Week 02 notes

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 sometimes called Bayes' Rule.

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 $P[R_1G_2] = P[R_1] \cdot P[G_2]$
    - Right side is $\frac{4}{11} \cdot \frac{7}{11}$
    - For $P[R_1G_2]$, have $4 \cdot 7$ ways to get R_1G_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 $P[R_1G_2]$
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: $P[R_1 G_2] = \frac{28}{110}$
   - Whereas, right side:

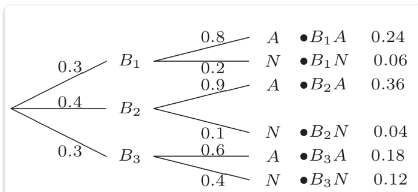
$$P[R_1] \cdot P[G_2] = \frac{4}{11} \cdot \frac{70}{110} = \frac{28}{121}$$

   - But $\frac{28}{110} \neq \frac{28}{121}$ so $P[R_1 G_2] \neq P[R_1] \cdot P[G_2]$ 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 $B_1 N$ 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:

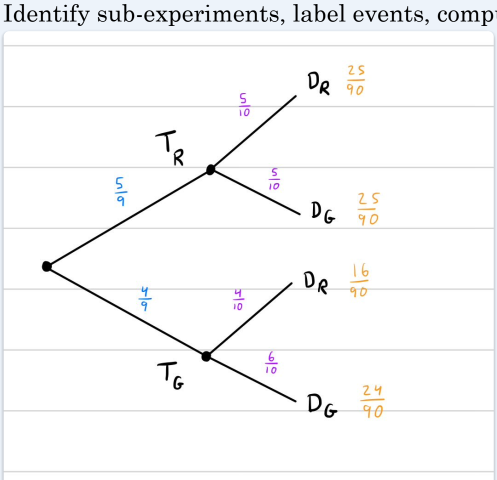
- You take a random marble from Bin 1 and put it in Bin 2 and shake Bin 2.
- Then you draw a random marble from Bin 2 and look at it.

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:

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