Contents

1 Statistical Hypothesis Testing

1.1 What is a Hypothesis?

Hello everyone! Here we come to the last module of our course, and today we are going to discuss the problem of hypothesis testing solved by methods of mathematical statistics. When building probability models of random experiments, we occasionally make some assumptions about population, probability of some event, population distribution, distribution parameters, etc. All these assumptions are logically called hypotheses. Why? Well, at the time of observation, we have no correct answers, but we do have hypotheses (or assumptions) that we want to accept or reject. For example, we measure the product sales volume and want to test the hypothesis that sales volume will be on average greater this year than the last year. We cannot compare mean values of the current and last year since the result could be random: a greater mean value could occur by chance. So, we have to go further into the question because the hypothesis is not about comparing the mean values of two samples (i.e. the sample means) but the expected values of two populations with unknown distribution and parameters.

We can consider another example. Say, let's get the answer to the question "Who is cooking dinner today?" by tossing a coin. Of course, the coin must be fair (that is, give a fair answer) with the probabilities of both heads and tails very close to 0.5. How to check it? For example, we can toss a coin 100 times and calculate how many heads and tails we've got and then compare these numbers. But will we ever get exactly 50 heads and 50 tails?

Most likely, we won't. Approximately equal numbers most probably explain why we want to conclude that the coin is fair. But there are pitfalls. After all, such a result could have occurred in a series of experiments when we used an unfair coin (say, the coin with the probability of getting heads equal to 0.3). On the contrary, if the number of heads obtained is far from 0.5, we are most likely to say that the coin is unfair. But there is no certainty that it is true: after all, anything can happen, as the experiment is random. So, the coin may well be fair in this case too. The thing is that, as a rule, it is impossible to make faultless conclusions about the distribution based on the sample of finite size, so we have to bear in mind that there is a chance to choose a wrong hypothesis. In probabilistic language, the last described experiment (with a coin) is a series of Bernoulli trials with unknown probability of success p in each test. Then, the hypothesis that we are testing can be written the following way: $p = 0.5$ (or, for example, $p \in (0.45, 0.55)$, to allow some kind of a logical defect).

So, let's define what a hypothesis is. Let sample $X = (X_1, X_2, ..., X_n)$ be taken from the distribution $\mathcal{P}$. Unless otherwise stated, we will assume that all observations (all random variables included in the sample) have the same distribution. It turns out that, in some cases, this assumption needs to be verified! The same goes for independence.

Definition 1.1.1 *Hypothesis* H *is a tentative assumption about the distribution of population* ξ.

Taking this into account, we can say that hypothesis H *is a tentative assumption about the distribution of observations.*

Let's introduce the concepts of simple and complex hypotheses.

Definition 1.1.2 *Hypothesis* H *is simple if it is associated with only one distribution of* $\mathsf{H} = \{\mathcal{P} = \mathcal{P}_1\}$. *Otherwise, hypothesis* H *is complex* $\mathsf{H} = \{\mathsf{P} \in \mathbb{P}\}$, *where* $\mathbb{P}$ *is a subset of the set of all distributions.*

Often, when there are only two hypotheses, one of them is called a null hypothesis, and another is called an alternative one. What's important is that the alternative hypothesis, which, of course, contradicts the null hypothesis, should not constitute the family of distributions entirely. Let's give typical examples (or describe typical challenges of hypothesis testing) to apply what we have just discussed.

Example 1.1.1 *A choice among several simple hypotheses is made:* $\mathsf{H}_1 = \{\mathcal{P} = \mathcal{P}_1\}$, $\mathsf{H}_2 = \{\mathcal{P} = \mathcal{P}_2\}$, ..., $\mathsf{H}_k = \{\mathcal{P} = \mathcal{P}_k\}$. *Say,*

$$\mathsf{H}_1 = \{\mathcal{P} = \mathsf{B}_{1/2}\},\ \mathsf{H}_2 = \{\mathcal{P} = \mathsf{U}_{0,1}\},\ \mathsf{H}_3 = \{\mathcal{P} = \mathsf{N}_{2,4}\}.$$

One more classical example.

Example 1.1.2 *There is a simple null hypothesis and a complex alternative hypothesis:* $\mathsf{H}_1 = \{\mathcal{P} = \mathcal{P}_1\}$ *and* $\mathsf{H}_2 = \{\mathcal{P} \in \mathbb{P}\}$, *where* $\mathbb{P}$ *is a subset of the family of all distributions that does not include* $\mathcal{P}_1$. *For example,* $\mathsf{H}_1 = \{\mathcal{P} = \mathsf{B}_{0.5}\}$ *and* $\mathsf{H}_2 = \{\mathcal{P} = \mathsf{B}_p,\ p < 0.5\}$.

Of course, both the null hypothesis and the alternative one can be complex, e.g.

Example 1.1.3 $\mathsf{H}_1 = \{\mathcal{P} = \mathsf{N}_{a,\sigma^2},\ a \in \mathbb{R}, \sigma > 0\}$ *and* $\mathsf{H}_2 = \{$*hypothesis* H_1 *is wrong*$\}$.

Often there is the so-called test of homogeneity

Example 1.1.4 *Let there be sample* $X_{11}, X_{12}, ..., X_{1n_1}$ *taken from distribution* $\mathcal{P}_1$, $X_{21}, X_{22}, ..., X_{2n_2}$, *from distribution* $\mathcal{P}_2$, *and so on,* $X_{k1}, X_{k2}, ..., X_{kn_k}$, *from distribution* $\mathcal{P}_k$. *We test complex null hypothesis* $\mathsf{H}_1 = \{\mathcal{P}_1 = \mathcal{P}_2 = ... = \mathcal{P}_k\}$ *against complex alternative hypothesis* $\mathsf{H}_2 = \{$*hypothesis* H_1 *is wrong*$\}$.

In the last example, an important problem is posed, i.e. the problem of checking whether the samples are taken from the same distribution or not. It is important in practice, as we need to know what kind of distribution we are studying based on a particular sample. This knowledge can help us to avoid incidents, such as predicting the clothes size of the basketball team players based on a randomly selected sample of children's height in a kindergarten. Well, but how to understand which hypothesis is better? Of course, with the help of a sample and the so-called criterion.

1.2 Criterion and Error Types

Definition 1.2.1 *Let's there be hypotheses* H_1, H_2, *...,* H_k *and sample* $X = (X_1, ..., X_n)$. *Then, criterion* $\delta = \delta(X_1, X_2, ..., X_n)$ *is a projection*

$$\delta : \mathbb{R}^n \to \{\mathsf{H}_1, \mathsf{H}_2, ..., \mathsf{H}_n\}.$$

So, the criterion is a sample function that returns a possible hypothesis. There are also the so-called randomized criteria that accept each hypothesis with some probability, but we will not consider them.

As discussed earlier, the criterion does not always return the true hypothesis because the sample (the finite number of observations of the population values) does not provide complete information about itself. That's why the so-called type $1, 2$, etc. errors arise. As a rule, we will consider two hypotheses (null and alternative). Due to this, we'll also simplify the definition.

Definition 1.2.2 *A type I error is said to occur if the criterion rejects true hypothesis* H_1. *A type II error is said to occur if the criterion rejects true hypothesis* H_2.

Along with the hypothesis testing errors, the error probabilities of type 1 and 2 also occur. Let's define them. So, the probability of type I error is

$$\alpha_1 = \mathsf{P}_{\mathsf{H}_1}(\delta(X) \neq \mathsf{H}_1) = \mathsf{P}_{\mathsf{H}_1}(\delta(X) = \mathsf{H}_2),$$

while the probability of type II error is

$$\alpha_2 = \mathsf{P}_{\mathsf{H}_2}(\delta(X) \neq \mathsf{H}_2) = \mathsf{P}_{\mathsf{H}_2}(\delta(X) = \mathsf{H}_1).$$

Note that, when saying, for example, that hypothesis H_1 is true and calculating $\mathsf{P}_{\mathsf{H}_1}(\cdot)$, we suppose that the sample distribution is exactly as that assumed by H_1, so we calculate the probability in accordance with this distribution. The same applies to H_2. Let's give a simple example.

Example 1.2.1 *Let a randomly chosen product of a company be defective with probability p. The quality assurance department also makes mistakes: it rejects a quality product with probability γ and issues a defective item with probability ε. It's reasonable to introduce two hypotheses for a randomly chosen product $\mathsf{H}_1 = \{a\ quality\ product\}$ and $\mathsf{H}_2 = \{a\ defective\ product\}$. Let's find the probabilities of type I and type II errors.*

$$\alpha_1 = \mathsf{P}_{\mathsf{H}_1}(\delta = \mathsf{H}_2) = \mathsf{P}_{a\ quality\ product}(QA\ rejected\ the\ product) = \gamma,$$

$$\alpha_2 = \mathsf{P}_{\mathsf{H}_2}(\delta = \mathsf{H}_1) = \mathsf{P}_{defective\ product}(product\ passed\ QA) = \varepsilon.$$

Well, let's briefly sum up. First, the statistical criterion gives no certain answer to whether the tested hypothesis is true or false. It only helps us to decide whether the sample data contradicts or does not contradict the tested hypothesis. Thus, it is reasonable to say that the hypothesis "is rejected" and "is not rejected" instead of "rejected" and "accepted" (although the latter term is also very popular). Then, there is one main hypothesis, and the rest are undesirable deviations from it. Hence, the conclusion "data contradicts the hypothesis" is much more significant than the conclusion "data does not contradict the hypothesis". Finally, we never know which hypothesis is a true one. Therefore, we must consider the error probabilities of the criterion. Let's discuss the latter in detail.

1.3 Significance Level and Power

Let's look at the case when two simple hypotheses on the observations distribution are considered:

$$\mathsf{H}_1 = \{\mathcal{P} = \mathcal{P}_1\}, \quad \mathsf{H}_2 = \{\mathcal{P} = \mathcal{P}_2\}.$$

In this case, any criterion δ takes no more than two values on sample $X = (X_1, X_2, ..., X_n)$ (recall that criterion δ is a function from $\mathbb{R}^n$ to $\{\mathsf{H}_1, \mathsf{H}_2\}$). Thus, we can assume that entire $\mathbb{R}^n$ breaks down into two parts S and $\mathbb{R}^n \setminus S$, that is

$$\mathbb{R}^n = S \cup (\mathbb{R}^n \setminus S),$$

and, in general form, the criterion δ takes the form

$$\delta(X) = \begin{cases} \mathsf{H}_1, & \text{if } X \in \mathbb{R}^n \setminus S \\ \mathsf{H}_2, & \text{if } X \in S \end{cases}.$$

Region S is often called a critical (or rejection) region.

Definition 1.3.1 *A level of significance δ is the probability of a type I error α_1:*

$$\alpha_1 = \alpha_1(\delta) = \mathsf{P}_{\mathsf{H}_1}(\delta \neq \mathsf{H}_1) = \mathsf{P}_{\mathsf{H}_1}(\delta = \mathsf{H}_2) = \mathsf{P}_{\mathsf{H}_1}(X \in S).$$

Thus, a significance level of a criterion is the probability that a sample falls into a critical region when hypothesis H_1 *is true.*

Such a characteristic as power is often considered together with the significance level.

Definition 1.3.2 *A power of criterion* δ *is* $1-\alpha_2$*, that is*

$$1-\alpha_2 = 1-\alpha_2(\delta) = 1-\mathsf{P}_{\mathsf{H}_2}(\delta \neq \mathsf{H}_2) = \mathsf{P}_{\mathsf{H}_2}(\delta = \mathsf{H}_2) = \mathsf{P}_{\mathsf{H}_2}(X \in S).$$

So, a power of a criterion is the probability that a sample falls in a critical region if hypothesis H_2 *is true.*

Note that the probabilities of the type I and type II errors are calculated under different assumptions about the distributions (hypothesis H_1 or H_2 is true). Thus, fixed relations, such as $\alpha_1 = 1-\alpha_2$ and similar, simply do not exist. Clearly, we want the significance level to be as low as possible because the significance level is the probability of the type I error, and errors must be fixed. However, it is also important to increase the power. After all, the greater the power, the smaller the type II error. Unfortunately, in practice, monitoring both parameters at the same time is not possible due to the lack of an explicit functional relationship between the power and significance level. In practice, a sufficiently small level of significance is selected, and we do not choose the power, so it turns out to have some value. It turns out that this is even the case. A decrease in the significance level usually leads to a decrease in the power, that is, as a rule, we cannot reduce the significance level and increase power at the same time. After all, by reducing the type I error, i.e. decreasing the probability of $\mathsf{P}_{\mathsf{H}_1}(X \in S)$, we will reduce the critical region of S, thereby reducing the power equal to $\mathsf{P}_{\mathsf{H}_2}(X \in S)$. Thus, some compromise must be maintained between the power and the significance level. Let's show this with a funny example. Assume we have a sample of size 1 from normal distribution $\mathsf{N}_{a,1}$. We consider two hypotheses: $\mathsf{H}_1 :\ a = 0$ and $\mathsf{H}_2 :\ a = 1$ and criterion

$$\delta(X_1) = \begin{cases} \mathsf{H}_1, & X_1 \le d \\ \mathsf{H}_2, & X_1 > d \end{cases}.$$

What is a type I error? It is the probability of not rejecting hypothesis H_2 when hypothesis H_1 is true, that is, $\alpha = \mathsf{P}_{\mathsf{H}_1}(\delta = H_2) = \mathsf{P}_{\mathsf{H}_1}(X_1 > d)$. What is a type II error? It is the probability of not rejecting hypothesis H_1 when hypothesis H_2 is true, that is, $\mathsf{P}_{\mathsf{H}_2}(\delta = \mathsf{H}_1) = \mathsf{P}_{\mathsf{H}_2}(X_1 \le d)$. As you can see, when the probability of the type I error (α) is decreasing, the probability of the type II error is increasing, which is decreasing the power of the criterion and vice versa.

Note that we fix only the type I error but not the type II error. Thus, it matters what hypothesis is considered a null hypothesis and what is an alternative

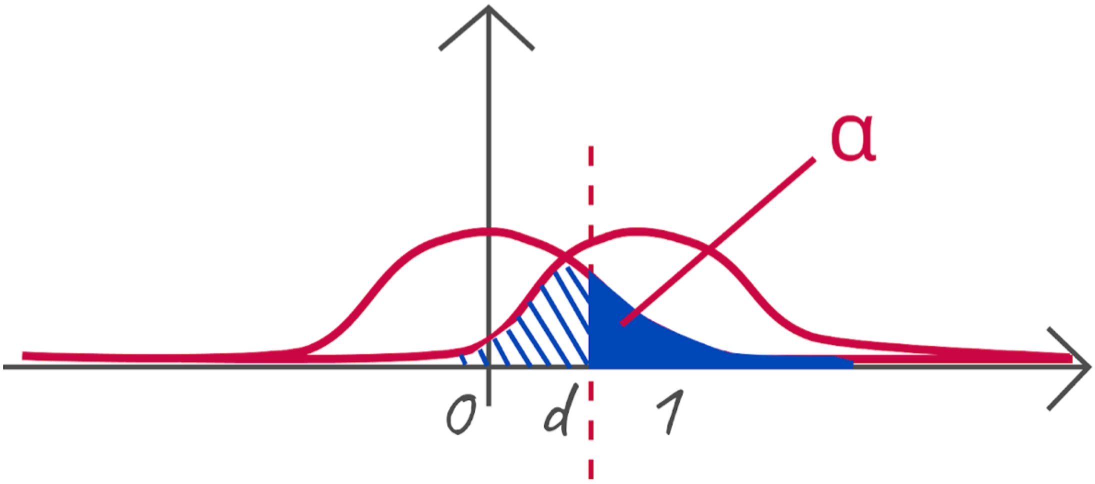

Figure 1: Type I and type II errors

hypothesis. For example, we are on the plane and have two hypotheses: $\mathsf{H}_1 = \{\text{the plane has no defective parts}\}$ and $\mathsf{H}_2 = \{\text{the plane has defective parts}\}$. Hypothesis testing produces two errors: reject the hypothesis that the plane has no defective parts if the plane has no defective parts (the type I error), and reject the hypothesis that the plane has defective parts if it has defective parts (the type II error). With this set of hypotheses, we control the error of the event. In the worst-case scenario, we will lose money due to ticket cancellation. However, we do not control the probability of the event that can lead to a possible plane crash.

In this case, the hypotheses should be formulated differently: $\mathsf{H}_1 = \{\text{the plane has defective parts}\}$ and $\mathsf{H}_2 = \{\text{the plane has no defective parts}\}$. Now we can control the probability of the event "reject the hypothesis that the plane has defective parts if it has defective parts". We will make the probability known and very small and stop worrying about the possible plane crash (with an unknown probability of occurring).

2 Goodness of Fit

2.1 Concept of Goodness of Fit

A topic of hypothesis testing deserves not just one module but the entire course. Anyway, we will discuss only the so-called goodness-of-fit tests and some of their generalizations.

Definition 2.1.1 *Goodness-of-fit tests are the criteria used to test simple null* *hypothesis* $\mathsf{H}_1 = \{\mathcal{P} = \mathcal{P}_1\}$, *when alternative complex hypothesis* $\mathsf{H}_2 = \{$*hypothesis* H_1 *is wrong*$\}$.

We will stick to the following principle. Let a function reflect the deviation of an empirical distribution (derived from sample $X = (X_1, X_2, ..., X_n)$) from the theoretical one. The distribution of this function differs significantly depending on whether the null hypothesis is true or not. It is reasonable that, depending on the value of this deviation "function", we can either reject or not reject the null hypothesis. Let's try to formulate an algorithm.
1. Let us define such a function $\rho(X)$ that:

- if hypothesis H_1 is true, then

$$\rho(X) \xrightarrow[n \to +\infty]{\mathsf{d}} Y \sim \mathcal{G},$$

  where $\mathcal{G}$ is a continuous distribution;

- if hypothesis H_1 is wrong, then

$$|\rho(X)| \xrightarrow[n \to +\infty]{\mathsf{P}} \infty.$$

2. Thus, for random variable Y having distribution $\mathcal{G}$, we can find constant C based on the condition that $\varepsilon = \mathsf{P}(|Y| \geq C)$. Let's formulate the criterion as follows:

$$\delta(X) = \begin{cases} \mathsf{H}_1, & \text{if } |\rho(X)| < C \\ \mathsf{H}_2, & \text{if } |\rho(X)| \geq C \end{cases}$$

The formulated criterion obeys the rule. If the absolute value of the deviation function is large on the given sample, then the alternative hypothesis should be chosen, and vice versa.

We can show that, as n increases, the significance level of our criterion tends to ε, and the power to unity. The last concept needs to be clarified because the alternative hypothesis consists not of the one distribution but a set thereof. In our case, the power for each specific distribution from the alternative hypothesis tends to unity.

Let's give specific examples of the criteria and their applications.

2.2 Kolmogorov Goodness-of-Fit Test

One of the most important tasks is to determine a distribution of population ξ based on sample $X = (X_1, X_2, ..., X_n)$. Kolmogorov goodness-of-fit test allows us to do it when the population distribution function is continuous.

Let $X = (X_1, X_2, ..., X_n)$ be a sample from distribution $\mathcal{P}$. Null hypothesis $\mathsf{H}_1 = \{\mathcal{P} = \mathcal{P}_1\}$ is tested against complex alternative hypothesis $\mathsf{H}_2 = \{\mathcal{P} \neq \mathcal{P}_1\}$.

If assumed distribution $\mathcal{P}_1$ has continuous distribution function F_1, it is convenient to use the so-called Kolmogorov goodness-of-fit test. Let's consider deviation function

$$\rho(X) = \sqrt{n} \sup_t |F_n^*(t) - F_1(t)|,$$

where F_n^* is an empirical distribution function derived from sample X.

We can show that the introduced function satisfies the conditions described above, but first let's recall the Kolmogorov's theorem.

Theorem 2.2.1 (Kolmogorov's theorem) *Let $X_1, ..., X_n$ be a sample from population ξ having continuous distribution function F_ξ. Then, for empirical distribution function $F_n^*(t)$, the following equation is satisfied*

$$Y_n = \sqrt{n} \cdot \sup_{t \in \mathbb{R}} |F_n^*(t) - F_\xi(t)| \xrightarrow[n \to +\infty]{\mathsf{d}} Y,$$

where random variable Y has the Kolmogorov distribution with distribution function

$$F_Y(t) = \begin{cases} \sum\limits_{i=-\infty}^{+\infty} (-1)^i e^{-2i^2t^2}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

Well, let's check all the points. Let $\rho(X) = \sqrt{n} \sup\limits_t |F_n^*(t) - F_1(t)|$, then

- If H_1 is true, all X_i have distribution $\mathcal{P}_1$. By Kolmogorov's theorem, the distribution of $\rho(X)$ converges to a random variable having the Kolmogorov's distribution.

- If H_1 is wrong, all random variables X_i have distribution $\mathcal{P}_2$ different from $\mathcal{P}_1$. Then, by the Glivenko–Cantelli theorem,

$$F_n^* \xrightarrow[n \to +\infty]{\mathsf{P}} F_2$$

for each $y \in \mathbb{R}$. However, since $\mathcal{P}_1 \neq \mathcal{P}_2$, then there is t_0 so that $F_1(t_0) \neq F_2(t_0)$, which means that

$$\sup_{t \in \mathbb{R}} |F_n^*(t) - F_1(t)| \geq |F_n^*(t_0) - F_1(t_0)| \xrightarrow[n \to +\infty]{\mathsf{P}} |F_2(t_0) - F_1(t_0)| > 0.$$

But then

$$\rho(X) = \sqrt{n} \sup_{t \in \mathbb{R}} |F_n^*(t) - F_1(t)| \xrightarrow[n \to +\infty]{\mathsf{P}} \infty.$$

How hypothesis testing is done? Let random variable Y have a distribution with the Kolmogorov's distribution function. This distribution is tabulated (the table can be found in the additional materials to the module). Thus, given a $\varepsilon > 0$, it is easy to find such C that $\varepsilon = \mathsf{P}(Y \geq C)$, and then the Kolmogorov goodness-of-fit test condition can be written as:

$$\delta(X) = \begin{cases} \mathsf{H}_1, & \rho(X) < C \\ \mathsf{H}_2, & \rho(X) \geq C \end{cases}.$$

An example of the application of the goodness-of-fit test to the data and step-by-step algorithms are given in the quiz for this part.

In addition to the Kolmogorov goodness-of-fit test, the Pearson's chi-squared test χ^2 is often used to test both parametric and non-parametric hypotheses. To familiarize yourself with this test, please refer to the additional materials.

2.3 Test of Homogeneity

As we have noted before, it is useful to know whether the data is taken from the same distribution or not. For example, does the height of basketball team players and the height of children in kindergarten obey the same distribution? Probably not, you would say. But the answer to this question is not always that obvious.

Thus, let us have two samples $X = (X_1, X_2, ..., X_n)$ and $Y = (Y_1, Y_2, ..., Y_m)$ from unknown distributions $\mathcal{P}$ and $\mathcal{G}$, respectively. We are going to test complex hypothesis $\mathsf{H}_1 = \{\mathcal{P} = \mathcal{G}\}$ against alternative hypothesis $\mathsf{H}_2 = \{\mathsf{H}_1 \text{ is wrong}\}$. We will also assume that distributions $\mathcal{P}$ and $\mathcal{G}$ have continuous distribution functions.

As in the Kolmogorov goodness-of-fit test, it it worth considering the deviation function:

$$\rho(X, Y) = \sqrt{\frac{mn}{m+n}} \sup_{t \in \mathbb{R}} |F_n^*(t) - G_m^*(t)|,$$

where function $F_n^*(t)$ is an empirical distribution function derived from sample X, and $G_m^*(t)$ is an empirical distribution function derived from sample Y.

The following theorem turns out to be valid.

Theorem 2.3.1 *If hypothesis* H_1 *is true, then*

$$\rho(X, Y) \xrightarrow[m,n \to +\infty]{\mathsf{d}} Y,$$

where random variable Y has the Kolmogorov distribution.

As in the previous point, given ε, let's find such C that $\varepsilon = \mathsf{P}(Y \geq C)$, Kolmogorov–Smirnov goodness-of-fit test takes the following form:

$$\delta(X, Y) = \begin{cases} \mathsf{H}_1, & \rho(X, Y) < C \\ \mathsf{H}_2, & \rho(X, Y) \geq C \end{cases}.$$

An example of the application of the goodness-of-fit test to the data and step-by-step algorithms are given in the quiz for this part.

2.4 Hypotheses Testing with Normal Distribution

One of the most important distributions is a normal one. Let's consider some hypotheses (and, of course, their testing methods), concerning the normal distribution parameters.

2.4.1 Comparing Two Means of Normally Distributed Samples with Equal Variances

So, let us have two independent samples $X = (X_1, X_2, ..., X_n)$ from distribution $\mathsf{N}_{a_1,\sigma^2}$ and $Y = (Y_1, Y_2, ..., Y_m)$ from distribution $\mathsf{N}_{a_2,\sigma^2}$. Variance σ^2 is the same for both distributions but unknown. Health care studies often require to verify the equality of the means of two independent samples having normal distributions to determine the presence or absence of the drug effect. This problem is a special case of the two-sample problem of homogeneity. If the variances are different, the problem is solved only in special cases, which we will not discuss.

Well, we are going to test hypothesis $\mathsf{H}_1 = \{a_1 = a_2\}$ using random variable

$$t_{n+m-2} = \sqrt{\frac{nm}{n+m}} \cdot \frac{\left(\overline{X} - a_1\right) - \left(\overline{Y} - a_2\right)}{\sqrt{\frac{(n-1)S_0^2(X)+(m-1)S_0^2(Y)}{n+m-2}}},$$

where $S_0^2(X)$ is unbiased sample variance derived from sample X, and $S_0^2(Y)$ is that derived from sample Y.

The following theorem turns out to be valid.

Theorem 2.4.1 *Random variable t_{n+m-2} has Student's t-distribution T_{n+m-2} with $(n + m - 2)$ degrees of freedom.*

If hypothesis H_1 is valid, it's reasonable to consider the deviation function:

$$\rho(X, Y) = \sqrt{\frac{nm}{n+m}} \cdot \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{(n-1)S_0^2(X)+(m-1)S_0^2(Y)}{n+m-2}}},$$

which, by the formulated theorem, has Student's t-distribution T_{n+m-2}. It's worth formulating the criterion the following way.

Given ε, we can find a quantile $C = \tau_{1-\varepsilon/2}$, that is $(1-\varepsilon/2)$ quantile of Student's t-distribution T_{n+m-2}. For such a quantile

$$\mathsf{P}(|t_{n+m-2}| > C) = \varepsilon$$

and Student's t-test has the same form as the goodness-of-fit tests:

$$\delta(X,Y) = \begin{cases} \mathsf{H}_1, & |\rho(X,Y)| < C \\ \mathsf{H}_2, & |\rho(X,Y)| \geq C. \end{cases}$$

An example of t-test application to the data and step-by-step algorithms are given in the quiz for this part.

2.4.2 Variances of Two Normally Distributed Samples

Even if the samples are taken from the normal distribution, we often don't know whether they have the same variances or not, and it is not clear whether the described Student's t-test can be applied. To verify this fact, the so-called Fisher's F-test is often used.

So, let's have two independent samples $X = (X_1, X_2, ..., X_n)$ from distribution $\mathsf{N}_{a_1,\sigma_1^2}$ and $Y = (Y_1, Y_2, ..., Y_m)$ from distribution $\mathsf{N}_{a_2,\sigma_2^2}$. We are going to check hypothesis $\mathsf{H}_1 = \{\sigma_1^2 = \sigma_2^2\}$.

Let's define the deviation function the following way:

$$\rho(X,Y) = \frac{S_0^2(X)}{S_0^2(Y)}$$

The following theorem is valid.

Theorem 2.4.2 *If hypothesis H_1 is true, then $\rho(X,Y)$ has F-distribution $\mathsf{F}_{n-1,m-1}$ with $(n-1, m-1)$ degrees of freedom.*

More information on F-distribution can be found in the additional materials to the module.

Let's formulate F-test. Let $f_{\varepsilon/2}$ and $f_{1-\varepsilon/2}$ be the corresponding quantiles of F-distribution $F_{n-1,m-1}$. Then, F-test is formulated as follows:

$$\delta(X,Y) = \begin{cases} \mathsf{H}_1, & f_{\varepsilon/2} \leq \rho(X,Y) \leq f_{1-\varepsilon/2}, \\ \mathsf{H}_2, & \text{otherwise} \end{cases}.$$

An example of the application of the goodness-of-fit test to the data and step-by-step algorithms are given in the quiz for this part.

2.4.3 Hypothesis on Mean of Normal Distribution with Known Variance

Let us have a sample from normal distribution N_{a,σ^2} with known variance σ^2. Simple hypothesis $\mathsf{H}_1 = \{a = a_0\}$ is tested against complex alternative hypothesis $\mathsf{H}_2 = \{a \neq a_0\}$.

Let's consider a deviation function

$$\rho(X) = \sqrt{n}\frac{\overline{X} - a_0}{\sigma}.$$

If hypothesis H_1 is true, then $\rho(X)$ has a standard normal distribution. If $\varepsilon > 0$, we'll choose $C = \tau_{1-\varepsilon/2}$ that is a quantile of level $(1-\varepsilon/2)$ of the standard normal distribution. Thus.

$$\varepsilon = \mathsf{P}_{\mathsf{H}_1}(|\rho(X)| \geq C)$$

and the criterion, like all goodness-of-fit test criteria, is written in the following form:

$$\delta(X) = \begin{cases} \mathsf{H}_1, & |\rho(X)| < C \\ \mathsf{H}_2, & |\rho(X)| \geq C. \end{cases}$$

An example of the criterion application to the data and step-by-step algorithms are given in the quiz for this part.

2.4.4 Hypothesis on Mean of Normal Distribution with Unknown Variance

Let's have a sample from normal distribution N_{a,σ^2} with unknown variance σ^2. We test simple hypothesis $\mathsf{H}_1 = \{a = a_0\}$ against complex alternative hypothesis $\mathsf{H}_2 = \{a \neq a_0\}$.

Let's consider a deviation function

$$\rho(X) = \sqrt{n}\frac{\overline{X} - a_0}{S_0(X)}.$$

If hypothesis H_1 is true, then $\rho(X)$ has Student's t-distribution T_{n-1}. If $\varepsilon > 0$, we choose $C = \tau_{1-\varepsilon/2}$ that is a quantile of level $(1-\varepsilon/2)$ of the distribution T_{n-1}, then

$$\varepsilon = \mathsf{P}_{\mathsf{H}_1}(|\rho(X)| \geq C)$$

and the criterion, like all goodness-of-fit test criteria, is written in the following form:

$$\delta(X) = \begin{cases} \mathsf{H}_1, & |\rho(X)| < C \\ \mathsf{H}_2, & |\rho(X)| \geq C. \end{cases}$$

An example of the criterion application to the data and step-by-step algorithms are given in the quiz for this part.

2.5 Summary

In this module at the end of the course, we've examined some of the approaches to hypothesis testing. We could also have discussed multivariate samples (not covered in our course), criterion comparing, as well as estimates comparing. The statistical apparatus is rich and extensive, but all these topics cannot be squeezed into a one-term course. We hope that the basic knowledge you've acquired in this course will help you to master the methods that you're interested in. Well, good luck and see you soon!