ECS509U - Probability & Matrices

Tassos Tombros

Week 3

Last week

- In Week 2 we covered:
  - Conditional probability
  - Event independence
  - Counting permutations
  - Counting combinations
  - Permutation/combination probability problems

Week 3: Learning Objectives

- **Review exercises from week 2**
- **Bayes Theorem**
- At the end of Week 3 you should be able to:
  - Understand the law of total probability
  - Apply the law to solving appropriate probability problems
  - Understand Bayes theorem
  - Apply Bayes theorem to solving probability problems

From last week

- **Permutations of *r* items chosen from *n* items: order matters, repetitions allowed:**

$$P(n,r) = n^r$$

- **Permutations of *r* items chosen from *n* items: order matters, no repetitions**

$$P(n,r) = \frac{n!}{(n-r)!}$$

- **Combinations of *r* items chosen from *n* items : order does not matter, no repetitions**

$$C(n,r) = \frac{n!}{(n-r)!r!}$$

- Note: **the relationship between P(*n*,*r*) and C(*n*,*r*)**
  - **C(n,r)=P(n,r)/r!**, because we are discounting (by dividing with *r*!) all the possible permutations of the *r* items we want to choose, since for C(*n*,*r*) we are not interested in the order in which the *r* items are selected

Permutations when some objects are the same

- In how many ways can you arrange the letters of the word "Mississippi"? (note that the arrangements do not have to make sense!!!)
  - 11!/(4!2!4!1!) = 34,650 ways
  - In the denominator we "discount" the possible number of permutations of the same letters: 4! for the 4 'i', 2! for the 2 p, 4! for the 4 s, and 1! for the 1 m
- In general, when we have ***n* the items being permuted, and n_1, n_2, …, n_k the number of each of the k types of objects being involved** (e.g. the sets of letters in the example), then the number of distinguishable arrangements is given by:

$$\frac{n!}{n_1!n_2!\ldots n_k!}, \text{where } n = n_1 + n_2 + \ldots + n_k$$

Probability problems

- Be aware of the difference
  - Some problems may ask you to calculate the **number of possible combinations or** of possible **permutations**
  - Whereas some other may ask you to calculate the **probability** that some specific **permutations or combinations** can occur
- Remember that normally the probability problems will require you to calculate a fraction:

$$p = \frac{\text{number of ways in which desired outcome can occur}}{\text{total number of all possible results of the 'experiment'}}$$

The law of total probability

- There are many probability problems where we **do not know the probability of an event B**, but **we know the probability of B given that some other event had occurred**
- An ice-cream seller has to decide whether to order more stock for the bank holiday weekend.

  If the weather is sunny, he has a 90% chance of selling all his stock, if the weather is cloudy 60%, and if it rains 20%.

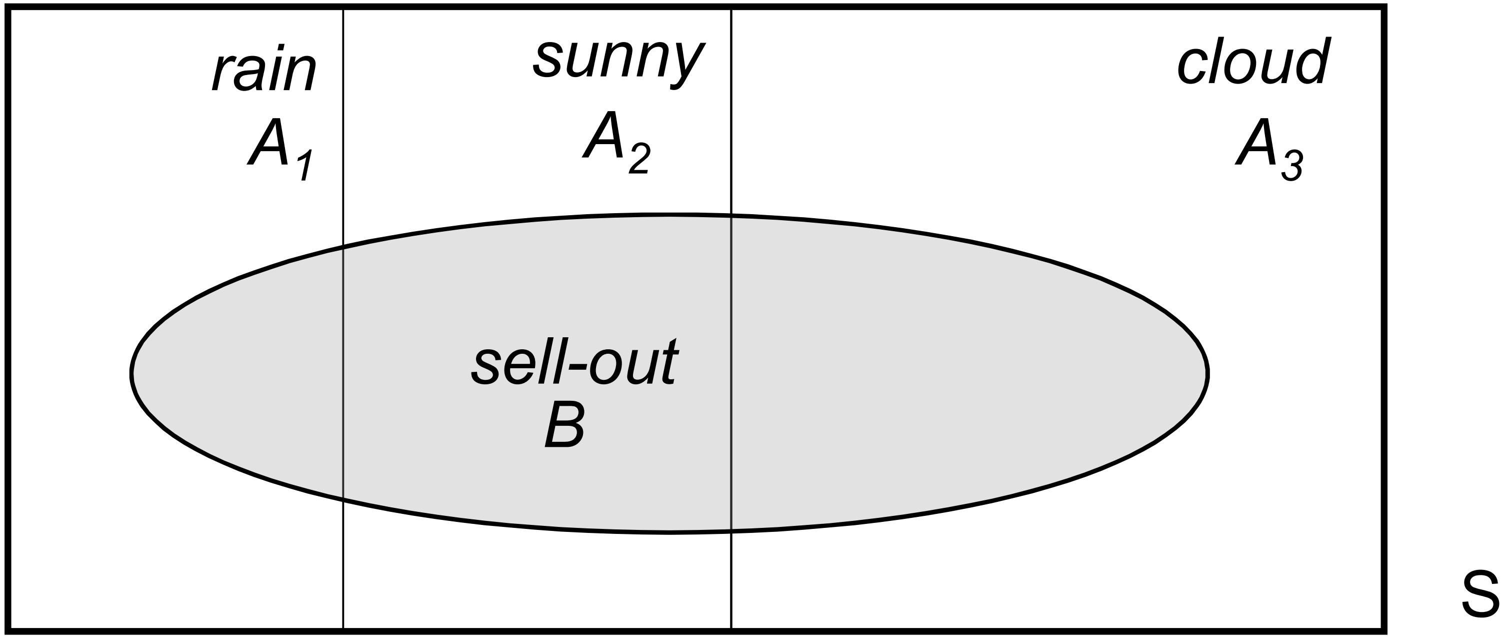
  According to the weather forecast, the probability of rain is 25%, of sunshine 30% and of cloud 45%.

  What is the overall probability that the salesman will sell all his stock?

The law of total probability

- In our example we are given:
  - P(sell-out|sunny)=0.9, P(sell-out|cloud)=0.6, P(sell-out|rain)=0.2
  - P(sunny)=0.3, P(cloud)=0.45, P(rain)=0.25
  - Note that **P(sunny) + P(cloud) + P(rain)=1**
  - We are asked to calculate P(sell-out)
- In such problems we can use the **Law of Total Probability** to find a solution

A graphical illustration



These two conditions must hold:

• $A_1 \cup A_2 \cup \ \ldots \cup A_n = S$ (the events A_1, A_2, … A_n) form a **partition of the sample space** S

• A_1, A_2, … A_n should be **pairwise disjoint (mutually exclusive)**: $A_i \cap A_j = \varnothing$ for all possible pairs of events A_i , A_j

The law of total probability

- If these conditions hold, then the law states that:

$$P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n) =$$

$$\sum_{i=1}^{n} P(A_i)P(B \mid A_i)$$

- In our example with the ice-cream:
  - P(sell-out)=P(rain)P(sell-out|rain) + P(cloud)P(sell-out|cloud) + P(sunny)P(sell-out|sunny) =>

    P(sell-out) = 0.25*0.2 + 0.45*0.6 + 0.3*0.9 = 0.59

A consequence of the law

- Let A and B be events, and suppose that P(A)≠0 and P(A)≠1. Then:

$$P(B) = P(A)P(B \mid A) + P(A')P(B \mid A')$$

- **Why?**
  - An event A and its complement A’ always satisfy the conditions for the law of total probability (they are mutually exclusive, and their union is the entire sample space **S**)

Exercise

- Suppose you have 3 restaurants to choose from for lunch on campus, R1, R2 and R3.

  R1, R2 and R3 get 50%, 30% and 20% of the business respectively.

  Suppose that you know that 70% of customers of R1, 60% of customers of R2 and 50% of customers of R3 are satisfied.

  What is the probability that someone who eats lunch on campus will be satisfied?

Sample solution: We are given: P(R1)=0.5, P(R2)=0.3, P(R3)=0.2, P(sat|R1)=0.7, P(sat|r2)=0.6, P(sat|r3)=0.5, and are asked to calculate P(sat). From the law of total probability,

P(sat)=P(R1)P(sat|r1)+P(R2)P(sat|R2)+P(R3)P(sat|R3), where all the values are known . . . P(sat)=0.63

Bayes Theorem

- There are many probability problems **where we are asked to calculate P(A|B) when we know P(B|A)**, some individual (marginal) probabilities and their complements
- In the ice-cream example, given that our salesperson sold all his ice-cream, what is the probability that the weather is sunny?
- We are given:
  - P(sell-out|sunny)=0.9, P(sunny)=0.3, and we also calculated that P(sell-out)=0.59
  - We are asked to calculate P(sunny|sell-out)

Bayes Theorem

- The theorem allows us to calculate P(A|B) given P(B|A)
- P(A|B) is called the **posterior probability**, because A has actually occurred first
  - in our example the weather was sunny and then the salesperson would have sold his ice-creams
- Bayes Theorem tells us that:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

  - the proof is actually **very** simple

In simpler words

- We have a ***prior belief*** about an event A, P(A)
  - e.g. about Crystal Palace winning the League
- Now you find out some new piece of information (event B) which is relevant to your belief
  - e.g. their star player will be out injured for the rest of the season
- You intuitively feel that you need to revise your belief in A downwards… but by how much?
- Bayes gives you the answer, if you have evidence for:
  - Prior belief about B, P(B) may be the proportion of premiership footballers sustaining a serious injury any given week
  - P(B|A), the probability of observing the new evidence given that the initial statement is true - e.g. the proportion of times teams have won the League when their star player was injured

Another version of Bayes

- It can be easily shown (based on the consequence of the law of total probability), that:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(A)P(B \mid A) + P(A')P(B \mid A')}$$

because as you recall from a few slides back:

$$P(B) = P(A)P(B \mid A) + P(A')P(B \mid A')$$

And another version of Bayes

- It can also easily be shown (based on the law of total probability), that:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{\sum_{i=1}^{n} P(A_i)P(B \mid A_i)}$$

because as you recall from a few slides back:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(A)P(B \mid A) + P(A')P(B \mid A')}$$

In our example . . .

- The posterior probability that the weather was sunny given that the stock was sold out, is given by:

$$P(sunny \mid sell-out) = \frac{P(sell-out \mid sunny)P(sunny)}{P(sell-out)} =$$

$$\frac{0.9 \times 0.3}{0.59} = 0.46 \text{ (increase from initial P(sunny) = 0.3)}$$

Exercise

- In a certain computer science class, 12% of the men and 4% of the women are taller than 6 feet. Furthermore, 20% of the students in the class are women. Suppose that a randomly selected student is taller than 6 feet. Find the probability *p* that the student is a woman.
- We are given: P(man)=0.8, P(woman)=0.2, P(tall | man) = 0.12, P(tall|woman)=0.04. We are asked to calculate P(woman|tall)

  From Bayes: **P(woman | tall) = P(tall|woman)P(woman) / P(Tall)**, from where it becomes obvious that we are missing the value for P(tall). We can calculate this from the law of total probability: **P(tall)= (woman)P(tall|woman) + P(man)P(tall|man)**=0.2x0.04+0.8x0.12 = 0.104

  we now have all the values for Bayes, we plug them in and find P(woman|tall) = 0.04x0.2 /0.104 = 0.0769

Applications of Bayes Theorem

- Bayes Theorem has found numerous applications in many fields, including Computer Science
  - Bayesian Networks
  - Bayesian Classifiers
    - spam filtering, web page classification (e.g. Yahoo-style hierarchies), object classification, etc.
  - Bayesian Machine Learning: Bayesian Inference / Bayesian Decision Theory
- More details in Week 5

Bayesian Spam Filtering

- In a very simple form, we would need to calculate probabilities of the form:

$$P(spam \mid words) = \frac{P(words \mid spam)P(spam)}{P(words)}$$

- For the conditional probabilities P(words|spam) we can collect from ‘training data’
  - emails that we know whether they are spam or not - we can count all kinds of statistics for these emails
  - in week 5 we will move one step further, and see how this formula is further broken down
- P(spam) is the prior probability that an email will be spam, we can estimate this based, again, on our training data

Summary of lecture

- In Week 3 we covered:
  - The law of total probability
  - Bayes Theorem
  - Their application to probability problems
- Don’t forget:
  - Work on the exercises for Friday **BEFORE** you come to the class
  - **Come only at your assigned time**