Does Your iPod *Really* Play Favorites?

Amy G. FROELICH, William M. DUCKWORTH, and Jessica CULHANE

Since the introduction of the first iPod portable music player (MP3 player) by Apple, Inc., users have questioned the randomness of the shuffle feature. Most evidence cited by users claiming to show nonrandom behavior in the shuffle feature is anecdotal in nature and not based on any systematic analysis of its randomness. This article reports on our attempt to investigate the shuffle feature on the iPod and to test its randomness through the use of probability and statistical modeling. We begin by reviewing the research on people's inability to perceive and understand both random and nonrandom behavior. Probability models are then developed, under the assumption of a random shuffle, for several of the most common types of events cited as evidence of a nonrandom shuffle. Under this null hypothesis of a random shuffle, several goodness-of-fit tests of one of the probability models are conducted using data collected from real iPods. No evidence to support user claims of a nonrandom shuffle was found. Finally, we conclude with some reflections on and ideas for incorporating these examples into undergraduate probability and statistics courses.

KEY WORDS: Goodness-of-fit test; Probability models; Randomness.

1. INTRODUCTION

Since the introduction of the first iPod portable music player by Apple, Inc. in October 2001, the small device has become a huge social phenomenon. At the original iPod product launch, Steve Jobs, CEO of Apple, Inc. stated "... iPod, a thousand songs in your pocket. This is a major, major breakthrough." (Levy 2006, p. 9). One of the amazing aspects about storing 1000 songs in your pocket is the ability to become your own disc jockey. This aspect is further enhanced by a feature built into the iPod software called "shuffle." The shuffle feature takes a list of songs, called a playlist, and rearranges them in a random order.

Since its introduction, users have questioned the randomness of the shuffle on the iPod. Most notable is an article by Steven Levy from *Newsweek* magazine titled "Does Your iPod Play Favorites?" (Levy 2005). In his article and subsequent book (Levy 2006), he reported anecdotal evidence of potential nonrandom behavior in the shuffle feature when using his iPod. A study of many iPod and technology related websites shows the same results; people believe the random shuffle feature on the iPod is not *really* random. Adding to the controversy is the refusal of Apple, Inc. to release the code used to produce these random permutations. On some websites, this controversy has become a full-blown conspiracy.

After reading the *Newsweek* article and other sources, we were very skeptical of the reported evidence of nonrandom behavior of the shuffle feature. As statisticians, we are very familiar with people's inability to understand randomness. There are many examples in the literature of how people's intuitive ideas of probabilities do not match reality. Songs also evoke emotions, which play a role in our inability to recognize random behavior.

With this background, we decided to look specifically at some of the reported evidence of nonrandom behavior and develop probability models for these events under the assumption of a random shuffle. We then used data collected from our own iPods to conduct goodness-of-fit tests under the null hypothesis of a random shuffle for one of these probability models. In Section 2 of this article, we provide a summary of the current research on the psychology of understanding random and nonrandom events. In Section 3, we report on the development of the probability models and then study some of the probabilities for events reported in the *Newsweek* article and book by Steven Levy. The results of several goodness-of-fit tests for one of these probability models are included in Section 4 of this article, along with the collected data. In Section 5 below, we show how these examples can be used and incorporated into undergraduate probability and statistics courses, and in Section 6, we offer final conclusions.

2. THE PSYCHOLOGY OF UNDERSTANDING RANDOMNESS

The concept of randomness is not easily understood. Most people feel that any randomly produced series should contain very few, if any, extended runs of the same event and should represent the long run expected frequencies of events. In actuality, randomly generated series may not display either of these properties in the short run. When iPod users notice that their shuffle feature seems to violate these properties by choosing

Amy G. Froelich is Associate Professor, Department of Statistics, Iowa State University, 3109 Snedecor Hall, Ames, IA 50011-1210 (E-mail: *amyf@iastate.edu*). William M. Duckworth is Associate Professor of Decision Sciences, Creighton University, College of Business, 2500 California Plaza, Omaha, NE 68178 (E-mail: *wmd@creighton.edu*). Jessica Culhane is Actuarial Analyst, Milliman, 1301 Fifth Avenue, Suite 3800, Seattle, WA 98101 (E-mail: *jessica.culhane@milliman.com*). Portions of this manuscript were completed as part of the third author's senior honors project under the direction of the first author at Iowa State University. The authors wish to thank Dr. W. Robert Stephenson for his assistance on the probabilities in Section 3.3 and the Editor, Associate Editor, and two anonymous reviewers for their helpful comments on earlier versions of this manuscript.

 DOI: 10.1198/tast.2009.07073

three songs in a row from the same album or five songs out of ten from the same artist, they conclude the shuffle is not random.

An aversion to runs is shown in studies where participants are asked to generate random binary series, like coin tosses. The series they create tend to have too many alterations and too few runs of two or three successes than what would be expected from a random process (Bakan 1960; Orr, Federspiel, and Maxwell 1972; Diener and Thompson 1985). For an infinite number of coin tosses, the limiting proportion of heads equals the limiting proportion of tails. Many people also expect this to be true over the short run, that is, over relatively few tosses of a coin. The belief is that occurrences in the short run different from this expectation should quickly correct themselves. This is an example of the gambler's fallacy, the belief that past occurrences of a random event will influence its future occurrences. So, if a gambler has tossed ten heads in a row, he may believe that tails are "due," or conversely, he may feel that heads are "hot."

When a few series do not represent the expected frequencies of events, people tend to label the event a "coincidence." The more personally meaningful or seemingly improbable a coincidence is, the more surprising it will seem (Falk 1989). With music, "personally meaningful" could mean anything from the first song at your wedding to the least favorite song in your library. So when a personally meaningful song is played in the first ten songs of a shuffle, it is more surprising than if the song had no personal meaning. If an event is more surprising, we have a better chance of remembering it, and, therefore, these surprising events will seem to occur too often to be random.

One example of the relative surprise of coincidences is commonly referred to as the birthday paradox. When asked, "What is the probability that two or more people out of a group of 25 will share the same birthday?" most people find it hard to believe the answer is more than 50%. This is so surprising because we tend to focus on our own birthday, which has personal meaning. Therefore, the question becomes "What is the probability that someone in a group of 25 people will share **my** birthday?" In this case, the chance is much smaller–less than 7% (Bennett 1998).

Because randomness and probability are counterintuitive to many people, it is a very difficult concept to teach. People cannot help but remember surprising coincidences, such as a shuffle with six songs in a row that begin with the letter "D." As Paul Kocher, CEO of Cryptography Research, notes in Levy's book, "Our brains aren't wired to understand randomness–there's even a huge industry that takes advantage of people's inability to deal with random distributions. It's called gambling." (Levy 2006, p. 191).

3. PROBABILITY MODELS FROM A RANDOM SHUFFLE

In this section, we develop several probability models for different outcomes from a shuffle under the assumption the shuffle feature is truly random. Most of the probability models can be found in standard textbooks like Ross (2006) and Wackerly, Mendenhall, and Scheaffer (2007).

The shuffle feature of any digital music player works by taking a collection of songs, called a playlist, and producing a random permutation. Each song will appear in the shuffled playlist only once and each shuffle of a playlist is independent from all others. This type of shuffle contrasts with a physical shuffling of objects (like a riffle shuffle of a deck of cards). See Baker and Diaconis (1992) for results on the randomness of the riffle shuffle.

Let N denote the number of songs in a playlist. The number of songs in the shuffled playlist the user listens to before selecting a different playlist or reshuffling the same playlist will be denoted as n. While $n \leq N$, n is almost always strictly less than N for longer playlists. Each song in a N song playlist can be classified according to one of several groupings, including Artist, Album, Date When Song was Added to iTunes, whether or not Song was purchased from the iTunes Store, etc. For a particular grouping (Artist, Album, etc.), let G_i denote the number of songs in the playlist belonging to the ith group and g denote the number of groups.

Assuming a random shuffle, the probability that any of the $N!$ shuffles will occur is $1/N!$. However, if the user only listens to the first n songs in the shuffled playlist, the probability that any of the $N!/(N-n)!$ shuffles will occur is equal to $(N-n)!/N!$.

3.1 Length of Time in One Shuffle Before a Particular Group Occurs

Both in the *Newsweek* article and in his book, Steven Levy describes a phenomenon he calls the Length of Time Before Steely Dan (LTBSD) Factor. At the time, Levy's iPod contained approximately 3,000 songs with approximately 50 of these songs belonging to the artist Steely Dan. To him, his first iPod seemed to prefer this musical group over other artists. He noted this by observing the first Steely Dan song appeared to occur very early in a shuffle, much earlier than he would have expected.

Assuming a random shuffle, the probability the first song of a shuffle would belong to the ith artist is equal to G_i/N. Let $X_j = 1$ indicate the jth song in the shuffle is from the ith artist and 0 otherwise and let T be defined as the minimum value of j so that $X_j = 1$. Then the distribution of T is negative hypergeometric (with $r = 1$) (Miller and Fridell 2007) so that

$$P(T = t) = \frac{\binom{N-t}{G_i-1}}{\binom{N}{G_i}}, \qquad t = 1, 2, \ldots, N - G_i + 1. \qquad (1)$$

Assuming 50 songs from Steely Dan in a playlist of $N = 3{,}000$ songs, in approximately 11% of all shuffles, the first Steely Dan song will occur at or before the seventh song in the shuffled playlist and in approximately 50% of all shuffles, the first Steely Dan song will occur at or before the 41st song in the shuffled playlist. Contrary to Levy's first impressions, the first song from Steely Dan will occur fairly early in a large percentage of shuffles.

3.2 Number of Songs From One Group in One Shuffle

Many comments from Levy and from other users of the iPod claim nonrandomness in the shuffle feature based on the number of songs from a particular grouping, such as Artist or Album, that occur during the first n songs in a shuffled playlist. Assuming a random shuffle, for a particular grouping, the num-

Table 1. Probabilities of hearing 0, 1, 2, or 3 songs from Steely Dan in the first 20, 40, or 60 songs in a shuffled playlist of $N = 3{,}000$ songs.

n	$P(Y_i = 0)$	$P(Y_i = 1)$	$P(Y_i = 2)$	$P(Y_i = 3)$
20	0.7138	0.2435	0.0387	0.0038
40	0.5083	0.3492	0.1146	0.0239
60	0.3611	0.3747	0.1873	0.0601

ber of songs Y_i from the ith group that appears in the first n songs of a N song playlist has a hypergeometric distribution with probability distribution function

$$P(Y_i = y) = \frac{\binom{G_i}{y}\binom{N-G_i}{n-y}}{\binom{N}{n}}$$
$$y = \max(0, G_i + n - N), \ldots, \min(G_i, n). \quad (2)$$

In the *Newsweek* article, Levy points again to the overrepresentation of songs from Steely Dan "whose songs always seemed to pop up two or three times in the first hour of play." Assuming $G_i = 50$ songs from Steely Dan out of a playlist of $N = 3{,}000$, the probabilities of getting zero, one, two, or three songs from Steely Dan in the first $n = 20, 40$, and 60 songs (corresponding to roughly 1, 2, and 3 hours of play) in the shuffled playlist are summarized in Table 1. For each value of n, the probabilities of hearing two or three songs from Steely Dan are all fairly low. In repeated shuffles, this outcome should not happen with any great regularity. Unfortunately, Levy did not perform a systematic study of this outcome, making it possible these "coincidences" did not occur as often as stated. This possibility is further reinforced by Levy's own description of the group's music as "terse, jazzy, and sometimes lyrically incomprehensible" (Levy 2006, p. 177). Although Levy's conclusions may have been influenced by a combination of "coincidences" and personally meaningful music, we cannot immediately refute his claims based on the probabilities.

3.3 Numbers of Songs From All Groups in One Shuffle

In 2004, Apple, Inc. introduced a new low-priced iPod called Shuffle, built entirely around the shuffle feature. The first Shuffle held approximately 120 songs. Since most users have many more than 120 songs in their digital music libraries, Apple uses a feature called Autofill to randomly select enough songs from a user's digital library to fill the capacity of the Shuffle. In his *Newsweek* article and book, Levy tested the Autofill feature and reported "The first few times…, I found some disturbing clusters in the songs chosen. More than once the 'random' playlist included three tracks from the same album! Since there are more than 3,000 tunes in my library, this seemed to defy the odds."

One way to look at this statement is to choose an album, look at the 120 songs selected to fill the Shuffle, and count how many songs are selected from the chosen album. The number of songs selected from this album follows the hypergeometric distribution of Equation (2). From a $G_i = 12$ song album with a total library of $N = 3{,}000$ songs, the probability of obtaining three or more songs from this particular album out of the $n = 120$ songs selected to fill the Shuffle is indeed small, approximately 1%. Looking at Levy's statement in this manner is equivalent to looking at the probability that someone in a room will share *your* birthday. Focusing on one possible outcome makes the probability of this event fairly small.

However, this is not the event Levy is describing. The event he describes is the event where there are three or more songs from *any* album in the $n = 120$ songs selected by the Autofill feature. This is the same as grouping the library of $N = 3{,}000$ songs by Album and looking at the maximum number of songs from any one album in the $n = 120$ songs selected to fill the Shuffle. In terms of the birthday example, his statement is the equivalent of looking at the probability that someone in a room will share *any* birthday.

Define random variables $Y_1, Y_2, \ldots, Y_g$ to be the numbers of songs selected to fill the $n = 120$ song Shuffle from each of the g Albums in the playlist. The joint distribution of $Y_1, Y_2, \ldots, Y_g$ follows a multivariate hypergeometric distribution (Johnson and Kotz 1969) with probabilities

$$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_g = y_g) = \frac{\binom{G_1}{y_1}\binom{G_2}{y_2}\cdots\binom{G_g}{y_g}}{\binom{N}{n}}. \quad (3)$$

The random variable described in Levy's statement is the maximum observation from a multivariate hypergeometric distribution ($\max(Y_1, Y_2, \ldots, Y_g)$) and the event described is the probability this maximum value will be three or more.

Published research on the multivariate hypergeometric distribution is focused on the values of $(Y_1, Y_2, \ldots, Y_g)$ producing the largest probability value, the mode of the distribution. A survey of the literature failed to produce any results for the probability distribution of the maximum observation from this distribution. The reason could be tied to the difficulty of directly calculating probabilities for the maximum observation from this distribution. Looking at a simplified case, assume the number of songs in the playlist is $N = 3{,}000$ and each album in the library contains $G_i = 12$ songs for a total of $g = 250$ albums. In this case, the maximum number of songs from any one album can range from 1 to 12. To calculate the desired probability that the maximum number of songs from any one album is three or more, we calculate the complement event; the probability the maximum number of songs selected from any one album is one or two.

For the maximum number of songs to be one, each of the $n = 120$ songs selected must each come from only one of the different $g = 250$ albums. This probability is

$$P(\max(Y_1, Y_2, \ldots, Y_{250}) = 1) = \frac{\binom{250}{120}\binom{12}{1}^{120}}{\binom{3000}{120}}$$
$$= 9.856 \times 10^{-15}. \quad (4)$$

For the maximum number of songs to be two, two songs could come from one album, or two songs could be selected from each of two albums, or two songs could be selected from each of three albums, etc. ending with two songs could come from each of 60 albums. The remaining songs must each come from a different album not already selected. This probability is

$$P(\max(Y_1, Y_2, \ldots, Y_{250}) = 2)$$
$$= \frac{\sum_{x=1}^{60} \binom{250}{x}\binom{250-x}{120-2x}\binom{12}{2}^x \binom{12}{1}^{120-2x}}{\binom{3000}{120}}$$
$$= 0.05554727. \quad (5)$$

This makes the desired probability

$$P(\max(Y_1, Y_2, \ldots, Y_{250}) \geq 3)$$
$$= 1 - P(\max(Y_1, Y_2, \ldots, Y_{250}) \leq 2)$$
$$= 0.9445. \quad (6)$$

Using this simplified example, the situation described in Levy's article would happen in 94.45% of all Autofills of the iPod Shuffle. Far from defying the odds, the event of observing a cluster of three of more songs from any album when using the Autofill feature is very much expected to occur. Just as in the birthday example, going from one particular outcome to any possible outcome leads to a much larger probability than expected.

Determining the probabilities of other maximum values becomes very difficult for even this simplified case, requiring more knowledge in combinatorics than most students in undergraduate probability and statistics courses generally possess. Instead of calculating these probabilities directly, a simulation program was written to estimate the probability distribution of the maximum observation from a multivariate hypergeometric distribution. The estimated probabilities based on 100,000 trials are listed in Table 2 below.

In reality, the digital music libraries of iPod users will not contain equal numbers of songs per albums or artists. This is especially true given the option of purchasing through iTunes a few favorite songs from a given album instead of owning the entire album. To determine how much the probability distribution of the maximum observation from the multivariate hypergeometric distribution would vary in more realistic settings, another simulation program was written to estimate this probability distribution function for a general user's library. For example, using the first author's library of $g = 81$ albums and $N = 1017$ songs, the estimated probability of getting three or more songs from any one album is approximately 1, the estimated probability of getting 8 or more songs from any one album is 0.06817, and the estimated probability of getting 9 or more songs from any one album is 0.01825. If you group songs by Artist instead of Album, the effect becomes more dramatic. Using the second author's library of $g = 109$ artists and $N = 856$ songs, the estimated probability of getting 3 or more songs from any one artist is approximately 1, the estimated probability of getting 17 or more songs from any one artist is 0.05996 and the estimated probability of getting 18 or more songs from any one artist is 0.02977.

Table 2. Estimated probabilities for the maximum observation of simplified multivariate hypergeometric distribution.

x	$P(\max Y_i \geq x)$	x	$P(\max Y_i \geq x)$
1	1.0000	5	0.0142
2	1.0000	6	0.0006
3	0.9453	7	0.0003
4	0.2116	8	0.0000

3.4 Number of Shuffles Required in Order to Hear One Particular Song

In his article in *Newsweek*, Levy contrasted the perceived favoritism shown by his iPod to Steely Dan songs versus the lack of favoritism for a particular song he purchased online. "Months after I bought *Wild Thing* from the iTunes store, I'm still waiting for my iPod to cue it up."

Clearly, if you listen to an entire shuffled playlist, you will hear each song on the playlist just once. However, this scenario almost never happens since users will listen only to the first n songs in an N song playlist before reshuffling the same playlist or choosing another. Assuming a random shuffle, the probability a particular song occurs in the first n songs of a N song playlist is n/N. Assuming each shuffle is independent, the number of shuffles S that occur until a particular song appears in the first n songs of the shuffle has a geometric distribution with probability n/N. Table 3 gives the 10th, 25th, 50th, 75th, and 90th percentiles for the random variable S assuming a $N = 3{,}000$ song playlist with varying values of n (30, 60, 90, 120). Depending on his listening habits, it would be entirely possible under a random shuffle to need many shuffles in order to hear *Wild Thing*.

4. GOODNESS–OF–FIT TESTS FOR NUMBER OF SONGS FROM ONE GROUP IN ONE SHUFFLE

Using several probability models, we are able to explain much of the anecdotal evidence of a nonrandom shuffle mentioned in Levy's *Newsweek* article and book. However, the seeming overrepresentation of a particular Artist early in the shuffled playlist cannot be refuted based on the probabilities alone. This perceived overrepresentation of certain groups early in a shuffled playlist has been reported not only by Levy, but by iPod users on several different websites. Usually, the groupings are by Artist or Album, but some users have claimed the shuffle favors songs more recently added to their iPods or songs purchased through Apple's iTunes store over songs from their CD collection.

To determine if there is any merit to these claims, we developed several goodness-of-fit tests for the number of songs from

Table 3. Percentiles from the distribution of the number of shuffles required to hear one song when listening to the first $n = 30, 60, 90,$ or 120 songs from a $N = 3{,}000$ song playlist.

	Percentile				
n	10th	25th	50th	75th	90th
30	10	28	68	137	229
60	5	14	34	68	113
90	3	9	22	45	75
120	2	7	16	33	56

one group appearing in the first n songs of a N song playlist. We then conducted each test using real data collected from iPods purchased by the authors. In order to develop and conduct these test, we assumed (a) the software on all computers and iPods tested have no defects and are no different than any other iPods available for purchase, and (b) to avoid listening to each song in its entirety and to complete the tests in a reasonable amount of time, the playlist must be shuffled as soon as the user selects it. The act of clicking past a song in the shuffled playlist without listening to the entire song does not change the order of the songs in the current shuffled playlist.

Under the null hypothesis of a random shuffle, the number of songs Y_i from a particular group appearing in the first n songs of a N song playlist has the hypergeometric distribution in Equation (2). Under this hypothesis, in s shuffles of the playlist, the expected number of times y songs will appear in the first n songs of the playlist is $E_i = s * P(Y_i = y)$. Pearson's chi-square goodness-of-fit test statistic is:

$$X^2 = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}. \tag{7}$$

Several outcomes y are combined in Equation (7) above so that the expected number E_i for each outcome i is at least 5. The p-value of the test is $P(\chi^2_{c-1} > X^2)$ and we will reject the null hypothesis when the p-value is less than $\alpha = 5\%$.

We conducted five separate goodness-of-fit tests. The first three tests were all conducted using the same $N = 240$ song playlist on the first author's iPod. For each of the three tests, the number of songs by a particular artist out of the first $n = 60$ songs in the playlist was recorded. The artists were selected based on the number of songs on the playlist. There were $G_i = 3$ songs in the playlist by Faith Hill, $G_i = 14$ songs by Queen and $G_i = 31$ songs by Jimmy Buffett, the most of any artist in the playlist. For each of the three tests, $s = 100$ shuffles were performed. The data collected in each of these three tests are given in Table 4. Combined outcomes are marked with one asterisk or with two asterisks. Included in the table are the test statistic Equation (7) and p-value for each test. In each case, we will fail to reject the null hypothesis and conclude the shuffle is random.

To test the perception of favoritism of recently added songs or songs purchased through Apple, the third author's iPod was used. For the first test, a playlist of $N = 40$ songs was created. Twenty of these songs were chosen randomly from a large number that were added on a single day shortly after the iPod was purchased, and the remaining 20 were chosen randomly from a large number that were added on a single day over a year later. For the second test, a different playlist of $N = 40$ songs was created, with 20 songs randomly chosen from all songs purchased from Apple's iTunes Store and 20 songs randomly chosen from all songs from the third author's CD collection. For each test, the number of songs appearing in the first $n = 10$ songs in the two groups (recently added songs for the first test and purchased songs for the second test) was recorded for $s = 200$ shuffles. The data collected in each of these two tests are given in Table 5. Combined outcomes are marked with one asterisk or with two asterisks. Included in the table are the test statistic found in Equation (7) and p-value for each test. In each case, we again fail to reject the null hypothesis and conclude the shuffle is random.

Thus, we failed to find any evidence to support the claim of users like Steven Levy of favoritism of certain groups in the shuffle.

5. CLASSROOM USES

Several of these examples have been used in the undergraduate probability and statistics courses at Iowa State University. These examples were well received by students, and several of them stayed after class or visited office hours to discuss ideas related to testing the shuffle feature or their personal impressions of the randomness of the shuffle. After teaching these courses for many years, the first author can state without reservation that no other examples or material has elicited this kind of response from students in these courses. (See Stefanski 2007 for thought-provoking examples to teach variable selection in linear regression courses.)

Table 4. Number of songs appearing in the first $n = 60$ songs of a $N = 240$ song playlist from three different artists in each of $s = 100$ shuffles.

Faith Hill ($G_i = 3$ songs)		Queen ($G_i = 14$ songs)		Jimmy Buffett ($G_i = 31$ songs)	
# of songs	# of shuffles	# of songs	# of shuffles	# of songs	# of shuffles
0	51	0*	1	0, 1, or 2*	0
1	32	1*	7	3*	1
2*	12	2	20	4*	6
3*	5	3	20	5	9
		4	23	6	14
		5	20	7	20
		6**	7	8	16
		7**	2	9	17
		8 or more**	0	10	11
				11	3
				12**	2
				13**	1
				14 or more**	0
$X^2 = 4.6539$		$X^2 = 3.31200$		$X^2 = 3.57408$	
p-value $= 0.0976$		p-value $= 0.6520$		p-value $= 0.8934$	

Table 5. Number of songs appearing in the first $n = 10$ songs of two $N = 40$ song playlists (songs recently added and songs purchased through iTunes) in each of $s = 100$ shuffles.

	Recently added	Purchased
# of songs	# of trials	# of trials
0*	0	0
1*	1	2
2*	4	7
3	19	18
4	40	47
5	61	52
6	43	42
7	24	25
8**	6	6
9**	2	1
10**	0	0
X^2	2.1154	2.9092
p-value	0.9088	0.8202

If the examples in this article are to be used in the classroom, the general idea of a random shuffle feature should be discussed first, so that all students have a general understanding of the topic. The statistical tests could be used directly from this article, or students could be asked to design and perform their own tests on a particular aspect of the random shuffle feature. If iPods are not available, the same type of study could be conducted using the shuffling code of a statistical package.

Finally, the analyses in this article could be used to start a discussion of several aspects of hypothesis tests. Setting the α level of a hypothesis test is based on prior belief in the truth of the null hypothesis, and is open to interpretation. From our viewpoint, there is no reason to think the software engineers at Apple, Inc. made an error in the shuffle code. Further, people are notoriously poor at detecting random and nonrandom patterns and determining probabilities of events intuitively. These reasons, coupled with the fact that music can be personally meaningful and cause a emotional response in the listener, would lead us to discuss setting the α level of any test of the random shuffle feature very low, at most 1%. One could also use these results to discuss the power of a hypothesis test. The statistical tests presented in this article are, of course, not a definitive proof of the randomness of the shuffle. Using a statistical package, students could estimate the power of a statistical test to detect a nonrandom shuffle by simulating using probabilities different than what is indicated by a random shuffle.

6. CONCLUSIONS

Random behavior is a difficult concept. The lack of understanding of random behavior often leads to misconceptions about what constitutes nonrandom behavior. Music also evokes strong emotions. Songs on a person's iPod are personally meaningful to them; they own the songs. The controversy about the shuffle feature combines a difficult and often misunderstood concept with personally meaningful events. Consequently, the existence of this controversy is not at all surprising.

Much of the evidence of nonrandom behavior reported by Steven Levy and others does not hold up when the probability models of the events are determined. Our results show the probability models for a random shuffle in many cases do not match the intuition of users. In addition, our statistical tests show the long-term occurrences of these events are within expectations under the assumption of a random shuffle.

[Received April 2007. Revised April 2009.]

REFERENCES

Bakan, P. (1960), "Response-Tendencies in Attempts to Generate Random Binary Series," *The American Journal of Psychology*, 73 (1), 127–131.

Baker, D., and Diaconis, P. (1992), "Trailing the Dovetail Shuffle to Its Lair," *The Annals of Applied Probability*, 2 (2), 294–313.

Bennett, D. (1998), *Randomness*, London: Harvard University Press.

Diener, D., and Thompson, W. B. (1985), "Recognizing Randomness," *The American Journal of Psychology*, 98 (3), 433–447.

Falk, R. (1989), "Judgment of Coincidences: Mine versus Yours," *The American Journal of Psychology*, 102 (4), 477–493.

Johnson, N. L., and Kotz, S. (1969), *Discrete Distributions*, New York: Wiley.

Levy, S. (2005), "Does Your iPod Play Favorites?" *Newsweek*, CXLV (5).

——— (2006), *The Perfect Thing*, New York: Simon & Schuster.

Miller, G. K., and Fridell, S. L. (2007), "A Forgotten Discrete Distribution? Reviving the Negative Hypergeometric Model," *The American Statistician*, 61 (4), 347–350.

Orr, W. F., Federspiel, C. F., and Maxwell, O. D. (1972), "Patterns in Random Chance," *The American Journal of Psychology*, 85 (1), 99–115.

Ross, S. (2006), *A First Course in Probability*, New Jersey: Pearson Prentice Hall.

Stefanski, L. A. (2007), "Residual (Sur)Realism," *The American Statistician*, 61 (2), 163–177.

Wackerly, D., Mendenhall, W., and Scheaffer, R. (2007), *Mathematical Statistics With Applications*, California: Thomson Brooks/Cole.