The purpose of this paper is to provide an introduction to basic ideas and methods of the theory of statistical testing of hypotheses. Only main notions and several classical examples are presented in detail, based on the Neyman-Pearson approach.

1. Introduction

Statistical testing of hypotheses - a major area of mathematical statistics, involves a theory and a set of methods for statistical testing of correspondences between experimental data on the one hand and hypotheses on their probability characteristics on the other. Any statistical problem can be formulated and solved in terms of statistical testing of hypotheses.
2. Statistical Hypothesis

According to the accepted terminology in the theory of statistical testing of hypotheses, any statement (assumption) about the distribution of the observed random element is called a statistical hypothesis or a hypothesis, and we note such hypotheses by \( H, H_0, H_1 \) etc., according to the situation. In mathematical statistics the results of an experiment are treated as the realization of a number of random variables, whether finite or infinite. The joint distribution of these random variables is not completely known or is unknown completely. If a statement determines completely this distribution we speak about a simple hypothesis, say \( H \) for example, otherwise we say that we have a composite hypotheses \( H \).

For example, let \( X \) be a random element taking values in a sample space \( (X, A) \). Suppose that we have in mind a family of distributions \( \mathcal{P} = \{P_\theta, \theta \in \Theta\} \), and we wish to verify the hypothesis \( H \) according to which the distribution of \( X \) belongs to this family. In this case a statistician says that he/she has to test the hypothesis \( H \), which determines the family \( \mathcal{P} \) of possible distributions of \( X \). If we have some a priori information about the distribution of \( X \), which renders the hypothesis \( H \) precise, that is, we can propose a null hypothesis \( H_0 \) according to which the distribution of \( X \) belongs to a subset \( \mathcal{P}_0 \) :

\[
\mathcal{P}_0 \subset \mathcal{P}, \quad \mathcal{P}_0 = \{P_\theta, \theta \in \Theta_0\}, \quad \text{where} \quad \Theta_0 \subset \Theta. \tag{1}
\]

In this case we obtain another interesting statistical problem: to test the null hypothesis \( H_0 \) versus the alternative hypothesis \( H_1 \), according to which the distribution of \( X \) belongs to the family of distributions

\[
\mathcal{P}_1 = \mathcal{P} \setminus \mathcal{P}_0 = \{P_\theta, \theta \in \Theta_1\},
\]

where \( \Theta_1 = \Theta \setminus \Theta_0 \). Concisely, we write that we need to test

\[
H_0 : \ P_\theta \in \mathcal{P}_0 \quad \text{versus} \quad H_1 : \ P_\theta \in \mathcal{P}_1. \tag{2}
\]

Often the same problem is written in terms of \( \theta \) by the following way:

\[
H_0 : \ \theta \in \Theta_0 \quad \text{against} \quad H_1 : \ \theta \in \Theta_1. \tag{3}
\]

Example 1. Let a random vector \( X = (X_1, \ldots, X_n) \) be observed, with components \( X_1, \ldots, X_n \) that are independent identically-distributed random variables subject to the normal law \( N(\theta, 1) \), with unknown mathematical expectation

\[
\theta = E_\theta X_1, \quad \theta \in \Theta = \mathbb{R}^1 = (-\infty, +\infty),
\]

while the variance is equal to 1, i.e. for any real number \( x \),

\[
P_\theta \{X_i \leq x\} = \Phi(x - \theta), \quad (i = 1, \ldots, n),
\]
where
\[ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt \]
is the distribution function of the standard normal law \( N(0, 1) \). Under these conditions it is possible to examine the problem of testing \( H_0 : \theta = \theta_0 \) against \( H_1 : \theta \neq \theta_0 \), where \( \theta_0 \) is a given number. In the given example, \( H_0 \) is simple, since \( \Theta_0 = \{ \theta_0 \} \), while \( H_1 \) is a composite two-sided hypothesis, since
\[ \Theta_1 = (-\infty, \theta_0) \cup (\theta_0, \infty) . \]

**Example 2.** We have the same data as in Example 1, and we wish to test
\[ H_0 : \theta \leq \theta_0 \text{ versus } H_1 : \theta > \theta_0 , \quad (4) \]
where \( \theta_0 \) is a given number. In this case both hypotheses \( H_0 \) and \( H_1 \) are one-sided composite hypotheses.

Formally, the competing hypotheses \( H_0 \) and \( H_1 \) are equivalent in the problem of choosing between them, and the question of which of these two non-intersecting and mutually-complementary sets from \( H \) should be called the null hypotheses is not vital and does not affect construction of the theory of statistical hypotheses testing itself. However, as a rule, the objective of study of the problem itself affects the choice of the null hypothesis, with the result that the null hypothesis is often taken to be that subset \( H_0 \) of the set \( H \) of all admissible hypotheses that in the researcher's opinion, bearing in mind the nature of the phenomenon in question, or in the light of any physical considerations, will best fit in with the expected experimental data. For this very reason, \( H_0 \) is often explained by the fact that, as a rule, \( H_0 \) has a simpler structure that \( H_1 \), as reflected in the researcher's preference to the simpler model. Of course, often a statistician is in a position when he/she is unable to construct two competing hypotheses; he/she is only in a position to study the initial hypothesis \( H \), which plays the role of the null hypothesis, \( H = H_0 \). In such case he/she has to construct the so-called goodness-of-fit test for testing \( H_0 \). One can consider also that in this situation a statistician tests \( H_0 \) against all other distributions (hypothesis \( H_1 \)), which are not in the family of distributions, determined by \( H_0 \).

For example, let \( X = (X_1, \ldots, X_n) \) be a sample, i.e. \( X_1, \ldots, X_n \) are independent identically distributed random variables, and we want to test a simple hypothesis
\[ H_0 : \mathbb{P}\{X_i \leq x\} = F_0(x), \quad x \in \mathbb{R}^1 , \quad (5) \]
where \( F_0 \) is given continuous distribution function. In a case when \( F_0(\cdot) = \Phi(\cdot) \) we test that our sample is taken from the standard normal distribution. As alternative \( H_1 \) for \( H_0 : X_i \sim F_0 \), one can consider, for example, the family all other distribution functions (may be discrete also). To test \( H_0 \) it is recommended to apply Kolmogorov test or Pearson
chi-square test. If $H_0$ is composite one can use Pearson chi-square test, for example. For more about these aspects see, for example, Greenwood and Nikulin (1996).

3. Statistical Test

In mathematical statistics the solution to the problem of testing $H_0$ against $H_1$ is given in terms of a statistical test constructed for this purpose. A statistical test is a decision rule according to which a decision

"the null hypothesis $H_0$ is true" or "the alternative hypothesis $H_1$ is true"

is taken on the basis of results of observations on $X = (X_1, \ldots, X_n)$ or on some statistic $T_n = T_n(X_1, \ldots, X_n)$, constructed for this problem.

In more general terms a statistical test to test $H_0$ against $H_1$ is based on the so-called critical function or the decision function $\varphi(\cdot)$ such that

$$0 \leq j(x) \leq 1, \quad x \in \mathcal{X}.$$  \hspace{1cm} (6)

Let us suppose that in the experiment was obtained $X = x$. According to the statistical test based on the given critical function $j(\cdot)$, the null hypothesis $H_0$ is rejected with probability $j(x)$ in favor of the alternative hypothesis $H_1$, and with probability $1 - j(x)$ the hypothesis $H_0$ is accepted. In other words, a statistical test based on a critical function $j(\cdot)$ is a ride which assigns to the observation $X = x$ a probability $j(x)$ that $H_0$ will be rejected and a probability $1 - j(x)$ that the alternative $H_1$ will be rejected. So one can note that the structure of any statistical test is completely determined by its critical function; different critical functions determine different tests. Sometimes we say $j$-statistical test or more shortly $j$-test to underline that the test is based on the critical function $j$. In general speaking $j(\cdot)$ could be any $\mathcal{B}$-measurable function, mapping the sample space $\mathcal{X}$ onto the interval $[0, 1]$.

4. Errors of the First and the Second Kind

The use of a statistical test leads either to a correct decision being taken or to one of the following two errors being made: rejection of $H_0$ and then acceptance of $H_1$, when in fact $H_0$ is correct (called the first kind error or the Type I error), or acceptance of $H_0$ when in fact $H_1$ is true (called the second kind error or the Type II error).

We have to note that if the null hypothesis is accepted it does not prove that it is true. We keep $H_0$ until it does not contradict evidently to the new data. To control these errors (to minimize at least one of them, for example) we need the next very important notions such as the power function, the power and the significance level of the test.
Bibliography


Kolmogorov A.N. (1933). *Sulla determinazione empirica di una legge di distribuzione*, Giorn.Inst. Ital. Attuari, V.4, 83-91. [In this paper is given the original demonstration of the famous theorem of Kolmogorov on the limit distribution of the empirical processes is given]


Zacks S. (1971). The theory of Statistical Inference. New York: John Wiley & Sons. [This presents the unified approach the theory of statistical estimation and the theory of statistical testing of hypotheses with applications in the sequential analysis]

Biographical Sketch