Math 345 - Notes
Probability & Statistics I
Axioms, independence, conditional probability
September 19, 2018

Probability – definition

There are three views of probability.

- **axiomatic approach:** each event E is associated to a probability $P(E)$ which measures the chance of this event occurring.
- **frequency interpretation:** if we repeat an experiment a large number of times, then the fraction of times the event E occurs will be close to $P(E)$. This is supported by the Law of Large Numbers.
- **measure of belief:** we associate a likelihood of occurrence to an event using our intuition and understanding of the world. This is not a good idea, since our intuition is often wrong.

For example, if we want to find the probability of a die landing on 6, we would reason as follows:

- **axiomatic approach:** since each outcome is equally likely, we assign probability $1/6$ to landing on 6.
- **frequency interpretation:** we roll the die many times and observe that the die lands on 6 about $1/6$ of the time, hence the probability of a 6 is $1/6$.
- **measure of belief:** each side looks the same, so intuitively, a die lands on 6 with probability $1/6$.

Roughly, a probability is a measure of the occurrence of an event. Formally, we will opt for the axiomatic approach and define probability as follows:

Def: A probability measure P is a function $P : \Omega \to [0, 1]$ satisfying the axioms:

(i) $0 \leq P(E) \leq 1$ for any event E.

(ii) $P(\Omega) = 1$

(iii) If $E_1, E_2, \ldots$ are *pairwise disjoint* (meaning $E_i \cap E_j = \emptyset$ for all $i \neq j$), then

$$P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k).$$

Notes:

- It might be easy to look at a diagram and interpret probability as the "weight" of an event, with the understanding that the "weight" of the entire sample space Ω is 1.
- $\emptyset$ refers to the emptyset (a set with no elements)

- $E \cap F$ denotes the *intersection* of E and F (the overlap of E and F)
- $E \cup F$ denotes the *union* of E and F (all outcomes that are in E, or F, or both)
- The complement E^c contains all outcomes that are NOT in E: $E^c = \Omega - E$.

Properties:

1. $P(\emptyset) = 0$
2. If E^c denotes the complement of E, then $P(E^c) = 1 - P(E)$.
3. If E and F are disjoint ($E \cap F = \emptyset$),
$$P(E \cup F) = P(E) + P(F).$$
4. If E and F are not disjoint ($E \cap F \neq \emptyset$),
$$P(E \cup F) = P(E) + P(F) - P(E \cap F). \tag{1}$$
This can be proved by noting that when we add the weights of E and F, we add the weight of the overlap $E \cap F$ twice, so we need to subtract one copy of it.

Examples:

1. Roll 2D6 (red and blue). Let A = red die lands on 6, B = blue die lands on 6, E = at least one die lands on 6. Then there are 36 possible outcomes for this experiment and
$$P(A) = \frac{6}{36} = \frac{1}{6}, \quad P(B) = \frac{6}{36} = \frac{1}{6},$$
$P(A \cap B) = \frac{1}{36}$ (since both dice land on 6) and $P(A \cup B) = \frac{11}{36}$ by looking at all outcomes when the red die, the blue die or both land on 6. Note that $E = A \cup B$, so then we found $P(E) = P(A \cup B) = \frac{11}{36}$ and
$$P(E^c) = 1 - P(E) = 1 - \frac{11}{36} = \frac{25}{36}$$
which makes sense because E^c is the event where no die lands on 6, which occurs in 25 outcomes. Let us also verify equation (1):
$$P(A \cup B) = \frac{11}{36}, \quad P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
2. [extra] Flip 3 coins. Let A = first two coins land on Heads, B = last two coins land on Heads. Then the sample space is
$$S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}.$$
We find the following events
$$A = \{HHH, HHT\}, \quad B = \{HHH, THH\}, \quad A \cup B = \{HHH, HHT, THH\}, \quad A \cap B = \{HHH\}$$
Therefore, $P(A) = 1/4$, $P(B) = 1/4$, $P(A \cup B) = 3/8$ and $P(A \cap B) = 1/8$. Note that (1) holds true, since
$$P(A) + P(B) - P(A \cap B) = \frac{1}{4} + \frac{1}{4} - \frac{1}{8} = \frac{3}{8} = P(A \cup B).$$

3. Here is another example for computing the probability of $A \cup B$.
   (a) Suppose we pick one card from a deck of 52. Let $A =$ it is a face card, $B =$it is a spade. Then $P(A) = 12/52$, $P(B) = 13/52$ and $P(A \cap B) = P(\text{face on spade card}) = 3/52$. Thus, $P(A \cup B) = (12 + 13 - 3)/52 = 22/52$.

       This makes sense, since counting all cards that are spades or face cards (or both), we can first take the 13 spades, and to those add the 9 remaining face cards not yet picked, to get 22 cards out of 52.

   (b) Suppose we pick one card from a deck of 52. Let $A =$ it is a heart, $B =$it is a spade. Then $P(A) = 13/52$, $P(B) = 13/52$ and $P(A \cap B) = P(\text{heart and spade card}) = 0$. Thus, $P(A \cup B) = (13 + 13)/52 = 1/2$, which seems reasonable since the hearts and spades make up half the deck of cards.

   Remark: It is very important to remember to subtract the overlap, otherwise the probabilities will be overestimated! When the events are disjoint (or mutually exclusive), the probability of the overlap will be zero.

Independence

Two events

Informally, two events are independent if the outcome of one has no influence on the outcome of the other. Furthermore if we know that A has occurred, it gives us no extra information on the occurrence of B. Some events can be easily be seen as independent or dependent. For example, if we flip two fair coins, the outcome of the first coin does not influence in any way the outcome of the second coin. However, if we remove cards from a deck of 52, one at a time, without replacement, the outcome of the second card will be influenced by what we removed from the deck before. Sometimes, we can not intuitively determine if two events are independent, in which case we use the following formal definition to decide.

Definition: We say the events A and B are **independent** if

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

If two events are *not independent* we call them **dependent.**

Remark: independence plays a big role in how we compute probabilities in multi-step experiments, such a flipping multiple coins, drawing cards without replacement from a deck etc.

Examples:

1. We roll 2D6. Let $A =$ first die lands on 6, $B =$ second die lands on 6. Intuitively, the events are independent. Let us verify:

   $$P(A) = 1/6, \;\; P(B) = 1/6, \;\; P(A \cap B) = 1/36 \;\; \Rightarrow \;\; P(A \cap B) = P(A)P(B).$$

2. [extra] We roll 2D6. Let $A =$ first die lands on 6.
   (a) Let $B =$ the sum of the dice is 7. Then $B = \{16, 25, 34, 43, 52, 61\}$ and $A \cap B = \{61\}$

       $$P(A) = 1/6, \;\; P(B) = 6/36, \;\; P(A \cap B) = 1/36 \;\; \Rightarrow \;\; P(A \cap B) = P(A)P(B).$$

       So A and B are independent.

(b) Let $B=$ the sum of the dice is 9. Then $B=\{36,45,54,63\}$ and $A\cap B=\{63\}$

$$P(A)=1/6,\ \ P(B)=4/36,\ \ P(A\cap B)=1/36\ \ \Rightarrow\ \ P(A\cap B)\neq P(A)P(B).$$

So A and B are dependent in this case.

3. Draw a card from a deck. Let $A=$ the card is an ace, $B=$ the card is a spade. Then $P(A)=1/13$, $P(B)=1/4$ and

$$P(A\cap B)=P(\text{the card is the ace of spade})=1/52.$$

Thus, $P(A\cap B)=P(A)P(B)$, so A and B are independent.

4. We draw two cards from a deck of 52. Let $B=$ second card is a spade.

(a) Let $A=$ the first card is an ace. Then $P(A)=1/13$ and

$$\begin{aligned} P(B) &= P(\text{1st card} = \spadesuit, \text{2nd card} = \spadesuit) + P(\text{1st card} \neq \spadesuit, \text{2nd card} = \spadesuit) \\ &= \frac{13\times 12}{52\times 51}+\frac{39\times 13}{52\times 51}=\frac{1}{4}. \end{aligned}$$

On the other hand,

$$\begin{aligned} P(A\cap B) &= P(\text{1st card} = \text{ace } \spadesuit, \text{2nd card} = \spadesuit) + P(\text{1st card} = \text{ace, not } \spadesuit, \text{2nd card} = \spadesuit) \\ &= \frac{1\times 12}{52\times 51}+\frac{3\times 13}{52\times 51}=\frac{1}{52}. \end{aligned}$$

We check that $P(A\cap B)=P(A)P(B)$, so A and B are independent.

(b) Let $A=$ the first card is a spade. Then $P(A)=1/4$ and $P(B)=1/4$ from (a), but

$$P(A\cap B)=P(\text{1st card} = \spadesuit, \text{2nd card} = \spadesuit)=\frac{13\times 12}{52\times 51}=\frac{1}{17}.$$

Thus, $P(A\cap B)\neq P(A)P(B)$ so A and B are dependent.

More than two events (optional)

We can define independence for more than 2 sets in the following way.

Definition: If $A_1, A_2, \cdots, A_n$ are n events, they are independent if

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)P(A_2)\cdots P(A_n).$$

The events are pairwise independent if $P(A_i\cap A_j)=P(A_i)P(A_j)$ for all pairs $i\neq j$.

Remark: If a set of events are independent, then they must also be pairwise independent. There could be events that are pairwise independent but not independent.

Examples:

1. Flip 20 coins. Let A_k denote the event the kth coin land on heads. These events are independent because

$$P(A_1)=\cdots=P(A_{20})=\frac{1}{2},\ \ P(A_1\cap A_2\cap\cdots\cap A_{20})=\frac{1}{2^{20}}\ \ \Rightarrow P(A_1\cap\cdots\cap A_{20})=P(A_1)P(A_2)\cdots P(A_{20}).$$

2. Let A=Al and Bob share a birthday, B=Bob and Cate share a birthday, and C=Cate and Al share a birthday. Then
$$P(A) = P(B) = P(C) = \frac{365}{365^2} = \frac{1}{365}$$
since there are 365 possible days for the common birthday and the number of ways for two people to be born is 365×365 (365 days possible for each). Now
$$P(A \cap B \cap C) = P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{365}{365^3} = \frac{1}{365^2}.$$
This follows from the fact that $A \cap B \cap C = A \cap B = B \cap C = C \cap A$ all mean "Al, Bob and Cate share a birthday". Therefore, these events are pairwise independent because
$$P(A \cap B) = P(A)P(B),\;\; P(A \cap C) = P(A)P(C),\;\; P(B \cap C) = P(B)P(C),$$
but they are not independent because $P(A \cap B \cap C) \neq P(A)P(B)P(B)$.

Conditional Probability

If two events are independent, the occurrence of one does not influence the probability of the other, but what if they are dependent? Let's consider an example.

Example: We flip two fair coins.

(a) I know the first lands on Heads. What is the probability both land on Heads?

When we flip two coins the possible outcomes are $\{TT, TH, HT, HH\}$, but once we know the first lands on Heads, the set of possible outcomes is reduced to $\{HT, HH\}$, both equally likely. Thus, the probability of HH is $1/2$.

(b) I know at least one coin lands on Heads. What is the probability both land on Heads?

The set of possible outcomes is reduced to $\{TH, HT, HH\}$, equally likely, so the probability of HH is $1/3$.

Remarks:

- Probabilities vary depending on the extra information that we have.
- The extra information reduces the state space, and we compute probabilities in this smaller space.
- To find probabilities once we know extra information means to **condition on** the extra information.

Definition: If A is an event of positive probability $P(A) > 0$, then *the probability of B once we know A*, or *probability of B given A* is defined as
$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$
We read $P(B|A)$ as "probability of B given A", or "probability of B conditioned on A".

Some important properties of conditional probability:

1. We assume $P(A) > 0$ because it makes no sense to condition on an event that has probability 0; furthermore, if $P(A) = 0$ the fraction in the definition is undefined!

2. Multiplying both sides of the definition by $P(A)$ gives the **Multiplication Formula**

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

   This formula will be very useful in computing two-step probabilities. Intuitively, for both A and B to occur, first A must occur, and then given that A occurs, B must occur too.

3. If A and B are independent, then

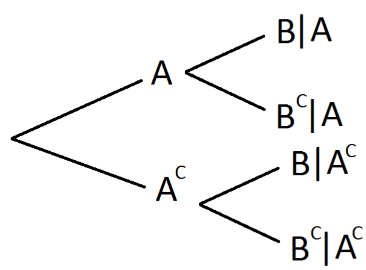
$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B),$$

   which makes sense since knowing A does not influence the probability of B in this case.

We can find the probability of some event B by considering the two possibilities of A occurring or A^c occurring (here A^c denotes the complement of the event A). Then

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

This is just a formal way to compute probabilities along the branches of a tree. Recall that we multiply horizontally in the tree (depth) and add vertically (breadth).



Here were derived **Bayes Formula**, which is at the base at Bayesian statistics:

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

Why is this useful? Suppose we know how the occurrence of A affects the probability of B, but we would like to find out how the occurrence of B affects the probability of A. Bayes Formula allows us to do exactly that.

Examples:

1. Pick two cards from a deck of 52. Find the probability that both cards are spades.

   Let A =first card is a spade and B =second card is a spade. Then we know $P(A) = 13/52$ and $P(B) = 1/4$. We also know that $P(B|A) = 12/51$ since conditioning on having picked a spade, there are 12 spades to pick from the remaining 51 cards. Therefore, using the multiplication formula:

$$P(A \cap B) = P(A)P(B|A) = \frac{13}{52} \times \frac{12}{51} = \frac{3}{51}.$$

   Note that A and B are not independent since $P(A)P(B) = \frac{1}{16}$.

2. Based on past experience, 70% of students pass the midterm, out of which 80% pass the final. Only 40% of students failing the midterm pass the final. How many students pass the final?

Let M denote that the student passed the midterm and F that the student passed the final. We want to find the probability that a student (randomly chosen) passed the final, that is $P(F)$. We know $P(F|M) = .8$ and $P(F|M^c) = .4$. We also know that 70% of students pass the midterm, i.e. $P(M) = .7$ Then

$$\begin{cases} M, \quad (\text{prob. } 0.7) \begin{cases} F, & (\text{prob. } 0.8) \\ F^c, & (\text{prob. } 0.2) \end{cases} \\ M^c, \quad (\text{prob. } 0.3) \begin{cases} F, & (\text{prob. } 0.4) \\ F^c, & (\text{prob. } 0.6) \end{cases} \end{cases}$$

$$P(F) = P(F \cap M) + P(F \cap M^c) = P(M)P(F|M) + P(M^c)P(F|M^c) = (.7)(.8) + (.3)(.4) = .68$$

3. Roll two dice. Let $A =$ sum is 8, and $B =$ first die lands on 3. Then

$$A = \{26, 35, 44, 53, 62\}, \quad B = \{31, 32, 33, 34, 35, 36\}.$$

Thus, $P(A) = 5/36$, $P(B) = 1/6$ and $P(A \cap B) = P(\text{pair } 35) = 1/36$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6} \quad \text{and} \quad P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/36}{5/36} = \frac{1}{5}.$$

Another way to solve this problem, once we listed the outcomes in A and B, is to note that

- conditioning on B, we are left with 6 outcomes out of which 35 occurs with probability 1/6;
- conditioning on A, we are left with 5 outcomes out of which 35 occurs with probability 1/5.

4. Five pennies are sitting on a table. One is a trick coin that has Heads on both sides, but the other four are normal. You pick up a penny at random and flip it four times, getting Heads each time. Given this, what is the probability you picked up the two-headed penny?

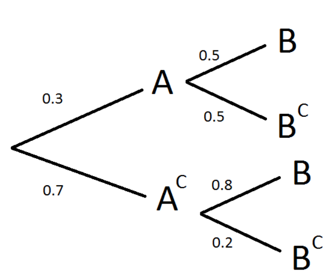
We have 2 cases: we pick a fair coin with probability 4/5 or the trick coin with probability 1/5.

$$\begin{cases} \text{fair coin}, \quad (\text{prob. } \tfrac{4}{5}) \begin{cases} HHHH, & (\text{prob. } \tfrac{1}{16}) \\ \text{something else}, & (\text{prob. } \tfrac{15}{16}) \end{cases} \\ \text{trick coin}, \quad (\text{prob. } \tfrac{1}{5}) \begin{cases} HHHH, & (\text{prob. } 1) \\ \text{something else}, & (\text{prob. } 0) \end{cases} \end{cases}$$

Let T be the event that we picked the trick coin and let H denote the even that we get four H in four coin tosses. Then

$$P(T|H) = \frac{P(T \cap H)}{P(H)} = \frac{P(H|T)P(T)}{P(H|T)P(T) + P(H|T^c)P(T^c)} = \frac{1 \cdot \frac{1}{5}}{1 \cdot \frac{1}{5} + \frac{1}{16} \cdot \frac{4}{5}} = \frac{4}{5}$$

5. Consider the following tree, with probabilities along the branches.



We read the tree as follows:

(a) $P(A) = 0.3$

(b) $P(A^c) = 0.7$

(c) $P(B|A) = 0.5$

(d) $P(B^c|A) = 0.5$

(e) $P(B|A^c) = 0.8$

(f) $P(B^c|A^c) = 0.2$

And we compute

(a) $P(A \cap B) = P(A)P(B|A) = (0.3)(0.5) = 0.15$

(b) $P(A^c \cap B) = P(A^c)P(B|A^c) = (0.7)(0.8) = 0.56$

(c) $P(B) = P(A \cap B) + P(A^c \cap B) = 0.15 + 0.56 = 0.71$

(d) $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.15}{0.71} = \dfrac{15}{71}$

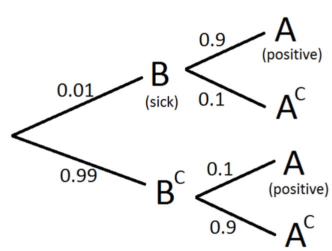
6. Suppose two events A and B have probabilities $P(A) = 0.2$, $P(B) = 0.3$ and $P(A|B) = 0.5$. We find

   (a) $P(A \cap B) = P(A|B)P(B) = (0.5)(0.3) = 0.15$

   (b) $P(B|A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{0.15}{0.2} = \dfrac{3}{4}$

7. [extra] Approximately 1% of women aged 40-50 have breast cancer. A woman with breast cancer has a 90% chance of a positive test from a mammogram, while a woman without has a 10% chance of a false positive result. What is the probability a woman has breast cancer given that she just had a positive test?

   Let B denote the event that a randomly selected woman 40-50 years old has breast cancer. Let A denote the event that the mammogram is positive for the disease.



   Then we know $P(B) = .01$, $P(A|B) = .9$ and $P(A|B^c) = .1$, where B^c denotes the complement of B, namely the event that the woman does not have breast cancer. Then $P(B^c) = 1 - P(B) = .99$ and we need to find $P(B|A)$.

$$
\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A \cap B) + P(A \cap B^c)} \\
&= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} \\
&= \frac{(.9)(.01)}{(.9)(.01) + (.1)(.99)} = \frac{9}{108}
\end{aligned}
$$

   Another way to look at this problem is to find the proportion of women who have breast cancer and a positive test, to the women who have a positive test. The probability of a woman to have breast cancer and a positive test is $(.01)(.9) = .009$, or .9% of women tested are in this category. A woman can have a positive test if she has breast cancer, with probability .009 as we saw above, or she can be disease free but have a false positive test with probability $(.99)(.1) = .099$ or 9.9% of population. So the percentage of the population with a positive test is 10.8%. Thus, given a positive test, the proportion of women who are sick is $\frac{.9}{10.8} = \frac{9}{108}$.

8. [extra] **The Monty Hall Problem** We play the following game show. The game show host (Monty Hall) shows us 3 doors. A car is hidden behind one door and goats behind the other two. We are asked to pick a door. For simplicity, let's assume we pick Door #1. Then the game show host opens Door #3 and reveals a goat. We are asked if we would like to switch our pick from Door #1 to Door #2. Should we switch? What is the probability of winning the car behind Door #1, now that we are left with only 2 doors? Is is 1/2? Or is it 1/3 as it was when we started the game? Or is it something else? To compute the probability that the car is behind Door #1, we look at 3 possible cases.

Case	Door #1	Door #2	Door #3	Host action
1	Car	Goat	Goat	reveal D#2 or D#3 (each with prob. 1/2)
2	Goat	Car	Goat	reveal D#3
3	Goat	Goat	Car	reveal D#2

Once we know *the host picked D#3*, we cannot be in Case 3. So we are left with two equally likely cases: Case 1 and Case 2. Note that it is twice as likely he picked Door #3 in Case 2 than in Case 1, so the probability of being in Case 1 (hence winning the car if we do not switch) is 1/3. More precisely,

$$P(\text{Case 1}|\text{reveal D\#3}) = \frac{P(\text{Case 1 \& reveal D\#3})}{P(\text{reveal D\#3})} = \frac{(1/3)(1/2)}{(1/3)(1/2) + (1/3)(1) + (1/3)(0)} = \frac{1}{3},$$

where we found

$$P(\text{reveal D\#3}) = P(\text{Case 1 \& reveal D\#3}) + P(\text{Case 2 \& reveal D\#3}) + P(\text{Case 3 \& reveal D\#3}).$$

Therefore switching at this point to Door #2 would increase our probability of winning the car to 2/3 from the 1/3 probability we started with. Intuitively, your first choice has 1/3 chance of winning and all the other probability of 2/3 is hidden behind the other unopened doors. Once a door is revealed, the door that remains keeps the 2/3 probability.

Now you can imagine what happens if we have more doors: you select one and the host reveals all other but one. What happens to your probability of winning if you stay with your initial choice? It stays the same, it does not improve. But switching brings you a big advantage. For example, if there are 10 doors, you pick the winner with probability 1/10. Now the game show host reveals 8 doors with goats and you have to decide whether to keep your door or switch. The other door has 9/10 chance of being a winner, so you should certainly switch. In fact, even if the host reveals only one door, you should still switch to increase your winning probability from 1/10 to 9/80.