Computer Science

CSC380: Principles of Data Science

Probability 3

Chicheng Zhang

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

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 $P(B|A) = P(B)$)

Announcements 2/12

- HW3 will be posted tomorrow (2/13)

  - You must make two submissions:
    - Your complete homework as a SINGLE PDF file by the stated deadline to gradescope, "HW3" entry. **Include your code and output of the code as texts in the PDF.**
    - Your codes to a separate submission to gradescope, "HW3 Code" entry: a single notebook file including codes for questions 1, 2.

- We highly recommend Piazza for class communications
  - This way all class staff can be on the same page in helping you with the problems

Quiz

- A roulette wheel has 38 wedges: one each with the numbers 1 through 36, one labeled 0 and one labeled 00. Each wedge has a pocket that a metal ball can land in. With each spin, the ball lands in one of these pockets.
- Consider spinning the roulette wheel once. Let Z be the event that the resulting number contains a zero. Let E be the event that the number is even (both 0 and 00 are considered even).
  1. Find $P(Z|E)$. Show your work!
  2. Are E and Z independent? How do you know?

Quiz

$$P(Z|E) = \frac{P(E,Z)}{P(E)}$$

$0, 00, 10, 20, 30$

$0, 00$ and 18 even numbers in {1,.., 36}

$$P(E,Z) = 5/38$$

$$P(E) = 20/38$$

$$\Rightarrow \; P(Z|E) = \frac{5}{20} = \frac{1}{4}$$

Quiz

Are *E* and *Z* independent?

Let's check if $P(Z|E) = P(Z)$..

$0, 00, 10, 20, 30$

$P(Z) = 5/38$

this does not equal $P(Z|E) = \frac{1}{4}$, so *E* and *Z* are not independent

Outline

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

- Summarizing distributions: mean and variance
- Example discrete random variables

- Continuous random variables
  - Probability density functions (PDF)
  - Examples

Random Variables

Random variables (RVs, r.v.'s)

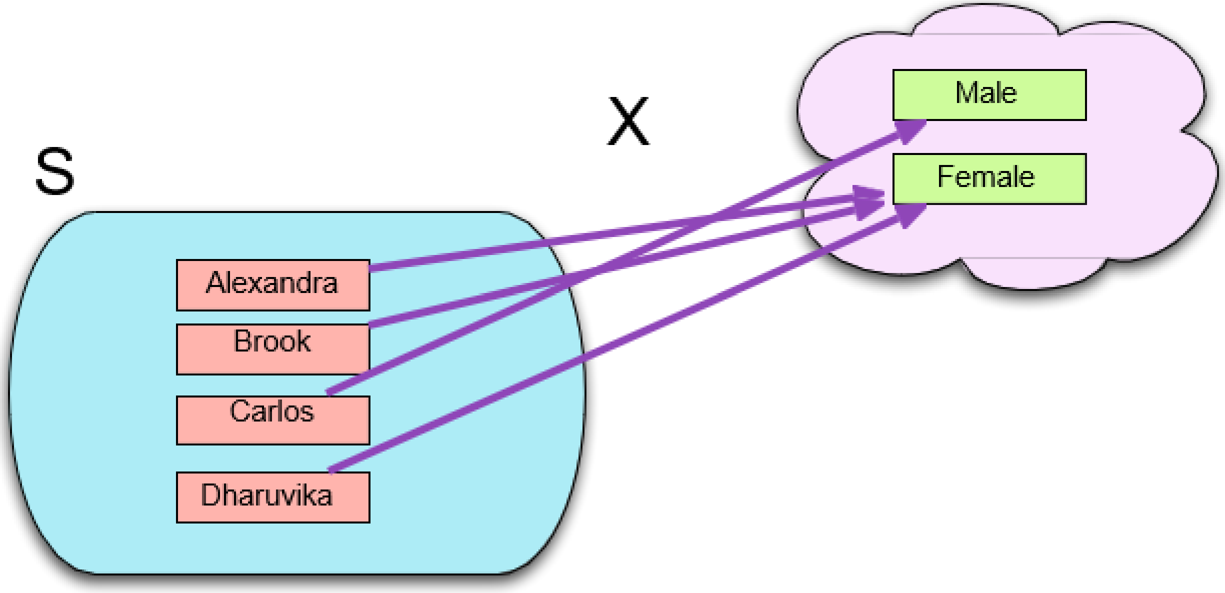
- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., gender, weight, cancer status, etc.) using a separate random variable.

Random Variable

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

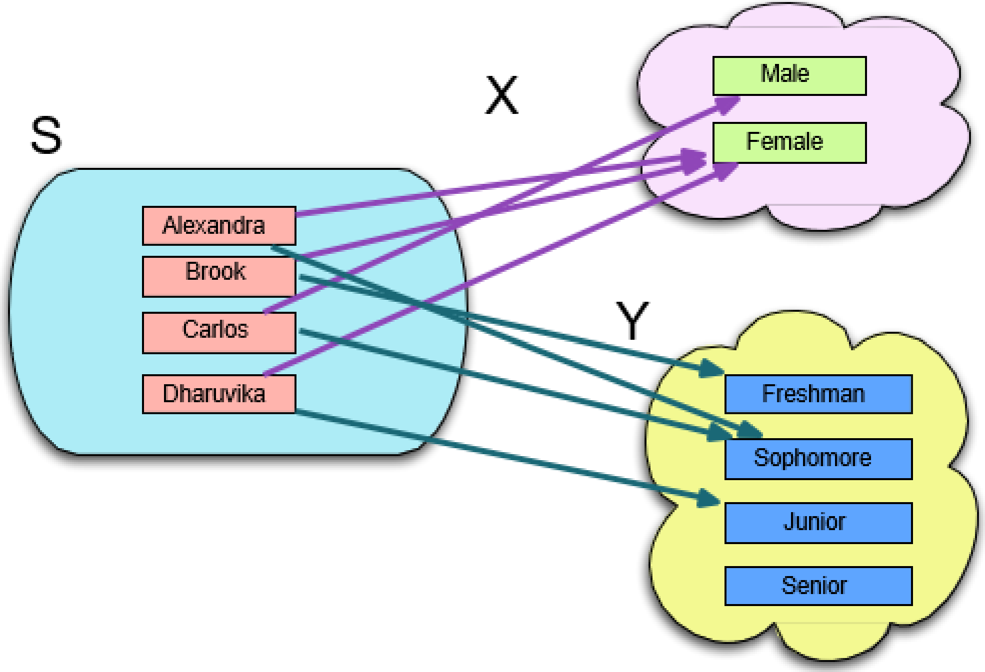
Each value of the random variable (e.g., male or female) has an associated probability.

Random Variable: Example



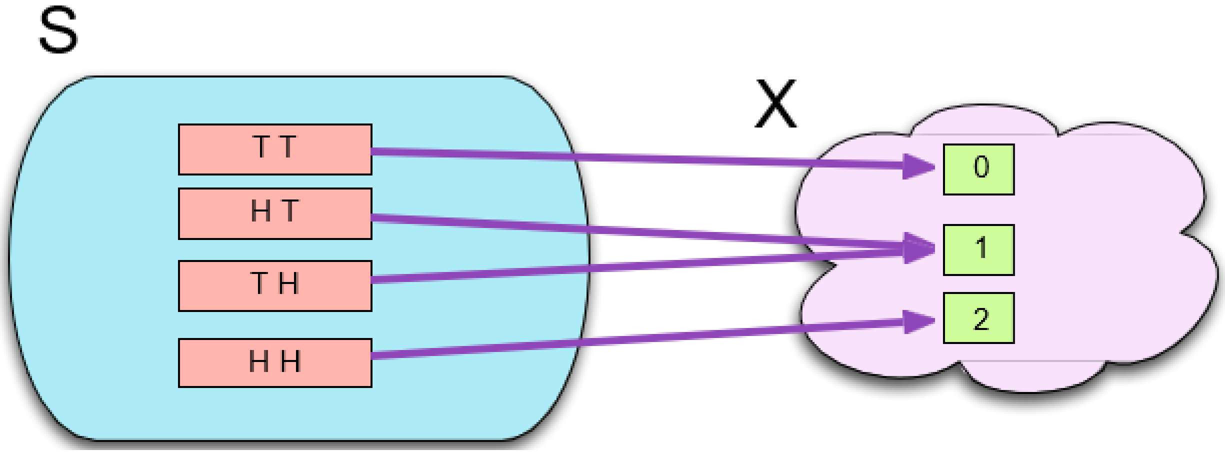
- X: people -> their genders

Random Variable: Example



- Y: people -> their class year

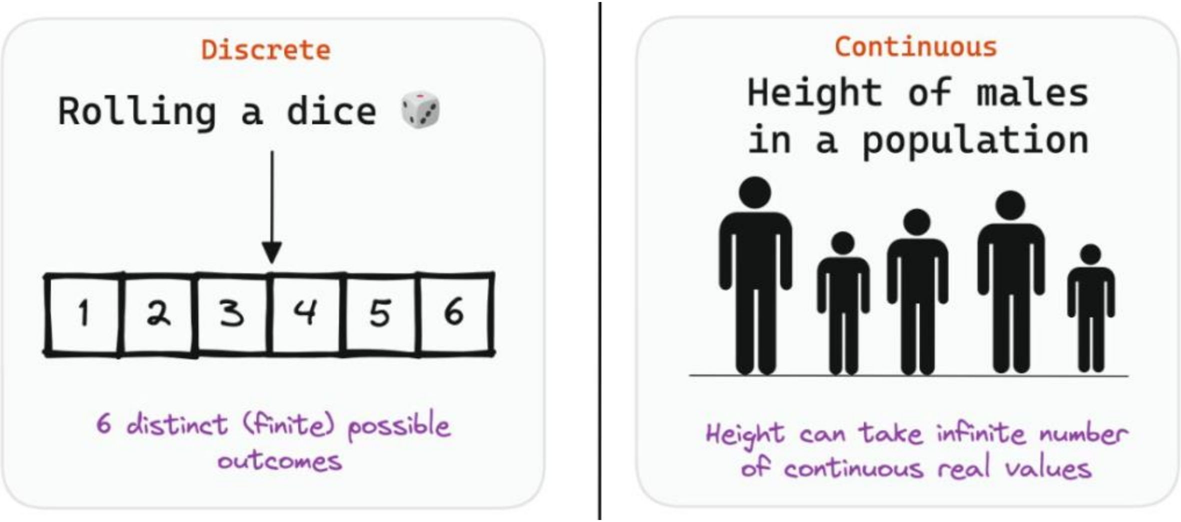
Random Variable: Example



- 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.



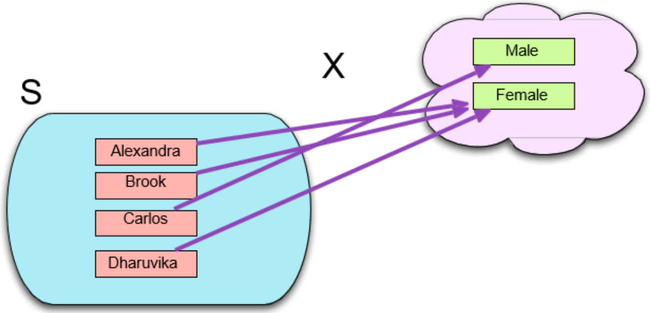
Distribution functions

Discrete distributions

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

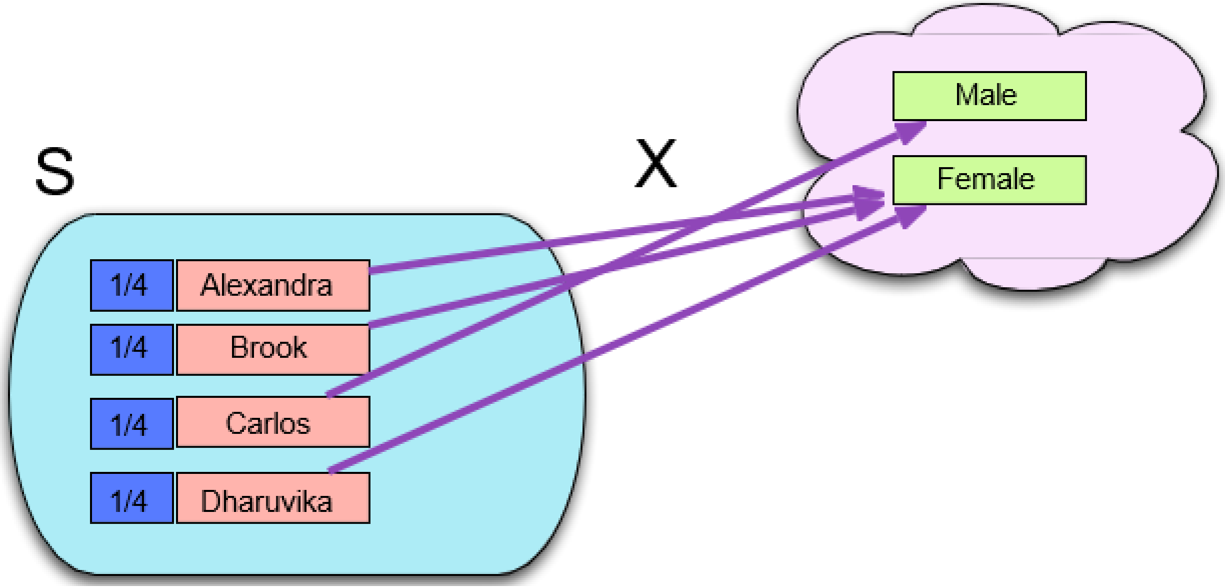
- The probability that the random variable takes a particular value comes from the probability associated with the set of individual outcomes that have that value.
    - This set is an event

- E.g. P(X = Female)



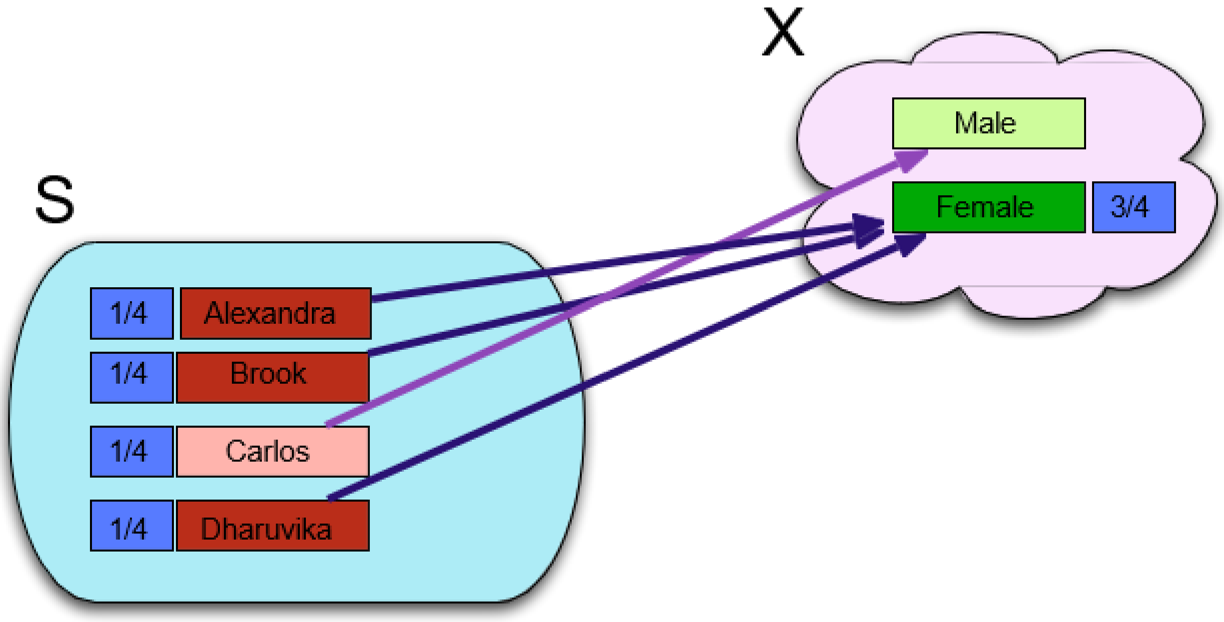
Discrete distributions

- How to find P(X = Female)?



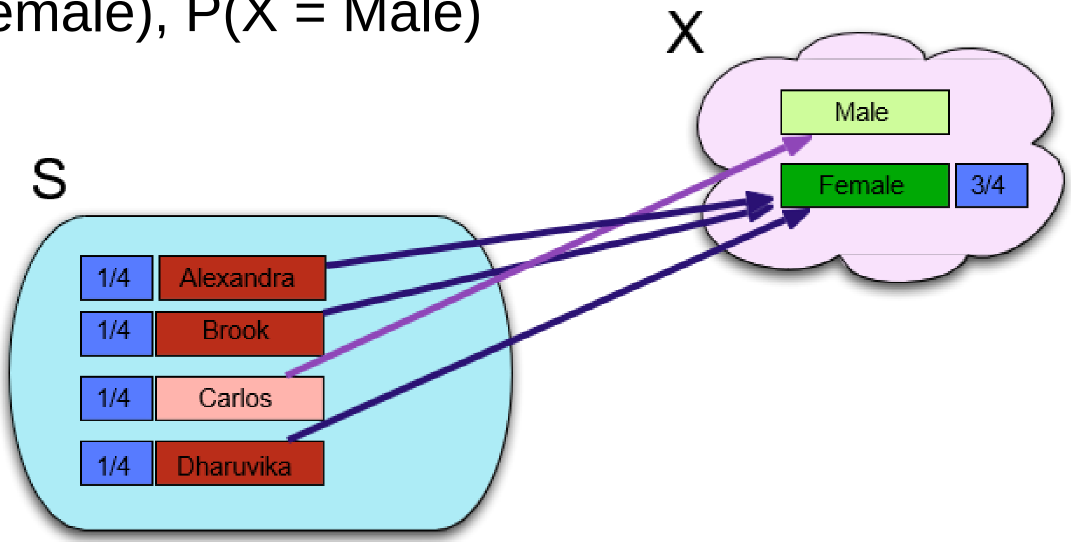
Discrete distributions

- How to find P(X = Female)?



Discrete distributions

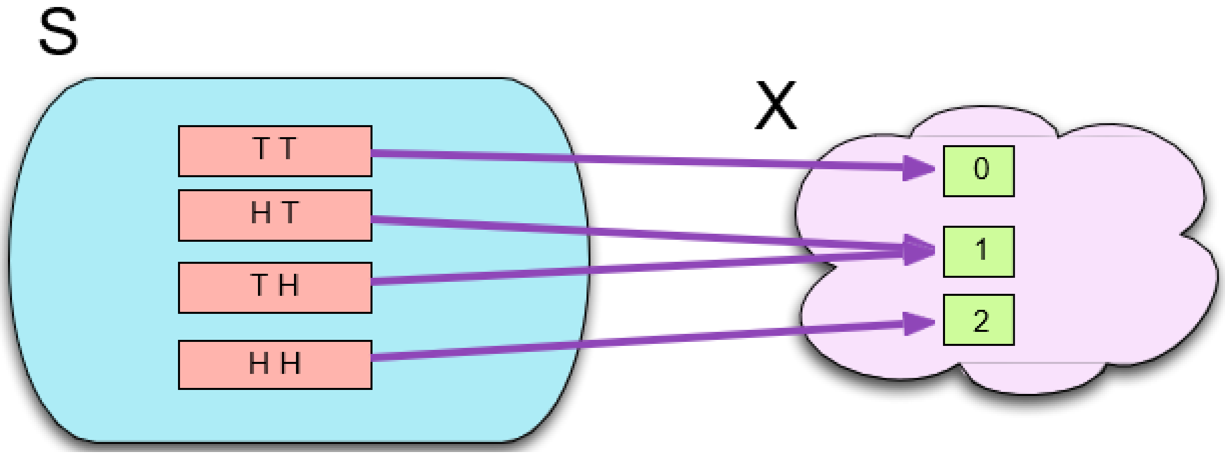
- What is the distribution of random variable X?
    - P(X = Female), P(X = Male)



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.

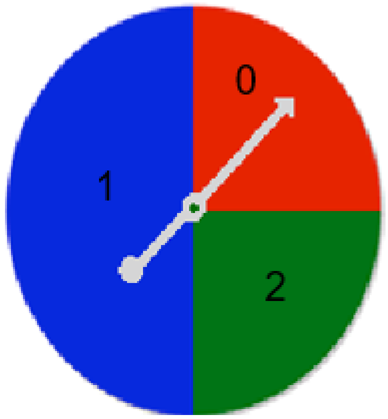
x	Male	Female
$P(X = x)$	$1/4$	$3/4$

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

- Some random variables have a PMF that has an algebraic formula (more like functions you’re used to), but we’ll get to those later.

Visualizing discrete distributions

- Very simple distributions can be visualized with a pie chart.
- Can imagine a spinner mechanism that lands on a slice according to its probability.
- But, like pie charts, this is limited in its ability to convey information.



The spike plot

- An alternative is the spike plot

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

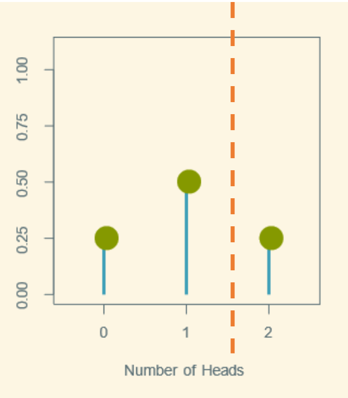
The cumulative distribution function (CDF)

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

The Cumulative Distribution Function

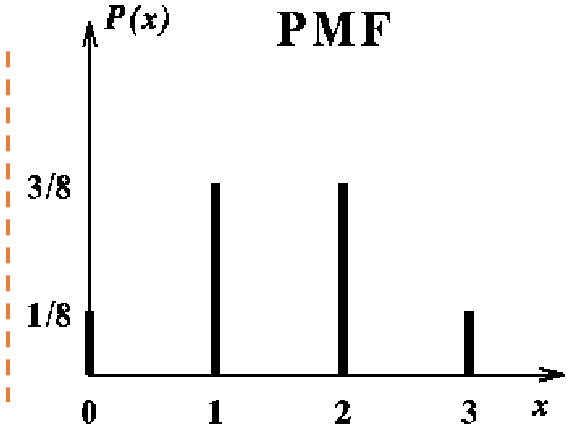
A random variable, X, can be characterized by its **cumulative distribution function**, F (or sometimes F_X if we need to be explicit), which takes values and returns *cumulative* probabilities:

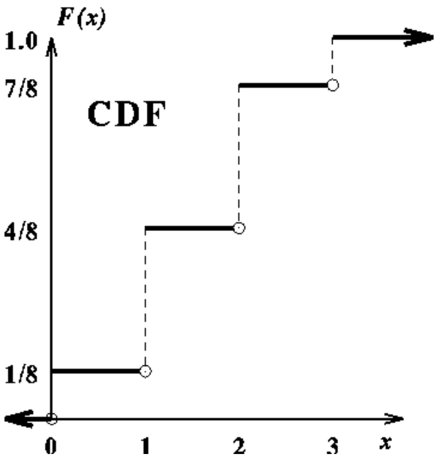
$$F(x) \text{ is defined to be } P(X \leq x)$$



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(-1)? F(-0.01)? F(0)?

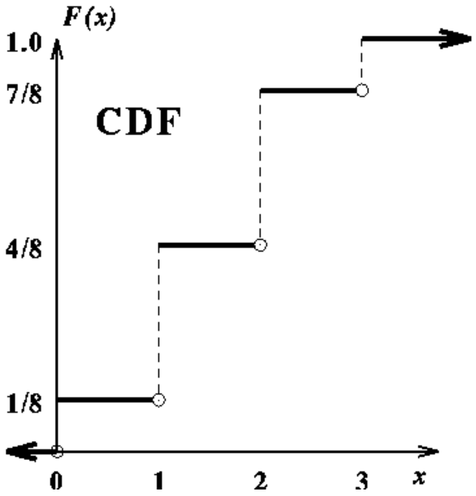




- For discrete random variables, F(x) *jumps* at locations with nonzero probability mass

Relating PMF to CDF

- How could we find f(x) from a cumulative distribution function F?



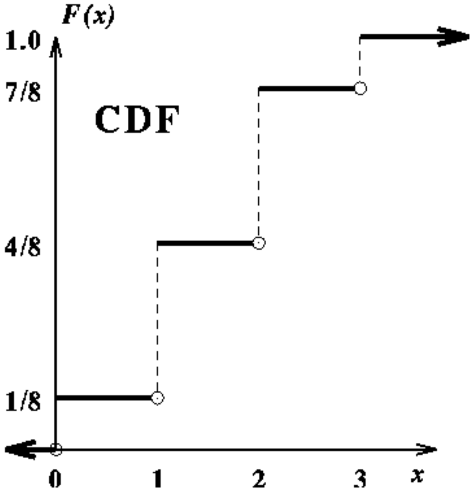
- Only need to focus on “jumps”
- For jump locations x, f(x) = F(x) – F(value below x)

Exercise: using CDF and PMF

Given the CDF F:

- How to calculate P(X > x)?
  - 1 – F(x)

- How about 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

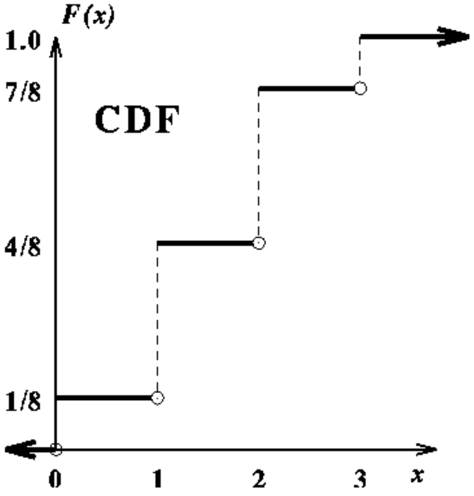
Given the CDF F:

- 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 $X + 5, 3X, X^2, \ldots,$ 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

Examples (cont'd)

- Suppose X has PMF

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

- What is the PMF of $Z = 3X$?

z	3	-3
$P(Z = z)$	0.5	0.5

- What is the PMF of $W = X^2$?

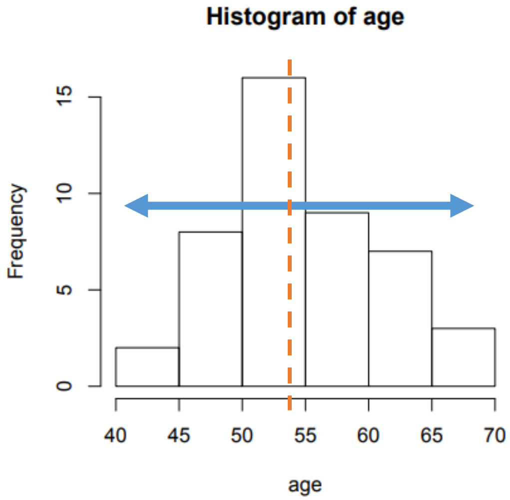
w	1
$P(W = w)$	1

Note: $\{W = 1\} = \{X = +1 \text{ or } X = -1\}$

Mean and Variance

Summarizing random variables

- As with data (lec. 2), it is useful to characterize the *center* and *spread* of a probability distribution.



- Let us ask questions like “what value do we expect to occur”, and “how confident are we in our prediction” (roughly, “how far off do we expect to be on average”)?

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; but the average is *weighted by the probabilities*.

- Makes sense if you think of probability as long-run proportion: in the long run, a value with a probability of 1/2 will occur twice as often as one with a probability 1/4, etc., so it should count twice as much in determining the 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

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

- Interpretation: how much net win do we expect in one 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

- Ans: 0 x 0.7 + 1 x 0.2 + 2 x 0.1 = 0.4

Expectation formula

- Given RV X and its PMF, how do we 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 just 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

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

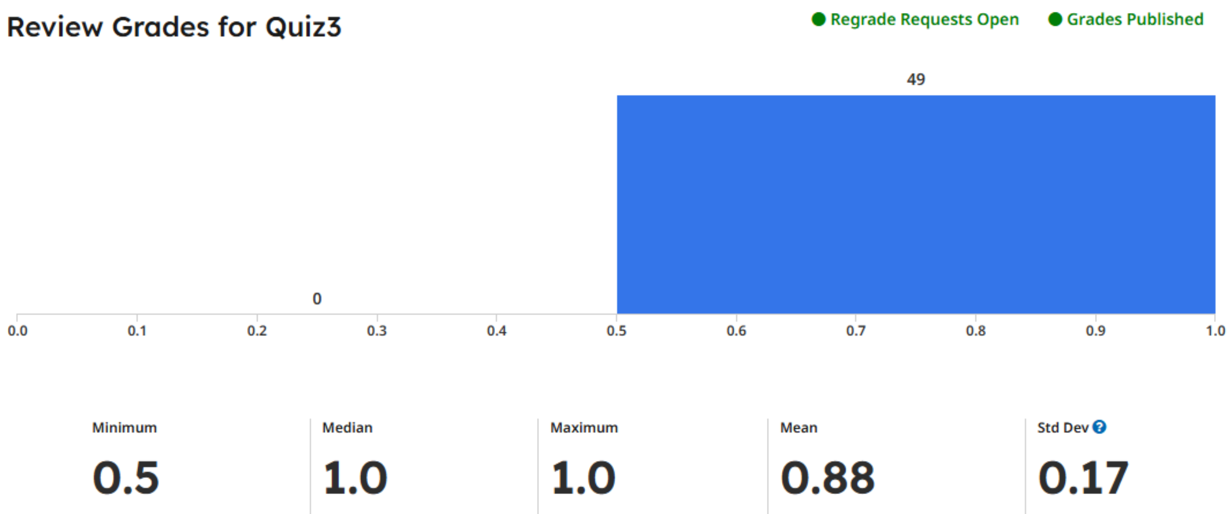
Standard deviation

- 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

Announcements 2/17

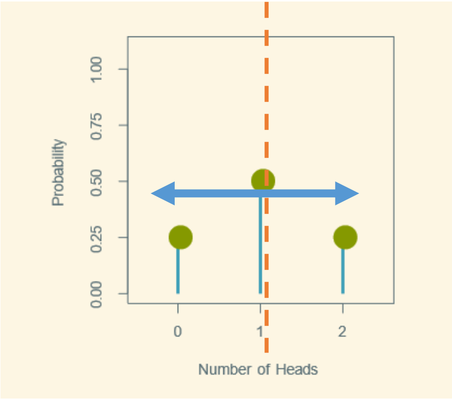
- Quiz 3 graded



- We will have a quiz next time (2/19)

Recap 2/17

- Discrete random variables
    - PMF, CDF
- Summary of the distributions



Expected Value of a Discrete Random Variable

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

Variance of a Discrete Random Variable

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

Expectation formula

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

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:

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

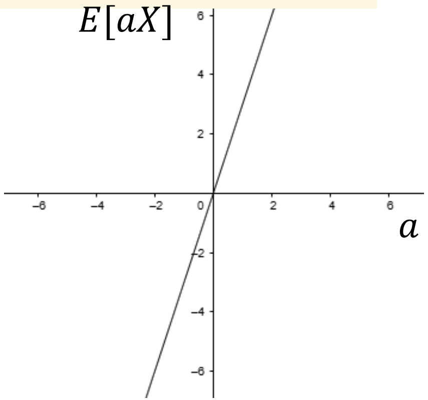
Properties of expectation

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

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

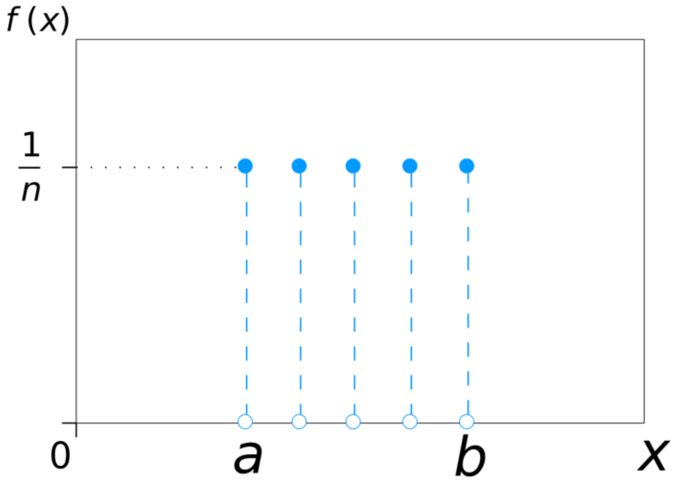
Uniform distribution over a set

Suppose X is the outcome of throwing a fair N-faced die. Its PMF is:

$$p(X = k) = \frac{1}{N}$$

for $k = 1, .., N$

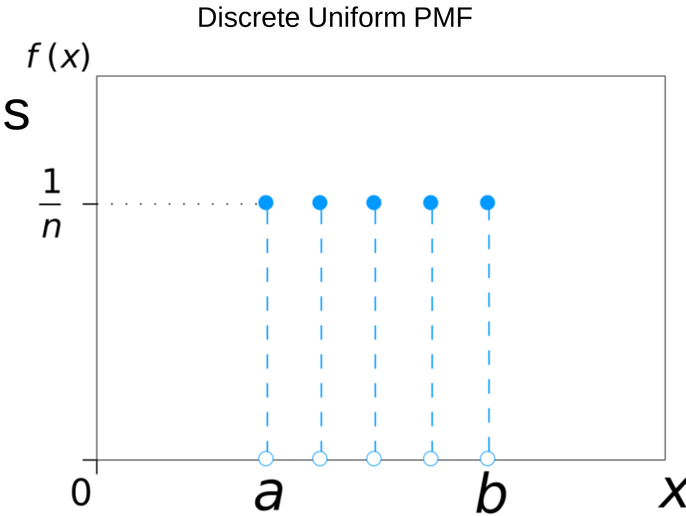
Discrete Uniform PMF



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 $X \sim \mathrm{Uniform}(S)$

- 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

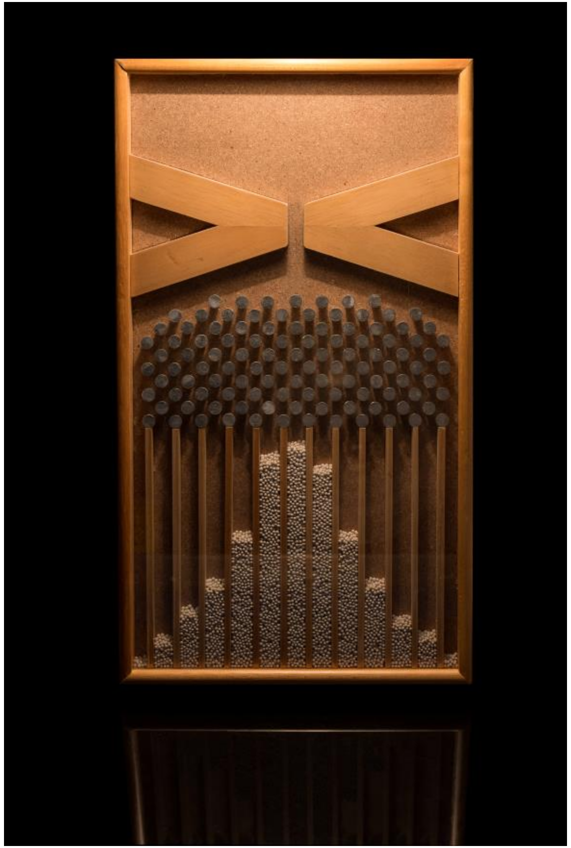
$$X \sim \text{Bin}(n, p)$$

Galton Boards

- Illustration of binomial distribution
- Bead has 10 chances hitting a peg
- 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:

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

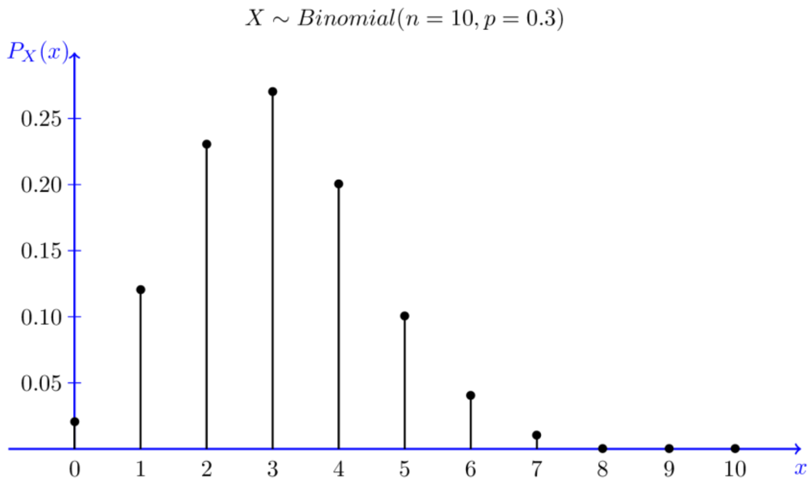


Binomial distribution

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

Facts:

- $\text{E}[X] = ?$
  - 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 $\text{Bernoulli}(p)$

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 a question in HW)

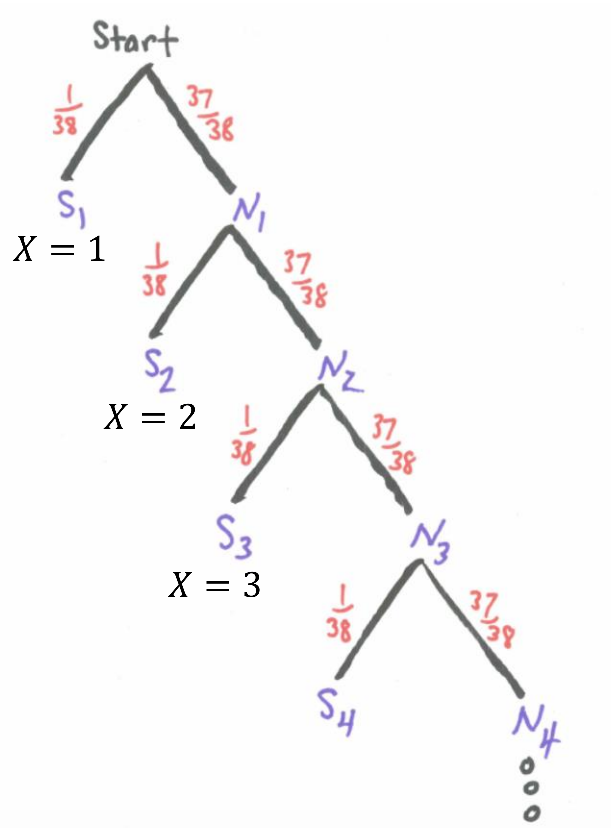
- Applications:
  - Call center: # calls before encountering first dissatisfied customer
  - Basketball: # shots before scoring the first
  - Networking: # attempts before a successful transmission
  - Gambling: # plays before first win

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

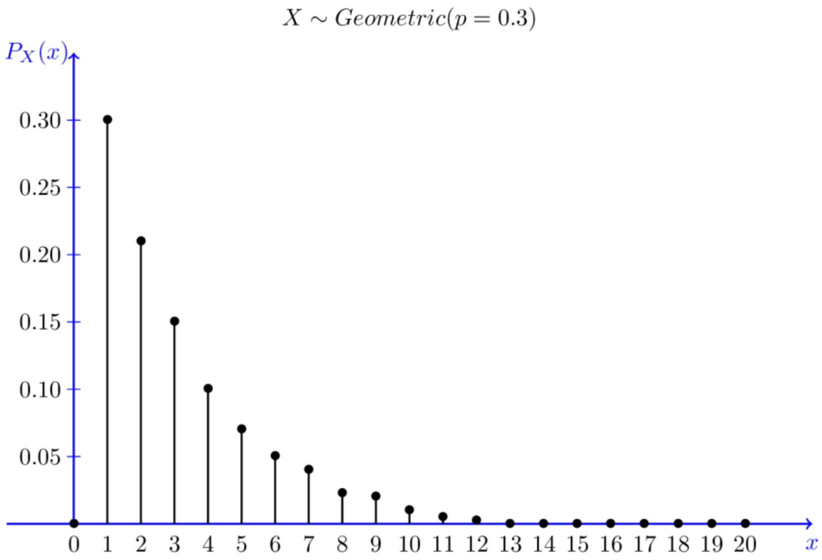
- In conclusion,

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

  for $x = 1,2, \ldots$

Fact:

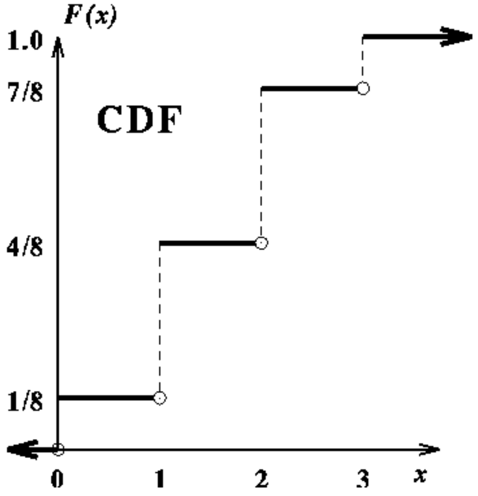
- $\mathrm{E}[X] = ?$
  - $\frac{1}{p}$
- $\mathrm{Var}[X] = \frac{1-p}{p^2}$
  - Smaller when p tends to 1



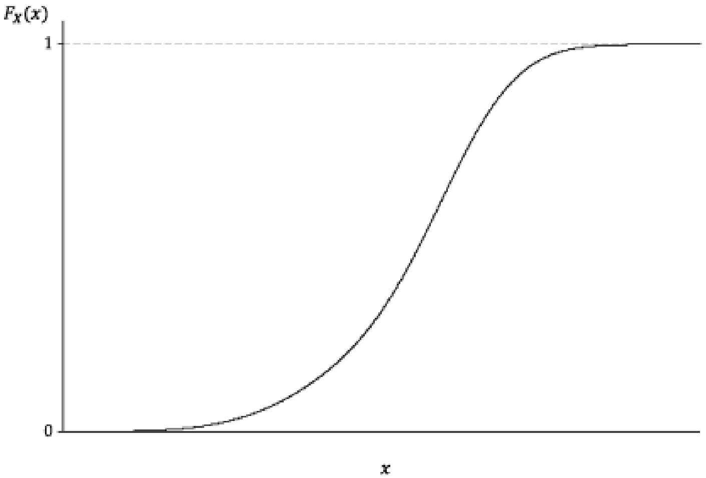
Continuous Random Variables

Continuous random variables

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



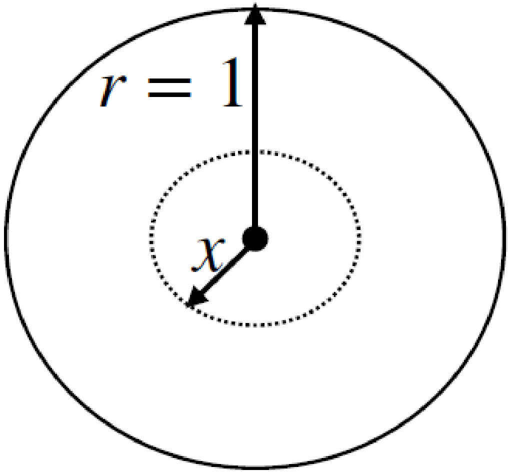
- Continuous random variables take values in a continuous set
- Their CDFs are continuous



Example: dart

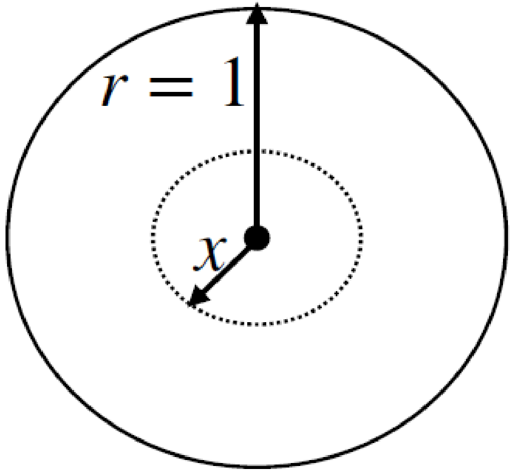
- Dartboard with radius 1; dart lands uniformly at random on the board

- X: distance to the center
- What is the CDF of X?
- $P(X \leq x) = \frac{\pi\, x^2}{\pi} = x^2$ 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: dart

- Dartboard with radius 1; lands uniformly at random on the board
- X: distance to 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 area of the circle of radius 0.3 is zero



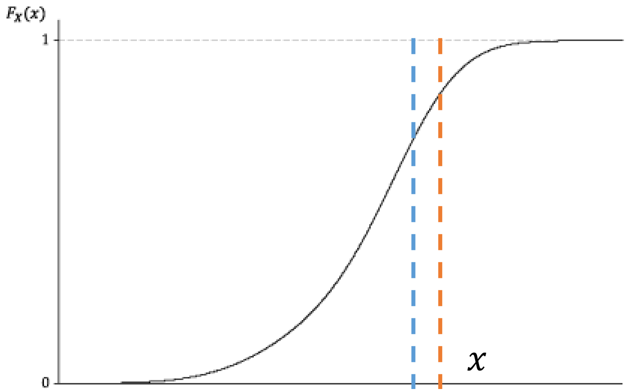
Continuous random variables

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

- In other words,
- P(X=x) = F(x) – F(value below x)

  is still true

- Maybe mind-blowing at first sight
  - All outcomes have zero probabilities?!



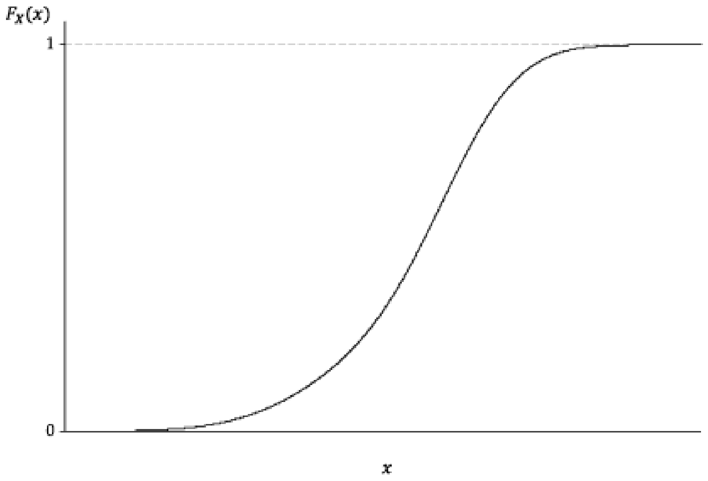
Value below x

Maybe it is not that weird..

- 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? $P(X = a) = P(X = b) = 0$

Announcements 2/19

For HW & Quiz & Exam, to get full credit:

- Show your steps

  "The size of the sample space is 36, the event E has 5 outcomes {(1,5), (2,4), (3,3), (4,2), (5,1)}, and thus the probability P(E) is |E|/|S| = 5/36"

- If you answer is a fraction, present its simplified form
  - E.g. 15/35 => 3/7

Quiz 4

- Fill in the missing value for the Probability Mass Function below

x	0	1	2	3
$f(x)$	0.20	0.45		0.15

PMF should sum to 1 => 0.2

- Fill in the missing value for the Cumulative Distribution Function below (not the same random variable as above).

x	3	5	7	9
$F(x)$	0.15	0.35	0.75	

Last point in the CDF should always be 1

- Fill in the missing values for the PMF and CDF below

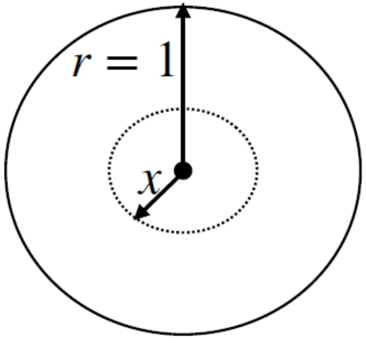
x	1	2	3	4	5
$f(x)$	0.10		0.40	0.15	0.10
$F(x)$	0.10	0.35	0.75		1.00

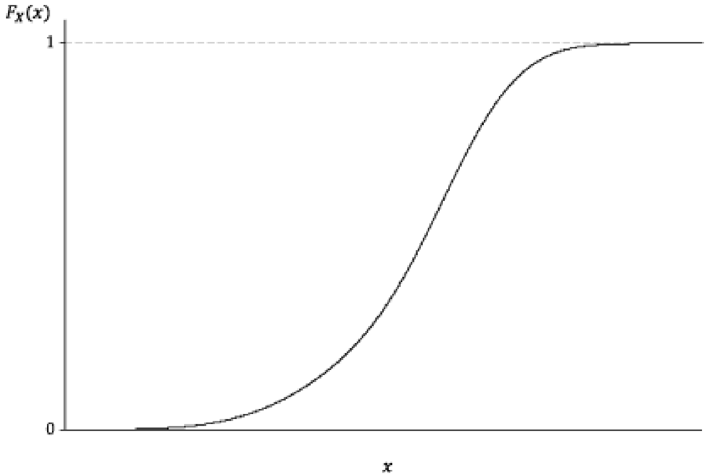
F(2) = f(1) + f(2) => f(2) = 0.25

F(4) = F(3) + f(4) => F(4) = 0.9

Recap: continuous random variables

- Continuous RVs are those whose CDFs are continuous (no jumps)
- E.g. X: distance to the center

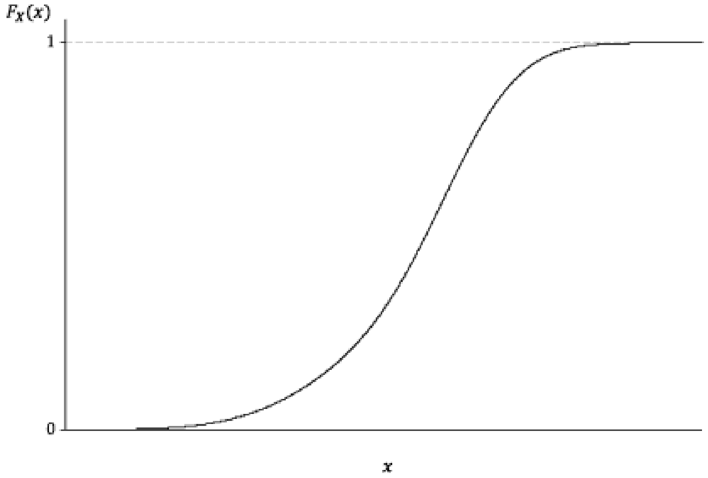




- Different from discrete RVs, the sum of the probabilities of all outcomes are not 1
  - In fact, $\mathrm{P}(X = x) = 0$ for all x

CDF for continuous RVs

- Suppose $F(x) = P(X \leq x)$ is the CDF of continuous RV X



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

- What is the analogue of PMF for continuous RVs?

- Can we define function f such that

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

- We cannot use $P(X = x)$, since it takes value 0 everywhere

- Need something else

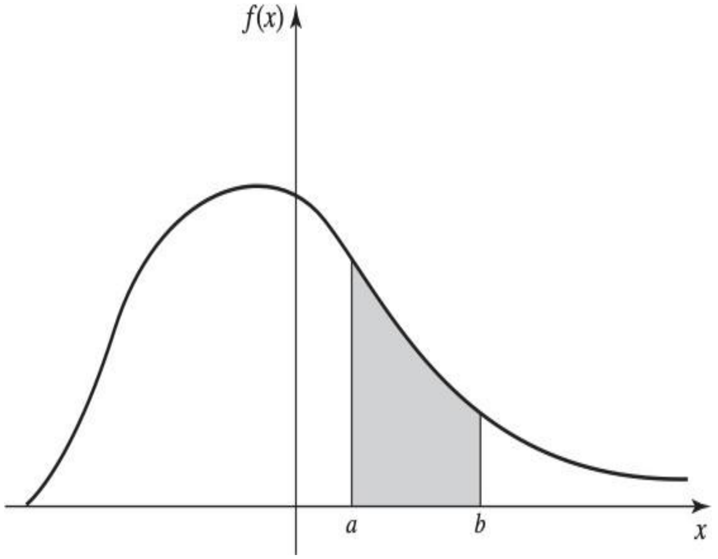
Math interlude: integration

- How to calculate 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

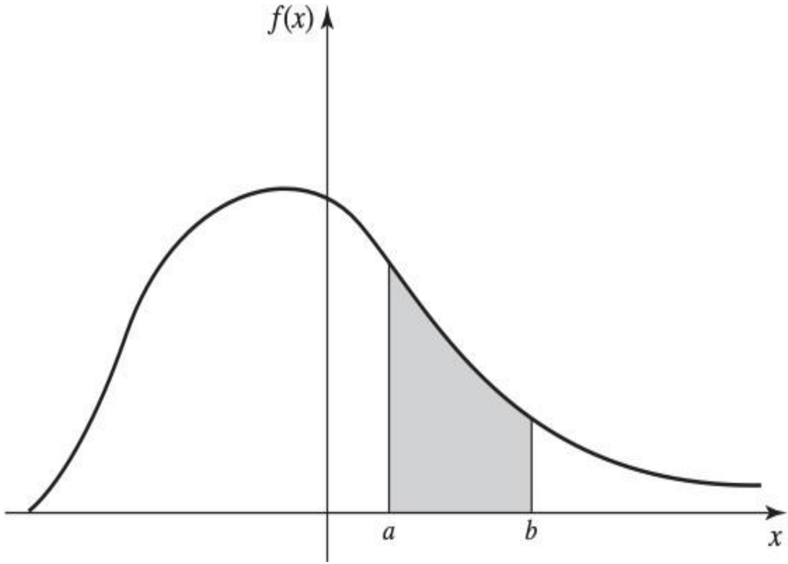
$$\int_a^b f(x)\,dx$$



Reads "the integral of f from a to b"

Math interlude: integration

Applications of integration



- x: time, $f(x)$: speed
- $\int_a^b f(x)\,dx$: total distance traveled

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

Math interlude: integration

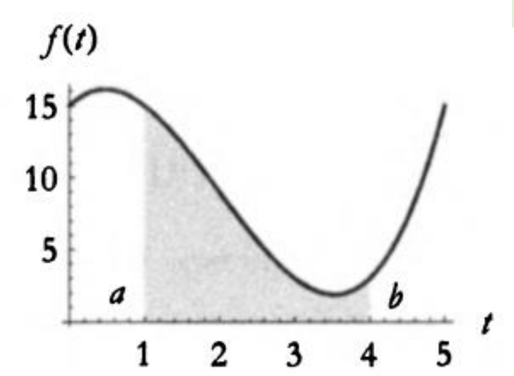
Why the weird $\int$ symbol?

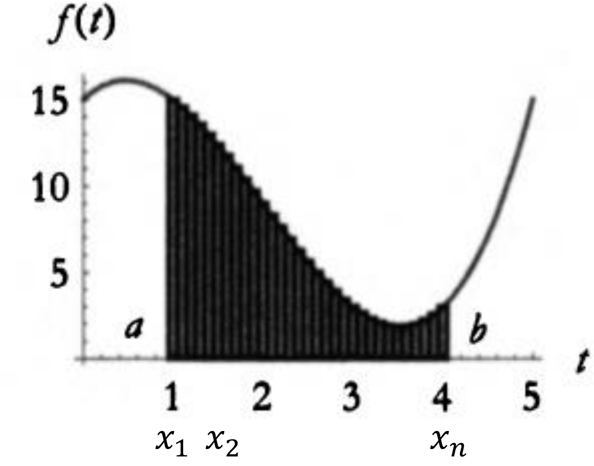
'$\int$ ' is a stylized version of 'S', representing sum

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

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

With the partition being finer, this tends to $\int_a^b f(x)\,dx$ (HW4)





Math interlude: integration

- How to calculate the area under the curve of $f(x)$, for $x \in [a, b]$?

- **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

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

- 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

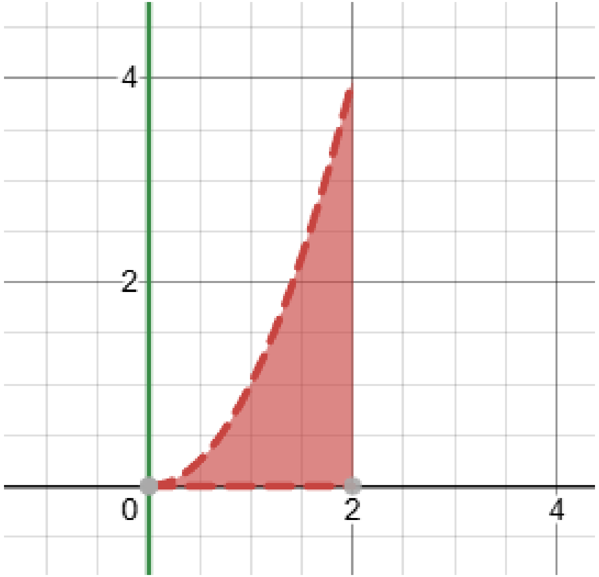
$$\begin{tikzcd} f \arrow[r, bend right, "\text{Antiderivative}"'] & F \arrow[l, bend right, "\text{Derivative}"'] \end{tikzcd}$$

- 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

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 = $F(2) - F(0) = \frac{8}{3}$

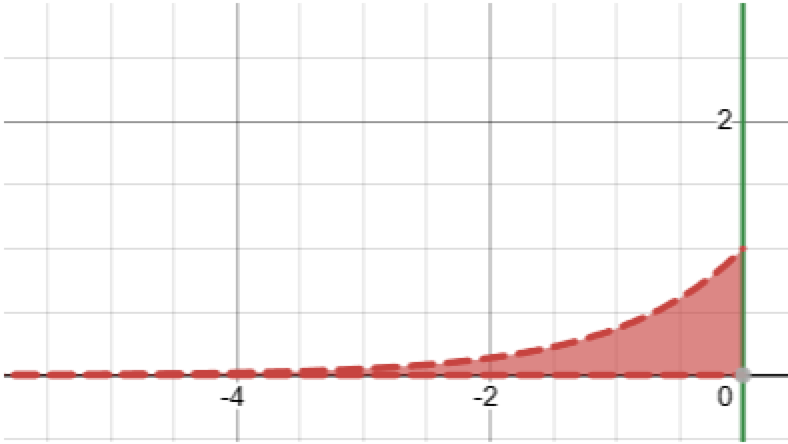


Math interlude: integration

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

- Step 1: find F, antiderivative of e^x
  - $F(x) = e^x$

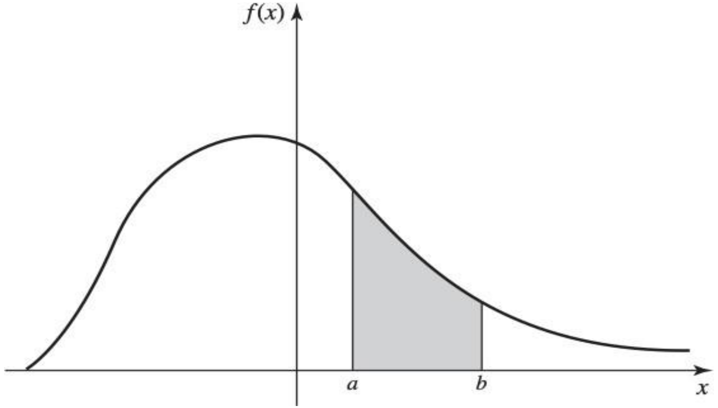
- Step 2: evaluate F at both end points
  - $F(0) = 1, F(-\infty) = 0$
  - Ans = $F(0) - F(-\infty) = 1$



Probability density function (PDF)

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

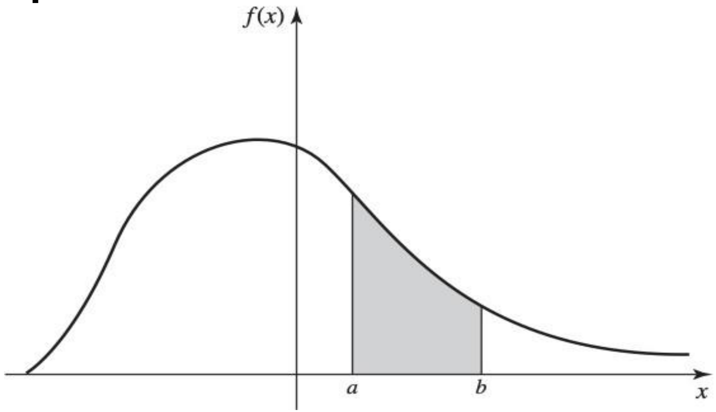
$$P(a \le X \le b) = \int_a^b f(x)\,dx$$



function f is called the *probability density function* of X

Probability density function (PDF)

- $f(x)$ measures how likely X takes value in the *neighborhood* of x
- graph of f: "histogram of infinite samples of X"

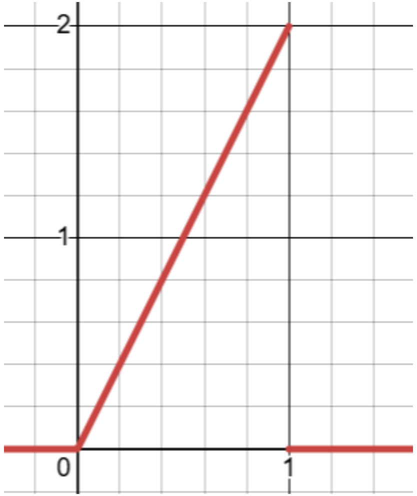


- Another view: X is drawn from a pile of sand
  - $f(x)$: height of sand a location x
  - A random sample X is drawn by choosing a grain of sand from the pile and return its location

Probability density function (PDF)

Example X: lifetime of a lightbulb, has PDF

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

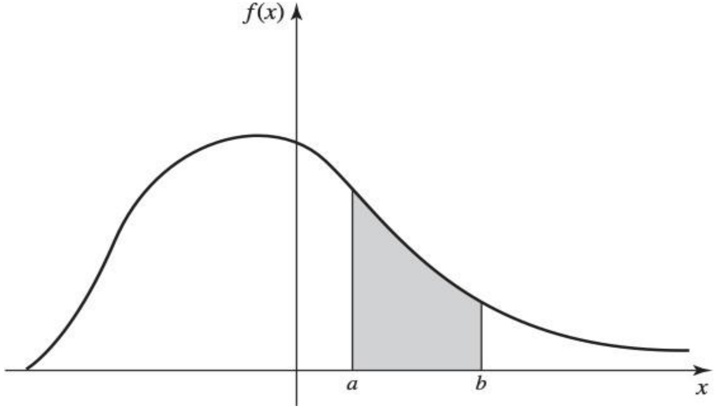


Find $\mathrm{P}(0.3 < X < 0.5)$

Soln This is equal to

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

Properties of PDF



- Nonnegativity: $f(x) \geq 0$ for all x

- Normalized:

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

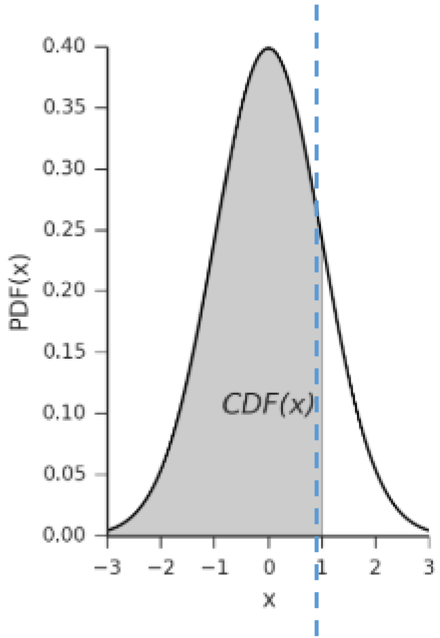
- Why?
    - The integral represents $P(-\infty \leq X \leq +\infty)$

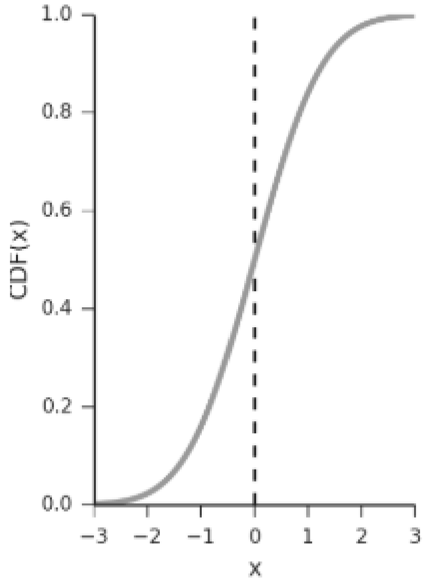
Relationship between PDF and CDF

- How to find CDF F based on PDF f?

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

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

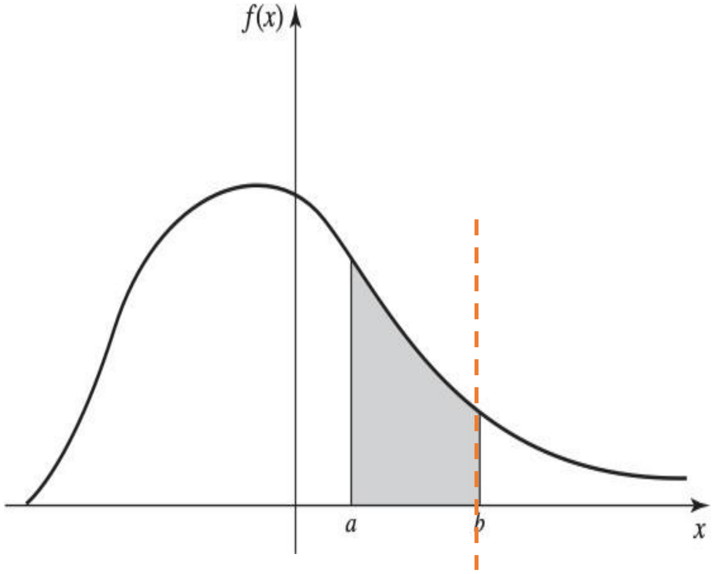




Math interlude: integration

Indefinite integral

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



Example

- x: time, $f(x)$: speed
- $F_a(b)$: displacement at time b (relative to time a)

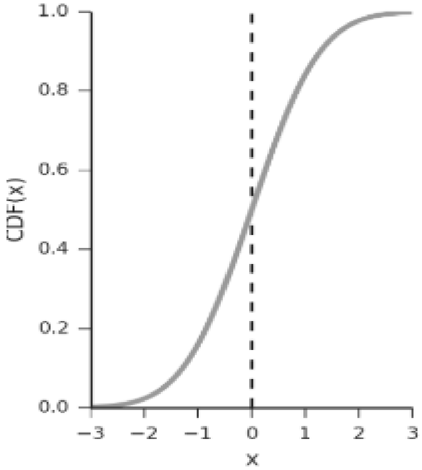
Fact:

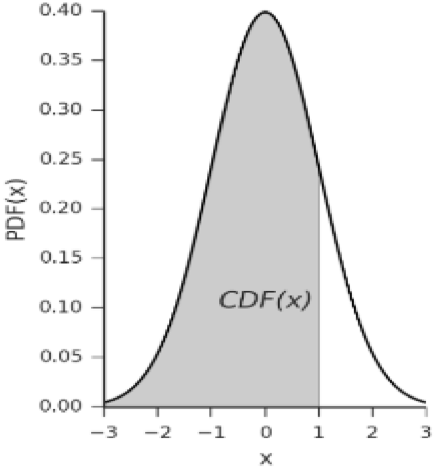
f is the derivative of F_a

(i.e., F_a is an antiderivative of f)

Relationship between PDF and CDF

- How to find PDF f based on CDF F? $F(b) = \int_{-\infty}^{b} f(x)\,dx$
- F is an indefinite integral of f => $f(x) = F'(x)$



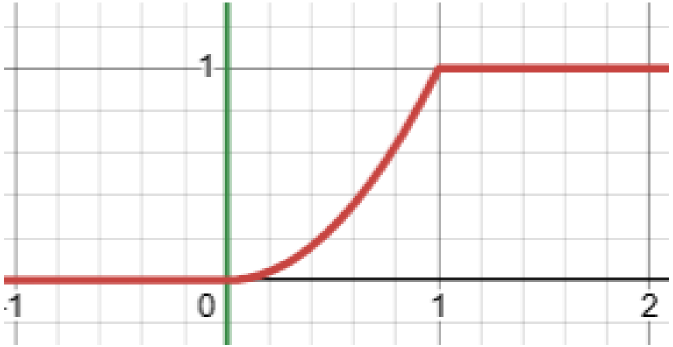


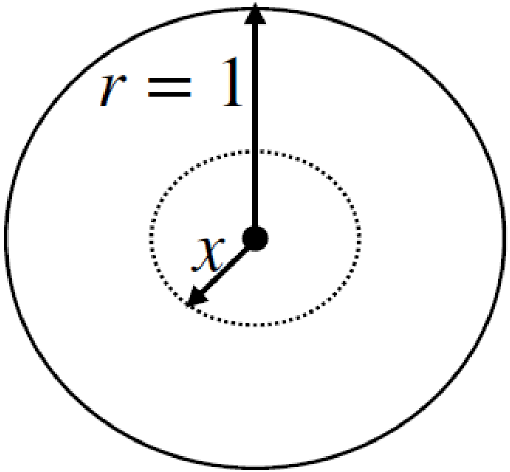
- F has large slope at x => $f(x)$ is large

Example: dart

- X: distance to the center

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



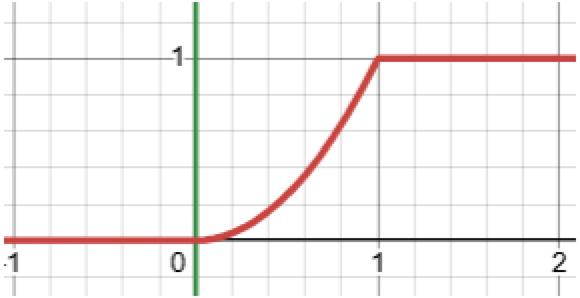


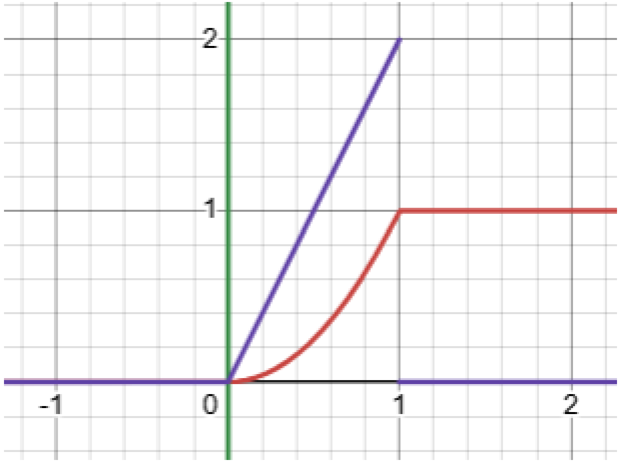
- What is the PDF of X?

Example: dart

What is the PDF of X?

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





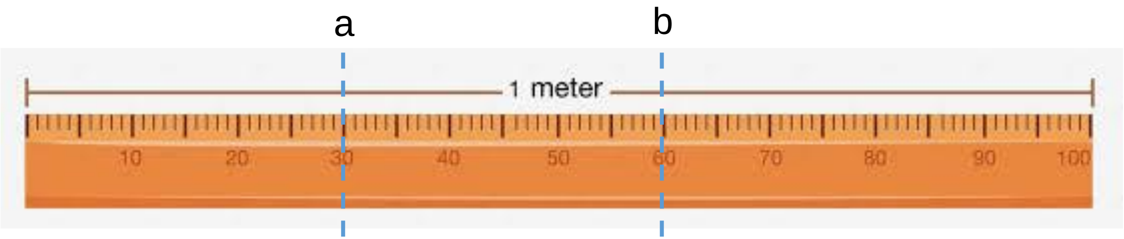
How to find the distribution of a continuous RV

We can follow this recipe:

- Step 1: Find its CDF $F(x) = P(X \leq x)$
- Step 2: find its PDF $f(x)$ by taking derivative of F

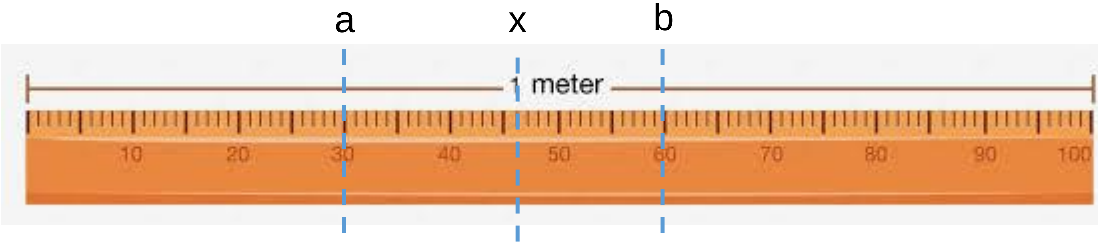
In-class activity: ruler

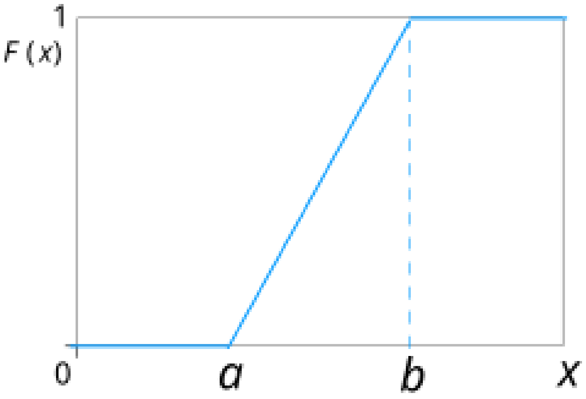
- We choose X uniformly at random from $[a, b]$, two points in a ruler
- Find the CDF and PDF of X



In-class activity: ruler

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





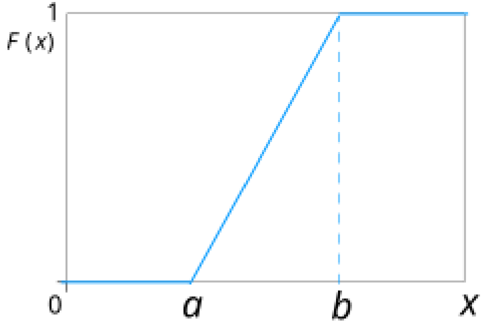
In-class activity: ruler

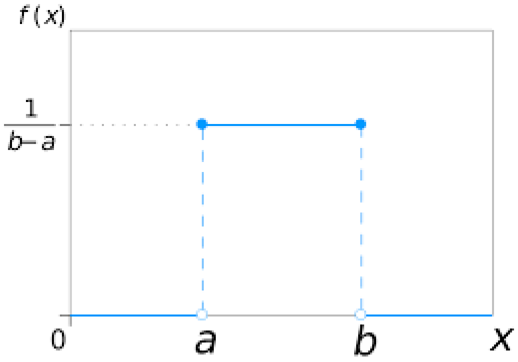
- Now that we have CDF F

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

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

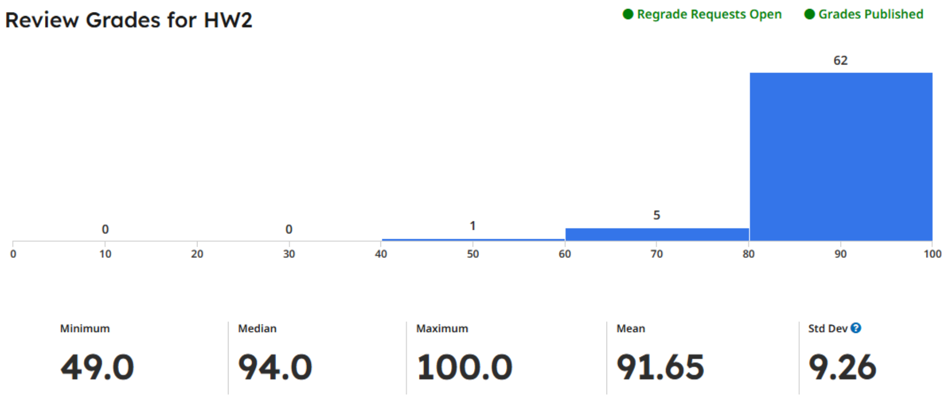
$$\text{Uniform}([a, b])$$





Announcements 2/24

- HW2 was graded last Wednesday

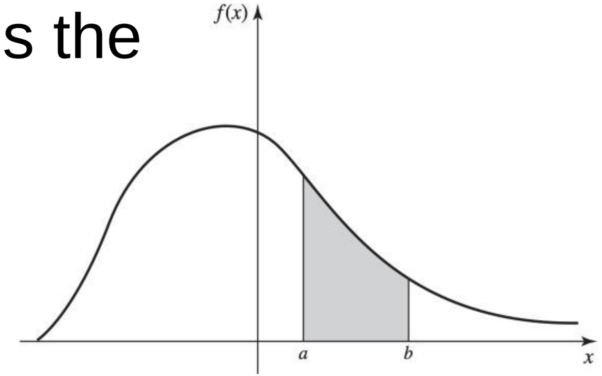


- We will have a quiz next lecture (2/26)
- I am planning to have a midterm review session in the 2nd half of next Monday's lecture (3/3)

Recap 2/24

- Probability density function characterizes the distribution of continuous RVs

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

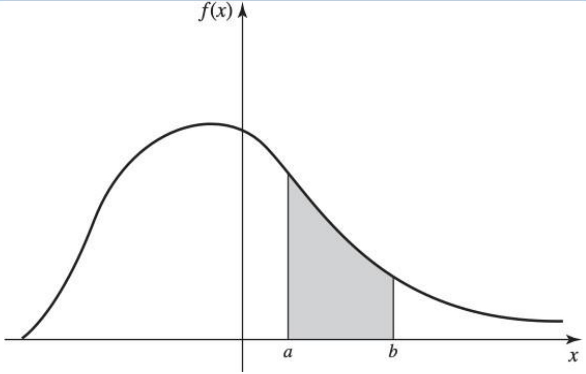


- $f(x) = F'(x)$, where F is CDF of X

Many ways to think of f:

- Histograms of samples of X
- X as a pile of sand – heights of the sand

Recap 2/24



- Is $f(x)$ equal to $P(X = x)$?
  - No -- $P(X = x) = 0$ always

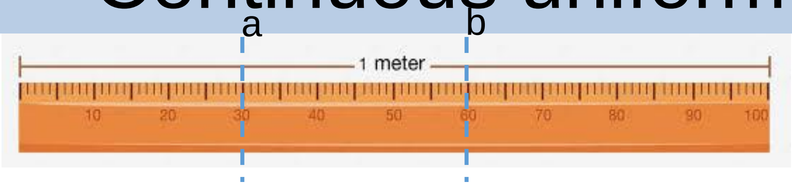
- Correct interpretation of $f(x)$: probability *density* (not probability)
  - Probability = probability density × size of region

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

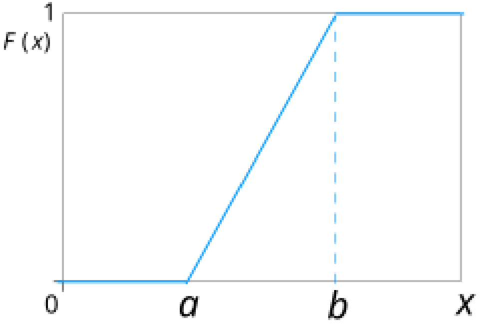
Plan 2/24

- Transformations of continuous RVs
- Summarizing continuous RVs: expectation, variance
- Useful continuous RVs

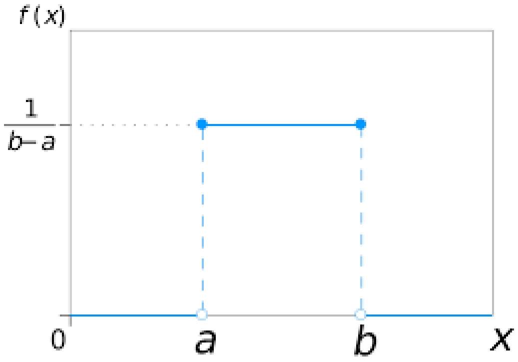
Continuous uniform distribution $\mathrm{Uniform}([a, b])$



- Its CDF F
  - $F(x) = 0, \; x < a$
  - $F(x) = \frac{x-a}{b-a}, \; x \in [a, b]$
  - $F(x) = 1, \; x > b$



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



Uniform distribution

numpy.random.uniform

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

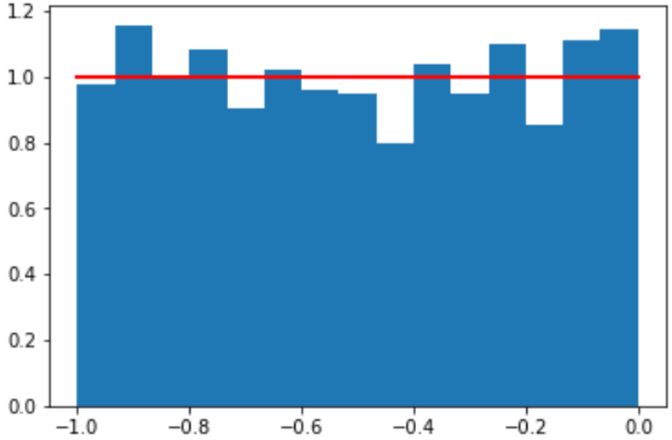
Draw samples from a uniform distribution.

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),

```
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()
```

redline: PDF of uniform distr.



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
  - X: seismic wave amplitude; $\log_{10}(X)$: Richter magnitude
- 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 $X \sim \mathrm{Uniform}([0,1])$. Find the distribution of $Y = X + b$.

Step 1: write down the CDF of X

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

Step 2: write down the CDF of Y

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

- $y < b$: 0
- $y \in [b, b+1]$: $y - b$
- $y > b + 1$: 1

Transformations of a continuous RV

Step 2: write down the CDF of Y

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

(do you recognize this CDF?)

Step 3: Take derivative to get the PDF of Y

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

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

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

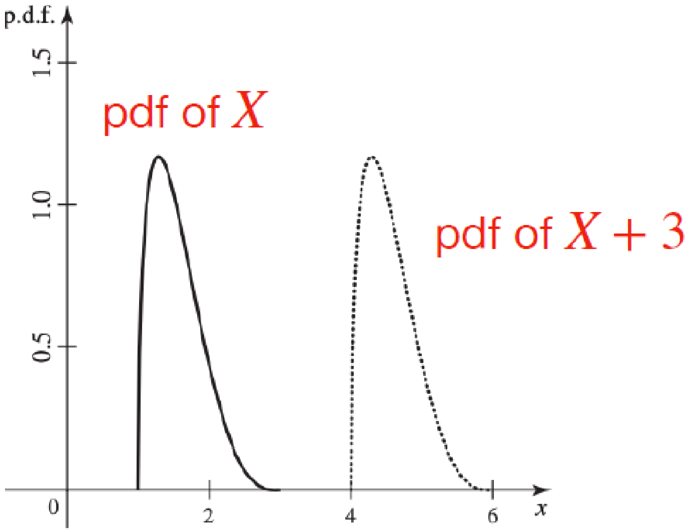
- Step 2: write down the CDF of Z

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

- Conclusion: $Z \sim \mathrm{Uniform}([0, a])$

Shifting a continuous RV

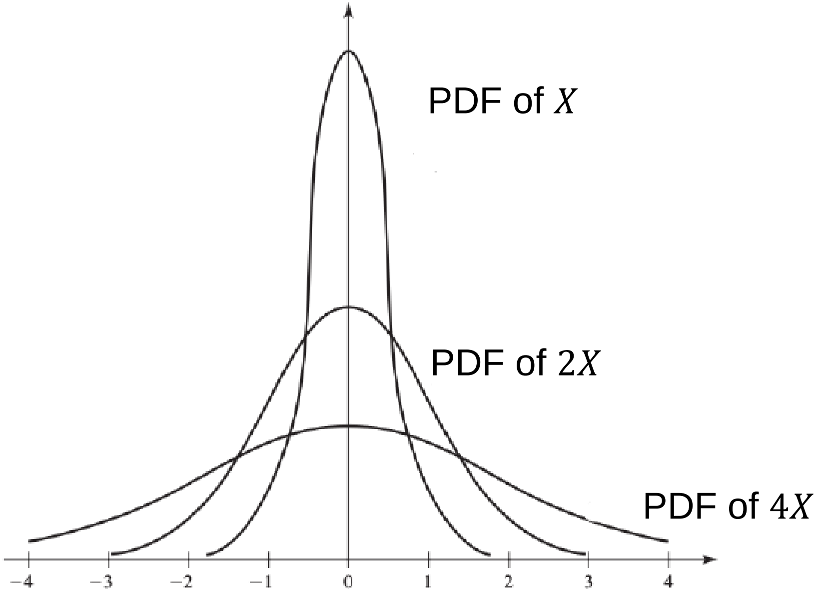
- In general:
- $X + b$ has a PDF that is a translation of X's PDF (by b units)



- $f_{X+b}(x) = f_X(x - b)$

Scaling a continuous RV

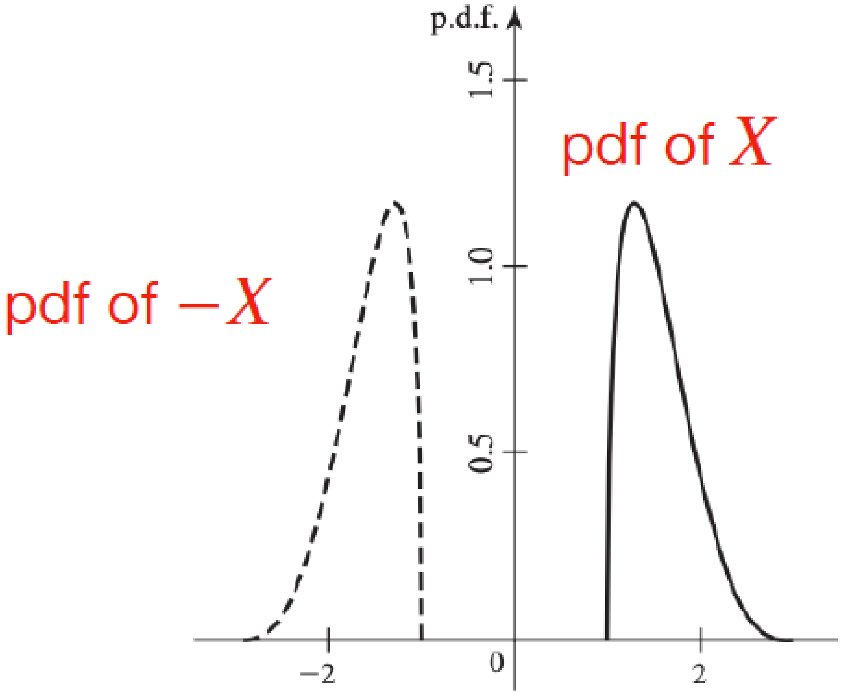
- aX's PDF is X's PDF stretched by a factor of a horizontally



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

Scaling a continuous RV

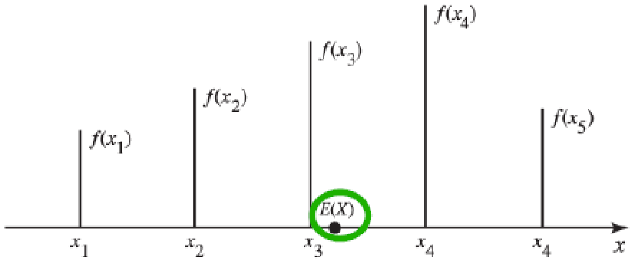
- Given X's PDF; what does $-X$'s PDF look like?

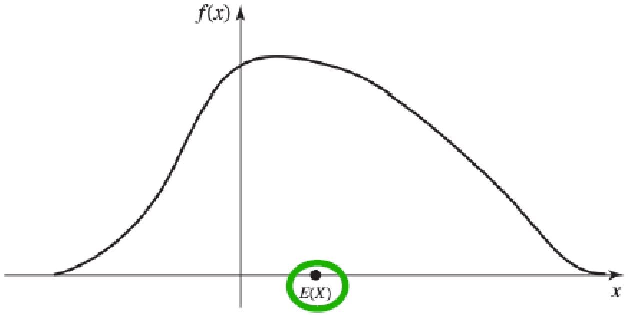


Summarizing Continuous Random Variables

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

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

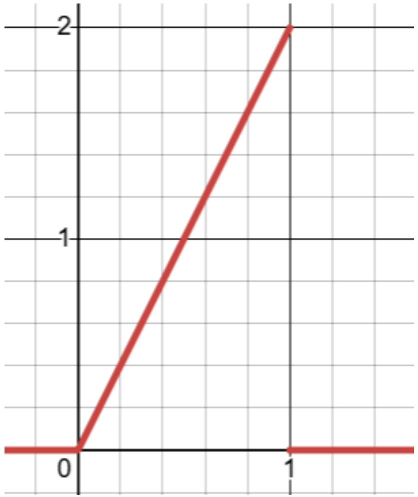
Continuous

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

Mean

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

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



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

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

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

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,

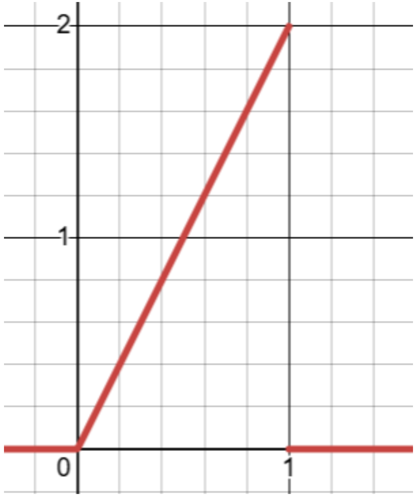
$$\mathrm{E}[r(X)] = \int r(x) f(x)\, dx$$

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

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

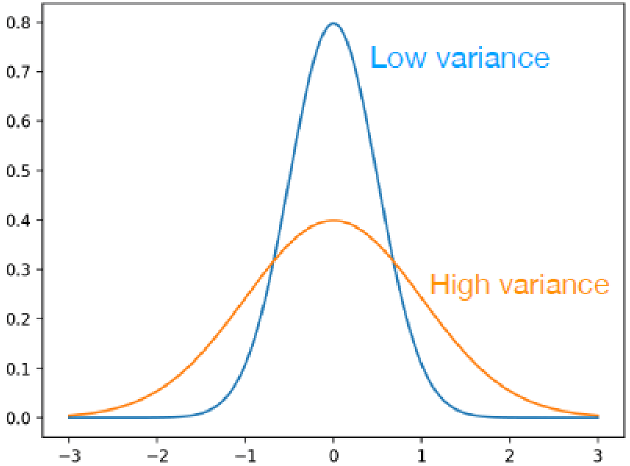
$$= \int_0^1 \sqrt{x}(2x)\, dx = \int_0^1 2x^{\frac{3}{2}}\, dx = \frac{4}{5}$$

Variance

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

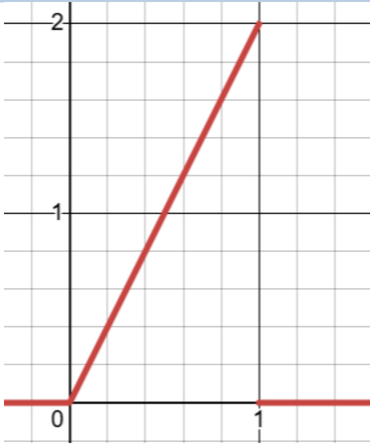
- Defn: $\mathrm{Var}(X) = \sigma^2 = \mathrm{E}[(X - \mu)^2]$

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 $\mathrm{E}[X^2]$

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

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

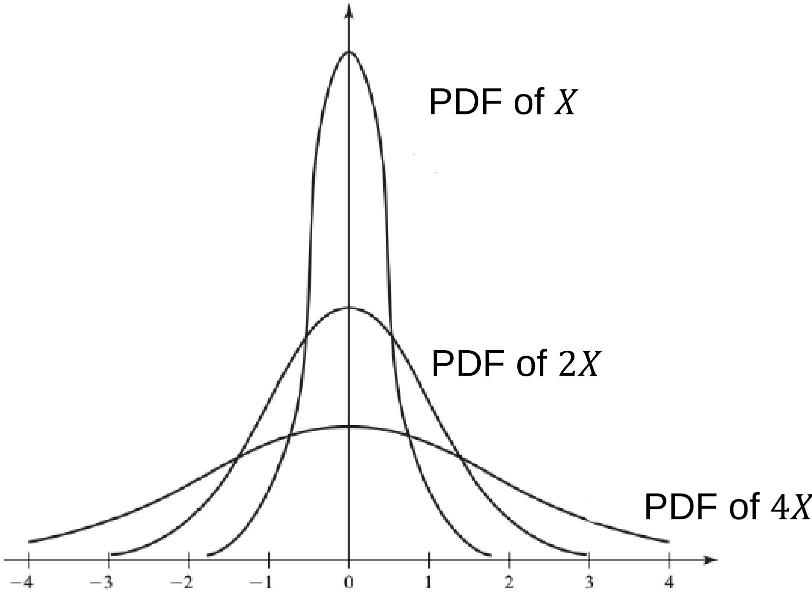
Properties of Mean & variance

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

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

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

- How about $\mathrm{E}[X + b]$ and $\mathrm{Var}[X + b]$?

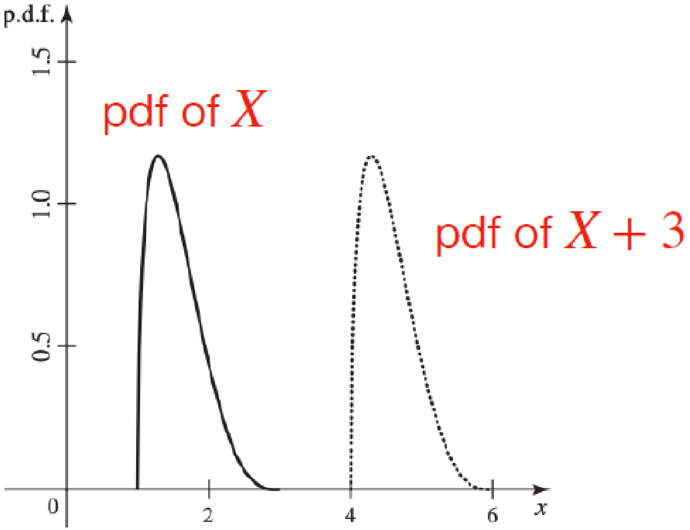


Properties of Mean & variance

How about $\mathrm{E}[X+b]$ and $\mathrm{Var}[X+b]$?

Fact

- $\mathrm{E}[X+b] = \mathrm{E}[X] + b$
- $\mathrm{Var}(X+b) = \mathrm{Var}(X)$



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:

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

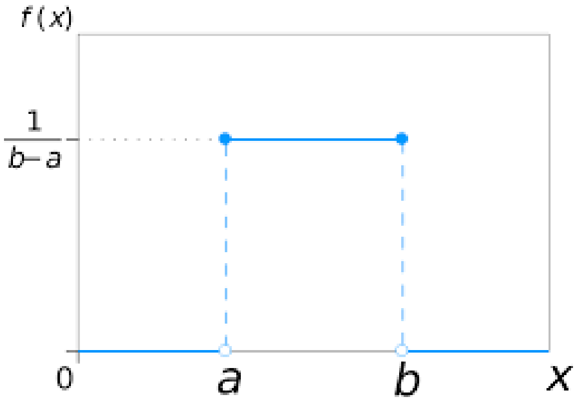
- $\mathrm{Var}[aX + b] = \mathrm{Var}[aX] = a^2 \cdot \mathrm{Var}[X]$

Useful Continuous Probability Distributions

Continuous Uniform Distribution

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

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



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

$$\text{Var}[X] = \frac{(b-a)^2}{12}$$

Exponential Distribution

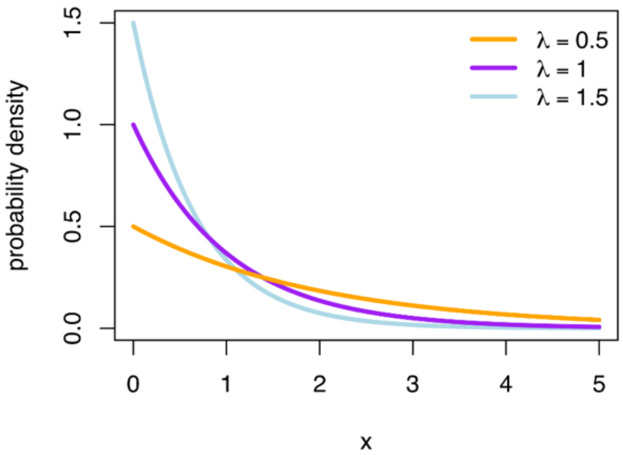
- Denoted as $X \sim \mathrm{Exp}(\lambda)$

- $f(x) = \lambda e^{-\lambda x}, x > 0$

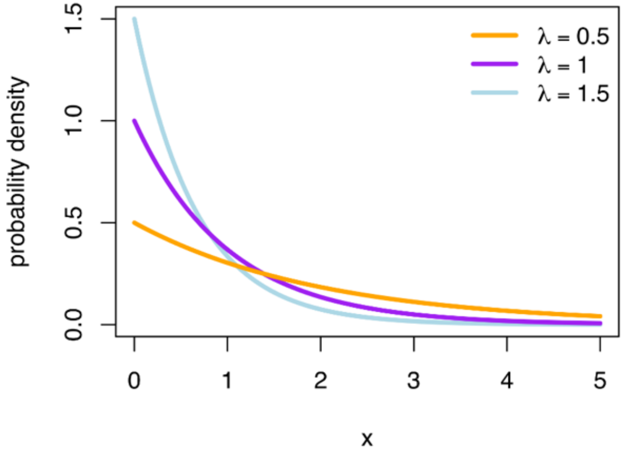
- λ: scale parameter

Examples:

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



Exponential Distribution



- $X \sim \mathrm{Exp}(\lambda)$

- Exponential distribution is the continuous analogue of geometric distribution!

- $\mathrm{E}[X] = \frac{1}{\lambda}$ (average life of lightbulb / particle)
- $\mathrm{Var}[X] = \left(\frac{1}{\lambda}\right)^2$

Exponential Distribution

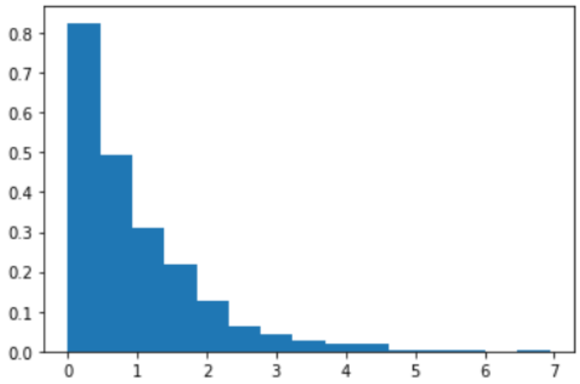
numpy.random.exponential

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

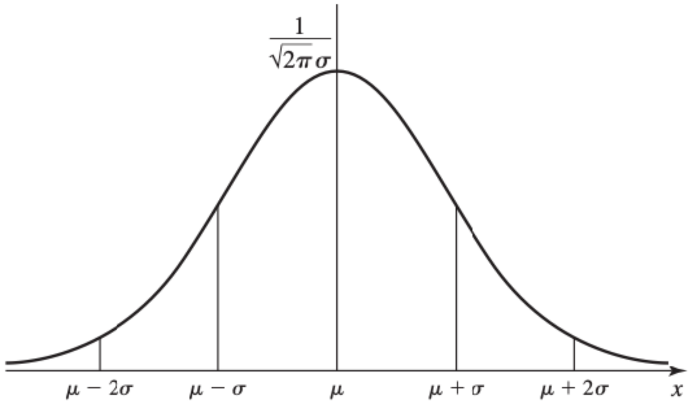
Gaussian (a.k.a. Normal) distribution with location μ and scale σ^2 parameters,

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

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

Perhaps *the most important* distribution in prob & stats

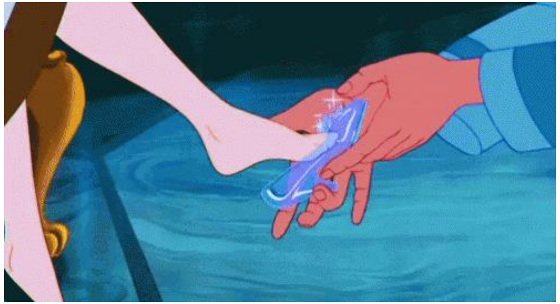
Does the shape of the curve ring a bell?

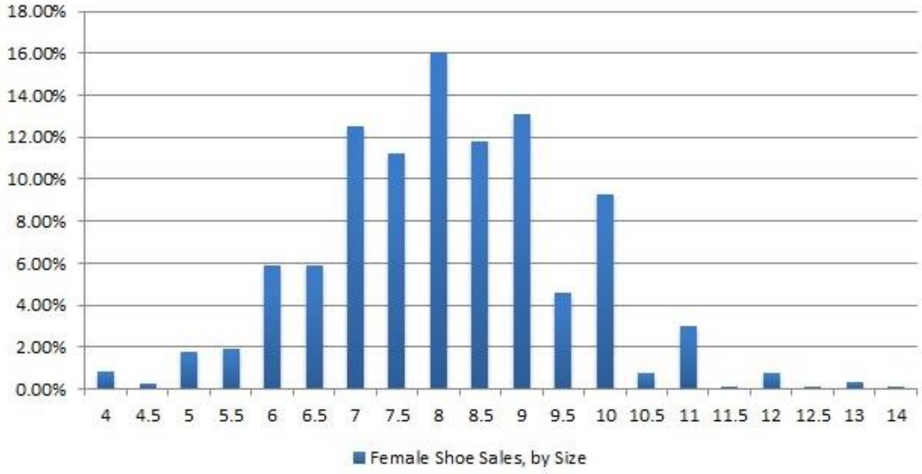


Similar to binomial distribution!

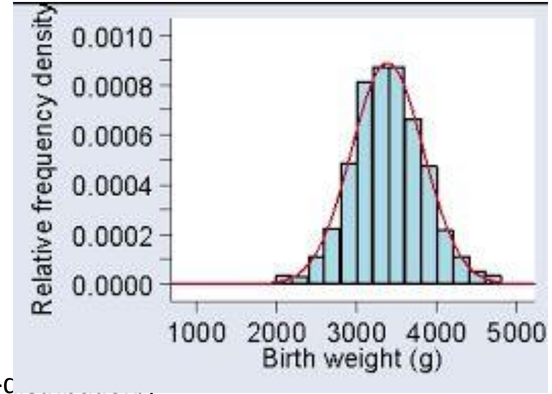
Distributions that follow Gaussian

Shoe size





Birth Weight



Q: Do they actually follow exact Gaussians?

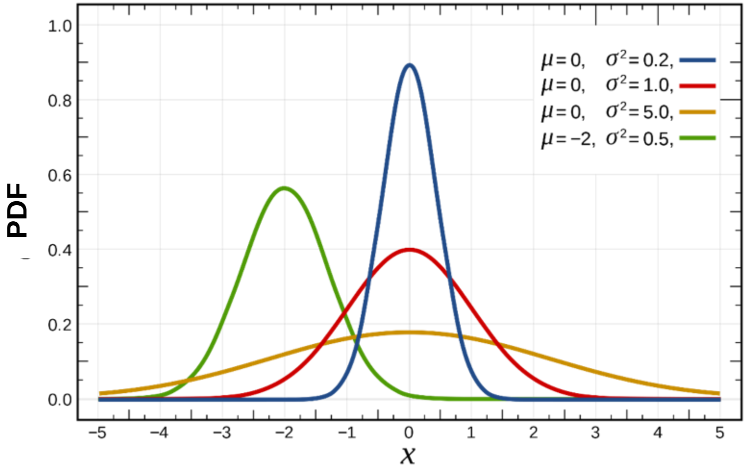
No exactly, but very close

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

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

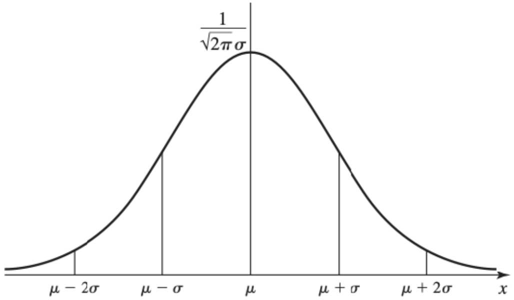
Gaussian Distribution

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?

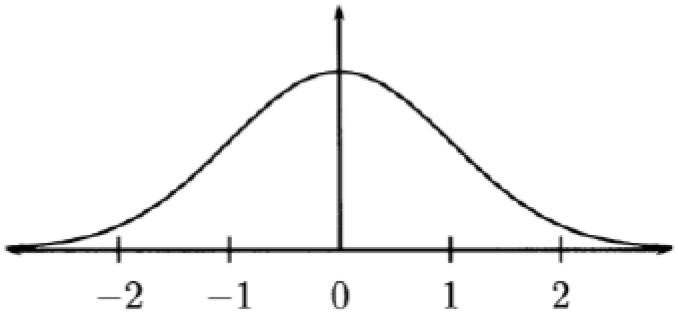
- $\mathrm{E}[Y] = \mathrm{E}[aX + b] = a\mu + b$
- $\mathrm{Var}[Y] = \mathrm{Var}[aX + b] = \mathrm{Var}[aX] = a^2\sigma^2$



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

The standard Gaussian distribution

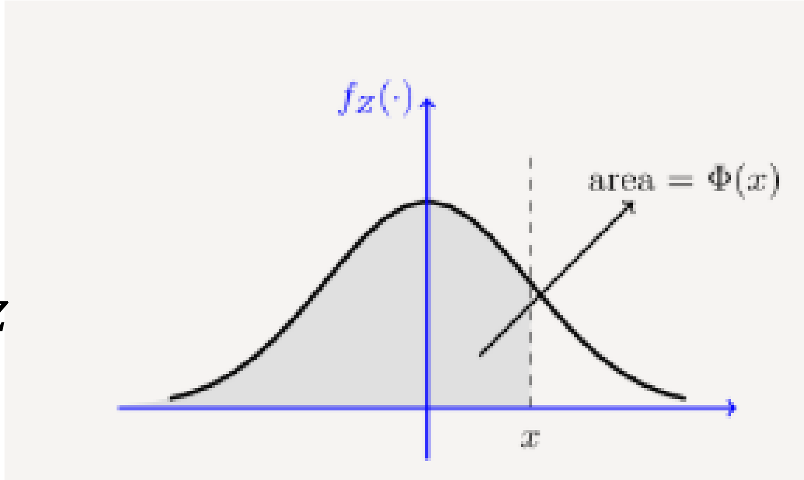
- Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$



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

The standard Gaussian distribution

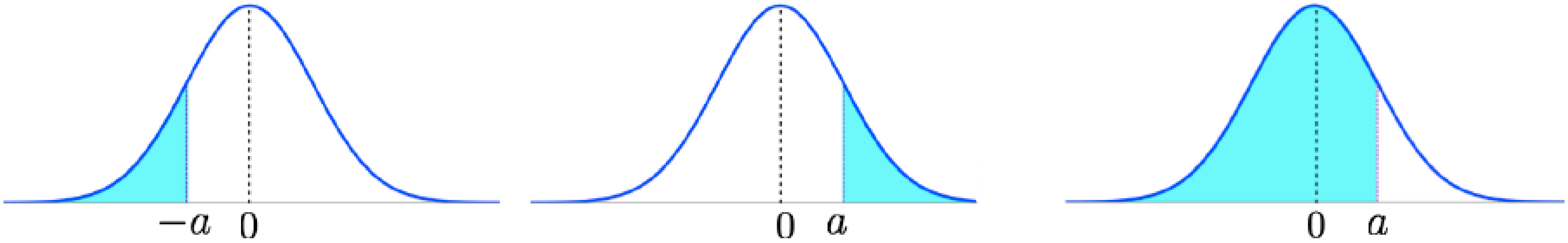
- PDF: $\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$
- CDF: $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)\, dz$



- Φ does not have a closed form, but it is a very important function
    - We can find the value of Φ by calling scipy.stats.norm.cdf

Calculating probabilities about Gaussians

- Symmetry of ϕ => $\Phi(-a) = 1 - \Phi(a)$



$\Phi(-a) = P(Z \le -a)$ $\quad = P(Z \ge a)$ $\quad = 1 - P(Z \le a) = 1 - \Phi(a)$

Calculating probabilities about Gaussians

- Suppose $X \sim N(5, 2^2)$, how can I calculate $P(1 < X < 8)$?

- From normal to standard normal
  - $X \sim N(\mu, \sigma^2)$

  $\Rightarrow X - \mu \sim N(0, \sigma^2)$

  $\Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

- We can write $\mathrm{P}(a < X < b)$ using $\mathrm{P}(c < Z < d)$, which in turn can be written in Φ. Here is how..

Calculating probabilities about Gaussians

- $\mathrm{P}(a < X < b)$

$$= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

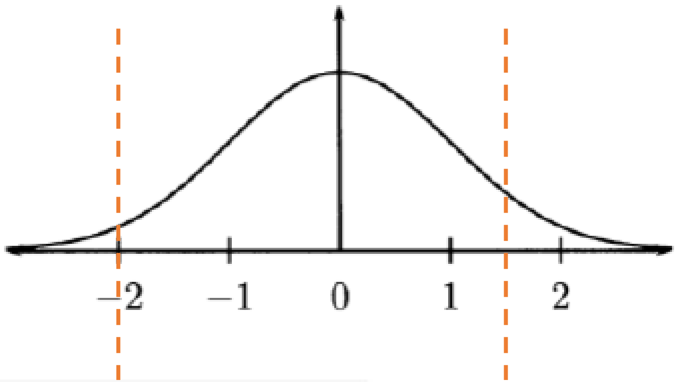
$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Example Suppose $X \sim N(5, 2^2)$, calculate $P(1 < X < 8)$

This is $\Phi\left(\frac{8-5}{2}\right) - \Phi\left(\frac{1-5}{2}\right) = \Phi(1.5) - \Phi(-2)$

Calculating probabilities about Gaussians

- $\Phi(1.5) - \Phi(-2)$
  $= \Phi(1.5) - (1 - \Phi(2))$

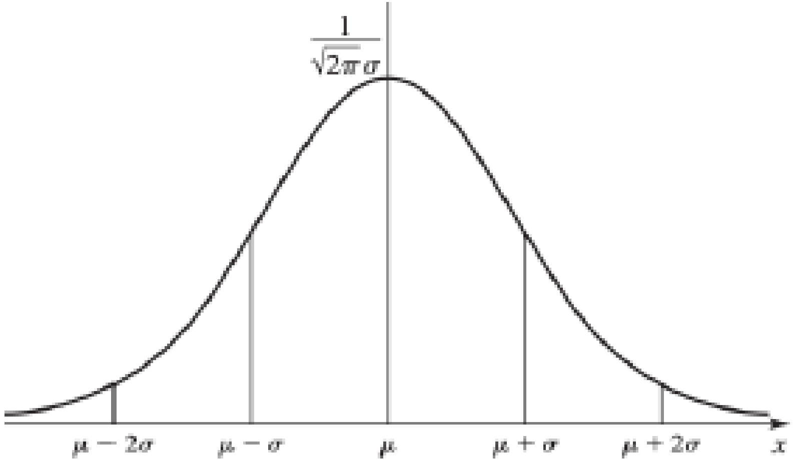



```
from scipy.stats import norm
print(norm.cdf(1.5)-(1-norm.cdf(2)))
```

```
0.9104426667829627
```

Calculating probabilities about Gaussians

- What is the probability that a Gaussian RV X is within 1 std of its mean? What about 2, 3?



- $\mathrm{P}(\mu - k\sigma \leq X < \mu + k\sigma)$

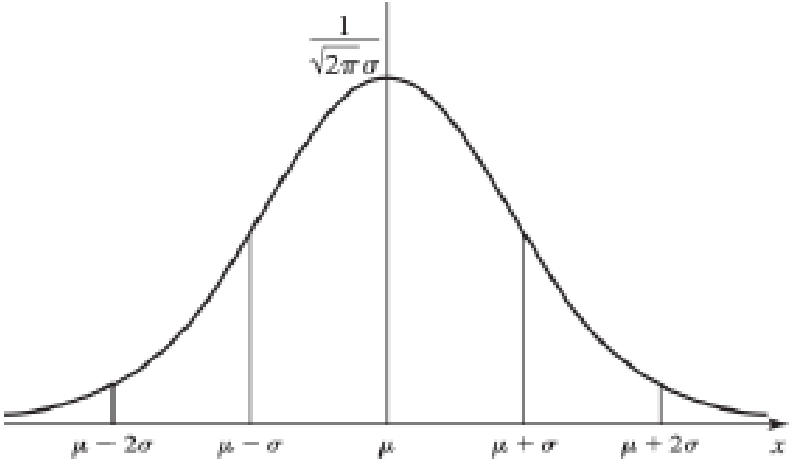
Calculating probabilities about Gaussians

- $p_k = \mathrm{P}(\mu - k\sigma \leq X < \mu + k\sigma)$

$= P(-k < Z < k)$

$= 2\Phi(k) - 1$

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994



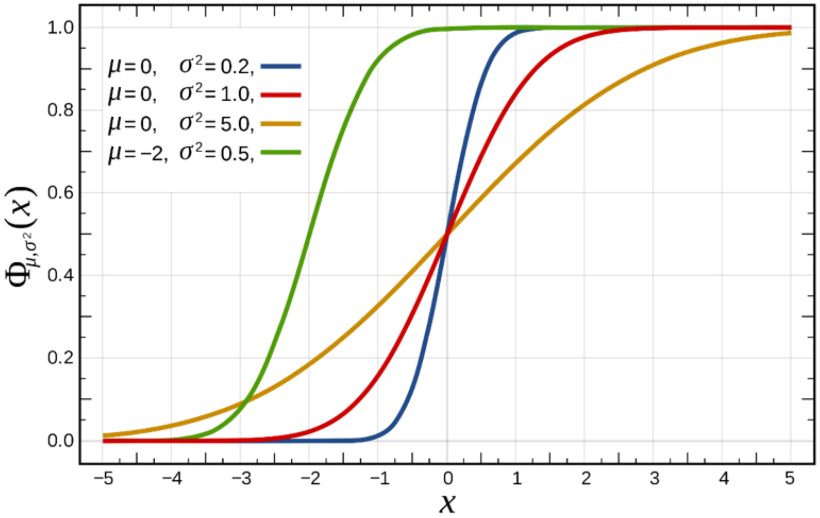
In words,

- With probability about 95%, X is within 2 std of its mean
- With overwhelming prob. (99.7%), X within 3 std of mean

CDF of Gaussian Distributions

- F: CDF of Gaussian $N(\mu, \sigma^2)$

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



- $F(x)$ changes fast when x starts to move away from μ
- F's "sensitive range" is about $[\mu - 3\sigma, \mu + 3\sigma]$

Backup