GRANULAR DENSITY IN THE EVOLVING UNIVERSITY PROBABILITY COURSE

J. Todd Lee
Elon University
2320 Campus Box
tlee@elon.edu

The growth of STEM disciplines in middle-sized universities creates a new reality for the purpose and clients in an introductory probability course. In the investigated class (16 students), the instructor faced the challenge of providing a conceptual framework that is productive for the learning and application of probability to a broad range of majors. The course serves as an upper-level elective for various STEM majors, a required course for math education majors, and a prerequisite for theoretical statistics required for statistics majors. The granular density metaphor (Lee & Lee, 2014) was used as a core content element to address the needed multi-purpose structure. This paper discusses the initial findings on the viability of applying this metaphor to various course topics from examining different student artifacts.

INTRODUCTION

When it comes to teaching much of the theoretical foundations for statistics and probability majors, smaller institutions often must rely upon a single course. The institution of this study at one point had a single theoretical course in probability and statistics taken by all mathematics, mathematics teaching licensure and pre-engineering students. The necessity of this was not only mandated by class size, but also by being an institute with a 4-hour course structure. With the introduction and rapid growth of a statistics major, the single theoretical foundation course was extended into two classes; one for probability theory and one for theoretical statistics. It is not clear that at the undergraduate level, further courses are necessary. In various statistics courses, as well as statistical content in allied courses, pertinent probability density functions are given and used prescriptively for different techniques implemented via R or SAS; the understanding of a PDF to successful student may be cursory at best. Often PDFs are seen (and taught) as part of the machinery that is not particularly central to interpreting more immediate probabilistic indicators like confidence intervals or p-values.

On the other hand, interpreting probability in application is often plagued with biases and vulnerable to immature conceptions, or even worse, misconceptions (see Batanero, Chernoff, Engel, Lee, & Sanchez, 2016 for descriptions of common biases and misconceptions). Perhaps even a modest increase in students' understanding of probability (and its use in models) would be helpful. When advanced undergraduates take a class in probability, there is an undeniable opportunity for strengthening the students' applied probabilistic reasoning *and* to engage in action research to investigate how instructional strategies afford or constrain students' probabilistic reasoning (McNiff, 2016) . It is an opportunity to not only dig deeply into the nuances of interpreting the application of probability, but to do so in areas beyond the application of small discrete or empirical distributions.

This paper reports on the ongoing action research of an undergraduate junior-level probability class of 16 students, consisting mostly of mathematics and statistics majors. A deliberate attempt was made in the curriculum to emphasize the philosophical nature of interpreting probabilistic models, with a focus on three quite recognizable interpretations of probability--classical, frequentist, and subjective (e.g., Batanero et al., 2016). The course also included an alternative fourth, granular-density approach developed in part by myself (Lee T. & Lee, H., 2014; Lee, H. & Lee, T., 2009). In this paper, I describe the deliberate construction of the curriculum of a probability course to promote the students' ability to interpret a probability in a way that is consistent with how we represent probabilities, especially visually. I am exploring how well this granular density interpretation has become a focal point in their discussion of probability, presenting preliminary results from in-class discussions and assignments/quizzes, as well several post-course interviews.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

CURRICULUM

In the years before the introduction of a statistics major, the course in question was a theoretical probability and statistics course that covered a broad swathe of material. The course started with the discrete mathematics of sets, combinatorics, and definitions and axioms for discrete probability. The material is very much in line with standards for U.S. pre-collegiate probability curriculum and the U.S. finite math collegiate courses still taught as a first courses in business mathematics. Then formally introducing the idea of random variable, we would move quickly through more advanced discrete probability distributions and onto some of the analogous continuous distributions. From here, joint distributions may be introduced, but certainly a move is made through sums of variables and the introduction of mean, variation and the limit theorems. With the theme of limits comes the notions of statistical estimators and bias; with a good deal of time spent on maximum likelihood estimators. It seems from recent inspection, that the past couple of decades have brought a change in the collegiate curriculum, with a lot more time spent with the development aimed for Bayesian statistics instead of a previously pure focus on developing confidence intervals and null hypothesis testing (Borovcnik & Kapadia, 2014). A perusal of U.S. textbooks and courses named "Introduction to Probability and Statistics" seem to follow this list, including the current online course out of M.I.T (https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/). Learning is a time-consuming activity; thus, in a typical course, there is little time while trying to "cover" this list of topics to focus on philosophical interpretations of probability (though a professor can certainly pay homage to it in their own mind.)

At the institution at which the current action research is situated, the later part of the typical material has been moved to a second foundations course aimed purely at statistics majors. Not only does this add much needed breathing room to learn the quite large amount of material focused on probability theory, but it also recognizes that theoretical probability is an interesting field and an applicable tool in areas outside of statistics; useful to students not aiming to be statisticians. The probability course is now aimed at ending with a rather rich moment-generating approach to the Central Limit theorem. Neither joint distributions nor more general variable transforms are covered; however, sums of variables are covered, as well probability bounds such as Chernoff's and Markov's.

Foundational Material

An important aspect of this work is that a great deal of time in the beginning of the course is devoted to developing probability as a mathematical construct of definitions and Kolmogorov's axioms, along with philosophical visions of how these tools are used as models. This includes an emphasis on the difference between the mathematical construct and its interpretations in application. I nudge (shove) students towards meta considerations of the difficulties we as humans have with intuiting probabilities, and how considering probability from different perspectives can be helpful (Borovcnik & Kapadia, 2014). One does not have to look far to find an abundance of simply stated problems that can evoke common misconceptions (e.g., Fischbein & Schnarch, 1997).

The start of the course builds with a succinct step by step story of probability, complete with vocabulary we are careful to use consistently:

- We have a *Random Experiment*, a black-box process, a single instance of which is called a *Trial*.
- A *Trial* produces a *Result*.
- A *Result* is one of a set of possible *Outcomes* forming the *Sample Space*.
- *Events* are most, if not all, of the different subsets of *Outcomes* in the *Sample Space*.
- The *Event Space* is the set of all events.

Later in the course we define a *Random Variable* to be an added step of assigning real numbers to the outcomes so that our sample space is a subset of the reals. It is nice *not* to do this from the beginning since we can avoid the connotations of "random" and the added layer to be considered when modeling a real-world situation.

As a separate object, we consider a Probability function defined over the domain of the event space with a range of [0,1]. Any probability function must meet Kolmogorov's axioms. With this we spend time in a pure mathematics mode, filling out the properties that a probability function must have and covering set operations, and we introduce modelling as we explore scenarios for sample spaces.

Common Interpretations of Probability

In first example scenarios used in the course we built a model up to an event space. In these scenarios, there were usually compelling circumstances for the students to immediately define a particular probability function. And most times it was easiest to define the function over the events of single outcomes, making it particular tough to impress the idea that nevertheless, the domain is the event space, not the sample space. This is an example of a nuanced issue that a professor can seemingly over invest in unless you have the experience to know what can happen later in the class. I was fairly happily able to avoid much of the student cognitive resistance that comes later on the issue of singleton events having probability zero in the continuous cases.

It is at this point where the class started exploring what we actually meant for the probability to represent when we rushed to give values to different events. Most of the starting examples were of small chance-makers like fair die or coins, where some underlying set of events were intuitively equiprobable by the students. A discussion and thought experiments are conducted to sense out that "equiprobable" has a strong intuitive meaning for us, that at the same time is fairly difficult to verbalize. Using that intuition, we introduce the classical Laplacian interpretation of probability that if we can form a sufficiently fine mesh of events in the sample space, each of which can be intuitively see as being equiprobable with the other such events, then a complete probability function can be constructed. Later, when we discuss limit theorems, we explore the idea that appropriate models can be further validated when comparing the mathematical limiting values with observed samples created from multiple trials. An emphasis was placed on the accepted model fuzziness of what is "equiprobable", a kind of payment that is most always made in modelling.

Next, the students demonstrated a high awareness of, and intuitions about, laws of large numbers long before we formally addressed it in class. It is the notion behind validating any of our interpretations by taking samples and comparing empirical relative frequencies with corresponding probabilities. I flipped this around and introduced the frequentist point of view that we can interpret probabilities as tendency values observed in relative frequencies of events in ever larger sets of results. Note that I do not use the word limit, because we know the fallacies that can come from thinking of relative frequencies having limits in the calculus sense, something that many statistics text authors, teachers, and even the mighty Wikipedia alludes to. These students all have one thing in common, their deep, and probably successful classes in several semesters of calculus, and they know, at least at an intuitive level, that a limiting sequence can be deterministically forced to be within any fixed distance of the limit value by using a single bounding value on size of index. This is simply *not true* for the law of large numbers, and I openly discuss the seemingly obvious, but ultimately false beliefs that lead to things like gambler's fallacies. In this foundation period, I simply stuck with the phrase "tendency value" instead of an early introduction to probabilistic limits. We discussed that is was the primary interpretation for standard inferential statistics, and the lack of having any certainty statements in inferential statistics was do greatly to the nuances of tendency value versus limiting value.

The third interpretation we discussed was the subjectivist view, with a warning that this name was often used to include "gut feelings" for probability, where we were interested in the epistemological view. The probabilities that we assign are interpreted as a measure of our knowledge about the experiment and each of the events. I found this tough for the students to be comfortable with, but there was eventually rises in quiz scores as we explored examples like a weighted die. The story is familiar; we address the perfect symmetry of a cubic die, and with that knowledge make a statement of each face being equiprobable. However, as we roll the die and observe a high frequency of 3s, we are gaining knowledge, and with that knowledge we may readdress the assigned probabilities. I do go back to Bayes' theorem and point out the conditional probabilities based on what we have seen, compelling us to change our assigned probabilities. We also address the "fuzziness" of this interpretation, i.e., that different people with different knowledge of the experiment may well disagree on the assigned probabilities. They can settle this discrepancy by sharing knowledge, which can very easily lead to a third model. This contradicts the idea of probability being an innate thing "in" the real experiment; I claim this is a great benefit as we want students not to think of models as what is "true" about a real-world system, but rather the model is a predictive tool and perhaps metaphor for the "Truth".

Finally, we summarized the previous interpretations, and did a great deal of formative assessments in class using an online quizzing that allows for anonymous answering and immediate display of class responses. It is the case that class showed a distinction between students who were doing what it takes to "keep up" with the class, and those who didn't. However, it worth noting that in the plethora of factors involved, it is nonetheless true that most students did poorly on the first time we quizzed anything. This is not just this class or my classes; in general teachers that do a lot of formative quizzing and re-quizzing see the real disconnect between where the class thinks it is at, and the actual truth. The re-quizzing and discussions do, however, prompt serious progress towards common understandings and vocabulary among the students.

Granular Density

Once students assessed well on the prior, we moved into the granular density interpretation and our first serious use of visualizing probabilities. My aim was not to reinforce prior conceptions of visualizations until we were ready to immediately move into probability density histograms instead of probability/relative frequency histograms. I acknowledge that most college students have common background experiences in using descriptive graphs for data analysis, and how they almost always use functional distance from the sample space axis as the value indicator (e.g., frequency bar graphs). However, this visual interpretation is seriously problematic when we move to continuous sample spaces. The widely used textbook trick of comparing probability/relative frequency histograms of bin width 1 to PDFs is mathematically correct, but seems almost counter-productive in transferring student's focus on how probabilities are being stored in area rather than height of bins. In prior work (H. Lee & T. Lee, 2009), a paradigm was introduced to promote a probability interpretation that naturally invokes area as best indicator of probability.

The granular density interpretation has a Popper's propensity approach, where I ask students to consider what attributes of a real world scenario makes it a good candidate for a probability model. For many of these cases, like even the simple coin flip, the experiment has a large phase space of multiple initial variables (e.g., speed, force, wind, initial position of coin) which all contribute in a complex manner to the final result of a trial. Consider that it is the variation in the phase space that leads to different results, because if we built a machine in a vacuum, that flipped a coin with the exact same physical conditions, the coin would land on the same side every time. Thus we could create an intermediate sample space of the phase space of contributing variables, and imagine dividing it up into a very fine mesh of regions such that any two points in the same region resulted in the same outcome and that the experiment was equiprobable across the set of regions. If we metaphorically associate these regions as tiny grains of "probability sand", the total of which is unit 1, then we can assign probabilities to the final event space as a distribution of the sand across the sample space.

For instance, if we have a coin where we are told that it is weighted to have a 0.7 probability of flipping to a heads, then we can illustrate this as a bar chart of heads and tails where 0.7 units of sand are piled over heads and 0.3 units of sand are piled over tails. All individual grains of sand are equally likely to be chosen at random. One trial can be illustrated by randomly picking exactly one grain of sand and lighting it up; whatever outcome the lit grain is attributed to is also the trial result.

And with this example, we see how the emphasis in visualizing probabilities is on area, as using the sand metaphor invokes thoughts of distributing a unit of sand across some subset of the real number line. Students did not seem to be bothered by considering sand as a 2-D construct, though I do wish to teach a class introducing joint distributions to carry over the sand metaphor via piles on top of rectangles in the plane.

Along with the interpretations discussions, the class progresses up through defining a random variable and working through some typical discrete distributions, with visualizations being point mass graphs. The timing is worked so that the granular density interpretation is discussed while reviewing the discrete distributions and then launching into continuous random variables, along with density bar charts, histograms and graphing probability density functions.

REFLECTION ON STUDENTS' THINKING

Throughout the class we had numerous meta-conversation about vocabulary, pronouns, context, etc. Our assessments used the vocabulary precisely, very much pushing the students of having a working understanding of even nuanced usage. This became most apparent in the heart of the moment-generating material where students where struggling to understand why we would we even care to have such a creature from the depths of Calculus II. We were in the latter parts of the semester, and I noted several times, students would back track on the vocabulary and self-correct their questions when I displayed the first note of confusion. Students would even back-handedly correct each other's questions. However, the most powerful sign was when the students were in small groups discussing something and they would continue to self-correct themselves on their usage. This is an important step.

Most importantly to me was how well students were developing their conceptualization of a probability density function. The students quickly grew to understand that I thought probability sand was an important concept and that they should try to use it, though their preference for calling it "magic sand" was not shared by me. Students knew the granular density approach was not standard in concept or language, because there was not the slightest hint of it in the course text material. The text was an open educational resource and available online. The students did make the mistake of invoking sand in places where it had nothing to do with anything, but at other times they would invoke it precisely in the right place, with the right conclusions.

In one instance the sand metaphor helped explain the use of "density" in a probability density function. The student asked why it was named such, and instead of going down the text book explanation, I asked the students to imagine the horizontal axis as a thin rod and we compressed the sand down along the rod to an equal thickness, though we might have to compress with much more force in some areas than others. Would the density of this "sand rod" be equally dense? Where would the density be higher? They appeared to understand this quickly.

There is an interesting event in class that left quite the impression on me. Long ago, when I first realized a standard approach statistics textbooks use to move from discrete bar charts to histograms, and finally to a PDF, I began telling the story in classes whenever it was appropriate (Lee, T., 1999). As I pointed out earlier, the idea is that *if* you set the width = 1 for each rectangle, in a relative frequency histogram, then you can over lap the relative frequency histogram with a PDF using the same vertical axis and they will be comparable. Yes, this is interesting to know if you are using software that can't produce density histograms, but as a transition tool, students just do not absorb the nuance and are even more likely to look to the height of the PDF for probabilities. Regardless, this has almost always been unimpressive to most students, as they are often puzzling over my comment of the height of the PDF not meaning anything. However, in this class, I simply described the act of overlapping an empirical relative frequency histogram with a corresponding PDF, then stopped and asked them to discuss why would we do this and what could we conclude. When reporting back, the seven groups agreed that the tops of the two graphs would "line up" if the sample had been large. The groups were roughly split in further reactions; half wanted to know if this implied that the PDF gave probabilities, as they were under the correct impression that was not the case. The other half of the class wanted to know if it was true that if they kept making the bin widths smaller, would the tops of the bins simply disappear? (No, they would stop at bars of height of 1, larger when there were exact overlaps.) Either way. I finally had a class that was not thinking of PDFs as probability functions. But what about after a few months after class?

As of the writing of this paper, interviews had been conducted with two students in the semester following the probability course. What follows are preliminary reflections on students' thinking. The interviews were one-on-one in a relaxed setting, with the students having access to a iPad Pro notes app with the questions typed one per page and an Apple pen to write and screencast recording the screen and sound in sync. In the interview, they were given three questions, with freeform voiced follow-up questions and discussions. The original three questions were

- What is a PDF?

- Give at least two different ways of explaining the following sentence to a friend, “We have a seemingly regular six-sided die that has probability 3/5’s of rolling a four.”
- You own a paper clip company and you are considering buying a machine that will produce the wiring for the paper clips. The machine’s manufacturer sent the following partial sketch of the probability density function for the wire diameter produced by the machine. What else (if anything) would you need to know on the sketch to decide if you are going to buy the machine? (The sketch contains to un marked axes and a bounded distribution on in the first quadrant.)

During the interviews, students confidently use the vocabulary, but the terms have become conflated with each other, especially in using “event” almost exclusively instead of “result” or “outcome”. I was careful to use the correct terms any time I spoke, but without emphasis. I did not notice this having an impact.

When the students faced the first and third question, they immediately mentioned probability sand and proceeded to create PDF curves and fill the area with dots representing the grains of sand. However, they were referring to these grains almost exclusively as representing different events, with little sign (even up to aggressive prodding), of showing a connection of the grains with a tiny range of conditions that lead to a single outcome. When the interviewing students discussed the grains in the sand/area below PDF curves, only once did a student use the individual grains “lighting” to represent the result of a single trial. I take it as confirmation of my fears that this part of the metaphor does not get taken into the students’ conceptualization. This does raise the question of what exactly *has* taken hold; what do the students really mean when they discuss the grains.

Probably where I expected (wanted) to see the granular density metaphor was in the second question, but this was not to be. One student gave a fairly solid explanation using a frequentist approach and then just could not give another explanation. The other student found this difficult, with the personal reasoning for difficulty being that they simply, deeply knew what this probability meant. I asked if the probability where changed to other values, would they continue to feel the same way, and they budged into a relative comparison of size, suggesting that if the 3/5s became much bigger, then getting a four would become “very likely”. This very same student, however, walked themselves slowly through a very compelling argument for the third question. They discussed needing to know how tightly the majority of the area was held within an interval centered around the desired diameter value. It was in this same discussion that at one point the student had sketched and named a normal PDF, even after they struggled to sketch and name any specific PDFs in the first question. For the next interviews I will definitely start digging more into how much a visualization can prompt the granular density metaphor. If either of these students had been additionally supplied a sketch of a possible PDF for the die values, would they then have kicked in with granular interpretations?

REFERENCES

Batanero, C., Chernoff, E., Engel, J., Lee, H. S., & Sanchez, E. (2016). *Research in teaching and learning probability: An International Congress of Mathematical Education Topical Surv*ey. (40 pages), Springer.

Borovcnik, M., & Kapadia, R. (2014). A historical and philosophical perspective on probability. In E. J Chernoff & B. Sriraman, (Eds.), *Probabilistic thinking: presenting plural perspectives* (pp. 7–34). New York: Springer.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for research in mathematics education*, 96-105.

Lee, H. S., & Lee, J. T. (2009). Reasoning about probabilistic phenomena: Lessons learned and applied in software design. *Technology Innovations in Statistics Education 3*(2), 1-25.

Lee, J. T. (1999). Discuss with your colleagues: It's all in the area. *Mathematics Teacher 92*(8), 670-675.

Lee, J. T., & Lee, H. S. (2014). Visual representations of empirical probability distributions when using the granular density metaphor. In *Proceedings of the International Conference on Teaching Statistics*. Flagstaff, AZ, July 2014.

McNiff, J. (2016). *You and your action research project* (4th Edition). Routledge.