An active learning approach to graduate level applied probability

Robin Pemantle

University of Pennsylvania

pemantle@math.upenn.edu

Midwest Probability Colloquium

22 October, 2022

Outline

1. The students and their needs
2. Curriculum: principles and meta-curriculum
3. Pedagogy: flipped classes, active learning, applied exercises
4. Assessment: oral exams, final projects

Who is the course for

- Ph.D. students in fields such as CS, genomics, engineering, biostat
- Masters students in Math and Applied Math
- A few advanced undergraduates from Math or sometimes Economics or Computer Science.
- Note: no Math or Stat Ph.D. students - they all take the regular course (for example, a full year course out of Durrett)

Backgrounds of these students

Some are mathematically strong enough to take the regular course. Especially we see students from China, Korea and some European countries with an excellent real analysis background.

Others are borderline on the real analysis background to take this course, and even more borderline for the regular course, but they really need a deeper and more formal understanding of this material if they are to have a hope of reading about it on their own and using it in their work.

Before the Applied Probability course existed, all these students would all take the regular course. It was very overloaded.

Aims of applied probability students

Students who take the Applied Probability course, are generally more interested in using probability in a field of application than in than in probability theory, *per se*. They will be a great asset to their research group if they can:

- Propose a stochastic model and guess how it turns out
- Recognize well studied probability models embedded in other processes
- Have some familiarity with basic tools (tail bounds, conditional expectations, large deviations, characteristic functions, Poisson and Dirichlet processes) and know when and how to use them
- Read a paper using probability formalism

Why this course

Other available courses:

- Regular two semester grad course, e.g., out of Durrett
- Standard one semester UG course, e.g., out of Ross
- Courses in other departments with Hoel-Port-Stone approach: do what you can that does not require technicalities

These courses all omit crucial skills:

- Writing a probability model for a verbal scenario
- Reading an article using significant probability theory
- Formulating conjectures about behavior of a system and having intuition about what this might be

What I do in this course

- Active learning
- Concentrate on firm understanding of constructions and definitions
- Ditch the less relevant (and often longer) proofs
- Emphasize the skills on the previous slide

The rest of the talk is about how these things can actually be achieved.

Disclaimer

I will talk about *principles*, using the course I developed as an example. All my course materials are freely available via Dropbox share.

Disclaimer: your students will be different from mine in subtle and unpredicatble ways. My materials can be a starting point but you will need to put in a fair amount of work the first couple of times, adjusting materials and creating new ones.

That's why the principles are important.

Pedagogical principle: Measure theory

Measure theory is needed when it aids understanding.

- Measure theoretic results are carefully defined and quoted, but no proofs are given. However: students should always know what formal object they are dealing with!
- In the beginning they are always allowed to think of $\mathcal{F}$ as the power set of Ω.
- Later, for conditional expectations, we discuss σ-fields in more detail.

Pedagogical principle: Measure theory

Measure theory is needed when it aids understanding.

- Measure theoretic results are carefully defined and quoted, but no proofs are given. However: students should always know what formal object they are dealing with!
- In the beginning they are always allowed to think of $\mathcal{F}$ as the power set of Ω.
- Later, for conditional expectations, we discuss σ-fields in more detail.

(Actually, this is how I teach the regular course as well.)

Pedagogical principle: proofs

- This course carries mathematics graduate credit. Therefore, the general approach is rigorous foundation.
- Proofs are to aid understanding. Students have very different needs, some needing details spelled out, others relying more on pictures, examples and analogies. Terminal course: not training for future, harder, proofier courses.
- Rule of thumb: there's no point in proving something if including the proof displaces the time needed for examples, applications, etc.
- Caveat: never cut corners on definitions or constructions.

Meta-curriculum

- Intuition: seeing in the mind's eye how random sequences behave
- Hands-on understanding: if you can't simulate it, you don't understand it
- Reading skills: reading articles and learning from the textbook on your own
- Speak the Language of Asymptotics

Meta-curriculum

- Intuition: seeing in the mind's eye how random sequences behave
- Hands-on understanding: if you can't simulate it, you don't understand it
- Reading skills: reading articles and learning from the textbook on your own
- Speak the Language of Asymptotics

- Intuition: always ask students to guess first. Sometimes it leads to a great conversation.
- Hands-on: some of the homework should be simulation.
- Reading skills: flipped textbook; term project.
- Language: hear me do it, now you do it.

Curriculum outline

A perturbation of the usual graduate course up through martingales.

- Measures, Lebesgue integral, conditional expectation (3 weeks, including a primer on asymptotics)
- Weak laws, Borel-Cantelli, Strong laws, maximal inequalities, 3-series theorem (2 weeks)
- Distributional convergence and tightness (1 week)
- Large deviations; Poisson processes; Arcsine laws; Characteristic functions / CLT (1 week each, total of 4 weeks)
- Martingales (2 weeks)

Fitting it all in

The meta-curriculum is demanding. How can one fit in these extra empahses, with a less prepared median student, and an active-learning discovery approach?

Skipping hard proofs helps but is nowhere near enough.

Need the students to do a complete pass through each chapter before the week starts. Therefore, this course is more work for the students than it looks like it will be. Best to let them know this up front!

Self-check questions enable students to know whether they are reading in enough depth. It doesn't hurt to give some of them some extra help on the topic, "How do I read a mathematics text?" This is a good place for an auxilliary optional video.

Flipped classes

- Mostly reading and self-check
- Some videos when topics require them
    - Probability triples: writing probability models
    - Asymptotic notation and how to use it
    - Conditional expectation
    - Types of convergence
    - Tightness and compacitifications
    - Large deviations: optimize upper bound; tilted measures
    - Poisson process: intuition behind the formalism
    - Stable laws: summing points of compensated Poisson processes
- Videos enrich rather than repeat what is easily gotten from the text

Examples

Examples of self-check

Measures and random variables

What measure theory do we need to know? Enough to be able to compute measures of interesting events, and expectations of interesting random variables (we haven't defined these yet).

1 σ-fields

Definition 1.1 (σ-field)**.** *A σ-field on a space Ω is a subset $\mathcal{F}$ of the set of all subsets of Ω, that is closed under countable set operations (unions, intersections and complements). In particular, $\mathcal{F}$ is a σ-field if*

(i) $\emptyset, \Omega \in \mathcal{F}$*;*

(ii) $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$*;*

(iii) $A_1, A_2, \ldots \in \mathcal{F}$ *implies* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$*.*

Elements of $\mathcal{F}$ are called **measurable sets**.

Example 1.2. The power set (set of all subsets) of Ω, denoted 2^Ω, is a σ-field.

Exercise 1. *What are all the σ-fields on the three-element space $\Omega := \{A, B, C\}$? Can you generalize this to any finite set?*

3.1 Poisson process axioms

Let $(S, \mathcal{S}, \mu)$ be a measure space with measure μ that is non-atomic and σ-finite. We would like points to spring up randomly, with each small patch of measure $d\mu$ to have probability $d\mu$ of containing one of the points. We also want the numbers of points springing up in the sets $A_1, \ldots, A_n$ to be independent if the sets $\{A_j : 1 \le j \le n\}$ are disjoint. What kind of formal object must we be dealing with? In our minds, we see a random collection of points (in time or space). It is an unordered collection, which gives us grief notationally. Rather than trying to construct it as a random set, it turns out much easier to construct it as a random measure. That is, we associate a sample set W with the measure taking a set A to $\nu_W(A) := \#(W \cap A)$. This measure has the property that $\nu_W\{x\}$ is either zero or one for every x. We call such a measure a **counting measure**. The correspondence is invertible: if ν is a finite measure on Borel sets taking values in $\mathbb{Z}^+$ then by repeated halving, we may find a finite set supporting ν. Thus ν is atomic and if $\nu\{x\} \le 1$ for every x, then we see $\nu = \nu_W$ where W is the support of ν. With this in mind, we have the following definition.

Definition 3.1. *Let μ be a σ-finite measure on the space $(S, \mathcal{S})$. A Poisson process with intensity μ is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with a map $N : \Omega \times \mathcal{S} \to \mathbb{Z}^+$ such that*

(i) for each $\omega \in \Omega$, the map $A \to N(\omega, A)$ is a counting measure on $(S, \mathcal{S})$;

(ii) If $A_1, \ldots, A_k \in \mathcal{S}$ are disjoint sets of finite measure then $N(\cdot, A_1), \ldots, N(\cdot, A_k)$ are independent Poisson variables with means $\mu(A_1), \ldots, \mu(A_k)$.

Exercise 4. *Let N be a Poisson process on $(\Omega, \mathcal{F}, \mathbb{P})$ with intensity $(S, \mathcal{S}, \mu)$. Given $T \in \mathcal{S}$, define the process N_T by $N_T(\omega, A) = N(\omega, A \cap T)$. Is N_T a Poisson process? If so, with what intensity? If not, which axiom is violated?*

Alternating classes - assume 2x90 minutes per week

[Weekend: students do the reading. It helps if you have a chat board and can monitor it, but even unmonitored, often students help each other.]

Mondays: assimilating content. Q & A on reading, mini-lectures, going over self-check exercises.
The first week, because there's no pre-reading, we discuss the mini-unit on asymptotics (posted for you).

Wednesdays: work on problem worksheets in small groups. Things work better when homework is lagged by one week.
The first week, we do a problem set on formal modeling (posted for you).

If you have enough students for a grader, best to have the grader help in class on days with group work.

What is Q & A like?

The most success I had with the Q & A portion of the course was the third, most recent time I taught it.

I invited any and all questions on the material *except* those on the self-check questions. Those were supposed to be done in advance, not just attempted but discussed, help sought, and figured out.

The result was some very interesting questions such as these.

- I've heard that Lebesgue integration is integrating in the y-variable instead of the x-variable. Can you explain?
- Given that the condition $t\mathbb{P}(X > t) \to 0$ is strictly weaker than $\mathbb{E}|X| < \infty$, is there a sharp tail bound condition for finite first moment?

Examples

Examples of worksheets

1. Let $(\Omega, \mathcal{F}_0, \mathbb{P}) = \left([-1,1], \mathcal{B}, \frac{1}{2}\mathbf{m}\right)$ and let $\mathcal{F}$ be the σ-field of all sets $A \cup (A \oplus 1)$ where A a Borel subset of $[-1,0]$ and $A \oplus 1$ is the translation $\{y+1 : y \in A\}$.

What is $\mathbb{E}(X \mid \mathcal{F})$, where $X(\omega) = \omega$?

Before this problem the students should have read the guess-and-check definition of conditional expectation and should be encouraged to have a copy on hand.

Also, beforehand, I like to give them a game version. You get to ask as many questions as you like of the format, "is ω in the set G?" If $G \in \mathcal{F}$ then I have to answer. After your questions, you have to give your best guess as to the mean of X.

Most groups didn't know how to start so we played the game. Many tried asking if $\omega \in G$ for some $G \notin \mathcal{F}$ so I told them I wouldn't answer and they had to wait till I came back around. No one made this mistake a second time.

Modeling

Perhaps the most important point: EMPHASIZE MODELING.

Students find this incredibly difficult!

The way to teach this: I do it, now you do it.

Aside: this problem seems universal in math education. Try asking your calculus students to write a differential equation rather than solve one! It is ubiquitous in K-12 education: students can't do word problems. The learning curve for this in applied probability is so steep at first that it is highly gratifying.

I also give a handout (link given a few slides later).

Example 1

1. Early on I asked students to create a formal probability model for this:

> *An insane gambler faces an infinite sequence of even money bets, betting every thing he has and letting his winings ride each time.*

Correcting their mis-steps and helping them to see where they could make more efficient choices was perhaps the greatest learning rate of any in-class activity.

Student reactions

An insane gambler faces an infinite sequence of even money bets, betting every thing he has and letting his winings ride each time.

- There was confusion as to what random variables they would eventually want to be able to define.
- Many began by trying to model the state at time n directly, but then saw (sometimes with help) that it would be better to model each wager and let the state be a function of that.
- The idea of creating some random variables that may never be used escaped some of the students at first.
- They were unsure as to when they had completely specified the probability measure.

Example 2

2. Try asking students to create a formal probability model for the birthday problem. What is the probability space? What random variables do you define and what are their interpretations? What events are we trying to understand?

It's good for students to see that this problem allows different reasonable choices, with different advanages.

- Should you create a different probability space for each value of N, the number of days in the year? (It's easier, and there's little benefit in not doing so.)
- Should you create a different probability space for each value of k, the number of people in the room? (It's not any easier and if you don't use a single probability space you can't define the first collision time.)

Birthday bonus

Bonus: it turns out the birthday problem is a good place to try out skills at asymptotic analysis.

The fact that $\tau/\sqrt{N}$ converges in distribution to a Rayleigh, where τ is the first collision time, is a straighforward application of some easy Taylor approximations.

How far into the tail does this work? In other words, How fast can g grow and still have $t \le g(N)$ imply

$$\mathbb{P}(\tau > t\sqrt{N}) \sim e^{-t^2/2}\,?$$

Poisson processes

Several steps are required for models involving Poisson processes.

First, students have to get the Poisson process formalism right (an exercise in reading comprehension, to be sure).

Secondly, they have to choose a model in which it is possible to define the relevant basic random variables and events.

Thirdly, they may have to construct some more complicated random variables, as in queuing processes or graphical models for interacting systems.

Q & A followed by guided exploration works well for this.

Simulation exercises

I give one problem set early on that involves simulation, then sprinkle simulations throughout the course (averaging maybe one small part of a problem in half the problem sets).

Materials in this talk, such as these, can be viewed at https://www2.math.upenn.edu/∼pemantle/5460.html

It starts with a brief explanation of a somewhat general Pólya-type urn process, then has for multi-part problems. The second is theoretical but the others involve simulating and conjecturing.

The actual simulation gives the students an idea of a thought experiment: how to "run the movie" in their mind.

Capstone problem on Poisson processes

Example: students are asked to create a probability model for **Lilly pad percolation**. Part of this is to define the event of a left to right crossing. Then they are asked to simulate it, including the (algorithmically nontrivial) task of automatically checking for a L-R crossing and tabulating the frequency with which one exists.

1. The technicalities of the construction are challenging, except to the computer science students.
2. This kind of exercise helps students learn to formulate conjectures and high level thoughts about what makes the model tick.

Training yourself

Certain elements of this pedagogy seem to run themselves. They usually go smoothly from Day 1.

- **Self-check questions:** work well if they're not too hard.
- **Q & A day:** Sometimes at the beginning, no one speaks up. But if you're quiet, patient, looking out over the audience, refraining from saying what questions you think they should ask, the questions usually start. Verbally reward for asking without patronizing (think a little about this beforehand) and the stream will continue. In the rare cases where no one wants to ask anything, skip to the self-check quiz and tell everyone they can go home!
- **Oral exams:** the only hard part is taking adequate notes.

The hard part

Supervising group work is the hardest aspect to learn to teach. Here are a few pointers.

- Use good problems: low threshold high ceiling.
- Things go better with well tested problems, but you will sometimes need to make up new ones. It's OK if these don't work so well the first time. Immediately after class write down your criticisms and adjust for next year. That goes for well tested problems too.
- Dynamics between you and the groups take years to perfect but you can get better fast if you have a good source of wisdom (next page) and a partner with whom you observe each other's classes.

Resources

The book "Building a Thinking Classroom" by Peter Liljedahl was written for K-12 teachers but almost everything in it can be applied to graduate students!

A set of my tips, written for calculus but applicable to this course, is posted for you at the site previously mentioned.

Some fundamentals are:

- Have the students work at whiteboards, not tables.
- Randomize groups and re-randomize two or three times
- Scan the room for groups that are stuck - no productive conversation - and learn good ways to unstick them
- Engage with their logic - use how to turn dead ends and contradictions into learning points

Principles

1. Assessments should reflect what is valued.
2. Assessments should encourage behavior you want to see.
3. Assessments should seem fair, non-arbitrary, and predicatable by the end of the course.
4. Assessment should not eat up too much of your time or theirs.

Reflect what is valued

Part of the learning is specific content, part is experiential. Accordingly the students should earn some significant part of their grade by coming to class and participating. For me, last time, it was 23%.

When a student was significantly late they lost half the attendance for that day. Usually this gets the point across the first time.

Very rarely do I have to point out that being a wall flower does not count as participation. It doesn't usually fix the problem, but it does allow you flexibility at the end to give a low score, leading to a low grade when you know it's deserved.

Incentivize

For this to work it is essential that students do the reading before class each week.

Self-check exercises therefore count for significant credit, also 23% in my last instantiation of this class.

The mechanism was a quiz at the end of each Monday Q & A session. This was an open-notes quiz, where you could simply copy what you'd already written. Therefore I could give two problems in 10 minutes. I usually excluded any of the self-check problems that were harder than average and tried not to be too predictable otherwise.

Homework (31%) was important, often teaching new concepts.

Predictability

Self-check and attendance create little variance, but homework is tough and gives the student a pretty good sense of where they stand.

Most semesters I gave a 15-minute oral midterm and a final, with the option of replacing the final by a term project. I felt that the oral exam was astonishingly accurate. As a result, the students were never surprised by their course grade and had already bought into the message it represented.

Oral exams

Actually this is something I started in my regular grad class because of the pandemic, but I like it a lot and it fits with the emphasis on ability to communicate.

Obviously feasibility depends on class size, but the tradeoffs aren't as bad as you might think.

I find that 15 minute oral exams suffice to get a very good idea of student's general levels of understanding.

Example: with 20 students, 5 hours of exams shoots my whole day, but then proctoring and grading 20 exams shoots at least a whole day.

How to conduct oral exams

- Random bank of 10 easier definition/construction questions and 10 harder "what happens here and why?" questions
- nearly equal difficulty within question banks
- Using Excel, create a randomized selection of two questions per student, one from each bank of 10
- Leave room on the spreadsheet to takes notes on students' performances

The oral exam format allows me to be liberal enough with hints to keep the pace going forward, so 5 minutes per question is not unreasonable.

My notes might say, "got it, but with a lot of prompting."

Optional final projects

Students can choose to do a term paper project instead of taking the final exam. To do so, they find a research paper they want to read, either from a list of papers I post on their Canvas site, or in their area of application, from their own pre-existing interests and exposure.

Project goals:

- Learn to read a research paper
- Extend knowledge by epsilon, via empirical work on real or synthetic data
- (start to) learn to write a publishable paper

For the Masters students, this is often **exactly** the training their degree program is supposed to provide.

Sample readings

E. Abbe.
Community detection and stochastic block models: recent developments.
J. Math. Learn. Res., 18(177):1–86, 2017.

William Aiello, Fan Chung, and Linyuan Lu.
A random graph model for massive graphs.
In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 171–180. ACM, New York, 2000.

William Aiello, Fan Chung, and Linyuan Lu.
A random graph model for power law graphs.
Experiment. Math., 10(1):53–66, 2001.

A. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286:509–512, 15 October, 1999.

William T. Barry, Andrew B. Nobel, and Fred A. Wright.
A statistical framework for testing functional categories in microarray data.
Ann. Appl. Stat., 2(1):286–315, 2008.

S. Bubeck.
Convex optimization: algorithms and complexity.
Found. Trends Mach. Learn., 8(3–4):231–358, 2015.

B. Cooper and M. Lipsitch.
The analysis of hospital infection data using hidden Markov models.
Biostatistics, 3(2):223–237, 2004.

Peer critiques

I will go into detail about the process by which I have learned to make it useful for them but not a time burden for me.

Each paper gets one peer critique. The critiques are graded (this year, 180 points for the paper itself, versus 50 points for the critique).

Critiquers get guidelines as to what to look for. Critques can miss the boat if they praise the paper for being intriguing while missing the vagueness or incomprehensibility of the writing.

Both parties learn a lot from the critques. Student evaluations give high marks to the term paper/critique process.

Process from beginning to end

- Select topic (may involve some conversation with me but not usually more than a few minutes)
- Turn in an outline (by about the 2/3 mark of the course)
- Rough draft
- Critique
- Final draft, due the last day of class, which is reserved for student presentations.

At each step, a couple of students miss the deadline and realize maybe they don't want to do a term project after all. This is fine!

30 minute rule: I spend no more than 30 minutes per student on their final project (except for grading it afterward, but usually I am pretty familiar with it by then).

If the outline is no good, they burn their 30 minutes with me fixing it up, and have do rely on the peer critique for improvement on revision. Ditto, if the rough draft is not good enough to send for peer review, they get half an hour of help, then rely on peer critique for the rest. Else, they get my comments on their rough draft as well as the peer review.

Accounting: 30 minutes includes the 15 that the oral final would have taken. In addition I don't hold office hours the last two weeks because of all the time I'm spending with students on their papers. The papers still cost me a little time, but are clearly worth it for students.

Summing up

- Students learn to model
- Students who do term projects learn a lot about the research process (short of solving a research problem)
- I learn a lot about the students; I know what their grade should be without having to add up the scores
- Students get a sense of connection between probability and their own field
- Cost: students are less equipped to do research in probability theory because they don't know the inner workings of classical theorems

THE END

Thanks for listening