Unit 01 notes

Events and outcomes

01 Theory

Events and outcomes – informally

- An **event** is a *description* of something that can happen.
- An **outcome** is a *complete description* of something that can happen.

All outcomes are events. An event is usually a *partial* description. Outcomes are events given with a *complete* description.

Here 'complete' and 'partial' are within the context of the **probability model**.

- ⚠ It can be misleading to say that an 'outcome' is an 'observation'.
    - 'Observations' occur in the *real world*, while 'outcomes' occur in the *model*.
    - To the extent the model is a good one, and the observation conveys *complete* information, we can say 'outcome' for the observation.

Notice:

- Because outcomes are *complete*, no two distinct outcomes could *actually happen* in a run of the experiment being modeled.

When an event happens, the *fact* that it has happened constitutes **information**.

Events and outcomes – mathematically

- The **sample space** is the *set of possible outcomes*, so it is the set of the complete descriptions of everything that can happen.
- An **event** is a *subset* of the sample space, so it is a *collection of outcomes*.

- For mathematicians: some "wild" subsets are not *valid* events. Problems with infinity and the continuum...

Notation

- Write S for the set of possible outcomes, $s \in S$ for a single outcome in S.
- Write $A, B, C, \cdots \subset S$ or $A_1, A_2, A_3, \cdots \subset S$ for some events, subsets of S.
- Write $\mathcal{F}$ for the collection of all events. This is frequently a *huge* set!
- Write $|A|$ for the **cardinality** or *size* of a set A, i.e. the *number of elements it contains*.

Using this notation, we can consider an *outcome itself as an event* by considering the "singleton" subset $\{\omega\} \subset S$ which contains that outcome alone.

02 Illustration

Example - Coin flipping

Flip a fair coin two times and record both results.

- *Outcomes:* sequences, like HH or TH.
- *Sample space:* all possible sequences, i.e. the set $S = \{HH, HT, TH, TT\}$.
- *Events:* for example:
  - $A = \{HH, HT\} =$ "first was heads"
  - $B = \{HT, TH\} =$ "exactly one heads"
  - $C = \{HT, TH, HH\} =$ "at least one heads"

With this setup, we may combine events in various ways to generate other events:

- *Complex events:* for example:
  - $A \cap B = \{HT\}$, or in words:

    $$\text{"first was heads" AND "exactly one heads"} = \text{"heads-then-tails"}$$

    Notice that the last one is a *complete description*, namely the *outcome* HT.
  - $A \cup B = \{HH, HT, TH\}$, or in words:

    $$\begin{gathered}\text{"first was heads" OR "exactly one heads"} \\ = \text{"starts with heads, else it's tails-then-heads"}\end{gathered}$$

Exercise - Coin flipping: counting subsets

Flip a fair coin five times and record the results.

How many elements are in the sample space? (How big is S?)
How many events are there? (How big is $\mathcal{F}$?)

Solution

03 Theory

New events from old

Given two events A and B, we can form new events using set operations:

$$A \cup B \quad \longleftrightarrow \quad \text{"event } A \text{ OR event } B\text{"}$$

$$A \cap B \quad \longleftrightarrow \quad \text{"event } A \text{ AND event } B\text{"}$$

$$A^c \quad \longleftrightarrow \quad \textbf{not} \text{ event } A$$

We also use these terms for events A and B:

- They are **mutually exclusive** when $A \cap B = \emptyset$, that is, they have *no elements in common*.
- They are **collectively exhaustive** $A \cup B = S$, that is, when they jointly *cover all possible outcomes*.

- In probability texts, sometimes $A \cap B$ is written "$A \cdot B$" or even (frequently!) "AB".

Rules for sets

Algebraic rules

- Associativity: $(A \cup B) \cup C = A \cup (B \cup C)$. Analogous to $(A+B)+C = A+(B+C)$.
- Distributivity: $A \cap (B \cup C) = (A \cap B) \ \cup \ (A \cap C)$. Analogous to $A(B+C) = AB + AC$.

De Morgan's Laws

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$
  In other words: you can distribute " ^{c} " but must simultaneously do a switch $\cap \leftrightarrow \cup$.

Probability models

04 Theory

Axioms of probability

A **probability measure** is a function $P : \mathcal{F} \to \mathbb{R}$ satisfying:

Kolmogorov Axioms:

- **Axiom 1:** $P[A] \geq 0$ for every event A
  (probabilities are not negative!)
- **Axiom 2:** $P[S] = 1$
  (probability of "anything" happening is 1)
- **Axiom 3:** additivity for any *countable collection* of *mutually exclusive* events:

$$P[A_1 \cup A_2 \cup A_3 \cup \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots$$
$$\text{when:} \quad A_i \cap A_j = \emptyset \quad \text{for all } \ i \neq j$$

- %& Notation: we write $P[A]$ instead of $P(A)$, even though P is a function, to emphasize the fact that A is a set.

Probability model

A **probability model** or **probability space** consists of a triple $(S, \mathcal{F}, P)$:

- S the sample space
- $\mathcal{F}$ the set of valid events, where every $A \in \mathcal{F}$ satisfies $A \subset S$
- $P : \mathcal{F} \to \mathbb{R}$ a probability measure satisfying the Kolmogorov Axioms

Finitely many exclusive events

It is a consequence of the Kolmogorov Axioms that additivity also works for finite collections of mutually exclusive events:

$$P[A \cup B] = P[A] + P[B]$$

$$P[A_1 \cup \cdots \cup A_n] = P[A_1] + \cdots + P[A_n]$$

Inferences from Kolmogorov

A probability measure satisfies these rules.
They can be deduced from the Kolmogorov Axioms.

- **Negation:** Can you find $P[A^c]$ but not $P[A]$? Use negation:
$$P[A] = 1 - P[A^c]$$
- **Monotonicity:** Probabilities grow when outcomes are added:
$$A \subset B \quad \gg\gg \quad P[A] \leq P[B]$$
- **Inclusion-Exclusion:** A trick for resolving unions:
$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$
(even when A and B are *not exclusive!*)

Inclusion-Exclusion

The principle of inclusion-exclusion generalizes to three events:

$$P[A \cup B \cup C] =$$

$$P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]$$

The same pattern works for any number of events!

The pattern goes: “include singles” then “exclude doubles” then “include triples” then ...

Include, exclude, include, exclude, include, ...

05 Illustration

Example - Lucia is Host or Player

The professor chooses three students at random for a game in a class of 40, one to be Host, one to be Player, one to be Judge. What is the probability that Lucia is either Host or Player?

Solution

1. Set up the probability model.
   - Label the students 1 to 40. Write L for Lucia’s number.
   - *Outcomes:* assignments such as $(H, P, J) = (2, 5, 8)$
     These are ordered triples with *distinct* entries in $1, 2, \ldots, 40$.
   - *Sample space:* S is the collection of all such distinct triples
   - *Events:* any subset of S
   - *Probability measure*: assume all outcomes are equally likely, so $P[(i, j, k)] = P[(r, l, p)]$ for all $i,\ j,\ k,\ r,\ l,\ p$
   - In total there are $40 \cdot 39 \cdot 38$ triples of distinct numbers.
   - Therefore $P[(i, j, k)] = \frac{1}{40 \cdot 39 \cdot 38}$ for any *specific* outcome (i, j, k).
   - Therefore $P[A] = \frac{|A|}{40 \cdot 39 \cdot 38}$ for any event A. (Recall $|A|$ is the *number* of outcomes in A.)

2. Define the desired event.
    - Want to find $P[\text{"Lucia is Host or Player"}]$
    - Define $A =$ "Lucia is Host" and $B =$ "Lucia is Player". Thus:

$$A = \big\{(L, j, k) \mid \text{any } j, k\big\}, \qquad B = \big\{(i, L, k) \mid \text{any } i, k\big\}$$

    - So we seek $P[A \cup B]$.
3. Compute the desired probability.
    - Importantly, $A \cap B = \emptyset$ (mutually exclusive).
      There are no outcomes in S in which Lucia is *both* Host and Player.
    - By *additivity*, we infer $P[A \cup B] = P[A] + P[B]$.
    - Now compute $P[A]$.
        - There are $39 \cdot 38$ ways to choose j and k from the students besides Lucia.
        - Therefore $|A| = 39 \cdot 38$.
        - Therefore:

$$P[A] \quad \gg\gg \quad \frac{|A|}{40 \cdot 39 \cdot 38} \quad \gg\gg \quad \frac{39 \cdot 38}{40 \cdot 39 \cdot 38} \quad \gg\gg \quad \frac{1}{40}$$

    - Now compute $P[B]$. It is similar: $P[B] = \frac{1}{40}$.
    - Finally compute that $P[A] + P[B] = \frac{1}{20}$, so the answer is:

$$P[A \cup B] \quad \gg\gg \quad P[A] + P[B] \quad \gg\gg \quad \frac{1}{20}$$

Example - iPhones and iPads

At Mr. Jefferson's University, 25% of students have an iPhone, 30% have an iPad, and 60% have neither.

What is the probability that a randomly chosen student has *some* iProduct? (Q1)

What about *both*? (Q2)

Solution

1. Set up the probability model.
    - A student is chosen at random: an *outcome* is the chosen student.
    - *Sample space* S is the set of all students.
    - Write $O =$ "has iPhone" and $A =$ "has iPad" concerning the chosen student.
    - All students are equally likely to be chosen: therefore $P[E] = \frac{|E|}{|S|}$ for any event E.
    - Therefore $P[O] = 0.25$ and $P[A] = 0.30$.
    - Furthermore, $P[O^c A^c] = 0.60$. This means 60% have "not iPhone AND not iPad".
2. Define the desired event.
    - Q1: desired event $= O \cup A$
    - Q2: desired event $= OA$
3. Compute the probabilities.
    - We do not believe O and A are exclusive.

- Try: apply inclusion-exclusion:

$$P[O \cup A] = P[O] + P[A] - P[OA]$$

- We know $P[O] = 0.25$ and $P[A] = 0.30$. So this formula, with given data, RELATES Q1 and Q2.
- Notice the complements in $O^c A^c$ and try *Negation*.
- *Negation:*

$$P[(OA)^c] = 1 - P[OA]$$

  DOESN'T HELP.
- Try again: *Negation:*

$$P[(O^c A^c)^c] = 1 - P[O^c A^c]$$

- And De Morgan (or a Venn diagram!):

$$(O^c A^c)^c \quad \gg\gg \quad O \cup A$$

- Therefore:

$$P[O \cup A] \quad \gg\gg \quad P[(O^c A^c)^c]$$

$$\gg\gg \quad 1 - P[O^c A^c] \quad \gg\gg \quad 1 - 0.6 = 0.4$$

- We have found Q1: $P[O \cup A] = 0.40$.
- Applying the RELATION from inclusion-exclusion, we get Q2:

$$P[O \cup A] = P[O] + P[A] - P[OA]$$

$$\gg\gg \quad 0.40 = 0.25 + 0.30 - P[OA]$$

$$\gg\gg \quad P[OA] = 0.15$$

Conditional probability

06 Theory

Conditional probability

The **conditional probability** of "B given A" is defined by:

$$P[B \mid A] = \frac{P[B \cap A]}{P[A]}$$

This conditional probability $P[B \mid A]$ represents the probability of event B taking place *given the assumption* that A took place. (All within the given probability model.)

By letting the actuality of event A be taken as a fixed hypothesis, we can define a *conditional probability measure* by plugging events into the slot of B:

$$P[- \mid A] = \frac{P[- \cap A]}{P[A]}$$

It is possible to verify each of the Kolmogorov axioms for this function, and therefore $P[- \mid A]$ itself defines a bona fide *probability measure*.

Conditioning

What does it really mean?

Conceptually, $P[B \mid A]$ corresponds to *creating a new experiment* in which we run the old experiment and record data *only those times that A happened*. Or, it corresponds to finding ourselves with *knowledge* or *data* that A happened, and we seek our best estimates of the likelihoods of other events, based on our existing model and the actuality of A.

Mathematically, $P[B \mid A]$ corresponds to *restricting* the probability function to outcomes in A, and *renormalizing* the values (dividing by $p[A]$) so that the total probability of all the outcomes (in A) is now 1.

The definition of conditional probability can also be turned around and reinterpreted:

Multiplication rule

$$P[AB] = P[A] \cdot P[B \mid A]$$

"The probability of A AND B equals the probability of A *times* the probability of B-given-A."

This principle generalizes to any events in sequence:

Generalized multiplication rule

$$P[A_1 A_2 A_3] = P[A_1] \cdot P[A_2 \mid A_1] \cdot P[A_3 \mid A_1 A_2]$$

$$P[A_1 \cdots A_n] = P[A_1] \cdot P[A_2 \mid A_1] \cdot P[A_3 \mid A_1 A_2] \; \cdots \; P[A_n \mid A_1 \cdots A_{n-1}]$$

The generalized rule can be verified like this. First substitute A_2 for B and A_1 for A in the original rule. Now repeat, substituting A_3 for B and $A_1 A_2$ for A in the original rule, and combine with the first one, and you find the rule for triples. Repeat again with A_4 and $A_1 A_2 A_3$, combine with the triples, and you get quadruples.

07 Illustration

Exercise - Simplifying conditionals

Let $A \subset B$. Simplify the following values:

$$P[A \mid B], \quad P[A \mid B^c], \quad P[B \mid A], \quad P[B \mid A^c]$$

Solution

Example - Coin flipping: at least 2 heads

Flip a fair coin 4 times and record the outcomes as sequences, like $HHTH$.

Let $A_{\geq 2}$ be the event that there are at least two heads, and $A_{\geq 1}$ the event that there is at least one heads.

First let's calculate $P[A_{\geq 2}]$.

Define A_2, the event that there were exactly 2 heads, and A_3, the event of exactly 3, and A_4 the event of exactly 4. These events are exclusive, so:

$$P[A_{\geq 2}] = P[A_2 \cup A_3 \cup A_4] \quad \gg\gg \quad P[A_2] + P[A_3] + P[A_4]$$

Each term on the right can be calculated by counting:

$$P[A_2] = \frac{|A_2|}{2^4} \quad \gg\gg \quad \frac{\binom{4}{2}}{16} \quad \gg\gg \quad \frac{6}{16}$$

$$P[A_3] = \frac{|A_3|}{2^4} \quad \gg\gg \quad \frac{\binom{4}{1}}{16} \quad \gg\gg \quad \frac{4}{16}$$

$$P[A_4] = \frac{|A_4|}{2^4} \quad \gg\gg \quad \frac{\binom{4}{0}}{16} \quad \gg\gg \quad \frac{1}{16}$$

Therefore, $P[A_{\geq 2}] = \frac{11}{16}$.

Now suppose we find out that "at least one heads definitely came up". (Meaning that we know $A_{\geq 1}$.) For example, our friend is running the experiment and tells us this fact about the outcome.

Now what is our estimate of likelihood of $A_{\geq 2}$?

The formula for conditioning gives:

$$P[A_{\geq 2} \mid A_{\geq 1}] = \frac{P[A_{\geq 2} \cap A_{\geq 1}]}{P[A_{\geq 1}]}$$

Now $A_{\geq 2} \cap A_{\geq 1} = A_{\geq 2}$. (Any outcome with at least two heads automatically has at least one heads.) We already found that $P[A_{\geq 2}] = \frac{11}{16}$. To compute $P[A_{\geq 1}]$ we simply *add* the probability $P[A_1]$, which is $\frac{4}{16}$, to get $P[A_{\geq 1}] = \frac{15}{16}$.

Therefore:

$$P[A_{\geq 2} \mid A_{\geq 1}] = \frac{11/16}{15/16} \quad \gg\gg \quad \frac{11}{15}$$

Example: Flip a coin, then roll dice

Flip a coin. If the outcome is heads, roll two dice and add the numbers. If the outcome is tails, roll a single die and take that number. What is the probability of getting a tails AND a number at least 3?

Solution

This "two-stage" experiment lends itself to a solution using the multiplication rule for conditional probability.

1. Label the events of interest.
   - Let H and T be the events that the coin showed heads and tails, respectively.
   - Let $A_1, \ldots, A_{12}$ be the events that the final number is $1, \ldots, 12$, respectively.
   - The value we seek is $P[TA_{\geq 3}]$.
2. Observe known (conditional) probabilities.
   - We know that $P[H] = \frac{1}{2}$ and $P[T] = \frac{1}{2}$.
   - We know that $P[A_5 \mid T] = \frac{1}{6}$, for example, or that $P[A_2 \mid H] = \frac{1}{36}$.
3. Apply "multiplication" rule.

- This rule gives:

$$P[TA_{\geq 3}] = P[T] \cdot P[A_{\geq 3} \mid T]$$

- We know $P[T] = \frac{1}{2}$ and can see by counting that $P[A_{\geq 3} \mid T] = \frac{2}{3}$.
- Therefore $P[TA_{\geq 3}] = \frac{1}{3}$.

Multiplication: draw two cards

Two cards are drawn from a standard deck (without replacement).

What is the probability that the first is a 3, and the second is a 4?

Solution

This "two-stage" experiment lends itself to a solution using the multiplication rule for conditional probability.

1. Label events.
    - Write T for the event that the first card is a 3
    - Write F for the event that the second card is a 4.
    - We seek $P[TF]$.
2. Write down knowns.
    - We know $P[T] = \frac{4}{52}$. (It does not depend on the second draw.)
    - Easily find $P[F \mid T]$.
        - If the first is a 3, then there are four 4s remaining and 51 cards.
        - So $P[F \mid T] = \frac{4}{51}$.
3. Apply multiplication rule.
    - Multiplication rule:

$$P[TF] = P[T] \cdot P[F \mid T]$$

$$P[TF] = \frac{4}{52} \cdot \frac{4}{51} \quad \gg\gg \quad \frac{4}{13 \cdot 51}$$

    - Therefore $P[TF] = \frac{4}{663}$

08 Theory

Division into Cases

For any events A and B:

$$P[B] = P[A] \cdot P[B \mid A] + P[A^c] \cdot P[B \mid A^c]$$

Interpretation: event B may be *divided along the lines of A*, with some of $P[B]$ coming from the part in A and the rest from the part in A^c.

Total Probability - Explanation

- First divide B itself into parts in and out of A:

$$B = B \cap A \bigcup B \cap A^c$$

- These parts are exclusive, so in probability we have:

$$P[B] = P[BA] + P[BA^c]$$

- Use the Multiplication rule to break up $P[BA]$ and $P[BA^c]$:

$$P[BA] \quad \gg\gg \quad P[A] \cdot P[B \mid A]$$

$$P[BA^c] \quad \gg\gg \quad P[A^c] \cdot P[B \mid A^c]$$

- Now substitute in the prior formula:

$$P[B] \quad \gg\gg \quad P[BA] + P[BA^c] \quad \gg\gg \quad P[A] \cdot P[B \mid A] + P[A^c] \cdot P[B \mid A^c]$$

This law can be generalized to any **partition** of the sample space S. A partition is a collection of events A_i which are *mutually exclusive* and *jointly exhaustive*:

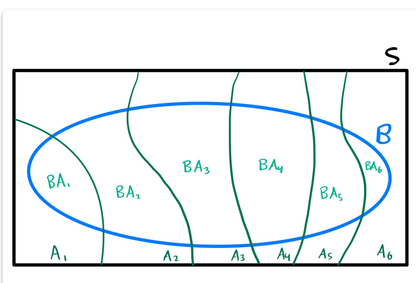
$$A_i \cap A_j = \emptyset, \qquad \bigcup_i A_i = S$$

The generalized formulation of Total Probability for a partition is:

Law of Total Probability

For a partition A_i of the sample space S:

$$P[B] = \sum_i P[A_i] \cdot P[B \mid A_i]$$



Division into Cases is just the *Law of Total Probability* after setting $A_1 = A$ and $A_2 = A^c$.

09 Illustration

Exercise - Marble transferred, marble drawn

Setup:

- Bin 1 holds five red and four green marbles.
- Bin 2 holds four red and five green marbles.

Experiment:

- You take a random marble from Bin 1 and put it in Bin 2 and shake Bin 2.
- Then you draw a random marble from Bin 2 and look at it.

What is the probability that the marble you look at is red?

Solution

Bayes' Theorem

10 Theory

Bayes' Theorem

For any events A and B:

$$P[B \mid A] = P[A \mid B] \cdot \frac{P[B]}{P[A]}$$

- ⚠ Bayes' Theorem is also called Bayes' Rule sometimes.

Bayes' Theorem - Derivation

Start with the observation that $AB = BA$, or event "A AND B" equals event "B AND A".

Apply the *multiplication rule* to each of order:

$$P[AB] = P[A] \cdot P[B \mid A]$$

$$P[BA] = P[B] \cdot P[A \mid B]$$

Equate them and rearrange:

$$P[AB] = P[BA] \quad \gg\gg \quad P[A] \cdot P[B \mid A] = P[B] \cdot P[A \mid B]$$

$$\gg\gg \quad P[B \mid A] = P[A \mid B] \cdot \frac{P[B]}{P[A]}$$

The main application of Bayes' Theorem is to calculate $P[A \mid B]$ when it is easy to calculate $P[B \mid A]$ from the problem setup. Often this occurs in **multi-stage experiments** where event A describes outcomes of an intermediate stage.

Note: these notes use *alphabetical order* A, B as a mnemonic for *temporal or logical order*, i.e. that A comes *first* in time, or that otherwise that A is the *prior* conditional from which it is easier to calculate B.

11 Illustration

Example - Bayes' Theorem - COVID tests

Assume that 0.5% of people have COVID. Suppose a COVID test gives a (true) positive on 96% of patients who have COVID, but gives a (false) positive on 2% of patients who do not have COVID. Bob tests positive. What is the probability that Bob has COVID?

Solution

1. Label events.
    - Event A_P: Bob is actually positive for COVID

- Event A_N: Bob is actually negative; note $A_N = A_P^c$
- Event T_P: Bob tests positive
- Event T_N: Bob tests negative; note $T_N = T_P^c$

2. Identify knowns.
   - Know: $P[T_P \mid A_P] = 96\%$
   - Know: $P[T_P \mid A_N] = 2\%$
   - Know: $P[A_P] = 0.5\%$ and therefore $P[A_N] = 99.5\%$
   - We seek: $P[A_P \mid T_P]$
3. Translate Bayes' Theorem.
   - Using $A = T_P$ and $B = A_P$ in the formula:

   $$P[A_P \mid T_P] = P[T_P \mid A_P] \cdot \frac{P[A_P]}{P[T_P]}$$

   - We know all values on the right except $P[T_P]$
4. Use Division into Cases.
   - Observe:

   $$T_P = T_P \cap A_P \bigcup T_P \cap A_N$$

   - Division into Cases yields:

   $$P[T_P] = P[A_P] \cdot P[T_P \mid A_P] + P[A_N] \cdot P[T_P \mid A_N]$$

   - Important to notice this technique!
     - It is a common element of Bayes' Theorem application problems.
     - It is frequently needed *for the denominator*.
   - Plug in data and compute:

   $$\gg\gg \quad P[T_P] = \frac{5}{1000} \cdot \frac{96}{100} + \frac{995}{1000} \cdot \frac{2}{100} \quad \gg\gg \quad \approx 0.0247$$

5. Compute answer.
   - Plug in and compute:

   $$P[A_P \mid T_P] = P[T_P \mid A_P] \cdot \frac{P[A_P]}{P[T_P]}$$

   $$\gg\gg \quad 0.96 \cdot \frac{0.005}{0.0247} \quad \gg\gg \quad \approx 19\%$$

Intuition - COVID testing

Some people find the low number surprising. In order to repair your intuition, think about it like this: roughly 2.5% of tests are positive, with roughly 2% coming from *false* positives, and roughly 0.5% from *true* positives. The true ones make up only 1/5 of the positive results!

(This rough approximation is by assuming $96\% = 100\%$.)

If *two* tests both come back positive, the odds of COVID are now 98%.

If *only people with symptoms* are tested, so that, say, 20% of those tested have COVID, that is, $P[A_P \mid T_P] = 20\%$, then one positive test implies a COVID probability of 92%.

Exercise - Bayes' Theorem and Multiplication: Inferring bin from marble

There are marbles in bins in a room:

- Bin 1 holds 7 red and 5 green marbles.
- Bin 2 holds 4 red and 3 green marbles.

Your friend goes in the room, shuts the door, and selects a random bin, then draws a random marble. (Equal odds for each bin, then equal odds for each marble in that bin.) He comes out and shows you a red marble.

What is the probability that this red marble was taken from Bin 1?

Solution

Independence

12 Theory

Two events are independent when information about one of them does not change our probability estimate for the other. Mathematically, there are three ways to express this fact:

Independence

Events A and B are **independent** when these (logically equivalent) equations hold:

- $P[B \mid A] = P[B]$
- $P[A \mid B] = P[A]$
- $P[BA] = P[B] \cdot P[A]$

- The last equation is symmetric in A and B.
  - Check: $BA = AB$ and $P[B] \cdot P[A] = P[A] \cdot P[B]$
  - This symmetric version is the preferred definition of the concept.

Multiple-independence

A *collection* of events $A_1, \ldots, A_n$ is **mutually independent** when every subcollection $A_{i_1}, \ldots, A_{i_k}$ satisfies:

$$P[A_{i_1} \cdots A_{i_k}] = P[A_{i_1}] \cdots P[A_{i_k}]$$

A potentially *weaker condition* for a collection $A_1, \ldots, A_n$ is called **pairwise independence**, which holds when all 2-member subcollections are independent:

$$P[A_i A_j] = P[A_i] \cdot P[A_j] \quad \text{for all } i \neq j$$

One could also define 3-member independence, or n-member independence. Plain 'independence' means *any*-member independence.

13 Illustration

Exercise - Independence and complements

Prove that these are logically equivalent statements:

- A and B are independent
- A and B^c are independent
- A^c and B^c are independent

Make sure you demonstrate both directions of each equivalency.

Solution

Example - Checking independence by hand

A bin contains 4 red and 7 green marbles. Two marbles are drawn.

Let R_1 be the event that the first marble is red, and let G_2 be the event that the second marble is green.

- (a) Show that R_1 and G_2 are independent if the marbles are drawn *with replacement*.
- (b) Show that R_1 and G_2 are not independent if the marbles are drawn *without replacement*.

Solution

(a) With replacement.

1. Identify knowns.
    - Know: $P[R_1] = \frac{4}{11}$
    - Know: $P[G_2] = \frac{7}{11}$
2. Compute both sides of independence relation.
    - Relation is $P[R_1 G_2] = P[R_1] \cdot P[G_2]$
    - Right side is $\frac{4}{11} \cdot \frac{7}{11}$
    - For $P[R_1 G_2]$, have $4 \cdot 7$ ways to get $R_1 G_2$, and 11^2 total outcomes.
    - So left side is $\frac{4 \cdot 7}{11^2}$, which equals the right side.

(b) Without replacement.

1. Identify knowns.
    - Know: $P[R_1] = \frac{4}{11}$ and therefore $P[R_1^c] = \frac{7}{11}$
    - We seek: $P[G_2]$ and $P[R_1 G_2]$
2. Find $P[G_2]$ using Division into Cases.
    - Division into cases:

    $$G_2 = G_2 \cap R_1 \bigcup G_2 \cap R_1^c$$

    - Therefore:

    $$P[G_2] = P[R_1] \cdot P[G_2 \mid R_1] + P[R_1^c] \cdot P[G_2 \mid R_1^c]$$

    - Find these by counting and compute:

    $$\gg\gg \quad P[G_2] = \frac{4}{11} \cdot \frac{7}{10} + \frac{7}{11} \cdot \frac{6}{10} \quad \gg\gg \quad \frac{70}{110}$$

3. Find $P[R_1G_2]$ using Multiplication rule.
   - Multiplication rule (implicitly used above already):

$$P[R_1G_2] = P[R_1] \cdot P[G_2 \mid R_1] \quad \gg\gg \quad \frac{4}{11} \cdot \frac{7}{10} \quad \gg\gg \quad \frac{28}{110}$$

4. Compare both sides.
   - Left side: $P[R_1G_2] = \frac{28}{110}$
   - Whereas, right side:

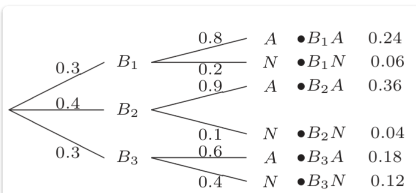
$$P[R_1] \cdot P[G_2] = \frac{4}{11} \cdot \frac{70}{110} = \frac{28}{121}$$

   - But $\frac{28}{110} \neq \frac{28}{121}$ so $P[R_1G_2] \neq P[R_1] \cdot P[G_2]$ and they are *not independent.*

Tree diagrams

14 Theory

A **tree diagram** depicts the components of a **multi-stage experiment**. Nodes, or *branch points*, represent sources of randomness.



An *outcome* of the experiment is represented by a *pathway* taken from the root (left-most node) to a leaf (right-most node). The branch chosen at a given node junction represents the outcome of the "sub-experiment" constituting that branch point. So a pathway encodes the outcomes of all sub-experiments.

Each branch from a node is labeled with a probability number. This is the probability that the sub-experiment of that node has the outcome of that branch.

- The probability label on some branch is the conditional probability of that branch, assuming the pathway from root to prior node.
  - In the example: $0.8 = P[A \mid B_1]$.
  - Therefore, branch labels from given node sum to 1. (Law of Total Probability)
- The probability of a given (overall) outcome is the *product* of the probabilities on each branch of the pathway to that outcome.
  - Makes sense, because (e.g.): $P[AB_1] = P[A] \cdot P[B_1 \mid A]$
  - More generally: remember that (e.g.): $P[ABCD] = P[ABC] \cdot P[D \mid ABC]$
  - This overall outcome probability may be written at the leaf.

One can also use a tree diagram to remember quickly how to calculate certain probabilities.

For example, what is $P[A]$ in the diagram?
Answer: add up the pathway probabilities (leaf numbers) terminating in A. That makes $0.24 + 0.36 + 0.18 = 0.78$

For example, what is $P[B_1 \mid N]$?
Answer: divide the leaf probability of B_1N by the total probability of N. That makes:

$$P[B_1 \mid N] = \frac{0.06}{0.06 + 0.04 + 0.12} \approx 0.27$$

15 Illustration

Example - Tree diagrams: Marble transferred, marble drawn

Setup:

- Bin 1 holds five red and four green marbles.
- Bin 2 holds four red and five green marbles.

Experiment:

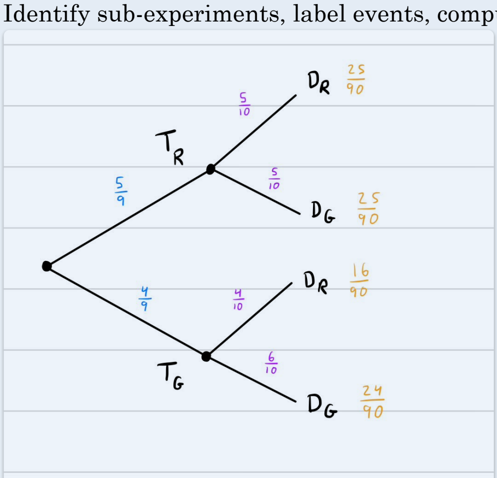
- You take a random marble from Bin 1 and put it in Bin 2 and shake Bin 2.
- Then you draw a random marble from Bin 2 and look at it.

Questions:

- (a) What is the probability you *draw* a red marble?
- (b) Supposing that you drew a red marble, what is the probability that a red marble was *transferred*?

Solution

1. Construct the tree diagram.
   - Identify sub-experiments, label events, compute probabilities:



2. For (a), compute $P[D_R]$.
   - Add up leaf numbers for D_R at leaf:

$$P[D_R] = \frac{25}{90} + \frac{16}{90} = \frac{41}{90}$$

3. For (b), compute $P[T_R \mid D_R]$.
   - Conditional probability:

$$P[T_R \mid D_R] = \frac{P[T_R D_R]}{P[D_R]}$$

   - Plug in data and compute:

$$\gg\gg \quad \frac{25/90}{41/90} \quad \gg\gg \quad \frac{25}{41}$$

- Interpretation: mass of desired pathway over mass of possible pathways.

Counting

16 Theory

In many "games of chance", it is assumed by symmetry principles that all outcomes are equally likely. From this assumption we infer the rule for $P[-]$:

$$P[A] = \frac{|A|}{|S|}$$

In words: the probability of event A is the number of outcomes in A divided by the number of possible outcomes.

When this formula applies, it is important to be able to count total outcomes, as well as outcomes satisfying various conditions.

Permutations

Permutations count the number of *ordered lists* one can form from some items. For a list of r items taken from a total collection of n, the number of permutations is:

$$\frac{n!}{(n-r)!}$$

To see where this comes from:
There are n choices for the first item, then $n-1$ for the second, then ... then $n-r+1$ for the r^{th} item. So the number is $n(n-1)(n-2)\cdots(n-r+1)$. Observe:

$$\frac{n!}{(n-r)!} = \frac{n(n-1)(n-2)\cdots(n-r+1)(n-r)(n-r-1)\cdots 1}{(n-r)(n-r-1)\cdots 1}$$

$$\gg\gg \quad n(n-1)(n-2)\cdots(n-r+1)$$

Combinations, binomial coefficient

Combinations count the number of *sets* (ignoring order) one can form from some items. We define a notation for it like this:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

This counts the number of sets of r distinct elements taken from a total collection of n items.

Another name for combinations is the **binomial coefficient**.

This formula can be derived from the formula for permutations. The possible permutations can be partitioned into combinations: each combination gives a set, and by specifying an ordering of elements in the set, we get a permutation. For a set of r elements taken from n items, there are $r!$ ways to put them into a specific order. So the number of permutations must be a factor of $r!$ greater than the number of combinations.

This notation, $\binom{n}{r}$, is also called the **binomial coefficient** because it provides the coefficients of a binomial expansion:

$$(x+y)^n = \sum_{i=1}^{n} \binom{n}{i} x^{n-i} y^i$$

For example:

$$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$$

There are also 'higher' combinations:

Multinomial coefficient

The general multinomial coefficient is defined by the formula:

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}$$

where $r_1 + r_2 + \cdots + r_k = n$.

The multinomial coefficient measures the number of ways to partition n items into sets with sizes r_1, r_2, …, r_k, respectively.

Notice that $\binom{5}{3,2} = \binom{5}{3}$ so we already defined these values $(k=2)$ with binomial coefficients. But with $k > 2$, we have new values. They correspond to the coefficients in multinomial expansions. For example $k=3$ gives coefficients for $(x+y+z)^n$.

17 Illustration

Exercise - Combinations: Counting teams with Cooper

A team of 3 student volunteers is formed at random from a class of 40. What is the probability that Cooper is on the team?

Solution

Example - Combinations: Groups with Haley and Hugo

The class has 40 students. Suppose the professor chooses 3 students Wednesday at random, and again 3 on Friday. What is the probability that Haley is chosen today and Hugo on Friday?

Solution

1. Count total outcomes.
    - Have $\binom{40}{3}$ possible groups chosen Wednesday.
    - Have $\binom{40}{3}$ possible groups chosen Friday.
    - Therefore $\binom{40}{3} \times \binom{40}{3}$ possible groups in total.
2. Count desired outcomes.
    - Groups of 3 with Haley are same as groups of 2 taken from others.
    - Therefore have $\binom{39}{2}$ groups that contain Haley.

- Have $\binom{39}{2}$ groups that contain Hugo.
- Therefore $\binom{39}{2} \times \binom{39}{2}$ total desired outcomes.

3. Compute probability.
   - Let E label the desired event.
   - Use formula:

$$P[E] = \frac{|E|}{|S|}$$

   - Therefore:

$$P[E] \quad \gg\gg \quad \frac{\binom{39}{2} \times \binom{39}{2}}{\binom{40}{3} \times \binom{40}{3}}$$

$$\gg\gg \quad \left(\frac{\frac{39 \cdot 38}{2!}}{\frac{40 \cdot 39 \cdot 38}{3!}} \right)^2 \quad \gg\gg \quad \left(\frac{3}{40} \right)^2$$

Example - Counting VA license plates

A VA license plate has three letters (with no I, O, or Q) followed by four numerals. A random plate is seen on the road.

- (a) What is the probability that the numerals are in increasing order?
- (b) What is the probability that at least one number is repeated?

Solution

(a)

1. Count ways to have 4 numerals in increasing order.
   - Any four distinct numerals have a single order that's increasing.
   - There are $\binom{10}{4}$ ways to choose 4 numerals from 10 options.
2. Count ways to have 3 letters in order except I, O, Q.
   - 26 total letters, 3 excluded, thus 23 options.
   - Repetition allowed, thus $23 \cdot 23 \cdot 23 = 23^3$ possibilities.
3. Count total plates with increasing numerals.
   - Multiply the options:

$$23^3 \cdot \binom{10}{4}$$

4. Count total plates.
   - Have $23 \cdot 23 \cdot 23$ options for letters.
   - Have $10 \cdot 10 \cdot 10 \cdot 10$ options for numbers.
   - Thus $23^3 \cdot 10^4$ possible plates.
5. Compute probability.
   - Let E label the event that a plate has increasing numerals.

- Use the formula:

$$P[E] = \frac{|E|}{|S|}$$

- Therefore:

$$P[E] \quad \gg\gg \quad \frac{23^3 \cdot \binom{10}{4}}{23^3 \cdot 10^4} \quad \gg\gg \quad \frac{\frac{10!}{4!6!}}{10000} \quad \gg\gg \quad \frac{21}{1000}$$

(b)

1. Count plates with at least one number repeated.
   - "At least" is hard! Try *complement*: "no repeats".
   - Let E^c be event that *no* numbers are repeated. All distinct.
   - Count possibilities:

   $$|E^c| = 23 \cdot 23 \cdot 23 \cdot 10 \cdot 9 \cdot 8 \cdot 7$$

   - Total license plates is still $23^3 \cdot 10^4$.
   - Therefore, license plates with *at least one number repeated*:

   $$|E| = |S| - |E|$$

   $$\gg\gg \quad 23^3 \cdot 10^4 - 23^3 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \quad \gg\gg \quad 60348320$$

2. Compute probability.
   - Desired outcomes over total outcomes:

   $$\frac{|E|}{|S|} \quad \gg\gg \quad \frac{60348320}{23^3 \cdot 10^4} \quad \gg\gg \quad 0.496$$