Bus 274:
Further Statistics For Business

Harald Schmidbauer

İSTANBUL BİLGİ ÜNİVERSİTESİ

About These Slides

- The present slides are not self-contained; they need to be explained and discussed. This will be done in the lectures.

- Even though being a "work in progress" and subject to revision, the slides constitute copyrighted material.
  If you want to reproduce or copy anything from the slides, please ask:

Harald Schmidbauer	**harald** at **hs-stat** dot **com**
Angi Rösch	**angi** at **angi-stat** dot **com**

- The slides were produced using LaTeX (www.latex-project.org) and R (www.R-project.org) on a GNU/Linux system, all of which are free and open source software (FOSS).

- R files used for this course are available upon request.

PART III:

Statistical Inference

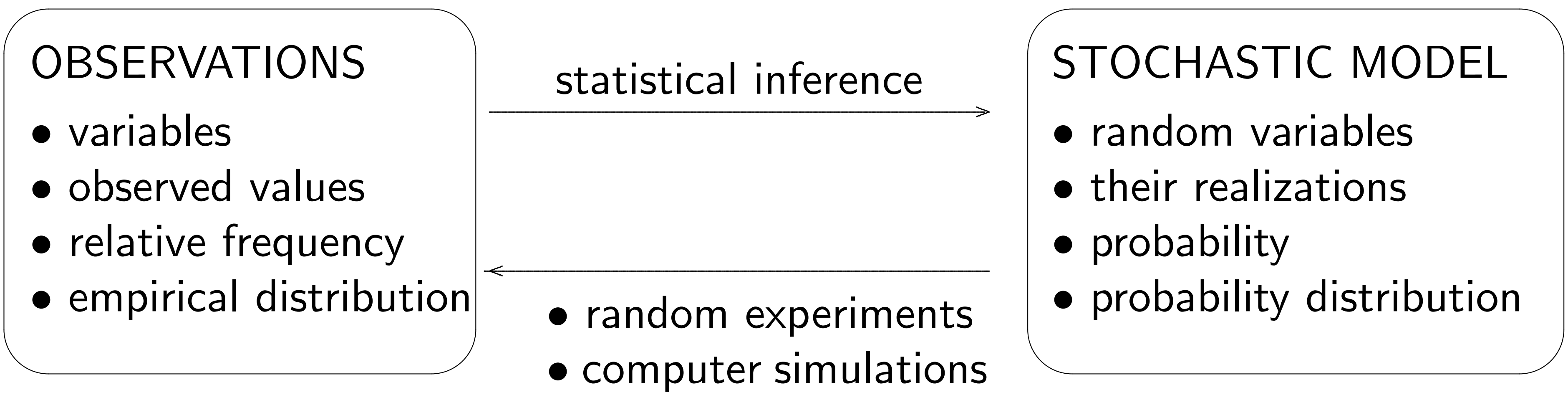
Part III: Statistical Inference

What is statistical inference?

- Statistical inference: Use empirical data (observations) to learn about unknown distributions (or unknown parameters).
- These unknown distributions characterize the population.
- Rests on the paradigm of inductive statistics.

Part III: Statistical Inference

Reminder: Statistical inference and other activities. . .



Part III: Statistical Inference

Methods of statistical inference.

The main methods of statistical inference are
(as far as parameters of distributions are concerned):

- point estimation of parameters,
- interval estimation of parameters,
- hypothesis tests concerning parameters.

Chapter 9:

Estimation

9.1 Introduction

Point estimation and interval estimation.

- Point estimation: The unknown parameter of a probability distribution is estimated, using a **single number** (a "point").
- Interval estimation: The unknown parameter of a probability distribution is estimated, using a **confidence interval** (an interval of numbers).

9.1 Introduction

Example: A public opinion poll — see also Chapter 5!

"Do you think New Orleans should be rebuilt?"

- Define: $p =$ share of American adults who say "YES"
- For any randomly selected person, we have a random variable:

$$X = \begin{cases} 1 & \text{if the person says "YES"} \\ 0 & \text{if the person says "NO"} \end{cases}$$

- Then, $P(X = 1) = p$.
- We need empirical evidence (i.e., data: a random sample of X) to learn about p!

9.1 Introduction

Example: A public opinion poll.

"Do you think New Orleans should be rebuilt?"

- There is empirical evidence:
  384 out of 609 randomly selected American adults said "YES". (According to CNN, 2005-09-08.)
- We can then estimate p:

$$\hat{p} = \frac{384}{609} = 63\%.$$

This is a *point estimate* — it might be better to use an *interval* to estimate p.

9.2 Point Estimation

The general situation.

- X: our variable of interest
- $X_1, X_2, \ldots, X_n$: random sample of X
- $x_1, x_2, \ldots, x_n$: realizations of $X_1, X_2, \ldots, X_n$
- The distribution of X depends on a parameter θ.
- The parameter θ is unknown and to be estimated from the data.

9.2 Point Estimation

The general situation.

- Given a distribution, and given a sample: How can we find an estimator for the parameter θ?
- A method is needed.
- We shall now see two methods:
  - the method of moments,
  - the maximum likelihood method.

9.2 Point Estimation

The method of moments.

- Set empirical k-th moment $=$ theoretical k-th moment.
- Solve for the unknown parameters.

In particular:

	empirical	theoretical
first moment:	$\bar{x}$	$\mathrm{E}(X)$
second (central) moment:	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$\mathrm{var}(X)$

9.2 Point Estimation

The method of moments — example: the binomial distribution.

- Variable of interest: $X \sim \mathrm{B}(1, p)$.
  (Example: public opinion poll.)

- It holds that $\mathrm{E}(X) = p$.

- Sample: $X_1, \ldots, X_n$

- Method of moments: Estimate

$$\hat{p} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

9.2 Point Estimation

The method of moments — example: the Poisson distribution.

- Let $X \sim \text{Po}(\lambda)$. Then $\text{E}(X) = \lambda$ and $\text{var}(X) = \lambda$.
- Sample: $X_1, \ldots, X_n$
- The method of moments is indifferent between

$$\begin{aligned} \hat{\lambda} &= \bar{X}, \\ \hat{\lambda} &= s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2. \end{aligned}$$

9.2 Point Estimation

Example: Football matches of Beşiktaş İstanbul.

- Define X = number of goals in a match.
- An appropriate model is: $X \sim \text{Po}(\lambda)$
- Using data from 170 matches, we computed:

$$\bar{x} = 2.96, \quad s^2 = 3.21$$

- How can we estimate λ?
- The method of moments is indifferent between

$$\hat{\lambda} = 2.96 \quad \text{and} \quad \hat{\lambda} = 3.21.$$

9.2 Point Estimation

The method of moments — example: the normal distribution.

- Let $X \sim \mathsf{N}(\mu, \sigma^2)$. Then $\mathrm{E}(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$.
- Sample $X_1, \ldots, X_n$.
- According to the method of moments, we estimate

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

9.2 Point Estimation

Maximum likelihood — an introductory example.

- Ten randomly selected people are asked whether they own a microwave.
- Three of them answer "yes".
- How likely is the empirical evidence

  "3 successes in 10 independent trials"

  as a function of success probability (here: the proportion) p?

9.2 Point Estimation

Maximum likelihood — an introductory example.

- This function (the likelihood function) is:

$$L(p; 10 \text{ trials, } 3 \text{ successes}) = \binom{10}{3} p^3 (1-p)^7$$

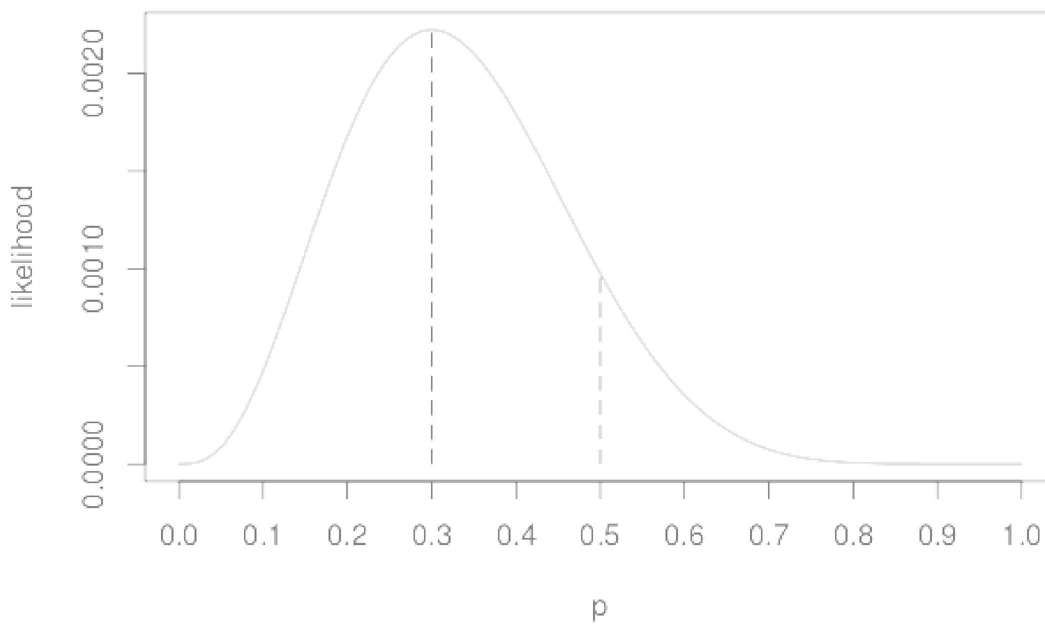
- We can use this function to estimate the unknown p.

- ML principle: To estimate p, use the value where L attains its maximum.

9.2 Point Estimation

Maximum likelihood — an introductory example.

- The function $p \mapsto L(p) = p^3(1-p)^7$:



- The ML principle says: Estimate $\hat{p} = 0.3$.

9.2 Point Estimation

Maximum likelihood — the definition of likelihood.

i) If X is discrete:

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P_\theta(X_i = x_i).$$

ii) If X is continuous with density f_θ:

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i).$$

9.2 Point Estimation

Maximum likelihood.

Method of Maximum Likelihood:

To estimate θ, use $\hat{\theta}$ such that the likelihood $L(\theta; x_1, x_2, \ldots, x_n)$ attains its maximum.

9.2 Point Estimation

Maximum likelihood — example: The binomial distribution.

Estimating an unknown share p. The likelihood function is

$$\begin{aligned} L(p) &= L(p; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P_p(X_i = x_i) \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i}. \end{aligned}$$

Take logarithms first:

$$\ln L(p) = \sum_{i=1}^{n} x_i \ln(p) + (n - \sum_{i=1}^{n} x_i) \ln(1-p)$$

Then maximize this expression.

9.2 Point Estimation

Maximum likelihood — example: The binomial distribution.

To maximize the expression

$$\ln L(p) = \sum_{i=1}^{n} x_i \ln(p) + (n - \sum_{i=1}^{n} x_i) \ln(1-p),$$

we equate the first derivative with zero and solve for p:

$$\frac{\mathrm{d}}{\mathrm{d}p} \ln L(p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0 \quad \Leftrightarrow \quad p = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Hence, the ML estimator of p is:

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

9.2 Point Estimation

Maximum likelihood — example: The Poisson distribution.

$X \sim \text{Po}(\lambda)$. The likelihood function is

$$L(\lambda) = L(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \mathrm{e}^{-\lambda}.$$

Take logarithms to maximize L:

$$\ln L(\lambda) = \sum_{i=1}^{n} \left(x_i \ln(\lambda) - \ln(x_i!) - \lambda \right),$$

equate the derivative with zero:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \ln L(\lambda) = \sum_{i=1}^{n} \left(\frac{x_i}{\lambda} - 1 \right) = 0 \Leftrightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i; \quad \text{ML estimator: } \hat{\lambda} = \bar{x}.$$

9.2 Point Estimation

Maximum likelihood — example: The normal distribution.

If only μ is to be estimated, the likelihood function is

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right].$$

In order to maximize L with respect to μ, we first take logs:

$$\ln L(\mu) = \sum_{i=1}^{n} \left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Then:

$$\frac{\mathrm{d}}{\mathrm{d}\mu} \ln L(\mu) = \sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2} = 0 \quad \Leftrightarrow \quad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i;$$ ML estimator: $\hat{\mu} = \bar{X}$.

9.2 Point Estimation

Maximum likelihood — example: The normal distribution.

If both μ and σ are unknown, the likelihood function is

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right].$$

In order to maximize L with respect to μ and σ^2, we first take logs:

$$\ln L(\mu,\sigma^2)=\sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi\sigma^2)-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

Maximization w.r.t. μ: as before; maximization w.r.t. σ^2:

$$\frac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2)=\sum_{i=1}^{n}\left(-\frac{1}{2\sigma^2}+\frac{(x_i-\mu)^2}{2\sigma^4}\right)=0 \quad\Leftrightarrow\quad \sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2.$$

9.2 Point Estimation

Maximum likelihood — example: The normal distribution.

- The ML estimator for σ^2 in this situation is therefore

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

- It holds that

$$E_{\sigma^2}(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \quad < \quad \sigma^2,$$

that is: $\hat{\sigma}^2$ is not unbiased. (Is this important?)

- This is why we often use

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

to estimate σ^2.

9.2 Point Estimation

Quality criteria for point estimators.

- Let $T = T(X_1, \ldots, X_n)$ be an estimator for θ
- Quality criteria: T should. . .
  - be unbiased.
  - approach the true θ as n grows large, i.e. be consistent.
  - have the smallest possible mean square error, i.e. be efficient.
  - extract all the information about θ from the sample, i.e. be sufficient.

9.2 Point Estimation

Advantages and disadvantages of a point estimate.

+ It is easy to understand.
+ It is easy to communicate.
– It is practically always wrong.
– It doesn't give any clue about the degree of accuracy.

Often better:

Use a confidence interval instead of a point estimate.

9.3 Confidence Intervals

Definition of a confidence interval.

- The distribution of X depends an an unknown parameter θ.
- Sample: $X_1, \ldots, X_n$.
- An interval $[C_1, C_2] = [C_1(X_1, \ldots, X_n), C_2(X_1, \ldots, X_n)]$ such that

$$P_\theta(\theta \in [C_1, C_2]) = 1 - \alpha$$

  is called a $(1 - \alpha) \cdot 100\%$ confidence interval for θ.
- $1 - \alpha$ is called the confidence level. (Often, $1 - \alpha = 0.95$.)

9.3 Confidence Intervals

Derivation of a confidence interval for p.

- We have seen that

$$\sum_{i=1}^{n} X_i \sim \mathsf{B}(n, p),$$

- Approximately according to the CLT:

$$\sum_{i=1}^{n} X_i \sim \mathsf{N}(np, np(1-p)), \quad \text{or:} \quad \hat{p} \sim \mathsf{N}(p, p(1-p)/n).$$

- This can be used to derive a confidence interval for p.

9.3 Confidence Intervals

Derivation of a confidence interval for p.

- Standardization:

$$P\left(-1.96 \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1.96\right) = 0.95.$$

- Manipulate this double inequality to isolate p:

$$P\left(\hat{p} - 1.96\sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{p(1-p)}{n}}\right) = 0.95$$

9.3 Confidence Intervals

Derivation of a confidence interval for p.

- Problem: p under the root is unknown. . .
- But: Substitute $\hat{p}$ for p under the root:

$$P\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 0.95.$$

- Again an approximation; slightly worse than the previous ones because p was replaced by its point estimate.

9.3 Confidence Intervals

Derivation of a confidence interval for p.

- We have obtained:

  An approximate 95% confidence interval for the unknown share p is:

$$\left[\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,,\;\; \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

9.3 Confidence Intervals

Example: A public opinion poll.

- Let's compute an approximate 95% confidence interval for the share of those who say New Orleans should be rebuilt.

- The 95% confidence bounds are given as

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.63 \pm 1.96\sqrt{\frac{0.63(1-0.63)}{609}} \approx 0.63 \pm 0.04.$$

- The approximate 95% confidence interval is therefore:

$$[0.59, 0.67]$$

9.3 Confidence Intervals

Some remarks about confidence intervals.

- Any value in the confidence interval is a plausible estimate for the unknown parameter.
- What **is** the confidence level? The distinction "Before/After" becomes again important!
- What should we do if we wish to obtain a "more precise" 95% confidence interval?
- What should we do if we wish to obtain a confidence interval with a higher confidence level? Is this desirable?

9.3 Confidence Intervals

Problem: Estimation of μ in the normal distribution.

- Let $X \sim \mathsf{N}(\mu, \sigma^2)$.

- Given a sample $X_1, \ldots, X_n$, how can we estimate μ?

- A point estimator is:

$$\hat{\mu} = \bar{X}$$

- Much better than a point estimator: a confidence interval!

9.3 Confidence Intervals

Derivation of a confidence interval for μ.

- We know:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathsf{N}(0, 1),$$

  which implies:

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +1.96\right) = 0.95$$

- As before, isolate μ.

9.3 Confidence Intervals

Derivation of a confidence interval for μ.

- Solving for μ, we obtain:
  A 95% confidence interval for the unknown μ is:

$$\left[\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}} \, , \quad \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

- The standard deviation of $\hat{\mu} = \bar{X}$ is $\frac{\sigma}{\sqrt{n}}$.

- The standard deviation of a point estimator is called the **standard error** of this estimator.

9.3 Confidence Intervals

Example: Cutting steel tubes.

- Cut-to-length operation generates tubes that have a normally distributed length (in inches) with mean μ and standard deviation σ.
- From previous operations, it is known that $\sigma = 0.1$, while μ is unknown, due to a new adjustment of the process.
- Sample of 15 tubes: 11.73, 12.02, 11.99, 11.86, $12.11, 12.11$, 12.02, 12.01, 11.89, 11.96, 12.12, 11.91, 11.98, 12.03, 11.95.
- With these data, $\bar{x} = 11.98$;
  95% confidence bounds: $11.98 \pm 1.96 \cdot 0.1/\sqrt{15}$.
- 95% confidence interval for μ: [11.93, 12.03]

9.3 Confidence Intervals

Derivation of a confidence interval for μ.

- What if σ^2 is unknown? Student's t distribution:

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}, \quad \text{where} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

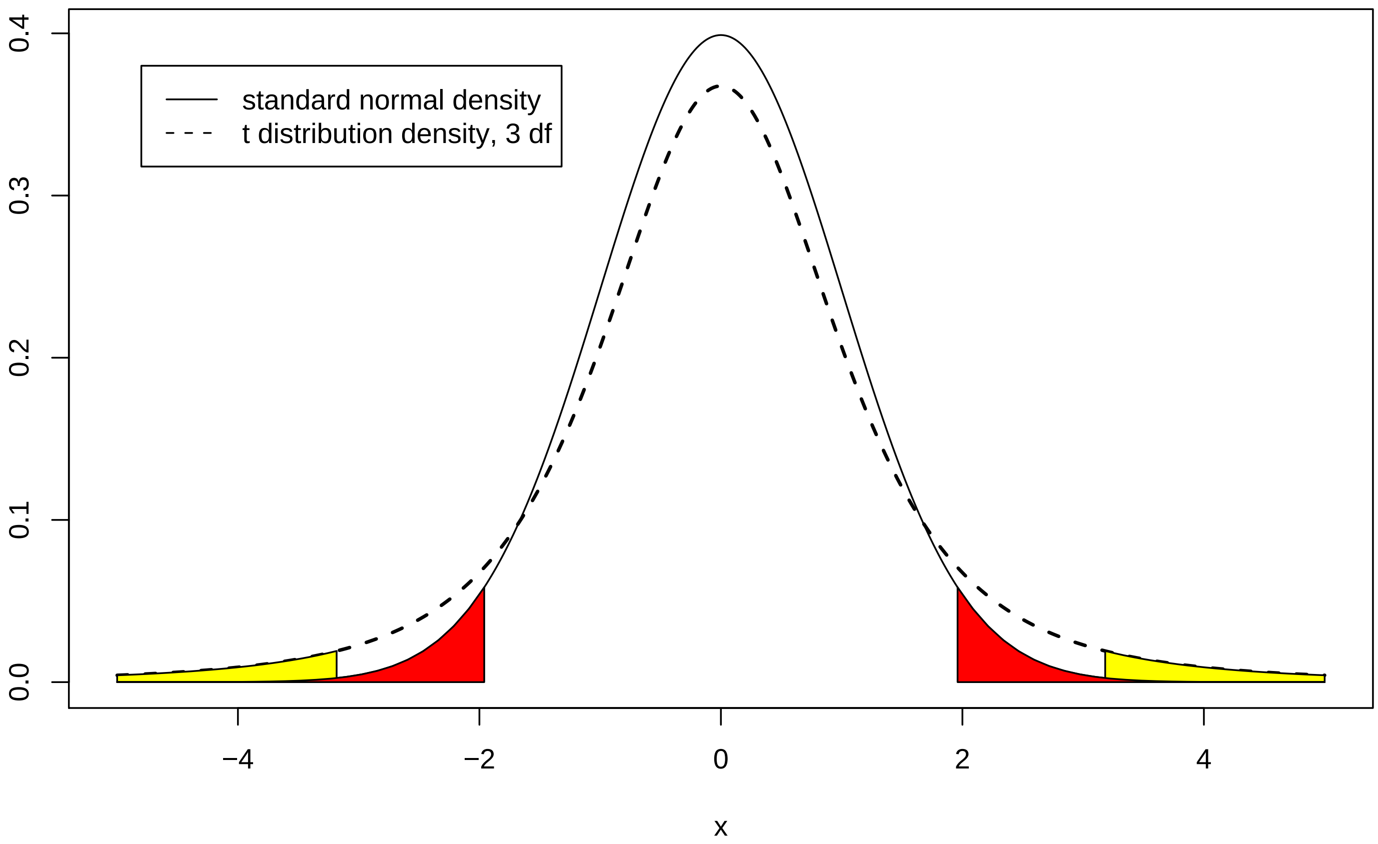
- This leads to a slightly larger 95% confidence interval for μ:

$$\left[\hat{\mu} - t_{n-1;0.975}\frac{s}{\sqrt{n}}\,,\quad \hat{\mu} + t_{n-1;0.975}\frac{s}{\sqrt{n}}\right],$$

where $t_{n-1;0.975}$ is the 97.5% quantile of the t distribution with $n-1$ degrees of freedom (df).

9.3 Confidence Intervals

N$(0, 1)$ and the t distribution.



9.3 Confidence Intervals

Problem: Estimation of σ^2 in the normal distribution.

- Let $X \sim \mathrm{N}(\mu, \sigma^2)$.
- Given a sample $X_1, \ldots, X_n$, how can we estimate σ^2?
- An unbiased point estimator is:

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Much better than a point estimator: a confidence interval!

9.3 Confidence Intervals

Derivation of a confidence interval for σ^2.

- We know:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$

which implies:

$$P\left(\chi^2_{n-1;0.025} \leq \frac{n-1}{\sigma^2}s^2 \leq \chi^2_{n-1;0.975}\right) = 0.95$$

where

$\chi^2_{n-1;0.025}$: 2.5% quantile of the χ^2_{n-1} distribution,
$\chi^2_{n-1;0.975}$: 97.5% quantile of the χ^2_{n-1} distribution.

9.3 Confidence Intervals

Derivation of a confidence interval for σ^2.

- Isolating σ^2 in the middle leads to:

$$P\left(\frac{(n-1)s^2}{\chi^2_{n-1;0.975}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1;0.025}}\right) = 0.95.$$

- A 95% confidence interval for σ^2 is therefore:

$$\left[\frac{(n-1)s^2}{\chi^2_{n-1;0.975}}\,,\;\frac{(n-1)s^2}{\chi^2_{n-1;0.025}}\right]$$

9.3 Confidence Intervals

Example: Monthly phone expenditure.

- We assume:

  X = monthly phone expenditure (in YTL)

  is a random variable $X \sim \mathsf{N}(\mu, \sigma^2)$ with unknown μ and σ^2.

- With data

$$x_1 = 126, \quad x_2 = 138, \quad x_3 = 116, \quad x_4 = 120$$

  we obtain the point estimate $s^2 = 92$.

- What is a 95% confidence interval for σ^2?

9.3 Confidence Intervals

Approximate confidence intervals.

- The general shape of approximate 95% confidence bounds for an unknown parameter θ is:

$$\hat{\theta} \pm 2 \cdot \text{ standard error of } \hat{\theta},$$

where $\hat{\theta}$ is a point estimator for θ.

9.3 Confidence Intervals

Example: Total expenditure of customers in a supermarket.

- Let

  $$X = \text{total expenditure of a customer at the supermarket}$$

- Goal: Find a confidence interval for the expectation of X.
- We found earlier, using 508 observations:

  $$\bar{x} = 15.43, \quad s_x^2 = 166.96, \quad s_x = 12.92$$

- Standard error of $\hat{\mu} = \bar{X}$: $\quad s_x/\sqrt{508} = 0.57$
- Approximate 95% confidence bounds: $15.43 \pm 2 \cdot 0.57$.
- Resulting confidence interval: $[14.29, 16.57]$

9.3 Confidence Intervals

Example: Analyzing returns on stocks.

	bvsp	dji	gdaxi
first day	2001-01-02	2001-01-02	2001-01-02
last day	2006-01-24	2006-01-24	2006-01-24
observations	1249	1271	1284
NAs	72	50	37
mean	0.08852	0.00575	0.00031
std error	0.05689	0.03471	0.04158
var	3.33644	1.26360	2.98375
std deviation	1.82659	1.12410	1.72735
skewness	−0.26851	0.14790	0.06951
std error	0.20456	0.26835	0.19960
kurtosis	1.67302	3.89917	2.83231
std error	1.07215	1.04682	0.52009

9.3 Confidence Intervals

Example: Analyzing returns on stocks.

Approximate 95% confidence intervals for the kurtosis are

Bovespa:	$[-0.47, 3.82]$
Dow-Jones:	$[1.81, 5.99]$
DAX:	$[1.79, 3.87]$

It turns out that Bovespa is different with respect to its kurtosis!