

# Conditional Probability



# Agenda

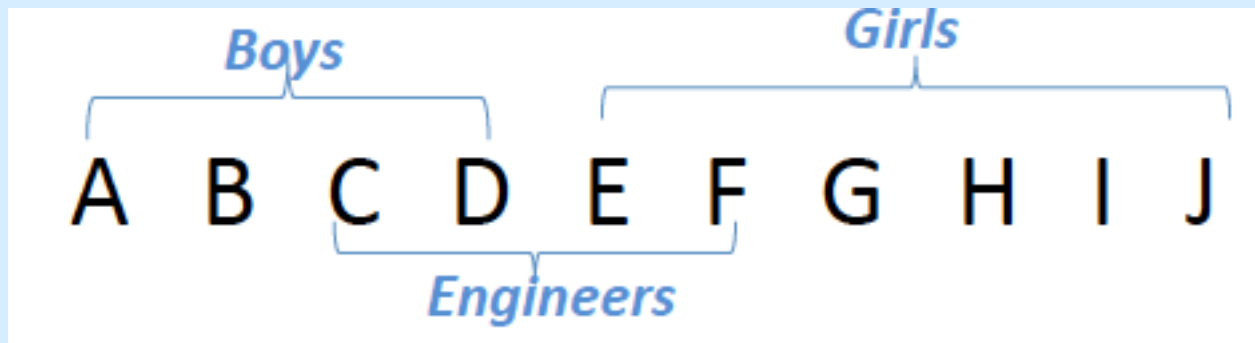


- Conditional Probability
- Multiplication Rule
- Independent and Dependent Events
- Bayes' Rule

# Job Interview



- For a job at a firm, 10 applicants appear for interview. We have some information about them:



- What is the sample space?
  - $S = \{A, B, C, D, E, F, G, H, I, J\}$
  - Let probability of each candidate getting job is  $p_1, p_2, p_3, \dots, p_{10}$



- Event B: Boy is selected
  - $B = \{A, B, C, D\}$  &  $P(B) = p_1 + p_2 + p_3 + p_4$
- Event G: Girl is selected
  - $G = \{E, F, G, H, I, J\}$  &  $P(G) = p_5 + p_6 + p_7 + p_8 + p_9 + p_{10}$
- Event E: Engineer is selected
  - $E = \{C, D, E, F\}$  &  $P(E) = p_3 + p_4 + p_5 + p_6$
- Selected candidate will be announced after a week. After 2 days, candidate start calling the firm. Firm spokesperson says that she cannot reveal much, but only thing she knows is that selected person is an engineer.
- Given this information, what is prob that a girl gets the job.

# Conditional Prob Defined

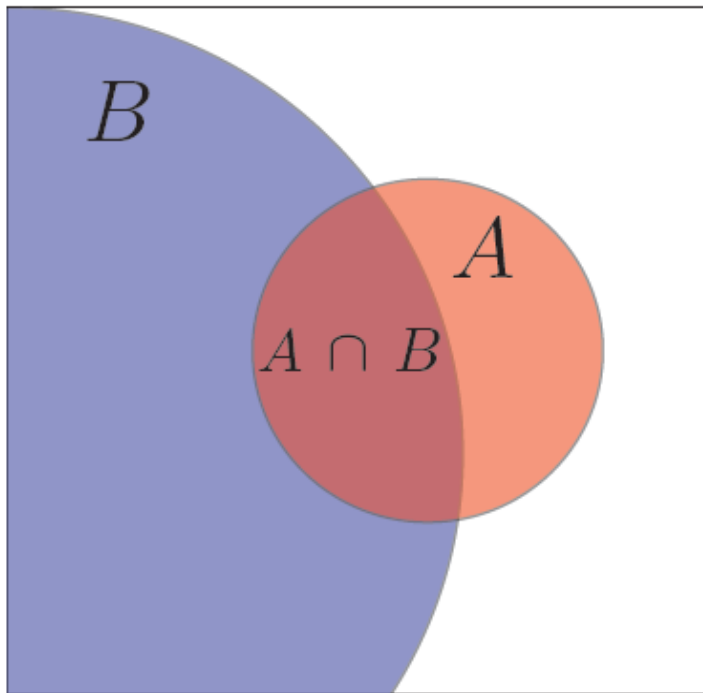


- New sample space is now  $S_1 = \{C,D,E,F\}$
- Prob that we need now is that a girl is selected given an engineer got the job i.e.  $P(G \text{ given } E \text{ occurred})$ 
  - This is called conditional probability
  - Denoted by  $P(G|E)$
  - $P(G|E) = (p_5/p_3+p_4+p_5+p_6) + (p_6/p_3+p_4+p_5+p_6) = (p_5+p_6)/(p_3+p_4+p_5+p_6)$
  - $p_5+p_6$  can also be written as  $P(G \cap E)$ ; Event  $G \cap E = \{E,F\}$
  - Event  $E = \{C,D,E,F\} \Rightarrow P(E)=p_3+p_4+p_5+p_6$
  - Thus,  $P(G|E) = P(G \cap E)/P(E)$  (given that  $P(E) > 0$ )

# Conditional Probability



- Prob of A given B
- $P(A|B) = P(A \cap B)/P(B)$  (given that  $P(B) > 0$ )



$A = A \cap B$		$B$			
↓		↓			
HHH	HHT	THH	THT		
HTH	HTT	TTH	TTT		

# One More Example



- Consumer approaches a bank for loan. Historical data suggests that default rate is 4%. Bank will decide on application in a week. Research department of bank mentions that default rates are different during boom economy and bust economy.
  - D be event that consumer defaults.  $P(D) = 0.04$
  - L be event that economy is bust.  $P(L) = 0.07$
  - Also, let  $P(D \cap L) = 0.03$
  - Thus,  $P(D|L) = P(D \cap L) / P(L) = 0.03/0.07 = 0.43$  !!!!

# Multiplication Rule



- We know that  $P(A|B) = P(A \cap B)/P(B)$
- Multiplication rule in probability states that
- $P(A \cap B) = P(A|B) P(B)$
- Also,  $P(A \cap B) = P(B|A) P(A)$

A consulting firm is bidding for two jobs, one with each of two large multinational corporations. The company executives estimate that the probability of obtaining the consulting job with firm A, event A, is 0.45. The executives also feel that if the company should get the job with firm A, then there is a 0.90 probability that firm B will also give the company the consulting job. What are the company's chances of getting *both* jobs?

We are given  $P(A) = 0.45$ . We also know that  $P(B | A) = 0.90$ , and we are looking for  $P(A \cap B)$ , which is the probability that both A and B will occur. From the equation we have  $P(A \cap B) = P(B | A)P(A) = 0.90 \times 0.45 = 0.405$ .

- Example taken from Complete Business Statistics book

# Independent Events



- Does knowing about A tell us anything about whether B happens or not
- If A does not have impact on B, then A and B are independent events.
- Assume default rate does not change with economy state i.e.  $P(D|L) = P(D)$ 
  - Default is independent from economy event
- Formally, A and B are independent if
- $P(A|B) = P(A)$
- $P(A \cap B) = P(A) P(B)$

# Independence vs Disjoint

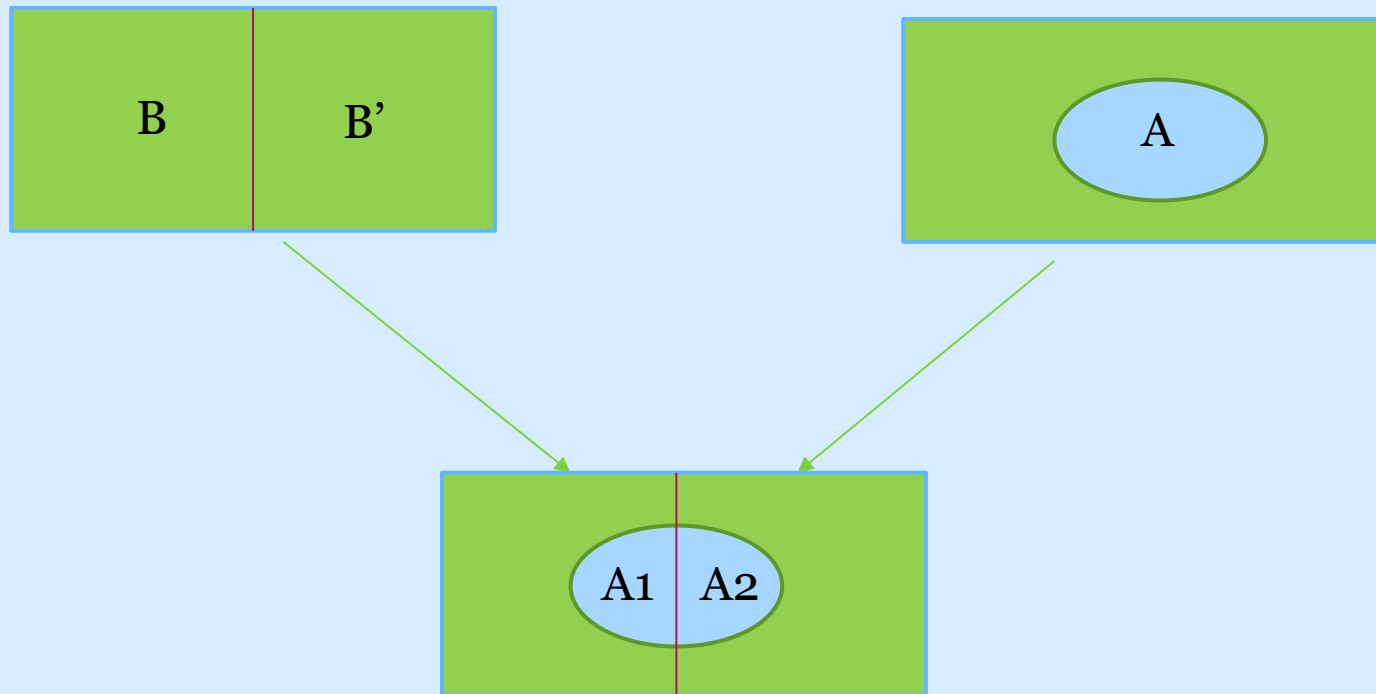


- Do not confuse independent events from disjoint events
- A and B are independent if  $P(A|B) = P(A)$
- A and B are disjoint if  $P(A \cap B) = 0$
- Consider card deck
  - C: card is a club
  - R: card is red
  - Independence  $P(R|C) = P(R)$ .
    - ✦  $P(R|C) = 0$   $P(R) = 1/2$  Thus, not independent
  - Disjoint  $P(C \cap R) = 0$  Thus, disjoint

# Law of total probability



- Consider sample space  $S$





- Now  $P(A) = P(A_1) + P(A_2)$
- We can say  $A_1 = A \cap B$
- $A_2 = A \cap B'$
- Thus,  $P(A) = P(A \cap B) + P(A \cap B')$
- Also,  $P(A \cap B) = P(A|B) P(B)$
- Thus,  $P(A) = P(A|B) P(B) + P(A|B') P(B')$
- This is law of total probability

# Example



- Imagine that two companies ( $X$  and  $Y$ ) supply fans.  $X$ 's fans work for over 2000 hours in 99% of cases, whereas  $Y$ 's fans work for over 2000 hours in 95% of cases. We know that  $X$  supplies 70% of the total fans. What is the probability that a fan I purchase will work for longer than 2000 hours?
- We want to find  $P(A)$  where  $A$  is event that fan works longer than 2000 hours.
- This can be solved using law of total probability



- $P(A) = P(A|B) P(B) + P(A|B') P(B')$ 
  - $P(A)$  = Prob that fan works more than 2000 hours
  - $P(B)$  = Prob that X sold the fan =  $7/10 = .7$
  - $P(B')$  = Prob that Y sold the fan =  $3/10 = .3$
  - $P(A|B)$  = Prob that fan will work more than 2000 hours given that it is sold by X =  $99/100$
  - $P(A|B')$  = Prob that fan will work more than 2000 hours given that it is sold by Y =  $95/100$
  - Putting everything together
  - $P(A) = .99*.7 + .95*.3 = .693 + .285 = .978$



# Bayes' Rule

*Probably the most important rule in Machine Learning*

# Bayes' Rule



- $P(A)$ : probability of event A occurring
- $P(A|B)$ : probability of A occurring given B occurred
- $P(B|A)$ : probability of B occurring given A occurred
- $P(A \cap B)$ : probability of **A and B** occurring simultaneously (joint probability of A and B)

Joint probability of A and B

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

- Bayes' rule allows to find  $P(A|B)$  from  $P(B|A)$ , i.e. to 'invert' conditional probabilities.

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

- Also,  $P(A) = P(A|B)*P(B) + P(A|B')*P(B')$

- Thus,  $P(B|A) = \frac{P(A|B)*P(B)}{P(A|B)*P(B) + P(A|B')*P(B')}$

# Example



- Imagine that you are interested in knowing if your friend's health is at risk given that he/she has an unhealthy lifestyle. Being a brilliant Data Scientist, you want to approach the problem scientifically (no emotions involved).
- Let A be the event that a person's health is at risk. Based on past data, you know that  $P(A) = 0.10$
- Let B be the event that a person has an unhealthy lifestyle. Based on past data, you know that  $P(B) = 0.05$
- Based on past data, you also know that for all those people with health problems, 7% have an unhealthy lifestyle. Thus,  $P(B|A) = 0.07$

# Example



- We are interested in knowing that given that your friend has an unhealthy lifestyle, what is the probability that their health is at risk.
- Essentially, we want to know  $P(A|B)$
- $P(A|B) = P(B|A) * P(A) / P(B)$
- $\Rightarrow (0.07 * 0.10) / 0.05$
- $\Rightarrow 0.14$

# Example



- A builder markets houses through brokers and online. Builder has data that 60% of prospective buyers who know about his houses know it through brokers, and 40% know through online website. He also knows that out of prospective buyers contacted through brokers, 25% actually end up buying a house; and out of those through who knew through online medium 20% end up buying. Which one is more effective medium for the builder (broker or online)?



- Let us setup the problem.
- B: Marketing done through broker
- O: Marketing done online
- H: House actually bought
- $P(B) = .6$ ,  $P(O) = .4$ ,  $P(H|B) = .25$ ,  $P(H|O) = .2$
- We need to know  $P(B|H)$  and  $P(O|H)$  to make decision
- We know that  $P(B|H) = P(B \cap H)/P(H)$ 
  - $\Rightarrow (P(H|B) P(B))/P(H) = .25 \cdot .6 / P(H) = .15 / P(H)$



- $P(H)$  is not given to us. But we know that
  - $P(H) = P(H|B)P(B) + P(H|O) P(O)$
  - $\Rightarrow .25*.6 + .2*.4 = .15 + .08 = .23$
- Thus,  $P(B|H) = .15/.23 = .66$
- Now, do the same for  $P(O|H)$
- $P(O|H) = (P(H|O) P(O))/P(H)$ 
  - $\Rightarrow .2*.4/.23 = .34$
- Thus, 66% of potential buyers marketed through brokers buy house, but only 34% buy who are marketed through online.
- Therefore brokers is more effective medium

# Example



- Let us consider Bayes' theorem one more time. Imagine that a person shows up at hospital for cancer test. What is the probability that the person has cancer if they test positive on an examination?
- 1% who participate in routine screening have cancer. 80% with cancer will get positive test results. 9.6% without cancer will also get positive results. A person had a positive result in a routine screening. What is the probability that the person actually has cancer?

# Example



- We can represent the situation in this manner

	Cancer (1%)	No Cancer (99%)
Test Positive	80%	9.6%
Test Negative	20%	90.4%

- 1% have cancer (99% do not)
- If a person has cancer, they are in first column. With 80% probability they will test positive. With 20% probability they will test negative.
- If a person does not have cancer, then they are in second column

# Example



- Now, if someone is tested positive, what are the chances they really have cancer?
- Let us represent it with this table

	Cancer (1%)	No Cancer (99%)
Test Positive	True Positive $1\% \times 80\% = 0.008$	False Positive $99\% \times 9.6\% = 0.095$
Test Negative	False Negative $1\% \times 20\% = 0.002$	True Negative $99\% \times 90.4\% = 0.894$

- Probability is likelihood of something happening i.e.  
desired event/all probabilities =  
 $0.008 / (0.008 + 0.095) = 7.8\%$

# Example



- We can also think of it in terms of Bayes' rule
- $P(A)$  = Prob of cancer = 1%
- $P(A')$  = 99%
- $P(B|A)$  = Prob of testing positive given cancer = 80%
- $P(B|A')$  = Prob of testing positive given not cancer = 9.6%
- $P(A|B)$  = Prob of cancer given positive test
- Thus,  $P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|A') * P(A')}$
- = 7.8%

# Naïve Bayes ML Algo



- One of the most important algorithms in machine learning is based on Bayes' rule
- It allows us to classify objects into one of several possible categories
- Suppose your friend meets you and says he had a great dinner with someone yesterday. You do not know who the person could be, so you would assign a 50% prob that person was a woman and 50% prob that person was a man.
- Probing further, your friend says the person had long hair. Now you would be more inclined to think the person was a woman (based on premise that women have longer hair).
- You have just updated your prob based on new information. This is the basic premise of Bayes' rule, and is one of the most powerful approaches.

# Terminology



- $P(A)$  is called **the *marginal* or *prior* probability of A** (since it is the probability of A *prior* to having any information about B)

Similarly:

- $P(B)$ : **the *marginal* or *prior* probability of B**
- $P(A|B)$  is called **the *likelihood* function for A given B**.
- $P(B|A)$ : **the *posterior* probability** of B given A (since it depends on having information about A)

## Bayes Rule

$$P(B|A) = P(A|B) * P(B) / P(A)$$

**prior** probabilities of B, A (“priors”)

“**posterior**” probability of B given A

“**likelihood**” function for B (for fixed A)



- The **prior** is the probability of the parameter and represents what was thought **before seeing the data**.
- The **likelihood** is the probability of the data given the parameter and represents **the data now available**.
- The **posterior** represents what is thought **given both prior information and the data just seen**.
- It relates to the conditional density of a parameter (**posterior probability**) with its unconditional density (**prior**, since depends on information present before the experiment).

# Naïve Bayes Formula



- Bayes Rule
  - $P(B|A) = P(A|B) * P(B) / P(A)$
- Naïve Bayes simplifies the formula to
  - $P(B|A) = P(A|B) * P(B)$  (For multiple classes of B. More on this later)
- Why?
  - We are interested in numerator only.
  - Denominator is constant – it does not depend on B
  - Naïve Bayes assumed conditional independence – all features are independent

# Let us build our first ML Algo



- Imagine 10 individuals and we know if they are cricketers or not

Person	Cricket er
1	Y
2	N
3	Y
4	Y
5	Y
6	N
7	N
8	Y
9	N
10	Y
11	???



- Given only that data, you would say prob of person 11 being cricketer is .6 (6/10).
  - This is prior probability, as we don't know anything more about person 11.
- Let us have some more data now

Smokes	Drinks	Runs	Cricketer
YES	YES	NO	YES
YES	NO	YES	NO
NO	YES	NO	YES
NO	NO	YES	YES
YES	YES	NO	YES
NO	NO	YES	NO
NO	YES	YES	NO
YES	YES	YES	YES
NO	NO	NO	NO
YES	YES	YES	YES
NO	YES	NO	



- Given additional info, we want to know posterior prob i.e. prob that person 11 is cricketer given certain characteristics about the person.
- For this, we need to compute conditional prob for each characteristic i.e.  $P(S|C)$  and  $P(S|C')$ . Similarly,  $P(D|C)$  and  $P(D|C')$ ; and  $P(R|C)$  and  $P(R|C')$
- Let us compute this using the data we have

		Smoke	Drink	Run
$P(X C)$	$C=Y(1)$	0.66	0.82	0.5
$P(X C')$	$C=N(0)$	0.25	0.25	0.75



- Let us also compute conditional prob for complements of each characteristic i.e.  $P(S'|C)$  and  $P(S'|C')$ . Similarly,  $P(D'|C)$  and  $P(D'|C')$ ; and  $P(R'|C)$  and  $P(R'|C')$
- It is simply  $1-P(X|C)$  and  $1-P(X|C')$

		Smoke	Drink	Run
$P(X' C)$	$C=Y(1)$	0.33	0.18	0.5
$P(X' C')$	$C=N(0)$	0.75	0.75	0.25



- Now, if we have to classify Person 11 as cricketer or not, let us see what are characteristics of person 11

Smokes	Drinks	Runs	Cricketer
NO	YES	NO	???

- Since we know characteristics and conditional probabilities, we need to compute  $P(C|\text{characteristics})$  and  $P(C'|\text{characteristics})$
- $P(C|\text{characteristics}) = (.33 * .82 * .5) * (0.6) = .081$
- $P(C'|\text{characteristics}) = (.75 * .25 * .25) * (0.4) = .018$
- Given the data,  $P(C|\text{characteristics}) > P(C'|\text{characteristics})$ 
  - => classify person 11 as Cricketer



- Congratulations!!! You just saw how Naïve Bayes Classifier algorithm works.
- It is one of the most popular algorithms in ML.
- We are already making use of our learnings.
- Let us implement this in R.
- How many lines of code do you expect to write this in
  - 100?
  - 200?
  - 500?