Computer Science

CSC380: Principles of Data Science

Probability 3

Xinchen Yu

Review: "probability cheatsheet"

Additivity:

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

Inclusion-exclusion rule: $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Law of total probability: For events $B_1, B_2, \ldots$ that partitions Ω, $$P(A) = \sum_i P(A \cap B_i)$$

Conditional probability: $$P(A|B) \coloneqq \frac{P(A \cap B)}{P(B)}$$ ($P(A|B) \neq P(B|A)$ 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 $P(B|A) = P(B)$)

Outline

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

Random Variables

Random variables (RVs)

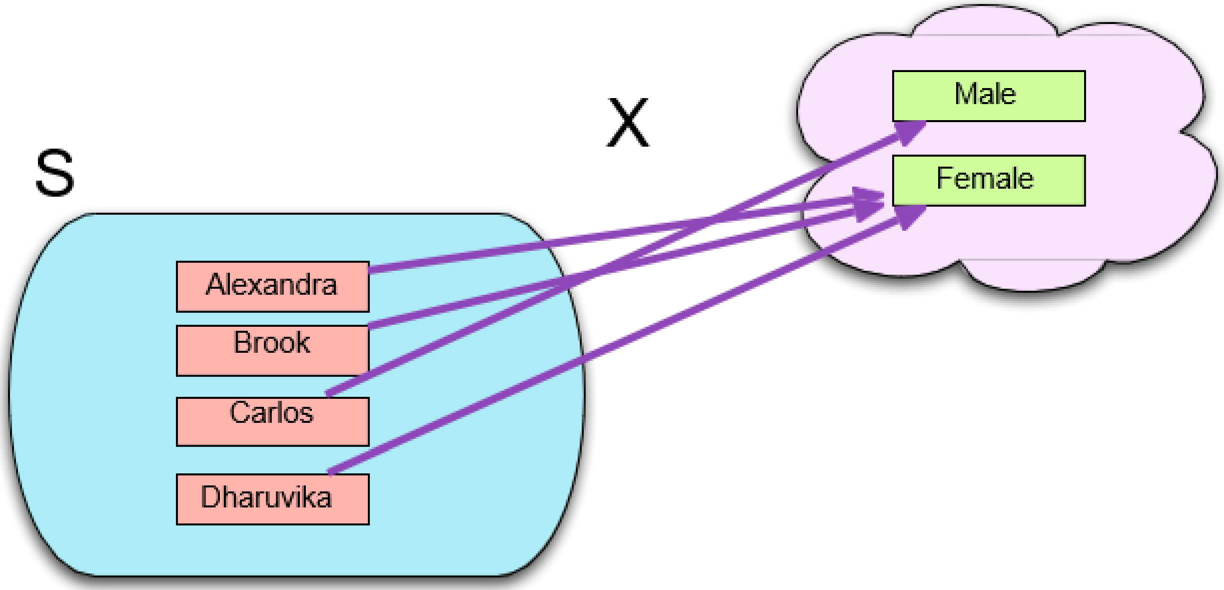
- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., gender, weight, cancer status, etc.) using a separate random variable.

Random Variable

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

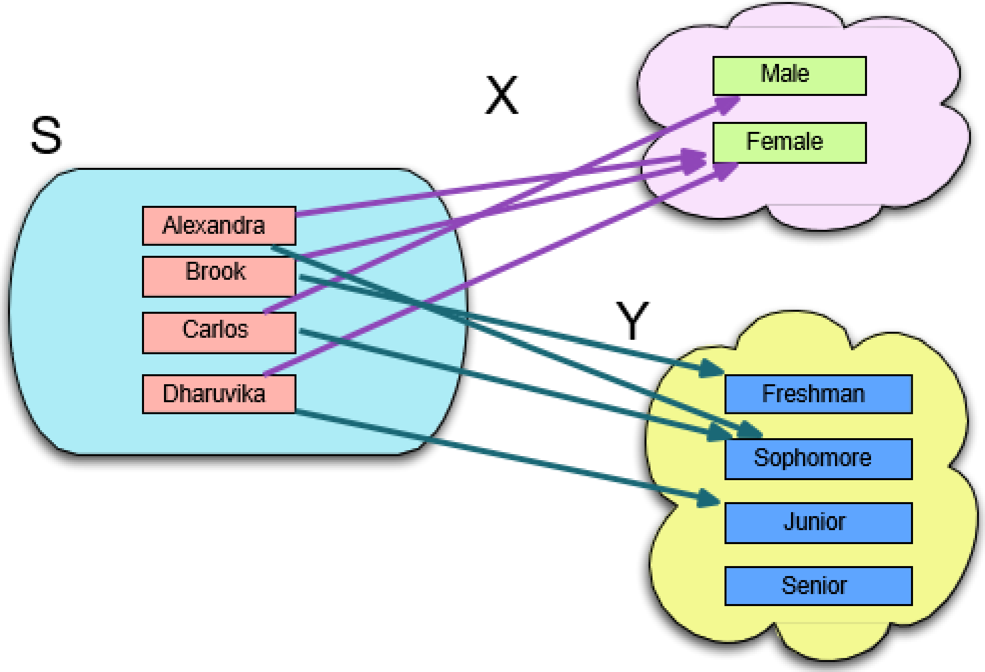
Each value of the random variable (e.g., male or female) has an associated probability.

Random Variable: Example



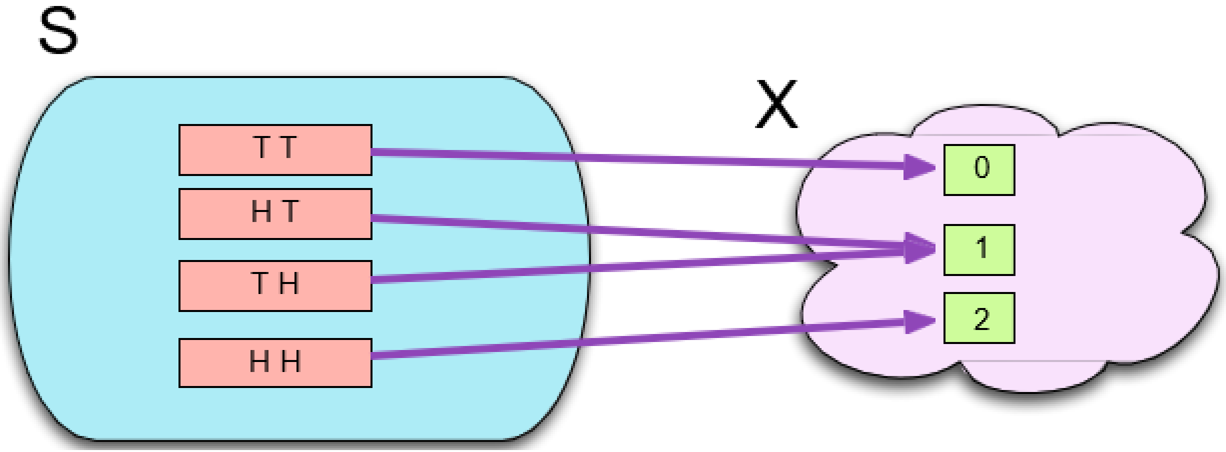
- X: people -> their genders

Random Variable: Example



- Y: people -> their class year

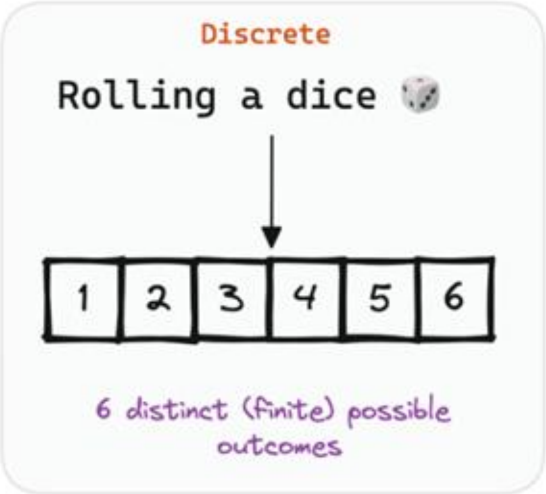
Random Variable: Example

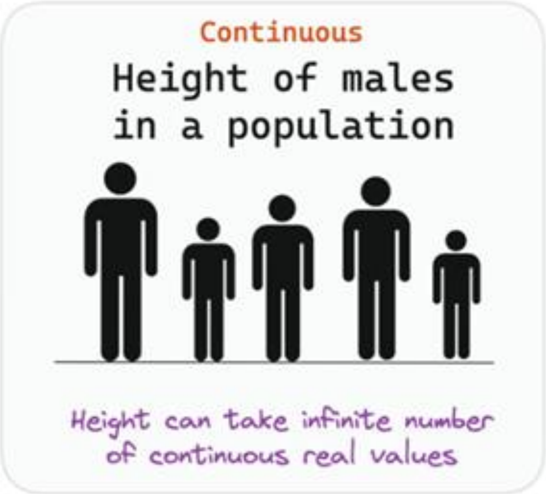


- X: sequence of coin flips -> Number of heads

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.





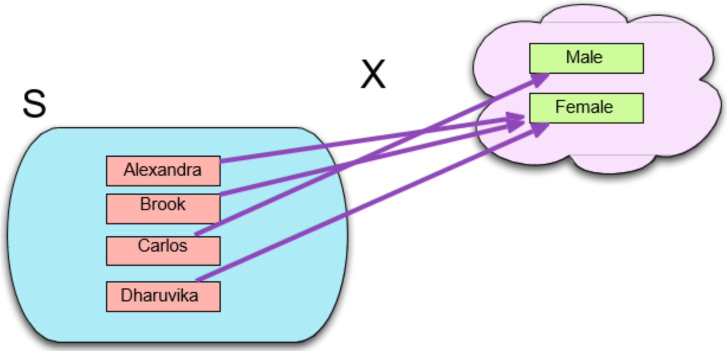
Distribution functions

Discrete distributions

- When a random variable is discrete, its *distribution* is characterized by the probabilities assigned to each distinct value.

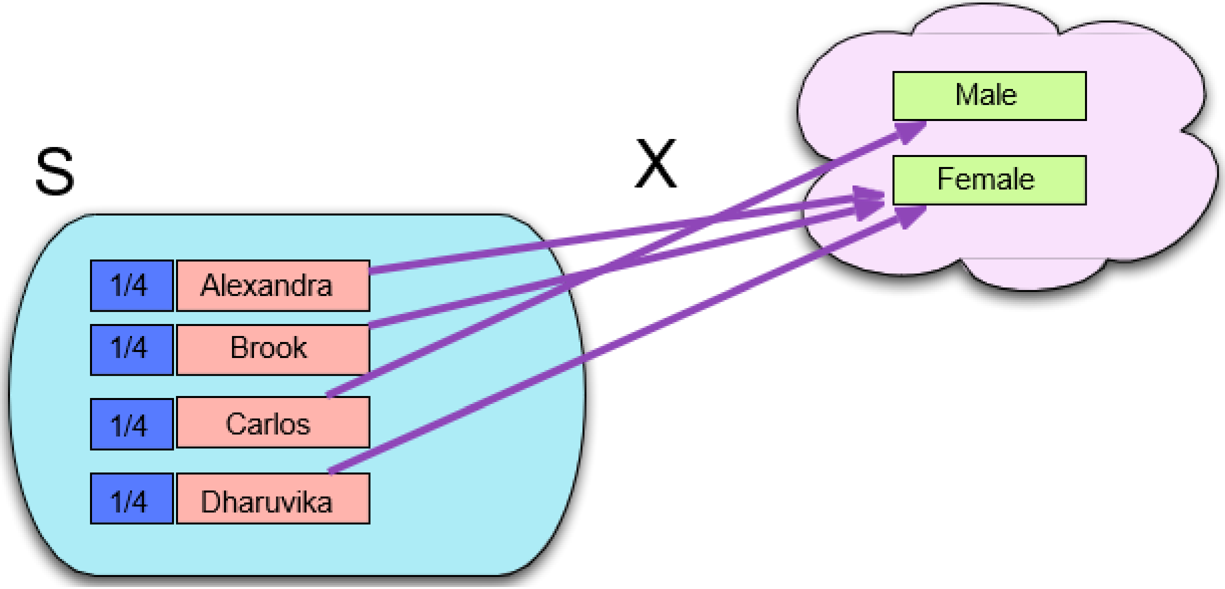
- The probability that the random variable takes a particular value comes from the probability associated with the set of individual outcomes that have that value.
    - This set is an event

- E.g. P(X = Female)



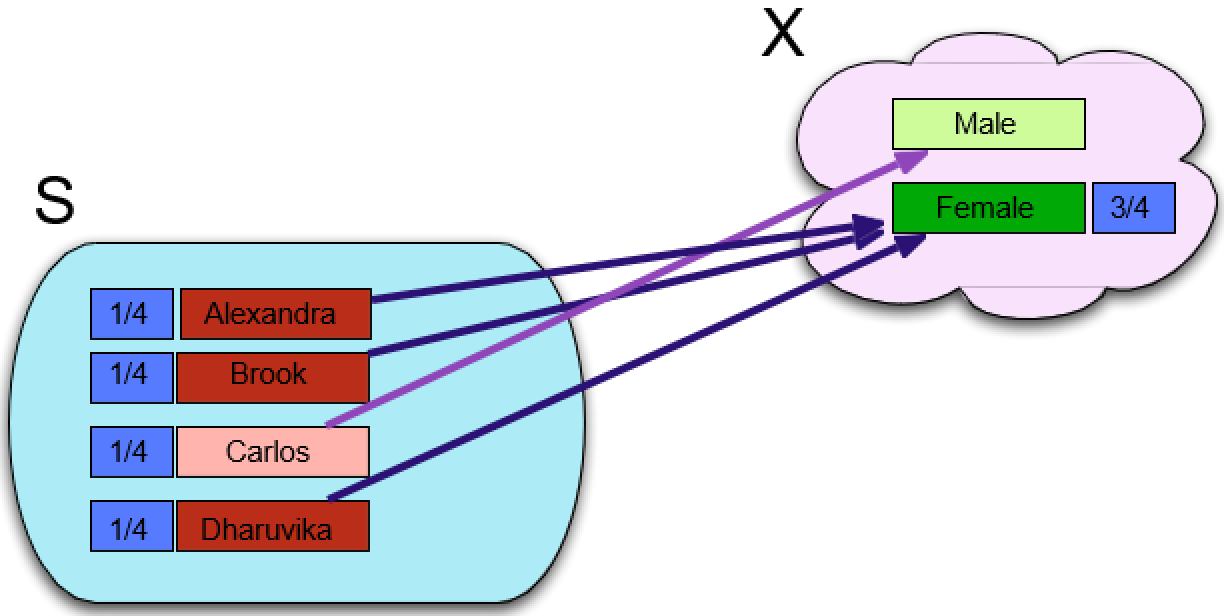
Discrete distributions

- How to find P(X = Female)?



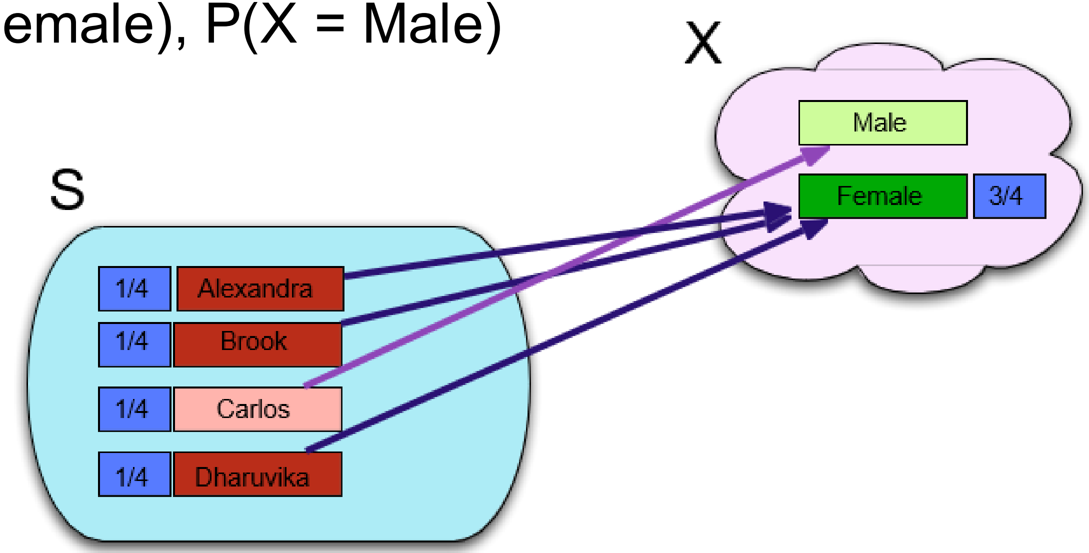
Discrete distributions

- How to find P(X = Female)?



Discrete distributions

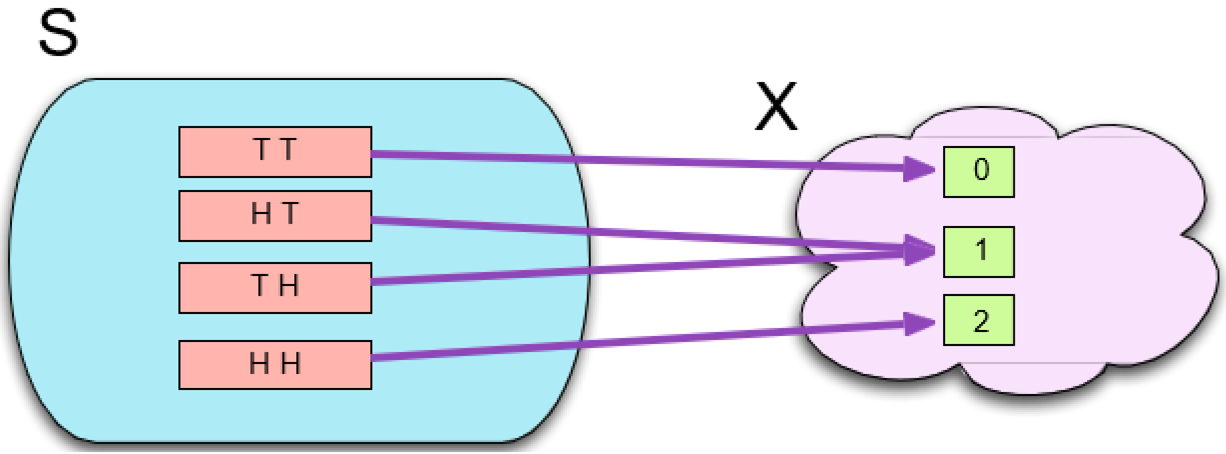
- What is the distribution of random variable X?
  - P(X = Female), P(X = Male)



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:

$$f(x) \text{ is } \textit{defined} \text{ to be } P(X = x)$$

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.

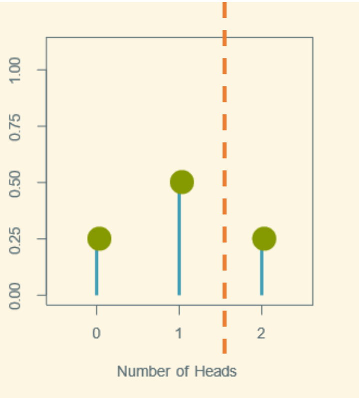
The cumulative distribution function (CDF)

- 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

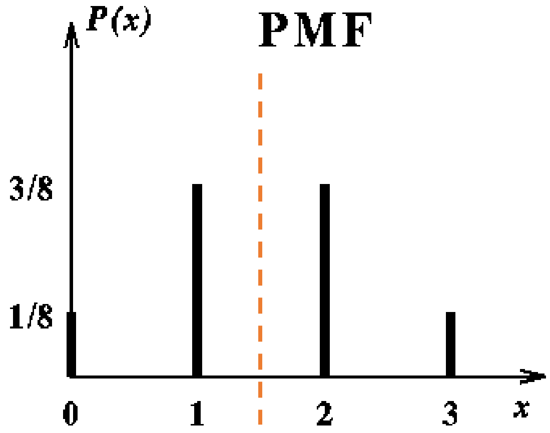
A random variable, X, can be characterized by its **cumulative distribution function**, F (or sometimes F_X if we need to be explicit), which takes values and returns *cumulative* probabilities:

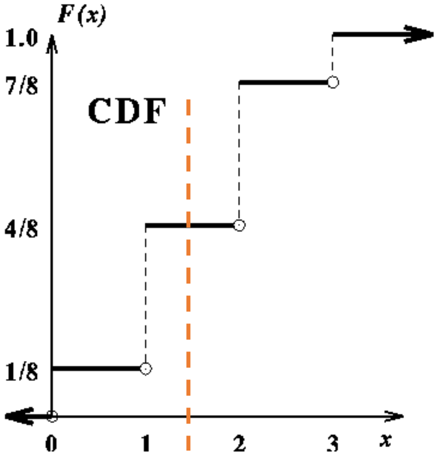
$$F(x) \text{ is defined to be } P(X \leq x)$$



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., $\mathrm{P}(\mathrm{X} \leq -0.1)$)? F(1.5)?





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

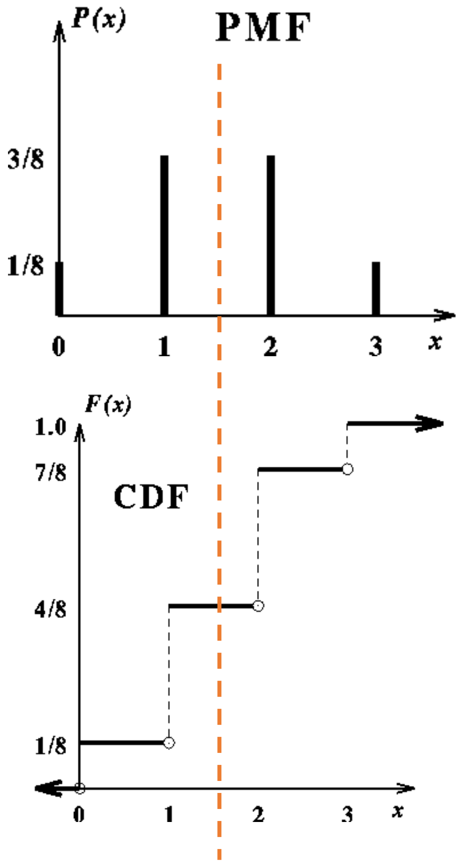
Relating PMF to CDF

- 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:

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



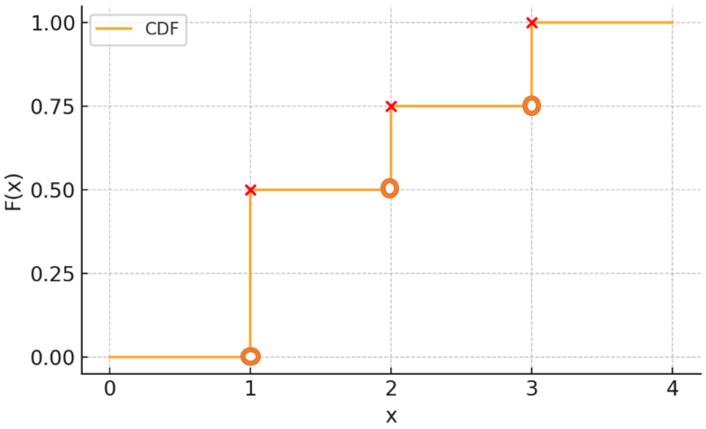
In-class activity

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

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

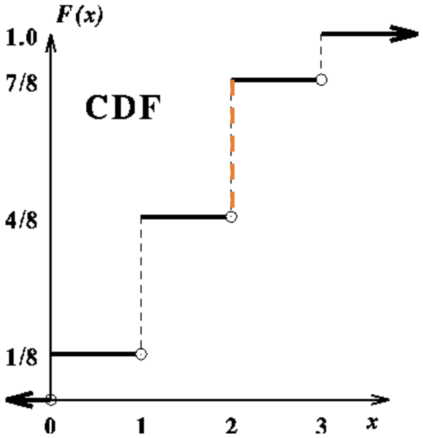
- Answer:

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



Relating CDF to PMF

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



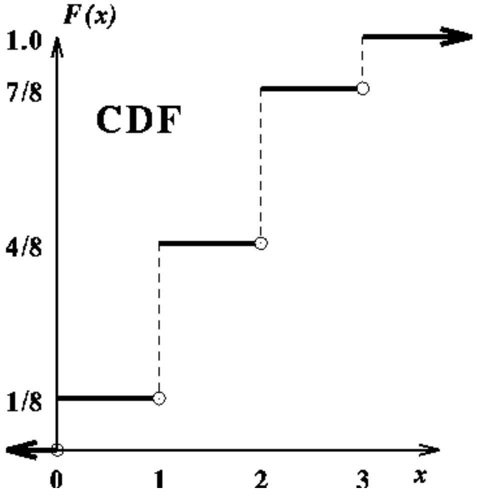
- Focus on "jumps": f(x) = F(x) – F(jump just below x)
  - $\mathrm{f}(2) = F(2) - F(1) = \frac{7}{8} - \frac{4}{8} = \frac{3}{8}$
  - $\mathrm{f}(2.1) = F(2.1) - F(2) = \frac{7}{8} - \frac{7}{8} = 0$
  - $\mathrm{f}(1.5) = F(1.5) - F(1) = \frac{4}{8} - \frac{4}{8} = 0$

Exercise: using CDF and PMF

Given the CDF F:

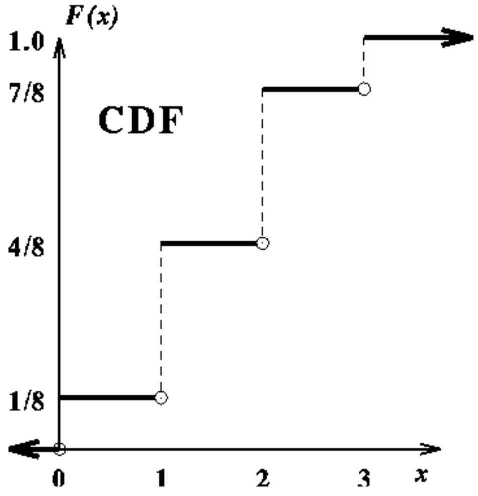
- How to calculate $P(X > x)$?
  - $P(X > x) = 1 – P(X \le x) = 1 – F(x)$

- How about $P(X \ge x)$?
  - $P(X \ge x) = 1 – P(X < x) = 1 – (P(X \le x) – P(X=x))$
  - $1 – F(x) + f(x)$
  - $f(x)$ 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

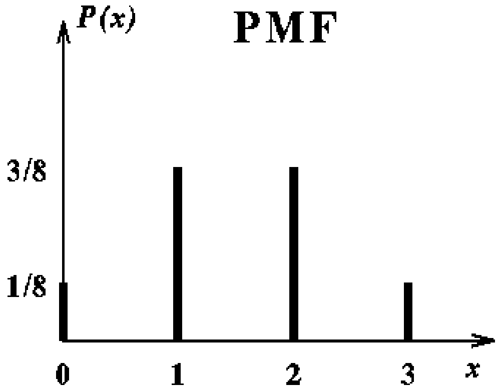


Using the formula:

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

Another way:

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



Exercise: using CDF and PMF

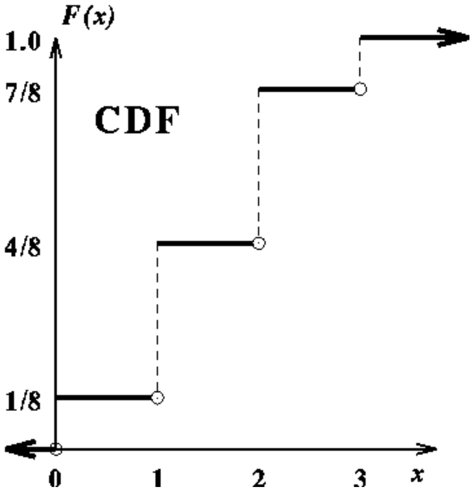
Given the CDF F:

- How to calculate P(a < X ≤ b)?

  = P(X ≤ b) - P(X ≤ 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 $X + 5, 3X, X^2, \ldots$, 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 $\mathrm{P}(f(X) = c)$

Examples

- Suppose X has PMF

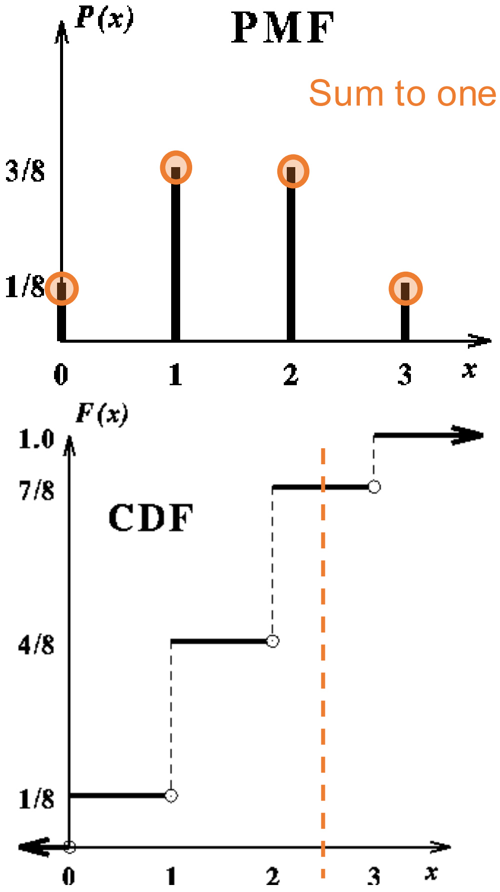
x	1	-1
$P(X = x)$	0.5	0.5

- 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

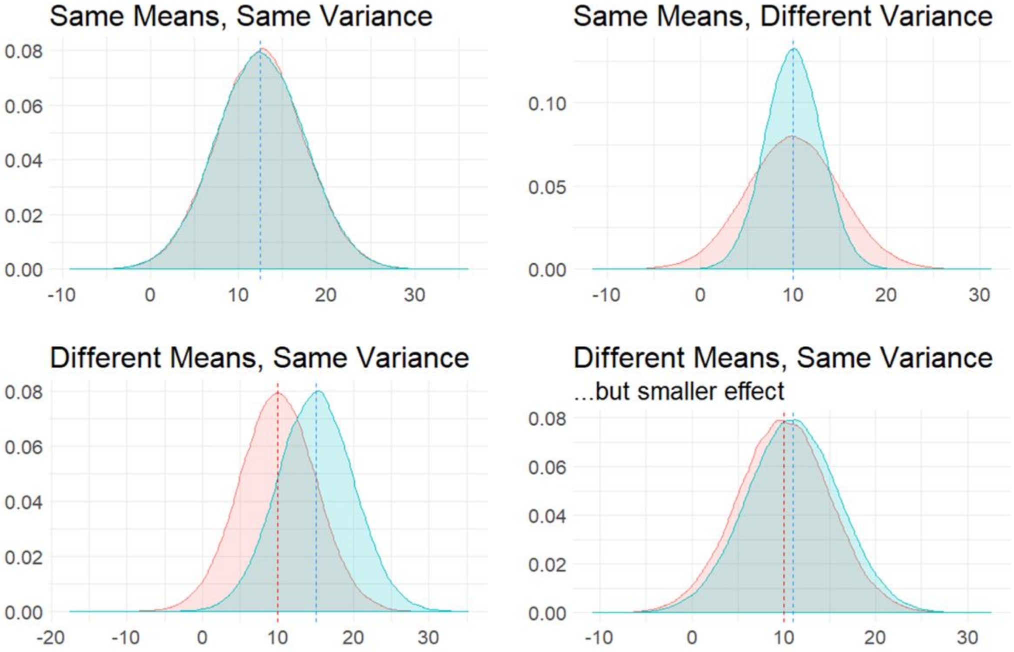
- RV: connects all outcomes to a property of interest
- A RV has a distribution, which assign a probability to each distinct value X can take
- For discrete RV X:
  - PMF: $\mathrm{f}(x)$ defined as $\mathrm{P}(\mathrm{X} = x)$
  - CDF: $\mathrm{F}(x)$ defined as $\mathrm{P}(\mathrm{X} \leq x)$
- Derive CDF from PMF, and vice versa
  - f(x) = F(x) – F(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 $\mathrm{P}(f(X) = c)$



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 $\mathrm{E}[X]$.

- 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 * ½ + 1 * ¼ + 5 * ¼ = 2.5

- Makes sense if you were to repeat the random process many times, the average of the observed values of X would approach $\mathrm{E}[X]$. 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($) for outcomes where we win, and −1($) 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

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

- 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

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

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 x 0.7 + 1 x 0.2 + 2 x 0.1 = 0.4

Expectation formula

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

Expectation formula

$$\mathrm{E}[f(X)] = \sum_{x} f(x) \cdot P(X = x)$$

Expectation formula: example

- Suppose X has PMF
- Find: $\mathrm{E}[X+5], \mathrm{E}[X^2]$

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

Expectation formula

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

- $\mathrm{E}[X+5] = (1+5) \times 0.5 + (-1+5) \times 0.5 = 5$
- $\mathrm{E}[X^2] = 1^2 \times 0.5 + (-1)^2 \times 0.5 = 1$

Variance

- The variance, written σ^2 or $\mathrm{Var}(X)$ or $\mathrm{E}[(X-\mu)^2]$ 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

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

- $\mathrm{E}[(X-\mu)^2]$ – 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

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

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

Exercise

- Find the mean and variance for the random variable with PMF given by

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

Ans:

- $\mu = 0 \times 0.7 + 1 \times 0.2 + 2 \times 0.1 = 0.4$
- $\sigma^2 = 0.4^2 \times 0.7 + 0.6^2 \times 0.2 + 1.6^2 \times 0.1$

  $= 0.44$
- For a random variable X, when is its σ^2 zero?

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

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

- 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:

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

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

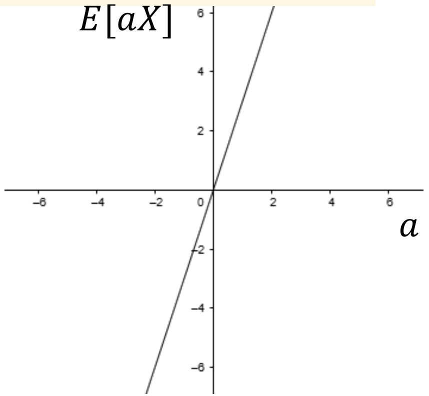
Properties of expectation

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

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

- Its new variance is

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

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

Properties of Variance

Alternative formula for finding variance

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

This sometimes simplifies calculations quite a bit

Example X has PMF

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

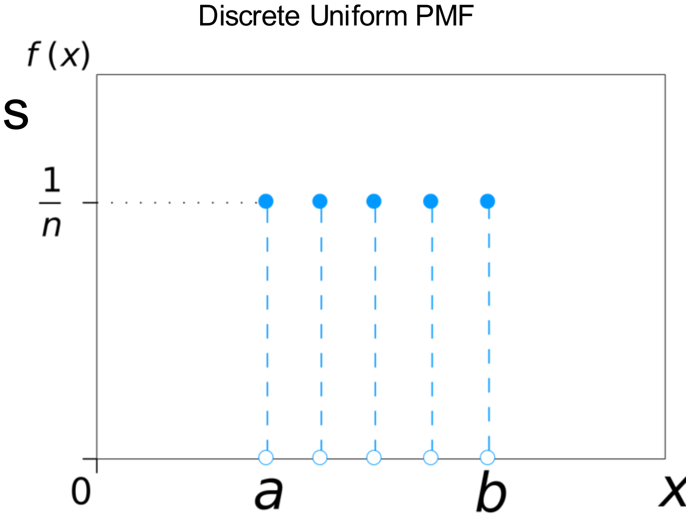
- $\mathrm{E}[X^2] = 1$
- $\mathrm{E}[X] = -\frac{2}{38}$
- $\Rightarrow \mathrm{Var}(X) = 1 - \left(\frac{2}{38}\right)^2 = 0.997$

Example Discrete Random Variables

Uniform distribution over a set

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

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



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

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

numpy.random

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?

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

- 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

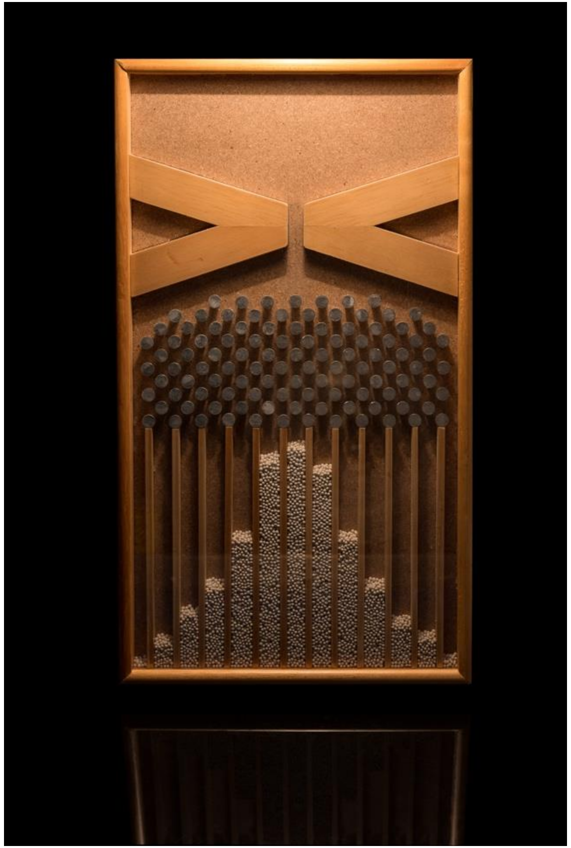
$$X \sim \mathrm{Bin}(n,p)$$

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

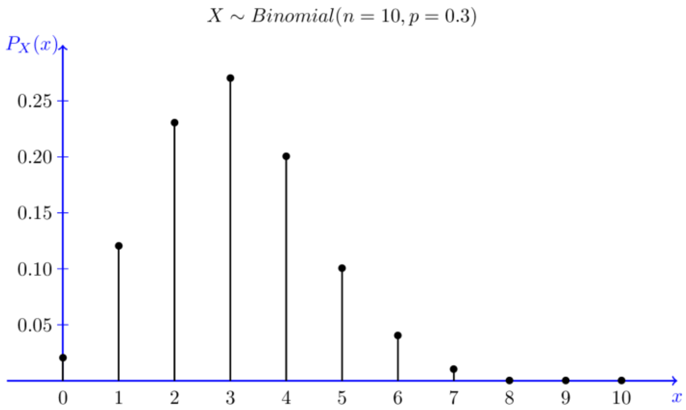
- Number of times it bounces to the left:

$$X \sim \mathrm{Bin}(10, 0.5)$$



Binomial distribution

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



Facts:

- $\text{E}[X] = E[n \cdot X_i] = n \cdot E[X_i] = np$

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

Bernoulli distribution

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

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

- This is called the Bernoulli distribution with parameter p, abbreviated as $\mathrm{Bernoulli}(p)$

- $\mathrm{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p$

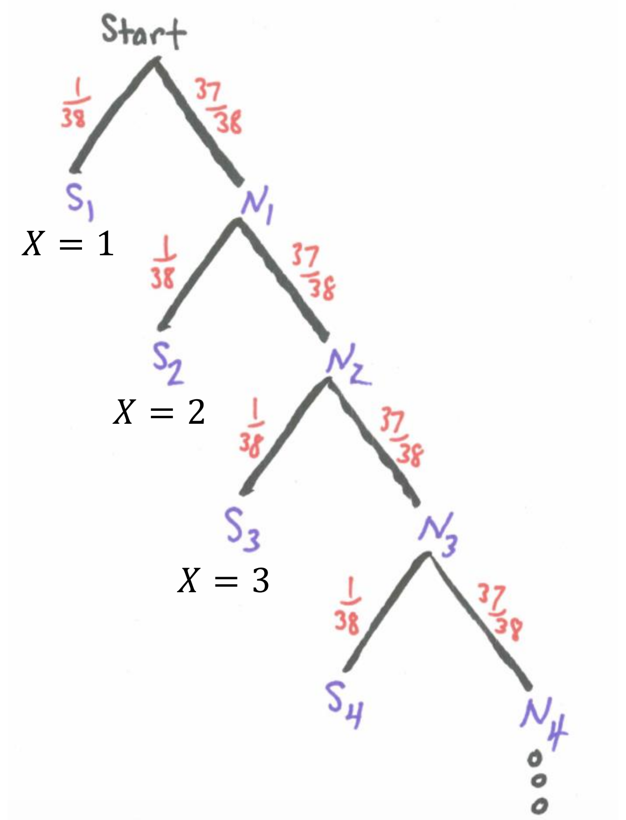
Geometric distribution

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



https://randombooks.org/geometric-distribution.html

Geometric distribution

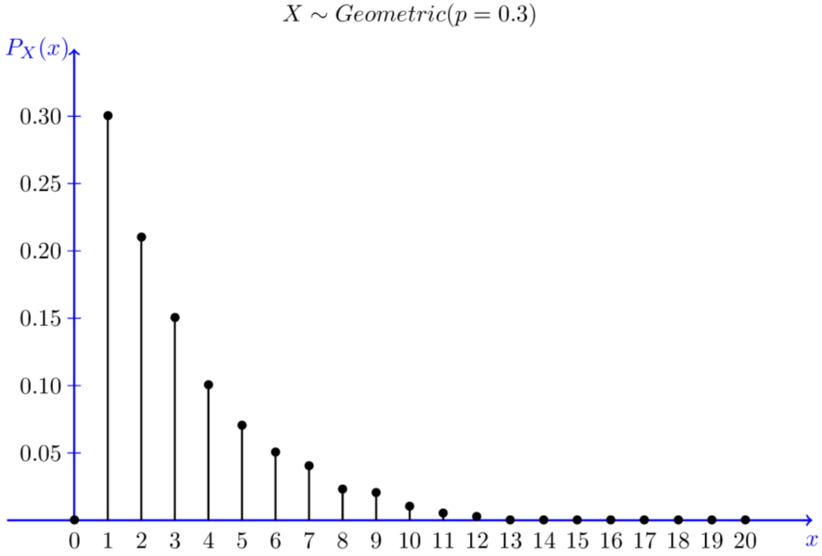
- In conclusion,

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

for $x = 1,2, \ldots$

Fact:

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



Recap

- Mean:
  - $\mu = \mathrm{E}[X] = \sum_x x \cdot P(X = x)$
  - $\mathrm{E}[f(X)] = \sum_x f(x) \cdot P(X = x)$
  - $\mathrm{E[a \cdot X] = a \cdot E[X]}$
- Variance:
  - $\mathrm{Var}(X) = \sigma^2 = \mathrm{E}[(X - \mu)^2] = \sum_x (x - \mu)^2 \cdot P(X = x)$
  - $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$
  - $\mathrm{Var}(a \cdot X) = a^2 \cdot \mathrm{Var}(X)$
- Example discrete RVs and their summary statistics (i.e., mean, variance)

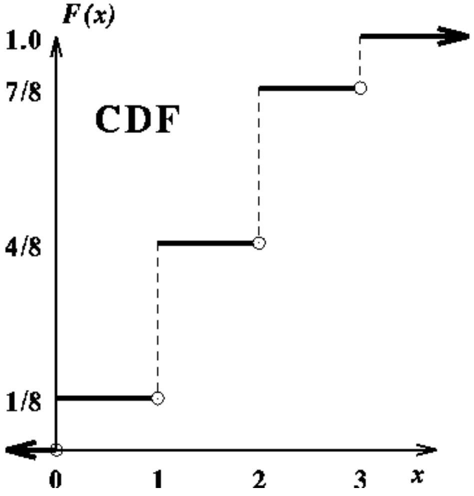
Continuous Random Variables

Plan

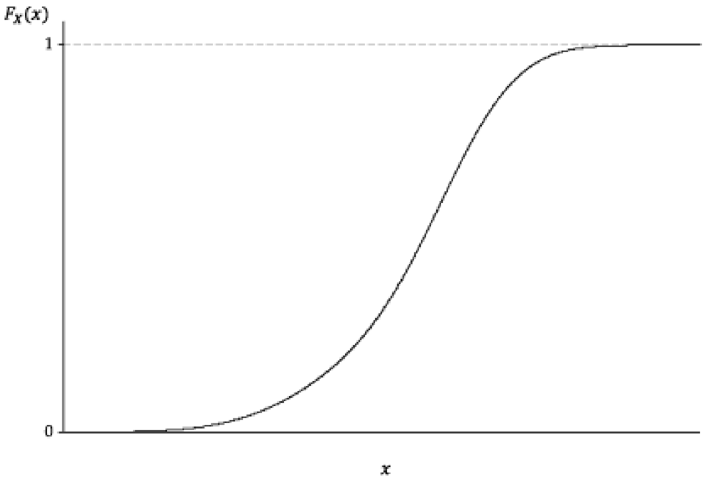
- Properties of CDF
- For continuous RV X, what is $\mathrm{P}(X = x)$?
- PDF and its properties
- Relation of CDF and PDF

Continuous random variables

- Discrete random variables (RVs) take values in a discrete set
- Their CDFs are discontinuous



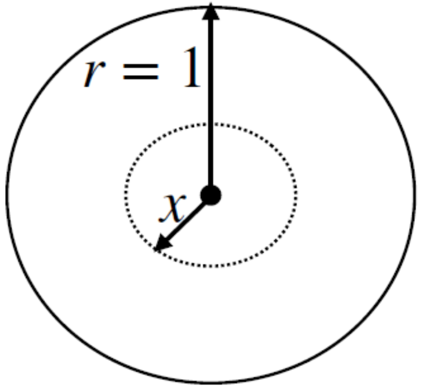
- Continuous RVs take values in a continuous set
- Their CDFs ($\mathrm{P}(X \leq x)$) are continuous



Example: throwing dart

Dartboard with radius 1; dart lands uniformly at random on the board. X is the distance to the center.

What is the CDF of X (the probability that the dart lands at a distance less than or equal to x from the center)?



- $P(X \leq x) = \frac{\pi\, x^2}{\pi\, 1^2} = x^2$ for $x \in [0,1]$

Thus,

- $F(x) = \begin{cases} 0, & x < 0 \\ x^2, & x \in [0,1] \\ 1, & x > 1 \end{cases}$

Example: throwing dart

Dartboard with radius 1; dart lands uniformly at random on the board. X is the distance to the center.

What is the CDF of X (the probability that the dart lands at a distance less than or equal to x from the center)?

- E.g. $P(X \leq 0.3) = 0.3^2 = 0.09$
- Can you find $P(X = 0.3)$?
  - $P(X = 0.3) = 0!$
  - The probability that lands at exactly a distance of 0.3 from the center is 0

Maybe it is not that weird..

Fact for a continuous random variable X, the probability that it takes a specific value x is 0.

Q1: Probability that your house water usage tomorrow is 20.58 gallon?

- 0

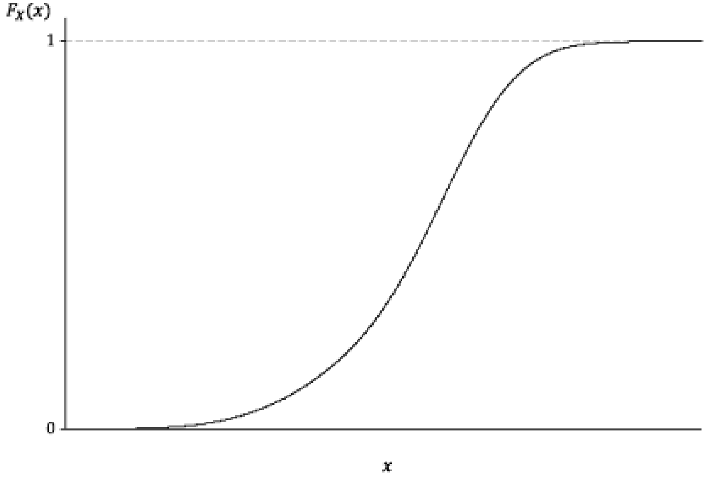
Q2: Probability that your house water usage tomorrow is between 20 and 25 gallon?

- A more useful question

CDF for continuous RVs

- Suppose F is the CDF of continuous random variable X

- What is $P(a < X \leq b)$?
  - $F(b) - F(a)$



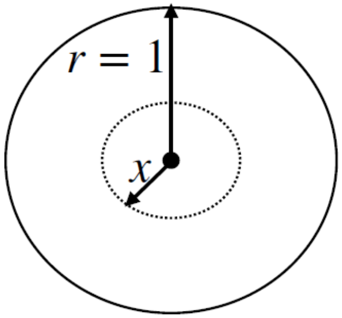
- What is $P(a \leq X \leq b)$?
  - Same!
  - $P(a < X < b)$, $P(a \leq X < b)$ also have the same value
  - Why? $P(X = a) = P(X = b) = 0$

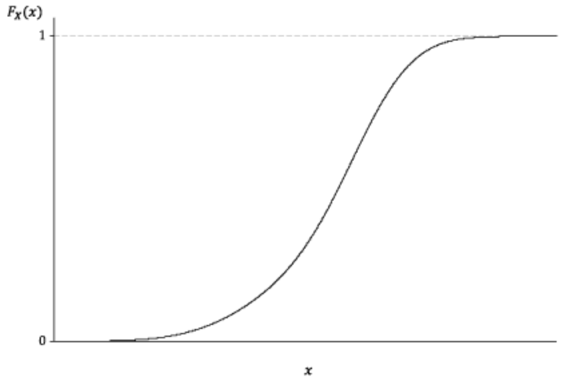
CDF for continuous RVs

- Continuous RVs are those whose CDFs are continuous (no jumps)
- For example, X is the distance to the center

F generally satisfies properties:

- F is continuous (no jumps)
- F is monotonically increasing
- F goes to 0 as $x \to -\infty$
  - Abbrev. $F(-\infty) = 0$
- F goes to 1 as $x \to +\infty$





Continuous random variables

- For discrete RVs, we have PMF and CDF.

- For continuous RVs, what is the analogue of PMF?

- Can we use $P(X = x)$ and sum over all x?
    - No, $P(X = x)$ is always 0

- Maybe we can define function f such that

$$P(a \leq X \leq b) = \text{“sum over } f(x),\ x \in [a, b]\text{”}$$

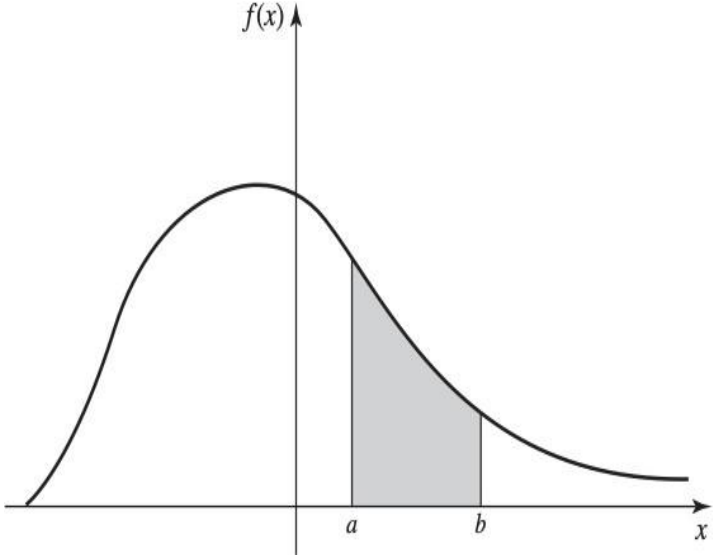
Math interlude: integration

- Summing over $f(x)$, $x \in [a, b]$ is the same as calculating the area under the curve of $f(x)$, for $x \in [a, b]$

- This problem is called integration,

and the area of interest is denoted by:

$$\int_a^b f(x)\, dx$$



Reads "the integral of f from a to b"

Math interlude: integration

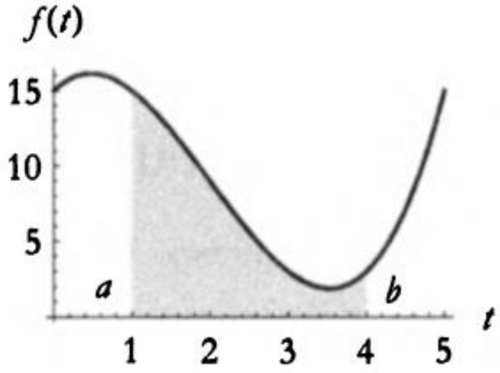
Why the weird $\int$ symbol?

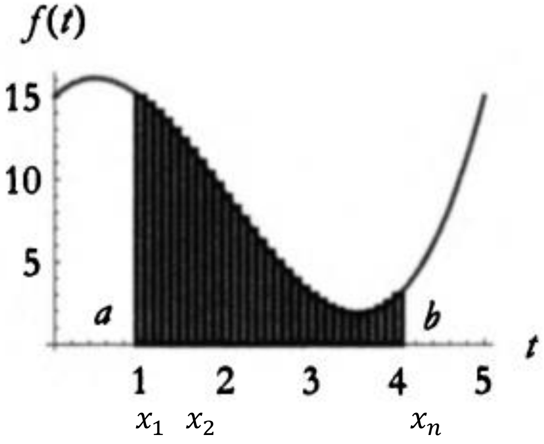
'$\int$ ' is a stylized version of 'S', representing sum

This comes from approximating the area using a series of small rectangles

$$\sum_{i=1}^{n} f(x_i)\,(x_{i+1} - x_i) \coloneqq \sum f(x)\,\Delta x$$

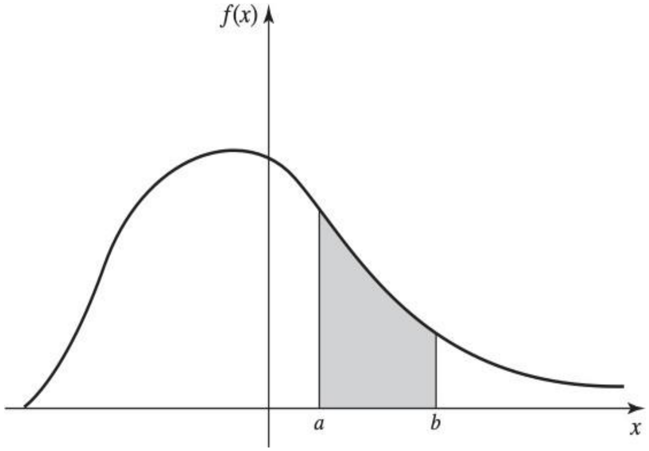
With the partition being finer, this tends to $\int_a^b f(x)\,dx$





Math interlude: integration

Applications of integration



- x: time, $f(x)$: speed
- $\int_a^b f(x)\,dx$: total distance traveled within time $[a, b]$, or displacement at time b (relative to time a)

- x: time (hour), $f(x)$: power consumption (in Watts)
- $\int_a^b f(x)\,dx$: total energy used (in Watt-hours)

Math interlude: integration

- How to calculate $F_a(b)$, in other words, $\int_a^b f(x)\,dx$?

- **Fact (Fundamental Theorem of Calculus, Newton-Leibniz)**

$\int_a^b f(x)\,dx$ can be calculated by:

  - Finding F, the antiderivative of f
  - Evaluate $F(b) - F(a)$ (abbrev. $F(x)|_a^b$)

- What is antiderivative?

Math interlude: integration

$$\begin{tikzcd}[column sep=huge]
f \arrow[r, bend right, "\text{Antiderivative}"'] & F \arrow[l, bend right, "\text{Derivative}"']
\end{tikzcd}$$

- Examples
  - $f(x) = x$ $\qquad F(x) = \frac{1}{2}x^2$
  - $f(x) = x^m$ $\qquad F(x) = \frac{x^{m+1}}{m+1} \quad (m \neq -1)$
  - $f(x) = \frac{1}{x}$ $\qquad F(x) = \ln x$

Math interlude: integration

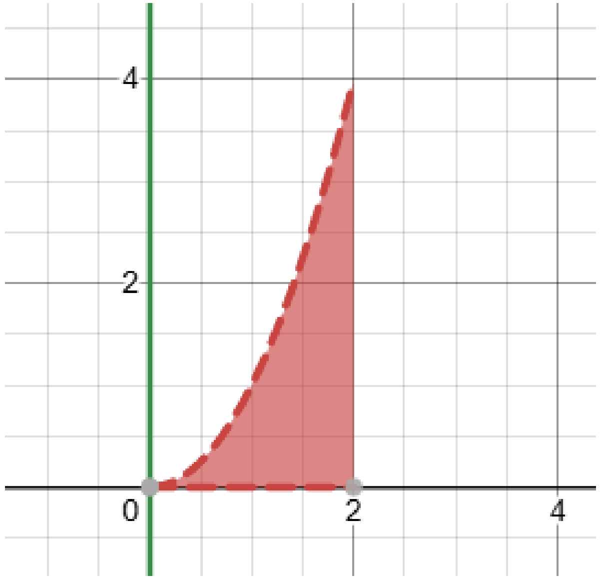
$$\begin{tikzcd} f \arrow[rr, bend right, "\text{Antiderivative}"'] & & F \arrow[ll, bend right, "\text{Derivative}"'] \end{tikzcd}$$

- f can have many antiderivatives
- Useful example
  - f: speed(time); F: distance(time)
- E.g. $f(x) = 1$
  - $F(x) = x, \;\; F(x) = x + 2$ are all valid antiderivatives
  - All antiderivatives of f are equal up to a constant
  - We use the shorthand $F(x) = x + C$ to emphasize this

Math interlude: integration

Example find $\int_0^2 x^2\, dx$

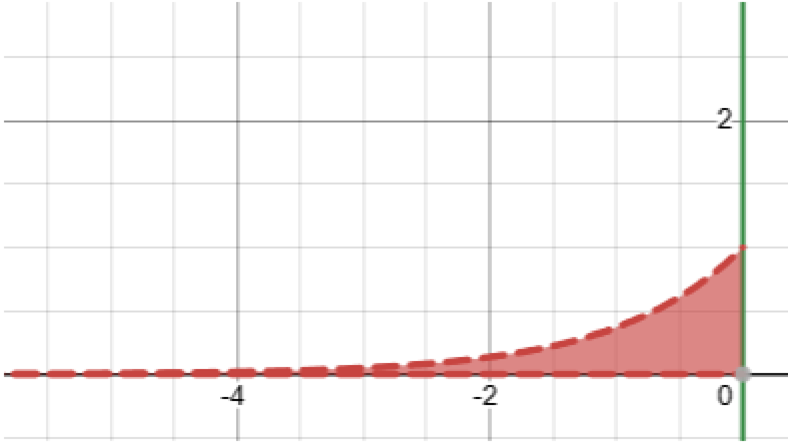
- Step 1: find F, antiderivative of x^2
  - $F(x) = \frac{x^3}{3}$
- Step 2: evaluate F at both end points
  - $F(2) = \frac{8}{3}$, $F(0) = 0$
  - Ans = $F(2) - F(0) = \frac{8}{3}$



Math interlude: integration

Example find $\int_{-\infty}^{0} e^x \, dx$

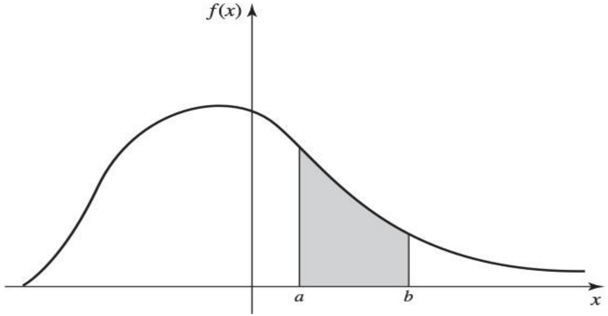
- Step 1: find F, antiderivative of e^x
  - $F(x) = e^x$
- Step 2: evaluate F at both end points
  - $F(0) = 1, F(-\infty) = 0$
  - Ans = $F(0) - F(-\infty) = 1$



Probability density function (PDF)

Fact For continuous random variable X, there is a function f_X (abbrev. f) such that for any a, b,

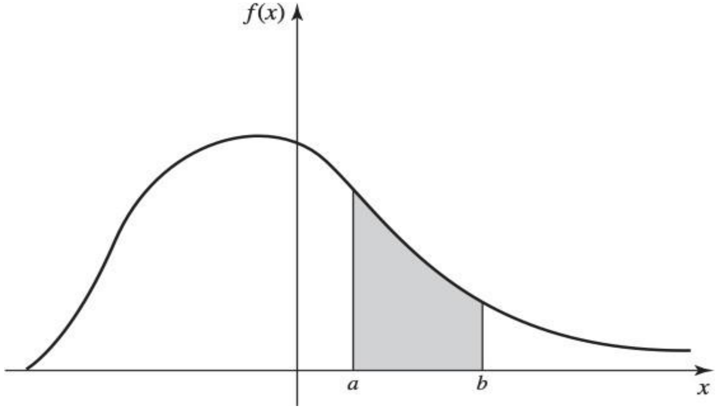
$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$



Function f is called the *probability density function (PDF)* of X.

$f(x)$ measures how likely X takes value in the *neighborhood* of x.

Properties of PDF



- Nonnegativity: $f(x) \geq 0$ for all x
  - But $\mathrm{P}(X = x) = 0$!
- Normalized:

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1$$

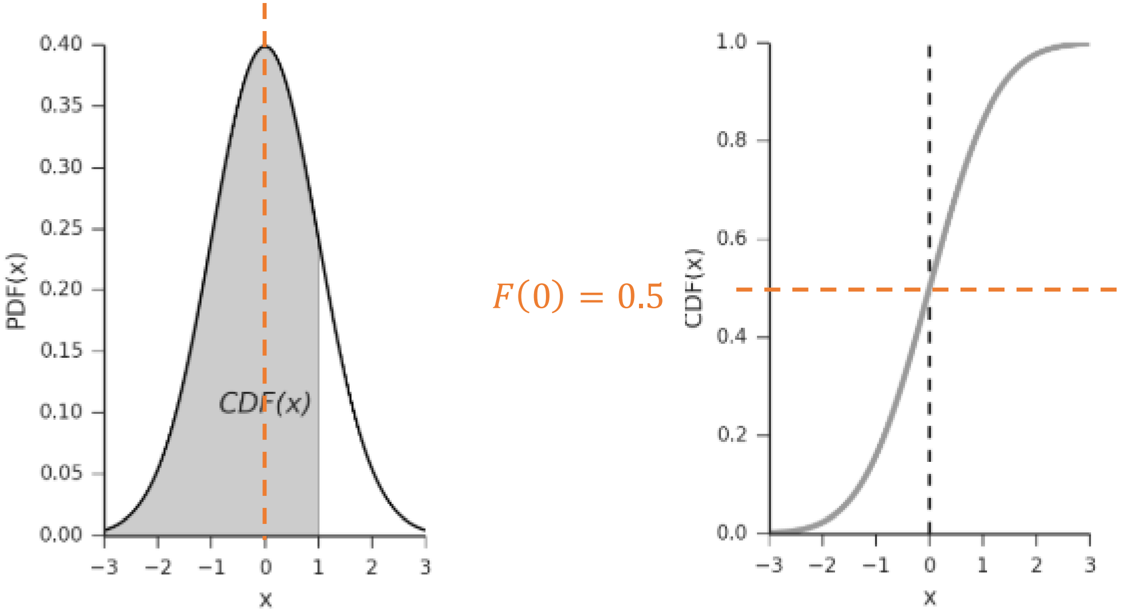
  - Reason: the integral represents $P(-\infty \leq X \leq +\infty)$

Relationship between PDF and CDF

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

- How to find CDF F based on PDF f?

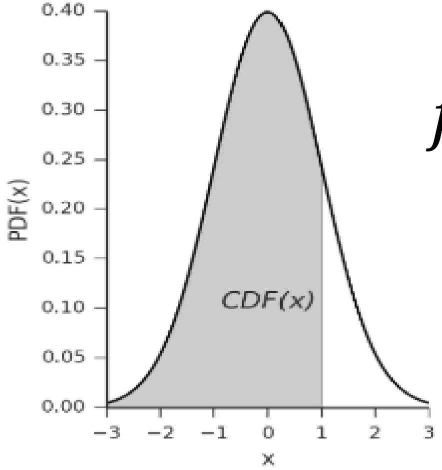
$$F(b) = P(X \le b) = \int_{-\infty}^{b} f(x)\,dx$$



area under the curve when $X \le 0$ *is* 0.5

Relationship between PDF and CDF

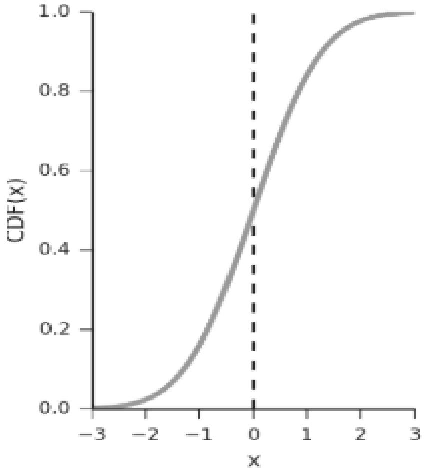
- F is an indefinite integral of f: $f(x) = F'(x)$
- F has large slope at x: $f(x)$ is large



Derivative

f F

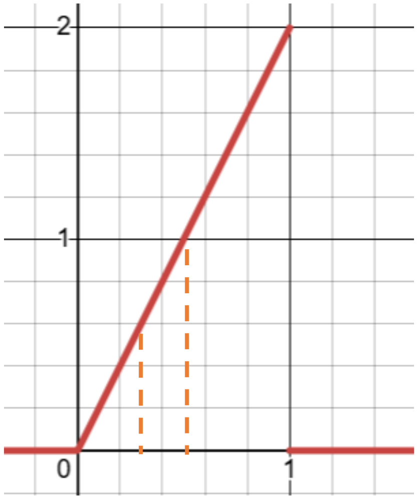
Antiderivative



Probability density function (PDF)

Example X: lifetime of a lightbulb, has PDF

$$f(x) = 2x,\ 0 < x < 1$$



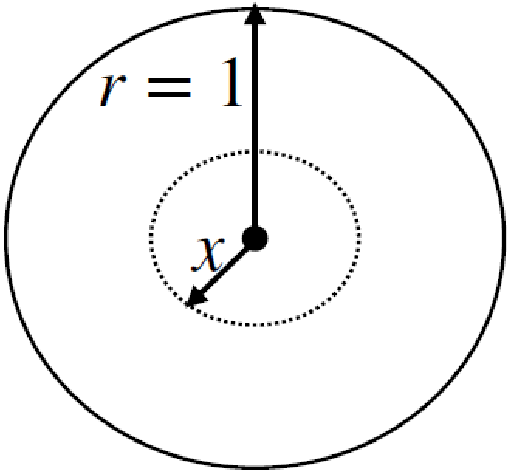
Find $\mathrm{P}(0.3 < X < 0.5)$

Soln This is equal to

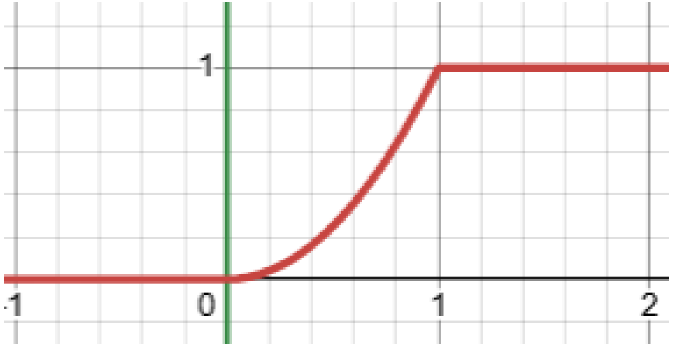
$$\int_{0.3}^{0.5} 2x\,dx = x^2\,\Big|_{0.3}^{0.5} = 0.5^2 - 0.3^2 = 0.16$$

Example: dart

- X: distance to the center, given CDF:



$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

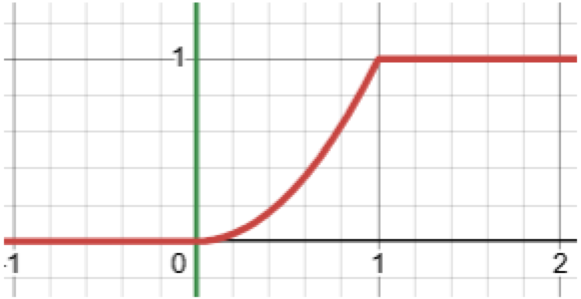


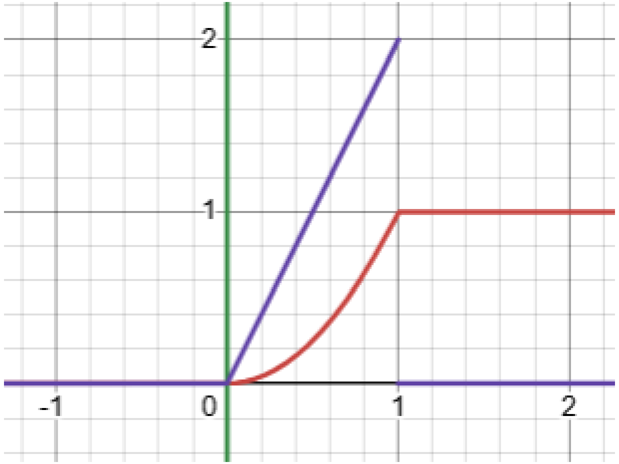
- What is the PDF of X?

Example: dart

What is the PDF of X?

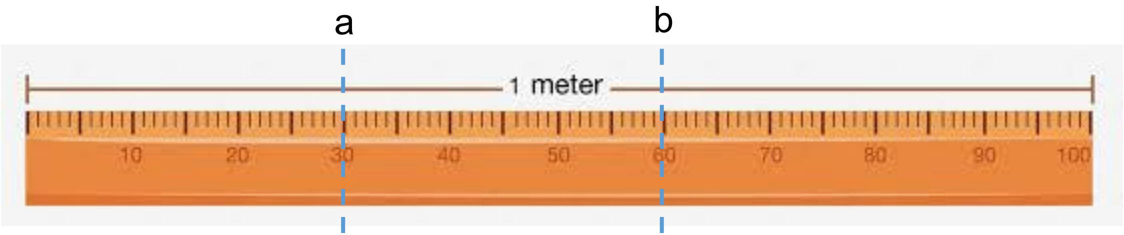
- $f(x)$ is the derivative of F
  - $f(x) = 0, \quad x < 0$
  - $f(x) = 2x, \; x \in [0,1]$
  - $f(x) = 0, x > 0$





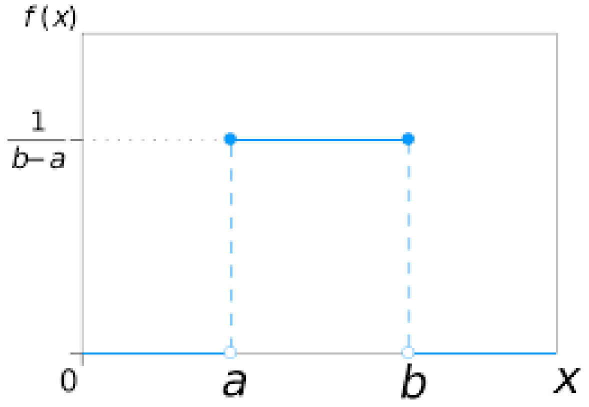
In-class activity: ruler

- We choose X uniformly at random from $[a, b]$, two points in a ruler. In other words, X can land anywhere between $[a, b]$ with equal likelihood.
- Find the PDF and CDF of X



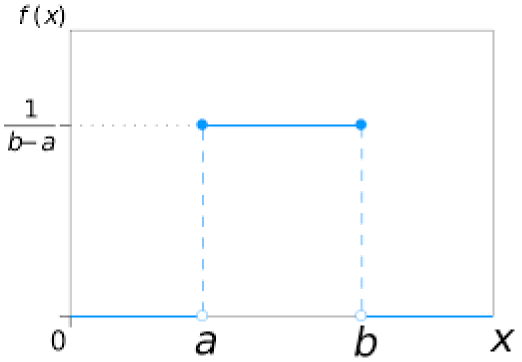
In-class activity: ruler

- We know $P(a \le X \le b) = 1$, and $f(x)$ is constant on [a, b]
- $f(x)$ is constant: say c
- $\int_a^b f(x)\,dx = \int_a^b c\,dx = cx \,|_a^b = cb - ca = 1$
- $c = \frac{1}{b-a}$
- So $f(x) = \frac{1}{b-a},\ x \in [a, b]$



In-class activity: ruler

- What is the PDF $f(x)$?
  - $f(x) = 0, \ x < a$
  - $f(x) = \frac{1}{b-a}, \ x \in [a, b]$
  - $f(x) = 0, \ x > b$



- This is also known as the *uniform distribution* over $[a, b]$, abbrev.

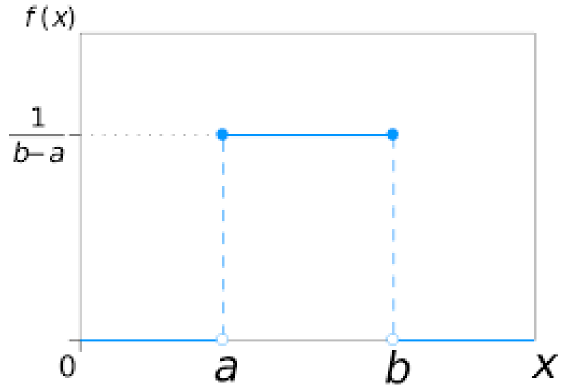
$$\text{Uniform}([a, b])$$

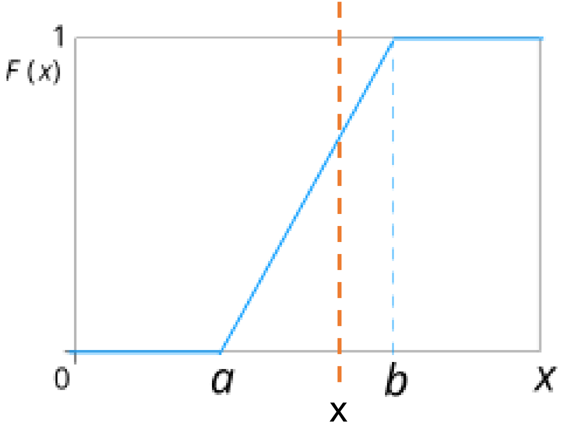
- What is the CDF $F(x) = P(X \leq x)$?

In-class activity: ruler

- What is the CDF $F(x) = P(X \leq x)$?
  - $F(x) = 0, \quad x < a$
  - $F(x) = \frac{x-a}{b-a}, \quad x \in [a, b]$
  - $F(x) = 1, \quad x > b$

$$F(x) = \int_a^x f(x)\,dt = \int_a^x \frac{1}{b-a}\,dt = \frac{1}{b-a}(x-a)$$



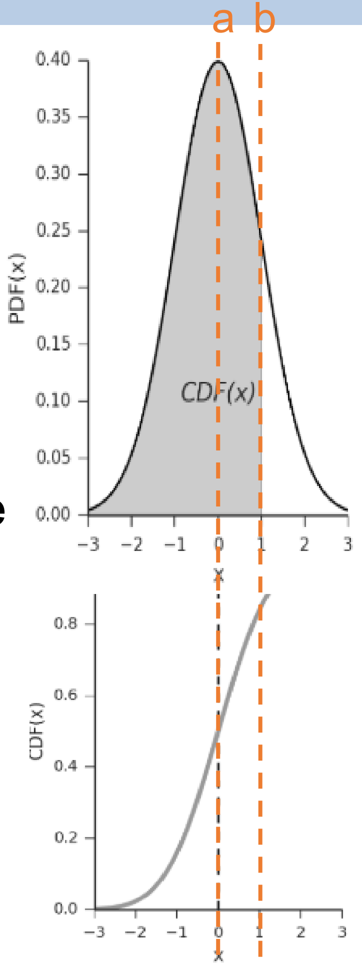


Recap

- Is $f(x)$ equal to $P(X = x)$?
  - No -- $P(X = x) = 0$ always
  - Correct interpretation: probability *density* (not probability)

- $P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a)$
  - the probability that a RV lies between a and b is given by the area under the PDF from a to b = the difference between the CDF values at b and a.
  - F is the antiderivative of f

- Are there real-world RVs that are neither discrete nor continuous?



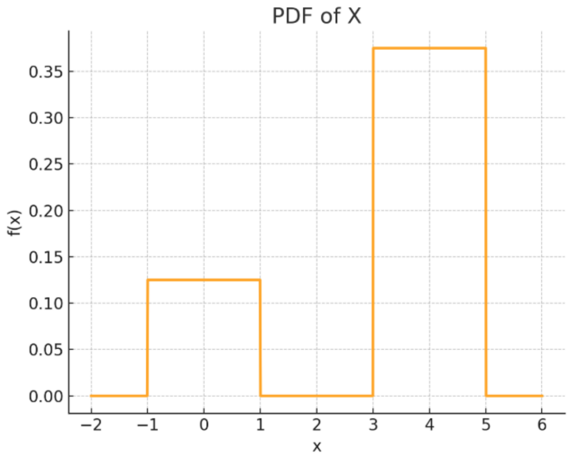
Plans

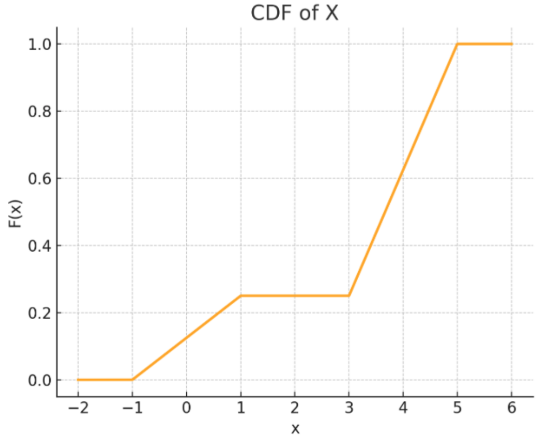
- Transformation of a continuous RV, its CDF and PDF
- Expectation and variance of continuous RVs
- Useful continuous probability distributions

In-class activity

- Given by the PDF of X, find its CDF.
- $$f(x) = \begin{cases} \frac{1}{8}, & x \in [-1, 1] \\ \frac{3}{8}, & x \in [3,5] \\ 0, & otherwise \end{cases}$$

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{x+1}{8}, & x \in [-1,1) \\ \frac{1}{4}, & x \in [1, 3) \\ \frac{3x-7}{8}, & x \in [3, 5) \\ 1, & x \geq 5 \end{cases}$$





In-class activity

$x \in [-1, 1]$:

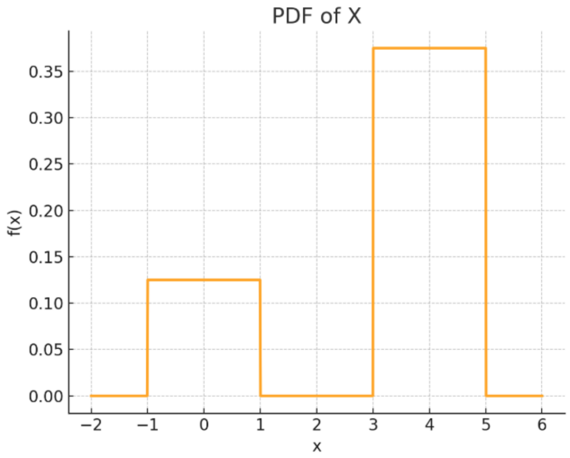
- $F(x) = \int_{-1}^{x} f(x)\,dx = \int_{-1}^{x} \frac{1}{8}\,dx = \frac{1}{8}\big(x - (-1)\big)$

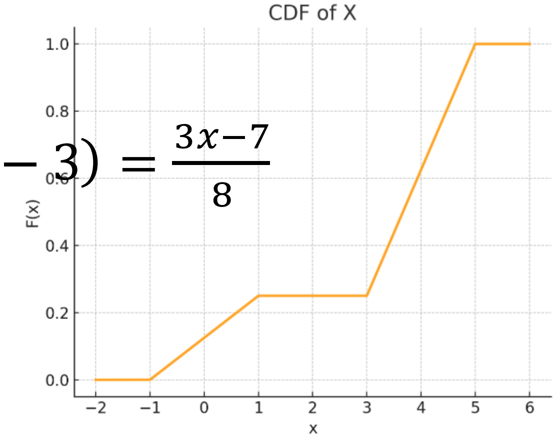
$x \in [1, 3)$:

- $F(x) = F(1) = \frac{2}{8}$

$x \in [3, 5)$:

- $F(x) = F(3) + \int_{3}^{x} f(x)\,dx = \frac{1}{4} + \int_{3}^{x} \frac{3}{8}\,dx = \frac{1}{4} + \frac{3}{8}(x - 3) = \frac{3x - 7}{8}$





Transformations of a continuous RV

- Given a continuous RV X and any transformation f, $f(X)$ is a random variable (e.g. $X+5, 3X, X^2$)
- Applications:
  - X: temperature tomorrow in Celsius, $1.8X+32$: temp in Fahrenheit

- How to find the distribution of $Y=f(X)$ based on that of X?
  - First, find Y's CDF
  - Take derivative to find Y's PDF

Transformations of a continuous RV

Example Suppose $X \sim \text{Uniform}([0,1])$. Find the distribution of $Y = X + b$.

Step 1: write down the CDF of X

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \in [0,1] \\ 0, & x > 1 \end{cases}$$

$$F(x) = P(X \le x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

Uniform([a, b]): $f(x) = \frac{1}{b-a},\ x \in [a, b]$

Step 2: write down the CDF of Y

$$\mathrm{P}(Y \le y) = P(X + b \le y) = P(X \le y - b) = F(y - b)$$

$$\mathrm{F}(\mathrm{y}) = \mathrm{P}(Y \le y) = \begin{cases} 0, & y < b \\ y - b, & y \in [b, b+1] \\ 1, & y > b + 1 \end{cases}$$

Transformations of a continuous RV

Step 2: write down the CDF of Y

$$\mathrm{P}(Y \leq y) = \begin{cases} 0, & y < b \\ y - b, & y \in [b, b+1] \\ 1, & y > b+1 \end{cases}$$

Step 3: Take derivative to get the PDF of Y

$$f(y) = \begin{cases} 0, & y < b \\ 1, & y \in [b, b+1] \\ 0, & y > b+1 \end{cases}$$

In summary, $Y \sim \mathrm{Uniform}([b, b+1])$

Transformations of a continuous RV: $X + b$

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \in [0,1] \\ 0, & x > 1 \end{cases}$$

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

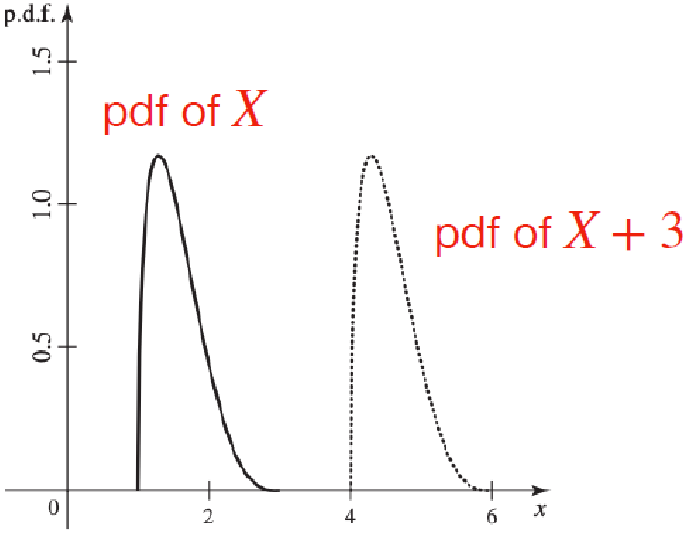
$$f(y) = \begin{cases} 0, & y < b \\ 1, & y \in [b, b+1] \\ 0, & y > b+1 \end{cases}$$

$$F(y) = \mathrm{P}(Y \leq y) = \begin{cases} 0, & y < b \\ y - b, & y \in [b, b+1] \\ 1, & y > b+1 \end{cases}$$

$X + b$ has a PDF that is a translation of X's PDF (by b units)

Shifting a continuous RV

- In general:
- $Y = X + b$ has a PDF that is a translation of X's PDF (by b units)



- $f_{X+b}(x) = f_X(x - b)$: $f_Y(4) = f_X(1)$:

In-class activity: scaling an RV

- **Example** Suppose $X \sim \mathrm{Uniform}([0,1])$. Find the distribution of $Z = aX$.

- Step 1: write down the CDF of X

- Step 2: write down the CDF of Z

- Step 3: Take derivative to get the PDF of Z

In-class activity: scaling an RV

- **Example** Suppose $X \sim \mathrm{Uniform}([0,1])$. Find the distribution of $Z = aX$.

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ x, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

Write down the CDF of Z

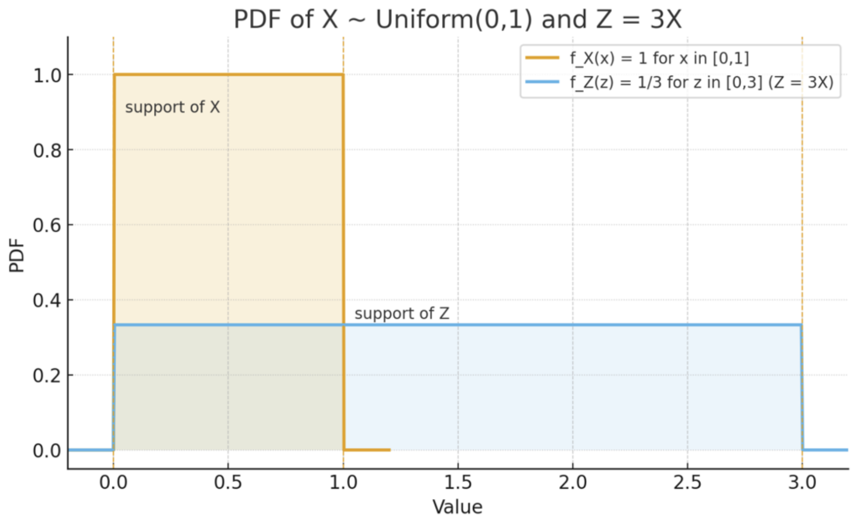
$$\mathrm{P}(Z \leq z) = P(aX \leq z) = P\left(X \leq \frac{z}{a}\right) = F(\frac{z}{a})$$

$$F(z) = P(Z \leq z) = \begin{cases} 0, & Z < 0 \\ \frac{z}{a}, & Z \in [0, a] \\ 1, & Z > a \end{cases}$$

In-class activity: scaling an RV

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \in [0,1] \\ 0, & x > 1 \end{cases}$$

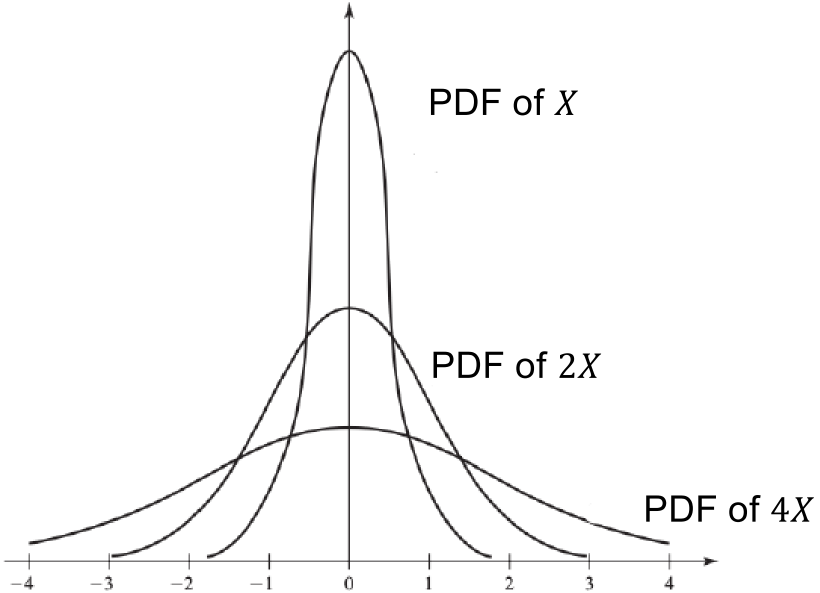
$$f(z) = \begin{cases} 0, & z < 0 \\ \frac{1}{a}, & z \in [0,a] \\ 0, & z > a \end{cases}$$



Conclusion: $Z \sim \text{Uniform}([0,a])$, aX's PDF is X's PDF stretched by a factor of a horizontally

Scaling a continuous RV

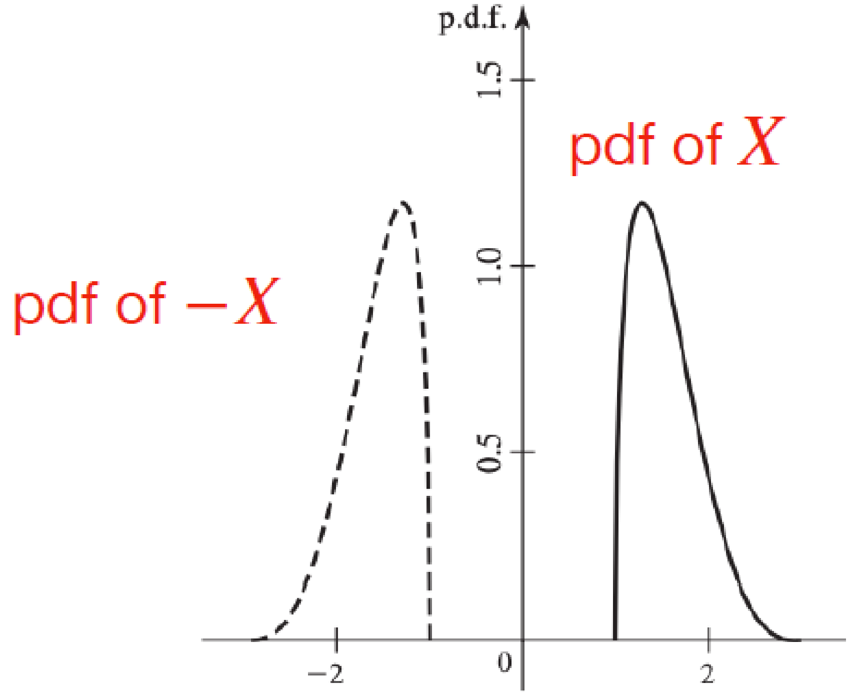
- $Z = aX$'s PDF is X's PDF stretched by a factor of a horizontally



- $f_{aX}(x) = \frac{1}{|a|} f_X\left(\frac{x}{a}\right) : f_{2X}(0) = \frac{1}{2} f_X(0)$

Scaling a continuous RV

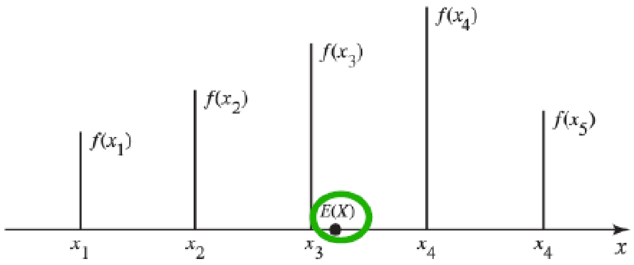
- Given X's PDF; what does $-X$'s PDF look like?

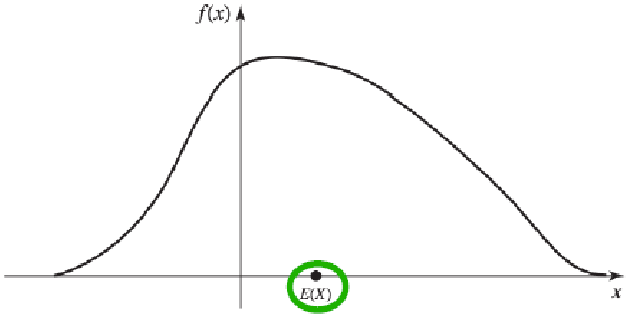


Summarizing Continuous Random Variables

Mean (aka Expected Value, Expectation)

- Weighted average of values of a random variable where weights are probabilities, denoted as μ, or $\mathrm{E}[X]$
- Expectation as center of gravity





Discrete

Continuous

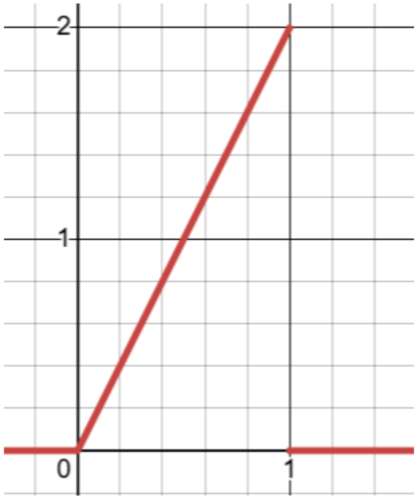
$$\mathrm{E}[X] = \sum_{x} x \cdot P(X = x)$$

$$\mathrm{E}[X] = \int x\, f(x)\, dx$$

Mean

Example X: Time until a lightbulb fails. Its pdf:

$$f(x) = 2x,\ 0 < x < 1$$



What is $\mathrm{E[X]}$?

$$\mathrm{E[X]} = \int_{\mathrm{R}} x\, f(x)\, dx$$

$$= \int_0^1 x(2x)\, dx = \int_0^1 2x^2\, dx = \frac{2}{3} x^3 \,\Big|_0^1 = \frac{2}{3}$$

Expectation formula

- How to find $\mathrm{E}[r(X)]$ given the probability distribution of X?
- For discrete RVs we saw:

$$\mathrm{E}[r(X)] = \sum_x r(x) \cdot P(X = x)$$

- For continuous RVs,

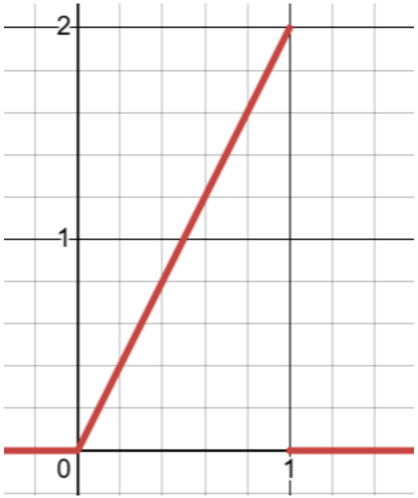
$$\mathrm{E}[r(X)] = \int r(x) f(x)\, dx$$

Rule of the lazy statistician: could also find it by first finding pdf of $r(X)$ which would require many further calculations. Lazy prefers easy.

Expectation formula

Example Assume the pdf of the previous example,

$$f(x) = 2x,\ 0 < x < 1$$



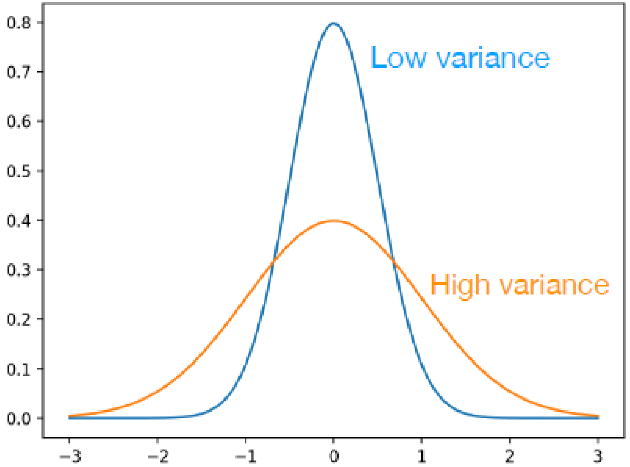
Find $\mathrm{E}[\sqrt{X}]$

$$\mathrm{E}[\sqrt{X}] = \int_{\mathrm{R}} \sqrt{x}\, f(x)\, dx$$

$$= \int_0^1 \sqrt{x}(2x)\, dx = \int_0^1 2x^{\frac{3}{2}}\, dx = \frac{4}{5} x^{\frac{5}{2}} \Big|_0^1 = \frac{4}{5}$$

Variance

- Variance of X measures how spread out the distribution of X is

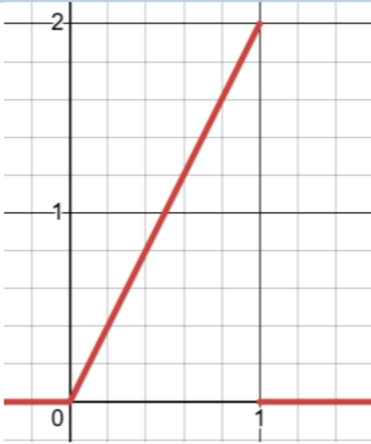


- Defn: $\mathrm{Var}(X) = \sigma^2 = \mathrm{E}[(X - \mu)^2]$

Mean of X

- Fact: $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$ continues to hold

Variance



Example Assume the pdf of the previous example,

$$f(x) = 2x,\ 0 < x < 1$$

Find $\mathrm{Var}(X)$.

Soln We saw before that $\mathrm{E}[X] = \frac{2}{3}$. Let's try to find $\mathrm{E}[X^2]$

$$\mathrm{E}[X^2] = \int_0^1 x^2(2x)\,dx = \frac{2}{4} = \frac{1}{2}$$

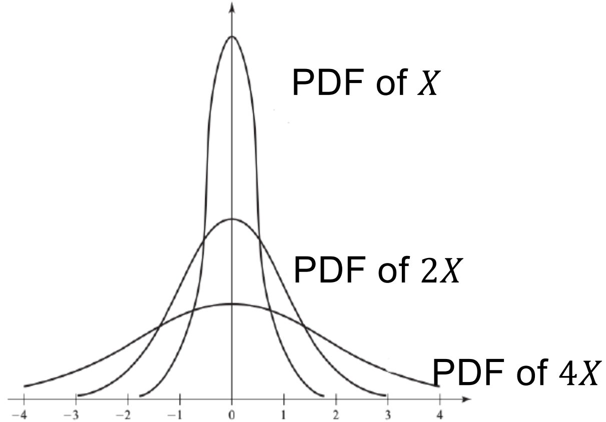
$$\mathrm{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 \approx 0.055$$

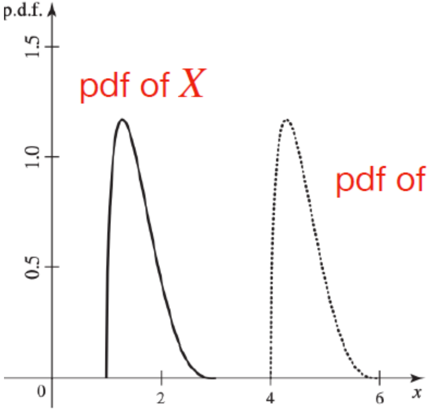
Properties of Mean & variance

How does aX's mean & variance relate to those of X?

Fact same as discrete RVs, for continuous RVs, it continues to hold :

- $\mathrm{E}[aX] = a\,\mathrm{E}[X]$
- $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
- $\mathrm{E}[X+b] = \mathrm{E}[X] + b$
- $\mathrm{Var}(X+b) = \mathrm{Var}(X)$





Properties of Mean & variance

- How about $\mathrm{E}[aX + b]$ and $\mathrm{Var}[aX + b]$?
- E.g. Celsius to Fahrenheit, $a = 1.8$, $b = 32$

- We can now combine the previous results to get:

- $\mathrm{E}[aX + b] = \mathrm{E}[aX] + b = a\mathrm{E}[X] + b$

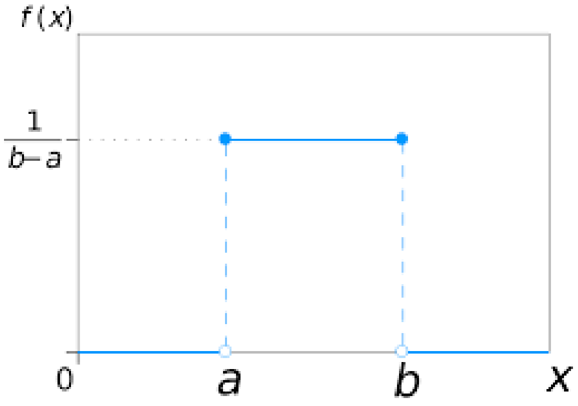
- $\mathrm{Var}[aX + b] = \mathrm{Var}[aX] = a^2 \cdot \mathrm{Var}[X]$

Useful Continuous Probability Distributions

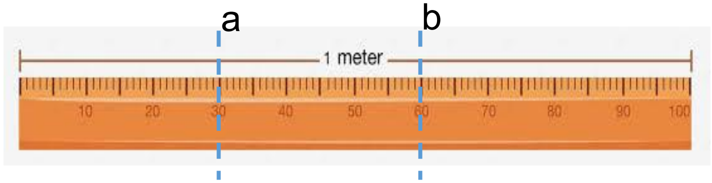
Uniform Distribution

- $X \sim \text{Uniform}([a, b])$

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & x \in [a, b] \\ 0, & x > b \end{cases}$$



- Mean: $\mathrm{E}[X] = \frac{a+b}{2}$
- Variance:
  - $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$
  - $\text{Uniform}([0,1])$ has a variance of $1/12$



Uniform distribution

numpy.random.uniform

numpy.random.uniform(*low=0.0, high=1.0, size=None*)

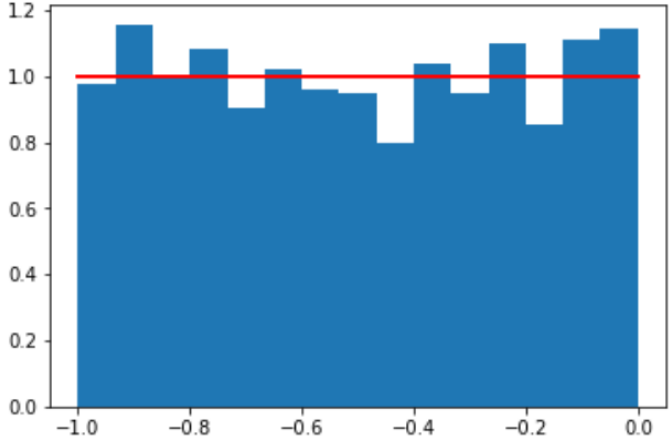
Draw samples from a uniform distribution.

Samples are uniformly distributed over the half-open interval `[low, high)` (includes low, but excludes high). In other words, any value within the given interval is equally likely to be drawn by **uniform**.

Example Draw 1,000 samples from a uniform distribution on [-1,0),

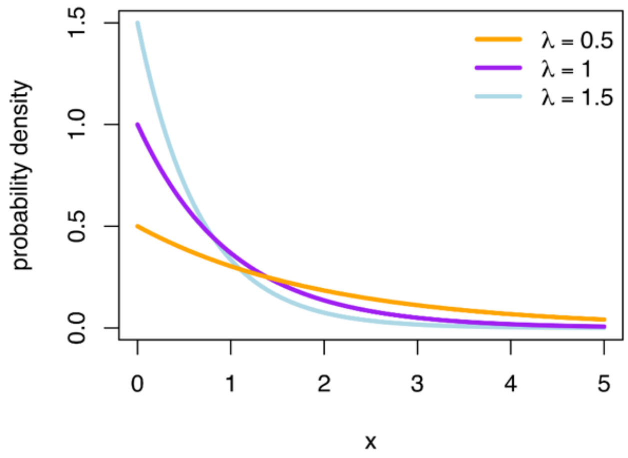
redline: PDF of uniform distr.

```
a = -1
b = 0
N = 1000
X = np.random.uniform(a,b,N)
count, bins, ignored = plt.hist(X, 15, density=True)
plt.plot(bins, np.ones_like(bins), linewidth=2, color='r')
plt.show()
```



Exponential Distribution

- Denoted as $X \sim \text{Exp}(\lambda)$
  - $f(x) = \lambda e^{-\lambda x}$, $x > 0$
  - λ: scale parameter
  - $\text{E}[X] = \frac{1}{\lambda}$
  - $\text{Var}[X] = \left(\frac{1}{\lambda}\right)^2$
  - the continuous analogue of geometric distribution



Examples:

- Time between geyser eruptions
- Lifetime of lightbulbs
- Time of radioactive particle decays

Exponential Distribution

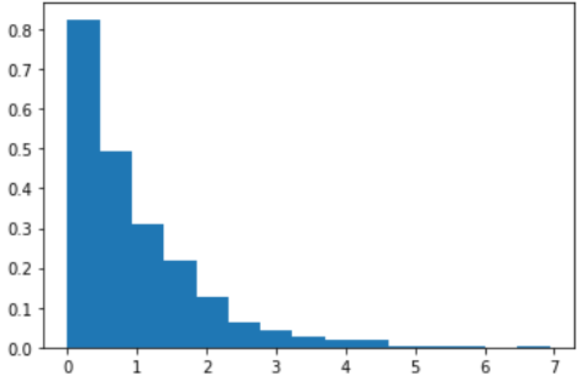
numpy.random.exponential

numpy.random.exponential(*scale=1.0*, *size=None*)

scale = λ

Example Draw 1,000 samples from exponential with $\lambda = 1.0$

```
lam = 1.0
N = 1000
X = np.random.exponential(lam, N)
count, bins, ignored = plt.hist(X, 15, density=True)
plt.show()
```



Gaussian Distribution

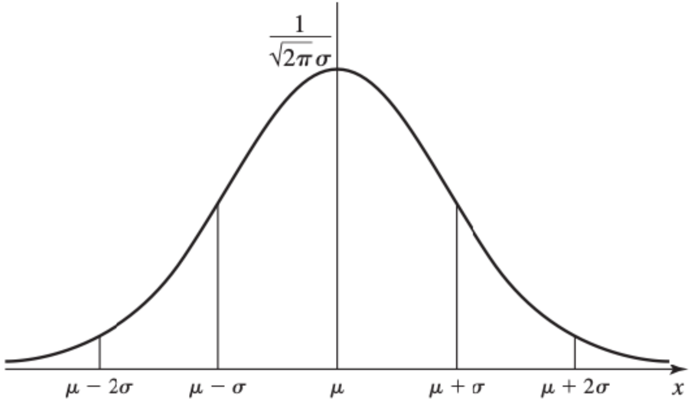
Gaussian (a.k.a. Normal) distribution with location μ and scale σ^2 parameters,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$\mathrm{E}[X] = \mu$
$\mathrm{Var}[X] = \sigma^2$

Abbreviated as $N(\mu, \sigma^2)$

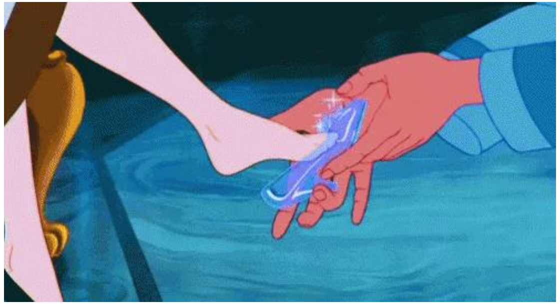
Perhaps *the most important* distribution in prob & stats

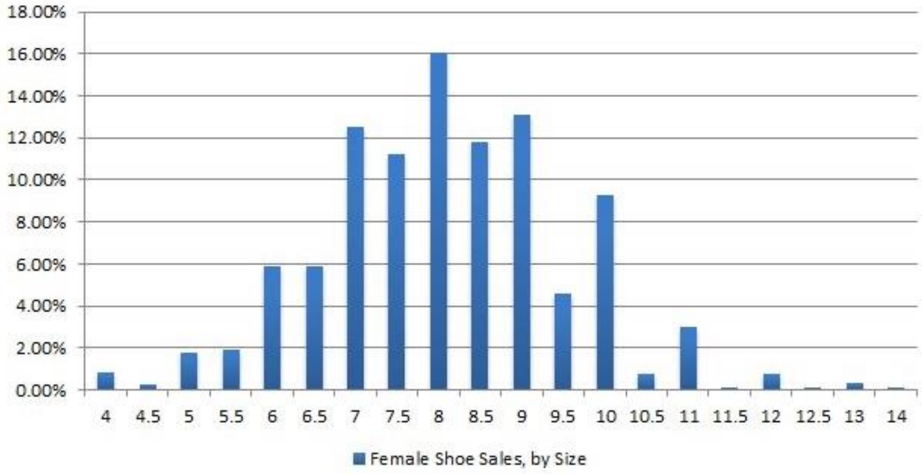


Similar to binomial distribution

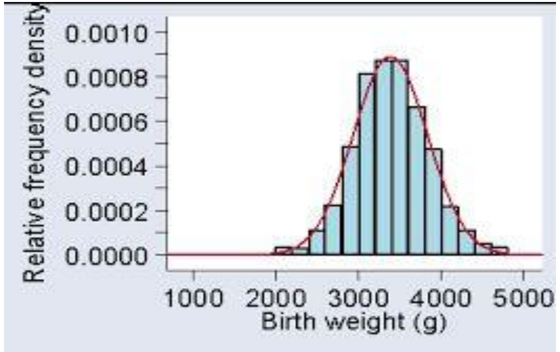
Distributions that follow Gaussian

Shoe size





Birth Weight



They do not actually follow exact Gaussians, but very close

(From https://studiousguy.com/real-life-examples-normal-distribution/)

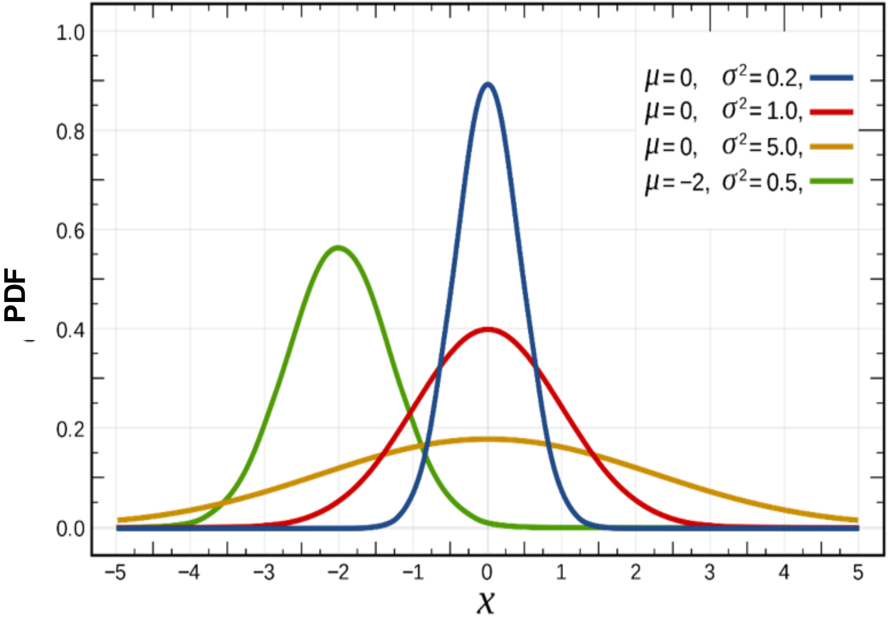
Gaussian Distribution

Observations:

- Larger $\sigma^2 \Rightarrow p(x)$ more "spread out"
- Larger $\mu \Rightarrow p(x)$'s center shifts to the right more

Fact if $X \sim N(\mu, \sigma^2)$

- $\mathrm{E}[X] = \mu$
- $\mathrm{Var}[X] = \sigma^2$



Gaussian Distribution

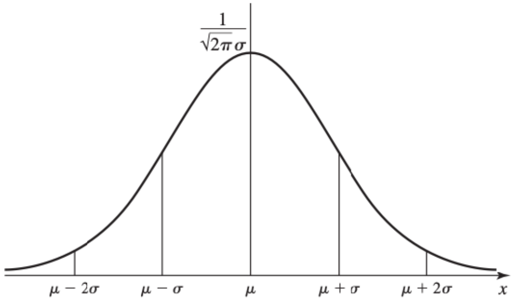
Linear transformations of Gaussian is still Gaussian

Fact if $X \sim N(\mu, \sigma^2)$, then $Y = aX + b$ is still Gaussian.

What are the parameters of Y's Gaussian distribution, $\mathrm{E}[Y]$, $\mathrm{Var}[Y]$?

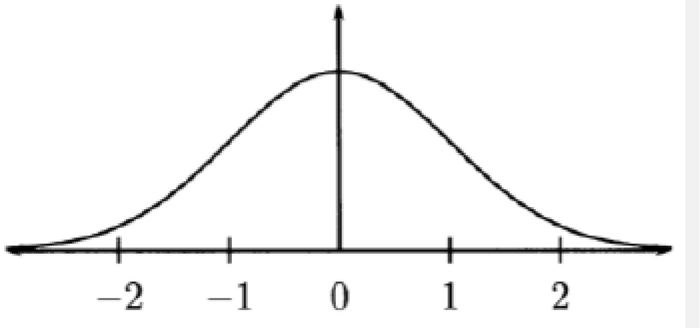
- $\mathrm{E}[Y] = \mathrm{E}[aX + b] = a\mu + b$
- $\mathrm{Var}[Y] = \mathrm{Var}[aX + b] = \mathrm{Var}[aX] = a^2\sigma^2$

- So, $Y \sim N(a\mu + b, a^2\sigma^2)$



The standard Gaussian distribution

- Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$

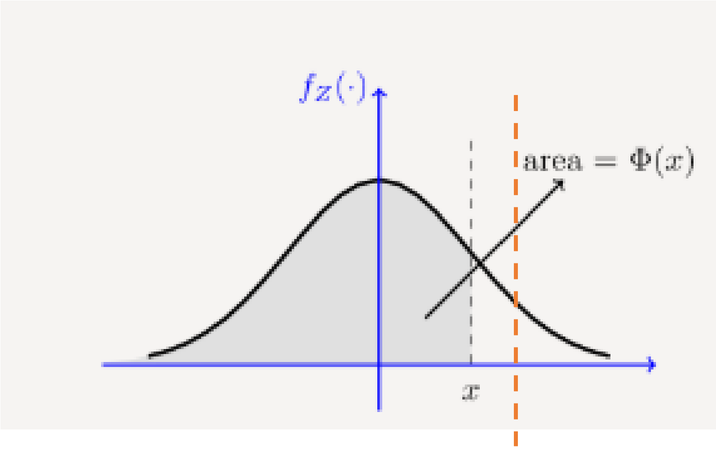


- Denoted by $Z \sim N(0,1)$
- Its PDF denoted by $\phi(z)$, and CDF denoted by $\Phi(z)$

The standard Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \mu = 0, \sigma = 1$$

- PDF: $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$
- CDF: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)\, dz$



$x = 1.5$

- We can find the value of Φ by calling scipy.stats.norm.cdf

```
from scipy.stats import norm
norm.cdf(1.5)
```