Nov 23, 2022

Engineering Statistics

Week 5: Random variable, probability distribution and expected value

Random variable

Continuous random variable

Random variables that can assume values corresponding to any of the points contained in an interval are called continuous.

Example

- If in the study of the ecology of a lake, we make depth measurements at randomly chosen locations, then X = the depth at such a location is a continuous random variable. Here, values that X can take between the minimum depth and the maximum depth in the region being sampled.

- If a chemical compound is randomly selected and its pH X is determined, then X is a continuous random variable because any pH value between 0 and 14 is possible.

Continuous random variable

- Although the probability distribution of a continuous random variable cannot be presented in tabular form, it can be stated as a formula.
- Such a formula would necessarily be a function of the numerical values of the continuous random variable X and as such will be represented by the functional notation $f(x)$.
- In dealing with continuous variables, $f(x)$ is usually called the probability density function, or simply the density function of X.
- If X is continuous then it is defined over a continuous sample space.

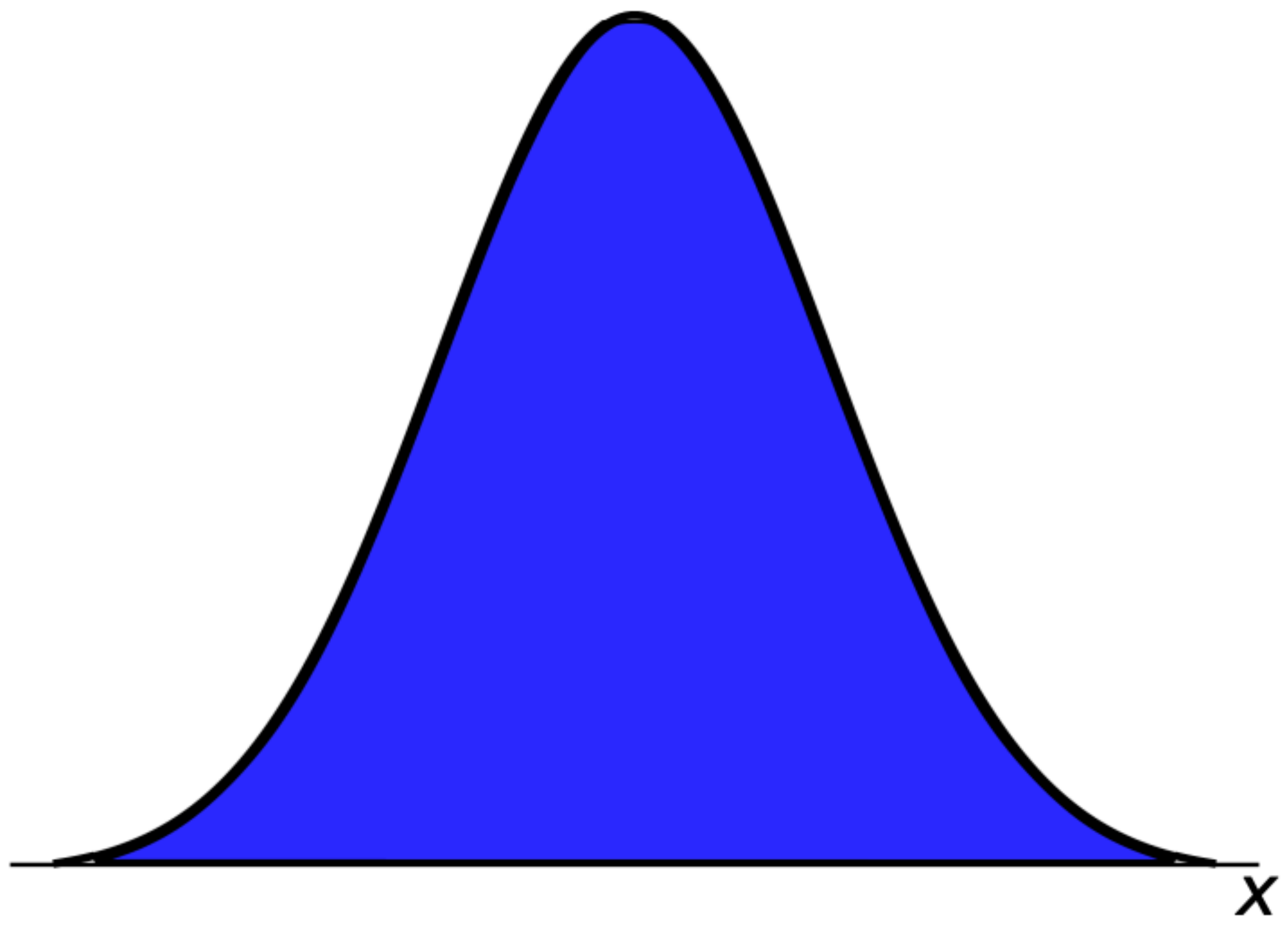
Probability distribution of a continuous random variable

Probability density function of a continuous variable

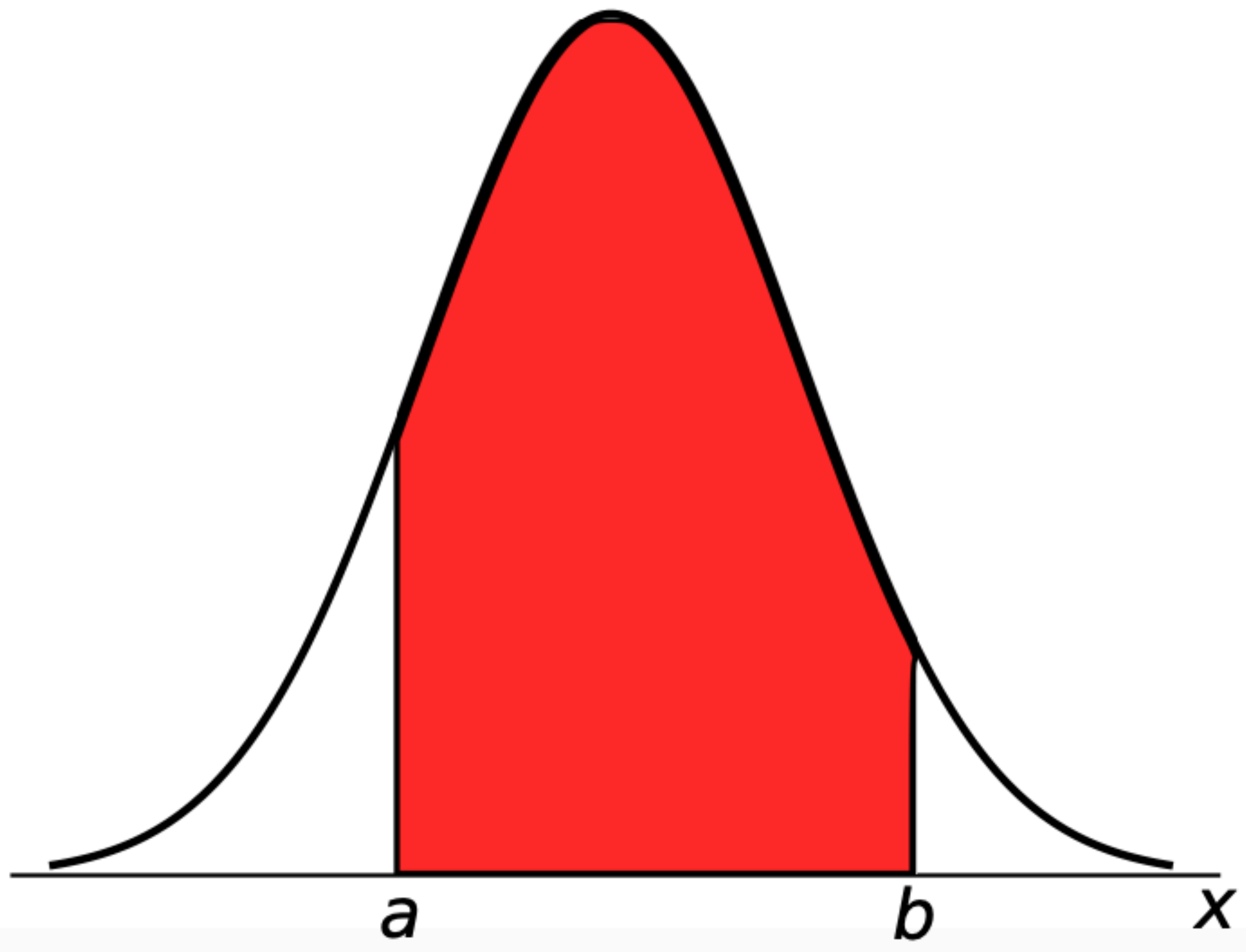
The probability density function (pdf) $f(x)$ describes the distribution of probability for a continuous random variable X. It has the following properties:

- $f(x) \geq 0$ for all x.
- The total area under the probability density curve is 1.
- $P(a \leq X \leq b)$ = area under the probability density curve between a and b.

Probability density function of a continuous variable



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



$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

Probability density function of a continuous variable

- With a continuous random variable, the probability that $X = x$ is always 0, i.e. $P(X = x) = 0$
- When determining the probability of an interval a to b, we need not be concerned if either or both endpoints are included in the interval. Since the probabilities of $X = a$ and $X = b$ are both equal to 0,

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

Example

Suppose that the error in the reaction temperature, in celcius, for a controlled laboratory experiment is a continuous random variable X having the function

$$f(x) = \frac{x^2}{3}, -1 < x < 2$$

- Verify that $f(x)$ is a pdf.
- Find $P(0 < X \leq 1)$.

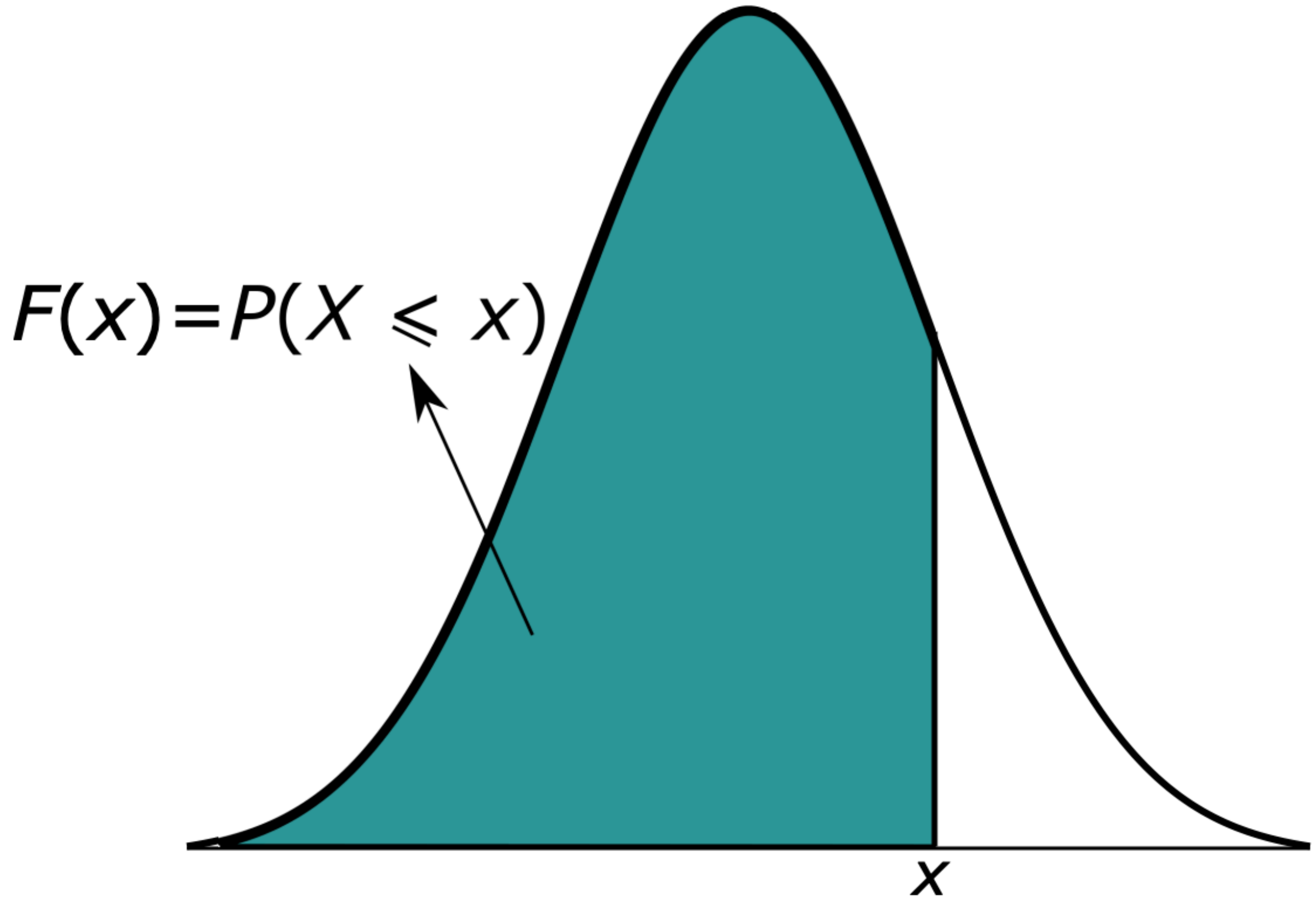
Cumulative density function of a continuous random variable

Cumulative density function of a continuous random variable

The cumulative density function (cdf) $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \text{ for all } -\infty < x < \infty.$$

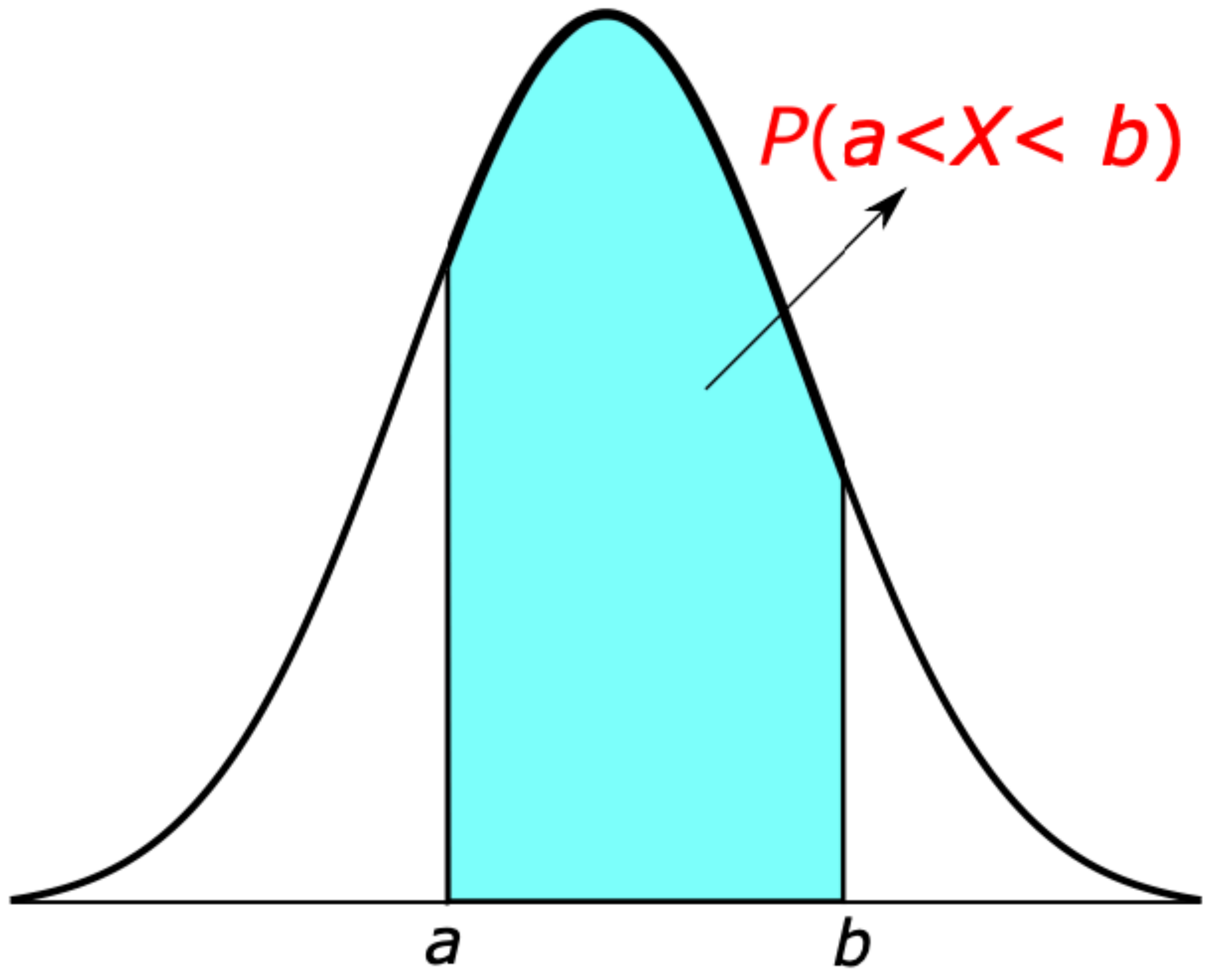
Cumulative density function of a continuous random variable



Cumulative density function of a continuous random variable

The cdf of a continuous random variable X can also be used for computing the probabilities associated with X. If the cdf of X is $F(x)$ then

$$P(a < X < b) = F(b) - F(a)$$



Example

Suppose that the error in the reaction temperature, in celcius, for a controlled laboratory experiment is a continuous random variable X having the function

$$f(x) = \frac{x^2}{3}, -1 < x < 2$$

Compute the probability $P(0 < X \leq 1)$ using this pdf.

Mean and variance of a continuous random variable

Mean and variance

Suppose that a continuous random variable X has the pdf $f(x)$. The mean and variance of the X are denoted by μ and σ^2, respectively and formulated as follows:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$\sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Example

Suppose that the error in the reaction temperature, in celcius, for a controlled laboratory experiment is a continuous random variable X having the function

$$f(x) = \frac{x^2}{3}, -1 < x < 2$$

Find the mean and variance of X.

Example

$$\mu = E(X) = \int_{-1}^{2} xf(x)dx = \int_{-1}^{2} x\frac{x^2}{3}dx = \int_{-1}^{2} \frac{x^3}{3}dx = \frac{x^4}{4*3}\Big|_{-1}^{2} = \left(\frac{2^4}{4*3}\right) - \left(\frac{-1^4}{4*3}\right) = \frac{16}{12} - \frac{1}{12} = 1.25$$

$$\sigma^2 = E(X-\mu)^2 = \int_{-1}^{2} (x-\mu)^2 f(x)dx = \int_{-1}^{2} (x-1.25)^2\frac{x^2}{3}dx = \frac{1}{3}\int_{-1}^{2} (x-1.25)^2 x^2 dx = \frac{1}{3}\int_{-1}^{2} x^4 - 2.5x^3 + 1.5625x^2 dx = \frac{1}{3}(33.5 + 9.375 + 4.6875) = 0.6375$$

Normal distribution

Normal distribution

- The normal distribution is the most important one in all of probability and statistics.
- Many numerical populations have distributions that can be fit very closely by an appropriate normal curve.
- Even when individual variables themselves are not normally distributed, sums and averages of the variables will under suitable conditions have approximately a normal distribution; this is the content of the **Central Limit Theorem**.

Normal distribution

A continuous random variable X is said to have a normal distribution with parameters μ and σ^2 if the probability density function of X is

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

where $-\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$. Here,

$$E(X) = \mu \text{ and } Var(X) = \sigma^2$$

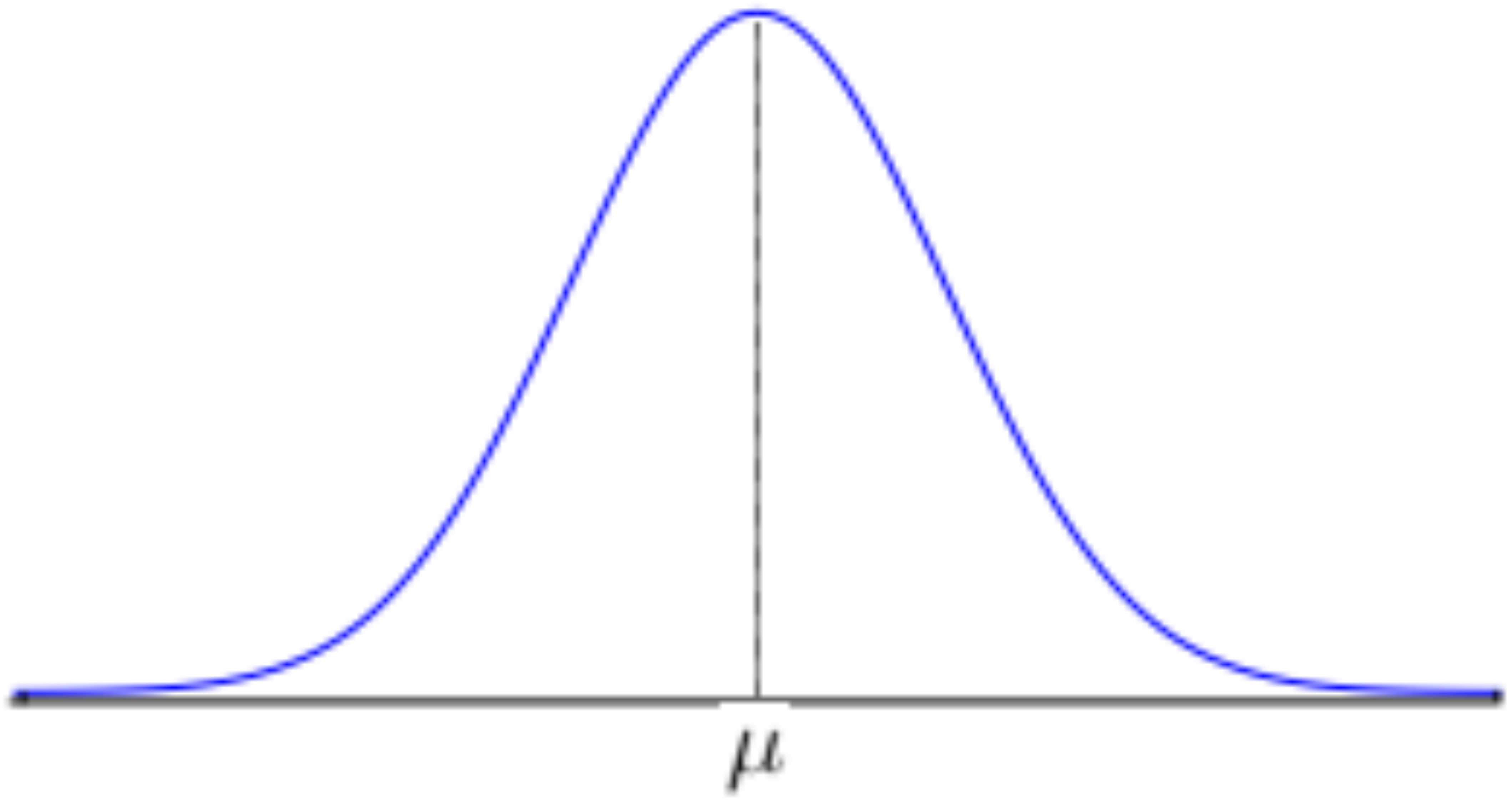
are mean and variance of the normal distribution, respectively.

Normal distribution

The random variable X having the probability distribution given is said to have normal distribution with parameters μ and σ^2. This is shortly denoted by:

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

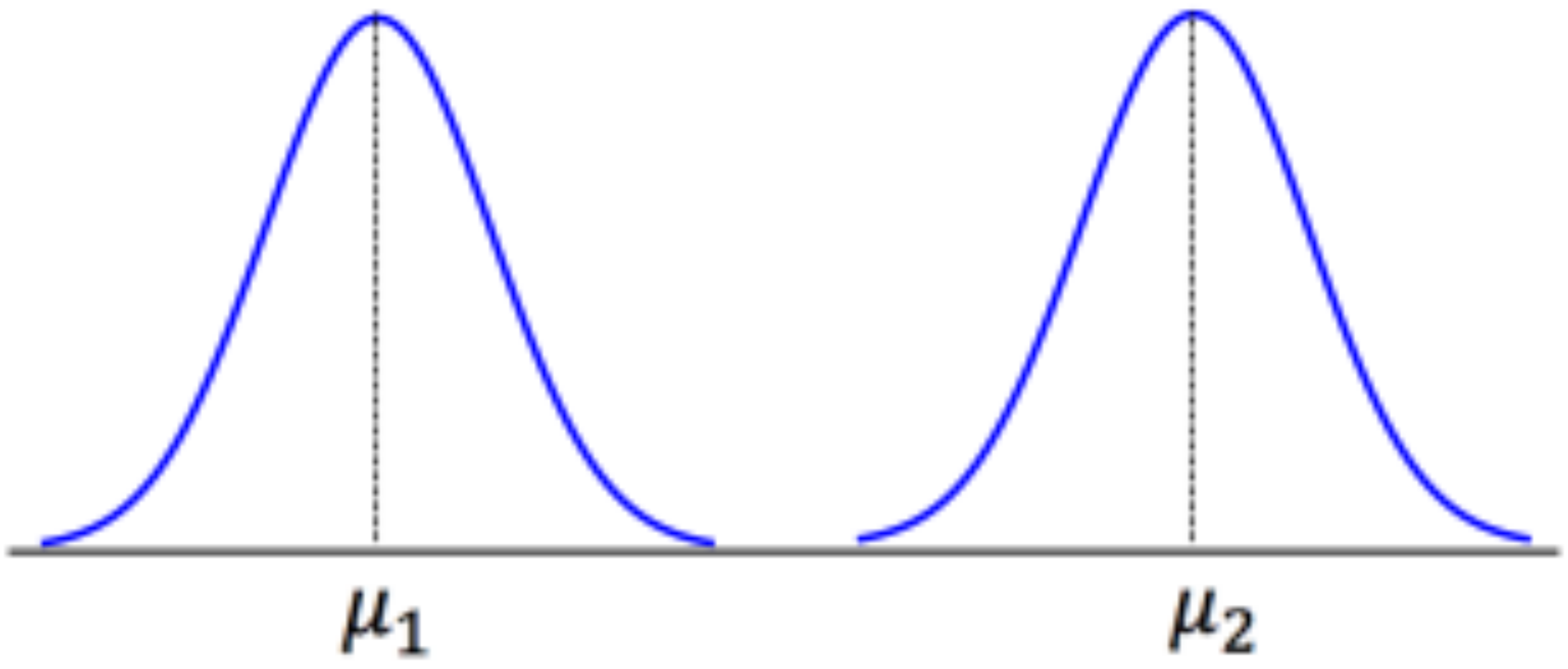
Normal distribution



The normal curve:

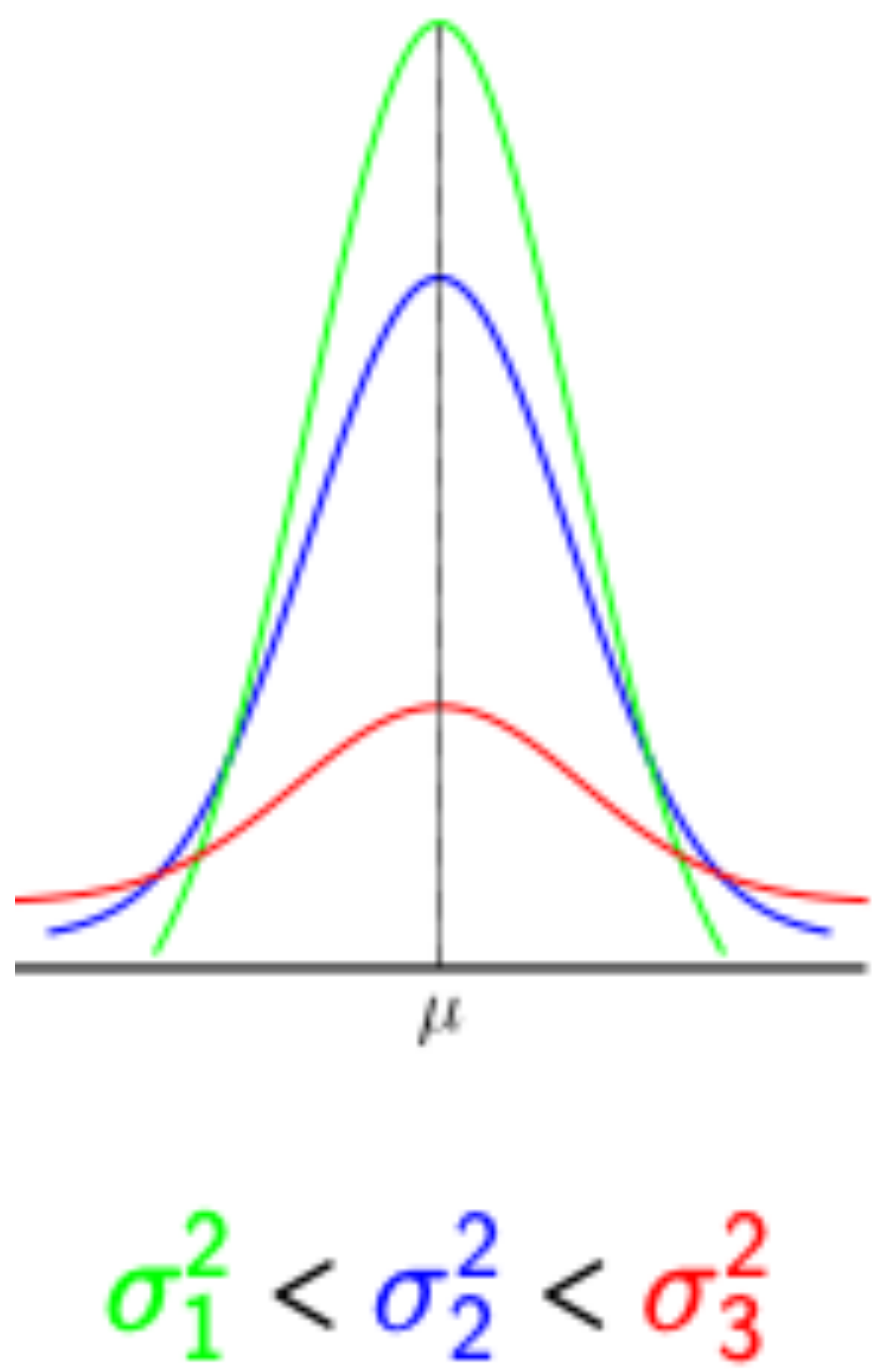
- is bell-shaped,
- is symmetric around μ.
- attains its maximum value at μ.

Normal distribution



μ controls the location of the distribution. A change of mean from μ_1 to a larger value μ_2 merely slides the bell-shaped curve along the axis until a new center is established at μ_2. There is no change in the shape of the curve.

Normal distribution



σ^2 controls the spread of the normal distribution. Large values of it yield graphs that are quite spread out about μ, whereas small values of σ yield graphs with a high peak above μ and most of the area under the graph quite close to μ.

Normal distribution

Normal distribution has two parameters:

- Location parameter: $\mu = E(X)$
- Scale parameter: $\sigma^2 = Var(X)$

Standard normal distribution

Standard normal distribution

The normal distribution with parameter values $\mu = 0$ and $\sigma^2 = 1$ is called the standard normal distribution. A random variable having a standard normal distribution is called a standard normal random variable and will be denoted by Z. The probability density function of Z is

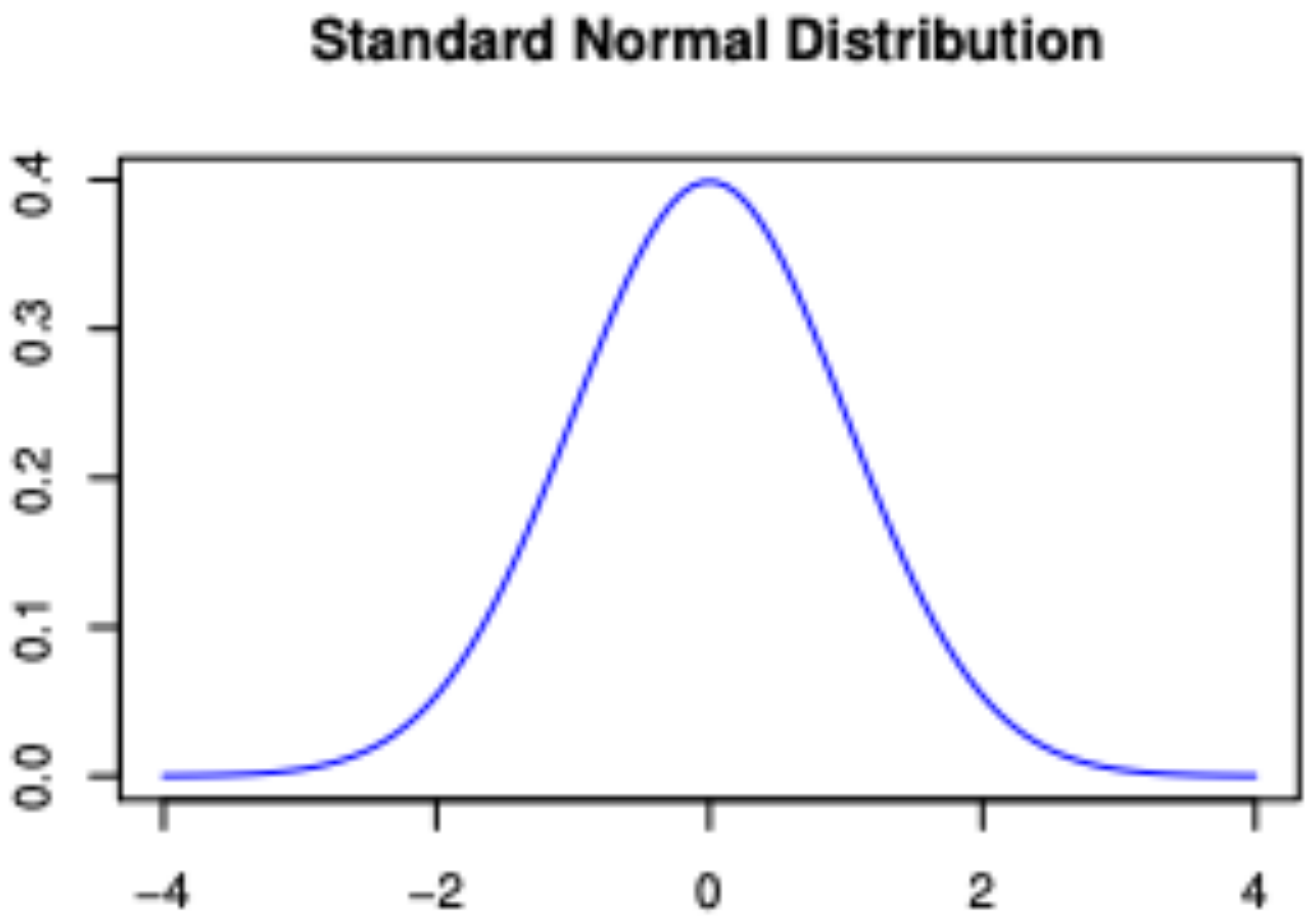
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty$$

Standard normal distribution

The random variable X having the probability distribution given is said to have standard normal distribution is shortly denoted by:

$$Z \sim N(0,1)$$

Standard normal distribution



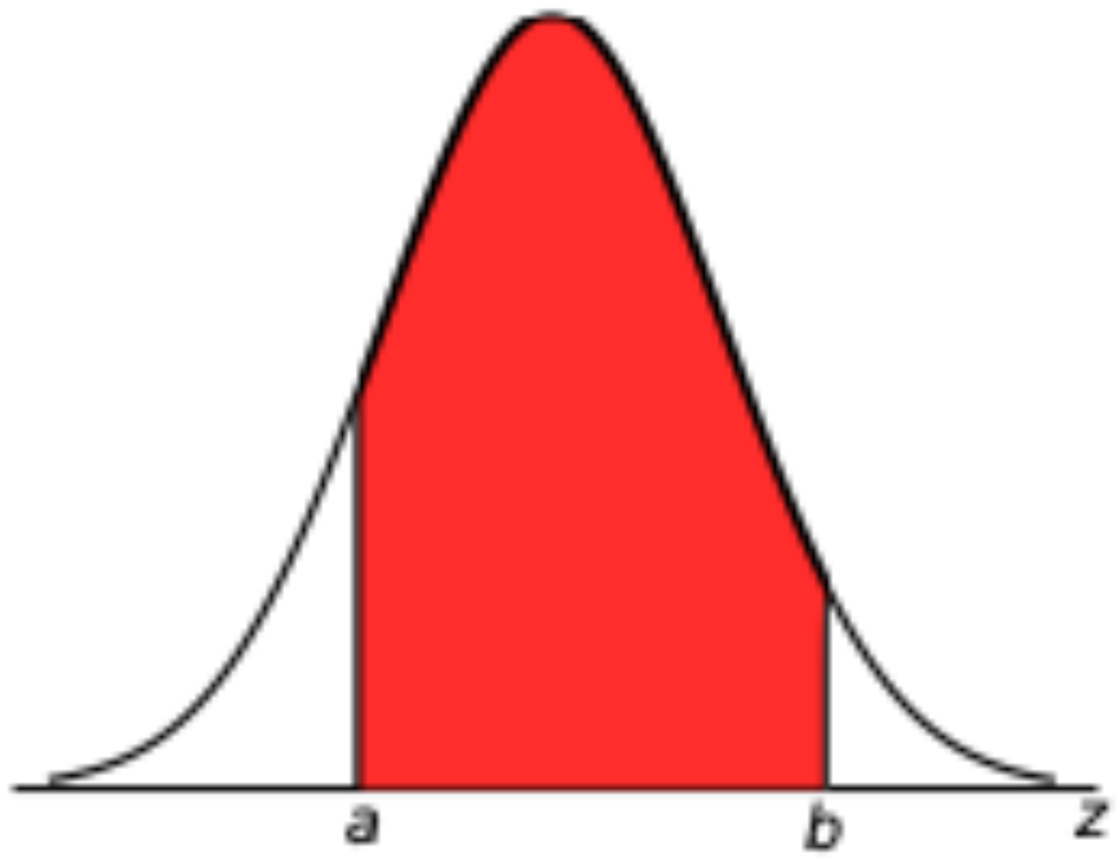
The standard normal curve:

- is bell-shaped,
- is symmetric around 0.
- attains its maximum value at 0.

Standard normal distribution

Calculating probabilities associated with standard normal distribution

Suppose that $Z \sim N(0,1)$, and we want to calculate the probability $P(a < Z < b)$. As indicated earlier, $P(a < Z < b)$ is area under the standard normal curve between a and b.

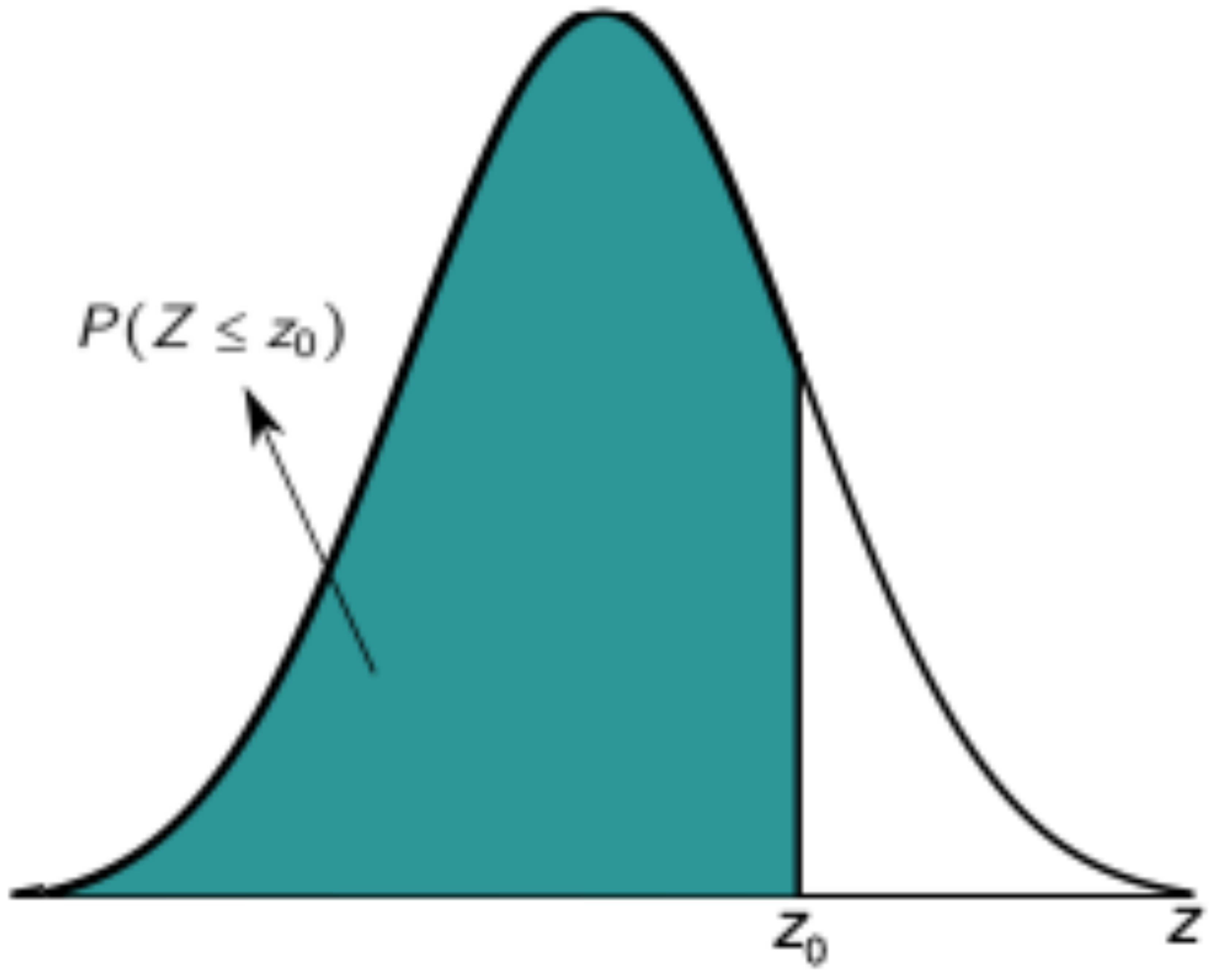


$$P(a < Z < b) = \int_a^b \phi(z)dz = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

Standard normal probabilities table

It is not possible to solve the integral explicitly. Therefore, we use **standard normal probabilities table**.

The standard normal table gives the area to the left of a specified value of z_0 as $\phi(z_0) = P(Z \leq z_0)$ = Area under curve to the left of z_0.



Standard normal probabilities table

Standard Normal Probabilities

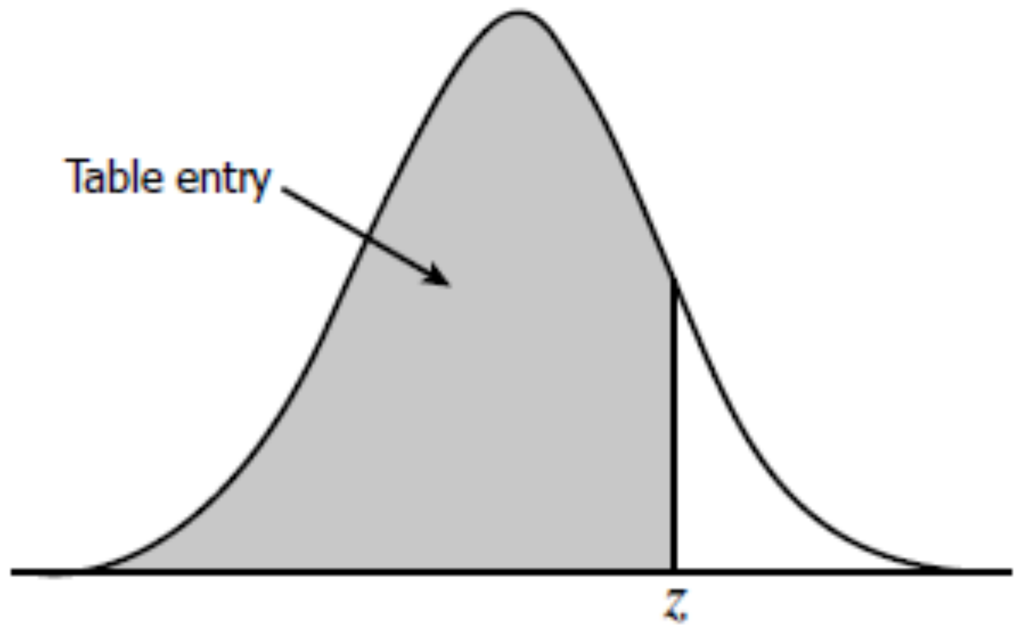


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

STANDARD NORMAL PROBABILITIES

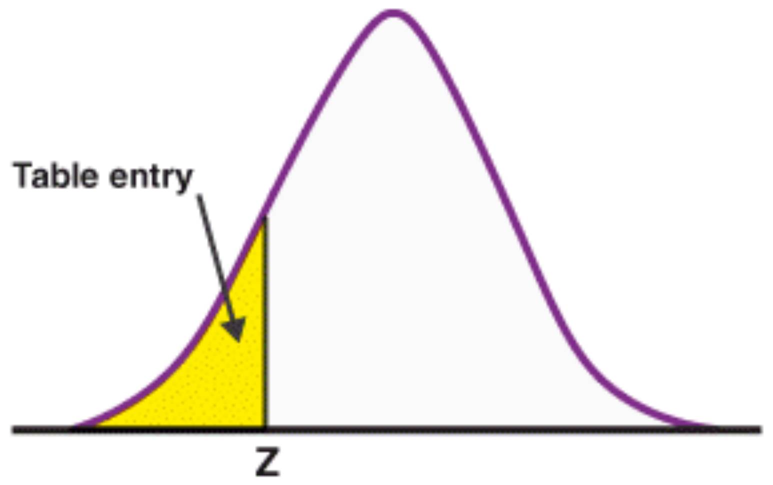


Table entry for z is the area under the standard normal curve to theleft of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143

Standard normal probabilities table

Some rules:

- $P(a \leq Z \leq b) = \phi(b) - \phi(a) = P(Z \leq b) - P(Z \leq a)$
- $P(Z \leq 0) = \phi(0) = 0.5$
- $P(z \geq -z_0) = 1 - \phi(z_0) = 1 - P(Z \leq z_0)$

Example

Let $Z \sim N(0,1)$. Compute the following probabilities.

- $P(-0.38 \leq Z \leq 1.85)$
- $P(Z \leq -1.64)$
- $P(Z \geq 1.96)$
- $P(|Z| \geq 2.58)$

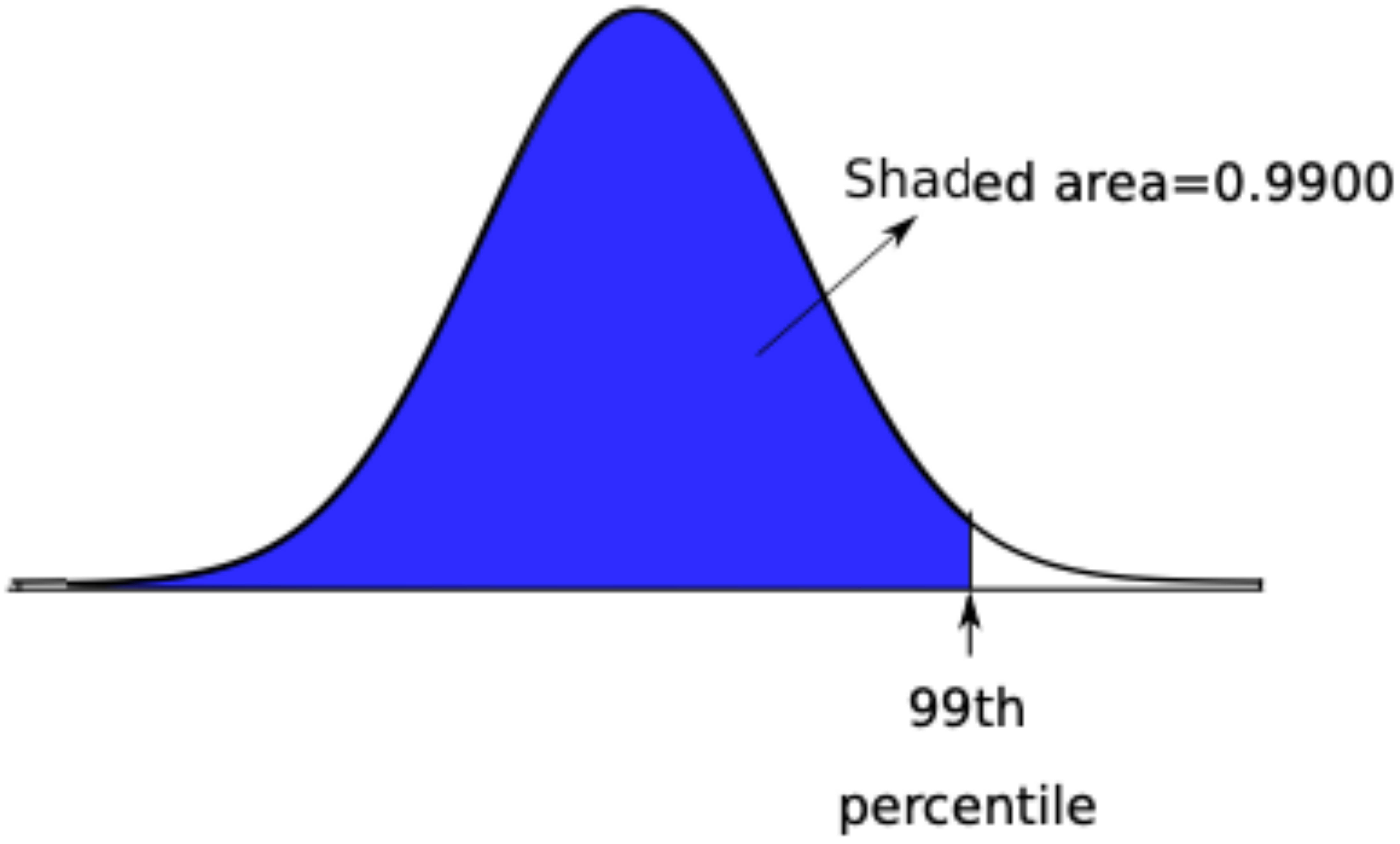
Percentiles of the standard normal distribution

For any p between 0 and 1, standard normal probabilities table can be used to obtain the $(100p)th$ percentile of the standard normal distribution.

$$\phi(z_0) = p \leftrightarrow z_0 = \phi^{-1}(p), \quad 0 < p < 1$$

Example

The 99th percentile of the standard normal distribution is that value on the horizontal axis such that the area under the z curve to the left of the value is 0.9900.



Example

Let $Z \sim N(0,1)$. Then,

- $P(Z \leq a) = 0.6255$ find a.
- $P(Z > b) = 0.0250$ find b.
- $P(Z < c) = 0.90$ find c.

Example

Let $X \sim N(60, 16)$ and we would like to compute $P(55 < X \leq 62)$. Do we need a probability table for $N(60, 16)$?

Fortunately, no new tables are required for probability calculations regarding the general normal distribution. Any normal distribution can be set in correspondence to the standard normal by a transformation called **standardization**.

Standardization

If X is distributed as $N(\mu, \sigma^2)$, then the variable

$$Z = \frac{x - \mu}{\sigma}$$

has the standard normal distribution. This transformation is called **standardization.**

Standardization

Standardization of the normal distribution allows us to cast a probability problem concerning X into one concerning Z.

$$X \sim N(\mu, \sigma^2) \to Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Standardization

To find the probability that X lies in a given interval, convert the interval to the Z scale and then calculate the probability by using the standard normal probabilities table.

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

Example #1

If X is distributed as $N(60,16)$, find $P(55 < X \leq 62)$.

Example #1

If X is distributed as $N(60,16)$, find $P(55 < X \leq 62)$.

$$P(55 < X \leq 62) = P\left(\frac{55-60}{\sqrt{16}} < \frac{X-60}{\sqrt{16}} \leq \frac{62-60}{\sqrt{16}}\right) = P\left(\frac{-5}{\sqrt{16}} < Z \leq \frac{2}{\sqrt{16}}\right)$$

$$= P\Big(-1.25 < Z \leq 0.5\Big) = F(0.5) - [1 - F(1.25)]$$

$$= 0.6915 - (1 - 0.8944) = 0.6915 - 0.1056 = 0.5859$$

Example #2

The number of calories in a salad on the lunch menu is normally distributed with mean $\mu = 200$ and standard deviation $\sigma = 5$. Find the probability that the salad you select will contain:

- More than 208 calories.
- Between 190 and 200 calories.

Example #3

The breakdown voltage of a randomly chosen diode of a particular type is known to be normally distributed. What is the probability that a diode's breakdown voltage is within,

- 1 standard deviation of its mean value?
- 2 standard deviation of its mean value?
- 3 standard deviation of its mean value?

Example #4

Suggest that the population of hours of sleep can be modeled as a normal distribution with mean $\mu = 7.2$ hours and standard deviation $\sigma = 1.3$ hours.

(a) Determine the probability assigned to sleeping less than 6.5 hours.

(b) Find the 70th percentile of the distribution for hours of sleep.

Example #5

Suppose the scores on a college entrance examination are normally distributed with a mean of 550 and a standard deviation of 100. A certain prestigious university will consider for admission only those applicants whose scores exceed the 90th percentile of the distribution.

Find the minimum score an applicant must achieve in order to receive consideration for admission to the university.

Approximating the binomial distribution

The normal approximation to the binomial distribution

- The binomial distribution pertains to the number of successes X in n independent trials of an experiment.
- When the success probability p is not too near 0 or 1 and the number of trials is large, the normal distribution serves as a good approximation to the binomial probabilities.
- Bypassing the mathematical proof, we concentrate on illustrating the manner in which this approximation works.

The normal approximation to the binomial distribution

Let $X \sim Binom(n, p)$. When np and $n(1-p)$ are both large, say, greater than 15, the binomial distribution is well approximated by the normal distribution having,

$$\mu = np \text{ and } \sigma^2 = np(1-p)$$

That is,

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

Example

In a large scale statewide survey concerning television viewing by children, about 40% of the babies a few months old were reported to watch TV regularly. In a future random sample of 150 babies in this age group, let X be the number who regularly watch TV. Approximate the probability that,

(a) X is between 52 and 71 both inclusive.

(b) X is greater than 67.

Course materials

You can download the notes and codes from:

https://github.com/mcavs/ESTUMatse_2022Fall_EngineeringStatistics

SCAN ME

Contact

Do not hesitate to contact me on:

https://twitter.com/mustafa__cavus

https://www.linkedin.com/in/mustafacavusphd/

mustafacavus@eskisehir.edu.tr