Bayes Rule
Independence

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 2.2, 2.3

Aaron Osgood-Zimmerman

Department of Statistics

Practice solution

Practice

Randomly pair 4 keys {a, b, c, d} with 3 locks {a, b, c}.
What is $\mathbb{P}$(at least one match)?

Let A denote the event that lock A is not matched with the a key. Similarly, B, C. Let D denote the event that key d is paired with any of the locks. Solve by calculating $1 - \mathbb{P}$(no keys match).

- $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B \cap C \cap D) + \mathbb{P}(A \cap B \cap C \cap D^c)$
- $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B \cap C|D) * \mathbb{P}(D) + \mathbb{P}(A \cap B \cap C|D^c) * \mathbb{P}(D^c)$
- Starting with the second term, if d key is not selected, then we only have keys a, b, c to permute among locks A, B, C without getting a match. There are only two ways to do this. Why?
- So, $\mathbb{P}(A \cap B \cap C|D^c) * \mathbb{P}(D^c) = \frac{2}{3*2*1} \times \frac{3*2*1}{4*3*2*1} = 2/24$
- As for the first term, if d is selected, then for each combination of sets of 3 keys, there are 3 ways to arrange them without a match. Why?
- So, $\mathbb{P}(A \cap B \cap C|D) * \mathbb{P}(D) = \frac{\binom{3}{2}*3}{\binom{3}{2}*3!} \times (1 - \frac{3*2*1}{4*3*2*1}) = 9/24$

$\mathbb{P}(\text{at least 1 match}) = 1 - \mathbb{P}(\text{no match}) = 1 - (9/24 + 2/24) = 13/24$

Practice solution

Practice (Example from a Wall Street job interview)

Let us play a Russian roulette. Here's the barrel of the gun, six chambers, all empty. Now watch me as I put two bullets into the barrel, into two adjacent chambers. I close the barrel and spin it. I put a gun to your head and pull the trigger.... Click. Lucky you!

Now I'm going to pull the trigger *one more time*. Which would you prefer: that I spin the barrel first or that I just pull the trigger?

Solution

- Let $A = \{\text{survival on the first try}\}$, $B = \{\text{survival on the second try}\}$.
- If you spin,
$$\mathbb{P}(B \mid A) = \mathbb{P}(B) = 4/6 = 2/3.$$
- If you do not spin,
$$\mathbb{P}(A \cap B) = 3/6 = 1/2$$
Why? There are 6 choices for two consecutive slots and 3 choices for two consecutive empty slots.
- Hence
$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{1/2}{2/3} = 3/4 > 2/3,$$
so do not spin!

Recap

Conditional Probabilities

- **Definition**
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{for } \mathbb{P}(B) > 0$$
- **Factorization rule**[1]
  - $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$
  - $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(C \mid A, B)\mathbb{P}(B \mid A)\mathbb{P}(A)$
  - $\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1, A_2) \cdots \mathbb{P}(A_n \mid A_1, \ldots, A_{n-1})$
- **Law of total probability**[1]
  - $\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c)$
  - $\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$ for $B_1, \ldots, B_n$ a partition of Ω

[1] Provided that the conditional probabilities are well defined, i.e., that the conditioning event has a non-zero prob.

Outline

Bayes' Formula

- Sometimes we might know one conditional probability, but we may want to learn something about the events if the *conditioning was reversed*
- For example, if a quick response covid test has been created and the tested on known covid positive patients (eg via RNA tests), we might know that *given the disease, the test comes back positive with 90% probability.*
- But, if you go and get the test and it comes back positive, you might want instead want to know: *"what's the probability that I have covid if the test is positive?"*
- that is we might want to go from $P(T+|D+)$ to $P(D+|T+)$

Bayes' formula

Example

Two urns. Urn I has 2 green balls and 1 red ball. Urn II has 2 red balls and 3 yellow balls.

Consider

1. Picking one of the two urns with equal prob.
2. Then sample one ball uniformly at random from the selected urn

Given that we got a red ball, what is the prob. that we picked it from urn I?

Solution

$$\begin{aligned}\mathbb{P}(\text{urn I} \mid \text{red}) &= \frac{\mathbb{P}(\text{red} \cap \text{urn I})}{\mathbb{P}(\text{red})} \\ &= \frac{\mathbb{P}(\text{red} \mid \text{urn I})\mathbb{P}(\text{urn I})}{\mathbb{P}(\text{red} \mid \text{urn I})\mathbb{P}(\text{urn I}) + \mathbb{P}(\text{red} \mid \text{urn II})\mathbb{P}(\text{urn II})} \\ &= \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{1}{2}} = \frac{5}{11}\end{aligned}$$

Bayes Formula

Goal

Get $\mathbb{P}(B|A)$, knowing $\mathbb{P}(A|B)$, $\mathbb{P}(A|B^c)$ and $\mathbb{P}(B)$, $\mathbb{P}(B^c)$.

Proposition (Bayes' formula)

For A, B, B^c such that $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(B^c) > 0$,

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

Testing COVID-19

Exercise

Test T:

- Prob. of True positive: $P(T=+\mid \text{covid})=90\%$.
- Prob. of True negative : $P(T=-\mid \text{healthy})=95\%$.

Assume the prevalence of COVID-19 in the population is $\approx 2\%$. ($\mathbb{P}(\text{covid})=2\%$)

- What is $P(\text{covid}\mid T=+)$ for someone at random in the population?
- Now consider the test is done on people that have symptoms and we know that for these people, $\mathbb{P}(\text{covid})=50\%$, what is $P(\text{covid}\mid T=+)$?

Solution

$$\begin{aligned}P(\text{covid}\mid +) &= \frac{P(+\mid \text{covid})P(\text{covid})}{P(+\mid \text{covid})P(\text{covid})+P(+\mid \text{heathy})P(\text{healthy})}\\ &= \frac{90\%\times 2\%}{90\%\times 2\%+(1-95\%)(1-2\%)}\approx 26.9\%.\end{aligned}$$

Surprising! So we use two tests, or look at people with symptoms (by the same computations): $\mathbb{P}(\text{covid}|+,\text{symptomatic})=94.7\%$

Bayes' formula

Using the law of total prob., Bayes' formula can be generalized to partitions.

Proposition (Bayes' formula)

Let $B_1, \dots, B_n$ be a partition of Ω with $\mathbb{P}(B_i) > 0$.

For any event A, with $\mathbb{P}(A) > 0$, and any $k \in \{1, \dots, n\}$

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}$$

Interpretation

- We have n possible hypothesis B_i that could lead to our observation A
- We know how A would be generated if B_i was true (we know $\mathbb{P}(A \mid B_i)$)
- We also have some **prior** prob. of each hypothesis ($\mathbb{P}(B_i)$)

Bayes formula gives the **posterior** prob. of the hypothesis
after having observed A, i.e. $\mathbb{P}(B_i \mid A)$

- This is the basis of the Bayesian viewpoint on probability:
  We start from some prior belief on events $\mathbb{P}(B_i)$,
  we update the beliefs using some observations A using Bayes formula

Bayes' formula

Exercise

3 coins: fair (F), moderately biased (M), heavily biased (H) s.t.,

$\mathbb{P}(\text{tail} \mid F) = 1/2 \quad \mathbb{P}(\text{tail} \mid M) = 3/5 \quad \mathbb{P}(\text{tail} \mid H) = 9/10$

You take a coin at random. The prob. of its type are *(prior probabilities)*

$\mathbb{P}(F) = 90/100 \quad \mathbb{P}(M) = 9/100 \quad \mathbb{P}(H) = 1/100$

You flip it 3 times and get 3 tails. What are the posterior probs?

Solution

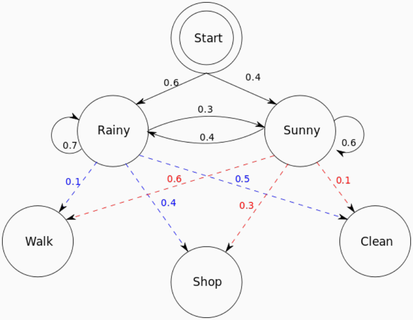
- $$\begin{aligned}\mathbb{P}(F \mid TTT) &= \frac{\mathbb{P}(TTT \mid F)\mathbb{P}(F)}{\mathbb{P}(TTT \mid F)\mathbb{P}(F) + \mathbb{P}(TTT \mid M)\mathbb{P}(M) + \mathbb{P}(TTT \mid H)\mathbb{P}(H)} \\ &= \frac{\frac{1}{2}^3 \cdot \frac{90}{100}}{\frac{1}{2}^3 \cdot \frac{90}{100} + \frac{3}{5}^3 \cdot \frac{9}{100} + \frac{9}{10}^3 \cdot \frac{1}{100}} \approx 0.81\end{aligned}$$
- $\mathbb{P}(M \mid TTT) \approx 0.14$, $\mathbb{P}(H \mid TTT) \approx 0.05$

Hidden Markov Models

The last exercise is an example of a Hidden Markov Model (HMM):

A statistical model where the system being modeled is assumed to be a Markov process, denoted X – with unobservable ("hidden") states. A HMM assumes that there is another observable process Y which depends on X. The goal is to learn about X by observing Y.

Two friends talk on the phone every day and play a game, guessing the weather based on what activity their friend is performing during the call

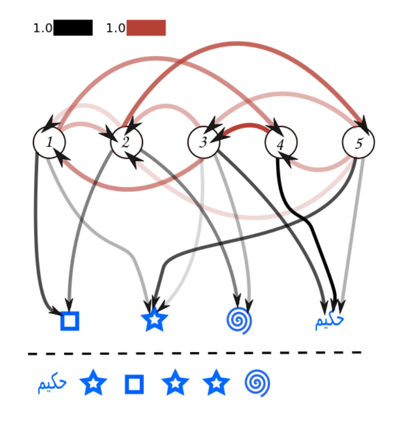


example taken from wiki

Hidden Markov Models

Given the observed sequence of events, we may be interested in the most likely sequence of states that could have produced it. Based on the arrows that are present in the diagram, the following state sequences are candidates:

5 3 2 5 3 2
4 3 2 5 3 2
3 1 2 5 3 2



We can find the most likely sequence by evaluating the joint probability of both the state sequence and the observations for each case (simply by multiplying the probability values, which here correspond to the opacities of the arrows involved).

example taken from wiki

Outline

Bayes formula

Independence

Independence

Motivation

- Intuitively, we would say that A is *independent* of B is the conditional prob. of A given B is the same as the unconditional prob. of A, i.e., $\mathbb{P}(A \mid B) = \mathbb{P}(A)$
- In that case $\mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A)$, that is, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Definition

Two events A and B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Example

You pick a point $z = (x, y)$ uniformly at random on the square $[0, 1] \times [0, 1]$

Let $A = \{$the x coordinate of the point z is btw 0.4 and 0.6$\}$

and $B = \{$the y coordinate of the point z is btw 0.7 and 0.8$\}$

Are A and B independent?

Solution

- Intuitively, knowing the x-coord does not change the prob. of the y-coord.
- Intuition is good but can be misleading... better to check
- We have $A \cap B = \{$the point z=(x,y) satisfies $0.4 \le x \le 0.6$ and $0.7 \le y \le 0.8\}$,

$$P(A \cap B) = \frac{|[0.4, 0.6] \times [0.7, 0.8]|}{|[0, 1] \times [0, 1]|} = \frac{0.2 \cdot 0.1}{1 \cdot 1} = 0.02$$

- We have

$$\mathbb{P}(A) = \frac{|[0.4, 0.6] \times [0, 1]|}{|[0, 1] \times [0, 1]|} = 0.2, \quad \mathbb{P}(B) = \frac{|[0, 1] \times [0.7, 0.8]|}{|[0, 1] \times [0, 1]|} = 0.1,$$

  so $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B)$ therefore the two events are independent
- This example gives you some kind of representation of independence: two events are independent if the surface of their intersection can be computed as a product of the surfaces of the events
- Overall independence simplifies the computation of probabilities, when it holds

Independence

Example
Consider picking a point $z = (x, y)$ uniformly at random on the triangle $\{(x, y) : x \geq 0, y \geq 0, x + y \leq 1\}$

Let $A = \{$the x coordinate of the point z is btw 0.4 and 0.6$\}$

and $B = \{$the y coordinate of the point z is btw 0.7 and 0.8$\}$

Are A and B independent?

Solution

- Here intuitively the answer will be no: the possible values of y depend on the x
- Rigorously we have that $A \cap B = \{(x, y) : 0.4 \leq x \leq 0.6, 0.7 \leq y \leq 0.8\}$, so for all $z = (x, y) \in A \cap B$, $x + y > 1$ so
$$\mathbb{P}(A \cap B) = 0$$
- On the other hand $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$.
- So $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$, the events are not independent.

Practice: repeat but with $A = \{(x, y) : x \leq 0.5\}$ and $B = \{(x, y) : y \leq 0.5\}$, you'll need to compute the respective areas of A, B, $A \cap B$, you will find they are dependent.

Independence

Be careful

Do not confound **disjoint events** and **independent events**.

For example A and A^c are disjoint,

but $0 = \mathbb{P}(A \cap A^c) \neq \mathbb{P}(A)\mathbb{P}(A^c) > 0$ for any A s.t. $0 < \mathbb{P}(A) < 1$.

Practice next lecture

Practice

A firm wants to know if can lower the prescribed 40hr work week and maintain the same productivity level from its employees.

- Realizing that asking if employees work less than the mandated 40hrs and the sensitivity of this issue, the personnel director decides to use a *randomized response survey*.
- Each employee is asked to flip a fair coin,
  - If head (H), answer the question "Do you carpool to work?"
  - If tail (T), answer the question "Have you worked less than 40hr weeks in the last month?"
- Out of 8000 responses, 4820 answered "YES" (assuming honesty)
- The company knows that 35% of its employees carpool to work.
- What is the probability that an employee (chosen at random) worked less than a 40hr week in the last month?