

CHARACTERISTICS OF STUDENTS' PROBABALISTIC REASONING IN A SIMULATION-BASED STATISTICS COURSE

Aaron Weinberg

Department of Mathematics, Ithaca College, Ithaca, NY, USA
aweinberg@ithaca.edu

Although students in traditional introductory college statistics courses see a frequentist definition of probability, they are rarely asked to use this definition, instead relying on technology to conduct parametric tests. In contrast, in a simulation- and randomization-based statistics course, the relative frequency of simulated outcomes becomes the central focus of the process of drawing inferences. In this study, eight college students who were enrolled in an introductory simulation-based statistics course were interviewed and asked to describe sampling distributions and make inferences; the results and analysis describes the ways they used and appeared to think about empirical probabilities. Although the students appeared to be able to make connections between various aspects of inferential reasoning, they also encountered difficulties that may be related to their focus on empirical probabilities.

INTRODUCTION

In recent years, there has been an increasing call for introductory statistics courses to focus on informal inference as an important precursor to understanding formal inference (e.g., Cobb, 2007). As part of this call, there have been numerous recommendations for students to use simulation and randomization-based methods to engage in activities to support the development of this inferential reasoning (e.g., Garfield, delMas & Zieffler, 2012).

In contrast to traditional approaches, the simulation-based pedagogy offers a basis for reconceptualizing the notion of probability and its role in drawing statistical inferences. In a traditional approach, students first learn probability rules (e.g., the addition and multiplication rules) then parametric probability models (e.g., the normal distribution) and then inference. In contrast, students in simulation-based classes use computer-based methods to construct empirical sampling distributions and then compute *empirical probabilities* to make inferences.

The goals of this paper are to discuss the empirical probability and sampling distributions; to present an outline of an instructional sequence that is designed to help students use this notion of probability to make statistical inferences; and to identify features of students' reasoning about sampling, probability, and inference after they participated in the instructional sequence.

BACKGROUND: EMPIRICAL PROBABILITY

Students typically encounter several notions of probability in a traditional introductory statistics class. Textbooks often initially define probability using a frequentist notion—for example, Moore (2010) defines it as “the proportion of times the outcome would occur in a very long series of repetitions” (p. 263). After presenting this definition, the books typically present classical probability by describing situations with equally likely outcomes (e.g. rolling a die or flipping a coin), and then describe axiomatic probability (e.g., the addition and multiplication rules), random variables, and probability distributions (e.g., the binomial and normal distributions).

As Shaughnessy (1992) notes, “the model of probability that we employ in a particular situation should be determined by the task we are asking our students to investigate, and by the types of problems we wish to solve” (p. 468). A simulation-based approach to introductory statistics requires that students conceptualize and compute probabilities as long-run relative frequencies as they make inferences. Students in such classes learn to reason following the trajectory shown in Figure 1: They first construct a simulation model (e.g., a list of the numbers one through six to model rolling a fair die) then they randomly sample from the model and compute a statistic; this process is governed by the relative frequencies with which the outcomes occur in the model. Next, students repeat the sampling process to create an empirical sampling distribution of statistics; this construction is governed by sampling variation. Finally, the students examine the relative frequency of particular outcomes in the sampling distribution to make an inference. Since these relative frequencies are made explicit in the probability model and the

empirical sampling distribution, the probabilities involved in this process can be thought of as *empirical probabilities*.

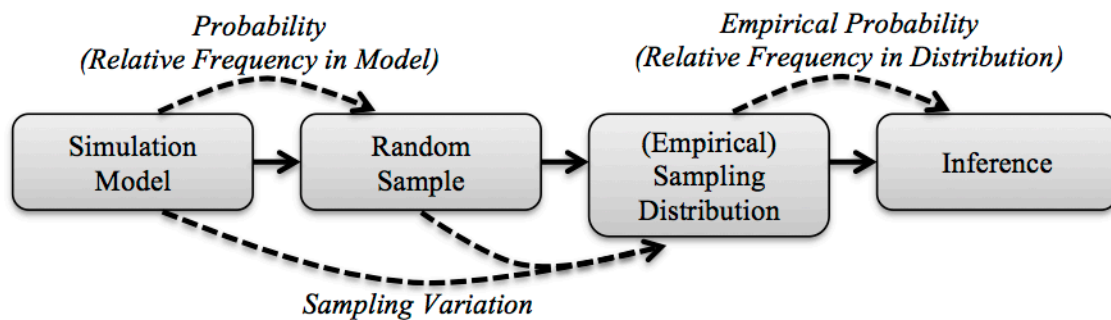


Figure 1. Trajectory of inferential reasoning in simulation-based pedagogy

PRIOR RESEARCH

The recent research in teaching and learning of probability has tended to focus on describing the ways students connect (or fail to connect) theoretically derived probabilities with the relative frequencies that are present in empirical data. For example, Lee, Angotti and Tarr (2010) and Rider and Stohl Lee (2006) had students use computer simulation methods to make inferences about whether or not dice was fair given data from rolling the dice. They examined the ways students understood and used ideas of sample size, variability, independence, and fairness within the context of the task and its various representations; they found that students who used the idea of long-run relative frequency and simulation methods were able to make “meaningful connections between the empirical data they had collected and theoretical probability distributions” (Rider & Stohl Lee, 2006 p. 5). In contrast, students who collected small samples and relied on formal statistical tests to make inferences appeared to not make such connections.

Pratt, Johnston-Wilder, Ainley, and Mason (2008) investigated the tensions students experienced between “local” and “global” perspectives, in which students respectively focused on the short-term patterns and long-term patterns in randomly-generated data. They found that students in some situations initially focused on local patterns and then, later, on global patterns; in other situations, students appeared to shift their attention back and forth between the local and global perspectives.

The students in these studies had (indirect) access to “actual” population parameters and based their inferential on the Law of Large Numbers rather than drawing conclusions using probability. This is in contrast to the scenarios students would encounter in an advanced statistics course in which they have access only to sample data and need to consider sampling variability when making inferences. Thus, there is still a need to investigate the ways students understand probability and inference in the context of resampling.

INSTRUCTIONAL SEQUENCE

The instructional sequence described here is based on the CATALYST curriculum materials (Garfield, delMas & Zieffler, 2012). The initial activities are designed to introduce students to the concept of probability as an empirical long-run relative frequency as well as to give them experience working with probability derived from a population model and as relative frequency in an empirical sampling distribution:

- In the first activity, students are given several pig-shaped dice and asked to assign point values to each “face” so as to make a die-rolling game “fair.” The students in the class collect and aggregate their statistics, and then are asked to describe probabilities of rolling individual faces and of rolling collections of faces (e.g., having three back-rolls in a 10-roll set). Following this, they are asked to decide whether a particular group’s collection of rolls appeared “unusual” or “surprising.”

- In the second activity, groups of students take scoops of various colored beans from a large bag (containing approximately 30,000 beans) and use their aggregated statistics to estimate the proportions of the bean colors. Following this, they are asked to imagine that they are running a carnival game in which contestants are trying to guess the proportion of colored beans in the bag (after taking a scoop) and to use the class' data to determine how often a contestant would guess within a certain range of the population parameter.

Following this introduction to the two types of probability in the inferential reasoning trajectory, the students begin building population models and setting up computer simulations to generate sampling distributions. Initially, they use the aggregated data from the first two activities to build these models; next, they are introduced to a “blind guessing” model (i.e. a null model). In each repeated-sampling simulation, they create a weighted list of outcomes, use computer tools to take numerous samples and compute statistics, and then make statistical inferences based on the relative frequencies in the empirical sampling distributions.

For example, in one activity, students are presented with a scenario in which infants in a research study are asked to choose between a “helper” toy and a “hinderer” toy (which had previously been observed engaging in, respectively, positive and negative interactions with other toys); the students are told that 14 out of 16 infants chose the “helper.” After constructing a null model (in which the babies are equally likely to select each toy), students use a computer to construct an empirical sampling distribution and answer questions such as the following:

- The “Key statistical question”: If 16 babies randomly select between the helper and hinderer toys, how unlikely is it to see the observed result—or one further from what we'd expect—just due to the natural variation from sample-to-sample?
- How would you find the probability of getting the observed result?
- What does your numerical value mean? Use the meaning of probability in your explanation.
- What would you conclude based on this probability? Does your numerical result make sense? Explain why or why not.

The rest of the activities in the instructional sequence are pedagogically similar to the one described above, but introduce means, comparing multiple populations, randomization, and bootstrapping. Throughout these activities, the students are expected to use the underlying idea that probability is an empirical long-run relative frequency to explain and interpret the results of their simulations and the inferences they make based on these results.

METHODS

Eight undergraduate students participated in the study; all were enrolled in a one-semester introductory statistics course with a target audience of mathematics majors and minors. Each student participated in two semi-structured interviews near the mid-point and end of the semester. In the interview, they were asked to work on four problems, which are outlined in Table 1.

The students were asked to think aloud as much as possible while working on the problems; the interviewer asked questions designed to challenge the student's reasoning and conclusions. The interviews were video-recorded (screen capture software was used for instances when the student was using the computer) and transcribed for analysis.

Analysis of the data was conducted using grounded theory (Strauss & Corbin, 1990). The students' utterances were read and categorized according to aspects of sampling, resampling, probability, distributions, relative frequencies, and statistical inference that they discussed and used as well as their reasoning when drawing conclusions and comparing distributions.

Table 1: Outline of interview questions

Problem 1: You have a population of men with a mean height of 69 inches. A clerk at a post office in a small town takes a 5-person sample each day and record the mean height; a clerk in a large town takes 50-person samples. Which clerk will record more days over 71 inches?
Problem 2: Researchers collect many samples of 50 Goodyear tires and compute the mean tread life of each; their results are displayed in a histogram [included in the question]. You collect a sample of Michelin 50 tires and find that it has a mean tread life of 6 years; is that evidence that the Michelin tires last longer than the Goodyear tires?
Problem 3: Researchers sent out 2600 identical resumes; half had “white sounding” names and the others had “black sounding” names. They received 121 positive responses for the white-sounding names and 87 for the black-sounding names. Then they ran 500 simulations under the assumption that 208 out of every 2600 names should receive a response and the results are displayed in a histogram [included in the question]. Should the researchers be concerned?
Problem 4: Given a histogram of a population of test scores and histograms of four potential sampling distributions [included in the question], which histogram(s) could represent sampling distributions with sample sizes of 4 and 50?

RESULTS AND ANALYSIS

Attending to Relative Frequencies

All of the students discussed the relative frequencies of outcomes in the sampling distribution while they were making inferences and drew conclusions about parameters by looking for relative frequencies of an observed statistic. For example, below is an excerpt from Brian’s response to Problem 2 in Interview 2:

Brian: From the data, it [getting a tread-life average of 6.6 years] would be surprising because you're looking at—it would be over here in this last bar [pointing to right-most bar]. And that's the upper, like, two-and-a-half percent, I guess you could say, of our data.

Coordinating Sampling Processes

In order to describe sampling distributions or make inferences, students needed to be able to discuss the processes of sampling from the population based on the relative frequencies of the population outcomes and then repeating this process multiple times. While doing this, they need to coordinate the shape, center, and spread of the population distribution with the center and size of the sample and then with the shape, center, and spread of the sampling distribution. All of the students in the study attempted to do this on most of the problems in the interview; some students were able to coordinate more processes and aspects than others.

For example, in the excerpt below from Problem 1 of the first interview, Nick made connections between the (implied) shape, center, and spread of the population distribution with the center of the sample and then with the center and spread of the sampling distribution while discussing the role of sample size:

Nick: So, the small post office is a pretty small sample, so you can have, like, a pretty wide variety with just five people, but if you have 50 people, a lot of them... because 69 inches is the average, so you're saying a lot of people are 69 or close to 69, otherwise you wouldn't get that as the average. So if you have a group of 50, there's going to be a fair amount of them with around 69 inches, but if you only have five then you could have a really tall guy and you might have one or two near the average but you could also have some short guys, and it's just more likely that your average will be further from the 69 inches.

Attending to Sampling Variation

Most of the students discussed the idea of sampling variation and the role it played in constructing the empirical sampling distribution. For example, in the excerpt below from Problem 1 in Interview 2, Kevin used this idea to describe the spread of the sampling distribution as he connected the sampling and resampling process:

Kevin: There's natural sampling variation from each day. Like each day is going to differ a little bit. It's possible that they'll both be close to each other. And I mean, they would vary for the reasons I said before, because the one has a smaller sample size, the mean is going to get thrown in different directions depending on who walks through the door. And I mean, that principle holds true for the sample of 50, but you would need a lot more of either short or tall people to throw the mean.

Attending Explicitly to Population Probabilities

Although most students connected the process of sampling with constructing a sampling distribution, they typically *didn't* attend to the probabilities in the population distribution when doing so. For example, in the excerpts below from Problem 1 in Interview 2, Brian's reasoning reflects a lack of consideration that a fixed population should lead to a fixed probability of selecting a "tall" person:

Brian: I think that the five—the smaller post office will record more days with an average over 71 because you have a smaller sample size so it would be easier for, let's say, two people to come in and be abnormally tall and it would skew the data way more because you have less data. So one skewed data point can affect the mean, because we're looking for the mean, more than it could affect it on the large post office.

By failing to attend to the probabilities in the population model as part of the sampling process, students may have difficulty constructing an accurate empirical sampling distribution. This is particularly problematic when they attempt to make inferences, since these probabilities form the basis for repeated sampling from a null model; attending to this null model is essential for interpreting the empirical probabilities in the sampling distribution.

Inference Only with Certainty

Some students in the study attended to the relative frequency of outcomes in the sampling distribution but, based on this, were only willing to base their inferences on whether or not an outcome appeared. For example, in the excerpts below from Problem 2 in Interview 1, Nick decided that one could draw a conclusion about the tread length of tires only if the observed statistic didn't appear in the histogram. Furthermore, Brian was able to discuss the role that sampling variation played in the resampling process and used this idea to justify his reasoning:

Nick: I don't think that's [a statistic of 6 years] enough evidence because you have Goodyear groups that last six years or more, so it's looking at one group, and.... I guess Goodyear is capable of producing a group of tires that last just as long as the Michelin tires, so you can't say that one is better than the other.

Inference Only from Multiple Samples

On both of the problems where they were asked to make an inference (problems 2 and 3), many students believed that you couldn't make an inference based on a single sample; instead, you needed to record multiple statistics to make an inference. In the excerpt below from Problem 2 in Interview 1, Howie asserted that you would need to have multiple observed statistics to draw a conclusion:

Howie: I mean it's one instance where they last longer, but I think we have to look at many groups of 50 to compare the different data sets that we have.

Distinguishing Probability from the Law of Large Numbers

Several students appeared to have difficulty distinguishing between probability as a long-run relative frequency and the idea of the law of large numbers. For example, Nick appeared to do this on Problem 1 in Interview 1:

Nick: So probability is... the long run relative frequency. So if you make a lot of observations, you take the number of times a certain outcome occurs and divide by the total number of times you made observations. So, in this case it would be each person's height divided by all the heights you got... and you throw in average, um... yeah that, hmm...

DISCUSSION

The students in this study discussed statistical inference in terms of the reasoning process they had experienced in the simulation-based classes by focusing on the relative frequencies of various outcomes in the sampling distributions. In addition, they were able to describe and coordinate various aspects of the reasoning trajectory to make sense of the sampling distributions. In particular, most of the students considered the effects of sampling variation in the repeated sampling process and discussed the ways this might relate to the values they observed in the sampling distribution. Despite this, most students had difficulty accounting for the probabilities in the population (or null) distribution as they described the sampling process.

There were three aspects of the students' inferential reasoning that might be distinct from the ways students in a traditional class might struggle with aspects of statistical inference. First, some students felt that it was only possible to draw a conclusion when the observed statistic had an empirical probability of zero in the sampling distribution (i.e. when the observed statistic didn't appear in the sampling distribution). Since students in a traditional class wouldn't have constructed an empirical sampling distribution from which to observe relative frequencies, it seems unlikely that they would want to reason from certainty rather than probability in the same way as the students in this study. Second, some students in the study believed that making an inference required that you actually draw repeated samples from the population and compare these to the sampling distribution; this might stem from a misapplication of the belief that repeated sampling is essential to the process of making an inference. Third, some students in the study had difficulty distinguishing the idea of a long-run relative frequency from the law of large numbers and, as a result, they confounded probabilities with averages.

These results suggest that the students in this study were able to conceive of probabilities as long-run relative frequencies in order to discuss sampling distributions and make inferences. This suggests that the students' experiences in the simulation- and randomization-based curriculum supported the development of these conceptions. However, as students develop this conception of probability, they also appeared develop particular misconceptions, and attending to these misconceptions will require new types of instructional intervention.

REFERENCES

- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1).
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM—The International Journal on Mathematics Education*, 44(7), 883-898.
- Lee, H. S., Angotti, R. L., & Tarr, J. E. (2010). Making connections between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal*, 9(1), 69-96.
- Moore, D. (2010). *The basic practice of statistics (5th Edition)*. New York: W. H. Freeman.
- Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, 7(2), 107-129.
- Rider, R., & Stohl Lee, H. (2006). *Differences in students' use of computer simulation tools and reasoning about empirical data and theoretical distributions*. Paper presented at the Seventh International Conference on Teaching Statistics (ICOTS-7), Ljubljana, Slovenia.
- Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics and learning* (pp. 465-494). New York: Macmillan.
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage Publications.