Chapter 2 Part 1: Conditional probability and independence

Jess Kunke

MATH/STAT 394: Probability I (Summer 2022 A-term)

Outline

Review of Chapter 1

Conditional probability

Bayes' formula

Independence

Random variables: an introduction

Outline

Review of Chapter 1

Conditional probability

Bayes' formula

Independence

Random variables: an introduction

Review of Chapter 1: for your practice

The midterm will cover Chapters 1-2. I'll suggest a similar review after Ch. 2

As you review the material, make sure you are familiar with the terms and notation we covered. You are not responsible for knowing about sigma-algebras or a precise definition of the word "measure".

1. What are the three components of a probability model? Propose a simple experiment and illustrate how you would define a probability model for that example.
2. Write in math: the probability of a union of pairwise disjoint events is the sum of the probabilities of the individual events.
3. What are the two conditions we have to meet in order to use $P(A) = |A|/|\Omega|$ to compute the probability of an event A?
4. Give an example of each of the three basic types of sampling mechanisms we covered and explain how to compute the probability of an event.
5. Give an example of a countably infinite set and an uncountably infinite set.
6. If you draw a number uniformly at random from the interval [0, 5], what is the probability you get a number greater than 3? What is the probability you get the number 3?

7. How do you compute the probability of an event when the sample space has uncountably many equally likely outcomes?
8. What is a partition and how can you use it to compute a probability? Give an example of a partition of the real numbers.
9. What is the relationship between the probability of an event and the probability of that event not happening? Give an example.
10. If $B \subseteq C$, what can you say about the relationship between $P(B)$ and $P(C)$? Give an example.
11. What is the inclusion-exclusion rule? Draw a Venn diagram and demonstrate an example of how you can use this rule in practice.
12. What is de Morgan's law?

You will not be expected to state or know the names of the rules/results we covered like de Morgan's law and the inclusion-exclusion rule, just what the mathematical statements are, why they work, and how to use them.

Starting Chapter 2 today

Let's dive in!

Outline

- Review of Chapter 1
- Conditional probability
- Bayes' formula
- Independence
- Random variables: an introduction

Motivation

Additional information constrains the possible outcomes. How do we update the probability model?

Motivation

Additional information constrains the possible outcomes. How do we update the probability model?

Example

Class will be held in one of three rooms, and you think Room 1 is twice as likely as each of Rooms 2 or 3, which are equally likely.

$$\Omega = \{1, 2, 3\}, \quad \frac{1}{2}P(1) = P(2) = P(3)$$

$$\implies P(1) = \frac{1}{2}, P(2) = P(3) = \frac{1}{4}.$$

Motivation

Additional information constrains the possible outcomes. How do we update the probability model?

Example

Class will be held in one of three rooms, and you think Room 1 is twice as likely as each of Rooms 2 or 3, which are equally likely.

$$\Omega = \{1, 2, 3\}, \quad \frac{1}{2}P(1) = P(2) = P(3)$$

$$\implies P(1) = \frac{1}{2}, P(2) = P(3) = \frac{1}{4}.$$

Now you learn that Room 3 is unavailable. Define the event B that class is in either Room 1 or 2:

$$B = \{1, 2\}$$

What is the new probability measure $\tilde{P}$?

Additional information constrains the possible outcomes. How do we update the probability model?

Example

Class will be held in one of three rooms, and you think Room 1 is twice as likely as each of Rooms 2 or 3, which are equally likely.

$$\Omega = \{1, 2, 3\}, \quad \frac{1}{2}P(1) = P(2) = P(3)$$

$$\implies P(1) = \frac{1}{2}, P(2) = P(3) = \frac{1}{4}.$$

Now you learn that Room 3 is unavailable. Define the event B that class is in either Room 1 or 2:

$$B = \{1, 2\}$$

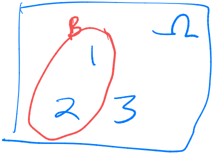
What is the new probability measure $\tilde{P}$?

have the

We know $\tilde{P}(3) = 0$ and let's suppose that we still think Rooms 1 and 2 ~~are~~ ~~equally likely~~. Since probabilities must sum to 1, $\cancel{\tilde{P}(1) = \tilde{P}(2) = \frac{1}{2}}$

same relative probabilities as before. $\tilde{P}(1) = \frac{2}{3}$, $\tilde{P}(2) = \frac{1}{3}$

Motivation



Before we have the information about Room 3,

$$P(B) = P(1) + P(2) = \frac{3}{4}.$$

Once we learn B has happened, we want to restrict to B (require that $P(B) = 1$) and leave the other relative probabilities intact, so we divide by $P(B)$:

$$1 = \frac{P(B)}{P(B)} = \frac{P(1) + P(2)}{P(B)}$$

Therefore we can define

$$\tilde{P}(1) = \frac{P(1)}{P(B)}, \quad \tilde{P}(2) = \frac{P(2)}{P(B)}.$$

$$= \frac{1/2}{3/4}$$

$$= 2/3$$

Conditional probability: definition

Definition

Let B be an event in the sample space Ω such that $P(B) > 0$. Then for all events A the **conditional probability of A given B** is defined as

$$P(A \mid B) = \frac{P(AB)}{P(B)}. \quad \leftarrow \quad P(A \cap B)$$

Notes:

- In this section we will often use the following shorthand notation for intersections: AB for $A \cap B$, ABC for $A \cap B \cap C$.
- $P(B)$ must be greater than zero; otherwise we are conditioning on something impossible

In our previous example,

$$\tilde{P}(1) = P(\text{Room } 1 \mid \underbrace{\text{Room } 1 \text{ or } 2}_{B}) = \frac{P(\{1\} \text{ and } \{1,2\})}{P(\{1,2\})} = \frac{P(\{1\})}{P(\{1,2\})} = \frac{P(1)}{P(B)}.$$

An example

Example

We have an urn containing 4 red and 6 green balls. We draw a sample of three without replacement. Find the probability that the sample contains exactly 2 red balls given that at least one ball in the sample is red.

Define $A = \{\text{exactly 2 red balls}\}$ $B = \{\text{at least 1 red ball}\}$

$A \subseteq B$

unord. w/o replacement

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

$$P(AB) = P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{3}{10}.$$

$$P(B) = 1 - P(B^c) = 1 - P(\text{no red}) = 1 - \frac{\binom{6}{3}}{\binom{10}{3}} = \frac{5}{6}.$$

$$P(A|B) = \frac{3/10}{5/6} = \frac{9}{25}.$$

Some properties

Be careful:

$$P(A \mid B) \neq P(B \mid A)$$

$$P(\text{suspect is guilty} \mid \text{evidence}) \neq P(\text{evidence} \mid \text{suspect is guilty})$$

Some properties

Be careful:

$$P(A \mid B) \neq P(B \mid A)$$

$$P(\text{suspect is guilty} \mid \text{evidence}) \neq P(\text{evidence} \mid \text{suspect is guilty})$$

Properties of conditional probability

1. $P(A \mid A) =?$ 1
2. $P(A \mid \Omega) =?$ $P(A)$
3. $P(A^c \mid A) =?$ 0
4. $P(A^c \mid B) =?$ $1 - P(A \mid B)$

Some properties

Be careful:

$$P(A \mid B) \neq P(B \mid A)$$

$$P(\text{suspect is guilty} \mid \text{evidence}) \neq P(\text{evidence} \mid \text{suspect is guilty})$$

Properties of conditional probability

1. $P(A \mid A) = ?$
2. $P(A \mid \Omega) = ?$
3. $P(A^c \mid A) = ?$
4. $P(A^c \mid B) = ?$

Generally:
Given B s.t. $P(B) > 0$, then $P(\cdot \mid B) : A \to P(A \mid B)$ is a probability measure

Multiplication rule

Sometimes conditional probability is easier to use if we multiply through by $P(B)$:

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

Multiplication rule: For n events $A_1, ..., A_n$, if the conditional probabilities below make sense for the problem, then

$$P(A_1 \cancel{,} ... \cancel{,} A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1 \cdots A_n).$$

An example

Example

We have an urn containing 8 red balls and 4 white balls. We draw two without replacement. What is the probability that both are red?

An example

Example

We have an urn containing 8 red balls and 4 white balls. We draw two without replacement. What is the probability that both are red?

Our previous approach: define the event of interest as $A = \{\text{both balls are red}\}$. Then we would compute $|A|$ and $|\Omega|$ and compute $P(A)$:

An example

Example

We have an urn containing 8 red balls and 4 white balls. We draw two without replacement. What is the probability that both are red?

Our previous approach: define the event of interest as $A = \{\text{both balls are red}\}$. Then we would compute $|A|$ and $|\Omega|$ and compute $P(A)$:

$$|A| = \binom{8}{2} = \frac{8!}{6!2!} = \frac{8 \cdot 7}{2}, \quad |\Omega| = \binom{12}{2} = \frac{12!}{10!2!} = \frac{12 \cdot 11}{2},$$

and

$$P(A) = \frac{8 \cdot 7}{12 \cdot 11} = \frac{14}{33}.$$

An example

Example

We have an urn containing 8 red balls and 4 white balls. We draw two without replacement. What is the probability that both are red?

Our previous approach: define the event of interest as $A = \{\text{both balls are red}\}$. Then we would compute $|A|$ and $|\Omega|$ and compute $P(A)$:

$$|A| = \binom{8}{2} = \frac{8!}{6!2!} = \frac{8 \cdot 7}{2}, \quad |\Omega| = \binom{12}{2} = \frac{12!}{10!2!} = \frac{12 \cdot 11}{2},$$

and

$$P(A) = \frac{8 \cdot 7}{12 \cdot 11} = \frac{14}{33}.$$

Using conditional probability: Let

$$R_1 = \{\text{first draw is red}\}, \quad R_2 = \{\text{second draw is red}\}.$$

Then $A = R_1 R_2$ and

$$P(R_1 R_2) = P(R_1)P(R_2|R_1) = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}.$$

Combining what we've learned

Example

You have two urns: Urn 1 contains 2 green balls and 1 red ball. Urn 2 contains 2 red balls and 3 yellow balls. You perform a two-stage experiment:

Stage 1: Select an urn with equal probability

Stage 2: Draw 1 ball uniformly at random from that urn

$P(\{red\}) = ?$

$\Omega = \{ \underbrace{\text{Urn 1 } G_1, \text{ Urn 1 } G_2, \text{ Urn 1 } R,}_{\{\text{Urn 1}\}} \underbrace{\text{Urn 2 } R_1, \text{ Urn 2 } R_2, \text{ Urn 2 } Y_1, \text{ Urn 2 } Y_2, \text{ Urn 2 } Y_3}_{\{\text{Urn 2}\}} \}$

not equally likely, But within each urn they are

$\{\text{Urn 1}\} \cup \{\text{Urn 2}\} = \Omega, \quad \{\text{Urn 1}\} \cap \{\text{Urn 2}\} = \emptyset$

$\Rightarrow \{\text{Urn 1}\}, \{\text{Urn 2}\}$ partition Ω.

partition

$P(\text{red}) = P(\{\text{red}\} \cap \{\text{Urn 1}\}) + P(\{\text{red}\} \cap \{\text{Urn 2}\})$

$= P(\text{red} \mid \text{Urn 1}) P(\text{Urn 1}) + P(\text{red} \mid \text{Urn 2}) P(\text{Urn 2})$

$= \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{2} = \frac{11}{30}.$

Combining what we've learned

Example

You have two urns: Urn 1 contains 2 green balls and 1 red ball. Urn 2 contains 2 red balls and 3 yellow balls. You perform a two-stage experiment:

Stage 1: Select an urn with equal probability

Stage 2: Draw 1 ball uniformly at random from that urn

$P(\{red\}) =?$

Extra practice: $P(\{yellow\}) =?$ $P(\{green\}) =?$

Law of total probability

This was an example of a very useful result called the **law of total probability**: If $B_1, ..., B_n$ is a partition of Ω with $P(B_i) > 0 \; \forall \, i = 1, ..., n$, then for any event A we have

$$P(A) = \sum_{i=1}^{n} P(AB_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Outline

Review of Chapter 1

Conditional probability

Bayes' formula

Independence

Random variables: an introduction

Motivation

Think about the two-stage urn problem again

- We answered $P(\text{red})$ by accounting for the possibilities that the ball comes from Urn 1 or Urn 2
- Now let's ask, *if you draw a red ball, what is the probability that it came from Urn 1?*

$\Longrightarrow$ conditional probabilities as evidence for comparing competing explanations for an observed event

$$\begin{aligned}
P(\text{Urn I} \mid \text{red}) &= && \text{cond. prob.} \\
&= && \text{cond. prob.} \\
&= && \text{law of total prob.} \\
&= && \text{(final answer)}
\end{aligned}$$

Motivation

Think about the two-stage urn problem again

- We answered $P(\text{red})$ by accounting for the possibilities that the ball comes from Urn 1 or Urn 2
- Now let's ask, *if you draw a red ball, what is the probability that it came from Urn 1?*

$\implies$ conditional probabilities as evidence for comparing competing explanations for an observed event

$P(B)\, P(A|B) = P(AB)$

$$
\begin{aligned}
P(\text{Urn } 1 \mid \text{red}) &= \frac{P(\text{red} \cap \text{Urn } 1)}{P(\text{red})} && \text{cond. prob.} \\
&= \frac{P(\text{red} \mid \text{Urn } 1)P(\text{Urn } 1)}{P(\text{red})} && \text{mult. rule} \\
&= \frac{P(\text{red} \mid \text{Urn } 1)P(\text{Urn } 1)}{P(\text{red} \mid \text{Urn } 1)P(\text{Urn } 1) + P(\text{red} \mid \text{Urn } 2)P(\text{Urn } 2)} && \text{law of total prob.} \\
&= \frac{\frac{1}{3}\cdot\frac{1}{2}}{\frac{1}{3}\cdot\frac{1}{2}+\frac{2}{5}\cdot\frac{1}{2}} = \frac{5}{11}. && \text{(final answer)}
\end{aligned}
$$

Bayes' formula

General statements of what we just did:

Bayes' formula: If $P(A), P(B), P(B^c) > 0$, then

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}.$$

Bayes' formula

General statements of what we just did:

Bayes' formula: If $P(A), P(B), P(B^c) > 0$, then

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}.$$

We can go further than B and B^c. If $B_1, ..., B_n$ partition the sample space Ω and $P(B_i) > 0$ for all i, then for any event A with $P(A) > 0$, and any $k = 1, ..., n$,

$$P(B_k \mid A) = \frac{P(AB_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

Bayes' formula

General statements of what we just did:

Bayes' formula: If $P(A), P(B), P(B^c) > 0$, then

$$P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}.$$

We can go further than B and B^c. If $B_1, ..., B_n$ partition the sample space Ω and $P(B_i) > 0$ for all i, then for any event A with $P(A) > 0$, and any $k = 1, ..., n$,

$$P(B_k \mid A) = \frac{P(AB_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

- A is our additional information
- $P(B_k)$ for $k = 1, ..., n$ are called **prior probabilities** (beliefs before additional information A is collected)
- $P(B_k \mid A)$ for $k = 1, ..., n$ are called **posterior probabilities** (beliefs updated based on A)

An example

Example

Test T:

- Prob. of True positive: $P(T = + \mid \text{covid}) = 90\%$.
- Prob. of True negative : $P(T = - \mid \text{healthy}) = 95\%$.

Assume the prevalence of COVID-19 in the population is $\approx 2\%$. ($P(\text{covid}) = 2\%$)

- What is $P(\text{covid} \mid T = +)$ for someone at random in the population?
- Now consider the test is done on people that have symptoms and we know that for these people, $P(\text{covid}) = 50\%$, what is $P(\text{covid} \mid T = +)$?

An example

Example

Test T:

- Prob. of True positive: $P(T = + \mid \text{covid}) = 90\%$.
- Prob. of True negative : $P(T = - \mid \text{healthy}) = 95\%$.

Assume the prevalence of COVID-19 in the population is $\approx 2\%$. ($P(\text{covid}) = 2\%$)

- What is $P(\text{covid} \mid T = +)$ for someone at random in the population?
- Now consider the test is done on people that have symptoms and we know that for these people, $P(\text{covid}) = 50\%$, what is $P(\text{covid} \mid T = +)$?

Solution:

$$
\begin{aligned}
P(\text{covid} \mid +) &= \frac{P(+ \mid \text{covid})P(\text{covid})}{P(+ \mid \text{covid})P(\text{covid}) + P(+ \mid \text{heathy})P(\text{healthy})} \\
&= \frac{90\% \times 2\%}{90\% \times 2\% + (1 - 95\%)(1 - 2\%)} \approx 26.9\%.
\end{aligned}
$$

An example

Example

Test T:

- Prob. of True positive: $P(T = + \mid \text{covid}) = 90\%$.
- Prob. of True negative : $P(T = - \mid \text{healthy}) = 95\%$.

Assume the prevalence of COVID-19 in the population is $\approx 2\%$. ($P(\text{covid}) = 2\%$)

- What is $P(\text{covid} \mid T = +)$ for someone at random in the population?
- Now consider the test is done on people that have symptoms and we know that for these people, $P(\text{covid}) = 50\%$, what is $P(\text{covid} \mid T = +)$?

Solution:

$$\begin{aligned} P(\text{covid} \mid +) &= \frac{P(+ \mid \text{covid})P(\text{covid})}{P(+ \mid \text{covid})P(\text{covid}) + P(+ \mid \text{heathy})P(\text{healthy})} \\ &= \frac{90\% \times 2\%}{90\% \times 2\% + (1 - 95\%)(1 - 2\%)} \approx 26.9\%. \end{aligned}$$

Surprising isn't it? In practice, either two independent tests are used which improve the result or we consider only people that have symptoms.

If you consider only people that have the symptoms, then by the same computations, $P(\text{covid} \mid +) = 95\%$.

Outline

Review of Chapter 1

Conditional probability

Bayes' formula

Independence

Random variables: an introduction

Motivation

- Conditional probability $P(A|B)$ quantifies the effect of event B on another event A
- Intuition: Two events A and B are independent if $P(A|B) = P(A)$
- Applying Bayes' formula (if $P(A) > 0$), this would imply that $P(B) = P(AB)/P(A)$, i.e. $P(AB) = P(A)P(B)$

Definition

Two events A and B are **independent** if

$$P(AB) = P(A)P(B).$$

Notes:

- This only makes sense when the two events are defined on the same sample space

Example

independent:
$P(AB) = P(A)\,P(B)$

Example

Consider an urn with 4 red and 7 green balls.
Sample in order two balls and define the events

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is green}\}$$

1. If the sampling is with replacement, are A and B independent?
2. If the sampling is without replacement, are A and B independent?

Example

Example

Consider an urn with 4 red and 7 green balls.
Sample in order two balls and define the events

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is green}\}$$

1. If the sampling is with replacement, are A and B independent?
2. If the sampling is without replacement, are A and B independent?

- Intuition
  - If you replace the balls, the sampling restart, and the events should be independent.
  - If you do not replace the ball, the second sampling will be affected by which ball you got first.

Example

Example

Consider an urn with 4 red and 7 green balls.
Sample in order two balls and define the events

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is green}\}$$

1. If the sampling is with replacement, are A and B independent?
2. If the sampling is without replacement, are A and B independent?

- Intuition
  - If you replace the balls, the sampling restart, and the events should be independent.
  - If you do not replace the ball, the second sampling will be affected by which ball you got first.
- With replacement:
  In this case,

$$|\Omega| = 11^2, \quad |A| = 4 \cdot 11, \quad |B| = 11 \cdot 7, \quad |A \cap B| = 4 \cdot 7$$

$$P(A \cap B) = \frac{4 \cdot 7}{11^2}, \quad P(A) \cdot P(B) = \frac{4 \cdot 11}{11^2} \frac{7 \cdot 11}{11^2}$$

  Hence $P(A \cap B) = P(A)P(B)$, so the events are independent.

Example

Example

Consider an urn with 4 red and 7 green balls.
Sample in order two balls and define the events

$$A = \{\text{first ball is red}\} \quad B = \{\text{second ball is green}\}$$

1. If the sampling is with replacement, are A and B independent?
2. If the sampling is without replacement, are A and B independent?

- Without replacement:
  In this case,
  $$|\Omega| = 11 \cdot 9, \quad |A| = 4 \cdot 10,$$
  $$|B| = |A \cap B| + |A \cap B^c| = 4 \cdot 7 + 7 \cdot 6 = 70, \quad |A \cap B| = 4 \cdot 7$$
  $$P(A \cap B) = \frac{4 \cdot 7}{11 \cdot 10} = \frac{28}{110}, \quad P(A) \cdot P(B) = \frac{4 \cdot 10}{11 \cdot 10} \frac{70}{11 \cdot 10} = \frac{28}{121}$$
  Hence $P(A \cap B) \neq P(A)P(B)$, so the events are not independent.

Caution: disjoint vs independent

Do not confound **disjoint events** and **independent events**.

For example A and A^c are disjoint,
but $0 = P(A \cap A^c) \neq P(A)P(A^c) > 0$ for any A s.t. $0 < P(A) < 1$.

Independence

If the fact that B occurs does not change the probability of A, then knowing that B does not occur should also have no impact on the probability of A.

This is formalized in the proposition below:

Independence

If the fact that B occurs does not change the probability of A, then knowing that B does not occur should also have no impact on the probability of A.

This is formalized in the proposition below:

Suppose A and B are independent. Then

- A^c and B are independent
- A and B^c are independent
- A^c and B^c are independent

Proof.
We will prove the first result above. We saw in Chapter 1 that $P(B) = P(A^cB) + P(AB)$. Rearranging,

$$\begin{aligned} P(A^cB) &= P(B) - P(AB) & \\ &= P(B) - P(A)P(B) & A, B \text{ independent} \\ &= [1 - P(A)]P(B) & \text{factoring} \\ &= P(A^c)P(B). & \text{def. of complement} \end{aligned}$$

Therefore, A and B^c are independent. The rest are left as practice. $\square$

Mutual independence

For more than two events, we generalize the notion of independence as follows:

Mutual independence

For more than two events, we generalize the notion of independence as follows:

Definition

Events $A_1, \ldots, A_n$ are **independent** (or **mutually independent)** if for any collection $A_{i_1}, \ldots, A_{i_k}$ with $2 \leq k \leq n$, $1 \leq i_1 < \ldots < i_k \leq n$,

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$$

Mutual independence

For more than two events, we generalize the notion of independence as follows:

Definition

Events $A_1, \ldots, A_n$ are **independent** (or **mutually independent)** if for any collection $A_{i_1}, \ldots, A_{i_k}$ with $2 \leq k \leq n$, $1 \leq i_1 < \ldots < i_k \leq n$,

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$$

Events $A_1, \ldots, A_n$ are **pairwise independent** if for any $i \neq j$,

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

Mutual independence

For more than two events, we generalize the notion of independence as follows:

Definition

Events $A_1, \ldots, A_n$ are **independent** (or **mutually independent)** if for any collection $A_{i_1}, \ldots, A_{i_k}$ with $2 \leq k \leq n$, $1 \leq i_1 < \ldots < i_k \leq n$,

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$$

Events $A_1, \ldots, A_n$ are **pairwise independent** if for any $i \neq j$,

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

- Clearly (mutual) independence $\Longrightarrow$ pairwise independence
- But **the reverse is false**
- Mutual independence requires that **any** collection of the variables satisfy the factorization.

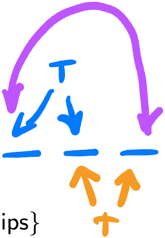
Mutual independence

if $P(A|B) = \frac{P(AB)}{P(B)} = P(A)$, then $\frac{P(AB)}{P(B)} = P(A)$

$P(AB) = P(A)P(B)$

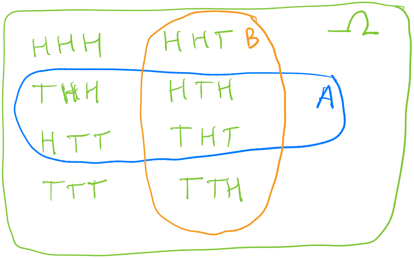
Example

Flip a fair coin three times. Let



$$A = \{\text{exactly one tails in the first two flips}\}$$
$$B = \{\text{exactly one tails in the last two flips}\}$$
$$C = \{\text{exactly one tails in the first and last flips}\}$$

1. Are A, B, C pairwise independent ? Yes.
2. Are A, B, C (mutually) independent?



$P(A) = \frac{|A|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}$

draw A & B

$P(A|B) = \frac{|AB|}{|B|} = \frac{2}{4} = \frac{1}{2}$
$= P(A)$

A & B indep.

$\Rightarrow$ by symmetry B & C indep., C & A indep

$P(AB) = \frac{|AB|}{|\Omega|} = \frac{2}{8} = \frac{1}{4}$, $P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = 1/4$

Mutual independence

Example

Flip a fair coin three times. Let

$$A = \{\text{exactly one tails in the first two flips}\}$$
$$B = \{\text{exactly one tails in the last two flips}\}$$
$$C = \{\text{exaxctly one tails in the first and last flips}\}$$

1. Are A, B, C pairwise independent ?
2. Are A, B, C (mutually) independent?

Solution:

1. Based on our example before, we see that they are pairwise independent.

Mutual independence

Example

Flip a fair coin three times. Let

$$A = \{\text{exactly one tails in the first two flips}\}$$
$$B = \{\text{exactly one tails in the last two flips}\}$$
$$C = \{\text{exaxctly one tails in the first and last flips}\}$$

$A \cap B = \{(T, H, T), (H, T, H)\}$

2 T in 1st & 3rd

0 T in 1st & 3rd

1. Are A, B, C pairwise independent ?
2. Are A, B, C (mutually) independent?

Solution:

1. Based on our example before, we see that they are pairwise independent.
2. Yet $A \cap B \cap C = \emptyset$ so

$$0 = P(A \cap B \cap C) \neq P(A)P(B)P(C) > 0$$

They are not (mutually) independent.

The following proposition simply generalizes the fact that if A and B are independent then their complements are also independent.

Mutual independence

The following proposition simply generalizes the fact that if A and B are independent then their complements are also independent.

If $A_1, \dots A_n$ are (mutually) independent, then for any collection $A_{i_1}, \dots, A_{i_k}$ with $2 \leq k \leq n$ and $1 \leq i_1 < \dots < i_k \leq n$, we have that

$$P(A^*_{i_1} \cap \dots \cap A^*_{i_k}) = P(A^*_{i_1}) \dots P(A^*_{i_k})$$

where each A^*_i is either A_i or A^c_i. $A_1, \dots, A_5$

e.g. $P(A_1 A_2 A_3^c A_4^c A_5) = P(A_1)P(A_2)P(A_3^c)P(A_4^c)P(A_5)$

Outline

Review of Chapter 1

Conditional probability

Bayes' formula

Independence

Random variables: an introduction

Motivation

Consider the example of rolling a die three times.

- So far we've asked questions like these:
  - What is the probability of rolling the same number more than once?
  - What is the probability that the sum of the rolls is at least 10?
- What about questions like these:
  - What is a "typical" die roll?
  - What die roll would we expect on average if we roll many times?
  - Suppose this is a carnival game and the prize you win is some function of the sequence of die rolls. What prize would you expect to win on average?

Random variables can make it easier to handle the first type of question, and they will enable us to address the second type of question

Random variables

Definition

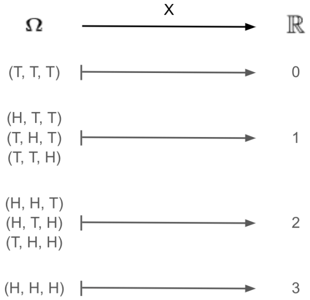
A **random variable** X on a sample space Ω is a function from Ω into the real numbers, or a real-valued function on Ω:

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

- Usually denoted by a capital letter

Example

Flip a coin three times. Define X = number of heads.



Events and random variables

We denote[1]

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}.$$

This gives us two ways to define an event:

- In our previous example, consider the event "the coin comes up heads once". We've seen that we can define this by which outcomes in Ω meet this criterion:

$$\{(H, T, T), (T, H, T), (T, T, H)\}$$

- We now see that we can also view this as the subset of Ω that is mapped by X to the real number 1:

$$\{\omega \in \Omega : X(\omega) = 1\}$$

Idea: we can measure probabilities in Ω by expressing them as the set of elements such that the random variable satisfies some equality/inequality

[1] For those familiar with the following terminology, the preimage in Ω of any subset B of the codomain (set of possible outputs/values) of a random variable X on Ω is an event.

A note on notation

Let X, Y be random variables.

Equivalent expressions of "both":

- intersection
- $P(X < 3, Y = 4) \quad \leftarrow$ comma means "and"
- $P(X < 3 \text{ and } Y = 4)$
- $P(X < 3) \cap P(Y = 4)$ intersection

Equivalent expressions of "either/or":

- union
- $P(X < 3 \text{ or } Y = 4)$ Union
- $P(X < 3) \cup P(Y = 4)$

Now that we have random variables, we can define the following as well:

Definition

Let X be a random variable.

- The **probability distribution** of X is the collection of probabilities $P\{X \in B\}$ for sets $B \subseteq \mathbb{R}$.
- X is a **discrete random variable** if $\exists$ a finite or countably infinite set $\{k_1, k_2, \ldots\}$ of real numbers such that

$$\sum_i P(X = k_i) = 1.$$

- The **probability mass function** or pmf of a discrete random variable X is the function p (or p_X) defined by

$$p(k) = P(X = k)$$

  for all possible values k of X.

Example 1: Finding the pmf of a discrete RV

Let's model daily stock price as a coin flip. Every time the coin comes up heads (tails), the price increases (decreases) and you earn \$1 (\$1). Define X as your net profit or loss after five flips. in dollars

lose

- What is the set of possible values of X?
- Define the probability mass function for X.

$X \in \{-5, -3, -1, 1, 3, 5\}$

5 H : \$5 $\binom{5}{5}$

$$P(X=k) = \begin{cases} \frac{1}{2^5} = \frac{1}{32} & k=5 \\ 5/32 & k=3 \\ & k=1 \\ & k=-1 \\ & k=-3 \\ & k=-5 \end{cases}$$

$= 1$

4H, 1T : \$4 − \$1 = \$3

$\binom{5}{4} = \binom{5}{1}$

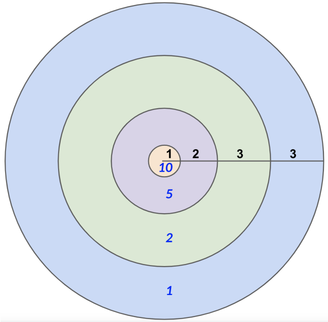
$$\sum_{k \in \{-5,-3,-1,1,3,5\}} P(X=k) = 1.$$

Example 2: Finding the pmf of a discrete RV

Suppose we have the target drawn below; the black numbers along the horizontal segment are the difference in radii of the circles, and the blue vertical numbers are the number of points you get if your arrow lands in that region. Suppose your arrow lands uniformly at random on the target, and define X to be the number of points you get.

- What are the possible values of X (the codomain of X)? $X \in \{1, 2, 5, 10\}$.
- Define the pmf of X. Verify that the probabilities sum to 1.

$$p(k) = P(X=k) = \begin{cases} & k=1 \\ & k=2 \\ & k=5 \\ & k=10 \end{cases}$$



$P(X=1) = 45/81$

$P(X=2) = 27/81$

$P(X=5) = P\left(\text{bullseye or purple}\right) - P(\text{bulls}) = \frac{9\pi}{81\pi} - \frac{1}{81} = \frac{8}{81}$

$P(X=10) = \frac{\text{area(bullseye)}}{\text{area(target)}} = \frac{\pi}{81\pi} = \frac{1}{81}$