Long Run Predictions

MARK F. SCHILLING

It is June of 1992, and the citizens of riot-torn Los Angeles have gone to the polls to cast their ballots in the local elections. The results are announced, and a judge orders a recount for one particularly closely contested race. Precinct by precinct, each batch of ballots is scrutinized by several officials and witnesses. It seems that one candidate's votes often occur in long runs, with several ballots in a row marked for the same individual. Could there have been sabotage?

Here is the voting record from one typical precinct:

CEEDAACCANAAEAADANNNA
CADDDANCBBADAAAAAECCN
ANABNAAABACNADACBENAN
ECAEBNAEDNNDNBAABAAAA
BABABAANAAAAAAAANADCA
ANANNNAAAAGEAAAAAADCA
AAAADNACABBBCNAAAACCA
DBANCAAAAADAACABNAAAN
ACACAFAABFAABCAAAACAN
FAABFBFDCBAAAAAAENNAA
AAAAADDAAAANDAAAEABANA
ANAAEBNAAAAAABABEECDAC

The seven candidates are indicated by the letters A through G; N represents a ballot with no vote marked. Notice that there are several long runs of votes for candidate A, including a series of eight in a row, seven in a row and six in a row. Do you think that the ballots were tampered with?

The fans cheered with enthusiasm as the Seattle Supersonics of the National Basketball Association trotted onto the court for the start of the 1990–91 basketball season. With talented young players such as Shawn Kemp and Gary Payton complementing an able crop of veterans, hopes were high for a stellar season. Unfortunately, inconsistency marked the team's play and the Supersonics wound up with a mediocre .500 record of 41 wins and 41 losses. Here is the game-by-game record of their performance (W = win, L = loss):

WWWLLLLWLLLLLLWLLWLWW
WWWWLLWWLWLLWLWWLWLL
WLWWLWLLLWLWWWWWLWLL
LLLWWLLLWWWLWWWWWLLWL

Several winning and losing streaks are evident, including one each of length six. Is it fair to say that the Seattle Supersonics of 1990–1991 were an unusually streaky team?

On August 18, 1913, at the famous Monte Carlo casino in Monaco, black came up 26 times in a row on a roulette wheel. As the run on black continued to grow, people began to bet larger and larger sums of money on red in the belief that another repeat of black was virtually impossible. The casino made several million francs that night. Was the wheel fixed? Could a run as long as 26 possibly be expected to occur on an *honest* roulette wheel?

To help answer these questions, let's compare these results to the sorts of runs that tend to occur in truly random sequences. For example, if an ordinary coin is tossed, say 250 times, how long is the longest run of consecutive heads likely to be? A simple argument can give us a rough answer: Except on the first toss, a run of heads can only begin directly after a toss showing tails. There should be around 125 tails in 250 tosses, each providing an opportunity for a head run to start. After about half of these tails the succeeding toss will be heads, giving around 63 head runs in all. Roughly half of the time, the first head will be followed by a second one,

so around 32 runs will be at least two heads long. Again, about half of these will contain at least another head. Thus we can expect around 16 head runs of length at least three, eight runs at least four heads long, four runs of length at least five, two of length six or more, and one run of seven heads or longer.

If many people each toss a coin 250 times, we can therefore expect most of them to obtain a head run of at least seven heads. In fact, precise calculations show that among all strings of 250 coin tosses, 87% will contain head runs of length at least six, 63% will have runs of seven or more, and a substantial 38% will possess runs that are at least eight heads long.

What about long runs of *either* heads *or* tails? Note that we can translate any string of 250 coin tossing outcomes into a new sequence of 249 elements in which we keep a record of whether the outcome of each toss after the first is the same as (S) or different from (D) the previous one. For instance, HTTHTTTHTHH...generates the sequence DSDDSSDDDS...Since D's and S's occur independently with probabil-

MARK SCHILLING is professor of mathematics at California State University-Northridge.

ity 1/2, we can apply the same analysis as above to predict the longest strings of S. The difference between 250 and 249 outcomes has a negligible effect on run lengths; since a string of k S's in a row corresponds to a string of either $k+1$ heads or $k+1$ tails, the results for head runs are simply shifted up by one. For instance, 63% of all sequences of 250 tosses of a fair coin will have *some* run at least eight elements long.

This is a surprising result. In fact, when asked to write down long sequences of heads and tails that would look like a typical random arrangement, most people are quite reluctant to include strings of more than four or five heads (or tails) in a row. It is therefore generally rather easy to distinguish a sequence simulated by a human from an actual random sequence by the failure of most simulated sequences to incorporate long enough runs!

A Formula for the Longest Run

We can easily generalize the informal arguments given above to predict probable run lengths for a sequence of n tosses of a coin in which the chance of heads on each toss, say p, is any value other than 0 or 1. Reasoning as before, there should be approximately $n(1-p)$ tails in the sequence, hence $n(1-p)$ possible starting points for a run of heads. Then about $n(1-p)p$ head runs of length one or more will occur, about $n(1-p)p^2$ head runs of length at least two, and so forth. In general, we will obtain about $N_R = n(1-p)p^R$ runs of length at least R. To find a reasonable value for the typical length of the *longest* run of heads, we can solve the equation $N_R = 1$ for R to obtain

$$R = log_{1/p}(n(1-p)). \qquad (1)$$

In the case of a fair coin ($p=0.5$), this reduces to $R=log_2(0.5n)=log_2(n)-1$, which then gives simply $R=log_2(n)$ for the longest run of heads *or* tails.

We can use (1) to predict typical longest run lengths in a wide range of situations whose structure, like coin tossing, can be modeled (at least approximately) by what statisticians call Bernoulli trials. Bernoulli trials are repetitive sequences with the same two possible outcomes for each event in the sequence. Each event's outcome is independent of the others, and the probabilities remain unchanged from trial to trial.

For example in the voting data, candidate A received $p=51.6\%$ of the $n=252$ votes in the precinct. Formula (1) gives $R=7.3$ votes, quite in line with the actual longest run of eight votes for A. Votes that are *not* for A show similar results, with two runs of seven in a row. So the observed long runs in the ballot data do *not* represent evidence of tampering.

In the basketball data, the chance of a Seattle win undoubtedly varied somewhat from game to game. We shall use $p=0.5$ since the team won exactly half of its $n=82$ games. Formula (1) predicts $R=6.4$ for the longest run of either wins or losses. The actual longest run was six. The Supersonics' 1990–1991 performance was not unusually streaky by this measure.

Figure 1 shows the approximate chances that the longest run will be longer (area to the right of zero) or shorter (area to the left of zero) than formula (1) predicts. The shaded region represents the central 95% of the curve's area.

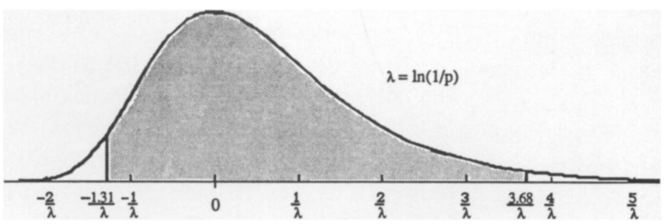

Figure 1. Approximate distribution of the length of the longest run minus its predicted value

Thus, for example, for coin tossing, $p=0.5$ gives $\lambda = ln(1/p) = .69$, so there is about a 95% chance that the length of the longest run of heads will be somewhere between $log_2(0.5n)-1.9$ and $log_2(0.5n)+5.3$. A surprising feature of the curve in Figure 1 is that its scale depends on p but *not* on n (unless n is very small). Remarkably, therefore, one can predict the length of the longest run with the same degree of accuracy for $n=100$ as for $n=1{,}000{,}000$!

Long Run Theory vs. Momentum

Most people attribute streaks of successful or unsuccessful performances in sports to "momentum." For example, a team that has won several games in a row is considered "hot," and is considered more likely to win its next game than its overall record would predict. The opposite kind of momentum is believed to hold for "cold" teams in losing streaks. Similar "hot" and "cold" periods are believed to affect baseball hitters, basketball shooters, and so forth. Many people believe that momentum also applies in gambling, producing both "lucky" and "unlucky" streaks during which betting should be increased or decreased accordingly.

One way to evaluate the case for "momentum" is to compare calculations from long run theory to actual records from human experience, such as those that have occurred in sports and gambling. Perhaps the most famous of these is Joe DiMaggio's accomplishment of managing at least one hit in 56 consecutive baseball games. Was DiMaggio a "hot" hitter whose proficiency during the streak was greatly increased over his normal ability? Or could such a streak have been predicted by long run theory? Although DiMaggio's record was clearly an exceptional accomplishment even after accounting for his superb ability to hit a baseball, the real question we need to ask is whether in the absence of momentum we could have expected anyone in the history of major league baseball to achieve a hitting streak as long as 56 games.

It seems highly probable that the record would be set by a player who is a very good hitter. If we place end-to-end career records of the top 20% of all major league baseball players from 1901 to the present, we obtain a string of about n = 500,000 player-games. (The possibility that the longest run in this list overlaps two different players' career records is fairly remote and will be ignored.) The overall batting average (hitting percentage) of these players is roughly .300; assuming four hitting opportunities per game, the chance that a .300 hitter would get *no* hits in a given game is $(1-.300)^4 = .24$, thus $p = 1 - .24 = .76$. Applying formula (1) gives $R = 43$, while the 95% prediction interval of 38 to 56 just reaches the DiMaggio record.

Similarly adjoining all of the spins of honest roulette wheels in history (estimating a total of half a billion spins) predicts a longest run of $R = 27$ of the same color. The remarkable run at Monte Carlo is in fact quite reasonable when viewed in this context.

Table 1 presents several prominent record streaks in sports and gambling and compares them to predictions derived from the long run theory above. In most cases, the figure used for p is an average, as the value typically varies among the trials. The values of n are also approximate, in some cases "ballpark figures" obtained from rough calculations. However, moderate changes in n do not greatly alter the predictions.

Event	Record	p	n	Predicted Longest Run	95% Prediction Interval
Baseball					
Consecutive games with hit (top 20% of batters)	56 (DiMaggio, 1941)	.76	500,000	43	(38, 56)
Consecutive hits (top 20% of batters)	12 (Higgins,1938; Dropo, 1952)	.30	2,000,000	12	(11, 15)
Consec. wins, team (top 20% of teams)	26 (NY Giants, 1916)	.60	55,000	20	(17, 27)
Consec. wins, pitcher (top 20% of pitchers)	24 (Hubbell, 1936-7)	.62	40,000	20	(17, 28)
Basketball					
Consec. wins, team (teams winning >70%)	33 (L.A. Lakers, 1971-2)	.73	6,000	23	(19, 35)
Consec. losses, team (teams winning <30%)	24 (Cl. Cavaliers, 1982)	.73	6,000	23	(19, 35)
Free Throws	97 (Williams, 1993)	.90	40,000	79	(66, 114)
Roulette Same Color	26 (Monte Carlo, 1913)	.48	5×10^8	27	(26, 32)
Craps Consec. passes	28 (Las Vegas, 1950)	.49	5×10^8	27	(25, 33)

Table 1

The record streaks in the sports categories of Table 1 are in some cases somewhat above the length predicted from formula (1), but all lie within the 95% prediction intervals. The runs in gambling, while at first glance amazing, are completely in line with what is expected for run lengths in very long strings of independent Bernoulli trials. Although a run of 26 in a row in roulette, for example, has only a 1 in 68,411,592 chance of occurring in a specified set of 26 rolls at Monte Carlo, our analysis has shown that it is quite a reasonable thing to have occurred on *some* wheel, at *some* time.

The data therefore refute the idea of momentum in games of chance, while giving less conclusive results for baseball and basketball. Although it can be argued that player attitudes and emotions must surely cause significantly longer runs in sports than those expected by chance alone, the empirical support for this claim is weak. Detailed statistical analyses indicate, in fact, that contrary to the strong prevailing opinions of fans, sports reporters, and the players themselves, momentum may be merely an illusion of the human mind [1-4].

The almost universal tendency to regard long runs as having underlying, non-random causes is probably due to the selectiveness of human perception, which is *geared* towards pattern recognition—streaks tend to stand out and be remembered—coupled with an unawareness of the extent to which long runs arise by chance alone in merely *random* progressions of data. When a randomly generated sequence is placed side by side with an *actual* sequence reflecting human performance, gambling outcomes, etc., the patterns in the two sequences are quite often indistinguishable. Additional information on the theory of longest runs can be found in [5] and in the references cited therein. ■

References

1. Albright, C. (1992). "Streaks and slumps," *OR/MS Today*, vol. 19, 94-95.
2. Albright, S.C. (1993). "A statistical analysis of hitting streaks in baseball (with discussion)," *Journal of the American Statistical Association*, vol. 88, no. 424, 1175-1183.
3. Gilovich, T. and Tversky, A. (1989). "The cold facts about the hot hand in basketball," *Chance: New Directions for Statistics and Computing*, vol. 2, no. 1, 16-21.
4. Gilovich, T. and Tversky, A. (1989). "The 'hot hand': statistical reality or cognitive illusion?," *Chance: New Directions for Statistics and Computing*, vol. 2, no. 4, 31-34.
5. Schilling, M.F. (1990). "The longest run of heads," *The College Mathematics Journal*, vol. 21, no. 3, 196-207.