



Computer
Science

CSC380: Principles of Data Science

Probability 3

Xinchen Yu

Review: “probability cheatsheet”

2

Additivity:

For any *finite* or *countably infinite* sequence of disjoint events E_1, E_2, E_3, \dots ,
$$P\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} P(E_i)$$

Inclusion-exclusion rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Law of total probability: For events B_1, B_2, \dots that partitions Ω ,

$$P(A) = \sum_i P(A \cap B_i)$$

Conditional probability:

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

$(P(A|B) \neq P(B|A) \text{ in general})$

Probability chain rule:

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

Law of total probability + Conditional probability:

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(B_i)P(A|B_i) = \sum_i P(A)P(B_i|A)$$

Bayes' rule:

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

Independence:

(definition) A and B are independent if $P(A, B) = P(A)P(B)$

(property) A and B are independent if and only if $P(A|B) = P(A)$ (or $P(B|A) = P(B)$)

- Random variables
- Distribution functions
 - probability mass functions (PMF)
 - cumulative distribution function (CDF)

Random Variables

Random variables (RVs)

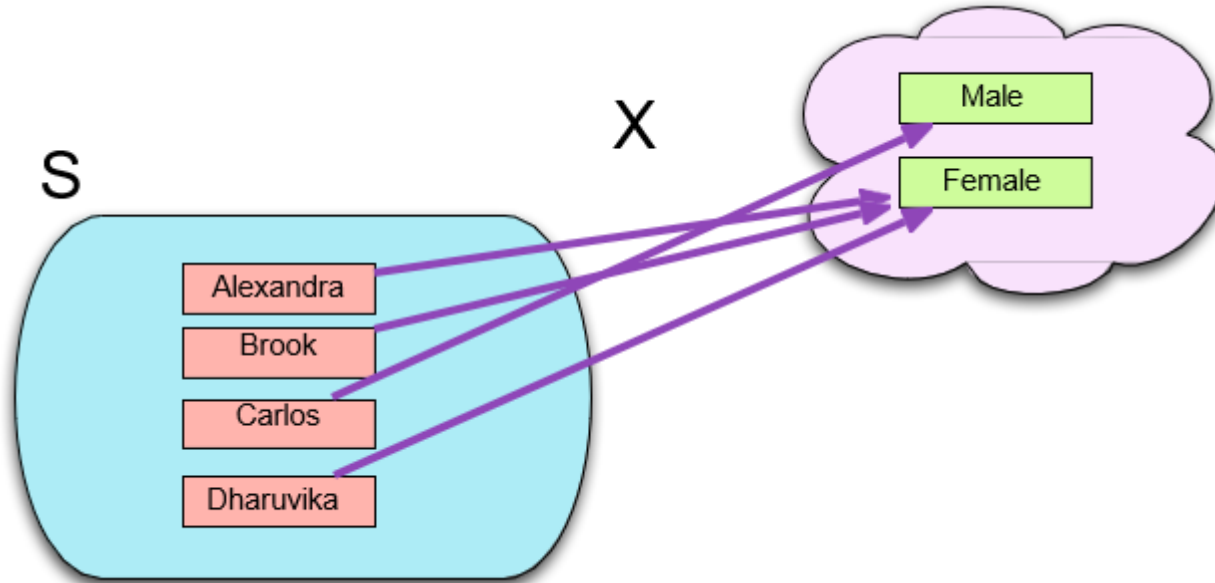
- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., gender, weight, cancer status, etc.) using a separate random variable.

Random Variable

A **random variable** connects each possible outcome in the sample space to some property of interest.

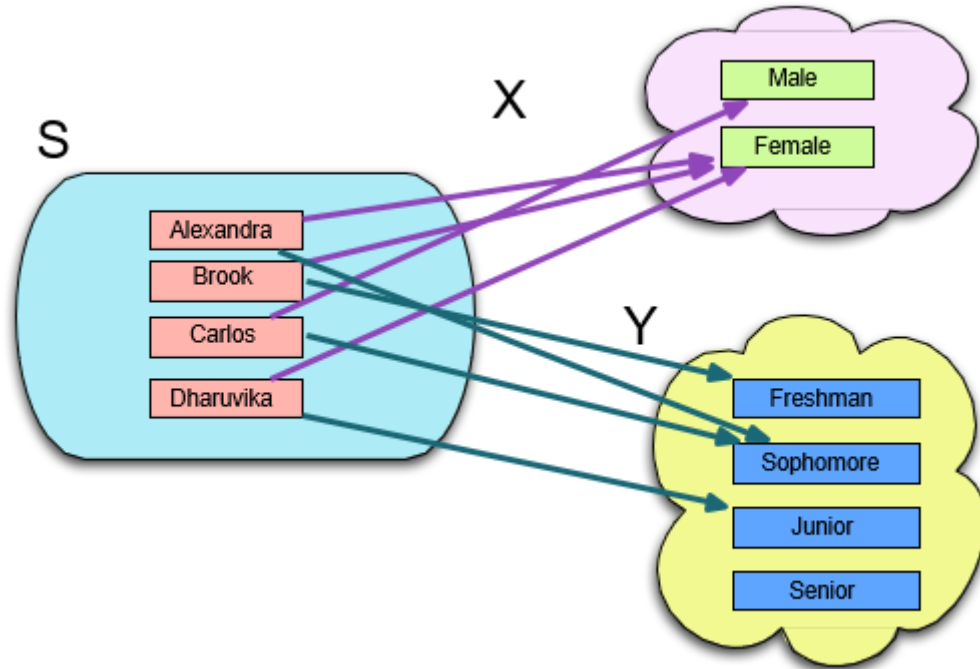
Each value of the random variable (e.g., male or female) has an associated probability.

Random Variable: Example



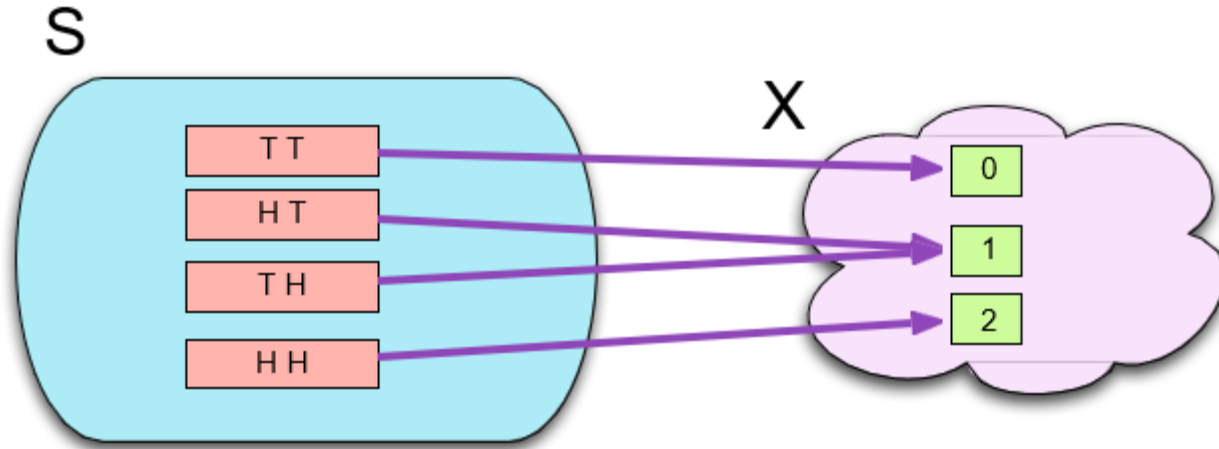
- X : people \rightarrow their genders

Random Variable: Example



- Y : people \rightarrow their class year

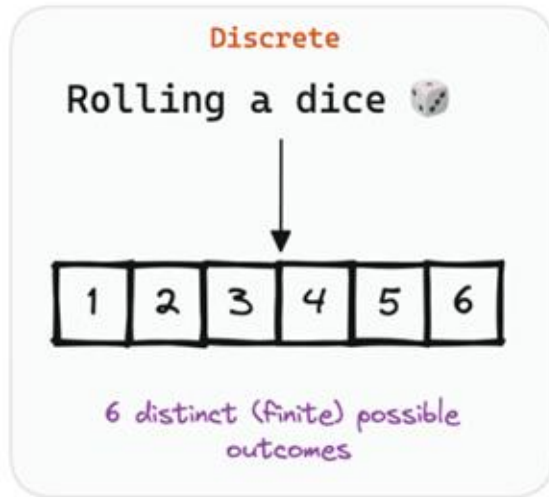
Random Variable: Example



- X : sequence of coin flips \rightarrow Number of heads

Types of Random Variables

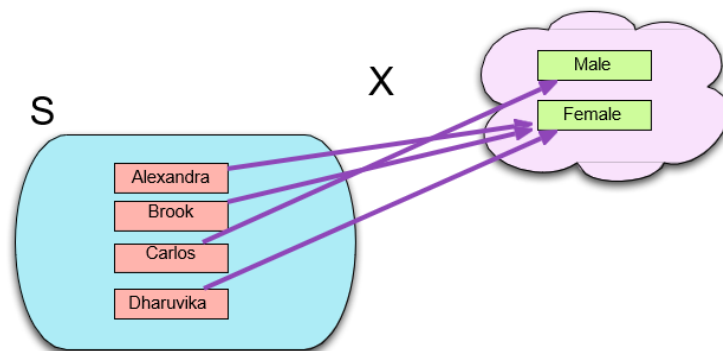
- Discrete random variable: takes a finite or countable number of distinct values.
- Continuous random variable: takes an infinite number of values within a specified range or interval.



Distribution functions

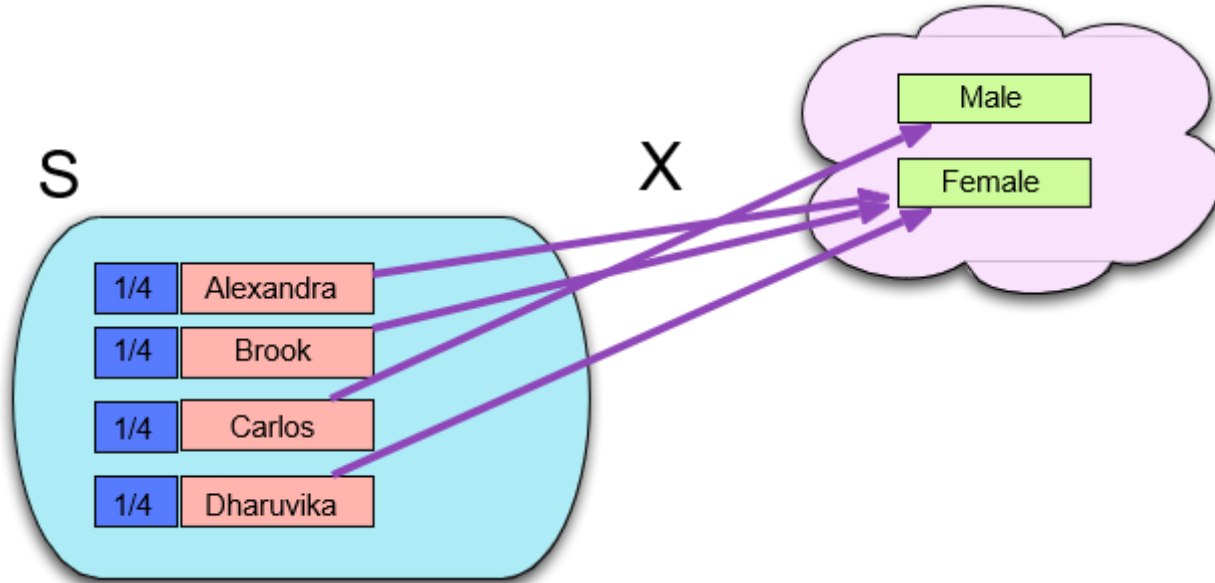
Discrete distributions

- When a random variable is discrete, its *distribution* is characterized by the probabilities assigned to each distinct value.
- The probability that the random variable takes a particular value comes from the probability associated with the set of individual outcomes that have that value.
 - This set is an event
- E.g. $P(X = \text{Female})$



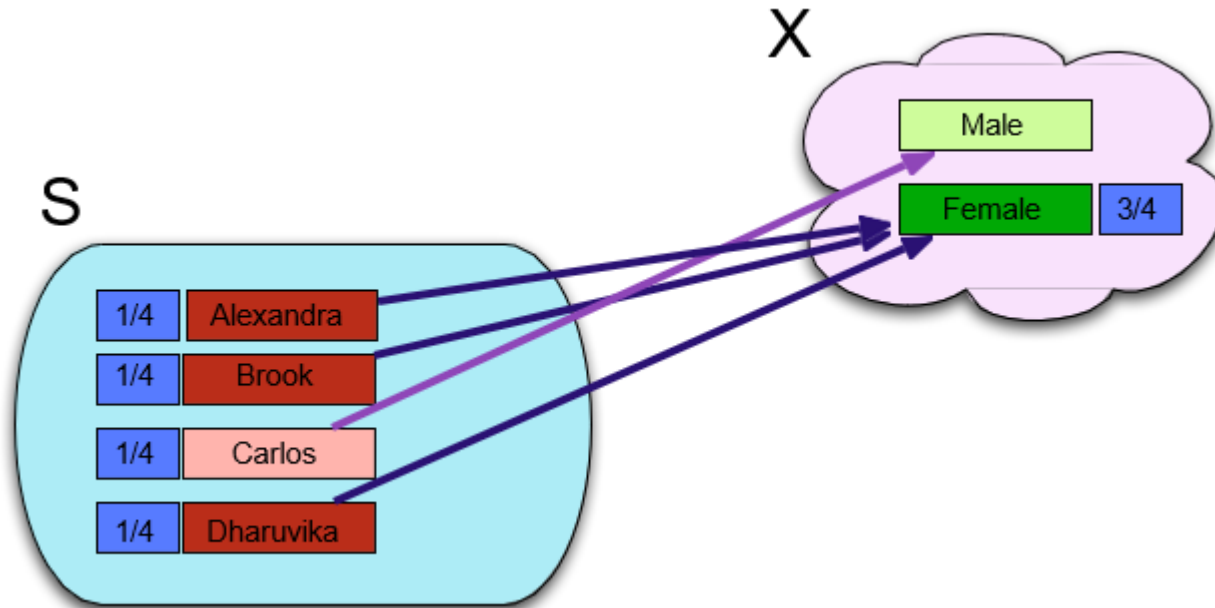
Discrete distributions

- How to find $P(X = \text{Female})$?



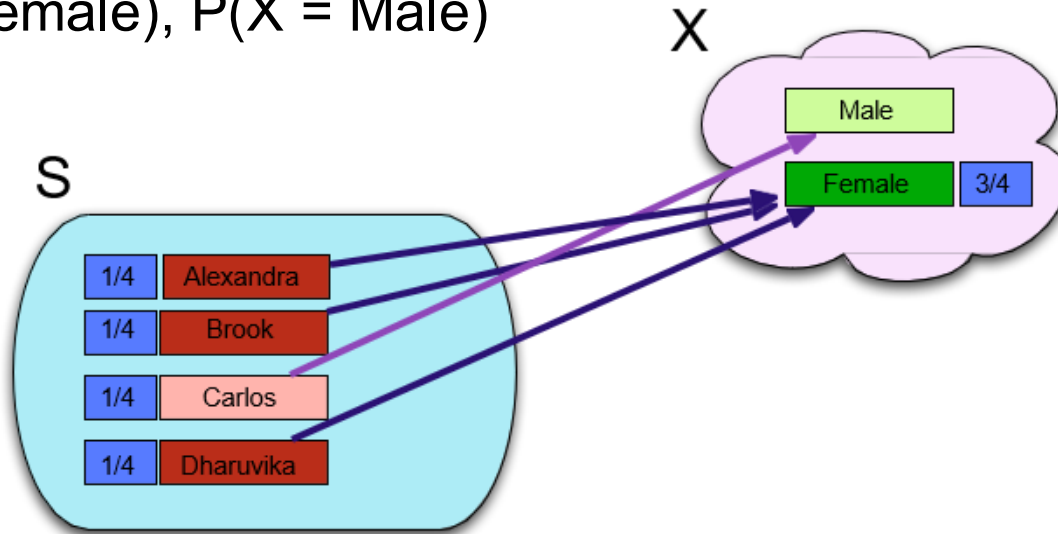
Discrete distributions

- How to find $P(X = \text{Female})$?



Discrete distributions

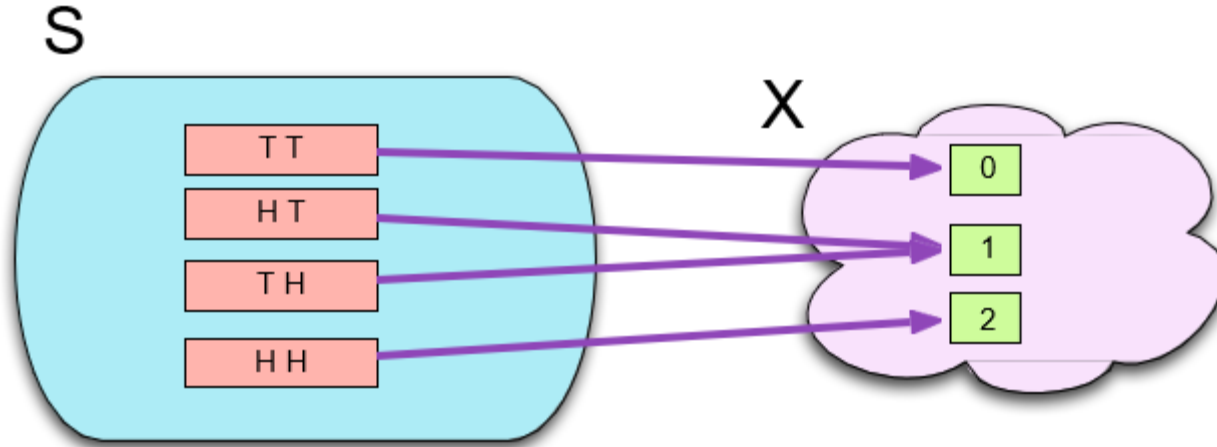
- What is the distribution of random variable X ?
 - $P(X = \text{Female})$, $P(X = \text{Male})$



x	Male	Female
$P(X = x)$	$1/4$	$3/4$

Discrete distributions

- What is the distribution of random variable X ?



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

Properties of Discrete Distributions

- We can write $P(X = x)$ to mean “The probability that the random variable X takes the value x ”.
- What must be true of these probabilities?

Properties of Discrete Distributions

1. Each $P(X = x)$ is a probability, so must be between 0 and 1.
2. The $P(X = x)$ must sum to 1 over all possible x values.

Probability Mass function (PMF)

The Probability Mass Function

A discrete random variable, X , can be characterized by its **probability mass function**, f (might sometimes write f_X if it's not clear from context which random variable we're talking about).

The PMF takes in values of the variable, and returns probabilities:

$f(x)$ is *defined* to be $P(X = x)$

PMF is a table

- Think of the PMF as a lookup table.

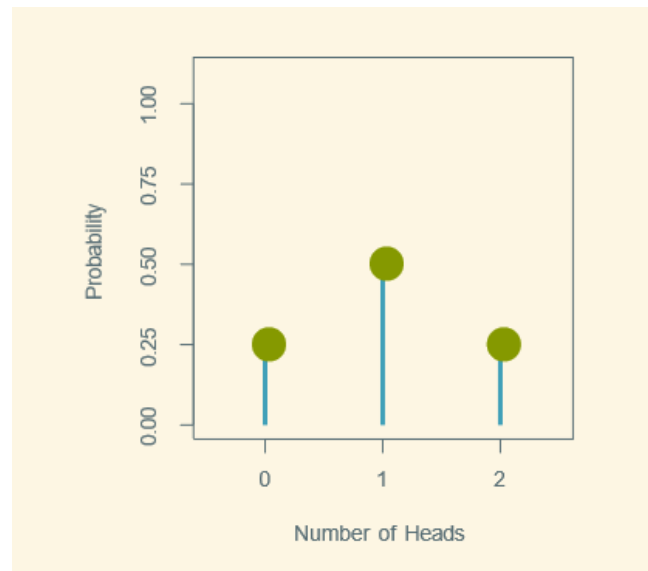
x	Male	Female
$P(X = x)$	$1/4$	$3/4$

- Best way to think of discrete random variables: they take various values, and each value has a certain probability of happening.

Visualizing discrete distributions: spike plot

Flip two coins at the same time, probability distribution of number of heads:

- Often use the spike plot
- Like a bar plot, but with probabilities, instead of frequencies or proportions, on the y-axis.



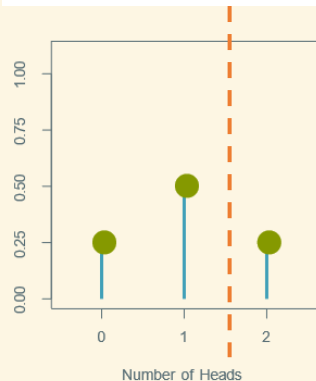
The cumulative distribution function (CDF)

- Often, we are interested in the probability of falling in some range of values.
- We can use the cumulative distribution function (CDF), which gives the “accumulated probability” up to a particular value.

The Cumulative Distribution Function

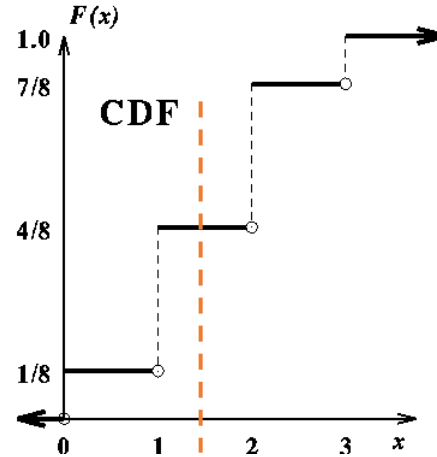
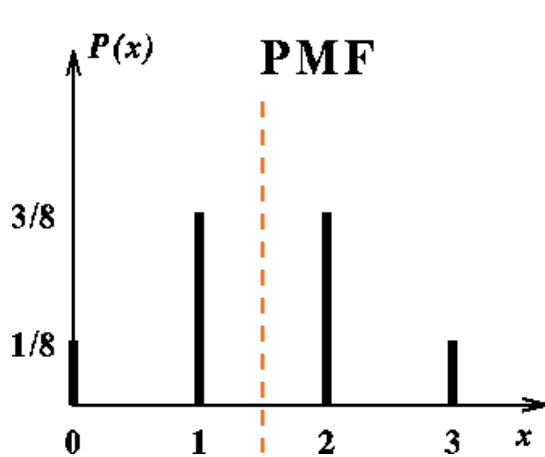
A random variable, X , can be characterized by its **cumulative distribution function**, F (or sometimes F_X if we need to be explicit), which takes values and returns *cumulative* probabilities:

$F(x)$ is defined to be $P(X \leq x)$



Relating PMF to CDF

- How can we calculate $F(x)$ from the PMF table f ?
 - Add up all the probabilities up to and including $f(x)$.
 - What is the value of $F(-0.1)$ (i.e., $P(X \leq -0.1)$)? $F(1.5)$?



- For discrete random variables, $F(x)$ *jumps* at locations with nonzero probability mass

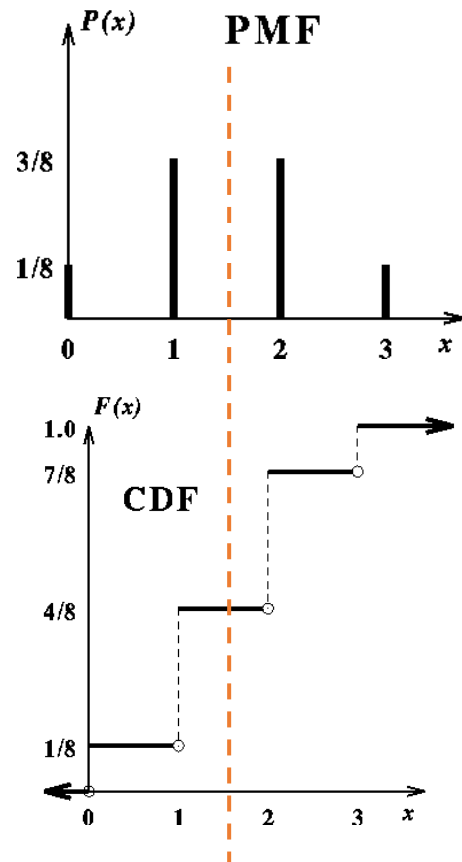
Relating PMF to CDF

- So the PMF of X is:

$$f(x) = \begin{cases} 1/8, & x = 0 \\ 3/8, & x = 1 \\ 3/8, & x = 2 \\ 1/8, & x = 3 \end{cases}$$

- We can write the CDF of X :

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{8}, & 0 \leq x < 1 \\ \frac{4}{8}, & 1 \leq x < 2 \\ \frac{7}{8}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$



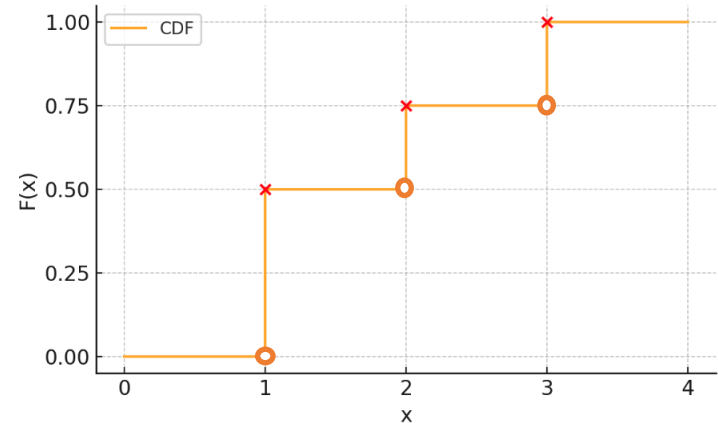
In-class activity

- Given by the PMF of X , find the CDF of X .

$$f(x) = \begin{cases} 1/2, & x = 1 \\ 1/4, & x = 2 \\ 1/4, & x = 3 \end{cases}$$

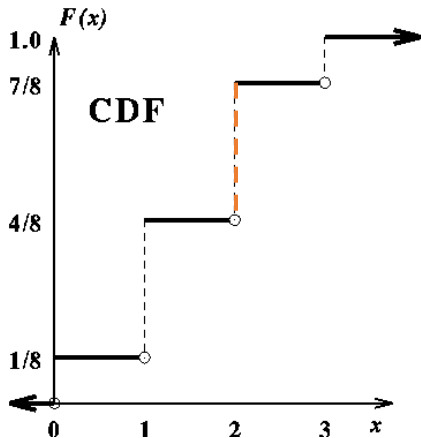
- Answer:

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{2}, & 1 \leq x < 2 \\ \frac{3}{4}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$



Relating CDF to PMF

- How could we find $f(x)$ from a cumulative distribution function F ? e.g., $f(2)$?

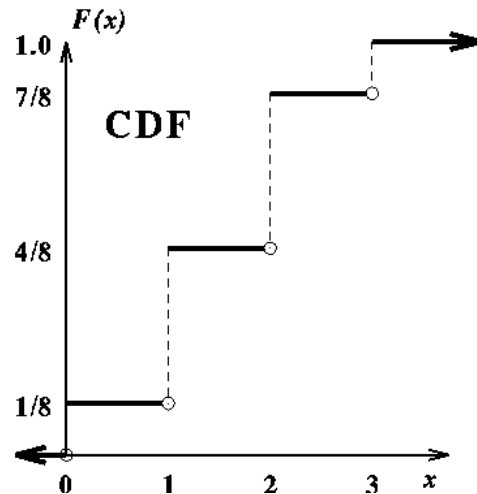


- Focus on “jumps”: $f(x) = F(x) - F(\text{jump just below } x)$
 - $f(2) = F(2) - F(1) = \frac{7}{8} - \frac{4}{8} = \frac{3}{8}$
 - $f(2.1) = F(2.1) - F(2) = \frac{7}{8} - \frac{7}{8} = 0$
 - $f(1.5) = F(1.5) - F(1) = \frac{4}{8} - \frac{4}{8} = 0$

Exercise: using CDF and PMF

Given the CDF F :

- How to calculate $P(X > x)$?
 - $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$
- How about $P(X \geq x)$?
 - $P(X \geq x) = 1 - P(X < x) = 1 - (P(X \leq x) - P(X=x))$
 - $1 - F(x) + f(x)$
 - $f(x)$ can be 0 or nonzero, depending on whether x is a jump



Exercise: using CDF and PMF

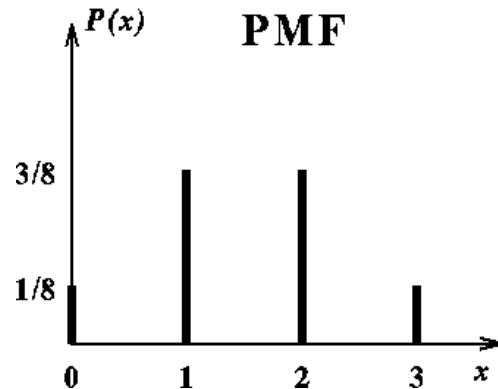
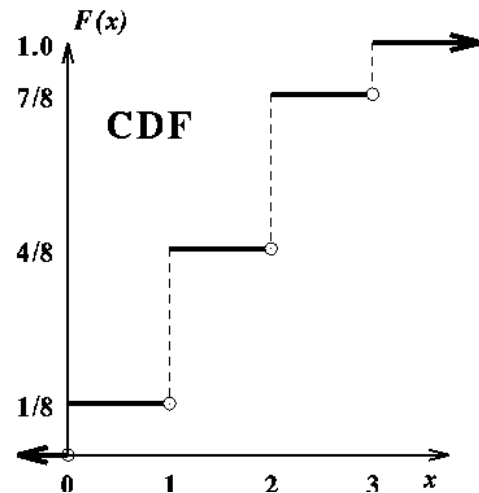
- What is $P(X \geq 2)$?
 - $P(X \geq x) = 1 - F(x) + f(x)$
 - $f(x)$ can be 0 or nonzero, depending on whether x is a jump

Using the formula:

$$\bullet \quad P(X \geq 2) = 1 - F(2) + f(2) = 1 - \frac{7}{8} + \frac{3}{8} = \frac{1}{2}$$

Another way:

$$\bullet \quad P(X \geq 2) = P(X = 2) + P(X = 3) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}$$



Exercise: using CDF and PMF

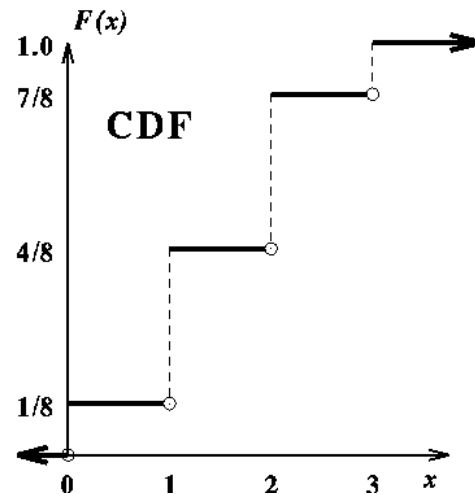
Given the CDF F :

- How to calculate $P(a < X \leq b)$?

$$= P(X \leq b) - P(X \leq a)$$

$$= F(b) - F(a)$$

- How to calculate $P(a < X < b)$?
 - (I'll leave this to you as an exercise..)



Transformations of random variables

- If X is a random variable, then $X + 5, 3X, X^2, \dots$, are all random variables
- Given any transformation function f , $f(X)$ is a random variable
- How to find the PMF of $f(X)$ based on that of X ?
 - First, find all values $f(X)$ can take
 - For each value c , try to find $P(f(X) = c)$

Examples

- Suppose X has PMF

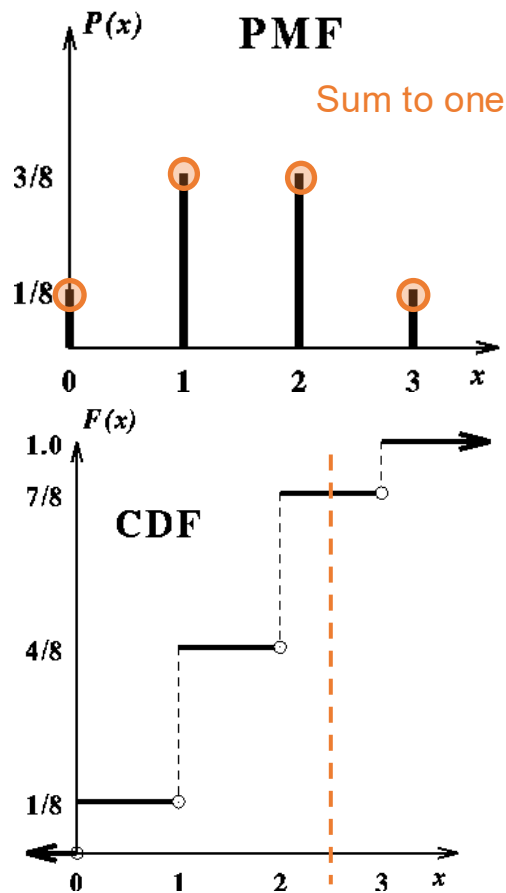
x	1	-1
$P(X = x)$	0.5	0.5

- What is the PMF of $Y = X + 5$?
 - Y can take values 6 and 4
 - $P(Y = 6) = P(X = 1) = 0.5$
 - $P(Y = 4) = P(X = -1) = 0.5$

y	6	4
$P(Y = y)$	0.5	0.5

Recap: RV, PMF and CDF

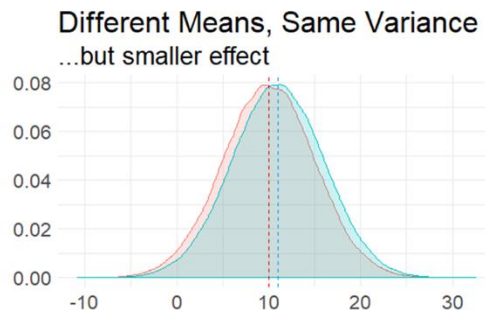
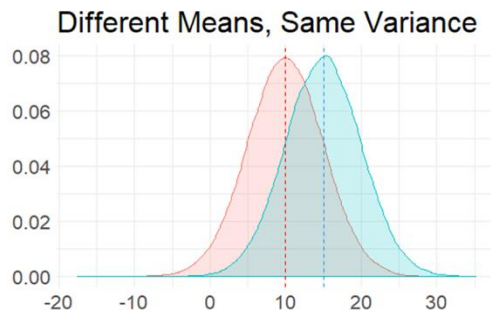
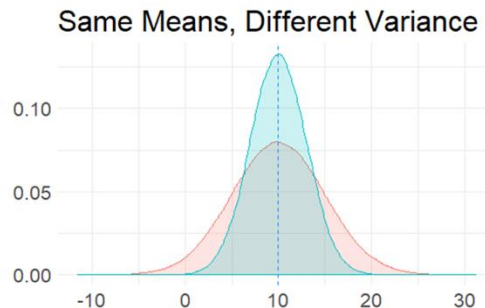
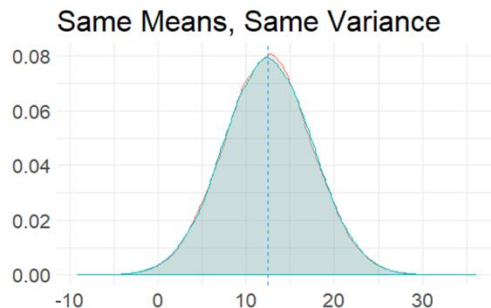
- RV: connects all outcomes to a property of interest
- A RV has a distribution, which assign a probability to each distinct value X can take
- For discrete RV X :
 - PMF: $f(x)$ defined as $P(X = x)$
 - CDF: $F(x)$ defined as $P(X \leq x)$
- Derive CDF from PMF, and vice versa
 - $f(x) = F(x) - F(\text{jump just below } x)$
 - $F(x)$: the total of all jumps (PMF values) at points less than or equal to x
- PMF of $f(X)$
 - First, find all values $f(X)$ can take
 - For each value c , try to find $P(f(X) = c)$



Mean and Variance

Summarizing random variables

- It is useful to characterize the *center* and *spread* of a probability distribution
 - “what value do we expect to occur?”, and
 - “how confident are we in our prediction?”



Mean (aka expectation, expected value)

- The mean of a random variable X is also called its *expected value*. Usually written as μ or $E[X]$.
- As with a sample mean, it represents an average over the possible values; and the average is *weighted by the probabilities*.
 - $(2 + 2 + 1 + 5)/4 = 2.5$
 - $2 * \frac{1}{2} + 1 * \frac{1}{4} + 5 * \frac{1}{4} = 2.5$
- Makes sense if you were to repeat the random process many times, the average of the observed values of X would approach $E[X]$. It doesn't mean this value will be observed directly—it's a weighted average.

Example: expected winnings at Roulette

- 38 outcomes (18 red, 18 black, 2 green: 0, 00) equally likely
- Suppose we bet on black. Define X which takes the value 1(\$ for outcomes where we win, and $-1(\text{\$})$ for outcomes where we lose.
- Its probability mass function is given by

x	-1	1
$P(X = x)$	20/38	18/38



Example: expected winnings at Roulette

- X's PMF is

x	-1	1
$P(X = x)$	20/38	18/38

- Its expected value is

$$\begin{aligned}\mu &= -1 \times P(X = -1) + 1 \times P(X = 1) \\ &= -\frac{2}{38}\end{aligned}$$

- expected value : if I play this game thousands of times, what is my average profit/loss per spin?

Example: expected winnings at Roulette

- In general we have:

Expected Value of a Discrete Random Variable

$$\mu \text{ (aka } E(X)) := \sum_x xP(X = x)$$

Summation is over all values X can take

- Ex: find the mean of the random variable with PMF

x	0	1	2
$P(X = x)$	0.7	0.2	0.1

- Answer: $0 \times 0.7 + 1 \times 0.2 + 2 \times 0.1 = 0.4$

Expectation formula

- Given RV X and its PMF, how to find $E[X + 5]$, $E[3X]$, etc?
- Idea 1: find the PMF of the transformed RV and use the definition of expectation
- Idea 2: use the following fact:

Expectation formula

$$E[f(X)] = \sum_x f(x) \cdot P(X = x)$$

Variance

- The variance, written σ^2 or $\text{Var}(X)$ or $E[(X - \mu)^2]$ is the “expected squared deviation” from the mean.
- It is a weighted average of the squared deviations corresponding to the individual values.

Variance of a Discrete Random Variable

$$\sigma^2 \text{ (aka } \text{Var}(X), \text{ aka } E((X - \mu)^2)) = \sum_x (x - \mu)^2 P(X = x)$$

- $E[(X - \mu)^2]$ – expectation of $(X - \mu)^2$, another RV

Example: Roulette

- X's PMF is

x	-1	1
$P(X = x)$	$20/38$	$18/38$

- Its expected value is $\mu = -\frac{2}{38}$

- Its variance is

$$\begin{aligned}\sigma^2 &= (-1 - \mu)^2 \cdot P(X = -1) + (1 - \mu)^2 \cdot P(X = 1) \\ &= \left(-1 - \left(-\frac{2}{38}\right)\right)^2 \times \frac{20}{38} + \left(1 - \left(-\frac{2}{38}\right)\right)^2 \times \frac{18}{38} \\ &= \dots \approx 0.997\end{aligned}$$

Standard deviation

- Just as with a sample, the standard deviation, σ , is the square root of the variance.
- E.g. in the roulette example, $\sigma = \sqrt{0.997} \approx 0.998$
 - In one spin, the “typical” variation of our balance is 0.998

Properties of expectation

- What will happen to the roulette game if we bet \$2 instead of \$1?
- The new PMF becomes
- The new expected winnings are then

x	-2	2
$P(X = x)$	20/38	18/38

$$\begin{aligned}\mu &= -2 \times P(X = -2) + 2 \times P(X = 2) \\ &= -\frac{4}{38}\end{aligned}$$

- What's the relationship between this value and the old expected value?
 - Doubling the individual values (w/o changing probs) doubles the expected value

Properties of expectation

- This works in general: if we change the values of a random variable by multiplying by a constant, the expectation gets multiplied by a constant.
- To see this, recall the expectation formula:

$$E[f(X)] = \sum_x f(x) \cdot P(X = x)$$

$$E[aX] = \sum_x ax P(X = x) = a \sum_x x P(X = x) = aE[X]$$

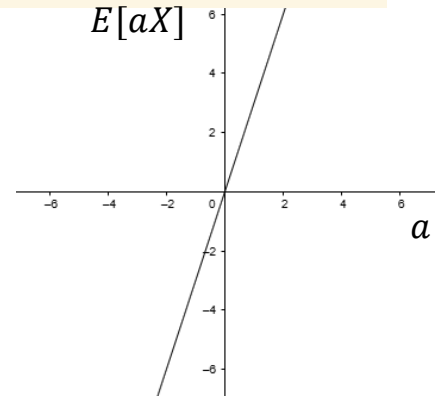
Properties of expectation

Property of Expectation

Multiplying a random variable by a constant scales the expected value by the same constant:

$$E(aX) = aE(X)$$

- Sometimes called “linearity of expectation”



Properties of Variance

- What will happen to the variance if we multiply every value of a random variable by a constant a ?
- This is as if we increase our bet in the roulette game

x	-2	2
$P(X = x)$	20/38	18/38

- Variance = expected *squared* deviation
- All squared deviations are scaled by a^2 , making variance also scaled by a^2

Properties of Variance

- Its old variance is

$$\begin{aligned}\sigma^2 &= (-1 - \mu)^2 \cdot P(X = -1) + (1 - \mu)^2 \cdot P(X = 1) \\ &= \left(-1 - \left(-\frac{2}{38}\right)\right)^2 \times \frac{20}{38} + \left(1 - \left(-\frac{2}{38}\right)\right)^2 \times \frac{18}{38} \\ &= \dots \approx 0.997\end{aligned}$$

- Its new variance is

$$\begin{aligned}\sigma^2 &= (-2 - 2\mu)^2 \cdot P(X = -2) + (2 - 2\mu)^2 \cdot P(X = 2) \\ &= 4 \times \left(-1 - \left(-\frac{2}{38}\right)\right)^2 \times \frac{20}{38} + 4 \times \left(1 - \left(-\frac{2}{38}\right)\right)^2 \times \frac{18}{38} \\ &= \dots \approx 4 \times 0.997\end{aligned}$$

Properties of Variance

Property of Variance

If the values of a random variable are multiplied by a constant, a , then the variance gets multiplied by a^2 .

- In other words, $\text{Var}(aX) = a^2 \text{Var}(X)$
- How would standard deviation change accordingly?
 - scaled by $|a|$ (!)

Properties of Variance

Alternative formula for finding variance

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

This sometimes simplifies calculations quite a bit

Example X has PMF

- $E[X^2] = 1$
- $E[X] = -\frac{2}{38}$
- $\Rightarrow \text{Var}(X) = 1 - \left(\frac{2}{38}\right)^2 = 0.997$

x	-1	1
$P(X = x)$	20/38	18/38

Example Discrete Random Variables

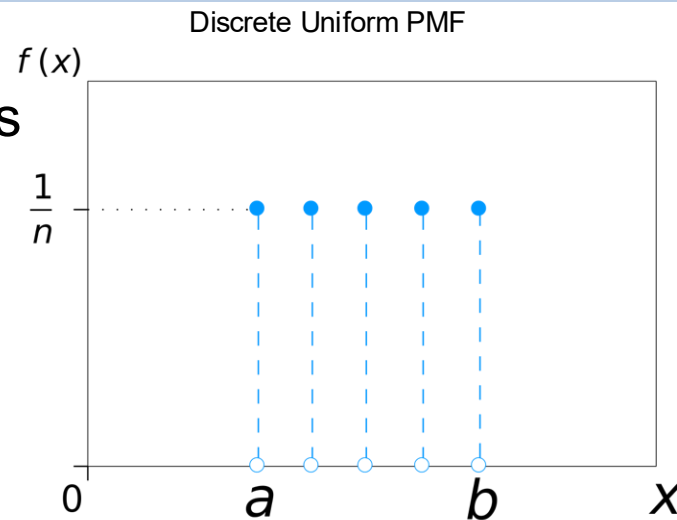
Uniform distribution over a set

More generally, consider $S = \{v_1, v_2, \dots, v_N\}$; X is drawn from the uniform distribution of S , then

$$P(X = k) = \begin{cases} \frac{1}{N} & \text{if } k \in \{v_1, v_2, \dots, v_N\} \\ 0 & \text{otherwise} \end{cases}$$

We denote this by $X \sim \text{Uniform}(S)$

- Selecting a student from a class
- Drawing a card from a shuffled deck
- Choosing a letter from the alphabet



To generate a sample from a uniform discrete distribution,

```
random.choice(a, size=None, replace=True, p=None)
```

Generates a random sample from a given 1-D array

```
numpy.random.choice([2,5,6])
```

Example output: 2

Binomial distribution

- Suppose we perform n repeated independent trials, each with success probability p , what is the distribution of the number of successes X ?

- What values can X take?

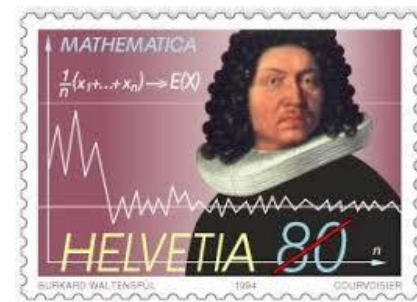
$$m = 0, 1, \dots, n$$

- We have seen that $P(X = m) =$

$$\binom{n}{m} \cdot p^m (1 - p)^{n-m}$$

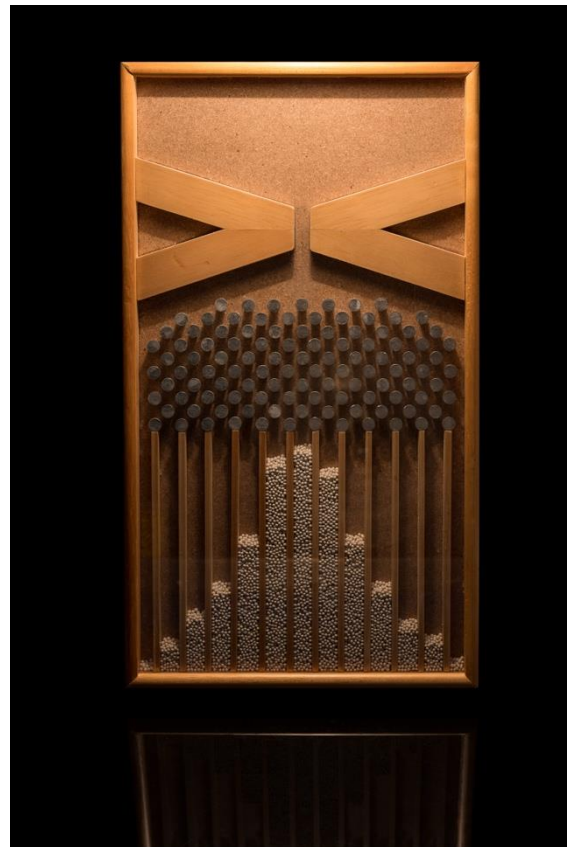
- In this case, X is said to be drawn from a *binomial distribution*, denoted by

$$X \sim \text{Bin}(n, p)$$



Galton Boards

- Illustration of binomial distribution
- Bead has 10 chances hitting pegs (10 rows of pegs)
- each time a peg is hit, bead randomly bounces to the left or the right with equal probabilities
- Number of times it bounces to the left:
$$X \sim \text{Bin}(10, 0.5)$$

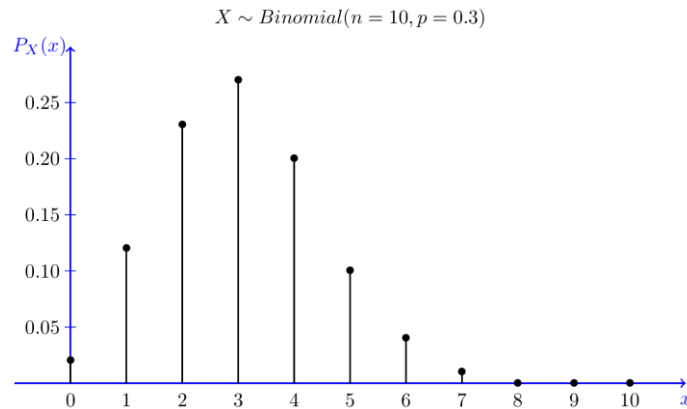


Binomial distribution

- $X \sim \text{Bin}(n, p)$
- X 's PMF is “Bell-shaped”

Facts:

- $E[X] = E[n \cdot X_i] = n \cdot E[X_i] = np$
- $\text{Var}[X] = np(1 - p)$
 - Small when p is close to 0 or 1



Bernoulli distribution

- What does $X \sim \text{Bin}(1, p)$ mean?

x	0	1
$P(X = x)$	$1-p$	p

- This is called the Bernoulli distribution with parameter p , abbreviated as Bernoulli(p)
- $E[X] = 0 \cdot (1 - p) + 1 \cdot p = p$

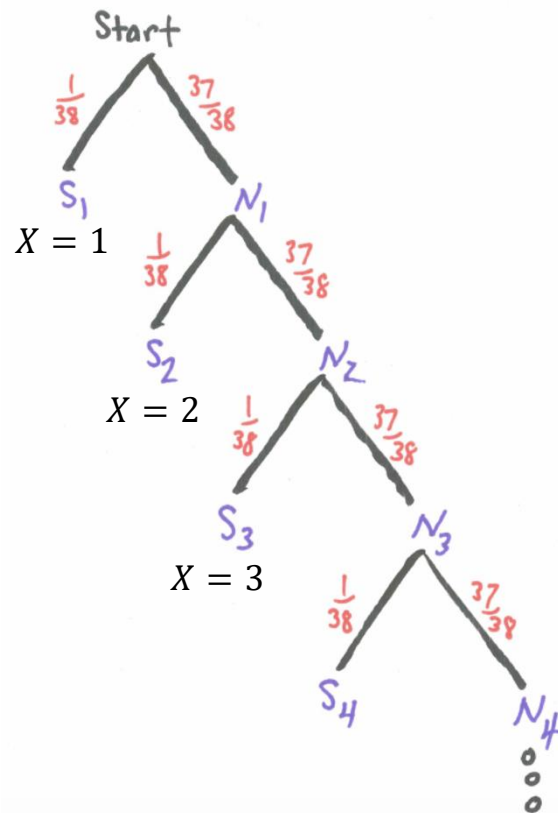


Geometric distribution

- Suppose we perform repeated independent trials with success probability p . What is the distribution of X , the number of trials needed to get a success? (related to Q4 in HW3)
- Applications:
 - Call center: # calls before encountering first dissatisfied customer
 - Basketball: # shots before scoring the first
 - Networking: # attempts before a successful transmission
 - Gambling: # plays before first win

Geometric distribution

- How to find $P(X = x)$?
- Let's draw a probability tree..
- Example: $p = \frac{1}{38}$ (roulette)
- $P(X = 1) = p$
- $P(X = 2) = (1 - p) p$
- $P(X = 3) = (1 - p)^2 p$
- ...



Geometric distribution

- In conclusion,

$$P(X = x) = p (1 - p)^{x-1}$$

for $x = 1, 2, \dots$

Fact:

- $E[X] = \frac{1}{p}$
- $\text{Var}[X] = \frac{1-p}{p^2}$
 - Smaller when p closes to 1

