CONCEPTUAL ISSUES IN QUANTIFYING EXPECTATION: INSIGHTS FROM STUDENTS' EXPERIENCES IN DESIGNING SAMPLING SIMULATIONS IN A COMPUTER MICROWORLD

Luis Saldanha
School of Mathematical and Statistical Sciences, Arizona State University,
United States of America
luis.saldanha@asu.edu

We report on part of a classroom teaching experiment that engaged a group of high school students in designing and running sampling simulation in a computer micro world. The simulation design activities provided a vehicle for engaging students with informal hypothesis testing and for fostering their (re)construal of contextualized situations as probabilistic experiments—that is, as a scheme of interconnected ideas involving an imagined population, a sample drawn from it, and repeated sampling as an imagistic basis for quantifying one's expectation of particular sampling outcomes under an assumption about the composition of the sampled population. Our report highlights challenges that students experienced and that shed light on aspects of quantifying one's expectation that a random sampling process will produce a particular type of outcome.

BACKGROUND

The use of simulation-based activities in statistics instruction over the last quarter century has been well documented (Mills, 2002). Mills' (2002) review of the literature on the use of computer simulations in statistics education reveals that a panoply of instructional approaches and software resources have been developed and used for teaching a variety of statistical topics, mostly at the post-secondary school level. Mills also points out that there is a paucity of empirical research into the impact of simulation-based instruction on students' understanding of statistical concepts. Nevertheless, several notable empirical studies of students' statistical reasoning or achievement in relation to their participation in instruction involving the use of computer simulations have been conducted (delMas, Chance & Garfield, 1999; Drier, 2000; Well, Pollatsek & Boyce, 1990; Kuhn, Hoppe, Lingnau & Wichmann, 2006; Pratt, 2000; Saldanha & Thompson, 2002, 2007; Sedlemeier, 1999; Stohl & Tarr, 2002). A subset of these studies (Drier, 2000; Kuhn et al., 2006; Pratt, 2000; Stohl & Tarr, 2002) engaged students with instructional tasks in which they actually *designed* components of simulations, in the context of modeling given stochastic situations within specially designed microworlds. We highlight two salient commonalities among this focused subset of studies that serve to frame the study we report here. The first commonality is the prominence of the idea of repeated sampling and long-run relative frequency as a basis for supporting students' understanding of ideas of likelihood, sample representativeness, and the relationship between increased sample size and reliability of the results of a simulated sampling experiment. Instructionally, these studies each engaged students within a microworld that provided opportunities for them to readily and easily experiment with and control the number of trials of a simulated experiment, which in turn reportedly supported their developing the documented understandings. The second commonality relates to the nature of the virtual simulation tools that students used and within which their constructive activity and learning occurred: the microworlds employed in each of these studies involved simulation tools that were virtual versions of canonical random devices such as spinners and dice. As such, these microworlds provided students with pre-determined models to use in the modeling process, thereby arguably providing them with a certain transparency to the virtual devices' real counterparts. Doerr and Pratt (2008) point to a rationale for this microworld design in terms of the idea of *exploratory modeling*; the intent is to engage learners with a pre-made model constructed by experts to represent the structure of knowledge in some content domain (stochastics in the above cases), and to provide them with opportunities to explore that domain through the testing of hypotheses and variation of parameters within the constraints of the pre-made model.

PURPOSE AND METHOD

The two aforementioned commonalities speak to both an overlap and a difference between these studies and ours. The overlap is in the importance of repeated sampling and long term relative frequency of outcomes as foundational to a coherent understanding of likelihood and representativeness. The difference is in the nature of the tasks and the microworld we employed in our study, which in turn reflects the different purpose of our research. Our study also engaged students in exploratory modeling activities, but it did so in the context of tasks and a microworld that did not involve the use of virtual versions of familiar stochastic devices. Instead of having students work with predetermined models in the modeling process, our tasks and microworld aimed to provide them with a metaphor for thinking about simulating stochastic experiments in terms of a population, a sample, and a repeatable method of selecting the former from the latter. Working within this set-up—in which the relationship between the modeled situation and the modeling tool was not so transparent—provided us an opportunity to investigate aspects of re-conceiving scenarios that were not apparently stochastic *as having a stochastic structure*. In this regard our study differs significantly from those cited above; a central goal of ours was to provide insight into two conceptual problems: 1) learning to model such scenarios explicitly in terms of conceiving a stochastic experiment, and 2) conceiving of expectation as a statistical quantity—important areas of reasoning that are not well represented in the stochastics education research literature. We moved to advance this goal by conducting a classroom teaching experiment that explored the thinking of a group of high school students as they engaged with tasks designed to foster their creation and use of computer simulations for making informal inferences and testing hypotheses in situations involving the construal of a stochastic experiment.

Eight students participated in the classroom teaching experiment conducted in their intact introductory statistics course during the fall semester at a suburban high school in the Southeastern United States. We engaged students in activities entailing the design of sampling simulations within the Prob Sim microworld (Konold & Miller, 1996). Prob Sim's user interface employs the metaphor of a "mixer" for a population and a "bin" for a sample; it enables the user to easily specify a population's composition and size, the size and selection method of a sample (with or without replacement), and the number of trials of the simulated sampling experiment to be conducted (see Figure 1, right panel). The software also displays the outcome of each simulated sampling experiment as a raw data list, and it provides summary analyses of the aggregated outcomes and displays their distribution as a relative frequency histogram. The simulation design activities asked students to explore questions of the type "is event x unusual?". Students were presented with scenarios in textual form that described a contextualized situation in non-statistical terms and they were asked to investigate whether there was reason to believe that a specified event of interest was unusual. Their task was to design a simulation, guided by the constraints of the Prob Sim interface, to run the simulation and then draw a conclusion on the basis of its results. The left panel of Figure 1 displays a representative example of a simulation design activity that we will discuss in the next section of the paper; the task prompts listed in this example embody the common structure of all such activities that students engaged with. The activities were designed to both promote students' conceptualizing probabilistic situations and engage them informally with the logic of statistical hypothesis testing. These activities built on the foundational idea of repeated sampling (as a strategy for exploring a hypothesis about particular kinds of sampling outcomes) that we had developed with students in a preceding phase of the experiment.

Activities unfolded over four 45-55-minute class periods held on consecutive school days. They typically involved whole-class discussions directed at having students think and describe how they would use Prob Sim to investigate and resolve the question raised in a given situation.

We conducted a retrospective analysis of the classroom discursive data (all sessions were videotaped) and students' written responses on a post-test, documenting evidence of their understandings and challenges they experienced in their efforts to think of expectation as a statistical quantity. Our analytical procedures involved generating and applying descriptive and explanatory codes, following Cobb and Whitenack's (1996) method of videotape analysis and triangulation with the written data.

INVESTIGATING "UNUSUALNESS"

Ephram works at a theatre, taking tickets for one movie per night at a theatre that holds 250 people. The town has 30, 000 people. He estimates that he knows 300 of them by name.

Ephram noticed that he often saw at least two people he knew. Is it in fact unusual that at least two people Ephram knows attend the movie he shows, or could people be coming because he is there?
(The theatre holds 250 people)

Assumptions for your investigation:

Method of investigation:

"Gut level" answer:

Result:

Conclusion:

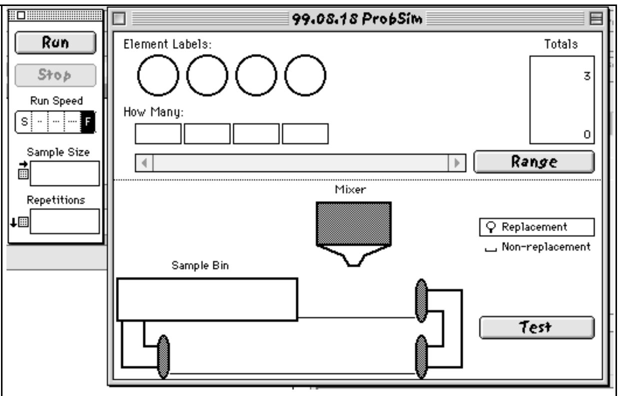

Figure 1. The "Movie Theatre" task (Konold, 2002),
involving the design of a sampling simulation (left panel) in the Prob Sim microworld (right panel)

RESULTS

This brief report summarizes the results around our second research problem—conceiving of expectation as a statistical quantity. The results generally illustrate students' experiences and struggles in their efforts to move from an intuitive sense that an event in question might be unusual to (re)conceiving a given situation in a manner that would enable them to quantify their expectation and to express that quantification in operational terms. The "Movie Theatre" task displayed in Figure 1 was preceded by classroom discussions about what it means to think of an event as being "unusual", in the context of simulating the likelihood of finding two people with the same birthday in a class of 25 people. The first two excerpts that follow are from this discussion on two consecutive days.

Excerpt 1 (Lesson 5):

1. I: […] So what does it mean for an even to be unusual?
2. Peter: You're not expecting it.
3. Kit: You don't expect it.
4. I: Yeah, it's unexpected, that in a large number of times that you do this (2-second pause) you, you, you expect to see it rarely. Ok? (2-second pause) All right, so in this particular case what does it mean that, to wonder if the event that's described is unusual?
5. Peter: You wouldn't expect 2 people in a cla—in a group of 25 people to have the same birthday.

Excerpt 2 (Lesson 6):

1. I: Uhh, in last class we looked at how, how, ways in which we could investigate whether or not something was unusual. All right? Now, what did it m—what did we mean by saying that something was unusual?
2. Luke: It's happens less than 50% of the time.
3. I: Or uhh, was that it? I mean, if it happened 49% of, like, is it unusual to get a tail if we toss a head?
4. Michelle: Unexpected
5. Luke: Yeah, it's unexpected.
6. I: All right, it's unexpected. And how, how would you quantify that?

7. Michelle: How would we what?
8. I: How would you quantify that, that it's unexpected?
9. Nicole: I mean—
10. David: It doesn't usually happen.

These excerpts illustrate how students' ideas about unusualness at this stage were largely intuitive and non-quantitative; their images were focused on individual occurrences not embedded within a sequence of trials of a repeatable experiment. Ideas of (relative) frequency were still largely absent from students' thinking and discourse. Students' descriptions suggest that they were interpreting the situation as it is described rather than re-conceiving and describing it in terms of a repeatable experiment that might support thinking about how to quantify expectation. In Excerpt 2, Luke's response (line 2) is interesting, as it appears to center on the idea of relative frequency. However, his focus was on a numerical criterion for unusualness. Despite the suggestion of underlying images of relative frequency in Luke's response, such imagery appears not to have been at the foreground of his thinking as indicated by his lack of a description of the *process* by which one might arrive at a choice of 50% as a proportion below which we agree to call an occurrence "unusual". Similarly, responses like David's "it doesn't usually happen" (line 10) amount to a non-quantitative rephrasing of "it's unexpected"; such responses were quite common in the earlier phases of the experiment.

The next excerpt is from the discussion around the Movie Theatre task (see Figure 1). It illustrates specific challenges of developing the mental operations entailed in a quantitative image of expectation and unusualness of an event.

Excerpt 3 (Lesson 6)

400. I: Let's all make sure that we know what's going on. What is, what is it that's at issue?

(7-second silence)

401. I: Kit?
402. Kit: Whether or not it's unusual for him to see at least 2 people that he knows.
403. I: And what does it mean, what does "unusual" mean?
404. Kit: Not expected.
405. I: Ok. Go on and quantify that.

(5-second silence)

406. I: It means if he were, if he were to do this many many times he would expect some small fraction of the time for this to happen, to see—see, keep, you gotta, I want you to keep putting this idea of repeating an event over and over and over again. (3-second pause). […] So now, Kit, once more: what is, what does it mean to be unexpected?
407. Kit: uhh a small fraction of the time, when it's done several times.
408. I: Ok, and what is the "it" in this case?
409. Kit: Uhh, the people—uhh seeing more than 2 people or 2 people one time per night, that he knows.

In Excerpt 3 Kit's attempt to characterize "unexpected" in quantitative terms (line 407), subsequent to the instructor's prompting (line 406), indicates that the idea of "a small fraction of the time" was salient for her, and that she might therefore have been mindful of a repetitive process. However, her description is somewhat tenuous and suggests an ambiguity in her thinking with regard to distinguishing the process being repeated and observing the event of interest; her response to the instructor's call for clarification of the referent for "it"(lines 407-409) indicates that she imagined that observing the outcome of interest repeatedly was itself the process being repeated. Thus, it would appear that at that point Kit was mindful of two images—repeating a process and observing the outcome of interest—but was not easily coordinating them to think of this scenario: *imagine having repeated the process of looking in the movie theater once per night, on many nights, and counting the number of those nights on which you saw at least two acquaintances*. Kit's ambiguous use of the referent "it" in this interchange (line 407) is subtle, but it is significant insofar as it speaks to the important issue of the coordination of images we hypothesize are entailed in construing the problem scenario in a manner that supports conceiving of expectation as a quantity. This entails a scheme involving the composition and coordination of two levels of imagery; one level entails an image of accumulating a number of repetitions of the observation process, another level entails accumulating a number of "successful detections".

Table 1 displays students' written responses to the following post-test question, crafted to query their ability to conceive an event's expectation (and its potential unusualness) as a statistical quantity and to describe such in operational terms:

> "What does it mean, in statistics, that an event is "unusual"? (We know that *unlikely*, *unexpected*, and *rare* are synonyms of unusual, so mentioning them will not answer the question. Please explain the meaning, don't just give synonyms.)"

Table 1. Students' responses to a post-test question.

Student	**Response**[1]
Nicole	An unusual event would be one the most unlikely to occur. (I.E.- In 3-card poker it is unusual to get a 3 of a kind.) It's that something that occurs that wasn't predicted.
Sue	That mean there is a only few percentage of event occur during the longrun collection of samples.
Kit	When something happens a small % of the time.
Sarah	Of all the samples taken, or items tried, the usual occurance happens the least or close to the least
Peter	That an event is not likely to happen. If 1000 samples are taken and unusual event will happen about 5% or less of the time.
Tina	It means that., How unusual is it for this to occur? That it does not occur/show up as often
David	The event unusual means in statistics that there is a lower percentage ["chance"] that something unusual will occur. Like something unusual will probably only occur 10% of the time.
Luke	In statistics the word usually relates to the occurance that something occurs. When someone says that the results are unusual, then they mean that the results don't come out like this on a common occurance.

These responses indicate that roughly half of the students were able to give what we consider sufficiently quantitative characterizations of what it means to say that an event is unusual. The responses offered by Sue, Sarah, and Peter are the only ones that entail explicit mention and coordination of the two essential ideas we discussed earlier: 1) an idea of collecting multiple samples or repeating a process multiple times *and* 2) an idea of infrequency (or low relative frequency).[2] They thus arguably demonstrate an emergent ability to think of expectation quantitatively and therefore suggest understandings consistent with that targeted in our instruction. The responses given by Kit, Tina, and David also refer explicitly to the second idea, but only implicitly to the first. We coded these three responses as *ambiguous* because we could not ascertain how elaborated were their images of "percent of the time" or "does not occur as often".[3] Finally, Nicole's and Luke's responses contain far less evidence of an ability to operationalize "unusualness" and were thus coded as *inconsistent* with the targeted understandings. Nicole's response amounts, in places, to a rephrasing using synonyms for "unusual", whereas Luke's response is circular in its use of the term. Although both refer to the idea of an event or occurrence being uncommon or not predicted, they do so in a manner that seems non-quantitative.

CONCLUSION

To summarize, the evidence we have presented and discussed here indicates how difficult it was for the students in our experiment to think of and characterize expectation and related ideas of likelihood and unusualness in quantitative terms. Moreover, the discussion excerpts provide indications that the students were not easily oriented toward construing the task situations (e.g., the Birthday scenario and the Movie Theatre scenario) in a manner amenable to conceiving of expectation as a quantifiable feature of those situations. We hypothesize that a key factor in this difficulty was the lack of explicit cues in the task situations that might have evoked their prior knowledge of repeated sampling as a basis for quantifying expectation. Indeed, these task situations were designed so as *not* to explicitly mention quantities that they might have taken as constituents, under appropriate composition, of expectation as a statistical quantity[4]. In retrospect these activities thus served as indicators that the students' prior knowledge was evidently not sufficiently elaborated or robust to provide a conceptual substrate for quantifying expectation in contexts where doing so was not explicitly cued. More specifically, our students experienced significant difficulties in generating and composing requisite images and mental operations for conceiving expectation as a statistical quantity. Our study not only highlights the importance of those operations, but it points to a need to better structure instruction so as to support students in developing them.

NOTES

1. These responses were copied verbatim from students' test papers; they have not been edited for grammar or punctuation.
2. We make this claim notwithstanding Sarah's reference to "usual" instead of "unusual" in her response, which we take as a minor error.
3. It is possible, perhaps plausible, that such expressions encapsulated an underlying idea of repeating a process for these students. However, given our instructional and research goals, we were deliberately conservative in coding these responses.
4. This is in contrast to, for instance, typical rate story problems in which explicit quantities are provided that may cue the reader to use them in determining a requested quantity.

REFERENCES

Cobb, P., & Whitenack, J. (1996). A method for conducting longitudinal analyses of classroom videorecordings and transcripts. *Education Studies in Mathematics*, *30*(3), 213-228.

delMas, R. C., Garfield, J., & Chance, B. L. (1999). A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning. *Journal of Statistics Education*, *7*(3). Online: www.amstat.org/publications/jse/secure/v7n3/delmas.cfm.

Doerr, H. M., & Pratt. D. (2008). Mathematics and mathematical modeling. In M. K. Heid & G. W. Blume (Eds.), *Research on Technology and the Teaching and Learning of Mathematics: Vol. 1. Research Syntheses* (pp. 259-285). Charlotte, NC: Information age Publishing.

Drier, H. S. (2000). Children's meaning-making activity with dynamic multiple representations in a probability microworld. In M. Fernandez (Ed.), *Proceedings of the twenty-second annual meeting of the North American chapter of the international group for the psychology of mathematics education* (pp. 691-696). Columbus, OH: ERIC Clearinghouse of Science, Mathematics, and Environmental Education.

Konold, C. (2002). *Prob Sim User Guide*. Amherst, MA: Author.

Konold, C., & Miller, C. (1996). *Prob Sim*. Computer Program. Amherst, MA.

Kuhn, M., Hoppe, U., Lingnau, A., & Wichmann, A. (2006). Computational modelling and simulation fostering new approaches in learning probability. *Innovations in Education and Teaching International*, *43*(2), 183-194.

Mills, J. (2002). Using Computer Simulation Methods to Teach Statistics: A Review of the Literature. *Journal of Statistics Education, 10*(1). Online: www.amstat.org/publications/jse/v10n1/mills.html.

Pratt, D. (2000). Making Sense of the Total of Two Dice. *Journal for Research in Mathematics Education*, *31*(5), 602-625.

Saldanha, L., & Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*(3), 257-270.

Saldanha, L. A., & Thompson, P. W. (2007). Exploring connections between sampling distributions and statistical inference: An analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education, 2*(3). Online: www.iejme.com.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications.* Mahwah, NJ: Lawrence Erlbaum.

Stohl, H., & Tarr, J. E. (2002). Developing notions of inference using probability simulation 3tools. *Journal of Mathematical Behavior, 21*, 319-337.

Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Journal of Organizational Behavior and Human Decision Processes, 47*, 289-312.