Is a basketball free-throw sequence nonrandom? A group exercise for undergraduate statistics students

Stephen C. Adolph
Harvey Mudd College, California, U.S.A.

Key Words: runs test, binomial, transition probabilities, contingency table, sequential independence, randomization

Abstract

I describe a group exercise that I give to my undergraduate biostatistics class. The exercise involves analyzing a series of 200 consecutive basketball free-throw attempts to determine whether there is any evidence for sequential dependence in the probability of making a free-throw. The students are given the exercise before they have learned the appropriate statistical tests, so that they can come up with ideas on their own. Students spend a full class period working on the problem, with my guidance and hints. In the next class period, we discuss how each student group approached the problem. I then present several alternative ways to analyze the data, including a runs test and a contingency table analysis of transition frequencies.

1. Introduction

Analyzing sequences of discrete and categorical data is important in many fields, including diverse areas of biology. This is particularly true in the rapidly emerging field of genomics, which involves statistical analysis of DNA sequences and related problems (Claverie and Notredame 2003). As another example, animal behavior researchers often analyze discrete time sequences of behaviors (Cane 1978; Haccou and Meelis 1992). The simplest sequential data set involves a series of Bernoulli trials, such as success or failure on a task, choice of one alternative vs. another, and whether it rained or not on each of a series of days.

Several statistical methods are available for testing sequences of dichotomous data for sequential independence (randomness). However, because this kind of data set is simple, it offers beginning statistics students the opportunity to brainstorm about possible analyses *before* they have learned the formal methods. In this paper, I describe an in-class exercise in which students are given a sequence of dichotomous data and asked to try to determine whether the sequence is random or not (i.e., whether each event is independent of previous events). I then summarize some of the common approaches taken by students, and present several complementary analyses of the data.

2. Class exercise

2.1 Level and background of the students

My Biostatistics course typically has 20-30 students, mostly college juniors and seniors. Most of the students are either Biology or Mathematical Biology majors; other majors have included Environmental Studies, Chemistry, Neuroscience and Mathematics. The mathematical background of the students is mixed. About half of the students have two years of college mathematics, including one semester each of multivariate calculus, linear algebra, ordinary differential equations, and introductory computer science, and a half-semester introduction to probability and statistics. The others have typically taken calculus but no other college mathematics, statistics, or computer science courses. However, the students who have less mathematical background generally have had more experience with collecting and analyzing field and laboratory data in their biology courses. This diversity in academic backgrounds leads to a wider and more interesting variety of approaches to this exercise and to the course in general.

2.2 Course topics prior to this exercise

I give this exercise about halfway through the 14-week semester. At this point the students are familiar with standard introductory concepts in frequentist statistics, including estimation, sampling distributions, and hypothesis testing, and have done both parametric and non-parametric analyses of continuous and discrete data. They have worked with one-, two-, and paired-sample tests. They are also familiar with the basic idea of randomization tests, and have done some simple randomization tests themselves.

During the week prior to this exercise, the students have been analyzing data involving discrete categorical variables, including work with binomial and Poisson distributions, contingency tables, and goodness-of-fit tests. We do this exercise before the students have learned any methods for analyzing temporal sequences of data (e.g., runs tests) because I want them to experience this as a novel problem.

2.3 The exercise

We use a full 50-minute class period for this exercise. I give the students a one-page handout showing a sequence of dichotomous data (Fig. 1) and a series of questions. The data represent a sequence of 200 consecutive basketball free-throw attempts, labeled as successes and failures. The data are real, although some students are surprised that the success rate was exactly 70% (140/200). To obtain the data, I attempted 200 consecutive basketball free-throws under standard conditions (basket rim 18 inches in diameter and 10 feet above the ground; horizontal distance from free-throw line to basket 15 feet; leather basketball 9 inches in diameter).
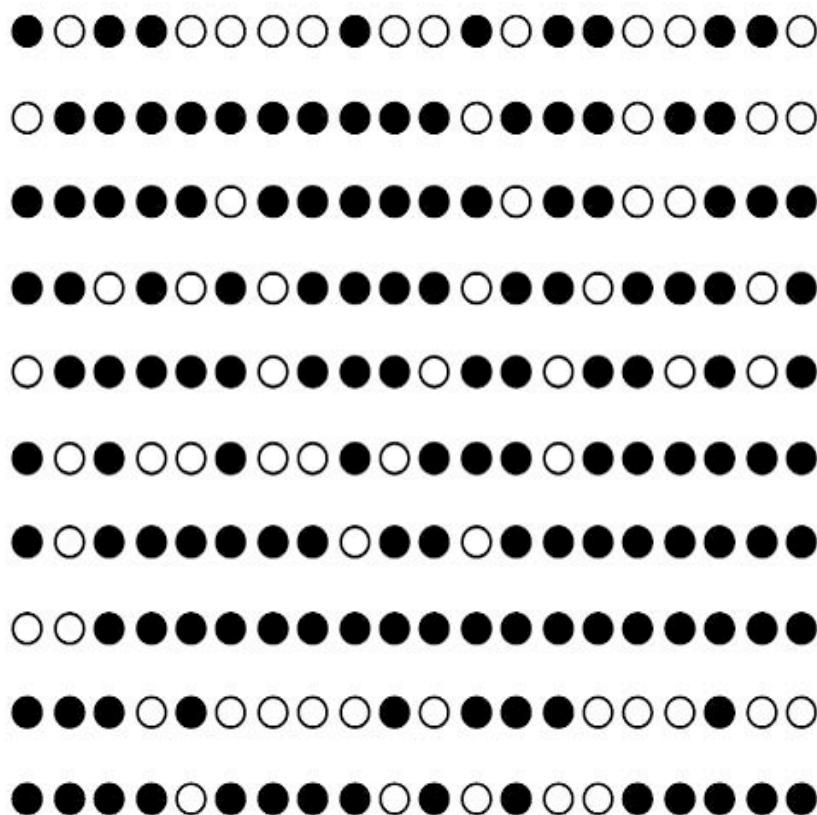
Figure 1. Sequence of 200 consecutive basketball free-throws by a single shooter. Black circle = success, open circle = failure. Sequence begins at upper left, and each row is read from left to right.

The students work in groups of 3-4. To guide their discussions, I encourage them to consider the following questions:

a. How might you determine whether there is a pattern to this sequence?

b. What is your null hypothesis?

c. Can you think of a statistic whose value would vary depending on whether there was a sequential pattern?

d. Is there more than one type of non-random, sequential pattern one could observe with data like these? (i.e., is there more than one way the data could deviate from a random sequence?)

While the students are working on the problem, I circulate around the room and interact with each group; I offer hints, feedback, and further questions. I encourage them not to try to apply any specific methods they have already learned, but instead to think about the problem from a fresh perspective. Because of this goal, the students are not allowed to consult their textbooks during the exercise, except to look up critical values for statistical tests they are already familiar with.

During the last 10-15 minutes of the class period, I ask each group to briefly describe to the rest of the class how they have approached the problem. In the four years that I have given this exercise, students have commonly taken one of several approaches:

a. Initially, some groups of students are tempted to test the hypothesis that the binomial probability of a success $p = 0.5$. This reflects the students' tendency to apply a familiar test to a new situation, even when this approach might not be ideal or appropriate. Another possibility is that they think that "random" means

totally unpredictable, which they identify with $p = 0.5$ (I thank an anonymous reviewer for suggesting this point). Usually, the students who initially consider this approach quickly recognize that this is not a particularly interesting question for this data set, although it is a simple test to perform. These students then move on to other ideas.

b. Many students are interested in the conspicuous streak of 21 successful free-throws. Some calculate the probability of such a streak based on the binomial parameter $p = 0.7$, which is indeed low: 0.00056. This presents an opportunity for discussing the likelihood of observing individually rare events within a larger set of observations. A focus on streaks often leads students to calculate the probabilities of streaks of different lengths (for both successes and failures), with the goal of comparing these to the observed data. Interestingly, while students often think of calculating the probability of a streak of a given length, they rarely consider the much simpler approach of counting the number of runs of any length.

c. Some students observe that there appear to be periods within the 200 trials that have relatively low success rates and other periods with high success rates. This usually leads them to divide the data into blocks of 5, 10 or 20 attempts, then test for heterogeneity among the blocks. This is the one analysis that yields a statistically significant result for this data set, although it does not address the central question of the exercise. Once students are headed down this path, I encourage them to finish, because it provides an opportunity to discuss the assumption of a constant (stationary) value of the binomial parameter p, and how a violation of this assumption would affect the occurrence of streaks.

d. Occasionally students will identify the possible extreme arrangements of these data, such as 60 consecutive failures followed by 140 consecutive successes. The other extreme involves a large number of very short runs (frequent alternation of successes and failures). When students have identified these extremes, the next step is to guide them towards the idea of a runs test.

e. On two separate occasions, students recognized that the sequence of successes and failures could be converted to four classes of transitions: make-make, make-miss, miss-make, and miss-miss. (One of these students had taken a course in stochastic processes, while the other came up with this approach on her own.) These students and their partners then tabulated the number of each type of transition. With a few hints, these groups then figured out that a standard test for independence for a contingency table was an appropriate test of the null hypothesis of temporal independence.

3. Statistical analyses

During the next class period, we revisit this problem. I recap the ideas that students came up with, then present the following alternative ways one could analyze these data. I try to connect these methods to the ideas that students developed, and provide a handout that summarizes these methods.

3.1 Runs test (including randomization)

The simplest analysis for temporal independence of binary data is the runs test (e.g., [Zar 1998](#)). In a runs test, each string of one or more of the same values is counted as a run. Here are the free-throw data converted into a sequence of runs (1 = one or more consecutive successes, 0 = one or more consecutive failures):

1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1

0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0

1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1

Note that there are 85 runs.

Our sample size is too large for the tables of critical values given by our course textbook so we use a normal approximation for u, the number of runs (Zar 1998, p. 584). The hypothesized mean number of

runs (under the null hypothesis that each trial is independent) is given by

$$\mu_u = \frac{2n_1n_2}{N} + 1$$

where n_1 is the total number of successes in the sample, n_2 is the total numbers of failures, and $N = n_1 + n_2$.

The hypothesized standard deviation under the null hypothesis is

$$\sigma_u = \sqrt{\frac{2n_1n_2(2n_1n_2 - N)}{N^2(N-1)}}$$

We then construct the test statistic

$$Z = \frac{|u - \mu_u| - 0.5}{\sigma_u}$$

and compare it to the standard normal distribution. For the free-throw data, $n_1 = 140$, $n_2 = 60$, $N = 200$, and $u = 85$. Therefore $\mu_u = 85$, $\sigma_u = 5.919$, and $Z = 0.084$ ($P = 0.933$; two-sided). Thus, 85 is not an unusual number of runs (in fact, 85 is the expected number), and we can conclude that this sequence does not show serial dependence.

An alternative way to assess the significance of the test statistic u is to use randomization to obtain the distribution of u under the null hypothesis (Manly 1997). During the semester, I frequently ask the students to imagine a randomization procedure that would be an appropriate alternative to a given statistical test. Therefore, these students are familiar with the idea of randomization, and have performed several simple randomization tests as exercises.

The 140 successes and 60 failures can be randomly shuffled, and the resulting number of runs counted. One set of 10,000 trials gives the runs distribution shown in Fig. 2.
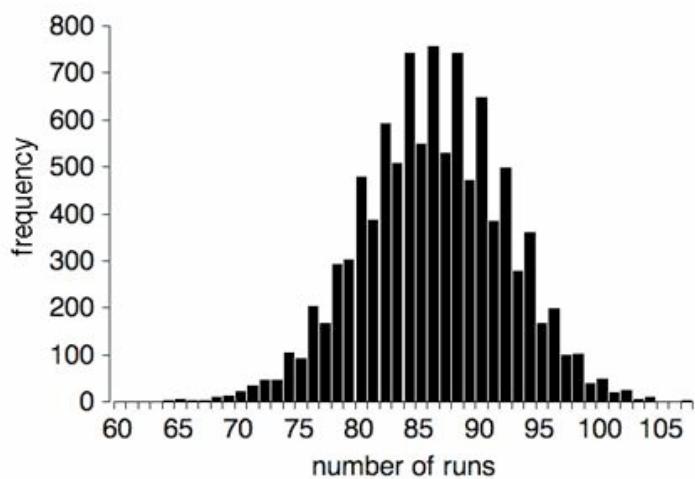
Figure 2. Frequency distribution of the number of runs, u, obtained by randomly rearranging the data shown in Fig. 1 (10,000 replicates).

The number of runs in the randomized samples ranged from 60 to 107, with a mean of 86.0 (median = 85), yielding $P = 1.0$ (2-sided) for our data. Thus, our result from the randomization procedure is essentially identical to the result from the normal approximation, in that the observed number of runs is very close to the center of the sampling distribution of the test statistic (under the null hypothesis). It is appropriate to assume a 2-sided alternative hypothesis, since successive free-throws could potentially be either positively correlated (resulting in fewer, longer runs) or negatively correlated (resulting in more, shorter runs). Students could perform this randomization procedure themselves, although I have not yet had my students do this.

One advantage of presenting the randomization approach is that students can see the whole distribution of runs under the null hypothesis. A possible extension to this exercise would be to have the students construct their own rearrangements of the data that exhibit either positive or negative serial dependence. They could then compare their number of runs to the null distribution in Fig. 2.

3.2 Longest-run analysis

Although I have not yet included this topic in my course, the run of 21 successes could be analyzed as a "longest-run" problem (e.g., Schilling 1990, Berresford 2002). Schilling (1990) reported that the longest run of successes, R_n, can be roughly approximated as

$$R_n \approx \log_{1/p}(nq).$$

For the free-throw data, assuming that $p = 0.7$, $R_n = 11.48$. Using the first three terms of Schilling's more detailed expression (Schilling 1990, eq. 5) yields $R_n \approx 12.58$. Schilling also gives an expression for the variance of the sampling distribution of longest runs of successes. These analyses would be interesting to students with sufficient background in probability theory.

For students without this background, such as most of my biostatistics students, the longest run problem could be explored via simulation or randomization methods. For example, students could write programs that randomly shuffled sequences of 140 successes and 60 failures and kept track of the longest streak of successes in each sequence. The resulting distribution of longest streaks could then be used to evaluate the probability of observing a run of 21 or more successes under the assumption that the trials are independent. I ran such a program and observed 193 streaks of 21 or more successes out of 10,000 trials; this corresponds to P = 0.0193. The mean longest streak was 12.30, close to Schilling's approximation.

3.3 Transitions (Markov analysis)

Another method of testing for sequential independence is to examine **transitions** from one value to the next. There are four possible transitions for free-throws: success to success, success to failure, failure to success, and failure to failure. If successive observations are independent, then P(success | success) = P(success | failure), and P(failure | success) = P(failure | failure).

These transitions are summarized in Table 1.

Table 1. Observed transitions in the series of free-throws shown in Fig. 1.

			To	
			Success	**Failure**
From	**Success**		97	42
	Failure		42	18

Note that there are 199 transitions, rather than 200.

Under the null hypothesis, the rows should be in the same proportions (within sampling error). Thus, this table can be analyzed as a 2×2 contingency table, using tests that the students are already familiar with. The expected values are shown in Table 2.

Table 2. Expected number of transitions under the null hypothesis that each free-throw is an independent event.

		To	
		Success	**Failure**
From	**Success**	97.09	42.91
	Failure	41.91	18.09

The expected values are almost identical to the observed values, yielding a very small chi-square value of 0.00093 ($P = 0.976$, 1 $d.f.$). Both the contingency table analysis and the runs test lead to the same conclusion: successive free-throws are independent.

This is the first exposure to stochastic processes for most of my students. If we had found that the probability of a making a free-throw was dependent only on the result of the previous free-throw, then the sequence would be an example of a Markov chain. In that case we could have estimated these conditional probabilities using Table 1. Bishop et al. (1975) provide extensive discussion and many examples of the construction and statistical analysis of transition matrices, including some biological applications.

Although we do not explore these methods in my course, I like to point out to my students that Markovian analyses have a wide range of applications in biology. One example is the analysis of nucleotide sequence data: one can analyze the frequencies with which A follows A, C, T, and G in a sequence, etc., and use this information for a variety of purposes. Similarly, one can analyze evolutionary changes in nucleotide base pairs or amino acids at a particular site using a Markovian approach (for examples of applications see Claverie and Notredame 2003; Allman and Rhodes 2004). In animal behavior, researchers have used Markov models to describe the temporal structure of bird songs and other types of behaviors (Dobson and Lemon 1979; Haccou and Meelis 1992).

3.4 Heterogeneity in *p* over time

The previous analyses addressed whether successive free-throws were independent. These analyses implicitly assumed that the overall probability of making a free-throw was constant (stationary) over the 200 attempts. However, *p* might change over time; it could increase as a result of experience, decrease due to fatigue, or simply fluctuate. We can test for stationarity of *p* by breaking up the data into blocks, and testing for heterogeneity in the frequency of successes vs. failures. Table 3 shows the data in blocks of 20.

Table 3. Outcome of free-throw trials broken into successive blocks of 20 (using data shown in Fig. 1).

Successes	Failures
9	11
15	5
16	4
14	6
14	6
13	7
17	3
18	2
9	11
15	5

Under the null hypothesis that $p = 0.7$, the expected frequencies for each block are 14 successes and 6 failures. The resulting chi-square value is 19.524 ($P = 0.021$, 9 $d.f.$). Thus, we can reject the null hypothesis that p is stationary over the 200 trials. Interestingly, p appears to fluctuate rather than exhibiting a steady increase or decrease.

3.5 Alternative data sets that exhibit serial dependence

Students find it helpful to see alternative arrangements of the data that are clearly non-random. Suppose that we observed this sequence, where $1 =$ success and $0 =$ failure:

1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1 1 0 0 0

(repeated 10 times)

Like the original data, this sequence has $n_1 = 140$, $n_2 = 60$, and $N = 200$. However, it has fewer runs: $u = 40$. (Fewer runs means they must be longer on average.) This value of u is very extreme: using the normal approximation, $Z = 7.52$, which is far out in the tail of the standard normal distribution ($P \ll 0.0001$). Similarly, in 10,000 runs of the randomization procedure (Fig. 2) the smallest value of u was 59, which occurred only once; 40 is much more extreme than 59. Thus, for this artificial data set we can reject the null hypothesis of serial independence. Instead, there is a strong pattern of serial dependence, which in this case is positive (a success is more likely following another success than it is following a failure).

On the other hand, consider this sequence:

1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 0

(repeated 10 times)

Again, there are the same number of successes and failures as the original data, but in this case $u = 120$ runs, yielding a value of $Z = 5.83$ (P < 0.0001; see also Fig. 2). So this sequence also exhibits significant serial dependence, but this time the dependence is **negative**: the probability of a success following a success (0.57) is lower than the overall probability of a success (0.7). Negative serial dependence causes many short runs.

Although the preceding two examples clearly illustrate the two kinds of serial non-independence, they are obviously highly artificial. Therefore, I also simulated a more realistic data set that exhibits positive serial dependence. This provides the students an opportunity to obtain a statistically significant result. I generated sequences of 200 free throws using the following conditional probabilities:

P(success | success) = 0.76

P(failure | success) = 0.24

P(success | failure) = 0.56

P(failure | failure) = 0.44

These conditional probabilities yield the same overall probabilities as the values estimated from the original data: P(success) = 0.7 and P(failure) = 0.3.

Here is one such sequence that has the same number of successes and failures as the original data set:

1 1 0 1 0 0 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 0 0 0

0 1 1 0 1 1 1 0 0 0 1 1 1 1 1 1 0 1 1 1 0 1 0 0 1

0 0 1 1 1 1 1 0 0 0 0 1 1 0 0 1 1 1 1 1 1 0 0 1 1

1 1 0 1 1 1 1 1 1 1 0 1 0 0 0 1 1 0 0 1 0 1 0 0

0 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 0 1 1 0 1 1

1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 0 1 1 1 1 0 1 1 1

1 1 1 1 1 0 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1

0 0 1 0 1 1 1 0 0 0 1 1 0 1 1 1 1 1 1 1 1 0 0 1

For these data, the number of runs = 69, which leads to $Z = 2.6187$ (P = 0.0088).

The transitions are shown in Table 4.

Table 4. Transitions in the simulated series of free-throws shown above.

		To	
		Success	**Failure**
From	**Success**	105	34
	Failure	34	26

Testing this table for independence of rows and columns yields a chi-square value of 7.0881 ($P = 0.0078$). So, both the runs test and the contingency table analysis detect the positive serial dependence in these simulated data. On the other hand, these data do not exhibit temporal heterogeneity in p: dividing the sequence into blocks of 20 and testing for heterogeneity yields a chi-square value of 13.33 ($P = 0.148$).

4. Conclusion

This data set is simple but allows a variety of different approaches. Students find the exercise to be conceptually challenging while at the same time mathematically simple. It is therefore ideal for my introductory class. Students are focused during the exercise, and the classroom atmosphere is lively. The exercise demonstrates that some problems may have more than one legitimate statistical approach, and that ideally these complementary analyses lead to similar conclusions. During the follow-up discussion they enjoy seeing the problem analyzed from several perspectives.

The exercise and subsequent discussion involve several fundamental concepts in statistical inference, including independence, hypothesis testing, normal approximations, and randomization. I feel that my students have a stronger understanding of each of these concepts after this exercise, although I have not yet formally assessed their learning.

The data are easy to collect, and I encourage other teachers to obtain their own data, perhaps from student volunteers. Students may also be interested in analyzing free-throw or field goal sequences of individual professional basketball players. These data can be obtained by examining the complete play-by-play records for each game that are available on the National Basketball Association's web site (www.nba.com). Gilovich et al. (1985) describe such an exercise (see also Tversky and Gilovich 1989). Other types of data, such as whether or not it rained on a given day, may be more likely to exhibit serial dependence.

5. Getting the Data

The file freethrows.dat contains the raw data. The file freethrows.txt is a documentation file containing a brief description of the dataset.

Acknowledgments

References

Allman E. S., and Rhodes, J. A. (2004), *Mathematical Models in Biology: An Introduction*, Cambridge: Cambridge University Press.

Berresford, G. (2002), "Runs in coin tossing: Randomness revealed," *College Mathematics Journal*, 33, 391-394.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge, Massachusetts: The MIT Press.

Cane, V. R. (1978), "Fitting low-order Markov-chains to behavior sequences," *Animal Behaviour*, 26, 332-338.

Claverie, J.-F., and Notredame, C. (2003), *Bioinformatics for Dummies*. New York: Wiley.

Dobson, C. W., and Lemon, R. E. (1979), "Markov sequences in songs of American thrushes," *Behaviour*, 68, 86-105.

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

Haccou, P., and Meelis, E. (1992), *Statistical Analysis of Behavioural Data: An Approach Based on Time-Structured Models*, Oxford: Oxford University Press.

Manly, B. F. J. (1997), *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 2nd ed., London: Chapman and Hall.

Schilling, M. (1990), "The longest run of heads," *College Mathematics Journal*, 21, 196-207.

Tversky, A., and Gilovich, T. (1989), "The cold facts about the 'hot hands' in basketball," *Chance*, 2, 16-21.

Zar, J. H. (1998), *Biostatistical Analysis*. 4th ed., Upper Saddle River, New Jersey: Prentice Hall.

Stephen C. Adolph
Department of Biology
Harvey Mudd College
301 Platt Blvd.
Claremont, CA 91711
U.S.A.
adolph@hmc.edu