Computer Science

CSC380: Principles of Data Science

Probability 3

Xinchen Yu

Outline

- Independence
- Random variables
- Distribution

Independence

Independence

- Informally, given two events A and B, they are <u>**independent**</u> if the probability of A is not affected by whether B is true or false (and vice versa)
  - E.g., A = “die1=1” and B=“die2=1” are independent.

    $\Longrightarrow$ the probability of die1 being 1 would not be changed just because die2=1.
- Mathematically, this can be written as $P(A|B) = P(A)$ or $P(B|A) = P(B)$.

- E.g., A = “die1=1” and B=“two dice sum to 6” are <u>**not independent**</u>.

quiz candidate

$\because$ P(A) = 1/6 = 0.166... . However, P(A|B) = 1/5 = 0.2

A = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)}

B = { (1,5), (2,4), (3,3), (4,2), (5,1)}

Independence

- Informally, given two events A and B, they are **independent** if the probability of A is not affected by whether B is true or false (and vice versa)
- Mathematically, this can be written as $P(A|B) = P(A)$ or $P(B|A) = P(B)$.

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

$$P(A,B) = P(A)P(B)$$

$A \perp B$: A and B are independent

Independence

[Def] Two events A and B are **independent** if

$$P(A, B) = P(A)P(B)$$

$A \perp B$ means A and B are independent

"joint probability is product of two marginal probabilities"

=> note: symmetric!

Also, a set of events $\{A_i\}_{i=1}^{n}$ (n can be ∞) are **mutually independent** if

for every J $\subseteq \{1, \ldots, n\}$, we have $P\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$

Random Events and Probability

Rolling two fair dice

(2,2) (3,5) (1,2)
(6,2)
(5,2) (5,3) (5,6)
(1,6) (2,4)
(3,4)
(4,3) (2,6) (4,2)
(6,4)
(2,1) (1,5) (5,1)
(3,2)
(4,6)
(4,4) (6,5) (6,1) (3,3)
(5,4)
(3,1) (1,4)
(4,5) (5,5) (6,3) (1,1)
(4,1)
(3,6) (2,5) (6,6) (1,3) (2,3)

Each outcome is equally likely by the **independence**
=> 1/36

Independence

- Ex) recall two fair dice
  - We took it for granted that P((1,1)) is 1/36.
  - But why is it true, really?

$$P(\text{die1}=1, \text{die2}=1) = P(\text{die1}=1) P(\text{die2}=1) = 1/6 \cdot 1/6$$

- E.g., two biased coin **C1** and **C2**. Suppose P(C1=H) = 0.3 and P(C2=H) = 0.4. Compute the probability of P(C1=H,C2=T).

0.3·0.6 = 0.18

Quiz candidate

Example: Dependent Coin Flips

- First coin (X1): fair coin
- Second coin (X2):
  - if X1=H, throw a **fair** coin.
  - If X1=T, throw an **unfair** coin P(H) = 0.2, P(T) = 0.8

- Q: Are X1=H and X2=H independent or not?

P(X1=H) = ____ 0.5

P(X2=H) = ____ = P(X2=H,X1=H) + P(X2=H,X1=T) = 0.25 + 0.1 = 0.35

P(X1=H, X2=H) = ____ 0.25

P(X1=H)*P(X2=H) = 0.175

Quiz candidate

Review

Axiom 3:

For any *finite* or *countably infinite* sequence of disjoint events $E_1, E_2, E_3, \ldots$, $P\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} P(E_i)$

Inclusion-exclusion rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Law of total probability: For events $B_1, B_2, \ldots$ that partitions Ω, $P(A) = \sum_i P(A \cap B_i)$

Conditional probability: $P(A|B) \coloneqq \dfrac{P(A \cap B)}{P(B)}$ ($P(A|B) \neq P(B|A)$ in general)

Probability chain rule: $P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C)$

Law of total probability + Conditional probability: $P(A) = \sum_i P(A \cap B_i) = \sum_i P(B_i)P(A|B_i) = \sum_i P(A)P(B_i|A)$

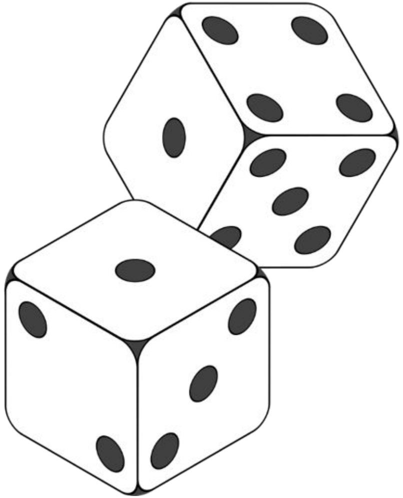
Bayes' rule: $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(B|A)P(A)}{P(B)}$

Independence: (definition) A and B are independent if $P(A, B) = P(A)P(B)$

(property) A and B are independent if and only if $P(A|B) = P(A)$ (or $P(B|A) = P(B)$)

Random Variables and Probability

Random Variables and Probability



Suppose we are interested in probabilities about the sum of two dice…

Option 1 Let E_i be event that the sum equals *i*

Two dice example:

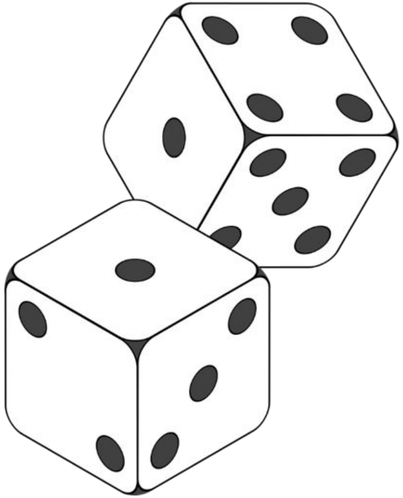
$$E_2 = \{(1,1)\} \quad E_3 = \{(1,2),(2,1)\} \quad E_4 = \{(1,3),(2,2),(3,1)\}$$

$$E_5 = \{(1,4),(2,3),(3,2),(4,1)\} \quad E_6 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

Enumerate all possible outcomes obtaining the desired sum.
Gets cumbersome for N>2 dice…

Random Variables and Probability

Suppose we are interested in probabilities about the <u>sum of dice</u>…



<u>Option 2</u> Give it a name

Let X be the sum of two dice.

We can say the event "X = i" to mean E_i.

X is called random variable.

$P(X=2) = 1/36$
$P(X=3) = 2/36$
$P(X=4) = 3/36$
...
$P(X=12) = 1/36$

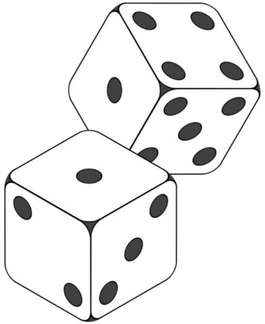
Random Variables and Probability

A random variable is a numerical description of the outcomes of a statistical experiment.

Example 1

- let X = sum of two dice;
- probability of X on different values:

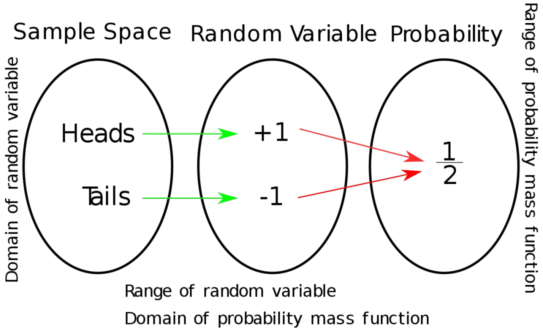
$$P(X = 2) = 1/36$$
$$P(X = 3) = 2/36$$
$$P(X = 4) = 3/36$$
$$\dots$$
$$P(X = 12) = 1/36$$



Example 2.

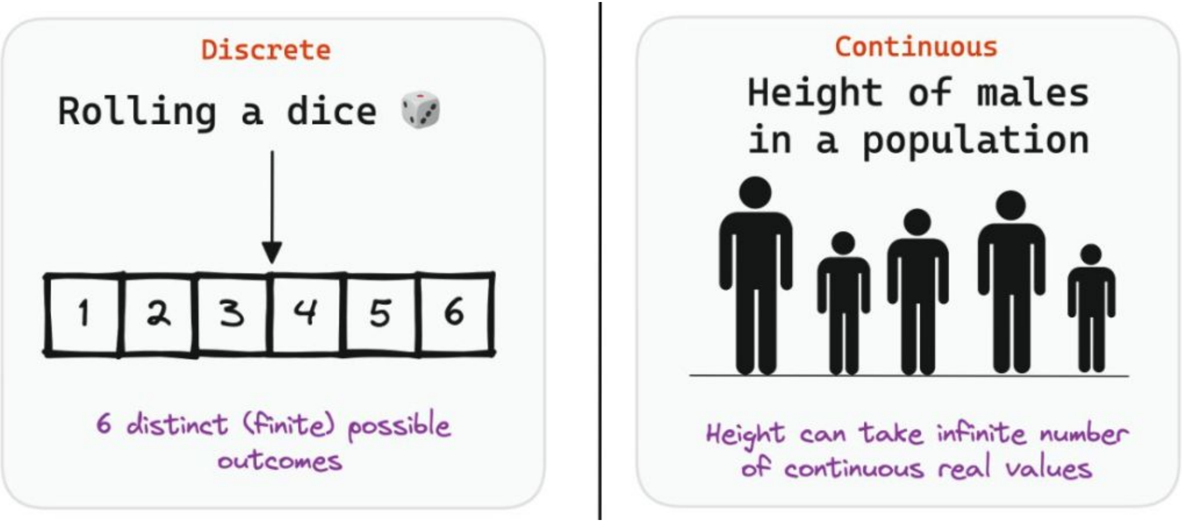
- let Y = outcomes of one coin toss;
- probability of Y on 1 (head) and -1 (tail):

$$P(Y = 1) = 1/2$$
$$P(Y = -1) = 1/2$$



Random Variables and Probability

- Discrete random variable: takes a finite or countable number of distinct values.
- Continuous random variable: takes an infinite number of values within a specified range or interval.



Random Variables and Probability

All the laws/rules about events applies to RVs.

The ***law of total probability*** for random variable is,

$$P(y) = \sum_i P(y, x_i) \qquad P(Y = y) = \sum_x P(Y = y, X = x)$$

for all x: P(X=x) >0

… you will also see people write down $p(Y) = \sum_x p(Y, X = x)$

This means $p(Y = y) = \sum_x p(Y = y, X = x)$ for all y

Random Variables and Probability

- I have three bags that each contain 100 marbles:
  - Bag A has 75 red and 25 blue marbles;
  - Bag B has 60 red and 40 blue marbles;
  - Bag C has 45 red and 55 blue marbles.

$$P(Y=y)=\sum_x P(Y=y,X=x)$$

I choose one of the bags at random and then pick a marble from the chosen bag, also at random. What is the probability that the chosen marble is red? (A: $X{=}1$, B: $X{=}2$, C: $X{=}3$)

$$P(Y=1|X=1)=0.75$$
$$P(Y=1|X=2)=0.60$$
$$P(Y=1|X=3)=0.45$$

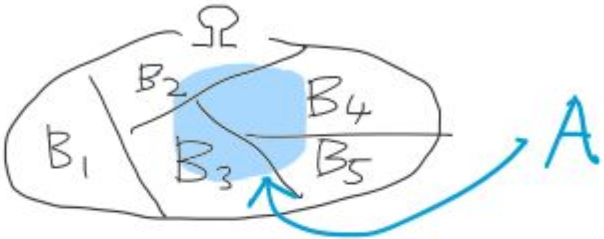
Y: pick a marble
X: choose a bag

$$P(\mathrm{X}=1)=\mathrm{P}(\mathrm{X}=2)=\mathrm{P}(\mathrm{X}=3)=\frac{100}{300}=1/3$$

$$\begin{aligned}P(Y=1)&=P(Y=1,X=1)+P(Y=1,X=2)+P(Y=1,X=3)\\&=P(Y=1|X=1)P(X=1)+P(Y=1|X=2)P(X=2)+P(Y=1|X=3)P(X=3)\\&=0.75\times\frac{1}{3}+0.60\times\frac{1}{3}+0.45\times\frac{1}{3}\\&=0.60\end{aligned}$$

Conditional Probability

$$P(Y) = \sum_x P(Y, X = x)$$



Also works for conditional probabilities,

$$p(Y \mid Z) = \sum_x p(Y, X = x \mid Z)$$

HW1, hint 3

Rule: Any rules about the probability still works for the conditional probabilities!!

(just make sure you add <u>the conditioning part</u> for every p()!)

Proof:

$$P(Y|Z) = \frac{P(Y,Z)}{P(Z)} = \frac{\sum_x P(Y,Z,X=x)}{P(Z)} = \frac{\sum_x P(Y,X=x|Z)P(Z)}{P(Z)} = \sum_x P(Y, X = x|Z)$$

Conditional Probability

Conditional probability $p(X \mid Y) = \frac{p(X,Y)}{p(Y)}$

Conditional probability version

$$p(X|Y,Z) = \frac{p(X,Y|Z)}{p(Y|Z)}$$

Proof:

$$p(X|Y,Z) = \frac{p(X,Y,Z)}{p(Y,Z)} = \frac{p(X,Y|Z)p(Z)}{p(Y|Z)p(Z)}$$

chain rule

Conditional Probability

Conditional probability version

Conditional probability $p(X \mid Y) = \frac{p(X,Y)}{p(Y)}$

$$p(X|Y,Z) = \frac{p(X,Y|Z)}{p(Y|Z)}$$

↑ there is no 'double' conditioning

Chain rule: $p(X,Y) = p(X|Y)p(Y)$

$$p(X,Y|Z) = p(X|Y,Z)p(Y|Z)$$

HW1, hint 4

Bayes rule: $p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}$

$$p(X|Y,Z) = \frac{p(Y|X,Z)p(X|Z)}{p(Y|Z)}$$

Proof:

$$p(X|Y,Z) = \frac{p(X,Y,Z)}{p(Y,Z)} = \frac{p(Y|X,Z)p(X,Z)}{p(Y,Z)} = \frac{p(Y|X,Z)p(X|Z)p(Z)}{p(Y|Z)p(Z)}$$

Tabular Calculations for Random Variables

Tabular representation of two binary RVs (joint probability)

Use K-by-K probability table for K-valued discrete RVs

e.g., X = disease, Y = test result

$$P(Y) = \sum_x P(Y, X = x)$$

X		y_1	y_2	P(X)
	x_1	0.04	0.36	0.4 ($P(x_1)$)
	x_2	0.30	0.30	0.6 ($P(x_2)$)
	P(Y)	0.34 ($P(y_1)$)	0.66 ($P(y_2)$)	

$P(y_1)=P(x_1,y_1)+P(x_2,y_1)$
$P(y_2)=P(x_1,y_2)+P(x_2,y_2)$
[i.e., sum down columns]

$P(X=x_1)=P(x_1,y_1)+P(x_1,y_2)$
$P(X=x_2)=P(x_2,y_1)+P(x_2,y_2)$
[i.e., sum across rows]

Tabular Calculations for Random Variables

e.g., X = disease, Y = test result

		Y: y_1	y_2
X	x_1	0.04	Censored!
	x_2	0.30	
		0.34	

We don't care about event $Y=y_2$

$P(y_1)$

$P(X \mid Y = y_1)$?

Tabular Calculations for Random Variables

		$Y=y_1$
X	x_1	0.04
	x_2	0.30
		0.34

$P(y_1)$

⇨

0.04 / 0.34
0.30 / 0.34

$P(X|y_1)$

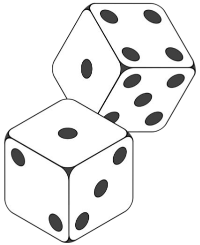
$$P(X=x_1 | Y = y_1) = P(x_1, y_1)/P(y_1)$$

These sum to one:
A conditional probability is still a 'probability'.

Independence

Definition *Two random variables* X *and* Y *are* ***independent*** *given if and only if*

$$p(X = x, Y = y) = p(X = x)p(Y = y)$$



for all values x and y, and we say $X \perp Y$.

- From now on, we will just write it down as $p(X, Y) = p(X)p(Y)$
- Property: X and Y are independent if and only if $p(X) = p(X|Y)$ (or $p(Y) = p(Y|X)$)

N RVs are independent if

$$p(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i)$$

Conditional Independence

Definition *Two random variables* X *and* Y *are conditionally independent given* Z *if and only if,*

$$p(X = x, Y = y \mid Z = z) = p(X = x \mid Z = z)p(Y = y \mid Z = z)$$

for all values x, y, and z, and we say that $X \perp Y \mid Z$.

N RVs conditionally independent, given Z, if and only if:

$$p(X_1, \ldots, X_N \mid Z) = \prod_{i=1}^{N} p(X_i \mid Z)$$

Caveat: $X \perp Y \neq X \perp Y|Z$

Discrete Distributions

Distribution and PMF

- If X is a random variable, then we can talk about its ‘distribution’
- **Distribution**: the set of values X can take and the probability assigned to each value.
- Examples:

X_1: unfair coin

value	prob.
1	0.2
2	0.8

X_2: unfair die

value	prob.
1	0.1
2	0.15
3	0.15
4	0.15
5	0.15
6	0.3

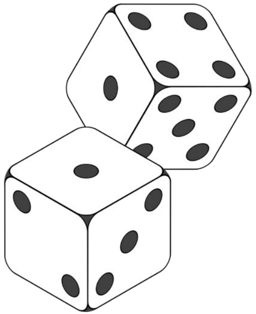
- Such a table can be viewed as a function f(x). This is called **probability mass function (PMF).**

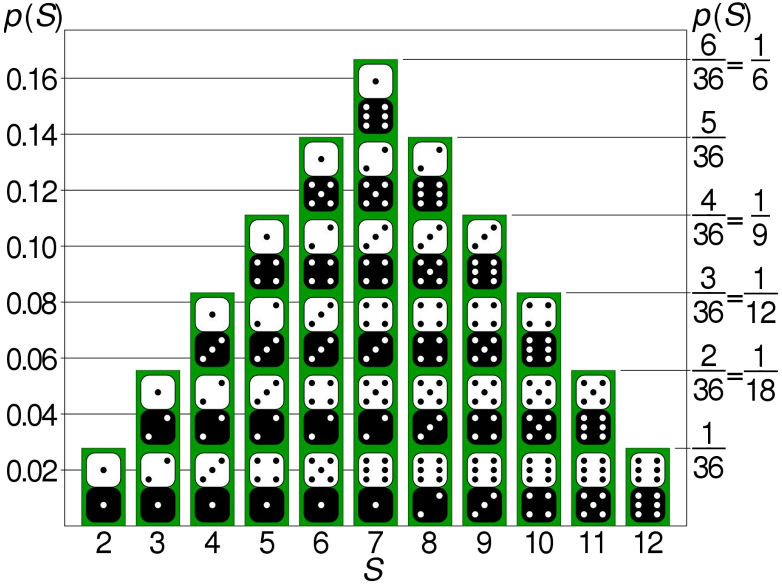
Distribution and PMF

Another example.

- let S = sum of two dice;
- probability of S on different values:

$$P(S=2) = 1/36$$
$$P(S=3) = 2/36$$
$$P(S=4) = 3/36$$
$$\ldots$$
$$P(S=12) = 1/36$$





PMF: $f_X(S) = \dfrac{\min(S-1, 13-S)}{36}$, for $S \in \{2,3,4,5,6,7,8,9,10,11,12\}$

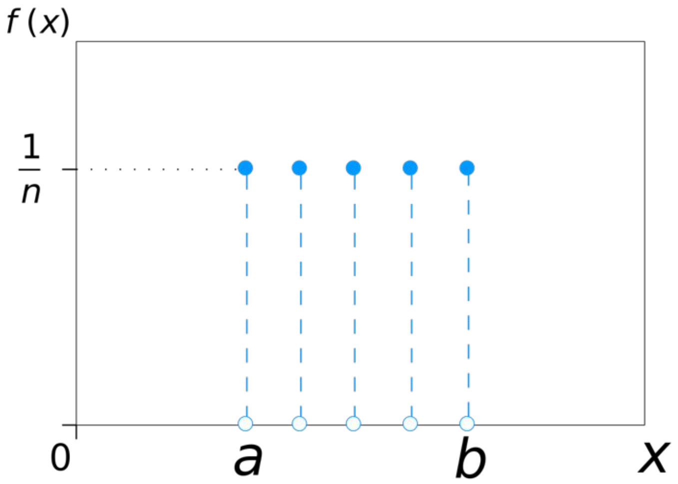
Uniform Distribution

Generalization of fair die with N-faced die. Its PMF is:

$$p(X = k) = \frac{1}{N}$$

More generally, we define a set of numbers $\{v_1, v_2, \ldots, v_N\}$

Discrete Uniform PMF



$$\text{Uniform}(X=k; \{v_1, v_2, \ldots, v_N\}) = \begin{cases} \frac{1}{N} & \text{if } k \in \{v_1, v_2, \ldots, v_N\} \\ 0 & \text{o.w.} \end{cases}$$

↑ it's like P(X=k) but being explicit about 'what' distribution X follows.

Bernoulli distribution

Bernoulli *a.k.a. the* ***coin flip*** *distribution on* *<u>binary</u>* *RVs* $X \in \{0, 1\}$

PMF: $p(X = x) = \pi^x (1 - \pi)^{1-x}$

Where π *is the probability of* ***success*** *(e.g., heads)*

used when you only have
2 possible outcomes

Example:

$$X \sim Bernoulli(0.2)$$

$$p(X = 0) = 0.2^0 \cdot (1 - 0.2)^{1-0} = 0.8$$

$$p(X = 1) = 0.2^1 \cdot (1 - 0.2)^{1-1} = 0.2$$

Binomial distribution

Suppose we flip N <u>independent</u> coins $X_1, X_2, \ldots, X_N$, what is the distribution over their sum $Y = \sum_{i=1}^{N} X_i$

Num. "successes" out of N trials

Num. ways to obtain k successes out of N

Binomial Dist.

$$p(Y = k) = \binom{N}{k} \pi^k (1 - \pi)^{N-k}$$

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 1}$$

Binomial distribution

Binomial Dist.

$$p(Y=k) = \binom{N}{k} \pi^k (1-\pi)^{N-k}$$

Why is this true?
Say N=5. Compute p(Y=3)

p(HTTHH) = $\pi(1-\pi)(1-\pi)\pi\pi$

p(TTHHH) =$(1-\pi)(1-\pi)\pi\pi\pi$

…

p(Y=3)=p(HTTHH, TTHHH, HHTTH,…., HHHTT)
=p(HTTHH) + p(TTHHH)+ ….+ p(HHHTT)
= $\binom{5}{3}\pi^3(1-\pi)^2$

The values are the same: $\pi^3(1-\pi)^2$!

By axiom 3, just add up $\pi^3(1-\pi)^2$ over all possible outcomes with the # of H is 3.
⇒ count: N choose k!

You'll use the same argument for HW1

Categorical distribution

Categorical Distribution on integer-valued RV $X \in \{1, \dots, K\}$ that takes $X = k$ with probability π_k.

$$p(X) = \prod_{k=1}^{K} \pi_k^{\mathbf{I}(X=k)} \quad \text{or} \quad p(X) = \sum_{k=1}^{K} \mathbf{I}(X=k) \cdot \pi_k$$

$$p(X = x) = \prod_{k=1}^{K} \pi_k^{\mathrm{I}(x=k)}$$

Indicator function

$$\mathbf{I}(A) := \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{othewise} \end{cases}$$

3 sided die example:

$$\begin{aligned} p(X = 2) &= \pi_1^{I(2=1)} \cdot \pi_2^{I(2=2)} \cdot \pi_3^{I(2=3)} \\ &= \pi_1^0 \cdot \pi_2^1 \cdot \pi_3^0 \\ &= \pi_2 \\ &= 0.1 \end{aligned}$$

$$\pi_1 = 0.2$$
$$\pi_2 = 0.1$$
$$\pi_3 = 0.7$$

Homework 1

Law of total probability for conditional probability $p(Y \mid Z) = \sum_x p(Y, X = x \mid Z)$

$$\boxed{P(W \mid S = (i,j)) = P(W, R_{i+j+1} = 1 \mid S = (i,j)) + P(W, R_{i+j+1} = 0 \mid S = (i,j))}$$

Chain rule $p(X,Y|Z) = p(X|Y,Z)p(Y|Z)$

$$\boxed{P(W, R_{i+j+1} \mid S = (i,j)) = P(W \mid R_{i+j+1}, S = (i,j))\boxed{P(R_{i+j+1} \mid S = (i,j))}} \rightarrow \tfrac{1}{2}$$

round i+j+1 you win and you have already win i rounds, opponents win j rounds = you win i+1, opponents win j

$$\boxed{P(W \mid R_{i+j+1} = 1, S = (i,j)) = P(W \mid S = (i+1,j))}$$

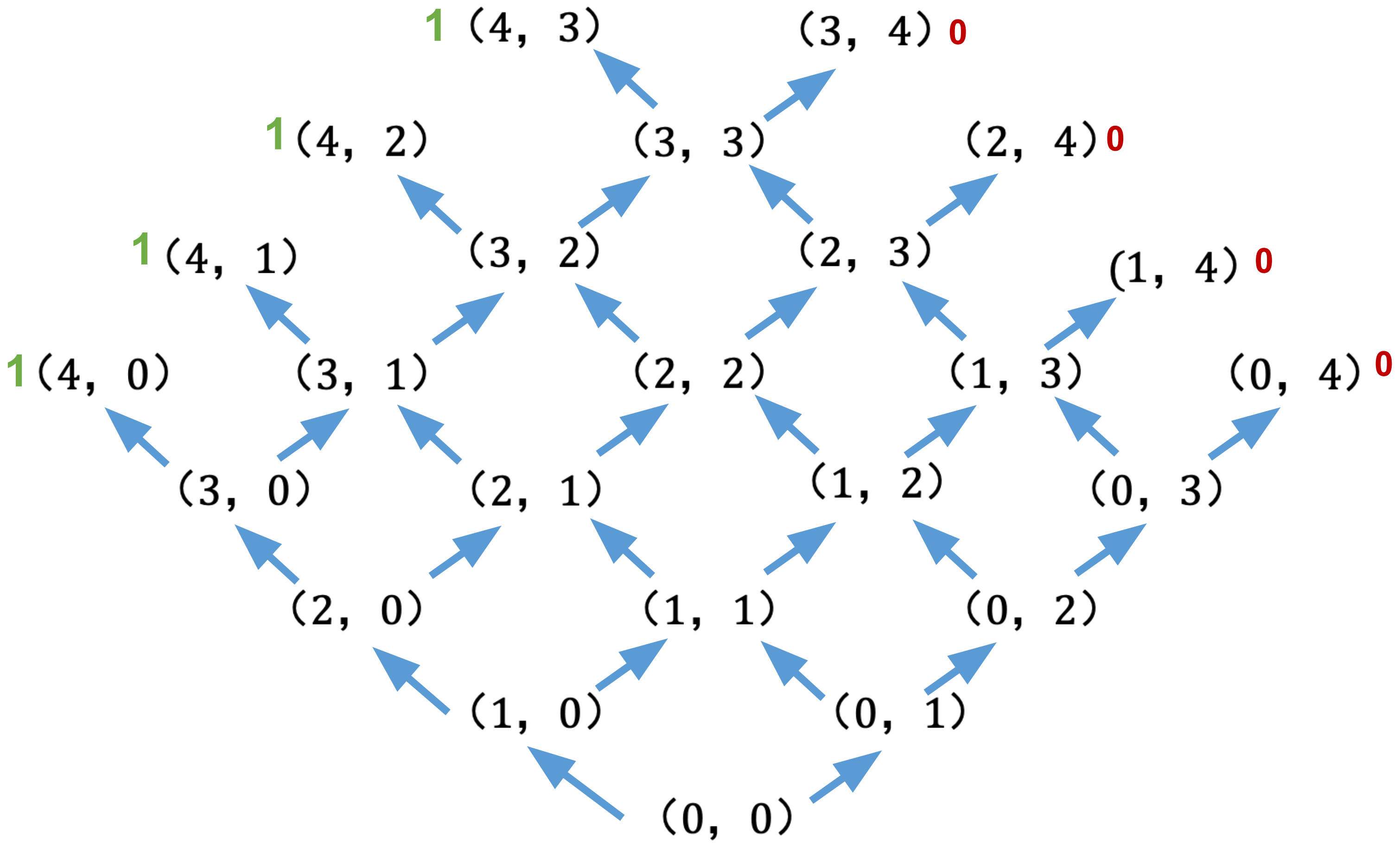
round i+j+1 you lose and you have already win i rounds, opponents win j rounds = you win i, opponents win j+1

$$\boxed{P(W \mid R_{i+j+1} = 0, S = (i,j)) = P(W \mid S = (i,j+1))}$$

We can get the probability of win in this round based on the probabilities of next round (recursive)

$$\boxed{P(W \mid S = (i,j)) = P(W \mid S = (i,j+1)) \times 1/2 + P(W \mid S = (i+1,j)) \times 1/2}$$

Homework 1



1 (4, 3)
(3, 4) 0
1 (4, 2)
(3, 3)
(2, 4) 0
1 (4, 1)
(3, 2)
(2, 3)
(1, 4) 0
1 (4, 0)
(3, 1)
(2, 2)
(1, 3)
(0, 4) 0
(3, 0)
(2, 1)
(1, 2)
(0, 3)
(2, 0)
(1, 1)
(0, 2)
(1, 0)
(0, 1)
(0, 0)