Mathematical Concepts (G6012)

Lecture 21

Thomas Nowotny
Chichester I, Room CI-105
Office hours: Tuesdays 15:00 - 16:45
T.Nowotny@sussex.ac.uk

Exam in January

- Fri 08 January, 09:30 SPORTCENTRE
  1 Hr 30 Mins

- Previous exams can be found here:

http://www.sussex.ac.uk/students/pastexams/search

MEQ

- The module evaluation questionnaires are open
- You can find them in your module resources on Sussex Direct
- I do read and consider every comment (but they are anonymous)

Lecture content for Thursday

- Two alternatives:
  - 1 hour introduction to Information Theory
  - Start with revisions
- I will ask for a show of hands at the end of today’s lecture

Last time: Hypothesis test

In hypothesis testing you set a "significance niveau" α , e.g. $\alpha = 0.05 = 5\%$

Then you calculate the probability P of your observation or **a more extreme one** if the hypothesis were true.

If your observation lies in the region of extreme observations (with respect to α) you reject the hypothesis.

Example chess game:

Chess game statistical test

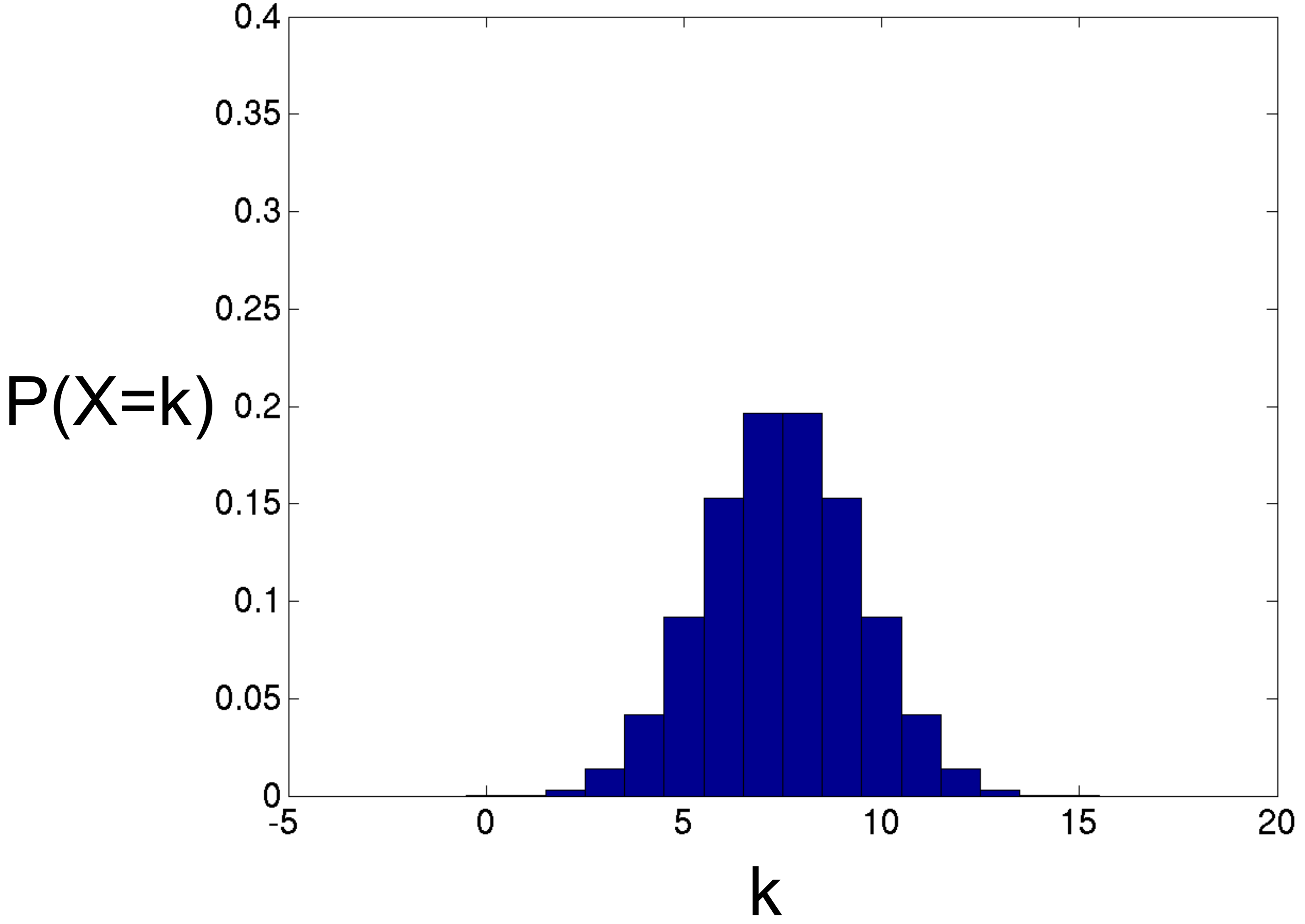
Probability space: $\Omega = \{0, 1\}$

$$P(\{0\}) = P(\{1\}) = \frac{1}{2}$$

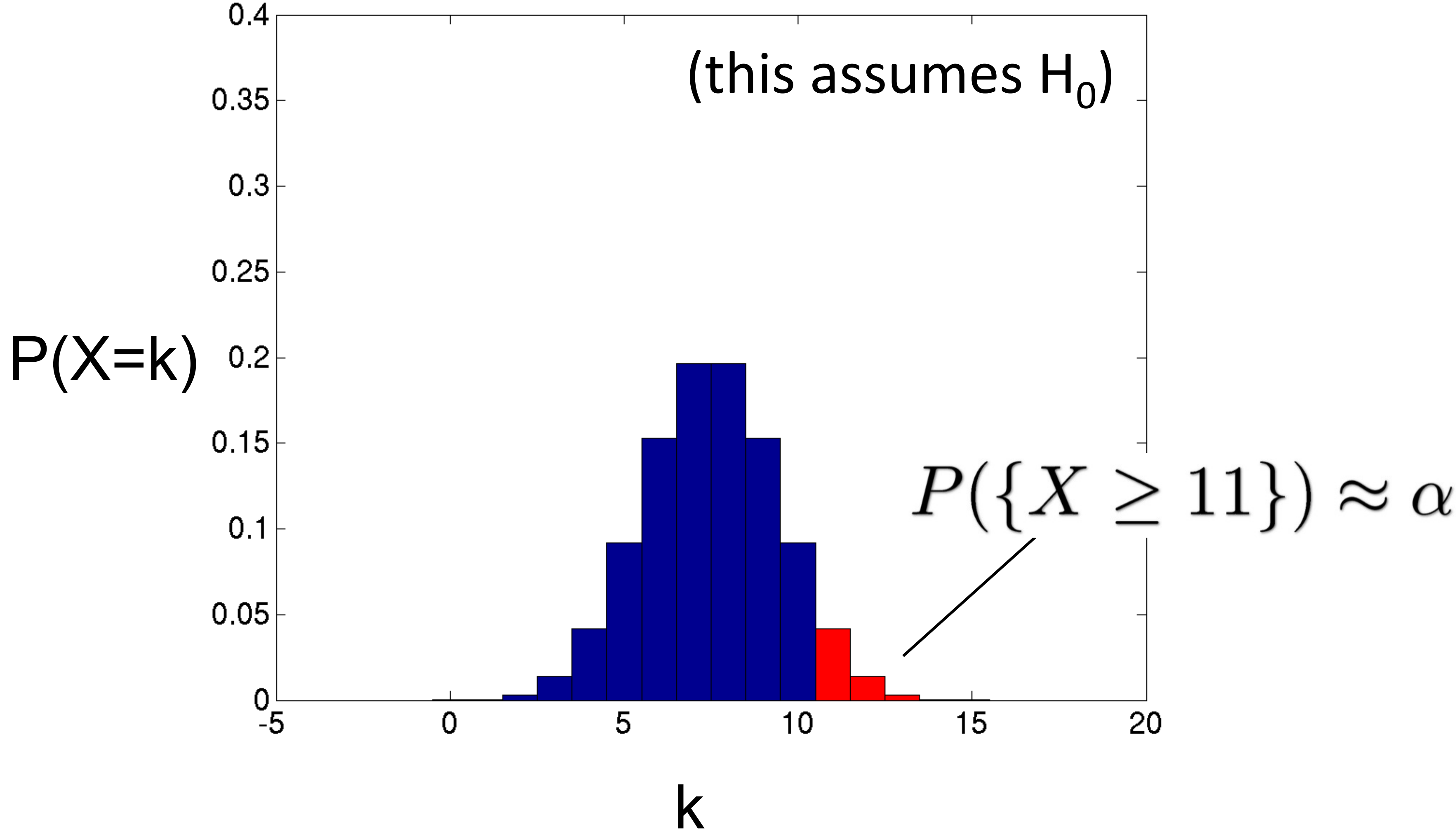
(This is the null hypothesis H_0)

We can calculate the probability for k wins from this:

Probability for k wins



Tail probability



Summary: Statistical tests so far

- Typically we test a **null-Hypothesis H_0**, typically about a property of the underlying probability space, e.g. **p**
- There is a **test statistic X** (some function of an observation) and a corresponding **probability distribution P(X=k)**
- We calculate the probability of an observed value **x** of the test statistic under the assumption that the null-Hypothesis is true
- Based on the this probability we reject the null-Hypothesis or "**do not reject**" it.
- There is no such thing as "accepting the null- Hypothesis"

Possible errors

- There are two possible errors we can make:

1. Errors of the first kind: The null-Hypothesis was true but we reject it accidentally. The probability for this type of error is limited by the significance level α

Possible errors

2. Errors of the second kind: The null-Hypothesis is false but we cannot reject it. The probability of this type of error is not controlled in basic test design and may be large*.
   This is one of the reasons why not rejecting the null-Hypothesis should not be interpreted as accepting it.
   This error is sometimes denoted β

* often it cannot be calculated (!)

Reporting significance (P-value)

- In the past the distributions of test statistics were taken from tables. To avoid imprecisions from this, scientists only reported $P < \alpha$ for a significance level α that was chosen up front.
- Nowadays, all distributions of test statistics can be calculated numerically to any precision – it is ok now to report observed P-values directly, e.g. $P=0.015$.

Testing two alternatives

- Example: Test for an infection
  - For infected patients, there is a $p_{w1}=0.15$ probability to observe a white blood cell and $p_{r1}=0.85$ probability to see a red one when examining cells in a blood sample
  - For healthy patients, these are $p_{w2}=0.1$ for white and $p_{r2}= 0.9$.
- Here, we want to test two Hypotheses against each other (H_0 (infected): $p_{w1}=0.15$ against H_1 (healthy): $p_{w2}=0.1$)

Testing two alternatives

- When counting n cells from an infected sample:

$$P(X_w = k) = b_{n,p_{w1}}(k)$$

- Let's use significance level $\alpha = 0.05$, and count n=100 cells.
- We calculate

k	$P(X_w \leq k)$
6	0.0047
7	0.0122
8	0.0275
9	0.0551
10	0.0994

Reject when x_w is 8 or smaller

Error of the second kind

- This is the error of a "false positive", i.e. not rejecting H_0 (infected), even though the subject is healthy

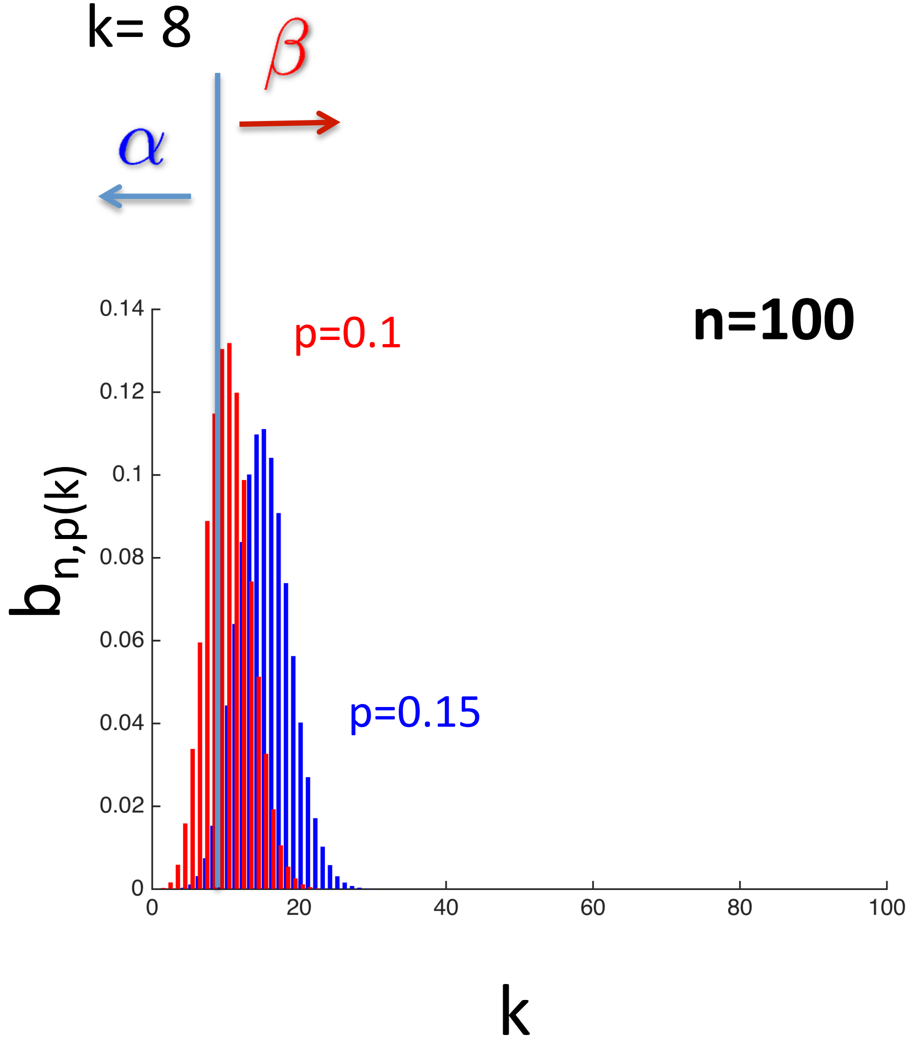
$$P(X'_w \geq 9) = \sum_{k=9}^{n} b_{n,p_{w2}}(k) \approx 0.6791 = 67.91\%$$

i.e. we would scare 67.9% of healthy people and send them on for further testing!

Visualisation

Decision boundary

k= 8



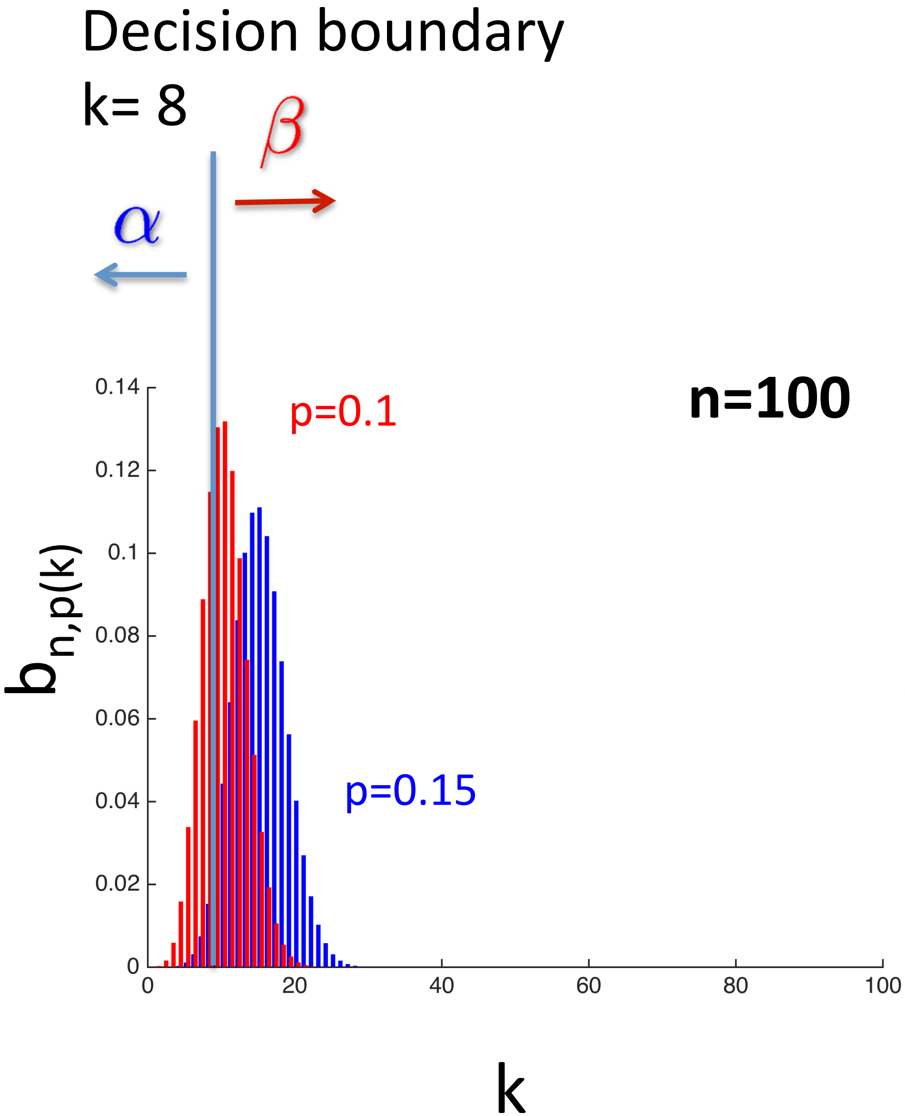
Increased sample size

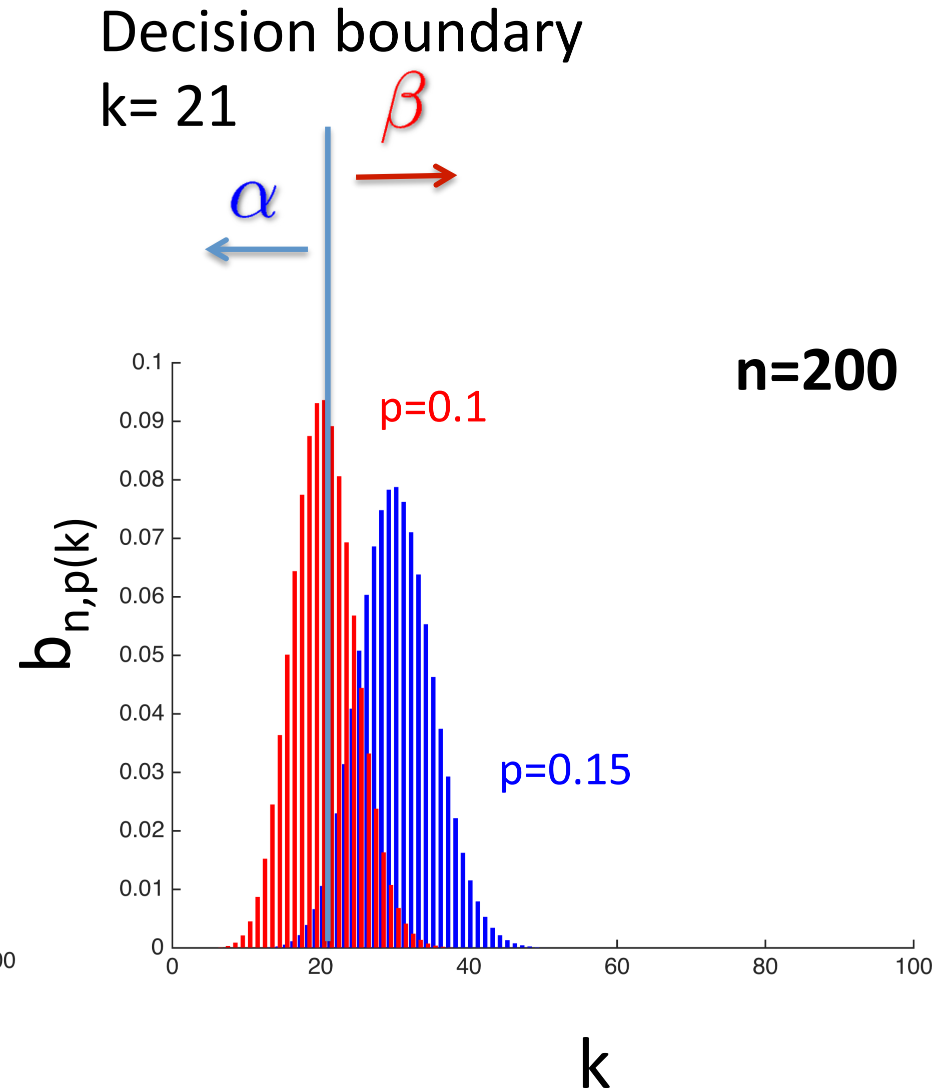
- If we do the same analysis but count 200 cells:
  - Reject for 21 or less white blood cells
  - Error of the second kind: 35.2%

- So, there is a trade-off between the errors for each given test size.

- The test is called “sensitive” or “powerful” if the error of the second kind is also small

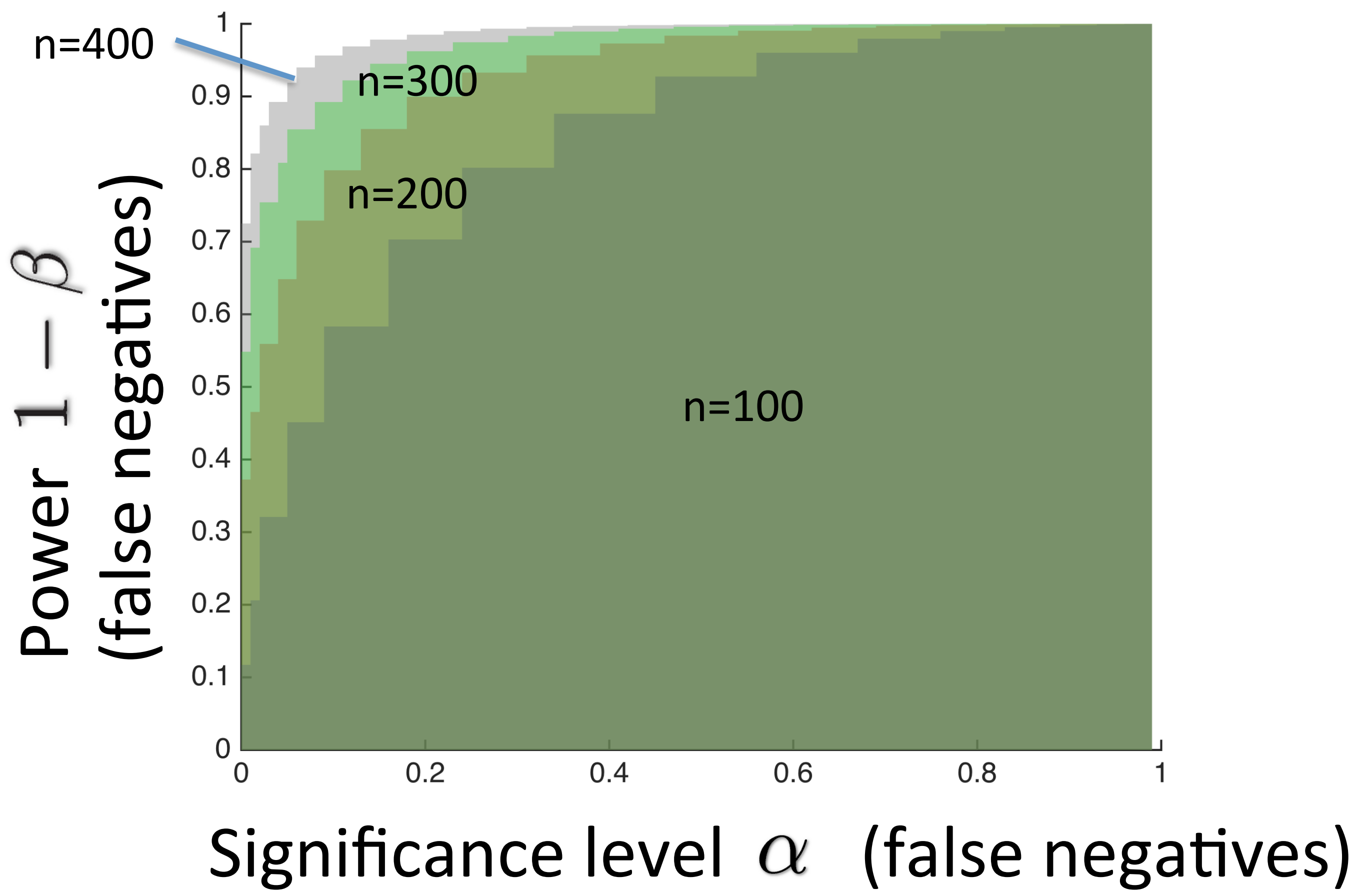
Visualisation





Area under the curve (AUC)

- The area under the α - β curve can be used as an indicator of the quality of a test:



So far: Binomial distribution

- Is correct for many applications:
  - Any Bernoulli processes (constant probability for success/failure, independent trials)
  - E.g.: games, gambling, many medical tests, elections (certain aspects), …
- However, it’s not easy to manipulate in practice
- For many other applications, the exact probability distribution is not known, e.g. repeated measurements of unknown quantities

Central limit Theorem

- For independent, identically distributed (i.i.d) random variables X_i with expectation μ and variance σ^2 the distribution of

$$s_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges to

$$\mathcal{N}(\mu, \sigma^2/n) = \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2/n}\Big)$$

(independent of the probability distribution of X_i)

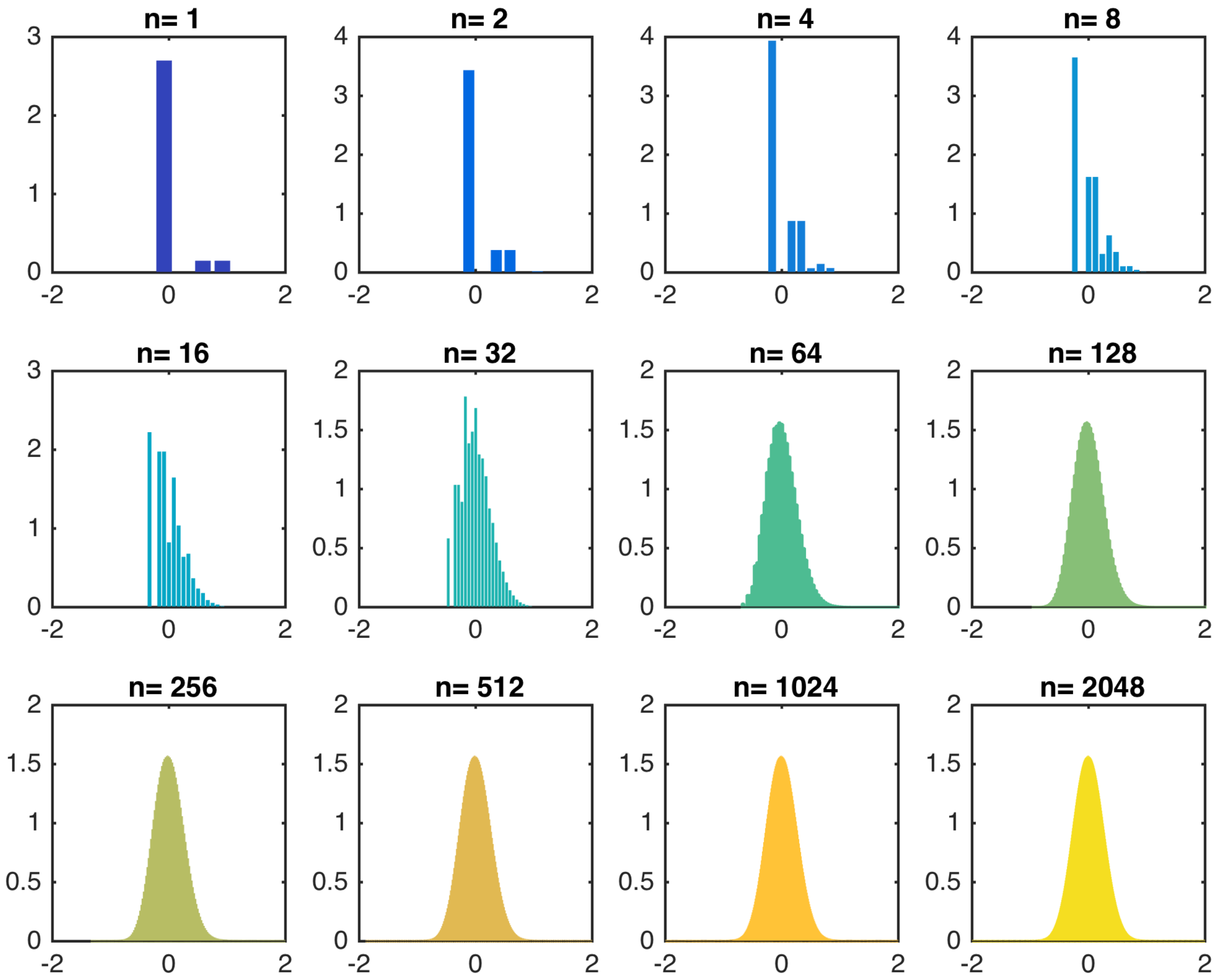
Normal distribution

- In other words, $s_n^* = \dfrac{s_n - \mu}{\sigma/\sqrt{n}}$ has always the same distribution for large n:

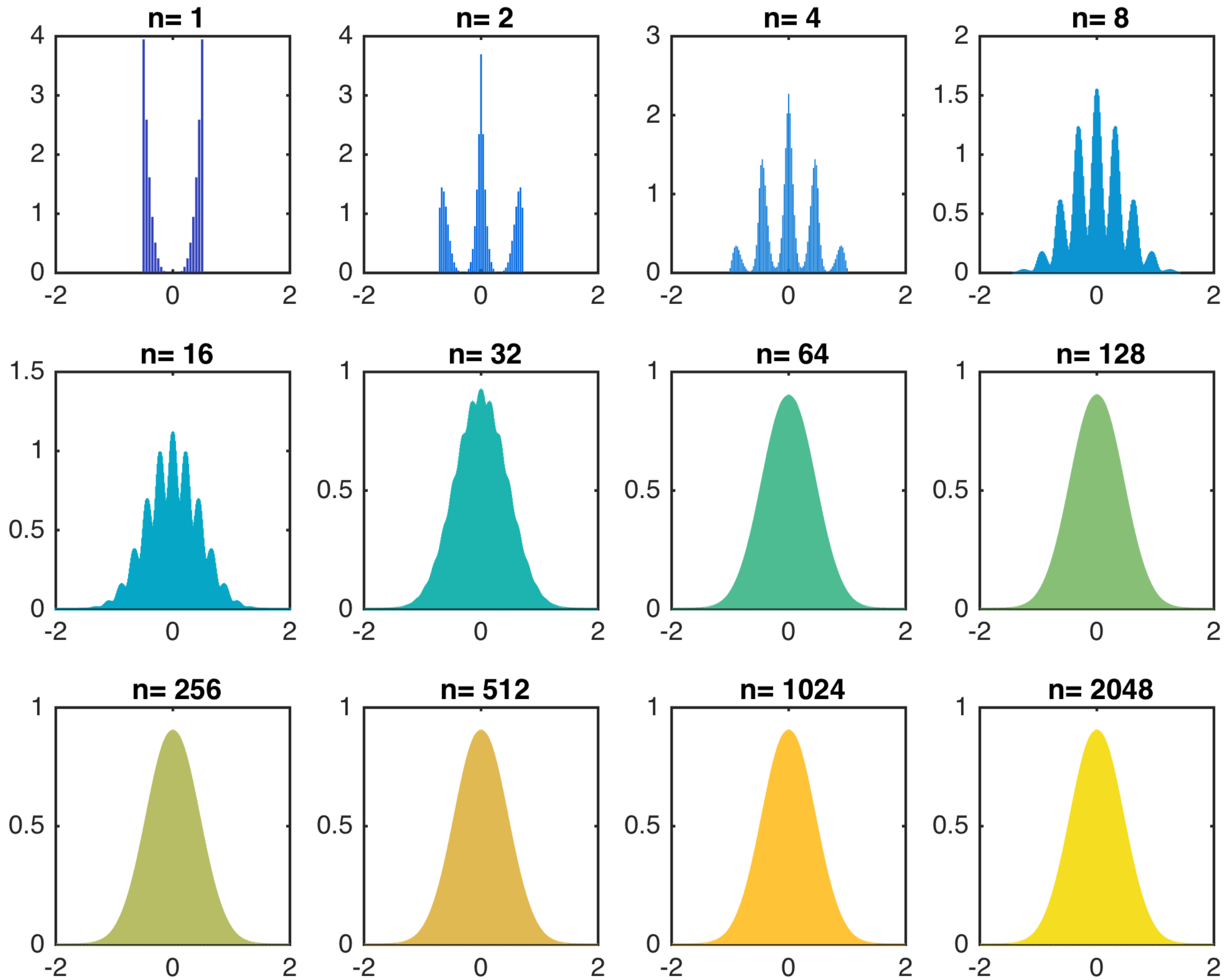
$$\mathcal{N}(0,1)(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

This is the so-called "Normal distribution".

Example 1



Example 2



Normal approximation

- In statistical testing it is generally accepted to use a normal distribution if
  - The samples are independent
  - the sample size n is larger than about 30.

Some concluding remarks about statistical tests

- A statistical test has
  - A Null Hypothesis H_0 that we would like to reject
  - A test statistic X (a function of an observed sample)
  - A probability distribution for the test statistic and hence the ability to calculate a "p value"
  - In practice the challenge is often to apply the right test:

Examples of tests

- Test for a specific expectation value:
  simple tail test
- Compare whether two groups of samples come from distributions with the same expectation (e.g. before and after an intervention):
  t-Test
- Decide whether there are meaningful groupings in a sample:
  F-test (or analysis of variance(ANOVA))
- …

Thursday Lecture

- Information theory
- Revisions I
- Abstentions