Math 3081 (Probability and Statistics)

Lecture #5 of 27 ~ July 13, 2021

Independence and Computing General Probabilities

- More With Independence
- Computing General Probabilities

This material represents §1.4.2-§1.4.3 from the course notes and problems 9-14 from WeBWorK 2.

Recall

Recall the definition of conditional probability:

Definition

If A and B are events and $P(B) > 0$, we define the conditional probability $P(A|B)$, the probability that A occurs given that B occurred, as $P(A|B) = P(A \cap B)/P(B)$.

We also defined independence.

Definition

We say that two events A and B are independent when $P(A \cap B) = P(A) \cdot P(B)$, which is equivalent to $P(A|B) = P(A)$ and to $P(B|A) = P(B)$. Two events that are not independent are said to be dependent.

We can use independence to compute probabilities of combinations of independent events as products.

Computing More Probabilities, I

We will now extend the ideas we have developed (of computing probabilities as products) to cover non-independent events.

- Our starting point is the rearrangement of the conditional probability formula: $P(A \cap B) = P(B|A) \cdot P(A)$.
- We interpret this formula as follows: if we want to find the probability that both A and B occur, first we compute the probability that A occurs, and then we multiply this by the probability that B also occurs given that A occurred.
- In many situations, it turns out to be much easier to compute the probabilities $P(A)$ and $P(B|A)$ separately by viewing the events of "choosing A" and then "choosing B given A" as being choices made in a sequence.
- This formula is the probability version of the multiplication principle, and we use it in the same way.

Computing More Probabilities, II

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

1. The first ball is red.
2. The second ball is red given that the first ball is red.
3. Both balls are red.
4. The first ball is purple and the second ball is red.
5. One ball is red and the other is purple.

Computing More Probabilities, II

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

1. The first ball is red.
2. The second ball is red given that the first ball is red.
3. Both balls are red.
4. The first ball is purple and the second ball is red.
5. One ball is red and the other is purple.

- Let R_1 (respectively P_1) be the event that the first ball is red (respectively purple) and R_2 (respectively P_2) be the event that the second ball is red (respectively purple).

Computing More Probabilities, III

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

1. The first ball is red.

Computing More Probabilities, III

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

1. The first ball is red.

- This event is simply R_1.
- If we use the sample space consisting only of the first ball drawn from the urn, then each of the 16 outcomes is equally likely, and 4 of them yield a red ball drawn.
- Thus, we see $P(R_1) = 4/16 = 0.25$.

Computing More Probabilities, IV

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

2. The second ball is red given that the first ball is red.

Computing More Probabilities, IV

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

2. The second ball is red given that the first ball is red.

- This event is $P(R_2|R_1)$.
- Imagine drawing the first red ball from the urn and then discarding it.
- Then, drawing the second ball is the same as drawing one ball from an urn containing 3 red balls and 12 purple balls.
- By the same logic as before, the probability that a red ball is drawn now is $P(R_2|R_1) = 3/15 = 0.2$.

Computing More Probabilities, V

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

3. Both balls are red.

Computing More Probabilities, V

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

3. Both balls are red.
   - This event is $P(R_1 \cap R_2)$.
   - By the intersection formula, we have $P(R_1 \cap R_2) = P(R_2|R_1) \cdot P(R_1) = (3/15) \cdot (4/16) = 1/20 = 0.05$.

Computing More Probabilities, V

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

3. Both balls are red.

- This event is $P(R_1 \cap R_2)$.
- By the intersection formula, we have $P(R_1 \cap R_2) = P(R_2|R_1) \cdot P(R_1) = (3/15) \cdot (4/16) = 1/20 = 0.05$.

You can think of this calculation $(4/16) \cdot (3/15)$ as working just like the multiplication principle:

- The probability that the first ball is red is $4/16$ (4 red balls from the total of 16).
- Then, once we make that selection, the probability that the second ball is red is $3/15$ (3 red balls from the total of 15).
- The overall probability is then simply the product of the probabilities arising from each choice in the sequence.

Computing More Probabilities, VI

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

4. The first ball is purple and the second ball is red.

Computing More Probabilities, VI

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

4. The first ball is purple and the second ball is red.

- This event is $P(P_1 \cap R_2) = P(R_2|P_1) \cdot P(P_1)$.
- For $P(P_1)$, we see that in choosing the first ball, there are 12 purple out of a total of 16, so $P(P_1) = 12/16 = 0.75$.
- Then, for $P(R_2|P_1)$, after one purple ball is chosen there are 4 red and 11 purple remaining, so the probability of selecting a red one is $4/15$.
- Thus, $P(P_1 \cap R_2) = P(R_2|P_1) \cdot P(P_1) = (12/16) \cdot (4/15) = 0.2$.

Computing More Probabilities, VII

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

5. One ball is red and the other is purple.

Computing More Probabilities, VII

Example: Suppose an urn contains 4 red balls and 12 purple balls. If two balls are randomly drawn from the urn without replacement, determine the probabilities of the following events:

5. One ball is red and the other is purple.

- This event is the union of the two mutually exclusive events $R_1 \cap P_2$ (first red then purple) and $P_1 \cap R_2$ (first purple then red).
- We already computed $P(P_1 \cap R_2) = (3/4) \cdot (4/15) = 1/5$.
- In the same way, we can find $P(R_1 \cap P_2) = (4/16) \cdot (12/15) = 1/5$.
- Thus, the probability of the original event is the sum $1/5 + 1/5 = 0.4$.

Computing More Probabilities, VIII

We can extend the logic in the last example for intersections of more than two events.

- For example, for three events A, B, C, we have $P(A \cap B \cap C) = P(C|A \cap B) \cdot P(A \cap B) = P(C|A \cap B) \cdot P(B|A) \cdot P(A)$.
- If we view these probabilities as a sequence of choices, then this formula tells us that we can compute the probability of $A \cap B \cap C$ by "choosing A", then "choosing B given A", then "choosing C given both A and B".
- The same idea extends to intersections of four or more events.
- The advantage to this approach is that it allows us to break complicated events down into simpler ones whose probabilities can be computed quickly using a small sample space.

Computing More Probabilities, IX

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

1. All four cards are kings.
2. All four cards are diamonds.
3. The first two cards are nines and the last two are aces.
4. The four cards are all different suits.

Computing More Probabilities, IX

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

1. All four cards are kings.
2. All four cards are diamonds.
3. The first two cards are nines and the last two are aces.
4. The four cards are all different suits.

For each of these, we think of selecting each card in order so that it satisfies the required condition, and then multiply all of the corresponding conditional probabilities.

Computing More Probabilities, X

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

1. All four cards are kings.

Computing More Probabilities, X

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

1. All four cards are kings.

- There are four kings in the deck, so the probability of the first card being a king is $4/52$.
- Once the first card is selected as a king, the probability of the second card also being a king is $3/51$ (3 kings out of 51 remaining cards).
- Similarly, the probabilities for the final two cards being kings are $2/50$ and $1/49$.
- Thus, the overall probability is $(4/52) \cdot (3/51) \cdot (2/50) \cdot (1/49) = 1/270725$.
- Note that we could also have solved this problem using counting principles (there are $\binom{52}{4} = 270725$ unordered hands, exactly one of which is the hand with four kings).

Computing More Probabilities, XI

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

2. All four cards are diamonds.

Computing More Probabilities, XI

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

2. All four cards are diamonds.

- There are 13 diamonds in the deck. So, by the same logic we just used, the probability of the first card being a diamond is $13/52$, then for the second card it is $12/51$, for the third card it is $11/50$, and for the last card it is $10/49$.
- Thus, the overall probability is $(13/52) \cdot (12/51) \cdot (11/50) \cdot (10/49) = 11/4165$.

3. The first two cards are nines and the last two are aces.

Computing More Probabilities, XI

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

2. All four cards are diamonds.

- There are 13 diamonds in the deck. So, by the same logic we just used, the probability of the first card being a diamond is 13/52, then for the second card it is 12/51, for the third card it is 11/50, and for the last card it is 10/49.
- Thus, the overall probability is $(13/52) \cdot (12/51) \cdot (11/50) \cdot (10/49) = 11/4165$.

3. The first two cards are nines and the last two are aces.

- By the same sort of approach, this probability is $(4/52) \cdot (3/51) \cdot (4/50) \cdot (3/49) = 6/270725$.

Computing More Probabilities, XII

<u>Example</u>: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

4. The four cards are all different suits.

Computing More Probabilities, XII

Example: Four cards are randomly drawn from a standard 52-card deck. Find the probabilities of these events:

4. The four cards are all different suits.

- For this event, we can think of selecting any card as the first card: probability 52/52.
- Then the next card must be a different suit, so there are 39 such cards out of 51 remaining: probability 39/51.
- The third card must be one of the two remaining suits (26 cards out of 50 total): probability 26/50.
- The last card must be the final remaining suit (13 cards out of 49 total): probability 13/49.
- Thus, the overall probability is $(52/52) \cdot (39/51) \cdot (26/50) \cdot (13/49) = 2197/20825 \approx 10.55\%$.

The Monty Hall Problem, I

Example (Monty Hall Problem): On a game show, a contestant chooses one of three doors: behind one door is a car and behind the other two are goats. The host opens one of the two unchosen doors to reveal a goat, and then offers the contestant the option of switching their choice from their original door to the remaining unopened door in the hopes of winning the car. Should the contestant accept the offer to switch doors?

The Monty Hall Problem, I

<u>Example</u> (Monty Hall Problem): On a game show, a contestant chooses one of three doors: behind one door is a car and behind the other two are goats. The host opens one of the two unchosen doors to reveal a goat, and then offers the contestant the option of switching their choice from their original door to the remaining unopened door in the hopes of winning the car. Should the contestant accept the offer to switch doors?

In this problem, we also make the following implicit assumptions:

- The car is randomly hidden behind one of the doors.
- The host knows what is behind each door.
- The host always opens a door that the contestant has not chosen that reveals a goat (randomly selecting between the two if the contestant has chosen the car).
- The host always offers the option to switch doors.

The Monty Hall Problem, II

We want to compute the probability that the car is behind the contestant's door.

- Suppose we label the contestant's door with the number 1, the door opened by the host with the number 2, and the remaining door with the number 3.
- Let P_1, P_2, P_3 be the events in which the prize is behind door 1, 2, or 3 respectively, and H_2 and H_3 be the events in which the host opens door 2 and door 3 respectively.
- Then the probability we want to compute is $P(P_1|H_2)$, the probability that the prize is behind door 1 given that the host opened door 2.
- From the conditional probability formula, $P(P_1|H_2) = P(P_1 \cap H_2)/P(H_2)$.
- So we are reduced to finding $P(P_1 \cap H_2)$ and $P(H_2)$.

The Monty Hall Problem, III

We want to find $P(P_1|H_2) = P(P_1 \cap H_2)/P(H_2)$.

- First, $P(P_1) = P(P_2) = P(P_3) = 1/3$ since the car is equally likely to be behind any of the doors at the start of the game.
- Also, $P(H_2) = P(H_3) = 1/2$ since by symmetry the host is equally likely to open door 2 or door 3.
- So we need only compute $P(P_1 \cap H_2) = P(H_2|P_1) \cdot P(P_1)$.
- But $P(H_2|P_1) = P(H_3|P_1) = 1/2$ since if the prize is behind door 1, the host is equally likely to open door 2 or door 3.
- Therefore, we get $P(P_1 \cap H_2) = (1/2) \cdot (1/3) = 1/6$.
- So, $P(P_1|H_2) = P(P_1 \cap H_2)/P(H_2) = (1/6)/(1/2) = 1/3$.

Our calculations show there is a $1/3$ probability that the prize is behind door 1 (the door chosen by the contestant), hence a $2/3$ probability that the prize is behind door 3 (the remaining unchosen door). In conclusion: it is better to switch!

The Monty Hall Problem, IV

This problem (the Monty Hall problem) is fairly infamous and was originally popularized in this form by vos Savant in 1990.

- Although she gave the correct answer to the puzzle in her solution, she evidently received many thousands of letters from readers who disagreed with the answer!
- It is a very common mistake to argue that because there are now only 2 doors remaining to choose between, the probability that the prize is behind each of them must be $1/2$.

The Monty Hall Problem, IV

This problem (the Monty Hall problem) is fairly infamous and was originally popularized in this form by vos Savant in 1990.

- Although she gave the correct answer to the puzzle in her solution, she evidently received many thousands of letters from readers who disagreed with the answer!
- It is a very common mistake to argue that because there are now only 2 doors remaining to choose between, the probability that the prize is behind each of them must be $1/2$.
- Another way to reason out the solution to the puzzle is to change the game's formulation from "the host opens one of the unchosen doors" to "the host offers the contestant the choice of having both unchosen doors".
- This new version is equivalent to the original (in terms of whether the contestant wins the car), but now it is very clear that switching is twice as likely to yield the prize!

The Monty Hall Problem, V

Here is another approach to the solution based on listing outcomes:

- Label the doors A, B, C, where the contestant selects door A.
- Before the host opens one of the other doors, the probability that the car is behind any one of these doors is $1/3$.

The Monty Hall Problem, V

Here is another approach to the solution based on listing outcomes:

- Label the doors A, B, C, where the contestant selects door A.
- Before the host opens one of the other doors, the probability that the car is behind any one of these doors is $1/3$.
- If the car is behind door A, then the host will open door B half the time and door C half the time. Thus overall, $P(\text{car A, host B}) = P(\text{car A, host C}) = 1/6$.
- If the car is behind door B, then the host will open door C every time, so $P(\text{car B, host B}) = 0$, $P(\text{car B, host C}) = 1/3$.
- Likewise, $P(\text{car C, host B}) = 1/3$, $P(\text{car C, host C}) = 0$.
- Now suppose the host opens door B. The total probability of this event is $1/6 + 1/3 = 1/2$, and so the conditional probability that the car is behind door A is $(1/6)/(1/2) = 1/3$.
- Likewise, if the host opens door C, the probability that the car is behind door A is again $(1/6)/(1/2) = 1/3$.

The Birthday Problem, I

Example: Assume that birthdays are randomly distributed among the 365 days in a non-leap year, and ignore February 29th.

1. What is the probability p_n that in a group of n people, some pair have the same birthday?
2. What is the smallest number of people required for which there will be at least a 50% chance of having two people with the same birthday?

Before we analyze this problem, take a moment to estimate how big you think the answer to the second question is. (Is 10 people enough to get a 50% chance? 50 people? 100 people? 200 people?)

The Birthday Problem, II

1. What is the probability p_n that in a group of n people, some pair have the same birthday?

- If we have 5 people, there are many ways that some pair of them could have the same birthday (e.g., the first two people, the last two people, the first and fourth, etc.).
- It is not so easy to answer the problem by counting in this manner.

The Birthday Problem, II

1. What is the probability p_n that in a group of n people, some pair have the same birthday?

- If we have 5 people, there are many ways that some pair of them could have the same birthday (e.g., the first two people, the last two people, the first and fourth, etc.).
- It is not so easy to answer the problem by counting in this manner.
- Instead, consider the complementary event: no two people have the same birthday, which is to say, all the birthdays are different. Now go through, one at a time:
- The first person may have any birthday: probability 1.
- The second person may have any birthday except the 1 already used: probability 364/365.
- The third person may have any birthday except the 2 already used: probability 363/365.

The Birthday Problem, III

1. What is the probability p_n that in a group of n people, some pair have the same birthday?

- In general, the kth person has $366 - k$ possible birthdays, so the probability of this event (conditioned on the previous ones) is $(366 - k)/365$.

The Birthday Problem, III

1. What is the probability p_n that in a group of n people, some pair have the same birthday?

- In general, the kth person has $366 - k$ possible birthdays, so the probability of this event (conditioned on the previous ones) is $(366 - k)/365$.
- Therefore, by our formula for the probability of the intersection of events, the probability that no two people have the same birthday is $\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \cdots \cdot \frac{366-n}{365}$.
- This means the answer to the original question is

$$p_n = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \cdots \cdot \frac{366-n}{365}.$$

The Birthday Problem, IV

1. What is the probability p_n that in a group of n people, some pair have the same birthday?

The formula is nice enough, but here are some actual values:

n	5	10	15	20	22	23
p_n	2.7%	11.7%	25.3%	41.1%	47.6%	50.7%
n	25	30	40	50	60	70
p_n	56.9%	70.6%	89.1%	97.0%	99.4%	99.9%

The Birthday Problem, IV

1. What is the probability p_n that in a group of n people, some pair have the same birthday?

The formula is nice enough, but here are some actual values:

n	5	10	15	20	22	23
p_n	2.7%	11.7%	25.3%	41.1%	47.6%	50.7%
n	25	30	40	50	60	70
p_n	56.9%	70.6%	89.1%	97.0%	99.4%	99.9%

In particular, the tables show that the answer to the second part

2. What is the smallest number of people required for which there will be at least a 50% chance of having two people with the same birthday?

is 23 people.

The Birthday Problem, V

The number we derived (23 people) tends to be surprisingly small to most people. We can give a heuristic argument for why this is actually the right order of magnitude:

- Observe that the desired event (some pair of people share a birthday) is the union of the events that one of the pairs of people in the group has the same birthday.
- The probability that any given pair shares a birthday is $1/365$.
- Although these events are not independent, they are moderately close to independent. With k people, there are $\binom{k}{2}$ such pairs, and so the rough probability of getting at least one match is $1 - (364/365)^{\binom{k}{2}} \approx 1 - \binom{k}{2}/365$.
- Setting this equal to 50% leads to an estimate $k \approx \frac{1}{2} + \sqrt{365} \approx 19.6$, not too far off the actual answer of 23.

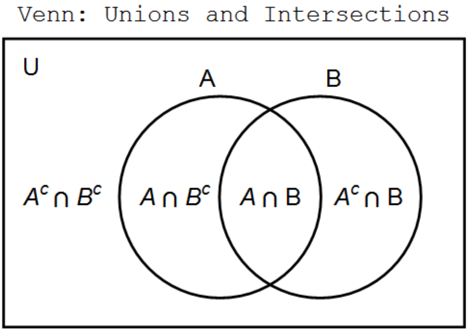
Probability With Venn Diagrams, I

Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find the following:

1. $P(A \cap B)$.
2. $P(A^c \cap B)$.
3. $P(B)$.
4. $P(A \cap B^c)$.
5. $P(A^c \cap B^c)$.
6. $P(A \cup B)$.
7. $P(A|B)$.
8. $P(A|B^c)$.
9. $P(A \cup B|A)$.
10. $P(A \cap B|A \cup B)$.
11. $P(A^c \cup B)$.
12. $P(A^c \cup B|A \cup B^c)$.

Probability With Venn Diagrams, II

When working with intersections and unions of events, it is very helpful to label the results on a Venn diagram like the one below:



We can label each region with its associated probability.

Probability With Venn Diagrams, III

Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find the following:

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

   - By the conditional probability formula, $P(A \cap B) = P(B|A) \cdot P(A) = 0.7 \cdot 0.6 = 0.42$.

2. $P(A^c \cap B)$.

   - Since $P(A) = 0.6$, we have $P(A^c) = 1 - 0.6 = 0.4$.
   - Then by the conditional probability formula, $P(A^c \cap B) = P(B|A^c) \cdot P(A^c) = 0.2 \cdot 0.4 = 0.08$.

Probability With Venn Diagrams, IV

Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find the following:

3. $P(B)$.

- As is readily seen from the Venn diagram, B is the union of the two mutually exclusive events $A \cap B$ and $A^c \cap B$.
- We computed $P(A \cap B) = 0.42$ and $P(A^c \cap B) = 0.08$ earlier, so we must have $P(B) = 0.42 + 0.08 = 0.5$.

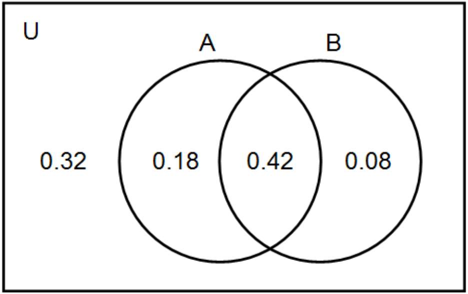
4. $P(A \cap B^c)$.

- As is readily seen from the Venn diagram, A is the union of the two mutually exclusive events $A \cap B$ and $A \cap B^c$.
- Therefore,
  $P(A \cap B^c) = P(A) - P(A \cap B) = 0.6 - 0.42 = 0.18$.

Probability With Venn Diagrams, V

Now that we have found the probabilities associated to three of the four regions in the Venn diagram, we can fill in the last one (on the outside) since the sum of all the probabilities must be 1:

Venn Probabilities, Completed



Probability With Venn Diagrams, VI

Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find the following:

5. $P(A^c \cap B^c)$.
   - We can read this probability directly from the Venn diagram (since it is the region on the outside), so $P(A^c \cap B^c) = 0.32$.
6. $P(A \cup B)$.
   - We can also read this probability directly from the Venn diagram as $0.18 + 0.42 + 0.08 = 0.68$.
   - Alternatively, we could use the union-intersection formula: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.5 - 0.42 = 0.68$.

Probability With Venn Diagrams, VII

Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find the following:

7. $P(A|B)$.

- From the conditional probability formula, $P(A|B) = P(A \cap B)/P(B) = 0.42/0.5 = 0.84$.

8. $P(A|B^c)$.

- From the conditional probability formula, $P(A|B^c) = P(A \cap B^c)/P(B^c) = 0.18/0.5 = 0.36$.

9. $P(A \cup B|A)$.

- $A \cup B$ is guaranteed to occur if A does, so the probability is 1.
- Alternatively, we can see that the intersection of these two events is simply A, so the conditional probability is $P(A \cup B|A) = P(A)/P(A) = 1$.

Probability With Venn Diagrams, VIII

Example: Suppose A and B are events such that $P(A) = 0.6$, $P(B|A) = 0.7$, and $P(B|A^c) = 0.2$. Find the following:

10. $P(A \cap B | A \cup B)$.
    - The intersection of these events is simply $A \cap B$, so $P(A \cap B | A \cup B) = P(A \cap B)/P(A \cup B) = 0.42/0.68 \approx 0.6176$.
11. $P(A^c \cup B)$.
    - We can read this off of the Venn diagram, or use the union-intersection formula: $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B) = 0.4 + 0.5 - 0.08 = 0.82$.
12. $P(A^c \cup B | A \cup B^c)$.
    - We can use the Venn diagram to see the intersection of these two events consists of the two regions $A \cap B$ and $A^c \cap B^c$, so its probability is $P(A \cap B) + P(A^c \cap B^c) = 0.42 + 0.32 = 0.74$.
    - Thus, $P(A^c \cup B | A \cup B^c) = 0.74/0.92 \approx 0.8043$.

This One Got Way Realer in 2020, I

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

1. What is the probability that a randomly chosen patient who tests positive actually has the disease?
2. What is the probability that a randomly chosen patient who tests negative actually doesn't have the disease?
3. What are the probabilities if everyone is tested twice?

This One Got Way Realer in 2020, I

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

1. What is the probability that a randomly chosen patient who tests positive actually has the disease?
2. What is the probability that a randomly chosen patient who tests negative actually doesn't have the disease?
3. What are the probabilities if everyone is tested twice?

- Let D, ND, $+$, and $-$ be the events of having the disease, not having the disease, testing positive, and testing negative.
- We are given $P(D) = 0.02$, $P(ND) = 0.98$, $P(+|D) = 0.97$ so that $P(-|D) = 0.03$, and $P(-|ND) = 0.99$ so that $P(+|ND) = 0.01$.

This One Got Way Realer in 2020, II

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

1. What is the probability that a randomly chosen patient who tests positive actually has the disease?

- We have $P(D) = 0.02$, $P(ND) = 0.98$, $P(+|D) = 0.97$, $P(-|D) = 0.03$, $P(-|ND) = 0.99$, $P(+|ND) = 0.01$.
- We want to compute $P(D|+) = P(D \cap +)/P(+)$.

This One Got Way Realer in 2020, II

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

1. What is the probability that a randomly chosen patient who tests positive actually has the disease?

- We have $P(D) = 0.02$, $P(ND) = 0.98$, $P(+|D) = 0.97$, $P(-|D) = 0.03$, $P(-|ND) = 0.99$, $P(+|ND) = 0.01$.
- We want to compute $P(D|+) = P(D \cap +)/P(+)$.
- First, $P(D \cap +) = P(+|D) \cdot P(D) = 0.97 \cdot 0.02 = 0.0194$.
- Also, we have $P(+) = P(D \cap +) + P(ND \cap +)$.
- Since $P(ND \cap +) = P(+|ND) \cdot P(ND) = 0.01 \cdot 0.98 = 0.0098$, this means $P(+) = 0.0194 + 0.0098 = 0.0292$.
- So, $P(D|+) = P(D \cap +)/P(+) = 0.0194/0.0292 \approx 0.6644$.

This One Got Way Realer in 2020, III

<u>Example</u>: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

2. What is the probability that a randomly chosen patient who tests negative actually doesn't have the disease?

- We have $P(D) = 0.02$, $P(ND) = 0.98$, $P(+|D) = 0.97$, $P(-|D) = 0.03$, $P(-|ND) = 0.99$, $P(+|ND) = 0.01$.
- Now we want $P(ND|-) = P(ND \cap -)/P(-)$.

This One Got Way Realer in 2020, III

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

2. What is the probability that a randomly chosen patient who tests negative actually doesn't have the disease?

- We have $P(D) = 0.02$, $P(ND) = 0.98$, $P(+|D) = 0.97$, $P(-|D) = 0.03$, $P(-|ND) = 0.99$, $P(+|ND) = 0.01$.
- Now we want $P(ND|-) = P(ND \cap -)/P(-)$.
- First, $P(ND \cap -) = P(-|ND) \cdot P(ND) = 0.99 \cdot 0.98 = 0.9702$.
- Also, $P(D \cap -) = P(-|D) \cdot P(D) = 0.03 \cdot 0.02 = 0.0006$.
- Thus, $P(-) = 0.9702 + 0.0006 = 0.9708$. Therefore, $P(ND|-) = P(ND \cap -)/P(-) = 0.9702/0.9708 \approx 0.9994$.

This One Got Way Realer in 2020, IV

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

3. What are the probabilities if everyone is tested twice?

- We analyse the cases separately.

This One Got Way Realer in 2020, IV

<u>Example</u>: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

3. What are the probabilities if everyone is tested twice?

- We analyse the cases separately.
- First suppose that the patient has the disease. Then the probability they obtain two positive tests is 0.97^2, so $P(D \cap ++) = 0.02 \cdot 0.97^2 = 0.018818$.
- The probability they obtain two negative tests is $P(D \cap --) = 0.02 \cdot 0.03^2 = 0.000018$.
- Now suppose that the patient does not have the disease.
- Using the same logic as above, we can compute that $P(ND \cap ++) = 0.98 \cdot 0.01^2 = 0.000098$ and $P(ND \cap --) = 0.98 \cdot 0.99^2 = 0.960498$.

This One Got Way Realer in 2020, V

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

3. What are the probabilities if everyone is tested twice?

- We calculated $P(D \cap ++) = 0.018818$, $P(D \cap --) = 0.000018$, $P(ND \cap ++) = 0.000098$ and $P(ND \cap --) = 0.960498$.
- Thus, the total probability of obtaining two positive tests is $P(D \cap ++) + P(ND \cap ++) = 0.018916$.
- So, if someone tests positive twice the probability they have the disease is $P(D| + +) = 0.018818/0.018916 \approx 0.9948$.
- Likewise, the total probability of having two negative tests is 0.960516, so $P(ND| - -) = 0.960498/0.960516 \approx 0.99998$.

This One Got Way Realer in 2020, VI

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

3. What are the probabilities if everyone is tested twice?

- We can also examine the case where the tests are inconclusive (one positive, one negative).

This One Got Way Realer in 2020, VI

Example: A medical test for a disease will detect the disease in 97% of test samples from patients who have the disease, but it also has a false positive rate of 1%. Suppose 2% of the population actually has the disease.

3. What are the probabilities if everyone is tested twice?

- We can also examine the case where the tests are inconclusive (one positive, one negative). As before, we can calculate $P(+-|D) = P(-+|D) = 0.97 \cdot 0.03 \cdot 0.02 = 0.000582$ while $P(+-|ND) = P(-+|ND) = 0.01 \cdot 0.99 \cdot 0.98 = 0.009702$.
- Thus, the total probability of getting one positive and one negative test is $2 \cdot 0.000582 + 2 \cdot 0.009702 = 0.020568$.
- In this situation, the conditional probability of having the disease is $2 \cdot 0.000582/0.020568 \approx 0.0566$ (about 5.66%), while the conditional probability of not having the disease is $2 \cdot 0.009702/0.020568 \approx 0.9434$ (about 94.34%).

This One Got Way Realer in 2020, VII

You may find it surprising that single positive test (which gives the correct result with at least "97% accuracy" in all scenarios) still only yields about a 2/3 probability of actually having the disease.

- Ultimately, the reason for this disparity is that a typical person is very unlikely to have the disease: so, even though the false positive rate is low, since the disease rate is also low, the total number of correct positives ends up being the same order of magnitude as the total number of false positives.
- Thus, even if a person tests positive once, it is still not especially likely that they have the disease, unless there is some reason to think that the person was not a randomly-selected member of the population.
- However, doing a second test (as we saw) vastly improves the accuracy of identifying people who have the disease, at least under the assumption that the two tests are independent.

Bayes' Formula, I

In this last example, we computed a conditional probability using the values of the conditional probabilities "in the other order".

- Specifically, we used $P(+|D)$ and $P(+|ND)$ to find $P(D|+)$, and we also used $P(-|D)$ and $P(-|ND)$ to find $P(ND|-)$.

The main idea in each case was to use the given information to compute the two necessary probabilities for finding the desired conditional probability.

Bayes' Formula, II

We can give a general formula for making calculations like these:

Theorem (Bayes' Formula)

If A and B are any events, then

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

More generally, if events $B_1, B_2, \dots, B_k$ are mutually exclusive and have union the entire sample space, then

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A|B_1) \cdot P(B_1) + \cdots + P(A|B_k) \cdot P(B_k)}.$$

This result is named after Rev. Thomas Bayes. We will do more with it next time.

Summary

We discussed collective independence and its uses in calculating probabilities of sequences of events

We discussed methods for computing probabilities of sequences of events using conditional probabilities.

We discussed the Monty Hall problem, the Birthday problem, and how to use Venn diagrams in probability problems.

Next lecture: Bayes' theorem, applications of probability.