CSE 312

Foundations of Computing II

Lecture 7: Random Variables

Announcements

- PSet 1 graded + solutions on canvas
- PSet 2 due tonight
- Pset 3 posted this evening
  - First programming assignment (naïve Bayes)
  - Extensive intro in the sections tomorrow
  - Python tutorial lesson on edstem

Review Chain rule & independence

Theorem. (Chain Rule) For events $A_1, A_2, \dots, A_n$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Definition. Two events A and A are (statistically) **independent** if

$$P(A \cap B) = P(A) \cdot P(B).$$

"Equivalently." $P(A|B) = P(A)$. $P(B) \neq 0$

One more related item: Conditional Independence

Definition. Two events A and B are **independent** conditioned on C if
$P(C) \neq 0$ and $P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$.

- If $P(A \cap C) \neq 0$, equivalent to $P(B|A \cap C) = P(B \mid C)$
- If $P(B \cap C) \neq 0$, equivalent to $P(A|B \cap C) = P(A \mid C)$

Plain Independence. Two events A and B are **independent** if

$$P(A \cap B) = P(A) \cdot P(B).$$

- If $P(A) \neq 0$, equivalent to $P(B|A) = P(B)$
- If $P(B) \neq 0$, equivalent to $P(A|B) = P(A)$

Example – ~~Throwing Dice~~ Flipping Coin

Suppose that Coin 1 has probability of heads 0.3
and Coin 2 has probability of head 0.9.
We choose one coin randomly with equal probability and flip that coin 3 times independently. What is the probability we get all heads?

$$P(HHH) = P(HHH \mid C_1) \cdot P(C_1) + P(HHH \mid C_2) \cdot P(C_2)$$ Law of Total Probability (LTP)

$$= P(H \mid C_1)^3 \, P(C_1) + P(H \mid C_2)^3 \, P(C_2)$$ Conditional Independence

$$= 0.3^3 \cdot 0.5 + 0.9^3 \cdot 0.5 = 0.378$$

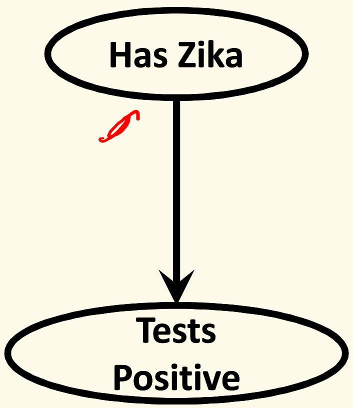
C_i = coin i was selected

Conditional independence and Bayesian inference in practice: Graphical models

- The sample space Ω is often the Cartesian product of possibilities of many different variables
- We often can understand the probability distribution P on Ω based on ***local properties*** that involve a few of these variables at a time
- We can represent this via a directed acyclic graph augmented with probability tables (called a Bayes net) in which each node represents one or more variables…

Graphical Models/Bayes Nets

- Bayes net for the Zika testing probability space (Ω, P)



Z	$\neg Z$
0.005	0.995

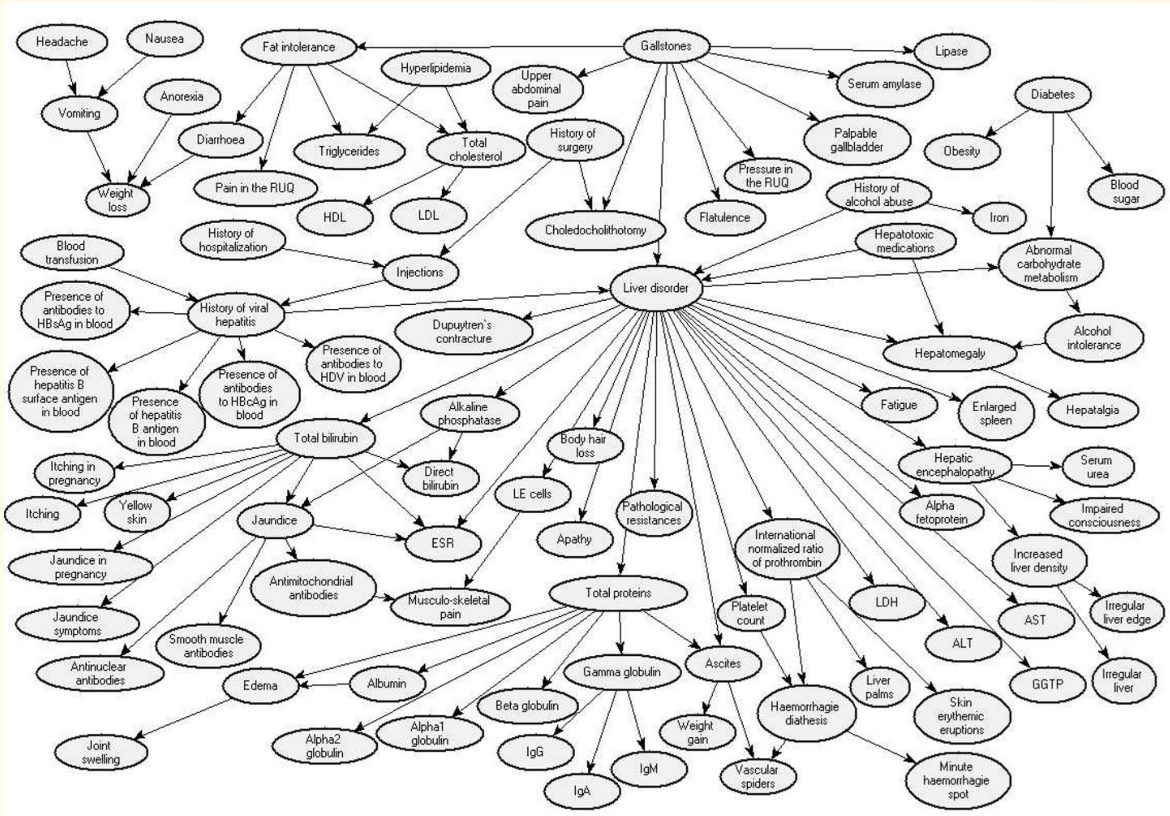
	T	$\neg T$
Z	0.98	0.02
$\neg Z$	0.01	0.99

$P(T|\neg Z)$

Conditional Probability Table:
- One column for each value of the variables at the node
- One row for each combination of values of immediate predecessors

Ω = Cartesian product of possible value assignments at all nodes.

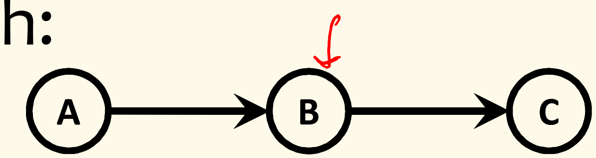
Graphical Models/Bayes Nets



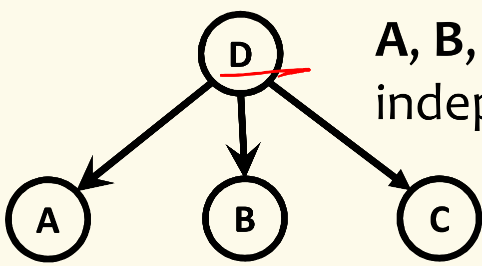
"A Bayesian Network Model for Diagnosis of Liver Disorders" – Agnieszka Onisko, M.S., Marek J. Druzdzel, Ph.D., and Hanna Wasyluk, M.D.,Ph.D.- September 1999.

Graphical Models/Bayes Nets

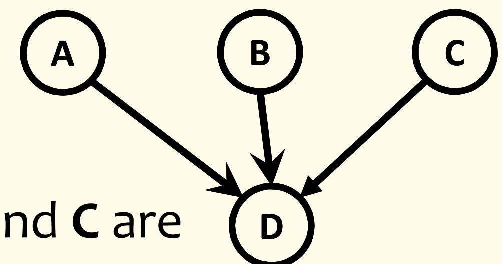
Bayes Net assumption/requirement

- The only dependence between variables is given by paths in the Bayes Net graph:
  - if only edges are 

    then **A** and **C** are *conditionally independent* given the value of **B**



A, B, C conditionally independent given **D**



A, B, and **C** are independent

Defines a unique global probability space (Ω, P)

Inference in Bayes Nets

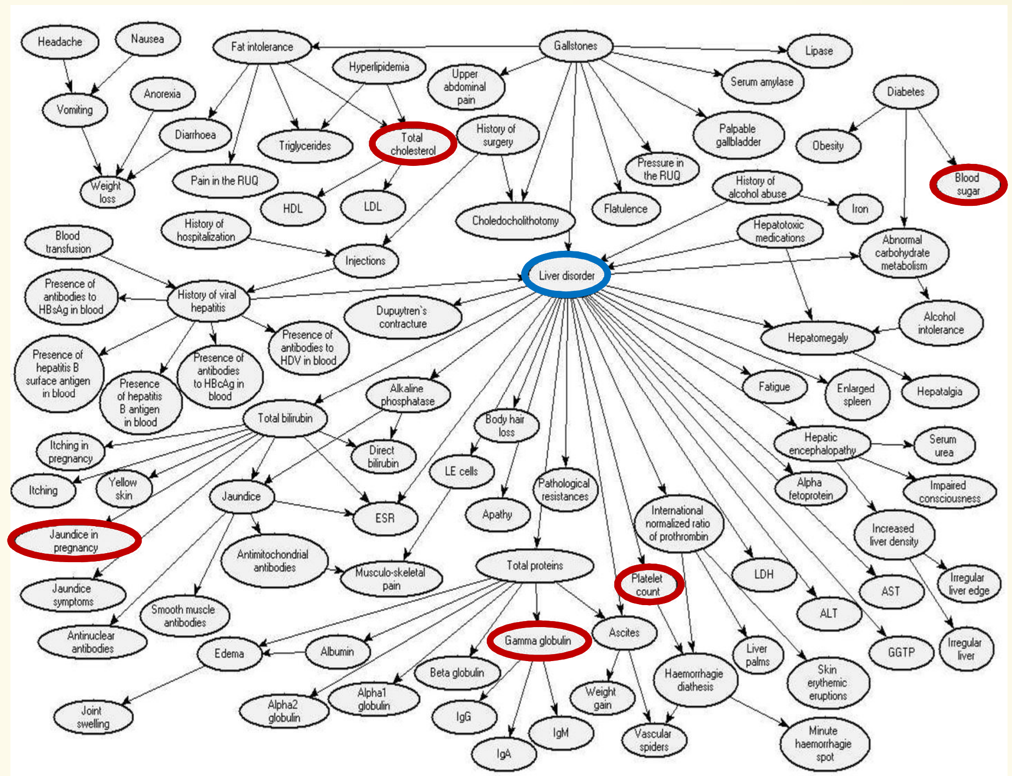
For much more see CSE 473

Given

- Bayes Net
  - graph
  - conditional probability tables for all nodes
- Observed values of variables at some nodes
  - e.g., clinical test results

Compute

- Probabilities of variables at other nodes
  - e.g., diagnoses



“A Bayesian Network Model for Diagnosis of Liver Disorders” – Agnieszka Onisko, M.S., Marek J. Druzdzel, Ph.D., and Hanna Wasyluk, M.D.,Ph.D.-September 1999.

Summary Chain rule & independence

Theorem. (Chain Rule) For events $A_1, A_2, \dots, A_n$,

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Definition. Two events A and A are (statistically) **independent** if

$$P(A \cap B) = P(A) \cdot P(B).$$

"Equivalently." $P(A|B) = P(A)$.

Definition. Two events A and B are **independent conditioned on** C if
$P(C) \neq 0$ and $P(A \cap B \mid C) = P(A \mid C) \cdot P(B \mid C)$.

Agenda

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- Expectation

Random Variables (Idea)

Often: We want to **capture quantitative properties** of the outcome of a random experiment, e.g.:

- *What is the total of two dice rolls?*
- *What is the number of coin tosses needed to see the first head?*
- *What is the number of heads among 2 coin tosses?*

Random Variables

Definition. A **random variable (RV)** for a probability space (Ω, P) is a function $X: \Omega \to \mathbb{R}$.

The set of values that X can take on is called its *range/support*

Two common notations: $X(\Omega)$ or Ω_X

Example. Two coin flips: $\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

$X(\Omega) = \{0,1,2\}$

$X =$ number of heads in two coin flips

$X(\text{HH}) = 2 \quad X(\text{HT}) = 1 \quad X(\text{TH}) = 1 \quad X(\text{TT}) = 0$

range (or support) of X is $X(\Omega) = \{0,1,2\}$

Another RV Example

20 different balls labeled 1, 2, …, 20 in a jar

- Draw a subset of 3 from the jar uniformly at random
- Let $X = \text{maximum of the 3 numbers on the balls}$
  - Example: $X(\{2,7,5\}) = 7$
  - Example: $X(\{15,3,8\}) = 15$

pollev.com/paulbeame028

How large is $|X(\Omega)|$?

A. 20^3
B. 20
C. 18
D. $\binom{20}{3}$

$X(\Omega) = \{3, 4, \ldots, 20\}$

min $X(\Omega)$

$|X(\Omega)| = 18$

Random Variables

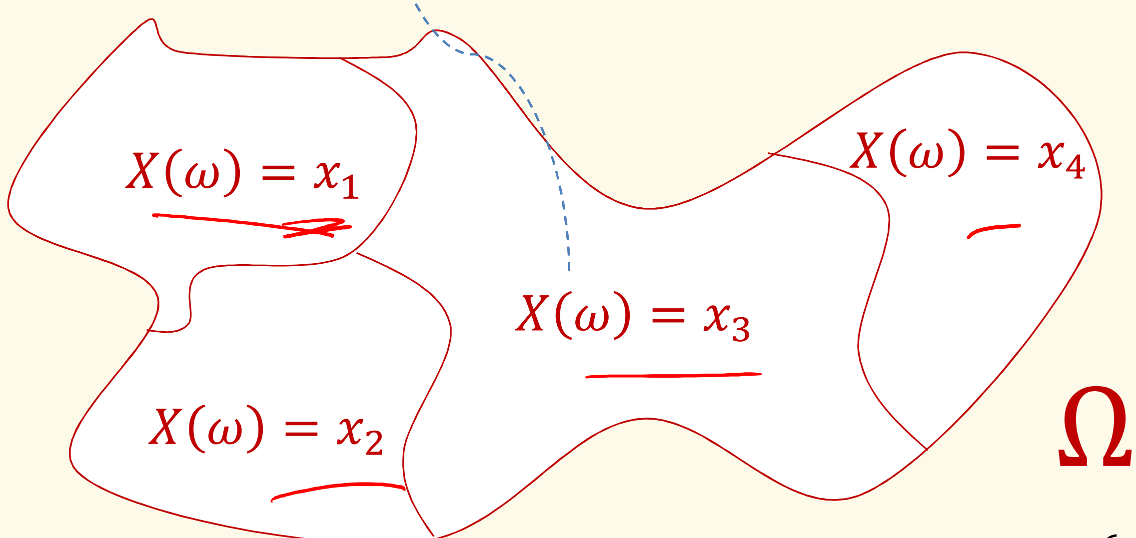
Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$$

We write $P(X = x) = P(\{X = x\})$

Random variables **partition** the sample space.

$$\sum_{x \in X(\Omega)} P(X = x) = 1$$



Random Variables

Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$$

We write $P(X = x) = P(\{X = x\})$

Example. Two coin flips: $\Omega = \{\mathrm{TT}, \mathrm{HT}, \mathrm{TH}, \mathrm{HH}\}$

$X =$ number of heads in two coin flips $\quad \Omega_X = X(\Omega) = \{0,1,2\}$

$$P(X=0) = \frac{1}{4} \quad P(X=1) = \frac{1}{2} \quad P(X=2) = \frac{1}{4}$$

The RV X yields a new probability distribution with sample space $\Omega_X \subset \mathbb{R}$!

Agenda

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- Expectation

Probability Mass Function (PMF)

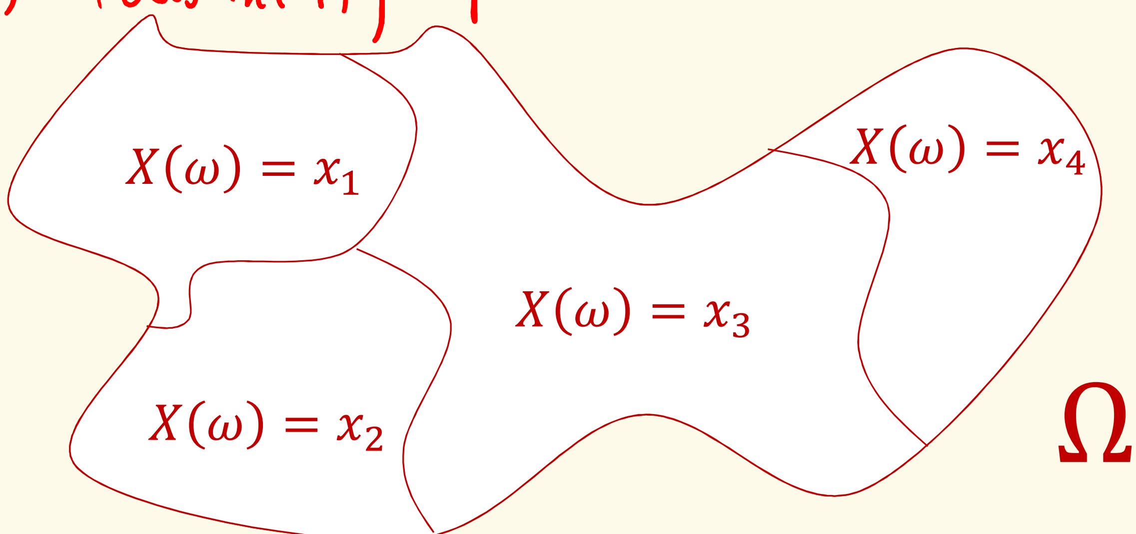
$p: \Omega \to \mathbb{R}$, ≥ 0, sum $= 1$

Definition. For a RV $X: \Omega \to \mathbb{R}$, the function $p_X: \Omega_X \to \mathbb{R}$ defined by $p_X(x) = P(X = x)$ is called the **probability mass function (PMF)** of X

(Ω_X, p_X) Probability space

Random variables **partition** the sample space.

$$\sum_{x \in X(\Omega)} P(X = x) = 1$$

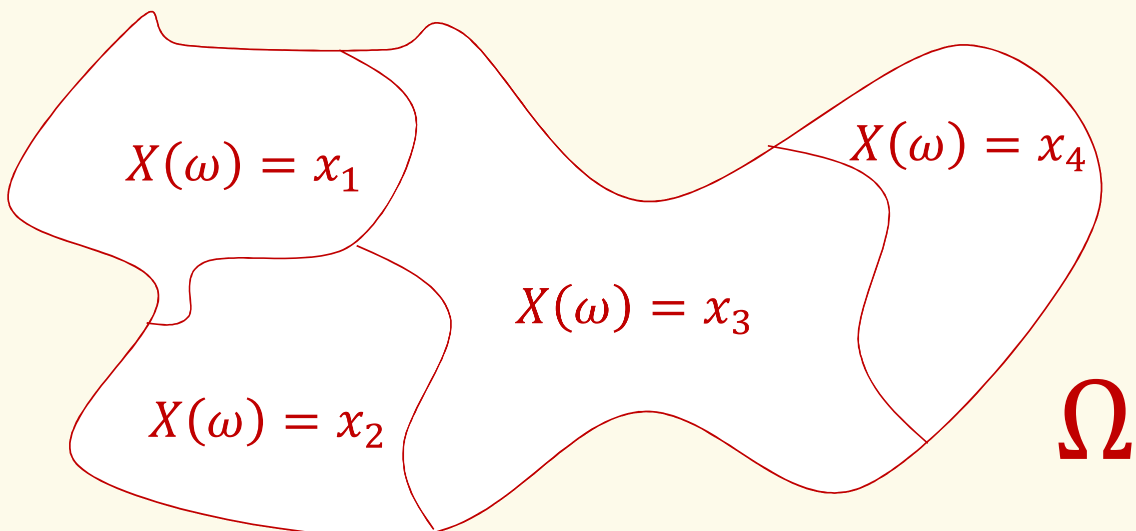


Probability Mass Function (PMF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the function $p_X: \Omega_X \to \mathbb{R}$ defined by $p_X(x) = P(X = x)$ is called the **probability mass function (PMF)** of X

Random variables **partition** the sample space.

$$\sum_{x \in \Omega_X} P(X = x) = 1$$

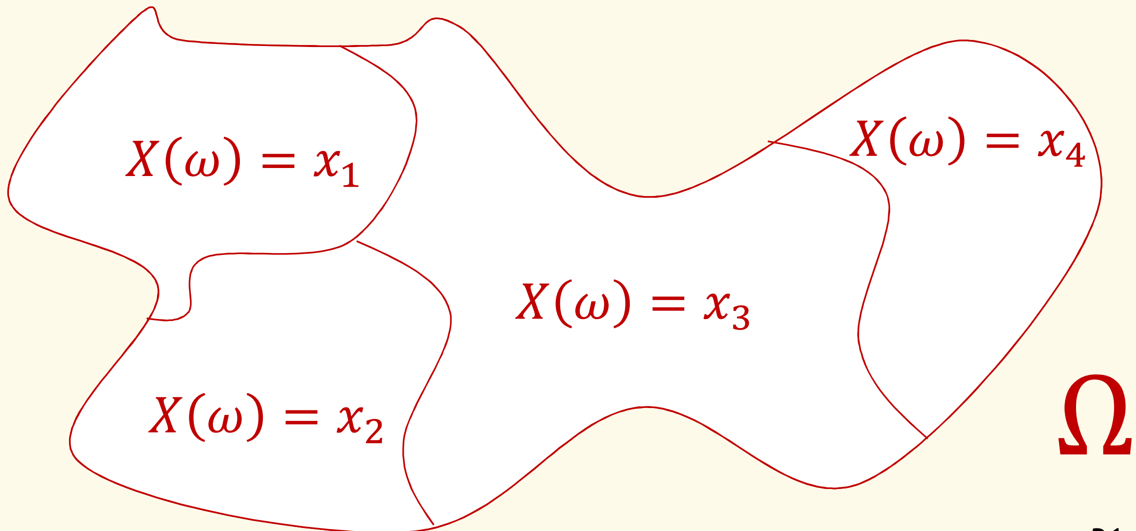


Probability Mass Function (PMF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the function $p_X: \Omega_X \to \mathbb{R}$ defined by $p_X(x) = P(X = x)$ is called the **probability mass function (PMF)** of X

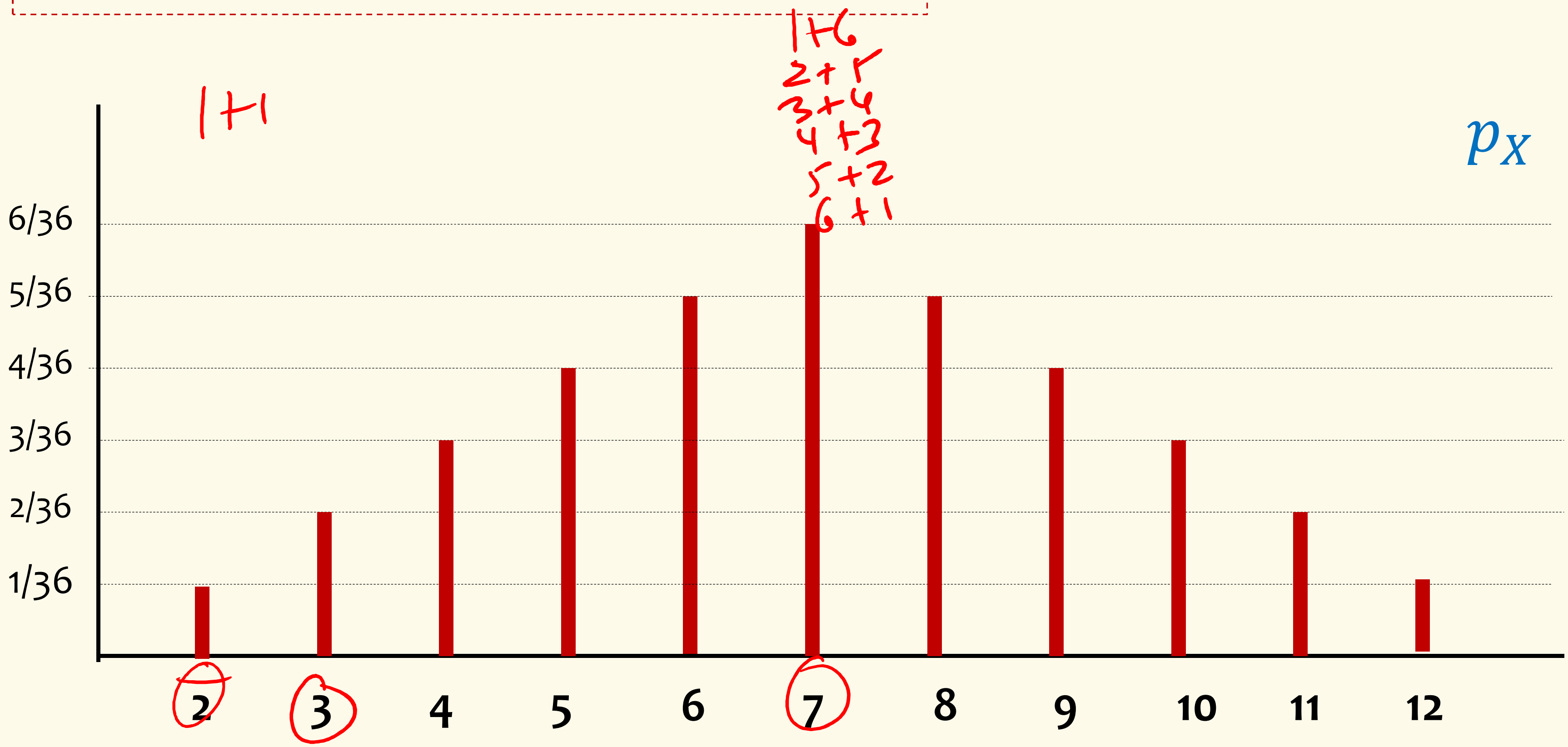
Random variables **partition** the sample space.

$$\sum_{x \in \Omega_X} p_X(x) = 1$$



Example – Two Fair Dice

X = sum of two dice throws



Example – Number of Heads

We flip n coins, independently, each heads with probability p

$$\Omega = \{\mathrm{HH}\cdots\mathrm{HH}, \mathrm{HH}\cdots\mathrm{HT}, \mathrm{HH}\cdots\mathrm{TH}, \ldots, \mathrm{TT}\cdots\mathrm{TT}\}$$

X = # of heads

$$p_X(k) = P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

# of sequences with k heads

Prob of sequence w/ k heads

gifak-net
h 2
h 1
b
a
S=a * h1
S=b * h2

Agenda

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- Expectation

Events concerning RVs

We already defined $P(X = x) = P(\{X = x\})$ where

$$\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$$

$X \leq 2.5$

Sometimes we want to understand other events involving RV X

– e.g. $\{X \leq x\} = \{\omega \in \Omega \mid X(\omega) \leq x\}$ which makes sense for any $x \in \mathbb{R}$

More generally…

– We could take any predicate $Q(\cdot)$ defined on the real numbers, and consider an event $\{Q(X)\} = \{\omega \in \Omega \mid Q(X(\omega)) \text{ is true}\}$

– If $Q(\cdot,\cdot)$ is a predicate of two real numbers and X and Y are RVs both defined on Ω then $\{Q(X, Y)\} = \{\omega \in \Omega \mid Q(X(\omega), Y(\omega)) \text{ is true}\}$

– The same thing works for properties of even more RVs

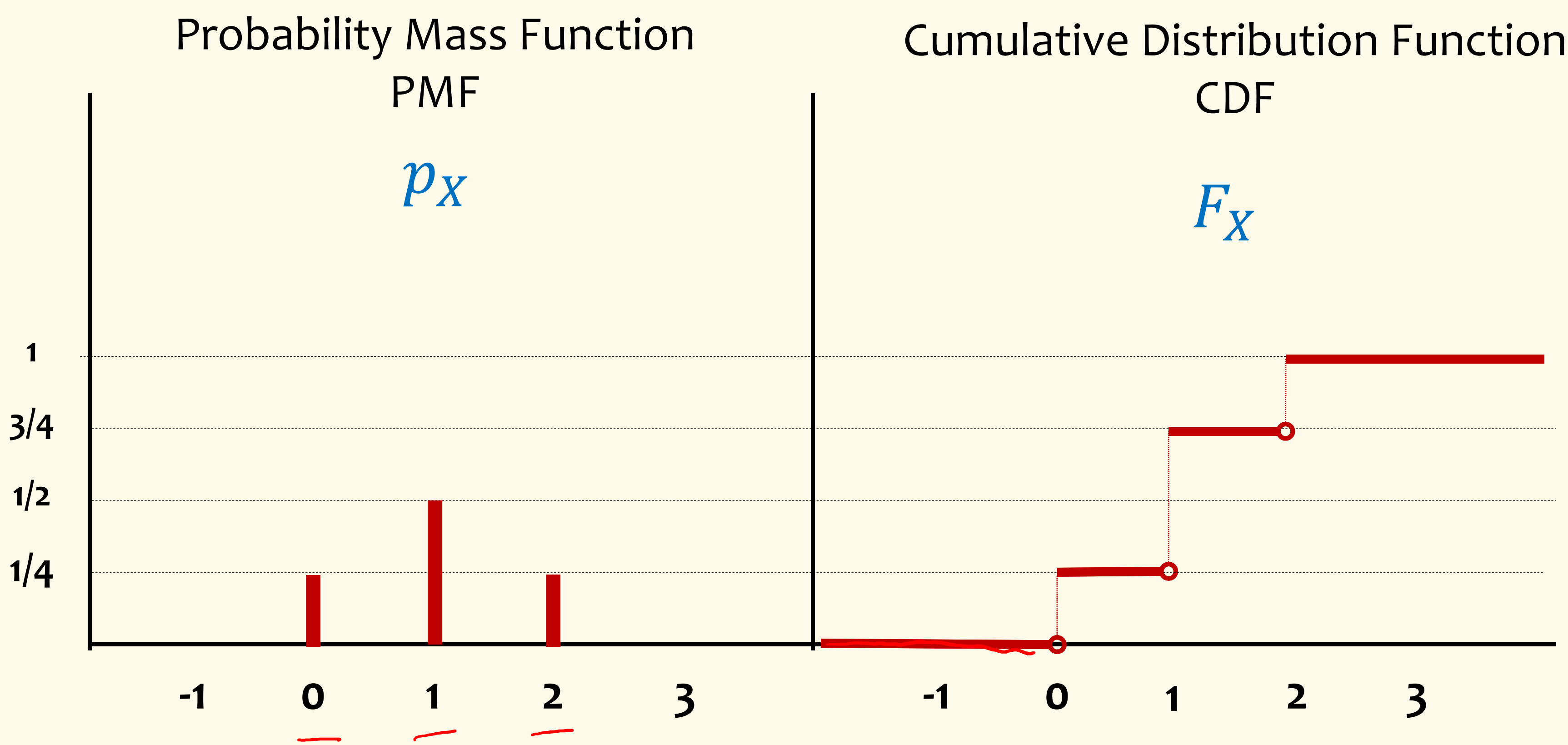
Cumulative Distribution Function (CDF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of X is the function $F_X: \mathbb{R} \to [0,1]$ that specifies for any real number x, the probability that $X \leq x$.

That is, F_X is defined by $F_X(x) = P(X \leq x)$

Example – Two fair coin flips

X = number of heads



Agenda

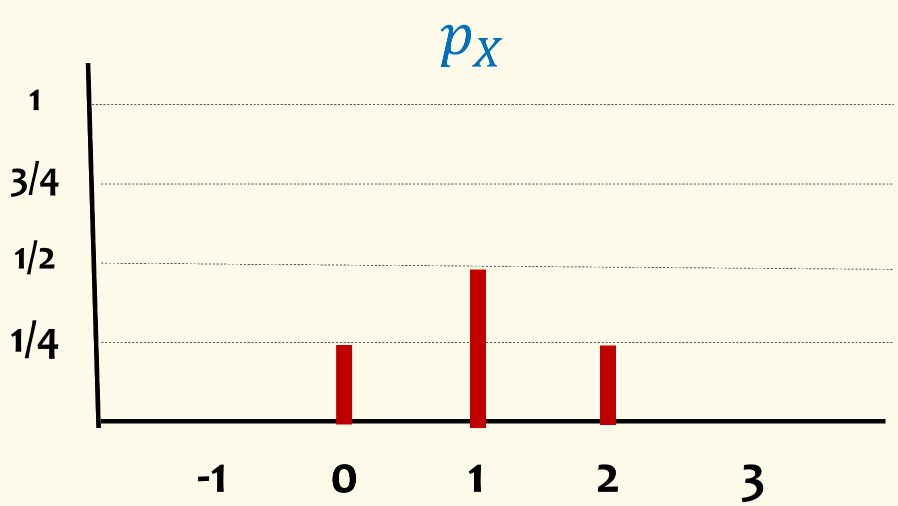
- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- Expectation

Expectation (Idea)

Example. Two fair coin flips

$\Omega = \{\mathrm{TT}, \mathrm{HT}, \mathrm{TH}, \mathrm{HH}\}$

$X =$ number of heads



- If we chose samples from Ω over and over repeatedly, how many heads would we expect to see per sample from Ω?
  – The idealized number, not the average of actual numbers seen (which will vary from the ideal)

Expected Value of a Random Variable

Definition. Given a discrete RV $X: \Omega \to \mathbb{R}$, the **expectation** or **expected value** or **mean** of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega)$$

or equivalently

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x) = \sum_{x \in \Omega_X} x \cdot p_X(x)$$

Intuition: "Weighted average" of the possible outcomes (weighted by probability)

Expected Value

Definition. The expected value of a (discrete) RV X is

$$\mathbb{E}[X] = \sum_x x \cdot p_X(x) = \sum_x x \cdot P(X = x)$$

Example. Value X of rolling one fair die

$$p_X(1) = p_X(2) = \cdots = p_X(6) = \frac{1}{6}$$

$$\mathbb{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

For the equally-likely outcomes case, this is just the average of the possible outcomes!