Computer Science

CSC380: Principles of Data Science

Probability Primer 3

Alon Efrat

Outline:
- Independence
- Random variables
- Distribution

Credit:
- Jason Pacheco,
- Kwang-Sung Jun,
- Chicheng Zhang
- Xinchen yu

1

1

Outline

2

- Independence
- Random variables
- Distribution

2

Independence

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)
  - E.g., A = "die1=1" and B="die2=1" are independent.

    ⟹ 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 **not independent**.

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

  quiz candidate

  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

5

- 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 $P(B|A) = P(B)$.

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

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

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

5

Independence

6

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

6

Random Events and Probability

Rolling two fair dice

(2,2) (3,5) (6,2) (1,6)
(5,2) (5,3) (4,2) (5,6)
(3,4) (1,2)
(4,3) (2,6) (2,4) (3,1)
(3,2) (6,4) (4,6) (2,1) (1,5) (5,1)
(4,4) (6,5) (5,4) (6,1) (1,1) (3,3) (1,4)
(4,5) (5,5) (6,3) (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?
  - To be rigorous,

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

  due to independence.

- 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

9

- 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

9

Review

10

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: $P(A|B) \coloneqq \frac{P(A \cap B)}{P(B)}$ ($P(A|B) \neq P(B|A)$ 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: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$

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 $P(B|A) = P(B)$)

10

Random Variables and Probability

11

11

Random Variables

My own perspective:
A random variable is actually a **function. This function assigns a numerical value to each outcome of the experiment.**

These values do not need to "make sense". Any definition is legit.
Two outcomes could have the same valued assign to them by a random variable.

Example: Tossing a fair coin. Lets defined a new random var Q.
If the coin shows Face, then we decide that **Q=15**.
If the coin shows HEAD, then we decide that **Q=3.**

Now we can ask what are the events that are correlated to each value of Q? What is the probability of these events ?
For example,
what is the probability P(Q=12) ?
What is the probability P(Q=15) ?

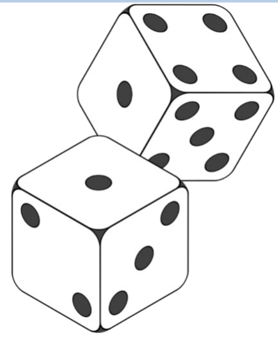
Harder: What is the probability P(Q<16) ? First find which event corresponds to.

12

12

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:

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

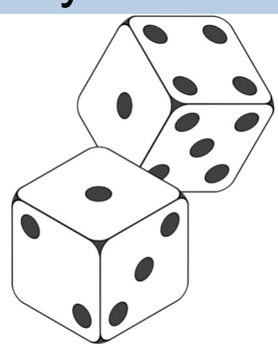
$$E_5 = \{(1,4),(2,3),(3,2),(4,1)\} \quad E_6 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

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

13

Random Variables and Probability

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



Option 2 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.

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

14

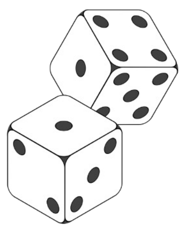
Random Variables and Probability

15

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

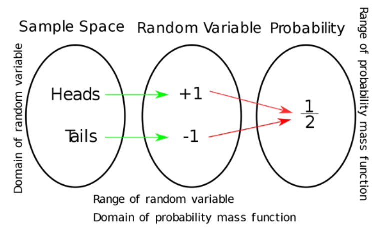
Example 1
- let X = sum of two dice;
- probability of X on different values:

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



Example 2.
- let Y = outcomes of one coin toss;
- probability of Y on 1 (head) and -1 (tail):

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

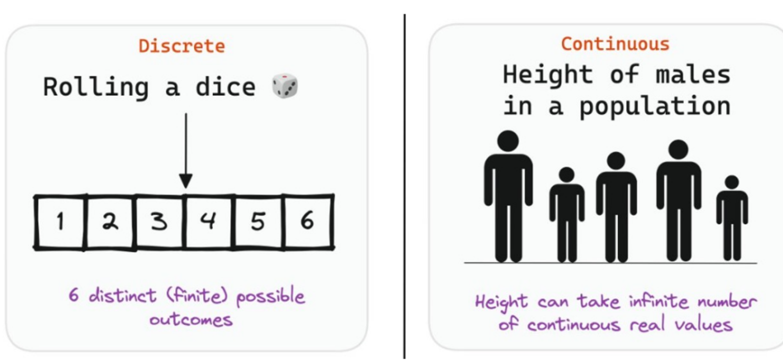


15

Random Variables and Probability

16

- A discrete random variable takes a finite or countable number of distinct values.
- A continuous random variable takes an infinite number of values within a specified range or interval.



16

Random Variables and Probability

17

- All the laws/rules about events applies to RVs.

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

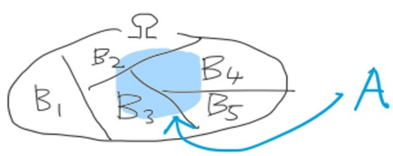
$P(y) = \sum_i P(y, x_i)$

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

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



17

Random Variables and Probability

18

- 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 UNIFORMLY at random and then pick a marble from the chosen bag, also UNIFORMLY at random. What is the probability that the chosen marble is **red**?

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

Y: pick a marble
X: choose a bag

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

Y=1 iff RED murble

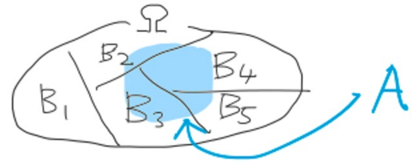
$$\begin{aligned} P(Y = 1) &= P(Y = 1, X = 1) + P(Y = 1, X = 2) + P(Y = 1, X = 3) \\ &= P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 2)P(X = 2) + P(Y = 1|X = 3)P(X = 3) \\ &= 0.75\times\frac{1}{3} + 0.60\times\frac{1}{3} + 0.45\times\frac{1}{3} \\ &= 0.60 \end{aligned}$$

18

Conditional Probability

19

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



Also works for conditional probabilities,

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

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

(just make sure you add the conditioning part for every p()!)

Proof:

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

19

Conditional Probability

20

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:

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

$$p(A|B) = \frac{p(A \cap B)}{p(B)}$$

20

Conditional Probability

21

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

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

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

Conditional probability version

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

↑ there is no 'double' conditioning

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

HW1, hint 4

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

Proof:

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

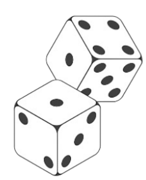
21

Independence

25

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

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

(Again, for all the possible values $x_1, \ldots, x_N$)

Note – this does not mean that they are pairwise independent

25

Conditional Independence

Definition *Two random variables X and Y are* <u>*conditionally independent*</u> *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

28

- 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).**

28

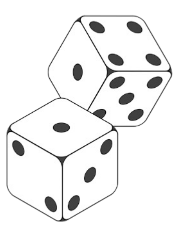
Distribution

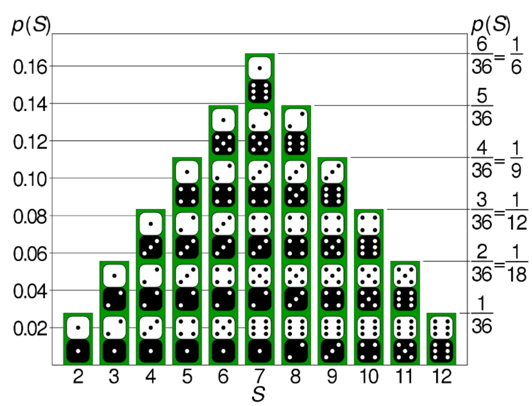
29

Another example.

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

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





PMF:- Probability Mass Function

29

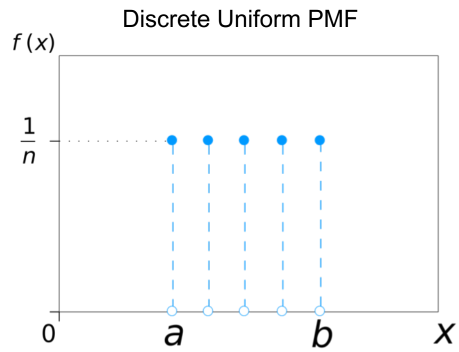
Uniform Distribution

30

Generalization of fair die with N-faced die. Its PMF (prob mass function) is:

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

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



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

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

30

Reminder: The number of combinations:

Assume we have 4 letters How many permutations ? Lets play. From the set ABCD, create a permutation π=$x_1x_2x_3x_4$ by first picking for x_4. There are 4 ways to pick x_4. After this choice we are left with only 3 choices for x_3, only 2 for x_3 and 1 for x_1 Altogether $4 \cdot 3 \cdot 2 \cdot 1 = 4!$

The same holds if we discuss n letters, π=$x_1x_2x_3x_4 \dots x_n$. The number of permutation is $n! = n(n-1)(n-2) \dots 1$.

What if the after setting the order, we decide that we only care which letters appear in the in the blue and which in the red of π=$x_1x_2x_3x_4x_5x_6$. We don't care about their order within the red and blue parts. Then the number of permutations *n!* should be divided by the number of permutations of red, and then by the number of permutation of blue. So the **binomial coefficient** is

$$\binom{n}{k} = \frac{n!}{k!\ (n-k)!}$$

31

31

Bernoulli distribution and Binomial dist.

32

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

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

Fair Coin: $\pi = \frac{1}{2}$.
So PMF= ½, ½.

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

Suppose we flip N independent 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. $p(Y = k) = \binom{N}{k} \pi^k (1 - \pi)^{N-k}$

Example:

Pay attention – really important

32

Binomial distribution

33

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

33

What does the PMF looks like when N increases

34

- https://www.geogebra.org/m/vxbve7s4

- Lets look at a population of frogs (many children to the same parents)
- All live in the same condition
- The weight of each frog= depends on N genes$(x_1, x_2 \dots x_N)$, each gene is either 0/1
- If $x_i = 0$, then the weight of the offspring stay “base weight” =100 grams
- If $x_i = 1$, then the weight of the offspring increased by 1 gram.
- $p(x_i = 1) = p(x_i = 0) = 0.5$
- Looking at the frogs population.
- How do we expect their weight to be distributed, if N=1 (only one gene) ?
- First interesting case: N=2, so $\mathrm{x}_1, \mathrm{x}_2 =$ {00,01,10,11}.
- For larger N, we have 2^N possible outcome.
- Let Y be the number of ones in $(x_1, x_2 \dots x_N)$. What are P(S=0), p(S=1), p(S=2) ?

34

What happened to the graph when N grows larger

35

Remember that in our case, $\pi = 0.5$

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

Use the gg app to see what happened when N becomes larger

35

Homework 1

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

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

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

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

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

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

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

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

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

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

36