Continuous probability distributions

(STATS201.stats 201 20: Probability and probability distributions)

Anna Helga Jónsdóttir
Sigrún Helga Lund

December 15, 2012

Probability distributions of random variables

- Many random variables follow probability distribution of known types.
- The probability distributions of random variables are continuous if the random variables are continuous.
- We will look closely at the most common continuous probability distribution:
  - **Normal distribution**
- And also look at three continuous distributions that we will use for statistical inference:
  - t-distribution
  - χ^2-distribution
  - F-distribution

Probability of continuous random variables

For continuous random variables holds that

$$P(X = x) = 0.$$

This equation tells us that the probability of a continuous random variable receiving any specific value is always zero, no matter what the value is. Therefore it holds that

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

when X is continuous. Remember that this does not generally apply to discrete random variables!

Formulas for continuous random variables

The following rules hold for continuous random variables and are often useful.

$$P(X > a) = 1 - P(X < a)$$

Density function, distribution function and density curve

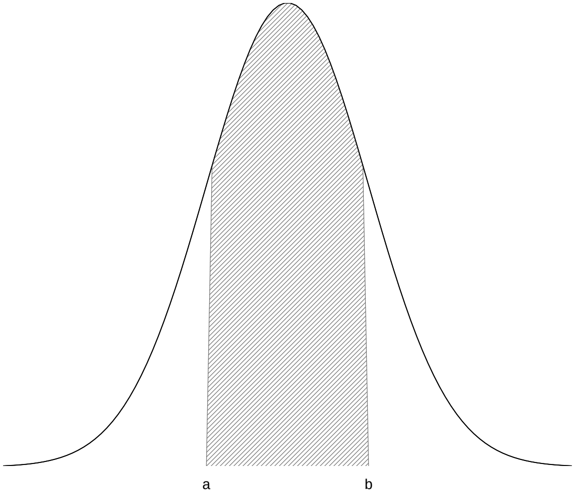
Density function, distribution function and density curve

Continuous distributions are described with a **density function**, denoted with $f(x)$. We use a so-called **distribution function**, which is the integral of the density function, to calculate the probability that a continuous random variable X receives a value that is less then a specific reference value x. The distribution function is denoted with $F(x)$ and can be written as

$$F(x) = P(X < x)$$

Density curve is described graphically with a **density curve**. The area under the density curve between two values a and b equals the probability that a random variable receives a value between a and b. The total area under the whole curve is always equal to 1.

Density function, distribution function and density curve



Normal distribution

- The Normal distribution is the most frequently used distribution within statistics.
- All sorts of phenomenas can be described with the normal distribution such as height, blood pressure, weight and so and so forth.
- We will also get to know the importanse of the normal distribution in lecture 090 when we study the central limit theorem.
- The density curve of the normal distribution is bell shaped and has two parameters that determine its shape.

Normal distribution

Normal distribution

The density function of the normal distribution is often denoted with $\phi(x)$ and may be written as

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

The function has two **parameters**, μ and σ. μ is the mean of the normal distribution and determines its location. σ^2 is the variance of the distribution and determines its spread. If a random variable X follows normal distribution with mean μ and variance σ^2 we write that $X \sim N(\mu, \sigma^2)$. The distribution function of the normal distribution is denoted with $\Phi(x)$.

Normal distribution

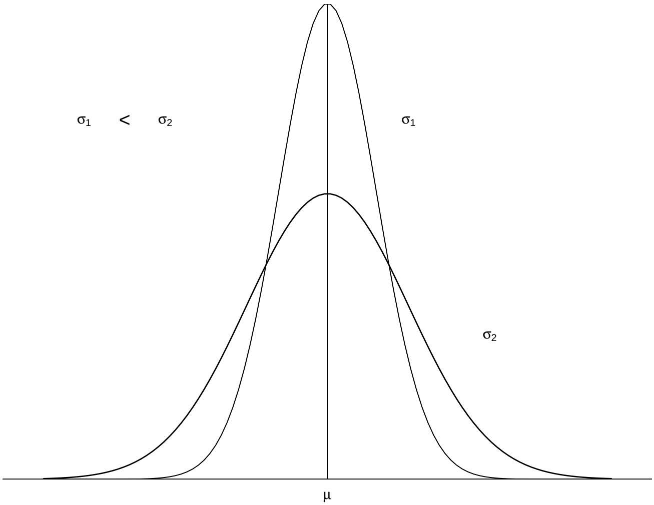


Figure: Two normal distributions with the same mean but different variances

Normal distribution

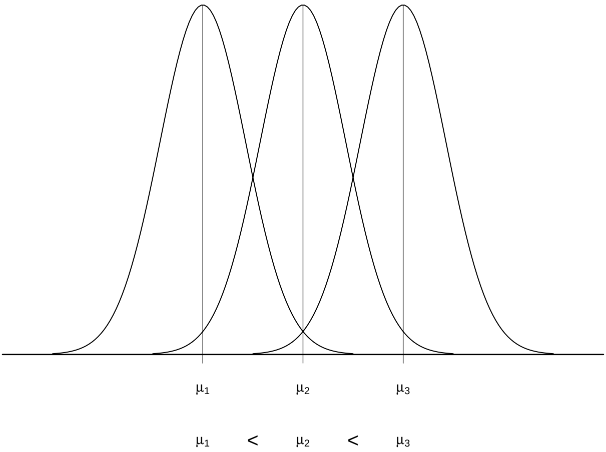


Figure: Three normal distributions with the same variance but different means.

The 68-95-99.7% rule

The 68-95-99.7% rule

For every normal distribution with mean μ and standard deviation σ holds that

- approx. 68% of the measurements will lie within one standard deviation from the mean
- approx. 95% of the measurements will lie within two standard deviations from the mean
- approx. 99.7% of the measurements will lie within three standard deviations from the mean

The 68-95-99.7% rule

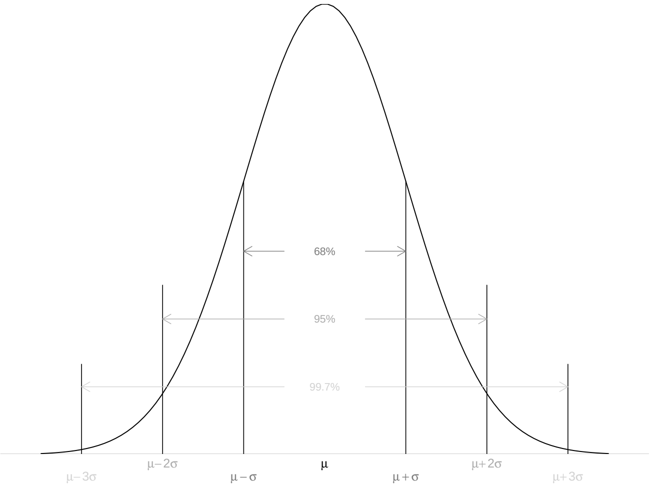


Figure: The 68-95-99.7% rule

Standardized normal distribution

Normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$ is called the standardized normal distribution.

Standardized normal distribution

If a random variable X follows normal distribution with mean μ, standard deviation σ and variance σ^2, written

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

then

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

follows a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$, written

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

Probability of normally distributed random variables

- The probability of normally distributed random variables can be calculated as the area under the density curve.
- If one wishes to find the probability that a normally distributed random variable lies on the interval from a to$b$ one need to integrate the density function from a to b. That is not done by hands, but with tables or computer software.

Before one can use the table, the normal distribution needs to be transformed to a standardized form. The table shows the probability

$$\Phi(z) = P(Z < z),$$

that is, it shows the probability that a random variable Z that follows standardized normal distribution will receive a value less then the number z, often called z-value. This can be though of as if the table is looking to the left.

Probability of normally distributed random variables

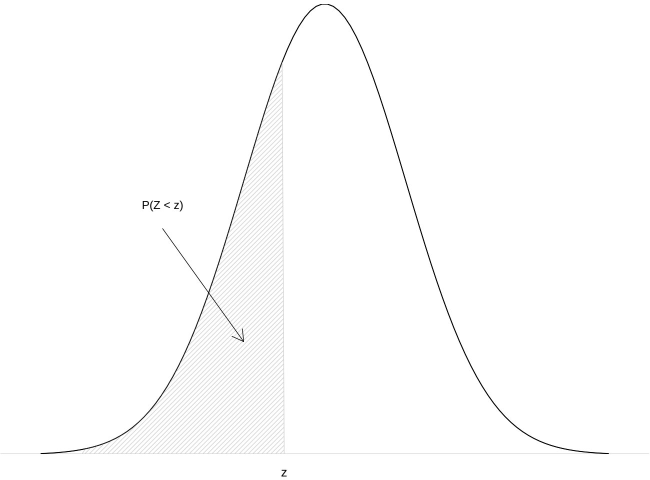


Figure: Normal distribution

Probability of normally distributed random variables

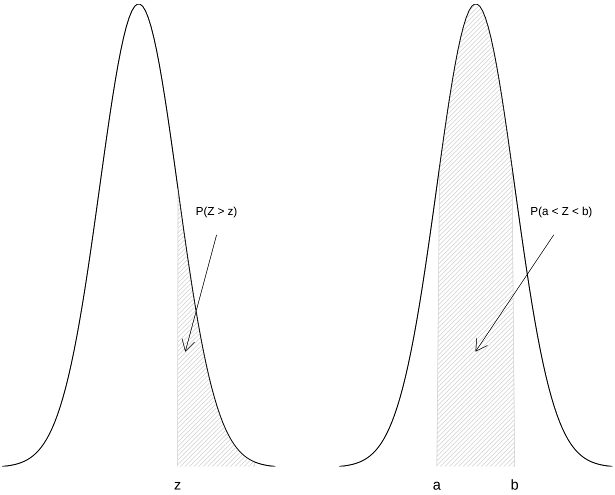


Figure: $P(Z > z)$ and $P(a < Z < b)$ where $Z \sim N(0, 1)$

Using the table for the standardized normal distribution

Using the table for finding probability

When finding the probability that belongs to a certain z-value:
If the value follows standardized normal distribution that value is the z-value itself. If it does not follow the standardized normal distribution we need to find a standardized z-value with

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

We find the z-value in the table (in bold) and the probability is the $\Phi(z)$ value on it's right side.

Using the table for the standardized normal distribution

Using the table for finding values

If we wish to find which z-value corresponds to a certain probability: We find the probability, or the probability closest to it, among the $\Phi(z)$-values in the table and the z-value is (in bold) on its left side.
If the value does not follow the standardized normal distribution we need to transform the z-value back to the original distribution, such that the value becomes

$$x = \mu + z\sigma$$

The relationship of X and Z

The relationship of X and Z

If a random variable X follows normal distribution with mean μ and variance σ^2 then

$$P(X \leq x) = P(Z \leq z)$$

Where $z = \frac{x-\mu}{\sigma}$ and Z follows the standardized normal distribution.

The syntax z_a

The syntax z_a

With z_a is denoted the z-value that is such that a random variable that follows standardized normal distribution has the probability a of receiving a values that is **less then** z_a. This can be written as:

$$a = P(Z < z_a).$$

where Z follows the standardized normal distribution. z_a is therefore the a-th percentile of the standardized normal distribution.

Normal probability plot

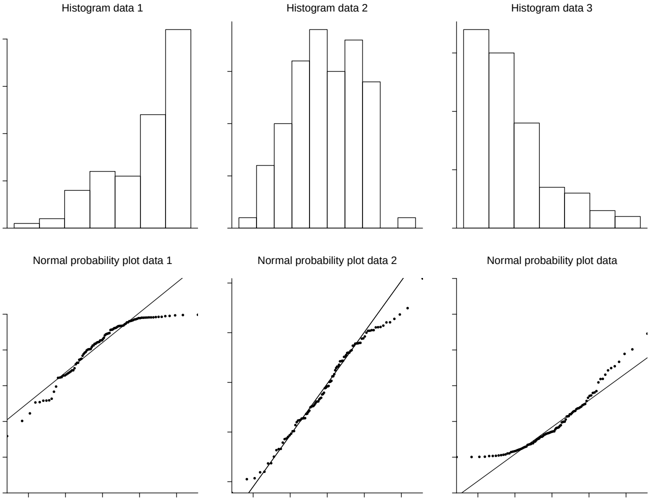
- Many statistical methods rely on that the measurements or some derived quantities follow a normal distribution.
- When applying these methods we need to ensure that this is the case.
- There are several methods to do so, a common one is the normal probability plot.
- The normal probability plot is a graphical method to investigate whether data follows a normal distribution or not.
- Normal probability plots are not drawn in statistical softwares, not by hand, but it is important to know how to interpret them.

Normal probability plot

Normal probability plot

If the points on the normal probability plot lie close the straight line that is shown on the plot and the end points on both sides do not bend critically up and/or down from the line then it is reasonable to assume that the data is normally distributed.

Normal probability plot



t-distribution

- The t-distribution, or the Student's t, is a continuous probability distribution that resembles the normal distribution.
- It is bell shape, symmetrical about the mean of the distribution, which is 0. The t-distribution is used for statistical inference.
- The t-distribution has one parameter, that is called the **degrees of freedom**. We use k to denote the number of degrees of freedom. A t-distribution with k degrees of freedom is denoted with $t_{(k)}$.

t-table

- We look up in the table after the number of degrees of freedom. The values we read are denoted with $t_{a,(k)}$.
- For $t_{a,(k)}$ holds that a random variable that follows t-distribution with k degrees of freedom has the probability a of receiving a values that is **less then** or equal to $t_{a,(k)}$.
- The column is determined by the a-value, but the line by the number of degrees of freedom.
- As the number of degrees of freedom grows, the more the t-distribution resembles the standardized normal distribution.

t-distribution

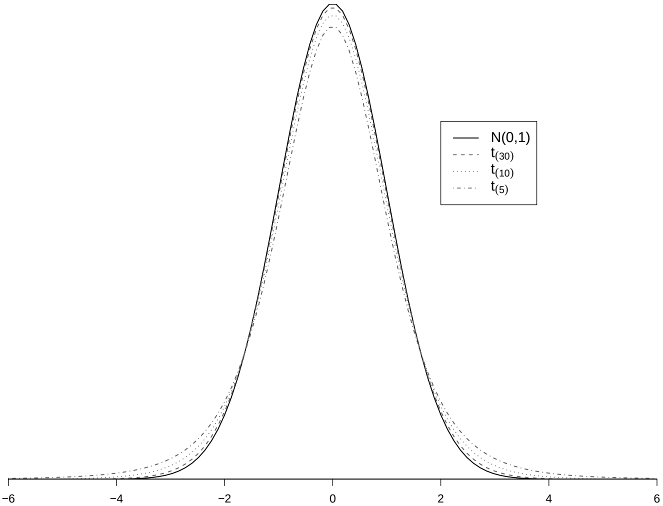


Figure: A few t-distributions

χ^2-distribution

- The χ^2-distribution, is a continuous probability distribution and commonly used for statistical inference.
- It is not symmetrical as the normal distribution.
- The χ^2-distribution has one parameter, the number of degrees of freedom, that is denoted with k.
- χ^2-distribution with k is denoted with $\chi^2_{(k)}$.
- The mean of the χ^2-distributions equals its number of degrees of freedom.

χ^2-table

- We look up in the table by the number of degrees of freedom. The values that we read from the table are denoted with $\chi^2_{a,(k)}$.
- For $\chi^2_{a,(k)}$ holds that a random variable that follows a χ^2-distribution with k degrees of freedom has the probability a of receiving a value that is **less then** $\chi^2_{a,(k)}$.
- We choose a column with the a-value and the line with the number of degrees of freedom.

χ^2-distribution

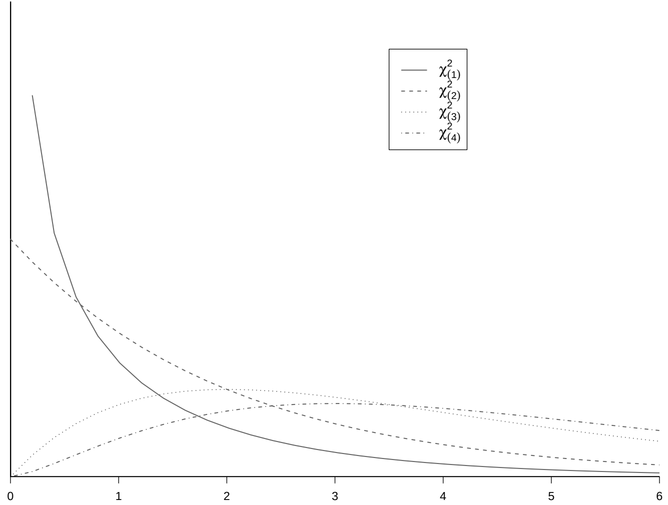


Figure: A few χ^2-distributions

F-distribution

- The F-distribution is a continuous probability distribution also commonly used for statistical inference.
- It is not symmetrical as the χ^2-distribution.
- The F-distribution has two parameters that are called the degrees of freedom and denoted with v_1 and v_2.
- F-distribution with v_1 and v_2 degrees of freedom is denoted with $F_{(v_1,v_2)}$.

F-table

- There are four tables for four different a-values, $a = 0.90$, $a = 0.95$, $a = 0.975$ and $a = 0.99$.
- The columns in the tables represent different values of v_1 and the lines different values of v_2.
- The values that are read from the table are denoted with $F_{a,(v_1,v_2)}$.
- For $F_{a,(v_1,v_2)}$ holds that a random variable that follows F-distribution with v_1 and v_2 degrees of freedom has the probability a of receiving a value that is **less then** $F_{a,(v_1,v_2)}$.

F-distribution

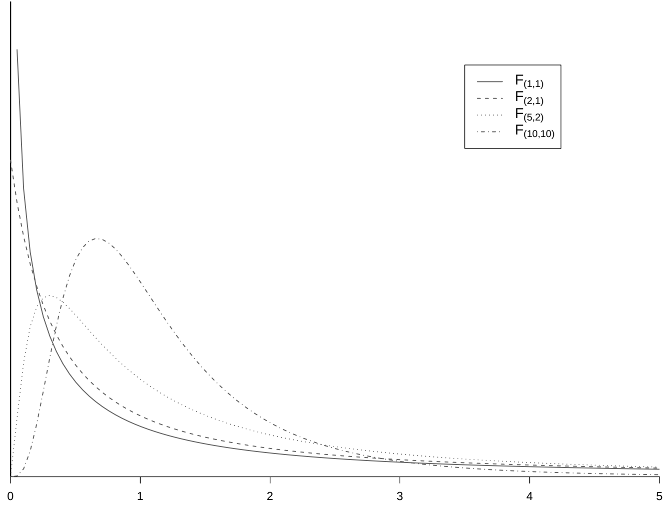


Figure: A few F-distributions