7. Continuous Probability Distributions

Cumulative distribution function

We define a *cumulative distribution function (cdf)* $\boldsymbol{F(x)}$ of a random variable $\boldsymbol{X}$ as

$$\boldsymbol{F(x) = P(X \leq x)}$$

Properties of the cdf

1. $\boldsymbol{F(-\infty) = 0}$

2. $\boldsymbol{F(+\infty) = 1}$

3. $\boldsymbol{F(x)} \uparrow$

Probability density function

If the cdf is differentiable then we can define a ***probability density function (pdf)*** by

$$f(x) = F'(x)$$

The probability density function satisfies the following conditions:

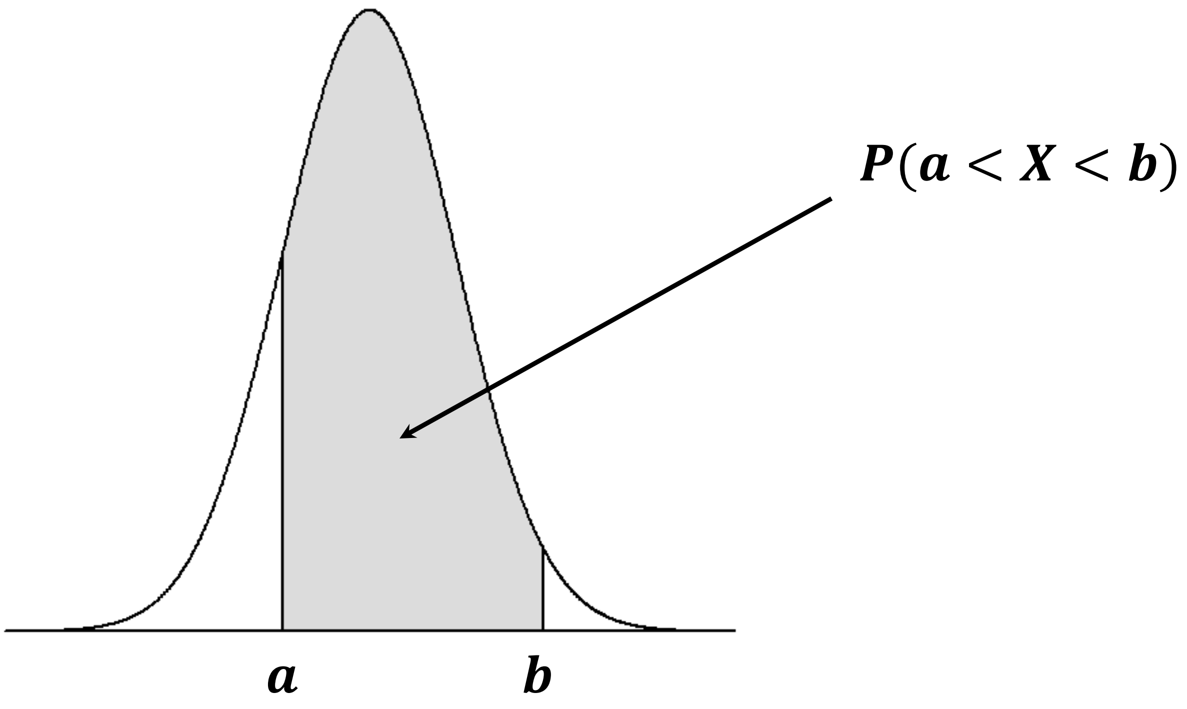
- $f(x)$ is non-negative,
- the total area under the curve representing $f(x)$ equals to 1.

Probability density function

The probability that $\boldsymbol{X}$ falls between $\boldsymbol{a}$ and $\boldsymbol{b}$ is found by calculating the area under the graph of $\boldsymbol{f(x)}$ between $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$\boldsymbol{P(a < X < b) = \int_a^b f(x)dx}$$

Probability density function: important picture



Expected value

The expected value of the random variable $\boldsymbol{X}$ is

$$\boldsymbol{E}(\boldsymbol{X}) = \int_{-\infty}^{+\infty} \boldsymbol{x}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{d}\boldsymbol{x}$$

Properties of the expected value

Properties of the expected value:

- $E(aX) = aE(X)$, if a is a constant
- $E(X+Y) = E(X) + E(Y)$
- $E(XY) = E(X)E(Y)$, if random variables X and Y are independent

Definition: **Random variables X and Y are *independent* iff**

$$P(X \leq a \; and \; Y \leq b) = P(X \leq a)P(Y \leq b)$$

for all possible a and b.

Expected value of *g(x)*

The expected value of the random variable $g(X)$ is given by

$$E\big(g(X)\big) = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

Variance

Variance of a continuous random variable is given by

$$\boldsymbol{Var(X) = E\big(X - E(X)\big)^2}$$

$$\boldsymbol{= \int_{all\ x} \big(x - E(X)\big)^2 f(x)dx}$$

or

$$\boldsymbol{Var(X) = E\big(X^2\big) - \big(E(X)\big)^2}$$

$$\boldsymbol{= \int_{all\ x} x^2 f(x)dx - \left[\int_{all\ x} xf(x)dx\right]^2}$$

Properties of the variance

Properties of the variance:

- $Var(aX) = a^2 Var(X)$
- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$, where
$$Cov(X,Y) = E(XY) - E(X)E(Y)$$
- If X and Y are independent, then $Cov(X,Y) = 0$ and, as a consequence,
$$Var(X+Y) = Var(X) + Var(Y)$$
- If $Var(X) = 0$, then X is a number, not a random variable.

Uniform distribution

Uniform Distribution

A random variable X is said to be uniformly distributed if its density function is

$$f(x) = \frac{1}{b-a}$$

when x is between a and b, and it is 0, otherwise.

The expected value and the variance of the uniform distribution are given by

$$E(X) = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$$

Normal distribution

This is the most important continuous distribution.

- Many random variables can be properly modeled as normally distributed.
- Many distributions can be approximated by a normal distribution.
- The normal distribution is the cornerstone distribution of statistical inference.

Normal distribution

Normal distribution

A random variable with mean μ and variance σ^2 is normally distributed if its probability density function is given by

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

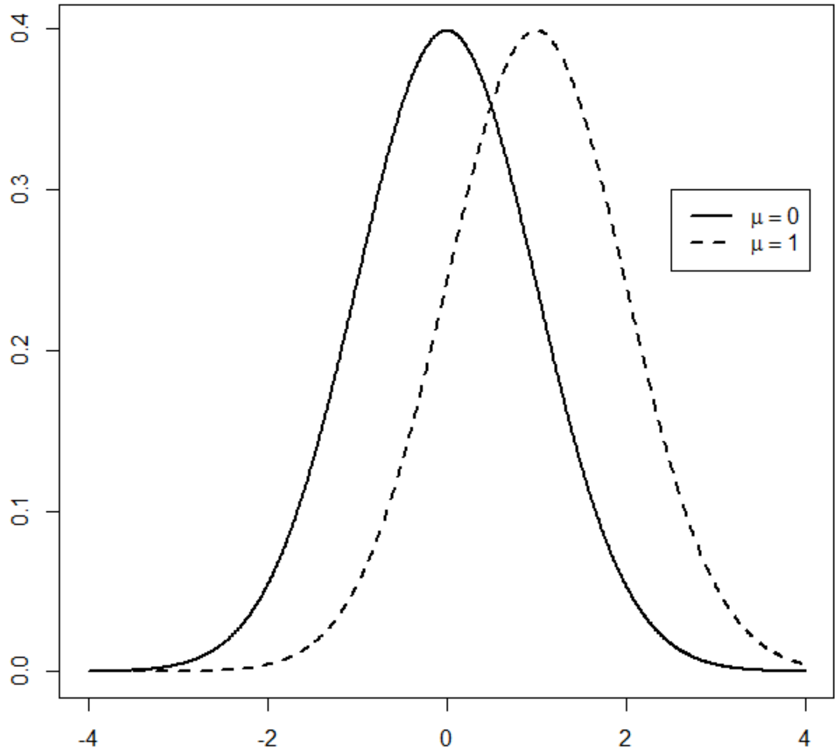
where $\pi = 3.14159$ and $e = 2.71828$.

Often instead of the phrase "*random variable X is normally distributed with mean μ and variance σ^2*" we will write

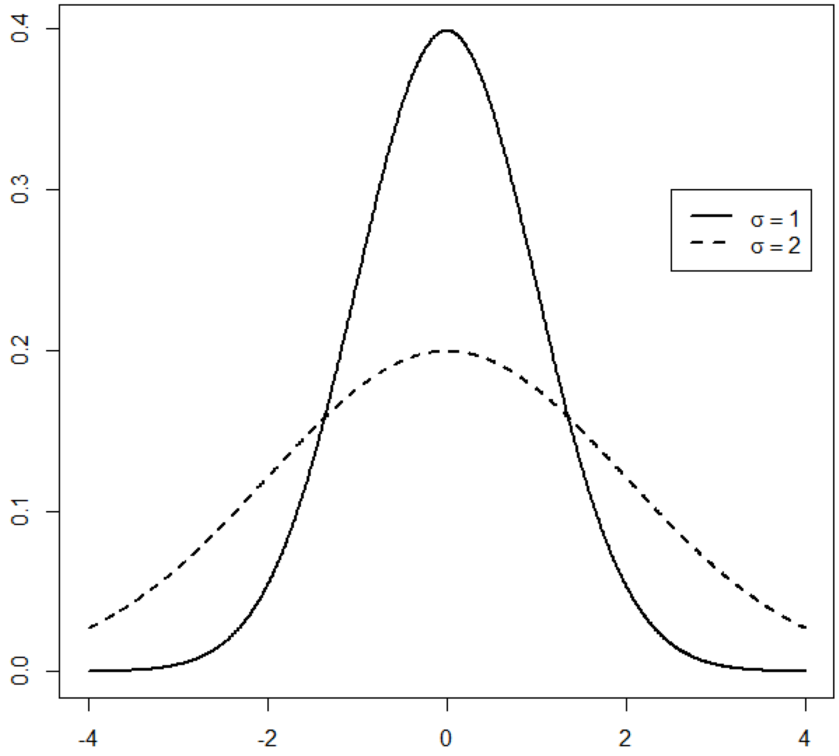
$$X \sim N(\mu, \sigma^2)$$

Normal random variable has a bell shaped distribution, symmetrical around mean μ.

How does μ affect the location of $f(x)$?



How does σ affect the shape of $f(x)$?



Functions of normal random variables

- If $X \sim N(\mu, \sigma^2)$, then $X - a \sim N(\mu - a, \sigma^2)$.

- If $X \sim N(\mu, \sigma^2)$, then a$X \sim N(a\mu, a^2\sigma^2)$.

- Finally, if X and Y are normally distributed *independent* random variables such that $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then

$$aX + bY \sim N\left(a\mu_X + b\mu_Y, a^2\,\sigma_X^2 + b^2\sigma_Y^2\right)$$

Finding normal probabilities

Two facts help calculate normal probabilities:

- The normal distribution is symmetrical about its mean.
- Any normal distribution with any mean and any variance (standard deviation) can be transformed into a normal distribution with mean 0 and variance 1. This distribution is called *standard normal distribution.* The random variables with this distribution will be denoted by Z.

Finding normal probabilities

- More specifically, any normal variable $\boldsymbol{X}$ with mean $\boldsymbol{\mu_X}$ and standard deviation $\boldsymbol{\sigma_X}$ can be transformed into the standard normal random variable:

$$\boldsymbol{Z = \frac{X - \mu_X}{\sigma_X}}$$

- By the properties mentioned above we have that $\boldsymbol{E(Z) = 0}$ and $\boldsymbol{Var(Z) = 1}$, that is, $\boldsymbol{Z \sim N(0, 1)}$. Therefore, once probabilities for $\boldsymbol{Z}$ are calculated (with help of calculus), probabilities related to any normal variable can found.
- The symmetry of the normal distribution makes it possible to calculate probabilities for negative values of the random variable $\boldsymbol{Z}$.

Finding normal probabilities: exercises

Exercise 1. Find the following probabilities.

1. $P(Z < 1.28)$
2. $P(Z < -1.96)$
3. $P(-1.96 < Z < 1.28)$
4. $P(Z > 1.28)$

Critical values of normal distribution

Given a number α such that $0 < \alpha < 1$, then *critical value of standard normal distribution*, denoted by z_{α}, represents the value for which the area under the standard normal curve to the right of z_{α} is equal to α, i.e. it satisfies the following equation:

$$P(Z > z_{\alpha}) = \alpha.$$

Finding critical values: exercises

Exercise 2. Find the following critical values.

1. $z_{.01}$
2. $z_{.05}$
3. $z_{.10}$

Normal distribution: exercises

Exercise 3. A diastolic blood pressure reading of less than 90 mm is considered normal; a reading of 90 or more indicates hypertension. Assume that the distribution of diastolic blood pressure readings of people 30 to 39 years old in the Framingham Heart Study is approximately normal with mean 79 and standard deviation 11. What proportion of this population has

1. Normal diastolic blood pressure?
2. Hypertensive diastolic readings?

Normal distribution: exercises

Exercise 4. The length of time it takes for a ferry to reach a summer resort from the mainland is approximately normally distributed with mean 120 minutes and standard deviation 12 minutes. Over many past trips, what proportion of times has the ferry reached the island in

1. Less than 105 minutes?
2. More than 125 minutes?
3. Between 110 minutes and 140 minutes?

Normal distribution: exercises

Exercise 5. Scores of males on the 1974 Mathematical Scholastic Aptitude Test (MSAT) were normally distributed with mean 500 and standard deviation 100.

1. What score indicates top 5%?
2. The middle 40% of the distribution is bounded by what two scores?
3. What percentage of scores is expected to be higher than 650?

Normal distribution: exercises

Exercise 6. Some auto companies design emission sensors so that they must be replaced after 100,000 miles. One such company found that the service life X (in months) of these sensors is approximately randomly normal with mean 48 months and standard deviation 9 months.

1. The company decided to guarantee the sensors for 3 years. What percentage of the sensors will not satisfy the guarantee?
2. The company decided to replace only 1% of all sensors. What should be the length (in months) of the guarantee?

Exponential distribution (optional)

The exponential distribution can be used to model

- the length of time between telephone calls
- the length of time between arrivals at a service station
- the life-time of electronic components

When the number of occurrences of an event follows the Poisson distribution, the time between occurrences follows the exponential distribution.

Exponential distribution (optional)

A random variable is exponentially distributed if its probability density function is given by

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

where λ is a parameter of the distribution, the average number of occurrences of the events. Also

$$E(X) = \frac{1}{\lambda}, Var(X) = \frac{1}{\lambda}, \quad \text{and } P(X > x) = e^{-\lambda x}$$