Causal Inference Recitation-Week 1

Alex Zhang

Sep 5, 2025

Logistics:

- Recitations are **mostly hands-on coding**.
- **Exceptions**
  - **This & next week:** stats/probability review
  - **Midterm & final weeks:** review sessions
- **Otherwise:** it's **code, code, code**—live demos + practice.
- **Exams are concept-focused: no coding questions** on the midterm or final.

All Causes Model

Basic model of our outcome of interest is $Y_i(s, u)$

The **All-Causes Model** is just our “language” for causal questions:

- **s** is the *state/treatment*; **u** is *everything else* that can influence outcomes; **Y** is the *outcome* we care about.
- Suppose we ask: *Does attending college improve early-career employment?* Let **s = 1** if someone attended college (0 otherwise); **Y = 1** if they’re employed within 6 months (0 otherwise); **u** bundles things like HS GPA, internships, family support, and local job market.
- Since s,u,Y vary across people, we model them as **random variables**.

Discrete (PMF) and Continuous R.V. (PDF):

Discrete random variables take on (for our purposes) a finite number of values. (e.g dice rolls)

$$p(x) = P(X = x) = \begin{cases} 1/6, & x \in \{1,2,3,4,5,6\} \\ 0, & \text{otherwise} \end{cases}$$

Distributions of discrete random variables are called probability mass functions(pmf).

- The sum of values of the pmf across all distinct values always adds up to 1

$$\textstyle\sum_{x=1}^{6} p(x) = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 6 \times (1/6) = 1.$$

- The value of a pmf for any one outcome is between 0 and 1, inclusive
- Simply put, one and only one outcome will always occur

Discrete (PMF) and Continuous R.V. (PDF):

Continuous random variables take on (for our purposes) a continuum of values.
(e.g Spinner angle)
Distributions of continuous random variables are called probability density functions.

- We always think of the probabilities of continuous rvs in terms of ranges of values
- Since the random variable can take an infinite number of values, the likelihood of any one realization is infinitely small

$$P(X=x) \neq f(x),\ P(X=x)=0$$

Discrete (PMF) and Continuous R.V. (PDF):

Properties of PDF:

The full set of values any R.V. takes on is called the support of the R.V. The integral of a pdf of a continuous R.V. over its entire support is always equal to 1.

- For a spinner, we can see the likelihood of the entire *support* is

$$
\begin{aligned}
P(0 < X < 360) &= \int_0^{360} f(x)dx \\
&= \int_0^{360} \frac{1}{360} dx \\
&= \frac{x}{360} \Big|_0^{360} \\
&= \frac{360}{360} - \frac{0}{360} \\
&= 1
\end{aligned}
$$

Discrete (PMF) and Continuous R.V. (PDF):

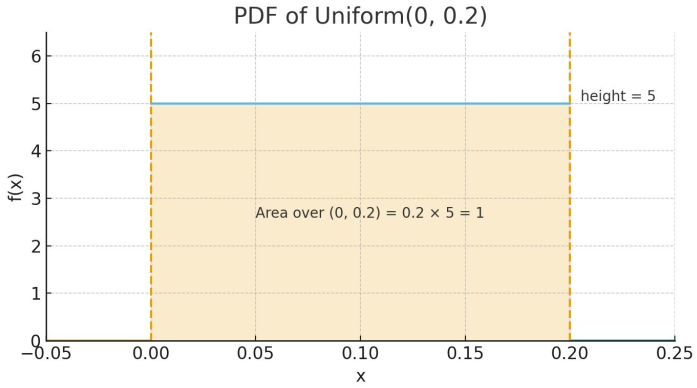
Properties of PDF:

While a pdf value can be larger than 1, the integral of the pdf over any range of values is strictly between 0 and 1.

Let $X \sim \mathrm{Uniform}(0, 0.2)$.

Then $f(x) = 5$ on $(0, 0.2)$ (note $5 > 1$) and 0 elsewhere.

- Small interval: $P(0.18 < X < 0.20) = \int_{0.18}^{0.20} 5\,dx = 5 \cdot 0.02 = 0.10$.
- Half the support: $P(0 < X < 0.10) = 5 \cdot 0.10 = 0.50$.
- Whole support: $P(0 < X < 0.20) = 5 \cdot 0.20 = 1$.
- Single point: $P(X = 0.18) = 0$.



Cumulative Distribution Function (and Parameters):

<u>Every random variable has a CDF</u>

We define $F(x) = P(X \leq x)$—think *probability up to here*.

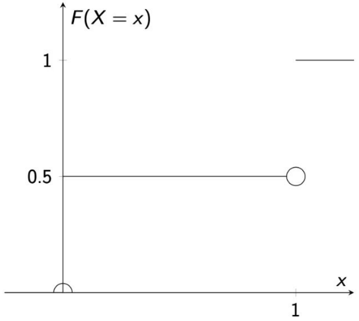
A function of a random variable is itself a random variable. The randomness gets “passed through" the function.

Cumulative distribution function(CDF) gives us the total (cumulative) probability of all outcomes up to a given one.

- From a PMF, we can get $F(x)$ by adding up the values of the pmf for all outcomes up to x.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.5 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

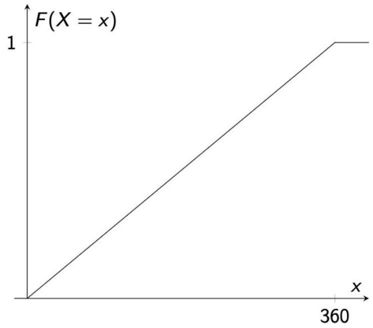
CDF for coin flip



- For a PDF, we can get $F(x)$ by integrating the pdf up to the x.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{360} & \text{if } 0 \leq x < 360 \\ 1 & \text{if } x \geq 360 \end{cases}$$

CDF for Uniform[0, 360]



Parameter 1: Expectations:

The expectation, or mean, tells us the average value we'd get if we drew the rv from its distribution many times.

- For a discrete R.V. that takes on k values:

$$E[X] = \sum_{i=1}^{k} x_i p(x_i)$$

$$\begin{aligned} E[X] &= 0 * P(X=0) + 1 * P(X=1) \\ &= 0 * 0.5 + 1 * 0.5 \\ &= 0.5 \end{aligned}$$

Expectation of a Coin Flip

- For a continuous R.V.:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

$$\begin{aligned} E[X] &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{x^2}{2(b-a)} \Big|_a^b \\ &= \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2} \end{aligned}$$

Expectation of a Uniform[a,b] R.V.

$$\begin{aligned} E[X] &= \int_0^{360} x \frac{1}{360} dx \\ &= \frac{x^2}{2 * 360} \Big|_0^{360} \\ &= \frac{360^2}{720} - \frac{0^2}{720} \\ &= \frac{129,600 - 0}{720} \\ &= 180 \end{aligned}$$

Expectation of a Spinner

Parameter 1: Expectations:

Expectation of Functions of R.V.

- For a discrete R.V. that takes on k values

$$E[g(X)] = \sum_{i=1}^{k} g(x_i)p(x_i)$$

- For a continuous R.V.

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Properties of Expectations

1. If X, Y are random variables and a, b are scalars, then
   - $E[a + bX] = a + bE[X]$
     - $E[a] = a$
   - $E[X + Y] = E[X] + E[Y]$
   - if $X \leq Y$, then $E[X] \leq E[Y]$
     - if $Y \geq 0$, then $E[Y] \geq 0$

NYU

Parameter 2: Variance:

The variance of a R.V. is a parameter measuring how disperse the distribution is.

$$\sigma_X^2 = Var(X) = E[(X - E[X])^2]$$

The units of the variance are the units of X squared.

Alternative Variance Formula:

$$Var(X) = E[X^2] - E[X]^2$$

The units of the variance are the units of X squared - slightly awkward to interpret

The **standard deviation**, the root of the variance, has the same units as X, which is easier to think about:

$$\sigma_X = Std\ Dev(X) = \sqrt{Var(X)}$$

$$\begin{aligned} E[(X - E[X])^2] &= E[X^2 - 2XE[X] + E[X]^2] && \text{(algebra)} \\ &= E[X^2] - E[2XE[X]] + E[X]^2 && \text{(property ii)} \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 && \text{(property i)} \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Properties of Variance:

For a, b scalars: $Var(a + bX) = b^2 Var(X) \rightarrow Var(a) = 0$