Computer Science 351

Conditional Probability and Independence

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #19

Learning Goals

Learning Goals:

- Review concepts concerning relationships between events, in an experiment, including
  - ***conditional probabilities,*** and
  - ***independence*** of sets of events.

  Additional useful results about these may also be introduced.

Conditional Probability

Let Ω be a sample space and let $A, B \subseteq \Omega$ be events such that $\mathrm{P}(B) > 0$. The ***conditional probability*** of A given B, denoted $\mathrm{P}(A \mid B)$, is

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

$\mathrm{P}(A \mid B)$ is not defined if $\mathrm{P}(B) = 0$.

Conditional Probability

Example: Consider the "Balls and Bins" example from the previous lecture, which involved two parameters:

- m — the number of balls
- n — the number of bins into which the balls must be placed.

Each outcome was represented as a *sequence*

$$(\alpha_1, \alpha_2, \ldots, \alpha_m)$$

where α_i is an integer such that $1 \leq \alpha_i \leq n$ for every integer i such that $1 \leq m$: This represents the situation where the i^{th} ball is placed in the α_i^{th} bin, for all such i.

Conditional Probability

The sample space is the set

$$\Omega = \{(\alpha_1, \alpha_2, \ldots, \alpha_m) \,|\alpha_i \in \mathbb{Z} \text{ and } 1 \leq \alpha_i \leq n \text{ for every integer } i \text{ such that } 1 \leq i \leq m\},$$

so that $|\Omega| = n^m$.

- It follows that if $\mathrm{P} : \Omega \to \mathbb{R}$ is the ***uniform probability distribution*** for this sample space, then

  $$\mathrm{P}(\vec{\alpha}) = n^{-m} = \frac{1}{n^m}$$

  for every m-tuple $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m) \in \Omega$.

Conditional Probability

Suppose that $m \geq 2$. Let us consider the following events.

- A: $\alpha_1 = 1$, that is, the first ball is placed in the first bin.
- B: $\alpha_1 = \alpha_2$, that is, the first and second balls are placed in the same bin.

Then A is the subset

$$\{(1, \alpha_2, \alpha_3, \ldots, \alpha_m) \mid \alpha_i \in \mathbb{Z} \text{ and } 1 \leq i \leq n \\ \text{for every integer } i \text{ such that } 2 \leq i \leq m\}$$

of Ω — so that $|A| = n^{m-1}$ and

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{n^{m-1}}{n^m} = \frac{1}{n}.$$

Conditional Probability

B is the subset

$$\{(\alpha_1, \alpha_2, \ldots, \alpha_m) \mid \alpha_i \in \mathbb{Z} \text{ and } 1 \leq i \leq n \text{ for every integer } i \text{ such that } 1 \leq i \leq m, \text{ and such that } \alpha_1 = \alpha_2\}$$

of Ω — so that $|B| = n^{m-1}$ and

$$\mathrm{P}(B) = \frac{|B|}{|\Omega|} = \frac{n^{m-1}}{n^m} = \frac{1}{n}.$$

Conditional Probability

$B \cap A$ is the subset

$$\{(1, 1, \alpha_3, \alpha_4, \ldots, \alpha_m) \mid \alpha_i \in \mathbb{Z} \text{ and } 1 \leq i \leq n \text{ for every integer } i \text{ such that } 3 \leq i \leq m\}$$

of Ω — so that $|B \cap A| = n^{m-2}$ and

$$\mathrm{P}(B \cap A) = \frac{|B \cap A|}{|\Omega|} = \frac{n^{m-2}}{n^m} = \frac{1}{n^2}.$$

Conditional Probability

Now

$$\begin{aligned} P(B \mid A) &= \frac{P(B \cap A)}{P(B)} \qquad \text{(by the definition of } P(B \mid A)) \\ &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{1/n^2}{1/n} \\ &= \frac{1}{n}. \end{aligned}$$

Conditional Probability Distribution

Once again, let Ω be a sample space and let $B \subseteq \Omega$ be event an event such that $\mathrm{P}(B) > 0$. Consider the function $\mathrm{P}_B : \Omega \to \mathbb{R}$ such that, for every outcome $x \in \Omega$,

$$\begin{aligned} \mathrm{P}_B(x) &= \mathrm{P}(\{x\} \mid B) \\ &= \frac{\mathrm{P}(\{x\} \cap B)}{\mathrm{P}(B)} \\ &= \begin{cases} \frac{\mathrm{P}(x)}{\mathrm{P}(B)} & \text{if } x \in B, \\ 0 & \text{if } x \notin B. \end{cases} \end{aligned}$$

Conditional Probability Distribution

Once again, let Ω be a sample space and let $B \subseteq \Gamma$ be event an event such that $\mathrm{P}(B) > 0$. Consider the function $\mathrm{P}_B : \Omega \to \mathbb{R}$ such that, for every outcome $x \in \Omega$,

$$\begin{aligned} \mathrm{P}_B(x) &= \mathrm{P}(\{x\} \mid B) \\ &= \frac{\mathrm{P}(\{x\} \cap B)}{\mathrm{P}(B)} \\ &= \begin{cases} \frac{\mathrm{P}(x)}{\mathrm{P}(B)} & \text{if } x \in B, \\ 0 & \text{if } x \notin B. \end{cases} \end{aligned}$$

Conditional Probability Distribution

Continuing our "Balls and Bins" example once again, recall that B is the event

$$\{(\alpha_1, \alpha_2, \ldots, \alpha_m) \mid \alpha_i \in \mathbb{Z} \text{ and } 1 \leq i \leq n \text{ for every integer } i \text{ such that } 1 \leq i \leq m \text{, and such that } \alpha_1 = \alpha_2\}$$

of Ω. It would follow that, for an outcome

$$\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m) \in \Omega,$$

that if $\vec{\alpha} \in B$ then

$$\mathrm{P}_B(\vec{\alpha}) = \frac{\mathrm{P}(\vec{\alpha})}{\mathrm{P}(B)} = \frac{n^{-m}}{n^{-1}} = n^{-(m-1)}.$$

Conditional Probability Distribution

Thus if $\vec{\alpha} \in \Omega$ then

$$\mathsf{P}_B(\vec{\alpha}) = \begin{cases} n^{-(m-1)} & \text{if } \vec{\alpha} \in B, \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Probability Distribution

Claim #1: If Ω, B, and functions $\mathrm{P}, \mathrm{P}_B : \Omega \to \mathbb{R}$ are as above — so that $\mathrm{P}(B) > 0$ — then P_B is a probability distribution.

Proof: Since P_B is a well-defined (total) function from Ω to $\mathbb{R}$, it is necessary and sufficient to show that it satisfies the properties given in the definition of a "probability distribution":

(a) $0 \leq \mathrm{P}_B(x) \leq 1$ for every outcome $x \in \Omega$, and

(b) $\sum_{x \in \Omega} \mathrm{P}_B(x) = 1$.

Conditional Probability Distribution

In order to establish property (a), let $x \in \Omega$. Then either $x \in B$ or $x \notin B$.

- If $x \notin B$ then $\mathrm{P}_B(x) = 0$, by the definition of P_B, so that $0 \leq \mathrm{P}_B(x) \leq 1$ in this case.
- On the other hand, if $x \in B$ then

$$\begin{aligned}
\mathrm{P}(B) &= \sum_{y \in B} \mathrm{P}(y) && \text{(by the definition of } \mathrm{P}(B)\text{)} \\
&= \mathrm{P}(x) + \sum_{y \in B \setminus \{x\}} \mathrm{P}(y) && \text{(since } x \in B\text{)} \\
&\geq \mathrm{P}(x) + \sum_{y \in B \setminus \{x\}} 0 \\
&= \mathrm{P}(x).
\end{aligned}$$

Conditional Probability Distribution

- Since $0 \leq \mathrm{P}(x) \leq \mathrm{P}(B)$, and $\mathrm{P}(B) > 0$,

$$0 \leq \frac{\mathrm{P}(x)}{\mathrm{P}(B)} \leq \frac{\mathrm{P}(B)}{\mathrm{P}(B)} = 1,$$

  that is, $0 \leq \mathrm{P}_B(x) \leq 1$ in this case too.
- Thus property (a) is satisfied in all possible cases.

Conditional Probability Distribution

In order to establish property (b) note that, since $B \subseteq \Omega$,

$$\begin{aligned}
\sum_{x \in \Omega} \mathrm{P}_B(x) &= \sum_{x \in B} \mathrm{P}_B(x) + \sum_{x \in \Omega \setminus B} \mathrm{P}_B(x) && \text{(splitting the sum)} \\
&= \sum_{x \in B} \frac{\mathrm{P}(x)}{\mathrm{P}(B)} + \sum_{x \in \Omega \setminus B} 0 \\
&= \sum_{x \in B} \frac{\mathrm{P}(x)}{\mathrm{P}(B)} \\
&= \frac{1}{\mathrm{P}(B)} \cdot \sum_{x \in B} \mathrm{P}(x) \\
&= \frac{1}{\mathrm{P}(B)} \cdot \mathrm{P}(B) && \text{(by the definition of } \mathrm{P}(B)\text{)} \\
&= 1.
\end{aligned}$$

Conditional Probability Distribution

Since properties (a) and (b) are both satisfied, it follows that the function

$$\mathrm{P}_B : \Omega \to \mathbb{R}$$

is a probability distribution, as claimed. □

Conditional Probability Distribution

Claim #2: Suppose that Ω, B, and the functions $\mathrm{P}, \mathrm{P}_B : \Omega \to \mathbb{R}$ are as given above. If $C \subseteq \Omega$ then

$$\mathrm{P}_B(C) = \mathrm{P}(C \mid B).$$

Proof: Recall that

$$C \setminus B = \{x \in \Omega \mid x \in C \text{ and } x \notin B\}$$

so that

$$(C \cap B) \cup (C \setminus B) = C$$

and

$$(C \cap B) \cap (C \setminus B) = \emptyset.$$

Conditional Probability Distribution

Now

$$
\begin{aligned}
P_B(C) &= \sum_{x \in C} P_B(x) \\
&= \sum_{x \in B \cap C} P_B(x) + \sum_{x \in C \setminus B} P_B(x) && \text{(splitting the sum)} \\
&= \sum_{x \in B \cap C} P_B(x) + \sum_{x \in C \setminus B} 0 && \text{(since } P_B(x) = 0 \text{ if } x \notin B\text{)} \\
&= \sum_{x \in B \cap C} P_B(x) \\
&= \sum_{x \in B \cap C} \frac{P(x)}{P(B)} && \text{(by the definition of } P_B(x)\text{)} \\
&= \frac{1}{P(B)} \cdot \sum_{x \in B \cap C} P(x).
\end{aligned}
$$

Conditional Probability Distribution

Continuing this derivation, we now have

$$\begin{aligned} P_B(C) &= \frac{1}{P(B)} \cdot \sum_{x \in B \cap C} P(x) \\ &= \frac{1}{P(B)} \cdot P(B \cap C) \\ &= \tfrac{P(B \cap C)}{P(B)} \\ &= P(C \mid B) \end{aligned}$$

as required to establish the claim. □

Conditional Probability Distribution

From now on, we will call the above probability distribution, P_B, the ***conditional probability distribution (defined from P) conditional on event B.***

- We will leave out "(defined from P)" when it is clear, from context, what the probability distribution, P, would be.

Conditional Probability Distribution

A variety of properties of *conditional probabilities* can be established using the fact that the conditional probability distribution is, indeed, a "probability distribution" itself.

Example: Recall, from Theorem #1 in the previous lecture, that if $A \subseteq \Omega$, for a sample space Ω, then the probability of the complement, $\overline{A}$, of the event A, is

$$\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A).$$

Now let $B \subseteq \Omega$ be an event such that $\mathrm{P}(B) > 0$.

Conditional Probability Distribution

Applying Theorem #1 from the previous lecture — with the conditional probability distribution P_B in place of the probability distribution P — we have that

$$\mathrm{P}_B(\overline{A}) = 1 - \mathrm{P}_B(A).$$

Now applying Claim #2 above — with each of the events A and $\overline{A}$ used in place of C — we see that

$$\mathrm{P}_B(\overline{A}) = \mathrm{P}(\overline{A} \mid B) \qquad \text{and} \qquad \mathrm{P}_B(A) = \mathrm{P}(A \mid B).$$

It now follows, by the above, that

$$\mathrm{P}(\overline{A} \mid B) = 1 - \mathrm{P}(A \mid B). \tag{1}$$

Conditional Probability Distribution

Another Example: Recall, by Theorem #2 from the previous lecture, that if $\mathrm{P} : \Omega \to \mathbb{R}$ is a probability distribution for a sample space Ω, and $A, B \subseteq \Omega$, then

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B).$$

- This property is often called the ***Inclusion-Exclusion Principle.***

Conditional Probability Distribution

Now let $C \subseteq \Omega$ be an event such that $\mathrm{P}(C) > 0$, so that the function $\mathrm{P}_C : \Omega \to \mathbb{R}$ is a probability distribution for Ω, as well.

- Applying the Inclusion-Exclusion Principle, with this probability distribution, we may now conclude that

$$\mathrm{P}_C(A \cup B) = \mathrm{P}_C(A) + \mathrm{P}_C(B) - \mathrm{P}_C(A \cap B)$$

— that is, (by Claim #2)

$$\mathrm{P}(A \cup B \,|\, C) = \mathrm{P}(A \,|\, C) + \mathrm{P}(B \,|\, C) - \mathrm{P}(A \cap B \,|\, C). \qquad (2)$$

The Law of Total Probability

Claim #3 (Law of Total Probability): Let Ω be a sample space and let $\mathrm{P} : \Omega \to \mathbb{R}$ be a probability distribution. Then, for any events A and B,

$$\mathrm{P}(A) = \mathrm{P}(A \mid B) \cdot \mathrm{P}(B) + \mathrm{P}(A \mid \overline{B}) \cdot \mathrm{P}(\overline{B}).$$

The Law of Total Probability

- $\mathrm{P}(A \mid B)$ has not been defined when $\mathrm{P}(B) = 0$ — but let us consider

$$\mathrm{P}(A \mid B) \cdot \mathrm{P}(B)$$

  to be 0 for any event A, whenever B is an event such that $\mathrm{P}(B) = 0$.

- The proof of Claim #3 is left as an ***exercise***. When completing this, it might be helpful to remember that

$$(A \cap B) \cup (A \cap \overline{B}) = A$$

  and

$$(A \cap B) \cap (A \cap \overline{B}) = \emptyset$$

  for all events $A, B \subseteq \Omega$.

Extended Partition Rule

The following is a generalization of the law of total probability.

Claim #4 (Extended Partition Rule): Let Ω be a sample space, let $\mathrm{P} : \Omega \to \mathbb{R}$ be a probability distribution, let k be a positive integer, and let A and $B_1, B_2, \ldots, B_k$ be events satisfying the following properties.

(a) $B_1, B_2, \ldots, B_k$ are *pairwise disjoint*. That is, $B_i \cap B_j = \emptyset$ for all integers i and j such that $1 \leq i, j \leq k$ and $i \neq j$.

(b) $A \subseteq B_1 \cup B_2 \cup \cdots \cup B_k$.

Then

$$\mathrm{P}(A) = \mathrm{P}(A \mid B_1) \cdot \mathrm{P}(B_1) + \mathrm{P}(A \mid B_2) \cdot \mathrm{P}(B_2) + \cdots + \mathrm{P}(A \mid B_k) \cdot \mathrm{P}(B_k).$$

Extended Partition Rule

- Suppose that A and $B_1, B_2, \ldots, B_k$ are as in the statement of the claim. Note that if $k \geq 3$ and

  $$\widetilde{B}_1 = B_2 \cup B_3 \cup \cdots \cup B_k$$

  then $B_1 \cap \widetilde{B}_1 = \emptyset$, since $B_i \cap B_j = \emptyset$ for every integer j such that $2 \leq k$ — and

  $$B_1 \cup \widetilde{B}_1 = B_1 \cup B_2 \cup \cdots \cup B_k$$

  — so that $A \cup B_1 \cup \widetilde{B}_1$.
- ***Exercise:*** Using the above, prove Claim #4 by induction on k.

Baye's Theorem

Claim #5 (Baye's Theorem): Let Ω be a sample space, let $\mathsf{P} : \Omega \to \mathbb{R}$, and let A and B be events such that $\mathsf{P}(A) > 0$ and $\mathsf{P}(B) > 0$. Then

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(B \mid A) \cdot \mathsf{P}(A)}{\mathsf{P}(B)}.$$

Proof: Recall that (by definition)

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

and

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

Baye's Theorem

Thus

$$\begin{aligned}
\frac{\mathrm{P}(B \mid A) \cdot \mathrm{P}(A)}{\mathrm{P}(B)} &= \frac{(\mathrm{P}(B \cap A)/\mathrm{P}(A)) \cdot \mathrm{P}(A)}{\mathrm{P}(B)} \\
&= \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(B)} \\
&= \mathrm{P}(A \mid B),
\end{aligned}$$

as claimed. □

Independence

Consider a sample space Ω, and probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, and a pair of events $A, B \subseteq \Omega$ such that $\mathrm{P}(B) > 0$.

- A is said to be ***attracted*** to B (under P) if $\mathrm{P}(A \mid B) > \mathrm{P}(A)$.
- A is said to be ***repelled*** by B (under P) if $\mathrm{P}(A \mid B) < \mathrm{P}(A)$.
- A is said to be ***indifferent*** to B (under P) if $\mathrm{P}(A \mid B) = \mathrm{P}(A)$.

We will leave out the phrase “under P” if the probability distribution being used is clear.

These technical terms are useful, but somewhat obscure.[1] The related term, defined next, is more widely used.

[1] I discovered them, for the first time, when looking at one of the introductions to probability theory that I discovered when preparing these lecture notes.

Independence

Events A and B are said to be ***independent*** if

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B).$$

- Note that if $\mathrm{P}(B) > 0$ then A and B are ***independent*** if and only if A is ***indifferent*** to B.

Independence

Example: Recall the example of “Shuffling a Deck of Playing Cards” from Lecture #18. In this case

- There are 52 ways to choose the first card (in the new order),
- Since the first card is no longer available, there are 51 ways to choose the second card,
- Since the first two cards are no longer available, there are 50 ways to choose the third card,

and so on, so that there are

$$52 \times 51 \times 50 \times \cdots \times 2 \times 1 = 52!$$

ways to shuffle the card in this deck — and the sample space, Ω, for this experiment has size $52! = 52 \times 51!$.

Independence

Consider the uniform distribution — so that we will use a probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$ such that

$$\mathrm{P}(x) = \tfrac{1}{52!}$$

for every outcome $x \in \Omega$. Consider, as well, the following events.

- A: The first card in the new order is a ***heart***.
- B: The *second* card in the new order is a ***heart***.

Independence

Consider the outcomes in Σ that belong to the event A.

- Since only 13 cards in the deck are hearts, there are now 13 ways to choose the first card.
- Once again, since the first card is no longer available, there are 51 ways to choose the second card. There are, then, 50 ways to choose the third card, and so on.

Thus the number of outcomes in event A is

$$|A| = 13 \times 51 \times 50 \times 49 \times \ldots 2 \times 1 = 13 \times 51!$$

so that (since P is the uniform distribution)

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|} = \frac{13 \times 51!}{52 \times 51!} = \frac{13}{52} = \frac{1}{4}.$$

Independence

Consider the outcomes in Σ that belong to the event B.

- ***Exercise:*** Prove that $\mathrm{P}(B) = \frac{1}{4}$, as well.
- You may be able to find more than one way to show this.
- If time permits, this will be discussed during the lecture presentation.

Independence

Consider the outcomes in Σ that belong to the event $A \cap B$ — so that the first two cards in the selected ordering are both hearts.

- Since 13 cards in the deck are hearts, there are 13 ways to choose the first card.
- Since 12 cards that are hearts are left over, after the first card has been chosen, there are 12 ways to choose the second card.
- There are 50 cards left after the first two cards have been chosen, so there are 50 ways to choose the third card.
- There are then 49 ways to choose the fourth card, 48 ways to choose the fifth card, and so on.

Independence

- Thus the number of outcomes in the event $A \cap B$ is

$$|A \cap B| = 13 \times 12 \times 50 \times 49 \times \cdots \times 2 \times 1 = 13 \times 12 \times 50!.$$

- Since P is the uniform probability distribution,

$$\mathrm{P}(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{13 \times 12 \times 50!}{52 \times 51 \times 50!} = \frac{1}{17}.$$

Independence

- Thus
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/17}{1/4} = \frac{4}{17}.$$
- Since
$$P(A) = \frac{1}{4} = \frac{4}{16} > \frac{4}{17} = P(A \mid B),$$
it follows that event A is ***repelled by*** event B.

Independence

Another Exercise: Consider the following events as well.

- C: The second card in the new order is a ***spade***.
- D: The second card in the new order is an ***ace*** (remembering that “ace” is not the name of a *card suit* like “heart” or “spade” is).

(a) Show that event A is ***attracted to*** event C.

(b) Show that event A is ***indifferent to*** event D. That is, show that events A and D are ***independent***.

Fixing Partial Random Choices

Informal Description

Suppose that a sample space Ω and probability distribution $\text{P} : \Omega \to \mathbb{R}$ is being used to model an experiment where results have several components — so that "partial results" can be considered.

- Suppose that one part of the result can be set — in every possible way — without the (conditional) probability of a given event A being changed.
- Then this part of the result can just be "arbitrarily set" in any way that you want to.The conditional probability of A, for this setting, will be the same as the (unconditional) probability of A, no matter which setting you picked.

Fixing Partial Random Choices

Formal Description

Claim #6: Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \rightarrow \mathbb{R}$. Let A be an event, and let $B_1, B_2, \ldots, B_k$ be pairwise disjoint events for a positive integer k (so that $B_i \cap B_j = \emptyset$ for all integers i and j such that $1 \leq i, j \leq k$ and $i \neq k$) such that

$$B_1 \cup B_2 \cup \cdots \cup B_k = \Omega.$$

Finally, suppose that

$$\mathrm{P}(A \mid B_i) = \mathrm{P}(A \mid B_2) = \cdots = \mathrm{P}(A \mid B_k) = q$$

for some real number q.

Then $\mathrm{P}(A) = q$ as well — so that the events A and B_i are ***independent*** for every integer i such that $1 \leq i \leq k$.

Fixing Partial Random Choices

Proof: It is given that $B_1, B_2, \ldots, B_k$ are pairwise disjoint and, since $B_1 \cup B_2 \cup \cdots \cup B_k = \Omega$, $A \subseteq B_1 \cup B_2 \cup \cdots \cup B_k$. It now follows by Claim #4, above, that

$$\begin{aligned} \mathrm{P}(A) &= \mathrm{P}(A \mid B_1) \cdot \mathrm{P}(B_1) + \mathrm{P}(A \mid B_2) \cdot \mathrm{P}(B_2) + \\ &\qquad \cdots + \mathrm{P}(A \mid B_k) \cdot \mathrm{P}(B_k) \\ &= q \cdot \mathrm{P}(B_1) + q \cdot \mathrm{P}(B_2) + \cdots + q \cdot \mathrm{P}(B_k) \\ &= q \cdot (\mathrm{P}(B_1) + \mathrm{P}(B_2) + \cdots + \mathrm{P}(B_k)) \end{aligned}$$

since $\mathrm{P}(A \mid B_i) = q$ for every integer i such that $1 \leq i \leq k$.

Fixing Partial Random Choices

Now, since $B_1, B_2, \ldots, B_k$ are pairwise disjoint, and since $B_1 \cup B_2 \cup \cdots \cup B_k = \Omega$, it can be established by induction on k that

$$\begin{aligned} \mathrm{P}(B_1) + \mathrm{P}(B_2) + \cdots + \mathrm{P}(B_k) &= \mathrm{P}(B_1 \cup B_2 \cup \cdots \cup B_k) \\ &= \mathrm{P}(\Omega) = 1 \end{aligned}$$

so that

$$\begin{aligned} \mathrm{P}(A) &= q \cdot (\mathrm{P}(B_1) + \mathrm{P}(B_2) + \cdots + \mathrm{P}(B_k)) \\ &= q \cdot 1 = q \end{aligned}$$

as claimed. Thus $\mathrm{P}(A) = \mathrm{P}(A \mid B_i) = q$ as well, and the events A and B_i are independent, for every integer i such that $1 \leq i \leq k$. □

Fixing Partial Random Choices

Example: Suppose that we are tossing a fair coin, three times, so that the sample space is

$$\Omega = \{(\mathtt{H},\mathtt{H},\mathtt{H}),(\mathtt{H},\mathtt{H},\mathtt{T}),(\mathtt{H},\mathtt{T},\mathtt{H}),(\mathtt{H},\mathtt{T},\mathtt{T}),\\ (\mathtt{T},\mathtt{H},\mathtt{H}),(\mathtt{T},\mathtt{H},\mathtt{T}),(\mathtt{T},\mathtt{T},\mathtt{H}),(\mathtt{T},\mathtt{T},\mathtt{T})\}.$$

Since we are tossing a *fair* coin the uniform probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$ is being used. That is, $\mathsf{P}(\sigma) = \frac{1}{|\Omega|} = \frac{1}{8}$ for every outcome $\sigma \in \Omega$.

Fixing Partial Random Choices

Consider the following events.

- B_1 is the event that the first toss is "$\mathtt{H}$" and the second toss is "$\mathtt{H}$", so that
$$B_1 = \{(\mathtt{H}, \mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{H}, \mathtt{T})\}.$$
- B_2 is the event that the first toss is "$\mathtt{H}$" and the second toss is "$\mathtt{T}$", so that
$$B_2 = \{(\mathtt{H}, \mathtt{T}, \mathtt{H}), (\mathtt{H}, \mathtt{T}, \mathtt{T})\}.$$
- B_3 is the event that the first toss is "$\mathtt{T}$" and the second toss is "$\mathtt{H}$", so that
$$B_3 = \{(\mathtt{T}, \mathtt{H}, \mathtt{H}), (\mathtt{T}, \mathtt{H}, \mathtt{T})\}.$$
- B_4 is the event that the first toss is "$\mathtt{T}$" and the second toss is "$\mathtt{T}$", so that
$$B_4 = \{(\mathtt{T}, \mathtt{T}, \mathtt{H}), (\mathtt{T}, \mathtt{T}, \mathtt{T})\}.$$

Fixing Partial Random Choices

Note that the events B_1, B_2, B_3 and B_4 are pairwise disjoint and that

$$B_1 \cup B_2 \cup B_3 \cup B_4 = \Omega$$

so that this set of events satisfy the conditions given (for "$B_1, B_2, \ldots, B_k$") in Claim #6.

Now consider the following event as well.

- *A* is the even that an *even* number of heads ("H") were tossed.

Fixing Partial Random Choices

Let us confirm that this event satisfies the conditions (for "A") in Claim #6.

- $A \cap B_1 = \{(\mathtt{H}, \mathtt{H}, \mathtt{T})\}$, so that

$$P(A \mid B_1) = \frac{P(A \cap B_1)}{P(B_1)} = \frac{|A \cap B_1|}{|B_1|} = \frac{1}{2}.$$

- $A \cap B_2 = \{(\mathtt{H}, \mathtt{T}, \mathtt{H})\}$, so that

$$P(A \mid B_2) = \frac{P(A \cap B_2)}{P(B_2)} = \frac{|A \cap B_2|}{|B_2|} = \frac{1}{2}.$$

Fixing Partial Random Choices

- $A \cap B_3 = \{(\mathtt{T}, \mathtt{H}, \mathtt{H})\}$, so that

$$\mathrm{P}(A \mid B_3) = \frac{\mathrm{P}(A \cap B_3)}{\mathrm{P}(B_3)} = \frac{|A \cap B_3|}{|B_3|} = \frac{1}{2}.$$

- $A \cap B_4 = \{(\mathtt{T}, \mathtt{T}, \mathtt{T})\}$, so that

$$\mathrm{P}(A \mid B_4) = \frac{\mathrm{P}(A \cap B_4)}{\mathrm{P}(B_4)} = \frac{|A \cap B_4|}{|B_4|} = \frac{1}{2}.$$

Setting $q = \frac{1}{2}$ we see that $\mathrm{P}(A \mid B_i) = q$ for every integer i such that $1 \leq i \leq 4$. It follows, by the above claim, that $\mathrm{P}(A) = q = \frac{1}{2}$ as well, and that the events A and B_i are independent, for every integer i such that $1 \leq i \leq 4$.

Mutual Independence

Once again, let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Let $A_1, A_2, \ldots, A_k \subseteq \Omega$ be events, for some integer $k \geq 2$.

Definition: The events $A_1, A_2, \ldots, A_k$ are ***mutually independent*** if

$$\mathrm{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathrm{P}(A_i) \tag{3}$$

for *every* S of $\{1, 2, \ldots, k\}$.

Note: The condition at line (3) is guaranteed to hold whenever $|S| \leq 1$, so this condition only needs to be considered when $|S| \geq 2$.

Mutual Independence

For example, if $k = 3$ and events A_1, A_2, A_3 are mutually independent, then each of the following equations is satisfied.

- $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \times \mathrm{P}(A_2)$.
- $\mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_3)$.
- $\mathrm{P}(A_2 \cap A_3) = \mathrm{P}(A_2) \times \mathrm{P}(A_3)$.
- $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_2) \times \mathrm{P}(A_3)$.

Mutual Independence

For example, if $k = 4$ and events A_1, A_2, A_3, A_4 are mutually independent, then each of the equations, shown on this and the next slide, is satisfied.

- $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \times \mathrm{P}(A_2)$.
- $\mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_3)$.
- $\mathrm{P}(A_1 \cap A_4) = \mathrm{P}(A_1) \times \mathrm{P}(A_4)$.
- $\mathrm{P}(A_2 \cap A_3) = \mathrm{P}(A_2) \times \mathrm{P}(A_3)$.
- $\mathrm{P}(A_2 \cap A_4) = \mathrm{P}(A_2) \times \mathrm{P}(A_4)$.
- $\mathrm{P}(A_3 \cap A_4) = \mathrm{P}(A_3) \times \mathrm{P}(A_4)$.

Mutual Independence

- $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$.
- $P(A_1 \cap A_2 \cap A_4) = P(A_1) \times P(A_2) \times P(A_4)$.
- $P(A_1 \cap A_3 \cap A_4) = P(A_1) \times P(A_3) \times P(A_4)$.
- $P(A_2 \cap A_3 \cap A_4) = P(A_2) \times P(A_3) \times P(A_4)$.
- $P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1) \times P(A_2) \times P(A_3) \times P(A_4)$.

Mutual Independence

Example: Once again, consider the experiment of tossing a fair coin three times, so that the sample space is

$$\Omega = \{(\mathtt{H},\mathtt{H},\mathtt{H}), (\mathtt{H},\mathtt{H},\mathtt{T}), (\mathtt{H},\mathtt{T},\mathtt{H}), (\mathtt{H},\mathtt{T},\mathtt{T}), (\mathtt{T},\mathtt{H},\mathtt{H}), (\mathtt{T},\mathtt{H},\mathtt{T}), (\mathtt{T},\mathtt{T},\mathtt{H}), (\mathtt{T},\mathtt{T},\mathtt{T})\}$$

and the probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$ is the uniform probability distribution — so that $\mathrm{P}(\sigma) = \frac{1}{|\Omega|} = \frac{1}{8}$ for each outcome $\sigma \in \Omega$.

Consider the following events.

- A_1: First coin toss is "$\mathtt{H}$".
- A_2: Second coin toss is "$\mathtt{H}$".
- A_3: Third coin toss is "$\mathtt{H}$".

Mutual Independence

Now

$$A_1 = \{(\mathtt{H}, \mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{H}, \mathtt{T}), (\mathtt{H}, \mathtt{T}, \mathtt{H}), (\mathtt{H}, \mathtt{T}, \mathtt{T})\},$$

so that

$$\mathrm{P}(A_1) = \frac{|A_1|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}.$$

Exercise: Confirm that $|A_2| = |A_3| = 4$, so that

$$\mathrm{P}(A_2) = \mathrm{P}(A_3) = \frac{1}{2}$$

as well.

In order to confirm that the events A_1, A_2 and A_3 are mutually independent we must establish properties (a), (b), (c) and (d), below.

Mutual Independence

(a) $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \times \mathrm{P}(A_2)$.

To see that this is true, note that $A_1 \cap A_2$ is the event "both the first and second coin tosses are $\mathtt{H}$", so that

$$A_1 \cap A_2 = \{(\mathtt{H}, \mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{H}, \mathtt{T})\}$$

and

$$\mathrm{P}(A_1 \cap A_2) = \frac{|A_1 \cap A + 2|}{|\Omega|} = \frac{2}{8} = \frac{1}{4},$$

while

$$\mathrm{P}(A_1) \times \mathrm{P}(A_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

as well. Thus $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \times \mathrm{P}(A_2)$, as required.

Mutual Independence

(b) $\mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_3)$.

(c) $\mathrm{P}(A_2 \cap A_3) = \mathrm{P}(A_2) \times \mathrm{P}(A_3)$.

Exercise: Modify the proof of property (a) to obtain similar proofs of properties (b) and (c).

Mutual Independence

(d) $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_2) \times \mathrm{P}(A_3)$.

To see that this is true, not that $A_1 \cap A_2 \cap A_3$ is the event "all three coin tosses are H", so that

$$A_1 \cap A_2 \cap A_3 = \{(\mathtt{H}, \mathtt{H}, \mathtt{H})\}$$

and

$$\mathrm{P}(A_1 \cap A_2 \cap) = \frac{|A_1 \cap A_2 \cap A_3|}{|\Omega|} = \frac{1}{8},$$

while

$$\mathrm{P}(A_1) \times \mathrm{P}(A_2) \times \mathrm{P}(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

as well. Thus $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_2) \times \mathrm{P}(A_3)$ as required.

Mutual Independence

Since all relevant conditions are satisfied, it follows that the events A_1, A_2 and A_3 are mutually independent.

Pairwise Independence

Once again, let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Let $A_1, A_2, \ldots, A_k \subseteq \Omega$ be events, for some integer $k \geq 2$.

Definition: The events $A_1, A_2, \ldots, A_k$ are ***pairwise independent*** if

$$\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i) \times \mathrm{P}(A_j) \tag{4}$$

for every pair of integers i and j such that $1 \leq i, j \leq k$ and $i \neq j$.

Note: Since

$$A_i \cap A_j = A_j \cap A_i \quad \text{and} \quad \mathrm{P}(A_i) \times \mathrm{P}(A_j) = \mathrm{P}(A_j) \times \mathrm{P}(A_i)$$

whenever $1 \leq i, j \leq k$ and $i \neq j$, it is sufficient to check the condition at line (4), for integers i and j such that $1 \leq i < j \leq k$, when pairwise independence is being checked.

Pairwise Independence

For example, if $k = 3$ and events A_1, A_2, A_3 are pairwise independent, then each of the following equations is satisfied.

- $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$.
- $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$.
- $P(A_2 \cap A_3) = P(A_2) \times P(A_3)$.

Pairwise Independence

For example, if $k = 4$ and events A_1, A_2, A_3, A_4 are pairwise independent, then each of the following equations is satisfied.

- $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$.
- $P(A_1 \cap A_3) = P(A_1) \times P(A_3)$.
- $P(A_1 \cap A_4) = P(A_1) \times P(A_4)$.
- $P(A_2 \cap A_3) = P(A_2) \times P(A_3)$.
- $P(A_2 \cap A_4) = P(A_2) \times P(A_4)$.
- $P(A_3 \cap A_4) = P(A_3) \times P(A_4)$.

Pairwise Independence

Example: Consider the experiment of tossing a fair coin *two* times — so that the sample space is

$$\Omega = \{(\mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{T}), (\mathtt{T}, \mathtt{H}), (\mathtt{T}, \mathtt{T})\}$$

and $\mathsf{P} : \omega \to \mathbb{R}$ is the uniform probability distribution, so that $\mathsf{P}(\sigma) = \frac{1}{|\Omega|} = \frac{1}{4}$ for every outcome $\sigma \in \Omega$.

Pairwise Independence

Consider the following events:

- A_1 is the event “the first toss is H”, so that

$$A_1 = \{(\mathtt{H}, \mathtt{H}), (\mathtt{H}, \mathtt{T})\}.$$

- A_2 is the event “the second toss is H”, so that

$$A_2 = \{(\mathtt{H}, \mathtt{H}), (\mathtt{T}, \mathtt{H})\}.$$

- A_2 is the event “the tosses are the same”, so that

$$A_3 = \{(\mathtt{H}, \mathtt{H}), (\mathtt{T}, \mathtt{T})\}.$$

Pairwise Independence

- Then
$$P(A_i) = \frac{|A_i|}{|\Omega|} = \frac{2}{4} = \frac{1}{2}$$
for each integer i such that $1 \leq i \leq 3$.
- In order to confirm that the events A_1, A_2 and A_3 are pairwise independent, it is necessary confirm that properties (a), (b) and (c), as given on the next slides, are all satisfied.

Pairwise Independence

(a) $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \times \mathrm{P}(A_2)$.

To see that this is true, note that

$$A_1 \cap A_2 = \{(\mathtt{H}, \mathtt{H})\}$$

so that

$$\mathrm{P}(A_1 \cap A_2) = \frac{|A_1 \cap A_2|}{|\Omega|} = \frac{1}{4},$$

while

$$\mathrm{P}(A_1) \times \mathrm{P}(A_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

Thus $\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \times \mathrm{P}(A_2)$, as required.

Pairwise Independence

Exercise: Modify the above argument to show that each of the following properties is satisfied, as well.

(b) $\mathrm{P}(A_1 \cap A_3) = \mathrm{P}(A_1) \times \mathrm{P}(A_3)$.

(c) $\mathrm{P}(A_2 \cap A_3) = \mathrm{P}(A_2) \times \mathrm{P}(A_3)$.

Since these properties are all satisfied, the above events A_1, A_2 and A_3 are all pairwise independent.

Mutual Independence and Pairwise Independence

Note: In order for these events to be ***mutually independent*** as well, it must also be true that

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3).$$

Now

$$A_1 \cap A_2 \cap A_3 = \{(\mathtt{H}, \mathtt{H})\},$$

so that

$$P(A_1 \cap A_2 \cap A_3) = \frac{|A_1 \cap A_2 \cap A_3|}{|\Omega|} = \frac{1}{4},$$

while

$$P(A_1) \times P(A_2) \times P(A_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}.$$

Thus $P(A_1 \cap A_2 \cap A_3) \neq P(A_1) \times P(A_2) \times P(A_3)$, and the events A_1, A_2 and A_3 are ***not*** mutually independent.

Mutual Independence and Pairwise Independence

- A comparison of the definitions of these terms is sufficient to confirm that, for every integer $k \geq 1$ and all events $A_1, A_2, \ldots, A_k \subseteq \Omega$,

  “if $A_1, A_2, \ldots, A_k$ are mutually independent then $A_1, A_2, \ldots, A_k$ are pairwise independent”.

- The above example shows, though, that pairwise independence *does not* always imply mutual independence.

Mutual Independence and Pairwise Independence

- Some references say that "$A_1, A_2, \ldots A_k$ are independent" when these events are ***mutually independent***, as defined above.
- The word "independent" will not be used in this way, in this course, because "pairwise independence" is also a useful property — which is not the same as "mutual independence", as shown above.