

Mathematical Statistics

Anna Janicka

Lecture IX, 23.04.2018

HYPOTHESIS TESTING

Plan for Today

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



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

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

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

Hypotheses H_0, H_1 :

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

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

(the hypotheses are mutually exclusive)



Statistical test, formally – cont.

A test of H_0 against H_1 :

Statistic $\delta : \mathbf{X} \rightarrow \{0,1\}$

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):

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

Region of acceptance:

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

$$K \cup A = \mathbf{X}, K \cap A = \emptyset$$

Statistical test, formally – cont. (2)

The critical region of a test usually takes the form

$$K = \{x \in \mathbf{X} : T(x) > c\}$$

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 K
- 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

$$X \sim B(400, p)$$

□ $H_0 : p = 1/2, \quad H_1 : p \neq 1/2$

□ 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

$P_{\theta}(K)$ for $\theta \in \Theta_0$ – probability of type I error

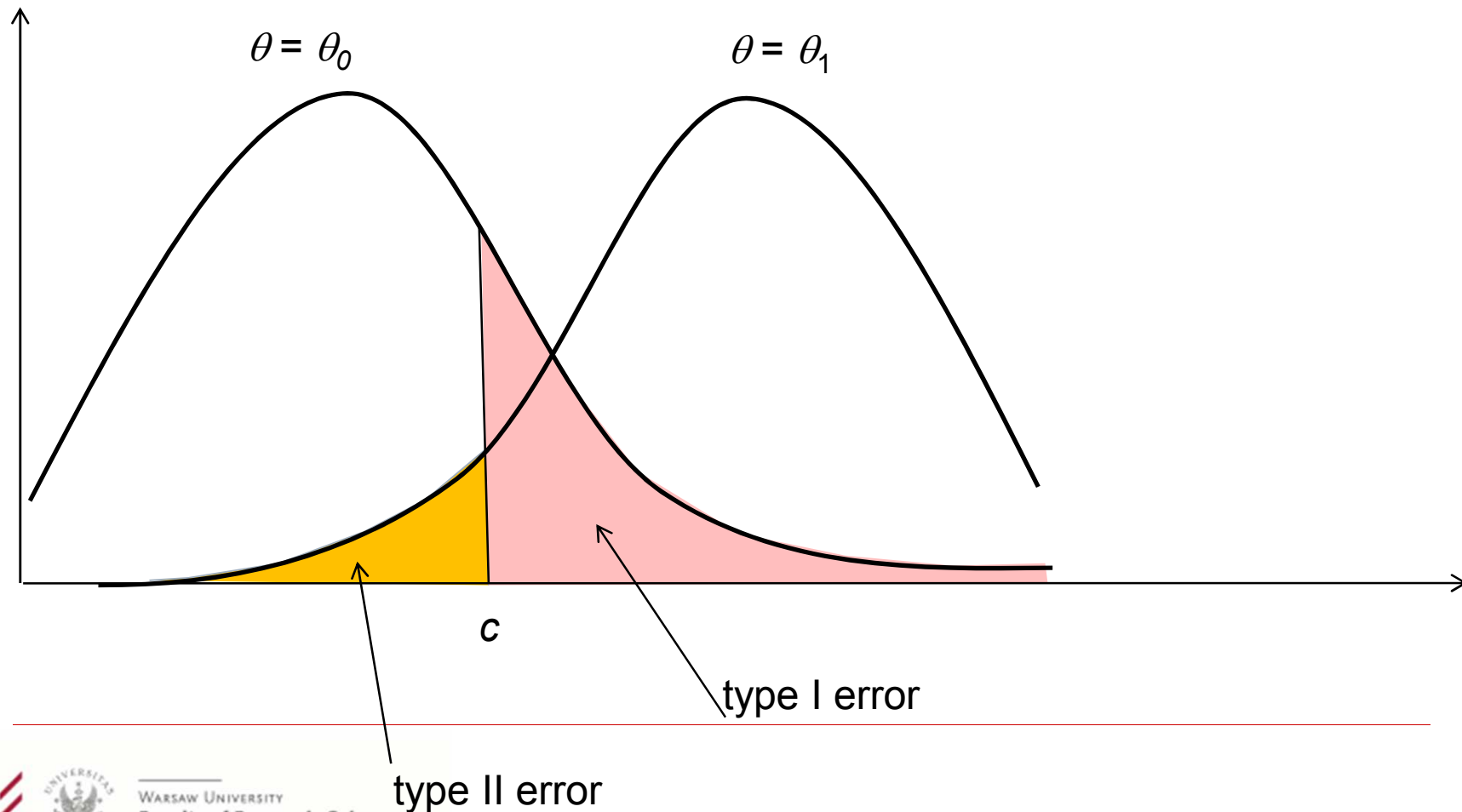
$P_{\theta}(A)$ for $\theta \in \Theta_1$ – probability of type II error

there is a trade-off between errors of 1st and 2nd type:
it's impossible to minimize both simultaneously



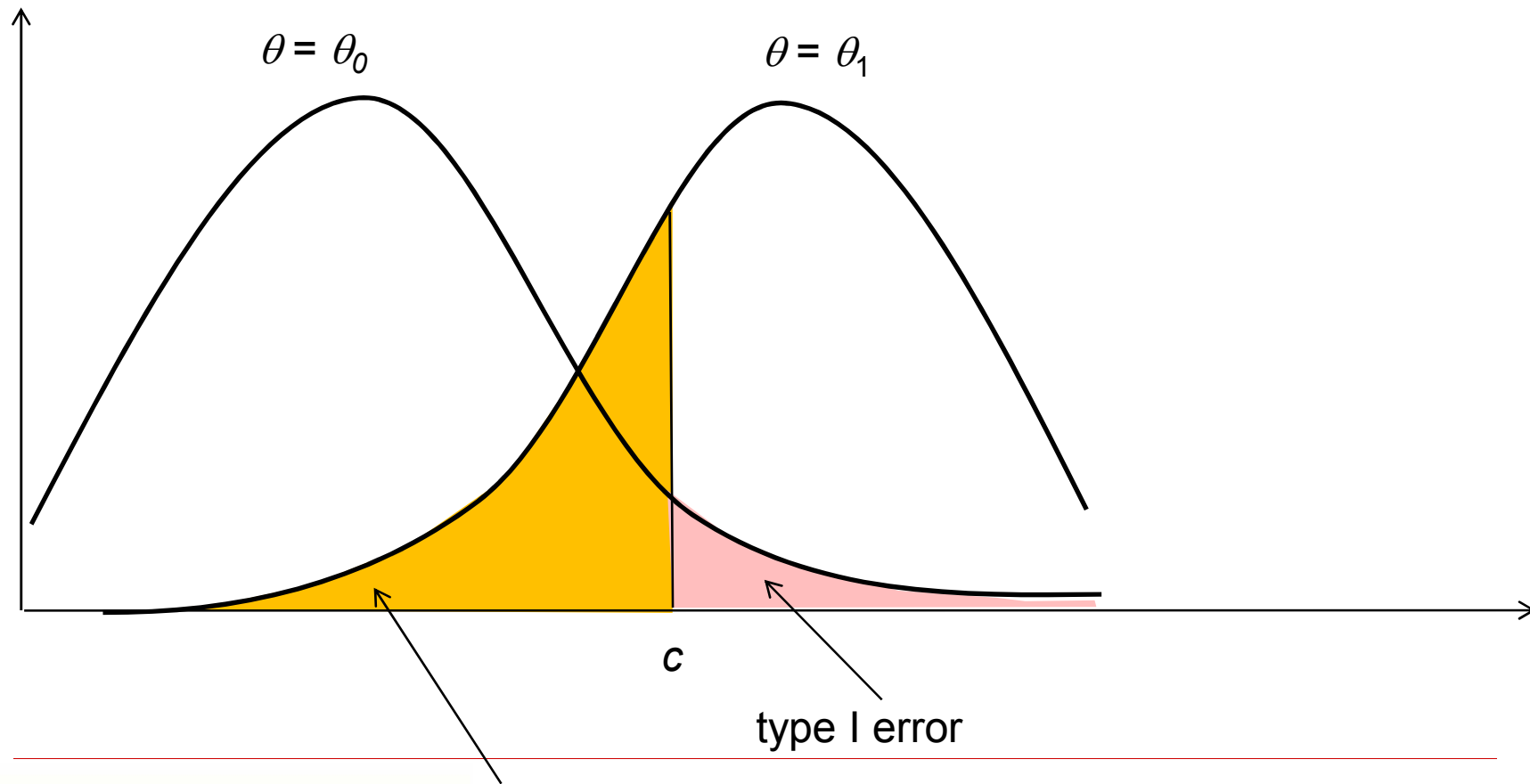
Type I and type II errors: graphical interpretation (1)

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 $P_\theta(K) \leq \alpha$.

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 $p = 1/2$)

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

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



$$K = \{0, 1, \dots, 174\} \cup \{226, 227, \dots, 400\}$$

Statistical test – example cont. (2).

p-value

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

We have:

$$P_{1/2}(|X - 200| > 20) \approx 0.05$$

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:

- $\text{p-value} < \alpha$ – reject the null hypothesis
- $\text{p-value} \geq \alpha$ – 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*.

□ $H_0 : p = 1/2, \quad H_1 : p > 1/2$

□ 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
 $H_0 : p \leq 1/2$



Statistical test – example cont. (4)

The choice of the alternative hypothesis

Again, from the de Moivre – Laplace theorem:

$$P_{1/2}(X - 200 > c) \approx 0.01 \text{ for } c \approx 23.3,$$

so for a significance level 0.01 we reject

$H_0 : p = 1/2$ in favor of $H_1 : p > 1/2$ 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 K
 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)

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

Function of the power of a test:

$$1-\beta : \Theta_1 \rightarrow [0,1] \text{ such that } 1-\beta(\theta) = P_{\theta}(K)$$

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 $H_0 : p = 1/2$ against $H_1 : p = 3/4$
with: $T(x) = X - 200$, $K = \{T(x) > 23.3\}$
(i.e. for a significance level $\alpha = 0.01$)

Power of the test:

$$\begin{aligned} 1-\beta(3/4) &= P(T(x) > 23.3 \mid p = 3/4) = P_{3/4}(X > 223.3) \\ &\approx 1-\Phi((223.3-300)/5\sqrt{3}) \approx \Phi(8.85) \approx 1 \end{aligned}$$

- But if $H_1 : p = 0.55$

$$\begin{aligned} 1-\beta(0.55) &= P(T(x) > 23.3 \mid p = 0.55) \approx 1-\Phi(0.33) \approx 1- \\ &0.63 \approx 0.37 \end{aligned}$$

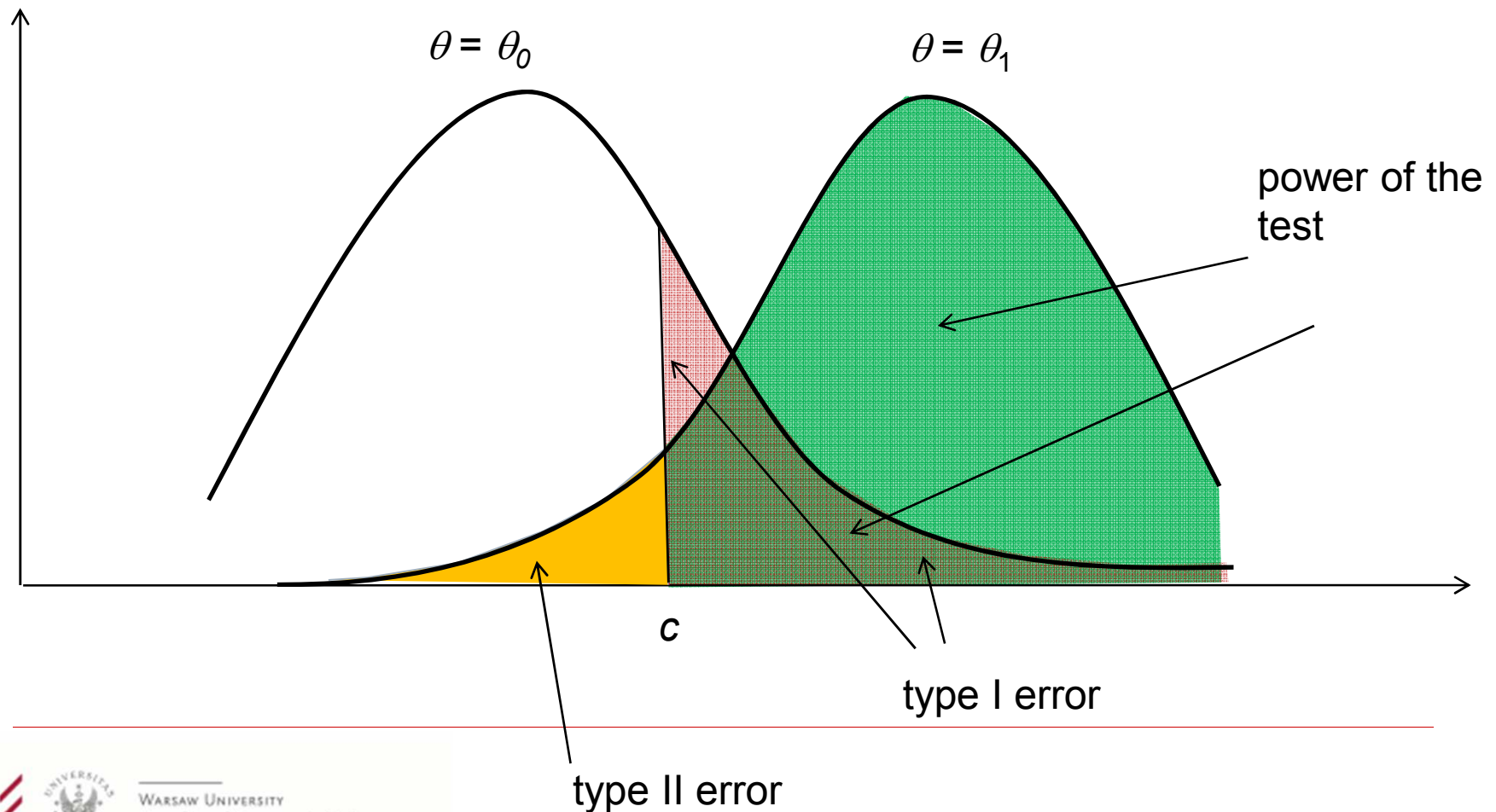
- And if $H_1 : p = 1/4$ for the same T we would get


$$1-\beta(1/4) = P(T(x) > 23.3 \mid p = 1/4) \approx 1-\Phi(14.23) \approx 0$$

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

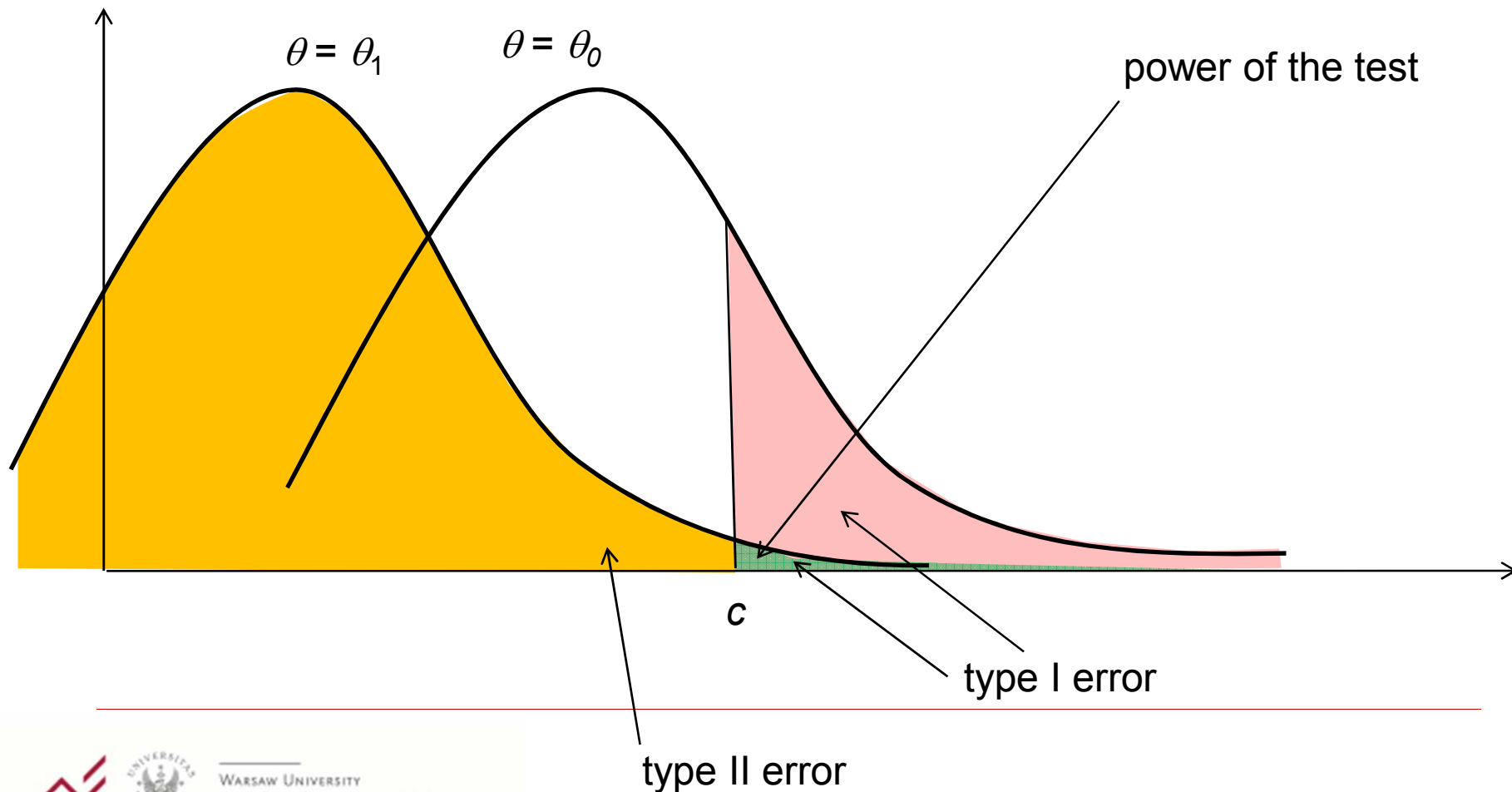
Power of the test: Graphical interpretation (1)

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



Sensitivity and specificity

Specificity – *true negative rate* (when in reality H_0 is not true)

Sensitivity – *true positive rate* (when in reality H_0 is true)

terms used commonly in diagnostic tests
(H_0 is having a medical condition)



Size of a test

sometimes we also look at the **size** of a test:

$$\sup_{\theta \in \Theta_0} P_{\theta}(K)$$

then we have:

significance level = α if the size of the test
does not exceed α .





WARSAW UNIVERSITY
Faculty of Economic Sciences