Bus 701:
Advanced Statistics

Harald Schmidbauer

İSTANBUL BİLGİ ÜNİVERSİTESİ

About These Slides

- The present slides are not self-contained; they need to be explained and discussed. They contain only a small part of the course Bus 701.

- 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.r** at **t-online** dot **de**

- The slides were produced using LaTeX and R (the R project; www.R-project.org) on a Linux system.

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

PART III:

Statistical Inference

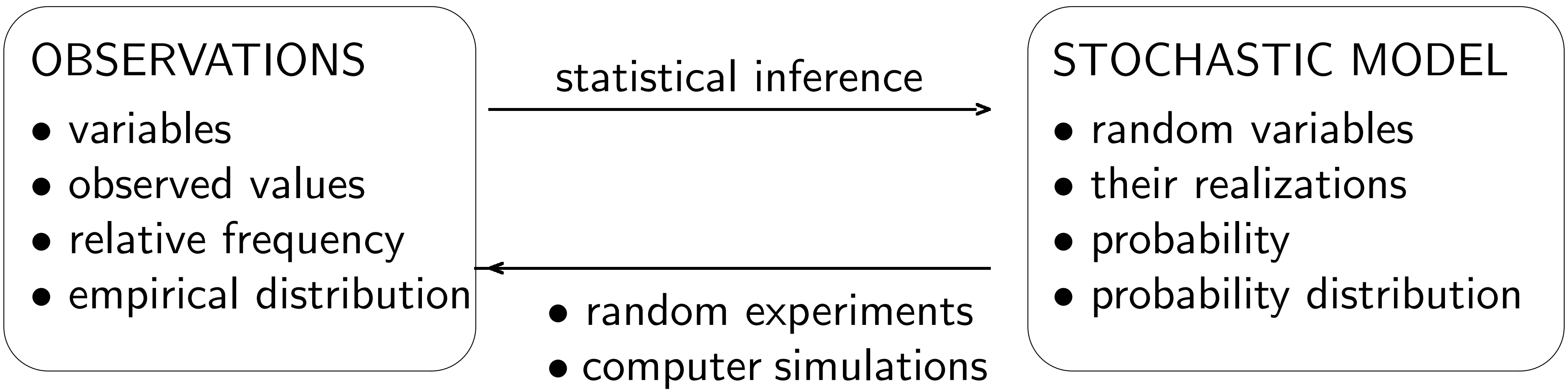
Part III: Statistical Inference

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. . .



Chapter 10:

Estimation

10.1 Introduction

Statistical inference:

- Point estimation
- Interval estimation
- Hypothesis testing

10.2 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:	$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$\mathrm{var}(X)$

10.2 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 $E(X) = p$.

- Sample: $X_1, \ldots, X_n$

- Method of moments: Estimate

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

10.2 Point Estimation

The method of moments — example: the Poisson distribution.

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

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

10.2 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: $X_1, \ldots, X_n$
- Method of moments:

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

10.2 Point Estimation

Maximum likelihood — an introductory example.

- Fifty randomly selected people are asked if they own a microwave.
- Ten of them say “yes”.
- Could the share of people in the entire population owning a microwave be, say, 40%?
- Maybe! But that’s quite unlikely. . .
- ML principle: Choose the parameter which is “most likely” to have produced the observed data.

10.2 Point Estimation

Maximum likelihood — the definition of likelihood.

- $X_1, X_2, \ldots, X_n$: 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.
- We have to distinguish between two cases: discrete and continuous X.

10.2 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:

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

10.2 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.

10.2 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:

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

10.2 Point Estimation

Quality criteria for point estimators.

- Let $T = T(X_1, \dots, X_n)$ 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.

10.2 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.

10.3 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, $1 - \alpha = 0.95$.)

10.3 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.

10.3 Confidence Intervals

Derivation of a confidence interval for p.

- Standardization:

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

- Manipulate this double inequality to isolate p:

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

10.3 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.

10.3 Confidence Intervals

Derivation of a confidence interval for p.

- We have obtained:

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

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

10.3 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:

$$[0.59, 0.67]$$

10.3 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?

10.3 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!

10.3 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.

10.3 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.

10.3 Confidence Intervals

Derivation of a confidence interval for μ.

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

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

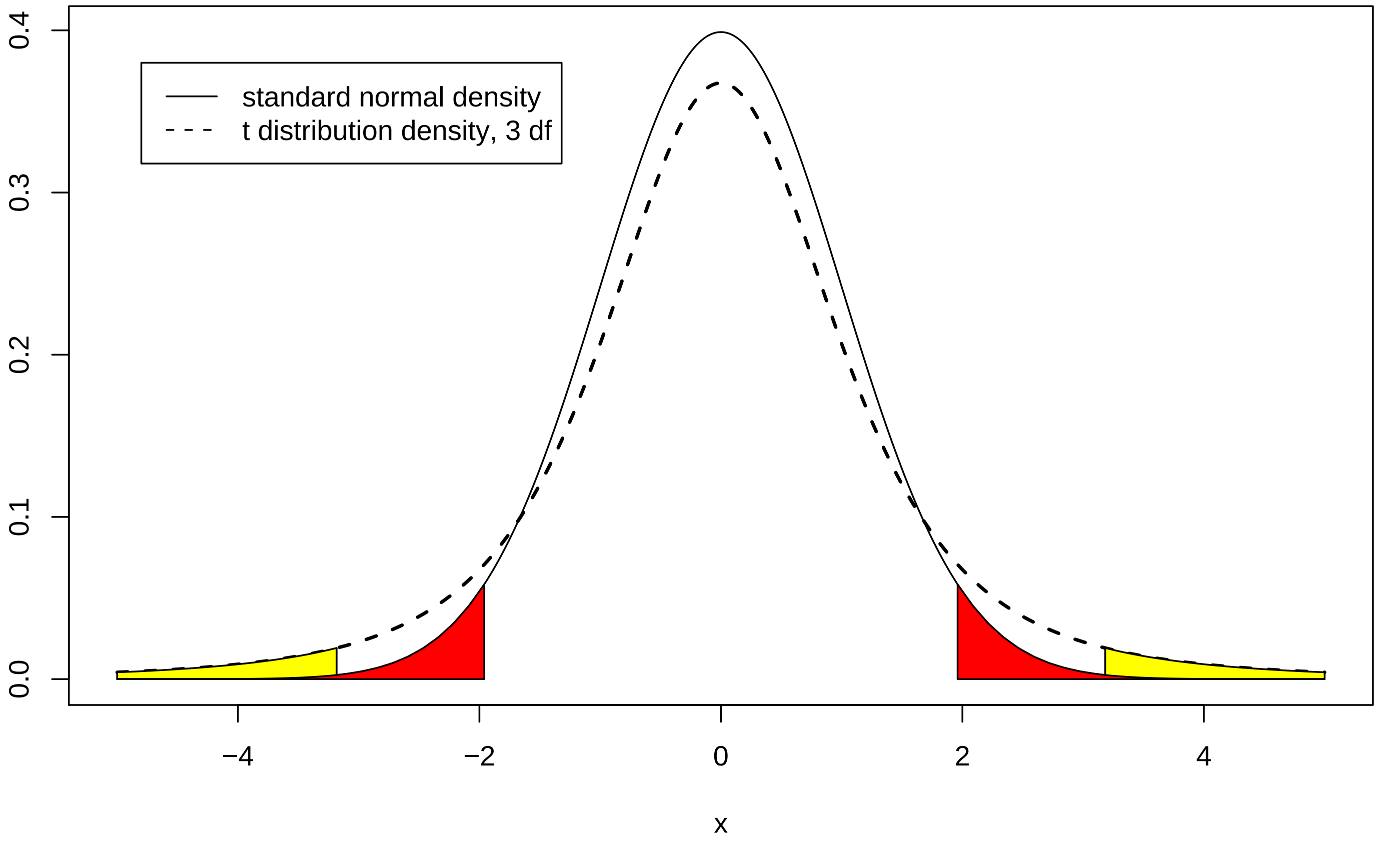
- This leads to a slightly larger 95% confidence interval for μ:

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

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

10.3 Confidence Intervals

$\mathsf{N}(0, 1)$ and the t distribution.



10.3 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.

10.3 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

10.3 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!