Accompanying Notes for Introduction and History of Probability Slides

Forward

The following material is provided for your education and to help give you perspective on where our academic discipline comes from. Do not worry about memorizing anything that follows, or studying it too carefully. It is provided in the spirit of giving you a welcoming first lecture and an enlightening introduction to the field.

What is probability?

So what is probability? And where does it come from? Perhaps the most salient source of probability in your daily life comes from your own efforts to understand random phenomena. What are the odds of dice coming up in your favor? What are the chances of two of us having the same birthday?

Whether or not the universe has true randomness[1], we can't say for certain, but regardless of the answer, modeling randomness with probability has become an incredibly fruitful and powerful human endeavor. Through modeling probability, this kind of quanta determining chance events, we can help to explain the long-term behavior of random events, and sometimes quite well.

How do probability and statistics differ?

Let's contrast two examples:

- Example 1: I give you a dataset on a perfect randomized clinical trial and I ask you the question: what was the effect of the drug comparing arms of the treatment? How confident are you in your answer? These are statistical questions, and they inherently depend on modeling assumptions, choices in methodology, etc.
- Example 2: Can you tell me what the probability of rolling k 1s in a series of n dice rolls is? $\rightarrow$ This has an exact answer. (The answer can be obtained by using the Binomial distribution, which you will be formally introduced to soon.)

One point of distinction between the two paradigms is that the theory of probability often has exact answers, though often these exact answers are in terms of probabilities and model parameters, while in statistics in practice, there is often an element of uncertainty around to what degree we believe the model assumptions hold, and often there are competing methods for the same problem that pose challenges and tradeoffs to the data analyst.

Some contrast these as deductive vs. inductive reasoning: in deductive reasoning, we start from premises and build up and up and up, creating more theory as we go. In inductive reasoning, we're often starting with the results, the data at hand, and trying to work backward to understand "what mechanism explains this data?"

Let's do an exercise

Consider the two sequences of coin-flips. Let's say one is produced by a student who has faked the results, and one is produced by recording the results of actual coin-flips. How might you tell the difference between the two?

Several key features you might look out for:

- Does it come up heads as often as tails?
- Is there any "extra structure" that shouldn't be there?
  - Like, it would be weird for it to be time-dependent.
  - We can ask: do heads come up after heads too often?

Think about the *morals* here: it *helps* to write down our assumptions as probabilistic statements. For example, we believe that:

- $P(\text{Heads}) = P(\text{Tails}) = 0.5$
- Each subsequent coin flip is independent of all past coin flips

---

Gambling is old. Really old.

One of the earliest *games of chance*, Senet, has physical archaeological evidence dating back to 3000 BC from Ancient Egypt[2]. Others, like Astragali, have featured in cave paintings from Mesopotamia and Grecian pottery dating back to 3600 BC.

Of course it would have helped to have any additional understanding of games of chance.

---

[1] Actually, it was one of our main characters who is credited with the first discussion of *causal determinism*: Pierre-Simon Laplace (1749-1827) wrote in 1814: "We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be the present to it." This idea later became known as Laplace's Demon.

[2] https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/13%3A_Games_of_Chance/13.01%3A_Introduction_to_Games_of_Chance

Liber de ludo aleae (The Book on Games of Chance)

Gerolamo Cardano (1501-1575) was a gambler[3]. On his 25-years of gambling, he said

> . . . and I do not mean to say only from time to time during those years, but I am ashamed to say it, everyday.

Perhaps as one might expect due to it being the 16th century, he made all kinds of mistakes in his book. One mistake he commonly made was called "reasoning on the mean;" With dice, the probability of a 6 coming up on a single roll is $1/6$, and in 3 independent trials, Cardano thought your probability of getting at least one six was $3 \cdot 1/6$ or $1/2$. It's not. Nowadays we could easily perform this calculation a number of ways (Binomial distribution, Inclusion Exclusion principle, or ... just count them!)

But he did get many things right. He wrote down what many consider the first definition of the term *probability*, which made reference to what he called a 'circuit,' though we'd now call that a sample space today. He understood that while probabilities (of independent events) multiply, odds do not. He discussed the concept of the Gambler's Ruin, i.e., that games with what we'd now call negative expected value will eventually bankrupt a player, as well as various gambling related paradoxes.

Unfortunately for Cardano, he could not separate his concept of Luck from Probability. He spoke of some mysterious force, the "authority of the Prince," which he thought could result in someone's throw "tending more in one direction than it should and less in another." Some scholars say that because he failed to recognize that such variability is inherent in games of chance, he relinquished his claim to founding the mathematical theory of probability.

Though Cardano wrote his text *Liber de ludo aleae* during the 1500s, it was not published until 1633, shortly before Blaise Pascal (1623 - 1662) and Pierre de Fermat (1601-1665)[4] developed the concept of expected values to address gambling related puzzles and problems more systematically. Specifically, they wanted to understand how a pot of winnings should be divided when ending a game prematurely, but one player is "up." This is known as the Problem of Points[5].

Christiaan Huygens (1629-1695) invented the pendulum clock, studied the rings of Saturn and other celestial objects, was an early proponent of the wave theory of light, and continued the intellectual work on expected values, culminating in his 1657 book *De ratiociniis in ludo aleae (On reasoning in games of chance).*

It should also be noted that starting around the work of Huygens and de Moivre is when we start to see application of probability theory to actuarial life tables, which is an especially important precursor to many later developments in epidemiology and public health. Much of this work laid the bedrock for the field of life insurance.

Jacob Bernoulli (1655 - 1705)[6] discovered e while studying compound interest, published significant work on combinatorics, and derived the first version of the law of large numbers (stating that the sample

[3] https://catalogimages.wiley.com/images/db/pdf/9781118063255.excerpt.pdf

[4] I just want to point out how central Fermat was to pre-calculus. Though he did not have the theory of infinitesimals as it was developed later by Isaac Newton (1642-1726) and Gottfried Wilhelm Leibniz (1646-1716), he did formally develop the concepts of the gradient and integral. He knew things like the rules for differentiation and integration of polynomials of degree greater than 1, though he did not realize the Fundamental Theorem of Calculus. That was the state of the art in 1636: in Fermat's text, *Method for Determining Maxima and Minima and Tangents to Curved Lines*, he described the technique for finding local extrema that we all use today: solving for where the gradient equals zero. A few useful dates to keep in mind: Though Newton was thinking about gravity and infinitesimals as early as 1666, his first published treatment on them was in his 1687 book *Philosophiae Naturalis Principia Mathematica (Mathematical Principles of Natural Philosophy).* The relevance to our fields are two-fold: first, it is that calculus is essentially a prerequisite for modern probability theory, especially for continuous random variables. After all, an intellectual descendent of Newton, Laplace, wrote of his *Théorie Analytique des Probabilités*, "one sees in this essay that the theory of probabilities is basically only common sense reduced to a calculus. It makes one estimate accurately what right-minded people feel by a sort of instinct, often without being able to give a reason for it." Second, a surprising amount of early statistics was born from interest in understanding planetary and celestial movement.

[5] The setup is that the game is won after 8 rounds, where in each round either player wins with equal odds. Essentially, Fermat reasoned that if one player requires $\geq r$ turns to win, the other player requires $\geq s$ turns, then the game will surely be decided after another $r+s-1$ turns. He reasoned that it does no harm to ignore the possibility of the game ending earlier, and tabulated the odds of each player winning in each of the 2^{r+s-1} possible continuations to show that using an expected value to divide the winnings was more fair than prior, more naive approaches.

[6] Jacob Bernoulli wanted his tombstone engraved with a logarithmic spiral and the phrase "Eadem mutate resurgo (Although changed, I rise again the same)." He wrote that the self-similar spiral "may be used as a symbol, either of fortitude and constancy in adversity, or of the man body, which after all its changes, even after death, will be restored to its exact and perfect self." Well, his loved ones clearly messed up, because they engraved an Archimedean spiral onto his tombstone instead of a logarithmic spiral. Sometimes you just can't win.

average is consistent for the population average) in *Ars Conjectandi*. His work was largely focused on discrete problems, and it is as a result of his work that we refer to a binary random variable as a Bernoulli trial.

Abraham de Moivre (1667-1754), friend of Newton, Edmond Halley, and James Stirling, wrote *The Doctrines of Chance: Or, a Method for Calculating the Probabilities of Events in Play* in 1718, in which he postulated the central limit theorem, and generalized Newton's generalized binomial theorem into the multinomial theorem.

To be quite specific about de Moivre's work on the central limit theorem, he included his unpublished work on a normal approximation to the Binomial distribution in *The Doctrines of Chance*, but largely credit for the more central limit theorem should go to Laplace for a more generalized, sophisticated statement and proof.

Notably, throughout his life, de Moivre remained poor. He tutored mathematics for money, never attaining a mathematics professorship despite his efforts, and made some money from playing chess at a coffeeshop (old Slaughter's Coffee House in London). Somewhat legendarily, in his older age he became increasingly lethargic, noting to himself that he was sleeping an extra 15 minutes each day, which he used to successfully predict the day of his death based on when he would require 24 hours of sleep. He died on the day he predicted he would, in London, on November 27th, 1754.

Pierre-Simon Laplace (1749-1827)[7] published his foundational work *Théorie analytique des probabilités (Analytic theory of Probability)* in 1812.

One fun tradition that Laplace seems to have adopted in earnest as one of its first prominent practitioners is the practice of mathematicians writing "One can clearly see, . . . " or "It is therefore obvious that . . . " before stating something extremely difficult to prove, or sometimes before going on to make a mistake. Apparently he used these phrases frequently in his work *Traité de mécanique céleste (Celestial Mechanics)*.

By the early 1800s, Laplace had access to basically fully developed calculus, as well as sophisticated works on differential equations by immediately previous mathematicians like Euler (1707 - 1783) and Lagrange (1736-1813). From that perspective, it's quite understandable how, motivated by applications of the Laplace transformation in differential equations, he was able to define the characteristic function and furnish the first general proof of the central limit theorem.

Do you know why we call the normal distribution the Gaussian?

Now, there's another development from the early 1800s I'd like to highlight that doesn't appear on the slides. Carl Friedrich Gauss (1777-1855), genius mathematician, geodesist, and astronomer, rose to fame in the scientific world after successfully predicting the location of Ceres which had been "lost" by the astronomical community.

On New Years' Day, 1801, Italian astronomer Joseph Piazzi first observed Ceres, which was a huge deal because the astronomical community had, at Kepler's insistence since the late 1500s/early 1600s, posited that there must be a celestial object between Mars and Jupiter because the gap was too great and it offended Kepler's "sense of proportionality." Piazzi observed Ceres until February 11th, when it became too close to the sun to be observable. Publications from October that year describe other astronomers dutifully waiting for the reemergence of Ceres, but no such observation had been made. This is the point at which Gauss becomes involved.

Having just finished his doctoral degree, 24 years old, Gauss took Piazzi's observational data and wanting to account for the potential of error in the measurements, assumed a normal error model and carried out a least squares procedure to determine a predicted position of Ceres. As a result, Ceres was rediscovered in January 1802, contributing to Gauss' earned fame.

Now skipping to the 20th century, Andrey Kolmogorov (1903-1987) is the last mathematician in the history of probability that we'll highlight. In 1933, Kolmogorov published *Foundations of the Theory of Probability*, which is regarded as the first axiomatic treatment of probability. Kolmogorov was so influential basically because he was the first mathematician with access to measure theory and advanced analysis including tools like functional analysis to work on probability.

Kolmogorov essentially transformed the field of probability from a patchwork of discrete math and calculus based tools into a fully-fledged mathematical discipline (on equal grounding to its contemporary 20th century developments in analysis, number theory, etc.), and his work on stochastic and limit processes created one of the sub-fields of probability that remains most fruitful to this day.

[7] Laplace is among the first credited with expressing the following idea, "The more extraordinary the event, the greater the need of its being supported by strong proofs."

Further Reading (just for fun)

- Teets D, Whitehead K. The Discovery of Ceres: How Gauss Became Famous. Mathematics Magazine. 1999 Apr;72(2):83-93. https://doi.org/10.2307/2690592
- Resnick, Brian. How Johann Carl Friedrich Gauss Rediscovered a Lost Dwarf Planet. Vox. 2018 Apr 30 [cited 2024 Aug 24]. Available from: https://www.vox.com/2018/4/30/17301292/johann-carl-friedrich-gauss-google-doodle
- Todhunter, Ian. A History of the Mathematical Theory of Probability: From the Time of Pascal to That of Laplace. Cambridge; London: Macmillan; 1865. Available from: https://archive.org/details/ofmathemahistory00todhrich/page/n15/mode/2up
- History of Probability. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/wiki/History_of_probability
- Fischer, Hans. A History of the Central Limit Theorem: From Classical to Modern Probability Theory. Springer Science & Business Media; 2010.
- Stigler, Stephen. The History of Statistics: The Measurement of Uncertainty before 1900. Cambridge, MA: Harvard University Press; 1986.
- Stigler, Stephen. Statistics on the Table: The History of Statistical Concepts and Methods. Cambridge, MA: Harvard University Press; 2002.
- Stigler, Stephen. Casanova's Lottery: The History of a Revolutionary Game of Chance. Chicago: University of Chicago Press; 2022.
- Sylla, Edith Dudley (translator). The Art of Conjecturing: Together with Letter to a Friend on Sets in Court Tennis. By Jacob Bernoulli. Baltimore: Johns Hopkins University Press; 2006.
- von Plato, Jan. Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective. Cambridge University Press; 1994.
- Hacking, Ian. The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference. 2nd ed. Cambridge University Press; 2006.
- Hacking, Ian. The Taming of Chance. Cambridge University Press; 1990.
- Salsburg, David. The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century. New York: Macmillan; 2002. 340 p.
- Nualart, David. "Kolmogorov and Probability Theory." *Arbor* CLXXVIII, no. 704 (August 2004): 607-619. https://core.ac.uk/download/pdf/268083255.pdf

Advice for Students

- A lot of people ask, "what can I do to maximize my experience in this program?" I think a useful starting place is to recognize that many aspects of whatever you've been doing must have clearly been working. You don't need to change yourself as much as you need to be open to what this experience brings you.
- These programs can be hard. Let them be hard. By which I mean, hard work can elicit a lot of emotions like frustration, feelings of inadequacy, and exhaustion. I find it useful to explicitly recognize that these are essentially symptoms of doing work that is demanding, time-intensive, and challenging, and acknowledge that there's a tradeoff being made: hard work now, for an investment in your future skills and opportunity to do meaningful work.
- There is no substitute for mindful practice.
- Emphasize the big picture and the motivation for sorting out the details will come.
- Understand the symptoms of burnout (e.g., nothing seems to get you excited or energized, you're feeling drained all the time), and understand its treatment: deliberately make time for and engage in activities that give you energy back.
- Boston is pretty far north — be aware of seasonal affective disorder, and have a plan for how your rest and recovery activities may need to change as the seasons change.
- This experience is a lot easier if you support each other. Lean on each other for support. Offer your support to each other, even when it doesn't seem like it's necessary.