Discrete Probability

Fall 2025

Instructor:

Ajit Rajwade

Topic Overview

- Some important terminology: sample space, event, probability
- Composition of events; mutual exclusion and independence
- Axioms of probability
- Principles of counting
- Conditional probability and Bayes’ theorem
- Some paradoxes!

What’s this topic all about?

- It will help answer questions like:

What’s the chance

- ❖ That a coin toss produces a head?
- ❖ That it produces at least 3 heads in a sequence of 10 tosses?
- ❖ That if you test positive for some disease, you actually have the disease
- ❖ That India will win the next 3 ODIs against Australia?
- ❖ That you will get an AA on all courses despite spending the semester on Whatsapp ☺?

Concept of Sample space

- Consider an experiment whose outcome is not known in advance.
- Example 1: A coin toss
- However, suppose we do know the complete set of possible outcomes – in this case Heads or Tails.
- The set of all possible outcomes of an experiment is called the **sample space**.

Concept of Sample space

- Example 2: Measurement of your body temperature (assume it's an integer) with a thermometer. What's the sample space?

-Let's say between 30 to 40 degrees Celsius, so the sample space = {30,31,…,39,40}

- Example 3: An experiment to randomly choose a student from the CSE 2025 batch at IITB and declare him/her the branch topper

-Sample space = set of all students in that batch

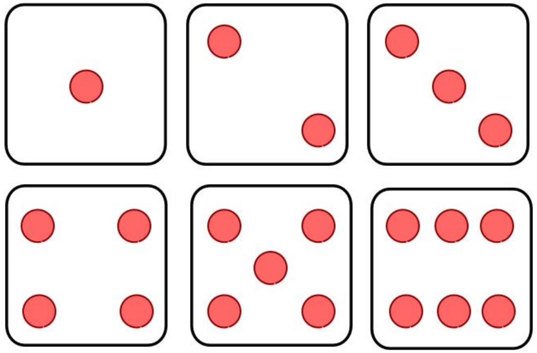
Concept of Sample space

- Example 4: Consider a four-country ODI series between India, Pakistan, Bangladesh and Australia. What is the set of rankings?

-Sample space = set of all permutations of the string IPBA

- Example 5: An experiment to roll a die

-Sample space = {1,2,3,4,5,6}



Digression: definition of a set

- A “set” is an informal English word, but here it has a precise mathematical meaning – you will learn more about it in a discrete structures course.

- A **set** is an *unordered* collections of elements, with each element occurring exactly *once*.

- Example: the set of all integers between 1 and 10 is given as A = {2,3,4,…,8,9}

Digression: definition of a set

- A set may have a finite number of elements (finite set) or an infinite number (infinite set).

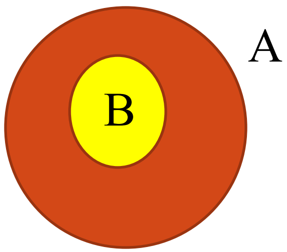
- For now, we will stick to finite sets.

Digression: definition of a subset

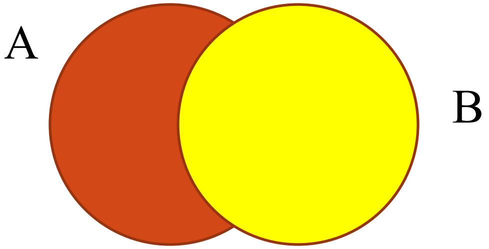
- Set B is called a **subset** of set A if B contains some or all of the elements of A, and nothing else. B can be possibly empty, in which case it is a **null** set or **empty** set.

- Example: the set of all integers between 1 and 10 is given as A = {2,3,4,…,8,9}. Examples of subsets of A - set of all even integers between 1 and 10, set of all integers between 1 and 10 divisible by 11 (this is a null set).

B is a subset of A



B is not a subset of A



Digression: definition of a subset

- Note: every set is a subset of itself!

- If B is a subset of A and B is not equal to A, then B is called a **proper subset** of A.

- "B is a subset of A" is denoted as: $B \subseteq A$

- "B is a proper subset of A" is denoted as: $B \subset A$

Concept of Event

- Any subset of the sample space is called an **event**.
- If the outcome of an experiment is contained in Event E, then we say E has *occurred*.
- In example 1, if $E = \{\mathrm{H}\}$, then E is the event that the coin produced a heads.
- In example 2, if $E = \{$set of temperatures from 33 to 37$\}$, then E is the event that the temperature was "normal" (i.e. not exceeding 37 and not less than 33)

Composition of Events

- Given event E, event E^c is the event that E did not occur. E^c is called the **complement** of E.

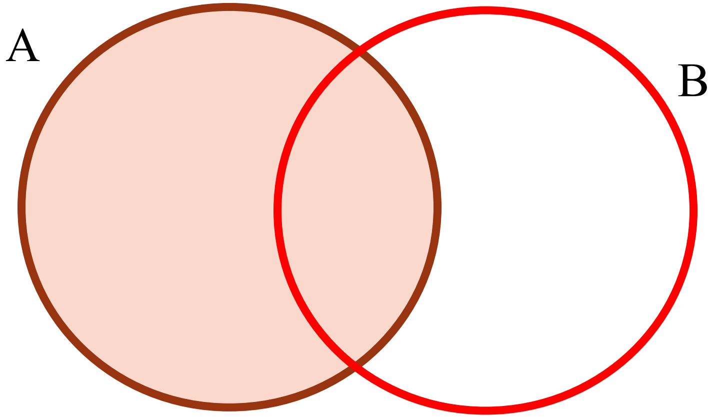
- Given events E and F, the event G that either E or F (or both) occur is called as the **union** of E and F, and denoted as $G = E \cup F$.

- Given events E and F, the event G that both E and F occur is called as the **intersection** of E and F, and denoted as $G = E \cap F$ or $G=EF$.

Composition of Events

- Union and intersection can be extended to handle any arbitrary number of events.

- If two events cannot occur together (for example?), then their intersection is a null set. Such events are called **mutually exclusive**.



This is called as a Venn diagram in set theory. The light shaded region bordered by red and brown curves is AB.

Composition of Events

- An event and its complement – are always mutually exclusive events.

- Let *F* be the event that a patient tests negative for a certain disease in a medical test. Let *G* be the event that (s)he tests positive for the same disease in the same test. Then *F* and *G* are mutually exclusive.

- Let *E* be the event that the sum of three consecutive dice throws was greater than or equal to 3. Let *F* be the event that the sum of three consecutive dice throws was greater than or equal to 4. Then *E* and *F* are NOT mutually exclusive. In fact *F* is a subset of *E*.

Properties of set operations

Commutative laws :

$$A \cup B = B \cup A$$

$$AB = BA$$

Distributive laws :

$$(A \cup B)C = AC \cup AB$$

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

Associative laws :

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

$$(AB)C = A(BC)$$

DeMorgan's laws :

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

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

Properties of set operations

DeMorgan's laws :

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

To prove $A^c \cup B^c = (A \cap B)^c$, we must show that

if any $x \in A^c \cup B^c$, then $x \in (A \cap B)^c$, and conversely .

$x \in A^c \cup B^c \to x \in A^c$ or $x \in B^c$. So $x \notin A \cap B$.

So $x \in (A \cap B)^c$.

Also if $x \in (A \cap B)^c$, then $x \notin A \cap B$, i.e. x cannot

lie in both A and B. So $x \in A^c$ or $x \in B^c$, i.e. $x \in A^c \cup B^c$.

Axioms of probability

- If an experiment is repeated several times under the same conditions, then the number of times a given event *E* occurs reaches some limiting value as the number of trials increases. This limiting frequency forms our notion of probability of the event denoted as *P*(*E*).
- For an event *E* from sample space *S*, we have:

Axiom 1 : $0 \le P(E) \le 1$

Axiom 2 :$P(S) = 1$

Axiom 3 :For mutually exclusive events $E_1, E_2, \ldots, E_n$, we have

$$P(\boxtimes_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$$

This strange symbol is intended to be set union.

Axioms of probability

- The notion of relative frequency of event *E* obeys the aforementioned axioms.
- Properties (can be proved by Venn diagrams):

$$P(A \cup B) = P(A) + P(B) - P(AB) \leq P(A) + P(B)$$

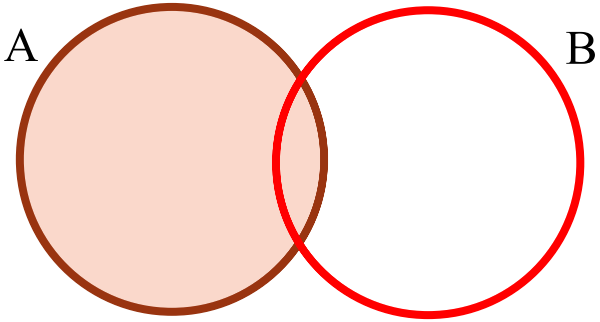
This implies

$$(1) P(A^c) = 1 - P(A)$$

$$(2) A \subseteq B \rightarrow P(A) \leq P(B)$$

Is the converse of (2) also true?

Food for thought: In terms of P(*A*) and P(*B*), what is the probability that exactly one of the events *A* or *B* occurs? That is, we have the condition that both *cannot* occur together.



Equally likely outcomes

- We will assume that each of the singleton outcomes in the sample space is equally likely.
- So, if the experiment is to roll a die, then all six faces will show up with equal probability.
- We will assume finite sample spaces for now.
- In such a case, the probability of an event *E* is given as:

$$P(E) = \frac{\text{Number of points in E}}{\text{Number of points in sample space}}$$

Principles of counting: motivating example

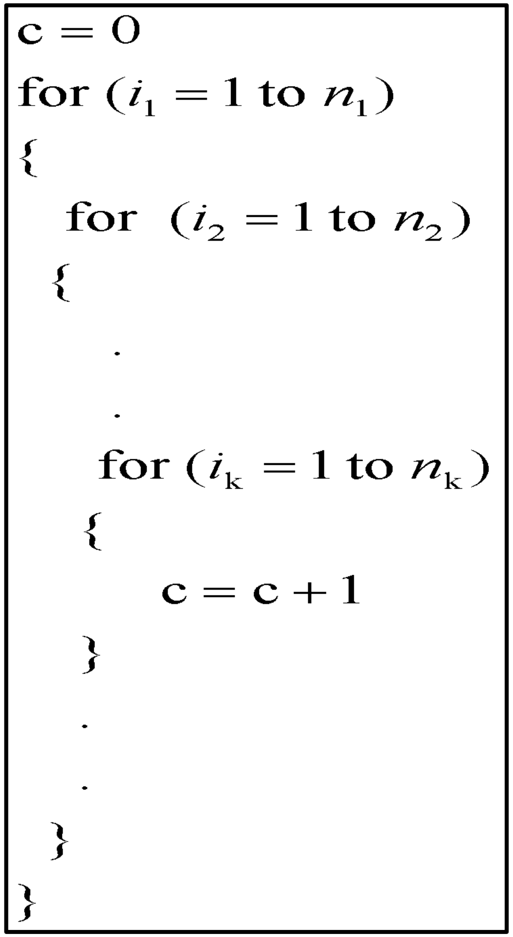
- They come in useful in solving problems on discrete probability.

- For example: Suppose a box contains 6 white and 5 black balls. If you draw two balls at random, what is the probability that one is white and the other is black?

Principles of counting

- **Product rule:** Suppose a procedure can be broken down into a sequence of k tasks. Suppose there are n_1 ways to do task 1, and for each of these there are n_2 ways to do task 2, …, and for each of these there are n_k ways to do task k. Then there are $n_1 n_2 \ldots n_k$ ways to do the entire procedure.

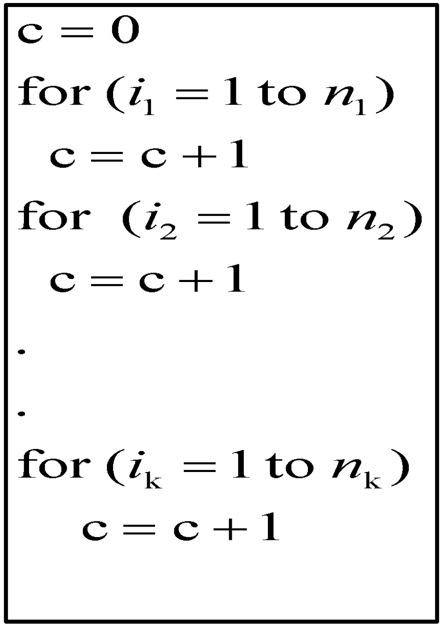
Consider three disjoint lists of project topics containing 20, 30, 40 projects respectively. A student has to pick one topic from each list. What is the total number of choices a student can make for the three projects? 20 x 30 x 40



```
c = 0
for ($i_1$ = 1 to $n_1$)
{
   for ($i_2$ = 1 to $n_2$)
   {
         .
         .
      for ($i_k$ = 1 to $n_k$)
      {
         c = c + 1
      }
      .
      .
   }
}
```

Principles of counting

- **Sum rule:** Suppose a procedure can be done either in n_1 ways or in n_2 ways … or in n_k ways, where the n_1 ways, n_2 ways, …, n_k ways are all completely disjoint. Then there are $n_1 + n_2 + \ldots + n_k$ ways to perform this task.



```
$c = 0$
for ($i_1 = 1$ to $n_1$)
  $c = c + 1$
for  ($i_2 = 1$ to $n_2$)
  $c = c + 1$
.
.
for ($i_k = 1$ to $n_k$)
    $c = c + 1$
```

Consider three disjoint lists of project topics containing 20, 30, 40 projects respectively. A student has to pick any **one** topic. What is the total number of projects the student can choose from? 20 + 30 + 40

Principles of counting: example

- For example: Suppose a box contains 6 white and 5 black balls. If you draw two balls at random, what is the probability that one is white and the other is black?

- There are two scenarios: (1) the first ball is white and second is black, or (2) vice versa.
- For (1), the probability that the white ball is picked is 6/11, and the probability that the black ball is picked is 5/10 (10 balls remain after the first white ball is picked). The overall probability is 30/110 (product rule).
- For (2), the probability that the black ball is picked is 5/11, followed by a 6/10 probability of picking a white ball, leading to an overall probability of 30/110 (product rule).
- The total probability is (30+30)/110 = 6/11 (sum rule).

Some interesting results: Boole's inequality

$$P(A \cup B) = P(A) + P(B) - P(AB) \leq P(A) + P(B)$$

Extension to n events :

$$P(\boxtimes_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

↔ Can you prove this?

This strange symbol is intended to be set union.

Some interesting results: Bonferroni's inequality

Questions such as the following crop up in many different applications:

According to reports by the National Institute of Nutrition, 70% of Indian children in the under 5 age bracket have iron deficiency, and 42.6% of them are underweight. What can you say about the probability that a randomly selected child in this age bracket is deficient in iron as well as underweight?

$$P(A \cup B) \leq P(A) + P(B)$$

At least 12.6%

$$P(A^c \cup B^c) \leq P(A^c) + P(B^c)$$

$$\therefore P((A \cap B)^c) \leq 2 - (P(A) + P(B))$$

$$\therefore 1 - P(A \cap B) \leq 2 - (P(A) + P(B))$$

$$\boxed{\therefore P(A \cap B) \geq P(A) + P(B) - 1}$$

This is called as Bonferroni's inequality

Some interesting results: Bonferroni's inequality

$$P(A \cup B) \le P(A) + P(B)$$

$$P(A^c \cup B^c) \le P(A^c) + P(B^c)$$

$$\therefore P((A \cap B)^c) \le 2 - (P(A) + P(B))$$

$$\therefore 1 - P(A \cap B) \le 2 - (P(A) + P(B))$$

$$\therefore P(A \cap B) \ge P(A) + P(B) - 1$$

This is called as Bonferroni's inequality

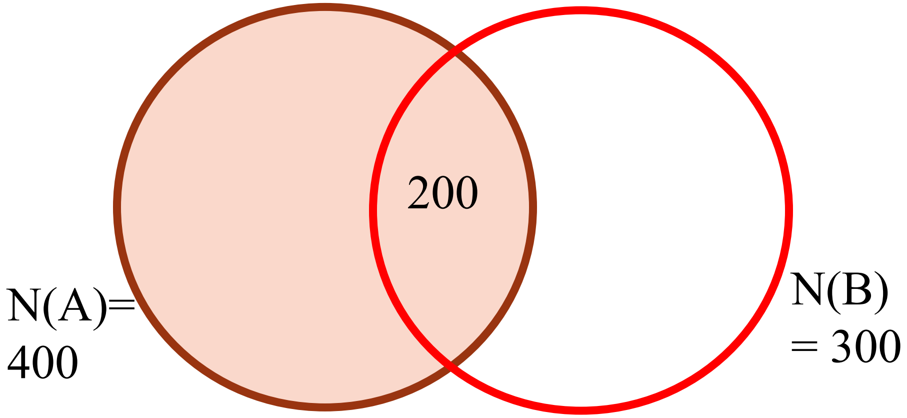
Extension to n events :

$$P(\boxtimes_{i=1}^{n} A_i) \ge 1 - n + \sum_{i=1}^{n} P(A_i)$$

This strange symbol is intended to be set intersection.

Conditional Probability

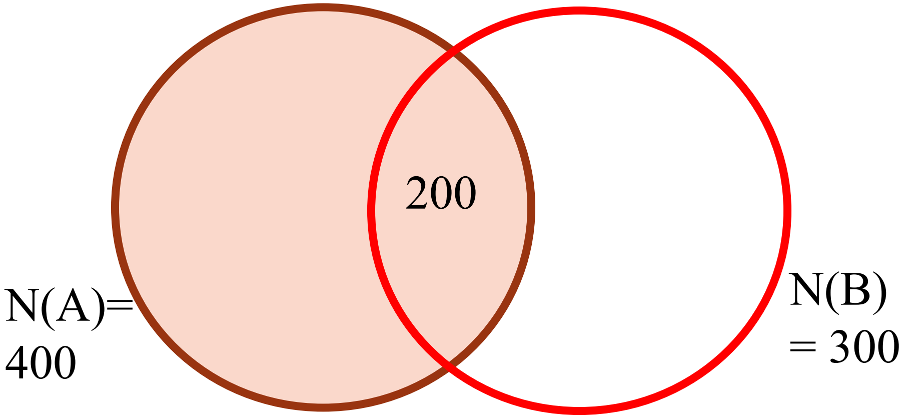
In a group of 1000 people, 400 have tested positive for tuberculosis using a recently designed method, and 300 are previously known to have the disease (maybe using some older method). There are 200 people who are previously known to have the disease *and* who tested positive using the new method. **Given that** person XYZ tests positive, what is the probability that he/she is previously known to have the disease?



We know that 400 people tested positive. Out of these 200 are known to have the disease. So the probability that you have the disease as per the older test given that you test positive with the new method is 0.5.

Conditional Probability

- In a group of 1000 people, 400 have tested positive for tuberculosis using a recently designed method, and 300 are previously known to have the disease (maybe using some older method).
- There are 200 people who are previously known to have the disease *and* who tested positive using the new method.
- *Given that person XYZ tests positive*, what is the probability that he/she is previously known to have the disease?

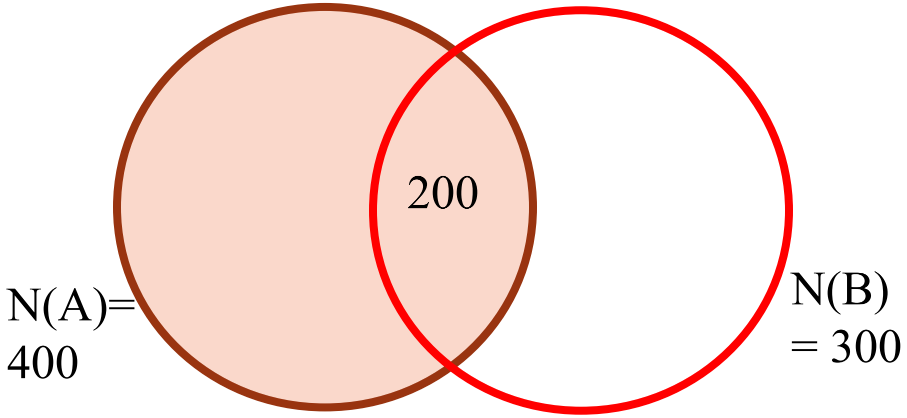


We know that 400 people tested positive. Out of these 200 are known to have the disease. So the probability that you have the disease as per the older test given that you test positive with the new method is 0.5. This is called the **conditional probability** of *B* given *A* and is denoted as P(*B*|*A*). It is calculated as

P(*B*|*A*) = P(*AB*)/P(*A*).

Conditional Probability

- In a group of 1000 people, 400 have tested positive for tuberculosis using a recently designed method, and 300 are previously known to have the disease (maybe using some other method).
- There are 200 people who are previously known to have the disease *and* who tested positive using the new method. *Given that person XYZ is previously known to have the disease*, what is the probability that (s)he will test positive?



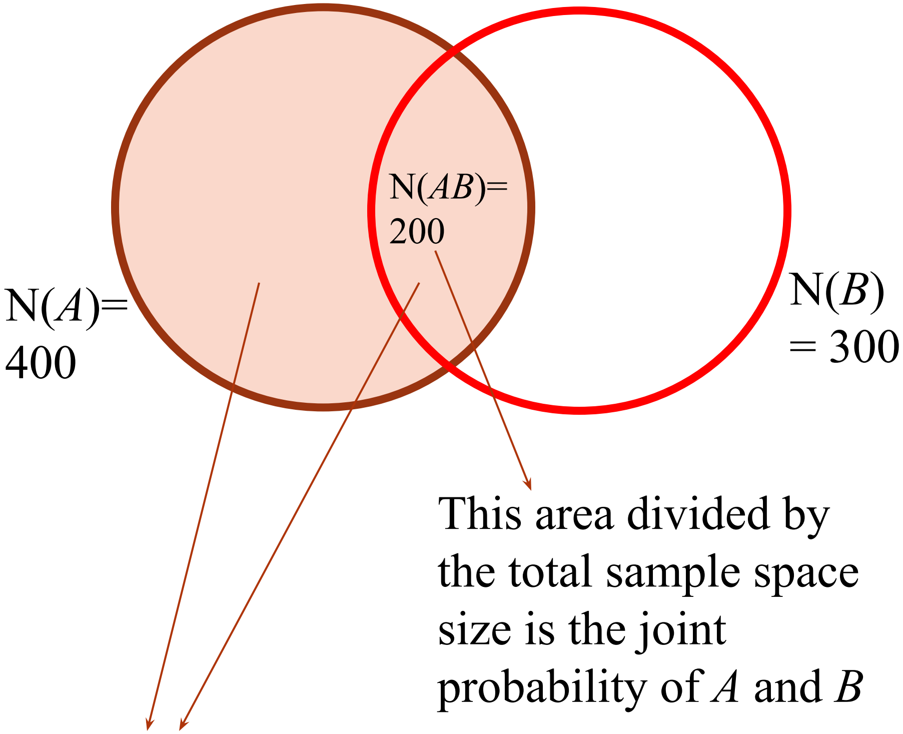
We know that 300 people have the disease. Out of these 200 test positive. So the probability that you test positive given that you already have the disease is 200/300 = 2/3.
$P(A|B) = P(AB)/P(B)$.

Joint probability

- The probability that events *A* and *B* both occur (in the same experiment) is called the **joint probability** of the events *A* and *B*. This is another word for the probability of the *intersection* of *A* and *B*.

- In the previous example, this means the probability of the person testing positive with the new method *and* the older method.

Conditional and Joint probability: what's the difference?



This area divided by the total sample space size is the joint probability of *A* and *B*

The ratio of the smaller area to the larger area (area of the full circle) is the conditional probability of *B* given *A*.

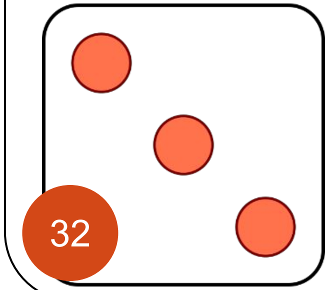
- Let the original sample space be *S*.
- In computing P(*B*|*A*), you assume that *A* has already occurred.
- Therefore your new sample space *S*'' for computing P(*B*|*A*) contains only those events which lie in *A*.
- For computing P(*AB*), the sample space is the entire *S*.
- Now can you compute P(*A*|*B*)?

Conditional and joint probability: what's the difference

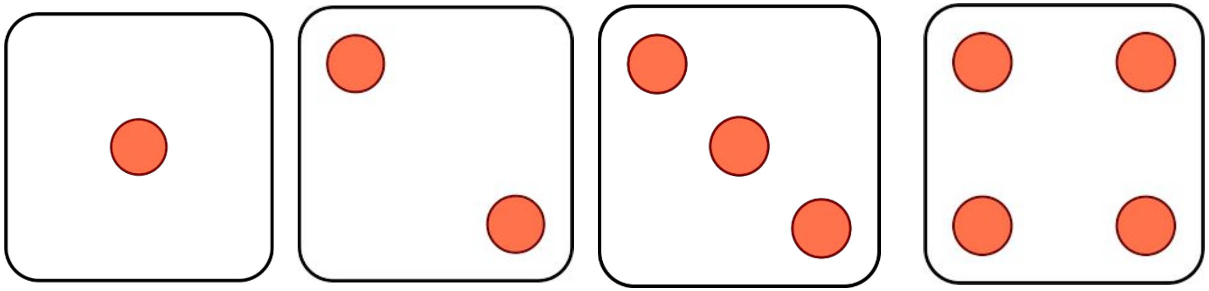
- Consider two consecutive rolls of a die. Given that the first die produced a 3, what's the probability that the sum of the two throws does not exceed 7?

- **Solution:** A = event that first throw produced a 3. $P(A) = 1/6$. B = event that sum does not exceed 7. We want $P(B|A) = P(AB)/P(A)$. Now, $P(AB) = 4/36$. So $P(B|A) = (4/36)/(1/6)=2/3$. Note that $P(B)=21/36$ (why?)

1st roll



2nd roll given that 1st roll is a 3



Conditional and joint probability: what's the difference

- India has a literacy rate of 74%. The state of Kerala has a literacy rate of 94% and constitutes 2.8% of India's population.

- What is the probability that:

❑ A randomly chosen Indian person is literate

❑ A randomly chosen Indian person is from Kerala

❑ A randomly chosen person from Kerala is literate

❑ A randomly chosen Indian person is from Kerala and is literate

❑ A randomly chosen Indian person is from Kerala if you knew already that (s)he was literate

Conditional and joint probability: what's the difference

- India has a literacy rate of 74%. The state of Kerala has a literacy rate of 94% and constitutes 2.8% of India's population.

- What is the probability that:

❑ A randomly chosen Indian person is literate $P(L)=0.74$

❑ A randomly chosen Indian person is from Kerala $P(K)=0.028$

❑ A randomly chosen person from Kerala is literate $P(L|K)=0.94$

❑ A randomly chosen person is from Kerala and is literate

$P(K, L)=P(L|K)P(K)=0.94*0.028$

❑ A randomly chosen person is from Kerala if you knew already that (s)he was literate $P(K|L)=P(K, L)/P(L)=0.94*0.028/0.74$

Concept of independence

- It is said that 40% of the Indian population is infected with TB bacteria (source). 0.02% of the Indian population suffers from oral cancer (source). Let's say that the incidence of TB and oral cancer (OC) have no known connections whatsoever. Given that a person is known to be infected by TB, what is the probability that (s)he is suffering from oral cancer?

- Answer: Since TB and OC have no correlation, $P(TB|OC) = P(TB)$. Also $P(OC|TB) = P(OC)$.

Concept of independence

- Events such as the ones described on the previous slide are called as **independent** events.

- A and B are independent events if and only if $P(A|B) = P(A)$ (equivalently $P(B|A) = P(B)$). That is, if A is independent of B, then B is independent of A.

- Now for independent events, $P(A|B) = P(A)$. But $P(A|B) = P(AB) /P(B)$ which means $P(AB) = P(A) P(B)$.

Concept of independence

- Thus if *A* and *B* are independent, the occurrence of *A* has no bearing on the probability of occurrence of *B* (and vice versa).

Independence of more than two events

- We say that $n > 2$ events are mutually independent if and only if for **every** subset A of $k \leq n$ events, we have:

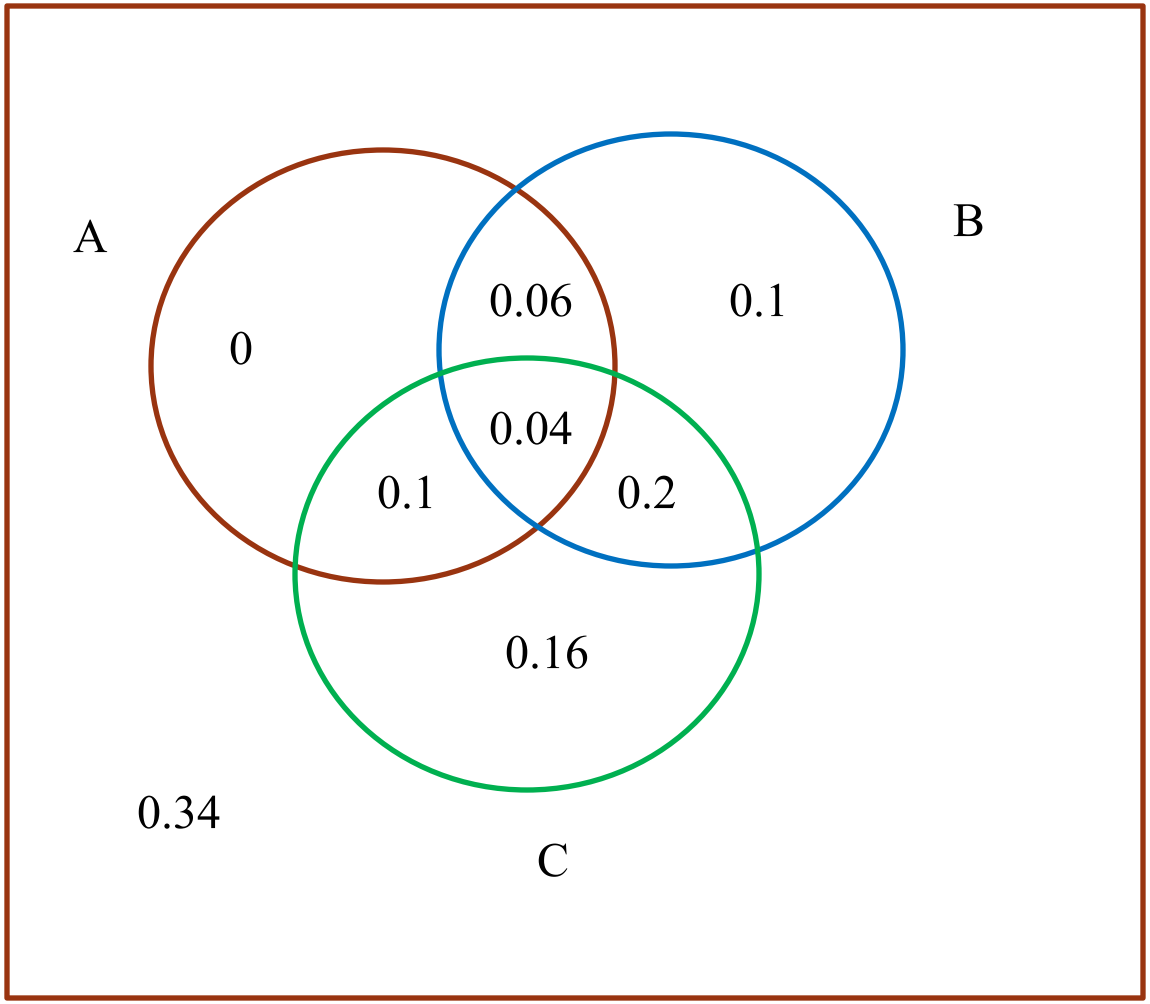
$$P(\mathop{\boxtimes}\limits_{i=1}^{k} A_i) = \prod_{i=1}^{k} P(A_i)$$

This strange symbol is intended to be set intersection.

- Note that only *n*-way independence of events does **not** imply that every pair of events are independent.

- Example: See next slide

http://www.engr.mun.ca/~ggeorge/MathGaz04.pdf



$$P(ABC) = 0.04 = P(A)P(B)P(C) = (0.2)(0.4)(0.5)$$

$$P(AB) = 0.1 \neq P(A)P(B)$$

Independence versus Mutual Exclusion

- If A and B are mutually exclusive, then $P(AB) = 0$.

- If A and B are independent, then $P(AB) = P(A)P(B) \neq 0$.

- The two are usually not the same! In fact, for mutually exclusive events, the occurrence of one *does* have an effect on that of the other.

Independence and mutual exclusion

- If A is independent of B, can we say that A is independent of B^c?

Let' s say A and B are independen t events.

$$A = AB \cup AB^c$$

$$P(A) = P(AB) + P(AB^c) - -why?$$

$$= P(A)P(B) + P(AB^c)$$

$$\therefore P(AB^c) = P(A) - P(A)P(B) = P(A)(1 - P(B))$$

$$= P(A)P(B^c)$$

Bayes Rule

$$P(AB) = P(A \mid B)P(B)$$

$$P(AB) = P(B \mid A)P(A)$$

$$\therefore P(A \mid B) = P(B \mid A)P(A) / P(B)$$

$$Extension:$$

$$P(A) = P(AB) + P(AB^c) =$$

$$P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$$

$$\therefore P(B \mid A) = \frac{P(AB)}{P(AB) + P(AB^c)}$$

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

Example: Bayes rule

- A drug test proposed by a company tests positive 99% of the time on drug consumers, and it tests negative 99% of the time on non-consumers. Let's say the drug is consumed by 0.5% of the people. If a person tests positive for the drug, what is the probability (s)he is a drug consumer?

- Let C = event that a person is a drug consumer.
- Let + = event that a person tests positive.

Example: the false positive paradox

$$P(C) = 0.005, P(C^c) = 0.995$$

$$P(+ \mid C) = 0.99$$

$$P(C \mid +) = P(+ \mid C)P(C) / P(+)$$

$$P(+) = P(+ \mid C)P(C) + P(+ \mid C^c)P(C^c)$$

$$= 0.99 * 0.005 + (1 - 0.99) * 0.995$$

$$P(C \mid +) = \frac{0.99 * 0.005}{0.99 * 0.005 + 0.01 * 0.995} \approx 33.22\%$$

An individual testing positive is most likely not a consumer – despite the apparent (99%) accuracy of the test! This is because the number of drug consumers is small and hence the factor 0.995 outweighs the consumer probability! This is called the **false positive paradox.** For fewer false positives, we need more than 99% accuracy on non-consumers (Eg: 99.99% yields P(C|+) = 0.5%).

More such examples on wikipedia.

The Birthday Paradox!

- Given n people in a room, what should be the least value of n such that the probability that at least 2 people in the room share the same birthday is greater than or equal to 99.9%?

- Each person can have his/her birthday on any of the 365 days. For n people, there are 365^n outcomes.

- The number of outcomes resulting in no two people sharing a birthday is (365)(364)(363)…($365-n+1$).

The Birthday Paradox!

- So required probability is

$1\text{-}(365)(364)(363)\ldots(365\text{-}n\text{+}1)/(365)^n \geq 0.999$ (given)

- This is satisfied for *n* as small as 70.
- For *n* = 20, it is around 41%.
- For *n* = 40, it is around 89%.

- For more information see the wikipedia article on the birthday paradox.