now consider whether any additional insight can be gained by considering the notion of 'form' to be represented by longer-than-random runs of successes or failures. Any insights depend critically on what is meant by 'success' and 'failure', but before turning to those questions we can note that, once definitions of success and failure have been established, there is a large body of established theory on the distribution of (statistical) runs that can be drawn upon, and provide a short summary of the part of that theory relevant to the study of batting form. Consider a string containing a random arrangement n_S successes and n_F failures, with $n = n_S + n_F$. Let R_{Sj} denote a random variable indicating the number of runs of successes of length j and R_{Fj} be another random variable indicating the number of runs of failures of length j. The total number of runs of successes (of any length) is then given by $R_S = \sum_j R_{Sj}$, and the total number of runs of failures (of any length) is given by $R_F = \sum_j R_{Fj}$. Then Mood [12] has shown that the expected numbers of runs of successes and failures of length j are

$$\mathrm{E}[R_{Sj}] = \frac{(n_F+1)^{(2)} n_S^{(j)}}{n^{(j+1)}} \tag{2}$$

$$\mathrm{E}[R_{Fj}] = \frac{(n_S+1)^{(2)} n_F^{(j)}}{n^{(j+1)}} \tag{3}$$

where $x^{(a)}$ denotes the factorial form $x(x-1)(x-2)\dots(x-a+1)$. A simple way to evaluate the presence of form is to compute the actual number of successful runs of length j and failed runs of length j and compare these empirical values against the expected values derived using (2) or (3).

We begin by considering a batsman to have had a successful innings if his score is greater than or equal to the median of his scores-to-date. Otherwise, the innings is a failure. We term this simple decision rule D_1. The median is preferred to the mean on the basis of its robustness to not-out scores and extremely high totals – typically batting scores are strongly skewed to the right. The median-to-date is used rather than the final career median (given in Table 3 in the appendix) on the basis that it seems more realistic that a player be evaluated relative to his performance thus far in his career rather than relative to his unknown performance in the future. For completeness, we have also considered success as being measured relative to the average-to-date, and to a fixed value of 40 runs. These results, which are provided in the appendix, are consistent with those obtained using the median, which are presented in this section. A possible weakness of the definition is that the resulting runs of success or failure can only be computed relative to that particular batsman's overall performance thus far in his career i.e. even if a batsman has never scored more than 20, he may still experience 'successful' innings. Nevertheless, we conjecture that many opinions about good or bad form do in fact take into account the overall quality of a batsman, and that the median is thus a reasonable benchmark for success or failure.

A more serious objection is that the definition D_1 does not take into account the fact that some innings are not completed i.e. that some scores are not-out scores. Although one may simply consider that not-out innings are evaluated in the same way as completed innings or that only completed innings affect form perceptions, both these presumptions are patently not true. A more realistic evaluation is that the effect of a not-out score depends on the size of the score already achieved, and in particular is not symmetric about the median. For example, a batsman who scores a small number of runs (e.g. 10), but does not complete his innings because the team achieves the winning or declaring total (or he runs out of batting partners), cannot be held responsible for his low score; but if a batsman scores a high score (e.g. 150) and does not complete the innings, the innings is likely to be judged a success (if anything, a greater success than if the same score had been achieved in a completed innings). There is a sense, therefore, in which a not-out innings can never inflict harm on a batsman's reputation – if he scores a low not-out score, the not-out saves him from failure; if he scores a high not-out score, success is redoubled. Formally, we extend the previous definition of success and failure D_1 so that if a score less than the median is a not-out score, it is not considered either a failure or a success. If a score is greater than or equal to the median, it is considered a success regardless of whether the innings was completed or not. We refer to this decision rule as D_2.

A third definition of success and failure attempts to introduce some well-known psychological realities into our evaluation of batting form. It is a well-established empirical fact that our interpretation of new information is often coloured by our existing beliefs, preconceptions, and expectations (e.g. [8, 4]). Data that is consistent with our beliefs is treated in a qualitatively different way from data that contradicts our beliefs. In the context of batting form, the same score may be viewed as either a success or a failure, depending on whether the batsman is perceived to be in a peak or slump of batting form.[1] We model this filtering of information in the following simple way. The most recent score of a batsman who is perceived to be 'in good form' is not evaluated relative to his current median score, but instead is compared to an adjusted score that is lower than the median and decreasing with the number of innings that he has been in good form for. It thus becomes relatively easier to register further successes as form is perceived to improve. Specifically, the current

[1]Treating incoming information differently based on prior beliefs is a legitimate strategy in many cases where a substantial body of experience has been built up, if done with caution. As Gilovich points out in [8], 'we are justified in allowing our beliefs and theories to influence our assessment of new information in direct proportion to how plausible and well-substantiated they are in the first place'.

score s_i is viewed as a success if it is greater than or equal to $s_M/\sqrt{t_G}$, where s_M is the current median and the quantity t_G measures the number of periods that batsman has been in good form for. Several definitions of t_G are possible; in this paper any scores above the median i.e. 'objectively' good innings according to D_2, increase t_G by one (and hence marginally increase the evaluation of the length of the run of good form), and any objectively poor innings that are viewed with bias as successful ($s_M/\sqrt{t_G} < s_i < s_M$) decrease t_G by one (and hence marginally decrease the evaluation of the length of the run of good form). Analogously, the current score of a player perceived to be in bad form is evaluated against an adjusted score that is higher than the median and increasing with the extent of the run of bad form, t_B. Specifically, the current score is viewed as a failure if it is less than or equal to $\sqrt{t_B}s_M$, where t_B is increased by one if $s_i < s_M$ and decreased by one if $s_M < s_i < \sqrt{t_B}s_M$. We refer to this decision rule as D_3. Though D_3 leaves open the possibility of a batsmen entering a never-ending run of bad form from which it becomes impossible to escape because $\sqrt{t_B}s_M$ becomes so large, this did not occur for any of the selected batsman and the exponent of 1/2 seemed to provide a reasonable picture of the discounting of form over time. Innings that are not completed are treated in the same way as described for D_2.

Table 2 shows the aggregated (over all 16 batsmen) number of runs of successes r_{Sj} and failures r_{Fj} for various run lengths j, for each of the definitions of success and failure established by D_1, D_2, and D_3. The actual number of runs is given first, with the expected number of runs given in parentheses. The final two rows of the table give the results of a χ^2 test of association between the observed and expected frequencies. The results make two important contributions to the literature on the perceptions of sequences of sporting successes – the 'hot-hand'. Firstly, the observed number of runs of successes and failures for

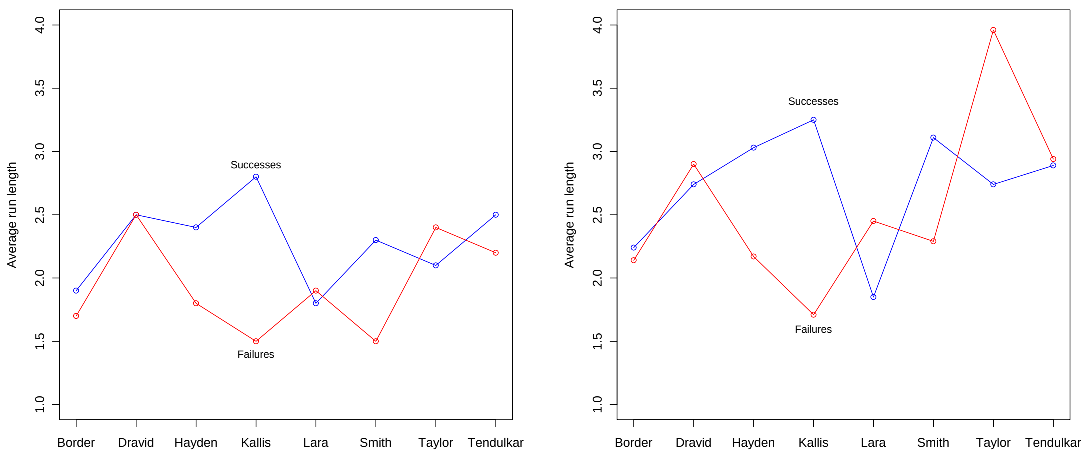
j	$r_{Sj}(D_1)$	$r_{Fj}(D_1)$	$r_{Sj}(D_2)$	$r_{Fj}(D_2)$	$r_{Sj}(D_3)$	$r_{Fj}(D_3)$
1–3	586 (589.9)	601 (617.4)	553 (563.5)	579 (603.9)	433 (566.5)	445 (592.9)
4–6	84 (85.5)	66 (64.8)	86 (88.4)	60 (58)	73 (82.8)	62 (63.7)
7–9	10 (13)	6 (7.1)	12 (14.6)	7 (5.8)	22 (13.7)	16 (7.7)
10–12	4 (2.1)	3 (0.8)	6 (2.6)	2 (0.6)	12 (2.6)	8 (1)
13+	0 (0.4)	0 (0.1)	0 (0.6)	0 (0.1)	5 (0.7)	6 (0.2)
χ^2	2.9	7.0	5.8	4.7	99.4	311.8
p	0.58	0.13	0.21	0.31	< 0.0001	< 0.001

Table 2: Actual and expected number of runs of successes r_{Sj} and failures r_{Fj} for various run lengths j under definitions of success D_1, D_2, and D_3

both D_1 and D_2 is remarkably similar to what would be expected from random sequences of the same numbers of successes and failures. That is, references to a batsman being in good or bad form appear to be highly exaggerated, particularly in the medium-length runs typified by $j = 4, 5$, and 6. There *are* some consistent indications that extremely long runs appear more often than expected, but these cases represent only a tiny proportion of all innings ($2 - 3\%$) and so cannot realistically be considered to the source of public and expert perceptions. Simple statistical tests of association reinforce these conclusions. The findings are supported by the results of similar analyses shown in Table 4 in the appendix, which replace the median s_M with either a batsman's average-to-date $\bar{s}$or a fixed value of 40 runs as a benchmark for defining success and failure. Taken as a whole, the results provide substantial evidence that people overestimate the prevalence and extent of form in sporting sequences, and suggest that most references by expert or armchair commentators to a run of bad or good batting form would be better phrased as entirely natural random fluctuations in fortune.

The second contribution made by the results is to show that it *is* possible to induce very clearly non-random runs of successes and failures by including a simple psychological model of how a cricket fan might view a batsman's current score through the lens of their current perceptions about whether the batsman is already in good or bad form. The average run length increases from 2.04 under D_1 to 2.07 under D_2 to 2.49 under D_3, and the distribution clearly has a heavier right-tail when definition D_3 is used. Previous research on the randomness of sporting sequences has tended toward more sophisticated models, both in terms of analytical procedures (e.g. [1]) and inclusion of further covariates that it was felt might be obscuring the underlying patterns of sporting form that were being perceived by fans and experts alike (e.g. [2]). The difficulty that these studies have had in detecting any definitive general evidence for non-random sequences together with our results indicating the ease with which non-random sequences can be induced by varying definitions of success and failure suggest, firstly, that notions such as 'batting form' and the 'hot-hand' seem at the moment to be far more likely to be *predominantly* perceptual phenomena than actual, empirically observable realities (even if there turns out to be a small observable component); and secondly, that these perceptions might be due more to differently encoding what constitutes a success and a failure than a perception of a raw sequence of scores (either binary hit/miss in basketball or number of runs scored in cricket). These are key points since they shift the direction of research away from the evaluation of actual sequences to the way in which sequences are encoded, and have much in common with the phenomenon of the 'near-miss' often encountered in gambling behaviour (e.g. [19], chapter 5), in which a loss is encoded as a win if it is 'close' to a win (even if this proximity is cognitive rather than statistical; for example a losing lottery ticket consisting of numbers 10, 21 and 35 may be considered a near-miss if the winning numbers are 9, 22 and 36).

A common response to the lack of evidence for any non-randomness in aggregate level data is the claim that the phenomenon of form exists, but only for some players who are particularly prone to playing in streaks (e.g. [11]). This is particularly true of studies in basketball and baseball. While we consider cricket to be less vulnerable to this accusation on the basis that form (as it is usually intended) is supposed to affect all or nearly all batsmen who play for any substantial period of time, it is still interesting to ask to what extent the batsmen in our sample differ with respect to the length of the runs of successes and failures that they go through. Figure 2(a) shows the average length of runs of successes and runs of failures for a subset of 8 batsmen using the definition of success and failure described by D_2, while Figure 2(b) shows the same quantities under the assumption of definition D_3.



(a) "Actual" form (b) "Perceived" form

Figure 2: Average length of runs of successes and failures for a subset of sampled batsmen (a) using definition D_2, (b) using definition D_3

Figure 2(a) shows, firstly, the importance of considering runs of successes and failures separately and, secondly, that some batsmen do show a greater tendency toward longer runs than other batsmen. The need to consider successes and failures separately is well illustrated by the case of Jacques Kallis, who possesses the longest average successful run (2.8 innings) but the shortest average run of failures (1.5 innings), a particularly attractive combination. Other batsmen, notably Allan Border, Rahul Dravid, and Brian Lara, have average run lengths that are more similar for both successes and failures. Among the eight batsmen, Border and Lara seem to be the ones least prone to long runs of either successes or failures, suggesting that they might have 'form' attributed to them less often than batsmen such as Dravid, Mark Taylor, and Sachin Tendulkar, who are subject to longer average run lengths. Though these differences are not enormous, most standard errors are of the order of 0.05 and the differences are thus significant at the 5% level. Also interestingly, runs of successes tend to be longer on average than runs of failures, though the reverse is true for two players: Brian Lara (for whom the difference is non-significant though) and Mark Taylor.

Finally, the average run length of 'actual' successes and failures shown in Figure 2(a) can be compared to the average run length of 'perceived' successes and failures in Figure 2(b). As mentioned previously, perceived runs tend to be substantially longer than the runs obtained using the D_2 definitions. However, the differences tend to be larger for certain batsmen. For example, the perceptions of the types of streaks experienced by some other players (Dravid, Lara, Taylor, and Tendulkar) differ substantially from what the more 'objective' averages indicate, and differ in the systematic way that losses tend to loom larger in the perceived evaluations than they should according to the average run lengths obtained under D_2. Taylor is a particularly strong example of this, and is perceived to be by some way the most form-prone batsmen, suffering from long periods of failure interspersed with moderately long periods of success. In fact, his perceived tendency toward long runs of failures turns out to be due to two runs of 19 and 20 failures. The latter string of failures may be of some interest since it is the longest of any of the streaks considered in this paper,and reads: $7, 25^*, 21, 10, 27, 37, 43, 36, 27, 16, 7, 10, 11, 2, 1, 16, 8, 13, 38, 5, 7,$ where the superscripted star indicates a not-out total. This sequence of failures, which lasted more than 18 months, was finally broken by a score of 129 against England in the first Ashes test of 1997.

4 Discussion and conclusions

In this paper we have considered perceived patterns in the sequences of scores that batsmen have achieved over the course of their careers, and compared these perceived patterns to any that are actually observable in the data. Widespread public and expert opinion suggests that we should find regions of this sequence where scores are generally low, indicating a period of poor batting form, and regions where the scores are generally high, indicating periods of good form. This notion of form can be captured in two statistical concepts: autocorrelation, and non-stationarity. The former can be investigated by examining whether scores in previous innings have any influence on scores in subsequent innings; the latter, by examining whether there are fewer statistical runs i.e. sequences of consecutive failures or consecutive successes, than a model based on a random process would predict. In both cases, we found only very limited evidence of any kind of batting form; specifically, we found that a weighted exponential sum of the last two batting scores had a marginally significant effect on the score in the next innings, and that extremely long runs of successes or failures were consistently but not significantly more prevalent in the observed sequences than would be expected under a hypothesis of randomness. There is thus very little empirical support for the notion of batting form *as it is generally understood*: as a phenomenon that exercises a kind of all-pervading influence on batting scores, affecting all batsmen at frequent intervals in their careers. In contrast, we did find conclusive evidence in favour of other general cricketing beliefs to the effect that batting is significantly more difficult in the second battings innings than in the first, and that scores tend to be higher when playing one of the so-called weaker test-playing nations (Bangladesh, Zimbabwe, New Zealand, and Sri Lanka). These beliefs are almost entirely uncontroversial; but they do draw attention to the fact that the belief in batting form, which is also generally regarded as uncontroversial, belongs to an entirely different class of beliefs, and needs to be treated more critically. This has natural and obvious implications for issues like team selection.

The second contribution of this paper is a more speculative attempt to explain just why it is that this belief is so widespread and pervasive. We showed how patterns that do constitute statistically verifiable runs of good and bad batting form can be induced by a simple psychological model which evaluates each score based on a prior belief that a batsman is already in good or bad form. We can thus show that a belief in form is entirely self-fulfilling: if one believes in form, successes and failures are constructed in such a way that statistically significant patterns really do exist. In summary, we have used this paper to show that one sequential pattern widely believed to exist in fact has no or very little empirical support, and have provided alternate explanations for the perception. The particular context for those beliefs and data has been provided by the sport of cricket, which provides a productive testing ground for research into random sequences and perceptions of those random sequences because data collection is so extensive and beliefs so widely held by experts and laypeople alike. There are a myriad of sequence-based beliefs in cricket, let alone other sports, and other areas of public life. Not all of these areas have the quality of statistical data that allows the underlying beliefs to be investigated and challenged, but the ones that do are invaluable as illustrations of statistical thinking and lapses thereof, and offer significant opportunity for future research.

A Additional tables

Player	Country	Innings	Runs	Not-outs	Average	Median
Allan Border	AUS	265	11184	44	50.6	31
Boeta Dippenaar	SA	62	1718	5	30.1	21.5
Daryl Cullinan	SA	115	4554	12	44.2	27
Rahul Dravid	IND	182	9174	22	57.3	35.5
Herchelle Gibbs	SA	144	5943	7	43.4	25
Matthew Hayden	AUS	159	7739	13	53.0	33
Jacques Kallis	SA	182	8430	29	55.1	34.5
Gary Kirsten	SA	176	7289	15	45.3	24.5
Brian Lara	WI	232	11953	6	52.9	33.5
Virender Sehwag	IND	87	4155	3	49.5	31
Graeme Smith	SA	96	4285	5	47.1	28.5
Mark Taylor	AUS	186	7525	13	43.5	27
Sachin Tendulkar	IND	217	10668	22	54.7	31
Inzamam Ul-Haq	PAK	198	8813	22	50.1	29.5
Mark Waugh	AUS	209	8029	17	41.8	26
Steve Waugh	AUS	260	10927	46	51.1	25.5

Table 3: A summary of key player statistics for the 16 batsmen used in the analysis of batting form

	Replace s_M with $\bar{s}$		Replace s_M with 40	
j	$r_{Sj}(D_2)$	$r_{Fj}(D_2)$	$r_{Sj}(D_2)$	$r_{Fj}(D_2)$
1–3	587 (593.4)	474 (481)	586 (605.2)	503 (529.8)
4–6	34 (31.1)	110 (114.1)	45 (44)	107 (101.6)
7–9	1 (1.6)	27 (26.9)	3 (3.2)	21 (19.5)
10–12	0 (0.1)	4 (6.3)	1 (0.2)	5 (3.8)
13+	0 (0)	5 (1.9)	0 (0)	3 (1)
χ^2	0.6	6.4	3.3	6.3
p	0.96	0.17	0.50	0.18

Table 4: Actual and expected number of runs of successes r_{Sj} and failures r_{Fj} for various run lengths j when performance is compared to the average-to-date or a fixed value of 40 runs

References

[1] J. Albert. Comment on "A statistical analysis of hitting streaks in baseball". *Journal of the American Statistical Association*, 88(424):1184–1188, 1993.

[2] S. Albright. A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association*, 88(424):1175–1183, 1993.

[3] G. Barr and B. Kantor. A criterion for comparing and selecting batsmen in limited overs cricket. *Journal of the Operational Research Society*, 55:1266–1274, 2004.

[4] D. Dearborn and H. Simon. Selective perception: A note on the departmental identifications of executives. *Sociometry*, 21(2):140–144, 1958.

[5] R. Dorsey-Palmateer and G. Smith. Bowler's hot hands. *The American Statistician*, 58(1):38–45, 2004.

[6] W. Elderton. Cricket scores and some skew correlation distributions (an arithmetical study). *Journal of the Royal Statistical Society*, 108(1/2):1–11, 1945.

[7] G. Gigerenzer. The psychology of good judgment: frequency formats and simple algorithms. *Journal of Medical Decision Making*, 16:273–280, 1996.

[8] T. Gilovich. *How we know what isn't so: The fallibility of human reason in everyday life*. Free Press, New York, 1991.

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

[10] A. Kimber and A. Hansford. A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society*, 156(3):443–455, 1993.

[11] P. Larkey, R. Smith, and J. Kadane. It's okay to believe in the "hot hand". *Chance*, 2(4):22–30, 1989.

[12] A. Mood. The distribution theory of runs. *Annals of Mathematical Statistics*, 11(4):367–392, 1940.

[13] L. Nelson. The Shewhart Control Chart – Tests for Special Causes. *Journal of Quality Technology*, 16(4):237–239, 1984.

[14] H. Stern and C. Morris. Comment on “A statistical analysis of hitting streaks in baseball”. *Journal of the American Statistical Association*, 88(424):1189–1194, 1993.

[15] A. Tversky and T. Gilovich. The “hot hand”: statistical reality or cognitive illusion? *Chance*, 2(4):31–34, 1989.

[16] A. Tversky and T. Gilovich. The cold facts about the “hot hand” in basketball. *Chance*, 2(4):16–21, 1989.

[17] A. Tversky and D. Kahneman. Belief in the law of small numbers. *Psychological Bulletin*, 76:105–110, 1971.

[18] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185:1124–1131, 1974.

[19] M. Walker. *The psychology of gambling.* Pergamon Press, Oxford, 1992.

[20] R. Wardrop. Statistical tests for the hot-hand in basketball in a controlled setting. Technical report, University of Wisconsin, 1999.

[21] G. Wood. Cricket scores and geometrical progression. *Journal of the Royal Statistical Society*, 108(1/2):12–40, 1945.