Computer Science 351

Probability Distributions

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #18

Learning Goal

Learning Goal:

- Review material about probability theory that was, ideally, introduced CPSC 251 (or a first course in Probability and Statistics that you completed instead).

  The current lecture will review ***probability distributions*** — but these might were not necessarily presented as formally in the prerequisite course, and terminology and notation might be slightly different.
- Examples might be different — because applications to computer science may be introduced — and there is probably a small amount of *new material* about probability theory at the end.

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

Tossing a Coin for a Fixed Number of Times

- Let k be a positive integer, and suppose we toss a coin k times. Then each ***outcome*** can be represented as a *sequence*

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

  with length k where $\alpha_i \in \{\mathtt{H}, \mathtt{T}\}$, and α_i is what is obtained for the i^{th} toss of the coin for each integer i such that $1 \leq i \leq k$.

- With this representation the ***sample space*** is

$$\Omega = \{(\alpha_1, \alpha_2, \ldots, \alpha_k) \mid \alpha_i \in \{\mathtt{H}, \mathtt{T}\} \text{ for } 1 \leq i \leq k\}.$$

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), $|\Omega| = 2^k$.

Example: Rolling a Die Once

Rolling a Die Once

- In this case — if we are interested in the side of the die at the top when it stops, and represent this by the number of dots on it — there are six ***outcomes*** that each belong to the ***sample space***

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

— a set with size six.

Example: Rolling a Die for a Fixed Number of Times

Rolling a Die for a Fixed Number of Times

- Let k be a positive integer, and suppose we roll a die k times. Then each ***outcome*** can be represented as a *sequence*

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

  with length k where $\alpha_i \in \{1, 2, 3, 4, 5, 6\}$, and α_i is the number of the dots visible at the top of the die when it stops rolling for the i^{th} time, for each integer i such that $1 \leq i \leq k$.
- With this representation the ***sample space*** is

$$\Omega = \{(\alpha_1, \alpha_2, \ldots, \alpha_k) \mid \alpha_i \in \{1, 2, 3, 4, 5, 6\} \text{ for } 1 \leq i \leq k\}.$$

Example: Rolling a Die for a Fixed Number of Times

- For example, if $k = 2$ then the sample space is

$$\begin{aligned}\Omega = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}\end{aligned}$$

— a set with size $6^2 = 36$.

- In general (that is, for arbitrary k), $|\Omega| = 6^k$.

Example: Shuffling a Deck of Playing Cards

Example: Shuffling a Deck of Playing Cards

- A standard deck of paying cards contains 52 cards[2].
  - Each card has one of four ***suits*** — either spaces ♠, clubs ♣, hearts ♡ or diamonds ♢.
  - Each card has one of thirteen ***ranks*** — either one of the integers from 1 to 10, of "Jack" J, "Queen" Q or "King" K.[3]
  - Each card can be identified by a sequence of two symbols — the symbol for its rank, followed by the symbol for its suit. For example, the "Ace" (card with rank 1) of Spades is represented as 1♠.

Let $\mathcal{C}$ be the set of all playing cards, so that $|\mathcal{C}| = 52$.

[2] Jokers will not be included here.

[3] Each card with rank 1 is also called an "Ace" A.

Example: Shuffling a Deck of Playing Cards

- When “shuffling a deck of playing cards” an ***outcome*** is an ordering, or ***permutation***, of the cards in $\mathcal{C}$. This can be represented as a sequence

$$(\alpha_1, \alpha_2, \ldots, \alpha_{52})$$

of 52 cards that includes exactly one copy of each of the cards in $\mathcal{C}$.

- Note that there are 52 possible choices of α_1. For each choice of α_1 there are 51 choices of α_2. For each pair of choices of α_1 and α_2 there are exactly 50 choices of α_3, and so on — so that the ***sample space*** Ω — the set of all permutations of the cards in $\mathcal{C}$ — is a set with size

$$52! = \prod_{i=1}^{52} i.$$

Example: Balls and Bins

A More Complicated Example: Balls and Bins

- Let *m* and *n* be positive integers. Suppose that you have *m* balls and *n* bins.[4] You wish to place each one of the balls into one of the bins — and you want to keep track of *which* balls got placed into each one of the bins.
- Each ***outcome*** can be 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 $1 \leq i \leq m$: This represents the situation where the i^{th} ball got placed into the α_i^{th} bin, for $1 \leq i \leq m$.

[4] In other descriptions of this problem, you have *n boxes* or *n baskets* instead of *n bins.*

Example: Balls and Bins

- With this representation, the ***sample space*** is

$$\Omega = \{(\alpha_1, \alpha_2, \dots, \alpha_m) \mid \alpha_i \in \mathbb{Z} \text{ and } 1 \leq i \leq n \text{ for } 1 \leq i \leq m\}$$

— a set with size n^m.

- For example, if $m = 3$ and $n = 2$ (so there are three balls and two bins) then

$$\Omega = \{(1,1,1),(1,1,2),(1,2,1),1,2,2), (2,1,1),(2,1,2),(2,2,1),(2,2,2)\}.$$

Example: Balls and Bins

- On the other hand, if $m = 2$ and $n = 3$ (so that there are two balls and three bins, instead) then

$$\Omega = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),\\ (3,1),(3,2),(3,3)\}.$$

- In a way, this "generalizes" the "Tossing a Coin for a Fixed Number of Times" example (if "Heads" is represented by 1 instead of $\mathtt{H}$, and "Tails" is represented by 2 instead of $\mathtt{T}$) — with m used as a parameter instead of k, and with n replacing 2.
- This also "generalizes" the "Rolling a Die for a Fixed Number of Times" example — with m used as a parameter instead of k, and with n replacing 6.

Example: Trying Until You Succeed

Once again, suppose that you toss a coin — so that the result of the coin toss could either be "heads" (represented by H) or "tails" (represented by T). However — this time — you keep trying until you get "heads" (and then stop, as soon as you do).

- For every positive integer n, one outcome is that you toss the coin exactly n times before you can stop (so you toss "tails" $n-1$ times and then toss "heads" after that). This outcome can be represented by the positive integer n.
- There is no guarantee that you ever toss "heads" at all — you might simply toss "tails" over and over again This outcome can be represented by "$+\infty$" ("positive infinity").
- Since there are no other outcomes, the sample space is

$$\Omega = \{n \in \mathbb{Z} \mid n \geq 1\} \cup \{+\infty\}.$$

Discrete Probability Theory

- In all of the examples considered so far, except for the last one, the sample space has been ***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 $f(i) = x$.[5]

[5] This is not the only way to define "countable sets" — but the other, commonly used, definition is equivalent to this one.

Recall, as well, that the set $\mathbb{N}$ is not defined the same way in all books or courses. In this course, $\mathbb{N}$ is the set of ***non-negative*** integers, $0, 1, 2, 3, \ldots,$ so that $0 \in \mathbb{N}$.

Discrete Probability Theory

- Every *finite* set is countable. The set $\mathbb{N}$ of natural numbers is an example of an *infinite* set that is countable — because the above property if we set $f : \mathbb{N} \rightarrow \mathbb{N}$ to be the identity function (so that $f(i) = i$ for all $i \in \mathbb{N}$).

Discrete Probability Theory

- Once again, consider the last example — which included the same space

$$\Omega = \{n \in \mathbb{Z} \mid n \geq 1\} \cup \{+\infty\}.$$

Consider the function $f : \mathbb{N} \to \Omega$ such that, for every integer n such that $n \geq 0$,

$$f(n) = \begin{cases} n & \text{if } n \geq 1, \\ +\infty & \text{if } n = 0. \end{cases}$$

- The function $f : \mathbb{N} \to \Omega$ is ***surjective***, as needed to show that the sample space, Ω, is countable.

Discrete Probability Theory

In this course we will almost always consider experiments where the sample space Ω is countable — so we will be studying a part of probability theory that is 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 a set of size one — that is, it is an event that only includes a single outcome.[6]

[6] Sometimes, people say that "elementary event" and "outcome" mean the same thing. That is not true when an "elementary event" is as defined here, because a set of size one is a different kind of thing than the single element of that set.

Events: An Example

Consider the experiment, "tossing a coin, three times". As noted above, the sample space is

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

The event[7] "$\mathtt{H}$ is tossed more often than $\mathtt{T}$" is the subset

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

[7] As this example may suggest, we will often describe or name an event using a property (true/false condition) that the outcomes in the event satisfy.

Events: Another Example

Consider the experiment of rolling a die, two times. As noted above, the sample space is

$$\begin{aligned}\Omega = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(2,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}\end{aligned}$$

The event "6 is rolled, at least once" is the subset

$$\{(1,6),(2,6),(3,6),(4,6),(5,6),\\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}.$$

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

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

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

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

by setting

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

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

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 $\mathrm{P} : \Omega \to \mathbb{R}$ such that

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

for every outcome $x \in \Omega$.

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

Uniform Distributions

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

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

Uniform Distributions

- ***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

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

Nonuniform Distributions

- ***Note:*** Some descriptions of probability theory (including introductory videos that you can find online) suggest that "uniform probability distributions" are the only ones that exist, or, at least, are the only ones of interest.
- This is not true! Sometimes it is important to consider the probability of various events when outcomes (included in the sample space) are *not* equally likely.

Nonuniform Distributions

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

- In particular suppose, once again, that we toss a coin three times, so that

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

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

Nonuniform Distributions

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

Nonuniform Distributions

- Once again consider the event "$\mathtt{H}$ is tossed more often than $\mathtt{T}$", 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 $\mathtt{H}$ is tossed more often than $\mathtt{T}$ 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 $\frac{1}{2}$.

An Experiment with an Infinite Sample Space

Once again, consider the experiment in which you toss a coin, over and over again, until “Heads” is tossed.

- As noted above experiment can be modelled using an ***infinite sample space***

  $$\Omega = \{n \in \mathbb{Z} \mid n \geq 1\} \cup \{+\infty\}$$

  where the elements of Ω represent the following situations.

  - For each positive integer k, k represents the outcome that “T” is tossed the first $k-1$ times, and then H is tossed after that.
  - $+\infty$ represents the outcome that you never get ‘H” at all. Instead, “T” is tossed, over and over again.

An Experiment with an Infinite Sample Space

Consider a function $\mathrm{P} : \Omega \to \mathbb{R}$ that is defined as follows.

- $\mathrm{P}(n) = 2^{-n}$ for every number $n \in \mathbb{Z}$ such that $n \geq 1$.
- $\mathrm{P}(+\infty) = 0$.

Then this is a total function from Ω to $\mathbb{R}$ such that $0 \leq \mathrm{P}(x) \leq 1$ for every outcome $x \in \Omega$, such that

$$\begin{aligned} \sum_{x \in \Omega} \mathrm{P}(x) &= \sum_{\substack{n \in \mathbb{Z} \\ n \geq 1}} \mathrm{P}(n) + \mathrm{P}(+\infty) \\ &= \sum_{\substack{n \in \mathbb{Z} \\ n \geq 1}} 2^{-n} + 0 = 1 \end{aligned}$$

(using a formula for the sum of a geometric series that you have, ideally, seen already). Thus the function P is a ***probability distribution*** for this experiment (and infinite sample space).

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***[8] of the event A, $\overline{A}$, is the set of outcomes that *are not* in A.

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

[8] 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 Complement of an Event

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

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

Proof: Let Ω, P and A be as in the statement of the claim. Then, since P is a probability distribution and $A \subseteq \Omega$,

$$\begin{aligned} 1 &= \sum_{x \in \Omega} \mathrm{P}(x) = \sum_{x \in A} \mathrm{P}(x) + \sum_{\substack{x \in \Omega \\ x \notin A}} \mathrm{P}(x) \\ &= \sum_{x \in A} \mathrm{P}(x) + \sum_{x \in \overline{A}} \mathrm{P}(x) \\ &= \mathrm{P}(A) + \mathrm{P}(\overline{A}). \end{aligned}$$

It follows that $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$, as claimed. $\square$

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 $A, B \subseteq \Omega$,

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

Proof: Let Ω, P, A and B be as in the statement of the claim. Let

$$A \setminus B = \{x \in A \mid x \notin B\}$$

and let

$$B \setminus A = \{x \in B \mid x \notin A\}.$$

Probability of the Union of Events

Then every event $x \in A$ belongs to exactly one of the sets $A \cap B$ and $A \setminus B$, so that

$$\begin{aligned} P(A) &= \sum_{x \in A} P(x) \\ &\qquad \text{(by the definition of the probability of an event)} \\ &= \sum_{x \in A \cap B} P(x) + \sum_{x \in A \setminus B} P(x) \qquad \text{(splitting the sum)} \\ &= P(A \cap B) + P(A \setminus B). \end{aligned} \tag{1}$$

Switching the roles of the events A and B, and applying the argument again, one can also show that

$$P(B) = P(A \cap B) + P(B \setminus A). \tag{2}$$

Probability of the Union of Events

Similarly, every event $x \in A \cup B$ belongs to exactly one of the sets $A \cap B$, $A \setminus B$, and $B \setminus A$. Thus

$$\begin{aligned} \mathrm{P}(A \cup B) &= \sum_{x \in A \cup B} \mathrm{P}(x) \\ &\quad \text{(by the definition of the probability of an event)} \\ &= \sum_{x \in A \cap B} \mathrm{P}(x) + \sum_{x \in A \setminus B} \mathrm{P}(x) + \sum_{x \in B \setminus A} \mathrm{P}(x) \\ &\quad \text{(splitting the sum)} \\ &= \mathrm{P}(A \cap B) + \mathrm{P}(A \setminus B) + \mathrm{P}(B \setminus A). \qquad (3) \end{aligned}$$

Probability of the Union of Events

Thus

$$
\begin{aligned}
&P(A \cup B) + P(A \cap B) \\
&\qquad = (P(A \cap B) + P(A \setminus B) + P(B \setminus A)) + P(A \cap B) \\
&\qquad\qquad \text{(by the equation at line (3))} \\
&\qquad = (P(A \cap B) + P(A \setminus B)) + (P(A \cap B) + P(B \setminus A)) \\
&\qquad\qquad \text{(reordering terms)} \\
&\qquad = P(A) + P(B) \\
&\qquad\qquad \text{(by the equations at lines (1) and (2)).}
\end{aligned}
$$

It now follows that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, as claimed. □

The Union Bound

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 $A, B \subseteq \Omega$,

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

Notice that, for a positive integer k such that $k \geq 3$, If $E_1, E_2, \ldots, E_k \subseteq \Omega$ then

$$E_1 \cup E_2 \cup \cdots \cup E_k = U_{k-1} \cup E_k$$

where

$$U_{k-1} = E_1 \cup E_2 \cup \cdots \cup E_{k-1}.$$

The Union Bound

This observation, and Corollary #3, can be used to establish the following by induction on k.

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

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

Why These Results Matter

- As probability theory is applied to solve problems in computer science, we will see that it is often useful to be able to compute — or, at least, find upper bounds for — the probabilities of various events.
- Sometimes the most easily understood way to do this is to express the event, that we are interested in, as
  - the *complement* of an event whose probability is easy to compute, or
  - the *union* of a finite number of events whose probabilities are easy to compute,

  and then apply Theorem #1, Theorem #2, or Theorem #4 to compute (or, in the last case, bound) the probability of the event we are interested in.