Lecture 2 - 2/2/2021

Topics: 1. conditional probability, independence
2. constructing and computing probability spaces
3. discrete random variables

Recall:
For today, $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ i.e. discrete & finite.

we always assume $\mathcal{F} = \mathcal{P}(\Omega)$.
$\Rightarrow$ An event A is any subset of Ω, i.e $A \subseteq \Omega$.

we define probability measure
$P(A)$ = probability/chance that an outcome occurs.

Note: the elements $\omega_i \in \Omega$ are the outcomes of an experiment.

Therefore, $(\Omega, \mathcal{F}, P)$ = probability space.

Our main questions:
1. How do we construct (useful) probability spaces, most particularly the probability measure? some examples?
2. How do we compute the probability $P(A)$ for $A \subseteq \Omega$?

Why is this hard?

An Answer to 1. : Exercise 1.17
$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$.

Define $P(\{\omega_1\}) = P(\omega_1) = a_1$, $P(\omega_2) = a_2, \dots, P(\omega_n) = a_n$
where we pick a_i s.t. $a_i \geq 0$, $\sum_{i=1}^{n} a_i = 1$.

Then for any A, $P(A) = \sum_{w_i \in A} P(w_i) = \sum_{w_i \in A} a_i$.

Ex: $\Omega = \{1, 2, 3, 4, 5, 6\}$

$P(1) = 1/2 = a_1$
$P(2) = 1/10 = a_2$
$P(3) = 1/10 = a_3$
$\vdots$
$P(6) = 1/10 = a_6$

$\Rightarrow P(\{\text{roll odd}\})$
$= P(\{1, 3, 5\})$
$= P(1) + P(3) + P(5)$
$= 1/2 + 1/10 + 1/10 = 7/10$.

why can't we always build probabilities and compute probabilities in this manner?.

Ex: I flip a coin 100 times. Let N_H be the number of H I flip. I roll N_H dice. I am interested in the probability that the sum of the dice is more than 300, (A).

$$\Omega = \{TTT\cdots T,\ THT\cdots T1, \cdots,\ HTT\cdots T1,\ HTT\cdots T2,\ \vdots\ HTT\cdots T6, \cdots\}$$

Hard to write Ω on paper, to describe it.

$\Rightarrow$ Need to define probabilities
$P(TT\cdots T) = a_1$ $\quad (\frac{1}{2})^{100}$
$P(HT\cdots T1) = a_2$ $\quad (\frac{1}{2})^{100}(\frac{1}{6})$
$P(HT\cdots T2) = a_3$
$\vdots$

what are the a_i?

so it is not practical to enumerate $a_1, \ldots, a_n$ (we don't even know what is n).

$\Rightarrow P(A) = \sum_{w_i \in A} P(w_i) = \sum_{w_i \in A} a_i$, $A = \{\text{dice roll sum to more than } 300\}$.

It is challenging to enumerate A ~~and $w_i \in A$~~.

The above A is too big to do this kind of computation, even on a super computer.

Comment: Probability spaces need to have structure or patterns, otherwise it's difficult to compute probabilities.

Today, two concepts help provide pattern & structure:

1. conditional probability
2. independence.

Def: Conditional Probability: Given Ω, P, two events A, B

$P(A|B)$ = probability that an outcome in A occurs if we know that an outcome in B occurred.

e.g. P(roll 1 | roll odd)

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

Def: Independence: Two events A, B are independent if $P(A\cap B) = P(A)P(B)$. Equivalently,

$P(A|B) = P(A)$, i.e. the fact that B occurred doesn't impact the probability of A.

e.g. Flip two coins. $A = \{\text{first flip } H\}$
$B = \{\text{second flip } H\}$.

$\Omega = \{HH, TH, HT, TT\}$

$P(HH) = 1/4$, $P(TH) = 1/4$, $P(HT) = 1/4$, $P(TT) = 1/4$.

Let's show A, B are independent. $\Rightarrow$ Need to show:

$$P(A \cap B) = P(HH) = 1/4$$
$$P(A) = P(\{HH, HT\}) = 1/4 + 1/4 = 1/2$$
$$P(B) = P(\{TH, HH\}) = 1/4 + 1/4 = 1/2$$

$$\Rightarrow P(A \cap B) = P(A)P(B) \Leftrightarrow \frac{1}{4} = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right).$$

Ex: Suppose I flip a coin repeatedly until I flip a H or until I flip 100 times. Assume that each coin flip is independent.

1. an event A that involves just the outcome of coin flip i is independent of all events that do not involve the outcome of coin flip i.

   e.g. $A = \{37^{th}$ coin flip is $H\}$
   $B = \{39^{th}$ coin flip is $T\}$ } independent.

2. any event involving a single coin flip has the probability given by a single coin flip probability space.

   e.g. $\Omega_c = \{H, T\}$, $P(H) = P(T) = \frac{1}{2}$ $\Rightarrow \Omega = \{H, TH, TTH, \dots, \overbrace{TT\cdots T}^{99}H, \underbrace{TT\cdots T}_{100}\}$.

The probability measure follows from (Ω_c, P) and independence.

$$P(TTH) \Leftarrow \begin{bmatrix} A_1 = \{\text{flip } T \text{ on } 1^{st} \text{ flip}\} \\ A_2 = \{\text{flip } T \text{ on } 2^{nd}\} \\ A_3 = \{\text{flip } H \text{ on } 3^{rd}\}. \end{bmatrix}$$

$$\begin{aligned} &\;\;\;\; \| \\ &P(A_1 \cap A_2 \cap A_3) \\ &\;\;\;\; \| \\ &P(A_1)P(A_2)P(A_3) \text{ by independence.} \\ &\;\;\;\; \| \\ &(1/2)(1/2)(1/2) = 1/8. \end{aligned}$$

<u>Note</u>: Due to the independence, we have structure of the large Ω using the constituent Ω.

↙ using independence and conditional probability.

<u>Ex</u>: Suppose I have 4 coins. The coins are biased with probability of H given by $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$ respectively. I pick one of the coins with uniform probability. And then I flip the coin as before (until H or 100). Assume the coin flips are ind. as before.

$$\begin{aligned} \Omega = \{ &C_1H, C_1TH, C_1TTH, \ldots, C_1T\cdots TH, C_1T\cdots T, \\ &C_2H, C_2TH, \ldots \\ &C_3H, C_3TH \cdots \\ &C_4H, C_4TH \cdots \}. \end{aligned}$$

$$P(C_2TH) = P(C_2)\cdot P(TH \mid C_2)$$

↑
pick coin 2 and then flip TH

$$= P(C_2) \cdot \frac{P(TH \cap C_2)}{P(C_2)} = P(TH \cap C_2).$$

The above helps us compute more easily.

$$\begin{aligned} P(C_2TH) &= P(C_2)\cdot P(TH/C_2) \\ &= 1/4 \cdot 2/3 \cdot 1/3 = \frac{2}{36} = \frac{1}{18}. \end{aligned}$$

Followup example:

$P(\{\text{I flip more than 50 times before I get an } H\})$.

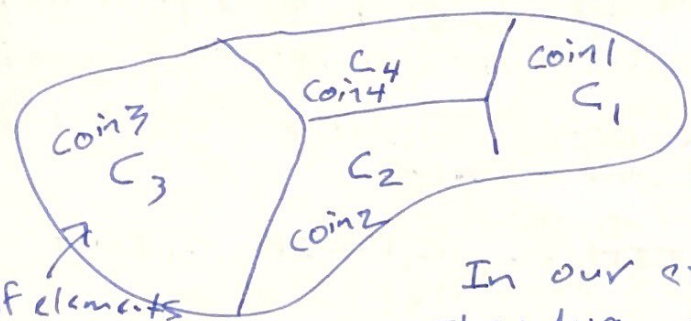
To start, assume I pick coin 2, C_2.

$$\Rightarrow P(\{\underbrace{T\cdots T}_{50}H, \underbrace{T\cdots T}_{51}H, \dots, \underbrace{T\cdots T}_{99}H\})$$

$$= P(\underbrace{TT\cdots T}_{50}H) + P(\underbrace{T\cdots T}_{51}H) + \dots + P(\underbrace{T\cdots T}_{99}H) = \left(\frac{2}{3}\right)^{50}\frac{1}{3} + \left(\frac{2}{3}\right)^{51}\frac{1}{3} + \dots + \left(\frac{2}{3}\right)^{99}\left(\frac{1}{3}\right)$$

$$= \frac{1}{3}\sum_{i=50}^{n}\left(\frac{2}{3}\right)^i$$

<u>Note</u>: when compute probabilities, it is useful to use conditional probability to split Ω so that there is a convenient structure.



Ω ← In each C_i, there is a nice consistency, which isn't true for all of Ω.

set of elements that contain coin 3.

In our example, our space has structure of 4 regions.

$$P(\{\text{I flip more than 50 times before I get } H\})$$

$$\begin{aligned} = P(A) &= P(\text{coin } 1)P(A \mid \text{coin } 1) && = \frac{1}{4}\left(\frac{1}{2}\sum_{i=51}^{99}\left(\frac{1}{2}\right)^i\right) \\ &+ P(\text{coin } 2)P(A \mid \text{coin } 2) && + \frac{1}{4}\left(\frac{1}{3}\sum_{i=51}^{99}\left(\frac{2}{3}\right)^i\right) \\ &+ P(\text{coin } 3)P(A \mid \text{coin } 3) && + \frac{1}{4}\left(\frac{1}{4}\sum_{i=51}^{99}\left(\frac{3}{4}\right)^i\right) \\ &+ P(\text{coin } 4)P(A \mid \text{coin } 4) && + \frac{1}{4}\left(\frac{1}{5}\sum(4/5)^i\right). \end{aligned}$$

The above is an example of the <u>Partition Theorem</u> (more on this in the homework).

<u>Note</u>: whenever we do these probabilities we are looking for some structure that makes it easier.

Ex: I flip a biased coin 100 times. The coin lands H with probability 1/3. What is the probability that I flip more than 55 H?

$\Omega = \{ TT\cdots T, HT\cdots T, \vdots \}$ $\Rightarrow P(\underbrace{HH}_{2}\underbrace{T\cdots T}_{98}) = P(H)P(H)P(T)\cdots P(T)$

$|\Omega| = 2^{100}$. $= \left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^{98}$.

$\Rightarrow P(\{\underbrace{\text{flip more than 55 } H}_{B}\})$. which outcomes have the same probability?

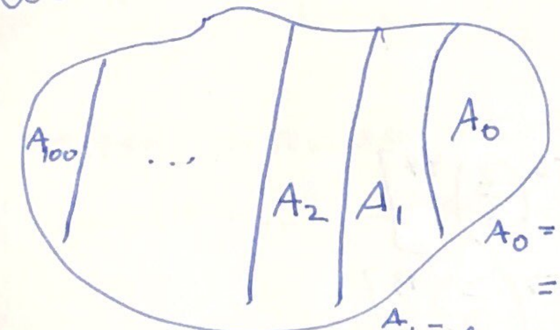
Note that the order doesn't matter in this example.

$P(\{k \text{ heads followed by } 100-k \text{ tails}\}) = \left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{100-k}$.

$P(\{\text{any outcome/element with } k\, H \text{ and } 100-k\ T\})$

$= \left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{100-k}$.

we observe the above gives us this schema:



Ω $\Rightarrow \Omega = A_1 \cup A_2 \cup \cdots \cup A_{100}$

$A_0 = \{\text{no } H\}$
$= \{\underbrace{T\cdots T}_{100}\}$

$A_1 = \{1H, 99T\} = \{HT\cdots T, THT\cdots T, \ldots\}$.

$B = A_{56} \cup A_{57} \cup \cdots \cup A_{100} \Rightarrow$

$P(B) = P(A_{56}) + P(A_{57}) + \cdots + P(A_{100})$.

$P(A_{56}) = \sum_{\omega_i \in A_{56}} P(\omega_i) = \sum_{\omega_i \in A_{56}} \left[\left(\frac{1}{3}\right)^{56}\left(\frac{2}{3}\right)^{44}\right] = \left[\left(\frac{1}{3}\right)^{56}\left(\frac{2}{3}\right)^{44}\right]\left(\sum_{\omega_i \in A_{56}} 1\right)$

Next PAGE $\Rightarrow$

$$= \left[\left(\frac{1}{3}\right)^{56}\left(\frac{2}{3}\right)^{44}\right] |A_{56}|$$

<u>Note</u>: When we compute the probability of an event A with elements with uniform probabilities, we need to compute the number of elements in A.

<u>Note</u>: There is a connection between probability and combinatorics, in that we can need to determine the sizes of things as above.

100 coin flips, 56 heads $\longrightarrow$ how many ways to do this?

$$\Rightarrow \binom{100}{56} = 100 \text{ choose } 56.$$

<u>Def</u>: $\binom{n}{j} =$ # of ways to choose j out of n possibilities $= \dfrac{n!}{(n-j)!\,j!}$

$$\Rightarrow |A_{56}| = \binom{100}{56} = \frac{100!}{44!\,56!}$$

$$\Rightarrow P\left(\left\{\begin{matrix}\text{more than} \\ 55\ H\end{matrix}\right\}\right) = \binom{100}{56}\left[\left(\frac{1}{3}\right)^{56}\left(\frac{2}{3}\right)^{44}\right] + \binom{100}{57}\left[\left(\frac{1}{3}\right)^{57}\left(\frac{2}{3}\right)^{43}\right] + \cdots + \binom{100}{100}\left[\left(\frac{1}{3}\right)^{100}\left(\frac{2}{3}\right)^{0}\right]$$

$$= \sum_{i=56}^{100} \binom{100}{i}\left(\frac{1}{3}\right)^{i}\left(\frac{2}{3}\right)^{100-i}$$

Random Variables:

Def: A discrete random variable is a function from a discrete ~~pr~~ sample space Ω of a probability space $(\Omega, \mathcal{F}, P)$ to the real numbers $\mathbb{R}$. Usually we denote random variables (r.v.) by a capital letter, e.g. X.

$$X : \Omega \to \mathbb{R}$$

$\mathbb{R} = (-\infty, \infty)$.

Ex: $\Omega = \{H, T\}$, $P(H) = P(T) = \frac{1}{2}$.

$X(H) = 1$, $X(H) = -57.3$
$X(T) = 0$, $X(T) = 22.6$

$P(X=1) = P(H) = 1/2$
$P(X=0) = P(T) = 1/2$

(I) $X = \begin{cases} 0 & \text{prob } 1/2 \\ 1 & \text{prob } 1/2 \end{cases}$

Note: we often don't care about the domain step Ω for X, and we just look at the example (I)

often discussed as probability mass function.
sometimes called distribution of X.

Ex: $\Omega = \{HH, TT, HT, TH\}$.

$P(HH) = (\frac{1}{3})^2$, $P(HT) = (\frac{1}{3})(\frac{2}{3})$, $P(TH) = (\frac{2}{3})(\frac{1}{3})$
$P(TT) = (\frac{1}{3})^2$.

$X = \#H \Rightarrow X(HH) = 2$, $X(HT) = 1$, $X(TH) = 1$
$X(TT) = 0$.

This is not the typical notation.

$$X = \begin{cases} 0 & \text{prob } \left(\frac{2}{3}\right)^2 \\ 1 & \text{prob } 2\left(\frac{1}{3}\right)\left(\frac{2}{3}\right) \\ 2 & \text{prob } \left(\frac{1}{3}\right)^2 \end{cases}$$

← distribution of X, pmf.

The above is the more typical notation. We usually ignore / throw out the underlying Ω. It is there but it isn't talked about often when we use r.v..