Math review

Machine Learning Summer Course 2020

Krishnakant Saboo

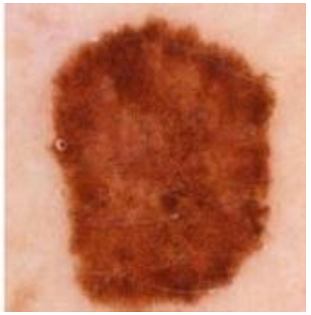
13th June 2020

Announcements

- We'll cover basic concepts from probability that will be useful to us
- Feedback form link: https://forms.gle/UxDTQgRMywszk37V9
- Assignment 1
  - Math problems
  - Coding environment setup and Python basics
- Office hours: Wednesday 6/17, 2000-2200 hrs CDT

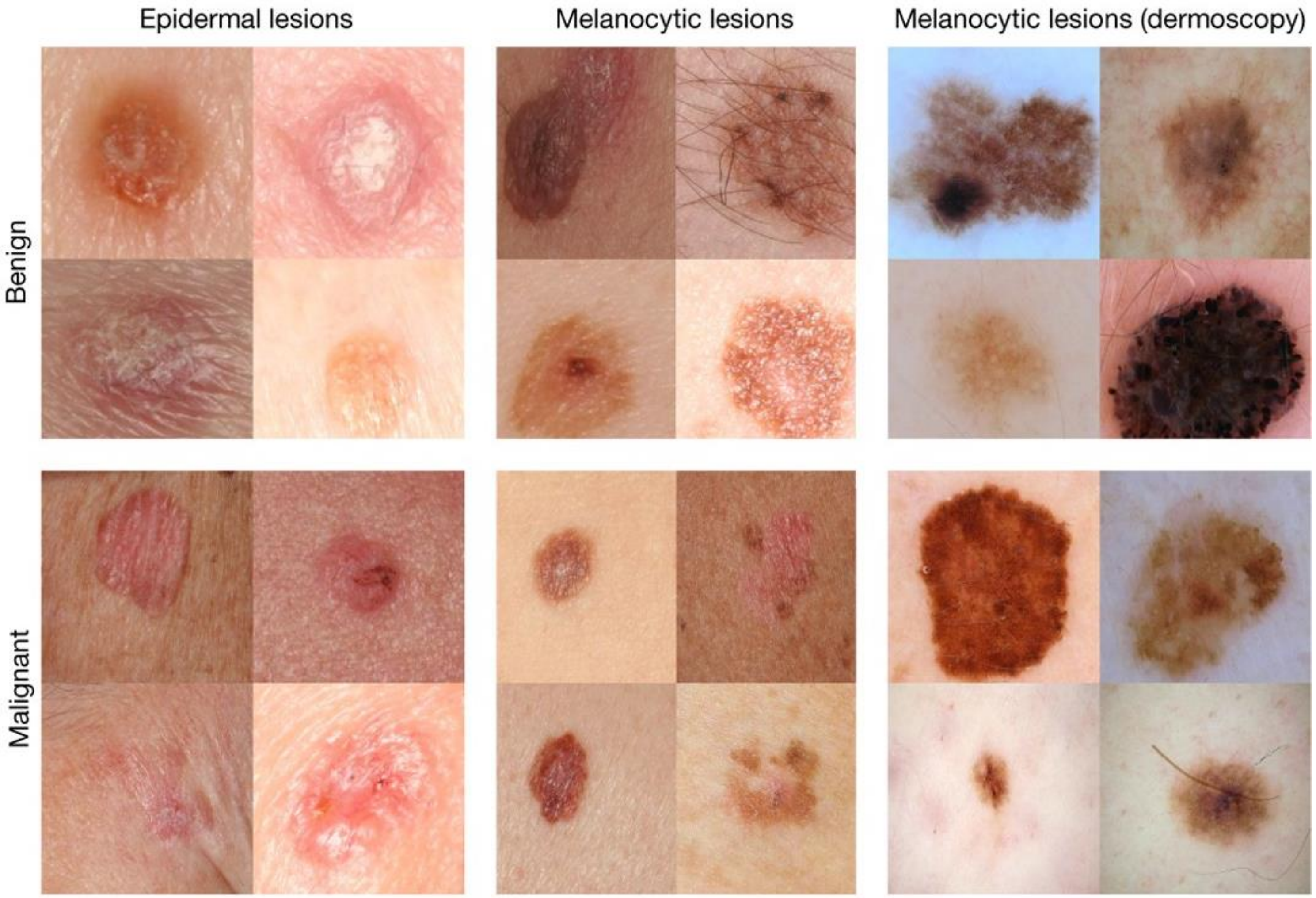
Machine Learning is everywhere around us!

ML for skin cancer detection



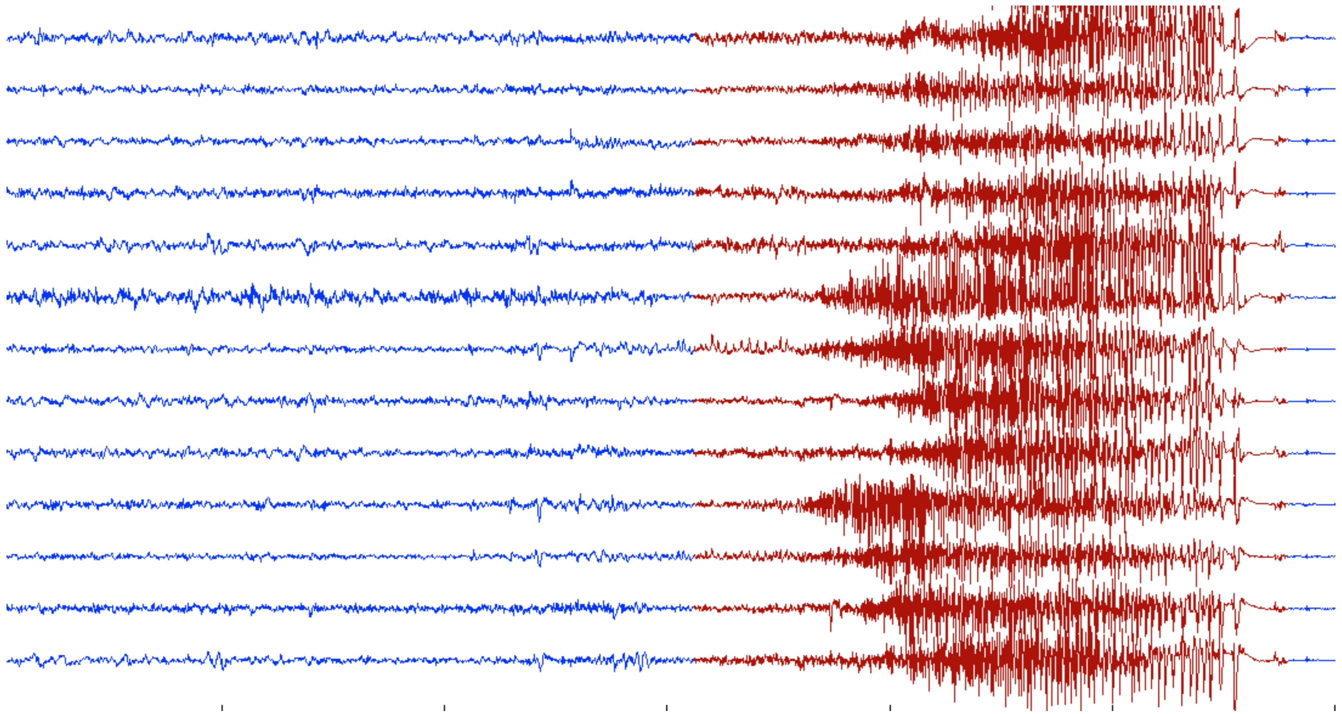
ML model

Benign/malignant

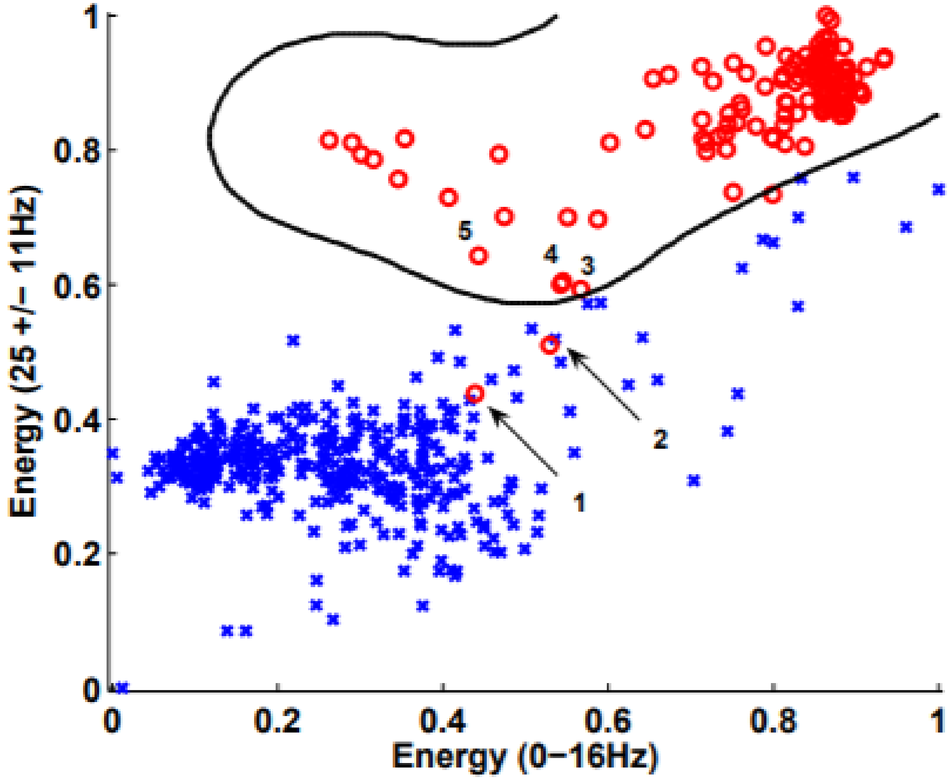


Source: https://www.nature.com/articles/nature21056

ML for seizure detection



EEG signal



Feature space

EEG signal → Feature extraction → ML model → Seizure begun yes/no

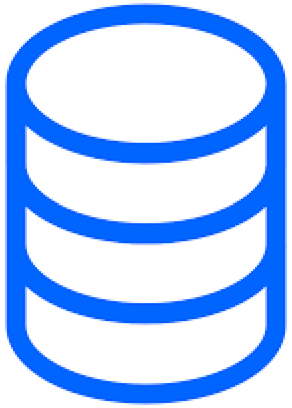
Source: https://icml.cc/Conferences/2010/papers/493.pdf

What is machine learning?

The field of machine learning is concerned with the automatic discovery of regularities in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data into different categories. [Bishop 2006]

Why now?

Computational resources



Large amounts of data

PyTorch

hadoop

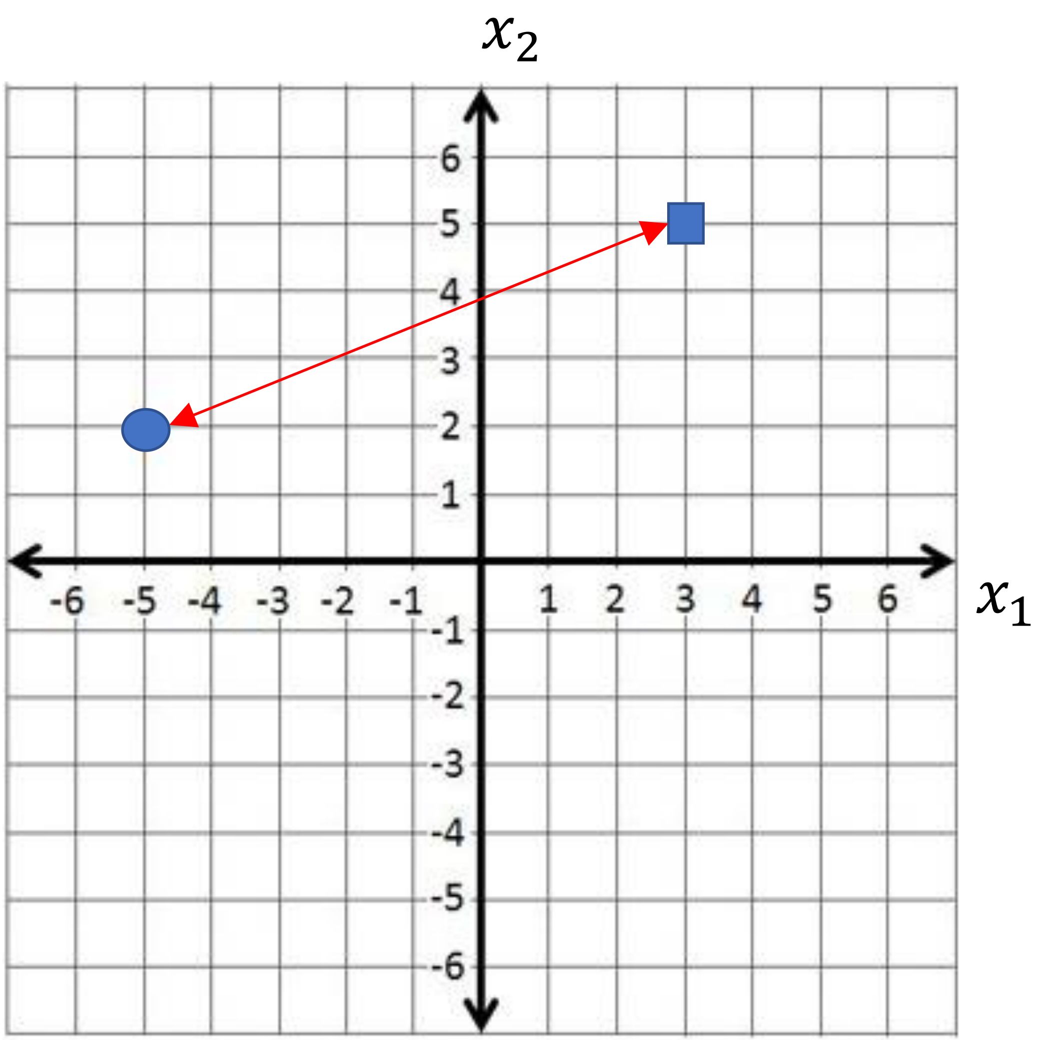
Platforms

Coordinate geometry

Points, dimensions, vectors

- Each axis corresponds to a ***dimension***
- Point is represented by an ordered pair $(\mathrm{x}_1, \mathrm{x}_2)$
- Distance between two points $a = (a_1, a_2)$ and $b = (b_1, b_2)$

$$D(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$



There can be more than 2 dimensions…

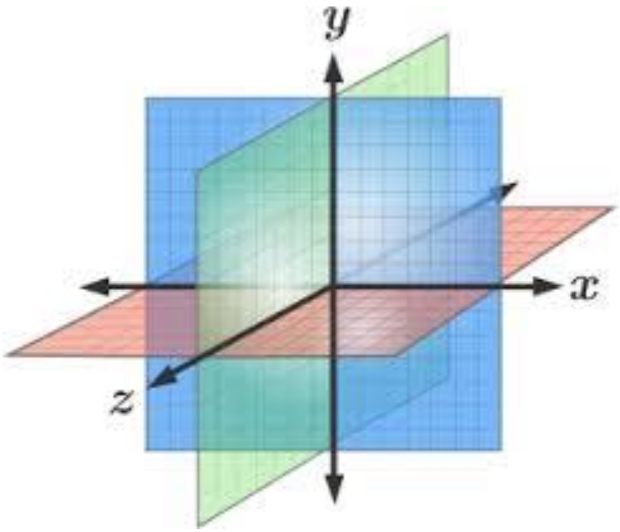
- Example: Dog=(weight, height, speed)
  - Dog1 = (25, 40, 5)
  - Dog2 = (15, 35, 8)
  - How different are Dog1 and Dog2?
- Distance between two points $a = (a_x, a_y, a_z)$ and $b = (b_x, b_y, b_z)$ is

$$D(a,b) = \sqrt{(a_x - b_x)^2 + \left(a_y - b_y\right)^2 + (a_z - b_z)^2}$$

- For a n dimensional space, $a = (a_1, \ldots, a_n)$, $b = (b_1, \ldots, b_n)$

$$D(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}$$

A tale of two dogs



3-dimensional space

Image source: https://brilliant.org/wiki/3d-coordinate-geometry-equation-of-a-plane/

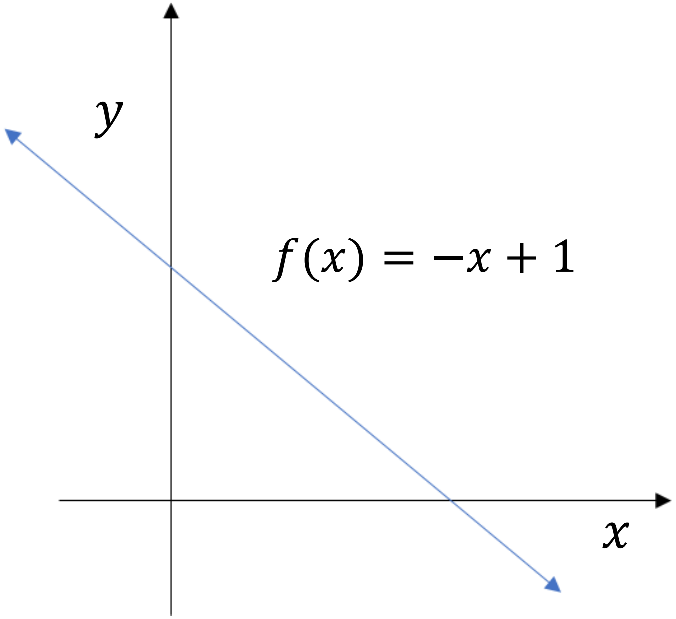
Function

Definition

Functions represent relations between different variables

Example:

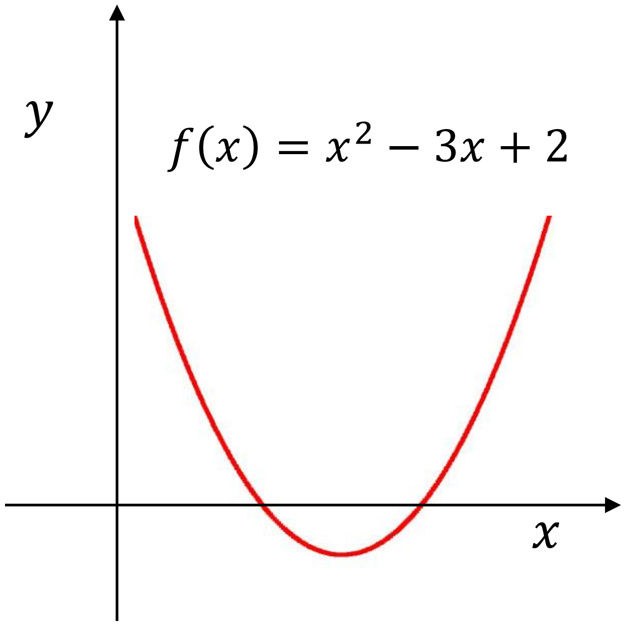
$$y = f(x) = -x + 1$$



Maps a value in the domain to the range

$$f\colon \mathbb{R} \to \mathbb{R}$$

Domain Range



Example: The number of COVID-19 cases (p) depends on the number of days (t) since first diagnoses

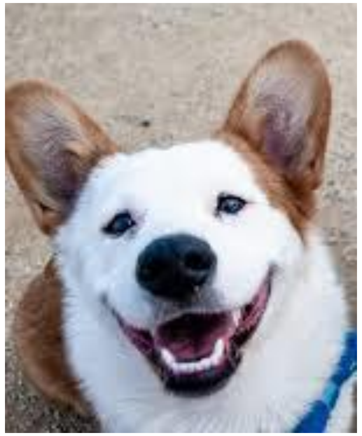
$$p = f(t) = 1.07^t$$

At this rate, the number of cases double every 10 days

Functions can take multiple variables too…

$$f: \mathbb{R}^N \to \mathbb{R}^M$$

Example: The amount of food (gms) and water (ml) the dogs need depends on their weight (in lbs), height (cms), and how fast they run (kmph). The relation is given as follows:



$$food = 3 * weight + 2 * height + 6 * speed$$

$$water = weight * height + 4 * speed$$

How much food do the dogs Dog1 and Dog2 need?

(Dog1) food = 3x25 + 2x40 + 6x5 = 185

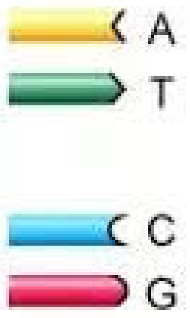
(Dog1) water = 25x40 + 4x5 = 1020

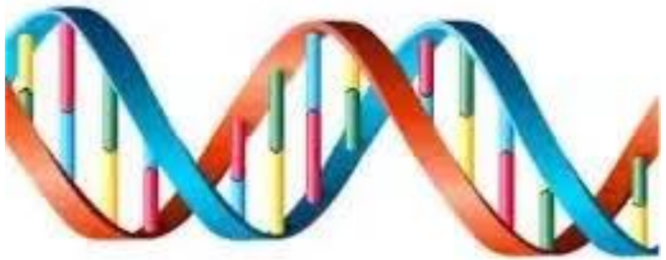
Probability

… and then there was life!

A biology grad student has been assigned the job of creating a new organism. His task is simple. He should create a DNA sequence using nucleotides A, C, T, G. He has a bag with many copies of each base; approximately the same number of copies of each base. The way he proceeds with his job is a follows:

1. He puts his hand in the bag and randomly picks a base (without seeing what he is picking)
2. He repeats the above procedure irrespective of the outcome of the previous experiment until he has picked out 4 bases.





Computing probability

1. What is the ***chance*** of the choosing A as the first base?

Since the quantity of each base (A,C,T,G) in the bag is the same, the chance of him picking an A is the same as that of picking a C ….

So, it is equally likely to pick any of the four bases. If we were talking about the chance (in %) of choosing a base, then the chance of picking A is 25%.

2. What is the ***probability*** of the choosing A as the first base?

The probability of picking A is 0.25.

Some definitions…

What is the ***probability*** of the choosing A as the first base?

- ***Random experiment*** is an experiment the outcome of which is not certain
  - For e.g., picking the first base
- ***Sample Space (S)*** is the totality of the possible outcomes of a random experiment
  - For e.g., {A,C,T,G}
- An ***event*** is a collection of certain sample points, i.e., a subset of the sample space
  - For e.g., {A}

Notation

What is the ***probability*** of the choosing A as the first base?

We denote a variable representing the first base as B_1. The probability of the first base being A is denoted as

$$P(B_1 = A)$$

From the previous analysis, we arrived at the answer $P(B_1 = A) = 0.25$

Axioms of probability

Let *S* be a sample space of a random experiment and *P(E)* be the probability of the event *E.* The probability function *P(.)* must satisfy the three following axioms:

- **(A1)** For any event E, $P(E) \geq 0$

  (probabilities are nonnegative real numbers)
- **(A2)** $P(S) = 1$

  (probability of a certain event, an event that must happen is equal 1)
- **(A3)** $P(E \cup F) = P(E) + P(F)$, whenever E and F are mutually exclusive events, i.e., $E \cap F = \phi$

  (probability function must be additive for mutually exclusive events)

Probability computation

What is the ***probability*** of the choosing A as the first base?

Recall that B_1: variable representing the first base
$P(B_1 = A)$: Probability of first base A
$P(B_1 = C)$: Probability of first base C
$P(B_1 = G)$: Probability of first base G
$P(B_1 = T)$: Probability of first base T

Event $B_1 = A$ and $B_1 = C$ are mutually exclusive. Same for other bases.
From A2 and A3: $P(B_1 = A) + P(B_1 = C) + P(B_1 = G) + P(B_1 = T) = 1$
Given that $P(B_1 = A) = P(B_1 = C) = P(B_1 = G) = P(B_1 = T)$

Solving the above equations gives us $P(B_1 = A) = 0.25$

Another way to arrive at the answer: Frequentist

What is the ***probability*** of the choosing A as the first base?

There are 4 possible options for the first base

Out of that, A happens in one

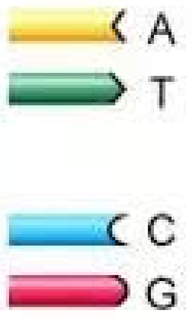
$$P(B_1 = A) = \frac{\#(B_1 = A)}{\#(B_1)} = \frac{1}{4}$$

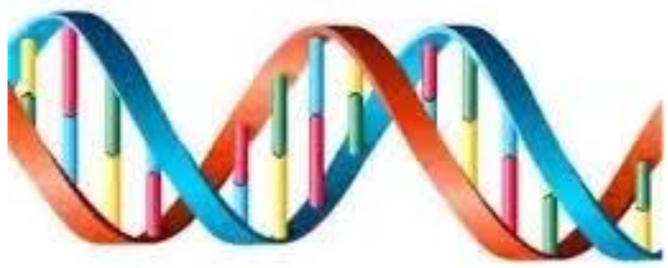
Counting expert

… and then there was life!

A biology grad student has been assigned the job of creating a new organism. His task is simple. He should create a DNA sequence using nucleotides A, C, T, G. He has a bag with many copies of each base; approximately the same number of copies of each base. The way he proceeds with his job is a follows:

1. He puts his hand in the bag and randomly picks a base (without seeing what he is picking)
2. He repeats the above procedure irrespective of the outcome of the previous experiment until he has picked out 4 bases.





Probability of a sequence

What is the probability of observing a sequence AA?

Let's try the frequentist way

$$P(B_1 = A, B_2 = A) = \frac{1}{16}$$

Probability of a sequence

What is the probability of observing a sequence AA?

Notice that:

$P(B_1 = A) = \frac{1}{4}, P(B_2 = A) = \frac{1}{4}$

$$P(B_1 = A)P(B_2 = A) = \frac{1}{16}$$

Could:

$P(B_1 = A, B_2 = A) = P(B_1 = A)P(B_2 = A)$?

Independence

If two event E and F are independent, then

$$P(E \ \cap F) = P(E)P(F)$$

If multiple event $E_1, E_2, \dots, E_n$ are independent, then

$$P(E_1 \cap E_2 \cap \cdots E_n) = P(E_1)P(E_2) \dots P(E_n)$$

Probability of a sequence

What is the probability of observing a sequence AAAA?

Let's try the frequentist way:

$$P(B_1 = A, B_2 = A, B_3 = A, B_4 = A) = \frac{1}{256}$$

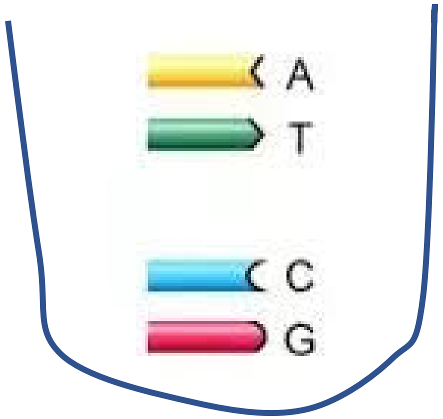
By independence:

$$P(B_1 = A, B_2 = A, B_3 = A, B_4 = A) = P(B_1 = A) \ldots P(B_4 = A) = \left(\frac{1}{4}\right)^4$$

But resources are limited…

The bag that the grad student has only 4 bases – one copy of each of A, C, T, G. The way he proceeds with his job is a follows:

1. He puts his hand in the bag and randomly picks a base (without seeing what he is picking)
2. He repeats the above procedure irrespective of the outcome of the previous experiment until he has picked out 4 bases.



First bases first…

What is the ***probability*** of the choosing A as the first base?

There are 4 possible options for the first base

Out of that, A happens in one

$$P(B_1 = A) = \frac{\#(B_1 = A)}{\#(B_1)} = \frac{1}{4}$$

Probability of impossible events

What is the probability of picking AA?

Since there is only one copy of A in the bag, it is impossible to create a sequence AA. So, $P(B_1 = A, B_2 = A) = 0$.

Conditional probability

What is the probability of observing AG i.e., $P(B_1 = A, B_2 = G)$?

List of possible 2 length sequences that can be picked from the bag:

{AC, AG, AT, CA, CG, CT, GA, GC, GT, TA, TC, TG} (12 possibilities)

Number of possibilities of interest = 1

Frequentist says: $P(B_1 = A, B_2 = G) = \frac{1}{12}$

Does multiplication rule apply always?

What is probability of observing G as the second base?

List of possible 2 length sequences that can be picked from the bag:

{AC, AG, AT, CA, CG, CT, GA, GC, GT, TA, TC, TG} (12 possibilities)

Number of possibilities of interest = 3

$$P(B_2 = G) = \frac{3}{12} = \frac{1}{4}$$

But, $P(B_1 = A)P(B_2 = G) = \frac{1}{4}.\frac{1}{4} = \frac{1}{16} \neq P(B_1 = A, B_2 = G)$

Observing the outcome of the first base changes things…

Once the first base (A) is removed from the bag, only 3 bases remain {G,C,T}. So, for the second experiment, there are only three equiprobable choices.

$$P(B_2 = G | B_1 = A) = \frac{1}{3}$$

Conditional probability

Conditional probability

- **Conditional Probability** of F given E, $P(F|E)$, defines the conditional probability of the event F given that the event E occurs and is given by:

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

if $P(E) \neq 0$ and ***is undefined otherwise.***

- A rearrangement of the above definition gives the following ***multiplication rule (MR):***

$$P(E,F) = \begin{cases} P(F)P(E|F), & if\ P(F) \neq 0 \\ P(E)P(F|E), & if\ P(E) \neq 0 \\ 0, & otherwise \end{cases}$$

Using conditional probability

What is the probability of observing AG i.e., $P(B_1 = A, B_2 = G)$?

$$P(B_1 = A, B_2 = G) = P(B_2 = G | B_1 = A) P(B_1 = A)$$

$$= \frac{1}{3} . \frac{1}{4} = \frac{1}{12}$$

Independence and conditional probability

$$P(B_1 = A, B_2 = G) = P(B_2 = G|B_1 = A)P(B_1 = A)$$

What happens if $P(B_2 = G|B_1 = A) = P(B_2 = G)$ i.e., the first experiment has no effect on the probability of the second event?

$$P(B_1 = A, B_2 = G) = P(B_1 = A)P(B_2 = G)$$

Independence: Events E and F are independent iff $P(F|E) = P(F)$

Conditional probability of interest could have other forms…

What is the probability that the first draw resulted in A if the second draw resulted in G?

Frequentist approach:

Cases when second draw is G = {AG, CG, TG}

$$P(B_1 = A | B_2 = G) = \frac{1}{3}$$

Bayes rule

What is the probability that the first draw resulted in A if the second base resulted in G?

We have already computed $P(B_1 = A), P(B_2 = G)$,
$P(B_2 = G|B_1 = A)$. Can we use these to make our life easier?

Bayes rules: $P(B_1 = A|B_2 = G) = \frac{P(B_2 = G|B_1 = A)P(B_1=A)}{P(B_2=G)} = \frac{\frac{1}{3} \cdot \frac{1}{4}}{\frac{1}{4}} = \frac{1}{3}$

Distributions

We need a convenient way of representing probabilities of various outcomes of an experiment.

- For e.g., first draw is A, result of coin toss, result of die roll, height of a dog..

A **probability distribution** is the mathematical function that gives the probabilities of occurrence of different possible **outcomes** for an experiment[1].

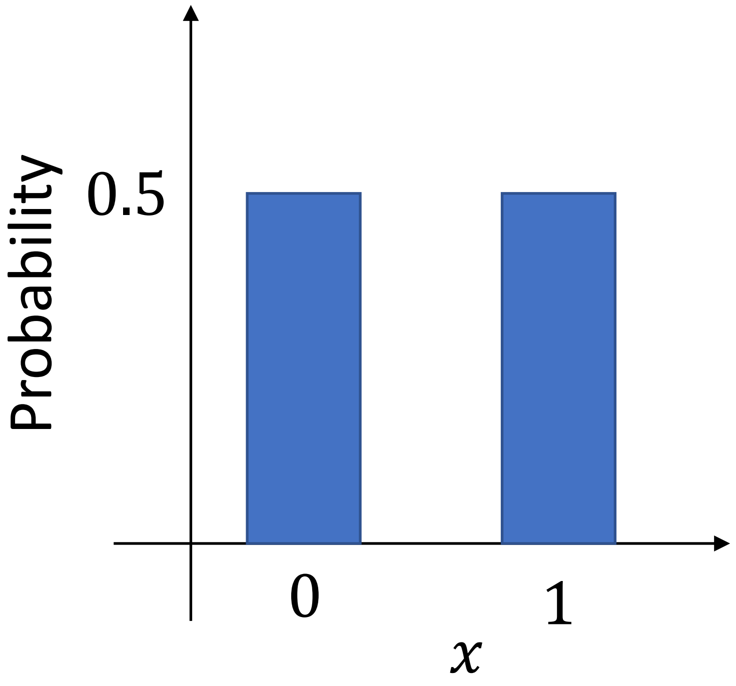
- First draw of grad student: $P(B_1 = A) = P(B_1 = C) = \cdots = \frac{1}{4}$
- Result of an unbiased coin toss: $P(Head) = P(Tail) = \frac{1}{2}$
- Result of an unbiased die roll: $P(1) = P(2) = \cdots P(6) = \frac{1}{6}$

[1] https://en.wikipedia.org/wiki/Probability_distribution

Bernoulli distribution

- Whenever the outcome is a yes/no – answer
  - Examples: diagnosis, coin toss, rain today
- Denote the outcomes by {0, 1}
- Probability out yes is given by p
  - $p = 0.5$ for unbiased coin
- Assume that random variable X denotes the outcome of a Bernoulli trial, then

$$P(X = 1) = p$$



Distribution of outcome for a fair coin toss.

Geometric distribution

- Image you are tossing a coin. You keep tossing it until you get H. You are interested in the number of trials it took to get a H.
- How many trials can there be? Denote p as probability of H and n as the trial at which H occurred
  - $n = 1, 2, 3, \dots$
- $P(X = n) = (1 - p)^{n-1} p$

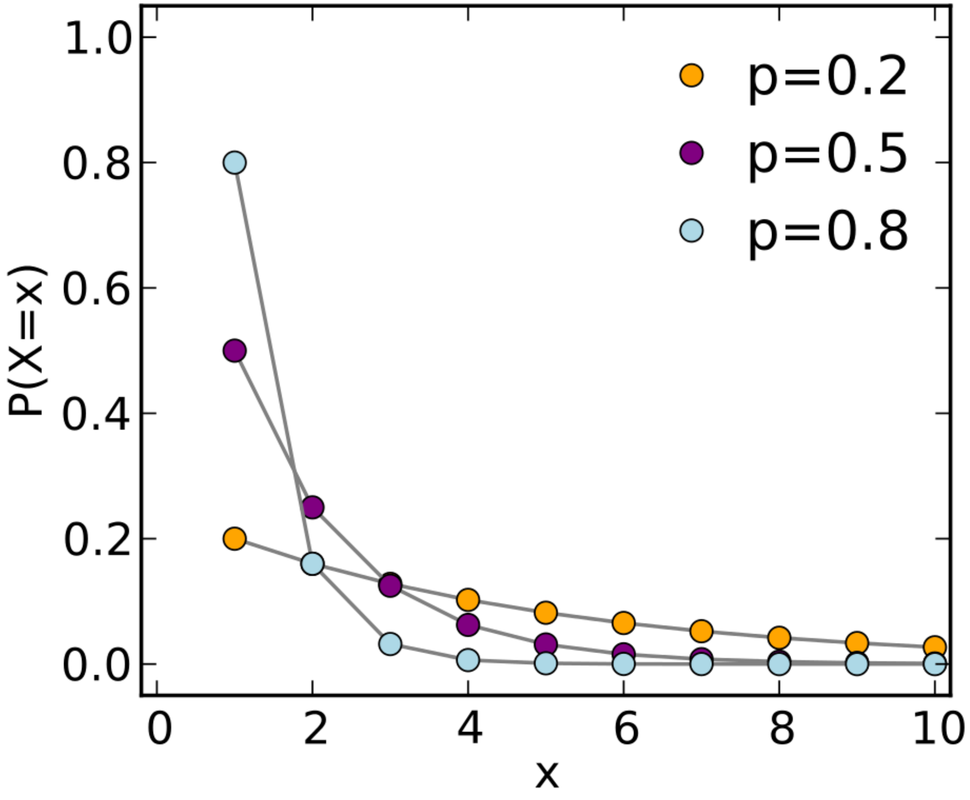


Image source: https://en.wikipedia.org/wiki/Geometric_distribution

PMF, CMF

- **Probability Mass Function (PMF)**

$$p(a) = P\{X = a\}$$

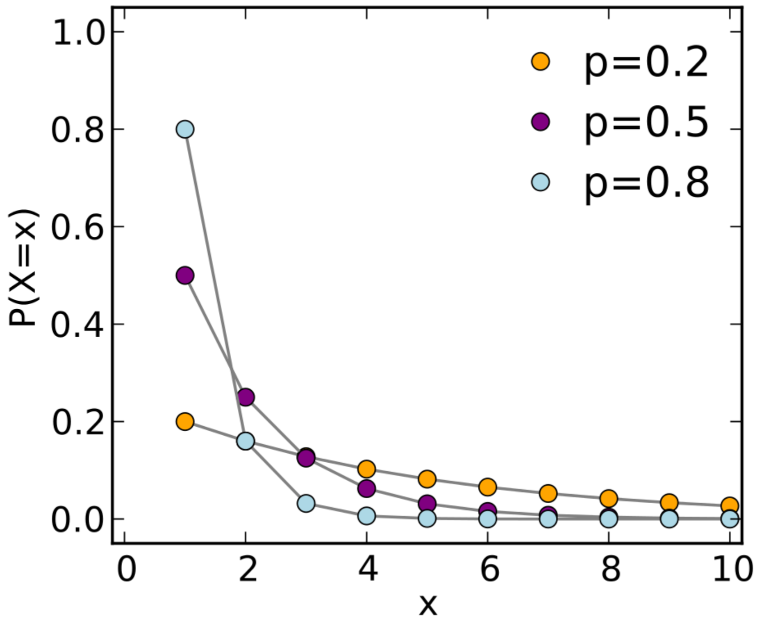
$$p(x) = \begin{cases} > 0, & x = x_1, x_2, \ldots \\ 0, & \text{for other values of } x \end{cases}$$

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

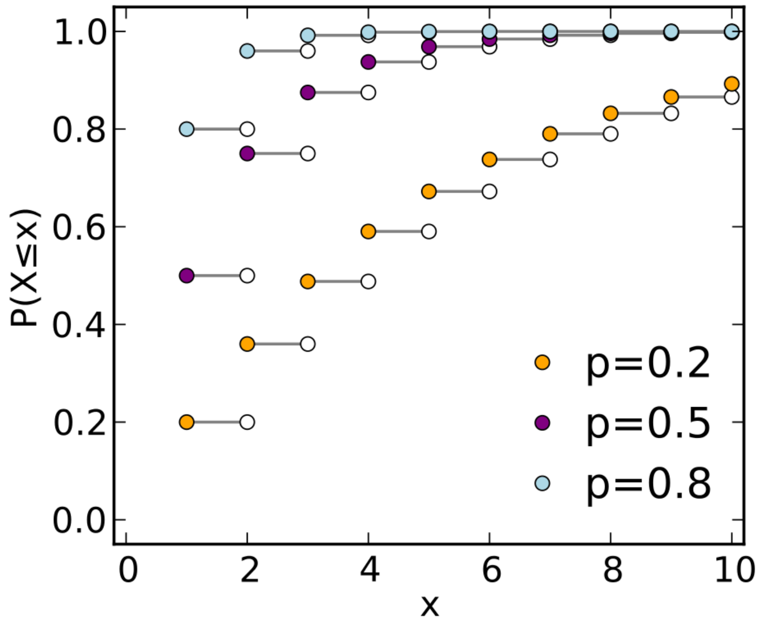
- **Cumulative Mass Function (CMF)**

$$F(a) = \sum_{all\ x_i \le a} p(x_i)$$

- Non-decreasing



PMF



CMF

Discrete and continuous random variables/distributions

- The distributions and examples we studied so far were dealing with **discrete variables**
- Many phenomena of interest have continuous values
  - Example: weight of the dog, length of a petal, amount of amyloid protein in the brain
- Such variables are **continuous variables** and characterized by a probability density function (**PDF**) and cumulative density function (**CDF**)

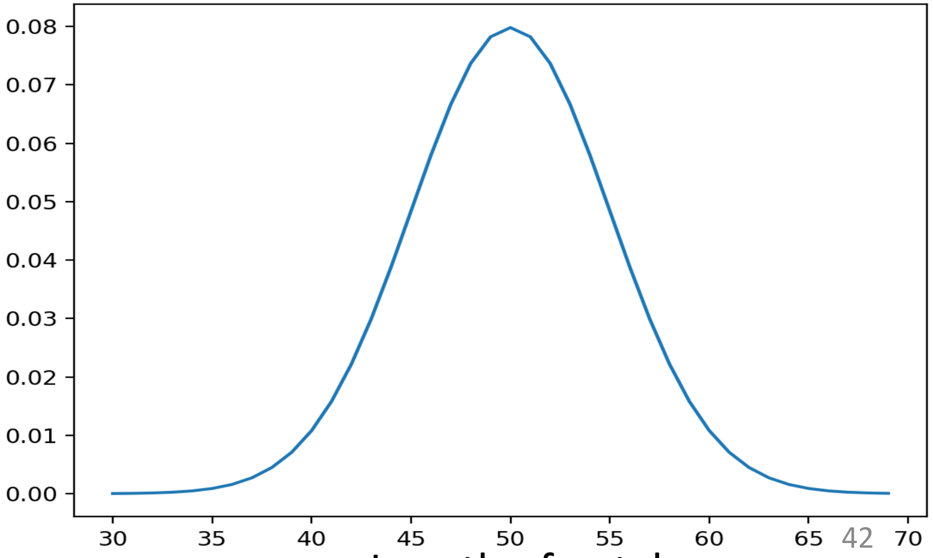
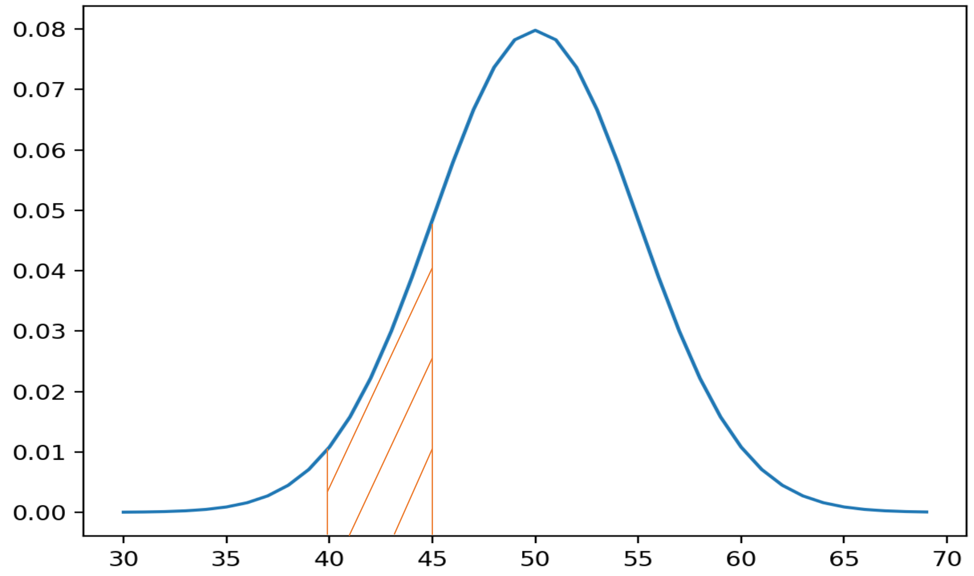


Image source: https://gilmour.com/jasmine-care

PDF, CDF

- What is the probability that the length of a petal is 45mm?
  - 0; it is a continuous variable, so probability of any specific value is zero
- What is the probability that the length of a petal is between 40 and 45 mm?
  - Compute the area under the PDF between 40 and 45 mm



PDF

PDF, CDF

Probability density function $\boldsymbol{f(x)}$

- *Relative likelihood* that the value of the random variable would equal that sample x

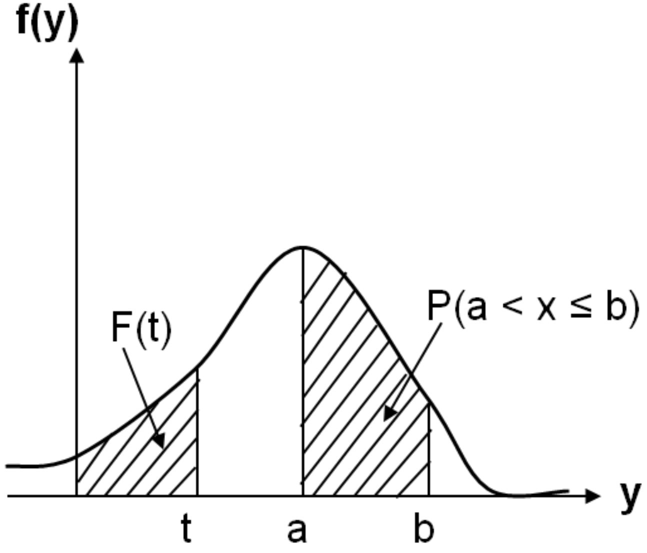
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$$

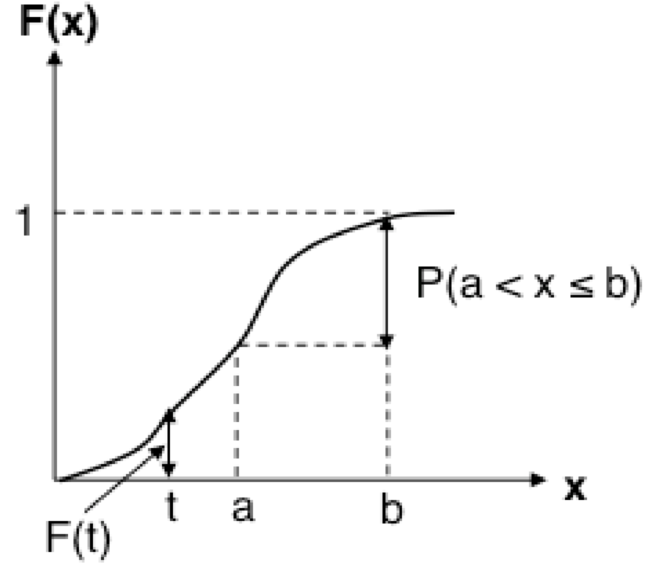
Cumulative density function $\boldsymbol{F(x)}$

- Probability that the random variable would be less than or equal to x

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



pdf



CDF

Gaussian distribution

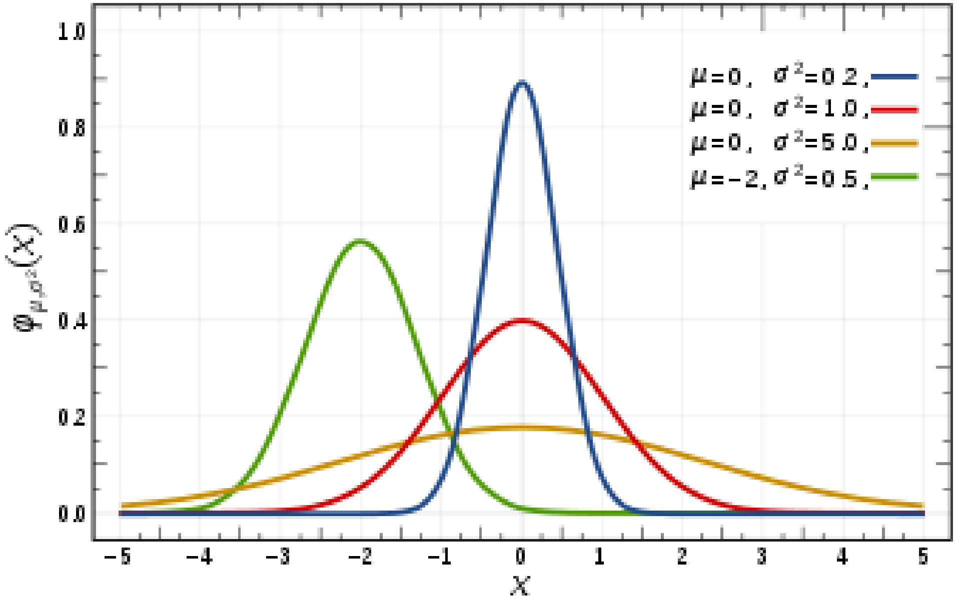
- Popular and important distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

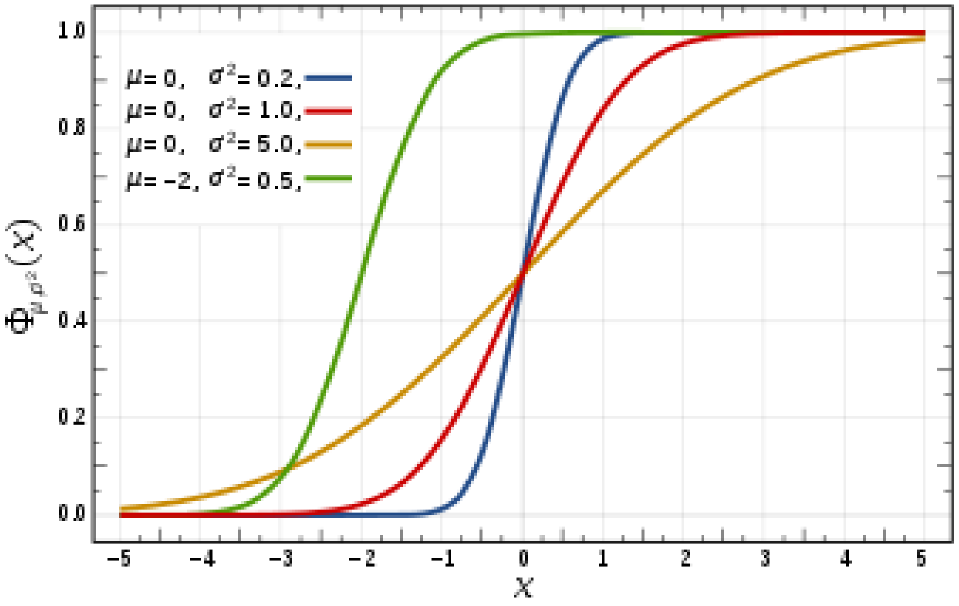
- Given by two parameters

Mean: μ

Variance: σ^2



PDF



CDF

Questions?