

Recitation/Discussion 2

Conditional Probability and Bayes' Theorem

accompaniment to MATH 170E: Introduction to Probability and Statistics 1: Probability
at the University of California, Los Angeles

Allen Lin

In this recitation, we will introduce independent and dependent events and use them to solve problems involving conditional probability. Then we introduce Bayes' theorem.

Exit Survey

Please fill out the exit survey.

This is so I can get feedback and know what to focus on next time!

Conditional Probability and Independence

Sometimes, we want to compute the probability of an event given that some other event occurred. This is widely applicable because in the real world events seldom occur without consequences to other events.

Definition 1. The *conditional probability of an event B given A* is defined by

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

Some intuition on this definition is in order. When one conditions on A , then A is the new sample space because we are only considering events where A has already taken place. Thus, when we ask for the probability that B occurs given A , in our new sample space A the desired set is $A \cap B$. Therefore, their quotient $P(A \cap B)/P(A)$ should be $P(B \mid A)$.

This definition motivates the following.

Corollary 2. The probability that two events A and B both occur is

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

This is a more general notion of the probability of the intersection of two events. Recall from last time that $P(A \cap B) = P(A)P(B)$ when A and B are independent.

Definition 3. Two events A and B are *independent* if and only if $P(A \cap B) = P(A)P(B)$ and *dependent* otherwise.

For example, in machine learning we often want to update the network to make predicts based on already-observed examples, allowing the algorithm to improve and "learn."

Of course, this doesn't happen when $P(A) = 0$.

Intuitively, this means that two events A and B are independent if and only if the probability of B does not change when conditioned on A . We make this notion more mathematically rigorous in the following.

Corollary 4. *Two events A and B are independent if and only if $P(B | A) = P(B)$.*

Also, $P(A | B) = P(A)$.

Proof. ($\implies$) Suppose that A and B are independent. If $A = \emptyset$, then there is nothing to prove. Suppose $A \neq \emptyset$, so $P(A) > 0$. Then

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

so $P(B | A) = P(B)$.

($\impliedby$) Suppose that $P(A | B) = P(A)$. Then

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

so A and B are independent. ■

The proof is symmetric about A and B , so we can switch the order of A and B and obtain that $P(A | B) = P(A)$ as well.

Exercise 1. Two cards are drawn successively and without replacement from an ordinary deck of 52 playing cards. Compute the probability of drawing

- (a) two hearts,
- (b) a heart on the first draw and a club on the second draw.

Solution. (a) Let A_1 be the event of drawing a heart on the first draw, and let A_2 be the event of drawing a heart on the second draw. Then

$$P(A_1 \cap A_2) = P(A_1)P(A_2 | A_1) = \frac{13}{52} \cdot \frac{12}{51};$$

here, $P(A_2 | A_1) = 12/51$ since out of the 51 remaining cards there are 12 hearts left.

- (b) Let A_1 be the event of drawing a heart on the first draw and a club on the second draw. Then

$$P(A_1 \cap A_2) = P(A_1)P(A_2 | A_1) = \frac{13}{52} \cdot \frac{13}{51};$$

here, $P(A_2 | A_1) = 13/51$ since out of the 51 remaining cards there are 13 clubs left.

Summary on Unions and Intersections of Events

Let A and B be events.

Union of Events

- When calculating $P(A \cup B)$, you can write

$$P(A \cup B) = P(A) + P(B)$$

only when the events are disjoint (i.e., mutually exclusive).

- If the **events are not disjoint**, then you are overcounting, and you need to correct it with the inclusion-exclusion principle that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

This extends to any arbitrary number of events.

Intersection of Events

- When calculating $P(A \cap B)$, you can only write

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

only when the events are independent.

- If the **events are not independent**, then your correction factor is

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

This extends to any arbitrary number of events.

Exercise 2. Let A and B be two events.

- If A and B are mutually exclusive, are A and B always independent? If not, can they ever be independent?
- If $A \subseteq B$, are A and B always independent? If not, can they ever be independent?

Solution. (a) Suppose that A and B are mutually exclusive. By definition, $P(A \cap B) = 0$. If A and B are independent, then

$$0 = P(A \cap B) = P(A)P(B).$$

This only happens when $P(A) = 0$ or $P(B) = 0$. So yes, A and B can be independent, but not always.

- Suppose $A \subseteq B$. Then $A \cap B = A$. If A and B are independent, then

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

so $P(A)(1 - P(B)) = 0$. This happens when $P(A) = 0$ or $P(B) = 1$. So yes, A and B can be independent, but not always.

The Birthday Problem

The probability that two events A and B both occur can be extended into any arbitrary number of events.

Corollary 5. Let $A_1, \dots, A_n$ be events. Then

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2 | A_1) \cdots P(A_n | A_1 \cap \cdots \cap A_{n-1}).$$

Exercise 3. Compute the probability that at least two people share the same birthday (not counting February 29) from a set of $n \leq 365$ randomly chosen people.

Hint. Let A_i be the event that the i th person has a different birthday from everyone before them and consider $\bigcap_{i=1}^n A_i$.

Solution. Let A be the event that at least two people share the same birthday. It's easier to compute the probability that no one shares the same birthday, or $P(A^c)$. Let A_i be the event that the i th person has a different birthday from everyone before them. Then $\bigcap_{i=1}^n A_i$ is the event that everyone has different birthdays, so

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1)P(A_2 | A_1) \cdots P(A_n | A_1 \cap \cdots \cap A_{n-1}).$$

We compute the conditional probabilities.

- $P(A_1) = 1$ since the event A_1 is automatically true.
- $P(A_2 | A_1) = 364/365$: Suppose that the first person has a unique birthday (A_1). Then for A_2 to be true given A_1 , the second person has 364 choices of birthdays.
- $P(A_3 | A_1 \cap A_2) = 363/365$. Suppose that the first and second person have unique birthdays ($A_1 \cap A_2$). Then for A_3 to be true given $A_1 \cap A_2$, the third person has 363 choices of birthdays.
- Continuing, we get $P(A_i | A_1 \cap \cdots \cap A_{i-1}) = (365 - (i - 1))/365$.

Thus,

$$\begin{aligned} P(A) &= 1 - P\left(\bigcap_{i=1}^n A_i\right) \\ &= 1 - P(A_1)P(A_2 | A_1) \cdots P(A_n | A_1 \cap \cdots \cap A_{n-1}) \\ &= 1 - 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - (n - 1)}{365}. \end{aligned}$$

Interestingly, when $n = 23$, $P(A) = 0.507297 > 1/2$. When $n = 50$, $P(A) = 0.97$, and when $n = 100$, $P(A) = 0.9999997$.

If there are $n > 365$ people, then at least two people share the same birthday by the pigeonhole principle.

Bayes' Theorem

Bayes' theorem is one of the most ubiquitous theorems with wide applications. Its motivation is simple: For two events A and B , often we can compute $P(A | B)$, but is there some relation between that and $P(B | A)$? A common scenario for this is when we want to find the probability of something unobservable given an observed event.

Theorem 6 (Bayes' theorem). Let $B_1, \dots, B_m$ be events that partition the sample space S , so $S = B_1 \cup \dots \cup B_m$ and the events are pairwise disjoint ($B_i \cap B_j = \emptyset$ for all $i \neq j$). Then for all $i \in \{1, \dots, m\}$,

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{i=1}^m P(B_i)P(A | B_i)}.$$

For example, suppose that you lost your phone. You know that your phone is at one of n locations, say $B_1, \dots, B_n$, and you have some guesses as to where your phone is, say that you believe that your phone is at B_i with $P(B_i)$ certainty. Luckily, you have another device that has a location tracker, but it's not the most accurate. Let's say that you look at the location tracker, which gives you a measure of how strong the cell signal is, and let's say A is the event that the cell signal has a strength higher than a threshold. Now after A happens, you want to update $P(B_i | A)$ to know where to prioritize your search. You scratch your head, but there is hope after all after you remember Bayes' theorem! The probability of the observation $P(A | B_i)$ is the probability that your cell signal is strong assuming that your phone is at location B_i , and your location tracker gives you that information! You use Bayes' theorem to calculate the probabilities $P(B_i | A)$, and then you go on your search.

We often see the more symmetrical version.

Theorem 7 (Bayes' theorem). Let A and B be two events. Then

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

Exercise 4. A rare disease affects approximately 1% of the population, and doctors are working on a test to detect the disease early. Someone with the disease has a 90% chance of a positive test result, and someone without the disease has a 10% chance of a negative test result. What is the probability that someone has the disease given that they had a positive test result?

Solution. Let B be the event that the person has the disease, and let A be the event of a positive test result. We know $P(B) = 0.01$, $P(A | B) = 0.9$, and $P(A | B^c) = 0.1$, and we want to calculate $P(B | A)$. We have

$$\begin{aligned} P(B | A) &= \frac{P(B)P(A | B)}{P(B)P(A | B) + P(B^c)P(A | B^c)} \\ &= \frac{0.01 \cdot 0.9}{0.01 \cdot 0.9 + 0.99 \cdot 0.1} = \frac{9}{108} \end{aligned}$$

by Bayes' theorem.

A worst nightmare scenario.

These are known as *priors*.

These are known as *posteriors*.

The answer is actually quite shocking!

$9/108 = 8.3\%$ is shockingly low.

Monty Hall Problem

Exercise 5. You are a contestant on a game show. There are three doors. Behind one of the door sits a nice car, and behind the other two there are two goats. You choose a door. The host opens one of the other doors, which shows a goat. The host gives you the opportunity of switching to the remaining door. Should you switch or keep your original choice?

Hint. Let B be the event that your door has the car, and let A be the event that the host opens one of the other two doors and reveals a goat.

Solution. You pick a door. Let B be the event that your door has the car, so B^c is the event that your door does not have the car. Before the host does anything, your belief is

This is your *prior* belief.

$$P(B) = \frac{1}{3} \quad \text{and} \quad P(B^c) = \frac{2}{3}.$$

Now let A be the event that the host opens one of the other two doors and reveals a goat. This is the observed information. Note that

$$P(A | B) = 1 \quad \text{and} \quad P(A | B^c) = 1$$

because given B or B^c , the host always chooses one of the other doors that has a goat behind it. You want to calculate $P(B | A)$. By Bayes' theorem,

This is your *posterior* belief.

$$\begin{aligned} P(B | A) &= \frac{P(B)P(A | B)}{P(B)P(A | B) + P(B^c)P(A | B^c)} \\ &= \frac{1/3}{1/3 + 2/3} \\ &= \frac{1}{3}. \end{aligned}$$

On the other hand,

$$\begin{aligned} P(B^c | A) &= \frac{P(B^c)P(A | B^c)}{P(B)P(A | B) + P(B^c)P(A | B^c)} \\ &= \frac{2/3}{1/3 + 2/3} \\ &= \frac{2}{3}. \end{aligned}$$

So now your posterior belief is that the door you originally chose has the car with probability $1/3$ and does not have the car with probability $2/3$. Thus you should switch!