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;

  biggest idea in the course
- First-Step Analysis for finding the probability that a process reaches some state, by conditioning on the outcome of the first step;

  Next Mon
- 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.* $A \subseteq \Omega$

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 $A_1, A_2, \ldots, A_n$,

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

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

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 the number of people in the class who ski, and divide by the total number of people in the class.

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

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 the total # females in the class.

$$\mathbb{P}(\text{female skis}) = \frac{\#\text{ female skiers in class}}{\text{total }\#\text{ females in class}}. \qquad \boxed{\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}}$$

new sample space: $A = \{\text{females}\}$

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 everybody to the set of females

or: conditioned on the event {female}.

We could write the above as:

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

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:

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

so:

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

and:

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

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

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

$$\mathbb{P}(B \mid A) = \frac{(\#\text{ in class who are both } B \text{ AND } A)/(\#\text{ in class})}{(\#\text{ in class who are } A)\ /\ (\#\text{ in class})}.$$

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

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

Think of $\mathbb{P}(B \mid A)$ as "probability of B, from WITHIN the A's".

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

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

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

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

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

$\boxed{\text{NOTE } \underline{\text{not}} \text{ standard notation!}}$

Rough working 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:

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

(Proof: Exercise).

The rules:

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

Examples:

1. Probability of a union. In general,

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

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

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

Easy!

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

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

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

(c) $\mathbb{P}(B \mid C \cap A)\mathbb{P}(C \mid A)$. ← c

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

Solution:

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

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

Answer c.

$\mathbb{P}_A \quad \mathbb{P}_\Omega \quad \mathbb{P}_A$ yuk!!

3. Which of the following is true?

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

Solution:

$$\mathbb{P}(\bar{B}\,|\,A) = \mathbb{P}_A(\bar{B})$$
$$= 1 - \mathbb{P}_A(B)$$

rough wk.

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

☺ Answer a.

4. Which of the following is true?

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

LHS RHS

Solution: $\text{LHS} = \mathbb{P}_\Omega(\bar{B}\cap A) = \mathbb{P}(\bar{B}\,|\,A)\,\mathbb{P}(A)$

rough wkg

$$= \mathbb{P}_A(\bar{B})\;\mathbb{P}(A)$$
$$= \{1 - \mathbb{P}_A(B)\}\;\mathbb{P}(A)$$
$$= \mathbb{P}(A) - \mathbb{P}(B\,|\,A)\,\mathbb{P}(A)$$
$$= \mathbb{P}(A) - \mathbb{P}(B\cap A)$$

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

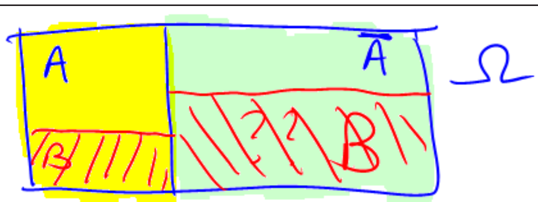
Answer: False: $\mathbb{P}(B\,|\,A) = \mathbb{P}_A(B)$. There is no easy way of converting from $\mathbb{P}_A$ to $\mathbb{P}_{\bar{A}}$. Probabilities on one sample space don't tell us anything about probabilities on a different sample space.

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!

$\mathbb{P}(B\,|\,A)$ = red as a fraction of yellow

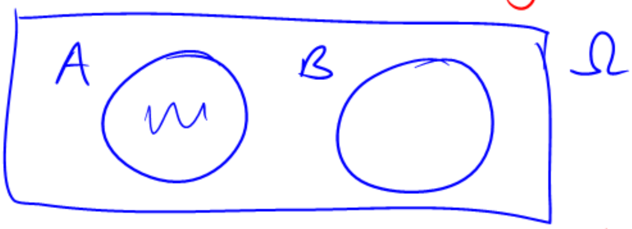


Basis for all of Stochastic Process Theory.

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 can not happen together. If A happens it excludes B from happening, and vice versa.

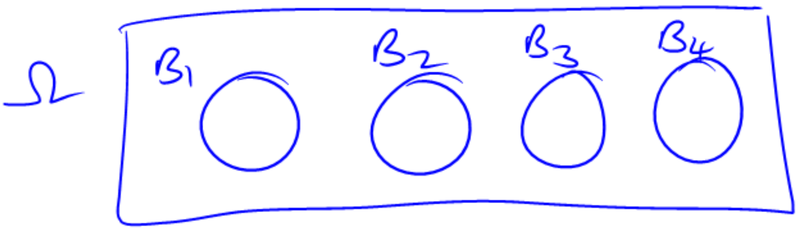


mutually exclusive events are NOT independent; strong dependence between them.

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

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

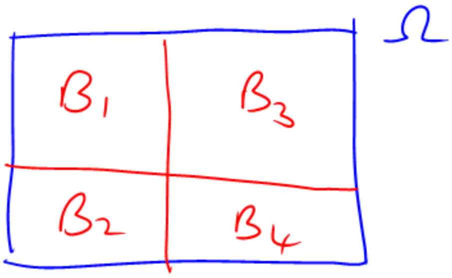
(1) no overlap between B's

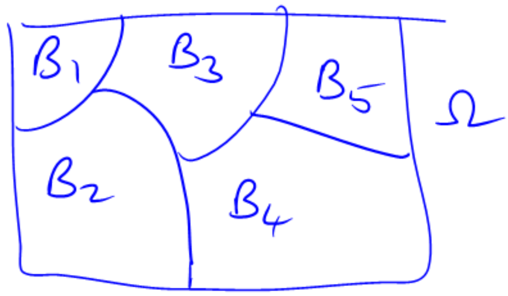
$$\textbf{and} \quad \bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots \cup B_k = \Omega.$$

(2) B's cover all outcomes in Ω.

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

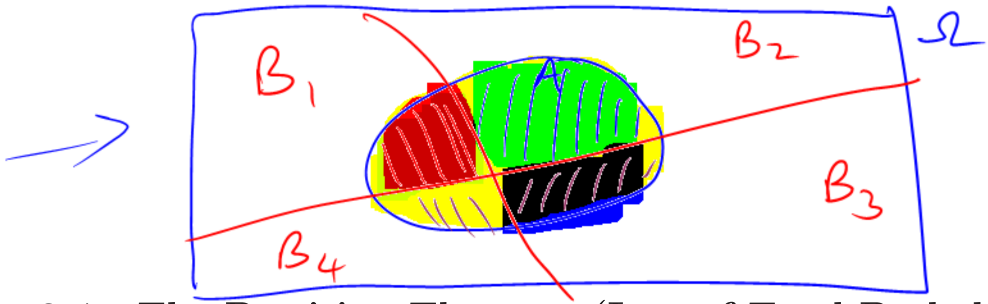
Examples:





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)

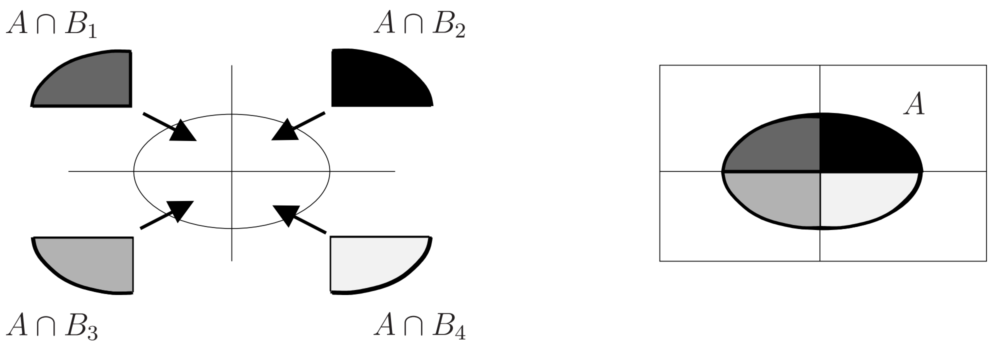
Let $B_1, B_2, \ldots, B_m$ form a partition of Ω. Then for any event A,

$$\boxed{\underset{\text{the whole}}{\mathbb{P}(A)} = \sum_{i=1}^{m} \mathbb{P}(A \cap B_i) = \underbrace{\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}_{\text{is the sum of its parts.}}}$$

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:

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

Extension of Bayes' Theorem

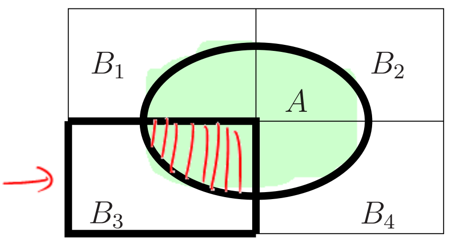
Suppose that $B_1, B_2, \ldots, B_m$ form a partition of Ω. By the Partition Theorem,

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

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

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

$B_j = B_3$



Ω

$\mathbb{P}(B_3 \mid A)$ represented by the ratio of red area to green area.

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

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

2. Information given: $\quad \Omega = \{\text{people being screened}\}$

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: $\quad \mathbb{P}(D \mid P)$.

$$\mathbb{P}(D \mid P) = \frac{\mathbb{P}(P \mid D)\,\mathbb{P}(D)}{\mathbb{P}(P)} \quad \text{(Bayes Thm).} \quad (*)$$

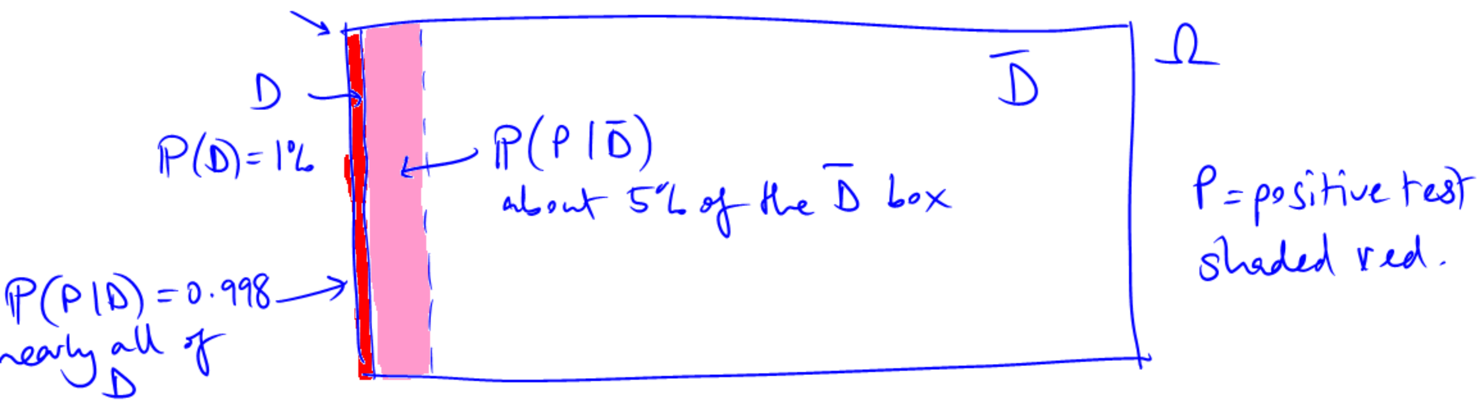
Now
$$\begin{aligned}\mathbb{P}(P \mid D) &= 1 - \mathbb{P}(\bar{P} \mid D) \\ &= 1 - \mathbb{P}(N \mid D) \\ &= 1 - 0.002 \\ &= 0.998 \quad (a)\end{aligned}$$

Also
$$\begin{aligned}\mathbb{P}(P) &= \mathbb{P}(P \mid D)\mathbb{P}(D) + \mathbb{P}(P \mid \bar{D})\mathbb{P}(\bar{D}) \\ &= 0.998 * 0.01 + 0.05(1 - 0.01) \\ &= 0.05948. \quad (b)\end{aligned}$$

Subst (a), (b) in (*):

$$\therefore \quad \mathbb{P}(D \mid P) = \frac{0.998 * 0.01}{0.05948} = 0.168.$$

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



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 often wish to know the probability of *eventually* reaching some particular state, given our current position.

Throughout this course, we will tackle this sort of problem using a technique called First-Step Analysis.

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:

Partition Theorem

$$\underbrace{\mathbb{P}(E)}_{\text{end up in } E \text{ eventually}} = \mathbb{P}(E \mid S_1)\,\mathbb{P}(S_1) + \mathbb{P}(E \mid S_2)\,\mathbb{P}(S_2) + \ldots + \mathbb{P}(E \mid S_k)\,\mathbb{P}(S_k).$$

1st step is to S_1 ... 1st step is to state S_k

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

Example: Tennis game at Deuce.

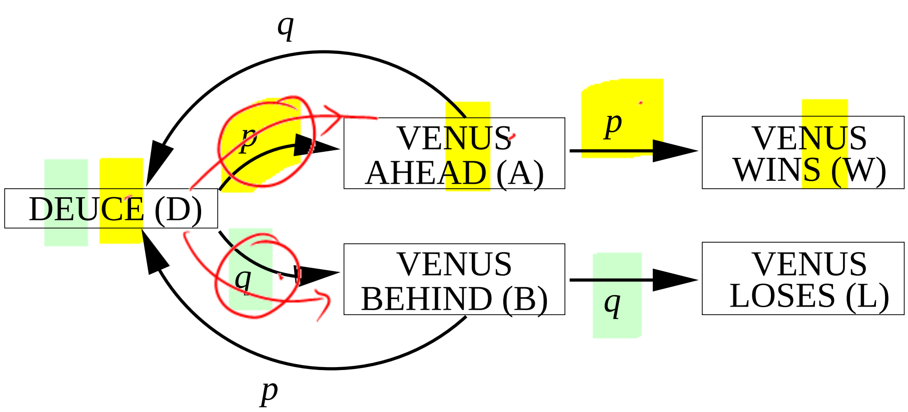
Venus and Serena are playing tennis, and have reached the score Deuce (40-40). (*Deuce* comes from the French word *Deux* for 'two', meaning that each player needs to win two consecutive points to win the game.)

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.



Use First-Step Analysis. The possible first steps starting from Deuce are:

1. Venus wins the next point (probability p): move to state A.
2. " loses " " " (probability q): move to state B.

Let V be the event that Venus wins EVENTUALLY, starting from Deuce,

so $v = \mathbb{P}(V \mid D_1)$. D_1 = at Deuce at time 1.

Starting from Deuce, the possible steps are to states A and B. So:

$$v = \mathbb{P}(\text{Venus wins} \mid D_1) = \mathbb{P}(V \mid D_1)$$

$$= \mathbb{P}_{D_1}(V) \quad \text{[rough notation: see p. 20].}$$

Partition Theorem →

$$= \mathbb{P}_{D_1}(V \mid A_2)\,\mathbb{P}_{D_1}(A_2) + \mathbb{P}_{D_1}(V \mid B_2)\,\mathbb{P}_{D_1}(B_2)$$

previous step D_1 is irrelevant once we know the next state A_2 or B_2. →

$$= \mathbb{P}(V \mid A_2)\, p + \mathbb{P}(V \mid B_2)\, q$$

$$= \mathbb{P}(V \mid A)\, p + \mathbb{P}(V \mid B)\, q. \quad (*)$$

Now we need to find $\mathbb{P}(V \mid A)$ and $\mathbb{P}(V \mid B)$. Use First-Step Analysis again.

$$\mathbb{P}(V \mid A_2) = \mathbb{P}_{A_2}(V \mid W_3)\, p + \mathbb{P}_{A_2}(V \mid D_3)\, q$$

$$= \underbrace{\mathbb{P}(V \mid W)}_{1}\, p + \underbrace{\mathbb{P}(V \mid D)}_{v}\, q$$

$$\mathbb{P}(V \mid A) = 1 * p + v q \quad (a)$$

Similarly, $\mathbb{P}(V \mid B_2) = \mathbb{P}_{B_2}(V \mid L_3)\, q + \mathbb{P}_{B_2}(V \mid D_3)\, p$

$$= 0 * q + v p$$

$$\mathbb{P}(V \mid B) = v p \quad (b)$$

Subst ⓐ and ⓑ into ⊛ :

$$v = (p + qv)\,p + (vp)\,q$$

$$v = p^2 + v.2pq$$

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

$$\therefore \boxed{v = \frac{p^2}{1-2pq}\,.}$$

Prob that Venus wins eventually, starting from Deuce.

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

So $1 - 2pq = p^2 + q^2$.

So the final probability that Venus wins the game is:

$$v = \frac{p^2}{1-2pq} = \frac{p^2}{p^2+q^2}$$

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.

First-step analysis as the Partition Theorem:

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

$$\mathbb{P}(\text{Venus wins eventually}) = v = \sum_{\text{first steps}} \mathbb{P}(V \mid \text{first step})\,\mathbb{P}(\text{first step})$$

First-step analysis is just the **Partition Theorem:**

The sample space is $\Omega = \{$all possible ROUTES from Deuce to the end$\}$.

An example of a sample point is: $D \to A \to D \to B \to D \to B \to L$

Another example is: $D \to B \to D \to A \to W$.

The **partition** of the sample space that we use in first-step analysis is:

$R_1 = \{$all possible routes Deuce to end that start with $D \to A\}$

$R_2 = \{$ " " " " " " " " " $D \to B\}$

Then first-step analysis simply states:

$$\mathbb{P}(V) = \mathbb{P}(V \mid R_1)\,\mathbb{P}(R_1) + \mathbb{P}(V \mid R_2)\,\mathbb{P}(R_2)$$

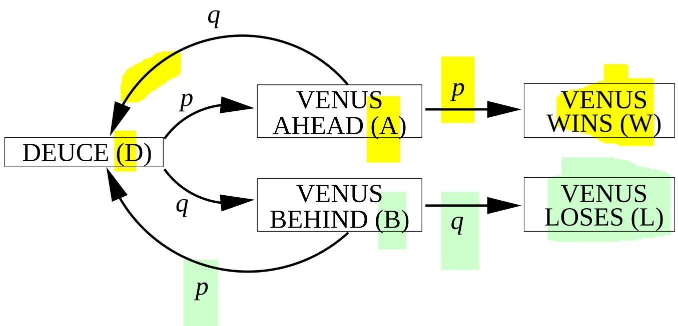
$$= \mathbb{P}_{D_1}(V \mid A_2)\,\mathbb{P}_{D_1}(A_2) + \mathbb{P}_{D_1}(V \mid B_2)\,\mathbb{P}_{D_1}(B_2).$$

Notation for quick solutions of first-step analysis problems

Defining a helpful notation is central to modelling with stochastic processes. Setting up well-defined notation helps you to solve problems quickly and easily. Defining your notation is one of the most important steps in modelling, because it provides the conversion from words (which is how your problem starts) to mathematics (which is how your problem is solved).

Several marks are allotted on first-step analysis questions for setting up a well-defined and helpful notation.

Use this way of solving eqns — NOT the notation on previous 2 pages.



Here is the correct way to formulate and solve this first-step analysis problem.

Need to find the probability that Venus wins eventually, starting from Deuce.

1. Define notation: let

$$v_D = \mathbb{P}(\text{Venus wins eventually} \mid \text{start at state } D)$$
$$v_A = \mathbb{P}(\text{Venus wins eventually} \mid \text{start at state } A)$$
$$v_B = \mathbb{P}(\text{Venus wins eventually} \mid \text{start at state } B)$$

all have same goal

all start in different places

2. First-step analysis:

$$v_D = p\,v_A + q\,v_B \qquad (a)$$
$$v_A = p * 1 + q\,v_D \qquad (b)$$
$$v_B = p\,v_D + q * 0 \qquad (c)$$

3. Substitute (b) and (c) into (a):

$$\Rightarrow v_D = p(p + q v_D) + q(p v_D)$$

$$\Rightarrow v_D(1 - pq - pq) = p^2$$

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

Much easier way of solving.

One of the foundational & most famous problems in probability.

2.7 Special Process: the Gambler's Ruin

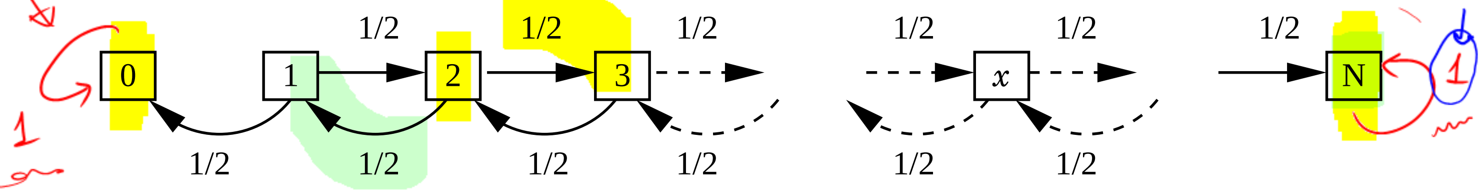

With permission, copyright Martin Ouellet

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 $\$N$, when she stops. What is the probability of the Gambler's Ruin, i.e. that the gambler ends up with \$0?

Ruin Win



Wish to find $\mathbb{P}(\text{ends with } \$0 \mid \text{starts with } \$x)$.

Define Notation: Let $p_x = \mathbb{P}(\text{ends with } \$0 \mid \text{starts with } \$x)$.

for $x = 0, 1, \ldots, N$.

Known Information: $p_0 = 1$ (definitely ruined if at state \$0)

$p_N = 0$ (already won, stop, so can't be ruined).

First-step Analysis: $p_x = \frac{1}{2} p_{x+1} + \frac{1}{2} p_{x-1}$ for $x = 1, 2, \ldots, N-1$. (*)

$\boxed{x-1} \xleftarrow{1/2} \boxed{x} \xrightarrow{1/2} \boxed{x+1}$

Eqn (*):

$$p_x = \frac{1}{2} p_{x+1} + \frac{1}{2} p_{x-1} \quad \text{for } x = 1, \ldots, N-1$$
$$p_0 = 1$$
$$p_N = 0$$

Solution of difference equation $(\star)$:

$$\begin{aligned} p_x &= \tfrac{1}{2}p_{x+1} + \tfrac{1}{2}p_{x-1} \quad \text{for } x = 1, 2, \ldots, N-1\,; \\ p_0 &= 1 \\ p_N &= 0. \end{aligned} \qquad (\star)$$

Boundary conditions

A & B solved using these.

We usually solve equations like this using the theory of 2nd-order difference equations. For this special case we will also verify the answer by two other methods.

1. Theory of linear 2nd order difference equations

Theory tells us that the general solution of $(\star)$ *is* $p_x = A + Bx$ *for some constants* A, B *and for* $x = 0, 1, \ldots, N$. *Our job is to find* A *and* B *using the boundary conditions:*

$p_x = A + Bx$ for constants A and B and for $x = 0, 1, \ldots, N$.

So $1 = p_0 = A + B * 0 = A \Rightarrow A = 1$

$0 = p_N = A + BN = 1 + BN$ (because $A=1$ just found)

$0 = 1 + BN$

$\Rightarrow B = -\frac{1}{N}$.

So our solution is: $\boxed{p_x = A + Bx = 1 - \frac{x}{N} \quad \text{for } x = 0, 1, \ldots, N.}$

For Stats 325, you will be told the general solution of the 2nd-order difference equation and expected to solve it using the boundary conditions.

For Stats 721, we will study the theory of 2nd-order difference equations. You will be able to derive the general solution for yourself before solving it.

Question: What is the probability that the gambler wins (ends with \$N), starting with \$x?

$$\mathbb{P}(\text{ends with } \$N) = 1 - \mathbb{P}(\text{ends with } 0)$$
$$= 1 - p_x$$
$$= \frac{x}{N} \quad \text{for } x = 0, 1, \ldots, N.$$

$\boxed{0}\ \boxed{1} \cdots \boxed{x} \cdots\cdots \boxed{N}$

$N = 30$

$x = 10$

$\mathbb{P}(\text{win}) = \frac{10}{30}$

2. Solution by inspection

The problem shown in this section is the ***symmetric*** Gambler's Ruin, where the probability is $\frac{1}{2}$ of moving up or down on any step. For this special case, we can solve the difference equation by inspection.

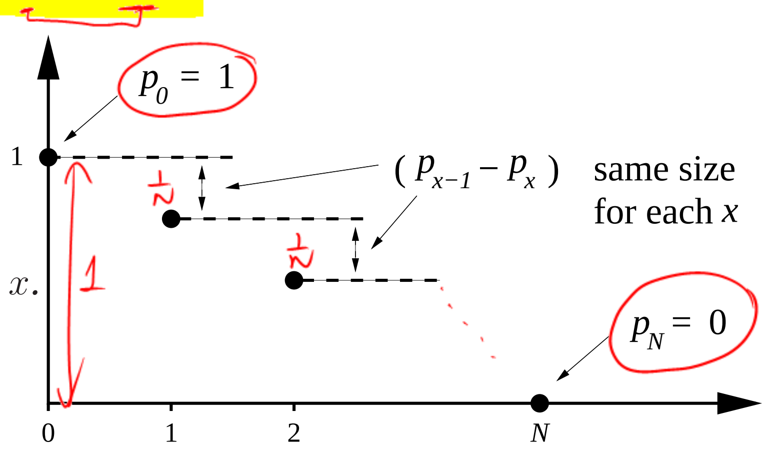
We have:

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

Rearranging: $$p_{x-1} - p_x = p_x - p_{x+1}.$$

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

There are N steps to go down from $p_0 = 1$ to $p_N = 0$.
Each step is the same size, because
$(p_{x-1} - p_x) = (p_x - p_{x+1})$ *for all x.*
So each step has size $1/N$,
$\Rightarrow \quad p_0 = 1$, $p_1 = 1 - 1/N$,
$p_2 = 1 - 2/N$, *etc.*



So

$$p_x = 1 - \frac{x}{N} \quad \textit{as before.}$$

3. Solution by repeated substitution.

In principle, all systems could be solved by this method, but it is usually too tedious to apply in practice.

Rearrange $(\star)$ *to give:*

$$
\begin{array}{rrcll}
 & p_{x+1} & = & 2p_x - p_{x-1} & \\
\Rightarrow \quad (x=1) & p_2 & = & 2p_1 - 1 & \textit{(recall } p_0 = 1\textit{)} \\
(x=2) & p_3 & = & 2p_2 - p_1 = 2(2p_1 - 1) - p_1 = 3p_1 - 2 & \\
(x=3) & p_4 & = & 2p_3 - p_2 = 2(3p_1 - 2) - (2p_1 - 1) = 4p_1 - 3 & \textit{etc} \\
 & & \vdots & & \\
\textit{giving} & p_x & = & xp_1 - (x-1) & \textit{in general,} \quad (\star\star) \\
\textit{likewise} & p_N & = & Np_1 - (N-1) & \textit{at endpoint.}
\end{array}
$$

Boundary condition: $p_N = 0 \quad \Rightarrow \quad Np_1 - (N-1) = 0 \quad \Rightarrow \quad p_1 = 1 - 1/N.$

Substitute in $(\star\star)$*:*

$$
\begin{aligned}
p_x &= xp_1 - (x-1) \\
&= x\left(1 - \tfrac{1}{N}\right) - (x-1) \\
&= x - \tfrac{x}{N} - x + 1 \\
p_x &= 1 - \tfrac{x}{N} \qquad \textit{as before.} \qquad \square
\end{aligned}
$$

2.8 Independence

Revision Read

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

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

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.

2.9 The Continuity Theorem

Read: New but subtle and only used a couple of times in course.

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

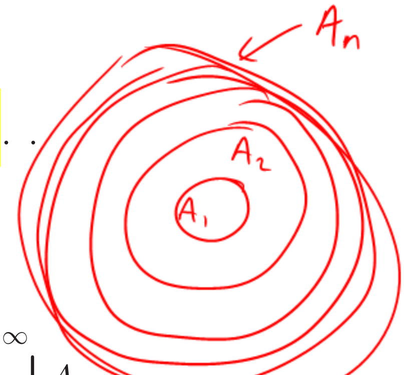
Theorem 2.9: The Continuity Theorem

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

$$\underbrace{\mathbb{P}}_{\text{prob of}}\Big(\underbrace{\lim_{n\to\infty} A_n}_{\text{limiting set}}\Big) = \underbrace{\lim_{n\to\infty} \mathbb{P}(A_n)}_{\text{limit of probabilities}}.$$



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

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

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

2.10 Random Variables

Revision: Read

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

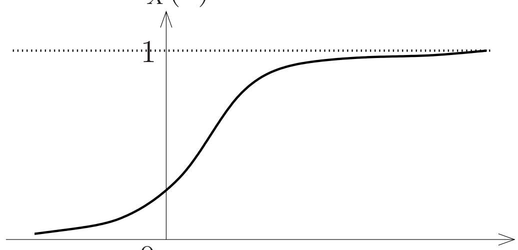
4) F_X is right-continuous: i.e. $\lim_{h \downarrow 0} F_X(x+h) = F_X(x)$.

2.11 Continuous Random Variables

Read

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* $X : \Omega \to [0, \infty)$.

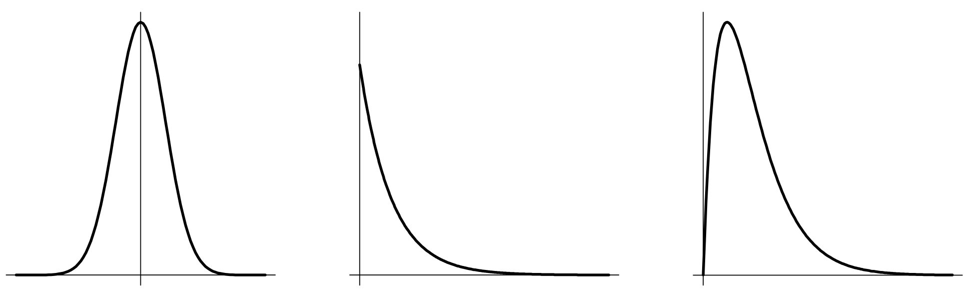


Probability Density Function for continuous random variables

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

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

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



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:

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

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

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

have a PDF, not a probability mass function

Calculating probabilities for continuous random variables

*

PDF = probability DENSITY function $= f_X(\cdot)$

CDF = cumulative distn function $= F_X(x) = \mathbb{P}(X \le x)$

X is continuous, $\mathbb{P}(X = x) = 0$ for ALL x.

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

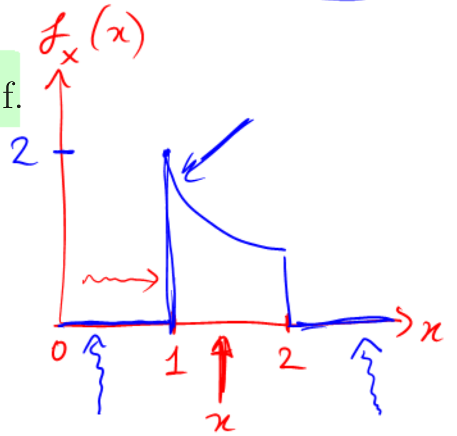
$\mathbb{P}(a \le X \le b) = F_X(b) - F_X(a)$ CDF

or

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

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 $\mathbb{P}(X \le 1.5)$.

a) $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du = \int_1^x 2u^{-2}\, du$ for $1 < x < 2$

$= \left[\frac{2u^{-1}}{-1} \right]_1^x = 2 - \frac{2}{x}$ for $1 < x < 2$.

So $F_X(x) = \begin{cases} 0 & \text{for } x \leq 1 \\ 2 - \frac{2}{x} & \text{for } 1 < x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$

b) $\mathbb{P}(X \leq 1.5) = F_X(1.5)$ by definition

$= 2 - \frac{2}{1.5}$

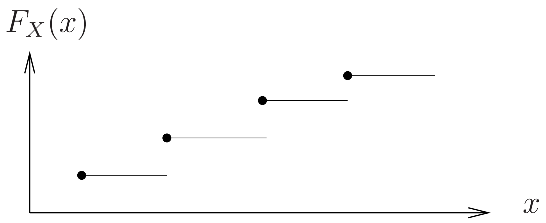
$= \frac{2}{3}$

2.12 Discrete Random Variables

Read: Revision.

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

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



Probability function

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

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

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

p. 42.

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

$$\mathbb{P}(A) = \sum_y \mathbb{P}(A \mid B_y)\mathbb{P}(B_y).$$

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

- Define event $A = \{X = x\}$
- Define events $B_0, B_1, B_2, \ldots$ as $B_y = \{Y = y\}$ for $y = 0, 1, 2, \ldots$ (or whatever values Y takes)
- Then, by the Partition Theorem,

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y)$$

2.13 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

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

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:

$$\text{Discrete } X, Y \text{ are indept} \iff \mathbb{P}(X = x \text{ AND } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \text{for ALL } x, y$$

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

Say we want to understand why $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ if A & B indept.

Suppose A happens $\frac{1}{2}$ the time: $\mathbb{P}(A) = \frac{1}{2}$

A & B indept $\Rightarrow$ B doesn't affect A.

So A should also happen on half the times that B happens.

$\hookrightarrow \mathbb{P}(A \cap B) = \frac{1}{2}\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B).$