STATISTICAL TESTING OF HYPOTHESES

Mikhail S. Nikulin
Université Victor Bordeaux 2, France

Keywords: Acceptance region, Bayesian approach, Binomial distribution, Chi-squared distribution, Chi-squared test, Composite hypothesis, Consistency, Critical function, Critical region, Critical value, Error of the first kind, Error of the second kind, Exponential family, Fuzzy data, Goodness-of-fit test, Hypothesis, Likelihood-ratio test, Non-parametric hypothesis, Kolmogorov test, Minimax approach, Neyman-Pearson Lemme, Neyman structure, Non-randomized test, Normal approximation, Poisson distribution, Parametric hypothesis, Power function, Power of test, Randomized test, Sample space, Sign test, Significance level, Size of test, Simple hypothesis, Sufficient statistic, Statistical hypothesis, Statistical testing hypotheses, Test, Unbiased Test, Uniformly most powerful test.

Contents

Summary

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 hypotheses*, 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* $(\mathcal{X}, \mathcal{A})$. 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 prioir* 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 $\mathcal{P}_0$:

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

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

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

Often the same problem is written in terms of θ by the following way:

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

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

$$\theta = \mathbf{E}_\theta X_1, \qquad \theta \in \Theta = \mathbf{R}^1 = (-\infty,\ +\infty),$$

while the variance is equal to 1, i.e. for any real number x,
$$\mathbf{P}_\theta\{X_i \le x\} = \Phi(x-\theta), \qquad (i = 1, \dots, n),$$

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 θ_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

$$H_0 : \quad \theta \leq \theta_0 \qquad \text{versus} \qquad H_1 : \quad \theta > \theta_0, \tag{4}$$

where θ_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

$$H_0 : \quad \mathbf{P}\{X_i \leq x\} = F_0(x), \qquad x \in \mathbf{R}^1, \tag{5}$$

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 = ($X_1, \dots, X_n$) or on some statistic $T_n = T_n(X_1, \dots, X_n)$, 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

$$0 \le j(x) \le 1, \qquad x \in \mathcal{X}. \tag{6}$$

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 he 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

Andersen P.K., Borgan O., Gill R. and Keiding N. (1993). *Statistical Model Based on Counting Processes*, 767 pp. New York, Berlin, Heidelberg: Springer-Verlag. [This book presents an excellent introduction in modern statistical methods in survival analysis]

Bagdonavičius V. and Nikulin M. (2000). *On goodness-of-fit for the linear transformation and frailty models*. Statistics and Probability Letters, Vol. 47, pp. 177-188. [This presents new statistical tests for testing validity of two important models in survival analysis and reliability]

Bagdonavičius V. and Nikulin M. (2001). *On goodness-of-fit for accelerated life models*. Comptes Rendus de l'Académie des Sciences de Paris,v.331, #1, pp. 171-176 [This present a new test for generalized Sedyakin model when accelerated experiments done under step-stresses versus the Cox proportional hazards model with time-varying stresses]

Bolshev L.N. and Smirnov N.V. *Tables of mathematics statistics*, Nauka, Moscow (1983). [This book gives a rich collection of statistical tables well adapted for different numerical computations on computers]

Cox D.R. and Hinkley P.Y. 1974. *Theoretical Statistics*, Chapman and Hall. [This book presents a good introduction in the theory of mathematical statistics]

Cox D.R. and Oakes D. (1984). *Analysis of Survival Data*, 201 pp. London, New York, Tokyo: Chapman & Hall. [This book gives an introduction in survival analysis]

Cramer H. (1946). *Mathematical methods of statistics*. Princeton: University Press. [Classical textbook in mathematical statistics in the second half of the twenty century]

Kolmogorov A.N. (1933). *Sulla determinazione empirica di una legge di distribuzione*, Giorn.lnst. 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]

Greenwood P. and Nikulin M. (1996). *A Guide to Chi-squared Testing*, 280 pp. New York, Chichester, Brisbane: John Wiley & Sons.[This monograph presents the theory of chi-squared type tests and its applications]

Hajek, J. and Sidák Z. (1967). *Theory of rank tests*. New York: Academic Press.[This monograph gives an introduction in the theory of rank tests]

Lehmann E.L. (1983) *Theory of Point estimation*. New York: John Wiley & Sons. [This monograph introduces all essential aspects of the classical theory of statistical estimation]

Lehmann E.L. (1988). *Testing statistical hypotheses*. New York: John Wiley & Sons. [This monograph introduces all essential aspects of the classical theory of statistical testing of hypotheses]

Meeker W.Q. and Escobar L.A. (1998). *Statistical Methods for Reliability Data*, 680pp. New York, Chichester, Weinheirn: John Wiley & Sons. [This monograph presents different modern statistical models, methods and approaches applied in reliability]

Neyman J. and Pearson E.S. (1933). On the problem of the most efficient tests of statistical hypotheses, *Phil. Trans. Roy. Soc. London Ser. A*, **A 231**, pp. 289-337. [This presents the original demonstration of the famous Neyman-Pearson lemma]

Nikulin M.S. (1989). A result of Bolshev's from the theory of the statistical hypothesis testing, *J. Soviet Math*. **44**, no. 3, pp. 522-529. *Zap. Nauchn. Sem. Inst. Steklov* **153** (1986), 129-137). [This paper investigates some generalizations of Neyman-Pearson lemma]

Rao C.R. (1973). *Linear Statistical Methods*. New York: John Wiley & Sons. [This presents a good collection of linear statistical methods]

Viertl R. (1996). *Statistical Methods for Non-Precise Data*, 191 pp. Boca Peaton, New York, London Tokyo: CRC Press. [This introduces some new approaches to use statistical methods to the situation of non-precise (fuzzy) data]

Voinov V. and Nikulin M. (1993). *Unbiased Estimators and Their Applications*. *Vol. 1: Univariate Case*, 560 pp. Dordrecht, Boston, London: Kluwer Academic Publishers. [This presents the theory, methods of computations and applications of unbiased estimators of parameters of univariate distributions]

Voinov V and Nikulin M. (1996). *Unbiased Estimators and Their Applications.Vol.2: Multivariate Case*, 262 pp. Dordrecht, Boston, London: Kluwer Academic Publishers. [This presents the theory, methods of computations and applications of unbiased estimators of parameters of multivariate distributions]

Wald A. (1950). *Sequential analysis*. New York: John Wiley & Sons. [This presents an introduction in the theory of statistical sequential analysis]

Wilks S.S. (1938). The large-simple distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statistics*, vol.9, pp. 60-62. [This presents the original demonstration of the limit distribution of the likelihood ratio statistic]

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

Mikhail Nikulin was born on 29.04.1944 in Saint Petersburg, Russia. He is Professor of Statistics and Director of the Laboratory " Mathematical Statistics and Its Applications}" at Victor Segalen University Bordeaux 2 and a member of the Laboratory of Statistical Methods in the Steklov Mathematical Institute, St. Petersburg, Russia. He obtained Ph.D in Theory of Probability and Mathematical Statistics (1973) from the Steklov Mathematical Institute in Moscow. He published more then 200 papers in probability theory and statistics. Co-author with V.Voinov of " Unbiased Estimators and Their Applications", vol. 1 : Univariate Case, (1993); vol. 2 : Multivariate Case (1996), Kluwer Academic Publishers, co-author with P.E.Greenwood of " A Guide to Chi-squared Testing", John Wiley & Sons (1996), co-author with D.K.Faddeev and I.F.Sokolovsky of "Elements of higher mathematics for high school students", Mir Publisher, Moscow, (1989), and coauthor with V.Bagdonavicius of “Accelerated Life Models”, Chapman and Hall,CRC (2001).