HW watch Wyner's Moneyball probability lectures

Agenda Discrete uniform distribution, Expected Value $\mathbb{E}$, Bernoulli, Binomial, linearity of expectation, Independence, Variance, Normal distribution, continuous probability, density, cdf, conditional probability, law of total probability, Bayes Rule,

$\boxed{\text{Dice}}$ Let X represent the roll of a dice,

$$X = \begin{cases} 1 & \text{with probability } 1/6 \\ 2 & \text{w.p. } 1/6 \\ 3 & \\ 4 & \\ 5 & \\ 6 & \end{cases}$$

$$X \sim \text{Discrete Uniform}(\{1,2,3,4,5,6\})$$

Q What is the average value of a dice roll?

$$3.5 = \frac{1+2+3+4+5+6}{6}$$

$$1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + \cdots + 6\left(\tfrac{1}{6}\right)$$

But why does 3.5 as the average make sense?

Frequency Argument

6 billion dice rolls

about 1 billion each will be 1, 2, 3, 4, 5, 6

* What if we have a new die Y

$$Y = \begin{cases} 1 & \text{w.p. } 0 \\ 2 & \text{w.p. } 2/6 \\ 3, 4, 5, 6 & \text{w.p. } 1/6 \end{cases}$$

Q What's the average value of a roll?
How to compute it?
Why?

$$2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= 3.67$$

Frequency argument:

6 billion dice Rolls

expect 1 billion each of 3, 4, 5, 6

2 billion 2's

$$\frac{2\cdot(2\text{ bi}\cancel{\text{llion}}) + 3\cdot(1\text{ bi}\cancel{\text{llion}}) + 4\cdot(1\text{ bi}\cancel{\text{llion}}) + 5\cdot(1\text{ bi}\cancel{\text{llion}}) + 6\cdot(1\text{ bi}\cancel{\text{llion}})}{6\text{ bi}\cancel{\text{llion}}}$$

$$= 2\cdot\left(\frac{2}{6}\right) + 3\cdot\left(\frac{1}{6}\right) + 4\cdot\left(\frac{1}{6}\right) + 5\cdot\left(\frac{1}{6}\right) + 6\cdot\left(\frac{1}{6}\right)$$

$$= 3.67$$

The <u>expected value</u> of a <u>random variable</u> X is $\boxed{\mathbb{E}X = \sum_{x} x\cdot\mathbb{P}(X=x)}$

$$X = \begin{cases} 2 & \text{w.p. } 2/6 \\ \begin{matrix} 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \text{w.p. } 1/6 \end{cases}$$

$$\mathbb{E}X = \sum_{x} x \cdot \mathbb{P}(X = x)$$

$$= \sum_{x \in \{2,3,4,5,6\}} x \cdot \mathbb{P}(X = x)$$

$$= 2 \cdot \mathbb{P}(X = 2) + 3 \cdot \mathbb{P}(X = 3)$$

$$+ 4 \cdot \mathbb{P}(X = 4) + 5 \cdot \mathbb{P}(X = 5)$$

$$+ 6 \cdot \mathbb{P}(X = 6)$$

$$= 3.67$$

$\boxed{\text{Coin Flip}}$

$$X = \begin{cases} 1 & \text{with probability } p \quad \text{(Heads)} \\ 0 & \text{w.p. } 1-p \quad \text{(Tails)} \end{cases}$$

$p \in [0,1]$

$$\boxed{X \sim \text{Bernoulli}(p)}$$

$$\mathbb{E}X = \sum_x x\, \mathbb{P}(X=x)$$

$$= \sum_{x \in \{0,1\}} x\, \mathbb{P}(X=x)$$

$$= 1 \cdot \mathbb{P}(X=1) + 0 \cdot \mathbb{P}(X=0)$$

$$= \mathbb{P}(X=1) = p.$$

Shaqtin' a fool

Shaq takes n free throws.
Say he makes each f.t. w.p. p.
The random variable X representing how many shots he makes is

$$X = \sum_{i=1}^{n} X_i, \quad X_i \overset{iid}{\sim} \text{Bernoulli}(p)$$

iid = independent and identically distributed

$$\boxed{X \sim \text{Binomial}(n, p)}$$

is the # of successes (1's) out of n trials
or the # of made free throws

if each trial (free throw) has prob. p

How many free throws will Shaq make on average?

$$\mathbb{E}X = \sum_{x=0}^{n} x \cdot \mathbb{P}(X=x)$$

$$= \sum_{x=0}^{n} x \cdot \binom{n}{x} p^{x} (1-p)^{n-x}$$

$$\mathbb{P}(X=x) = \binom{n}{x} p^{x} \cdot (1-p)^{n-x}$$

$$p \cdot p \cdot \ldots p \cdot (1-p) \cdot (1-p) \cdots (1-p)$$

$$\underbrace{1 \quad 1 \quad \ldots \quad 1}_{x} \quad \underbrace{0 \quad 0 \quad \ldots \quad 0}_{n-x}$$

and

$$\binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x}$$

$$\underbrace{0\ 0\ 0\ 0}_{n-x}\qquad\underbrace{1\ 1\ 1\ 1\ 1}_{x}$$

$$\boxed{\begin{array}{l}\underline{\text{Theorem (Linearity of Expectation)}}\\[1em] \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]\\[1em] \text{and}\quad \mathbb{E}[cX] = c\,\mathbb{E}[X]\end{array}}$$

$$X \sim \text{Binomial}(n, p)$$

$$X = \sum_{i=1}^{n} X_i, \qquad X_i \sim \text{Bernoulli}(p)$$

$$X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

$$\mathbb{E}X = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

$$= \sum_{i=1}^{n} p = \underbrace{p + p + \cdots + p}_{n \text{ times}} = n \cdot p$$

Show $\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}[X_i]$

Show $n=2$: $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$

$$\mathbb{E}(X+Y) = \sum_{x,y} (x+y)\, \mathbb{P}(X=x, Y=y)$$

, means "and"

$$= \sum_{x,y} x\, \mathbb{P}(X=x, Y=y) + \sum_{y,x} y\, \mathbb{P}(X=x, Y=y)$$

$$= \sum_{x,y} x\, \mathbb{P}(X=x, Y=y) + \sum_{x,y} y\, \mathbb{P}(X=x, Y=y)$$

$$= \sum_x x \underbrace{\sum_y \mathbb{P}(X=x, Y=y)}_{=\mathbb{P}(X=x)} + \sum_y y \sum_x \mathbb{P}(X=x, Y=y)$$

$$= \sum_x x\, \mathbb{P}(X=x) + \sum_y y\, \mathbb{P}(Y=y)$$

$$= \mathbb{E}X + \mathbb{E}Y$$

$$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$$

A <u>linear function</u> f satisfies

$$f(c \cdot x + y) = c f(x) + f(y)$$

X and Y are <u>independent</u> Random variables means

$$\mathbb{P}(X = x \text{ and } Y = y)$$
$$= \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

Frequency argument:

1 billion pairs of free throws each have p probability of success

$\approx p$ billion made 1st shots → of these, $\approx p$ fraction will have made 2nd shot
→ of these, $\approx 1-p$ fraction will have missed second shot

$\approx 1-p$ billion missed 1st shots

(made, made) $p^2 = p \cdot p$ billion

(made, miss) $p \cdot (1-p)$ bil

(miss, made) $(1-p) \cdot p$ bil

(miss, miss) $(1-p)^2$ billion

Fact If X, Y are independent then $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y$

$np \cdot np$

$n^2 p^2$

We now know how to compute the expected value of a random variable X.

e.g. average # free throws made in n trials $= np$

But what about the spread or deviation from the average?

$$np \pm 2$$

vs. $np \pm 1$ trillion

The **variance** of a random variable X is the expected value of its squared deviation from its own mean,

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

Standard deviation

$$\boxed{SD(X) = \sqrt{\mathrm{Var}(X)}}$$

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

$$\mathbb{E}X^2 = \sum_x x^2\, \mathbb{P}(X=x)$$

$$(\mathbb{E}X)^2 = \left(\sum_x x\, \mathbb{P}(X=x)\right)^2$$

$$\mathrm{VAR}(X) := \mathbb{E}\left[(X-\mathbb{E}X)^2\right]$$

$$= \mathbb{E}\left[X^2 - 2\,\mathbb{E}X\cdot X + (\mathbb{E}X)^2\right]$$

Linearity of Expectation

$$= \mathbb{E}[X^2] - \mathbb{E}\left[2\cdot\mathbb{E}X\cdot X\right] + \mathbb{E}\left[(\mathbb{E}X)^2\right]$$

Linearity of Expectation

$$= \mathbb{E}[X^2] - 2\cdot\mathbb{E}X\cdot\mathbb{E}[X] + (\mathbb{E}X)^2$$

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

$$= \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

The variance of Shaq's # of made free throws is

$$\mathrm{Var}(X) = \underbrace{\mathbb{E}X^2}_{?} - \underbrace{(\underbrace{\mathbb{E}X}_{n\cdot p})^2}_{(n\cdot p)^2}, \quad X \sim \mathrm{Binom}(n,p)$$

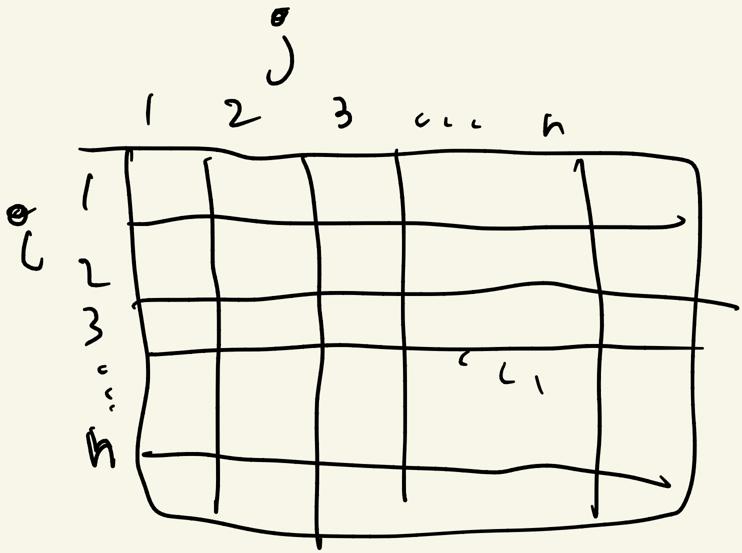
$$\mathbb{E}(X^2) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right)^2, \quad X_i \sim \mathrm{Ber}(p)$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i\right)\right]$$

$$= \mathbb{E}\left[(X_1 + X_2 + \dots + X_n) \cdot (X_1 + X_2 + \dots + X_n)\right]$$

$$= \mathbb{E}\begin{bmatrix} X_1 \cdot X_1 + X_1 \cdot X_2 + X_1 \cdot X_3 + \dots + X_1 \cdot X_n \\ + X_2 \cdot X_1 + X_2 \cdot X_2 + \dots + X_2 \cdot X_n \\ + \dots \end{bmatrix}$$

$$= \mathbb{E}\left[\sum_{\substack{i=1 \text{ to } n \\ j=1 \text{ to } n}} X_i \cdot X_j\right]$$



how many pairs (i, j) are there when $i \neq j$?

$n^2 - n$

$$= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i \cdot X_j\right]$$

$i \neq j$, X_i ind. X_j

linearity of $\mathbb{E}$

$$= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] + \mathbb{E}\left[\sum_{i \neq j} X_i \cdot X_j\right]$$

Linearity of Expectation

$$= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i \neq j} \mathbb{E}[X_i X_j]$$

Recall $X_i \overset{iid}{\sim} \text{Bernoulli}(p)$

iid = independent
identically distributed

$$\mathbb{E}X_1^2 = \mathbb{E}X_2^2 = \dots = \mathbb{E}X_n^2$$

$\mathbb{E}(X_i \cdot X_j)$ is the same for all $i \neq j$

$$= n \cdot \mathbb{E}(X_1^2) + \underbrace{(n^2 - n)} \cdot \mathbb{E}(X_1 \cdot X_2)$$

$$= n \cdot \mathbb{E}(X_1^2) + (n^2 - n) \cdot \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)$$

by independence

$$X_1 \sim \mathrm{Ber}(p)$$

$$\mathbb{E}X_1 = p$$

$$\mathbb{E}X_1^2 = \sum_x x^2 \cdot \mathbb{P}(X = x)$$

$$= \sum_{x \in \{0,1\}} x^2 \cdot \mathbb{P}(X = x)$$

$$= 0^2 \cdot \mathbb{P}(X=0) + 1^2 \cdot \mathbb{P}(X=1)$$

$$= \mathbb{P}(X=1) = p.$$

$$= n \cdot p + (n^2 - n) \cdot p \cdot p$$

$$\mathbb{E}X^2 = np + (n^2 - n)p^2$$

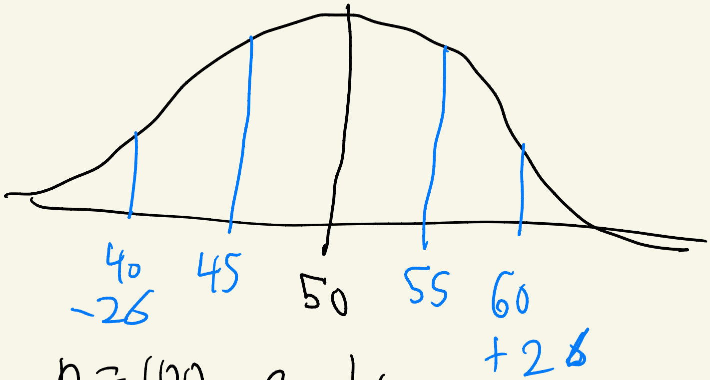
$$\mathrm{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(np - p + 1 - np)$$

$$= np(1-p)$$

$$\boxed{\begin{aligned} SD(X) &= \sqrt{np(1-p)} \\ \mathbb{E}X &= np \end{aligned}}$$



$n = 100, \ p = 1/2$

$\mathbb{E}X = 50$

$\sigma = sd(X) = \sqrt{100 \cdot \frac{1}{2} \cdot \frac{1}{2}}$

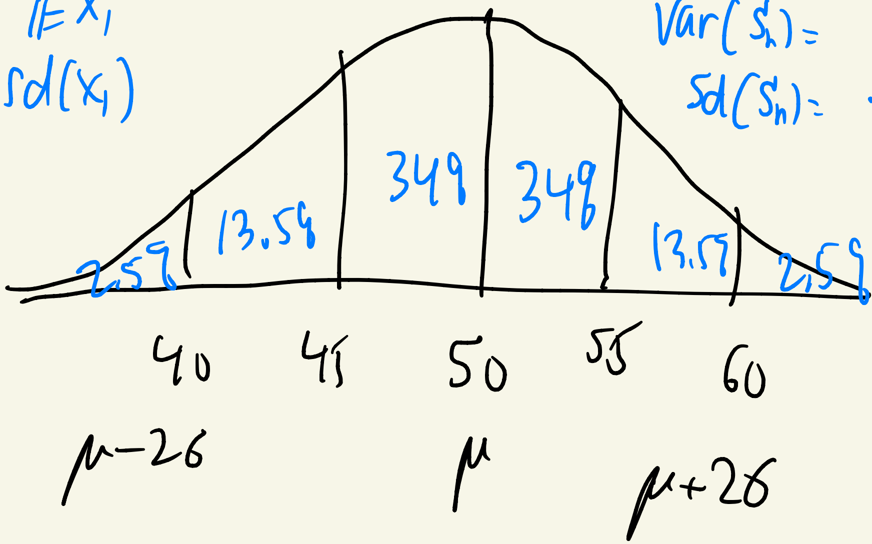
$= \sqrt{25} = 5$

<u>Central Limit Theorem</u> the (Re-scaled) Sum of iid Random variables with mean $\mu = \mathbb{E}X_i$ and s.d. $\sigma = \sqrt{\mathrm{VAR}(X_i)}$, written $S_n = \sum_{i=1}^{n} X_i$, converges to a standard Normal Random variable $Z \sim N(0,1)$,

$$\boxed{\mathbb{P}(a \le Z \le b) = \lim_{n \to \infty} \mathbb{P}\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right)}$$

X_1
$\mu = \mathbb{E}X_1$
$\sigma = sd(X_1)$

$\mathbb{E}S_n = n\mu = np$
$Var(S_n) = np(1-p)$
$sd(S_n) = \sqrt{n}\sqrt{p(1-p)}$



<u>Q</u> But what Really <u>is</u> the normal distribution?

Discrete probability:

$$\mathbb{P}(X = x) \quad \text{defines a random variable}$$

<u>discrete</u> means a <u>countable</u> set of possible outcomes Ω

e.g. dice $\Omega = \{1,2,3,4,5,6\}$

\# made free throws $\Omega = \{1,2,3,\dots,n\}$

$|\mathbb{N}| = \omega$ $\{1, 2, 3, \dots\} = \mathbb{N}$

$|\mathbb{R}| = 2^{\omega}$ not countable $\mathbb{R} = \{$ all decimal numbers $\}$

Want to prove $|\mathbb{R}| > |\mathbb{N}|$

going to do this by:

$|\mathbb{R}| = |\mathbb{N}| \xrightarrow{\text{then}}$ something false would have to be true

Suppose $|\mathbb{R}| = |\mathbb{N}|$

then you could list out all the real numbers

1. 9 1 2 3 4 3 2 2 2 1 . . .
2. 8 1 2 2 2 2 2 2 2
3. 8 1 3 2 2 3 2 2 2 . . .

0

0

0

0, 1, 2

824....

$|\mathbb{R}| > |\mathbb{N}|$

Discrete probability:

$\mathbb{P}(X = x)$ defines a random variable

Continuous probability:

set of outcomes Ω is <u>not</u> countable

e.g. $\Omega = \mathbb{R}$, $\mathbb{R}^+$

if $\Omega = \mathbb{R}$

then $\mathbb{P}(X = x)$ doesn't ~~always~~ make sense

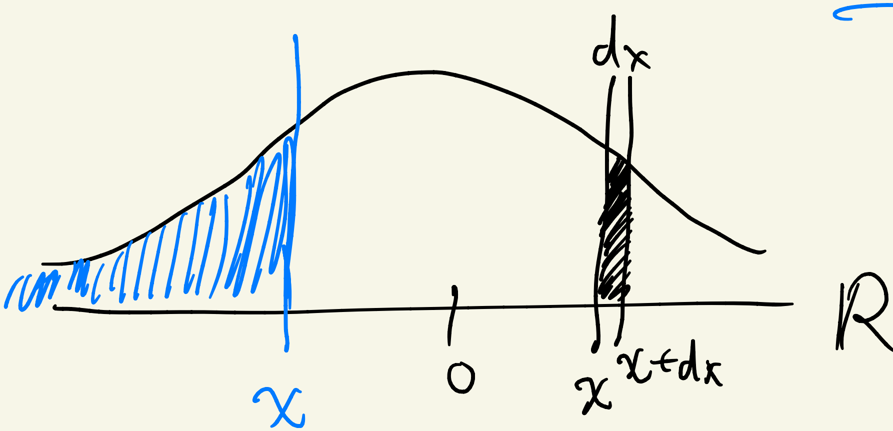
because $\sum_x \mathbb{P}(X = x) = 1$

in discrete

Instead, we have <u>Calculus</u>,

$\mathbb{P}(X \in [x, x+dx)) =: f(x)$

f is the density



is the analog of $\mathbb{P}(X = x)$ from discrete prob.

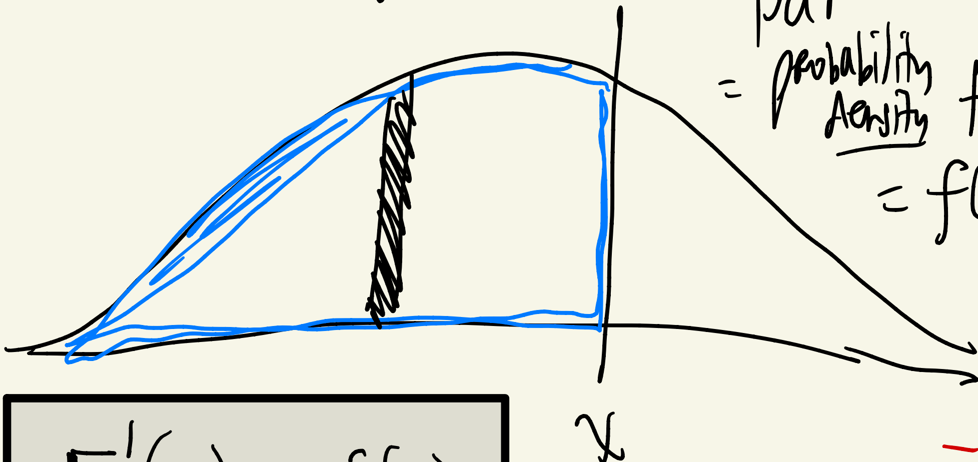
We also have the CDF

Cumulative distribution function

$$F(x) = \mathbb{P}(X \le x)$$

$$= \int_{-\infty}^{x} \mathbb{P}(X \in [x, x+dx))\, dx$$

$$= \int_{-\infty}^{x} f(x)\,dx$$



pdf = probability density function $= f(x)$

$$\boxed{F'(x) = f(x)}$$

$$e^{-x^2}$$

the Normal distribution $\mathcal{N}(\mu, \sigma^2)$

has density / pdf $\boxed{\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$

and CDF $\Phi(x) = \int_{-\infty}^{x} \phi(x)\,dx$

$$S_n = \sum_{i=1}^{n} X_i -$$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = Z_n$$

$$\phi(t) = \mathbb{E}\, e^{itZ_n} \qquad i = \sqrt{-1}$$

$$\downarrow$$

$$e^{-t^2}$$

Beta

gamma

exponential

hypergeometric

geometric

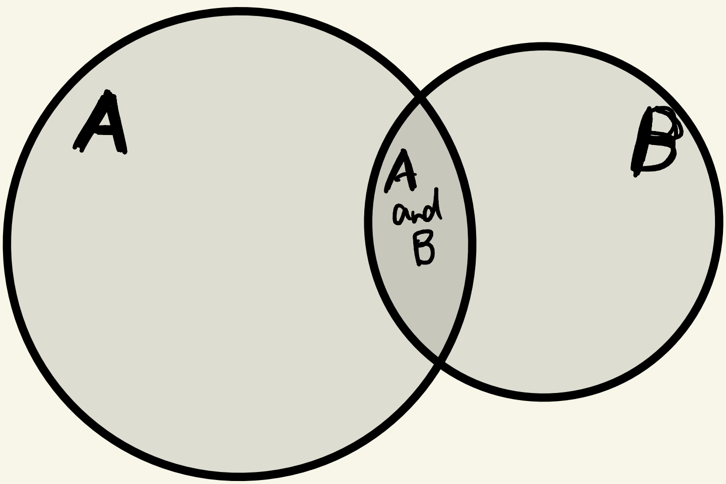
laplace poisson

Bayes Rule

Conditional probability

$A|B$ "A given B" "A conditional on B"

def $\boxed{\mathbb{P}(A|B) := \dfrac{\mathbb{P}(A,B)}{\mathbb{P}(B)}}$



$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B,A)}{\mathbb{P}(B)}$$

$$= \boxed{\frac{\mathbb{P}(B|A)\,\mathbb{P}(A)}{\mathbb{P}(B)} = \mathbb{P}(A|B)}$$

Sometimes $A|B$ is unknown but $B|A$ is known

A occurs before B

$B|A$ easier to work with than $A|B$

Shaq $X \sim \text{Binomial}(n, p)$, $n = 100$. Then we observe him actually take n free throws, making $x = 54$.

We know $\mathbb{P}(X = 54 | p)$

$$\mathbb{P}(p \mid X=54) = \frac{\mathbb{P}(X=54 \mid p) \cdot \cancel{\mathbb{P}(p)}}{\cancel{\mathbb{P}(X=54)}}$$

<u>Thm</u> If A, B are independent then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A,B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$

2 diceRolls X, Y

$$\mathbb{P}(X=2 \mid X+Y=5)$$

$$= \frac{\mathbb{P}(X+Y=5 \mid X=2)\, \mathbb{P}(X=2)}{\mathbb{P}(X+Y=5)}$$

$$= \frac{\mathbb{P}(Y=3) \cdot \mathbb{P}(X=2)}{\mathbb{P}(X+Y=5)}$$

- expectation $\mathbb{E}$
- independence
- variance
- density, CDF
- Bayes Rule