Lecture 13: Introduction to Probability

CS 5002: Discrete Math

Tamara Bonaci, Adrienne Slaughter

Northeastern University

December 6, 2018

What is a Tree?

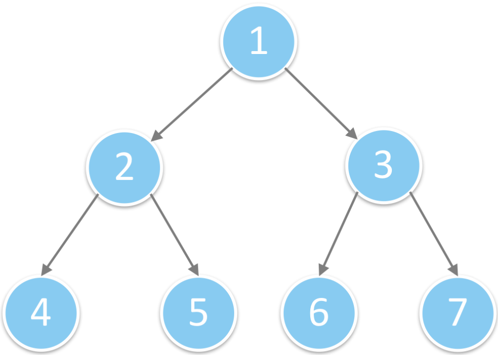
  

Tree - a directed, acyclic structure of linked nodes

- **Directed** - one-way links between nodes (no cycles)
- **Acyclic** - no path wraps back around to the same node twice (typically represents hierarchical data)

Tree Terminology

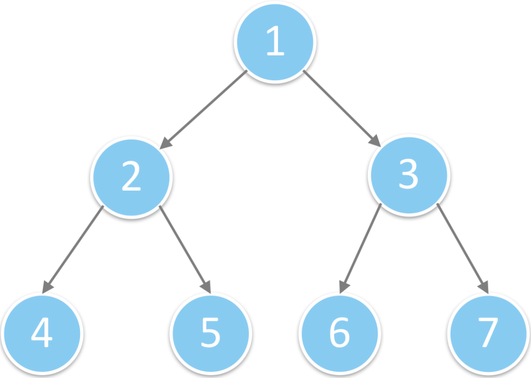
- **Tree** - a directed, acyclic structure of linked nodes
- **Node** - an object containing a data value and links to other nodes
- **Edge** - directed link, representing relationships between nodes
  - All the blue circles



Special Trees

- Binary Tree
- Binary Search Tree
- Balanced Tree
- Binary Heap/Priority Queue
- Red-Black Tree

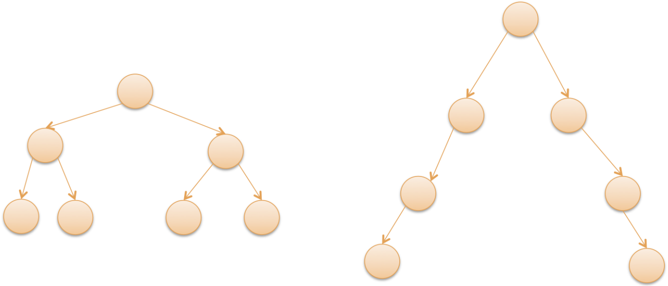
Binary tree - a tree where every node has at most two children



Binary Search Trees

- **Binary search tree (BST)** - a tree where nodes are organized in a **sorted** order to make it easier to search
- At every node, you are guaranteed:
- All nodes rooted at the left child are smaller than the current node value
- All nodes rooted at the right child are smaller than the current node value

Balancing Tree

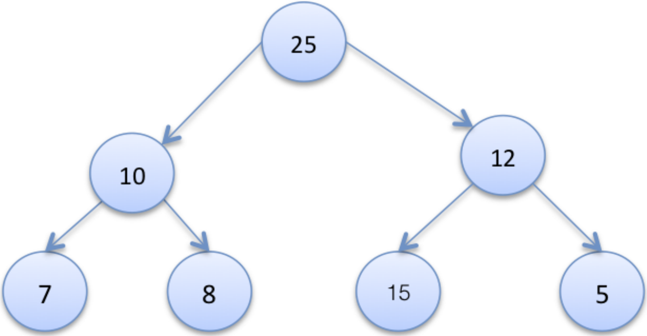


- **Observation:** it is not enough to balance only root, all nodes should be balanced.
- **The balancing condition**: the heights of all left and right subtrees differ by at most 1

BFS Example

Find element with value 15 in the tree below using BFS.

- BFS: traverse all of the nodes on the same level first, and then move on to the next (lower) level

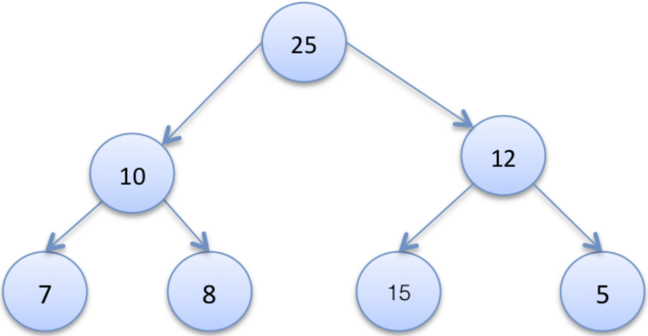


25 – 10 – 12 – 7 – 8 – 15 – 5

DFS Example

Find element with value 15 in the tree below using DFS.

- DFS: traverse one side of the tree all the way to the leaves, followed by the other side

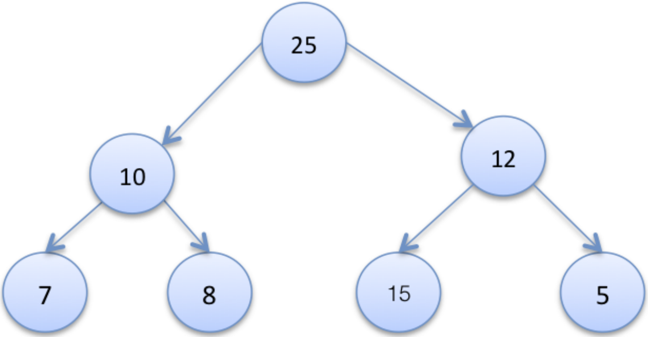


25 – 10 – 7 –8 – 12 – 15 – 5

Tree Traversals Example

Traverse the tree below, using:

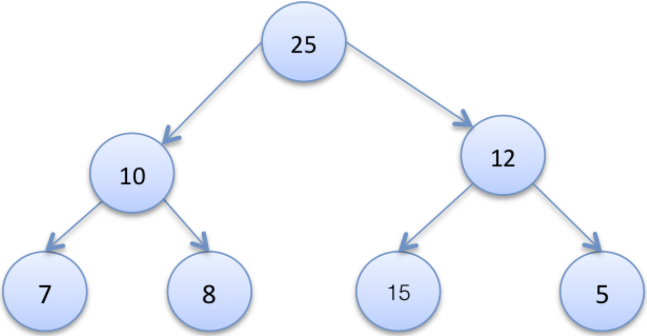
- Pre-order traversal: 25 – 10 – 7 – 8 – 12 – 15 – 5



Tree Traversals Example

Traverse the tree below, using:

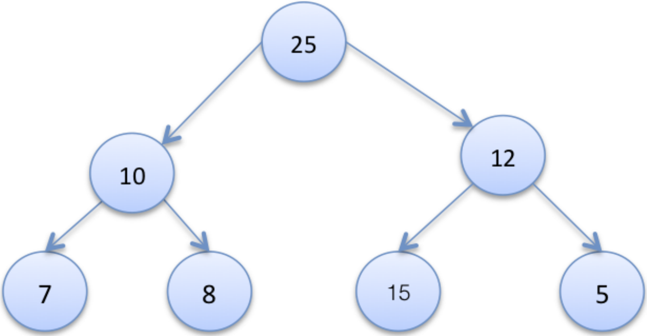
- Pre-order traversal: 25 – 10 – 7 – 8 – 12 – 15 – 5
- In-order traversal: 7 – 10 – 8 – 25 – 15 – 12 – 5



Tree Traversals Example

Traverse the tree below, using:

- Pre-order traversal: 25 – 10 – 7 – 8 – 12 – 15 – 5
- In-order traversal: 7 – 10 – 8 – 25 – 15 – 12 – 5
- Post-order traversal: 7 – 8 –10 – 15 –5 – 12 – 25



What is Graph?

Formal Definition:

- A graph G is a pair (V, E) where
- V is a set of vertices or nodes
- E is a set of edges that connect vertices

Simply put:

- A graph is a collection of nodes (vertices) and edges
- Linked lists, trees, and heaps are all special cases of graphs

Terminology: Undirected Graph

- Two vertices u and v are ***adjacent*** in an undirected graph G if $\{u, v\}$ is an edge in G
  - edge $e = \{u, v\}$ is ***incident*** with vertex u and vertex v
- The ***degree*** of a vertex in an undirected graph is the number of edges incident with it
  - a self-loop counts twice (both ends count)
  - denoted with $deg(v)$

Terminology: Directed Graph

- Vertex u is ***adjacent to*** vertex v in a directed graph G if (u, v) is an edge in G
  - vertex u is the initial vertex of (u, v)
- Vertex v is ***adjacent from*** vertex u
  - vertex v is the ***terminal*** (or end) vertex of (u, v)
- Degree
  - ***in-degree*** is the number of edges with the vertex as the terminal vertex
  - ***out-degree*** is the number of edges with the vertex as the initial vertex

Kinds of Graphs

- directed vs undirected
- weighted vs unweighted
- simple vs non-simple
- sparse vs dense
- cyclic vs acyclic
- labeled vs unlabeled

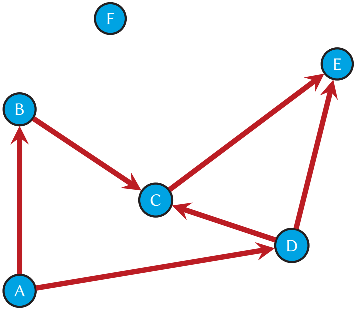
Graph Representations

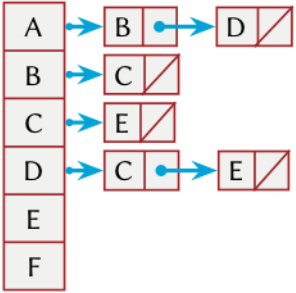
Two ways to represent a graph in code:

- **Adjacency List**
  - A list of nodes
  - Every node has a list of adjacent nodes
- **Adjacency Matrix**
  - A matrix has a column and a row to represent every node
  - All entries are 0 by default
  - An entry $G[u, v]$ is 1 if there is an edge from node u to v

Adjacency List

For each v in V, $L(v)$ = list of w such that (v, w) is in E:



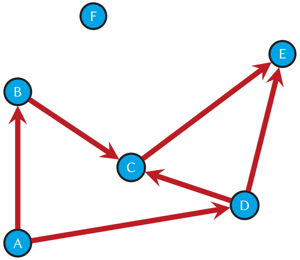


Storage space:

$$a|V| + b|E|$$

$$a = \text{sizeof}(node)$$

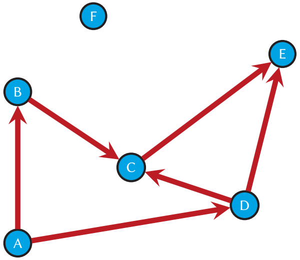
$$b = \text{sizeof}(\text{ linked list element})$$



$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{cccccc} A & B & C & D & E & F \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Storage space: $|V|^2$

Adjacency Matrix



$$
\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \\ F \end{array}
\begin{array}{c}
\begin{array}{cccccc} A & B & C & D & E & F \end{array} \\
\begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
\end{array}
$$

Storage space: $|V|^2$

Does this matrix represent a directed or undirected graph?

Comparing Matrix vs List

1. Faster to test if (x, y) is in a graph?
2. Faster to find the degree of a vertex?
3. Less memory on small graphs?
4. Less memory on big graphs?
5. Edge insertion or deletion?
6. Faster to traverse the graph?
7. Better for most problems?

1. adjacency matrix
2. adjacency list
3. adjacency list (m+n) vs (n2)
4. adjacency matrices (a little)
5. adjacency matrices O(1) vs O(d)
6. adjacency list
7. adjacency list

Analyzing Graph Algorithms

- Space and time are analyzed in terms of:
  - Number of vertices $m = |V|$
  - Number of edges $n = |E|$
- Aim for polynomial running times.
- But: is $O(m^2)$ or $O(n^3)$ a better running time?
  - depends on what the relation is between n and m
  - the number of edges m can be at most $n^2 \leq n^2$.
  - connected graphs have at least $m \geq n - 1$ edges
- Stil do not know which of two running times (such as m^2 and n^3) are better,
- Goal: implement the basic graph search algorithms in time $O(m + n)$.
  - This is linear time, since it takes $O(m + n)$ time simply to read the input.
- Note that when we work with connected graphs, a running time of $O(m + n)$ is the same as $O(m)$, since $m \geq n - 1$.

Single-Source Shortest Path

Input Directed graph with non-negative weighted edges, a starting node s and a destination node d

Problem Starting at the given node s, find the path with the lowest total edge weight to node d

Example A map with cities as nodes and the edges are distances between the cities. Find the shortest distance between city 1 and city 2.

Djikstra's Algorithm: Overview

- Find the "cheapest" node— the node you can get to in the shortest amount of time.
- Update the costs of the neighbors of this node.
- Repeat until you've done this for each node.
- Calculate the final path.

Djikstra's Algorithm: Formally

```
DJIKSTRA(G, w, s)
INITIALIZE-SINGLE-SOURCE(G, s)
S = ∅
Q = G.V
while Q ≠ ∅
    u = EXTRACT-MIN(Q)
    S = S ∪ {u}
    for each vertex v ∈ G.Adj[u]
        RELAX (u, v, w)
```

DJIKSTRA(G, w, s)

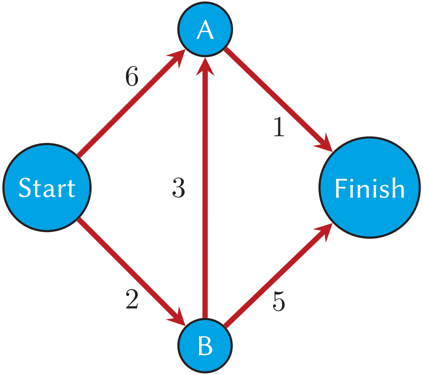
▷ G is a graph
▷ w is the weighting function such that $w(u, v)$ returns the weight of the
▷ s is the starting node
for each vertex $u \in G$
    $u.d = w(s, u)$ ▷ where $w(s, u) = \infty$ if there is no edge (s, u).
$S = \emptyset$ ▷ Nodes we know the distance to
$Q = G.V$ ▷ min-PriorityQueue starting with all our nodes, ordered by dis
while $Q \neq \emptyset$
    $u =$ EXTRACT-MIN(Q) ▷ Greedy step: get the closest node
    $S = S \cup \{u\}$ ▷ Set of nodes that have shortest-path-distance found
    for each vertex $v \in G.Adj[u]$
        RELAX (u, v, w)

RELAX(u, v, w)

▷ u is the start node
▷ v is the destination node
▷ w is the weight function

Djikstra's: A walkthrough

- Find the "cheapest" node— the node you can get to in the shortest amount of time.
- Update the costs of the neighbors of this node.
- Repeat until you've done this for each node.
- Calculate the final path.

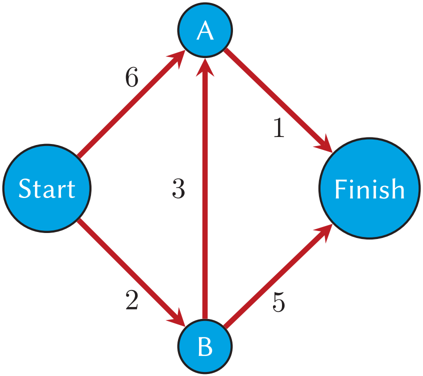


Breadth First Search: distance = 7

Step 1: Find the cheapest node

1. Should we go to A or B?
   - Make a table of how long it takes to get to each node from this node.
   - We don't know how long it takes to get to Finish, so we just say infinity for now.

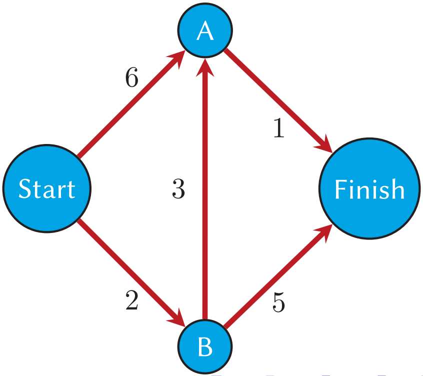
Node	Time to Node
A	6
B	2
Finish	∞



Step 2: Take the next step

1. Calculate how long it takes to get (from Start) to B's neighbors by following an edge from B
   - We chose B because it's the fastest to get to.
   - Assume we started at Start, went to B, and then now we're updating Time to Nodes.

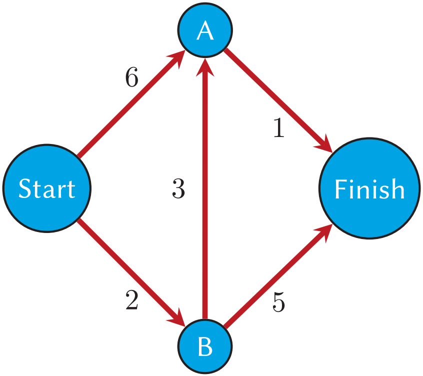
Node	Time to Node
A	$\cancel{6}5$
B	2
Finish	$\cancel{\infty}$ 7



Step 3: Repeat!

1. Find the node that takes the least amount of time to get to.
   - We already did B, so let's do A.
   - Update the costs of A's neighbors
     - Takes 5 to get to A; 1 more to get to Finish

Node	Time to Node
A	$\cancel{6}5$
B	2
Finish	$\cancel{7}6$



Section 2

Random Experiments and Sample Spaces

Random Experiment

Definition

Random experiment is an experiment whose outcome cannot be determined in advance, because it is unknown.

Some examples:

- Tossing a die

Random Experiment

Definition

Random experiment is an experiment whose outcome cannot be determined in advance, because it is unknown.

Some examples:

- Tossing a die
- Measuring the amount of rainfall in Seattle in December

Random Experiment

Definition

Random experiment is an experiment whose outcome cannot be determined in advance, because it is unknown.

Some examples:

- Tossing a die
- Measuring the amount of rainfall in Seattle in December
- Counting the number of sunny days in Seattle in December

Random Experiment

Definition

Random experiment is an experiment whose outcome cannot be determined in advance, because it is unknown.

Some examples:

- Tossing a die
- Measuring the amount of rainfall in Seattle in December
- Counting the number of sunny days in Seattle in December
- Selecting a random sample of CS 5002 students, and observing who is left-handed

Random Experiment

Definition

Random experiment is an experiment whose outcome cannot be determined in advance, because it is unknown.

Some examples:

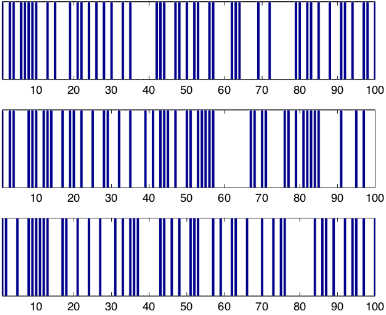
- Tossing a die
- Measuring the amount of rainfall in Seattle in December
- Counting the number of sunny days in Seattle in December
- Selecting a random sample of CS 5002 students, and observing who is left-handed
- Randomly choosing ten people, and recording which languages they speak

Random Experiment - Coin Tossing

Coin tossing

- The most fundamental stochastic experiment
- A coin is tossed, and an outcome is observed:
  - Head – H
  - Tail – T

Example: Three experiments showing a fair coin tossed 100 times



Sample Space

Definition

Sample space, Ω, of a random experiment is the set of all possible outcomes of the experiment.

Some examples:

- Tossing two coins consequtively: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$

Sample Space

Definition

Sample space, Ω, of a random experiment is the set of all possible outcomes of the experiment.

Some examples:

- Tossing two coins consequtively: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$
- The lifetime of some electronic device (in days):
  $\Omega = \mathbb{R}_+ = 0.5, 1, 1.1, \ldots$

Sample Space

Definition

Sample space, Ω, of a random experiment is the set of all possible outcomes of the experiment.

Some examples:

- Tossing two coins consequtively: $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$
- The lifetime of some electronic device (in days): $\Omega = \mathbb{R}_+ = 0.5, 1, 1.1, \ldots$
- The number of submitted responses for the optional quiz: $\Omega = \{0,1,2,\ldots,75\} = \mathbb{Z}_+$

Sample Space

Definition

Sample space, Ω, of a random experiment is the set of all possible outcomes of the experiment.

Some examples:

- Tossing two coins consequtively: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$
- The lifetime of some electronic device (in days):
  $\Omega = \mathbb{R}_+ = 0.5, 1, 1.1, \ldots$
- The number of submitted responses for the optional quiz:
  $\Omega = \{0, 1, 2, \ldots, 75\} = \mathbb{Z}_+$
- The weight of ten selected people:
  $\Omega = \{(x_1, x_2, \ldots, x_{10}), x_i \geq 0, i = 1, \ldots, 10\} = \mathbb{R}_+^{10}$

Event

Definition

Event is any subset of the sample space.

Some examples:

- The event that the sum of two dice is 10 or more:
  $A = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\}$

Event

Definition

Event is any subset of the sample space.

Some examples:

- The event that the sum of two dice is 10 or more:
  $A = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\}$
- The event that someone does not have a birthday on October 5:
  $B = \{$Birthday on Jan 1, Birthday on Jan 2, $\ldots$, Birthday on Oct 4, Birthday on Oct 6, $\ldots$, Bithday on Dec 31$\}$

Event

Definition

Event is any subset of the sample space.

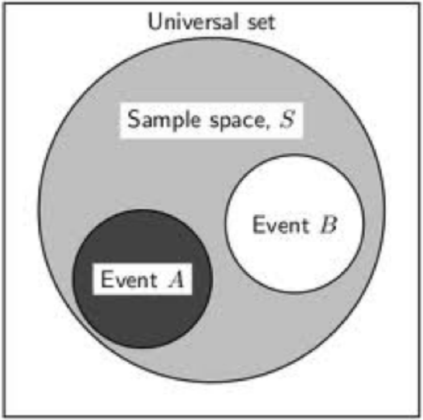
Some examples:

- The event that the sum of two dice is 10 or more: $A = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\}$
- The event that someone does not have a birthday on October 5: $B = \{$Birthday on Jan 1, Birthday on Jan 2, $\ldots$, Birthday on Oct 4, Birthday on Oct 6, $\ldots$, Bithday on Dec 31$\}$
- The event that out of fifty selected people, at least five are taller than 5.8": $C = \{5.2'', 5.4'', 5.8'', 6.0'', 6.1'', 6.3'', 5.9'', \ldots\}$

Event and Sets

Definition

Event is any subset of the sample space.



Events and Sets

Since events are sets, we can apply the usual set operations to them:

- The set $A \cup B$ (A union B) is the event that A or B or both occur,

Events and Sets

Since events are sets, we can apply the usual set operations to them:

- The set $A \cup B$ (A union B) is the event that A or B or both occur,
- The set $A \cap B$ (A intersection B) is the event that A and B both occur,

Events and Sets

Since events are sets, we can apply the usual set operations to them:

- The set $A \cup B$ (A union B) is the event that A or B or both occur,
- The set $A \cap B$ (A intersection B) is the event that A and B both occur,
- The set A^c (A complement) is the event that A does not occur,

Events and Sets

Since events are sets, we can apply the usual set operations to them:

- The set $A \cup B$ (A union B) is the event that A or B or both occur,
- The set $A \cap B$ (A intersection B) is the event that A and B both occur,
- The set A^c (A complement) is the event that A does not occur,
- If $A \subset B$ (A is a subset of B) then event A is said to **imply event** B.

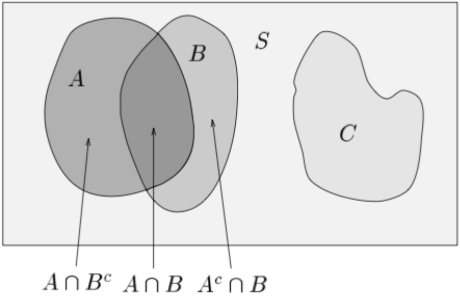
Events and Sets

Since events are sets, we can apply the usual set operations to them:

- The set $A \cup B$ (A union B) is the event that A or B or both occur,
- The set $A \cap B$ (A intersection B) is the event that A and B both occur,
- The set A^c (A complement) is the event that A does not occur,
- If $A \subset B$ (A is a subset of B) then event A is said to **imply event** B.
- Two events A and B which have no outcomes in common, that is, $A \cap B =$, are called **disjoint events**.

Event and Sets

Since events are sets, we can apply the usual set operations to them:



Random Experiments and Sample Spaces - Summary

- **Random experiment** - an experiment whosse outcome cannot be determined in advance, because it is unknown
- **Sample space,** Ω - set of all possible outcomes of a random experiment
- **Event** - any subset of the sample space

Section 3

Probability

Probability - Frequentist Viewpoint

To compute the probability of an event $A \subset \Omega$, count the number of occurences of A in N random experiments

Then $P(A) = \lim_{n \to \infty} \frac{N(A)}{N}$

For example, if we toss a coin 100 times, and observe 51 heads, then the probability of a head can be found as: $P(H) \approx \frac{51}{100}$

Probability

Definition

A probability $\mathbb{P}$ is a rule (function) that assigns a positive number to every event from the sample space, and satisfies the following **axioms of probability**:

- **Axiom 1:** $\mathbb{P}(A) \geq 0$

Probability

Definition

A probability $\mathbb{P}$ is a rule (function) that assigns a positive number to every event from the sample space, and satisfies the following **axioms of probability**:

- **Axiom 1:** $\mathbb{P}(A) \geq 0$
- **Axiom 2:** $\mathbb{P}(\Omega) = 1$

Probability

Definition

A probability $\mathbb{P}$ is a rule (function) that assigns a positive number to every event from the sample space, and satisfies the following **axioms of probability**:

- **Axiom 1:** $\mathbb{P}(A) \geq 0$
- **Axiom 2:** $\mathbb{P}(\Omega) = 1$
- **Axiom 3: The sum rule:** For any sequence of disjoint events, $A_1, A_2, \ldots,$ it holds that:

$$\mathbb{P}(\cup_i Ai) = \sum_i \mathbb{P}(A_i)$$

Probability - Examples

Consider the experiment where we throw a fair die. How do we define Ω and $\mathbb{P}$?

Consider the experiment where we throw a fair die. How do we define Ω and $\mathbb{P}$?
Obviously, $\Omega = \{1, 2, 3, 4, 5, 6\}$. We can now define $\mathbb{P}$ as:

$$\mathbb{A} = \frac{|A|}{6}$$

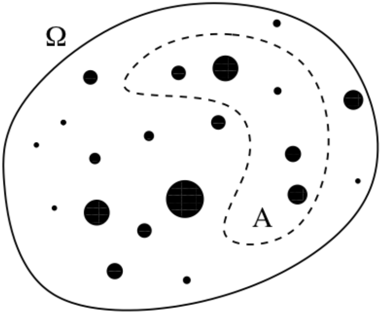
where A represents some possible event from the sample space. For example, all even number on the die.

Elementary Event and Discrete Sample Space

In many random experiments, sample space is **countable**:

$$\Omega = \{a_1, a_2, \ldots, a_n\}$$

Such a sample space is called a **discrete sample space**.

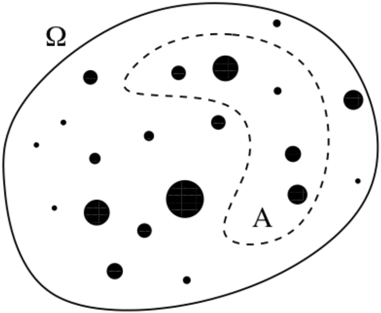


Discrete Sample Space

In many random experiments, sample space is **countable**:

$$\Omega = \{a_1, a_2, \ldots, a_n\}$$

Such a sample space is called a **discrete sample space**.



Discrete Sample Space and Elementary Events

In many random experiments, sample space is **countable**:

$$\Omega = \{a_1, a_2, \ldots, a_n\}$$

Such a sample space is called a **discrete sample space**.
The easiest way to specify the probability on a discrete sample space is to first specify the probability of all **elementary events** a_i, and then to define:

$$\mathbb{P}(A) = \sum_{a_i \in A} p_i, \ \ \forall A \subset \Omega$$

Discrete Sample Space and Equally Likely Events

Theorem

If sample space Ω has a finite number of outcomes, and all outcomes are equally likely, then the probability of every event A is equal to:

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

Discrete Sample Space and Counting

What does counting have to do with random experiments?

Section 4

Conditional Probability and Independance

Thought Experiment

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Some examples:

- Today is a sunny day on Kauai, Hawaii. What does that tell us about the weather in Seattle?

Thought Experiment

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Some examples:

- Today is a sunny day on Kauai, Hawaii. What does that tell us about the weather in Seattle?
- Today is sunny and cold around NEU Seattle campus. What does that tell us about the weather in Seattle today?

Thought Experiment

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Some examples:

- Today is a sunny day on Kauai, Hawaii. What does that tell us about the weather in Seattle?
- Today is sunny and cold around NEU Seattle campus. What does that tell us about the weather in Seattle today?
- Today is sunny and cold around NEU Seattle campus. What are the chances that it is raining on UW main campus?

Conditional Probability

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Let's now generalize:

- Let's assume that event B has occurred.

Conditional Probability

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Let's now generalize:

- Let's assume that event B has occurred.
- If B has occurred, we know that the outcome lies in B.

Conditional Probability

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Let's now generalize:

- Let's assume that event B has occurred.
- If B has occurred, we know that the outcome lies in B.
- That means that event A will occur if and only if $A \cap B$ occurs.

Conditional Probability

Let Ω be a sample space associated with a random experiment. Let A, B be two events. How do probabilities change when we know that some event $B \in \Omega$ has occured?

Let's now generalize:

- Let's assume that event B has occurred.
- If B has occurred, we know that the outcome lies in B.
- That means that event A will occur if and only if $A \cap B$ occurs.
- The relative probability of A occuring in those circumstance is given as:

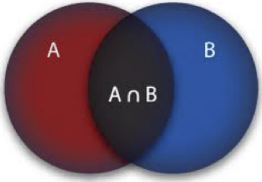
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Conditional Probability

Definition

Let Ω be a sample space associated with a random experiment. Let A, B be two events. The **conditional probability** of event A, given event B is defined as:

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



Conditional Probability - Example

Example: We throw two dice. Given that observed sum is equal to 10, what is the probability that one die is equal to 6?

Solution:

Let B be the event that the sum of dots on two dice is 10:

$$B = \{(4,6),(5,5),(6,4)\}$$

Conditional Probability - Example

Example: We throw two dice. Given that observed sum is equal to 10, what is the probability that one die is equal to 6?

Solution:

Let B be the event that the sum of dots on two dice is 10:

$$B = \{(4,6),(5,5),(6,4)\}$$

Let A now be an event that one die is showing 6:

$$A = \{(1,6),(2,6),(3,6),(4,6),(5,6)\}$$

Conditional Probability - Example

Example: We throw two dice. Given that observed sum is equal to 10, what is the probability that one die is equal to 6?
Solution:

Let B be the event that the sum of dots on two dice is 10:

$$B = \{(4,6),(5,5),(6,4)\}$$

Let A now be an event that one die is showing 6:

$$A = \{(1,6),(2,6),(3,6),(4,6),(5,6)\}$$

Now, $A \cap B = \{(4, 6, (6,4))\}$.

Conditional Probability - Example

Example: We throw two dice. Given that observed sum is equal to 10, what is the probability that one die is equal to 6?
Solution:

Let B be the event that the sum of dots on two dice is 10:

$$B = \{(4,6),(5,5),(6,4)\}$$

Let A now be an event that one die is showing 6:

$$A = \{(1,6),(2,6),(3,6),(4,6),(5,6)\}$$

Now, $A \cap B = \{(4,6,(6,4))\}$.
Since all elementary events are equally likely, we can finally write:

$$\mathbb{P}(A|B) = \frac{\frac{2}{36}}{\frac{3}{36}} = \frac{2}{3}$$

Product Rule

By the definition of conditional probability, we have:

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

We can generalize this to n intersections $A_1 \cap A_2 \cap \cdots \cap A_n$. This gives us the **product (chain) rule of probability**.

The Chain (Product) Rule

By the definition of conditional probability, we have:

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

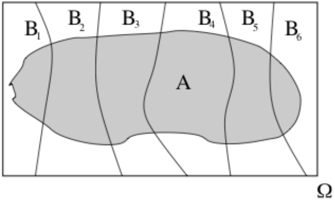
Theorem

Let $A_1, A_2, \ldots, A_n$ *be a sequence of events with* $\mathbb{P}(A_1, \ldots, A_{n-1}) > 0$. *Then:*

$$\mathbb{P}(A_1, \ldots, A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2)\ldots\mathbb{P}(A_n|A_1A_2\ldots A_{n-1})$$

The Law of Total Probability

Suppose that we can partition the sample space Ω into n **disjoint** partitions, $B_1, B_2, \ldots, B_n$.



The Law of Total Probability

Suppose that we can partition the sample space Ω into n **disjoint** partitions, $B_1, B_2, \ldots, B_n$.
Using the third axiom of probability (the sum rule), and the definition of conditional probability, we can derive the **law of total probability:**

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

The Bayes' Rule

A simple, yet important outcome of the definition of conditional probability is the **Bayes' rule**:

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

Generalization: Let's say we are given a probability $\mathbb{P}(A|B_1)$ and we need to compute $\mathbb{P}(B_1|A)$. Let $B_1, B_2, \ldots, B_k$ be disjoint events, covering the whole sample space. Then:

$$\begin{aligned}
\mathbb{P}(B_1|A) &= \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\mathbb{P}(A)} \\
&= \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\sum_{i=1}^{k} \mathbb{P}(A \cap B)} \\
&= \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\sum_{i=1}^{k} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}
\end{aligned}$$

Independence

- An important concept in probability and statistics

Independence

- An important concept in probability and statistics
- Conceptually, it models the **lack of information** between events

Independence

- An important concept in probability and statistics
- Conceptually, it models the **lack of information** between events
- Two events A and B are said to be independent, if the knowledge that A has occurred does not change the probability that B occurs

Independent Events

Definition

Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Stated differently, events A and B are independent is $\mathbb{P}(A|B) = \mathbb{P}(A)$ and $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Independance and Mutual Exclusiveness

Note: two events A and B being independent necessarily means that these events are not **mutually exclusive**:

$$A \cap B \neq \emptyset$$

Example: Consider simultaneous flips of two fair coins.

- **Event A:** First Coin was a head.

Independance and Mutual Exclusiveness

Note: two events A and B being independent necessarily means that these events are not **mutually exclusive**:

$$A \cap B \neq \emptyset$$

Example: Consider simultaneous flips of two fair coins.

- **Event** A: First Coin was a head.
- **Event** B: Second coin was a tail.

Independance and Mutual Exclusiveness

Note: two events A and B being independent necessarily means that these events are not **mutually exclusive**:

$$A \cap B \neq \emptyset$$

Example: Consider simultaneous flips of two fair coins.

- **Event** A: First Coin was a head.
- **Event** B: Second coin was a tail.
- Events A and B are independent events, and they are not mutually exclusive.

Independance and Mutual Exclusiveness

Note: two events A and B being independent necessarily means that these events are not **mutually exclusive**:

$$A \cap B \neq \emptyset$$

Example: Consider simultaneous flips of two fair coins.

- **Event** A: First Coin was a head.
- **Event** B: Second coin was a tail.
- Events A and B are independent events, and they are not mutually exclusive.
- On the other hand, let's define **event** C as: **first coin was a tail**. Then A and C are not independent. They are **mutually exclusive** - knowing that one has occurred (e.g. heads) implies the other cannot (tails).

Section 5

Random Variables

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

Note 1: Although random variable is a function, we typically denote it simply as X.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

Note 1: Although random variable is a function, we typically denote it simply as X.

Note 2: It is very common to directly define probabilities over the range of the random variable.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

Some examples of random variables:

- The number of defective microchips out of 100 inspected.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

Some examples of random variables:

- The number of defective microchips out of 100 inspected.
- The number of bugs in some computer program.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

Some examples of random variables:

- The number of defective microchips out of 100 inspected.
- The number of bugs in some computer program.
- The number of rainy days in Seattle in December.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

Some examples of random variables:

- The number of defective microchips out of 100 inspected.
- The number of bugs in some computer program.
- The number of rainy days in Seattle in December.
- The amount of time needed for one communication packet to be transferred from your computer to the server hosting our course website.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

- The set of possible values some random variable X can take is called the **range** of X.

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

- The set of possible values some random variable X can take is called the **range** of X.
- Based upon the countability of the range, we distinguish between:

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \rightarrow \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

- The set of possible values some random variable X can take is called the **range** of X.
- Based upon the countability of the range, we distinguish between:
  - **Discrete random variables** - have a countable number of possible values

Random Variable

Definition

A random variable is a **mapping** or a **function** from the sample space Ω to the real line: $X : \Omega \to \mathbb{R}$.
Let $x \in \mathbb{R}$. Then it holds that:

$$\mathbb{P}(X(w) = x) = \mathbb{P}(\{w \in \Omega : X = x\})$$

- The set of possible values some random variable X can take is called the **range** of X.
- Based upon the countability of the range, we distinguish between:
  - **Discrete random variables** - have a countable number of possible values
  - **Continuous random variables** - have an uncountable number of possible values

Discrete Random Variables - Examples

Discrete random variables - have a countable number of possible values

- Number of sunny days in Seattle in Summer

Discrete Random Variables - Examples

Discrete random variables - have a countable number of possible values

- Number of sunny days in Seattle in Summer
- Number of students who pass CS 5002 with grade A

Discrete Random Variables - Examples

Discrete random variables - have a countable number of possible values

- Number of sunny days in Seattle in Summer
- Number of students who pass CS 5002 with grade A
- Number of delievered packages currently on my porch

Discrete Random Variables - Examples

Discrete random variables - have a countable number of possible values

- Number of sunny days in Seattle in Summer
- Number of students who pass CS 5002 with grade A
- Number of delievered packages currently on my porch
- Number of problems on the final exam

Continuous Random Variables - Examples

Continuous random variables - have an uncountable number of possible values

- Rainfall measurement in Seattle in December

Continuous Random Variables - Examples

Continuous random variables - have an uncountable number of possible values

- Rainfall measurement in Seattle in December
- Snowfall measurement in Whistler in December

Continuous Random Variables - Examples

Continuous random variables - have an uncountable number of possible values

- Rainfall measurement in Seattle in December
- Snowfall measurement in Whistler in December
- The lifetime of my new TV

Continuous Random Variables - Examples

Continuous random variables - have an uncountable number of possible values

- Rainfall measurement in Seattle in December
- Snowfall measurement in Whistler in December
- The lifetime of my new TV
- The average lifetime of the most expensive smart phone currently on the market

Section 6

Probability Distributions

Probability Distribution

Let X be some random variable. We would like to specify the probabilities of various kinds of events. For examples:

- Probability of an event $X = x$

.

Probability Distribution

Let X be some random variable. We would like to specify the probabilities of various kinds of events. For examples:

- Probability of an event $X = x$
- Probability of an event $\{x \leq X \leq b\}$

.

Probability Distribution

Let X be some random variable. We would like to specify the probabilities of various kinds of events. For examples:

- Probability of an event $X = x$
- Probability of an event $\{x \leq X \leq b\}$

If we can specify all probabilities involving X, then we can specify the **probability distribution of** X.

Cumulative Distribution Function (cdf)

Definition

The **cumulative distribution function (cdf)** of some random variable X is the function $F : \mathbb{R} \to [0, 1]$, defined as:

$$F(x) := \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$$

The following properties of cdf F are a direct consequence of the axioms of probability:

- $0 \leq F(x) \leq 1$.

Cumulative Distribution Function (cdf)

Definition

The **cumulative distribution function (cdf)** of some random variable X is the function $F : \mathbb{R} \to [0, 1]$, defined as:

$$F(x) := \mathbb{P}(X \leq x), \ \ x \in \mathbb{R}$$

The following properties of cdf F are a direct consequence of the axioms of probability:

- $0 \leq F(x) \leq 1$.
- F is increasing: $x \leq y \to F(x) \leq F(y)$

Cumulative Distribution Function (cdf)

Definition

The **cumulative distribution function (cdf)** of some random variable X is the function $F : \mathbb{R} \to [0, 1]$, defined as:

$$F(x) := \mathbb{P}(X \leq x), \ \ x \in \mathbb{R}$$

The following properties of cdf F are a direct consequence of the axioms of probability:

- $0 \leq F(x) \leq 1.$
- F is increasing: $x \leq y \to F(x) \leq F(y)$
- $\lim_{x \to \infty} F(x) = 1$

Cumulative Distribution Function (cdf)

Definition

The **cumulative distribution function (cdf)** of some random variable X is the function $F : \mathbb{R} \to [0, 1]$, defined as:

$$F(x) := \mathbb{P}(X \leq x), \;\; x \in \mathbb{R}$$

The following properties of cdf F are a direct consequence of the axioms of probability:

- $0 \leq F(x) \leq 1.$
- F is increasing: $x \leq y \to F(x) \leq F(y)$
- $\lim_{x\to\infty} F(x) = 1$
- F is right-continuous.

Discrete Distributions and Probability Mass Function

Definition

Some random variable X has a **discrete probability distribution** if X is a discrete random variable.

Every discrete random variable has associated with it a **probability mass function** which outputs the probability of the random variable taking a particular value.

Discrete Distributions and Probability Mass Function

Every discrete random variable has associated with it, a **probability mass function** which ouputs the probability of the random variable taking a particular value.

Definition

Given some discrete random variable X, we can define its **probability mass function (pmf)**, f_x as:

$$f(x) = \mathbb{P}(X = x)$$

Note: the easiest way to specify the probability distribution of a discrete random variable is to specify its pmf.

Discrete Distributions and Probability Mass Function

Definition

Given some discrete random variable X, we can define its **probability mass function (pmf)**, f_x as:

$$f(x) = \mathbb{P}(X = x)$$

Note: the easiest way to specify the probability distribution of a discrete random variable is to specify its pmf.

Example: Let X denote a random variable corresponding to the role of a die. Then $P(X = i) = \frac{1}{6}$ is the pmf associated with X.

Section 7

Expectation

Expectation

Although all the probability information of some random variable is contained in its cdf and its pmf, it is often useful to consider other numerical characteristics of that random variable, including:

- Expectation

Expectation

Although all the probability information of some random variable is contained in its cdf and its pmf, it is often useful to consider other numerical characteristics of that random variable, including:

- Expectation
- Variance

Expectation

Although all the probability information of some random variable is contained in its cdf and its pmf, it is often useful to consider other numerical characteristics of that random variable, including:

- Expectation
- Variance
- Standard derivation

Expectation

Definition

Let X be a discrete random variable with pmf f. The **expectation** of X, denoted as $\mathbb{E}(X)$ is defined as:

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x) = \sum_x xf(x)$$

Note: The expectation is not necessarily a possible outcome of some random experiment. It is the weighted average of the values that a random variable can take (where the weights are given by the probabilities).

The Linearity of Expectation

Definition

Let X be a discrete random variable with pmf f. The **expectation** of X, denoted as $\mathbb{E}(X)$ is defined as:

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x) = \sum_x xf(x)$$

Theorem

If X is a discrete random variable with pmf f, then for any real-valued function g it holds that:

$$\mathbb{E}(g(X)) = \sum_x g(x)f(x)$$

The Linearity of Expectation

Theorem

If X is a discrete random variable with pmf f, then for antyreal-valued function g it holds that:

$$\mathbb{E}(g(X)) = \sum_{x} g(x)f(x)$$

An important consequence of the previous theorem is the fact that expectation is a linar function, and it holds that:

- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

The Linearity of Expectation

Theorem

If X is a discrete random variable with pmf f, then for antyreal-valued function g it holds that:

$$\mathbb{E}(g(X)) = \sum_x g(x) f(x)$$

An important consequence of the previous theorem is the fact that expectation is a linar function, and it holds that:

- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
- $\mathbb{E}(g(X) + h(X)) = \mathbb{E}g(X) + \mathbb{E}h(X)$

The Variance and Standard Derivation

Definition

The **variance** of a random variable X, denotes as $\mathsf{Var}(X)$ is defined as:

$$\mathsf{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$$

The following properties hold for variance:

- The squared root of the variance is called **standard derivation**.

The Variance and Standard Derivation

Definition

The **variance** of a random variable X, denotes as $\mathrm{Var}(X)$ is defined as:

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$$

The following properties hold for variance:

- The squared root of the variance is called **standard derivation**.
- $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

The Variance and Standard Derivation

Definition

The **variance** of a random variable X, denotes as $\text{Var}(X)$ is defined as:

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$$

The following properties hold for variance:

- The squared root of the variance is called **standard derivation**.
- $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$

Section 8

Some Imporant Discrete Distributions

Bernoulli Distribution

Definition

Some random variable X has a **Bernoulli distribution** with success probability p if X can only assume two possible values, 0 and 1, with probabilities:

$$\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$$

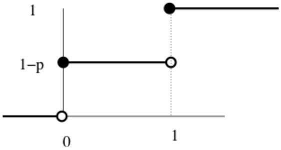
We write $X \approx \text{Ber}(p)$.

Example: Single coin toss experiment.

Bernoulli Distribution

Some properties:

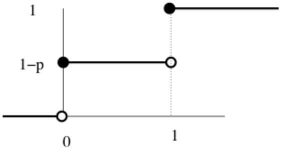
- The cdf given below.



Bernoulli Distribution

Some properties:

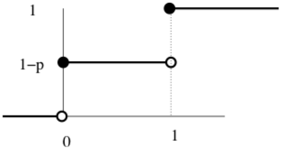
- The cdf given below.
- The expectation: $\mathbb{E}(P) = 0\mathbb{P}(X = 0) + 1\mathbb{P}(X = 1) = p$



Bernoulli Distribution

Some properties:

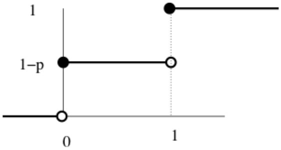
- The cdf given below.
- The expectation: $\mathbb{E}(P) = 0\mathbb{P}(X = 0) + 1\mathbb{P}(X = 1) = p$
- Variance: $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = p(1 - p)$



Bernoulli Distribution

Some properties:

- The cdf given below.
- The expectation: $\mathbb{E}(P) = 0\mathbb{P}(X = 0) + 1\mathbb{P}(X = 1) = p$
- Variance: $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = p(1-p)$
- One of the most important distributions in the probability theory.



Binomial Distribution

Definition

Consider the sequence of n random events with two possible outcomes, for example coin tosses, where the probability of a head is p. If X is the random variable that counts **the number of heads**, then we say that X has a **binomial distribution**, with parameters n and p, and write $X \approx \text{Bin}(n, p)$. The probability mass function (pmf) of X is given as:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - o)^{n-x}, x = 0, 1, \ldots, n$$

Binomial Distribution

Definition

The probability mass function (pmf) of $X \approx \text{Bin}(n, p)$ is given as:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - o)^{n-x}, x = 0, 1, \ldots, n$$

Some properties:

- The expecation of random variable $X \approx \text{Bin}(n, p)$ is $\mathbb{E}(X) = np$.

Definition

The probability mass function (pmf) of $X \approx \text{Bin}(n, p)$ is given as:

$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - o)^{n-x}, x = 0, 1, \ldots, n$$

Some properties:

- The expecation of random variable $X \approx \text{Bin}(n, p)$ is $\mathbb{E}(X) = np$.
- The variance of random variable $X \approx \text{Bin}(n, p)$ is Var$)(X) = np(1 - p)$.

Geometric Distribution

Definition

Consider again the sequence of n random events with two possible outcomes, for example coin tosses, where the probability of a head is p. If X is the random variable that counts **the number tosses needed before we see the first head**, then we say that X has a **geometric distribution**, with parameter p, and write $X \approx \text{Geo}(p)$. The probability mass function (pmf) of X is given as:

$$f(x) = P(X = x) = (1 - p)^{x-1}p, \;\; x = 1, 2, 3, \ldots$$

Geometric Distribution

Definition

The probability mass function (pmf) of $X \approx \text{Geo}(p)$ is given as:

$$f(x) = P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots$$

Some properties:

- The expecation of random variable $X \approx \text{Geo}(p)$ is $\mathbb{E}(X) = \frac{1}{p}$.

Geometric Distribution

Definition

The probability mass function (pmf) of $X \approx \text{Geo}(p)$ is given as:

$$f(x) = P(X = x) = (1-p)^{x-1}p, \;\; x = 1, 2, 3, \ldots$$

Some properties:

- The expecation of random variable $X \approx \text{Geo}(p)$ is $\mathbb{E}(X) = \frac{1}{p}$.
- The variance of random variable $X \approx \text{Geo}(p)$ is $\text{Var}(X) = np(1-p)$.

Geometric Distribution

Definition

The probability mass function (pmf) of $X \approx \text{Geo}(p)$ is given as:

$$f(x) = P(X = x) = (1-p)^{x-1}p, \;\; x = 1, 2, 3, \ldots$$

Some properties:

- The expecation of random variable $X \approx \text{Geo}(p)$ is $\mathbb{E}(X) = \frac{1}{p}$.
- The variance of random variable $X \approx \text{Geo}(p)$ is $\text{Var}(X) = np(1-p)$.
- **Memoryless property** - the fact that we have already had k failures does not make the event of getting a success in the next trial(s) any more likely.

Section 9

Stochastic Processess

Stochastic Processes

Definition

A **stochastic** or **random process** is typically defined as an **indexed collection of random variables**, and denotes as $\{X_i\}_{i=1}^{N}$.

Stochastic Processes

Definition

A **stochastic** or **random process** is typically defined as an **indexed collection of random variables**, and denotes as $\{X_i\}_{i=1}^{N}$.

Some examples:

- Weather around Seattle

Stochastic Processes

Definition

A **stochastic** or **random process** is typically defined as an **indexed collection of random variables**, and denotes as $\{X_i\}_{i=1}^{N}$.

Some examples:

- Weather around Seattle
- Stock prices

Stochastic Processes

Definition

A **stochastic** or **random process** is typically defined as an **indexed collection of random variables**, and denotes as $\{X_i\}_{i=1}^{N}$.

Some examples:

- Weather around Seattle
- Stock prices
- Customers entering and existing store

Stochastic Processes

Definition

A **stochastic** or **random process** is typically defined as an **indexed collection of random variables**, and denotes as $\{X_i\}_{i=1}^{N}$.

Some examples:

- Weather around Seattle
- Stock prices
- Customers entering and existing store
- Operation of telecommunication networks

Independent and Identically Distributed Stochastic Processes

Definition

A stochastic process is said to be **independent and identically distributed (iid)** if each of its random variables X_i has the same cdf, and all random variables are independent, in a sense that:

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\ldots\mathbb{P}(X_n = x_n$$