STEVENS
INSTITUTE OF TECHNOLOGY
1870

MA 540: Introduction to Probability Theory

Lecture 3:Conditional Probability and Independence

Benjamin Leinwand

Motivating Example

- Let’s say you’re on a dating app, and you’re shown a new person’s profile
- You’ve seen a lot of profiles, and your general rate of matching with somebody is p
  - Without knowing anything about the next person’s profile, expect to match with rate p
- What if you know a piece of information about the next person? How would it change your probability of matching with them?
  - Educational attainment
  - Income
  - Height
  - Favorite color
  - Hobbies
  - Grandfather’s first name

Conditional Probability

- Sometimes knowing event *A* occurred can tell you whether or not event *B* occurred
  - What if $A \subseteq B$?
  - What if $A \cap B = \emptyset$?

- Other times, knowing that event *A* occurred can change the probability event *B* occurred
  - Example: We won't match with every person who shares one of our hobbies, but we are more likely to match with someone who does than with someone who doesn't

- If P(B) > 0 then

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

Conditional vs Unconditional probability

- If we roll 2 dice, what is the probability of the event "the sum is 8?"
  - There are 36 possible outcomes, all with equal probability
  - {(2,6), (3,5), (4,4), (5,3), (6,2)} are the only outcomes in the event
  - 5/36

- If we know the first die is a 3, what is the probability of the event "the sum is 8?"
  - There are now only 6 possible outcomes in the event "the first die is a 3"
  - {(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)}
  - Of these, only {(3, 5)} sums to 8
  - Since all outcomes have same probability, the conditional probability is 1/6 ≠ 5/36

- Compare to result from formula: if event A is "the sum is 8?" and event B is "the first die is a 3"

  - $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{6/36} = 1/6$

Example

- Joe is 80 percent certain that his missing key is in one of the two pockets of his hanging jacket, being 40 percent certain it is in the left-hand pocket and 40 percent certain it is in the right-hand pocket. If a search of the left-hand pocket does not find the key, what is the conditional probability that it is in the other pocket?

Example 2.2

- In the card game bridge, the 52 cards are dealt out equally to 4 players—called East, West, North, and South. If North and South have a total of 8 spades among them, what is the probability that East has 3 of the remaining 5 spades?

Conditional probability is in fact a probability

1. $0 \leq P(A|B) \leq 1$
2. $P(S|B) = 1$
3. Letting $A_1, A_2, A_3 \ldots$ be mutually exclusive events (that is $A_i \cap A_j = \emptyset$ if $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i \,|B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

- Proof on board

Multiplication rule + proof

Letting $A_1, A_2, \ldots, A_n$ be events,

$P(A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n) = P(A_1) * P(A_2|A_1) P(A_3|A_1 \cap A_2) * P(A_n|A_1 \cap A_2 \cap \ldots \cap A_2)$

- Proof:

$$P(A_1) * \frac{P(A_1 \cap A_2)}{P(A_1)} * \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} * \cdots * \frac{P(A_1 \cap A_2 \cap \cdots \cap A_{n-1})}{P(A_1 \cap A_2 \cap \cdots \cap A_{n-2})} * \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n)}{P(A_1 \cap A_2 \cap \cdots \cap A_{n-1})}$$

$$= \quad P(A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n)$$

Bayes rule

- A disease has a rate of 1/10000. A test screening for this diseases has a 99% accuracy rate –that is, if you have the disease, you will test positive 99% of the time, and if you don't have the disease, you will test negative 99% of the time.

- If someone receives a positive test, what's their probability of having the disease?

Bayes rule

- A disease has a rate of 1/10000. A test screening for this diseases has a 99% accuracy rate –that is, if you have the disease, you will test positive 99% of the time, and if you don't have the disease, you will test negative 99% of the time.

- If someone receives a positive test, what's their probability of having the disease?

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

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

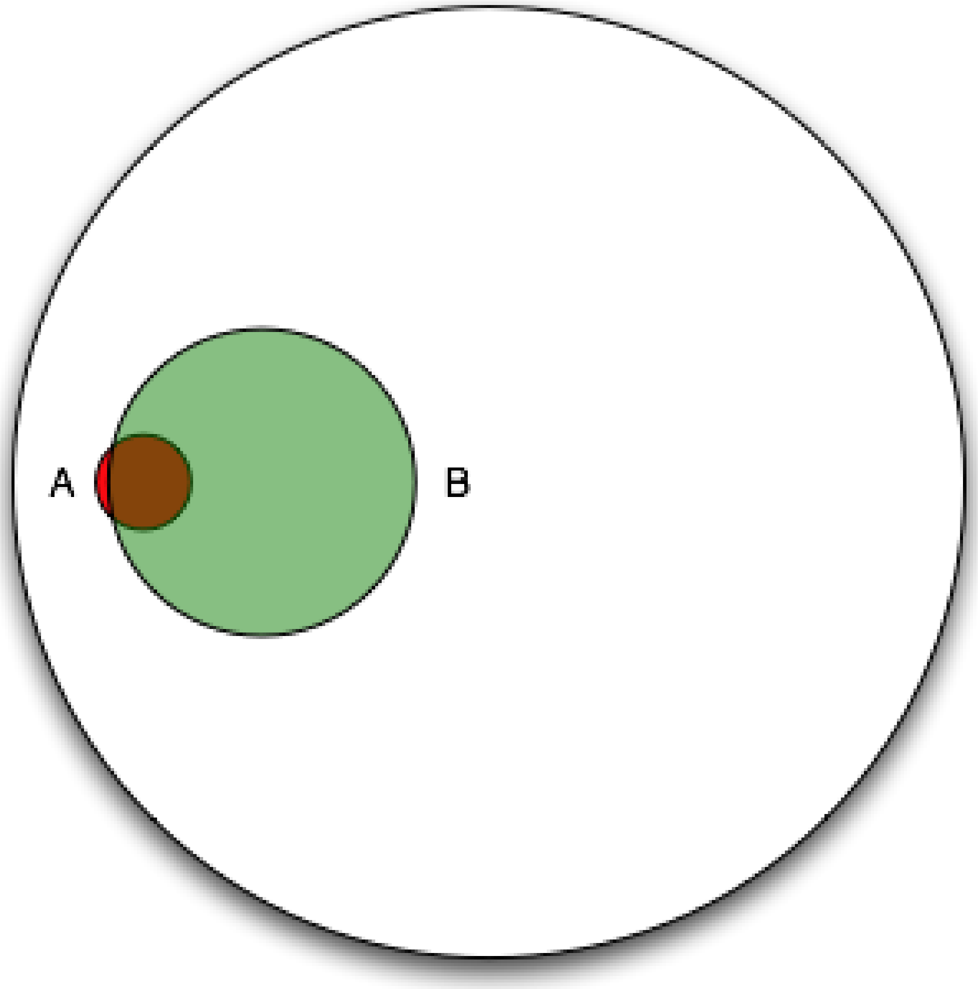
$$\frac{\frac{99}{100} * \frac{1}{10000}}{\left(.99*\frac{1}{10000}\right)+\left(.01*\frac{9999}{10000}\right)} = .0098$$

Example

- A bin contains 3 types of disposable flashlights. The probability that a type 1 flashlight will give more than 100 hours of use is .7, with the corresponding probabilities for type 2 and type 3 flashlights being .4 and .3, respectively. Suppose that 20 percent of the flashlights in the bin are type 1, 30 percent are type 2, and 50 percent are type 3.
- (a) What is the probability that a randomly chosen flashlight will give more than 100 hours of use?
- (b) Given that a flashlight lasted more than 100 hours, what is the conditional probability that it was a type j flashlight, j = 1, 2, 3?

Bayes rule as Venn Diagram

- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)}$
- P(A) is much smaller than P(B), but both are relatively small
- P(B|A) is large
- However, P(A|B) is still small



Bayes rule as a table

	Positive Test	Negative Test
Has disease	.99*.0001	.01*.0001
No disease	.01 * .9999	.99 * .9999

	Positive Test	Negative Test
Has disease	99	1
No disease	9999	989901

P(B|A) in this case is the top left cell divided by the sum of the left column

People seem to find Bayes rule unintuitive in practice. Even using the table above is tricky

When we translate into the expected number of people in each group, it may become clearer

Bayes Rule as a tree

- When a patient goes through breast cancer screening there are two competing claims: patient had cancer and patient doesn't have cancer. If a mammogram yields a positive result, what is the probability that patient actually has cancer?

P(disease|positive)

$$= \frac{.000099}{.000099+.009999}$$

Example – tech interview question

- Seattle is famous for raining. Let’s assume the probability it's raining in Seattle on any given day is 1/4. You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes, it is raining”. What is the probability that it's actually raining in Seattle?

Example – tech interview question

- Seattle is famous for raining. Let's assume the probability it's raining in Seattle on any given day is 1/4. You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes, it is raining". What is the probability that it's actually raining in Seattle?

- P(A) = P(It's raining)= ¼
- P(B) = all friends say it's raining
- P(A|B) = P(B|A)P(A)/P(B)
- P(B|A) = 2/3 * 2/3 *2/3
- P(A and B) = 2/3 * 2/3 *2/3 *1/4
- P(A and B^C) = 1/3 * 1/3 * 1/3 *3/4

$$\frac{2/3 * 2/3 * 2/3 * 1/4}{(2/3 * 2/3 * 2/3 * 1/4) + (1/3 * 1/3 * 1/3 * 3/4)}$$

$$\frac{8/108}{11/108} = 8/11$$

Example – tech interview question 2

- There are 100 coins in a box. The first coin is two-headed. The rest are fair coins. When one coin was picked at random from the box and tossed, it landed heads.
- What is the probability that the selected coin was the two-headed coin?

Example – tech interview question 2

- There are 100 coins in a box. The first coin is two-headed. The rest are fair coins. When one coin was picked at random from the box and tossed, it landed heads.
- What is the probability that the selected coin was the two-headed coin?

$$\frac{1 * \left(\frac{1}{100}\right)}{\left(1 * \left(\frac{1}{100}\right)\right) + \left(\frac{1}{2} * \left(\frac{99}{100}\right)\right)} = \frac{1/100}{\frac{101}{200}} = 2/101$$

- There's an intuitive explanation here, too. There are 101 total faces showing heads. 2 of those faces are on the two-headed coin

Example 3 – Monty Hall

- Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

Example 3 – Monty Hall

- Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

- P(door 1 wins) = (1/3)
- P(open door 3 | door 1 wins) = ½
- P(open door 3 | door 2 wins) = 1
- P(open door 3 | door 3 wins) = 0

$$\mathrm{P}(door\ 1\ wins\ |3rd\ door\ opened) = \frac{(\frac{1}{3} * \frac{1}{2})}{\left(\frac{1}{3} * \frac{1}{2}\right) + \left(\frac{1}{3} * 1\right) + \left(\frac{1}{3} * 0\right)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Simpson’s paradox

	Full Population, $N = 52$			Men (M), $N = 20$			Women $(\neg M)$, $N = 32$		
	Success (S)	Failure $(\neg S)$	Success Rate	Success	Failure	Success Rate	Success	Failure	Success Rate
Treatment (T)	20	20	50%	8	5	$\approx 61\%$	12	15	$\approx 44\%$
Control $(\neg T)$	6	6	50%	4	3	$\approx 57\%$	2	3	$\approx 40\%$

- P(Success) = 26/52 = .5
- P(Success| Treatment) = 20/40 = .5
- P(Success| No treatment) = 6/12 = .5
- P(Success | Man, Treatment) = 8/13 = 61%
- P (Success | Man, No Treatment) = 4/7 = 57%
- P(Success | Woman, Treatment) = 12/27 = 44%
- P (Success | Woman, No treatment) = 2/3 = 40%

Simpson’s paradox

	Full Population, N = 52			Men (M), N = 20			Women (¬M), N = 32		
	Success (S)	Failure (¬S)	Success Rate	Success	Failure	Success Rate	Success	Failure	Success Rate
Treatment (T)	20	20	50%	8	5	≈ 61%	12	15	≈ 44%
Control (¬T)	6	6	50%	4	3	≈ 57%	2	3	≈ 40%

- Very strange! Unconditional probability of success is 50%.
- Conditional probability only on treatment or no treatment sees no difference between treatment and no treatment
- Once we condition on gender, Treatment is better for both genders!
- Probabilities can appear misleading. Look at the raw numbers

Simpson's paradox – simple example

- Over the course of 2 baseball seasons, 2 batters had the following outcomes

- Season 1:
  - Player A: 4 hits on 10 at bats — P(hit) = .4
  - Player B: 350 hits on 1000 at bats — P(hit) = .35

- Season 2:
  - Player A: 300 hits on 1000 at bats — P(hit) = .3
  - Player B: 2 hits on 10 at bats — P(hit) = .2

- Combined over both seasons
  - Player A: 304 hits on 1010 at bats
  - Player B: 357 hits on 1010 at bats

Bayes rule in estimation

- Let's say we have 99 coins which have different probabilities of landing heads: 1/100, 2/100, … , 99/100. Let's say we randomly select one coin and flip it. What's the probability of getting a heads?
  - To get the unconditional probability, we can see it as a sum of conditional probabilities
  - $P(A) = P(A|C_1)P(C_1) + P(A|C_2)P(C_2) + \ldots + P(A|C_n)P(C_n)$ -- Law of total probability

- Now let's say we have picked a single coin, and observed several k flips already, where j flips landed heads. Can we update the probability of getting a heads on the next flip?
  - We can explicitly calculate the probability of having chosen any given coin by seeing how likely it is that we have observed j heads in k flips
  - For the next flip, the new probability can be expressed as

  $P(\text{next flip is heads}) = P(A|C_1)P(C_1|1^{st}\text{ k flips}) + P(A|C_2)P(C_2|1^{st}\text{ k flips}) \ldots + P(A|C_n)P(C_n|1^{st}\text{ k flips})$

  Where $P(C_i|1\text{st k flips}) = P(1^{st}\text{ k flips }|C_i)\, P(C_i) / P(1^{st}\text{ k flips})$ by Bayes Law.

- Note that we will weight those coins with probabilities closer to j/k higher than those with very different probabilities

Bayes rule in estimation

- There is some latent variable, over which we have some prior distribution
  - The latent variable is which coin. The probability here is uniform over 99 different coins
- The probability of the event depends on the value of the latent variable, but we don't know the value of the latent variable
- However, we can use the observed results to update our probabilistic beliefs about which latent variable we have selected
- Once we have updated our belief about the latent variable, we can recalculate the probability of the event using the law of total probability
- Real example: You meet a new person and don't know if they'll like your jokes. At first, you don't know if they're serious, friendly, mean, stupid, etc. However, after a few attempts, you can get a sense of whether they will like your jokes.

Advanced topic: Topic modeling (Natural Language Processing)

- If we have *M* different documents on *K* different topics, where document *m* contains N_m words, and we want to cluster the document to be about the appropriate topic or topics.
- General modeling approach:
  - Prior on whether some topic will be covered in a document
    - More common topics should be given more weight than arcane ones
    - This prior can vary depending on the document
  - Prior on whether some word relates to some topic, hopefully should only relate to few topics
    - “Finance” has to do with money so weight should be concentrated there.
    - “Bank” can be a financial institution or a river bank, so perhaps the weight should be split between topics
- Update the topics in each document based on the observed words, and update the weight of the words in each topic based on whether they appear in documents about that topic. Iterate

Example leading to odds

- At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose, however, that a new piece of evidence which shows that the criminal has a certain characteristic (such as lefthandedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?

Example leading to odds

- At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose, however, that a new piece of evidence which shows that the criminal has a certain characteristic (such as lefthandedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?

$$\begin{aligned} P(G|C) &= \frac{P(GC)}{P(C)} \\ &= \frac{P(C|G)P(G)}{P(C|G)P(G) + P(C|G^c)P(G^c)} \\ &= \frac{1(.6)}{1(.6) + (.2)(.4)} \\ &\approx .882 \end{aligned}$$

Odds

- The odds of event A are given by

$$\frac{P(A)}{P(A^C)} = \frac{P(A)}{1 - P(A)}$$

- How much more likely is P(A) to occur than to not occur?
- If P(A) = .8, the odds are 4 to 1
- In terms of modeling binary outcomes, it’s useful to know odds since, after exponentiation, they don’t suffer from boundary effects

Updating odds

- If event A has probability P(A), the odds are P(A)/ (1-P(A))

- Now what if we can condition on a second event we know has occurred. How do the odds change?
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ $\quad$ $P(A^C|B) = \frac{P(B|A^C)P(A^C)}{P(B)}$

- Note both these equations have P(B) in the denominator, since we're conditioning on the same new information

- New odds of A = $\frac{P(B|A)P(A)}{P(B|A^C)P(A^C)} = \frac{P(B|A)}{P(B|A^C)} \frac{P(A)}{P(A^C)}$
  - Can be seen as the original odds multiplied by the ratio of the conditional probabilities

Example

- An urn contains two type A coins and one type B coin. When a type A coin is flipped, it comes up heads with probability 1/4, whereas when a type B coin is flipped, it comes up heads with probability 3/4. A coin is randomly chosen from the urn and flipped. Given that the flip landed on heads, what are the odds that it was a type A coin? What is the probability it was a type A coin?

Example

- An urn contains two type A coins and one type B coin. When a type A coin is flipped, it comes up heads with probability 1/4, whereas when a type B coin is flipped, it comes up heads with probability 3/4. A coin is randomly chosen from the urn and flipped. Given that the flip landed on heads, what are the odds that it was a type A coin? What is the probability it was a type A coin?

Use the formula we just learned! $\frac{P(B|A)}{P(B|A^C)} \frac{P(A)}{P(A^C)}$

Odds of A unconditionally: $\frac{2/3}{1/3}$ = 2

Ratio of conditional probabilities: $\frac{1/4}{3/4} = \frac{1}{3}$

New odds: 2/3

New probability: if odds = P/(1-p) then P = odds/(1+odds) = $\frac{2/3}{5/3} = \frac{2}{5}$

Law of total probability

- Let $B_1, B_2, \ldots, B_n$ be mutually exclusive and collectively exhaustive events. That is
  - $B_i \cap B_j = \emptyset$ for $i \neq j$ and
  - $\bigcup_{i=1}^{n} B_i = S$

- Then for event A
  - $A = \bigcup_{i=1}^{n} A \cap B_i$
  - $P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$

- P(A) is equal to a weighted average of $P(A|B_i)$, each term being weighted by the probability of the event on which it is conditioned

Independence

- Two events A and B are said to be independent if

P(A and B) = P(A) P(B)

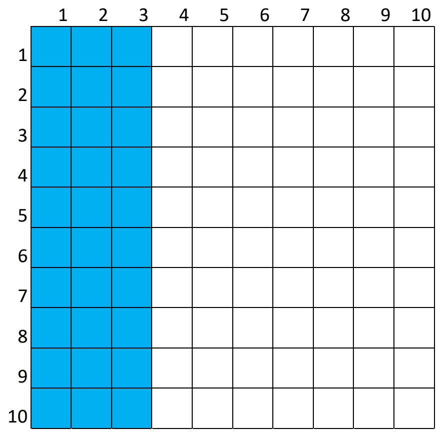
- Practically, this also means that
  - P(A|B) = P(A and B) / P(B) = P(A)P(B)/P(B) = P(A)
  - P(B|A) = P(A and B) / P(A) = P(A)P(B)/P(A) = P(B)

- Knowledge of whether or not A has occurred does not change the probability of whether B has occurred and vice versa

- This is not the same as disjointness in general
  - When the events are disjoint, usually knowing one of the events occurred is extremely informative
  - Disjointness is the same as independence when P(A) = 0 or P(B) = 0, or both = 0

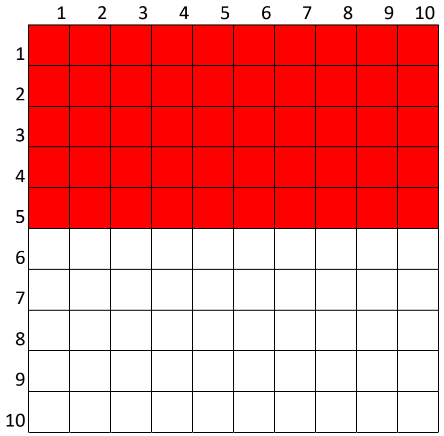
Geometric intuitive example

Event A: Box in the first 3 columns

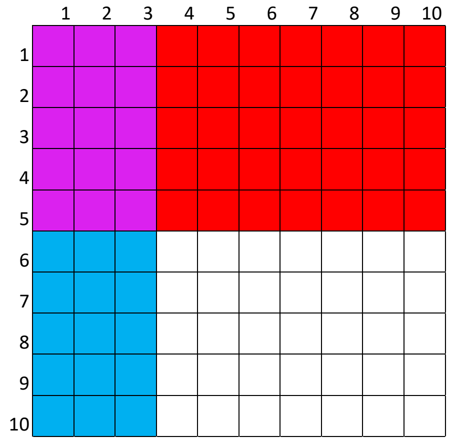


What is P(A)?

Event B: Box in the first 5 rows



What is P(B)?



$P(A \cap B)$ = **Purple**
$P(A \cap B^C)$ = **Blue**
$P(B \cap A^C)$ = **Red**
$P(A^C \cap B^C)$ = **White**

What is P(A and B)?
What is P(A or B)?

What is P(B|A)?
How does it compare to $P(B|A^C)$?
Try the same using (A|B)

Disjoint vs Independence Examples

- Independence: P(A and B) = P(A) * P(B)
- Disjoint: P(A and B) = 0

Example 1

- P(A) = .4
- P(B) = .8
- P(A and B) = .35

Are A and B disjoint?

Are A and B independent?

Example 2

- P(A) = .6
- P(B) = .9
- P(A and B) = .54

Are A and B disjoint?

Are A and B independent?

Example 3

- P(A) = .6
- P(B) = 0
- P(A and B) = ??

Are A and B disjoint?

Are A and B independent?

Disjoint vs Independence Examples

- Independence: P(A and B) = P(A) * P(B)
- Disjoint: P(A and B) = 0

Example 1

- P(A) = .2
- P(B) = .1

What would P(A and B) need to be for A and B to be disjoint?

What would P(A and B) need to be for A and B to be independent?

Example 2

- P(A) = .75
- P(B) = .75

What would P(A and B) need to be for A and B to be independent?

What would P(A and B) need to be for A and B to be disjoint? Is this possible?

Example 4C

- Suppose that we toss 2 fair dice. Let E denote the event that the sum of the dice is 6, and F denote the event that the first die equals 4. Are these events independent?

- Suppose that we toss 2 fair dice. Let G denote the event that the sum of the dice is 7, and F denote the event that the first die equals 4. Are these events independent?

Independence and complements

- If E and F are independent, then so are E and F^c.

Proof:

Assume that E and F are independent. Recall $E = EF \cup EF^c$

EF and EF^c are obviously mutually exclusive, so $P(E) = P(EF) + P(EF^c)$

$\rightarrow P(E) = P(E)P(F) + P(EF^c)$

$P(E) - P(E)P(F) = P(EF^c)$

$P(E)*(1 - P(F)) = P(EF^c)$

$P(E)*P(F^c) = P(EF^c)$

By the same logic, E^c is independent of both F and F^c

Independence of more than 2 events

Definition

Three events E, F, and G are said to be independent if

$$P(EFG) = P(E)P(F)P(G)$$
$$P(EF) = P(E)P(F)$$
$$P(EG) = P(E)P(G)$$
$$P(FG) = P(F)P(G)$$

- The events E1, E2, … , En are said to be independent if :

for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$ of these events, $r \le n$ $P(E_{1'}E_{2'} \cdots E_{r'}) = P(E_{1'})P(E_{2'}) \cdots P(E_{r'})$

- An infinite set of events is independent if every finite subset of those events is independent

Counterexample

- Two fair dice are thrown. Let E denote the event that the sum of the dice is 7. Let F denote the event that the first die equals 4 and G denote the event that the second die equals 3.

- P(E) = P({(1, 6), (2, 5), (3, 4),(4, 3), (5, 2), (6, 1)}) = 6/36 = 1/6
- P(F) = 1/6
- P(G) = 1/6
- P(E and F) = P({(4, 3)}) = 1/36
- P(E and G) = P({(4, 3)}) = 1/36
- P(F and G) = P({(4, 3)}) = 1/36
- While the pairwise probabilities all fit the independence criteria,

P(E and F and G) = P({(4, 3)}) = 1/36 ≠ P(E)P(F)P(G) so E F and G aren't all independent.

- This fits the intuitive explanation that if we know any 2 events have occurred, we're certain the 3^{rd} has, too

Independence over events formed from independent events

- If E, F, and G are independent, then E will be independent of any event formed from F and G. For instance, E is independent of $F \cup G$, since

$P[E(F \cup G)] = P(EF \cup EG)$ by distributivity

$= P(EF) + P(EG) - P(EFG)$ by the addition rule

$= P(E)P(F) + P(E)P(G) - P(E)P(FG)$ by independence

$= P(E)[P(F) + P(G) - P(FG)] = P(E)P(F \cup G)$ by the addition rule

Graph coloring – Probabilistic reasoning without finding an explicit solution

- The complete graph having n vertices is defined to be a set of n points (called vertices) in the plane and the $\binom{n}{2}$ lines (called edges) connecting each pair of vertices. Suppose now that each edge in a complete graph having n vertices is to be colored either red or blue. For a fixed integer k, a question of interest is, Is there a way of coloring the edges so that no set of k vertices has all of its $\binom{k}{2}$ connecting edges the same color?

Gambler's ruin

- Two gamblers, A and B, bet on the outcomes of successive flips of a coin. On each flip, if the coin comes up heads, A collects 1 unit from B, whereas if it comes up tails, A pays 1 unit to B. They continue to do this until one of them runs out of money. If it is assumed that the successive flips of the coin are independent and each flip results in a head with probability p, what is the probability that A ends up with all the money if he starts with i units and B starts with N – i units?

Sequential conditioning

- Suppose there are n mutually exclusive and exhaustive possible hypotheses, with initial (sometimes referred to as prior) probabilities $P(H_i)$ such that $\sum_{i=1}^{n} P(H_i) = 1$. Now, if information that the event E has occurred is received, then the conditional probability that H_i is the true hypothesis (sometimes referred to as the updated or posterior probability of H_i) is

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}$$

- Suppose now that we learn first that E_1 has occurred and then that E_2 has occurred. Then, given both pieces of information, the conditional probability that H_i is the true hypothesis is

$$P(H_i|E_1E_2) = \frac{P(E_1E_2|H_i)P(H_i)}{\sum_j P(E_1E_2|H_j)P(H_j)}$$

- Can we compute $P(H_i \mid E_1 E_2)$ by using the right side of the first equation with $E = E_2$ and with $P(H_j)$ replaced by $P(H_j| E_1)$, j = 1, … , n. That is, when is it legitimate to regard $P(H_j \mid E_1)$ as the prior probabilities and then use the first equation to compute the posterior probabilities?

Only legitimate when E_1 and E_2 are conditionally independent given H_j for j = 1,2, … , n

$$P(E_1E_2|H_j) = P(E_2|H_j)P(E_1|H_j), \quad j = 1,\ldots,n$$

$$\begin{aligned} P(H_i|E_1E_2) &= \frac{P(E_2|H_i)P(E_1|H_i)P(H_i)}{P(E_1E_2)} \\ &= \frac{P(E_2|H_i)P(E_1H_i)}{P(E_1E_2)} \\ &= \frac{P(E_2|H_i)P(H_i|E_1)P(E_1)}{P(E_1E_2)} \\ &= \frac{P(E_2|H_i)P(H_i|E_1)}{Q(1,2)} \end{aligned}$$

$$Q(1,2) = \frac{P(E_1E_2)}{P(E_1)}$$

$$1 = \sum_{i=1}^{n} P(H_i|E_1E_2) = \sum_{i=1}^{n} \frac{P(E_2|H_i)P(H_i|E_1)}{Q(1,2)}$$

$$Q(1,2) = \sum_{i=1}^{n} P(E_2|H_i)P(H_i|E_1)$$

$$P(H_i|E_1E_2) = \frac{P(E_2|H_i)P(H_i|E_1)}{\sum_{i=1}^{n} P(E_2|H_i)P(H_i|E_1)}$$

Upshot: After observing one event (e.g. the result of one experiment), we can treat our posterior probabilities as our new prior probabilities, and just update those based on new results!

Homework 3

- Due 10/4
- Problems from textbook Chapter 3:
  - 3.38
  - 3.47
  - 3.90
- Theoretical problems from textbook Chapter 3:
  - 3.8
  - 3.9
  - 3.10