

CMSC 250: Probability and Expected Value

Justin Wyss-Gallifent

April 20, 2023

1	Basics of Probability	2
	1.1 Experiments, Sample Spaces, and Events	2
	1.2 Counting Elements in Events and Sample Spaces	3
	1.3 Formal Definition of Probability	4
2	Expected Value	6
3	Conditional Probability and Independent Events	7
4	Two Handy Conditional Probability Formulas	9
5	Bayes' Theorem	11

1 Basics of Probability

1.1 Experiments, Sample Spaces, and Events

Definition 1.1.1. The *sample space* is the set (a set!) of all outcomes of an experiment, an *outcome* being a possible result.

Example 1.1. If we roll a die and write down the value then the sample space is $S = \{1, 2, 3, 4, 5, 6\}$

Example 1.2. If we randomly permute the three letter-word CAT then the sample space is $S = \{CAT, CTA, ACT, ATC, TCA, TAC\}$

Example 1.3. If we flip a coin and roll a 4-sided die in that order then the sample space is $S = \{(H, 1), (H, 2), (H, 3), (H, 4), (T, 1), (T, 2), (T, 3), (T, 4)\}$.

Definition 1.1.2. An *event* is a subset of the sample space. Usually this consists of a subset of outcomes in which we are interested.

Example 1.4. Suppose we roll two distinct dice and we are interested in those results for which the sum is 7. This will be our event A . Then we have:

$$S = \{(x, y) \mid 1 \leq x \leq 6, 1 \leq y \leq 6\}$$
$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

Note 1.1.1. It's worth pausing to comment that the use of the word "distinct" is there to ensure that $(1, 2)$ and $(2, 1)$ are considered different outcomes. If we had said "roll two dice" then arguably we might consider them the same. While that may not be an issue in some contexts it matters when computing probability because getting a 1 and a 2 (ignoring order) is twice as likely as getting two 1s.

Definition 1.1.3. For a set T we define $N(T)$ to be the cardinality of T . We can also write $|T|$ if we want, it's just that typically with probabilities all our sets are finite and so this new notation keeps that clear.

Definition 1.1.4. If all the outcomes in the sample space S of an experiment are equally likely then we define the *probability* of an event A to be:

$$P(A) = \frac{N(A)}{N(S)} = \frac{|A|}{|S|}$$

Example 1.5. Considering the example from earlier of rolling two distinct dice and looking at when the sum is 7. The 36 outcomes of rolling two dice are equally likely and hence the probability of getting a sum of 7 is:

$$P(A) = \frac{N(A)}{N(S)} = \frac{6}{36} = \frac{1}{6}$$

Example 1.6. Suppose we randomly permute the letters in the word PYTHON with all outcomes equally likely. What is the probability that the P and the Y will not be adjacent? We saw in earlier notes that there are $6! = 720$ ways to permute the letters in the word PYTHON. We also calculated that they are not adjacent in 480 of those. Thus the probability is:

$$P(A) = \frac{480}{720} = \frac{2}{3}$$

1.2 Counting Elements in Events and Sample Spaces

Oftentimes the most challenging part of calculating probabilities is simply counting the number of elements in the sample space and in the event subset.

In the notes on combinatorics we discussed methods for doing this in certain circumstances related to permutations and combinations but it's worth adding a few more tools to our toolbox.

Theorem 1.2.1. For events (in fact for subsets) A and B we have:

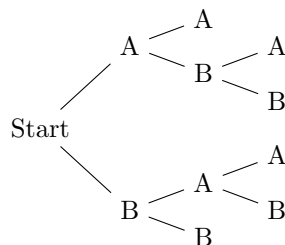
$$N(A \cup B) = N(A) + N(B) - N(A \cap B)$$

Proof. The reason for this is that when we count the items in A and in B separately we count those in the intersection twice, thus we have to subtract them to fix the issue. *QED*

Example 1.7. If $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{5, 6, 7, 8\}$ then $A \cap B = \{5, 6\}$ and so $N(A \cup B) = 6 + 4 - 2 = 8$.

Definition 1.2.1. A *possibility tree* is a tree which diagrams all the outcomes in an experiment which involves specific steps.

Example 1.8. Suppose two teams A and B play some games in which one wins and one loses. The plan is to play two games initially. If one team has won a majority then they stop, otherwise they play a tie-breaker. Draw the possibility tree for this experiment and write down the sample space.



The sample space is then:

$$S = \{AA, ABA, ABB, BAA, BAB, BB\}$$

Note 1.2.1. Note that no information is given in the above situation as to whether the outcomes are equally likely. This could be the case but really this depends on the teams, how good they are, etc. Without further information it's unclear.

1.3 Formal Definition of Probability

Definition 1.3.1. Suppose S is the sample space of an experiment. A probability function P from the set of all events in S to \mathbb{R} satisfies the following three axioms:

- (a) For any event $A \subseteq S$ we have $0 \leq P(A) \leq 1$.
- (b) We have $P(\emptyset) = 0$ and $P(S) = 1$.
- (c) If A and B are disjoint events then $P(A \cup B) = P(A) + P(B)$.

Note 1.3.1. What this all says really is that the probability of each outcome must be between 0 and 1 inclusive and that they must all add to 1.

Where these probabilities come from is not the issue in the definition. They could have come from equally likely outcomes, like we've seen, or they could come from research or other experimentation or knowledge.

Example 1.9. I have a sneaky coin and its probabilities can be given in this table:

Outcome	Probability
H	0.6
T	0.4

Example 1.10. I go to the store to get a pizza. The probabilities of me buying each type of pizza are listed in this table:

Outcome	Probability
Cheese	0.2
Bacon	0.25
Spinach	0.4
Pineapple	0.15

From the definition we get two useful facts:

Theorem 1.3.1. We have:

- (a) $P(A^c) = 1 - P(A)$.
 (b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Example 1.11. We saw earlier that if we randomly permute the letters in PYTHON that there is a $2/3$ probability that the P and Y will not be adjacent. Consequently the probability that they will be adjacent is:

$$1 - \frac{2}{3} = \frac{1}{3}$$

Example 1.12. Suppose there is a probability of 0.1 that I'll have a cheese pizza, a probability of 0.5 that I'll have a pineapple pizza, and a probability of 0.25 that I'll have both. What is the probability that I'll have one or the other or both?

If we think of A as the set of events in which I get a cheese pizza and B as the set of events in which I get a pineapple pizza then we know $P(A) = 0.1$, $P(B) = 0.5$, and $P(A \cap B) = 0.25$. We want $P(A \cup B)$ and so:

$$P(A \cup B) = 0.1 + 0.5 - 0.25 = 0.35$$

Example 1.13. Suppose we randomly permute the letters in PYTHON. There are $6!$ different and equally likely outcomes. What is the probability that a permutation starts with P or ends with ON, or both?

Define the following events:

- Let A be the set of outcomes in which the permutation starts with P. There are $5!$ of these, all with the form P????, so the probability of getting one is $P(A) = 5!/6! = 1/6$.
- Let B be the set of outcomes in which the permutation ends with ON. There are $4!$ of these, all with the form ???ON, so the probability of getting one is $P(B) = 4!/6! = 1/30$.

Observe that $A \cap B$ is the set of outcomes in which the permutation both starts with a P and ends with ON. There are $3!$ of these, all with the form

P??ON, so the probability of getting one is $P(A \cap B) = 3!/6 = 1/120$.

Consequently the probability of starting with P or ending with ON (or both) equals:

$$P(A \cup B) = \frac{1}{6} + \frac{1}{30} - \frac{1}{120}$$

2 Expected Value

Definition 2.0.1. If an experiment X consists of numerical outcomes a_1, a_2, \dots, a_n and if the probabilities are p_1, p_2, \dots, p_n respectively then the *expected value* of the experiment is:

$$E(X) = \sum_{i=1}^n p_i a_i$$

We'll just write E when there's only one experiment to deal with.

The meaning is that if the experiment was repeated over and over and if the average overall outcome were calculated each time then that average overall outcome would approach E .

In practice this means that if the experiment is repeated a very large number of times that the average outcome should be about E .

Despite the term, it does not mean that we actually expect to get that value.

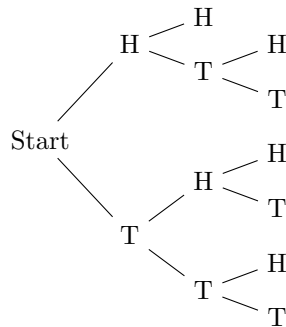
Example 2.1. If we roll a die the outcomes are 1 through 6 each with a probability of $1/6$. The expected value is then:

$$\frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{6}(3) + \frac{1}{6}(4) + \frac{1}{6}(5) + \frac{1}{6}(6) = 3.5$$

Thus in the long-term if we keep repeating this experiment the average result will be 3.5.

Example 2.2. Suppose you play a game. You flip a coin until either you have flipped three times or two H have come up in a row. If two H came up at the start you win \$5, otherwise you calculate the number of H minus the number of T and either win or lose that dollar amount accordingly. What is the expected value of this game?

A possibility tree can help us see the outcomes:



The probabilities for each of these are then listed in the following table in which we have also listed your win/loss amount:

Outcomes	Win/Loss	Probability
<i>HH</i>	+\$5	1/4
<i>HTH</i>	+\$1	1/8
<i>HTT</i>	-\$1	1/8
<i>THH</i>	+\$1	1/8
<i>THT</i>	-\$1	1/8
<i>TTH</i>	-\$1	1/8
<i>TTT</i>	-\$3	1/8

Here is a table with just the win/loss amounts and their probabilities:

Win/Loss	Probability
+\$5	1/4
+\$1	1/8 + 1/8 = 1/4
-\$1	1/8 + 1/8 + 1/8 = 3/8
-\$3	1/8

The expected value is then:

$$\frac{1}{4}(+5) + \frac{1}{4}(+1) + \frac{3}{8}(-1) + \frac{1}{8}(-3) = \frac{3}{4}$$

Note 2.0.1. In a game if the numerical outcomes are dollar figures then the expected value tells you your average per-game winnings and if the result is positive the game is ostensibly worth playing.

3 Conditional Probability and Independent Events

Definition 3.0.1. Given events $A, B \subseteq S$ the probability of A given B is the probability that A occurs given that we know that B occurs. This can be calculated via:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Note that if all events in S are equally likely then:

$$P(A|B) = \frac{N(A \cap B)}{N(B)}$$

Note 3.0.1. The fact that $P(A|B)$ and $P(B|A)$ represent and mean completely different values is one of the greatest sources of misunderstanding of data in the world today.

Example 3.1. Suppose all events in $S = \{1, 2, 3, 4, 5, 6\}$ are equally likely. If $A = \{1, 2, 5\}$ and $B = \{1, 2, 3, 4\}$ find $P(A|B)$.

Observe that:

$$P(A|B) = \frac{N(A \cap B)}{N(B)} = \frac{2}{4} = \frac{1}{2}$$

Stop and think about it - this makes sense, if event B happens it means the outcome is 1,2,3, or 4. Assuming this, the chance of getting 1,2, or 5 is 1/2, occurring when we get 1 or 2.

Again, this assumes equal likelihood.

Observe also that:

$$P(B|A) = \frac{N(B \cap A)}{N(A)} = \frac{2}{3}$$

Note that these are different!

Example 3.2. Suppose we randomly permute the letters in the word PYTHON. If the second letter is a P, what is the probability that the P and Y are adjacent?

If we put:

- A = The event where the P and Y are adjacent
- B = The event where the second letter is a P

Then we want $P(A|B)$. We know that:

$$P(A|B) = \frac{N(A \cap B)}{N(B)}$$

So we need to know two things:

- How many permutations have a P second and P and Y adjacent?

This happens only when we get $YP????$ or $PY????$ and there are $4! = 24$ ways each of these can happen, thus $2(4!) = 48$ ways total.

- How many permutations have a P second?

This happens when we get $?P????$ and there are $5! = 120$ ways this can happen.

Thus

$$P(A|B) = \frac{48}{120} = 0.4$$

Definition 3.0.2. Two events $A, B \subseteq S$ are *independent* if the fact that one occurs does not affect whether the other occurs. This means that:

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

Two events are *dependent* if they are not independent.

Note 3.0.2. It turns out that either both of these are equal or neither is. This means we can't have the weird situation where one is equal and the other not.

Note 3.0.3. Independent and disjoint events are not the same thing! Independent means $P(A|B) = P(A)$ and $P(B|A) = P(B)$, meaning that whether one occurs does not determine whether the other does.

Disjoint means $A \cap B = \emptyset$, meaning that both cannot happen.

If two nonempty events are disjoint then they are dependent. This is because if they are disjoint then both cannot happen, which then means that if one occurs, the other cannot, and this means they are dependent.

However if two nonempty events are not disjoint then they can be either dependent or independent.

Example 3.3. If $S = \{1, 2, 3, 4\}$, $A = \{1, 2, 3\}$, and $B = \{2, 3, 4\}$ then clearly A and B are not disjoint. Then observe that $P(A|B) = P(A \cap B)/P(B) = 0.5/0.75 = 2/3$ and $P(A) = 0.75$. Thus they are dependent.

Example 3.4. If $S = \{1, 2, 3, 4\}$, $A = \{1, 3\}$, and $B = \{1, 2\}$ then clearly A and B are not disjoint. Then observe that $P(A|B) = P(A \cap B)/P(B) = 0.25/0.5 = 0.5$ and $P(A) = 0.5$. Thus they are independent.

4 Two Handy Conditional Probability Formulas

First observe that the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Gives us:

$$P(A \cap B) = P(A|B)P(B)$$

Example 4.1. Suppose an urn contains five red balls and seven green balls. You take out one ball and then another (no replacement). What is the probability that the first is green and the second is red?

If we assign:

A = The event where the second is red

B = The event where the first is green

Then we want $P(A \cap B)$. Well:

$$P(A \cap B) = P(A|B)P(B)$$

For $P(A|B)$ observe that if B is true then the first is green and before the second one is removed there are five red and six green and so the probability that the second is red is $5/11$.

For $P(B)$ this is just the probability that the first is green and this is $7/12$.

Thus:

$$P(A \cap B) = P(A|B)P(B) = \left(\frac{5}{11}\right) \left(\frac{7}{12}\right)$$

Second observe that for events A and B an outcome in A is either in B or it isn't. The events $A \cap B$ and $A \cap B^c$ are disjoint, and thus we have:

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

Using the first formula then, we have:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Example 4.2. Suppose an urn contains five red balls and seven green balls. You take out one ball and then another (no replacement). What is the probability that the the second is red?

Note the difference between this and the earlier example. Now we don't care whether the first is green.

We still assign:

A = The event where the second is red

B = The event where the first is green

Now we want $P(A)$. Well:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

We did the first summand earlier so look at the second.

For $P(A|B^c)$ observe that if B is false then the first is red and before the second one is removed there are four red and seven green and so the probability that the second is red is $4/11$.

For $P(B^c)$ this is just the probability that the first is red and this is $5/12$.

Thus together:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = \left(\frac{5}{11}\right)\left(\frac{7}{12}\right) + \left(\frac{4}{11}\right)\left(\frac{5}{12}\right)$$

5 Bayes' Theorem

Bayes' Theorem establishes a connection between $P(A|B)$ and $P(B|A)$.

Theorem 5.0.1. In its most basic form:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Since $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$ from earlier, this is often written as:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Proof. It's easy to prove Bayes' Theorem:

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{P(A \cap B)}{P(B)} \cdot \frac{P(B)}{P(A)} \\ &= P(A|B) \cdot \frac{P(B)}{P(A)} \\ &= \frac{P(A|B)P(B)}{P(A)} \end{aligned}$$

QED

Here is a classic example of how Bayes' Theorem can tell you something wildly unexpected.

Example 5.1. Suppose a brand-new certain medical test for a virus yields either a positive or negative result. Data is collected and the following are calculated:

- If a person has the virus then the test has a 95% probability of indicating positive.
- If a person does not have the virus then the test has a 2% probability of indicating positive (this is a false positive).
- Several years of other tests have suggested that 3% of the population has the virus.

The first two bullet points suggest that this is a pretty good test. So let's ask a question. Suppose you get tested and the test is positive, what is the probability that you have the virus?

You would think that for a "pretty good test" the probability would be high, you wouldn't want a "pretty good test" so say "positive" when you're not!

Let's assign:

$P(A)$ = The probability that you test positive

$P(B)$ = The probability that you have the virus

You are looking for $P(B|A)$.

With our assignment our three bullet points then say:

- $P(A|B) = 0.95$
- $P(A|B^c) = 0.02$
- $P(B) = 0.03$ and therefore $P(B^c) = 0.97$.

Then by Bayes' Theorem we have:

$$\begin{aligned}
P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \\
&= \frac{(0.95)(0.03)}{(0.95)(0.03) + (0.02)(0.97)} \\
&= \frac{285}{479} \\
&\approx 0.5950
\end{aligned}$$

So this means that there's only a 59.50% probability that you actually have the virus!

Example 5.2. You're trying to refine a piece of software which translates foreign languages. The software has analyzed a large amount of data in this foreign language and has come to the following conclusions:

- 40% of the time a noun will be followed by a verb.
- 20% of the time a non-noun will be followed by a verb.
- 10% of all words are nouns.

In order to refine the software you're testing how well it can guess at previous words. Given that a word is a verb, what is the probability that the word before it was a noun?

Let's assign:

$P(A)$ = The probability that a word is a verb

$P(B)$ = The probability that the previous word is a noun

You are looking for $P(B|A)$.

With our assignment our three bullet points then say:

- $P(A|B) = 0.4$
- $P(A|B^c) = 0.2$
- $P(B) = 0.1$ and therefore $P(B^c) = 0.9$.

Then by Bayes' Theorem we have:

$$\begin{aligned}
P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \\
&= \frac{(0.4)(0.1)}{(0.4)(0.1) + (0.2)(0.9)} \\
&= \frac{2}{11} \\
&\approx 0.1818
\end{aligned}$$

So this means that there's an 18.18% probability that the word before a verb is a noun.

Example 5.3. You are building a classifier $f(x)$ which takes an item x and classifies it as in one of two populations, P_1 or P_2 . You want to get $f(x) = 1$ if $x \in P_1$ and $f(x) = 2$ if $x \in P_2$.

In order to test your classifier you obtain a training set of data which consists of 1000 items of which 553 are in P_1 and 447 are in P_2 .

You test your classifier by applying it to each item in the training set and you find:

$$\begin{aligned}
P(f(x) = 1 | x \in P_1) &= 0.91 \\
P(f(x) = 1 | x \in P_2) &= 0.02
\end{aligned}$$

You also suspect that your training set is a good sample of the real world and so you suggest that for x in the real world that:

$$\begin{aligned}
P(x \in P_1) &= 0.553 \\
P(x \in P_2) &= 0.447
\end{aligned}$$

Now then, if you take $f(x)$ to the real world and test it on some x , if $f(x) = 1$ what is the probability that $x \in P_1$?

By Bayes' Theorem we find:

$$\begin{aligned} P(x \in P_1 | f(x) = 1) &= \frac{P(f(x) = 1 | x \in P_1)P(x \in P_1)}{P(f(x) = 1 | x \in P_1)P(x \in P_1) + P(f(x) = 2 | x \in P_2)P(x \in P_2)} \\ &= \frac{(0.91)(0.553)}{(0.91)(0.553) + (0.02)(0.447)} \\ &\approx 0.9825 \end{aligned}$$

Not bad!