Chapter 2: Probability

The aim of this chapter is to revise the basic rules of probability. By the end of this chapter, you should be comfortable with:

- conditional probability, and what you can and can't do with conditional expressions;
- the Partition Theorem and Bayes' Theorem;
- First-Step Analysis for finding the probability that a process reaches some state, by conditioning on the outcome of the first step;
- calculating probabilities for continuous and discrete random variables.

2.1 Sample spaces and events

Definition: A **sample space**, Ω, is a *set of possible outcomes of a random experiment.*

Definition: An **event**, A, is a *subset of the sample space.*

This means that event A is simply *a collection of outcomes.*

Example:

Random experiment: Pick a person in this class at random.
Sample space: $\Omega = \{$*all people in class*$\}$
Event A: $A = \{$*all males in class*$\}$.

Definition: Event A **occurs** if *the outcome of the random experiment is a member of the set A.*

In the example above, event A occurs if *the person we pick is male.*

2.2 Probability Reference List

The following properties hold for all events A, B.

- $\mathbb{P}(\emptyset) = 0$.
- $0 \leq \mathbb{P}(A) \leq 1$.
- **Complement:** $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
- **Probability of a union:** $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

  For three events A, B, C:

  $$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)\,.$$

  If A and B are **mutually exclusive**, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

- **Conditional probability:** $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- **Multiplication rule:** $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$.
- **The Partition Theorem:** if $B_1, B_2, \ldots, B_m$ form a partition of Ω, then

  $$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) \quad \text{for any event } A.$$

  As a special case, B and $\overline{B}$ partition Ω, so:

  $$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \\ &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\mathbb{P}(\overline{B}) \quad \text{for any } A,\ B. \end{aligned}$$

- **Bayes' Theorem:** $\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}$.

  More generally, if $B_1, B_2, \ldots, B_m$ form a partition of Ω, then

  $$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)} \quad \text{for any } j.$$

- **Chains of events:** for any events $A_1, A_2, \ldots, A_n$,

  $$\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_2 \cap A_1) \ldots \mathbb{P}(A_n \mid A_{n-1} \cap \ldots \cap A_1).$$

$\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. $\dfrac{6}{50}$

2.3 Conditional Probability

Suppose we are working with sample space $\Omega = \{\text{people in class}\}$. I want to find the proportion of people in the class who ski. What do I do?

Count up #people in class who ski, and divide by the total #people in class.

$$\mathbb{P}(\text{person skis}) = \frac{\#\text{skiers in class}}{\#\text{people in class}}$$

Now suppose I want to find the proportion of *females* in the class who ski. What do I do?

Count up #females in the class who ski, and divide by total #females in class.

$$\mathbb{P}(\text{female skis}) = \frac{\#\text{female skiers in class}}{\#\text{females in class}}.$$

By changing from asking about everyone to asking about females only, we have:

- restricted attention to the set of females only.

or: reduced the sample space from the set of everyone to the set of females only,

or: conditioned on the event {females}.

We could write the above as:

$$\mathbb{P}(\text{skis} \mid \text{female}) = \frac{\#\text{female skiers in class}}{\#\text{females in class}}.$$

Conditioning is like changing the sample space: we are now working in a new sample space of females in class.

In the above example, we could replace 'skiing' with *any* attribute B. We have:

$$\mathbb{P}(\text{skis}) = \frac{\#\text{ skiers in class}}{\#\text{ class}}; \qquad \mathbb{P}(\text{skis} \mid \text{female}) = \frac{\#\text{ female skiers in class}}{\#\text{ females in class}};$$

$\mathbb{P}(B \mid \text{female})$

so:

$$\mathbb{P}(B) = \frac{\#\text{ } B\text{'s in class}}{\#\text{ in class}}$$

and:

$$\mathbb{P}(B \mid \text{female}) = \frac{\#\text{ female } B\text{'s in class}}{\#\text{ females in class}}$$

$$= \frac{\#\text{ in class who are both } B \text{ and female}}{\#\text{ in class who are female}}.$$

Likewise, we could replace 'female' with any attribute A:

$$\mathbb{P}(B \mid A) = \frac{\#\text{ class who are } B \text{ and } A}{\#\text{ class who are } A}$$

This is how we get the definition of conditional probability:

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \text{ and } A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.$$

By conditioning on event A, we have changed the sample space to the set of A's only.

Definition: Let A and B be events on the same sample space: so $A \subseteq \Omega$, $B \subseteq \Omega$.
The **conditional probability of event B, given event A**, is

$$\boxed{\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}.}$$

vertical bar

Multiplication Rule: (Immediate from above). For any events A and B,

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) = \mathbb{P}(B \mid A)\,\mathbb{P}(A) = \mathbb{P}(B \cap A)}$$

Conditioning as 'changing the sample space'

The idea that "Conditioning" = "changing the sample space" can be very helpful in understanding how to manipulate conditional probabilities.

Any 'unconditional' probability can be written as a conditional probability:

$$\mathbb{P}(B) = \mathbb{P}(B \mid \Omega)$$

Writing $\mathbb{P}(B) = \mathbb{P}(B \mid \Omega)$ just means that we are looking for the probability of event B, out of all possible outcomes in the set Ω.

In fact, the symbol $\mathbb{P}$ ***belongs*** to the set Ω: it has ***no meaning without*** Ω. To remind ourselves of this, we can write

$$\mathbb{P} = \mathbb{P}_\Omega$$

Then $\mathbb{P}(B) = \mathbb{P}(B \mid \Omega) = \mathbb{P}_\Omega(B)$.

Similarly, $\mathbb{P}(B \mid A)$ means that we are looking for the probability of event B, out of all possible outcomes in the set A.

So A is just another sample space. Thus we can manipulate conditional probabilities $\mathbb{P}(\,\cdot \mid A)$ just like any other probabilities, as long as we always stay inside the same sample space A.

The trick: Because we can think of A as just another sample space, let's write

$$\mathbb{P}(\,\cdot \mid A) = \mathbb{P}_A(\,\cdot\,)$$

Note: not standard notation! Rough work only

Then ***we can use*** $\mathbb{P}_A$ ***just like*** $\mathbb{P}$***, as long as we remember to keep the*** A ***subscript on EVERY*** $\mathbb{P}$ ***that we write.***

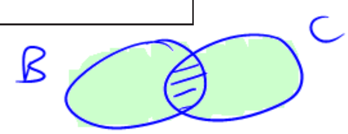
This helps us to make quite complex manipulations of conditional probabilities without thinking too hard or making mistakes. There is only one rule you need to learn to use this tool effectively:

$$\mathbb{P}_A(B \mid C) = \mathbb{P}(B \mid C \cap A). \text{ for any } A, B, C$$

(Proof: Exercise).

The rules:

$$\boxed{\begin{array}{c} \mathbb{P}(\,\cdot \mid A) = \mathbb{P}_A(\,\cdot\,) \\ \mathbb{P}_A(B \mid C) = \mathbb{P}(B \mid \underline{C \cap A}) \text{ for any } A,\ B,\ C. \end{array}}$$



Examples:

1. Probability of a union. In general,

$$\mathbb{P}_A(B \cup C) = \mathbb{P}_A(B) + \mathbb{P}_A(C) - \mathbb{P}_A(B \cap C)$$

So, $\mathbb{P}_A(B \cup C) = \mathbb{P}_A(B) + \mathbb{P}_A(C) - \mathbb{P}_A(B \cap C)$ [Rough wk]

Thus,

$$\mathbb{P}(B \cup C \mid A) = \mathbb{P}(B \mid A) + \mathbb{P}(C \mid A) - \mathbb{P}(B \cap C \mid A).$$

2. Which of the following is equal to $\mathbb{P}(B \cap C \mid A)$?

(a) $\underline{\mathbb{P}(B \mid C \cap A)}$. (c) $\mathbb{P}(B \mid C \cap A)\mathbb{P}(C \mid A)$. *

(b) $\dfrac{\mathbb{P}(B \mid C)}{\mathbb{P}(A)}$. (d) $\underline{\mathbb{P}(B \mid C)\mathbb{P}(C \mid A)}$.

Solution:

$$\left.\begin{aligned} \mathbb{P}(B \cap C \mid A) &= \mathbb{P}_A(B \cap C) \\ &= \mathbb{P}_A(B \mid C)\,\mathbb{P}_A(C) \end{aligned}\right\} \text{Rough.}$$

$$= \mathbb{P}(B \mid C \cap A)\,\mathbb{P}(C \mid A) \quad \text{Std notation}$$

$$= \text{answer (c).}$$

3. Which of the following is true?

(a) $\mathbb{P}(\overline{B}\,|\,A) = 1 - \mathbb{P}(B\,|\,A)$. (b) $\mathbb{P}(\overline{B}\,|\,A) = \mathbb{P}(B) - \mathbb{P}(B\,|\,A)$.

$\mathbb{P}(A\,|\,A)$

Solution:

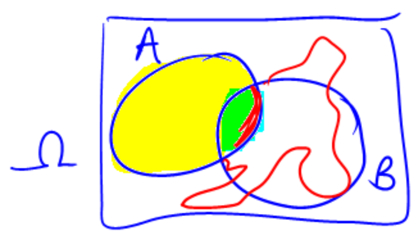
$$\mathbb{P}(\overline{B}\,|\,A) = \mathbb{P}_A(\overline{B})$$
$$= 1 - \mathbb{P}_A(B)$$
$$= 1 - \mathbb{P}(B\,|\,A) = \text{answer (a)}.$$

4. Which of the following is true?

(a) $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(A) - \mathbb{P}(B \cap A)$. (b) $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(B) - \mathbb{P}(B \cap A)$.

Solution: $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(\overline{B}\,|\,A)\,\mathbb{P}(A)$ $*$

$$= \{1 - \mathbb{P}(B\,|\,A)\}\,\mathbb{P}(A)$$
$$= \mathbb{P}(A) - \mathbb{P}(B\,|\,A)\,\mathbb{P}(A)$$
$$= \mathbb{P}(A) - \mathbb{P}(B \cap A).$$
$$= \text{answer (a)}.$$



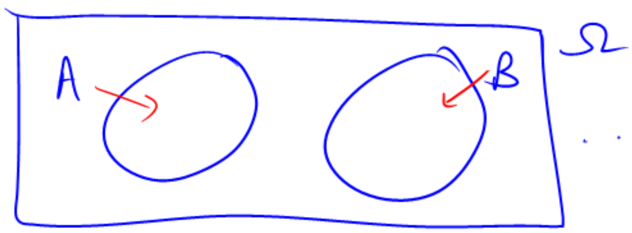
5. True or false: $\mathbb{P}(B\,|\,A) = 1 - \mathbb{P}(B\,|\,\overline{A})$? ✗

Answer: False: $\mathbb{P}(B\,|\,A) = \mathbb{P}_A(B)$. Once we've got $\mathbb{P}_A$, we are stuck with it! Doesn't make sense to add $\mathbb{P}_A$ and $\mathbb{P}_{\overline{A}}$ together.

Exercise: if we wish to express $\mathbb{P}(B\,|\,A)$ in terms of only B and $\overline{A}$, show that
$$\mathbb{P}(B\,|\,A) = \frac{\mathbb{P}(B) - \mathbb{P}(B\,|\,\overline{A})\mathbb{P}(\overline{A})}{1 - \mathbb{P}(\overline{A})}.$$
Note that this does not simplify nicely!

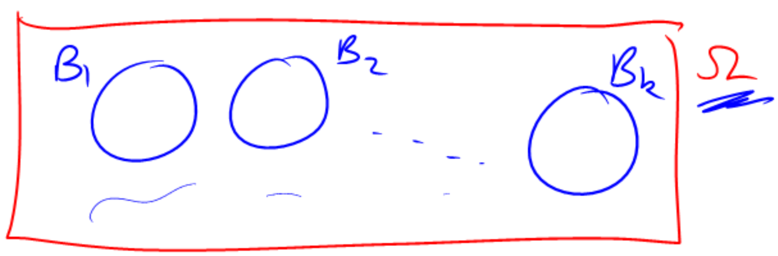
2.4 The Partition Theorem (Law of Total Probability)

Definition: Events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \phi$.
This means A and B cannot happen together. If A happens, it excludes B from happening, and vice versa.



If A and B are mutually exclusive, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
For all other A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Definition: Any number of events $B_1, B_2, \ldots, B_k$ are **mutually exclusive** if every pair of the events is mutually exclusive: ie. $B_i \cap B_j = \phi$ for all i, j with $i \neq j$.



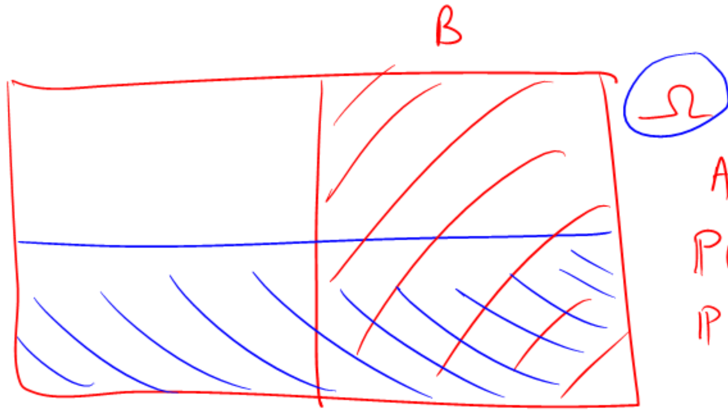
Definition: A **partition** of Ω is a collection of mutually exclusive events whose union is Ω.

That is, sets $B_1, B_2, \ldots, B_k$ form a partition of Ω if

$$B_i \cap B_j = \emptyset \text{ for all } i, j \text{ with } i \neq j,$$

$$\textbf{and} \quad \bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots \cup B_k = \Omega.$$

$B_1, \ldots, B_k$ form a partition of Ω if they have no overlap and collectively cover all possible outcomes.

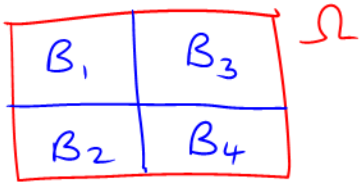


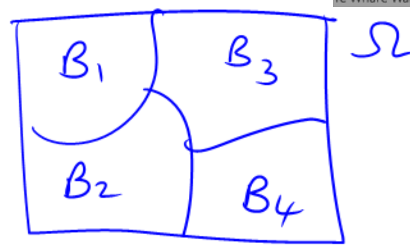
A & B indept.

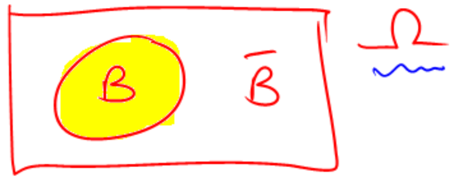
$\mathbb{P}(A) = 0.5$

$\mathbb{P}(A \mid B) = \mathbb{P}(A)$

Examples:

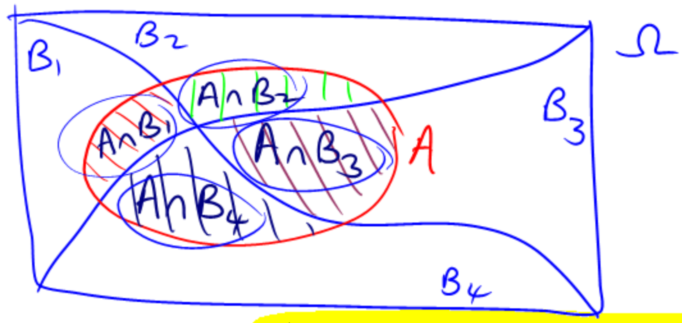






Partitioning an event A

Any set A can be partitioned: it doesn't have to be Ω.
In particular, if $B_1, \ldots, B_k$ form a partition of Ω, then $(A \cap B_1), \ldots, (A \cap B_k)$ form a partition of A.



Theorem 2.4: The Partition Theorem (Law of Total Probability)

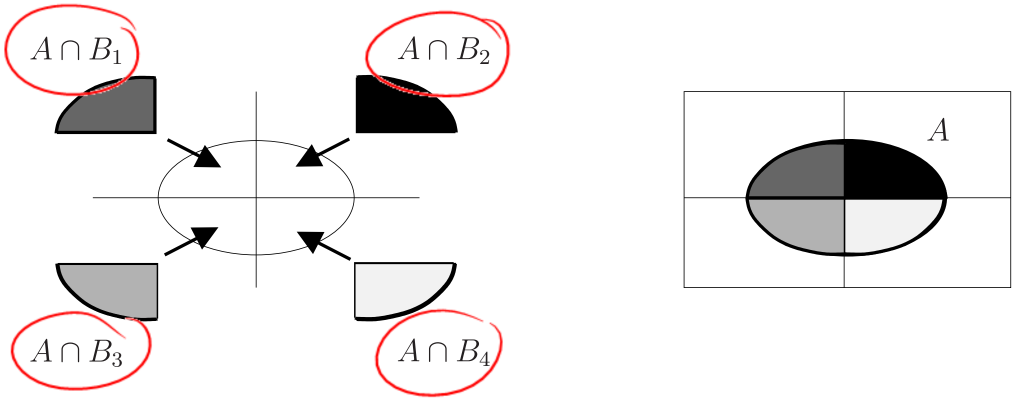
Let $B_1, B_2, \ldots, B_m$ form a partition of Ω. Then for any event A,

$$\boxed{\mathbb{P}(A) = \underbrace{\sum_{i=1}^{m} \mathbb{P}(A \cap B_i)}_{\text{understanding}} = \underbrace{\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}_{\text{using}}.}$$

Both formulations of the Partition Theorem are very widely used, but especially the conditional formulation $\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$.

Intuition behind the Partition Theorem:

The Partition Theorem is easy to understand because it simply states that "the whole is the sum of its parts."



$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) + \mathbb{P}(A \cap B_4).$$

2.5 Bayes' Theorem: inverting conditional probabilities

Bayes' Theorem allows us to "invert" a conditional statement, ie. to express $\mathbb{P}(B \mid A)$ in terms of $\mathbb{P}(A \mid B)$.

Theorem 2.5: Bayes' Theorem

For any events A and B:

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

Proof:

$$\begin{aligned} \mathbb{P}(B \cap A) &= \mathbb{P}(A \cap B) \\ \mathbb{P}(B \mid A)\mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) \qquad \text{(multiplication rule)} \\ \therefore \quad \mathbb{P}(B \mid A) &= \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}. \qquad \square \end{aligned}$$

Extension of Bayes' Theorem

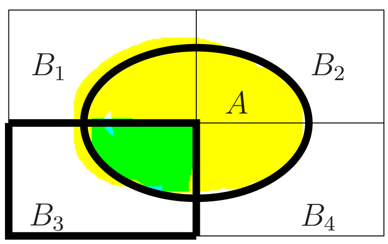
Suppose that $B_1, B_2, \ldots, B_m$ form a partition of Ω. By the Partition Theorem,

$$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)$$

Thus, for ***any single partition member*** B_j, put $B = B_j$ in Bayes' Theorem to obtain:

$$\boxed{\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_j)\,\mathbb{P}(B_j)}{\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)}}$$

e.g. $j = 3$



"given"
"within"

$\mathbb{P}(B_3 \mid A) = ?$

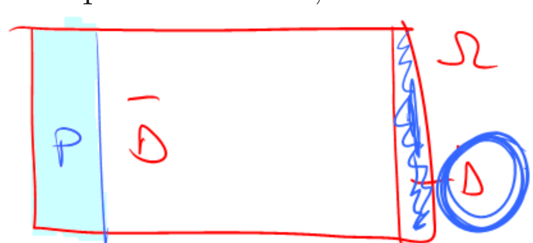
the probability of event B_3 WITHIN the event A.

Special case: $m = 2$

Given any event B, the events B and $\overline{B}$ form a partition of Ω. Thus:

$$\boxed{\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\,\mathbb{P}(B)}{\mathbb{P}(A \mid B)\,\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\,\mathbb{P}(\overline{B})}}$$

Example: In screening for a certain disease, the probability that a healthy person wrongly gets a positive result is 0.05. The probability that a diseased person wrongly gets a negative result is 0.002. The overall rate of the disease in the population being screened is 1%. If my test gives a positive result, what is the probability I actually have the disease?



1) Define Events:

$\to \Omega = \{\text{people taking test}\}$.

$D = \{\text{have disease}\}$ $\quad \bar{D} = \{\text{don't have disease}\}$

$P = \{\text{positive test}\}$ $\quad N = \bar{P} = \{\text{negative test}\}$

2) Information Given:

- False positive rate is 0.05 $\Rightarrow$ $\mathbb{P}(P \mid \bar{D}) = 0.05$
- False negative rate is 0.002 $\Rightarrow$ $\mathbb{P}(N \mid D) = 0.002$
- Disease rate is 1% $\Rightarrow$ $\mathbb{P}(D) = 0.01$.

3) Looking for $\mathbb{P}(D \mid P)$:

Bayes Thm $\Rightarrow$

$$\mathbb{P}(D \mid P) = \frac{\mathbb{P}(P \mid D)\,\mathbb{P}(D)}{\mathbb{P}(P)}$$

$$= \frac{\{1 - \mathbb{P}(\bar{P} \mid D)\}\, 0.01}{\mathbb{P}(P)}$$

$$= \frac{\{1 - \mathbb{P}(N \mid D)\} * 0.01}{\mathbb{P}(P \mid D)\mathbb{P}(D) + \mathbb{P}(P \mid \bar{D})\mathbb{P}(\bar{D})}$$

$$= \frac{\{1 - 0.002\} * 0.01}{\{1 - 0.002\} * 0.01 + 0.05 * 0.99}$$

$$= 0.168$$

Given a positive test, my chance of having the disease is only 16.8%.

1st order process

Probability

The Big Idea.

0.2

2.6 First-Step Analysis for calculating probabilities in a process

In a stochastic process, what happens at the next step depends upon the current state of the process. We are often interested to know the probability of *eventually* reaching some particular state, given our current position.

Now

End 1

Throughout this course, we will tackle this sort of problem using a technique called

End 2

The idea is to consider all possible first steps away from the current state. We derive a system of equations that specify the probability of the eventual outcome given each of the possible first steps. We then try to solve these equations for the probability of interest.

First-Step Analysis depends upon ***conditional probability and the Partition Theorem.*** Let $S_1, \ldots, S_k$ be the k possible first steps we can take away from our current state. We wish to find the probability that event E happens eventually. First-Step Analysis calculates $\mathbb{P}(E)$ as follows:

Venus wins

$$\mathbb{P}_N(E) = \mathbb{P}_N(E \mid S_1)\mathbb{P}_N(S_1) + \ldots + \mathbb{P}_N(E \mid S_k)\mathbb{P}_N(S_k)$$

N = "Now" = "Current state" $\mathbb{P}_N(\cdot) = \mathbb{P}(\cdot \mid N)$ shorthand.

Here, $\mathbb{P}(S_1), \ldots, \mathbb{P}(S_k)$ give the probabilities of taking the different first steps $1, 2, \ldots, k$.

Example 1: Tennis game at Deuce.

Venus and Serena are playing tennis, and have reached the score Deuce (40-40). (*Deuce* comes from the French word *Deux* for 'two', meaning that each player needs to win two consecutive points to win the game.)

For each point, let:

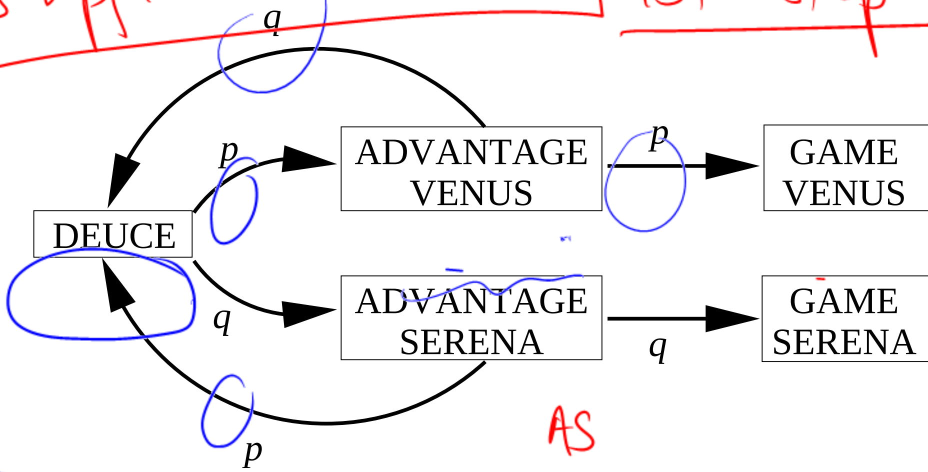
$$p = \mathbb{P}(\text{Venus wins point}), \qquad q = 1 - p = \mathbb{P}(\text{Serena wins point}).$$

Assume that all points are independent.

Let v be the probability that Venus wins the game eventually, starting from Deuce. Find v. $v = \mathbb{P}(\text{VW eventually} \mid \text{start at Deuce})$.

Understand partioning by this approach

1st-step analysis



Use first-step analysis. The possible first steps, starting from Deuce, are:

(1) Venus wins next point (probability p) : State AV

(2) Serena wins next point (" q) : State AS

Thus: $v = \mathbb{P}(VW \mid \text{Deuce}) = \mathbb{P}_D(VW)$

[partition theorem] $= \mathbb{P}_D(VW \mid AV)\,\mathbb{P}_D(AV) + \mathbb{P}_D(VW \mid AS)\,\mathbb{P}_D(AS)$

$\mathbb{P}(AV \mid D)$ in standard notation: $\mathbb{P}$(move to AV in one step, starting at D)

$= \mathbb{P}(VW \mid AV)\,p + \mathbb{P}(VW \mid AS)\,q \quad (*)$

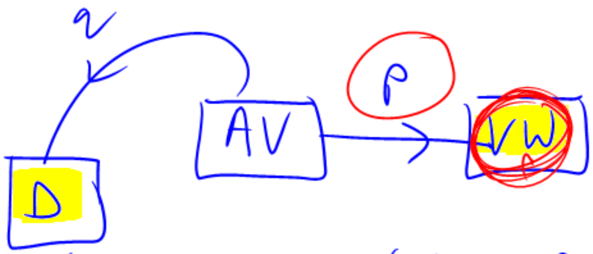
can drop D subscript now, because it only matters that we're in state AV, or AS, it doesn't matter how we got there (from D).

$\mathbb{P}(VW \mid AV_2 \cap D_1) = \mathbb{P}(VW \mid AV)$

Now we need to find $\mathbb{P}(VW \mid AV)$ and $\mathbb{P}(VW \mid AS)$.

Use first-step analysis again.

$$\mathbb{P}(VW \mid AV) = \mathbb{P}_{AV}(VW)$$



$$= \underbrace{\mathbb{P}_{AV}(VW \mid VW)}\ \underbrace{\mathbb{P}_{AV}(VW)}_{\mathbb{P}(VW \text{ start at } AV)} + \underbrace{\mathbb{P}_{AV}(VW \mid D)}_{=v}\ \underbrace{\mathbb{P}_{AV}(D)}_{q}$$

partition Thm.

$$= 1 * p + v * q$$

↖ back to the beginning!

$$\therefore \mathbb{P}(VW \mid AV) = p + qv. \quad (a)$$

Similarly, $\mathbb{P}(VW \mid AS) = \mathbb{P}_{AS}(VW)$

$$= \underbrace{\mathbb{P}_{AS}(VW \mid SW)}\ \mathbb{P}_{AS}(SW) + \underbrace{\mathbb{P}_{AS}(VW \mid D)}\ \underbrace{\mathbb{P}_{AS}(D)}$$

$$\mathbb{P}(VW \mid AS) = 0 * q + v * p \quad (b)$$

Substitute (a) & (b) into (*):

$$v = \overset{(a)}{(p + qv)}\, p + \overset{(b)}{(pv)}\, q$$

$$v = p^2 + 2pqv$$

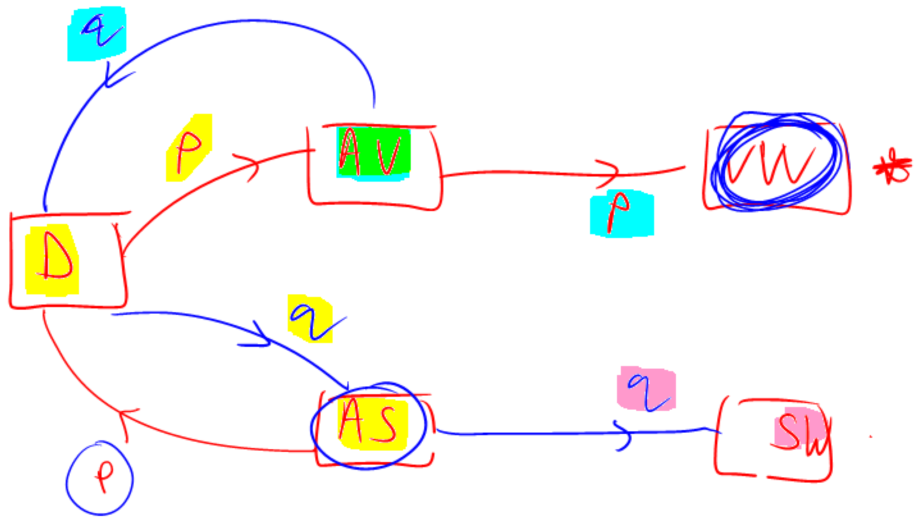
(1 eqn, 1 unknown
solve for v)

$$v(1 - 2pq) = p^2$$

$$\therefore \quad v = \frac{p^2}{1 - 2pq}$$

Quicker Method (Use this one!)

Read the first-step analysis straight off the diagram.



① Define notation: $V_D = \mathbb{P}(\text{Venus wins eventually} \mid \text{start at state } D)$

$V_{AV} = \mathbb{P}(\quad " \quad " \quad " \quad \mid \text{start at } AV)$

$V_{AS} = \mathbb{P}(\quad " \quad " \quad " \quad \mid \text{start at } AS)$

$$\therefore \quad V_D = p\,V_{AV} + q\,V_{AS}$$

$$V_{AV} = p * 1 + q\,V_D$$

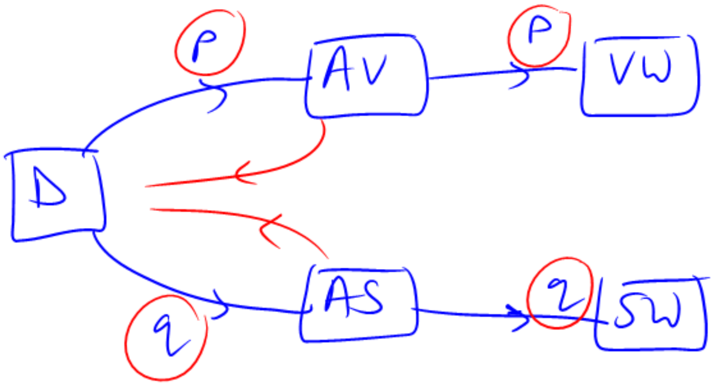
$$V_{AS} = q * 0 + p\,V_D$$

Use this approach

$$\therefore \quad V_D = p(p + q V_D) + q p V_D \Rightarrow V_D(1 - 2pq) = p^2$$

$$\therefore \quad V_D = \frac{p^2}{1 - 2pq}$$

Same answer for $V = V_D = \mathbb{P}(\text{Venus wins} \mid \text{start at } D)$.



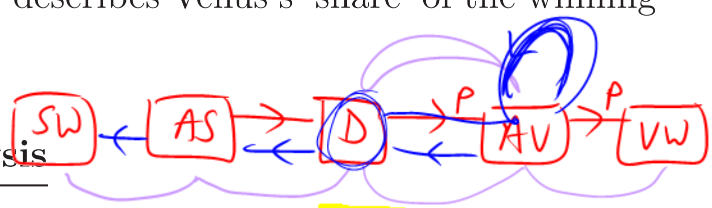
Note: Because $p + q = 1$, we have: $1 = 1^2 = (p+q)^2 = p^2 + q^2 + 2pq$

$$\therefore \quad 1 - 2pq = p^2 + q^2$$

So the final probability that Venus wins the game is:

$$v = \frac{p^2}{1-2pq} = \frac{p^2}{p^2+q^2} \qquad v = \mathbb{P}(\text{venus wins} \mid \text{somebody wins})$$

Note how this result makes intuitive sense. For the game to finish from Deuce, either Venus has to win two points in a row (probability p^2), or Serena does (probability q^2). The ratio $p^2/(p^2 + q^2)$ describes Venus's 'share' of the winning probability.



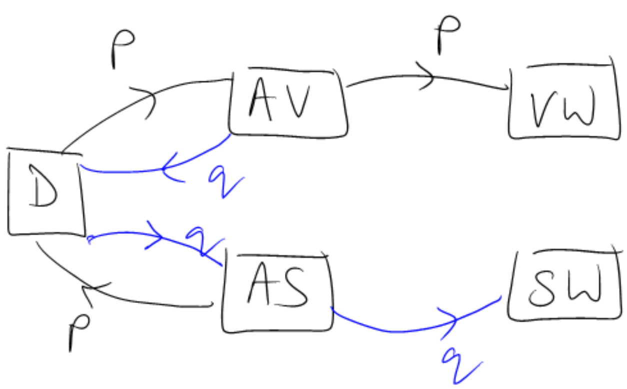
Alternative approach: two-step analysis

The diagram above is a simple one, and we can see that the *only* way of winning from Deuce is to take two steps away from Deuce. In situations like these, it is worth considering whether a two-step approach is possible.

For a two-step approach, we look at the possible pairs of steps we can take out of the state Deuce.

Pair of steps (partition)	Probability of this pair of steps	Outcome	$\mathbb{P}(VW \mid \text{outcome})$
$D \to AV \to VW$	p^2	Venus Wins	1
$D \to AS \to SW$	q^2	Serena wins	0
$D \to AV \to D$	pq	Back to start	v
$D \to AS \to D$	pq	Back to start	v
	$\mathbb{P}(\text{partition})$		$\mathbb{P}(\text{outcome} \mid \text{partition})$

Why does a 2-step analysis work for this diagram?



Possible pairs of steps:

D → AV → VW

D → AV → D

D → AS → D

D → AS → SW

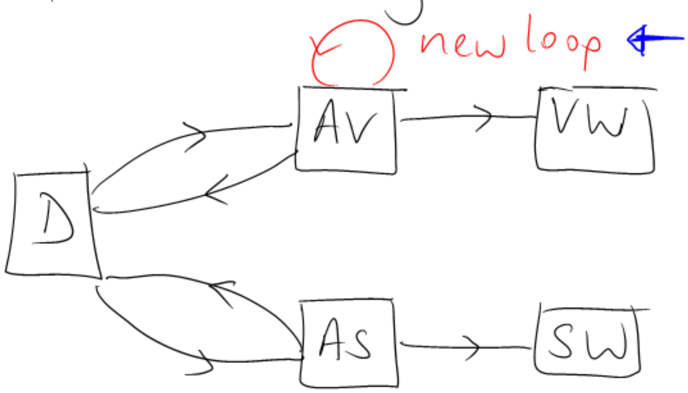
If we partition over possible pairs of steps out of Dence, we get a simplified system:

Outcome of a pair of steps	$\mathbb{P}$(steps)	$\mathbb{P}$(VW | steps)
D	$2pq$	v
VW	p^2	1
SW	q^2	0

all known except v

If our diagram was just a little more complicated, we wouldn't gain any simplicity from a 2-step analysis: it would just make things more complicated.

E.g.



This is unknown so we have to do another partition; we haven't gained anything from a 2-step analysis

Possible pairs of steps	Outcome	P(VW | steps)
D → AS → D or D → AV → D	D	v
D → AV → AV	AV	unknown
D → AV → VW	VW	1
D → AS → SW	SW	0

Our approach to finding $v = \mathbb{P}(\text{Venus wins})$ can be summarized as:

$$\mathbb{P}(\text{Venus wins}) = v = \sum_{\substack{\text{partition}\\\text{members}}} \mathbb{P}(VW \mid \text{partition member})\,\mathbb{P}(\text{partition member})$$

$$v = 1 * p^2 + 0 * q^2 + v * pq + v * pq$$

$$\therefore\ v = p^2 + 2pqv$$

$$v(1-2pq) = p^2$$

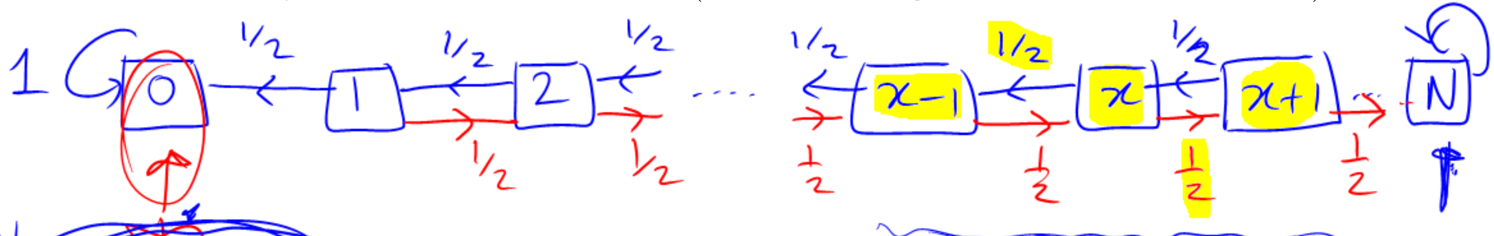
$$\therefore\ v = \frac{p^2}{1-2pq} \quad \text{as before.}$$

☺

Example 2: **Gambler's Ruin.**
This is a famous problem in probability. A gambler starts with $\$x$. She tosses a fair coin repeatedly.

If she gets a Head, she wins $1. If she gets a Tail, she loses $1.

The coin tossing is repeated until the gambler has either $\$0$ or $\$N$. What is the probability of the Gambler's Ruin (i.e. that the gambler ends up with $0)?



$\mathbb{P}(\text{Ruin})$?

* Define Notation: Let $p_x = \mathbb{P}(\text{eventual ruin} \mid \text{start at state } x)$

goal stays same

index start state

start state changes

$$p_x = \frac{1}{2} p_{x+1} + \frac{1}{2} p_{x-1} \qquad (*)$$

True for $x = 1, 2, \ldots, N-1$ with boundary conditions

$$p_0 = 1 \qquad p_N = 0$$

Possible first steps out of state x	Prob of this step	Outcome of this step	$\mathbb{P}(\text{ruin} \mid \text{outcome})$
Head	$\frac{1}{2}$	go to state $x+1$	p_{x+1}
Tail	$\frac{1}{2}$	go to state $x-1$	p_{x-1}

Partition Theorem

$$\mathbb{P}(\text{ruin} \mid \text{start at } x) = \mathbb{P}_x(R) \qquad R = \text{ruin}$$

$$= \mathbb{P}_x(R \mid \text{Head})\,\mathbb{P}_x(\text{Head}) + \mathbb{P}_x(R \mid \text{Tail})\,\mathbb{P}_x(\text{Tail})$$

$$= p_{x+1} * \frac{1}{2} + p_{x-1} * \frac{1}{2}$$

Solution of difference equation (*):

We have $p_x = \frac{1}{2} p_{x+1} + \frac{1}{2} p_{x-1}$

Rearrange: $\cancel{\frac{1}{2}} p_x + \cancel{\frac{1}{2}} p_x = \cancel{\frac{1}{2}} p_{x+1} + \cancel{\frac{1}{2}} p_{x-1}$

$$p_{x-1} - p_x = p_x - p_{x+1}$$

Boundaries: $p_0 = 1$, $p_N = 0$.

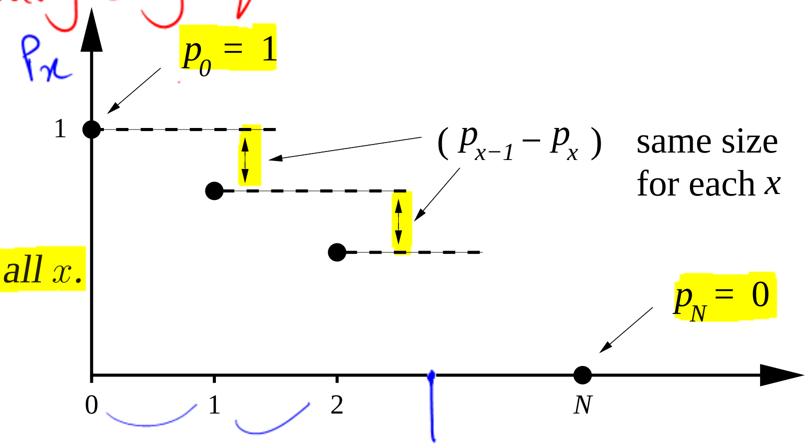
There are several ways of solving this equation. unusually easy eqn to solve!

1. By inspection

There are N steps to go down from $p_0 = 1$ to $p_N = 0$.
Each step is the same size, because

enables common sense solution → $(p_{x-1} - p_x) = (p_x - p_{x+1})$ for all x.
So each step has size $1/N$,
$\Rightarrow \quad p_0 = 1,\ p_1 = 1 - 1/N,$
$\quad p_2 = 1 - 2/N$, etc.

So $\boxed{p_x = 1 - \dfrac{x}{N}.}$



2. Theory of linear 2nd order difference equations *General Method.

Theory tells us that the general solution of $(\star)$ is $p_x = A + Bx$ for some constants A, B.

$p_0 = A + B \times 0 = 1$
$p_N = A + BN = 0 = 1 + BN = 0$

Boundary conditions:

$$p_0 = A + B \times 0 = 1 \quad \Rightarrow \quad A = 1$$
$$p_N = A + B \times N = 1 + BN = 0 \quad \Rightarrow \quad B = \frac{-1}{N}.$$

So $\boxed{p_x = A + B\,x = 1 - \dfrac{x}{N}}$ as before.

3. Repeated substitution: rearrange $(\star)$ to give

$$p_{x+1} = 2p_x - p_{x-1}$$

Start: $p_0 = 1$ (known)

$x=1$ $\quad p_2 = 2p_1 - 1$

$x=2$ $\quad p_3 = 2p_2 - p_1 = 2(2p_1 - 1) - p_1 = 3p_1 - 2$

etc etc In general, $p_x = x p_1 - (x-1)$

At endpoint, $p_N = N p_1 - (N-1) = 0$ known

$$\Rightarrow \quad p_1 = 1 - \frac{1}{N}$$

$\therefore \quad p_x = 1 - \frac{x}{N}$ in general, as before (Exercise).

$\mathbb{P}(A) = \frac{1}{2}$

2.7 Independence

Definition: Events A and B are **statistically independent** if and only if

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

This implies that A and B are statistically independent if and only if

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

Note: If events are *physically* independent, they will also be statistically indept.

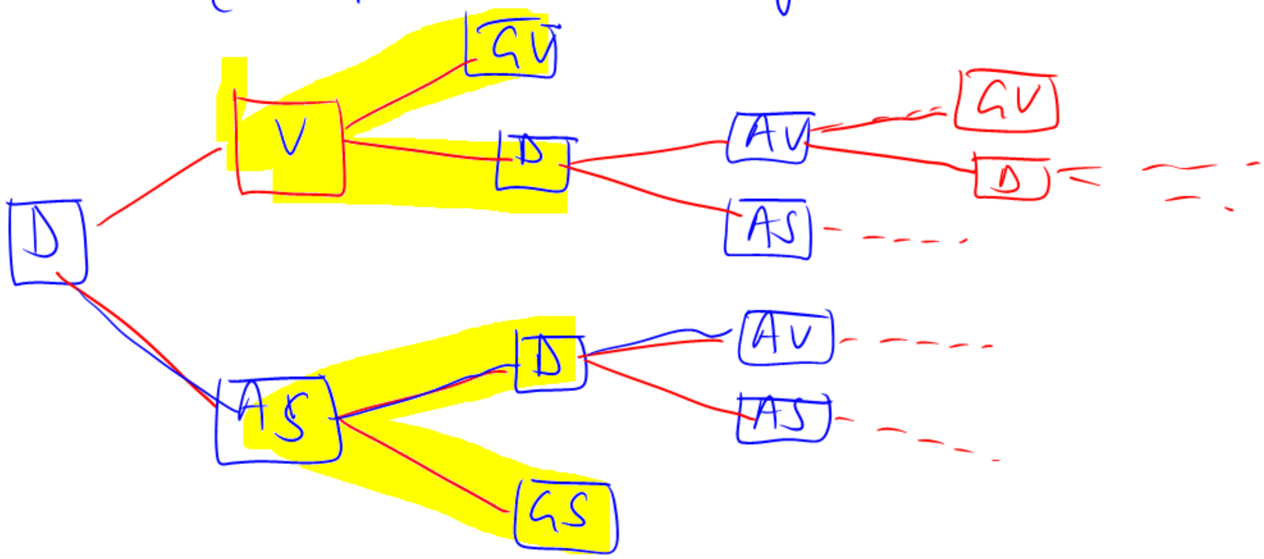
For interest: more than two events

Definition: For more than two events, $A_1, A_2, \ldots, A_n$, we say that $A_1, A_2, \ldots, A_n$ are **mutually independent** if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \quad \textit{for ALL finite subsets } J \subseteq \{1, 2, \ldots, n\}.$$

Tennis game

$\Omega = \{$all possible routes from Deuce to end$\}$



$\Omega = \{$ all routes from D to end $\}$

e.g. blue highlight = 1 outcome in my sample space.

Event $B_1 = \{$outcome passes thru' AV @ time 1$\}$

$B_2 = \{$ " " " AS @ " 1$\}$

B_1 & B_2 partition Ω,

So First Step Analysis works by saying:

$$\mathbb{P}_D(\text{VW eventually}) = \mathbb{P}_D(VW \mid B_1)\,\mathbb{P}_D(B_1) + \mathbb{P}_D(VW \mid B_2)\,\mathbb{P}_D(B_2)$$

Partition Thm.

Example: events A_1, A_2, A_3, A_4 are mutually independent if

i) $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all i, j with $i \neq j$; AND *

ii) $\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i)\mathbb{P}(A_j)\mathbb{P}(A_k)$ for all i, j, k that are all different; AND

iii) $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\mathbb{P}(A_4)$.

Note: For mutual independence, it is ***not*** enough to check that $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$. Pairwise independence does not imply mutual independence.

Definition: Events, $A_1, A_2, \ldots, A_n$ are **pairwise independent** if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \textit{for all } i \neq j.$$

Pairwise independence does NOT imply mutual independence. See example in Stats 210 notes.

2.8 The Continuity Theorem

The Continuity Theorem states that probability is a *continuous set function*:

Theorem 2.8: The Continuity Theorem

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P_n$$

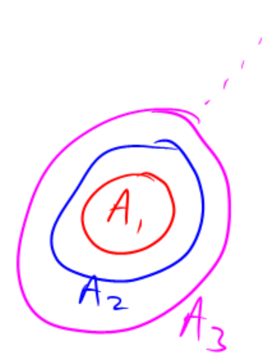
a) Let $A_1, A_2, \ldots$ be an *increasing sequence of events*: i.e.

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq A_{n+1} \subseteq \ldots .$$

Then

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

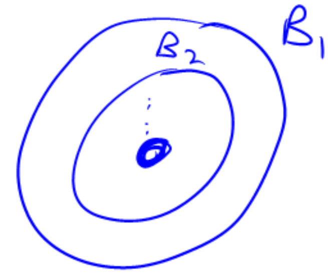
Note: because $A_1 \subseteq A_2 \subseteq \ldots$, we have: $\displaystyle\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$



Sets $A_1 \subseteq A_2 \subseteq A_3 \subseteq A_4 \subseteq \ldots$ *

Real #s $P_1 \leq P_2 \leq P_3 \leq P_4 \ldots$ *

$\mathbb{P}(A_1)$ $\mathbb{P}(A_2)$ $\ldots$



b) Let $B_1, B_2, \ldots$ be a *decreasing sequence of events*: i.e.

$$B_1 \supseteq B_2 \supseteq \ldots \supseteq B_n \supseteq B_{n+1} \supseteq \ldots .$$

Then

$$\mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Note: because $B_1 \supseteq B_2 \supseteq \ldots$, we have: $\displaystyle\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n$.

Proof (a) only: for (b), take complements and use (a).

Define $C_1 = A_1$, and $C_i = A_i \backslash A_{i-1}$ for $i = 2, 3, \ldots$. Then $C_1, C_2, \ldots$ are mutually exclusive, and $\bigcup_{i=1}^n C_i = \bigcup_{i=1}^n A_i$, and likewise, $\bigcup_{i=1}^\infty C_i = \bigcup_{i=1}^\infty A_i$.

Thus

$$\mathbb{P}(\lim_{n\to\infty} A_n) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(C_i) \quad (C_i \text{ mutually exclusive})$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(C_i)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} C_i\right)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n). \quad \square$$

2.9 Random Variables

X : outcomes $\mapsto$ real numbers

Definition: A **random variable**, X, is defined as *a function from the sample space to the real numbers:* $X : \Omega \to \mathbb{R}$.

A random variable therefore assigns a real number to every possible outcome of a random experiment.

A random variable is essentially a rule or mechanism for generating random real numbers.

The Distribution Function

Definition: The **cumulative distribution function** of a random variable X is given by

$$F_X(x) = \mathbb{P}(X \leq x).$$

$F_X(x)$ is often referred to as simply the **distribution function**.

Properties of the distribution function

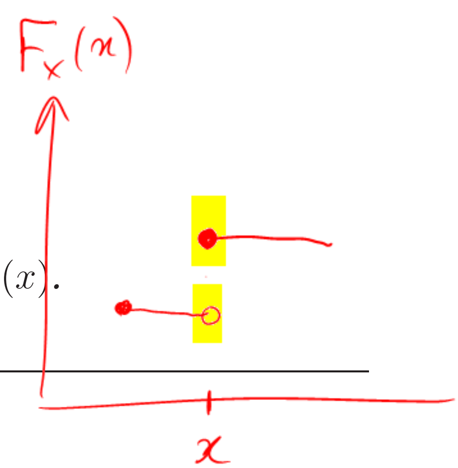
1) $F_X(-\infty) = \mathbb{P}(X \leq -\infty) = 0.$

   $F_X(+\infty) = \mathbb{P}(X \leq \infty) = 1.$

2) $F_X(x)$ is a non-decreasing function of x:
   if $x_1 < x_2$*, then* $F_X(x_1) \leq F_X(x_2)$.

3) If $b > a$, then $\boxed{\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).}$

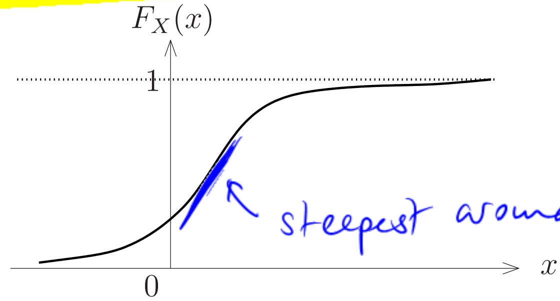
4) F_X is right-continuous: i.e. $\lim_{h \downarrow 0} F_X(x + h) = F_X(x)$.



2.10 Continuous Random Variables

Definition: The random variable X is **continuous** if *the distribution function* $F_X(x)$ *is a* *continuous function.*

In practice, this means that a continuous random variable *takes values in a continuous subset of* $\mathbb{R}$*: e.g.* $X : \Omega \to [0, 1]$ *or* $X : \Omega \to [0, \infty)$.



steepest around here, highest pdf here.

Probability Density Function for continuous random variables

Definition: Let X be a continuous random variable with continuous distribution function $F_X(x)$. The **probability density function (p.d.f.)** of X is defined as

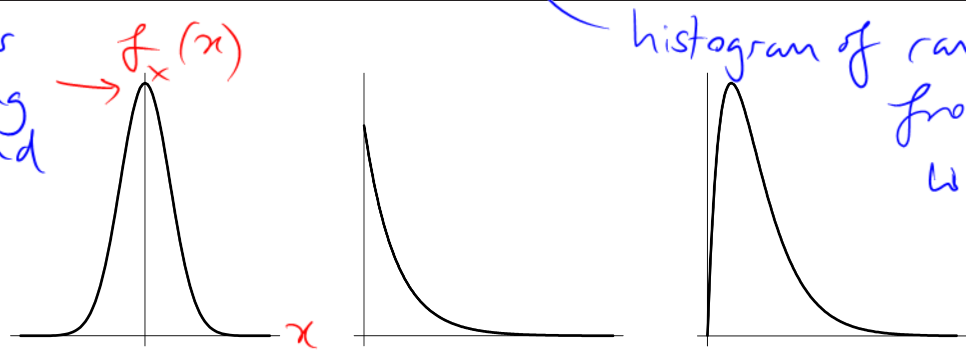
$$f_X(x) = F'_X(x) = \frac{d}{dx}(F_X(x))$$

$f_X(x)$: how fast is probability accumulating close to x?

> The pdf, $f_X(x)$, gives the *shape* of the distribution of X.

histogram of random values from the distn would look like $f_X(x)$.

probability is accumulating fastest around here (the peak)



Normal distribution   Exponential distribution   Gamma distribution

By the Fundamental Theorem of Calculus, the distribution function $F_X(x)$ can be written in terms of the probability density function, $f_X(x)$, as follows:

$$f_X(x) = \frac{dF}{dx}$$

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(u)\,du}$$

0.0001
0.000100001

$X = 4.53$? No 4.530000001

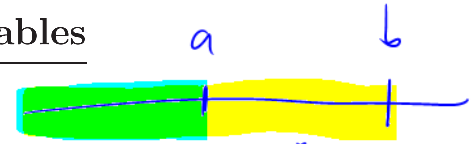
Endpoints of intervals

For continuous random variables, every point x has $\mathbb{P}(X = x) = 0$. This means that the endpoints of intervals are not important for continuous random variables.

Thus, $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$.

This is ***only*** true for ***continuous*** random variables.

Calculating probabilities for continuous random variables

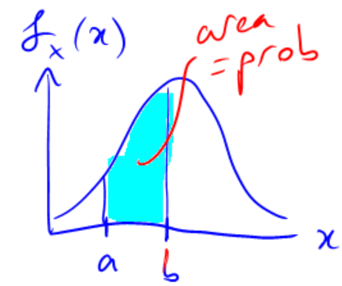


To calculate $\mathbb{P}(a \leq X \leq b)$, use ***either***

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$$

or

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$



Example: Let X be a continuous random variable with p.d.f.

$$f_X(x) = \begin{cases} 2x^{-2} & \text{for } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function, $F_X(x)$.

(b) Find $\mathbb{P}(X \leq 1.5)$.

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

$x : -\infty \to x$
nonsense

a) $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$

$= \int_1^{x} 2u^{-2}\,du$

NOT $\int_{-\infty}^{x} 2u^{-2}\,du$

$$= \int_1^x 2u^{-2}\,du = \left[2\frac{u^{-1}}{-1} \right]_1^x$$

$$= 2 - \frac{2}{x} \quad \text{for } 1 < x < 2 .$$

$$\therefore F_X(x) = \begin{cases} 0 & \text{for } x \le 1 \\ 2 - \frac{2}{x} & \text{for } 1 < x < 2 \\ 1 & \text{for } x \ge 2 \end{cases}$$

b) $\mathbb{P}(X \le 1.5) = F_X(1.5) = 2 - \frac{2}{1.5} = \frac{2}{3}$.

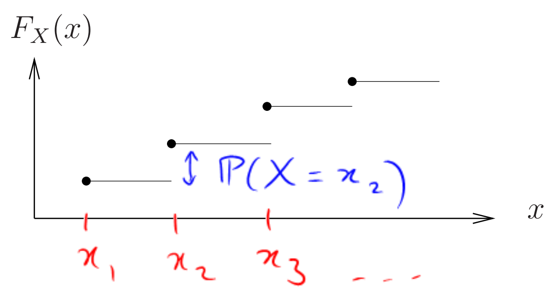
2.11 Discrete Random Variables

$\mathbb{R}$ not countable! 0.465

Definition: The random variable X is **discrete** if X *takes values in a finite or countable subset of* $\mathbb{R}$*: thus,* $X : \Omega \to \{x_1, x_2, \ldots\}$.

$\mathbb{N}$ $\mathbb{Z}$

When X is a discrete random variable, the distribution function $F_X(x)$ is a *step function.*



Sudden accumulation of probability at $x_1, x_2, \ldots$

Probability function

Definition: Let X be a discrete random variable with distribution function $F_X(x)$. The **probability function** of X is defined as

$$f_X(x) = \mathbb{P}(X = x).$$

721 : Thur 21st Oct 5pm
Test

Help class : 5pm Thur 26th Aug.

Endpoints of intervals

For discrete random variables, *individual points can have* $\mathbb{P}(X = x) > 0$.

This means that *the endpoints of intervals ARE important for discrete random variables.*

For example, if X takes values $0, 1, 2, \ldots$, and a, b are integers with $b > a$, then

$\mathbb{P}(a \le X \le b) = \mathbb{P}(a-1 < X \le b) = \mathbb{P}(a \le X < b+1) = \mathbb{P}(a-1 < X < b+1)$.

Calculating probabilities for discrete random variables

To calculate $\mathbb{P}(X \in A)$ for any countable set A, use

$$\mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x).$$

Partition Theorem for probabilities of discrete random variables

Recall the Partition Theorem: for any event A, and for events $B_1, B_2, \ldots$ that form a ***partition*** of Ω,

$$\mathbb{P}(A) = \sum_y \mathbb{P}(A \mid B_y)\,\mathbb{P}(B_y) \quad \ast$$

We can use the Partition Theorem to find probabilities for random variables. Let X and Y be **discrete** random variables.

- Define event A as $A = \{X = x\}$ the event that X takes value x.
- Define event B_y as $B_y = \{Y = y\}$ for $y = 0, 1, 2, \ldots$
- Partition Theorem $\Rightarrow$

$$\ast \quad \mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x \mid Y = y)\,\mathbb{P}(Y = y).$$

2.12 Independent Random Variables

Random variables X and Y are independent if they have no effect on each other. This means that the probability that they both take specified values simultaneously is the product of the individual probabilities.

Definition: Let X and Y be random variables. The **joint distribution function** of X and Y is given by

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

Definition: Let X and Y be any random variables (continuous or discrete). X and Y are **independent** if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \text{ for ALL } x, y \in \mathbb{R}.$$

If X and Y are ***discrete***, they are independent if and only if their joint probability function is the product of their individual probability functions:

Discrete X, Y are independent

$$\iff \mathbb{P}(X = x \text{ AND } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \text{ for ALL } x, y \in \mathbb{R}.$$

$$\iff f_{X,Y}(x, y) = f_X(x) f_Y(y) \text{ for ALL } x, y.$$