ESE 402/542 Recitation 1: Probability Review

Thomas Zhang
ttz2@seas.upenn.edu

A Short Ramble: "Theory" vs "Intuition"

- "Theory": based on these rigorous logical/mathematical tools, I can demonstrate infallibly that under the conditions I assumed, my assertion is true, whether or not it "makes sense" in my mind.
- "Intuition": based on my time and experience with related objects, I have a good idea of what to expect to see, whether or not I can pin down a mathematical formulation for that vision.

When tackling practical problems, one often starts with vague intuitions and personal experiences to draw *hypotheses* about the underlying natural events. Over time, one looks toward and (hopefully) develops theory to uncover and systematize our understanding of these phenomena in order to draw *conclusions*.

A Short Ramble: "Theory" vs "Intuition"

- Statistics follows such a development cycle through and through. Clearly motivated by practical needs and natural occurrences, followed by centuries of math to describe them...
- *This is all to say...* in a statistics class taught with data science as the motivating application, the end-goals are all firmly practical. Just need a mathematical language to approach them.
- I believe you have all the intuition in the world for analyzing data. Perhaps around two decades of experience :)
- The mathematical language in question: probability.

A “Short” Ramble: Conclusion

- Probability is a means to an end. Sometimes imposing, other times cumbersome, but a means through and through. We don’t do math for the sake of math here (sorry pure math folks).
- Ideally, we should be able to make sense of every mathematical expression in this class.
- In reality, a little bit of opaque algebra every may crop up every so often, but hindsight should prove to be 20/20–everything should make sense eventually.

The Longer Ramble: Probability Review

Topics:

1. Notation
2. What is a random variable?
3. PDF (PMF) & CDF
4. Expectation and Variance

Notation

- In general, lower-case alphabetical letters (x, y, z) denote non-random variables, capitalized alphabetical letters (X, Y, Z) denote random variables. Some lower-case letters, conventionally near the start e.g. a, b, c, or Greek e.g. γ, η, ν, are typically used to denote *constants* (think placeholders for numbers).
- An example statement in probability: $\mathbb{P}\left[X = x | Y = y\right] = c$. Translation: the probability ($\mathbb{P}$) that X is equal to x, given ($|$) that Y is equal to y, is c.
- IID: Independent and Identically Distributed. Often used in contexts dealing with multiple copies of a random variable, e.g. "$X_1, \ldots, X_n$ are IID gaussian random variables".

Notation: Logic

- $\in$: set inclusion. $\subseteq$: subset (strict subset denoted $\subset$). $\cap$: intersection of two sets. $\cup$: union of two sets.
- $\implies$, $\iff$: "implies", and "if and only if" (abbr. iff)
- $\exists, \forall$: "there exists", and "for all"
- Colon ":" sometimes used to denote "such that". Sometimes, people abbreviate "s.t." instead of using colons to prevent confusion.
- An example "logical" statement (formal logic folks don't hate):

  $$\exists x : \mathbb{P}[X \leq x] > y \; \forall y \in S.$$

  Translation: there exists ($\exists$) x such that (:) the probability that $X \leq x$ is greater than y for all ($\forall$) y in ($\in$) set S.
- In practice, we will use a lot more plain English than this

Random Variables

- Definition: variable whose values (in this class, real values) depend on the outcomes of a random phenomenon. Mathematically, random variable $X : S \to \mathbb{R}$ is a function from the set of outcomes S to (a subset of) real values.
- Example: coinflip. Random outcomes: heads and tails. Random variable maps heads to $+1$ and tails to -1.
- Note 1: It might sometimes be confusing to distinguish the random variable from the random outcome (e.g. dice roll), and it might ultimately philosophical question–don't think about it too much.
- Note 2: Some non-intuitive things can actually be thought of as random variables themselves, e.g. sample mean, sample variance, since their values depend on random outcomes!
  - Non-intuitive example: conditional expectation. $\mathbb{E}[X|Y = y]$. This is a random variable that depends on the outcome that generates Y.

PDFs, PMFs, and CDFs

- Natural characterization of a random variable. Since RV maps *random* outcomes to real values, it makes sense to assign a probability to each real value to characterize the random chance we attain those values. This is a PDF (Probability Density Function). PDF is usually denoted $p(x)$. Properties:

$$p(x) \geq 0 \; \forall x \in \mathbb{R}$$
$$\int_{-\infty}^{\infty} p(x)\, dx = 1, \text{ equiv. } \sum p(x) = 1.$$

- CDF (Cumulative DF) of r.v. X is defined as the function $F(x) = \mathbb{P}[X \leq x]$. Can be thought of (for all intents and purposes) as integral (equiv. sum) of PDF up to x:

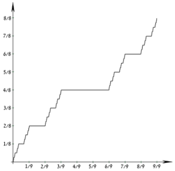
$$F(x) = \int_{-\infty}^{x} p(x)\, dx.$$

  Naturally, $F(x)$ will lie between $[0, 1]$. Sometimes it is more accurate to say PDF is derivative of CDF...

PDFs, PMFs, and CDFs

Intuitively, PDFs and CDFs seem to be unique identifiers of a random variable. After all, if two random variables attain the same values with the same probabilities, they're the same right?

- A Spooky *Gotcha*: this is actually only true for CDFs. Two random variables are equivalent iff their CDFs are the same function.
- Why not PDFs? *Most of the time*, it is true for PDFs too. However, the rough wisdom is "almost everything is integrable, not everything is differentiable". PDFs, which are formally the derivative of the CDF, *MAY NOT EXIST*!!!



- We won't ever have to think about that, though :)

Expectation

Intuition

- Want to characterize the "average" value a random variable takes, or equivalently, what value the random variable takes *in expectation*.
- Natural approach to writing down a formula for expectation: weight each value the RV takes with probability and then sum (integrate) over all values.

Formalism

- Following the intuitive description, we can write

$$\mathbb{E}[X] = \sum_x x \cdot p(x)$$

$$\text{or } \mathbb{E}[X] = \int_x x \cdot p(x)\, dx.$$

- We will often use μ or μ_X to denote the expectation of X.

Variance

Intuition

- Want to characterize the how concentrated around the mean an RV is, equivalently, how much *variation* we see in the values it takes.
- Natural approach to writing down a formula for expectation: average "distance" from the mean the random variable is, i.e. how far x is from μ, weighted by the probability of attaining x.

Formalism

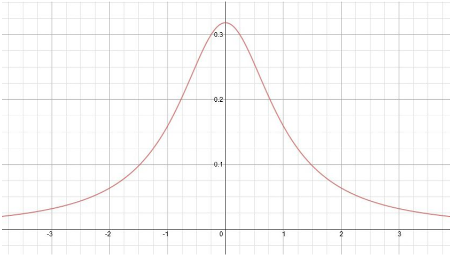
- Following the intuitive description, we can try

$$\begin{aligned} \mathrm{var}\left(X\right) &= \int_x (x-\mu)^2 \cdot p(x)\, dx \\ &= \mathbb{E}\left[(X-\mu)^2\right]. \end{aligned}$$

- Why mean-square distance? In practice, square is differentiable and nice. Certainly not the unique measure of dispersion though, e.g. interquartile range, median absolute deviation etc.

Expectation and Variance: A Cautionary Tale

- Sometimes intuition will clash with mathematical reality.
- Example: (standard) Cauchy Distribution



PDF: $p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$

- What is its expectation? I mean, it's symmetric around 0, so...

Expectation and Variance: A Cautionary Tale

- The expectation doesn't exist.
- Does the math check out?

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} xp(x)\,dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{x}{1+x^2}\,dx.\end{aligned}$$

Uh oh, $\frac{x}{1+x^2} \approx \frac{1}{x}$ isn't very friendly to integration...we're SOL.