Mathematical Statistics

Anna Janicka

Lecture IX, 26.04.2021

CONFIDENCE INTERVALS – cont.

HYPOTHESIS TESTING

Plan for Today

1. Confidence intervals – cont.
2. A statistical hypothesis
3. A statistical test
4. Type I and type II errors
5. Significance level, *p*-value
6. Testing scheme
7. Power of a test

Most commonly used models for CI

- *Model I (normal): CI for the mean, variance known*
- *Model II (normal): CI for the mean, variance unknown*
- *Model II (normal): CI for the variance*
- *Model III (asymptotic): CI for the mean*
- *Model IV (asymptotic): CI for the fraction*
- Asymptotic model: CI based on MLE

CI for the mean – Model III – reminder

Asymptotic model: $X_1, X_2, \ldots, X_n$ are an IID sample from a distr. with mean (μ) and variance, n – large.

Approximate CI for μ, for a confidence level $1-\alpha$:

$$\left[\overline{X} - u_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + u_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

where $u_{1-\alpha/2}$ is a quantile of rank $1-\alpha/2$ from the N(0,1) distribution, $S = \sqrt{S^2}$ for the unbiased estimator of the variance S^2.

Justification: from CLT, when $n \to \infty$ we have

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \xrightarrow{D} N(0,1)$$

CI for the fraction – Model IV – reminder

Asymptotic model: $X_1, X_2, \ldots, X_n$ are an IID sample from a two-point distribution, n – large.

$$P_p(X=1) = p = 1 - P_p(X=0)$$

Approximate CI for p, for a confidence level $1-\alpha$:

$$\left[\hat{p} - u_{1-\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + u_{1-\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right]$$

where $u_{1-\alpha/2}$ is a quantile of rank $1-\alpha/2$ from the N(0,1) distribution

CI for the fraction – Model IV, properties

- Assessment error: $$d = u_{1-\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$
- Sample size allowing to obtain a given *precision* (error) *d*:

$$n \geq \frac{\hat{p}(1-\hat{p})u_{1-\alpha/2}^{2}}{d^{2}}$$

if we do not know anything about *p*, we need to consider the worst scenario where *p*=1/2: $$n \geq \frac{u_{1-\alpha/2}^{2}}{4d^{2}}$$

e.g. $1{,}645^2/(4 * 0{,}025^2) \approx 1082$

CI on the base of the MLE – Asymptotic model

Asymptotic model: $X_1, X_2, \ldots, X_n$ are an IID sample from a distr. with unknown parameter θ, n – large.

If $\hat{\theta} = MLE(\theta)$ is asymptotically normal with an asymptotic variance equal to $1/I_1(\theta)$, i.e.

$$(\hat{\theta} - \theta)\sqrt{n} \xrightarrow{D} N(0, 1/I_1(\theta))$$

and if $I(\hat{\theta}) = MLE(I(\theta))$ is consistent, and we have:

$$(\hat{\theta} - \theta)\sqrt{nI(\hat{\theta})} \xrightarrow{D} N(0,1)$$

Approximate CI for θ, for a confidence level 1-α :

$$\left[\hat{\theta} - u_{1-\alpha/2}\frac{1}{\sqrt{nI_1(\hat{\theta})}}, \hat{\theta} + u_{1-\alpha/2}\frac{1}{\sqrt{nI_1(\hat{\theta})}}\right]$$

where $u_{1-\alpha/2}$ is a quantile of rank 1-$\alpha/2$ from N(0,1)

CI on the base of the MLE – Asymptotic model, general case

Asymptotic model: $X_1, X_2, \ldots, X_n$ are an IID sample from a distr. with unknown parameter θ, n – large.

If $g(\hat{\theta}) = g(MLE(\theta))$ is asymptotically normal with an asymptotic variance equal to $(g'(\theta))^2 / I_1(\theta)$, i.e.

$$(\hat{\theta} - \theta)\sqrt{n} \xrightarrow{D} N(0, (g'(\theta))^2 / I_1(\theta))$$

and if $I(\hat{\theta}) = MLE(I(\theta))$ is consistent, and we have:

$$(\hat{\theta} - \theta)\sqrt{nI(\hat{\theta})} \xrightarrow{D} N(0,1)$$

Approximate CI for $g(\theta)$, for a confidence level $1-\alpha$:

$$\left[g(\hat{\theta}) - u_{1-\alpha/2} \frac{|g'(\hat{\theta})|}{\sqrt{nI_1(\hat{\theta})}}, g(\hat{\theta}) + u_{1-\alpha/2} \frac{|g'(\hat{\theta})|}{\sqrt{nI_1(\hat{\theta})}} \right]$$

where $u_{1-\alpha/2}$ is a quantile of rank $1-\alpha/2$ from N(0,1)

CI on the base of the MLE – Example

Let $X_1, X_2, \ldots, X_n$ be an IID sample from a Poisson distr. with unknown parameter θ, n – large.

$\hat{\theta} = MLE(\theta) = \overline{X}$ is asymptotically normal (CLT) with an asymptotic variance equal to $1/I_1(\theta) = \theta$

$\hat{I}(\theta) = 1/\hat{\theta}$ behaves well.

Approximate CI for θ, for a confidence level 1-α :

$$\left[\overline{X} - u_{1-\alpha/2}\frac{\sqrt{\overline{X}}}{\sqrt{n}}, \overline{X} + u_{1-\alpha/2}\frac{\sqrt{\overline{X}}}{\sqrt{n}}\right]$$

where $u_{1-\alpha/2}$ is a quantile of rank 1-$\alpha/2$ from N(0,1)

For example, if for n=900 we had $\overline{X} = 4$, then the 90% CI for θ would be $\approx \left[4 - 1.645\sqrt{4/900}, 4 + 1.645\sqrt{4/900}\right] \approx [3.89, 4.11]$

CI on the base of the MLE – Example cont.

If we wanted to approximate the probability of the outcome = 0, we would look for $g(\theta) = e^{-\theta}$

$$g(\hat{\theta}) = g(MLE(\theta)) = e^{-\overline{X}}$$

And the *approximate* CI for $g(\theta)$, for a confidence level 1-α :

$$\left[e^{-\overline{X}} - u_{1-\alpha/2} \frac{\sqrt{\overline{X}}}{\sqrt{n}} e^{-\overline{X}}, e^{-\overline{X}} + u_{1-\alpha/2} \frac{\sqrt{\overline{X}}}{\sqrt{n}} e^{-\overline{X}} \right]$$

where $u_{1-\alpha/2}$ is a quantile of rank 1-$\alpha/2$ from N(0,1)

For example, if for n=900 we had $\overline{X} = 4$, then the 90% CI for $g(\theta)$ would be

$$\approx \left[e^{-4} - 1.645\sqrt{4/900}\, e^{-4}, e^{-4} + 1.645\sqrt{4/900}\, e^{-4} \right] \approx [0.016, 0.020]$$

A statistical hypothesis

a statement regarding the probability distribution governing the phenomenon of interest (the random variable observed)

Aim: we want to draw conclusions about the validity of the hypothesis based on observed values of the random variable

Examples of statistical hypotheses

- $X_1, X_2, ..., X_n$ are a sample from an exponential distribution
- $X_1, X_2, ..., X_n$ are a *sample from a normal distribution* (assumption) with param (5, 1)
- $EX_i = 7$ (the expected value of the distr is 7)
- Var $X_i > 1$ (the variance of the distribution exceeds 1)
- $X_1, X_2, ..., X_n$ are independent
- $EX_i=EY_j$ ($X_1, X_2, ..., X_n$ and $Y_1, Y_2, ..., Y_m$ have the same expected value)

Types of hypotheses

- hypothesis
  - parametric: concerning the value of distribution parameters
  - nonparametric: concerning other properties of the distribution

- hypothesis
  - simple: specifies a single distribution
  - composite: specifies a family of distributions

Null and alternative hypotheses

Null hypothesis: “basic”, denoted H_0

Alternative hypothesis: hypothesis which is accepted if the null is rejected, denoted H_1

e.g.:

- $H_0 : \lambda = 1, \quad H_1 : \lambda \neq 1$
- $H_0 : \lambda = 1, \quad H_1 : \lambda = 2$
- $H_0 : \lambda = 1, \quad H_1 : \lambda > 1$

Null and alternative hypotheses – cont.

The null and alternative hypotheses do not have equal status.

Null hypothesis: a statement, perhaps based on existing theory, deemed true until there appear observations very hard to reconcile with the statement. Speculative hypothesis.

Alternative hypothesis: the possibility taken into account when we are forced to reject the null hypothesis

Statistical test

A procedure, which for any sample of observations (any possible set of values) leads to one of two decisions:

- reject the null hypothesis (in favor of the alternative)
- do not reject the null hypothesis

reject H_0

no grounds to reject H_0

Statistical test, formally

Point of departure: statistical model

- $X = (X_1, X_2, \ldots, X_n)$ – vector of observations $\in \boldsymbol{X}$
- $X \sim P_\theta$, $\{P_\theta : \theta \in \Theta\}$ – a family of distributions

Hypotheses H_0, H_1 :

- $H_0 : \theta \in \Theta_0$
- $H_1 : \theta \in \Theta_1$

such that $\Theta_0 \cap \Theta_1 = \varnothing$

(the hypotheses are mutually exclusive)

Statistical test, formally – cont.

A test of H_0 **against** H_1 :

Statistic $\delta : \boldsymbol{X} \to \{0,1\}$

the value 1 is interpreted as rejection of H_0 (in favor of H_1) and 0 as not rejecting H_0

Region of rejection (critical region):

$C = \{x \in \boldsymbol{X} : \delta(x) = 1\}$ – set of values for which we reject H_0;

Region of acceptance:

$A = \{x \in \boldsymbol{X} : \delta(x) = 0\}$ – set of values for which we do not reject H_0

$C \cup A = \boldsymbol{X}$, $C \cap A = \varnothing$

Statistical test, formally – cont. (2)

The critical region of a test usually takes the form

$$C = \{x \in \boldsymbol{X} : T(x) > c\}$$

for a selected statistic *T* (**test statistic**) and a value *c* (**critical value**)

Equivalent descriptions of a test:

- specification of *T* and *c*
- specification of *C*
- specification of δ

in many cases by a **critical region** one means the range of values of the statistic, and not the range of observed values

Statistical test – example

We want to verify whether a coin is symmetric

We toss the coin 400 times

$X \sim B(400, p)$

- $H_0 : p = ½$, $H_1 : p \neq ½$
- Some results may suggest rejection of H_0:
  - $|X – 200| < c$ – do not reject H_0.
  - $|X – 200| \geq c$ – reject H_0 in favor of H_1.

i.e. $T(x) = |x – 200|$

→ how do we choose *c*?

Type I and type II errors

There is always a possibility of error due to randomness of observations

decision	In reality we have	
	H_0 true	H_0 false
reject H_0	**Type I error**	OK
do not reject H_0	OK	**Type II error**

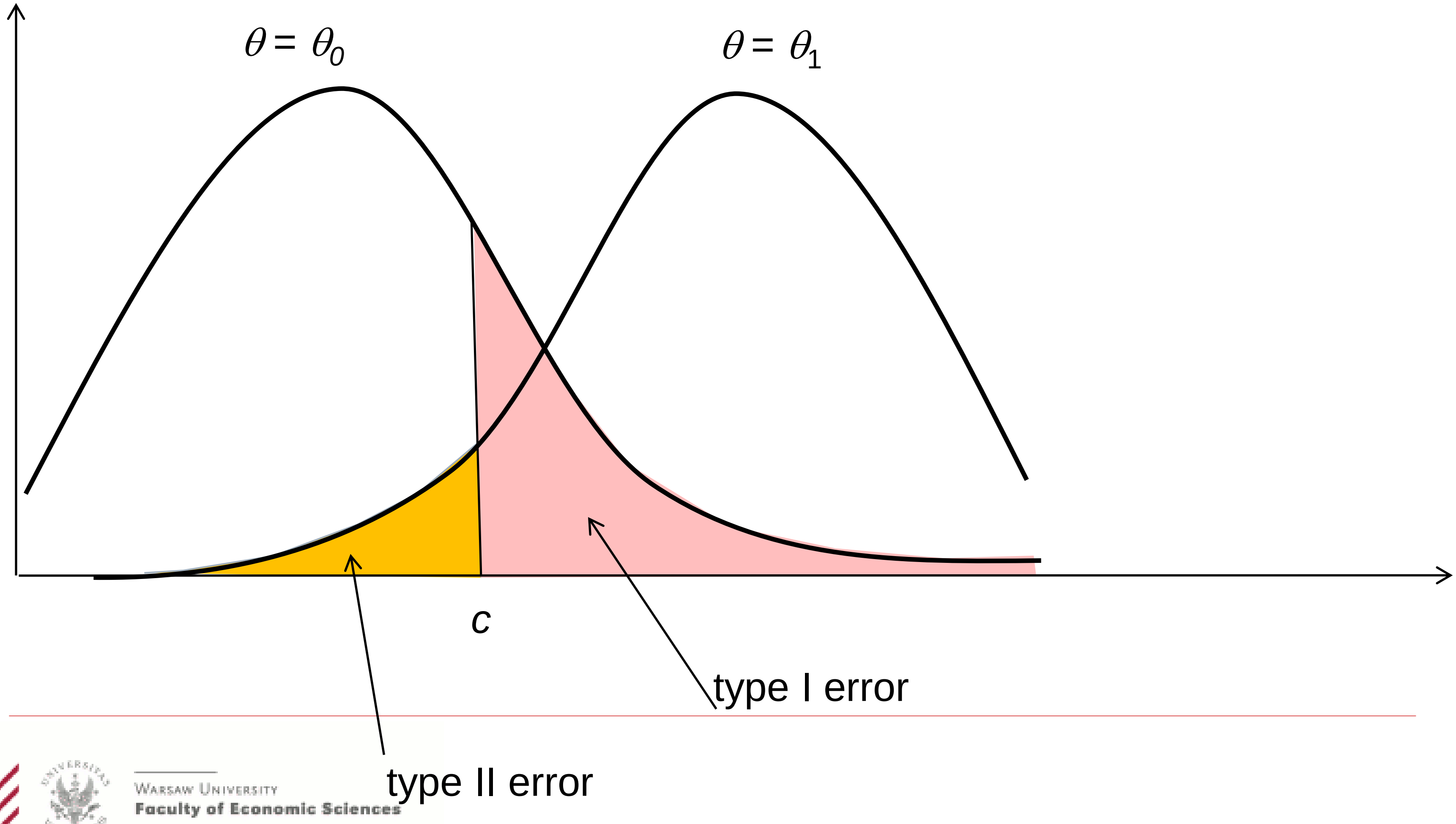
$P_\theta(C)$ for $\theta \in \Theta_0$ – probability of type I error

$P_\theta(A)$ for $\theta \in \Theta_1$ – probability of type II error

there is a trade-off between errors of Ist and IInd type: it's impossible to minimize both simultaneously

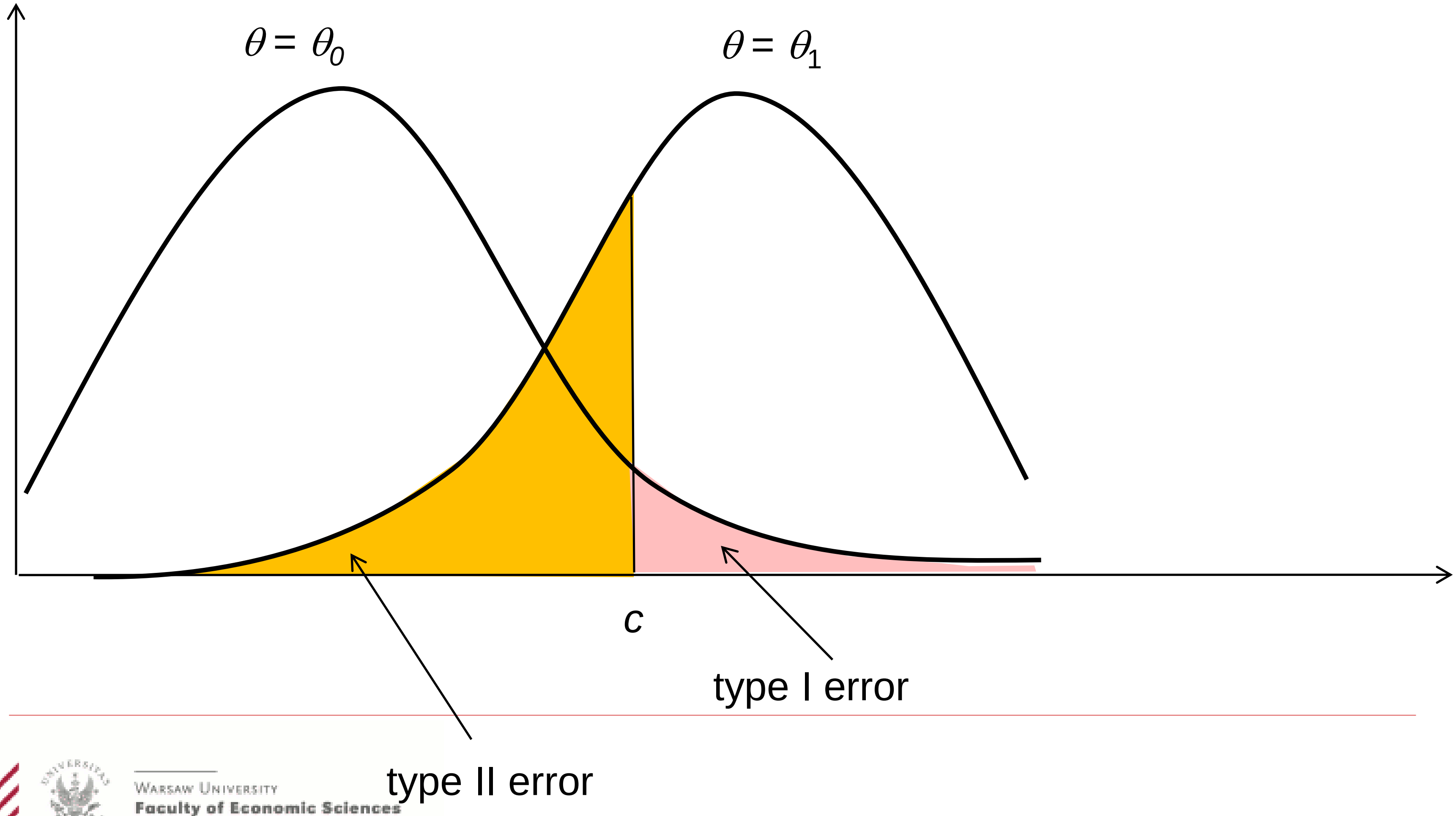
Type I and type II errors: graphical interpretation (1)

distributions of the test statistic T assuming that the null and alternative hypotheses are true



Type I and type II errors: graphical interpretation (2)

distributions of the test statistic T assuming that the null and alternative hypotheses are true



Significance level

A test has a **significance level** α, if for any $\theta \in \Theta_0$ we have $P_\theta(C) \leq \alpha$.

Usually: we look for tests with minimal probability of type II error for a given level of significance α, usually = 0.1 or 0.05 or 0.01

Type I error usually more important – not only conservatism

Statistical test – example cont.
Finding the critical range

We want: significance level $\alpha = 0.01$

We look for c such that (assuming $p = ½$)

$$P(|X - 200| > c) = 0.01$$

From the de Moivre-Laplace theorem for large n!

$$P(|X - 200| > c) \approx 2\,\Phi(-c/10), \text{ to get}$$

$$= 0.01 \text{ we need } c \approx 25.8$$

For a significance level approximately 0.01 we reject H_0 when the number of tails is lower than 175 or higher than 225

$$C = \{0,1,\ldots,174\} \cup \{226, 227,\ldots, 400\}$$

Statistical test – example cont. (2). p-value

Slightly different question: what if the number of tails were 220 (T = 20)?

We have:

$$P_{½}(|X - 200| > 20) \approx 0.05$$

p-value: probability of type I error, if the value of the test statistic obtained was the critical value

So: p-value for T = 20 is approximately 0.05

p-value

p-value – probability of obtaining results *at least as extreme* as the ones obtained

(contradicting the null at least as much as those obtained)

decisions:

- p-value < α – reject the null hypothesis
- p-value ≥ α – no grounds to reject the null hypothesis

Statistical test – example cont. (3)
The choice of the alternative hypothesis

For a different alternative...

For example, we lose if tails appear *too often*.

- $H_0 : p = ½$, $H_1 : p > ½$
- Which results would lead to rejecting H_0?
  - $X – 200 \leq c$ – do not reject H_0.
  - $X – 200 > c$ – reject H_0 in favor of H_1.

  i.e. $T(x) = x – 200$

we could have
$H_0: p \leq ½$

Statistical test – example cont. (4)
The choice of the alternative hypothesis

Again, from the de Moivre – Laplace theorem:

$P_{½}(X – 200 > c) \approx 0.01$ for $c \approx 23.3$,

so for a significance level 0.01 we reject $H_0 : p = ½$ in favor of $H_1 : p > ½$ if the number of tails is at least 224

What if we got 220 tails?

p-value is equal to ≈ 0.025; do not reject H_0

Scheme of conducting a statistical test

1. Definition of the statistical model
2. Posing hypotheses: H_0 and H_1
3. Choice of significance level α
4. Choice of the test statistic T / defining the critical region C
5. Decision: depends on whether the value of the test statistic falls into the critical region (or based on comparison of the p-value and α)

Power of the test (for an alternative hypothesis)

$P_\theta(C)$ for $\theta \in \Theta_1$ – power of the test (for an alternative hypothesis)

Function of the power of a test:

$1-\beta : \Theta_1 \to [0,1]$ such that $1-\beta(\theta) = P_\theta(C)$

Usually: we look for tests with a given level of significance and the highest power possible.

Statistical test – example cont. (5)
Power of the test

- We test $H_0 : p = ½$ against $H_1 : p = ¾$
  with: $T(x) = X – 200$, $C = \{T(x) > 23.3\}$
  (i.e. for a significance level $\alpha = 0.01$)
  Power of the test:
  $1-\beta(¾) = P(T(x) > 23.3 \mid p = ¾) = P_{¾}(X>223.3)$
  $\approx 1-\Phi((223.3-300)/5\sqrt{3}) \approx \Phi(8.85) \approx 1$
- But if $H_1 : p = 0.55$

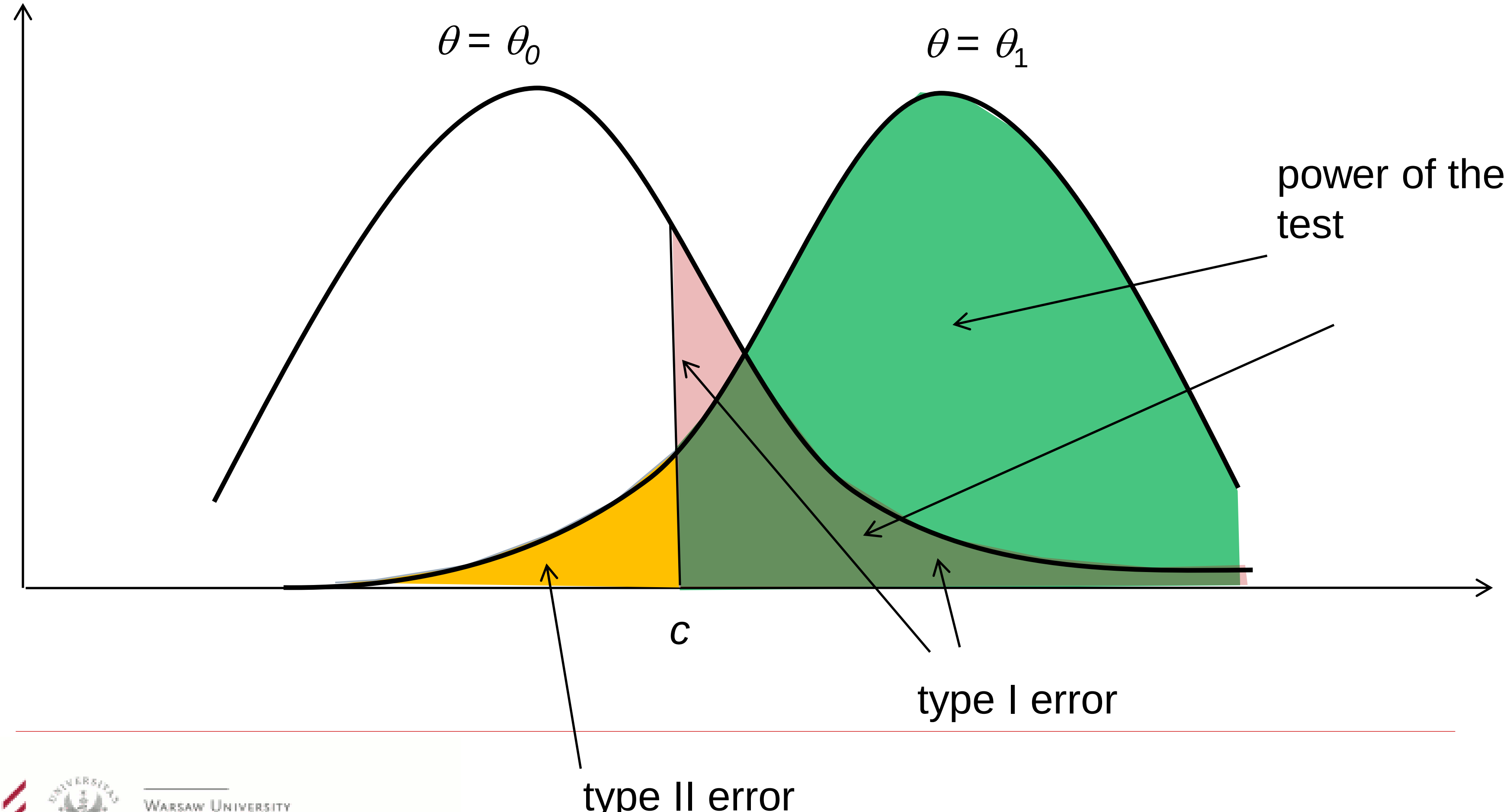
$1-\beta(0.55) = P(T(x) > 23.3 \mid p = 0.55) \approx 1-\Phi(0.33) \approx 1-0.63 \approx 0.37$

- And if $H_1 : p = ¼$ for the same T we would get

$1-\beta(¼) = P(T(x) > 23.3 \mid p = ¼) \approx 1-\Phi(14.23) \approx 0$

Power of the test: Graphical interpretation (1)

distributions of the test statistic T assuming that the null and alternative hypotheses are true



Power of the test: Graphical interpretation (2)

distributions of the test statistic T assuming that the null and alternative hypotheses are true