Nested Expectation

Over X
Using $P(X|Y=y)$

$$E[X] = E\big[\underbrace{E[X|Y=y]}_{\text{as a function of } y}\big]$$

Step back a little:

$$E[X] = \sum_x x P(X=x)$$

$$E[X] = \sum_x x\, p(x) \quad \text{[simplified notation]}$$

when it's clear from context.

$$E[X|Y=y] = \sum_x x P(X=x|Y=y) = \text{function of } y = f(y)$$

$$E[f(y)] = \sum_y f(y) P(y)$$

$$E[X] = \sum_x x\, p(x) = E\left[E[X \mid Y=y]\right]$$

a little complicated $\quad P(x) = \sum_y P(x \mid Y=y)\, P(y)$

An example: Assume X is a geometric R.V. where

$$P(X=k \mid p) = p(1-p)^{k-1} \qquad k \geq 1$$

conditioned on p.

what is $P(X=k)$?

$$E[X] = \sum_k k\, P(k)$$

! I don't have $P(k)$

$$E[X] = E\Big[\underbrace{E[X \mid p]}_{\text{Geometric R.V.}}\Big] = E[1/p]$$

For instance : $p = \begin{cases} 0.25 & 1/3 \\ 0.5 & 1/3 \\ 0.75 & 1/3 \end{cases}$ (uniform)

$$E[1/p] = \sum_p 1/p \, P(p) = \frac{1}{0.25}\times\frac{1}{3} + \frac{1}{0.5}\times\frac{1}{3} + \frac{1}{0.75}\times\frac{1}{3} = \frac{22}{9}$$

<u>Side Remark:</u> Expected value of geometric R.V.

$$P(X=k) = p(1-p)^{k-1} \quad k \geq 1$$

$$E[X] = \sum_{k=1}^{\infty} k\, p(1-p)^{k-1} = \frac{p}{1-p} \underbrace{\sum_{k=1}^{\infty} k(1-p)^{k}}_{(1-p)/p^2} = \frac{1}{p}$$

Let's find $P(X=k)$

$$P(X=k) = \sum_p \underline{P(X=k \mid p)}\, P(p)$$

$$\sum_p p(1-p)^{k-1}\left(\frac{1}{3}\right) = \frac{1}{3}\left[0.25\times 0.75^{k-1} + 0.5\times 0.5^{k-1} + 0.75\times 0.25^{k-1}\right]$$

$$E[X] = \sum_x x \frac{1}{3}\left[\quad \cdots\cdots \quad \right]$$

$$= \frac{1}{3} \underbrace{\sum_x x \times 0.25 \times 0.75^{k-1}}_{\frac{1}{0.25}} + \frac{1}{3} \underbrace{\sum_x x \times 0.5 \times 0.5^{k-1}}_{\frac{1}{0.5}} + \frac{1}{3} \underbrace{\sum_x x \times 0.75 \times 0.25^{k-1}}_{\frac{1}{0.75}}$$

$$\sum_x x \cdot \underbrace{p \cdot (1-p)^{k-1}}_{P(X=k)}$$

$P(X=k) \sim$ Geometric with param p

Variance:

$$\sigma_X^2 = E[(X - \underbrace{E[X]}_{\text{mean of } X})^2]$$ positive deviation from the mean

standard deviation σ_X is the sqrt of the above

$$\sigma_X^2 = E[X^2 + E[X]^2 - 2XE[X]]$$

$$= E[X^2] + E[E[X]^2] - E[2XE[X]]$$

constant

$$= E[X^2] + E[X]^2 - 2\underbrace{E[X]E[X]}_{E[X]^2}$$

$$= E[X^2] - E[X]^2$$

Some Known Means & Variances:

	Mean	Variance
Uniform:	$\frac{n+1}{2}$	$(n^2-1)/12$
Binomial:	np	$np(1-p)$
Geometric:	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Bernoulli:	p	$p(1-p)$

Properties of variance.

- Variance of a constant is 0.

- $\sigma_{aX}^2 = a^2\sigma_X^2$

- $\sigma_{X+Y}^2 = E[(X+Y)^2] - E[X+Y]^2$

(use linear property of expectation)

$= E[X^2+Y^2+2XY] - (E[X]+E[Y])^2$

$\vdots$

$= \sigma_X^2 + \sigma_Y^2 + 2(E[XY] - E[X]E[Y])$

X & Y are independent, this is 0

X and Y are independent $\Rightarrow$ $\underbrace{\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2}_{\text{uncorrelated}}$

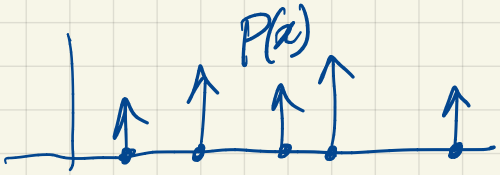
Random Variable (Continuous setting).

A random variable X is continuous if it's domain, which is the set of values x that X can take, is continuous.

e.g. X can take any real value between 0 and 1.

Discrete Case

Probability Mass Function.



Interpretation:

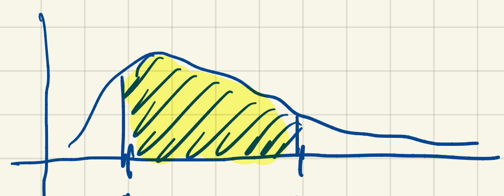
Every x has prob. $p(x)$

$$P(a \leq X \leq b) = \sum_{x \in [a,b]} P(x)$$

$$\sum_{x} P(x) = 1$$

Continuous Case

Probability "density" function



Interpretation

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

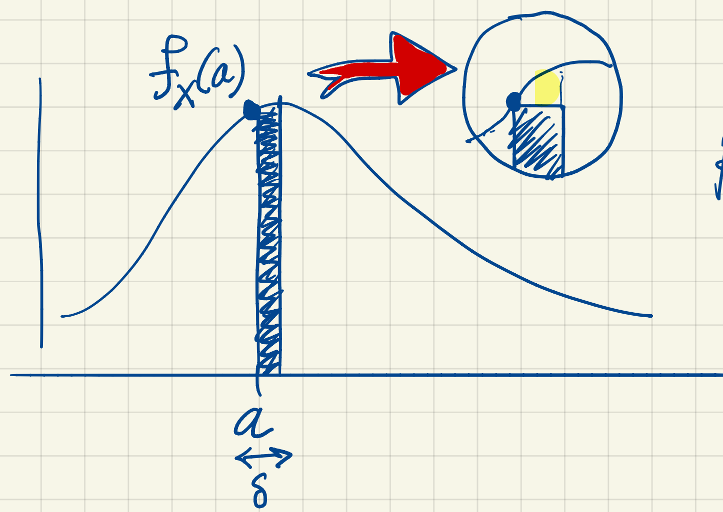
$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$$

For a continuous R.V. X what is

$$P(X=x) = P(x \le X \le x) = \int_x^x f_X(x)\,dx = 0$$

- This doesnot mean x never shows up!

- Not all zeros are the same.



$$P(a \le X \le a+\delta) = \int_a^{a+\delta} f_X(x)\,dx$$

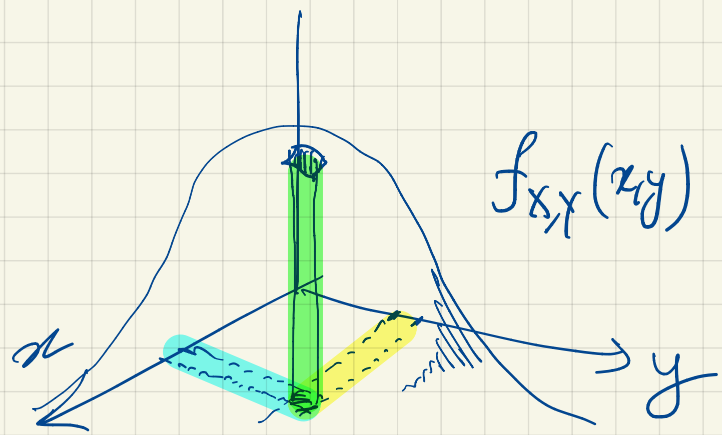
$$\approx f_X(a)\cdot\delta$$

$$\frac{P(X=a)}{P(X=b)} = \lim_{\delta\to 0} \frac{P(a\le X\le a+\delta)}{P(b\le X\le b+\delta)} \approx \frac{f_X(a)\,\delta}{f_X(b)\,\delta} = \frac{f_X(a)}{f_X(b)}$$

Let's carry over some concepts to continuous setting:

- Expectation: $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\,dx$

- Independence: X and Y independent $\iff \underbrace{f_{X,Y}(x,y)}_{\text{Joint density}} = f_X(x) f_Y(y)$



$$\iint f_{X,Y}(x,y)\,dx\,dy = 1$$

- Conditional density: $f_{X|Y=y}(x) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

Law of total prob:

$$f_Y(y) = \int f_{Y|X}(y)\, f_X(x)\, dx$$

Bayes Rule:

A specific x

$$f_{X|Y=y}(x) = \frac{f_{Y|X}(y)\, f_X(x)}{f_Y(y) = \int f_{Y|X}(y)\, f_X(x)\, dx}$$

"Sum" for all x

For simplicity, we drop subscripts when clear from context.

$$f(y) = \int f(y|x)\, f(x)\, dx$$

$$f(x|y) = \frac{f(y|x)\, f(x)}{\int f(y|x) f(x) dx}$$