W02 Notes

Bayes' Theorem

01 Theory

Bayes' Theorem

For any events A and B:

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

Note: *Bayes' Theorem* is sometimes called *Bayes' Rule*.

Bayes' Theorem - Derivation

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

Apply the *multiplication rule* to each product:

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

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

Equate them and rearrange:

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

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

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

02 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 known data:

- 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:

Set $A = T_P$ and $B = A_P$ in Bayes':

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

We know all values on the right except $P[T_P]$

---

(4) Denominator: apply Total Probability (Division into Cases):

Observe that $T_P \cap A_P$ and $T_P \cap A_N$ are *exclusive* events, and that:

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

Therefore:

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

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) Plug in and compute:

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

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

Intuition - COVID testing

Some people find this 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. Only 1/5 of all the positive results are true ones!

(This rough approximation assumes that $96\% \approx 100\%$.)

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%.

Practice exercise

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 >

(1) Label events:

- Event B_1: friend chooses Bin 1.
- Event B_2: friend chooses Bin 2.
- Event R: friend draws a red marble.
- Event G: friend draws a green marble.

Answer will be $P[B_1 \mid R]$.

---

(2) Identify knowns:

- Know $P[R \mid B_1] = 7/12$.
- Know $P[G \mid B_1] = 5/12$.
- Know $P[R \mid B_2] = 4/7$.
- Know $P[G \mid B_2] = 3/7$.
- Know $P[B_1] = P[B_2] = 1/2$.

---

(3) Apply Bayes' Theorem for $P[B_1 \mid R]$::

$$P[B_1 \mid R] = P[R \mid B_1] \cdot \frac{P[B_1]}{P[R]}$$

Division into Cases for the denominator:

$$P[R] = P[B_1] \cdot P[R \mid B_1] + P[B_2] \cdot P[R \mid B_2]$$

---

(3) Plug in data and compute:

$$P[R] \quad \gg\gg \quad \frac{1}{2}\cdot\frac{7}{12}+\frac{1}{2}\cdot\frac{4}{7} \quad \gg\gg \quad \frac{97}{168}$$

$$P[B_1 \mid R] \quad \gg\gg \quad \frac{7}{12}\cdot\frac{1/2}{97/168} \quad \gg\gg \quad \frac{49}{97}$$

Independence

03 Theory

Two events are independent when information about one of them does not change our probability estimate for the other.

Independence

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

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

Note that 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 of independence.

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.

04 Illustration

Practice exercise

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 >

(1) Show that $P[AB] = P[A] \cdot P[B] \iff P[AB^c] = P[A] \cdot P[B^c]$

Assume $P[AB] = P[A] \cdot P[B]$ and show $P[AB^c] = P[A] \cdot P[B^c]$.

Divide A into the B cases:

$$P[A] = P[AB] + P[AB^c]$$

Apply the assumption:

$$\gg\gg \quad P[A] = P[A] \cdot P[B] + P[AB^c]$$

Algebra:

$$\gg\gg \quad P[A]\big(1 - P[B]\big) = P[AB^c]$$

Negation rule:

$$\gg\gg \quad P[A] \cdot P[B^c] = P[AB^c]$$

Assume $P[AB^c] = P[A] \cdot P[B^c]$ and show $P[AB] = P[A] \cdot P[B]$.

In the above sequence, apply *this* assumption to break up the *second* term instead.

---

(2) Show that $P[AB^c] = P[A] \cdot P[B^c]$ and $P[A^cB^c] = P[A^c] \cdot P[B^c]$ are equivalent.

To do this, simply notice that in the first equivalence we can replace A with B^c and B with A. Use $AB = BA$ too.

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.

Identify knowns:

- Know: $P[R_1] = \frac{4}{11}$
- Know: $P[G_2] = \frac{7}{11}$

Now compute both sides of independence relation:

$$P[R_1G_2] \;=\; P[R_1]\cdot P[G_2]$$

The right side is $\frac{4}{11}\cdot\frac{7}{11}$.

For $P[R_1G_2]$, we 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. This is a bit harder.

(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 Total Probability (Division into Cases):

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

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

Find RHS factors by counting, then 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:

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

---

(4) Compare both sides:

Left side: $P[R_1G_2]=\frac{28}{110}$.

Right side:

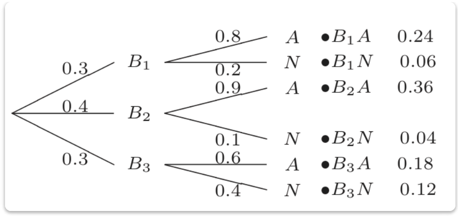
$$P[R_1]\cdot P[G_2] \;=\; \frac{4}{11}\cdot\frac{70}{110}=\frac{28}{121}$$

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

05 Theory

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



An *outcome* of the experiment is represented by a *complete path* taken from the root (left-most node, only one option) to a leaf (right-most node, many options). The branch chosen at a given node represents the outcome of a "sub-experiment." So a complete path encodes the outcomes of all sub-experiments along the way.

Each branch emanating from a node is labeled with a probability value. This is the probability that the sub-experiment of that node has the outcome of that branch. (In the example, $0.8 = P[A \mid B_1]$.) This is also the conditional probability of the branch's right node, given its left node as known.

Therefore, **branch values from any given node must sum to 1**.

The probability of a given outcome is the *product* of the probabilities along each branch of the path from the root to that outcome.

For example, for outcome AB_1, we have $P[AB_1] = P[A] \cdot P[B_1 \mid A]$.

Generally, remember that

$$\begin{aligned} P[ABCD] &= P[ABC] \cdot P[D \mid ABC] \\ &= P[AB] \cdot P[C \mid AB] \cdot P[D \mid ABC] \\ &= P[A] \cdot P[B \mid A] \cdot P[C \mid AB] \cdot P[D \mid ABC] \end{aligned}$$

This overall outcome probability may be written at the final leaf. (Not to be confused with the branch value of the last branch.)

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 path probabilities for all paths terminating in A. We obtain:

$$0.24 + 0.36 + 0.18 \quad \gg\gg \quad 0.78$$

For example, what is $P[B_1 \mid N]$?

- Answer: divide the leaf probability of B_1N by the total probability of N. We obtain:

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

06 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:

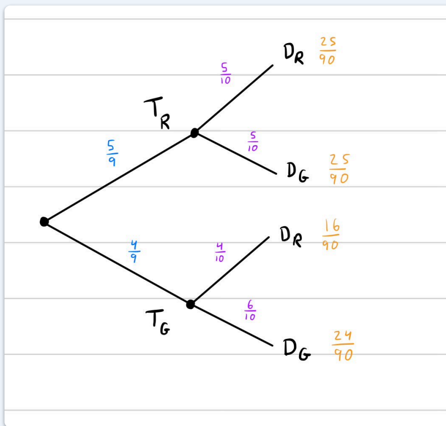
(a) What is the probability you *draw* a red marble?

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

Solution

Construct the tree diagram:

Identify sub-experiments, label events, compute probabilities:



(a) Compute $P[D_R]$:

Add up leaf numbers for D_R at leaf:

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

(b) Compute $P[T_R \mid D_R]$:

Conditional probability definition:

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

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

Interpretation: value of desired path over value of all possible paths.

Counting

07 Theory

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

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

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

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

Permutations

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

$$P(n,r) \;=\; \frac{n!}{(n-r)!}$$

Why is this formula true?

There are n choices for the first item. Then $n-1$ for the second item, after the first has been chosen and removed from the set of possibilities. Then $n-2$ for the third, then ..., then $n-r+1$ for the r^{th} item. So the total number of possibilities is the product:

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

We can express this with factorials using a technical observation:

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

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

Combinations

Combinations count the number of *subsets* (ignoring order) one can form from some items. For a subset of r items taken from a total collection of n items, the number of combinations is:

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

This formula can be derived from the formula for permutations.

The set of possible permutations can be *partitioned* into combinations: each combination determines a subset. By additionally specifying an *ordering* of the elements in a chosen subset, we obtain a permutation. For a given subset of r elements taken from n items, there are $r!$ ways to determine an ordering of them in a list. Therefore, the number of permutations must be a factor of $r!$ times the number of combinations. (For every combination of size r, there are $r!$ ways to order the items in a list.)

The notation $\binom{n}{r}$ is also called the **binomial coefficient** because it provides the coefficient values 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 that give **multinomial coefficients**:

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 $r_i \in \mathbb{N}$ and $r_1 + r_2 + \cdots + r_k = n$.

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

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

08 Illustration

Practice exercise

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

There are $\binom{39}{2}$ teams that include Cooper, and $\binom{40}{3}$ teams in total. So we have:

$$P = \frac{39!}{2!37!} \Big/ \frac{40!}{3!37!} = \frac{3}{40}$$

Example - Combinations: Groups with Haley and Hugo

A UVA class has 40 students. Suppose the professor chooses 3 students on 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:

- We have $\binom{40}{3}$ possible groups chosen Wednesday.
- We have $\binom{40}{3}$ possible groups chosen Friday.

Therefore $\binom{40}{3} \times \binom{40}{3}$ possible groups in total. (Product of possibilities.)

---

(2) Count desired outcomes:

The possible groups of 3 that include Haley can be counted by counting the subgroups of 2 formed of the *other* students in Haley's group.

- Therefore we have $\binom{39}{2}$ groups that contain Haley.
- Similarly, we have $\binom{39}{2}$ groups that contain Hugo.

Therefore we count $\binom{39}{2} \times \binom{39}{2}$ total desired outcomes.

---

(3) Compute probability:

Let E label the desired event. By the counting rule:

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

Evaluate using our data:

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

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

☰ Example - Counting VA license plates

VA license plates have 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 occur in increasing order?

(b) What is the probability that at least one number is repeated?

Solution

(a) Numerals in increasing order.

(1) 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.

---

(2) Count ways to have 4 numerals that occur in increasing order:

There are $\binom{10}{4}$ ways to choose 4 distinct numerals from 10 options.

For each choice of four distinct numerals, there is exactly one ordering that's increasing.

Therefore, there are $\binom{10}{4}$ ways to have 4 numerals that occur in increasing order:

---

(3) Count ways to have 3 letters (excluding I, O, Q) that occur in order.

There are 26 total letters, 3 are excluded, thus 23 options for each letter.

Repetition is allowed, thus we have $23 \cdot 23 \cdot 23 = 23^3$ total ways.

---

(3) Compute probability:

Total count of desired plates (taking the product of possibilities):

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

Let E label the event that a plate has increasing numerals. The counting formula for probability is:

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

Evaluate using our data:

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

(b) At least one number repeated.

"At least" is hard to work with! Lots of ways that can happen.

Try the *complement* event, which is much simpler: "no repetition"

Let E^c be event that *no* numbers are repeated. All are distinct. Then E is our desired event.

Count the possibilities:

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

The total number of plates is still $23^3 \cdot 10^4$.

Therefore, license plates with *at least one number repeated*:

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

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

Desired outcomes over total outcomes:

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

Counting out 4 teams

A board game requires 4 teams of players. How many configurations of teams are there out of a total of 17 players if the number of players per team is 4, 4, 4, 5, respectively.

Solution

This is just the multinomial coefficient with this data:

n	r_1	r_2	r_3	r_4
17	4	4	4	5

So we have:

$$\#\text{teams} = \frac{n!}{r_1!\,r_2!\,r_3!\,r_4!} \quad \ggg \quad \frac{17!}{4!\,4!\,4!\,5!} \quad \ggg \quad 214414200$$