Mathematical Statistics

Anna Janicka

Lecture X, 29.04.2019

HYPOTHESIS TESTING II:
COMPARING TESTS

Plan for Today

0. Definitions – reminder and supplement
1. Comparing tests
2. Uniformly Most Powerful Test
3. Likelihood ratio test: Neyman-Pearson Lemma
4. Examples of tests for simple hypotheses and generalizations

Definitions – reminder

We are testing H_0: $\theta \in \Theta_0$ against H_1: $\theta \in \Theta_1$

K – critical region of the test, the set of outcomes for which we reject H_0, $K = \{x \in \mathrm{X} : \delta(x) = 1\}$

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

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**

Statistical test example (is the coin symmetric?) reminder: finding the critical range

We want: significance level $\alpha = 0.01$

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

$$P(|X - 200| > c) = 0.01 \qquad (n=400)$$

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

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

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

For a different alternative...

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

- $H_0 : p = ½, \quad H_1 : p > ½$
- 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.
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

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:

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

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

Statistical test – example cont.
Power of the test

- We test $H_0 : p = \frac{1}{2}$ against $H_1 : p = \frac{3}{4}$

  with: $T(x) = X - 200$, $K = \{T(x) > 23.3\}$

  (i.e. for a significance level $\alpha = 0.01$)

  Power of the test:

  $1-\beta(\frac{3}{4}) = P(T(x) > 23.3 \mid p = \frac{3}{4}) = P_{\frac{3}{4}}(X>223.3)$

  $\approx 1-\Phi((223.3-300)/5\sqrt{3}) \approx \Phi(8.85) \approx 1$

- But if $H_1 : p = 0.55$

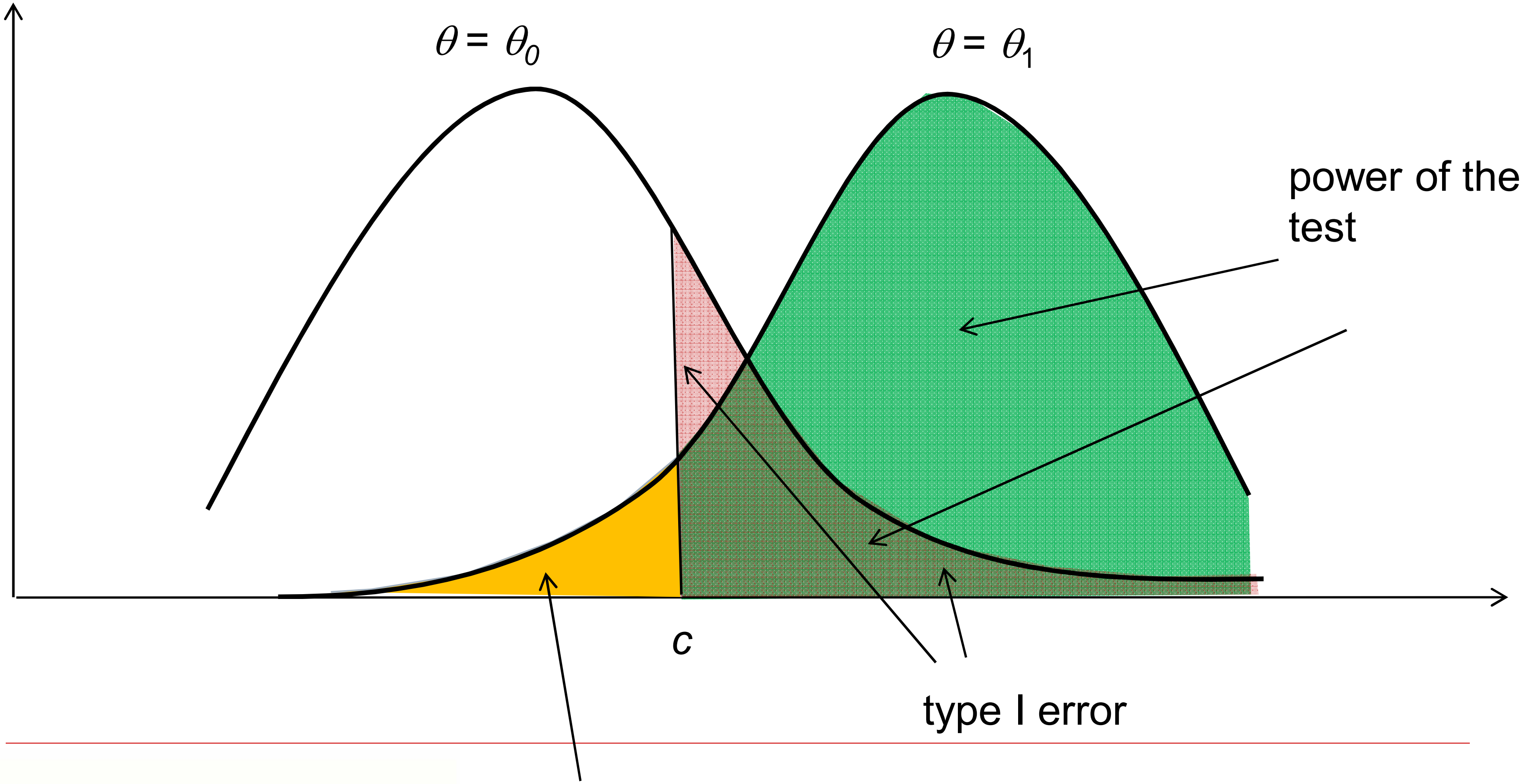
$1-\beta(0.55) = P(T(x) > 23.3 \mid p = 0.55) \approx 1-\Phi(0.33) \approx 1-0.63 \approx 0.37$

- And if $H_1 : p = \frac{1}{4}$ for the same T we would get

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

Power of the test: Graphical interpretation (1)

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



$\theta = \theta_0$

$\theta = \theta_1$

power of the test

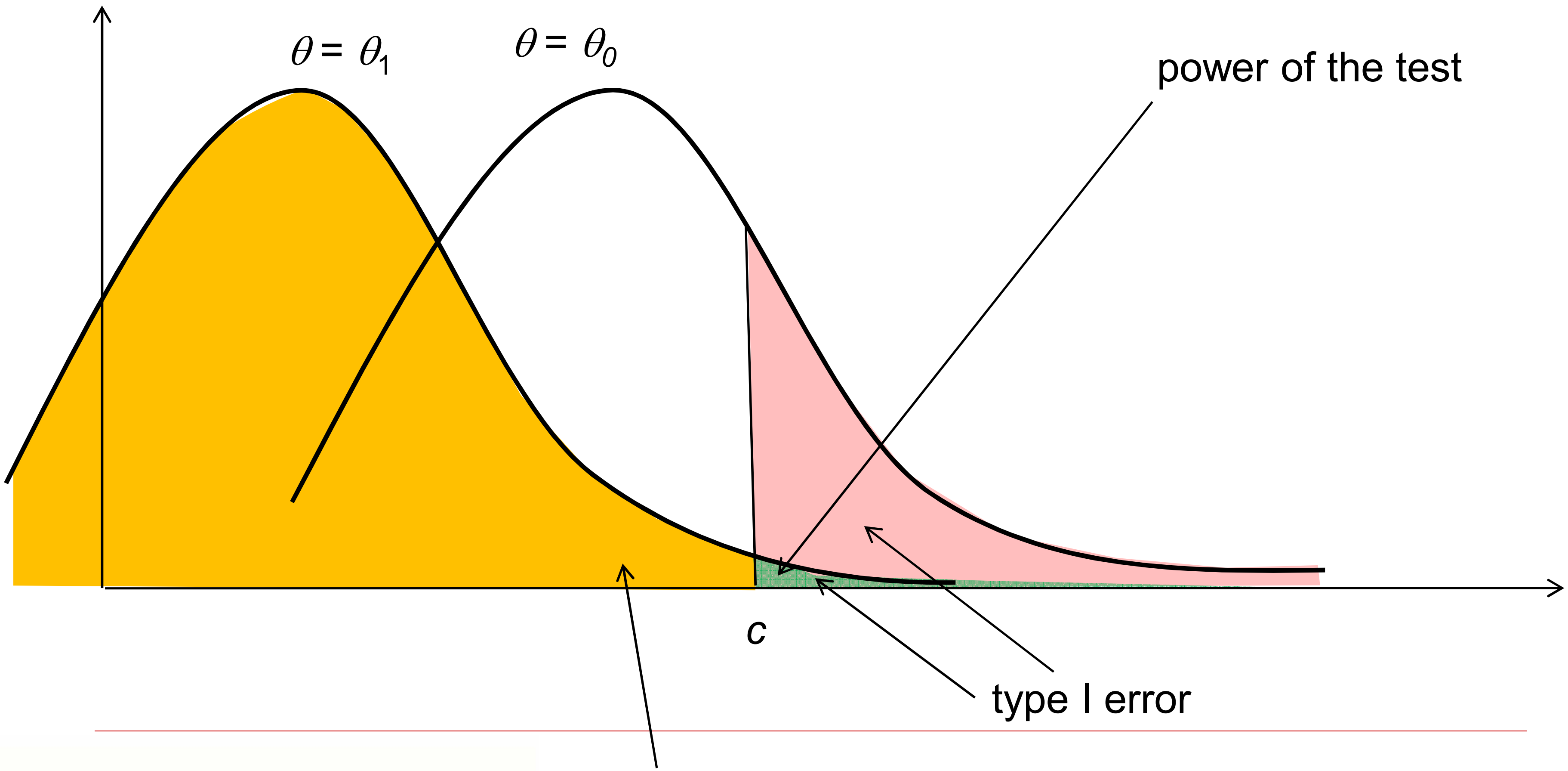
c

type I error

type II error

Power of the test:
Graphical interpretation (2) – a very bad test

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



$\theta = \theta_1$

$\theta = \theta_0$

power of the test

c

type I error

type II error

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.

Comparing tests

How do we chose the best test?

- for given null and alternative hypotheses
- for a given significance level

→ the test which is *more powerful* is better

Comparing the power of tests

$X \sim P_\theta$, $\{P_\theta : \theta \in \Theta\}$ – family of distributions

We test $H_0: \theta \in \Theta_0$ against $H_1: \theta \in \Theta_1$

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

with two tests with critical regions K_1 and K_2; both at significance level α.

The test with the critical region K_1 is **more powerful** than the test with critical region K_2, if

$$\forall \theta \in \Theta_1 : P_\theta(K_1) \geq P_\theta(K_2)$$

$$\text{and } \exists \theta_1 \in \Theta_1 : P_{\theta_1}(K_1) > P_{\theta_1}(K_2)$$

Uniformly most powerful test

For given H_0: $\theta \in \Theta_0$ and H_1: $\theta \in \Theta_1$:

δ^* is a **uniformly most powerful test** (UMPT) at significance level α, if

1) δ^* is a test at significance level α,

2) for any test δ at significance level α, we have, for any $\theta \in \Theta_1$:

$$P_\theta(\delta^*(X)=1) \geq P_\theta(\delta(X)=1)$$

i.e. the power of the test δ^* is not smaller than the power of any other test of the same hypotheses, for any $\theta \in \Theta_1$

if Θ_1 has one element, the word *uniform* is redundant

Uniformly most powerful test – alternative form

For given H_0: $\theta \in \Theta_0$ and H_1: $\theta \in \Theta_1$:

A test with critical region K^* is a **uniformly most powerful test** (UMPT) at significance level α, if

1) The test with critical region K^* is a test at significance level α, i.e.

$$\text{for any } \theta \in \Theta_0 \text{: } P_\theta(K^*) \leq \alpha,$$

2) for any test with critical region K at significance level α, we have for any $\theta \in \Theta_1$:

$$P_\theta(K^*) \geq P_\theta(K)$$

Testing simple hypotheses

We observe X. We want to test

H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$.

(two simple hypotheses)

We can write it as:

H_0: $X \sim f_0$ against H_1: $X \sim f_1$,

where f_0 and f_1 are *densities* of distributions defined by θ_0 and θ_1 (i.e. P_0 and P_1)

Likelihood ratio test for simple hypotheses. Neyman-Pearson Lemma

Let
$$K^* = \left\{ x \in X : \frac{f_1(x)}{f_0(x)} > c \right\}$$

such that $P_0(K^*) = \alpha$ and $P_1(K^*) = 1 - \beta$

Then, for any $K \subseteq X$:

if $P_0(K) \le \alpha$, then $P_1(K) \le 1 - \beta$.

(i.e.: the test with critical region K^* is the most powerful test for testing H_0 against H_1)

In many cases, it is easier to write the test as

$$K^* = \{x: \ln f_1(x) - \ln f_0(x) > c_1\}$$

Likelihood ratio test: we compare the likelihood ratio to a constant; if it is bad we reject H_0

Neyman-Pearson Lemma – Example 1

Normal model: $X_1, X_2, \dots, X_n$ are an IID sample from $N(\mu, \sigma^2)$, σ^2 is known

The most powerful test for

$$H_0: \mu = 0 \text{ against } H_1: \mu = 1.$$

$\mu_0 < \mu_1$

At significance level α :

$$K^* = \left\{ (x_1, x_2, \dots, x_n) : \overline{X} > \frac{u_{1-\alpha}\sigma}{\sqrt{n}} \right\}$$

For obs. 1.37; 0.21; 0.33; -0.45; 1.33; 0.85; 1.78; 1.21; 0.72 from $N(\mu, 1)$ we have, for $\alpha = 0.05$:

$$\overline{X} \approx 0.82 > \frac{1.645 \cdot 1}{\sqrt{9}} \approx 0.54$$

→ we reject H_0

Neyman-Pearson Lemma – Example 1 cont.

Power of the test

$$P_1(K^*) = P\left(\bar{X} > {}^{1.645\sigma}\!\Big/_{\sqrt{n}} \,\Big|\, \mu = 1\right) = \ldots.$$

$$= 1 - \Phi\left(1.645 - {}^{\mu_1 \cdot \sqrt{n}}\!\Big/_{\sigma}\right) \qquad \approx 0.91$$

If we change α, μ_1, n – the power of the test....

Neyman-Pearson Lemma: Generalization of example 1

The same test is UMP for H_1: $\mu > 0$ and for

$$H_0: \mu \leq 0 \text{ against } H_1: \mu > 0$$

more generally: under additional assumptions about the family of distributions, the same test is UMP for testing

$$H_0: \mu \leq \mu_0 \text{ against } H_1: \mu > \mu_0$$

Note the change of direction in the inequality when testing

$$H_0: \mu \geq \mu_0 \text{ against } H_1: \mu < \mu_0$$

Neyman-Pearson Lemma – Example 2

Exponential model: $X_1, X_2, \ldots, X_n$ are an IID sample from distr $\exp(\lambda)$, $n = 10$.

MP test for

$$H_0: \lambda = \tfrac{1}{2} \text{ against } H_1: \lambda = \tfrac{1}{4}.$$

At significance level $\alpha = 0.05$:

$$K^* = \left\{(x_1, x_2, \ldots, x_{10}) : \sum x_i > 31.41\right\}$$

E.g. for a sample: 2; 0.9; 1.7; 3.5; 1.9; 2.1; 3.7; 2.5; 3.4; 2.8:

$\Sigma = 24.5$ → no grounds for rejecting H_0.

$$\exp(\lambda) = \Gamma(1, \lambda) \qquad \Gamma(a, \lambda) + \Gamma(b, \lambda) = \Gamma(a + b, \lambda) \qquad \Gamma(n/2, 1/2) = \chi^2(n)$$

Neyman-Pearson Lemma – Example 2'

Exponential model: $X_1, X_2, ..., X_n$ are an IID sample from distr exp(λ), $n = 10$.

MP test for

$$H_0\colon \lambda = \tfrac{1}{2} \text{ against } H_1\colon \lambda = \tfrac{3}{4}.$$

At significance level $\alpha = 0.05$:

$$K^* = \left\{(x_1, x_2, ..., x_{10}) : \sum x_i < 10.85\right\}$$

E.g. for a sample: 2; 0.9; 1.7; 3.5; 1.9; 2.1; 3.7; 2.5; 3.4; 2.8:

$\Sigma = 24.5$ → no grounds for rejecting H_0.

$$\exp(\lambda) = \Gamma(1, \lambda) \qquad \Gamma(a, \lambda) + \Gamma(b, \lambda) = \Gamma(a + b, \lambda) \qquad \Gamma(n/2, 1/2) = \chi^2(n)$$

Example 2 cont.

The test

$$K^* = \left\{(x_1, x_2, \ldots, x_{10}) : \sum x_i > 31.41\right\}$$

is UMP for H_0: $\lambda \geq$ ½ against H_1: $\lambda <$ ½

The test

$$K^* = \left\{(x_1, x_2, \ldots, x_{10}) : \sum x_i < 10.85\right\}$$

is UMP for H_0: $\lambda \leq$ ½ against H_1: $\lambda >$ ½