FIAS Frankfurt Institute for Advanced Studies

FIAS Summer School

THEORETICAL NEUROSCIENCE & COMPLEX SYSTEMS

Scientific Directors: Wolfgang Maass • Christoph von der Malsburg

Gordon Pipa • Wolf Singer • Jochen Triesch • Misha Tsodyks

4 - 26 August 2007

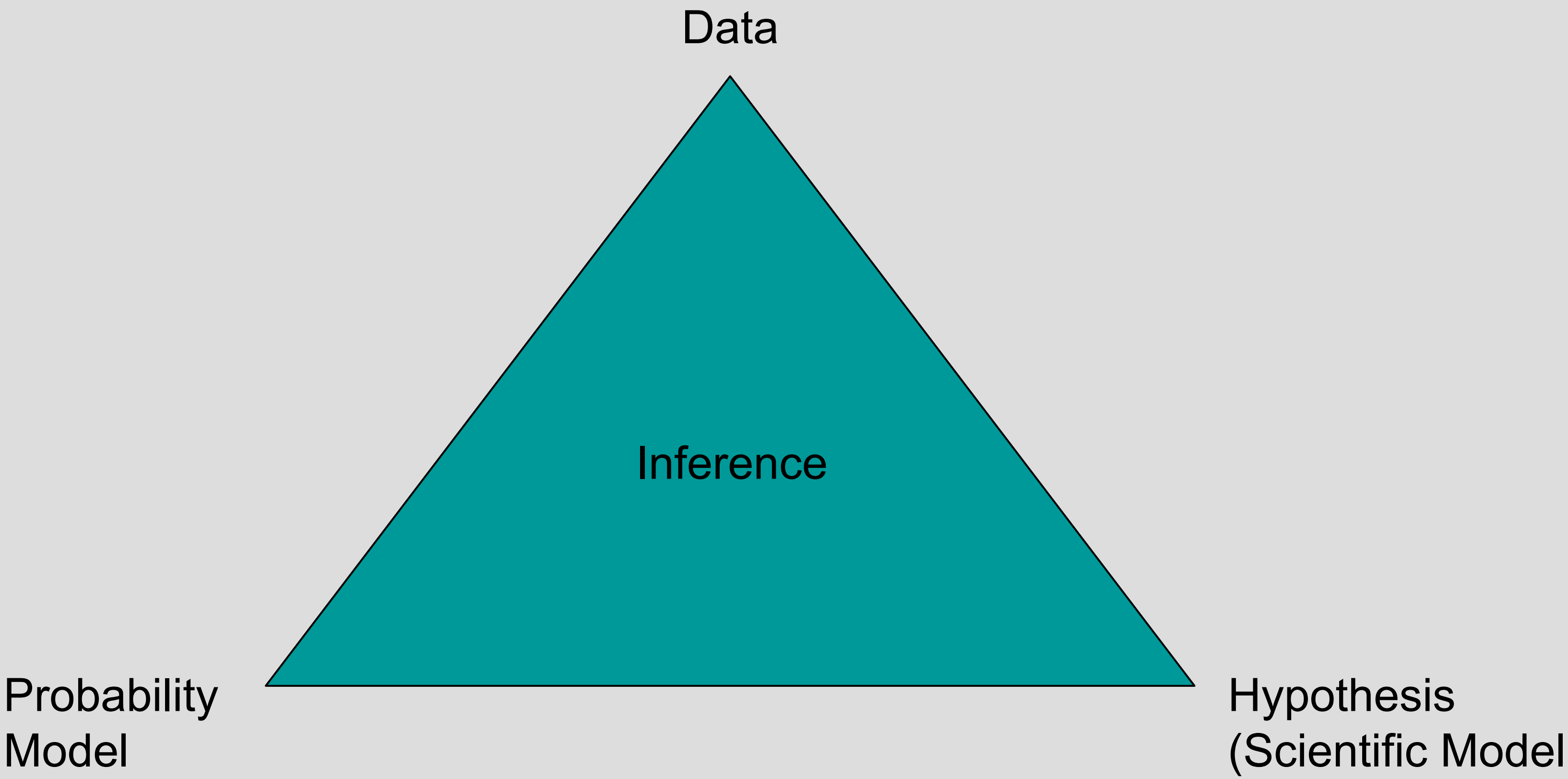
Frankfurt/M, Germany

Gordon Pipa

Frankfurt Institute for Advanced Studies &
Max-Planck for Brain Research

pipa@fias.uni-frankfurt.de

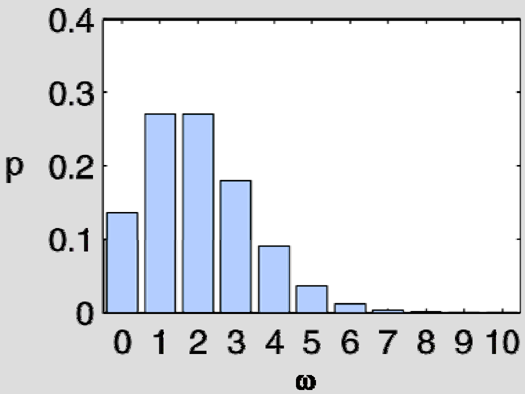
1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping


Data
Inference
Probability Model
Hypothesis (Scientific Model)

Two General Categories of Statistical Inference:

1. ***Estimation***
   - Point Estimates
   - Interval Estimates

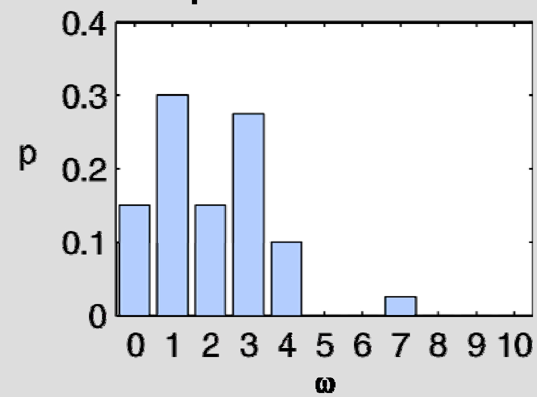
real distribution



experiment :

- set of n measurements (ω_i),
  with i=1..n $F \rightarrow \omega = (\omega_1, \omega_2, \ldots, \omega_n)$

empirical distribution



$$F = F(\boldsymbol{\omega})$$

$$\hat{F} = \hat{F}(\boldsymbol{\omega})$$

- **Nominal**

  A categorical variable places an individual into one of several groups or categories.

  (e.g. red, green, blue / long, short)

- **Ordinal**

  A quantitative variable that can be ranked. Real value is not considered

  (e.g. longest, second longest, … shortest)

- **Interval**

  A quantitative variable takes numerical values for which arithmetic operations such as adding and averaging make sense.

  (e.g. reddishness / length)

- **Population**:
  The entire group of individuals or potential outcomes that we want information about.
  (e.g. all Students at this course, or all potential outcomes of an dicing 1,2,3,4,5,6)

- **Sample**:
  A part of the population that we actually examine in order to gather information.
  (just the experimental people in this course, or 20 samples of dicing)

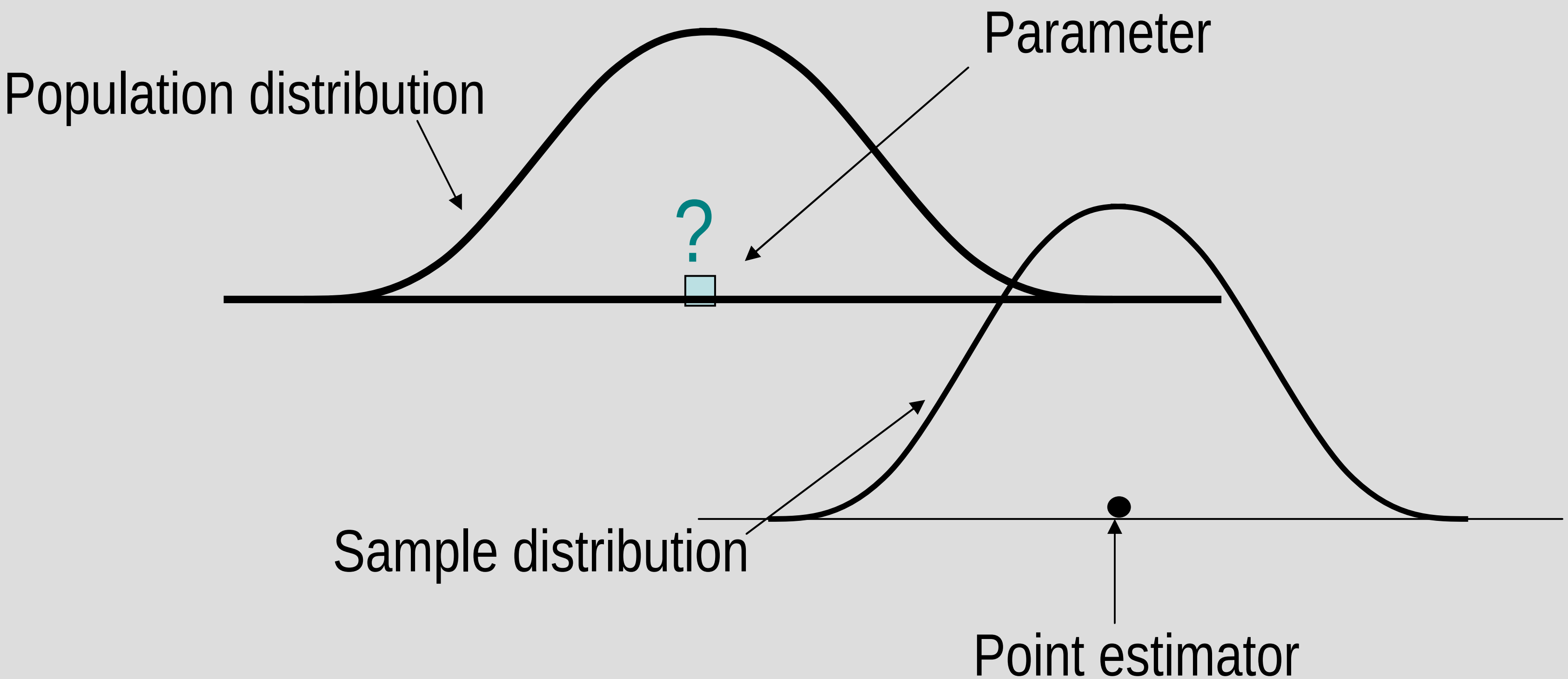
- **Sample size**:
  Number of observations/individuals in a sample.

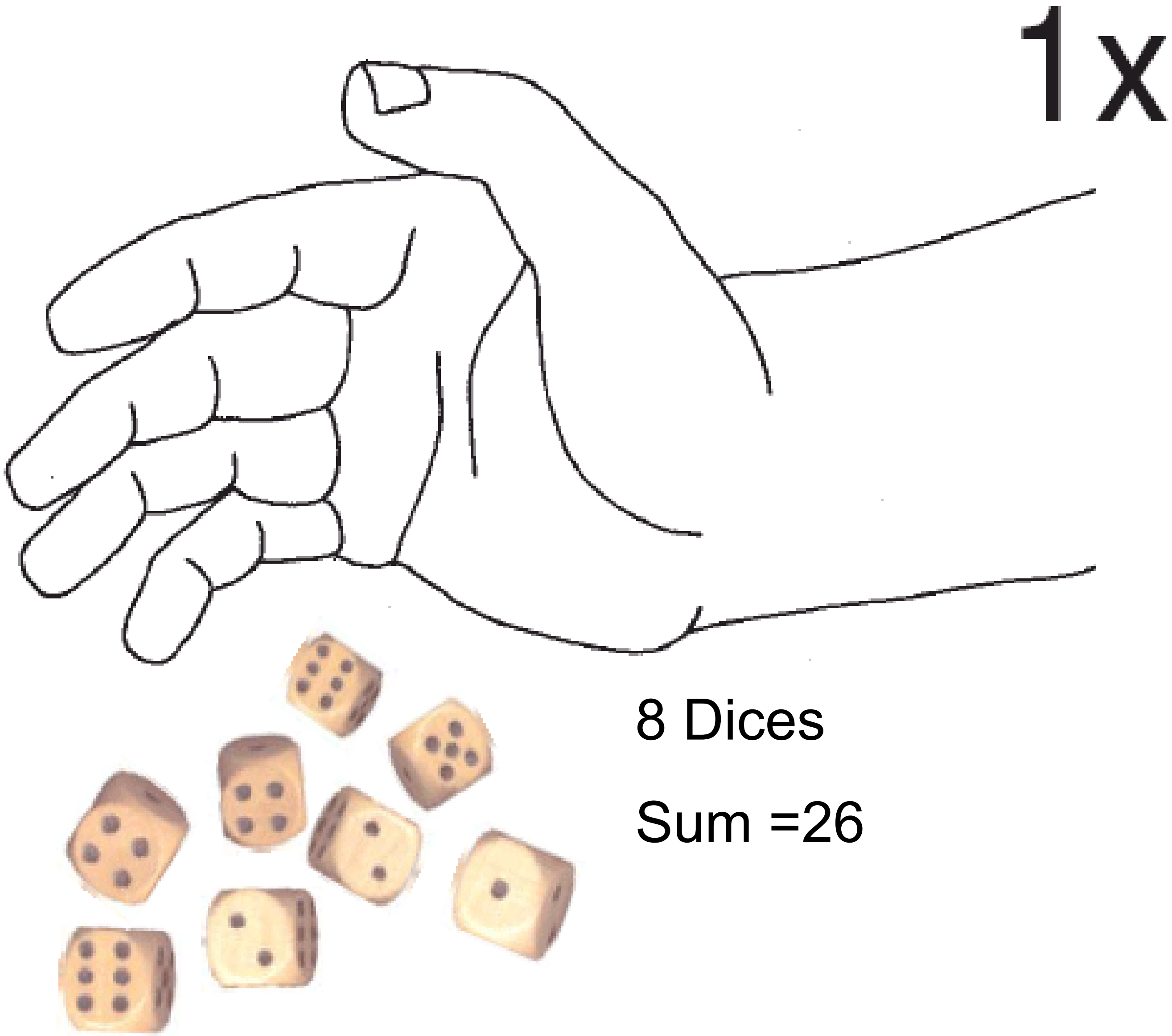
- **Statistical inference**:
  To make an inference about a population based on the information contained in a sample.
  (e.g. Nationality, or Background in Neuroscience)

Outline

1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping

A point estimator draws inference about a population by estimating the value of an unknown parameter using a single value or a point.

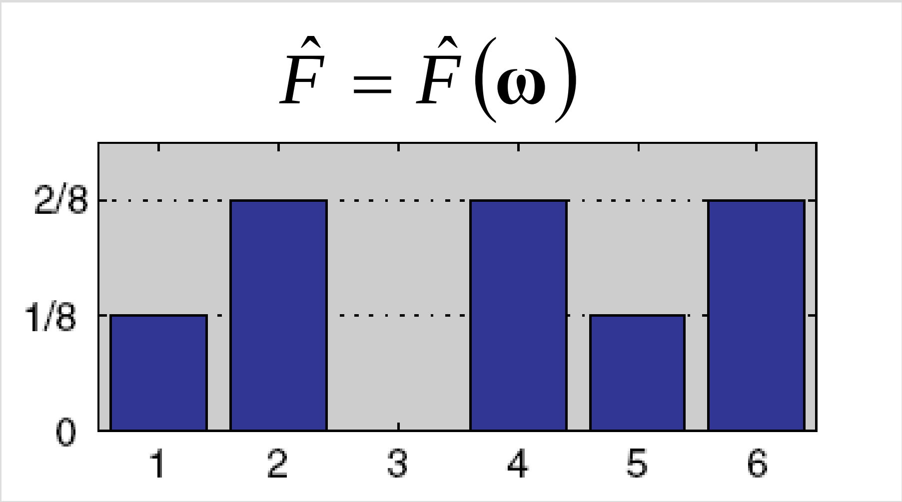




8 Dices

Sum =26

Population distribution is equally distributed (fair dices !)

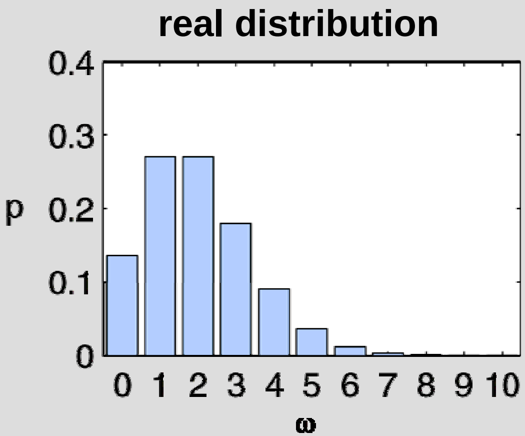


- 8 sample
- Test-statistic: average

$$\overline{\omega} = \frac{1}{8}\sum_{i=1}^{8}\omega_i$$

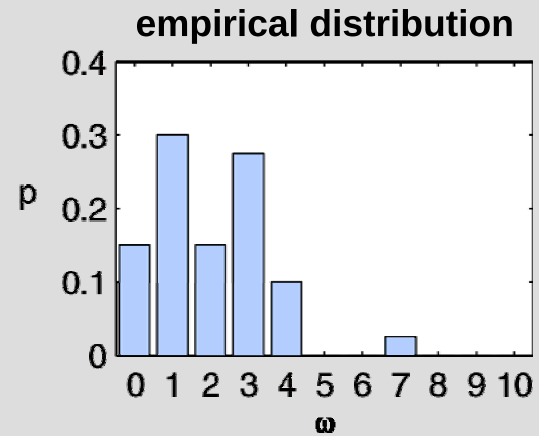
- Interval statistic

$$\overline{\omega} = 3.25$$



experiment :

- set of n measurements (ω_i), with i=1..n $F \rightarrow \omega = (\omega_1, \omega_2, \ldots, \omega_n)$



$$F = F(\omega)$$

$$\hat{F} = \hat{F}(\omega)$$

- Generally speaking, we want to quantify some population parameter (e.g. the expected value = population mean μ)

- A ***point estimate*** is a single numerical summary from a *random sample* that is used to estimate a population parameter (e.g. Estimate μ by $\bar{X}$)

- We generally like the point estimate to be unbiased (for infinite number of samples $\mu = \bar{X}$)

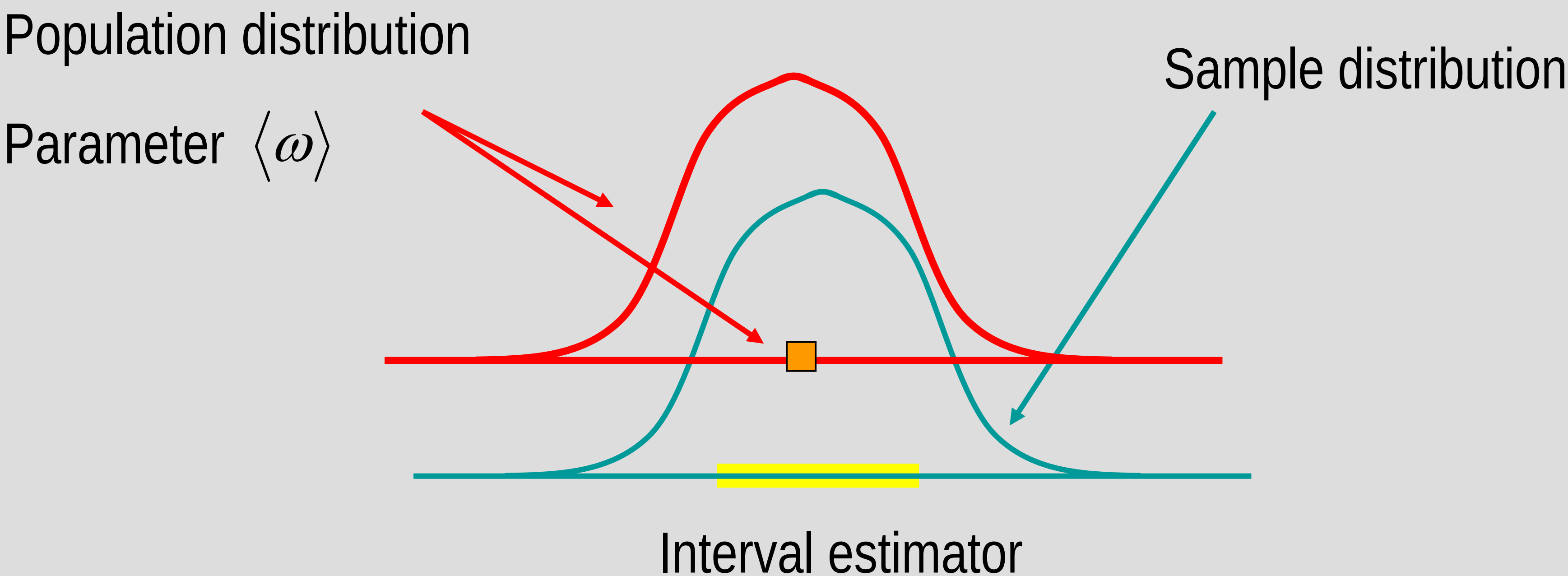
- One problem with point estimates is that they give no indication of precision.

Drawbacks from Point Estimates

- How different is the estimate from the true parameter?
- How reliable is our estimate?
- How confident are we with our estimate?

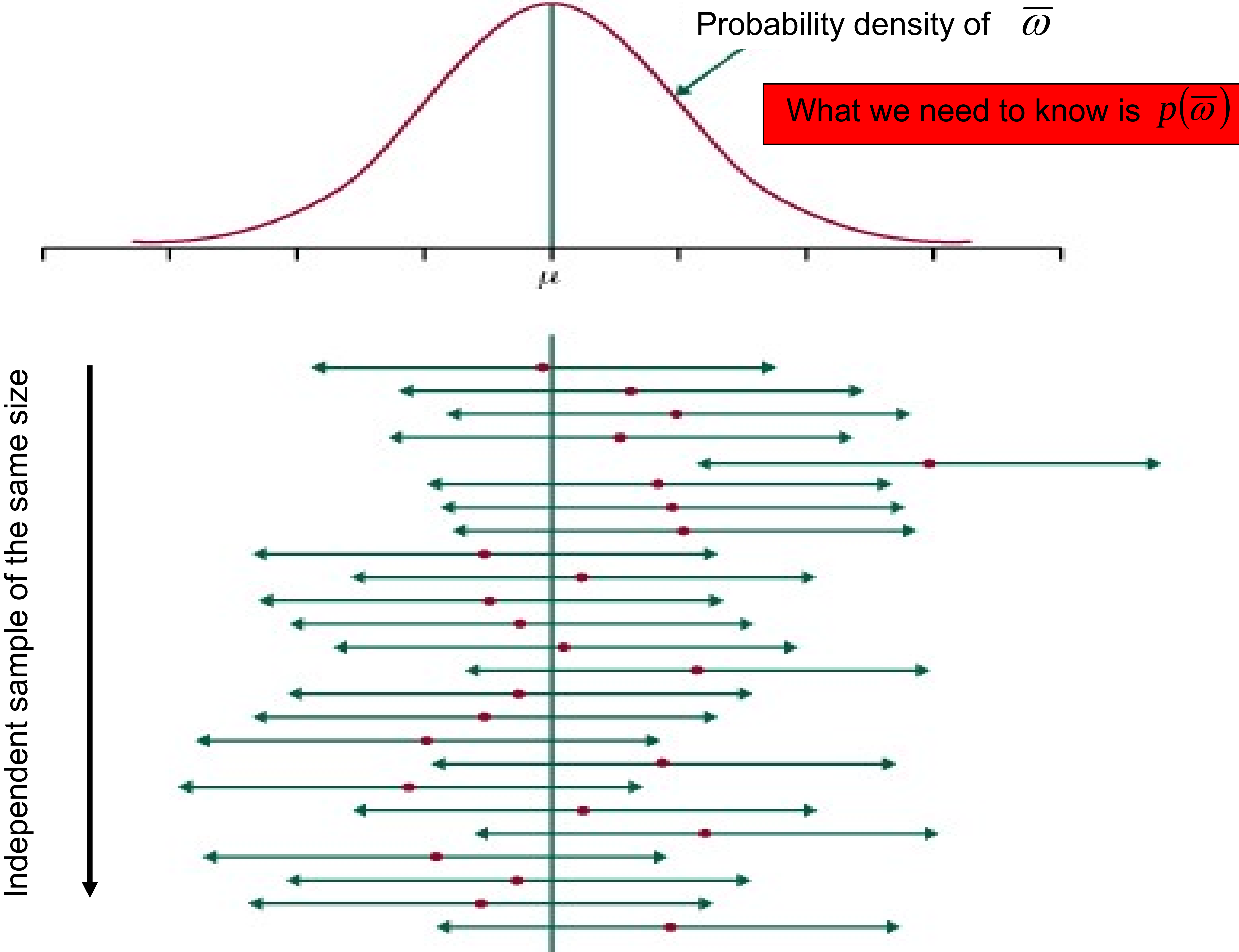
1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping

An interval estimator draws inferences about a population by estimating the value of an unknown parameter using an interval.



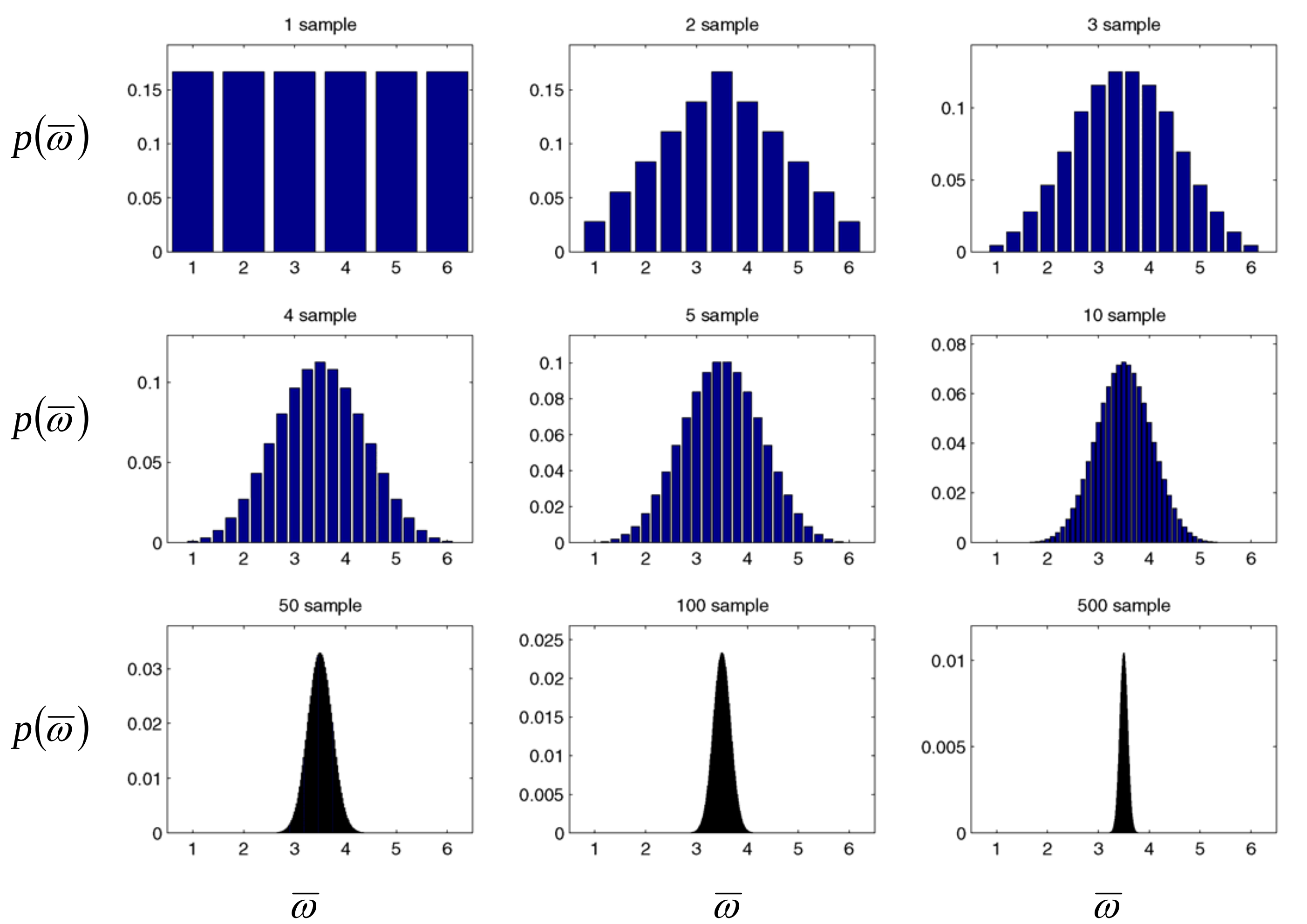
$\overline{\omega}$

- A confidence interval has the form: point estimate ± margin of error
- The point estimate is our guess for the value of the unknown parameter.
- The margin of error shows how accurate we believe our guess is, based on the sampling distribution of the estimate.
- The confidence level shows how confident we are that the procedure will catch the true population parameter, usually mean.



Probability density of $\overline{\omega}$
What we need to know is $p(\overline{\omega})$
μ
Independent sample of the same size

Central Limit theorem



$$\overline{\omega} = \frac{1}{N} \sum_{i=1}^{N} \omega_i$$

With increasing N the distribution of $\overline{\omega}$ approaches a normal distribution

Dilemma:

Population distribution is unknown but needed to estimate the certainty of average value.

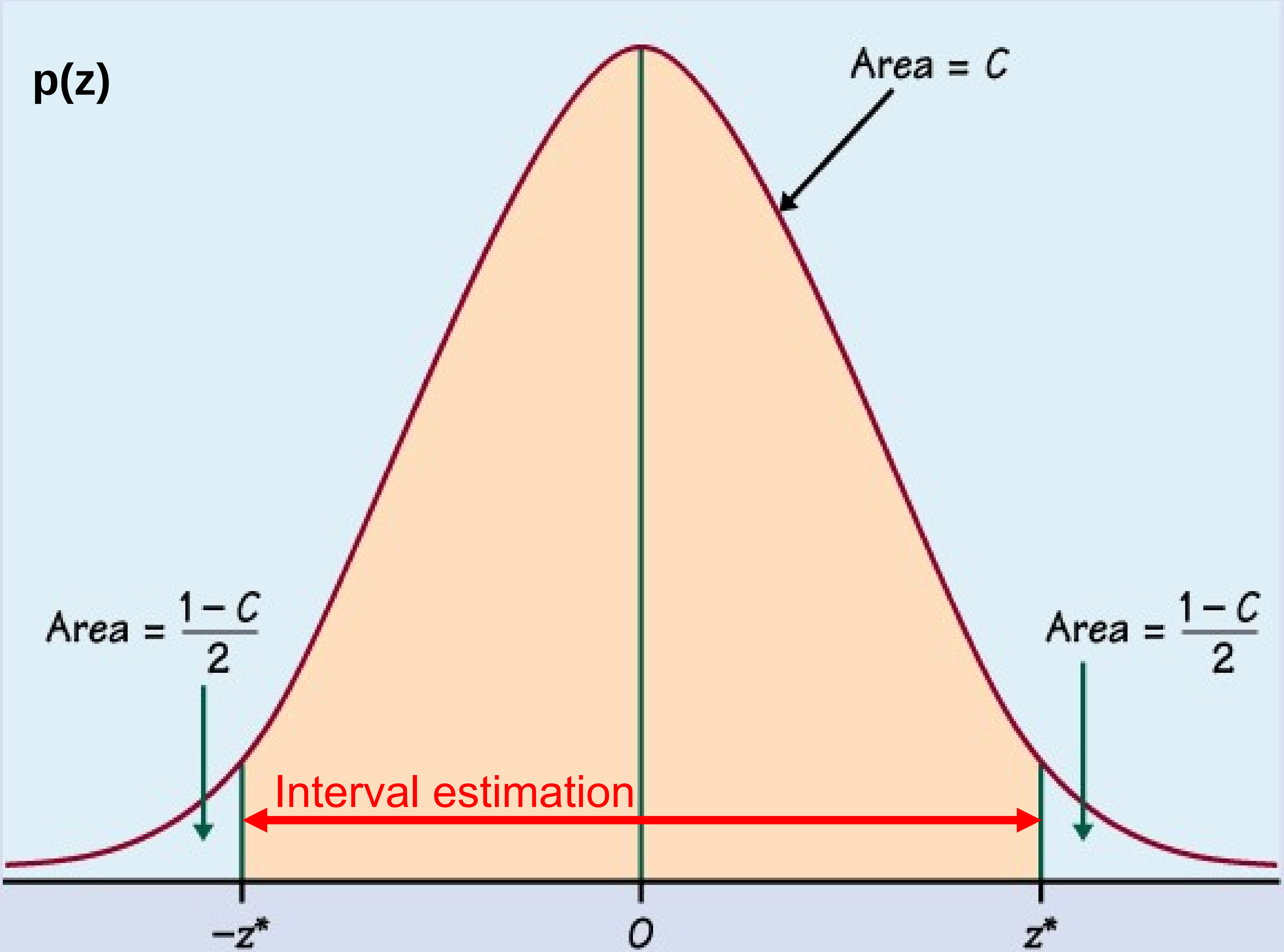
Solution:

For most cases the population distributions the distribution of the average is well approximated by an normal (Gaussian) distribution.

How is a confidence interval produced from a sampling distribution?

- To estimate μ, a sample of size n is drawn from the population, and its mean $\overline{X}$ is calculated.
- Under certain conditions, $\overline{X}$ is normally distributed (or approximately normally distributed.), thus

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \sim N(0,1) \text{ with } \sqrt{n}\sigma_{\bar{x}} = \sigma$$


p(z)
Area = C
Area = $\frac{1-C}{2}$
Area = $\frac{1-C}{2}$
Interval estimation
$-z^*$
0
z^*

We know

C: Confidence level

$$P(-z_* \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le z_*) = C.$$

$$P(\mu - z_* \frac{\sigma}{\sqrt{n}} \le \overline{x} \le \mu + z_* \frac{\sigma}{\sqrt{n}}) = C.$$

This leads to the relationship

$$P(\overline{x} - z_* \frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + z_* \frac{\sigma}{\sqrt{n}}) = C.$$

- Use the Standard Error (S) to construct a ***confidence interval***

- **General Form**:

  Lower bound < Point Estimate < Upper bound

- The margin of error is a function of:
  - the population standard deviation
  - the confidence level
  - the sample size.

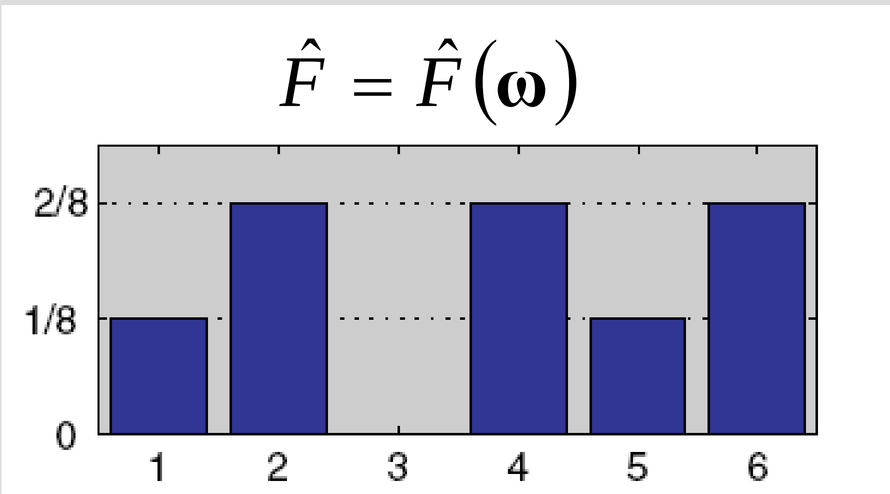
- If everything else remains the same, then
  - The *larger* the sample size, the *narrower* the CI.
  - The *higher* the confidence level, the *wider* the CI;
  - The *larger* the population SD, the *wider* the CI.

$$P\left(\boxed{\bar{x} - \boxed{z_*}\frac{\boxed{\sigma}}{\sqrt{\boxed{n}}}} < \mu < \boxed{\bar{x} + \boxed{z_*}\frac{\boxed{\sigma}}{\sqrt{\boxed{n}}}}\right) = C$$

Lower and upper bound

Confidence level	α	$\alpha/2$	Z_*
0.90	0.10	0.05	1.645
0.95	0.05	0.025	1.96
0.98	0.02	0.01	2.33
0.99	0.01	0.005	2.575

Population distribution is equally distributed (fair dices !)



- 8 sample
- Test-statistic: average

$$\overline{\omega} = \frac{1}{8}\sum_{i=1}^{8}\omega_i = 3.25$$

$$\sigma_{\omega} = 1.90 \rightarrow$$

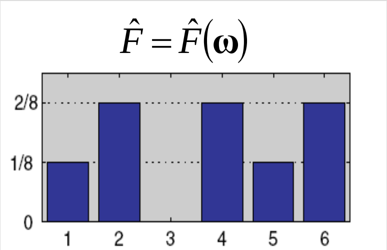
$$\sigma_{\overline{\omega}} = \frac{1}{\sqrt{8}}1.90 = 0.67$$

$$\overline{\omega} - 2 * \sigma_{\overline{\omega}} \leq \mu \leq \overline{\omega} + 2 * \sigma_{\overline{\omega}} \text{ with } \mathrm{z}^{*} = 2$$

$$3.25 - 2 * 0.67 \leq \mu \leq 3.25 - 2 * 0.67$$

$$1.91 \leq \mu \leq 4.59$$

1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping



General Idea:

Inference based on rejection of a Model

H_0: Null-Hypothesis:

This is the hypothesis that should be tested: $\langle \omega \rangle = 3.5$

Attention there are
only 1 to 6 spots per dice
1+2+3+4+5+6=21
21/6 = 3.5

Procedure 1:

First select a Confidence level and compute the compute the confidence interval. If the value assumed by H_o is not element of the confidence interval reject H_o

Procedure 2:

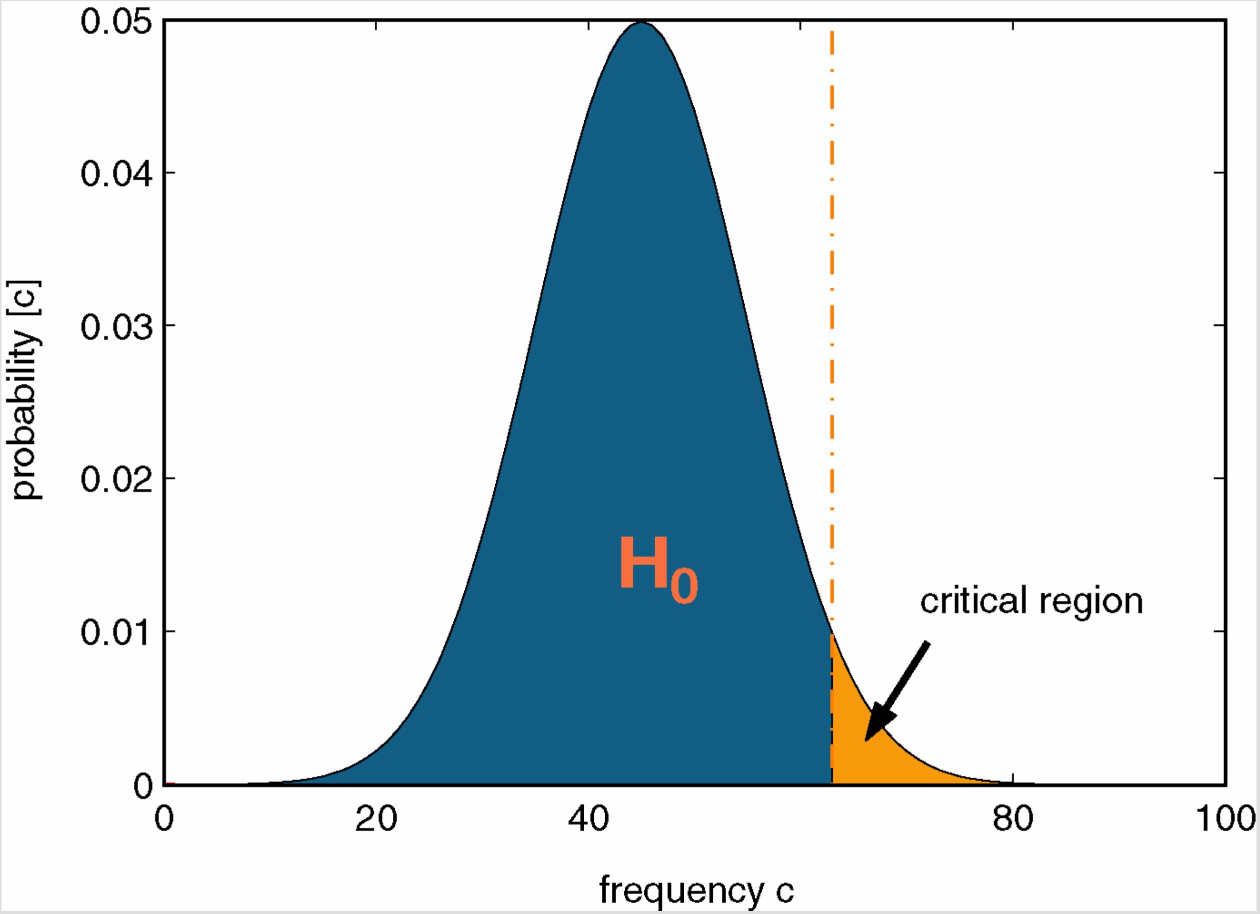
Compute the probability that the observation can be explained by H_o. If the probability is lower than critical value reject H_o

Definitions

- Test level defines the critical value
- Significance level defines the probability that H_0 is true

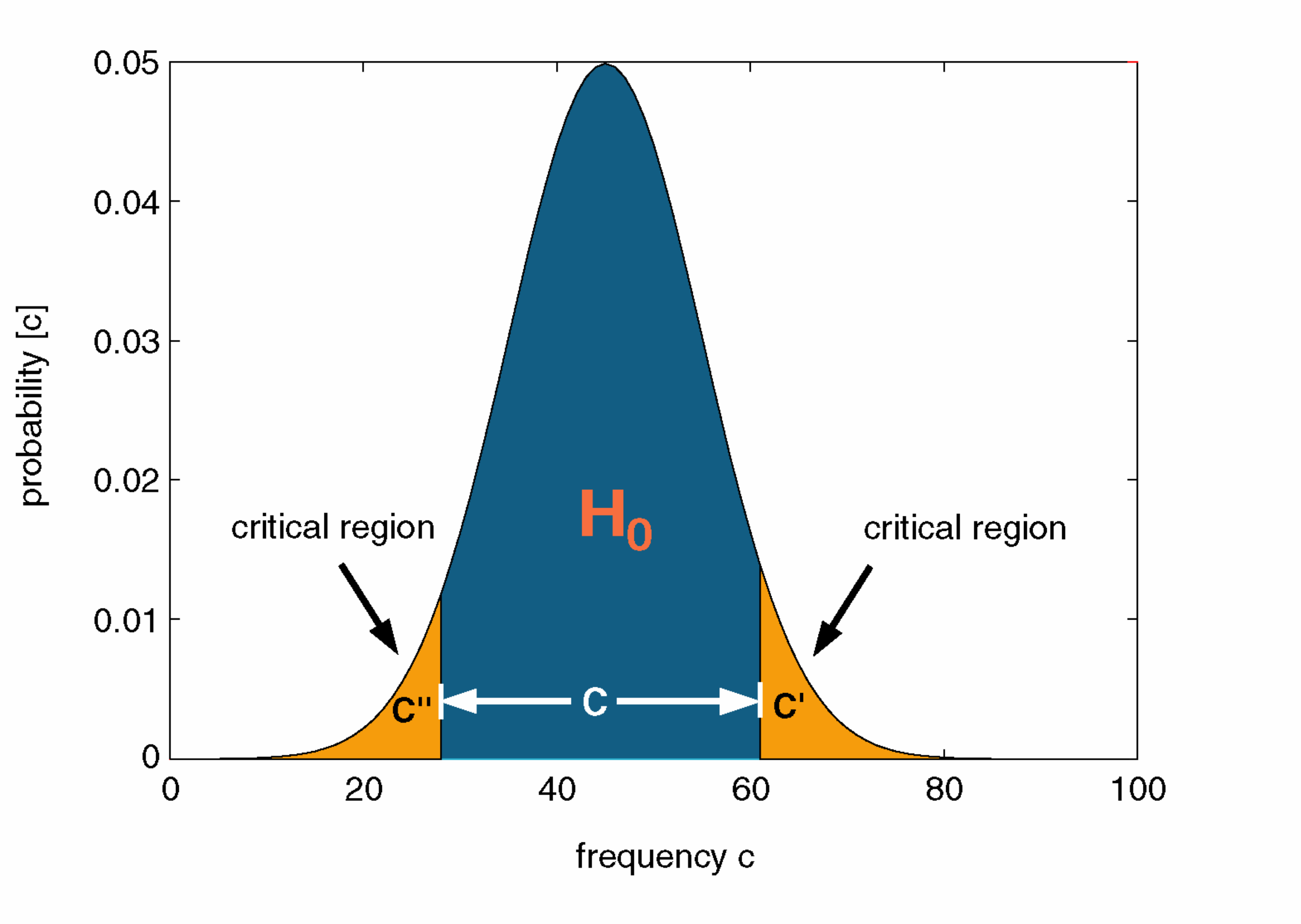
- A ***Type I Error or α Error*** occurs when we incorrectly reject the null hypothesis. The probability of a Type I Error is

$$\alpha = P\left(\text{reject } H_0 \mid H_0 \text{ is true}\right)$$



- A ***Type I Error or α Error*** occurs when we incorrectly reject the null hypothesis. The probability of a Type I Error is

$$\alpha = P\left(\text{reject } H_0 \mid H_0 \text{ is true}\right)$$

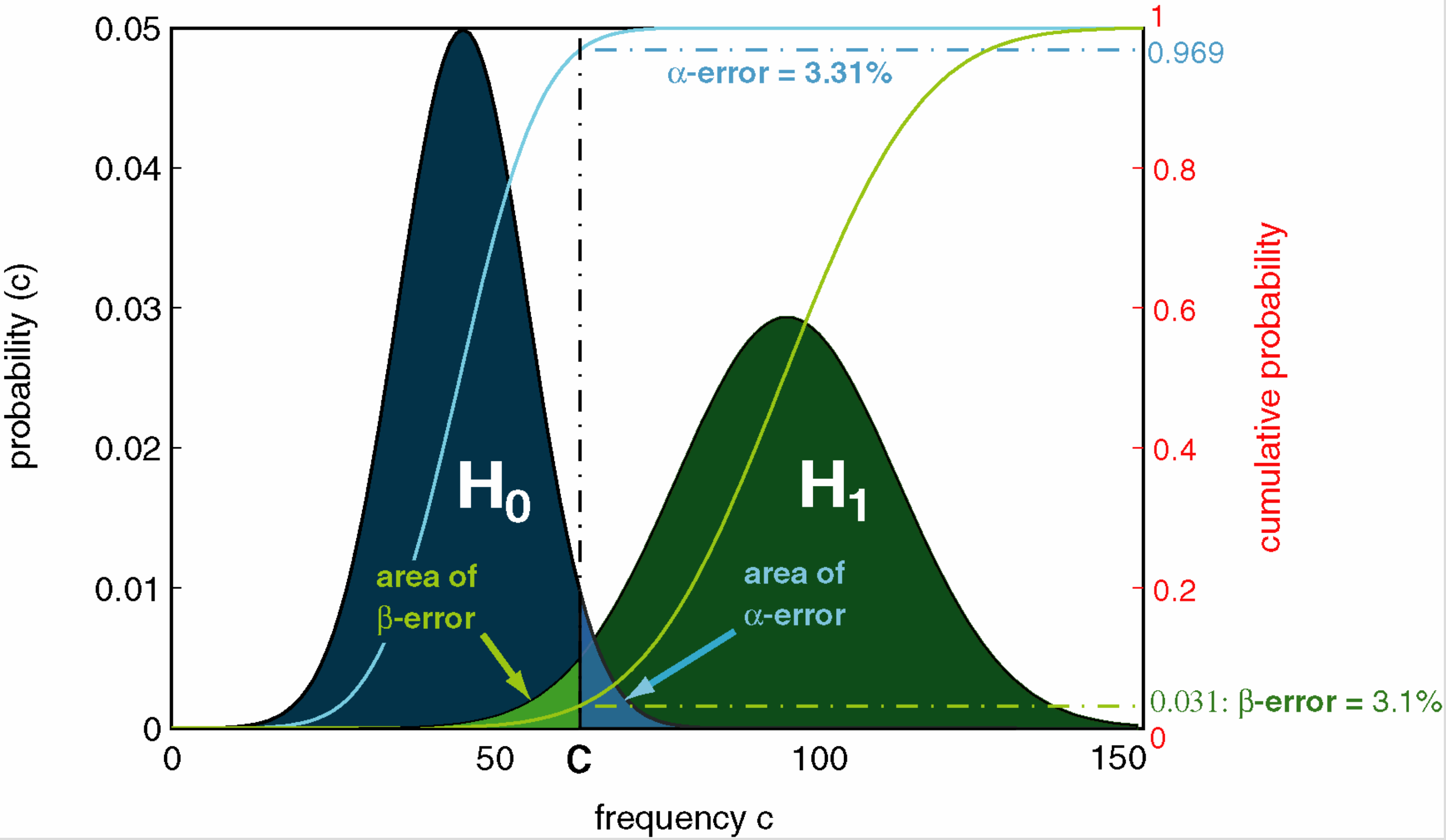


Describing *P-value*

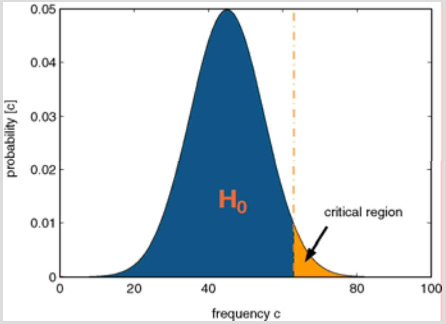
- If the *P-value* is less than 1%, there is ***overwhelming evidence*** that supports the alternative hypothesis.
- If the *P-value* is between 1% and 5%, there is ***strong evidence*** that supports the alternative hypothesis.
- If the *P-value* is between 5% and 10% there is ***weak evidence*** that supports the alternative hypothesis.
- If the *P-value* exceeds 10%, there is ***no evidence*** that supports of the alternative hypothesis.

- A ***Type II Error* or β Error** occurs when we incorrectly fail to reject the null hypothesis. The probability of a Type II Error is

$$\beta = P\left(\text{do not reject } H_0 \mid H_0 \text{ is false}\right)$$



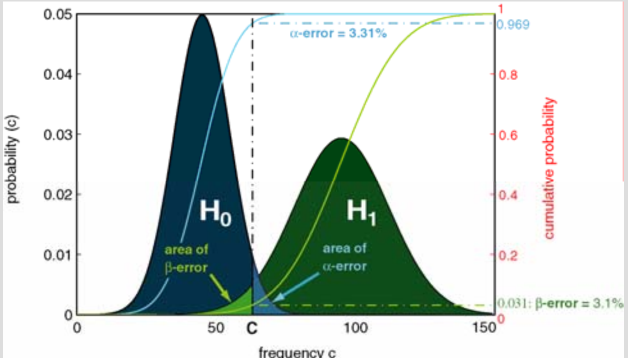
- A ***Type I Error or α Error*** occurs when we incorrectly reject the null hypothesis. The probability of a Type I Error is



$$\alpha = P\left(\text{reject } H_0 \mid H_0 \text{ is true}\right)$$

- A ***Type II Error* or β Error** occurs when we incorrectly fail to reject the null hypothesis. The probability of a Type II Error is

$$\beta = P\left(\text{do not reject } H_0 \mid H_0 \text{ is false}\right)$$



General procedure to test if the parameter is in a specific region

1. Look at the data & check assumptions.
2. State the null and alternative hypotheses.
3. Specify the desired significance level.
4. Specify the test statistic and its sampling distribution under the null.
5. Form a decision rule.
6. Compute the statistic and draw a conclusion.

Interval Statistics

1. **T-test (one sample) :**
   Tests is expected value μ_1 from the sampled population is equal μ_2

Ordinal Statistics

1. **Mann Whitney (one sample) :**
   Tests if the median μ_1 from the sampled population is equal μ_2

Why T-test and not standard confidence interval based on z

The standard deviation is underestimated for small n. The T-test corrects for this by accounting for different sample sizes (degree of freedom).

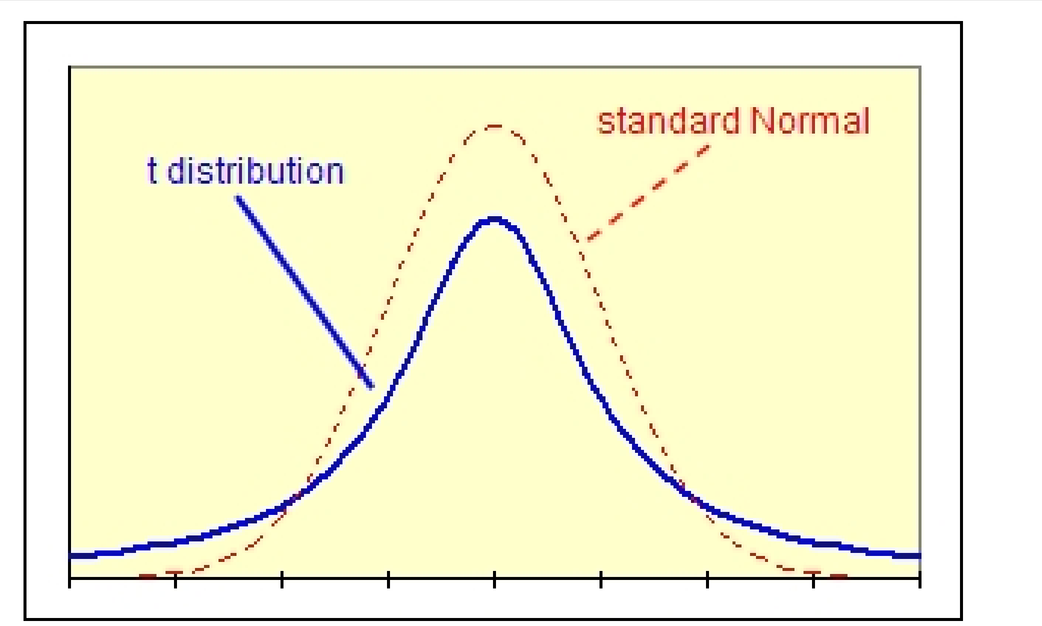
The test statistic z is NOT normal distributed:

$$z = \frac{\bar{x} - \mu}{s / \sqrt{n}} \qquad \cancel{\sim \mathbf{N(0,1)}}$$

Rather, it is distributed as a Student's t-distribution with n-1 degrees of freedom

$$\frac{\bar{x} - \mu}{s / \sqrt{n}} \sim \mathbf{t}(\mathbf{n\text{-}1})$$

- The t-distribution is completely characterized by only one parameter, the "degrees of freedom".
- The degrees of freedom is equal to n (the sample size) minus 1.
- The t-distribution looks like a standard normal, but has heavier tails. As n increases, the t-distribution begins to converge to a standard normal. You can explore the t-distribution on the sample Excel worksheet "Distributions"



T-test

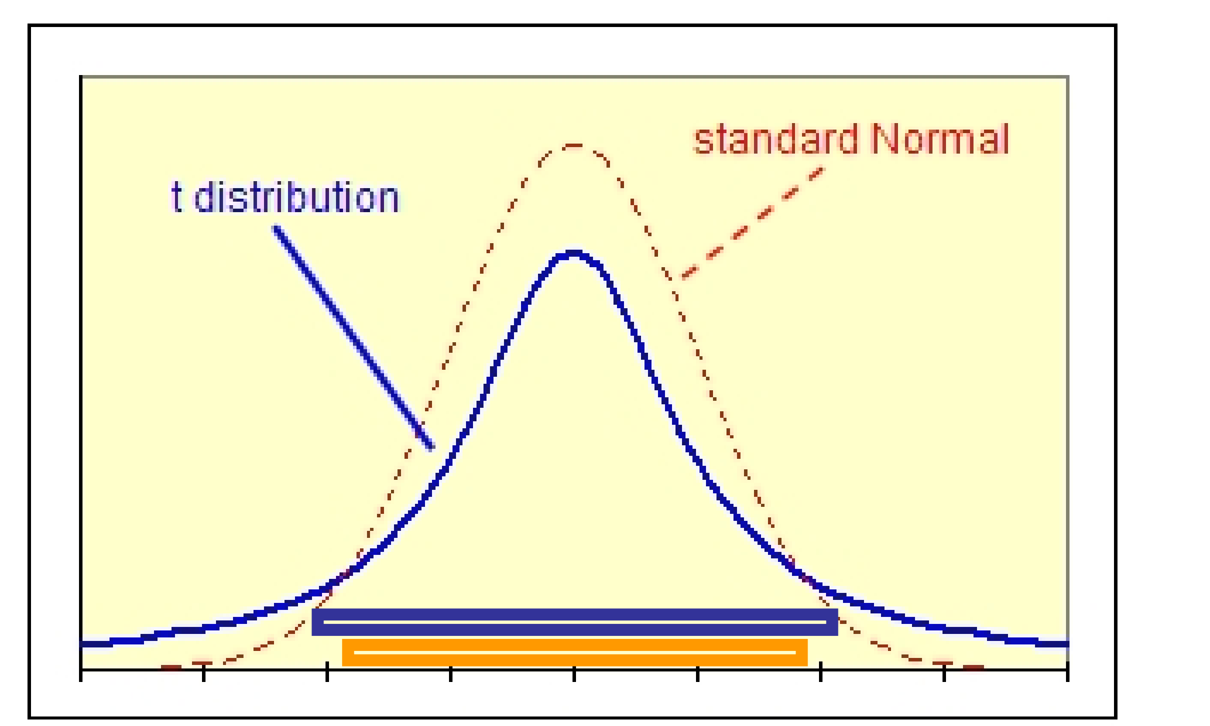
- So instead of:

$$P(\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

- We have:

$$P(\bar{x} - \boxed{t_{1-\frac{\alpha}{2},n-1}} \frac{s}{\sqrt{n}} < \mu < \bar{x} + \boxed{t_{1-\frac{\alpha}{2},n-1}} \frac{s}{\sqrt{n}}) = 1 - \alpha$$

- We want this to be set to α=0.05 (for a 95 percent confidence interval:
- So what are the t-values associated with setting alpha=0.05?
- It is not 1.96 like for the z
- Why? Because although the t-distribution is close to the standard normal, it is not the same as the standard normal (heavier tails) !!!



1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping

Why Ordinal Statistics

In case one is interested in qualitative studies rather than in quantitative studies ordinal statistics are more robust.

Example: A new teaching method is introduced

Interval Statistics:

Q: Is the average difference of scores (before-after) significantly different from 0 ? → Paired T-test

Ordinal Statistics:

Q: Do the new technique change scores → Wilcoxon Test

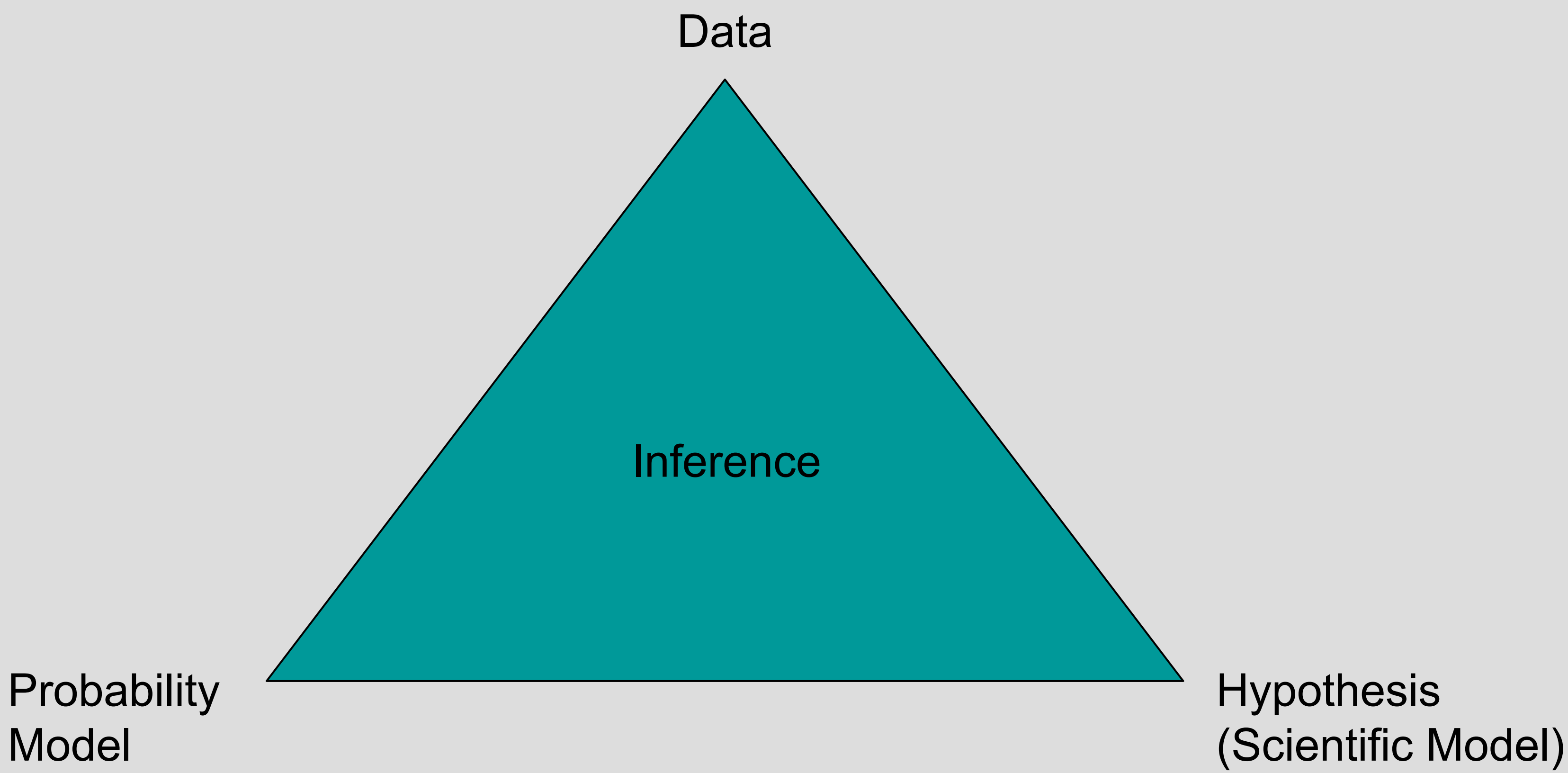
- Often want to perform more than one test on a data set

- If each test uses the same significance level then

  P(at least one Type I error) > α

1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping

Maximum Likelihood



Hypothesis defines class of the model, while data fits the parameters !

- Treat θ as a random variable that can be measured

$$\max_{\theta} p(\theta|d,m) = \max_{\theta} \frac{p(d|\theta,m)p(\theta|m)}{p(d|m)} = \max_{\theta} \frac{likelihood * prior}{evidence}$$

with m=model, θ=*parameter of the model,* d=*data*

Likelihood measures the match between data and the predicted model

The prior introduces and advanced belief about which values of the parameters are reasonable

The evidence measures how well one of many models can describe the data

- If we set p(θ)=p(d)=1 we use no prior and belief → Maximum Likelihood

$$\max_{\theta} p(\theta|d) = \max_{\theta} p(d|\theta) = \max_{\theta} likelihood$$

Maximum Likelihood in words:

We select a model that is parameterized by θ. *Than we maximize the probability that the model is describing the data by selecting the most optimal set of θ.*

→ ***We fit the model to the data***

$$\max_{\theta} p(\theta|d) = \max_{\theta} p(d|\theta) = \max_{\theta} likelihood$$

- Example:
  - Let $f(x, \Theta)$ be given by a Gaussian distribution function.
  - Let $\Theta = \mu$ be the mean of the Gaussian. We want to use our data+MLM to find the mean, μ.
  - We want the best estimate of α from our set of n measurements $\{x_1, x_2, \dots x_n\}$.
  - Let's assume that σ is the same for each measurement.
  - The likelihood function for this problem is: $f(x_i, \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}$

$$L = \prod_{i=1}^{n} f(x_i, \theta) =$$

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

Trick : log(p(d|Θ)) → Product becomes a sum

$$\max_\theta p(\theta|d) = \max_\theta p(d|\theta) = \max_\theta likelihood$$

We want to find the α that maximizes the log likelihood function:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(x_i - \alpha)^2}{2\sigma^2} \right] = 0$$

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (x_i - \alpha)^2 = 0$$

factor out σ since it is a constant

$$\sum_{i=1}^{n} 2(x_i - \alpha)(-1) = 0$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} {}_i\alpha = 0$$

don't forget the factor of n

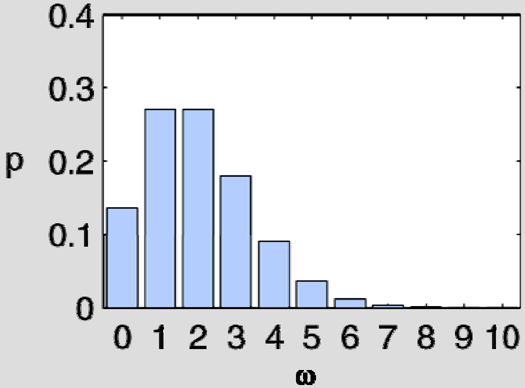
$$\sum_{i=1}^{n} x_i = n\alpha$$

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Average!

1. Concepts in statistical inference
2. Point estimates
3. Interval Estimates
4. Hypothesis Testing
5. Robustness of Tests
6. Maximum Likelihood
7. Bootstrapping

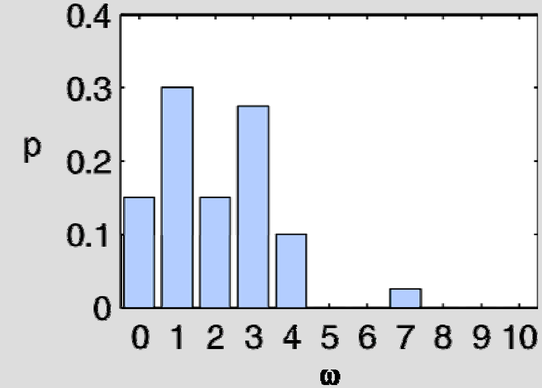
real distribution



experiment :

- estimation of distribution parameter
- set of n measurements (ω_i), with i=1..n $F \to \omega = (\omega_1, \omega_2, \ldots, \omega_n)$

empirical distribution



$\Phi(\overline{\omega}) = \Theta(F)$ with $\sigma(\Phi)$

$F \cong \hat{F}$

plug-in principle

$\Phi^*(\overline{\omega}^*) = \Theta(\hat{F})$ with $\sigma^*(\Phi^*)$

estimation of parameter of F by resampling of the empirical distribution :

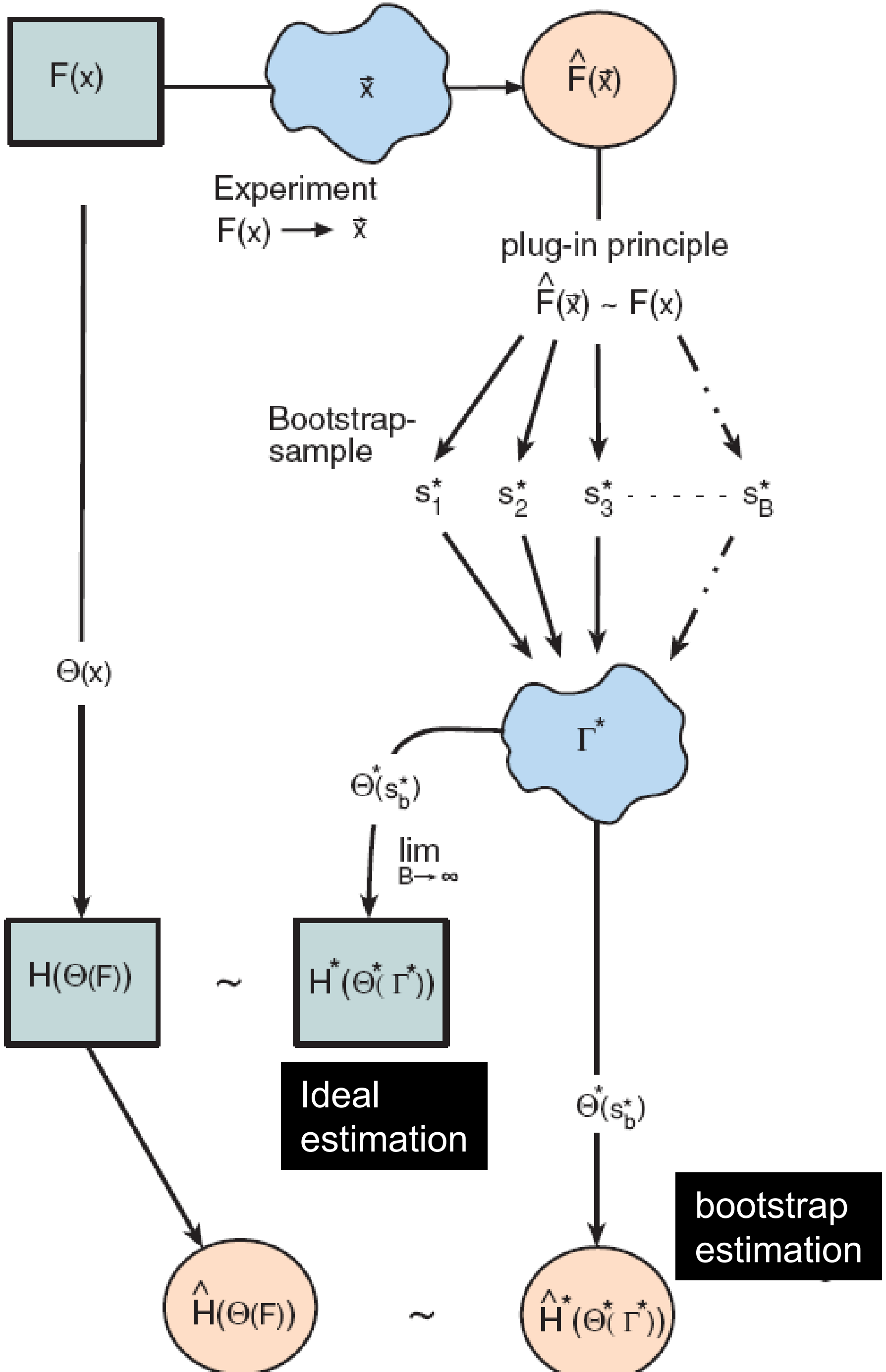
Resampling ⇔ **Experiment based on**

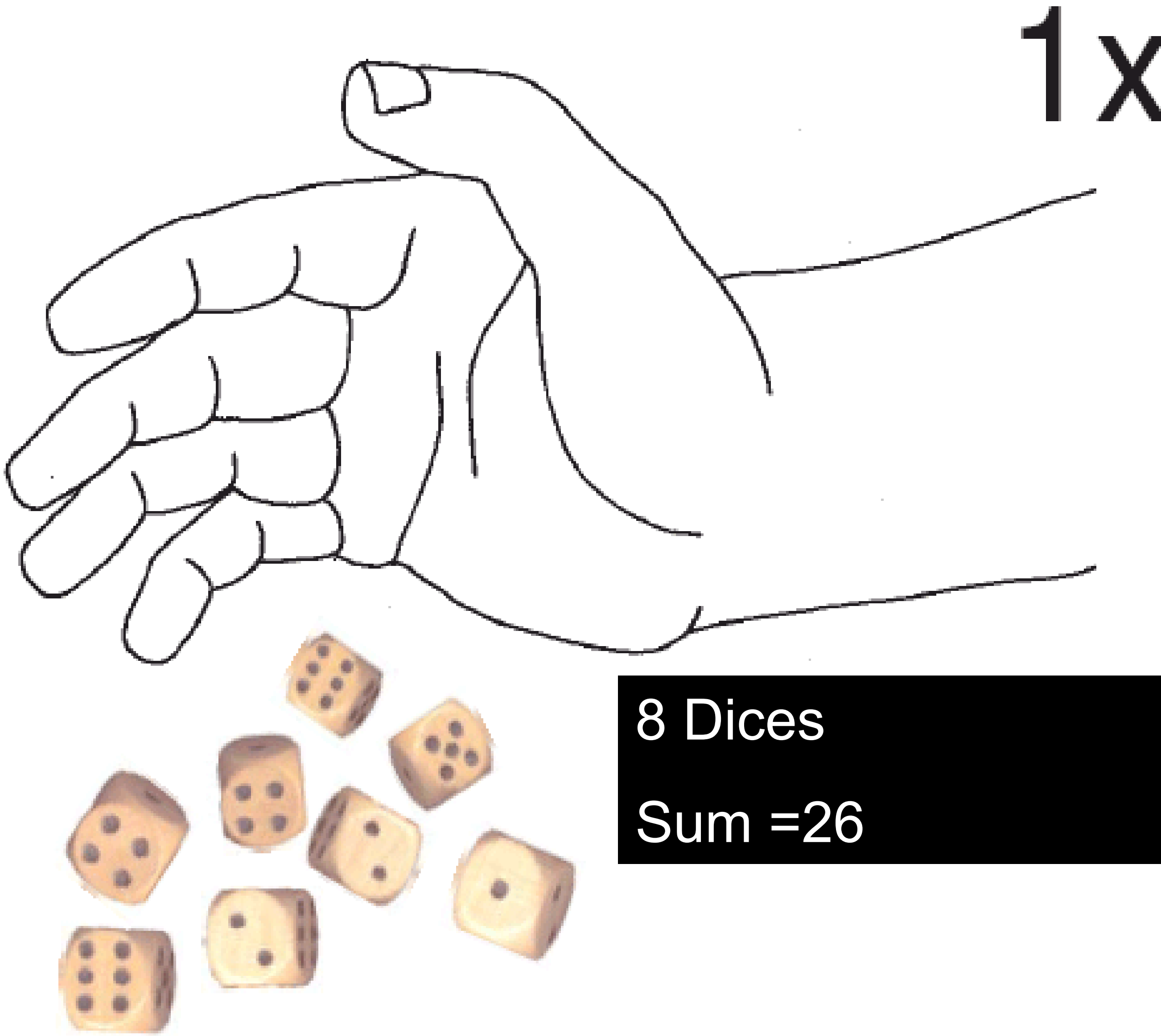
sample : $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$

empirical distribution : $\hat{F} = \hat{F}(\omega)$

Bootstrap sample : $\omega_b^* = (\omega^*_1, \omega^*_2, \ldots, \omega^*_n)_b$

Bootstrap estimation : $\overline{\omega}^* = \langle \omega_b^* \rangle = \Theta(\hat{F})$ with $\sigma(\overline{\omega}^*)$


F(x)
$\vec{x}$
$\hat{F}(\vec{x})$
Experiment
F(x) → $\vec{x}$
plug-in principle
$\hat{F}(\vec{x}) \sim F(x)$
Bootstrap-
sample
s_1^* s_2^* s_3^* - - - - - s_B^*
$\Theta(x)$
Γ^*
$\Theta^*(s_b^*)$
$\lim_{B \to \infty}$
$H(\Theta(F))$
~
$H^*(\Theta^*(\Gamma^*))$
Ideal
estimation
$\Theta^*(s_b^*)$
bootstrap
estimation
$\hat{H}(\Theta(F))$
~
$\hat{H}^*(\Theta^*(\Gamma^*))$

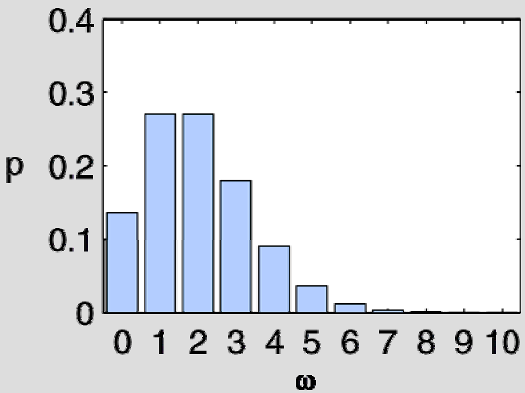


8 Dices

Sum =26

Bootstrapping : the plug-in-principle

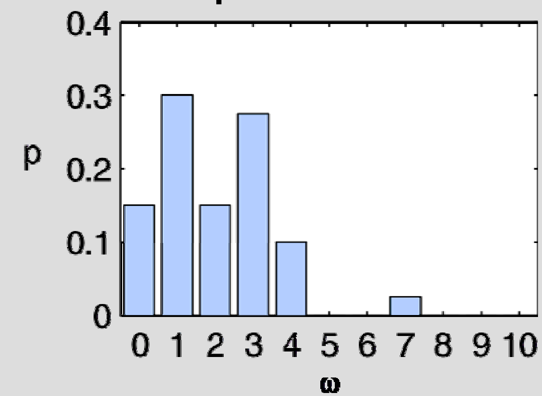
real distribution



experiment :

- estimation of distribution parameter
- set of n measurements (ω_i),
  with i=1..n $\quad F \rightarrow \omega = (\omega_1, \omega_2, \ldots, \omega_n)$

empirical distribution



$$\Phi(\overline{\omega}) = \Theta(F) \quad \text{with} \quad \sigma(\Phi)$$

$$F \cong \hat{F}$$

plug-in principle

$$\Phi^*(\overline{\omega}^*) = \Theta(\hat{F}) \quad \text{with} \quad \sigma^*(\Phi^*)$$

estimation of parameter of F by resampling of the empirical distribution :

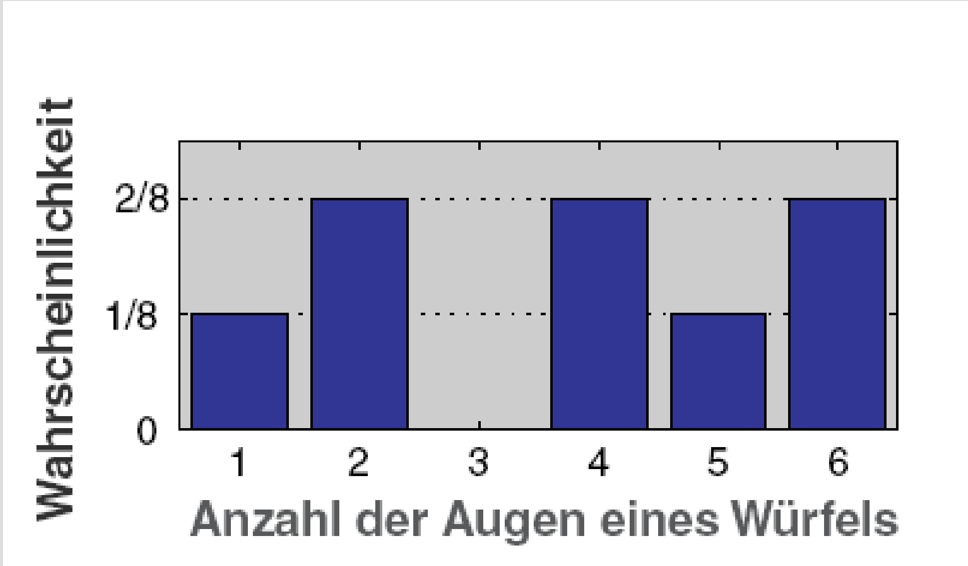
Resampling $\Leftrightarrow$ **Experiment based on**

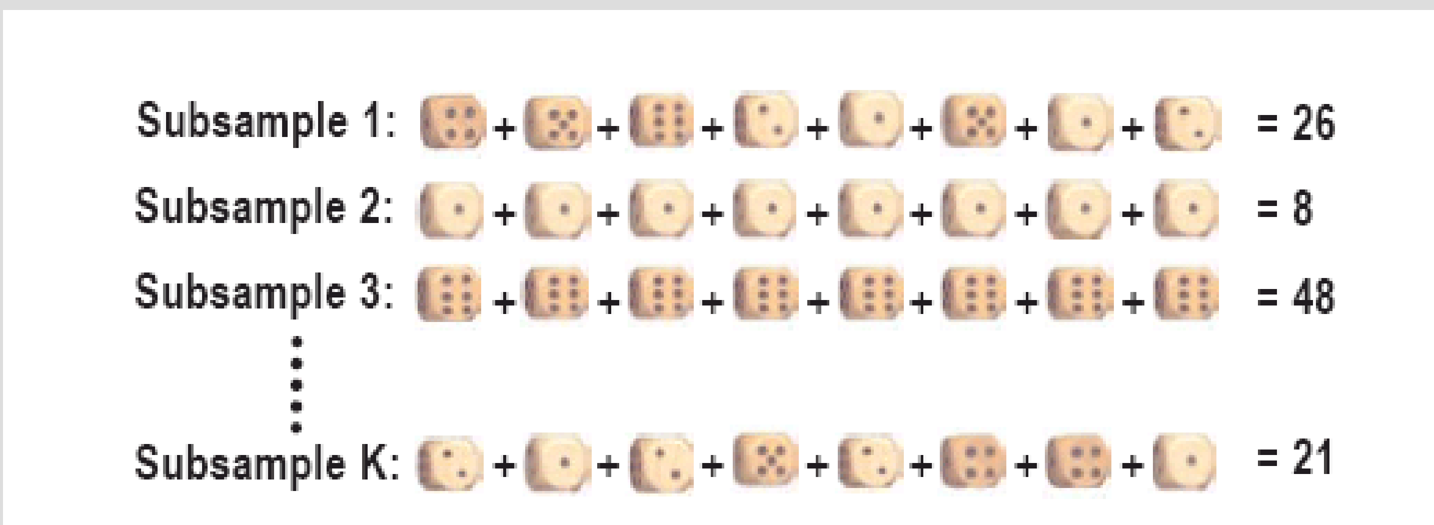
sample : $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$

empirical distribution : $\hat{F} = \hat{F}(\boldsymbol{\omega})$

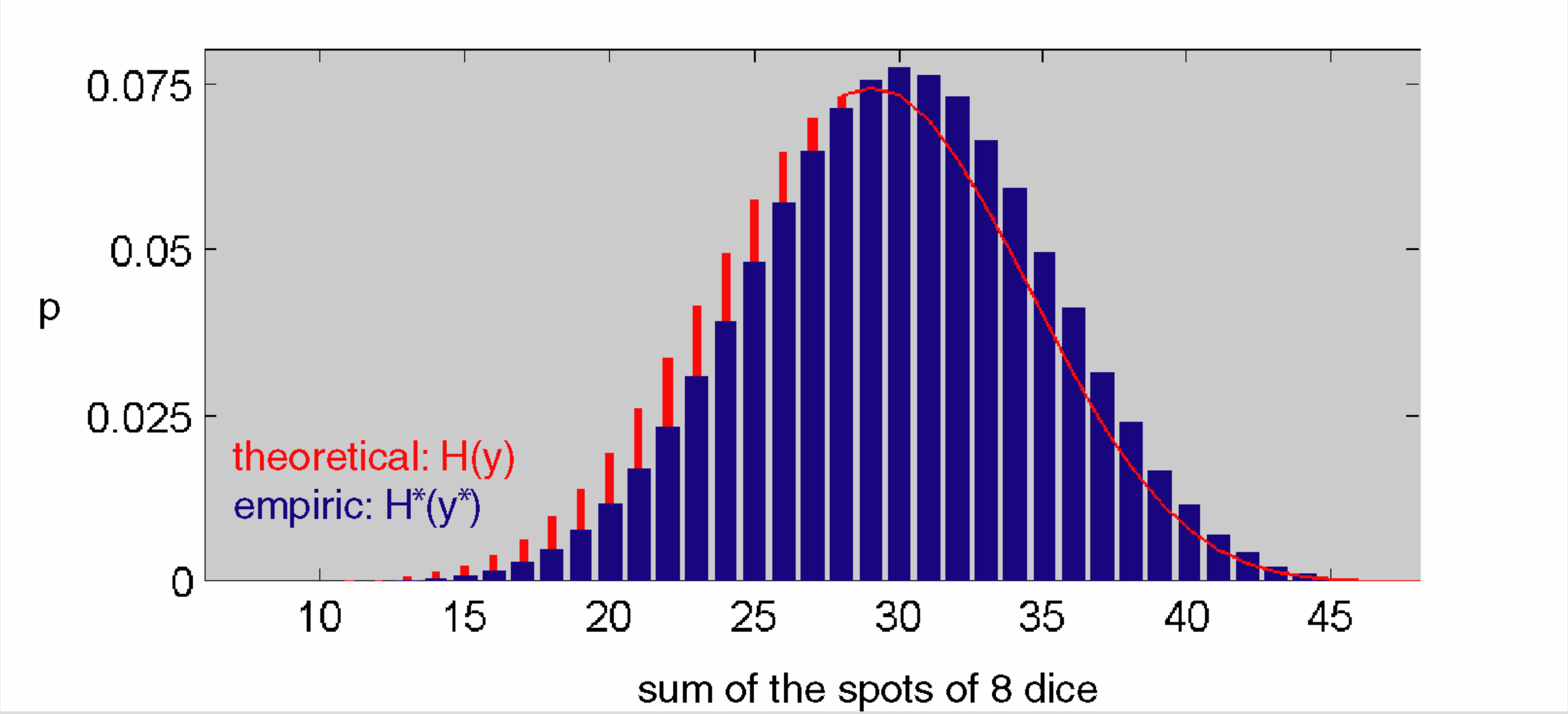
Bootstrap sample : $\omega_b{}^{*} = (\omega^*{}_1, \omega^*{}_2, \ldots, \omega^*{}_n)_b$

Bootstrap estimation : $\overline{\omega}^* = \langle \boldsymbol{\omega}_b^* \rangle = \Theta(\hat{F}) \quad \text{with} \quad \sigma(\overline{\omega}^*)$


Wahrscheinlichkeit
2/8
1/8
0
1
2
3
4
5
6
Anzahl der Augen eines Würfels


Subsample 1: + + + + + + + = 26
Subsample 2: + + + + + + + = 8
Subsample 3: + + + + + + + = 48
Subsample K: + + + + + + + = 21

Bootstrapped distribution of the mean



Properties of Bootstrapping:

1. Allows to derive any Statistics without having an analytical approach
2. Statistically robust since number of assumptions is reduced in comparison to model approaches
3. Computationally more demanding

Can be applied for

1. Estimation of confidence intervals and significance
2. Estimation of residuals for Maximum likelihood estimates
3. On any odd statistic that is not analytically treatable.

The end