Computer Science 351

Reintroduction to Discrete Probability Theory

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #19

Learning Goal

Learning Goal:

- Review material about ***probability distributions*** from an introduction to ***discrete probability theory***.
- Extend this with by introducing material (which you *might* have seen before) concerning ***conditional probabilities*** and the ***independence of events***.

Experiments — Definition and Classical Examples

An ***experiment*** is a procedure (or process) that yields one of a given set of possible ***outcomes***. The set of possible outcomes of the experiment — which we will often name Ω — is called the ***sample space***.

Classical Examples

- Tossing a coin once
- Tossing a coin, for a fixed number of times
- Rolling a die[1] once
- Rolling a die, for a fixed number of times
- Shuffling a deck of playing cards

[1] This is the singular form of the word *dice.*

Tossing a Coin Once

Example: Tossing a Coin Once

- In this case — since we won't worry about the coin landing on its edge — there are two ***outcomes***, *heads* or *tails*. In the future we will represent these as $\mathtt{H}$ and $\mathtt{T}$ respectively.
- The ***sample space*** is, then, the set

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

with size two.

Example: Tossing a Coin for a Fixed Number of Times

- For example, if $k = 3$, then

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

  so that $|\Omega| = 2^3 = 8$ in this case.
- In general (that is, for arbitrary k), $|\Omega| = 2^k$.

Discrete Probability Theory

- In most of the examples in this course, sample spaces will be ***finite***.
- Experiments with *infinite* same spaces can sometimes be useful too. We will say that a set Ω is ***countable*** if there is a total function

$$f : \mathbb{N} \to \Omega$$

  that is ***surjective*** (or "onto":) For every value $x \in \Omega$ there exists a natural number i such that $f(i) = x$.
- Every finite set is countable. Some — but not all — infinite sets are countable too.
- In this course we will (almost always) consider experiments where the sample space, Ω, is countable — studying a part of probability theory called ***discrete probability theory***.

Events

When we consider experiments (modelled by sets of outcomes, called sample spaces) we are often interested in various *properties* or *things that can happen*.

- Whether or not such a property is satisfied generally depends on the experiment's *outcome*.
- An ***event*** is a subset of the experiment's sample space Ω. Events are used to model the kinds of "things that are interested in" that will be studied.
- An ***elementary event*** is an event with size one — that is, it is an event that only includes a single outcome.

Probability Distributions

Consider an experiment with sample space Ω. A ***probability distribution*** is a (total) function

$$\mathsf{P} : \Omega \to \mathbb{R}$$

such that $0 \leq \mathsf{P}(x) \leq 1$ for every outcome $x \in \Omega$, and such that

$$\sum_{x \in \Omega} \mathsf{P}(x) = 1.$$

Uniform Distributions

If Ω is a finite set then the ***uniform probability distribution*** (for Ω) defines the probability of every outcome to be the same: This is the function $\mathsf{P} : \Omega \to \mathbb{R}$ such that

$$\mathsf{P}(x) = \frac{1}{|\Omega|}$$

for every outcome $x \in \Omega$.

Exercise: Prove that this function is a probability distribution (for an experiment with same space Ω).

Probabilities of Events

For any set Ω, let $\mathcal{P}(\Omega)$ denote the set of all ***subsets*** of Ω.

- ***Example:*** If $\Omega = \{1, 2, 3\}$ then

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

  where $\emptyset = \{\}$ is the ***empty set*** — so that

$$|\mathcal{P}(\Omega)| = 8 = 2^3 = 2^{|\Omega|}.$$

- $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$ for ***every*** finite set Ω.
- Thus, if Ω is a sample space for an experiment, then $\mathcal{P}(\Omega)$ is the set of all ***events*** (for this experiment).

Probabilities of Events

A probability distribution P (on an experiment with a countable sample space) is "extended" to get a function

$$\mathsf{P} : \mathcal{P}(\Omega) \to \mathbb{R}$$

by setting

$$\mathsf{P}(A) = \sum_{x \in A} \mathsf{P}(x)$$

for every event $A \subseteq \Omega$ (that is, for all $A \in \mathcal{P}(\Omega)$).

Probabilities of Events: Uniform Distributions

Suppose that $A \subseteq \Omega$ is an event. Then, if P is the uniform distribution for Ω, then

$$\begin{aligned} \mathsf{P}(A) &= \sum_{x \in A} \mathsf{P}(x) \\ &= \sum_{x \in A} \frac{1}{|\Omega|} \\ &= \frac{1}{|\Omega|} \sum_{x \in A} 1 \\ &= \frac{1}{|\Omega|} \cdot |A| \\ &= \frac{|A|}{|\Omega|}. \end{aligned}$$

Example: Tossing a Coin Three Times

- ***Example:*** Consider, again, the experiment "tossing a coin, three times". The corresponding sample space is a finite set, Ω, such that $|\Omega| = 2^3 = 8$.
- As noted above, the event "$\mathtt{H}$ is tossed more often than $\mathtt{T}$", is the subset

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

- It follows that — ***assuming the uniform probability distribution*** — the probability that $\mathtt{H}$ is tossed more often than $\mathtt{T}$ is

$$\frac{|A|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}.$$

Example: Tossing a Coin Three Times

Example: *Tossing a biased coin, for a fixed number of times.*

- Suppose, once again, that we toss a coin three times, so that

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

- *This time,* suppose that *tossing heads is more likely than tossing tails* — so that we are now using a *different* probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$.

Example: Tossing a Coin Three Times

- Suppose, in particular, that
  - $\mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{H})) = \frac{8}{27}$.
  - $\mathrm{P}((\mathtt{H},\mathtt{H},\mathtt{T})) = \mathrm{P}((\mathtt{H},\mathtt{T},\mathtt{H})) = \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{H})) = \frac{4}{27}$.
  - $\mathrm{P}((\mathtt{H},\mathtt{T},\mathtt{T})) = \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{T})) = \mathrm{P}((\mathtt{T},\mathtt{H},\mathtt{H})) = \frac{2}{27}$.
  - $\mathrm{P}((\mathtt{T},\mathtt{T},\mathtt{T})) = \frac{1}{27}$.
- Note that P is a total function from Ω to $\mathbb{R}$ and that $0 \leq \mathrm{P}(x) \leq 1$ for every outcome $x \in \Omega$.
- ***Exercise:*** Confirm that $\sum_{x \in \Omega} \mathrm{P}(x) = 1$.
- Since it is certainly not the "uniform probability distribution", it follows that P is an example of a ***nonuniform probability distribution*** for this experiment.

Example: Tossing a Coin Three Times

- Consider the event "H is tossed more often than T", that is, the event

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

- Under this nonuniform probability distribution the probability that H is tossed more often than T is

$$\begin{aligned}\mathsf{P}(A) &= \sum_{x\in A}\mathsf{P}(x)\\ &= \mathsf{P}((\mathtt{H},\mathtt{H},\mathtt{H})) + \mathsf{P}((\mathtt{H},\mathtt{H},\mathtt{T})) + \mathsf{P}((\mathtt{H},\mathtt{T},\mathtt{H}))\\ &\qquad + \mathsf{P}((\mathtt{T},\mathtt{H},\mathtt{H}))\\ &= \frac{8}{27} + \frac{4}{27} + \frac{4}{27} + \frac{4}{27}\\ &= \frac{20}{27}\end{aligned}$$

— while the probability of this event under the uniform probability distribution for this experiment was $\frac{1}{2}$.

Probability of the Complement of an Event

If Ω is a sample space for an experiment and $A \subseteq \Omega$ is an event, then the ***complement***[2] of the event A, $\overline{A}$, is the set of outcomes that *are not* in A.

$$\overline{A} = \{x \in \Omega \mid x \notin A\}.$$

Theorem #1: Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, and let $A \subseteq \Omega$. Then the probability of the complement, $\overline{A}$, of the event A is

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

[2] It is also OK if you use A^C to represent the complement of A, as we did for *languages*. These are both commonly used as the name for this set.

Probability of the Union of Events

Theorem #2: Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Then, for any events $A, B \subseteq \Omega$,

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

Since $\mathrm{P}(A \cap B) \geq 0$ for all events $A, B \subseteq \Omega$, Theorem #2 implies the following.

Corollary #3: Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. Then, for any events $A, B \subseteq \Omega$,

$$\mathrm{P}(A \cup B) \leq \mathrm{P}(A) + \mathrm{P}(B).$$

Probability of the Union of Events

This can be used to establish the following more general result.

Theorem #4 (Union Bound): Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$, let k be a positive integer, and let $E_1, E_2, \ldots, E_k \subseteq \Omega$. Then

$$\mathrm{P}(E_1 \cup E_2 \cup \cdots \cup E_k) \leq \sum_{i=1}^{k} \mathrm{P}(E_i).$$

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 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

Theorem #5: If Ω, B, and functions $\mathsf{P}, \mathsf{P}_B : \Omega \to \mathbb{R}$ are as above — so that $\mathsf{P}(B) > 0$ — then P_B is a probability distribution.

Theorem #6: Suppose that Ω, B, and the functions $\mathsf{P}, \mathsf{P}_B : \Omega \to \mathbb{R}$ are as given above. If $C \subseteq \Omega$ then

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

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, above, 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 this result — 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 Theorem #6 above — with each of the events A and $\overline{A}$ used in place of C — we see that

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

It now follows, by the above, that

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

Conditional Probability Distribution

Another Example: Recall, by Theorem #2, above, 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 $\mathsf{P}(C) > 0$, so that the function $\mathsf{P}_C : \Omega \to \mathbb{R}$ is a probability distribution for Ω, as well.

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

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

— that is, (by Theorem #6)

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

The Law of Total Probability

Theorem #7 (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 Theorem #7 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.

Theorem #8 (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

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

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 Theorem #8 by induction on k.

Baye's Theorem

Theorem #9 (Baye's Theorem): Let Ω be a sample space, let $\mathrm{P} : \Omega \to \mathbb{R}$, and let A and B be events such that $\mathrm{P}(A) > 0$ and $\mathrm{P}(B) > 0$. Then

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

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.[3] The related term, defined next, is more widely used.

[3] 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.

Fixing Partial Random Choices

Informal Description

Suppose that a sample space Ω and probability distribution $P : \Omega \to \mathbb{R}$ 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

Theorem #10: Let Ω be a sample space with probability distribution $\mathrm{P} : \Omega \to \mathbb{R}$. 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 $1 \leq i \leq k$.

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

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

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.

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

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.

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

Mutual Independence 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.