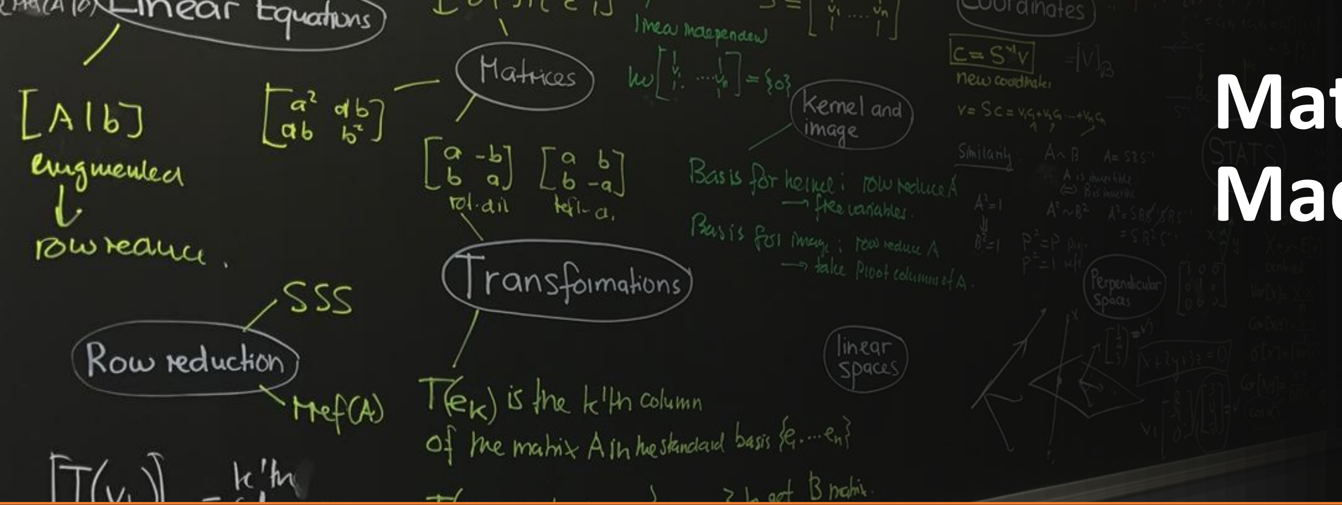

Mathematical Foundations for Machine Learning and Data Science

Probability Theory

Dr. Zubair Khalid

Department of Electrical Engineering
School of Science and Engineering
Lahore University of Management Sciences

https://www.zubairkhalid.org/ee212_2021.html

LUMS
A Not-for-Profit University

Outline

- *Motivation*
- *Overview of sets*
- *Probability Models*
- *Axioms of Probability*
- *Conditional Probability*
- *Independence*
- *Combinatorics*
- *Binomial Probability*

Probability Theory - Applications

- Speech Recognition System

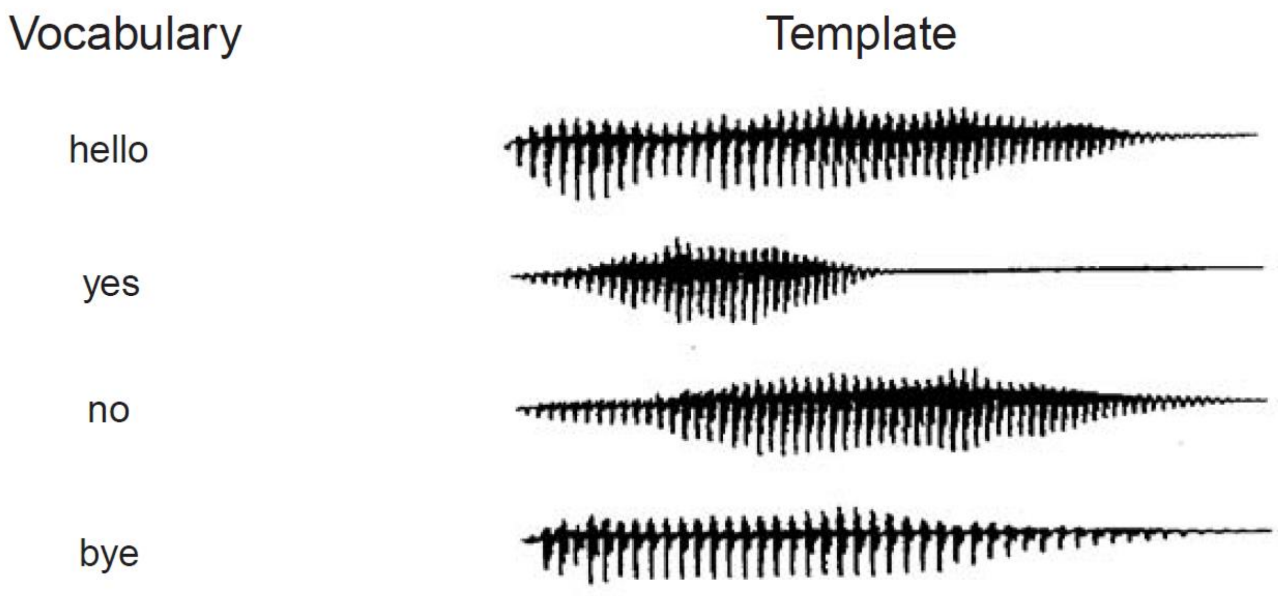

Fig 1: Speech templates

Probability Theory - Applications

- Speech Recognition System
  - There exists uncertainty in the templates of different speakers
  - Probability theory addresses the uncertainty

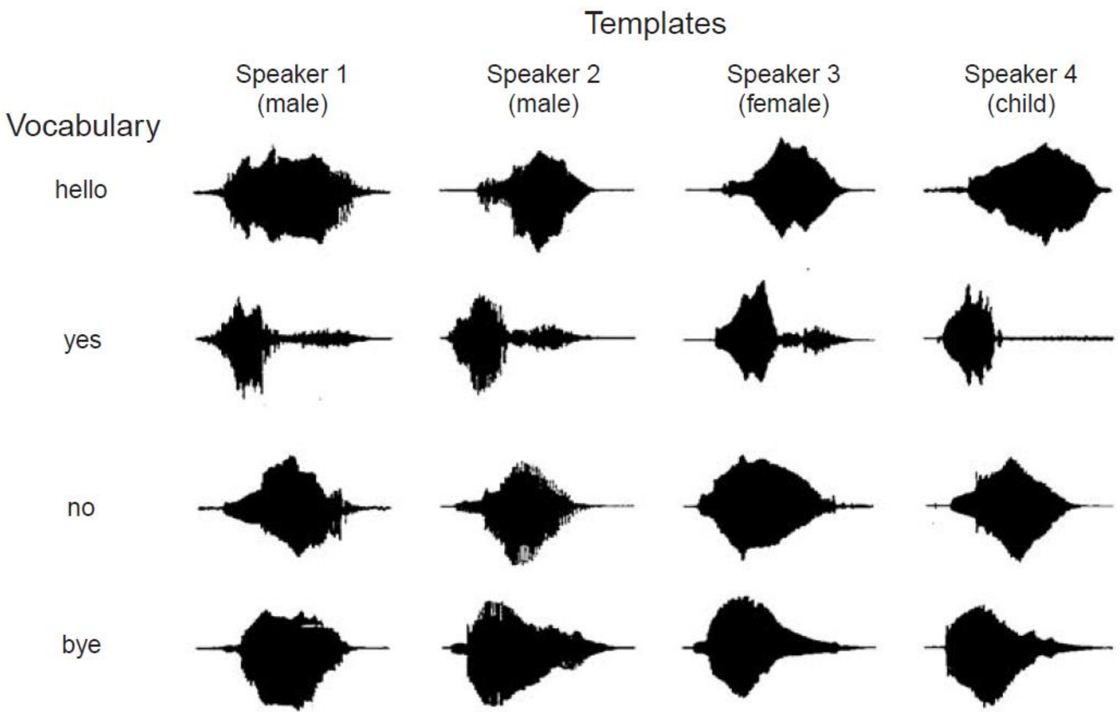

Fig 2: Variation in templates for different speakers

Probability Theory - Applications

- Radar Detection
  - Uncertainty involves in the presence of the target
  - Estimates the target
  - Probability theory deals with uncertainty in the estimation

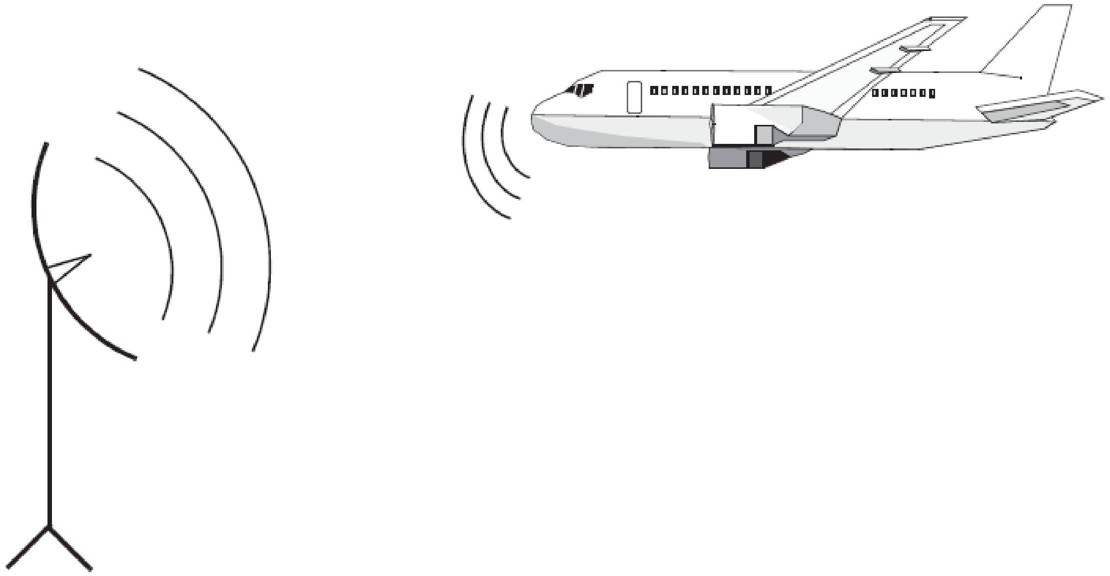

Fig 3: Radar detection system

Probability Theory - Applications

- Wireless Communication System
  - Noise modeling
  - Interference
  - Power estimation

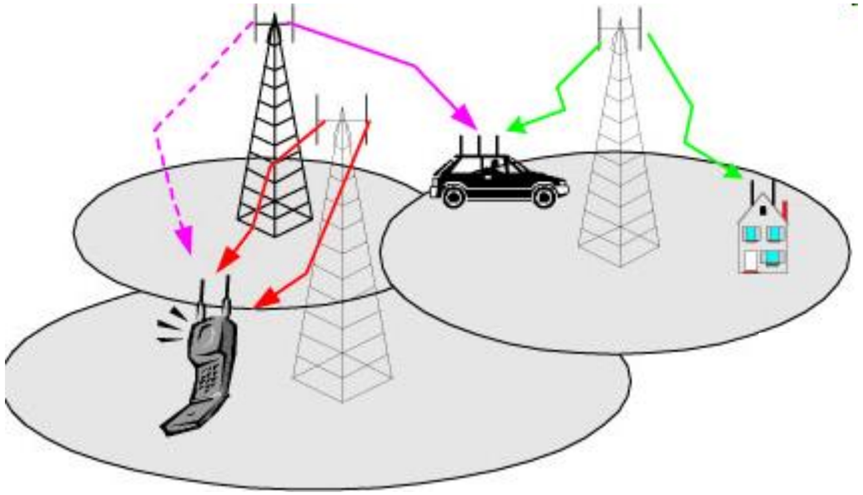

Fig 4: Wireless communication system

Basic Definitions (1)

Relative frequency

- Consider an experiment that can result in M possible outcomes O_1 , O_2 … O_M

- Let $N_n(O_i)$ denotes the number of times ,O_i occurred in n trials

- Relative frequency of outcome: $\frac{N_n(O_i)}{n}$

- When number of trials n becomes large, the relative frequency converge to some limiting value.

- This behaviour is known as statistical regularity.

Basic Definitions (2)

- **Sample Space:** set of all possible distinct **outcomes** of an experiment.
  - Outcome: ω
  - Sample space: Ω

- **Events:** Collection of outcomes are called events. Usually denoted by capital letters. Every event is a subset of sample space.

Examples:

1. Rolling two dies together.
   - sample space?
   - event: the sum of numbers on two dies = *6*?

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Sample space

2. Noise voltage can be between *0* and *5* volts.
   1. sample space?
   2. event: the noise voltage is between *2* and *3* volts?

Language of Sets – Review

- $\omega \in \Omega$, $A \subset \Omega$. (A is the event).
- Events A and B are equal, i.e., $A = B$ if $A \subset B$ and $B \subset A$.
- A^c: Complement of A. Given by $A^c := \{\omega \in \Omega : \omega \notin A\}.$
- Empty set or Null set is denoted by Ø and represents no point in Ω .
- Union of two sets: $A \cup B := \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}.$
- Intersection of two sets: $A \cap B := \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$
- Disjoint or mutually exclusive sets: $A \cap B = \varnothing$
- Set difference operation: $B \setminus A := B \cap A^c$

LUMS
A Not-for-Profit University

Set Identities

- Associative laws:

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

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

- Distributive laws:

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

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

- De Morgan's laws:

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

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

Venn Diagram

- Graphical representation of relationship between sets
- Rectangle represents sample space
- Ellipsoids represents events

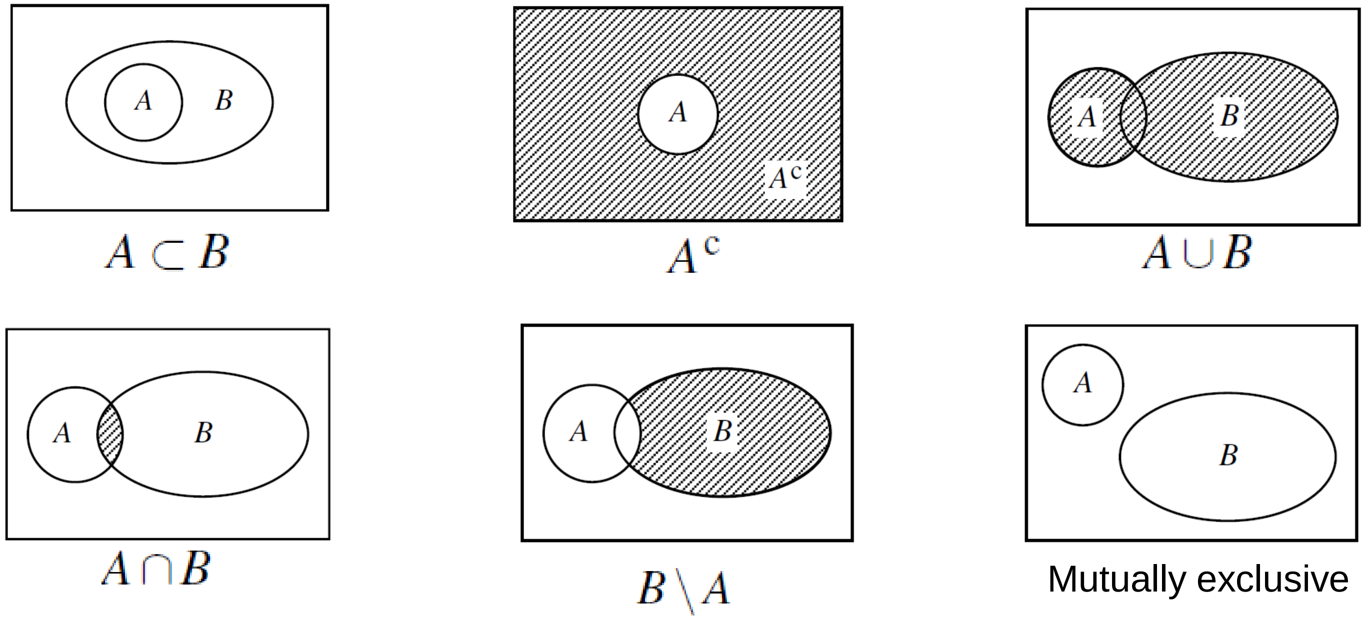

Fig 5: Venn diagram illustration

Countable and Uncountable Sets

- |*A*| denotes the number of points in set |*A*| and is called cardinality of set *A.*
- A nonempty set *A* is said to be countable if elements of *A* can be enumerated;

$$A = \bigcup_{k=1}^{\infty} \{a_k\}$$

- Every finite set is countable (No requirement that a_k to be distinct).

Example: *A* = {a,e,i,o,u}.

- Empty set is also countable.
- If the set is not finite, we call it countably infinite. Examples?
- If the set is not countable, we call it uncountable or uncountably infinite.

Probability Models

- Mathematical modelling of the problem.
- Basic model if each outcome is equally likely: $P(A) = \frac{|A|}{|\Omega|}$
- Example: Fair die
- If outcomes are not equally likely;

$$P(A) = \frac{|A|}{|\Omega|} = \sum_{\omega \in A} \frac{1}{|\Omega|} = \sum_{\omega \in A} p(\omega)$$

- $p(\omega)$ probability for each outcome $\omega \in \Omega$

Properties:

$P(\emptyset) = 0$

$P(A) \geq 0$

if $P(A \cap B) = \emptyset$, then
$P(A \cup B) = P(A) + P(B)$

$P(\Omega) = 1$

Probability Models

***Example* 1.10.** Construct a sample space Ω and probability P to model an unfair die in which faces 1–5 are equally likely, but face 6 has probability $1/3$. Using this model, compute the probability that a toss results in a face showing an even number of dots.

Probability Models

***Example* 1.12.** A single card is drawn at random from a well-shuffled deck of playing cards. Find the probability of drawing an ace. Also find the probability of drawing a face card.

Probability Models

***Example* 1.13.** Suppose that we have two well-shuffled decks of cards, and we draw one card at random from each deck. What is the probability of drawing the ace of spades followed by the jack of hearts? What is the probability of drawing an ace and a jack (in either order)?

Axioms of Probability

Given a nonempty set Ω, called the sample space, and a function P defined on the subsets of Ω, we say P is a probability measure if the following four axioms are satisfied:

1. The empty set $\emptyset$ is called the impossible event. $P(\emptyset) = 0$.
2. For any event $A \subset \Omega, P(A) \geq 0$.
3. If $A1$, $A2$, ... are events that are mutually exclusive, that is, $A_n \cap A_m = \emptyset$ for $n \neq m$, then

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n).$$

4. $P(\Omega) = 1$, Ω is a sure event.

Properties of Probability

- Finite version of axiom 3:

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n).$$

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

- $A \subset B \Rightarrow P(A) \leq P(B)$ — Monotonicity property

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ — Inclusion-exclusion property

Conditional Probability

- **Motivation:** Conditional probability, important concept in probabilistic modeling, allows us to update probabilistic models when additional information is revealed.
- Probability of event A given event B;

$$P(A|B) = \frac{\text{outcomes in A and B}}{\text{outcomes in B}} = \frac{\text{outcomes in A and B}}{\text{total}} \frac{\text{total}}{\text{outcomes in B}} = \frac{P(A \cap B)}{P(B)}$$

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

- **The law of total probability**

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

- **Baye's Rule**

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

Examples

Law of total probability

Example:

Due to an Internet configuration error, packets sent from Karachi to Quetta are routed through Lahore with probability $3/4$. Given that a packet is routed through Lahore, suppose it has conditional probability $1/3$ of being dropped. Given that a packet is not routed through Lahore, suppose it has conditional probability $1/4$ of being dropped. Find the probability that a packet is dropped.

Examples

Baye's Rule

Example:
Find the conditional probability that a packet is routed through Lahore given that it is not dropped.

Generalized Laws

- **The law of total probability:**

If $B1, B_2, \ldots$ denotes a sequence of pairwise mutually exclusive sets and $\sum_n P(B_n) = 1$, then,

$$P(A) = \sum_n P(A|B_n)P(B_n)$$

- **Baye's Rule**

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_n P(A|B_n)P(B_n)}$$

Next

- Independence
- Combinatorics
- Examples
- Binomial Probabilities

Independence

In probability theory, if events A and B satisfy $\mathsf{P}(A|B) = \mathsf{P}(A|B^c)$, we say A does not depend on B. This condition says that

$$\Rightarrow \qquad \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A \cap B^c)}{\mathsf{P}(B^c)}$$

Using $\mathsf{P}(A) = \mathsf{P}(A \cap B) + \mathsf{P}(A \cap B^c)$ and $\mathsf{P}(B^c) = 1 - \mathsf{P}(B)$

$$\Rightarrow \qquad \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A) - \mathsf{P}(A \cap B)}{1 - \mathsf{P}(B)}$$

$$\Rightarrow \qquad \mathsf{P}(A \cap B)[1 - \mathsf{P}(B)] = \mathsf{P}(B)[\mathsf{P}(A) - \mathsf{P}(A \cap B)]$$

$$\Rightarrow \qquad \mathsf{P}(A \cap B) - \mathsf{P}(A \cap B)\mathsf{P}(B) = \mathsf{P}(A)\mathsf{P}(B) - \mathsf{P}(A \cap B)\mathsf{P}(B)$$

$$\Rightarrow \qquad \boxed{\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)}$$

Independence

- Mutually exclusive events not to be confused with independent events. **(Different)**

- Interpretation of independence between events:

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A)\,\mathsf{P}(B)}{\mathsf{P}(B)} = \mathsf{P}(A)$$

- If A and B are independent events, A^c and B, A and B^c, A^c and B^c are also independent events.

- Independence of more than 2 events:

$$\mathsf{P}(A_i \cap A_j \cap A_k) = \mathsf{P}(A_i)\,\mathsf{P}(A_j)\,\mathsf{P}(A_k)$$

- Mutually independent

- Pairwise independent

- Mutually independence implies pairwise independence but not the other way.

Independence

***Example* 1.23.** An Internet packet travels from its source to router 1, from router 1 to router 2, and from router 2 to its destination. If routers drop packets independently with probability p, what is the probability that a packet is successfully transmitted from its source to its destination?

Combinatorics

- Combinatorics: branch of mathematics that deals with systematic counting and arrangement methods.

- Counting problems:
  - ordered sampling with replacement
  - ordered sampling without replacement
  - unordered sampling without replacement

Ordered sampling with replacement

***Example* 1.28.** Let A, B, and C be finite sets. How many triples are there of the form (a,b,c), where $a \in A$, $b \in B$, and $c \in C$?

***Solution*.** Since there are $|A|$ choices for a, $|B|$ choices for b, and $|C|$ choices for c, the total number of triples is $|A| \cdot |B| \cdot |C|$.

Similar reasoning shows that for k finite sets $A_1, \ldots, A_k$, there are $|A_1| \cdots |A_k|$ k-tuples of the form $(a_1, \ldots, a_k)$ where each $a_i \in A_i$.

If $|A_1| \ldots |A_k| = n$, there are n^k number of k- tuples.

Ordered sampling without replacement

- Means that the item chosen from any set is not replaced. The choice from the group affects the remaining choices.

***Example* 1.32** (ordered sampling without replacement). A computer virus erases files from a disk drive in random order. If there are n files on the disk, in how many different orders can $k \le n$ files be erased from the drive?

Solution. There are n choices for the first file to be erased, $n-1$ for the second, and so on. Hence, there are

$$n(n-1)\cdots(n-[k-1]) = \frac{n!}{(n-k)!}$$

different orders in which files can be erased from the disk.

Given a set A, we let A^k denote the set of all k-tuples $(a_1,\ldots,a_k)$ where each $a_i \in A$. We denote by A_*^k the subset of all k-tuples with *distinct* entries. If $|A| = n$, then $|A^k| = |A|^k = n^k$, and $|A_*^k| = n!/(n-k)!$.

- Often stated as: The number of **permutations** of k items chosen from n items.

Unordered sampling without replacement

- Too complicated in the book.
- Unordered; when order is not important: (1,2,3) is same as (3,2,1) or (2,1,3)…
- Given n elements and choose k tuples, we have: $\frac{n!}{(n-k)!}$
- Out of $\frac{n!}{(n-k)!}$ tuples, each k- tuple have k! same arrangement or permutation.
- Total k- tuples with different permutation: $\frac{n!}{k!(n-k)!} = \binom{n}{k}$
- Often stated as: The number of **combinations** of k items chosen from n items.

Combinatorics - Summary

- **Draw sample size of *k* out of *n* objects.**

Method	No. of outcomes
ordered samples **with** replacement	n^k
ordered samples **without** replacement	$\frac{n!}{(n-k)!}$
unordered samples **without** replacement	$\frac{n!}{k!(n-k)!}$

Combinatorics - Examples

Problem 76

In a pick-4 lottery game, a player selects four digits, each one from $0, \ldots, 9$. If the four digits selected by the player match the random four digits of the lottery drawing in any order, the player wins. If the player has selected four distinct digits, what is the probability of winning?

Binomial Probabilities

- When there are only two events, win or loss, success or failure.

- Examples:
  - Tossing of a coin: head (win) or tail (loss)
  - Rolling die: getting 4 (win) or getting other than 4 (loss)

***Example* 1.43** (binomial probabilities). A certain coin has probability p of turning up heads. If the coin is tossed n times, what is the probability that k of the n tosses result in heads? Assume tosses are independent.

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Binomial Probabilities

Example

The probability of a student to score 90% in EE 212 is 0.7 (**Assume it is true**). If 7 students appear in an exam, find the probability of exactly 4 getting 90% marks.

Next

- Discrete random variable
- Probability mass function
- Multiple random variable
- Independence of random variables
- Joint probability mass function
- Expectation