STA237: Probability, Statistics and Data Analysis I

Section 0101 Lecture 2

Lijia Wang

Department of Statistical Sciences
University of Toronto

Introduction to Lecture 2

Lecture 2 coverage:

- Events, Outcomes, and set operators (complements, unions, intersections)
- Define probability function and the probability axioms
- Inclusion-Exclusion Principle: 2 events

Lecture 2 learning outcomes:

- Use set notation and Venn disatrams to represent events in a given sample space
- Distinguish between mutually exclusive (i.e., disjoint) and independent events
- Understand the probability function
- State the three probability axioms and use axioms to relate probabilities of sets of events
- Apply inclusion-exclusion principle ('Addition Rule') to find probabilities involving unions of events

Outline

Discrete sample space

Definition (discrete sample space)

We say a sample space is *discrete* if it is either finite or countably infinite

Question: What does countably infinite mean?

Countably infinite

Definition (Countably infinite)

A set is *countably infinite* if the elements of the set can be arranged as a sequence.

Remark:

- And all countably infinite sets can be put in one-to-one correspondence with the natural numbers.

Example on countably infinite set

- **Example:** The natural numbers $1, 2, 3, \ldots$ is the classic example of a countably infinite set.
- **Counter example:**
  - The set of all real numbers is an infinite set that is not countably infinite. It is called uncountable.
  - An interval of real numbers, such as $(0, 1)$, the numbers between 0 and 1, is also uncountable.

discrete sample space

We assume for the next several lectures that the sample space is discrete.

- If the sample space is finite, it can be written as $\Omega = \{\omega_1, \ldots, \omega_k\}$.
- If the sample space is countably infinite, it can be written as $\Omega = \{\omega_1, \omega_2, \ldots\}$.

Remark: Probability on uncountable spaces will require differential and integral calculus and will be discussed later on for this term.

Properties of probabilities and event operations

We can conduct operations on probabilities / events like performing those with sets:

- Events can be combined together to create new events using the connectives "or", "and" and "not".
- These correspond to the set operations union, intersection, and complement.

Before that, we provide a brief review on set notations and operations.

To begin, we need some terminologies and notation from the set theory.

- Capital letters denote the sets of objects: A, B, etc.
- If set A consists of objects a_1, a_2, a_3, we write $A = \{a_1, a_2, a_3\}$
- S denotes the universal set which is the set of all possible objects.
- ϕ denotes the null or empty set ($\phi = \{\}$) which is the set without any object.

S will be used to denote sample space in event operations.

Set operation: subset

e.g. $A = \{2\}$

$B = \{\text{even numbers}\}$

event A is a subset of event B

event A happen $\Rightarrow$ event B must happen

the contrary may not hold.

Definition (Set operation: subset)

For any two sets A and B, A is a *subset* of B if every object in A is in B.
Notation: $A \subseteq B$.

- Example: If $B = \{a_1, a_2, a_3, a_4\}$ and $A = \{a_1, a_2, a_3\}$ then A is a subset of B.
- Note: The null set is a subset of every set.

Set operation: union

In event setting

union either event A or B or both happens.

$A = \{\text{obtain } 1\}$ $B = \{\text{obtain } 2\}$ $A \cup B = \{\text{obtain either } 1 \text{ or } 2\}$

Definition (Set operation: union)

The *union* of A and B is the set of all objects in A or B or both.
Notation: $A \cup B$.

- Example: If $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$ then $A \cup B = \{a_1, a_2, b_1, b_2\}$.

Set operation: intersection

In set operations, $A \cap B$ means event A and B must happen together in this experiment.

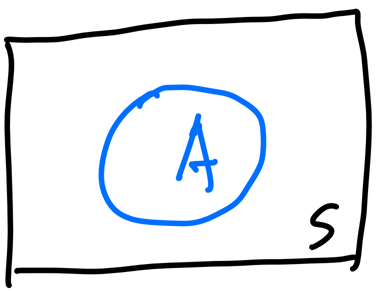
Definition (Set operation: intersection)

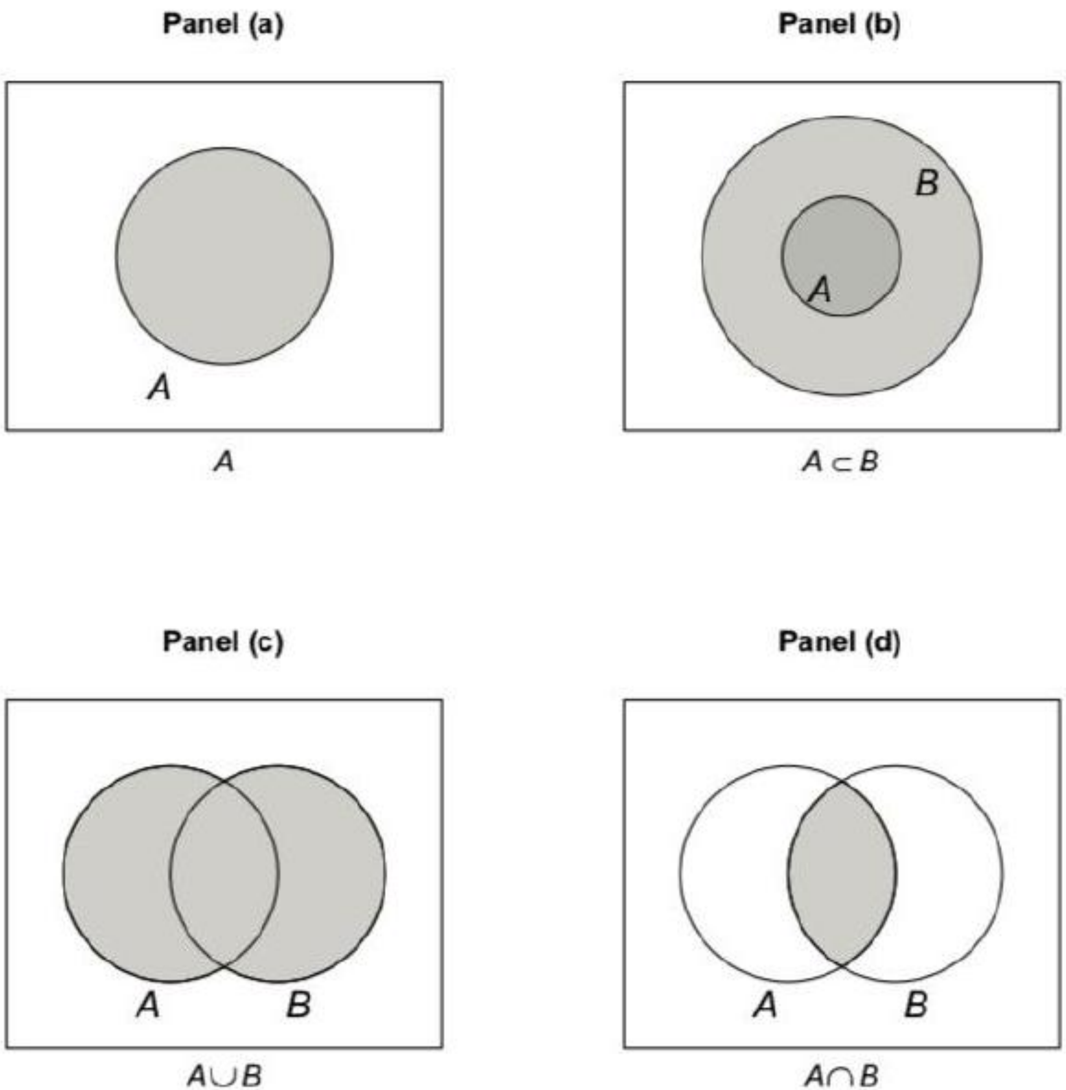
The *intersection* of sets A and B is the set of all objects in both A and B.
Notation: $A \cap B$ or just AB.

- Example: If $B = \{a_1, a_2, a_3, a_4\}$ and $A = \{a_1, a_2, a_3\}$ then $A \cap B = \{a_1, a_2, a_3\}$.
- The key word for union is or while the key word for intersection is and.

Venn Diagram

The Venn Diagram is a very useful tool:





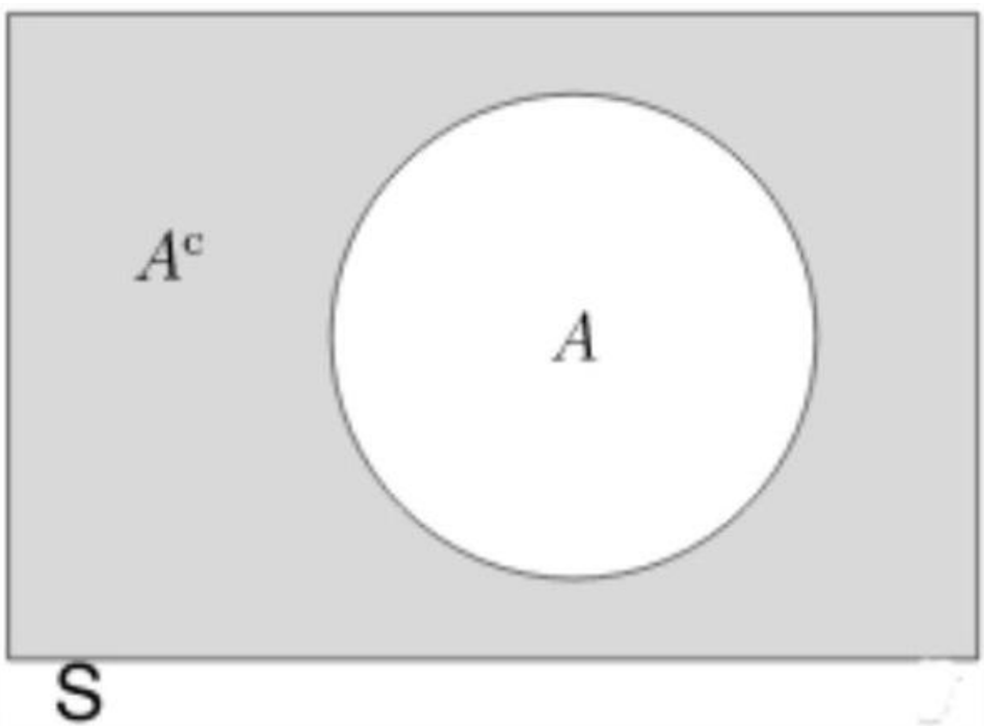
Set operation: complement

Definition

If A is a subset of S, then the complement of A is the set of all objects in S that are not in A. It is denoted by A^c.

- Note that $A \cup A^c = S$
- By definition, $A \cap A^c = \phi$

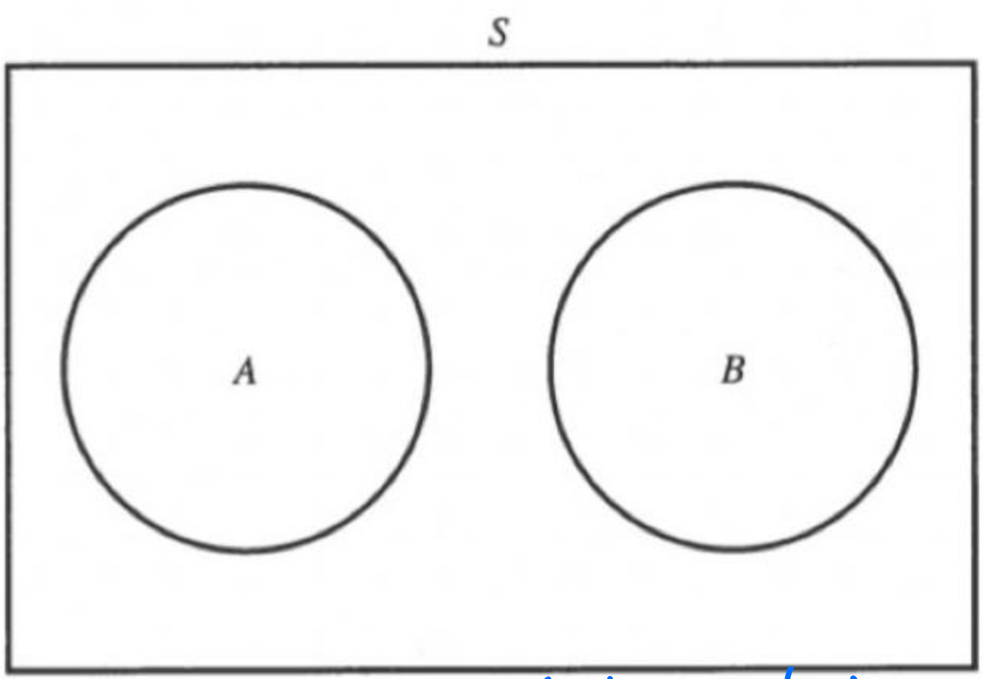
$A = \{1\}$ $A^c = \{2, 3, 4, 5, 6\}$.



Set operation: disjoint

Definition

Two sets A and B are said to be disjoint or mutually exclusive if they have no object in common. That is, A and B are disjoint if $A \cap B = \phi$



disjoint / mutually exclusive } same

Two sets without overlapping
$\Rightarrow$ disjoint, mutually exclusive.

Disjoint vs independent

Mutually exclusive / disjoint:

- Two events A and B are mutually exclusive / disjoint if the events cannot both occur or occur simultaneously as an outcome of the experiment.
- In a Venn Diagram, A and B would be disjoint if they have no overlapping

Independent:

- Two events A and B are independent if the occurrence of one event does not alter the probability of occurrence of the other in any way.

mutually exclusive $\Rightarrow$ highly "dependent".

Event operations

Event operations following the same logics as set operations:

Table: TABLE 1.2. Events and sets.

Description	Set notation
Either A or B or both occur	$A \cup B$
A and B	AB
Not A	A^c
A implies B; A is a subset of B	$A \subseteq B$
A but not B	AB^c
Neither A nor B	$A^c B^c$
At least one of the two events occurs	$A \cup B$
At most one of the two events occurs	$(AB)^c = A^c \cup B^c$

Important laws for set operations

- Commutative Laws:

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

Set operation

as multiplication.

- Associative Law :

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

- Distributive Laws:

$a \times (b+c)$

$= (a \times b) + (a \times c)$

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

$A \cap (B \cup C)$

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

DeMorgan's Laws

- DeMorgan's Laws (for two sets A and B):

$$(A \cup B)^c = A^c \cap B^c \qquad A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

- DeMorgan's Laws (general case for the set $\{A_1, A_2, \ldots A_n\}$):

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} {A_i}^c \qquad (A_1 \cup A_2 \cup A_3)^c$$
$$= A_1^c \cap A_2^c \cap A_3^c$$
$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} {A_i}^c$$

Outline

Probability function

In a random experiment with sample space Ω, the probability of an event A, denoted as $P(A)$ is a function that assigns to event A a numerical value that measures the chance that event A will occur.

There are three axioms that must hold for probability functions:

1. $P(A) \geq 0$ negative value doesn't make sense
2. $P(\Omega) = 1$ Eventually, one of the outcomes must happen.
3. For a set of disjoint (i.e. mutually exclusive) events $A_1, A_2, \ldots, A_n$ in Ω,

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

e.g.

$A_1 = \{1\}$ $A_2 = \{2\}$

$A_1 \cup A_2 = \{1, 2\}$ $p(A_1 \cup A_2) = p(A_1) + p(A_2)$

Probability function

Definition (Probability function)

Given a random experiment with discrete sample space Ω, a *probability function* P is a function on Ω with the following properties:

1. $P(\omega) \geq 0$, for all $\omega \in \Omega$.
2. $\sum_{\omega \in \Omega} P(\omega) = 1$
3. For all events $A \subseteq \Omega$, $P(A) = \sum_{\omega \in A} P(\omega)$

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

$A = \{\text{obtain } 1,2,3\} = \{1,2,3\} \subseteq \Omega$

$P(A) = P(\{1,2,3\}) = P(1) + P(2) + P(3)$

Probability function on discrete sample space

- In the case of a finite sample space $\Omega = \{\omega_1, \ldots, \omega_k\}$, the second condition becomes

$$\sum_{\omega \in \Omega} P(\omega) = P(\omega_1) + \cdots + P(\omega_k) = 1$$

- And in the case of a countably infinite sample space $\Omega = \{\omega_1, \omega_2, \ldots\}$, this gives

$$\sum_{\omega \in \Omega} P(\omega) = P(\omega_1) + P(\omega_2) + \cdots = \sum_{i=1}^{\infty} P(\omega_i) = 1.$$

Remark on probability function

Remark:

- The first axiom guarantees that the probability function is always non-negative
- The second axiom implies probabilities sum to 1
- The third axiom says that the probability of an event is the sum of the probabilities of all the outcomes contained in that event.

Example on probability function

Example: Suppose that a college has six majors: biology, geology, physics, dance, art, and music. The percentage of students taking these majors are 20, 20, 5, 10, 10, and 35 , respectively, with double majors not allowed. Choose a random student. What is the probability they are a science major?

event A : "science major"

$$A = \{\text{biology}, \text{geology}, \text{physics}\}$$

$$p(A) = p(\text{biology}) + p(\text{geology}) + p(\text{physics}) = 45\%$$

Outline

Theorem (The Additive Law of Probability)

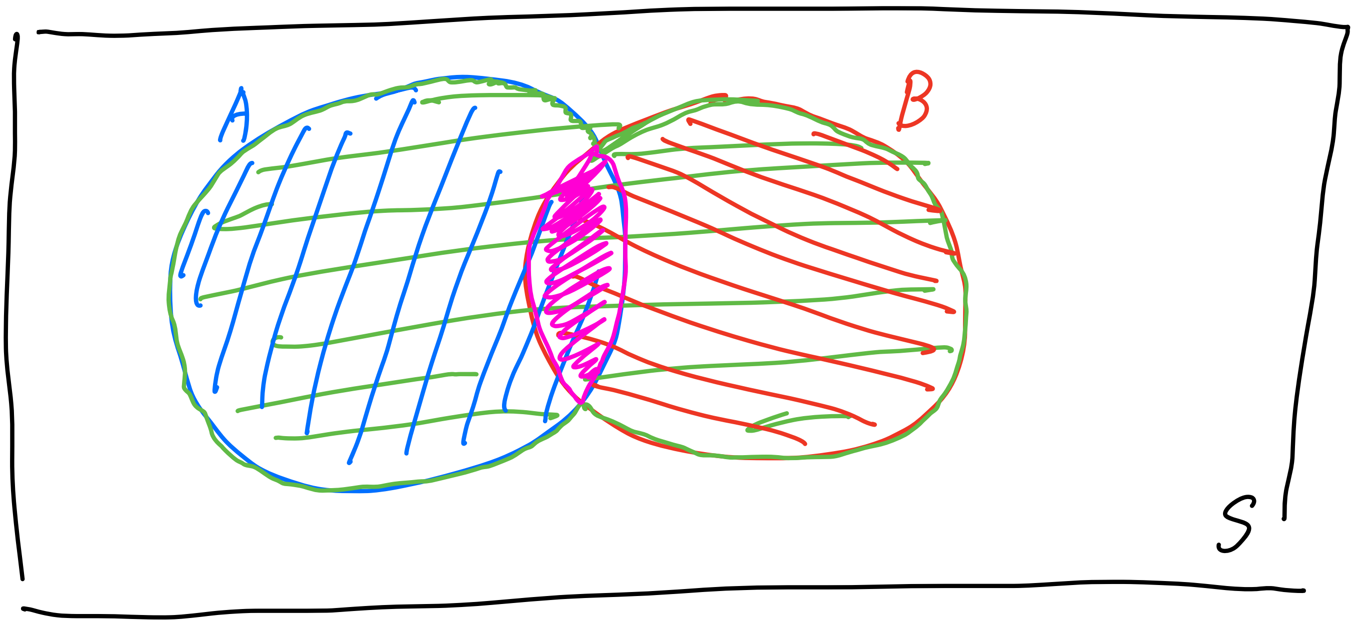
The probability of the union of two events A and B is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Remark:

- Note that this theorem is also called the *Inclusion-Exclusion Principle*.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



The addition rules (general case)

The addition rules can be extended to more general cases:

Theorem (The Additive Law of Probability (general case))

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \dots$$
$$+ (-1)^{r+1} \sum_{i_1<i_2<\dots i_r} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_r}) + \dots$$
$$+ (-1)^{n+1} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n})$$

Example: 3-way A_1, A_2, A_3

$$
\begin{aligned}
p(A_1 \cup A_2 \cup A_3) = \; & p(A_1) + p(A_2) + p(A_3) \\
& - p(A_1 \cap A_2) - p(A_1 \cap A_3) - p(A_2 \cap A_3) \\
& + p(A_1 \cap A_2 \cap A_3)
\end{aligned}
$$

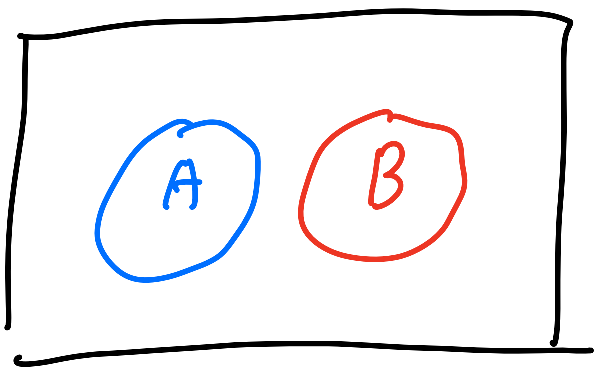
The addition rules for disjoint events

If the two events are mutually exclusive / disjoint, we have the following:

Theorem (Addition rule for mutually exclusive events)

If A and B are mutually exclusive events, then

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B).$$



Extention of the addition rule for mutually exclusive events

Suppose $A_1, A_2, \ldots$ is a sequence of pairwise mutually exclusive events. That is, A_i and A_j are mutually exclusive for all $i \neq j$. Then

$$P(\text{ at least one of the } A_i \text{ 's occurs }) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The complement rule and hierarchical rule

The following two important properties are derived from the aforementioned rules:

Theorem (The complement rule)

For an event A,

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

Theorem (The hierarchical rule)

If A implies B (i.e., $A \subseteq B$), then larger event, larger probability

$\{1\}$ $\{1,2\}$

$$P(A) \leq P(B).$$

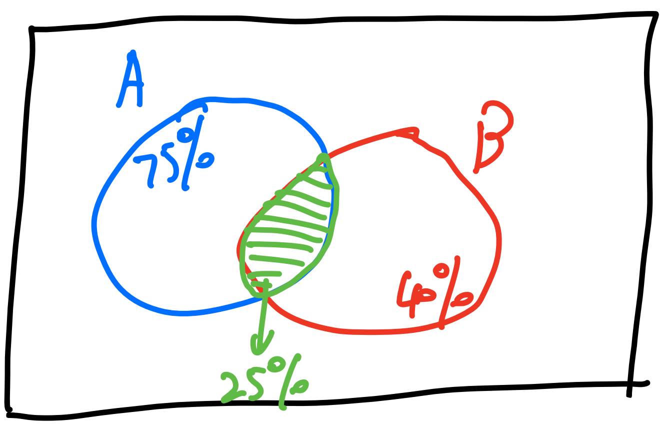
Example on probability calculation

event B

event A

Example: In a city, suppose 75% of the population have brown hair, 40% have brown eyes, and 25% have both brown hair and brown eyes. A person is chosen at random from the city. What is the probability that they

1. Have brown eyes or brown hair?
2. Have neither brown eyes nor brown hair?



$$P(A \cup B) = 75\% + 40\% - 25\%$$
$$= 90\%$$

Outline

Equally likely outcomes

The simplest probability model for a finite sample space is that all outcomes are equally likely.

- If Ω has k elements, then the probability of each outcome is $1/k$, as probabilities sum to 1 . That is, $P(\omega) = 1/k$, for all $\omega \in \Omega$.
- Suppose A is an event with s elements, with $s \leq k$. As $P(A)$ is the sum of the probabilities of all the outcomes contained in A,

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{k} = \frac{s}{k} = \frac{\text{Number of elements of } A}{\text{Number of elements of } \Omega}.$$

- In other words, probability with equally likely outcomes reduces to counting elements in A and Ω.

e.g. rolling a die $\Omega = \{1,2,3,4,5,6\}$

event $A = \{1,2,3\}$

$P(A) = \dfrac{\text{number of elements in } A}{\text{number of elements in } \Omega} = \dfrac{3}{6}$

Probability as Relative Frequency

The procedures described can be summarized into the following:

Theorem (Probability as Relative Frequency)

In cases where the sample space consists of equally likely elements, we can find this probability of event A by calculating the relative frequency of the event A in Ω :

$$P(A) = \frac{\#\ \textit{of outcomes in}\ A}{\textit{Total}\ \#\ \textit{of outcomes in the random experiment}} = \frac{n(A)}{n(\Omega)}$$

This is valid only if each element in Ω is equally likely.

The multiplication principle

Breakfast (3) Dinner (4)
Cereal rice
eggs pelking duck
Pancake steak
fish
12

The multiplication principle

- Consider a simple case: if there are m ways for one thing to happen, and n ways for a second thing to happen, there are $m \times n$ ways for both things to happen.
- More generally-and more formally-consider an n-element sequence $(a_1, a_2, \ldots, a_n)$. If there are k_1 possible values for the first element, k_2 possible values for the second element, ..., and k_n possible values for the nth element, there are $k_1 \times k_2 \times \cdots \times k_n$ possible sequences.

Remark:

- Note that this is also called the *fundamental principle of counting*.