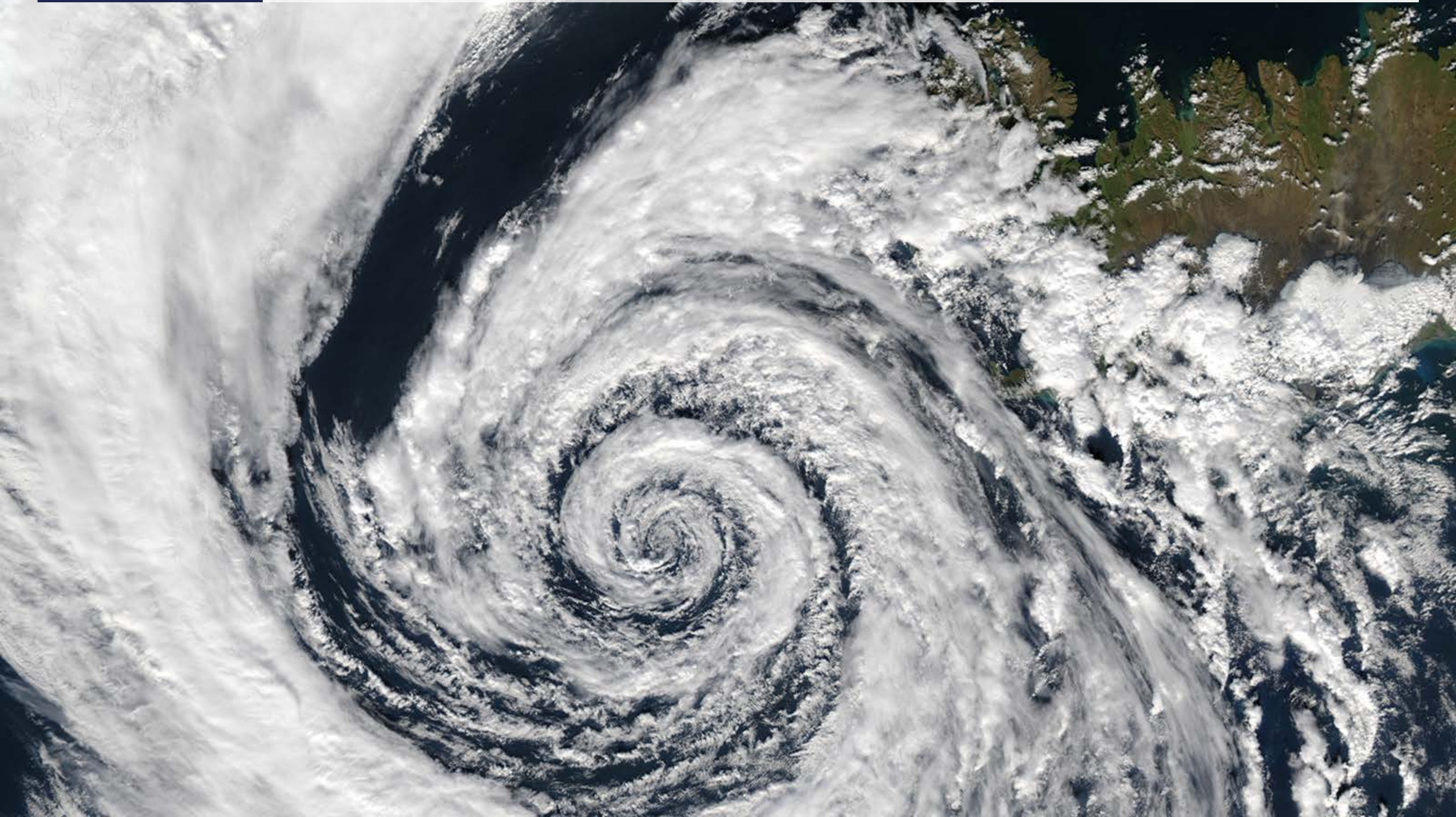
Statistics

Confidence Intervals

Name teacher

TU Delft

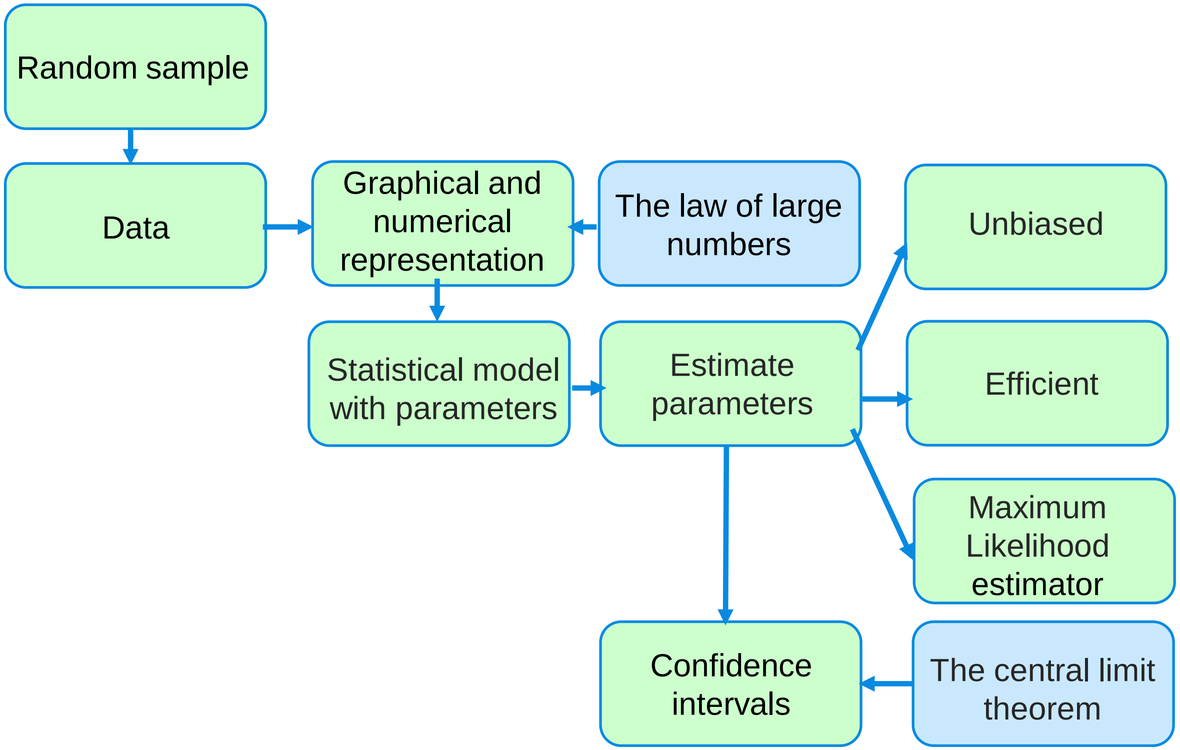


Learning objective

After this class you can

- **Create and interpret confidence intervals in various settings.**
- **Determine the sample size to achieve a given confidence level.**

Statistics


Random sample
Data
Graphical and numerical representation
The law of large numbers
Unbiased
Statistical model with parameters
Estimate parameters
Efficient
Maximum Likelihood estimator
Confidence intervals
The central limit theorem

Before this class (week 3.8 lesson 1):

Watch prelecture '*Confidence Intervals*'

Book: sections 23. 1,2,4 and 24. 1,2,3,4

Programme

- Critical values
- Confidence interval
- Confidence intervals with special properties
- Practice
- Half sided Confidence intervals

Suppose your scale weighs with a standard deviation of 0.5 kg. If for one person three measurings give 70.1, 70.2 and 70.5 kg. Give a 99% confidence interval for this person's 'true' weight.

A) (69.84, 70.70)

B) (68.82, 71.72)

C) (69.15 , 71.38)

D) (69.52, 71.01)

Critical values of the normal distribution

Definition

The critical value z_p is the value such that

$$P(Z \geq z_p) = p$$

Often used values:

p	0.05	0.025	0.01
z_p	1.645	1.960	2.326

Confidence interval for the mean of a normal distribution; known σ.

Theorem:

Suppose $X_1, X_2, \ldots, X_n$ have independent normal distributions with parameters μ and σ^2.

Then $\quad \mathrm{P}\left(\bar{X}_n - \frac{z_{\frac{1}{2}\alpha}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{\frac{1}{2}\alpha}\sigma}{\sqrt{n}}\right) = 1 - \alpha$

Definition

If in the above situation $x_1, x_2, \ldots, x_n$ is a realization of the RV's X_i, and $\gamma = 1 - \alpha$ then $\left(\bar{x}_n - z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}}\right)$ is called a 100γ per cent confidence interval for μ.

Confidence interval for μ (for normal data) if σ is unknown

Idea: instead of $\dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ standardized mean use $\dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$. 'studentized' mean

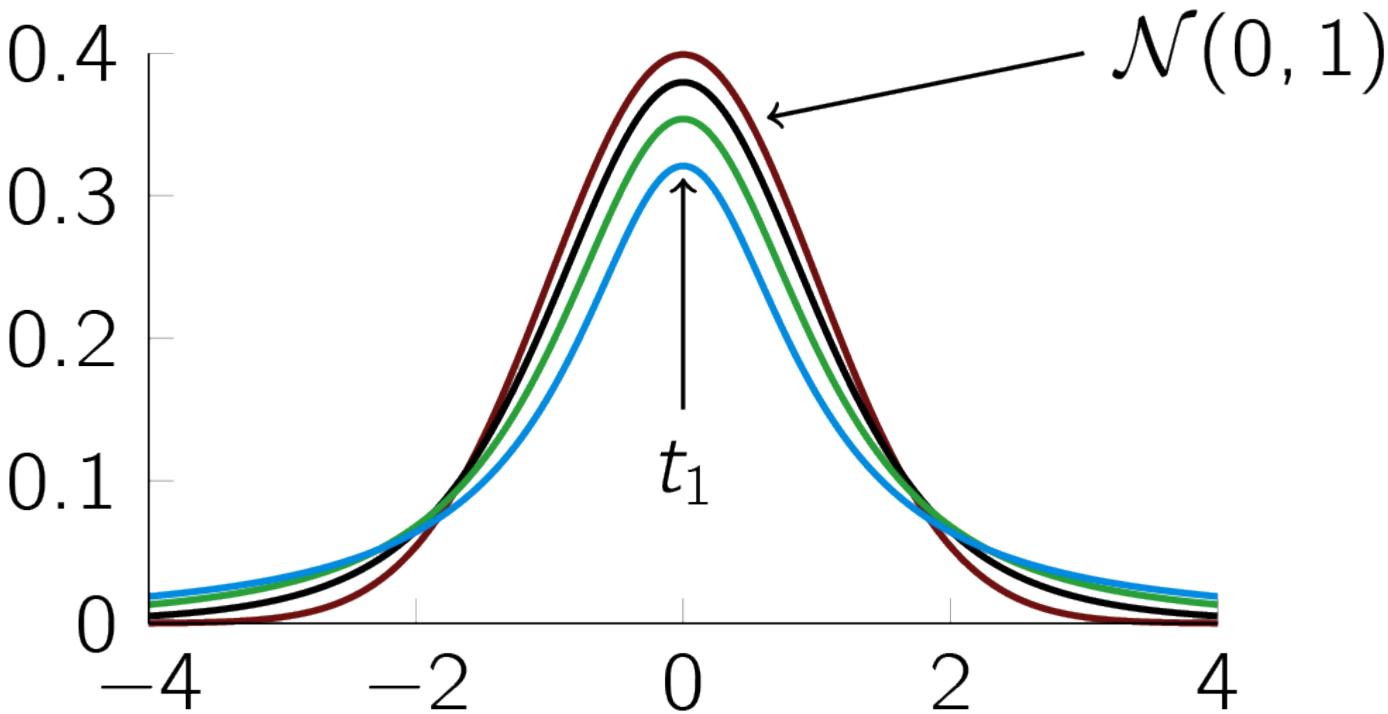
Theorem:

Suppose $X_1, X_2, \ldots, X_n$ have independent normal distributions with parameters μ and σ^2.

Then $T_n = \dfrac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$ has a

Student distribution with $n-1$ degrees of freedom.

The densities of Student distributions t_1, t_2, t_5 versus the standard normal density.



The higher n, the more t_n approaches the normal distribution. In general: t_n has ‘fatter tails’.

Confidence interval for μ (for normal data) if σ is unknown

Theorem:

Suppose $X_1, X_2, \ldots, X_n$ have independent normal distributions with parameters μ and (unknown) σ^2.

Then $$\mathrm{P}\left(-t_{n-1,\frac{1}{2}\alpha} \leq \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \leq t_{n-1,\frac{1}{2}\alpha}\right) = 1-\alpha,$$

where $t_{n-1,\frac{1}{2}\alpha}$ is a critical value of the Student distribution (with $n-1$ degrees of freedom).

Note that this can be rewritten as

$$\mathrm{P}\left(\bar{X}_n - t_{n-1,\frac{1}{2}\alpha}\frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{n-1,\frac{1}{2}\alpha}\frac{S_n}{\sqrt{n}}\right) = 1-\alpha$$

Confidence interval for μ (for normal data) if σ is unknown

If in the above situation $x_1, x_2, \ldots, x_n$ is a realization of the RV's X_i, and $\gamma = 1 - \alpha$ then

$$\left(\bar{x}_n - t_{n-1,\frac{1}{2}\alpha} \frac{s_n}{\sqrt{n}},\ \bar{x}_n + t_{n-1,\frac{1}{2}\alpha} \frac{s_n}{\sqrt{n}} \right)$$

is called a

100γ per cent confidence interval for μ.

Confidence interval for μ (for normal data) for large samples

If n is large, the law of large number implies that S_n^2 is a good approximation of σ^2, and also the t_n-distribution with a large n is close to the standard normal distribution.

Thus

$$P\left(\bar{X}_n - t_{n-1,\frac{1}{2}\alpha}\frac{S_n}{\sqrt{n}} \le \mu \le \bar{X}_n + t_{n-1,\frac{1}{2}\alpha}\frac{S_n}{\sqrt{n}}\right) \approx P\left(\bar{X}_n - z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n + z_{\frac{1}{2}\alpha}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

and so

$$\left(\bar{x}_n - z_{\frac{1}{2}\alpha}\frac{s_n}{\sqrt{n}}, \bar{x}_n + z_{\frac{1}{2}\alpha}\frac{s_n}{\sqrt{n}}\right)$$

is approximately a $(1-\alpha)100$ per cent confidence interval for μ.

Exercises

Ten bags of cement have an average weight of 93.5 kg. Assume the weights can be considered as a realization of a random sample from a normal distribution with unknown parameters. If the sample standard deviation of the ten weights is 0.75 kg, find
a 95% confidence interval for the expected weight of a bag.

In a paper a 95% confidence interval (1.6, 7.8) for the parameter μ of an $N(\mu, \sigma^2)$ distribution is reported.
(The researchers used the method with the studentized mean, and the sample size was $n = 16$.)

a. What is the mean of the (unknown) dataset?

b. What would be the 99% confidence interval?

Half-sided confidence interval

Let's look at the case with an unknown variance.

Once it is known that

$$\mathrm{P}\left(\mu \in \left(\bar{X}_n - t_{n-1,\frac{1}{2}\alpha}\frac{S_n}{\sqrt{n}},\ \bar{X}_n + t_{n-1,\frac{1}{2}\alpha}\frac{S_n}{\sqrt{n}}\right)\right) = 1-\alpha,$$

how would you construct a (stochastic) interval of the form $(L(X_1, \ldots, X_n), \infty)$ that contains μ with probability $(1-\alpha)$?

Exercise (bags of cement revisited: 10 bags, average weight 93.5 kg, sample SD 0.75 kg).
Find a 95% confidence interval (a, ∞) for the mean weight.

Confidence interval for a proportion p

The probability p of an event can estimated by a relative frequency. How can you construct a confidence interval for p?

Often used **setting**:
to find out about the fraction of a (large) population that has a certain behaviour or opinion, take a (relatively) small survey.

Confidence interval for a proportion *p*

Suppose X has a $\text{Bin}(n,p)$ distribution, and n is large.

Then $\dfrac{X - np}{\sqrt{np(1-p)}}$ is approximately standard normal.

Given a realization x, an approximate $100(1-\alpha)$ confidence interval for p is given by

$$\left(\hat{p} - z_{\frac{1}{2}\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\frac{1}{2}\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right),$$

where $\hat{p} = \dfrac{x}{n}$, a relative frequency.

Suppose a survey of 500 people gives 222 persons that vote for A, and 278 persons that vote for B. Let p_A be the fraction in favor of A of the whole population. A 95% confidence interval for p_A is (approximately) given by

A) (0.42, 0.47)

B) (0.40 , 0.49)

C) (0.43, 0.46)

D) (0.37, 0.52)

Suppose a measuring device measures with errors that are normally distributed with mean 0 and standard deviation 0.5 ‘unit’. How many measurements do you need for a 95% confidence interval of length < 0.2?

A) 16

B) 25

C) 54

D) 96

For next class (week 3.8 lesson 2):

Complete MyStatlab assignments and book exercises 25.1, 26.4 and 26.7

Watch prelectures *'Testing hypothesis'*

Book: section 25.1

After this class you can

- **Set up and carry out statistical hypothesis tests**
- **Interpret the results of statistical hypothesis tests**

Statistics

Good luck!