STATS201.stats 201 20 Probability and probability distributions

Anna Helga Jónsdóttir
Sigrún Helga Lund

December 16, 2012

Contents

1 Probability

1.1 Randomness

- Descriptive statistics describe the sample that we have obtained
- Statistical inference uses the sample to draw conclusions about the whole population.
- The variables that we measure are influenced by some randomness.
- We therefore look at every measurement as a random phenomena.
- In this lecture we look closer at random phenomena.

1.2 Events, outcomes and outcome space.

Outcome and outcome space
Every random phenomena has certain possible **outcomes**. The set of all possible outcomes is the **outcome space** an is denoted with Ω.

Event
An **Event** is a particular outcome or a set of particular outcomes of a random phenomena.

Which are the possible outcomes when a coin is tossed twice?

The possible outcomes are four, so Ω has four elements:

1. First heads, then tails
2. First heads, then heads
3. First tails, then tails
4. First tails, then heads

1.3 Disjoint events

> **Disjoint events**
> We say that events A and B are **disjoint** if they contain no common outcome.

If we toss a coin twice, are the following events disjoint?

A: To get at least one tail

B: To get no tails

Since we cannot have none and one tail at the same time the events are disjoint.

1.4 Union and intersection of events

> **Union of events**
> The **union** of events A and B is denoted $A \cup B$. It is the set of all outcomes that are in **either** A or B or **both of them**.

> **Intersection of events**
> The **Intersection** of events A and B is denoted $A \cap B$. It is the set of al outcomes that are in **both** A and B. If A and B are disjoint, then their intersection is **empty**.

Let A be the event: "outcome on the interval (-3,4)" and B be the event "outcome in the interval (2,8)". Fin the union and the intersection of the events A and B.

The union of A and B, $A \cup B$ is "outcome in the interval (-3,8)".

The intersection of A and B, $A \cap B$ is "outcome in the interval (2,4)".

1.5 The complement of an event

> **Complement of an event**
> The **Complement** of an event A is denoted A^C. It s the set of all outcomes in Ω that are **not it** A.

We toss a coin twice. Let A and B be the events:

A: To have at lest one tail.

B: To have two tails.

Find the union, intersection and the complement of the events.

The intersection of *A* and *B* is { „tail, tail“ }.

The union of *A* and *B* is { „tail, tail“, „tail, heads“, „heads, tails“ }.

The complement of *A* is {„heads, heads“}.

The complement of *B* is { „heads, heads“, „tails, heads“, „heads, tails“ }.

1.6 Probability

Probability
The **probability** of a certain outcome of a certain outcome of a random phenomena is the proportion of the cases when that the random phenomena gets that outcome when the phenomena is repeated often enough. This ratio can be at minimum **zero** and at maximum **one**.

Probability of an event
The **probability of an event** *A*, denoted $P(A)$, is the probability that the observed outcome will be in *A*.

1.7 Equally likely outcomes

Equally likely outcomes
Equally likely outcomes are only defined for random phenomena with **finite** Ω. Then the probability of every outcome in Ω is the same.

Probability of events when all outcomes are equally likely
If all of the outcomes of a random phenomena are **equally likely**, then the probability of an event *A* are:

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } \Omega}$$

What is the probability of getting one "tail" when two coins are flipped?

The event „to get exactly one tail“, has two outcomes. Ω has in total four equally likely outcomes the the probability is 2/4 or 50 %.

1.8 Formulas

Formulas

1. $P(\Omega) = 1$
2. $P(A^C) = 1 - P(A)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. If A and B are disjoint, $P(A \cup B) = P(A) + P(B)$

1.9 Conditional probability

Conditional probability
With $P(A|B)$ we denote the probability that event A occurs, given that event B has occurred. The probability of $P(A|B)$ can be calculated with

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) > 0.$$

Probability of intersection of events

$$P(A \cap B) = P(A|B)P(B), \quad \text{if } P(B) > 0.$$

1.10 Independent events

Independent events
We say that events A and B are **independent** if the probability that an event A occurs does not change even though the event B has occurred and vice versa.

Probability of independent events
If A and B are independent, then

$$P(A \cap B) = P(A) \cdot P(B)$$

Beta has 8 balls in a bag, three red and five white. She takes one ball and then returns it and then draws another ball.

1. What is the probability that she draws tw red balls?
2. What is the probability that the balls have different color?

1. Since Beta returns the first ball the the draws are independent. The probability of getting two red balls are:
$$\frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64}$$
approx. 14%.
2. Leta A be the event "first red, then white" and B be the event "first white, then red". Then the probability of getting balls with different color equal to $A \cup B$. Since A and B are disjoint we get:
$$P(A \cup B) = P(A) + P(B) = \frac{3}{8} \cdot \frac{5}{8} + \frac{3}{8} \cdot \frac{5}{8} = \frac{15}{64} + \frac{15}{64} = \frac{30}{64}$$
approx. 47 %.

2 Random variables

2.1 Random variable

Random variable
Random variable describes the outcome of a variable before it is measured

Syntax for random variables
Random variables are denoted with **capital** letters, often X

Values that a random variable **has received** are denoted with **lower-case** letters, often x

The same letter is always used for a random variable and the value it has received.

2.2 Discrete and continuous random variables

Discrete random variables
Discrete random variables describe discrete variables. They have a finite set of possible outcomes on every limited interval.

Continuous random variables
Continuous random variables describe continuous variables. They can obtain any outcome on some interval.

2.3 Syntax for the probability of random variables

Syntax for the probability of random variables

$P(X \leq a)$**:** Denotes the probability that the outcome of a random variable X will be **less or equal** then the value a.

$P(X \geq a)$**:** Denotes the probability that the outcome of a random variable X will be **greater or equal** then the value a.

$P(a \leq X \leq b)$**:** Denotes the probability that the outcome of a random variable X will be **between** $a$and b, both values included

$P(X = a)$**:** Denotes the probability that the outcome of a random variable X will be **exactly** the value a.

2.4 Probability distribution of random variables

Probability distribution random variables
The **Probability distribution** of a random variable is a rule that tells us which values a random variable can receive and furthermore:

$P(X = a)$ for all values a it can receive if the probability distribution is **discrete**.

$P(a \leq X \leq b)$ for all values a and b if the probability distribution is **continuous**.

The probability distribution of a random variable gives us all available information possible of the random variable!
Why do you think that we define the probability distribution in a different manner depending on whether the random variable is discrete or continuous?

2.5 Types of probability distributions

Types of probability distributions
The randomness of many of the variables that we investigate are similar by nature.

Then the random variables that they describe behave similarly.

As a consequence, they will have similar probability distributions.

Then we say that the probability distributions of the random variables are of the same type.

2.6 Parameter

Parameter
Every type of probability distribution is described with numbers that are called the **parameters** of the probability distribution.

Different parameters describe different probability distributions
.

Normally the parameters are only one or two.

If we know the type of the probability distribution of a random variable, the values of its parameters give all information available about the random variables.

2.7 Short summary

- One can talk about the probability that a random variable receives certain values.
- That probability is described by the probability distribution of the random variables, that give all information available about the random variables.
- Many random variables have probability distributions of a known type.
- Every type of probability distribution is described with numbers that are called parameters.
- To every type of probability distributions belong certain parameters and they are normally one or two.
- If we know the type of the probability distribution of a random variable, the values of its parameters give all information available about the random variables.

2.8 Independent random variables

Independent random variables
We say that two random variables are **independent** if the outcome of one random variable does not affect the outcome of the other random variable.

Dependent random variables
We say that two random variables are **dependent** if they are not independent, that is, if the outcome of one random variable does not affect the outcome of the other random variable or vice versa.

Independent and identically distributed random variables
We say that random variables $X_1, \ldots, X_n$ are **independent** if each of them is independent to all of the others and **identically distributed** if they all have the same probability distribution.

2.9 Expected value of a random variable

Expected value of a random variables
Expected value of a random variable is the **true** mean of the random variable. It is either denoted with μ or $E[X]$. It is also called **population mean** when appropriate.

Law of large numbers
As the number of measurements of a random variable X grows, the mean of the measurements, denoted $\bar{x}$, gets closer to the **expected value** of the random variable, denoted μ or $E[X]$.

2.10 Expected value of a discrete random variable

Expected value of a discrete random variable
If a random variable is **discrete** its expected value is the weighted mean of all of its possible outcomes, where the weight of each outcome is the probability that the random variable receives that outcome.

Formula for the expected value of a discrete random variable
If a random variable X is discrete, then its expected value is

$$\mu = \sum x_i \cdot P(X = x_i)$$

where we sum over all possible outcomes of the random variable.

What is the expected value of the random variable "the sum from tossing two dies"?

When we add the outcomes from tossing two dies the outcomes are not equally likely. The probabilities are

Value	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The expected value is:

$$\sum x_i \cdot P(X = x_i) =$$
$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36}$$
$$+ 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7.$$

2.11 Variance of random variables

Variance of random variables, *Var*[*X*]
As random variables have true means, they also have a **true variance**. It is either denoted with σ^2, or $Var[X]$. It is also called the **population variance** when appropriate.

Formula for the variance of a discrete random variable
The **variance** of a discrete random variable is

$$\sigma^2 = \sum (x_i - \mu)^2 \cdot P(X = x_i)$$

where we sum over all possible outcomes of the random variable.

What is the variance of the random variable "tossing a die"?

The expected value is equal to 3.5.

The variance is:

$$\sum (x_i - \mu)^2 \cdot P(X = x_i) = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} = 2.92$$

3 Discrete probability distributions

3.1 Probability distributions of random variables

- Many random variables have probability distributions of a known type.
- The probability distributions of random variables are discrete if the random variables are discrete and continuous if not.
- Let us look at the two most common discrete probability distributions:
  - **The binomial distribution**
  - **The Poisson distribution**
- We will see how these two probability distributions can be used to describe several random phenomena.

3.2 Mass function

Mass function
Discrete probability distributions are described with a **mass function** and we will use it to calculate that probability of certain outcomes of discrete random variables. We denote the mass function with $f(x)$ and it can be written as

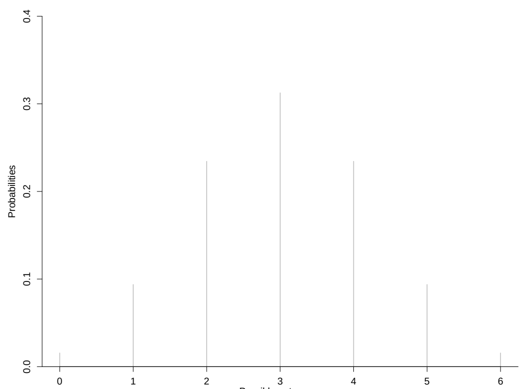
$$f(x) = P(X = x).$$

The following holds for the mass function:

$$f(x) \geq 0$$
$$\sum_{\text{yfir öll x}} f(x) = 1.$$

We use barplots to represent mass functions graphically.

3.3 Barplot of a mass function



3.4 Formulas for discrete random variables

Formulas fyrir discrete random variables
When calculating probabilities for a discrete random variable X calculations can often be simplified by "turning around" the probabilities

$$P(X \leq k) = 1 - P(X > k)$$
$$P(X < k) = 1 - P(X \geq k)$$
$$P(X \geq k) = 1 - P(X < k)$$
$$P(X > k) = 1 - P(X \leq k)$$

where k can be any number in the outcome space of X.

3.5 Bernoulli trial

Bernoulli trial
Every event in a group of repeated events is classified as a **Bernoulli trial** if the following holds:

1. Every event has only two possible outcomes. These outcomes are traditionally called **success** and **failure**. An event is successful if the outcome is a **success** and unsuccessful if its outcome is a **failure**.

2. The probability of a success are the same for every event. The probability of a failure is therefore the same for all events as the probability of a failure is always 1 minus the probability of a success.

3. An outcome in one event does not influence the outcome of another event, that is the events are independent.

3.6 The binomial distribution

- We are often interested in calculating how many successful events are among a set of Bernoulli trials.
- We would for example want to calculate the probability of receiving two sixes (which would be the success) when a dice is thrown five times.
- We view the total number of successful events as a random variable X.
- It has a known probability distribution that is called the **binomial distribution** and it is described with the parameters n which is the total number of Bernoulli trials that are conducted, and p which is the probability that is the probability of success within the Bernoulli trials.

3.7 The binomial distribution

The binomial distribution
Let the random variable X denote the number of successful events from n Bernoulli trials. Then X follows a binomial distribution with the parameters n and p, written $X \sim B(n,p)$, where p is the probability of success within each event.

The probability that the random variable X receives the value $k \in 0,1,2,\ldots n$ can be calculated with the mass function of the binomial distribution:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0,1,2,\ldots n$$

$\binom{n}{k}$ the binomial coefficient. It is the probability of receiving k positive outcomes in n events and calculated with

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Where $k! = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot (1)$. Notice that $0! = 1$.

Benni likes to toss coins. Calculate the probability that he will get exactly two "heads" when he tosses a coin four times.

We let X represent the number of heads. X follows a binomial distribution with $n = 4$, $p = 0.5$, $X \sim B(4, 0.5)$.

We start by finding the value of the binomaial coefficient:

$$\binom{n}{k} = \binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot (2 \cdot 1)} = 6$$

and then we use the mass function of the binomial distribution to calculate the probability:

$$P(X = 2) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{4}{2} 0.5^2 (1-0.5)^{4-2} = 6 \cdot 0.5^2 \cdot 0.5^2 = 0.3750.$$

The probability is 37.5 %.

3.8 The binomial distribution

- We have now seen that the probability that the random variable X receives a certain value k can be calculated.
- In addition to calculating $P(X = k)$ we are often interested in calculating
  - $P(a \leq X \leq b)$ or $P(a < X < b)$
  - $P(X \leq k))$ or $P(X < k)$
  - $P(X \geq k))$ or $P(X > k)$
- We can calculate all of those probabilities by using the formula for the mass function of a binomial distribution along with the rules on slide 3.4.

Helga throws a coin 10 times. We use X to represent the number of "heads", $X \sim B(10, 0.5)$.

a) What is the probability that Helga gets between 4 and 6 heads?

b) What is the probability that Helga gets 3 or less heads?

c) What is the probability that Helga gets 8 or more heads?

d) What is the probability that Helga gets more than 2 heads?

a) $P(4 \leq X \leq 6)$

We need to add the probabilities that she will geet 4, 5 and 6 heads or

$$P(4 \leq X \leq 6) = P(X = 4) + P(X = 5) + P(X = 6).$$

$$P(4 \leq X \leq 6) = 0.2051 + 0.2461 + 0.2051 = 0.6563.$$

b) $P(X \leq 3)$

We need to add the probability of getting 0, 1, 2 and 3 heads:

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

$$P(X \leq 3) = 0.0010 + 0.0098 + 0.0439 + 0.1172 = 0.1719.$$

c) $P(X \geq 8)$.

We need to ass the probability of getting 8, 9, or 10 heads:

$$P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10)$$

$$P(X \geq 8) = 0.0439 + 0.0098 + 0.0010 = 0.0547$$

d) $P(X > 2)$.

We can calculate the probability as:

$$P(X > 2) = P(X = 3) + P(X = 4) + ... + P(X = 10)$$

but it is easier to rewrite the probability and get:

$$P(X > 2) = 1 - P(X \leq 2) = 1 - (P(X = 2) + P(X = 1) + P(X = 0)).$$

$$P(X > 2) = 1 - (0.0439 + 0.0098 + 0.0010) = 0.9453.$$

3.9 Expected value and variance of a binomial distribution

Expected value and variance of a binomial distribution
If X follows a binomial distribution, $X \sim B(n, p)$ then

$$\mathrm{E}[X] = np$$
$$\mathrm{Var}[X] = np(1-p)$$

John is going to toss a die 900 times. We use X to denote the number of times a four comes up, $X \sim B(900, 1/6)$. Find $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

$$\begin{aligned} \mathrm{E}[X] &= np = 900 \cdot 1/6 = 150 \\ \mathrm{Var}[X] &= np(1-p) = 125. \end{aligned}$$

3.10 The Poisson distribution

- The Poisson distribution is often used to describe the number of random phenomena that occur within a **certain unit** but the number of possible outcomes has no upper limit.
- The units can be **time intervals**, **spatial intervals** or some **physical object**.
- As an example we can mention the number of phone calls an office receives every minute, the number of reindeers per each square kilometer or the number of typos on each page.

3.11 The Poisson distribution

The Poisson distribution
The Poisson distribution has one parameter that is called λ. If X follows a Poisson distribution with the parameter λ the probability that the random variable X receives a value k, $k = 0, 1, 2, \ldots$ with the mass function of the Poisson distribution:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

We write $X \sim \text{Pois}(\lambda)$. The sample space of X is $\Omega = \{0,1,2,...\}$. The parameter λ is the expected value of the random variable X, that is, its true mean. It describe how many successful outcomes we expect on average per each unit.

3.12 The Poisson distribution

- We have now seen that the probability that a random variable X that follows the Poisson distribution receives a certain value k can be calculated with the mass function of the Poisson distribution.
- We are often interested in calculating other probabilities:
  - $P(a \leq X \leq b)$ or $P(a < X < b)$
  - $P(X \leq k)$ or $P(X < k)$
  - $P(X \geq k)$ or $P(X > k)$.

We can calculate all of these probabilities with the mass function of the Poisson distribution.

3.13 Changing units

- When calculating the probability that a random variable that follows a Poisson distribution receives a certain value, we often have given the value of λ in another unit then the one we wish to use.
- We could for example know that the number of car incidents in Reykjavik every week, but we wish to know the number of incidents per day. Then λ need to be adjusted to a new unit.
- If the new unit is a times the old unit, then

$$\lambda_{\text{new}} = a \cdot \lambda_{\text{old}}$$

Where λ_{old} is the "old λ" and λ_{new} is the "new λ" adjusted to a new unit.

Anna is in a hurry and wonders how long time she has to wait in line in a supermarket. Average number of customers to get help there are 1.5 per minute.

Find the probability that:

a) 3 customers arrive at the resister in one minute.

b) 4 customers arrive at the resister in two minutes.

c) No more than 2 customers arrive at the register in one minute.

d) At least 1 customer arrives at the register in one minute.

a) We know that the average number of customers in one minute is 1.5, so $\lambda = 1.5$.

$$P(X=3) = \frac{e^{-1.5}1.5^3}{3!} = 0.1255.$$

b) We know that the average number of customers is per minute is 1.5 = λ. We need the average number of customers in two minutes. Since the average number of customers in one minute is 1.5 we expect on average $2 \cdot 1.5 = 3$ customers to arrive in two minutes.We use $\lambda = 3$ and get:

$$P(X=4) = \frac{e^{-3}3^4}{4!} = 0.1680$$

c)

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= 0.2231 + 0.3347 + 0.2510 \\ &= 0.8088. \end{aligned}$$

d)

$$P(X \geq 1).$$

We need to rewrite the probability:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) \\ &= 1 - P(X=0) \\ &= 1 - 0.2231 \\ &= 0.7769. \end{aligned}$$

3.14 The expected value and variance of a Poisson distribution

Expected value and variance of a Poisson distribution
If X follows a Poisson distribution, $X \in \text{Pois}(\lambda)$ then

$$\text{E}[X] = \lambda$$
$$\text{Var}[X] = \lambda.$$

Let X denote a random variable that follows a Poisson distribution with $\lambda = 2$. Find the expected value and variance of X.

$$\begin{aligned} \text{E}[X] &= \lambda = 2 \\ \text{Var}[X] &= \lambda = 2. \end{aligned}$$

4 Continuous probability distributions

4.1 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

4.2 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.

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

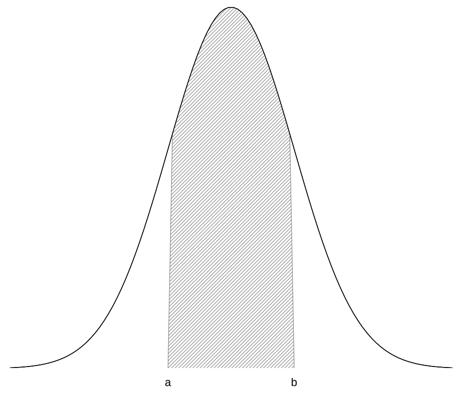
$$P(a < X < b) = P(X < b) - P(X < a).$$

4.3 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.



4.4 Density function, distribution function and density curve

4.5 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.

4.6 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 $\Phi(x)$.

4.7 Normal distribution

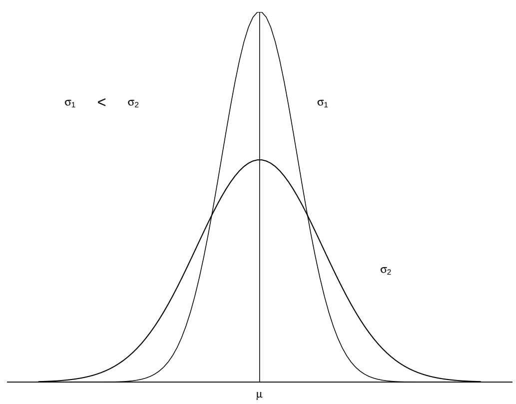


Figure 4: Two normal distributions with the same mean but different variances

4.8 Normal distribution

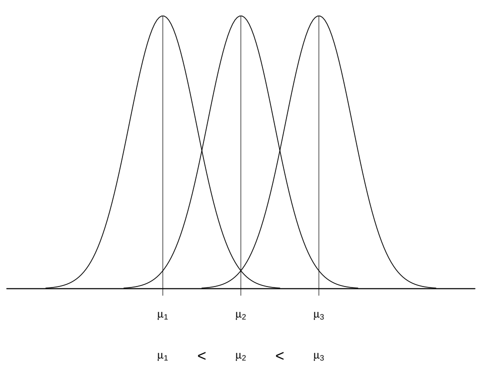


Figure 5: Three normal distributions with the same variance but different means.

4.9 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

4.10 The 68-95-99.7% rule

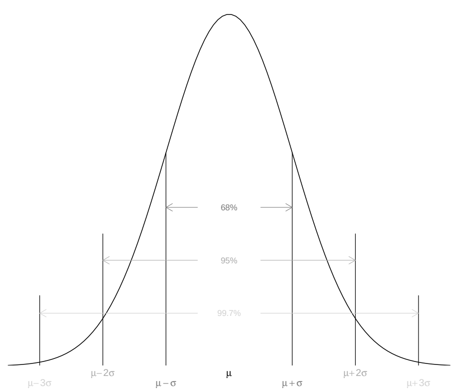


Figure 6: The 68-95-99.7% rule

4.11 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

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

4.12 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 distributionwill 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.

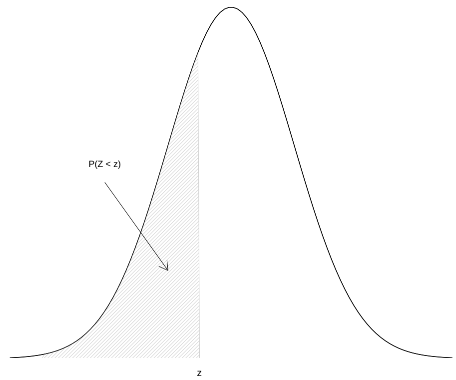


Figure 7: Normal distribution

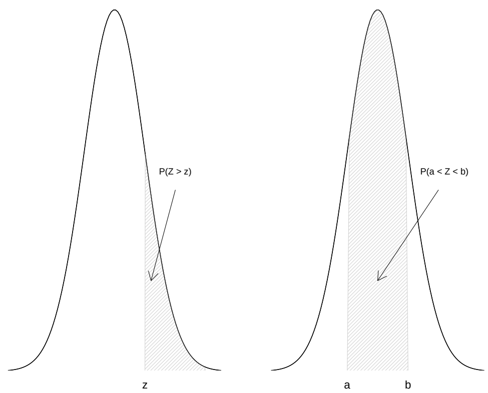


Figure 8: $P(Z > z)$ and $P(a < Z < b)$ where $Z \sim N(0,1)$

4.13 Probability of normally distributed random variables

4.14 Probability of normally distributed random variables

4.15 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.

Let us assume that the test scores of students in US on the SAT test follow normal distribution with mean 1026 and standard deviation 209. We use X to denote the test scores, $X \sim N(1026, 209^2)$.

a) Find the probability that student chosen at random has lower grade than 720, that is $P(X \leq 720)$.

b) Calculate the probability that a student, chosen at random, has higher score that 820, that is $P(X \geq 820)$.

c) Calculate the probability that a randomly chosen student has a grade on the interval 720 og 820, that is $P(720 \leq X \leq 820)$.

a)

$$P(X < 720).$$

We need to starty by standardizing:

$$\frac{720 - 1026}{209} = -1.46$$

$$P(X < 720) = P(Z < -1.46)$$

We use the normal distribution table and get 0.0721, that is approx. 7 %.

b)

$$P(X > 820).$$

We need to start by standardising:

$$\frac{820 - 1026}{209} = -0.99$$

$$P(X > 820) = P(Z > -0.99) = 1 - P(Z < -0.99)$$

We use the normal distribution table and get 1-0.1611 = 0.8389, that is approx. 84%.

c)

$$P(720 < X < 820).$$

We use the standardised values from a) and b)

$$P(720 < X < 820) = P(-1.46 < Z < -0.99) = P(Z < -0.99) - P(Z < -1.46)$$

We use the normal distribution table and get:

$$0.1611 - 0.0721 = 0.089$$

that is approx. 9%.

4.16 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

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

Which does a student need to get at minimum in order to be in the top 10% of the students?

Now we need the z-value. Remember that the table looks to the left so we need to find which z-value corresponds to 90% (then there are 10% above). We see that the z-value is 1.28.

The grades do not follow that standard normal distribution so we need to transform the z-value:

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

so

$$\text{Minimum grade} = 1026 + 1.28 \cdot 209 = 1292.5 \text{ points.}$$

4.17 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.

4.18 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.

Notið töflu staðlaðrar normaldreifingar til að finna $z_{0.95}$.

Hér þekkjum við líkurnar en okkur vantar z-gildið. Við finnum því 0.95 meðal Φ-gildanna í töflunni og lesum z-gildið við hlið þess. Í töflunni má sjá að $z_{0.95} = 1.64$.

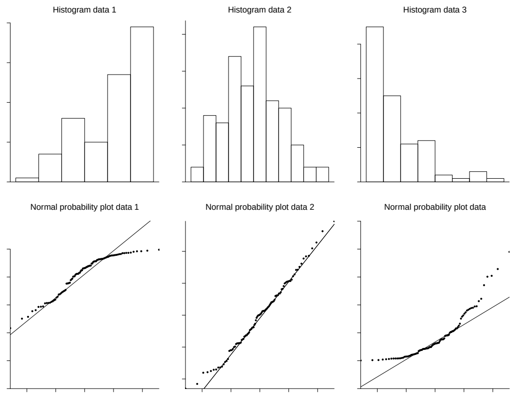
4.19 Normal probability plot

- 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.

4.20 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.

4.21 Normal probability plot



4.22 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 $t_{(k)}$.

4.23 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.

Find $t_{0.95,(17)}$.

We use a t-table with $a = 0.95$ (column) and $k = 17$ (line) and get that $t_{0.95,(17)} = 1.740$.

4.24 t-distribution

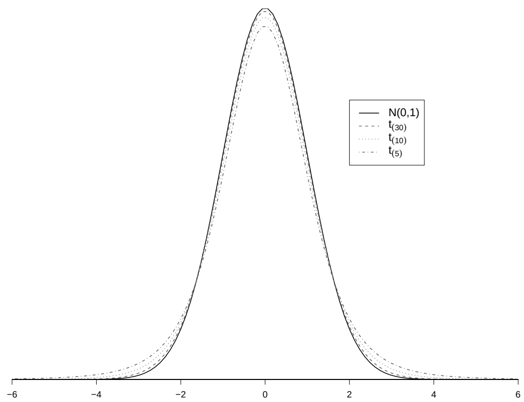


Figure 9: A few t-distributions

4.25 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.

4.26 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.

Find $\chi^2_{0.95,(4)}$.

We use χ^2-table. We chose $a = 0.95$ (column) and $k = 4$ (line) and get that $\chi^2_{0.95,(4)} = 9.488$.

4.27 χ^2-distribution

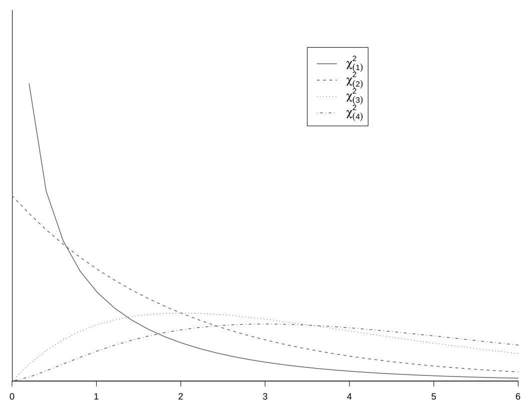


Figure 10: A few χ^2-distributions

4.28 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 $F_{(v_1,v_2)}$.

4.29 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** $F_{a,(v_1,v_2)}$.

Find $F_{0.95,(7,12)}$.

We use the F-table where $a = 0.95$. We chose a column $v_1 = 7$ and line $v_2 = 12$ and get $F_{0.95,(7,12)} = 2.913$.

4.30 F-distribution

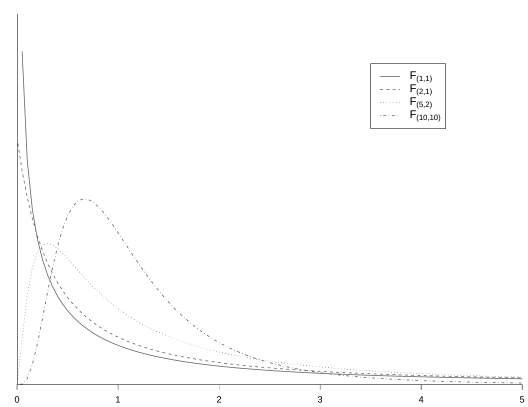


Figure 11: A few F-distributions

5 A statistic

5.1 Statistics

Statistic
A **statistic** is a number that is calculated by some method from our data.

- We look at our measurements as random variables because the outcome can change each time the experiment is repeated.
- Statistics are calculated from our measurements.
- If the outcomes change, the statistics can also change!
- That means that statistics are in fact random variables!

5.2 Sampling distribution

Sampling distribution
Every statistic is a random variable and has therefore some probability distribution. That distribution is called the **sampling distribution** of the statistic.

The sampling distribution depends on

- The **probability distribution of the measurements** that the statistic is calculated for.
- The **number of measurements**.

When certain criteria are fulfilled the sampling distribution of some statistics follow certain known types. Statistical inference normally relies on that fact.

5.3 Example

Let X_1 and X_2 be random variables that describe the outcome when a dice is thrown.

Below are shown all possible outcomes of the statistic $X_1 + X_2$.

					(6,1)					
				(5,1)	(5,2)	(6,2)				
			(4,1)	(4,2)	(4,3)	(5,3)	(6,3)			
		(3,1)	(3,2)	(3,3)	(3,4)	(4,4)	(5,4)	(6,4)		
	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)	
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)
2	3	4	5	6	7	8	9	10	11	12

5.4 Example

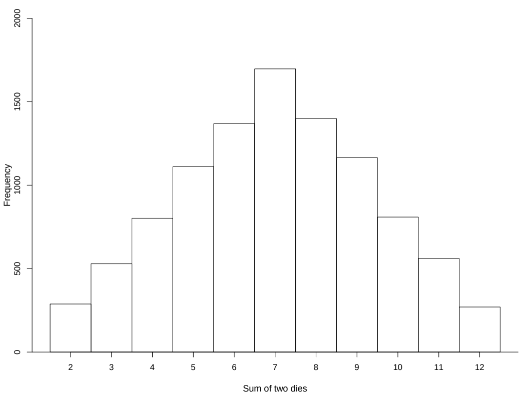


Figure 12: Simulation of 10000 throws of two dice.

5.5 Expected value of the sum of two random variables

Formulas for the expected value of random variables
If X and Y are two random variables, then

$$E[X+Y] = E[X]+E[Y]$$
$$E[X-Y] = E[X]-E[Y]$$

What is the expected value of the sum of two die tosses?

The expected value when tossing one die is 3.5. Let X and Y be random variables describing one die toss, then

$$E[X+Y] = E[X]+E[Y] = 3.5+3.5 = 7$$

5.6 Variance of the sum of two random variables

Formulas for the variance of random variables
If X and Y are two independent random variables then

$$Var[X+Y] = Var[X] + Var[Y]$$
$$Var[X-Y] = Var[X] + Var[Y]$$

What is the variance of the result of tossing two dies?

We use X and Y to represent tossing a single die. The variance of X and Y is 2.92 (calculated in the lecture on random variables).

$$Var[X+Y] = Var[X] + Var[Y] = 2.92 + 2.92 = 5.84,$$

5.7 Expected value and variance of the mean

Expected value and variance of the mean
If $X_1, \ldots, X_n$ are independent and identically distributed random variables with expected value $E[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$, then the following holds for the mean of them, denoted $\bar{X}$:

$$E[\bar{X}] = \mu$$
$$Var[\bar{X}] = \frac{\sigma^2}{n}$$

5.8 Standard error

Standard error
If $\bar{X}$ is the mean of $X_1, \ldots, X_n$, independent and identically distributed random variables with variance $Var[X_i] = \sigma^2$, then their **standard error** is

$$\sigma/\sqrt{n}$$

It is the standard deviation of the mean of the measurements.

5.9 The probability distribution of the mean of normally distributed random variables

The probability distribution of the mean of normally distributed random variables
If $X_1, \ldots, X_n$ are normally distributed random variables with mean μ and variance σ^2, then $\bar{X}$ follows also a normal distribution, with mean μ and variance σ^2/n.

That is if $X_i \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N(\mu, \sigma^2/n)$.

Grades of students on a standardized test follow a normal distribution with $\mu = 5$ and $\sigma = 2$. What is the probability distribution of the average of 10 randomly chosen students?

The average grade of 10 students follows a normal distribution with $\mu = 5$ and $\sigma^2/n = 4/10 = 0.4$.

5.10 Central limit theorem

Central limit theorem
If $X_1, \ldots, X_n$ are independent and identically distributed variables then $\bar{X}$ follows normal distribution with mean μ and variance σ^2/n

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

if n is large enaugh.

Notice that we do not need to know the probability distribution of the measurements!

5.11 Central limit theorem

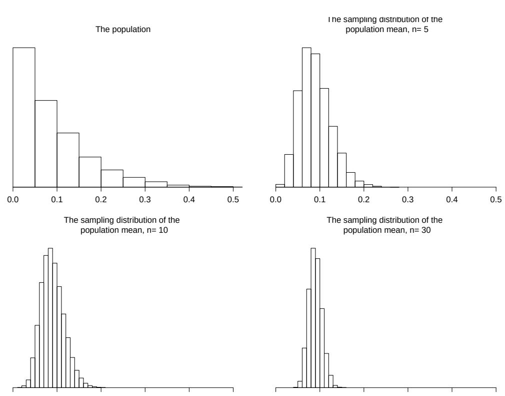


Figure 13: The sampling distribution of a mean when the random variables follow a very skewed distribution.

5.12 Central limit theorem

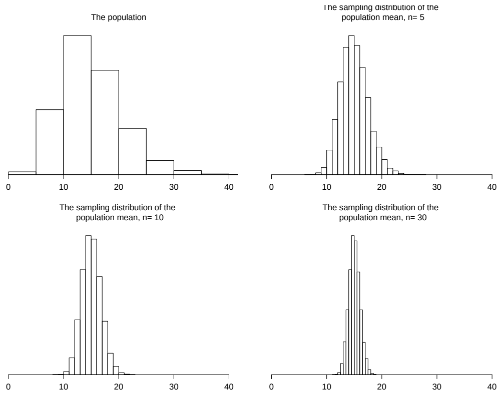


Figure 14: The sampling distribution of a mean when the random variables follow a slightly skewed distribution.

5.13 Estimators and test statistics

There are two groups of important statistics.

- **Estimators** estimate the parameters of the probability distribution that the random variables follow.
  Example: Estimator that estimates μ when the measurements are normally distribution.
  Example: Estimator that estimates p when the measurements are binomially distributed.

- **Test statistics** allow us to make statistical inference.
  Example: Test statistic that allows us to infer whether the variance of two population is the same.
  Example: Test statistic that allows us to infer whether the mean of a population differs from 20.