2.2–2.3 Introduction to Probability and Sample Spaces

Prof. Tesler

Math 186
Winter 2019

Course overview

Probability: Determine likelihood of events

- Roll a die. The probability of rolling 1 is $1/6$.

Descriptive statistics: Summarize data

- Mean, median, standard deviation, . . .

Inferential statistics: Infer a conclusion/prediction from data

- Test a drug to see if it is safe and effective, and at what dose.
- Poll to predict the outcome of an election.
- Repeatedly flip a coin or roll a die to determine if it is fair.

Bioinformatics

- We'll apply these to biological data, such as DNA sequences and microarrays.

Related courses

- **Math 183:** Usually uses the same textbook and chapters as Math 186. Focuses on the examples in the book. The mathematical content is the same, but Math 186 has extra material for bioinformatics.
- **Math 180ABC plus 181ABC:** More in-depth: a year of probability and a year of statistics.
- **CSE 103, Econ 120A, ECE 109:** One quarter intro to probability and statistics, specialized for other areas.
- **Math 283:** Graduate version of this course. Review of basic probability and statistics, with a lot more applications in bioinformatics.

2.2 Sample spaces

- Flip a coin 3 times. The possible *outcomes* are

$$HHH \quad HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH \quad TTT$$

- The *sample space* is the set of all possible outcomes:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

- The size of the sample space is
  $N(S) = 8$ Our book's notation
  $|S| = 8$ A more common notation in other books

- We could count this by making a $2 \times 2 \times 2$ table:

$$\begin{aligned}
&2 \text{ choices for the first flip}\\
&\times\, 2 \text{ choices for the second flip}\\
&\times\, 2 \text{ choices for the third flip}\\
&= 2^3 = 8
\end{aligned}$$

- The number of strings $x_1\,x_2\,\ldots\,x_k$ or sequences $(x_1, x_2, \ldots, x_k)$ of length k with r choices for each entry is r^k.

Rolling two dice

- Roll two six-sided dice, one red, one green:

		green 1	2	3	4	5	6
red	1	$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
	2	$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
	3	$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
	4	$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
	5	$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
	6	$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

- The sample space is

$$\begin{aligned} S &= \{(1,1),(1,2),\ldots,(6,6)\} \\ &= \left\{\,(i,j)\in\mathbb{Z}^2 \;:\; 1\leqslant i\leqslant 6,\;\; 1\leqslant j\leqslant 6\,\right\} \end{aligned}$$

where $\mathbb{Z}=$ integers $\quad \mathbb{Z}^2=$ ordered pairs of integers

- $N(S)=6^2=36$

DNA sequences

- A *codon* is a DNA sequence of length 3, in the alphabet of nucleotides { A, C, G, T }:

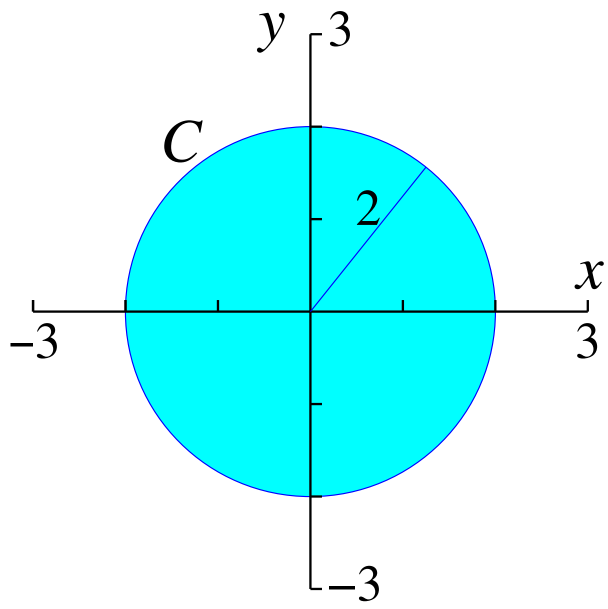
$$S = \{\text{ AAA, AAC, AAG, AAT}, \ldots, \text{TTT } \}$$

- How many codons are there?

$$N(S) = 4^3 = \boxed{\mathbf{64}}$$

A continuous sample space

Consider this disk (filled-in circle):



$$S = \{ (x, y) \in \mathbb{R}^2 \ : \ x^2 + y^2 \leqslant 2^2 \}$$

Complications

- The sample space is infinite and continuous.
- The choices of x and y are dependent. E.g.:

  at $x = 0$, we have $-2 \leqslant y \leqslant 2$;

  at $x = 2$, we have $y = 0$.

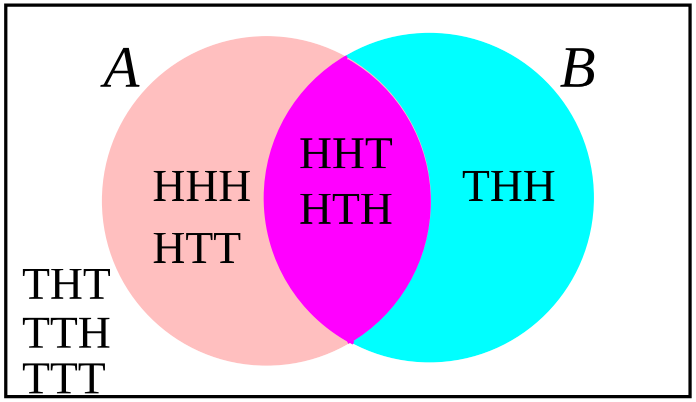
Events

- Flip a coin 3 times. The sample space is
  $$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$
- An *event* is a subset of the sample space ($A \subset S$):
  - $A =$ "First flip is heads" $= \{HHH, HHT, HTH, HTT\}$
  - $B =$ "Two flips are heads" $= \{HHT, HTH, THH\}$
  - $C =$ "Four flips are heads" $= \emptyset$ (*empty set* or *null set*)
- We can combine these using set operations. For example,
  "The first flip is heads or two flips are heads"
  $$\begin{array}{rcccccl} A = \{ & HHH, & HHT, & HTH, & HTT & & \} \\ B = \{ & & HHT, & HTH, & & THH & \} \\ \hline A \cup B = \{ & HHH, & HHT, & HTH, & HTT, & THH & \} \end{array}$$

Using set operations to form new events

A = "First flip is heads"
$= \{HHH, HHT, HTH, HTT\}$

B = "Two flips are heads"
$= \{HHT, HTH, THH\}$



Union: All elements that are in A or in B

- $A \cup B = \{HHH, HHT, HTH, HTT, THH\}$
- "A or B": "The first flip is heads or two flips are heads"
- This is *inclusive or*: one **or both** conditions are true.

Intersection: All elements that are in both A and in B

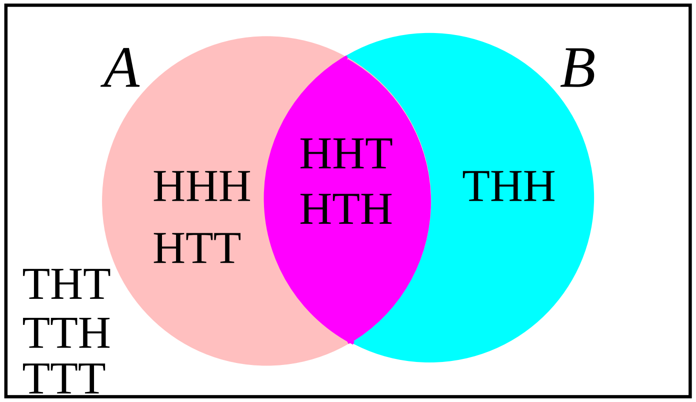
- $A \cap B = \{HHT, HTH\}$
- "A and B": "The first flip is heads and two flips are heads"

Complement: All elements of the sample space not in A

- $A^c = \{THT, TTH, TTT, THH\}$
- "Not A": "The first flip is not heads"

Venn diagram and set sizes

$$A = \{HHH, HHT, HTH, HTT\}$$
$$B = \{HHT, HTH, THH\}$$
$$A \cup B = \{HHH, HHT, HTH, HTT, THH\}$$
$$A \cap B = \{HHT, HTH\}$$



Relation between sizes of union and intersection

- Notice that
$$\begin{array}{ccccccc} N(A\cup B) & = & N(A) & + & N(B) & - & N(A\cap B) \\ 5 & = & 4 & + & 3 & - & 2 \end{array}$$
- $N(A)+N(B)$ counts everything in the union, but elements in the intersection are counted twice. Subtract $N(A\cap B)$ to compensate.

Size of complement

$$\begin{array}{ccccc} N(B^c) & = & N(S) & - & N(B) \\ 5 & = & 8 & - & 3 \end{array}$$

Algebraic rules for set theory

Commutative laws
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Associative laws
$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

One may omit parentheses in $A \cap B \cap C$ or $A \cup B \cup C$.
But don't do that with a mix of $\cup$ and $\cap$.

Distributive laws
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
These are like $a(b+c) = ab + ac$

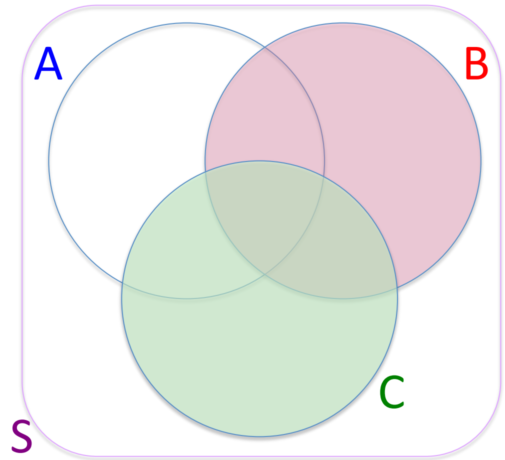
Complements
$$A \cup A^c = S$$
$$A \cap A^c = \emptyset$$

De Morgan's laws
$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

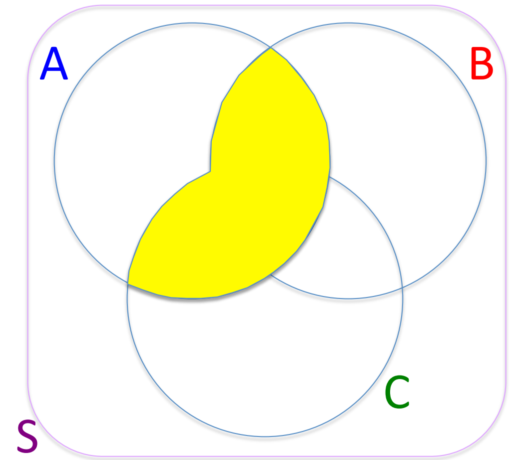
Distributive laws

Visualizing identities using Venn diagrams: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

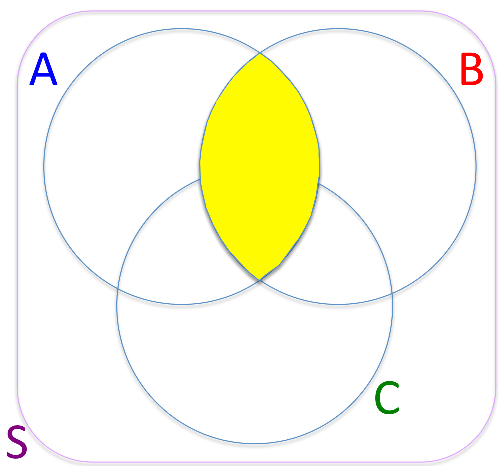
$B \cup C$



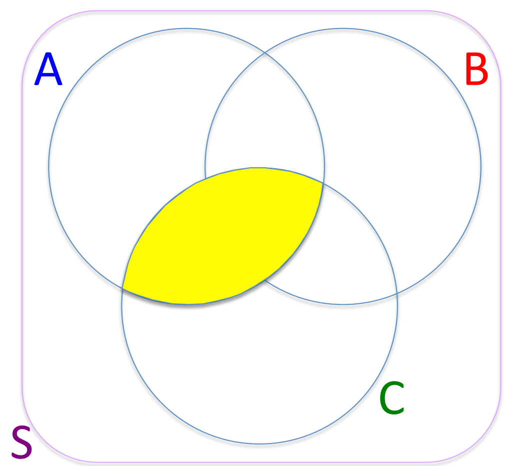
$A \cap (B \cup C)$



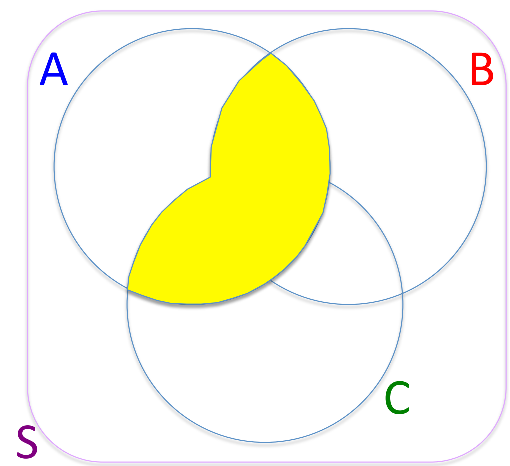
$A \cap B$



$A \cap C$



$(A \cap B) \cup (A \cap C)$



Mutually exclusive sets

- Two events are *mutually exclusive* if their intersection is $\emptyset$.

$$\begin{aligned} A &= \text{“First flip is heads”} = \{HHH, HHT, HTH, HTT\} \\ B &= \text{“Two flips are heads”} = \{HHT, HTH, THH\} \\ C &= \text{“One flip is heads”} = \{HTT, THT, TTH\} \end{aligned}$$

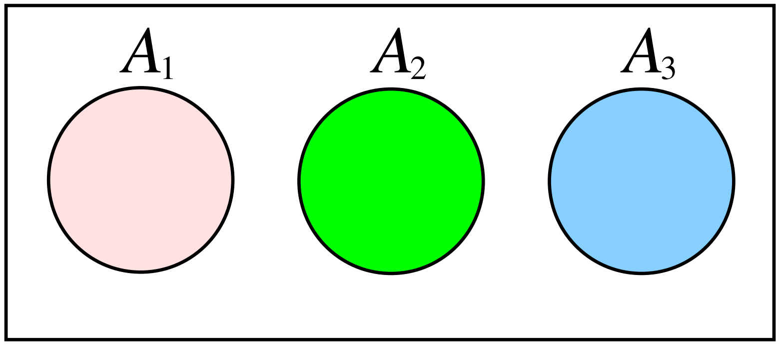
  A and B are not mutually exclusive, since $A \cap B = \{HHT, HTH\} \neq \emptyset$.
  B and C are mutually exclusive, since $B \cap C = \emptyset$.

- For mutually exclusive events, since $N(B \cap C) = 0$, we get:

$$N(B \cup C) = N(B) + N(C)$$

- Events $A_1, A_2, \ldots$ are *pairwise mutually exclusive* when

$$A_i \cap A_j = \emptyset \quad \text{for} \quad i \neq j.$$



2.3 Probability functions

Historically, there have been several ways of defining probabilities.
We'll start with *Classical Probability*:

Classical probability

- Suppose the sample space has n outcomes ($N(S) = n$) and all of them are equally likely.
- Each outcome has a probability $1/n$ of occurring:
$$P(s) = 1/n \qquad \text{for each outcome } s \in S$$
- An event $A \subset S$ with m outcomes has probability m/n of occurring:
$$P(A) = \frac{m}{n} = \frac{N(A)}{N(S)}$$

Example: Rolling a pair of dice

- $N(S) = n = 36$
- $P(\text{first die is 3}) = P\big(\{(3,1),(3,2),\ldots,(3,6)\}\big) = \frac{6}{36}$
- $P(\text{the sum is 8}) = P\big(\{(2,6),(3,5),(4,4),(5,3),(6,2)\}\big) = \frac{5}{36}$

Classical probability

Drawbacks

- What if outcomes are not equally likely?
- What if there are infinitely many outcomes?

Empirical probability

Use long-term frequencies of different outcomes to estimate their probabilities.

- Flip a coin a lot of times. Use the fraction of times it comes up heads to estimate the probability of heads.
- 520 heads out of 1000 flips leads to estimating $P(\text{heads}) = 0.520$.

- This estimate is only approximate because
  - Due to random variation, the numerator will fluctuate.
  - Precision is limited by the denominator. 1000 flips can only estimate it to three decimals.
- More on this later in the course in Chapter 5.3.

Empirical probability

- *E. coli* has been sequenced:

Position:	1	2	3	4	5	6	7	8	9	10	⋯
Base:	A	G	C	T	T	T	T	C	A	T	⋯

- On the forwards strand:

# A's	1,142,136	$P(A) = \frac{1,142,136}{4,639,221} \approx 0.2461913326$
# C's	1,179,433	$P(C) \approx 0.2536578878$
# G's	1,176,775	$P(G) \approx 0.2542308288$
# T's	1,140,877	$P(T) \approx 0.2459199508$
Total	4,639,221	1

- Sample space: set of positions $S = \{1, 2, \ldots, 4639221\}$
- Event A is the set of positions with nucleotide A (similar for C, G, T).
  $A = \{1, 9, \ldots\}$ $\quad C = \{3, 8, \ldots\}$ $\quad G = \{2, \ldots\}$ $\quad T = \{4, 5, 6, 7, 10, \ldots\}$
- Simplistic model: the sequence is generated from a biased 4-sided die with faces A, C, G, T.

Axiomatic probability

A definition of a probability function P based on events and the following axioms is the most useful.
Each event $A \subset S$ is assigned a probability that obeys these axioms:

Axioms for a finite sample space

- For any event $A \subset S$: $P(A) \geqslant 0$
- The total sample space has probability 1: $P(S) = 1$
- For mutually exclusive events A and B: $P(A \cup B) = P(A) + P(B)$

Additional axiom for an infinite sample space

- If $A_1, A_2, \ldots$ (infinitely many) are pairwise mutually exclusive, then
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$
- $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$ is like $\sum$ notation, but for unions.

Axiomatic probability — additional properties

Additional properties of the probability function follow from the axioms.

$P(A^c) = 1 - P(A)$

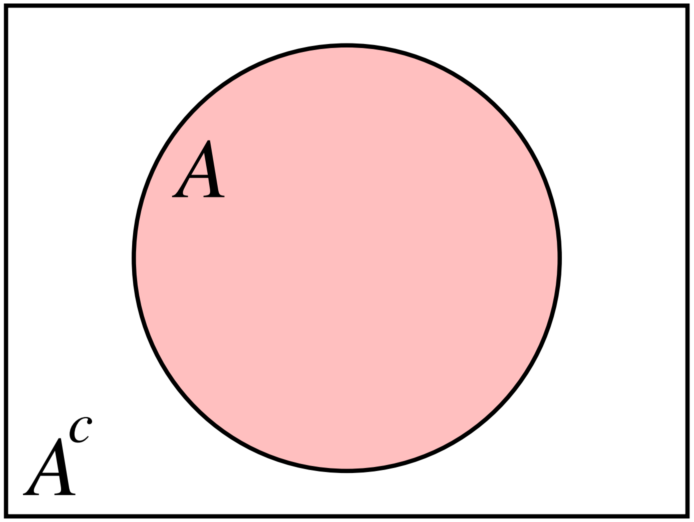
Example: $P(\text{die roll} = 1) = \frac{1}{6}$ $P(\text{die roll} \neq 1) = 1 - \frac{1}{6} = \frac{5}{6}$

Proof.

A and A^c are mutually exclusive, so $P(A \cup A^c) = P(A) + P(A^c)$.

Also, $P(A \cup A^c) = P(S) = 1$.

Thus, $P(A^c) = 1 - P(A)$. □



$P(\emptyset) = 0$

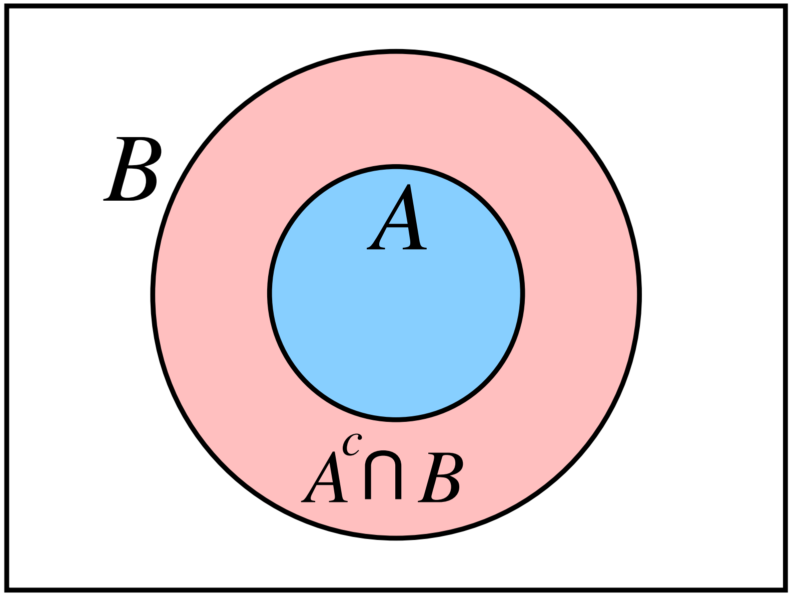
Proof: $P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0$

Axiomatic probability — additional properties

If $A \subset B$ then $P(A) \leqslant P(B)$

Proof: Write $B = A \cup (A^c \cap B)$.

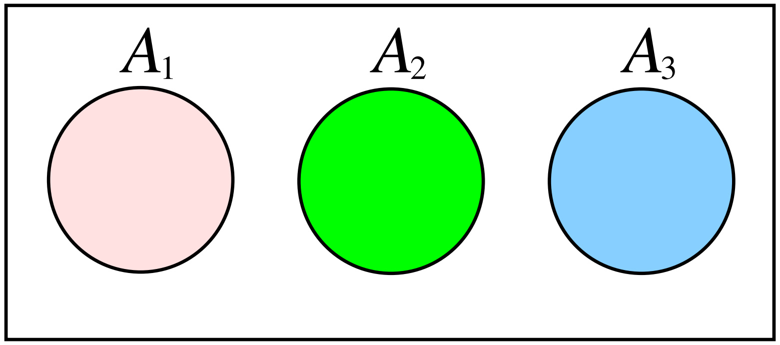
A and $A^c \cap B$ are mutually exclusive, so
$$P(B) = P(A) + P(A^c \cap B).$$
The first axiom gives $P(A^c \cap B) \geqslant 0$, so
$$P(B) \geqslant P(A).$$



$P(A) \leqslant 1$

Proof: $A \subset S$ so $P(A) \leqslant P(S) = 1$.

Axiomatic probability — additional properties



Additive Law

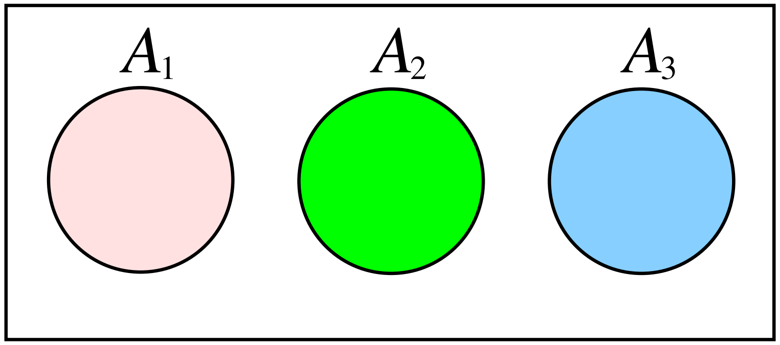
If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

- $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$ is like $\sum$ notation, but for unions.
- Prove for all integers $n \geqslant 1$ using induction, based on:

$$\begin{aligned} P\big((A_1 \cup \cdots \cup A_n) \cup A_{n+1}\big) &= P(A_1 \cup \cdots \cup A_n) + P(A_{n+1}) \\ &= \big(P(A_1) + \cdots + P(A_n)\big) + P(A_{n+1}). \end{aligned}$$

Axiomatic probability — additional properties



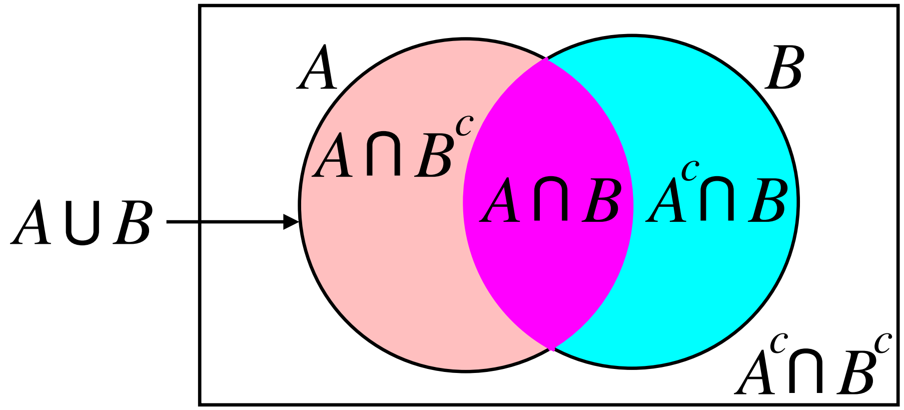
Induction proves the Additive Law for positive integers $n = 1, 2, \ldots$, but not for $n = \infty$, so we have to introduce an additional axiom for that:

Additional axiom for an infinite sample space

If $A_1, A_2, \ldots$ (infinitely many) are pairwise mutually exclusive, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Axiomatic probability — additional properties



De Moivre's Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$
$$P(B) = P(A \cap B) + P(A^c \cap B)$$

$$\begin{aligned} P(A) + P(B) &= P(A \cap B) + \big(P(A \cap B) + P(A \cap B^c) + P(A^c \cap B)\big) \\ &= P(A \cap B) + P(A \cup B) \end{aligned}$$

Axiomatic probability — additional properties

Additional properties of the probability function follow from the axioms:

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- If $A \subset B$ then $P(A) \leqslant P(B)$
- $P(A) \leqslant 1$
- *Additive Law:* If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive ($A_i \cap A_j = \emptyset$ for all $i \neq j$) then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

  (The first three axioms only lead to a proof of this for finite n, but not $n = \infty$, so $n = \infty$ has to be handled by an additional axiom.)
- *De Moivre's Law:* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Generalizing Sigma ($\sum$) notation

Sum over sets instead of over a range of consecutive integers

- Consider a biased n-sided die, with faces $1, \ldots, n$.
- The probability of face i is q_i, with $0 \leqslant q_i \leqslant 1$, and $q_1 + \cdots + q_n = 1$:

$$\sum_{i=1}^{n} q_i = 1$$

- For $n = 8$, the probability of an even number is:

$$P(\text{even number}) = \sum_{s \in \{2,4,6,8\}} q_s \quad = q_2 + q_4 + q_6 + q_8$$

- For a 26-sided die with faces A,B,...,Z, the total probability is

$$P(\{A, B, \ldots, Z\}) = \sum_{s \in \{A,B,\ldots,Z\}} q_s \quad = q_A + q_B + \cdots + q_Z = 1,$$

and the probability of a vowel is

$$P(\{A, E, I, O, U\}) = \sum_{s \in \{A,E,I,O,U\}} q_s \quad = q_A + q_E + q_I + q_O + q_U \, .$$

Finite discrete sample space

- Each outcome has a probability between 0 and 1 and the probabilities add up to 1:

$$0 \leqslant P(s) \leqslant 1 \quad \text{for each } s \in S \qquad\qquad \sum_{s \in S} P(s) = 1$$

- For an event $A \subset S$, define $P(A) = \sum_{s \in A} P(s)$.

Examples

- A biased n-sided die (previous slide).
- For flipping a coin 3 times,
  $P(\{HHT, HTH, THH\}) = P(HHT) + P(HTH) + P(THH)$.

DNA

- Generate a random DNA sequence by rolling a biased ACGT-die.
- Deviations from what's expected from random rolls of a dice are used to detect structure in the DNA sequence, like genes, codons, repeats, etc.