(-1) A Correction

Regrades open Wednesday at 8 pm, not Monday

⓪ More tips on problem solving

Every challenging mathematical problem is its own little world. To solve the problem, you need to become an expert on that world.

Some strategies

① Set the parameters to extreme values and see what happens.
E.g. try setting $P(A)=0$ or $P(A)=1$. Look at the complete graph and the graph with no edges, etc.

② Try out small examples where you can brute force the calculations. E.g. $n=2$ or $n=3$. Make up easy numbers

③ Make a guess about what's going on & perform "experiments" to try to refute it.

0.5 Recap

Previously on CS 70...

1. Independent events $P(A \cap B) = P(A)P(B)$

   "Knowing A doesn't tell you anything about B"

2. Intersections of events $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$

   Easy if A, B, C mutually independent

3. Unions of events

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

↳ inclusion-exclusion

Union bound: $P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$

↳ Looks silly but useful surprisingly often

When you can't calculate exactly, estimate/upper bound.

① Random variables

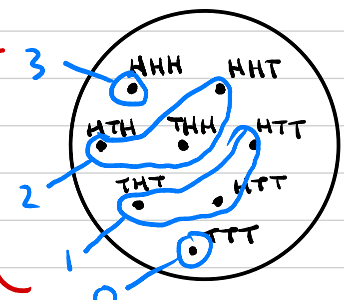
Common situation: want to measure some quantity of the outcomes in a sample space and answer questions like:

random variable

- Possible values and their probabilities Distribution
- Typical value of the quantity Expected value
- Is the quantity usually close to its "typical value"? Variance

Example Flip a fair coin 3 times & count the number of heads

Throws away a lot of information abt the outcomes



Possible values	Probability
3	1/8
2	3/8
1	3/8
0	1/8

distribution

<u>Def</u> Suppose Ω is a sample space. A **random variable** is a function $X : \Omega \to \mathbb{R}$

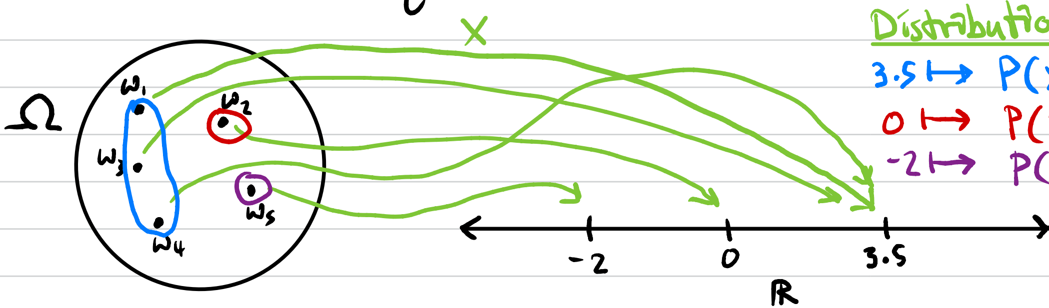
So for each outcome $\omega \in \Omega$, X tells you a number

$$\boxed{P(X = a)} = P(\{\omega \in \Omega \mid X(\omega) = a\})$$
$= $ probability that X is equal to a

The **distribution of X** is the function

$$\begin{aligned} \text{range}(X) &\longrightarrow [0,1] \\ a &\longmapsto P(X=a) \end{aligned}$$

that tells us for each possible value of X, the probability that X is equal to that value.



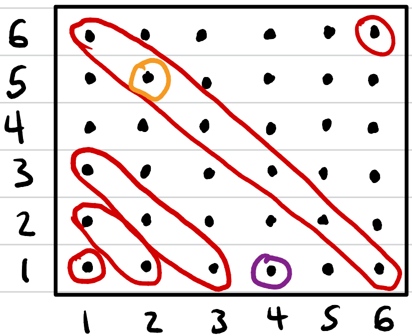
<u>Distribution of X</u>

$3.5 \mapsto P(X = 3.5) = P(\{\omega_1, \omega_3, \omega_4\})$

$0 \mapsto P(X = 0) = P(\{\omega_2\})$

$-2 \mapsto P(X = -2) = P(\{\omega_5\})$

Example Roll two fair dice and let X = sum of the two rolls

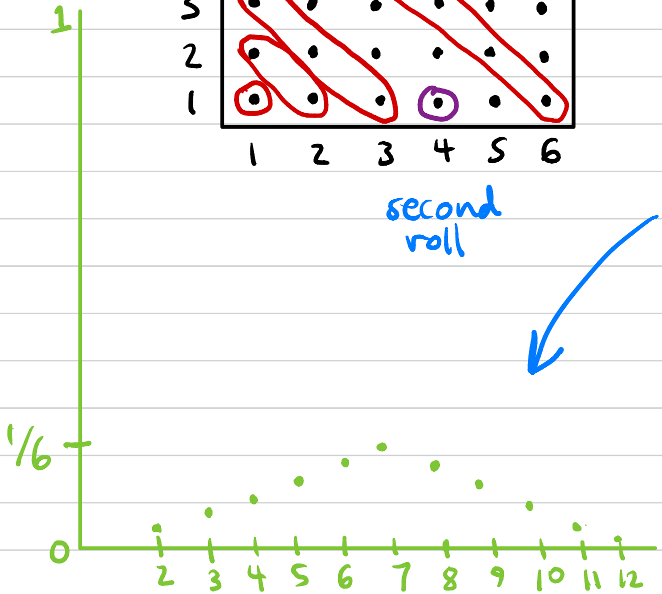


$X((1,4)) = 5$

$X((5,2)) = 7$

Distribution

$$\left\{\begin{array}{l}
2 \longmapsto P(X=2) = 1/36 \\
3 \longmapsto P(X=3) = 2/36 \\
4 \longmapsto P(X=4) = 3/36 \\
\vdots \\
7 \longmapsto P(X=7) = 6/36 \\
\vdots \\
12 \longmapsto P(X=12) = 1/36
\end{array}\right.$$



② Simple Random variables with fancy names

Some types of random variables show up so often that we give them fancy names

To be honest I never remember these names

Actually, technically we name the distributions of those random variables

2.1 Bernoulli Distribution

Have a biased coin which is heads with probability p. Flip it once. Define a random variable X by

$$X = \begin{cases} 1 & \text{if coin is heads} \\ 0 & \text{if coin is tails} \end{cases}$$

Distribution of X

$$P(X = i) = \begin{cases} p & \text{if } i = 1 \\ 1-p & \text{if } i = 0 \end{cases}$$

Write $X \sim \text{Bernoulli}(p)$ to summarize this situation

This is called the Bernoulli distribution with parameter p and we say "X is sampled from the Bernoulli distribution with parameter p"

I think it should be called the "Bernoulli random variable"

Fancy name, simple concept

Comment: How to get a simple concept in math named after you?
It helps to live in the 1600s

Example Roll a fair 6-sided die. Define a random variable X by

$$X = \begin{cases} 1 & \text{if roll is 1 or 2} \\ 0 & \text{else} \end{cases}$$

X is a random variable with Bernoulli(p) distribution.

Question What is p?

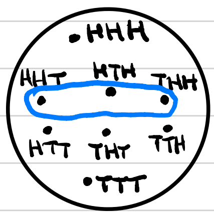
Answer $p = P(X=1) = \frac{2}{6} = \frac{1}{3}$

The point If all you care about is X, this situation is indistinguishable from flipping a coin which is heads with probability $1/3$

2.2 Binomial distribution

Have a biased coin which is heads with probability p. Flip it n times. Let X = number of heads.

Example $n=3$, $p=1/3$



$$
\begin{aligned}
P(X=2) &= P(HHT)+P(HTH)+P(THH) \\
&= P(H)P(H)P(T)+P(H)P(T)P(H)+P(T)P(H)P(H) \\
&= (1/3)(1/3)(2/3)+(1/3)(2/3)(1/3)+(2/3)(1/3)(1/3) \\
&= 3(1/3)^2(2/3)
\end{aligned}
$$

flips are independent

Distribution of X

$$P(X=k) = \sum_{\substack{\omega \text{ s.t.} \\ X(\omega)=k}} P(\omega) = \underbrace{\binom{n}{k}}_{\text{number of outcomes with } k \text{ heads}} \underbrace{p^k(1-p)^{n-k}}_{\text{probability of each outcome with } k \text{ heads}}$$

This is called the **binomial distribution with parameters n and p.** Written **$X \sim \mathrm{Bin}(n,p)$**

Binomial because of $\binom{n}{k}$ ("binomial coefficient")

<u>Example</u> When data is sent over the internet, it is broken up into small chunks called packets.

Each packet is sent separately and usually a few of the packets fail to reach the intended destination

How many packets successfully reach the destination?

"memory-less erasure channel"

This is often modelled as a random variable with binomial distribution $\text{Bin}(n, p)$ where

$n =$ # of packets

$p =$ probability that a single packet succeeds

What assumptions are we making?

① Success of packets are independent of each other (!)

② All packets have the same chance of success

③ Joint distributions

<u>Def</u> Suppose Ω is a sample space and $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are two random variables. The **joint distribution** of X and Y is the function

$$\begin{aligned} \text{range}(X) \times \text{range}(Y) &\longrightarrow [0,1] \\ (a,b) &\longmapsto P(X=a \text{ and } Y=b) \end{aligned}$$

intersection of 2 events

This tells you, for each pair of a possible value a of X and b of Y, the probability that $X=a$ and $Y=b$ occur simultaneously

<u>Example</u> Flip a fair coin 3 times.

X = # of heads

$$Y = \begin{cases} 1 & \text{if } 1^{st} \text{ flip is H} \\ 0 & \text{if } 1^{st} \text{ flip is T} \end{cases}$$

ω	$X(\omega)$	$Y(\omega)$
HHH	3	1
HHT	2	1
HTH	2	1
THH	2	0
HTT	1	1
THT	1	0
TTH	1	0
TTT	0	0

<u>Joint distribution</u>

X \ Y	0	1
3	0	1/8
2	2/8	1/8
1	2/8	1/8
0	1/8	0

3.1 Marginal distributions

Suppose we have 2 random variables and we know their joint distribution. How do we find the distribution of X by itself?

"Marginal distribution for X"

Example Suppose we have random variables X and Y

Joint distribution of X & Y:

X \ Y	1	2	3	distribution of X
-2	0	$1/12$	$2/12$	$3/12$
0	$1/12$	$1/12$	$3/12$	$5/12$
1.5	$1/12$	0	0	$1/12$
3	0	$1/12$	$2/12$	$3/12$
distribution of Y	$2/12$	$3/12$	$7/12$	

Question What is $P(X=-2)$?

Answer By the law of total probability,

$$P(X=2) = P(X=-2 \wedge Y=1) + P(X=-2 \wedge Y=2) + P(X=-2 \wedge Y=3)$$
$$= 0 + 1/12 + 2/12 = 3/12 = 1/4$$

Suppose we have 2 random variables and we know their joint distribution. How do we find the distribution of X by itself?

"Marginal distribution for X"

Prop For any random variables X and Y,

$$P(X=a) = \sum_{b \in \mathrm{range}(Y)} P(X=a \text{ and } Y=b)$$

"Marginalizing a joint distribution"

pf Law of total probability

3.2 Independence of random variables

Question Given distribution of X and distribution of Y, how can you compute the joint distribution of X and Y?

Answer Trick question! It depends on the relationship between X and Y

But there is one situation where this is easy to do

Def Random variables X and Y are **independent** if for all $a \in \text{range}(X)$ and $b \in \text{range}(Y)$,

$$P(X=a \text{ and } Y=b) = P(X=a)P(Y=b)$$

↳ Events $X=a$ and $Y=b$ are independent

Intuitive idea: Knowing the value of X tells you nothing new about the value of Y

Example Roll 2 fair dice

X = value of 1st roll

Y = value of 2nd roll

Z = sum of the two rolls

Question X and Y independent?

Answer Yes (basically by assumption)

$$P(X=a \text{ and } Y=b) = \frac{1}{36} = \frac{1}{6}\cdot\frac{1}{6} = P(X=a)P(Y=b)$$

Question X and Z independent?

Answer No!

$$P(X=1 \text{ and } Z=3) = P(\text{1st roll is } 1 \text{ \& 2nd roll is } 2)$$
$$= 1/36$$

$$P(X=1)P(Z=2) = 1/6\cdot(1/36) = 1/216 \neq 1/36$$

Intuitively, if you know sum is 3, it's more likely that 1st roll is 1

When you have > 2 random variables

Just like events,

Pairwise independence ≠ Mutual independence

<u>Def</u> Random variables $X_1, X_2, \dots, X_n$ are **pairwise independent** if for all $i \neq j$, X_i and X_j are independent

<u>Def</u> Random variables $X_1, X_2, \dots, X_n$ are **mutually independent** if for all $I \subseteq \{1, 2, \dots, n\}$ and all $\langle a_i \in \text{range}(X_i) \rangle_{i \in I}$

$$P\left(\bigcap_{i \in I} (X_i = a_i)\right) = \prod_{i \in I} P(X_i = a_i)$$

"independent & identically distributed"

<u>Common abbreviation</u> **$X_1, X_2, \dots, X_n$ are i.i.d. random variables** means the X_i are mutually independent and all have the same distribution as each other

Note : "Independent" usually means "mutually independent"

3.3 Combining random variables

Given random variables X, Y it is common to form a new random variable by combining X and Y in some way

Example Flip a fair coin 3 times

$X = \#$ of heads

$$Y = \begin{cases} 1 & \text{if } 1^{st} \text{ flip is heads} \\ 0 & \text{if } 1^{st} \text{ flip is tails} \end{cases}$$

Define $Z = X^2 + Y$

ω	$X(\omega)$	$Y(\omega)$	$Z(\omega)$
HHH	3	1	$3^2+1=10$
HHT	2	1	$2^2+1=5$
HTH	2	1	$2^2+1=5$
THH	2	0	$2^2+0=4$
HTT	1	1	$1^2+1=2$
THT	1	0	$1^2+0=1$
TTH	1	0	$1^2+0=1$
TTT	0	0	$0^2+0=0$

Question Are X, Y, Z mutually independent?

Answer No! If you know X, Y you know Z.

$P(X=3, Y=1, Z=5) = 0$

$P(X=3)\,P(Y=1)\,P(Z=5) = \frac{1}{8} \cdot \frac{1}{2} \cdot \frac{1}{4}$

↳ Actually X, Y also not independent

4 Expected value

Question I flip a fair coin 3 times and pay you \$1 for each H. How much would you pay me to play this game?

Answer Any amount less than \$1.50. Why?

Def If X is a random variable, the expected value of X, written $E(X)$, is defined by

$$E(X) = \sum_{a \in \text{range}(X)} a \cdot P(X=a)$$

Note: Expected value $\neq$ average. But they are similar

If you do something with expected value a and you repeat it many times (independently) then in the long run the average of those times will be close to a.
We'll see this more precisely next week

Def If X is a random variable, the expected value of X, written $E(X)$, is defined by

$$E(X) = \sum_{a \in \text{range}(X)} a \cdot P(X=a)$$

Example Flip a coin 3 times. $X = \#$ of heads

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3)$$
$$= 0 \cdot (1/8) + 1 \cdot (3/8) + 2 \cdot (3/8) + 3 \cdot (1/8)$$
$$= (3+6+3)/8 = 12/8 = \boxed{1.5}$$

Prop $E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$ ← Can sum over outcomes instead of over possible values of X

pf
$$E(X) = \sum_{a \in \text{range}(X)} a \cdot P(X=a)$$
$$= \sum_{a \in \text{range}(X)} a \cdot \Big(\sum_{\omega \in \Omega \text{ s.t. } X(\omega)=a} P(\omega) \Big)$$
$$= \sum_{a \in \text{range}(X)} \sum_{\omega \in \Omega \text{ s.t. } X(\omega)=a} \big(\cancel{a}^{\,X(\omega)} \cdot P(\omega) \big)$$
$$= \sum_{\omega \in \Omega} \big(X(\omega) \cdot P(\omega) \big)$$

Example $X \sim Bin(n, p)$ ← Coin with p chance of heads. Flip it n times
X = # of heads

$$E(X) = \sum_{a \in range(X)} a \cdot P(X = a)$$

$$= \sum_{k=0}^{n} k \cdot P(X = k)$$

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

$$= ??$$

Enter, the hero: Linearity of expectation

⑤ Linearity of Expectation

no assumptions about $X_1, \dots, X_n$ at all!

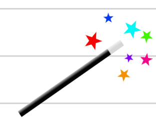
Thm For any random variables $X_1, X_2, \dots, X_n$

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

pf.

$$\begin{aligned}
E(X_1 + X_2 + \dots + X_n) &= \sum_{\omega \in \Omega} (X_1 + \dots + X_n)(\omega) \cdot P(\omega) \\
&= \sum_{\omega \in \Omega} (X_1(\omega) + \dots + X_n(\omega)) \cdot P(\omega) \\
&= \sum_{\omega \in \Omega} (X_1(\omega) P(\omega) + \dots + X_n(\omega) P(\omega)) \\
&= \left(\textstyle\sum_{\omega \in \Omega} X_1(\omega) P(\omega)\right) + \dots + \left(\textstyle\sum_{\omega \in \Omega} X_n(\omega) P(\omega)\right) \\
&= E(X_1) + \dots + E(X_n)
\end{aligned}$$

The magic wand of probability theory.



Expected value of binomial distribution, revisited

Have a coin which is heads with probability p
Flip it n times
X = # of heads

For $i \in \{1, 2, 3, \ldots, n\}$ define random variables

$$X_i = \begin{cases} 1 & \text{if flip } i \text{ is H} \\ 0 & \text{else} \end{cases}$$

"indicator variable"

Note: $X = X_1 + X_2 + \ldots + X_n$

$E(X_i) = 1 \cdot p + 0 \cdot (1-p) = p$

$E(X) = E(X_1 + X_2 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = n \cdot p$

linearity of expectation

<u>Cor</u> $\displaystyle\sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} = n \cdot p$ ← can also prove this by taking derivative of $(x + (1-p))$