Hypotheses. Type I and II errors

Statistics

Anton Afanasev

Higher School of Economics

DSBA 221

January 20, 2024

Seminar Overview

Quiz

Suppose that X is a random observation from a uniform distribution on the interval $(0, \delta)$, where $\delta > 1$ and that one wants to estimate $\theta = \mathsf{P}(X > 1) = 1 - 1/\delta$. Consider the following estimator T of θ:

$$T = \begin{cases} 1, & X > 1, \\ 0, & X \leq 1. \end{cases}$$

1. Is T an unbiased estimator?
2. Find the mean squared error of the estimator.

Problem 1

The coin was tossed 10 times, and 8 heads were observed. Can the coin be considered fair? (Use significance level $\alpha = 0.05$.)

Null hypothesis

- Null hypothesis H_0 is a claim we want to reject.
- It is also a claim of no effect, a claim of no relationship.

Example (Problem 1 null hypothesis)

- There is a sample of Bernoulli trials $\{X_1, \ldots, X_{10}\}$:

$$X_i \sim \text{Bernoulli}(p),$$

  where p is a probability to get heads after one toss.
- The claim:

$$H_0 : p = \frac{1}{2}.$$

- We have to come up with some rejection rule to figure out how close the null hypothesis is to reality.

Test statistic

- A null hypothesis can be rejected (or not rejected) based on a quantitative experiment result, which is called test statistic.
- $T(X_1, \ldots, X_n)$ – test statistic from a sample $\{X_1, \ldots, X_n\}$.

Example (Problem 1 test statistic)

- Possible test statistic:

$$T(X_1, \ldots, X_{10}) = X_1 + \ldots + X_{10},$$

- The closer value of $T(X_1, \ldots, X_{10})$ to 10 (or 0), the less likely it is for H_0 to be true.

- But what is the strict boundary for rejection? Is $T(X_1, \ldots, X_{10}) = 8$ good enough to reject H_0?

Type I error

- There is always a possibility to reject H_0 falsely.

Example (Problem 1 false rejection)

- A fair coin might give 10 heads out of 10 tosses.
- According even to the most conservative rejection rules, H_0 will be rejected.
- The probability to falsely reject H_0:

$$\mathsf{P}(\{1, \ldots, 1\}) = \frac{1}{2^{10}}.$$

- Such mistake is called Type I error, denoted as α.
- In order to set a strict rule when to reject H_0, we need to decide which Type I error we are ready to allow:

$$\alpha = \mathsf{P}(T(X_1, \ldots, X_n \mid \mathrm{H}_0) \geq t_{\text{crit}}),$$

thus getting a boundary value t_{crit} of a test statistic.

Type I error

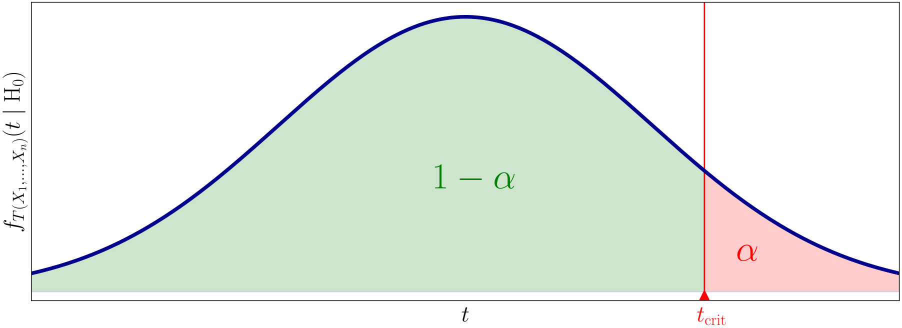

Figure: Type I error for a test statistic $T(X_1, \ldots, X_n)$.

Rejection rule

- A critical value t_{crit} of the test statistic separates critical region from a confidence region.
  If the result of an experiment (a value of the test statistic t_0 under H_0) falls into critical region, H_0 is rejected. When critical region is in the right tail:

$$t_0 > t_{\text{crit}} \quad \Longrightarrow \quad \mathrm{H}_0 \text{ is rejected.}$$

- A probability of the test statistic to have value t_0 or less probable (under H_0) is called p-value:

$$p\text{-val} = \mathsf{P}(T(X_1, \ldots, X_n \mid \mathrm{H}_0) \geq t_0).$$

  Another equivalent rejection rule:

$$p\text{-val} < \alpha \quad \Longrightarrow \quad \mathrm{H}_0 \text{ is rejected.}$$

- α is also called a significance level of the test.

Rejection rule

H_0 is rejected

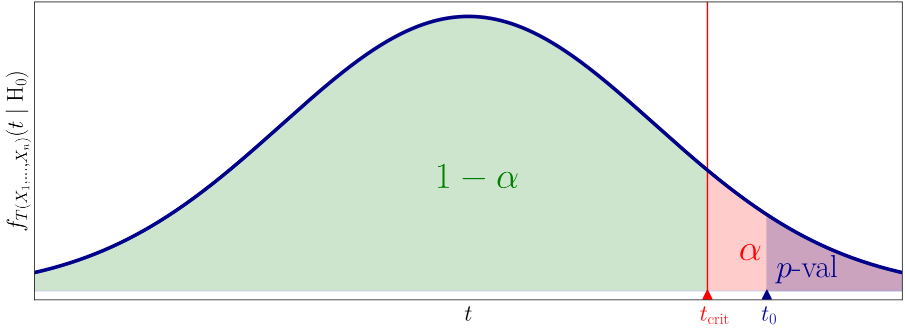

Figure: H_0 is rejected according to value of t_0 or p-val.

Rejection rule
H_0 is not rejected

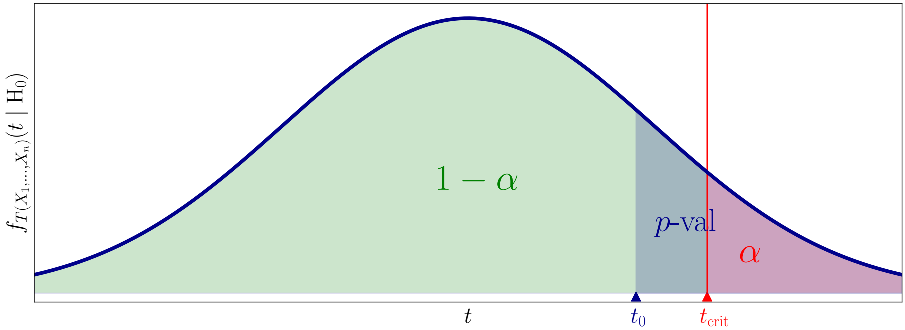

Figure: H_0 is not rejected according to value of t_0 or p-val.

Alternative hypothesis

- Alternative hypothesis H_1 is a claim we want to prove (by rejecting H_0).
- Usually there are 3 options for an alternative to H_0:

$$\mathrm{H}_0 : \theta = \theta_0 \quad \text{vs} \quad \begin{array}{l} \mathrm{H}_1 : \theta > \theta_0, \\ \mathrm{H}_1 : \theta \neq \theta_0, \\ \mathrm{H}_1 : \theta < \theta_0. \end{array}$$

- H_1 shows the direction of critical region with regards to H_0.

Example (Problem 1 alternative hypothesis)

- In Problem 1 we assumed right-sided alternative:

$$\mathrm{H}_0 : p = \frac{1}{2} \quad \text{vs} \quad \mathrm{H}_1 : p > \frac{1}{2}.$$

Type II error

- Failing to reject incorrect H_0 is called Type II error:

$$\beta = \mathsf{P}(T(X_1, \ldots, X_n \mid \mathrm{H}_1) < t_{\text{crit}}).$$

- Summary for type I and II errors:

	H_0 is true	H_1 is true
H_0 is not rejected	confidence level $1-\alpha$	Type II error β
H_0 is rejected	Type I error α	power of the test $1-\beta$

- Power of the test as a function of a parameter of population θ is often denoted as $K(\theta)$.
- The dependency of Type II error β on θ is graphed as OCC – operating characteristic curve.

Problem 2

Junior researcher Angela is presenting her half-year project about speed of blood clotting in front of the Head of her laboratory. That speed is normally distributed with population standard deviation $\sigma = 3$ minutes. Angela is very nervous and in the very responsible moment she has forgotten the resultant value of true population speed μ – it's either 6 or 9 minutes. Presentation slide claims that they had 16 observations and the sample mean is 7 minutes, so Angela assumes that the correct value is 6.

1. Find the critical value $\overline{x}_{\text{crit}}$ of sample mean for a hypothesis $\mathrm{H}_0 : \mu = 6$, which would guarantee that it's true within 95% confidence level.
2. Find the p-value for the hypothesis H_0.
3. Preserving confidence level from part 1, find the probability of Type II error, using aforementioned alternative $\mathrm{H}_1 : \mu = 9$.

Type I and II errors

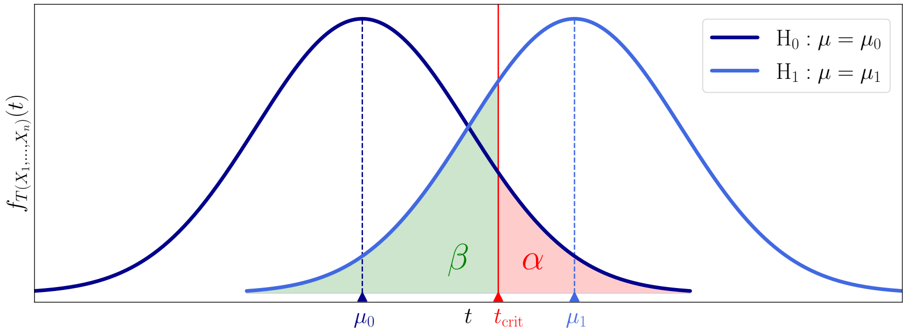

Figure: Type I and II errors for fixed H_0 and H_1.

Type I and II errors

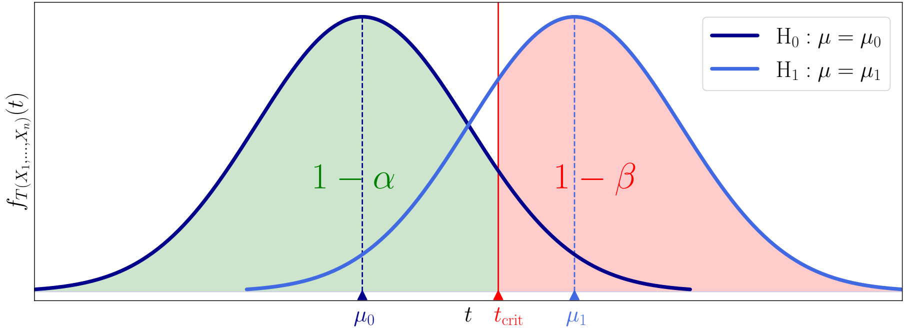

Figure: Confidence level and power of the test for fixed H_0 and H_1.

Problem 3

A firm manufacturing memory chips found that if everything was going right, 10% of them were defective. If the production process was in trouble then 40% were defective. Firm's quality control office tests four memory chips each hour. If two or more of four were defective, production would be shut down to look for trouble.

1. What is Type I and Type II errors here?
2. What is probability that production will be shut down if everything was going right?
3. What is the probability of the missed alarm?

Problem 3

4. Suppose, you have additional information:
   - A. Production goes "out of control" about 10% of hours.
   - B. Testing of one chip costs \$10.
   - C. Missed alarm costs \$10000.
   - D. False alarm costs \$2000.

   Calculate the expected total cost associated with faulty production. Is it better to use another decision rule for detecting the trouble (shut the production if at least one of four tested chips is defective)?

5. New manager suggested testing 100 chips and shutting down production if more than 25 were defective. Calculate expected total cost associated with faulty production for this (part 3) decision rule. Compare with the first two.

Problem 4

You have a coin with $\mathsf{P}(\text{tail}) = p$. You test the null hypothesis that the coin is a fair one. You flip the coin 5 times and if number of tails is 2 or 3, you do not reject the null hypothesis, otherwise you suppose the coin is biased.

1. What is significance level of the test?
2. Plot the OCC function and the power function of the test.
3. Find values of the power function at the $p = 0.3$ and $p = 0.7$.

Problem 4

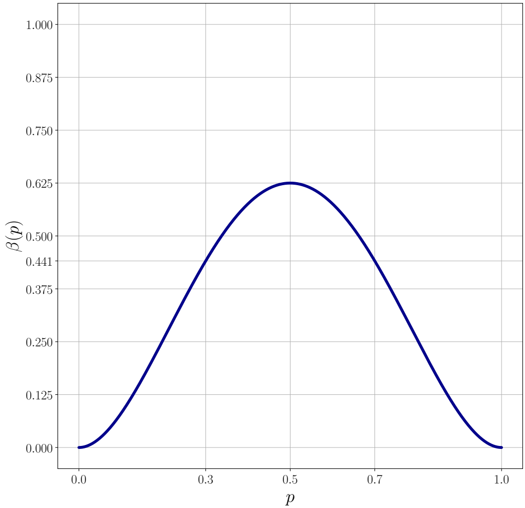

Figure: OCC of the test.

Problem 4

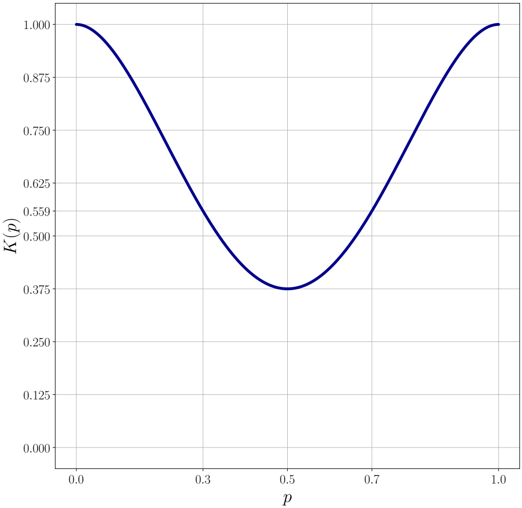

Figure: Graph of the power of the test.

Problem 5

Random variable X has normal distribution $\mathcal{N}\left(\mu, \sigma^2\right)$. Let σ be equal to 25 and sample size be equal to 100. You test null hypothesis $\mathrm{H}_0 : \mu = 100$ against the alternative hypothesis $\mathrm{H}_1 : \mu < 100$. Significance level of the test is 10%. You reject the null hypothesis if $\overline{X} < c$. Plot the power function of the test.

Problem 5

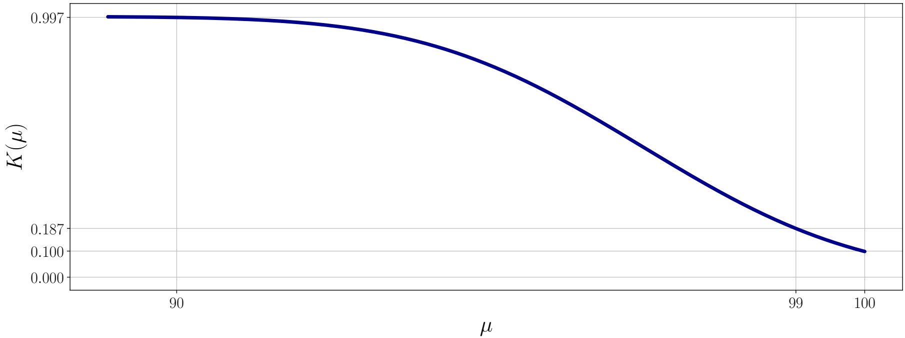

Figure: Graph of the power of the test.

Problem 6

Let X be $\mathcal{N}\left(\mu, 10^2\right)$. To test $\mathrm{H}_0 : \mu = 80$ against alternative $\mathrm{H}_1 : \mu > 80$ the critical region $\overline{x} > x_c = 83$ was chosen for the sample of size $n = 25$.

1. What is the power function $K(\mu)$ of this test?
2. What is the significance level of this test?
3. What are the values $K(80), K(83), K(86)$?
4. Sketch the graph of the power function and the OCC function.
5. What is the p-value corresponding to $\overline{x} = 83.41$?

Problem 6

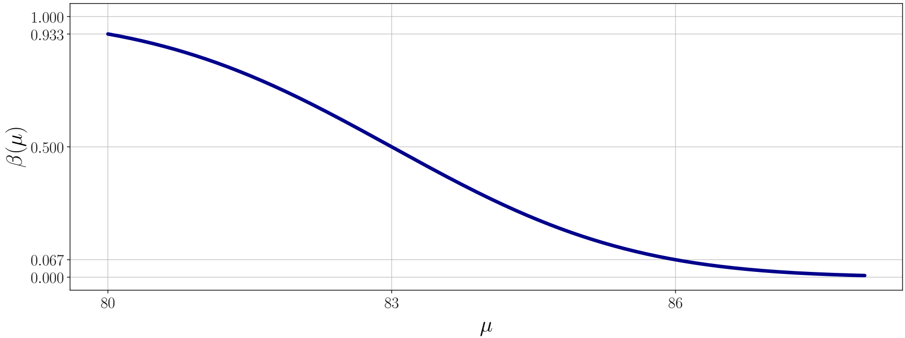

Figure: OCC of the test.

Problem 6

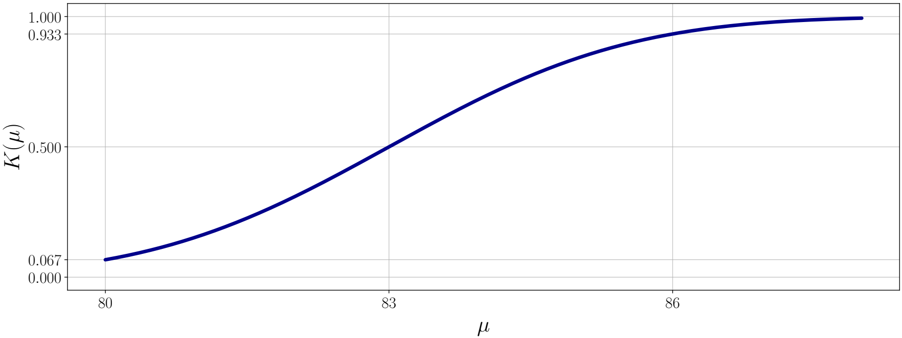

Figure: Graph of the power of the test.

Problem 7

Let X be a Bernoulli random variable with parameter p.
We would like to test the null hypothesis $\mathrm{H}_0 : p \leq 0.4$ against the alternative hypothesis $\mathrm{H}_1 : p > 0.4$.

For the test statistic let's use $Y = \sum_{i=1}^{n} X_i$, where $\{X_1, \ldots, X_n\}$ is a random sample of size n from this Bernoulli distribution.
Let the critical region be of the form $C = \{y : y \geq c\}$.

1. Let $n = 100$. On the same set of axes, sketch the graphs of the power function corresponding to the three critical regions: $C_1 = \{y : y \geq 40\}$, $C_2 = \{y : y \geq 50\}$, $C_3 = \{y : y \geq 60\}$. Use normal approximation to compute the probabilities.
2. Let $C_n = \{y : y \geq 0.45n\}$. On the same set of axes, sketch the graphs of the power function corresponding to the three samples of size 10, 100, and 1000.

Problem 7

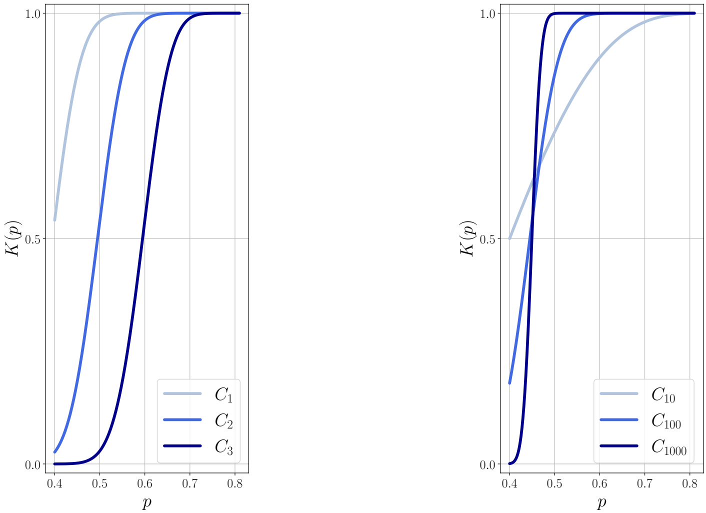

Figure: Graphs of the power of the test.

That's all Folks