Probability Theory

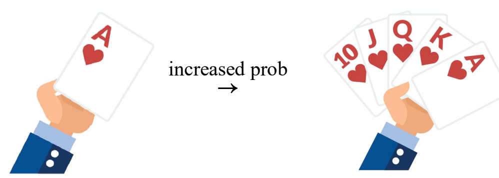
Instructor: Jodin Morey   moreyj@lemoyne.edu

Previous Lecture

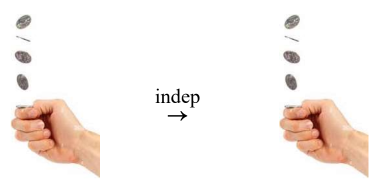
- Conditional Prob: Prior prob, Posterior prob
- Bayes' Rule
- LOTP
- Conditional Probs are Probs

§2.5 - Independence (indep) of Events

Sometimes knowing an event occurred $P(A) = 1$, changes the prob of another event B occurring.

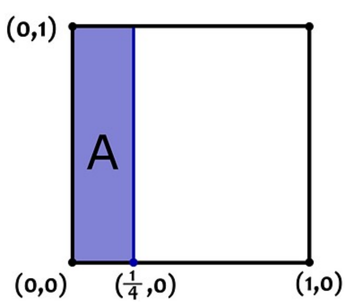


When two events are not linked this way, they are called independent.

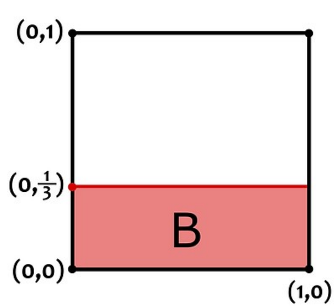


So: A and B are *indep* if learning B occurred gives us no info that would change the prob for A occurring (and vice versa).

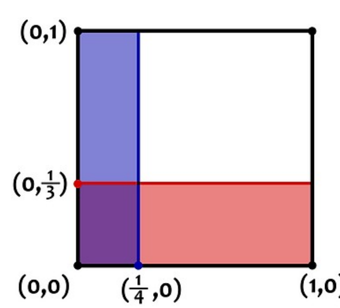
Visualization:



$P(A) = \frac{1}{4}$



$P(B) = \frac{1}{3}$



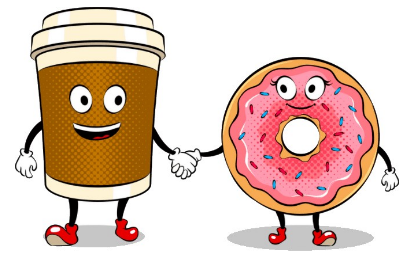
$P(A|B) = P(A) = \frac{1}{4}$

$P(A \cap B) = P(A)P(B) = \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}$

We put A and B on different axes so that the occurrence of each has no effect on the prob of the other.

Def (**Indep of Two Events**): A and B are indep if: $P(A \cap B) = P(A)P(B)$.

Note: if $P(A) > 0$ and $P(B) > 0$, this is equivalent to $P(A|B) = P(A)$ and $P(B|A) = P(B)$.



Ex (**Loyalty Program**): A local café offers a loyalty program where customers earn points by purchasing a coffee or donut. On any given day, the prob that a customer orders a

- Coffee is $P(C) = 0.6$.
- Donut is $P(D) = 0.4$.

Assume the decisions to order a coffee or a donut are indep.

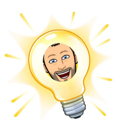
a) What's the prob that a customer orders both a coffee and a donut?

$$P(C \cap D) = P(C)P(D) = (0.6)(0.4) = 0.24.$$

a) What's the prob that a customer orders at least one of these two items?

$$P(C \cup D) = P(C) + P(D) - P(C \cap D)$$

$$= 0.6 + 0.4 - 0.24 = 0.76.$$



Knowing two events are indep allows us to calculate both the intersection and union.

Def (**Indep of Three Events**). Events A, B, and C are indep if *all* of the following hold:

- $P(A \cap B) = P(A)P(B)$,
- $P(A \cap C) = P(A)P(C)$,
- $P(B \cap C) = P(B)P(C)$,
- $P(A \cap B \cap C) = P(A)P(B)P(C)$.

The first three bullet pts are called being **pairwise indep**.

But being pairwise indep does not imply indep (still need last bullet pt). (this generalizes to n events)

Activity: 5b

Def (**Conditional Indep**). A and B are conditionally indep given E if $P(A \cap B|E) = P(A|E)P(B|E)$.

☣ It's easy to make blunders by confusing indep and conditional indep. (can't just insert "$|E$" *everywhere*)

Two events can be:

- indep, but not conditionally indep given E. $P(A,B) = P(A)P(B) \nRightarrow P(A,B|E) = P(A|E)P(B|E)$
- conditionally indep given E, but not indep. $P(A,B|E) = P(A|E)P(B|E) \nRightarrow P(A,B) = P(A)P(B)$
- conditionally indep given E, but not indep given E^c. $P(A,B|E) = P(A|E)P(B|E) \nRightarrow P(A,B|E^c) = P(A|E^c)P(B|E^c)$

(examples of each in the book)

Ex (**Why is the Baby Crying**?). A certain baby cries if and only if she's hungry, tired, or both.
Let C be event the baby is crying, H be she's hungry, and T be she's tired.
Assume none of the probs are equal to 0 or 1, and let H and T be indep.

(a) Find $P(C)$, in terms of $P(H)$ and $P(T)$.

Since H and T are indep, we have

$$P(C) = P(H \cup T) = P(H) + P(T) - P(H \cap T) = P(H) + P(T) - P(H)P(T).$$

(b) Expand $P(H|C)$, $P(T|C)$, and $P(H,T|C)$ using Bayes' rule.

$$P(H|C) = \frac{P(C|H)P(H)}{P(C)} = \frac{P(H)}{P(C)}, \qquad \text{(she cries if hungry)}$$

$$P(T|C) = \frac{P(C|T)P(T)}{P(C)} = \frac{P(T)}{P(C)}, \qquad \text{(she cries if tired)}$$

$$P(H,T|C) = \frac{P(C|H,T)P(H,T)}{P(C)} = \frac{P(H)P(T)}{P(C)}.$$

(c) Are H and T conditionally indep given C? Explain in two ways:
algebraically using the quantities from (b), and w/an intuitive explanation in words.

No, H and T are not conditionally indep given C, since $P(H,T|C) = \frac{P(H)P(T)}{P(C)} < \frac{P(H)P(T)}{P(C)^2} = P(H|C)P(T|C)$.

If the baby is crying but not hungry, she must be tired! (learning one, $P(H|C)$, gives info about the other, $P(T|C)$) $\square$

With §2.5, you can now accomplish HW 3

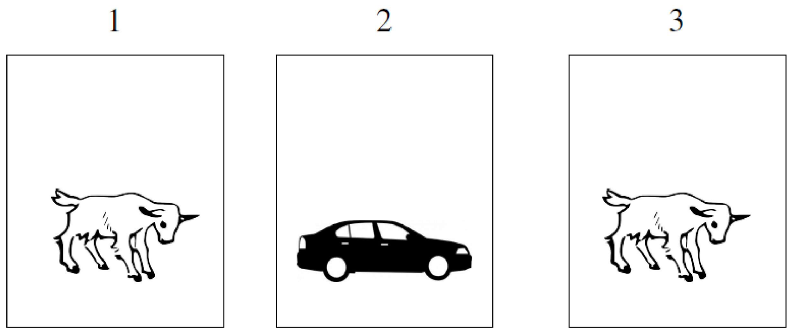
---

§2.7 Conditioning as a Problem-Solving Tool

Strategy. condition on what you wish you knew.

Wishful Thinking. Some problems would be made easier if we knew whether E happened or not.
So we condition on E and E^c, and consider these separately, then combine them using LOTP.

Ex (**Monty Hall**). A game show where you are shown three doors.
There are goats behind two doors, and a car behind the remaining door.
You choose a door. Then, the host of the show opens one of the unchosen doors, revealing a goat.
You are then given the opportunity to either stay with your initial door, or switch to the other closed-door.
Is it better to stay, or to switch?



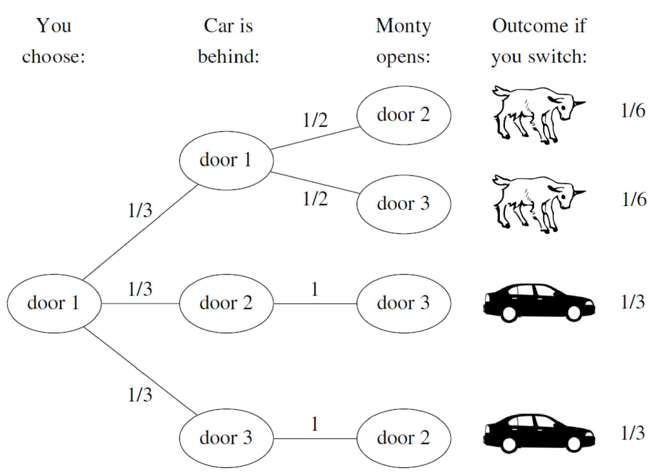
https://www.mathwarehouse.com/monty-hall-simulation-online/

First off, notice that you have a $\frac{1}{3}$ chance of choosing the door with the car on your first pick.

When you are offered the option to switch, many people first believe it doesn't matter if you switch.
They think, "Since there are only two unopened doors, one with the goat and one with a car, you have a 50 percent chance of getting the car, regardless of which door you choose."

But this is incorrect.

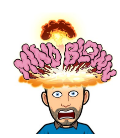
Observe the following decision tree (using multiplication rule along the way).
Without loss of generality, assume you initially choose door 1.



If car is behind 2 or 3, then switching strategy guarantees you'll get the car. (happens 2/3 of the time)

If car is behind door 1, then when you switch, you're switching away from the car, and you're guaranteed not to get the car.

So, switching doubles your chances of getting the car from $\frac{1}{3}$ to $\frac{2}{3}$! But why?



The 2nd time you're choosing, the host has had to reveal information.
You've forced him to choose a door that doesn't have a car behind it.
This information is what changes the odds.

Conditioning on where the car is, the prob you'll get the car if using the switching strategy when choosing door 1 first is:

$$P\left(\text{get car}\right) = P\left(\text{get car}|C_1\right)P(C_1) + P\left(\text{get car}|C_2\right)P(C_2) + P\left(\text{get car}|C_3\right)P(C_3),$$

where C_i represents the event that the car's behind door i.

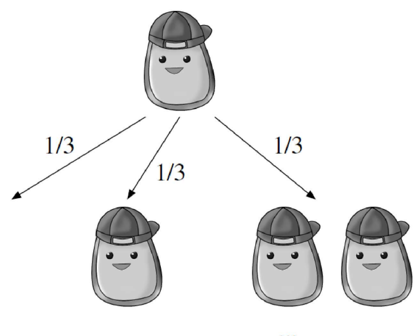
So, $P(\text{get car}) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$.

This follows from the tree diagram above, and the fact that the car is equally likely ($\frac{1}{3}$) to be behind each of the doors. □

Strategy. Condition on the First Step

First-Step Analysis. In problems with a recursive structure, it can often be useful to condition on the first step of the experiment. The next two examples apply this strategy, which we call first-step analysis.

Ex (**Branching process**). A single amoeba, Bobo, lives in a pond. After one minute Bobo will either die, split into two amoebas, or stay the same, with equal probability. In subsequent minutes all living amoebas will behave the same way, independently. What is the probability that the amoeba population will eventually die out?



Solution. Let D be the event that the population eventually dies out; we want to find $P(D)$. We proceed by conditioning on the outcome of the first step: let B_i be the event that Bobo turns into i amoebas after the first minute, for $i = 0, 1, 2$. We know $P(D|B_0) = 1$ and $P(D|B_1) = P(D)$ (if Bobo stays the same, we're back to where we started). If Bobo splits into two, then we just have two indep versions of our original problem! We need both of the offspring to eventually die out, so $P(D|B_2) = P(D)^2$. Now we have exhausted all the possible cases and can combine them with LOTP:

$$P(D) = P(D|B_0) \cdot \tfrac{1}{3} + P(D|B_1) \cdot \tfrac{1}{3} + P(D|B_2) \cdot \tfrac{1}{3}$$

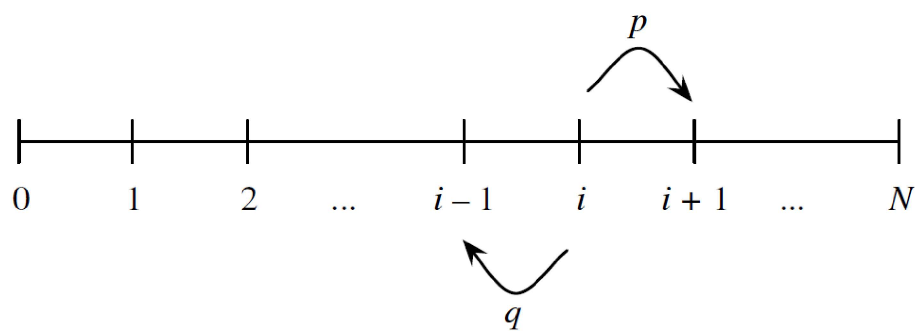
$$= 1 \cdot \tfrac{1}{3} + P(D) \cdot \tfrac{1}{3} + P(D)^2 \cdot \tfrac{1}{3}.$$

Solving for $P(D)$ gives $P(D) = 1$: the amoeba population will die out with prob 1. ■

Ex (**Gambler's ruin**). Two gamblers, A and B, make a sequence of \$1 bets. In each bet, A has probability p of winning, and B has probability $q = 1 - p$. A starts with i dollars and gambler B starts with $N - i$; total wealth between the two remains constant since every time A loses a dollar, the dollar goes to B, and vice versa.

We can visualize this game as a random walk on the integers between 0 and N, where p is the prob of going right in a given step. Imagine a person who starts at position i and, at each time step, moves one step to the right w/prob p and a step left w/prob $q = 1 - p$. The game ends when either A or B is ruined, i.e., when the random walk reaches 0 or N.

What's the prob that A wins (walks away with all the money)?



Solution. We recognize that this game, has a recursive structure.

After the first step, it's exactly the same game, except that A's wealth is now either $i+1$ or $i-1$.

Let p_i be the probability that A wins the game, given that A starts with i dollars.

We will use first-step analysis to solve for the p_i. Let W be the event that A wins the game.

By LOTP, conditioning on the outcome of the first…

Harvard Video (1st half):
youtube.com/watch?v=fDcjhAKuhqQ&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=7

§2.6 - Coherency of Bayes' Rule

Coherence. if we receive multiple pieces of info and wish to update our probs to incorporate all info, it doesn't matter whether we update **sequentially** (taking each piece of evidence into account one at a time), or **simultaneously** (using all evidence at once).

Ex. (**Testing for a Rare Disease**). Suppose Fred tested positive for "conditionitis" using a diagnostic test T_1. Let D be the event of having the disease. It's known that one in 100 people have the disease. The prob of the test being positive if a person has the disease is 0.9. And the prob of a false positive is 0.2. What's the prob that Fred has conditionitis given he also tests positive for a second test T_2?

Notation: $P(D) = 0.01, \qquad P(T_i|D) = 0.9, \qquad P(T_i|D^c) = 0.2.$

We wish to compute: $P(D|T_1,T_2)$.

One-step method (**Simultaneous Conditioning**):

Using Bayes' Thm directly:

$P(D|T_1,T_2) = \frac{P(T_1,T_2|D)P(D)}{P(T_1,T_2)}$

Since the tests are indep given D:

$P(T_1, T_2|D) = P(T_1|D)P(T_2|D) = 0.9 \cdot 0.9 = 0.81.$

$P(T_1, T_2|D^c) = P(T_1|D^c)P(T_2|D^c) = 0.2 \cdot 0.2 = 0.04.$

Using LOTP for the denominator: $P(T_1, T_2) = P(T_1, T_2|D)P(D) + P(T_1, T_2|D^c)P(D^c)$

$= (0.81 \cdot 0.01) + (0.04 \cdot 0.99) = 0.0081 + 0.0396 = 0.0477.$

Thus, $P(D|T_1, T_2) = \frac{0.81 \cdot 0.01}{0.0477} \approx 0.17.$

Two-step method (Sequential Conditioning):

Step 1: Conditioning on T_1:

$P(D|T_1) = \frac{P(T_1|D)P(D)}{P(T_1)}$, where

$P(T_1) = P(T_1|D)P(D) + P(T_1|D^c)P(D^c)$

$= (0.9 \cdot 0.01) + (0.2 \cdot 0.99) = 0.009 + 0.198 = 0.207.$

So, $P(D|T_1) = \frac{0.009}{0.207} \approx 0.0435.$

Step 2: Conditioning on T_2

$P(D|T_2, T_1) = \frac{P(T_2|D)P(D|T_1)}{P(T_2|T_1)}$, where

$P(T_2|T_1) = P(T_2|D, T_1)P(D|T_1) + P(T_2|D^c, T_1)P(D^c|T_1)$

$= P(T_2|D)P(D|T_1) + P(T_2|D^c)P(D^c|T_1)$ (if we know whether you have the disease, the previous test results provide no useful info)

$= (0.9 \cdot 0.0435) + (0.2 \cdot (1 - 0.0435))$

$= 0.03915 + 0.1913 = 0.23045$

So, $P(D|T_1, T_2) = \frac{0.9 \cdot 0.0435}{0.23045} = \frac{0.03915}{0.23045} \approx 0.17.$

So it doesn't matter if we condition simultaneously or sequentially.

Therefore, you can choose to do whichever is easiest in the given situation!

§2.8 - Pitfalls and Paradoxes

Ex (**Prosecutor's Fallacy**). [see book]

Ex (**Defense Attorney's Fallacy**). A woman has been murdered. The husband's put on trial. The husband has a history of abusing his wife. The defense attorney argues that evidence of abuse should be excluded on grounds of irrelevance, since only 1 in $10,000$ abusive husbands subsequently murder their wives. Should the judge grant the defense attorney's motion to bar this evidence from trial?

Suppose 1-in-$10,000$ figure is correct. Furthermore, assume: 1 in 10 husbands abuse their wives, 1 in 5 murdered wives were murdered by their husbands, and 50% of husbands who murder their wives previously abused them. Assume that if the husband of a murdered wife is not guilty of the murder, then prob that he abused his wife reverts to the unconditional probability of abuse.

Let A be event that the husband commits abuse against his wife, and let G be "the husband is guilty." The defense's argument is $P(G|A) = 1 = 10,000$, so guilt is still extremely unlikely conditional on a previous history of abuse. However, this fails to condition on a crucial fact: in this case, **we know that the wife was murdered**. Therefore, the relevant probability is not $P(G|A)$, but $P(G|A,M)$, where M is event "wife was murdered." Bayes' rule with extra conditioning gives:

$$P(G|A,M) = \frac{P(A|G,M)P(G|M)}{P(A|G,M)P(G|M)+P(A|G^c,M)P(G^c|M)} = \frac{0.5\cdot 0.2}{0.5\cdot 0.2+0.1\cdot 0.8} = \frac{5}{9}.$$

So posterior probability of guilt, $P(G|A,M)$, is over $5,000 \times P(G|A)$.
Conditioning on evidence of abuse increases prob of guilt from $P(G|M) = 0.2$ to $P(G|A,M) \approx 0.56$, so husband's history of abuse gives important info, contrary to defense attorney's argument.

We must condition on **all** the evidence. ■

Ex (**Simpson's paradox**). Two doctors, Dr. Hibbert & Dr. Nick, each perform two types of surgeries: heart surgery and Band-Aid removal. Each surgery can be either a success or a failure. The two doctors' respective records are given in the following tables:

	Heart	Band-Aid
Success	70	10
Failure	20	0

Dr. Hibbert

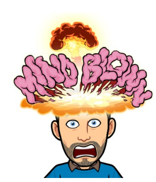
	Heart	Band-Aid
Success	2	81
Failure	8	9

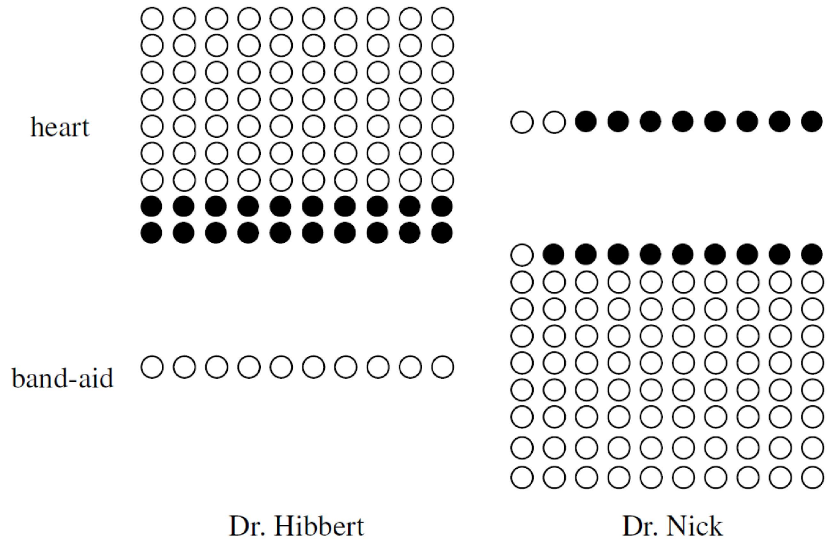
Dr. Nick

Dr. Hibbert had a higher success rate than Dr. Nick in heart surgeries: 70 out of 90 versus 2 out of 10.

Dr. Hibbert also had a higher success rate in Band-Aid removal: 10 out of 10 versus 81 out of 90.

If we aggregate across both types of surgeries to compare overall surgery success rates, Hibbert was successful 80 out of 100 surgeries. Nick was successful 83/100 surgeries: Dr. Nick's overall success rate is higher!





What's happening is that Dr. Hibbert is performing a greater number of heart surgeries, which are inherently riskier than Band-Aid removals.

Notationally: For events A, B, and C, we say that we have a Simpson's paradox if:

$$P(A|B,C) < P(A|B^c,C) \qquad P(A|B,C^c) < P(A|B^c,C^c), \qquad \text{but} \qquad P(A|B) > P(A|B^c).$$

In our example, A is event of a successful surgery, B is Dr. Nick as surgeon, and C is a heart surgery.

Harvard Video: youtube.com/watch?v=PNrqCdslGi4&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=8

What did we learn?

- Indep of Events (conditional indep)
- Conditioning as a Problem-Solving
  Wishful Thinking/First Step Analysis
- Coherency of Bayes' Rule
- Pitfalls and Paradoxes