CS 107, Probability, Spring 2020
Lecture 27

Michael Poghosyan
mpoghosyan@aua.am

AUA

08 April 2020

Content

- Some important Continuous rv.s
    - The Normal Distribution

The Normal Distribution

Normal Distribution

Distribution Name: $\mathcal{N}$;
Parameters: μ, σ^2 (or, sometimes, μ, σ, $\sigma > 0$)

The Normal (Gaussian) Distribution

We say that the r.v. X has a Normal (or Gaussian) Distribution with the Mean μ and the Standard Deviation σ (or, with the Mean μ and Variance σ^2), and we will write

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

if its PDF is given by

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Standard Normal Distribution

Important Case: If X is a Normal r.v. with the Mean 0 and Variance 1, i.e.

$$X \sim \mathcal{N}(0,1),$$

then we call X to be a **Standard Normal r.v.**
The PDF of the Standard Normal r.v. is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}.$$

The CDF of the Standard Normal r.v. is given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt, \qquad x \in \mathbb{R}.$$

Exercise: Prove that $\varphi(x)$ is a PDF of some r.v.

Standard Normal Distribution

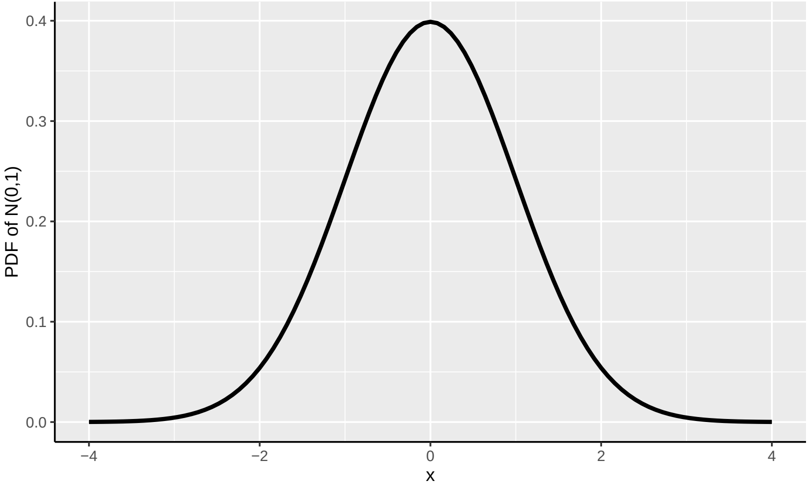

Figure: The PDF of $\mathcal{N}(0, 1)$

Standard Normal Distribution

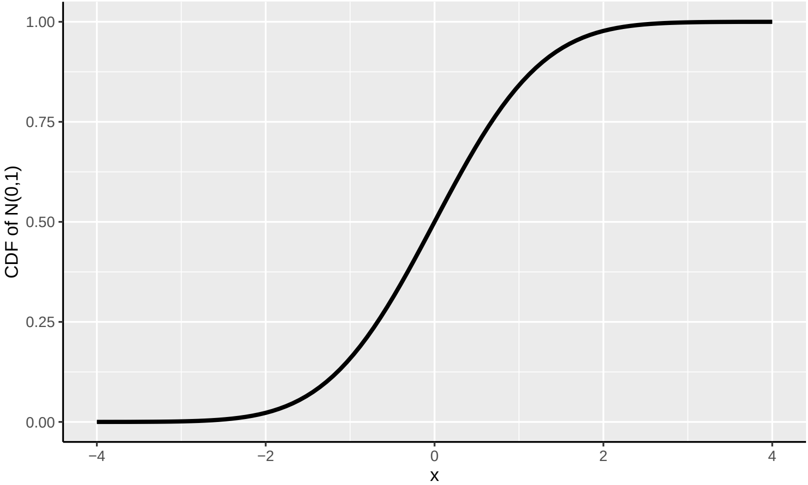

Figure: The CDF of $\mathcal{N}(0,1)$

Normal Distribution: Note

Note: In different texts, say, on ML, the PDF of $X \sim \mathcal{N}(\mu, \sigma^2)$ is denoted by

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Note: The other name of the Normal Distribution is the Gaussian Distribution.

Note: Again, in different texts, also in most of the software, the parameters of the Normal distribution are the Mean and the Standard Deviation, μ and σ, i.e., people are using the $\mathcal{N}(\mu, \sigma)$ form for the Normal Distribution with Mean μ and SD σ. But we will use $\mathcal{N}(\mu, \sigma^2)$ notation in our lectures: so when we write $X \sim \mathcal{N}(3, 4)$, this means that X is a Normal r.v. with the Mean $\mu = 3$, and the Variance $\sigma^2 = 4$, i.e., the Standard Deviation is $\sigma = 2$.

Normal Distribution: Example

Example 27.1: Assume $X \sim \mathcal{N}(-4, 4)$.

a. Find the Mean of X;
b. Find the Variance of X;
c. Find the Standard Deviation of X.

Parameters of the Normal PDFs, graphically

Now, about the parameters of $\mathcal{N}(\mu, \sigma^2)$ on its PDF graph:

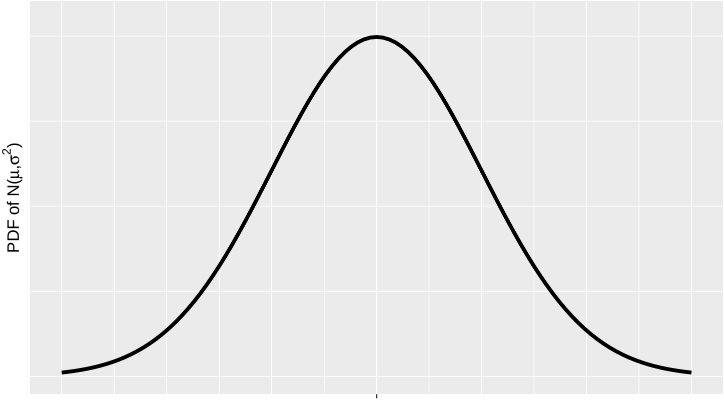

Figure: The PDF

Parameters of the Normal PDFs, graphically

Now, about the parameters of $\mathcal{N}(\mu, \sigma^2)$ on its PDF graph:

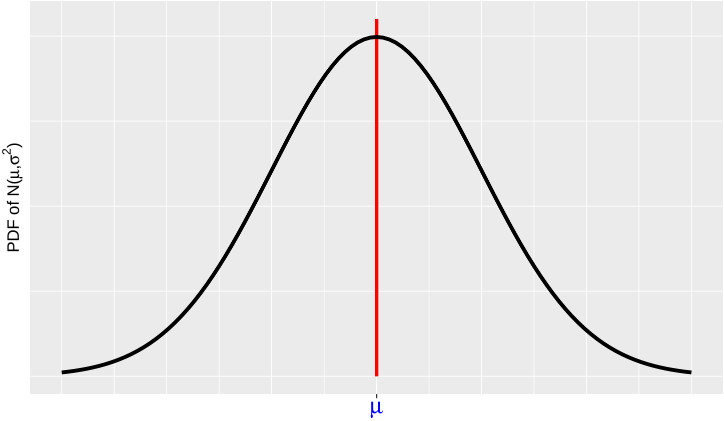

Figure: $x = \mu$ is its symmetry axis

Parameters of the Normal PDFs, graphically

Now, about the parameters of $\mathcal{N}(\mu, \sigma^2)$ on its PDF graph:

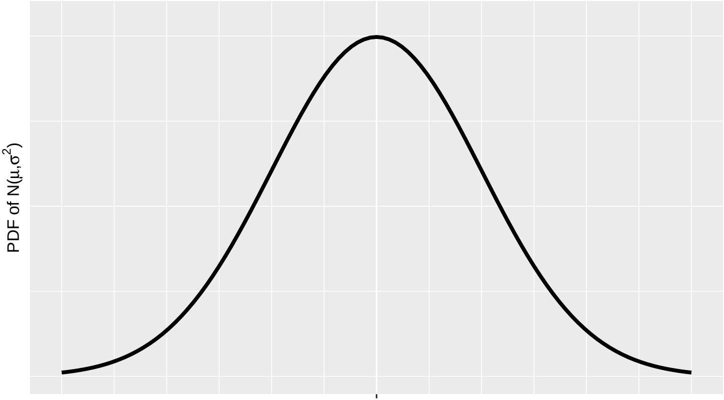

Figure: The PDF

Parameters of the Normal PDFs, graphically

Now, about the parameters of $\mathcal{N}(\mu, \sigma^2)$ on its PDF graph:

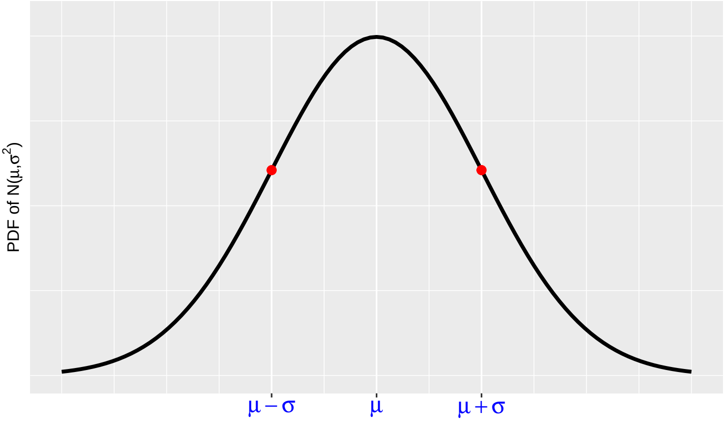

Figure: $\sigma =$ half of the distance between the infl. points

Parameters of the Normal PDFs, graphically

Now, about the parameters of $\mathcal{N}(\mu, \sigma^2)$ on its PDF graph:

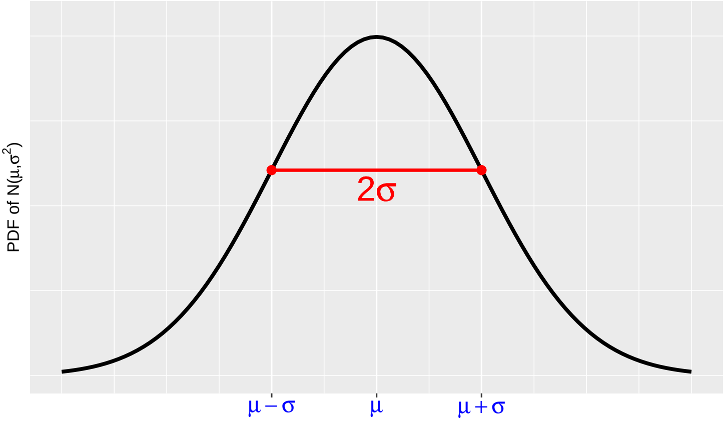

Figure: $\sigma =$ half of the distance between the infl. points

Comparison of Normal PDFs

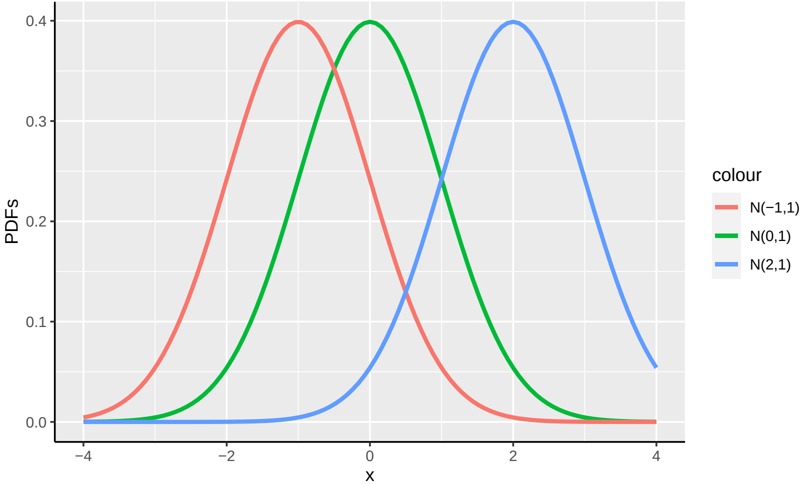

Figure: PDFs of $\mathcal{N}(-1,1)$, $\mathcal{N}(0,1)$, $\mathcal{N}(2,1)$

Comparison of Normal PDFs

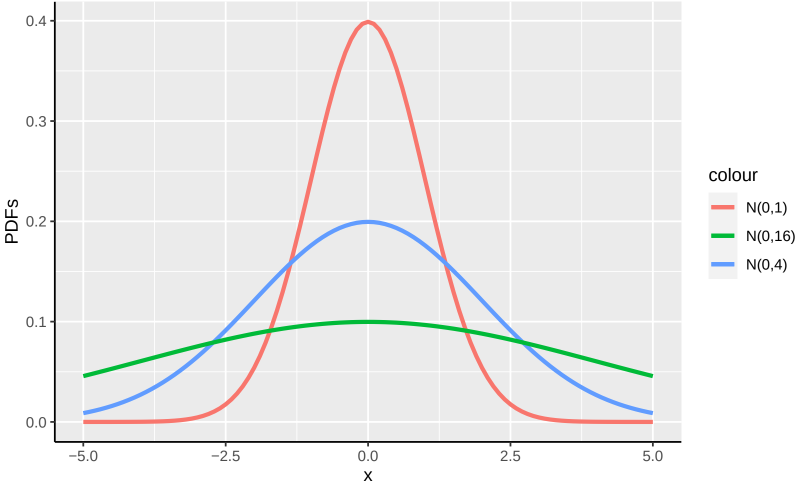

Figure: PDFs of $\mathcal{N}(0,1)$, $\mathcal{N}(0,2^2)$, $\mathcal{N}(0,4^2)$

Comparison of Normal PDFs

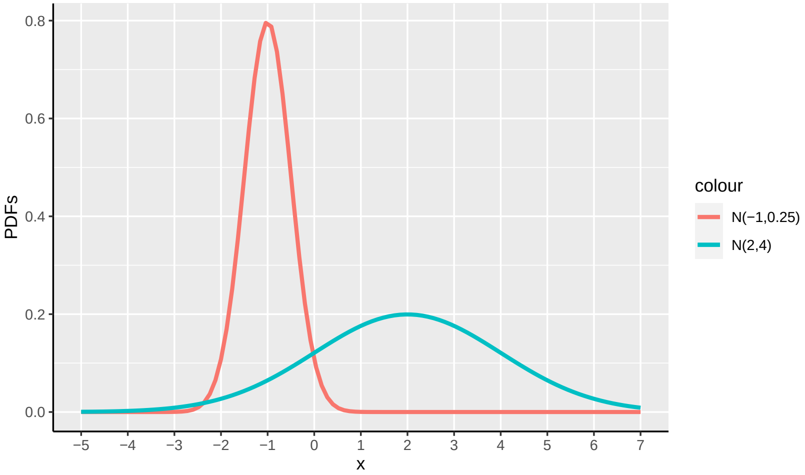

Figure: PDFs of $\mathcal{N}(-1, 0.5^2)$, $\mathcal{N}(2, 2^2)$

Some Properties

Assume f is the PDF of $X \sim \mathcal{N}(\mu, \sigma^2)$, F is its CDF, $\varphi(x)$ and $\Phi(x)$ are, respectively, the PDF and CDF of $Z \sim \mathcal{N}(0,1)$. Then

- $\varphi(x)$ is an even function, so its graph is symmetric around the Y-axis;
- The maximum point of φ is 0, the maximum value is $\varphi(0) = \frac{1}{\sqrt{2\pi}}$;
- φ has 2 inflection points: -1 and $+1$;
- $\Phi(-x) = 1 - \Phi(x)$;
- $f(x)$ is symmetric around the line $x = \mu$;
- The maximum point of f is μ, the maximum value is $f(\mu) = \frac{1}{\sigma \cdot \sqrt{2\pi}}$;
- f has 2 inflection points: $\mu - \sigma$ and $\mu + \sigma$.

Normal Distribution: Example

Example 27.2: Assume $X \sim \mathcal{N}(-1, 3)$.

a. Write down the PDF of X;
b. Write down the CDF of X;
c. Plot approximately, by a hand, the PDF of X;
d. Plot the PDF of X, using **R**;
e. Plot the CDF of X, using **R**.

Normal Distribution: 3σ rule

Assume $X \sim \mathcal{N}(\mu, \sigma^2)$. Then

- $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) \approx 0.682$
- $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.954$
- $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997$

and also

$$\mathbb{P}(\mu - 1.96\sigma < X < \mu + 1.96\sigma) \approx 0.95.$$

This is important for Statistics: if I am modeling some quantity by a Normal distribution, then I am 95% sure that its values are within 1.96 Standard Deviations from the Mean.

Normal Distribution: 3σ rule

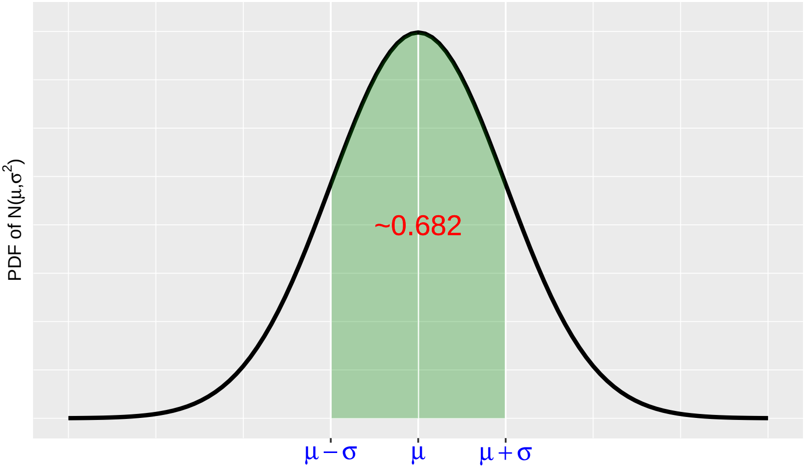

Figure: $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) \approx 0.682$

Normal Distribution: 3σ rule

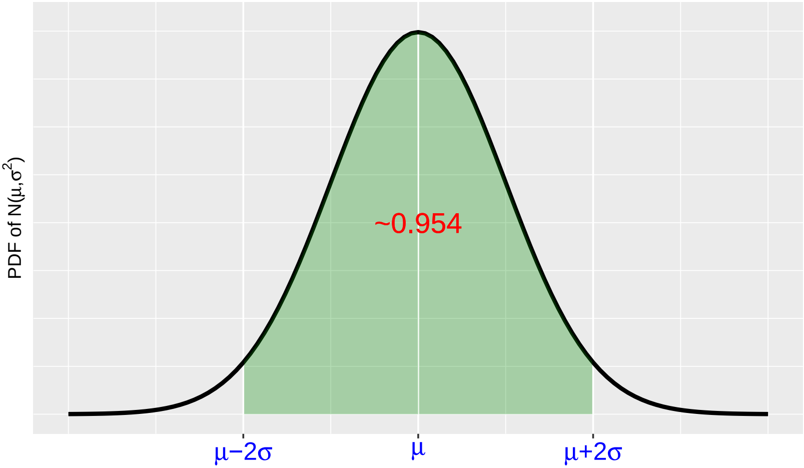

Figure: $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.954$

Normal Distribution: 3σ rule

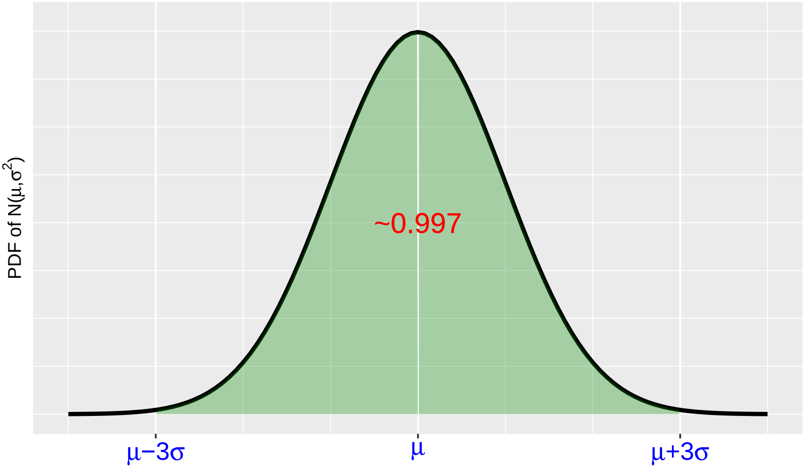

Figure: $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997$

Normal Distribution: 1.96σ rule

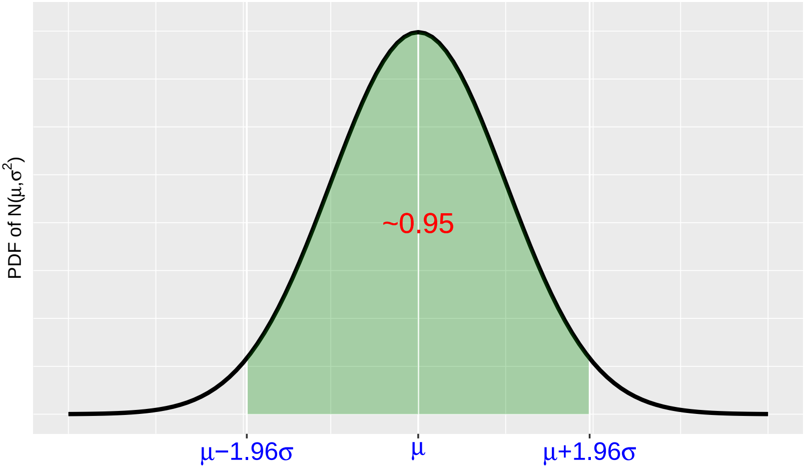

Figure: $\mathbb{P}(\mu - 1.96\sigma < X < \mu + 1.96\sigma) \approx 0.95$

Note: Bell-shaped Curves

Note: Not every r.v., the PDF of which looks like a bell-shaped curve, is a Normal r.v. . For example, the Standard Cauchy Distribution, $Cauchy(0, 1)$ is given by the PDF

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}, \qquad x \in \mathbb{R},$$

and the graph of its PDF is again bell-shaped.

Normal vs Cauchy Distribution

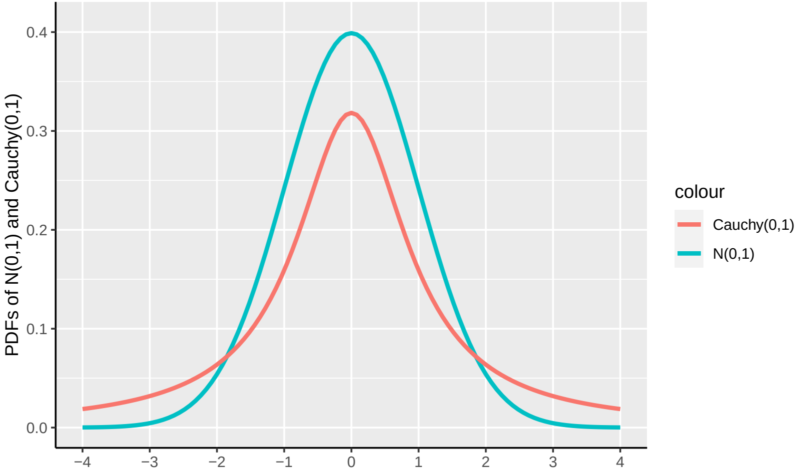

Figure: PDFs of $\mathcal{N}(0,1)$ and $Cauchy(0,1)$

From $\mathcal{N}(\mu, \sigma^2)$ to $\mathcal{N}(0, 1)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Y = X - \mu \sim \mathcal{N}(0, \sigma^2);$$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

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

so $Z = \dfrac{X - \mu}{\sigma}$ is a Standard Normal Variable. This transformation is called a **Standardization** (and sometimes, Normalization).

From $\mathcal{N}(0,1)$ to $\mathcal{N}(\mu,\sigma^2)$

- If $X \sim \mathcal{N}(0,1)$, then

$$Y = \mu + \sigma \cdot X \sim \mathcal{N}(\mu, \sigma^2)$$

- If $X \sim \mathcal{N}(\mu,\sigma^2)$, and if $Y = a \cdot X + b$, then

$$Y \sim \mathcal{N}(a \cdot \mu + b, a^2 \cdot \sigma^2).$$

Normal Distribution: On Standardization

Why is Standardization important: In older days, people were using the Standard Normal Tables to calculate probabilities for Normal r.v.s. The Idea was the following: assume we want to calculate $\mathbb{P}(a \le X \le b)$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. Then we Standardize X, in the following way:

$$\mathbb{P}(a \le X \le b) = \mathbb{P}(a-\mu \le X-\mu \le b-\mu) = \mathbb{P}\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$

Now we denote

$$Z = \frac{X-\mu}{\sigma}, \qquad \text{then} \qquad Z \sim \mathcal{N}(0,1).$$

So

$$\mathbb{P}(a \le X \le b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Normal Distribution: On Standardization

Another important application is in Statistics: say, we assume our data is coming from a Normal Distribution with Mean μ and Standard Deviation σ, i.e. our data is generated from $\mathcal{N}(\mu, \sigma^2)$. Assume we have a datapoint (observation) x, supposedly from our distribution. Then we can calculate the z-**score** for x:

$$z = \frac{x - \mu}{\sigma}.$$

If, say $|z| > 1.96$, then we can, with 95% confidence, say, that the observation x is not from our distribution.

Note: Note that I am using small x for the datapoint, since x is not a r.v., it is a supposed (possible) value of our r.v.

Note: Above I was using μ and σ, assuming that I know their values. Usually, this is not the case, and in that case we are taking their estimates. (Wait for the Statistics Course!)

Normal Distribution: On Standardization

The third important application is in Data Science, in Data Preprocessing: if you have a dataset(s) of numbers $x : x_1, x_2, ...$ (say, for different features of your observations), in many cases you want to Normalize them, to bring to some "common" interval. Say, you want to compare the effects of that numbers on something, or not to allow your program to ignore some part of your data (say, if you have some very large numbers for one feature, and also very small numbers for the other one, your program can just neglect that small numbers).

Normal Distribution: On Standardization

The Normalization can be done in various ways, say:

- Bringing x to $[0, 1]$: you just calculate

$$y_i = \frac{x_i - \min_k x_k}{\max_k x_k - \min_k x_k}, \qquad \forall i;$$

- Standardization: calculate

$$z_i = \frac{x_i - \mu}{\sigma}, \qquad \forall i;$$

  where μ and σ are either theoretical (if you know them) or estimated Mean and Standard Deviation for your dataset x.

Normal Distribution: Example

Example 27.3: Assume $X \sim \mathcal{N}(3, 10)$.

a. Standardize X;
b. What is the Distribution of that Standardized r.v.?
c. Express $\mathbb{P}(-2 \leq X < 10)$ in terms of a probability with a Standard Normal r.v. , and in terms of the CDF of a Standard Normal Distribution $\Phi(x)$;
d. Calculate the value of the above probability by using Standard Normal Tables and **R**.

Example 27.4: Assume $X \sim \mathcal{N}(0,1)$.

a. Transform X (linearly) to obtain a r.v. $Y \sim \mathcal{N}(-3,6)$;
b. Generate, in **R**, 10000 random numbers from the Standard Normal Distribution, transform them using the above transformation, plot their relative frequency histogram, and add the theoretical PDF of $\mathcal{N}(-3,6)$.

Normal Distribution: Example

Example 27.5: Assume $X \sim \mathcal{N}(0,1)$. Calculate

a. $\mathbb{P}(X \in \{0,1,2\})$;

b. $\mathbb{P}(X < 0)$;

c. $\mathbb{P}(X^2 < 1)$;

d. $\mathbb{P}(X > 3)$;

e. $\mathbb{P}(X < 3 | X > 0)$.

Normal Distribution: Example

Example 27.6: Assume that the heights of women in Armenia are Normally distributed with the Mean 158.1cm and Standard Deviation 5.7cm[1].

- Calculate the Probability that the (randomly chosen) woman height will be smaller than 158.1cm.
- Calculate the Probability that the (randomly chosen) woman height will be larger than 170cm.

[1] See `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0018962`

Normal Distribution: Example

Example 27.7: Prove the $68 - 95 - 99.7$ rule (3σ rule): if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

a. $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) \approx 0.682$;
b. $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.954$;
c. $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.997$.

Quantiles of the Normal Distribution

Now, let's talk about the quantiles of the Normal Distribution: they are very important in Statistics. Let us consider a r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$.

Now, we fix a number $\alpha \in (0, 1)$. The α-**quantile** of the Distribution $\mathcal{N}(\mu, \sigma^2)$ is the number $q = q_\alpha \in \mathbb{R}$ satisfying

$$\mathbb{P}(X \leq q) = \alpha.$$

In words, the α-quantile q_α of $\mathcal{N}(\mu, \sigma^2)$ is a point that divides the values of a r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$ into two parts: the probability that X will be less than q_α is exactly α.

Quantiles of the Normal Distribution

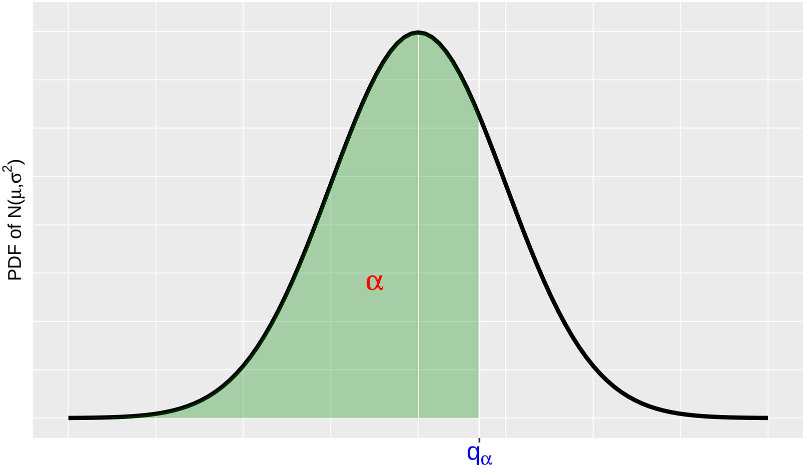

Figure: The shaded area is α

Quantiles of the Normal Distribution

Example 27.8: Assume $X \sim \mathcal{N}(-2, 4)$.

a. Find the 0.5-quantile of $\mathcal{N}(-2, 4)$;
b. Find the 0.7-quantile of $\mathcal{N}(-2, 4)$;
c. Find the 30% quantile[2] of $\mathcal{N}(-2, 4)$;
d. Which order (level) quantile is $q = 5$ for $\mathcal{N}(-2, 4)$?

[2] This is the same as the 0.3-quantile.

Quantiles of the Normal Distribution

Example 27.9: Assume $X \sim \mathcal{N}(0,1)$, $\alpha \in (0,1)$ and q_α is the α-quantile of $\mathcal{N}(0,1)$.

a. Calculate $\mathbb{P}(X > q_\alpha)$;

b. Express $q_{1-\alpha}$, the $(1-\alpha)$-level quantile of $\mathcal{N}(0,1)$, in terms of q_α;

c. Assume we want to find a number a such that

$$\mathbb{P}(-a \leq X \leq a) = 1 - \alpha.$$

Express a in terms of the quantiles of our distribution.

Normal Distribution: Note

Note: Small note about which kind of quantities can be modeled through Normal Distribution. In the literature you will find different quantities, like human heights, weights, SAT scores, some measurement errors, ... The explanation is usually the following: we will learn soon the CLT stating that the sum of independent r.v.s having the same distribution is approximately Normally distributed. So if quantities can be described as

$$\text{Some Mean} + \text{different independent sources of errors}$$

then it is reasonable to model them using the Normal Distribution.

Normal Distribution: Note

A cautionary note: Please note that if $X \sim \mathcal{N}(\mu, \sigma^2)$ for some μ and σ, then

$$\mathbb{P}(X < 0) > 0.$$

So for any Normal r.v., there is a positive probability that it will take negative values!! And if we model, say, weight of a person W using a Normal Distribution, then we will obtain that there is a nonzero probability that the weight is negative. But, of course, if the Mean is large and SD is not so large, then this probability will be veeeery small, will be negligible.

Some facts about the Normal Distribution

Examples of Normal Distribution

You can find many examples of Normal Distribution in the real world.

Figure: Hallgrímskirkja, church in Reykjavík, Iceland

Other examples of Normal Distribution

Figure: Normal Table

Other examples of Normal Distribution

Other examples of Normal Distribution

Other examples of Normal Distribution

Surely you know what us **curving**. And, of course, you know that the Probability that we will do curving for Probability is 0 $\ddot\smile$.

But, what is curving, in fact? And what is the **curve** we are talking about? Of course, our new friend the Bell-Shaped Gaussian Curve!

See `https://www.youtube.com/watch?v=vqNExEhXHvc`

Some Serious Things $\smile$̈

Fact: Gauss introduced the Normal Distribution in 1809 for modeling observational errors in the astronomy;

Fact: In 1997, the Nobel Prize in Economic Sciences was awarded to Robert Merton and Myron Scholes for the **Black-Scholes options pricing model**. And the Black-Scholes formula involves the CDF of the Standard Normal Distribution: see, e.g., `https://en.wikipedia.org/wiki/Black-Scholes_model`.

Fact: The Normal Distribution is very important in Prob and Stat not only because many quantities can be modeled by the Normal Distribution, but also because of the Central Limit Theorem (CLT). We will talk about CLT later, but at this moment let me show the Galton Board: see Link 1 or Link 2 or Link 3.

Some Serious Things $\smile$

Fact: Entropy is an important notion in the Information Theory. It measures the uncertainty, the information contained in the r.v. For example, if X is a discrete r.v. with values x_i and corresponding probabilities p_i, then the Entropy of X, in bits, is given by (we take $0 \cdot \log_2 0 = 0$)

$$\mathcal{H}(X) = -\Big(p_1 \cdot \log_2 p_1 + p_2 \cdot \log_2 p_2 + \dots\Big) = -\sum_k p_k \cdot \log_2 p_k.$$

Some Serious Things $\smile$̈

Say, for the r.v. $X \sim Bernoulli(p)$,

$$X \sim \left(\begin{array}{cc} 0 & 1 \\ 1-p & p \end{array} \right),$$

and the Entropy is equal to

$$\mathcal{H}(X) = -\Big(p \cdot \log_2 p + (1-p) \cdot \log_2(1-p)\Big)$$

Then this Entropy will be minimal, if $p = 0$ or $p = 1$, and will be maximal, if $p = \frac{1}{2}$. I.e., the information, uncertainty in the X is minimal, if $p = 0$ or 1, and is maximal, if $p = \frac{1}{2}$.

Some Serious Things $\smile$

Now, for a continuous r.v. X, with the PDF $f(x)$, the Entropy is defined as

$$\mathcal{H}(X) = -\int_{-\infty}^{+\infty} f(x)\log_2 f(x)\,dx.$$

Now, one of the nice characterizations of the Normal Distribution: among all Distributions with the mean μ and variance σ^2, the one with the Maximum Entropy is the Normal Distribution $\mathcal{N}(\mu, \sigma^2)$. So if we know about the r.v. X only its mean and variance, then X will have the maximum uncertainty, maximum information, iff $X \sim \mathcal{N}(\mu, \sigma^2)$.