Lecture: Conditional Distributions   PWD: Mirror

Last Time: More discrete distributions Expectations & Properties

Recall when X, Y discrete, their joint PMF is P_{XY}

Cond. Prob

$$P(A \mid C) := \frac{P(A \cap C)}{P(C)}$$

Conditional Distribution

special case of cond. prob.

- Conditional PMF

$$P_{X|Y}(x|y) := \frac{P_{XY}(x,y)}{P_Y(y)} = \frac{P(\{X=x\} \cap \{Y=y\})}{P(\{Y=y\})}$$

- Given a rv X, we can define a cond. prob. wrt it:
  - Since $P_{X|Y}(\cdot|y)$ is a PMF for each y with $P_Y(y) > 0$, we can take the expectation of X wrt it:

$$\mathbb{E}[X \mid Y=y] := \sum_{x \in \mathcal{X}} x \, P_{X|Y}(x|y)$$

fcn of Y for each y

- usually we just write $\mathbb{E}[X \mid Y]$ to denote this evaluated at a r.v. Y.

- $\mathbb{E}[X \mid Y]$ is itself a r.v.

$\mathbb{E}[X \mid Y = Y(\omega)]$

☆ Most Important Property of Conditional Expectation: ☆

Tower Property:

$\forall$ functions f,

$$\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)\,\mathbb{E}[X \mid Y]]$$

Q/ Why does this hold?

A /
$$\sum_y P_Y(y) f(y) \sum_x x P_{X|Y}(x|y)$$
$$= \sum_{x,y} P_{XY}(x,y) f(y) x$$
$$= \mathbb{E}[f(Y) X]$$

ex: iterated Expectation

- to compute r.v. X can use cond. exp.

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$$

⇝ Tower prop. $f(y) = 1$

ex

- Let $N \geq 0$ be integer valued r.v. Lets flip a fair coin N times & let $X = \#$ heads.

⇝ 2 sources of randomness → # of flips
→ # Heads

$$\mathbb{E}[X] = \mathbb{E}[\underbrace{\mathbb{E}[X \mid N]}_{\frac{N}{2} = N\,P(H) = N \cdot \frac{1}{2} = \frac{N}{2}}]$$
$$= \mathbb{E}\left[\frac{N}{2}\right]$$
$$= \frac{1}{2}\mathbb{E}[N]$$

Recall
$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}[X])^2$$

Conditional Variance

"given my observation, how must uncertainty do I have around X"?

- def Conditional variance:

$$\mathrm{Var}(X \mid Y=y) := \sum_x P_{X|Y}(x|y)\,(x - \mathbb{E}[X \mid Y=y])^2$$

- Just like Cond Exp, we write

$\mathrm{Var}(X \mid Y)$ to denote the r.v. evaluated at Y

Thm: Law of Total Variance

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y])$$

this is a version of the Pythagorean thm

Proof:

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$
$$= \mathbb{E}[\mathbb{E}[X^2|Y]] - (\mathbb{E}[\mathbb{E}[X|Y]])^2$$

$$\text{Var}(X|Y=y) = \underbrace{\mathbb{E}[X^2|Y=y]}_{2^{nd}\text{ moment}} - \underbrace{(\mathbb{E}[X|Y=y])^2}_{\text{mean squared}}$$

linearity of $\mathbb{E}$

$$= \mathbb{E}[\text{Var}(X|Y) + (\mathbb{E}[X|Y])^2] - (\mathbb{E}[\mathbb{E}[X|Y]])^2$$
$$= \mathbb{E}[\text{Var}(X|Y)] + \underbrace{\mathbb{E}[(\mathbb{E}[X|Y])^2] - (\mathbb{E}[\mathbb{E}[X|Y]])^2}_{\text{Var}(\mathbb{E}[X|Y])}$$

$$= \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y])$$

ex

- Let $N \geq 0$ be integer valued r.v. Lets flip a fair coin N times & let X = # heads.

$$\text{Var}(X) = \mathbb{E}[\underbrace{\text{Var}(X|N)}_{\frac{N}{4}}] + \underbrace{\text{Var}(\underbrace{\mathbb{E}[X|N]}_{\frac{N}{2}})}_{\frac{1}{4}\text{Var}(N)}$$

- as soon as we fix N, we have binomial distribution & X is sum of indep. coin flips
  ↳ $\text{Bin}(N, \frac{1}{2}) \rightarrow$

$$= \frac{1}{4}\mathbb{E}[N] + \frac{1}{4}\text{Var}(N)$$

Continuous Random Variables / Distributions

Note (N.B.):

- RVs need not be discrete or continuous or combinations thereof
- For a RV X, we can describe its cumulative distribution fcn (CDF) via:

$$F_X(x) := P\{X \leq x\} \qquad x \in \mathbb{R}$$

these describe distributions in general → given F_X can constr. an RV satisfying

① F_X is non-decreasing

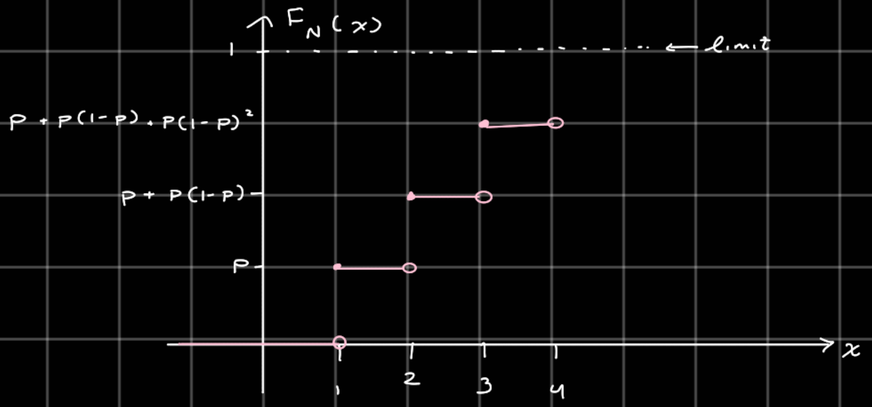
② $F_X(x) \xrightarrow{\text{approaches}} \begin{cases} 0 & x \to -\infty \\ 1 & x \to \infty \end{cases}$

↳ related to continuity property

③ F_X is continuous from the right

Ex:

$N \sim Geom(p)$



def: a RV X has continuous distribution if $\exists$ a fcn f_X s.t:

$$\underbrace{F_X(x) = \int_{-\infty}^{x} f_X(u)\,du \qquad \forall x \in \mathbb{R}}_{\text{Property: "}F_X\text{ is absolutely continuous"}}$$

f_X is called the density of X (PDF: Probability Density fcn)

↳ for f_X to be a density it must

① be non-negative ($f_X \geq 0$)

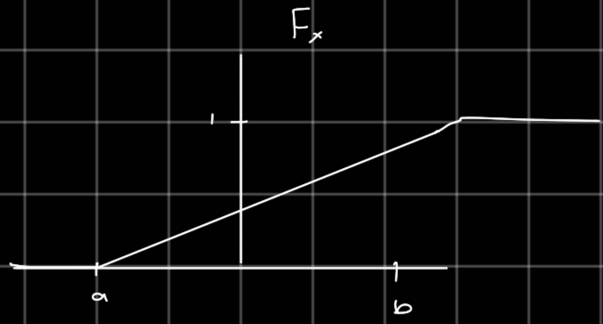
② integrate to 1 bc prob must Σ to 1 ($\int f_X\,dx = 1$)

. Continuous RVs are good for modeling "analog" signals from the real world

- eg: - time we wait until bus arrives (Exponential distribution)
  - voltage across resistor (Gaussian distribution)
  - phase of a received wireless signal (uniform)

- Continuous Distributions usually described by their density:
  - eg:
    - $X \sim \text{Unif}(a,b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & o/w \end{cases}$$



- $X \sim \text{Exp}(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & o/w \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & o/w \end{cases}$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

F_X has no closed form (this is why we usually look at densities)

Jointly CTS RVs:

- We say $X_1, X_2, \cdots X_n$ are jointly cts if

$$F_{X_1 X_2 \cdots X_N}(x_1, x_2, \cdots, x_n) := P\{X_1 \leq x_1, X_2 \leq x_2, \cdots X_n \leq x_n\}$$

can be expressed as an iterated integral:

$$= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{X_1 X_2 \cdots X_n}(u_1, \ldots, u_n)\, du_1 \ldots du_n \quad \text{for some } f_X : \mathbb{R}^n \to \mathbb{R}$$

- ex: let a dart land uniformly at random on 2D dartboard of radius $r > 0$. Let (X, Y) be x-y coords of the dart

not the case if board was a square (?)

area of dartboard

$$f_{XY}(x, y) = \begin{cases} \frac{1}{\pi r^2} & x^2 + y^2 \le r^2 \\ 0 & \text{o/w} \end{cases}$$

on the board

these RVs aren't indep bc say you know $y = 1$, x must be $= 0$ (bc it's on/in the circle)

bc we don't miss the board

Independence: r.v.'s X, Y are independent if

$$F_{XY}(x, y) = F_X(x) F_Y(y)$$

→ X, Y cts & indep $\iff$ $f_{XY}(x, y) = f_X(x) f_Y(y)$

Expectation

- For X cts rv:

$$\mathbb{E}[X] = \int x f_X(x)\, dx$$

- More generally: (LOTUS)

$$\mathbb{E}[g(X_1, \dots, X_n)] = \int \cdots \int g(X_1, \dots, X_n) f_{X_1 \dots X_n}(x_1 \cdots x_n)\, dx_1 \cdots dx_n$$

- ex:

$$\text{Var}(X) = \int (X - \mathbb{E}[X])^2 f_X(x)\, dx$$

- ex: Calculations w/ Unif distr:

$X \sim \text{Unif}(a, b)$

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b - a}\, dx = \frac{1}{2} \frac{b^2 - a^2}{b - a} = \frac{1}{2} \frac{(b-a)(b+a)}{(b-a)}$$

$$= \frac{1}{2}(b + a)$$

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{12}(b - a)^2$$

- Back to dartboard:

↳ let $R = \sqrt{X^2 + Y^2}$ (distance from dart to center)

$$P(R \le r/2) = P(X^2 + Y^2 \le \frac{r^2}{4})$$

$$= \mathbb{E}\left[\mathbb{1}_{\{x^2+y^2 \leq r^2/4\}} \right]$$

indicator

$$= \frac{1}{\pi r^2} \int 1_{\{x^2+y^2 \leq r^2/4\}} \, dx \, dy$$

$$= \frac{1}{4}$$