Ch2 - Probability Theory - 1

(Probability Definitions, Conditional Probability and Independence)

Statistics For Business and Economics - I

Shaikh Tanvir Hossain

East West University, Dhaka
Last Updated December 24, 2023

Outline

Outline

Math Recap - Sets and Related Ideas

Sets and Related Ideas

- A set is any *collection of items or objects* thought of as a whole, where the members of the sets are called ***elements***.
- If S is a set and a is an element, we can write $a \in S$. The notation $\in$ means "belongs to".
- For example, think about the set of even numbers between 1 and 11, if we enumerate then we can write this set as $S = \{2, 4, 6, 8, 10\}$. Note in this case, $2, 4, 6, 8, 10$ are the elements, so we can write $2 \in S$, $4 \in S$, and so on...
- Note that the same set can also be written as

$$S = \{x : x \text{ is an even number between 1 and 11}\}$$

- It means "S is a set of x, such that x is an even number between 1 and 11" (note the symbol ":" means "such that"). This is another way of writing sets. It is called *set builder notation*.
- Sometimes you will also see a slightly different notation "$|$" instead of ":". For example, we can write the same set as

$$S = \{x \mid x \text{ is an even number between 1 and 11}\}$$

Sets and Related Ideas

- Don't be scared of notations / symbols.
- Notations are just symbols to represent different things.
- For example we have already used the symbol ":" for "such that". So this is a notation.
- There will be many notations that we will use time to time, but please don't get worried when you see notations.
- In the end notations mean somethings, and your job is to understand what does a particular notation or symbol mean.

Sets and Related Ideas

- Here are some examples of sets,
  - $S = \{a, b, c\}$, the set S is a set of three letters, here a, b and c are called *elements* of sets. Here $a \in S$, $b \in S$ and also $c \in S$. But note that $d \notin S$.
  - $T = \{\text{Dhaka}, \text{Khulna}, \text{Barishal}\}$, the set T is a set of three cities, again Dhaka, Khulna and Barishal are the elements of the set T
  - Sets of real numbers $\mathbb{R}$. Just take all the real numbers like 0, 1, -2, 1.5. 10,000,.... and put them in a set. We will see some sets of numbers soon.
  - Think about the set of even numbers between 1 and 11, if we enumerate then we can write this set as $S = \{2, 4, 6, 8, 10\}$.
  - But the same set can also be written as $S = \{x : x \text{ is an even number between 1 and 11}\}$. It means "$S$ is a set of x, such that x is an even number between 1 and 11" (note the symbol ":" means "such that").
  - In the first three examples, we used enumeration method to write sets, in the last example we also used set builder method. Can you see the difference? Benefits?

Sets and Related Ideas

- **Empty set:** When a set has no element, then we call this an *empty set*. This set is denoted by $\emptyset$ or $\{\}$.
- **Equal Sets:** Two sets X and Y are equal if they contain exactly the same elements. and we write $X = Y$
- **Subset:** If we have two sets X and Y, and all the elements of a set X are also elements of the set Y, then X is a called a *subset* of Y, and we use the notation $X \subset Y$. Note that in this case Y is also called a *superset* of X.
- **Proper Subset:** Note when we write $X \subset Y$, then the set X may have exactly same elements as Y or may have less than Y. If all the elements in set X are in a set Y, but not all the elements of Y are in X, then X is called a *proper subset* of Y. In this case X must have less number of elements. The notation is $X \subsetneq Y$.
- Here is an example, suppose we have following sets,

$$A = \{a, b, c\}, \quad B = \{a, b, c\}, \quad C = \{b, c\} \text{ and } D = \{c\}$$

- Then we can see that $A = B$, but $A \neq C$ and also $B \neq C$ and also $C \neq D$.
- Also note $A \subset B$, and also $B \subset A$, $C \subset B$ and also $D \subset C$, and also $C \subsetneq B$ (*Question:* Is it correct to write $D \in C$, Ans: No, why?)

Sets and Related Ideas

- **Union of two sets:** The union of two sets A and B is the set of elements that belong *either* set A *or* set B or *both* of the sets. The notation we will use is $A \cup B$. So this means

$$A \cup B = \{x : x \in A \text{ OR } x \in B\}$$

- **Intersection of two sets:** The intersection of two sets A and B is the set of elements that belong to *both* A and B. We will use the notation $A \cap B$. So

$$A \cap B = \{x : x \in A \text{ AND } x \in B\}$$

- **Difference between two sets:** The difference (sometimes also called relative difference) of A and B is the set of elements that *belong to* A but *not belong to* B. The notation is $A \setminus B$. We can write,

$$A \setminus B = \{x : x \in A \text{ and } x \notin B\}$$

- **Product of two sets:** If A and B are sets, then the *Cartesian product* of A and B, is the *set* of all *ordered pairs* (a, b) such that $a \in A$ and $b \in B$. The notation is $A \times B$, and we can write.

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$$

- Note in the Cartesian Product ordering is important. For two sets A and B, $A \times B$ is not same as $B \times A$ (look at the next example).
  Example 2.1: Suppose we have following sets,
  $$A = \{a, b, c\}, \quad B = \{a, b, c\}, \quad C = \{b, c\},$$
  $$D = \{c\}, \quad E = \{1, 2\}$$
  Now $A \cup B = \{a, b, c\}$, $C \cup D = \{b, c\}$, $C \cap D = \{c\}$, $B \setminus C = \{a\}$.

  Let's think about Cartesian Products,
  $$A \times E = \{a, b, c\} \times \{1, 2\} = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\} \text{ but}$$
  $$E \times A = \{1, 2\} \times \{a, b, c\} = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}$$
- Note that, *ordering matters* for the product of sets, so $(a, 1) \neq (1, a)$. So $A \times E \neq E \times A$.

Idea of the Universal Set

- Often (depending upon the problem) we have a *universal set*, and usually we denote this set with U. Once we have a universal set, then we can find a complement of any set (which is a subset of the specified universal set). If $A \subset U$, then $A^c = U \setminus A$ (for complement sometimes there is another notation $\bar{A}$)
- Suppose we have following sets,

$$U = \{a, b, c, 1, 2\}, \quad A = \{a, b, c\}, \quad B = \{a, b, c\}, \quad C = \{b, c\}$$
$$D = \{c\}, \quad E = \{1, 2\}$$

- Now note that $A^c = \{1, 2\}$. Calculate B^c and C^c

Sets and Related Ideas

We can have a visual understanding of the set operations (union, intersection, complement, difference) using a diagram called *Venn diagram*.

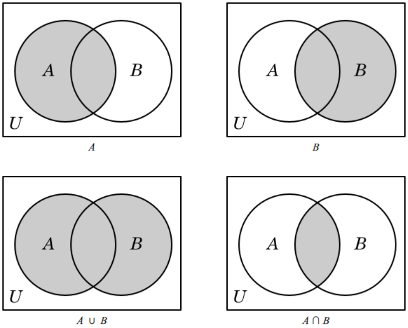


Figure 1: Venn diagram for $A \cup B$, $A \cap B$

Sets and Related Ideas

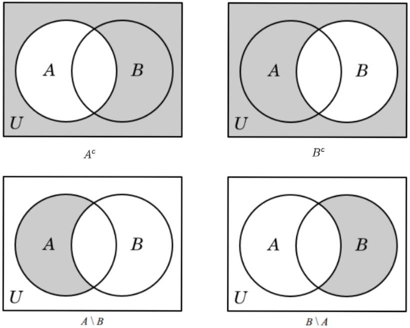


Figure 2: Venn diagram for A^c, B^c, $A \setminus B$, and $B \setminus A$

- The idea of taking Unions, Intersections, Cartesian Product and also Difference are called *set operations*, and these idea can be easily extended to more than two sets.
- For example if we have a *sequence of sets* $A_1, A_2, A_3, \ldots$, then we can write,

$$A_1 \cup A_2 \cup A_3 \cup \ldots \text{ or maybe } A_1 \cap A_2 \cap A_3 \cap \ldots$$

- There are two laws you should remember, we will learn them now without any proof!

Sets and Related Ideas

- Associative Law (interpret this law in words)-

$$A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

- Distributive Law (interpret this law in words) -

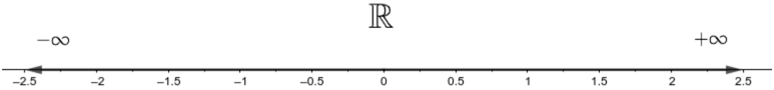
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

- There is another law which helps to connect intersections and unions via complements. This is known as *Demorgan's Law*. This law says

$$(A \cap B)^c = A^c \cup B^c$$
$$(A \cup B)^c = A^c \cap B^c$$

- We will do some math problems using these laws.
  - There are different *sets of numbers* in mathematics.
    - *Set of Real numbers*, we use the notation $\mathbb{R}$ to denote this set. This set include all numbers that you can possibly think about*. This is a huge set which is *uncountable* and of course *infinite*.
    - *Set of Natural numbers*, we use the notation $\mathbb{N}$. This set include all positive integer numbers $1, 2, 3, 4, \ldots,$. This is a *countable set* but an infinite set, so this is a *countably infinite* set†.
    - *Set of Integer numbers*, we use the notation $\mathbb{Z}$. This set include all the positive and negative integer numbers $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$. This is also a *countably infinite set*. Note that $\mathbb{N} \subsetneq \mathbb{Z}$

  - *Set of Rational numbers*, we use the notation $\mathbb{Q}$. This set include the numbers which can be written as fractions p/q, where p and q are both integers. This set has numbers like $2/3, 10/3$ and also all positive and negative integers are also part of this set (why?). This is also a *countably infinite set*. $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q}$
  - *Set of Irrational numbers*. Everything that is NOT Rational but in $\mathbb{R}$ is part of this set, for example $\sqrt{2}$, We can write this set with $\mathbb{R} \setminus \mathbb{Q}$.
  - This means we can write $\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$
- The set of real numbers $\mathbb{R}$ can be also visualized in the numberline, here is the numberline that you are probably familiar with



- At the center, we have the number 0 (this is often called the origin or center), at left the number goes to $-\infty$ and right it goes to ∞.
- We can point any number that belongs to $\mathbb{R}$ in the numberline. Here we showed only a few.

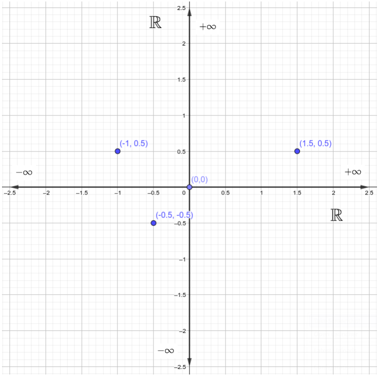
Sets and Related Ideas

- We can also have different kinds of intervals in $\mathbb{R}$, which are also subsets of $\mathbb{R}$. For example we can construct following intervals (here a and b can be any number in $\mathbb{R}$)

Notation	Set description	Type	Picture
(a, b)	$\{x \mid a < x < b\}$	Open	
$[a, b]$	$\{x \mid a \le x \le b\}$	Closed	
$[a, b)$	$\{x \mid a \le x < b\}$	Half-open	
$(a, b]$	$\{x \mid a < x \le b\}$	Half-open	
(a, ∞)	$\{x \mid x > a\}$	Open	
$[a, \infty)$	$\{x \mid x \ge a\}$	Closed	
$(-\infty, b)$	$\{x \mid x < b\}$	Open	
$(-\infty, b]$	$\{x \mid x \le b\}$	Closed	
$(-\infty, \infty)$	$\mathbb{R}$ (set of all real numbers)	Both open and closed	

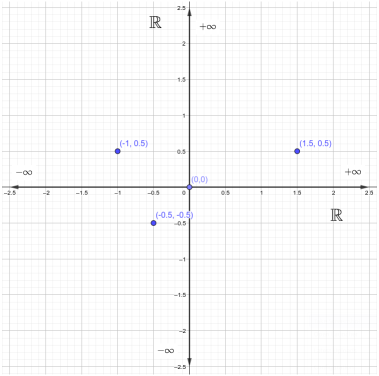
- Intervals like (a, b) is called *open intervals*, intervals like $[a, b]$ *closed intervals*, intervals like $(a, b]$ or $[a, b)$ are called *half-open* intervals.

- Recall the idea of Cartesian product? Can you think about the Cartesian product $\mathbb{R} \times \mathbb{R}$? Although it is impossible to write the set $\mathbb{R} \times \mathbb{R}$, but we can visualize it.
- Just put another numberline vertically on top of the horizontal one, then we will have something which is known as *Cartesian Coordinate* or *$x - y$ Coordinate* or *$x - y$ plane*
- Now here we have a horizontal axis, known as x-axis, and the vertical axis, known as y-axis.



Sets and Related Ideas

- Here we can show any pair of numbers (x, y), where the first number is on the x-axis and the second is on the y-axis, and togther we can locate the point (x, y) on this $x - y$ plane.
- We have showed the center, which is at $(0, 0)$, and also other three points, $(-1, 0.5)$, $(1.5, 0.5)$ and $(-0.5, -0.5)$



*except complex numbers, which we don't need now!

†Are you wondering what does countable mean? ...

Math Recap - Functions and Related Ideas

Functions and Related Ideas

- If you have taken any math courses, whether in college or courses like MAT100 or MAT110, you have definitely seen functions.
- For example following are all examples of functions,
  - $f(x) = x^2$
  - $f(x) = 2x + 1$
  - $f(x) = 3x^3 + 2x^2 + 1$
- It is very easy to understand a function, there is always *an input* and and *an output*, and the function specifies this input-output relation. Important is for each input there is only one output, there cannot be more than one.
- Following picture might be helpful

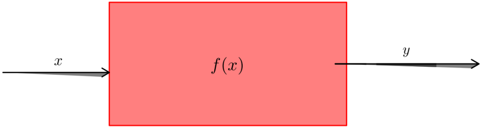


Figure 3: A function, we write $y = f(x)$ specifies a process here x can be viewed as an input and y or $f(x)$ is the output.

Functions and Related Ideas

- Formal definition of definition is similar, it just makes the definition more precise.

Definition 2.2: (Function)

Given any two sets A and B, a function $f : A \to B$ is a *mapping* between the elements of A and B such that the following condition is satisfied

- For every element of A there is a unique element in B.

In this case, the set A is called the *domain* of the function f and B is called the *codomian* of the function.

- It is important to mention that although in the definition we wrote one condition, that is *"For every element of A there is a unique element in B."*, this actually means two points,
  - First, Since we are saying "For *every* element of A...", the word "every" here automatically means all elements of A needs to be used for mapping .
  - Second, when we say for each element of A, there must be a unique element in B. This means it will never happen that a single point from A is mapped to two different points in B.

- Let's see some examples, suppose we have two sets $A = \{p, q, r, s\}$ and $B = \{1, 2, 3, 4, 5\}$. Here the domain is A and the codomain is B. Question - Is the following mapping a function? Answer - Yes it is, how?

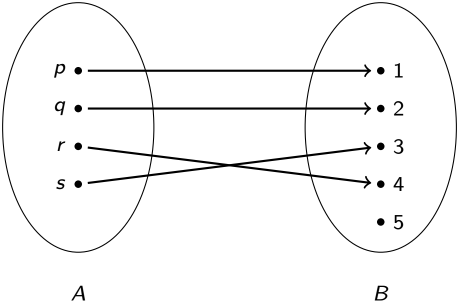


Figure 4: This is a mapping between two sets A and B. In general mapping can be any relation between the two sets. However this mapping is a function (check the condition) and we can call this function f, also we can write $f : A \to B$.

- Following is not a function. It is important to understand that here there is a viloation of the condition. In particular it violates the second point in page 23. So it is NOT a function.

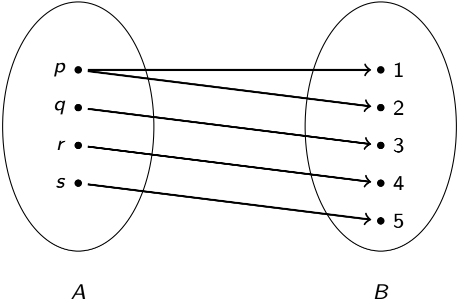


Figure 5: Note that this violates the condition, because for the element p we don't have a unique element in B, rather we have two elements 1 and 2.

- Following is a function, and there is no problem with the condition.

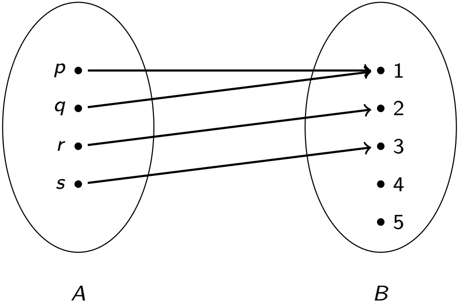


Figure 6: Here although both p and q are mapped to the same elements but still it does not violate the condition.

- For the notation of functions, usually we use the letters f, g, h, etc. Sometimes when we write many functions we also use index $1, 2, 3, \ldots,$. For example $f_1, f_2, f_3, \ldots,$ and so on.
- Always remember when we write $f : X \to Y$, this means f is a function, X is the domain and Y is the codomain.

Functions and Related Ideas

- Using the function notation for the last example we can write

$$f(p) = 1,$$
$$f(q) = 2,$$
$$f(r) = 4 \text{ and}$$
$$f(s) = 3$$

- and we can write $f : A \to B$.
- There is another set in function, which is called *range of a function*. Range is simply the subset of the co-domain which is used the in the mapping. So for the last example the set $B = \{1, 2, 3, 4, 5\}$ is the *codomain* and $\{1, 2, 3, 4\}$ is the *range*.
- Question - Note we did not use all the elements in B in Figure 3 to represent a function. Is this a problem with the definition of a function? (Answer is NO, why?)
- Question - For these two sets can you draw some mappings which are not functions? (Try this now, Hint: Just intentionally violate the two points mentioned in page 23.)

Functions and Related Ideas

- The functions that you already know or have seen, for example,
  - 1. $f(x) = x^2$
  - 2. $f(x) = 2x + 1$
  - 3. $f(x) = 3x^3 + 2x^2 + 1$

  are all examples functions where we used *algebraic expressions*. We write functions in this way when the domain and codomain are infinite or uncountable sets.
- Note that for above three functions,
  - 1. $f(x) : \mathbb{R} \to \mathbb{R}$, domain and codomain - $\mathbb{R}$ and range $\mathbb{R}_{\geq 0}$
  - 2. $f(x) : \mathbb{R} \to \mathbb{R}$, domain, codomain and range $\mathbb{R}$
  - 3. $f(x) : \mathbb{R} \to \mathbb{R}$, domain, codomain and range $\mathbb{R}$

- Here the functions are mapping between two huge sets, so we cannot draw pictures like Figure 3. But definitely when we have a *graph of the functions* then we can see also see the connections between the domian $\mathbb{R}$ and the co-domain $\mathbb{R}$.

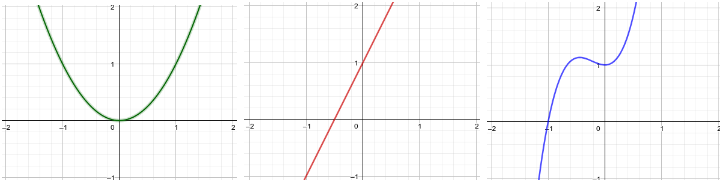
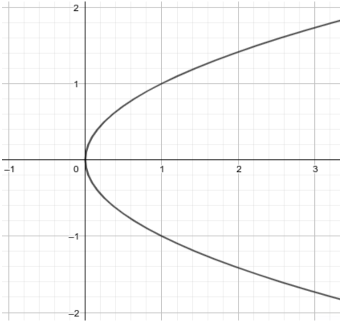


Figure 7: From left - $f(x) = x^2$, $f(x) = 2x + 1$ and $f(x) = 3x^3 + 2x^2 + 1$

- Question - If we draw any line on the $x - y$ coordinate, is it always going to be a function?
- For example, is the following a function? We can write this equation as $x = y^2$



- NO! why? Do a vertical line test.

Functions and Related Ideas

- An important type of functions is called *polynomial function*
- A polynomial of degree n is a function of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

- where the a 's are real numbers (sometimes called the coefficients of the polynomial), and x is the input of the function
- Although this general formula might look quite complicated, particular examples are much simpler.
- For example,

$$f(x) = 2 + x^2 + 3x^3$$

  is a polynomial of degree 3 , as 3 is the highest power of x in the formula. This is called a *cubic function*
- And

$$f(x) = 1 + x^5 + x^7$$

  is a function with polynomial of degree 7 , as 7 is the highest power of x.

- Notice here that we don't need every power of x up to 7: we need to know only the highest power of x to find out the degree.
- Following is a polynomial of degree 2 , as 2 is the highest power of x. This is called a *quadratic function*.

$$f(x) = 4 + 2x + 3x^2$$

- And is a polynomial of degree 1, this is called *linear function*.

$$f(x) = 2 + 3x$$

- The plotting of these functions using a software called *Geogebra* is really easy, just go to `https://www.geogebra.org/calculator` and plot the functions. I will show this on the class.

Math Recap - Counting Methods

Math Recap - Counting Methods

Multiplication rule

Counting Methods - Multiplication rule

- The *multiplication rule* helps to solve some counting problems when we have a process with more than one parts or steps.
- With this we can count *how many ways the entire process can be performed*.
- We will explain the method with some simple examples.

Counting Methods - Multiplication rule

Example 2.3: (Multiplication Rule)

- Suppose we have three cities, A, B and C. We need to go from A to C via B?
- If there are 2 ways we can go from A to B and 3 ways we can go from B to C, then how many ways we go from A to C via B?
- First note that the process has two parts, the first part we have 2 possible ways and in the second part we have 3 possible ways.
- So the whole process can be performed in $2 \times 3 = 6$ possible ways.
- For the multiplication problems, the tree diagram (figure on the right) might be useful to visualize.

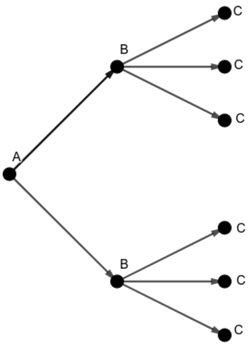


Figure 8: Tree diagram of the problem. *How many ways we can go from city A*
The answer is $2 \times 3 = 6$

Example 2.4: (Multiplication Rule)

- Suppose a retail store sells windbreaker jackets in small (S), medium (M), large (L), and extra large (XL). All are available in color "blue" or "red". If a customer wants to buy how many options/choices does he have?
- Applying multiplication rule, we get in total there are $4 \times 2 = 8$ possible choices. We can actually list them

$$\{(S, \text{Blue}), (S, \text{Red}), (M, \text{Blue}), (M, \text{Red}), (L, \text{Blue}), (L, \text{Red}), (XL, \text{Blue}), (XL, \text{Red})\}$$

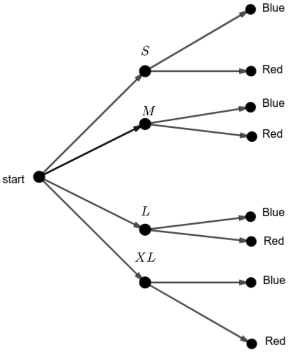


Figure 9: Tree diagram of the problem. *How many combined choices are ther*
$4 \times 2 = 8$

Counting Methods - Multiplication rule

- In the above 2 examples that we have discussed have only 2 parts in the process, but it is possible to have more than 2 parts or process with many parts.
- In this case we can directly count the total number of ways, and there is no need to draw the tree diagram.

Example 2.5: (Multiplication Rule)

- Suppose a coin is tossed 6 times, how many possible outcomes are there.
- Actually there will be $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^6 = 36$ possible outcomes. Can you list all possible outcomes. For example one outcomes is $HTTHHH$ (can you think about the tree diagram here?)
- Can you draw the tree diagram (yes but this is cumbersome)?

Example 2.6: (Multiplication Rule)

- Suppose we have a 3 digit combination lock where each digit can be from 0 to 9. How many possible combination locks we can set?
- There will be $10 \times 10 \times 10 = 10^3 = 1000$ possible combination locks.
- Can you draw the tree diagram (yes but this is cumbersome)?

Math Recap - Counting Methods

Permutation

Counting Methods - Permutation

- We can think about the combination problem in the following way, think about three empty boxes and then we want to know how many possible ways we can fill the boxes?
- For the first box we have 10 possible options, for the second we also have 10, and for the third we also have 10. This gives the following picture

$$\boxed{10} \times \boxed{10} \times \boxed{10}$$

- This means we have $10 \times 10 \times 10 = 1000$ possible options for locks.
- This is the same problem but we are thinking now with boxes, rather than tree diagram.

- Now consider the last combination lock problem, but now suppose we don't want any repetition. This means the same digit cannot appear more than once.
- In this problem we want all three digits are different. For example we don't want to count $0, 0, 1$ or $1, 1, 1$ as possible count.
- We can solve this problem using the box idea.
- For this problem the first place for the lock has 10 digits, the second place for the lock has 9 digits, and the third place for the lock has 8 digits.

$$\boxed{10} \times \boxed{9} \times \boxed{8}$$

- So we have $10 \times 9 \times 8 = 720$ possible combinations.
- So problem solved, this is a multiplication problem.

Counting Methods - Permutation

- Now the last answer is also an answer of another problem called *the counting problem where order matters*, this is also known as *ordering problem*.
- For ordering problem, we don't have to think about the combination lock, the idea is if we have 10 digits $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$, we ask how many ways we can *order* any 3 digits.
- The answer is the famous *formula for permutation*

$$^{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 10 \times 9 \times 8$$

- Notice the answer is same.
- So ordering problem is a multiplication problem.
- SideNote: In any ordering problem when we say "ordering matters", this means when we are counting we are treating $1, 2, 3$ and $2, 1, 3$ as a separate count.
- SideNote: The word "permutation" in English just means "rearrangement". For example if we have three letters a,b,c then a another permutation (or ordering) is b,a,c. So when we ask total number of permutations, this is same asking total number of arrangements or total number of orderings.

- In general if someone asks you "if we have n objects then how many ways we can order k of them if we pick one at a time?" the answer is

$$^{n}P_{k} = \frac{n!}{(n-k)!}$$

- Or you can think with the boxes, in this case you can think n objects and k empty boxes and we are trying to fill them one by one.
- SideNote: What if we have n objects and n boxes, then $^{n}P_{n} = n!$ (here we used $0! = 1$) This means $n!$ gives the total number of orderings when we want to order n objects and we have n empty boxes.
- Let's do some examples.

Counting Methods - Permutation

- Ques: If we have 5 letters a, b, c, d, e, then
  - a) How many ways we can order them?
  - b) How many ways we can order 3 of them (taking one at a time)?
- The answer of a) is $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$.
- To answer b), we can follow one of the two approaches
  - **Empty box approach (perhaps this is more intuitive):** We have 3 empty boxes, so for the first box we have 5 options, for the second box we have 4 and for the third we have 3. So in total we have $5 \times 4 \times 3 = 60$. So there are 60 possible ways we can order them
  - **Directly applying the formula:** Since this is a direct ordering problem we can apply the permutation formula, ${}^5P_3 = \frac{5!}{(5-3)!} = 60$.
- Here are all 60 permutations if we pick 3 letters out of 5.

a, b, c	b, a, c	c, a, b	a, c, b	b, c, a	c, b, a
a, b, d	b, a, d	d, a, b	a, d, b	b, d, a	d, b, a
a, b, e	b, a, e	e, a, b	a, e, b	b, e, a	e, b, a
a, c, d	c, a, d	d, a, c	a, d, c	c, d, a	d, c, a
a, c, e	c, a, e	e, a, c	a, e, c	c, e, a	e, c, a
a, d, e	d, a, e	e, a, d	a, e, d	d, e, a	e, d, a
b, c, d	c, b, d	d, b, c	b, d, c	c, d, b	d, c, b
b, c, e	c, b, e	e, b, c	b, e, c	c, e, b	e, c, b
b, d, e	d, b, e	e, b, d	b, e, d	d, e, b	e, d, b
c, d, e	d, c, e	e, c, d	c, e, d	d, e, c	e, d, c

Counting Methods - Permutation

- Note that ordering matters. Now it should be clear to you what does it mean - Take the first row where we have different permutations of the letter a, b, c. In total there are $3 \times 2 = 6$ permutations (look at row 1), ordering matters means when we count, we count all 6 of them.
- Similarly in every row we have 6 permutations of three letters and when we count we count all of them.
- What if we treat them as a single count?

Math Recap - Counting Methods

Combination

Counting Methods - Combination

- The answer to the question in the last line of the last slide is the answer of a *combination problem*.
- In the combination problem we just select, and we count all permutations as one count.
- For example, in the last letter problem, if we ask *"how many ways we can select 3 letters out of 5 letters", the answer is 10*.
- Notice the word "select" here.
- Here "select" means, we are just selecting and we don't care about orders now.
- Now how did we get 10? Just count one for each row where we show all possible permutations. Since there are 10 rows we have 10 possible combinations.
- Let's see the definition now and hopefully things will be clear.

Definition 2.7: (Combinations)

If we have a set of n elements. Each *subset* of size k chosen from this set is called a *combination of n elements taken k at a time*. We denote the number of *distinct* such combinations by the symbol ${}^nC_k$. And we can count this number by

$$^nC_k = \frac{n!}{k!(n-k)!}$$

- Notice the last formula can be written as

$$^{n}C_k = \frac{^{n}P_k}{k!}$$

- How does this formula come? We can explain this via permutation. The idea is let's think permutations as being constructed in two steps or two parts.
- *Step 1* A combination of k elements is chosen out of n, this is $^{n}C_k$
- *Step 2* those k elements are arranged in a specific order within themselves. This is $k!$
- Now we can use multiplication rule and we see that

$$^{n}P_k = {}^{n}C_k \times k!$$

- From here we get our formula

$$^{n}C_k = \frac{^{n}P_k}{k!}$$

- So we can say that the number of distinct subsets of size k that can be chosen from a set of size n is $^{n}C_k$.
- Or if someone asks you *"how many ways you can select k objects from n?"*, then the answer is $^{n}C_k = \frac{^{n}P_k}{k!} = \frac{n!}{k!(n-k)!}$
- There is another notation for the combination and that is $\begin{pmatrix} n \\ k \end{pmatrix}$
- So $\begin{pmatrix} n \\ k \end{pmatrix}$ is same as $^{n}C_k$, it means *"how many distinct ways we can select k objects from n?"*

There is a very useful application of ${}^nC_k$, we this call *Binomial Theorem*, here is the theorem.

Theorem 2.8: (Binomial Theorem.)

For all numbers x and y and each positive integer n,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

where $\binom{n}{k}$ is same as ${}^nC_k$, so this is possible number of combinations of k objects out of n. In this case this is also known as *Binomial co-efficient*

Random Experiment

Random Experiment

Probability theory starts from Random Experiment. Here is the definition,

Definition 2.9: (Experiment and Event)

A *random experiment* is any process, real or hypothetical, in which *before performing* the experiment we can identify all possible outcomes but we don't know exactly which outcome will come.

- The *set* of all possible outcomes is called *sample space of the experiment*. We will use the notation ω to denote a single outcome and Ω to denote the sample space, this means $\Omega = \{\omega : \omega \text{ is an outcome of the experiment}\}$
- Any subset of the sample space is called an *event of the experiment*.

- Note that the definition says before the experiment is performed we know all possible outcomes, but we do not know which outcome will come (so there is a lack of information or uncertainty!).
- Also another important thing, usually we can perform the same experiment more than once. When we perform the experiment a single time, we call it a *trial* of the experiment.
- Let's see some specific examples.
- Sidenote: Here both Ω and ω are Greek letters, see https://en.wikipedia.org/wiki/Omega. This is pronounced as "Oh-may-gaa". Ω is the upper-case and ω is the lower-case
- Here are some examples of Random Experiment.

Random Experiment

- Tossing a coin. The sample space is $\Omega = \{\mathrm{H}, \mathrm{T}\}$
- Tossing two coins. The sample space is $\Omega = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T}\})$ (use multiplication rule to calculate the total number of possible outcomes)
- Throwing a die - The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Throwing two dice - The sample space is $\Omega = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 6), (2, 6), (2, 1), \ldots, (6, 6)\}$ (use multiplication rule to calculate the total number of possible outcomes, here total number of possible outcomes is 36.)

- Another important example of random experiment is sampling,
- Sampling - The current population of Bangladesh is about $168, 000, 000$. Suppose we *randomly pick a sample* of 100 people so that it is a "good" representative of the population. This is a random experiment, because we don't know which 100 people will come in our sample, but we know the sample space Ω. It is the set of all people in Bangladesh. The sample in this case is called a *random sample*.
- It is important to note that in Statistics the bigger set from which we take our sample is called *population*. This may or may not mean literally population of a country. This could be something else. It depends upon the what problem we are trying to solve.
- In Statistics we are often interested to know about the population, or some characteristics about the population (for example average income of the population) but what happens is we cannot access to the population, so try to get a random sample and then use that sample to say something about the population (we will see more about this later in our course!).

Random Experiment

- We will come back to this. But for now just take the lesson that, *random sampling* is a very very important kind of random experiment. In fact most of the data that we analyze is a result of some kind of *random sampling*.

Figure 10: Throwing dices, tossing a single coin and sampling from a population, all are examples of random experiment!

- Once we know the sample space Ω, we can actually form different subsets of Ω, and think about different *events*. Recall an *events* is simply a subset of the sample space, so in principle everything that we have learned about Sets could be directly applicable when we are talking about Events.

Random Experiment

- For example, for the coin toss experiment, when the sample space is $\Omega = \{\mathrm{H}, \mathrm{T}\}$, can you think about *all possible events* (think about all possible subsets)? Yes, in this case, the answer is easy,

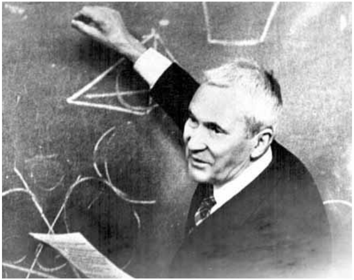
$$\{\mathrm{H}\}, \{\mathrm{T}\}, \{\mathrm{H}, \mathrm{T}\}, \emptyset$$

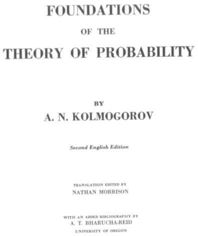
  - $\{\mathrm{H}\}$ is an event since $\{\mathrm{H}\} \subset \Omega$. Here event $\{\mathrm{H}\}$ means only head is appearing.
  - Similarly $\{\mathrm{T}\}$ is an event, it means only tail is appearing.
  - $\{\mathrm{H}, \mathrm{T}\}$ is also an event, since, it satisfies the definition of a subset. Note $\{\mathrm{H}, \mathrm{T}\} = \Omega$
  - Ques- What does the event $\{\mathrm{H}, \mathrm{T}\}$ mean? Ans: It means *any one* of the outcomes will appear, we can write $\{\mathrm{H}, \mathrm{T}\} = \{\mathrm{H}\} \cup \{\mathrm{T}\}$
  - It might look strange why $\emptyset$ is a subset of Ω. The answer is, it satisfies the definition of a subset. Recall, the set A is a subset of the set B if and only if every / all element of A is also an element of B. If A is the empty set then A has no elements and so all of its elements (there are none) belong to B no matter what set B we have. So, the empty set $\emptyset$ is a subset of every set. And in this case $\emptyset \subset \Omega$. Ques- *What does the event $\emptyset$ mean?* Ans: It means, nothing is appearing, so it is an impossible event.

Probability Definitions

Probability Definitions

- Although all of us might have some intuitive understanding of probability, but the history of Mathematics tells us that the modern definition of probability came not so long ago.
- The Russian Mathematician Andrey Nikolaevich Kolmogorov (1903-87) laid the mathematical foundations of probability theory and the theory of randomness.





FOUNDATIONS
OF THE
THEORY OF PROBABILITY

BY
A. N. KOLMOGOROV

Second English Edition

TRANSLATION EDITED BY
NATHAN MORRISON

WITH AN ADDED BIBLIOGRAPHY BY
A. T. BHARUCHA-REID
UNIVERSITY OF OREGON

- His monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung - Foundations of the Theory of Probability*[‡], published in 1933 first introduced the Probability Theory in a rigorous way using fundamental axioms.

Probability Definitions

- We will see the *Axiomatic approach* of defining probability later, first let's see the *Classical approach* and *Frequentist approach* of defining probability.
- In all definitions we will calculate probability for events. For example if the set A is an event (i.e., $A \subset \Omega$) then we will calculate $\mathbb{P}(A)$, this is going to be a number in $[0, 1]$.

Definition 2.10: - Classical Definition of Probability

If an experiment has n *equally likely* outcomes, and there is an event A where the number of outcomes is n_A, then the *probability of the event A* is,

$$\mathbb{P}(A) = \frac{n_A}{n} \tag{1}$$

- So when we are thinking about the event A, the classical definition says we can calculate the probability by,

$$\mathbb{P}(A) = \frac{\text{number of outcomes in the event } A}{\text{number of outcomes in the sample space or total number of outcomes}}$$

- Let's apply the classical definition and calculate probabilities of some events of an experiment.

  Example 2.11: (Applying the classical definition to calculate probabilities)
  Suppose our Probability is throwing a *balanced die*. Note here *balanced die* means the outcomes *are all equally likely*. Here we have $\Omega = \{1, 2, 3, 4, 5, 6\}$, so $n = 6$. Let A be the event that *an even number occurs*. This means

  $$A = \{2, 4, 6\}$$

  We want to calculate $\mathbb{P}(A)$. Here we have three outcomes for the event A (or associated with the event A), so $n_A = 3$, this means

  $$\mathbb{P}(A) = \frac{n_A}{n} = \frac{3}{6} = \frac{1}{2}$$

Example 2.12: (Applying the classical definition to calculate probabilities)
Suppose we toss 2 coins. Assume that all the outcomes are equally likely (fair coins).

- (a) What is the sample space?
- (b) Let A be the event that at least one of the coins shows up heads. Find $\mathbb{P}(A)$.

Example 2.13: (Applying the classical definition to calculate probabilities)
Now suppose we toss 3 coins. Assume that all the outcomes are equally likely (fair coins).

- (a) How many elements are there in the sample space? Can you write one random element?
- (b) Let A be the event that we have heads in all 3 coins. Find $\mathbb{P}(A)$.
- (c) Let B be the event that we have exactly one head and 2 tails. Find $\mathbb{P}(B)$.

Example 2.14:(Applying the classical definition to calculate probabilities)
Suppose in a city license plates have six characters: 3 letters followed by 3 numbers. Answer following questions,

- a) How many distinct such plates are possible?
- b) How many distinct plates are possible if the license plate contains no duplicate letters or numbers?
- c) Given that all sequences of six characters are equally likely, what is the probability that a randomly selected license plate for a new car will contain no duplicate letters or numbers?

- Ans of a) We can apply multiplication rule, there are $26^3 = 17,576$ different ways to choose the letters and $10^3 = 1000$ ways to choose the numbers, so we have $26^3 \times 10^3 = 17,576 \times 1000 = 17,576,000$ number of distinct plates. This means Ω consists the set of of all $17,576,000$ possible license plates, so here $n = 17,576,000$

- Ans of b) Let's denote the event with A where we do not have any duplicates with numbers or digits. This means set A has license plates with no duplicate letters or number.

  Now, no duplicate letters means there are $26 \times 25 \times 24 = 15,600$ ways to choose the letters. And then, no duplicate numbers mean there are $10 \times 9 \times 8 = 720$ ways to choose the numbers. From the multiplication principle, the number of outcomes in the event A is $15,600 \times 720 = 11,232,000$, so $n_A = 11,232,000$.

- Ans of c) So now we can calculate the probability of happening the event A,

$$\mathbb{P}(A) = \frac{11,232,000}{17,576,000} = .64$$

Probability Definitions

- We will solve more examples in the practice sheet, now let's discuss the issues with the classical definition.
- There are essentially two major problems with the classical definition of probability
  - Assumption of equally likely outcomes (how do we know this?). For example if we have a biased coin, then how do we calculate probability.
  - Finite sample space issues (sample space can be very large, e.g., $\Omega = \mathbb{R}$)
- Another definition is known as the *Frequency definition* of probability

Definition 2.15: - Frequency Definition of Probability

The probability of an event A is the relative proportion of outcomes if we perform the experiment *under identical condition* for a large number of times.

- So for example if our experiment is tossing a single coin, the probability of appearing heads is the number of times heads will appear if we perform this experiment almost infinite number of times.
- Frequency definition does not have equally likely outcomes assumption, but the issue is we need to perform the experiment *under identical conditions*, and this is often not possible.

Probability Definitions

- So in terms of the definition, the axiomatic definition does not have these issues, rather it's an abstract definition where we will define probability as a *set function*.

Definition2.16: - Axiomatic Definition of Probability

For a random experiment, if we have a sample space Ω and then we can define probability as a *set function* P such that for any event $A \subset \Omega$

- 1. $\mathbb{P}(A) \geq 0$
- 2. $\mathbb{P}(\Omega) = 1$.
- 3. For *pairwise disjoint but countable number of events* $A_1, A_2, \dots$ we have

$$\mathbb{P}\left(A_1 \cup A_2 \cup A_3 \dots\right) = \mathbb{P}\left(A_1\right) + \mathbb{P}\left(A_2\right) + \mathbb{P}\left(A_3\right) + \dots$$

- Let's explain each of these axioms (in class discussion).
- Note that, unlike the other definition, the Axiomatic definition does not tell us any ways to calculate probabilities, it only defines probability as a function.
- This means as long as any set function satisfies above three axioms, we will consider that function a probability function. Sometimes Probability function is also called *Probability measure*.

Probability Definitions

- With this definition, now we can show that the following rules of calculating probability

Theorem 2.17: (Probability Calculus)

P is a probability function and A is any event, then we can show that

- a. $\mathbb{P}(\emptyset) = 0$
- b. $\mathbb{P}(A) \leq 1$
- c. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

Proof:
First we will prove c. First note using Venn diagram, we can see that

$$\Omega = A \cup A^c$$

This means A and A^c makes a *partition* of the sample space Ω (What is a partition? It simply means if we take union of disjoint sets will get the whole set) Now we will apply the axioms,

$$\begin{aligned} \Omega &= A \cup A^c \\ \implies \mathbb{P}(\Omega) &= \mathbb{P}(A \cup A^c), \text{ [apply the second axiom]} \\ \implies \mathbb{P}(\Omega) &= \mathbb{P}(A) + \mathbb{P}(A^c) \text{ [}A\text{ and }A^c\text{ are disjoint, so apply the third axiom]} \\ \implies 1 &= \mathbb{P}(A) + \mathbb{P}(A^c) \text{ [Apply first axiom]} \end{aligned}$$

So last line means $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. So we have shown c

Since $0 \leq \mathbb{P}(A^c) \leq 1$ and we know $1 = \mathbb{P}(A^c) + \mathbb{P}(A)$, it means we must have $\mathbb{P}(A) \leq 1$, so this means (b) holds.

To prove (a), we use a similar argument like *c* First note,

$$\begin{aligned} \mathbb{P}(\Omega \cup \emptyset) &= \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) \text{ [since } \Omega \text{ and } \emptyset \text{ are disjoint and } \Omega = \Omega \cup \emptyset \text{, we apply third axiom]} \\ \mathbb{P}(\Omega) &= \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) \\ 1 &= 1 + \mathbb{P}(\emptyset) \text{ [apply second axiom]} \end{aligned}$$

so we have $\mathbb{P}(\emptyset) = 0$.

□

Probability Definitions

- As a side note here are the formal definitions of disjoint, pairwise disjoint and partition

Definition 2.18: (Disjoint, Pairwise Disjoint and Partition)

- Two events A and B are *disjoint* (or also called *mutually exclusive*) if $A \cap B = \emptyset$.
- The sequence of events $A_1, A_2, \ldots$ are *pairwise disjoint* (or *pairwise mutually exclusive*) if $A_i \cap A_j = \emptyset$ for any $i \neq j$.
- If $A_1, A_2, \ldots$ are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = \Omega$, then the collection $A_1, A_2, \ldots$ forms a *partition* of Ω.

Partition means it will break the sample space in disjoint parts. These concepts are easy to understand if we draw the Venn Diagrams.

Probability Definitions

Theorem 2.19: (More Probability Calculus)

If P is a probability function and A and B are any events, then

- a. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
- b. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

- It is possible to prove the above claims using the definition only, but we will skip it for now. But let's try to understand the theorem intuitively on board.
- $\mathbb{P}(A \cap B)$ is called the *joint probability*, because this calculates the probability of happening both events. On the other hand $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are called *marginal probabilities*.
- Note that if $\mathbb{P}(A \cap B) = 0$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. But in general we cannot write this, we have to use Theorem 2.9 (a)
- Also note, axiomatic definition doesn't tell us how to calculate probabilities, we only know some rules.
- When we have a *countable and finite sample space* then there is a nice rule to assign/calculate probability of an event A, following theorem gives us this rule.
- You have already applied this rule for the equally likely case. But now we don't need "equally likely assumption".

Theorem 2.20: (A rule to assign probabilities of events for a finite sample space)

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite sample space and let $\mathbb{P}(\{\omega_i\}) = p_i$, for $i = 1, 2, \ldots, n$ such that following two conditions hold

1. $p_i \geq 0$ for all $i = 1, 2, \ldots, n$
2. $\sum_{i=1}^{n} p_i = 1$

If for any event A, we can define $\mathbb{P}(A)$ by

$$\mathbb{P}(A) = \sum_{\{i:\omega_i \in A\}} p_i$$

Also for $\emptyset$ we have $\mathbb{P}(\emptyset) = 0$, then we can show that $\mathbb{P}$ is a probability function (this means all axioms are satisfied).

The above theorem remains true if Ω is a countable set, it means we can apply this theorem when we have $\Omega = \{\omega_1, \omega_2, \ldots\}$

Probability Definitions

- Let's see an application of this theorem. Suppose an experiment has five outcomes: $\omega_1, \omega_2, \omega_3, \omega_4, \omega_5$. Then if we know

$$\mathbb{P}(\{\omega_1\}) = p_1 = 0.2$$
$$\mathbb{P}(\{\omega_2\}) = p_2 = 0.3$$
$$\mathbb{P}(\{\omega_3\}) = p_3 = 0.2$$
$$\mathbb{P}(\{\omega_4\}) = p_4 = 0.1$$
$$\mathbb{P}(\{\omega_5\}) = p_5 = 0.2$$

  The theorem says we can calculate probabilities for any events. For example, we can calculate $P(\{\omega_1, \omega_2\})$
- First note the sample space is, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$. Now, let's calculate $P(\{\omega_1, \omega_2\})$. If we apply the theorem we have,

$$P(\{\omega_1, \omega_2\}) = p_1 + p_2 = 0.2 + 0.3 = 0.5$$

- Can you calculate the probability $P(\{\omega_1, \omega_2\})$, if we assume equally likely assumption?

‡Go to https://www.kolmogorov.com/Foundations.html to see the scanned version of the English translation.

Conditional Probability

Conditional Probability

- Now we will start with an important concept called *Conditioning*.
- *Conditioning is the soul of Statistics* (Joe Blitzstein, Harvard Stat 110).
- All of the probabilities that we have dealt so far are unconditional probabilities. A sample space was defined and all probabilities were calculated with respect to that sample space.
- However in many instances, we have new information.
- When we calculate the probabilities with updated information, we call it *Conditional Probability*.
- Now, when we have new information, ideally we need to update the sample space. But there is a problem, in many cases we are not in a position to *update the sample space*, hence we need new probability calculation but based original sample space. This is idea of the formula for conditional probability!
- Let's see the formula or the definition,

Definition 2.21: (Conditional Probability)

If A and B are events in Ω, and $\mathbb{P}(B) > 0$, then the *conditional probability* of A given B is defined as

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2}$$

- Here is an example

  Example 2.22: (Conditional Probability)

- It is very common for patients with episodes of depression to have a recurrence (or relapse) within two to three years. Suppose we have studied 3 treatments for depression: *imipramine, lithium carbonate, and a combination*. As is traditional in such studies (called clinical trials), there is also a group of patients who received a *Placebo*A placebo is a treatment that is supposed to be neither helpful nor harmful. Some patients are given a placebo so that they will not know that they did not receive one of the other treatments. None of the other patients knew which treatment or placebo they received.

- In this example, we shall consider 150 patients who entered the study. They were divided into the four groups (3 treatments and placebo) and followed to see how many had recurrences of depression. Following table summarizes the results (recall this is just a contingency table or crosstabulation).

	Treatment group				
Response	Imipramine	Lithium	Combination	Placebo	Total
Relapse	18	13	22	24	77
No relapse	22	25	16	10	73
Total	40	38	38	34	150

Conditional Probability

- Here are couple of questions (For all questions assume, equally likely case, this means all patients have same probability of getting selected)
  - a) What is the probability that a randomly selected patient had a relapse?
  - b) What is the probability that a randomly selected patient received a placebo?
  - c) What is the probability that a randomly selected patient received placebo and also had a relapse?
  - d) *Conditioning on the fact that a patient received placebo* (or if we know that the patient received a placebo), what is the probability that the patient had a relapse?

- Suppose A is the set of patients who had a relapse. Then calculating with the equally likely assumption, $\mathbb{P}(A)$ can be calculated with

$$\mathbb{P}(A) = \frac{\#\text{ of patients who had relapse}}{\#\text{ total patients}} = \frac{77}{150}$$

- Let B be the event that the patient received placebo, then we can also calculate,

$$\mathbb{P}(B) = \frac{\#\text{ of patients who received placebo}}{\#\text{ total patients}} = \frac{34}{150}$$

- So far we calculated the *marginal probabilities* of A and B.
- Now we can also calculate the joint probability $\mathbb{P}(A \cap B)$.
- Where $A \cap B$ is the event where a randomly selected patient received a placebo and also had a relapse.

$$\mathbb{P}(A \cap B) = \frac{24}{150}$$

Conditional Probability

- With this, just applying the formula we can also calculate $\mathbb{P}(A|B)$, which calculates the conditional probability,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{24/150}{34/150} = \frac{24}{34}$$

- Note that there is a difference between the event $A \cap B$ and the conditional event $A \mid B$
- Interesting to note is, the conditional probability can also be calculated with $24/34$, this is the calculation with the updated sample space includes only placebo patients. In this case we don't need to apply the formula $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, we can directly do the calculation as,

$$\frac{\#\text{patients who had a relapse out of the patients who received placebo}}{\#\text{ of patients who received placebo}}$$

- The formula is for when we use the original sample space and we want to calculate the conditional probability.

Conditional Probability

- We can also calculate $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B) = 10/34$. This is always possible for conditional probability.
- So conditional probability function will act like a probability function. But it will give us an updated calculation with respect to the new sample space.
- So you can say that, the intuition of conditional probability calculation is - our original sample space Ω has been updated to B, and then all further calculations are updated with respect to their relation to B.
- Ques: What happens to conditional probabilities of disjoint sets? Suppose A and B are disjoint, so $A \cap B = \emptyset$ and $\mathbb{P}(A \cap B) = 0$. It then follows that $\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A) = 0$. This means nothing will be updated for A if A and B are disjoint sets.
- Ques: When does it happen that $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ (this means, unconditional probability = conditional probability) ?
- Note that this happens when $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$, since

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

Conditional Probability

- The conditional probability definition gives a way to calculate the probabilities of a joint event. Note that using the formula in (2) we can easily get.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) \tag{3}$$

- This is sometimes called *the multiplication rule of conditional probability* (do not confuse this with the multiplication rule for counting!)
- Now using the same idea in (2) we can also calculate

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \text{ given that } \mathbb{P}(A) \neq 0 \tag{4}$$

- From here using the multiplication idea, we get

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

- So now we have a different way of writing $\mathbb{P}(A|B)$,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{5}$$

- The last formula where we "turned around" the conditional probabilities is called *Bayes' Rule*, this is after the name of Sir Thomas Bayes.
- So the Baye's Rule is

Theorem 2.23: (Bayes' Rule)

Let A and B be two events on the sample space Ω, and assume that $\mathbb{P}(B) > 0$, then we have

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{6}$$

Conditional Probability

- Note for three sets A_1, A_2, A_3 using Conditional Probability we can also calculate,

$$\mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)}$$

- Using the multiplication rule of conditional probability, we get

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2 \cap A_3) &= \mathbb{P}(A_1 \cap A_2) \times \mathbb{P}(A_3 \mid A_1 \cap A_2) \\ &= \mathbb{P}(A_1) \times \mathbb{P}(A_2|A_1) \times \mathbb{P}(A_3 \mid A_1 \cap A_2)\end{aligned}$$

- You can extend this formula for more than 2 events, but I am skipping the general version, see **?** for details.

- Now we will learn another law, which is called *Law of Total Probability*. This law is very important and it is an application of *partition*.

Theorem 2.24: (Law of Total Probability)

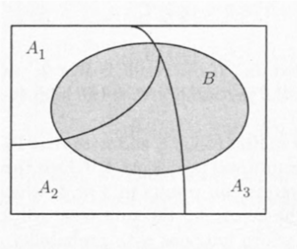
Let $A_1, \ldots, A_n$ be events that form a partition of the sample space Ω and assume that $\mathbb{P}(A_i) > 0$, for all i. Then, for any event B, we have

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1)\,\mathbb{P}(A_1) + \cdots + \mathbb{P}(B \mid A_n)\,\mathbb{P}(A_n) \tag{7}$$

- We will skip the general proof, but let's understand the theorem for a simpler case.

Conditional Probability

- Consider the following Venn diagram,



- Here A_1, A_2 and A_3 forms a *partition*. Recall a partition is a sequence of sets which splits the entire sample space.
- If A_1, A_2 and A_3 forms a partition of Ω and B is a set, then we can write,

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \tag{8}$$

- All sets are disjoint, so using the third axiom of Definition 2.16 we have,

$$\mathbb{P}(B) = \mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B) + \mathbb{P}(A_3 \cap B)$$

- Now using conditional probabilities we have

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}\left(A_1 \cap B\right) + \mathbb{P}\left(A_2 \cap B\right) + \mathbb{P}\left(A_3 \cap B\right) \\ &= \mathbb{P}\left(B \mid A_1\right)\mathbb{P}\left(A_1\right) + \mathbb{P}\left(B \mid A_2\right)\mathbb{P}\left(A_2\right) + \mathbb{P}\left(B \mid A_3\right)\mathbb{P}\left(A_3\right)\end{aligned}$$

- So this is the Law of Total Probability given in Theorem 2.24, but we explained it for three sets. You can extend the idea generally for n sets,

$$\mathbb{P}(B) = \mathbb{P}(B \mid A_1)\mathbb{P}(A_1) + \cdots + \mathbb{P}(B \mid A_n)\mathbb{P}(A_n)$$

- Now we can apply Bayes' rule here. First note applying simple Bayes' rule for set A_1 and B we get,

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1)}{\mathbb{P}(B)}$$

- Now we apply the law of total probability for $\mathbb{P}(B)$

$$\begin{aligned}\mathbb{P}(A_1 \mid B) &= \frac{\mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1)}{\mathbb{P}(B \mid A_1)\,\mathbb{P}(A_1) + \cdots + \mathbb{P}(B \mid A_n)\,\mathbb{P}(A_n)}\end{aligned}$$

- This is what we call the general version of the Bayes' rule or the Bayes' rule with law of total probability.
- Now we write the general version.

Theorem 2.25: (Bayes' Rule with Law of Total Probability)

Let $A_1, A_2, \ldots, A_n$ be events that form a partition of the sample space Ω, and assume that $\mathbb{P}(A_i) > 0$, for all i. Then, for any event B such that $\mathbb{P}(B) > 0$, we have

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)}{\mathbb{P}(B)}$$
$$= \frac{\mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i)}{\mathbb{P}(B \mid A_1)\,\mathbb{P}(A_1) + \cdots + \mathbb{P}(B \mid A_n)\,\mathbb{P}(A_n)}$$

Example 2.26: (Problem on Conditioning)
Suppose we have the following test results for a class of students. There are only two possible Grades, Grade A or Grade B, otherwise students fail the course. The course is offered combinedly for Bachelor and Master level students. We have the following data.

		GRADES			
		A_1	A_2	A_3	Totals
LEVEL	B_1	352	197	251	800
	B_2	150	161	194	505
	Totals	502	358	445	1305

Note in the table A_1, A_2 and A_3 means Grade A, B and F and B_1, B_2 means Masters and Bachelors students.

More Problems on Conditioning

- Calculate the joint probabilities and write down in a table.
  First Calculate each of the joint probabilities, and plug the value in the table. Here is the joint probability table

		GRADES			
		A_1	A_2	A_3	Totals
LEVEL	B_1	0.27	0.15	0.19	0.61
	B_2	0.12	0.12	0.15	0.39
	Totals	0.39	0.27	0.34	1

- Given that a student is a Bachelor student what is the Probability that he got grade A?
- Given that a student failed what is the Probability that he is a Masters student?

Example 2.27: (Bayes Rule and Conditioning)
Now suppose we change the question a bit. We don't know the joint table of numbers of joint probability table but we have following information.
We know 61% are Master student and 39% are Bachelor students. Out of the Master students, 44% got A, 25% got B and 31% Failed. So with this information only find out given if the student got A, what is the probability that he is a Master student.
In this case, we are asking, $\mathbb{P}(B_1 \mid A_1)$.

Independence

Independence

- The idea of independent events is very easy, we need to check whether joint probability is same as the product of marginal probability.

Theorem 2.28: (Independence of two events)

Let A and B be two events from the sample space Ω, we say A and B are independent if

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

- Here is one interpretation, since multiplication of probabilities will always be smaller, you can think when the events are independent than their joint probability will be very small.
- The idea of independence can also be explained via the conditional probability. This is what we mentioned in page 77. Recall, when A and B are independent we have

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

- So the unconditional probability or the marginal probability of A is same as the conditional probability. So learning B makes no difference to the probability of A.

Independence

- We can easily extend this concept to more than two events, the idea is then we need all subsets of the events to be independent.
- As an example, in order for three events A, B, and C to be independent, the following four relations must be satisfied:

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

and

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

- This idea is what we call *mutual independence*.

Definition 2.29: (Independent Events).

The k events $A_1, \ldots, A_k$ are independent (or mutually independent) if, for every subset the joint probability of the events can be written as a product of marginal probabilities.

- You will see some problems in the problem set.

References

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th edition.