Nonparametric Tests for Complete Data

Vilijandonas Bagdonavicius
Julius Kruopis
Mikhail S. Nikulin
Non-parametric Tests for Complete Data
Non-parametric Tests for Complete Data

Vilijandas Bagdonavičius
Julius Kruopis
Mikhail S. Nikulin
Table of Contents

Preface .................................................. xi

Terms and Notation ................................. xv

Chapter 1. Introduction ............................. 1
  1.1. Statistical hypotheses ...................... 1
  1.2. Examples of hypotheses in non-parametric
      models ........................................ 2
     1.2.1. Hypotheses on the probability distribution
            of data elements .......................... 2
     1.2.2. Independence hypotheses ............... 4
     1.2.3. Randomness hypothesis ................. 4
     1.2.4. Homogeneity hypotheses ............... 4
     1.2.5. Median value hypotheses ............... 5
  1.3. Statistical tests ......................... 5
  1.4. $P$-value .................................... 7
  1.5. Continuity correction .................... 10
  1.6. Asymptotic relative efficiency .......... 13

Chapter 2. Chi-squared Tests .................... 17
  2.1. Introduction ............................... 17
  2.2. Pearson's goodness-of-fit test: simple hypothesis . 17
2.3. Pearson’s goodness-of-fit test: composite hypothesis .......................... 26
2.4. Modified chi-squared test for composite hypotheses .......................... 34
   2.4.1. General case ........................................... 35
   2.4.2. Goodness-of-fit for exponential distributions .......................... 41
   2.4.3. Goodness-of-fit for location-scale and shape-scale families .......... 43
2.5. Chi-squared test for independence ................................................. 52
2.6. Chi-squared test for homogeneity .................................................. 57
2.7. Bibliographic notes ................................................................. 64
2.8. Exercises .................................................................................. 64
2.9. Answers .................................................................................. 72

Chapter 3. Goodness-of-fit Tests Based on Empirical Processes ............... 77
   3.1. Test statistics based on the empirical process ............................... 77
   3.2. Kolmogorov–Smirnov test ......................................................... 82
   3.3. $\omega^2$, Cramér–von-Mises and Andersen–Darling tests ............... 86
   3.5. Two-sample tests ................................................................ 98
   3.5.1. Two-sample Kolmogorov–Smirnov tests .................................. 98
   3.5.2. Two-sample Cramér–von-Mises test ..................................... 103
   3.6. Bibliographic notes ................................................................. 104
   3.7. Exercises ............................................................................. 106
   3.8. Answers .............................................................................. 109

Chapter 4. Rank Tests ........................................................................... 111
   4.1. Introduction ........................................................................... 111
   4.2. Ranks and their properties ....................................................... 112
   4.3. Rank tests for independence ..................................................... 117
   4.3.1. Spearman’s independence test ............................................. 117
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.2. Kendall’s independence test</td>
<td>124</td>
</tr>
<tr>
<td>4.3.3. ARE of Kendall’s independence test with respect to Pearson’s independence test under normal distribution</td>
<td>133</td>
</tr>
<tr>
<td>4.3.4. Normal scores independence test</td>
<td>137</td>
</tr>
<tr>
<td>4.4. Randomness tests</td>
<td>139</td>
</tr>
<tr>
<td>4.4.1. Kendall’s and Spearman’s randomness tests</td>
<td>140</td>
</tr>
<tr>
<td>4.4.2. Bartels–Von Neuman randomness test</td>
<td>143</td>
</tr>
<tr>
<td>4.5. Rank homogeneity tests for two independent samples</td>
<td>146</td>
</tr>
<tr>
<td>4.5.1. Wilcoxon (Mann–Whitney–Wilcoxon) rank sum test</td>
<td>146</td>
</tr>
<tr>
<td>4.5.2. Power of the Wilcoxon rank sum test against location alternatives</td>
<td>153</td>
</tr>
<tr>
<td>4.5.3. ARE of the Wilcoxon rank sum test with respect to the asymptotic Student’s test</td>
<td>155</td>
</tr>
<tr>
<td>4.5.4. Van der Warden’s test</td>
<td>161</td>
</tr>
<tr>
<td>4.5.5. Rank homogeneity tests for two independent samples under a scale alternative</td>
<td>163</td>
</tr>
<tr>
<td>4.6. Hypothesis on median value: the Wilcoxon signed ranks test</td>
<td>168</td>
</tr>
<tr>
<td>4.6.1. Wilcoxon’s signed ranks tests</td>
<td>168</td>
</tr>
<tr>
<td>4.6.2. ARE of the Wilcoxon signed ranks test with respect to Student’s test</td>
<td>177</td>
</tr>
<tr>
<td>4.7. Wilcoxon’s signed ranks test for homogeneity of two related samples</td>
<td>180</td>
</tr>
<tr>
<td>4.8. Test for homogeneity of several independent samples: Kruskal–Wallis test</td>
<td>181</td>
</tr>
<tr>
<td>4.9. Homogeneity hypotheses for k related samples: Friedman test</td>
<td>191</td>
</tr>
<tr>
<td>4.10. Independence test based on Kendall’s concordance coefficient</td>
<td>204</td>
</tr>
<tr>
<td>4.11. Bibliographic notes</td>
<td>208</td>
</tr>
<tr>
<td>4.12. Exercises</td>
<td>209</td>
</tr>
<tr>
<td>4.13. Answers</td>
<td>212</td>
</tr>
</tbody>
</table>
Chapter 5. Other Non-parametric Tests 215

5.1. Sign test 215

5.1.1. Introduction: parametric sign test 215

5.1.2. Hypothesis on the nullity of the medians of the differences of random vector components 218

5.1.3. Hypothesis on the median value 220

5.2. Runs test 221

5.2.1. Runs test for randomness of a sequence of two opposite events 223

5.2.2. Runs test for randomness of a sample 226

5.2.3. Wald–Wolfowitz test for homogeneity of two independent samples 228

5.3. McNemar’s test 231

5.4. Cochran test 238

5.5. Special goodness-of-fit tests 245

5.5.1. Normal distribution 245

5.5.2. Exponential distribution 253

5.5.3. Weibull distribution 260

5.5.4. Poisson distribution 262

5.6. Bibliographic notes 268

5.7. Exercises 269

5.8. Answers 271

APPENDICES 275

Appendix A. Parametric Maximum Likelihood Estimators: Complete Samples 277

Appendix B. Notions from the Theory of Stochastic Processes 281

B.1. Stochastic process 281

B.2. Examples of stochastic processes 282

B.2.1. Empirical process 282

B.2.2. Gauss process 283

B.2.3. Wiener process (Brownian motion) 283

B.2.4. Brownian bridge 284

B.3. Weak convergence of stochastic processes 285
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.4. Weak invariance of empirical processes</td>
<td>286</td>
</tr>
<tr>
<td>B.5. Properties of Brownian motion and Brownian bridge</td>
<td>287</td>
</tr>
<tr>
<td>Bibliography</td>
<td>293</td>
</tr>
<tr>
<td>Index</td>
<td>305</td>
</tr>
</tbody>
</table>
Preface

Testing hypotheses in non-parametric models are discussed in this book. A statistical model is non-parametric if it cannot be written in terms of a finite-dimensional parameter. The main hypotheses tested in such models are hypotheses on the probability distribution of elements of the following: data homogeneity, randomness and independence hypotheses. Tests for such hypotheses from complete samples are considered in many books on non-parametric statistics, including recent monographs by Maritz [MAR 95], Hollander and Wolfe [HOL 99], Sprent and Smeeton [SPR 01], Govindarajulu [GOV 07], Gibbons and Chakraborti [GIB 09] and Corder and Foreman [COR 09].

This book contains tests from complete samples. Tests for censored samples can be found in our book Tests for Censored Samples [BAG 11].

In Chapter 1, the basic ideas of hypothesis testing and general hypotheses on non-parametric models are briefly described.

In the initial phase of the solution of any statistical problem the analyst must choose a model for data analysis. The correctness of the data analysis strongly depends on the choice
of an appropriate model. Goodness-of-fit tests are used to check the adequacy of a model for real data.

One of the most-applied goodness-of-fit tests are chi-squared type tests, which use grouped data. In many books on statistical data analysis, chi-squared tests are applied incorrectly. Classical chi-squared tests are based on theoretical results which are obtained assuming that the ends of grouping intervals do not depend on the sample, and the parameters are estimated using grouped data. In real applications, these assumptions are often forgotten. The modified chi-squared tests considered in Chapter 2 do not suffer from such drawbacks. They are based on the assumption that the ends of grouping intervals depend on the data, and the parameters are estimated using initially non-grouped data.

Another class of goodness-of-fit tests based on functionals of the difference of empirical and theoretical cumulative distribution functions is described in Chapter 3. The tests for composite hypotheses classical statistics are modified by replacing unknown parameters by their estimators. Application of these tests is often incorrect because the critical values of the classical tests are used in testing the composite hypothesis and applying modified statistics.

In section 5.5, special goodness-of-fit tests which are not from the two above-mentioned classes, and which are specially designed for specified probability distributions, are given.

Tests for the equality of probability distributions (homogeneity tests) of two or more independent or dependent random variables are considered in several chapters. Chi-squared type tests are given in section 2.5 and tests based on functionals of the difference of empirical distribution functions are given in section 3.5. For many alternatives, the
most efficient tests are the rank tests for homogeneity given in sections 4.4 and 4.6–4.8.

Classical tests for the independence of random variables are given in sections 2.4 (tests of chi-square type), and 4.3 and 5.2 (rank tests).

Tests for data randomness are given in sections 4.3 and 5.2.

All tests are described in the following way: 1) a hypothesis is formulated; 2) the idea of test construction is given; 3) a statistic on which a test is based is given; 4) a finite sample and (or) asymptotic distribution of the test statistic is found; 5) a test, and often its modifications (continuity correction, data with \textit{ex aequo}, various approximations of asymptotic law) are given; 6) practical examples of application of the tests are given; and 7) at the end of the chapters problems with answers are given.

Anyone who uses non-parametric methods of mathematical statistics, or wants to know the ideas behind and mathematical substantiation of the tests, can use this book. It can be used as a textbook for a one-semester course on non-parametric hypotheses testing.

Knowledge of probability and parametric statistics are needed to follow the mathematical developments. The basic facts on probability and parametric statistics used in the the book are also given in the appendices.

The book consists of five chapters, and appendices. In each chapter, the numbering of theorems, formulas, and comments are given using the chapter number.

The book was written using lecture notes for graduate students in Vilnius and Bordeaux universities.
We thank our colleagues and students at Vilnius and Bordeaux universities for comments on the content of this book, especially Rūta Levulienė for writing the computer programs needed for application of the tests and solutions of all the exercises.

Vilijandas Bagdonavičius
Julius Kruopis
Mikhail Nikulin
Terms and Notation

\[ ||A|| \] – the norm \((\sum_i \sum_j a_{ij}^2)^{1/2}\) of a matrix \(A = [a_{ij}]\);

\(A > B\) \((A \geq B)\) – the matrix \(A - B\) is positive (non-negative) definite;

\(a \lor b\) \((a \land b)\) – the maximum (the minimum) of the numbers \(a\) and \(b\);

ASE – the asymptotic relative efficiency;

\(B(n, p)\) – binomial distribution with parameters \(n\) and \(p\);

\(B^{-}(n, p)\) – negative binomial distribution with parameters \(n\) and \(p\);

\(Be(\gamma, \eta)\) – beta distribution with parameters \(\gamma\) and \(\eta\);

cdf – the cumulative distribution function;

CLT – the central limit theorem;

\(\text{Cov}(X, Y)\) – the covariance of random variables \(X\) and \(Y\);

\(\text{Cov}(X, Y)\) – the covariance matrix of random vectors \(X\) and \(Y\);
\( EX \) – the mean of a random variable \( X \);
\( \mathbf{E}(X) \) – the mean of a random vector \( X \);
\( \mathbf{E}_\theta(X), \mathbf{E}(X|\theta), \mathbf{Var}_\theta(X), \mathbf{Var}(X|\theta) \) – the mean or the variance of a random variable \( X \) depending on the parameter \( \theta \);
\( \mathcal{E}(\lambda) \) – exponential distribution with parameters \( \lambda \);
\( F(m, n) \) – Fisher distribution with \( m \) and \( n \) degrees of freedom;
\( F(m, n; \delta) \) – non-central Fisher distribution with \( m \) and \( n \) degrees of freedom and non-centrality parameter \( \delta \);
\( F_\alpha(m, n) \) – \( \alpha \) critical value of Fisher distribution with \( m \) and \( n \) degrees of freedom;
\( F_T(x) \) (\( f_T(x) \)) – the cdf (the pdf) of the random variable \( T \);
\( f(x; \theta), f(x|\theta) \) – the pdf depending on a parameter \( \theta \);
\( F(x; \theta), F(x|\theta) \) – the cdf depending on a parameter \( \theta \);
\( G(\lambda, \eta) \) – gamma distribution with parameters \( \lambda \) and \( \eta \);
\( \text{iid} \) – independent identically distributed;
\( LN(\mu, \sigma) \) – lognormal distribution with parameters \( \mu \) and \( \sigma \);
\( \text{LS} \) – least-squares (method, estimator);
\( \text{ML} \) – maximum likelihood (function, method, estimator);
\( N(0, 1) \) – standard normal distribution;
\( N(\mu, \sigma^2) \) – normal distribution with parameters \( \mu \) and \( \sigma^2 \);
\( N_k(\mathbf{\mu}, \Sigma) \) – \( k \)-dimensional normal distribution with the mean vector \( \mathbf{\mu} \) and the covariance matrix \( \Sigma \);

\( \mathcal{P}(\lambda) \) – Poisson distribution with a parameter \( \lambda \);

pdf – the probability density function;

\( \mathbf{P}\{A\} \) – the probability of an event \( A \);

\( \mathbf{P}\{A|B\} \) – the conditional probability of event \( A \);

\( \mathbf{P}_\theta\{A\}, \mathbf{P}\{A|\theta\} \) – the probability depending on a parameter \( \theta \);

\( \mathcal{P}_k(n, \pi) \) – \( k \)-dimensional multinomial distribution with parameters \( n \) and \( \pi = (\pi_1, \ldots, \pi_k)^T, \pi_1 + \ldots + \pi_k = 1 \);

rv – random variable

\( S(n) \) – Student’s distribution with \( n \) degrees of freedom;

\( S(n; \delta) \) – non-central Student’s distribution with \( n \) degrees of freedom and non-centrality parameter \( \delta \);

\( t_\alpha(n) \) – \( \alpha \) critical value of Student’s distribution with \( n \) degrees of freedom;

\( U(\alpha, \beta) \) – uniform distribution in the interval \((\alpha, \beta)\);

UMP – uniformly most powerful (test);

UUMP – unbiased uniformly most powerful (test);

\( \text{Var}X \) – the variance of a random variable \( X \);

\( \text{Var}(\mathbf{X}) \) – the covariance matrix of a random vector \( \mathbf{X} \);

\( W(\theta, \nu) \) – Weibull distribution with parameters \( \theta \) ir \( \nu \);

\( X, Y, Z, \ldots \) – random variables;
\( \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \ldots \) – random vectors;

\( \mathbf{X}^T \) – the transposed vector, i.e. a vector-line;

\[ ||\mathbf{x}|| = \text{the length} \ (\mathbf{x}^T \mathbf{x})^{1/2} = (\sum_i x_i^2)^{1/2} \] of a vector \( \mathbf{x} = (x_1, \ldots, x_k)^T \);

\( X \sim N(\mu, \sigma^2) \) – random variable \( X \) normally distributed with parameters \( \mu \) and \( \sigma^2 \) (analogously in the case of other distributions);

\( X_n \xrightarrow{P} X \) – convergence in probability \( (n \to \infty) \);

\( X_n \xrightarrow{a.s.} X \) – almost sure convergence or convergence with probability 1 \( (n \to \infty) \);

\( X_n \xrightarrow{d} X, F_n(x) \xrightarrow{d} F(x) \) – weak convergence or convergence in distribution \( (n \to \infty) \);

\( X_n \xrightarrow{d} X \sim N(\mu, \sigma^2) \) – random variables \( X_n \) asymptotically \( (n \to \infty) \) normally distributed with parameters \( \mu \) and \( \sigma^2 \);

\( X_n \sim Y_n \) – random variables \( X_n \) and \( Y_n \) asymptotically \( (n \to \infty) \) equivalent \( (X_n - Y_n \xrightarrow{P} 0) \);

\( x(P) \) – \( P \)-th quantile;

\( x_P \) – \( P \)-th critical value;

\( z_{\alpha} \) – \( \alpha \) critical value of the standard normal distribution;

\( \Sigma = [\sigma_{ij}]_{k \times k} \) – covariance matrix;

\( \chi^2(n) \) – chi-squared distribution with \( n \) degrees of freedom;

\( \chi^2(n; \delta) \) – non-central chi-squared distribution with \( n \) degrees of freedom and non-centrality parameter \( \delta \);

\( \chi^2_{\alpha}(n) \) – \( \alpha \) critical value of chi-squared distribution with \( n \) degrees of freedom.
1.1. Statistical hypotheses

The simplest model of statistical data is a simple sample, i.e. a vector $X = (X_1, \ldots, X_n)^T$ of $n$ independent identically distributed random variables. In real experiments the values $x_i$ of the random variables $X_i$ are observed (measured). The non-random vector $x = (x_1, \ldots, x_n)^T$ is a realization of the simple sample $X$.

In more complicated experiments the elements $X_i$ are dependent, or not identically distributed, or are themselves random vectors. The random vector $X$ is then called a sample, not a simple sample.

Suppose that the cumulative distribution function (cdf) $F$ of a sample $X$ (or of any element $X_i$ of a simple sample) belongs to a set $\mathcal{F}$ of cumulative distribution functions. For example, if the sample is simple then $\mathcal{F}$ may be the set of absolutely continuous, discrete, symmetric, normal, Poisson cumulative distribution functions. The set $\mathcal{F}$ defines a statistical model.

Suppose that $\mathcal{F}_0$ is a subset of $\mathcal{F}$.
The statistical hypothesis $H_0$ is the following assertion: the cumulative distribution function $F$ belongs to the set $\mathcal{F}_0$. We write $H_0 : F \in \mathcal{F}_0$.

The hypothesis $H_1 : F \in \mathcal{F}_1$, where $\mathcal{F}_1 = \mathcal{F}\setminus\mathcal{F}_0$ is the complement of $\mathcal{F}_0$ to $\mathcal{F}$ is called alternative to the hypothesis $H_0$.

If $\mathcal{F} = \{F_\theta, \theta \in \Theta \subset \mathbb{R}^m\}$ is defined by a finite-dimensional parameter $\theta$ then the model is parametric. In this case the statistical hypothesis is a statement on the values of the finite-dimensional parameter $\theta$.

In this book non-parametric models are considered. A statistical model $\mathcal{F}$ is called non-parametric if $\mathcal{F}$ is not defined by a finite-dimensional parameter.

If the set $\mathcal{F}_0$ contains only one element of the set $\mathcal{F}$ then the hypothesis is simple, otherwise the hypothesis is composite.

### 1.2. Examples of hypotheses in non-parametric models

Let us look briefly and informally at examples of the hypotheses which will be considered in the book. We do not formulate concrete alternatives, only suppose that models are non-parametric. Concrete alternatives will be formulated in the chapters on specified hypotheses.

#### 1.2.1. Hypotheses on the probability distribution of data elements

The first class of hypotheses considered in this book consists of hypotheses on the form of the cdf $F$ of the elements of a sample.

Such hypotheses may be simple or composite.
A simple hypothesis has the form $H_0 : F = F_0$; here $F_0$ is a specified cdf. For example, such a hypothesis may mean that the $n$ numbers generated by a computer are realizations of random variables having uniform $U(0,1)$, Poisson $\mathcal{P}(2)$, normal $N(0,1)$ or other distributions.

A composite hypothesis has the form $H_0 : F \in \mathcal{F}_0 = \{F_\theta, \theta \in \Theta\}$, where $F_\theta$ are cdfs of known analytical form depending on the finite-dimensional parameter $\theta \in \Theta$. For example, this may mean that the salary of the doctors in a city are normally distributed, or the failure times of TV sets produced by a factory have the Weibull distribution.

More general composite hypotheses, meaning that the data verify some parametric or semi-parametric regression model, may be considered. For example, in investigating the influence of some factor $z$ on the survival time the following hypothesis on the cdf $F_i$ of the $i$-th sample element may be used:

$$F_i(x) = 1 - \left\{1 - F_0(x)\right\}^{\exp\{\beta z_i\}}, i = 1, \ldots, n$$

where $F_0$ is an unknown baseline cdf, $\beta$ is an unknown scalar parameter and $z_i$ is a known value of the factor for the $i$-th sample element.

The following tests for simple hypotheses are considered: 

- chi-squared tests (section 2.2) and tests based on the difference of empirical and cumulative distribution functions (sections 3.2 and 3.3).

The following tests for composite hypotheses are considered: general tests such as chi-squared tests (sections 2.3 and 2.4, tests based on the difference of non-parametric and parametric estimators of the cumulative distribution function (section 3.4), and also special tests for specified families of probability distributions (section 5.5).
1.2.2. **Independence hypotheses**

Suppose that \((X_i, Y_i)^T, \ i = 1, 2, \ldots, n\) is a simple sample of the random vector \((X, Y)^T\) with the cdf \(F = F(x, y) \in \mathcal{F}\); here \(\mathcal{F}\) is a non-parametric class two-dimensional cdf.

An *independence hypothesis* means that the components \(X\) and \(Y\) are independent. For example, this hypothesis may mean that the sum of sales of managers \(X\) and the number of complaints from consumers \(Y\) are independent random variables.

The following tests for independence of random variables are considered: *chi-squared independence tests* (section 2.5) and *rank tests* (sections 4.3 and 4.10).

1.2.3. **Randomness hypothesis**

A randomness hypothesis means that the observed vector \(x = (x_1, \ldots, x_n)^T\) is a realization of a simple sample \(X = (X_1, \ldots, X_n)^T\), i.e. of a random vector with independent and identically distributed (iid) components.

The following tests for randomness hypotheses are considered: *runs tests* (section 5.2) and *rank tests* (section 4.4).

1.2.4. **Homogeneity hypotheses**

A homogeneity hypothesis of two independent simple samples \(X = (X_1, \ldots, X_m)^T\) and \(Y = (Y_1, \ldots, Y_n)^T\) means that the cdfs \(F_1\) and \(F_2\) of the random variables \(X_i\) and \(Y_j\) coincide. The homogeneity hypothesis of \(k > 2\) independent samples is formulated analogously.

The following tests for homogeneity of independent simple samples are considered: *chi-squared tests* (section 2.6), *tests*
based on the difference of cumulative distribution functions (section 3.5), rank tests (sections 4.5 and 4.8), and some special tests (section 5.1).

If $n$ independent random vectors $X_i = (X_{i1}, ..., X_{ik})^T$, $i = 1, ..., n$ are observed then the vectors $(X_{1j}, ..., X_{nj})^T$ composed of the components are $k$ dependent samples, $j = 1, ..., k$. The homogeneity hypotheses of $k$ related samples means the equality of the cdfs $F_1, ..., F_k$ of the components $X_{i1}, ..., X_{ik}$.

The following tests for homogeneity of related samples are considered: rank tests (sections 4.7 and 4.9) and other special tests (sections 5.1, 5.3 and 5.4).

1.2.5. Median value hypotheses

Suppose that $X = (X_1, ..., X_n)^T$ is a simple sample of a continuous random variable $X$. Denote by $M$ the median of the random variable $X$. The median value hypothesis has the form $H : M = M_0$; here $M_0$ is a specified value of the median.

The following tests for this hypothesis are considered: sign tests (section 5.1) and rank tests (section 4.6).

1.3. Statistical tests

A statistical test or simply a test is a rule which enables a decision to be made on whether or not the zero hypothesis $H_0$ should be rejected on the basis of the observed realization of the sample.

Any test considered in this book is based on the values of some statistic $T = T(X) = T(X_1, ..., X_n)$, called the test statistic. Usually the statistic $T$ takes different values under the hypothesis $H_0$ and the alternative $H_1$. If the statistic $T$ has a tendency to take smaller (greater) values under
the hypothesis $H_0$ than under the alternative $H_1$ then the hypothesis $H_0$ is rejected in favor of the alternative if $T > c$ ($T < c$, respectively), where $c$ is a well-chosen real number.

If the values of the statistic $T$ have a tendency to concentrate in some interval under the hypothesis and outside this interval under the alternative then the hypothesis $H_0$ is rejected in favor of the alternative if $T < c_1$ or $T > c_2$, where $c_1$ and $c_2$ are well-chosen real numbers.

Suppose that the hypothesis $H_0$ is rejected if $T > c$ (the other two cases are considered similarly).

The probability

$$\beta(F) = P_F\{T > c\}$$

doing the rejection of the hypothesis $H_0$ when the true cumulative distribution function is a specified function $F \in \mathcal{F}$ is called the power function of the test. When using a test, two types of error are possible:

1. The hypothesis $H_0$ is rejected when it is true, i.e. when $F \in \mathcal{F}_0$. Such an error is called a type I error. The probability of this error is $\beta(F)$, $F \in \mathcal{F}_0$.

2. The hypothesis $H_0$ is not rejected when it is false, i.e. when $F \in \mathcal{F}_1$. Such an error is called a type II error. The probability of this error is $1 - \beta(F)$, $F \in \mathcal{F}_1$.

The number

$$\sup_{F \in \mathcal{F}_0} \beta(F) \quad [1.1]$$

is called the significance level of the test.

Fix $\alpha \in (0, 1)$. If the significance level does not exceed $\alpha$ then for any $F \in \mathcal{F}_0$ the type I error does not exceed $\alpha$. 
Usually tests with significance level values not greater than \( \alpha = 0.1; 0.05; 0.01 \) are used.

If the distribution of the statistic \( T \) is absolutely continuous then, usually, for any \( \alpha \in (0, 1) \) we can find a test based on this statistic such that the significance level is equal to \( \alpha \).

A test with a significance level not greater than \( \alpha \) is called \textit{unbiased}, if

\[
\inf_{F \in \mathcal{F}_1} \beta(F) \geq \alpha
\]

This means that the zero hypothesis is rejected with greater probability under any specified alternative than under the zero hypothesis. Let \( T \) be a class of test statistics of unbiased tests with a significance level not greater than \( \alpha \).

The statistic \( T \) defines the \textit{uniformly most powerful unbiased} test in the class \( T \) if \( \beta_T(F) \geq \beta_{T^*}(F) \) for all \( T^* \in T \) and for all \( F \in \mathcal{F}_1 \).

A test is called \textit{consistent} if for all \( F \in \mathcal{F}_1 \)

\[
\beta(F) \to 1, \quad \text{as} \quad n \to \infty
\]

This means that if \( n \) is large then under any specified alternative the probability of rejecting the zero hypothesis is near to 1.

1.4. \textit{P-value}

Suppose that a simple statistical hypothesis \( H_0 \) is rejected using tests of one of the following forms:

1) \( T \geq c \); 2) \( T \leq c \); or 3) \( T \leq c_1 \) or \( T \geq c_2 \); here \( T = T(X) \) is a test statistic based on the sample \( X = (X_1, \ldots, X_n)^T \).

We write \( P_0[A] = P\{A|H_0\} \).
Fix $\alpha \in (0, 1)$. The first (second) test has a significance level not greater than $\alpha$, and nearest to $\alpha$ if the constant $c = \inf\{s : P_0\{T \geq s\} \leq \alpha\}$ ($c = \sup\{s : P_0\{T \leq s\} \leq \alpha\}$).

The third test has a significance level not greater than $\alpha$ if $c_1 = \sup\{s : P\{T \leq s\} \leq \alpha/2\}$ and $c_2 = \inf\{s : P\{T \geq s\} \leq \alpha/2\}$.

Denote by $t$ the observed value of the statistic $T$. In the case of the tests of the first two forms the $P$-values are defined as the probabilities

$$pv = P_0\{T \geq t\} \quad \text{and} \quad pv = P_0\{T \leq t\}.$$ 

Thus the $P$-value is the probability that under the zero hypothesis $H_0$ the statistic $T$ takes a value more distant than $t$ in the direction of the alternative (in the first case to the right, in the second case to the left, from $t$).

In the third case, if

$$P_0\{T \leq t\} \leq P_0\{T \geq t\}$$

then

$$pv/2 = P_0\{T \leq t\}$$

and if

$$P_0\{T \leq t\} \geq P_0\{T \geq t\}$$

then

$$pv/2 = P_0\{T \geq t\}$$

So in the third case the $P$-value is defined as follows

$$pv = 2 \min\{P_0\{T \leq t\}, P_0\{T \geq t\}\} = 2 \min\{F_T(t), 1 - F_T(t-\cdot)\}$$

where $F_T$ is the cdf of the statistic $T$ under the zero hypothesis $H_0$. If the distribution of $T$ is absolutely continuous and symmetric with respect to the origin, the last formula implies

$$pv = 2 \min\{F_T(t), F_T(-t)\} = 2F_T(-|t|) = 2\{1 - F_T(|t|)\}$$
If the result observed during the experiment is a rare event when the zero hypothesis is true then the $P$-value is small and the hypothesis should be rejected. This is confirmed by the following theorem.

**Theorem 1.1.** Suppose that the test is of any of the three forms considered above. For the experiment with the value $t$ of the statistic $T$ the inequality $pv \leq \alpha$ is equivalent to the rejection of the zero hypothesis.

**Proof.** Let us consider an experiment where $T = t$. If the test is defined by the inequality $T \geq c$ ($T \leq c$) then $c = \inf\{s \colon P_0\{T \geq s\} \leq \alpha\}$ ($c = \sup\{s \colon P_0\{T \leq s\} \leq \alpha\}$) and $P_0\{T \leq t\} = pv$ ($P_0\{T \geq t\} = pv$). So the inequality $pv \leq \alpha$ is equivalent to the inequality $t \geq c$ ($t \leq c$). The last inequalities mean that the hypothesis is rejected.

If the test is defined by the inequalities $T \leq c_1$ or $T \geq c_2$ then $c_1 = \sup\{s \colon P_0\{T \leq s\} \leq \alpha/2\}$, $c_2 = \inf\{s \colon P_0\{T \geq s\} \leq \alpha/2\}$ and $2 \min\{P_0\{T \leq t\}, P_0\{T \geq t\}\} = pv$. So the inequality $pv \leq \alpha$ means that $2 \min\{P_0\{T \leq t\}, P_0\{T \geq t\}\} \leq \alpha$.

If $P_0\{T \leq t\} \geq P_0\{T \geq t\}$, then the inequality $pv \leq \alpha$ means that $P_0\{T \geq t\} \leq \alpha/2$. This is equivalent to the inequality $t \geq c_2$, which means that the hypothesis is rejected. Analogously, if $P_0\{T \leq t\} \geq P_0\{T \geq t\}$ then the inequality $pv \leq \alpha$ means that $P_0\{T \leq t\} \leq \alpha/2$. This is equivalent to the inequality $t \leq c_1$, which means that the hypothesis is rejected. So in both cases the inequality $pv \leq \alpha$ means that the hypothesis is rejected.

$\Delta$

If the critical region is defined by the asymptotic distribution of $T$ (usually normal or chi-squared) then the $P$-value $pv_a$ is computed using the asymptotic distribution of $T$, and it is called the asymptotic $P$-value.
Sometimes the $P$-value $pv$ is interpreted as random because each value $t$ of $T$ defines a specific value of $pv$. In the case of the alternatives considered above the $P$-values are the realizations of the following random variables:

$$1 - F_T(T^-), \quad F_T(T) \quad \text{and} \quad 2 \min\{F_T(T), 1 - F_T(T^-)\}$$

1.5. Continuity correction

If the distribution of the statistic $T$ is discrete and the asymptotic distribution of this statistic is absolutely continuous (usually normal) then for medium-sized samples the approximation of distribution $T$ can be improved using the continuity correction [YAT 34].

The idea of a continuity correction is explained by the following example.

**Example 1.1.** Let us consider the parametric hypothesis: $H: p = 0.5$ and the alternative $H_1: p > 0.5$; here $p$ is the Bernoulli distribution parameter. For example, suppose that during $n = 20$ Bernoulli trials the number of successes is $T = 13$. It is evident that the hypothesis $H_0$ is rejected if the statistic $T$ takes large values, i.e. if $T \geq c$ for a given $c$. Under $H_0$, the statistic $T$ has the binomial distribution $B(20; 0.5)$. The exact $P$-value is

$$pv = P\{T \geq 13\} = \sum_{i=13}^{20} C_{20}^i (1/2)^{20} = I_{1/2}(13, 8) = 0.131588$$

Using the normal approximation

$$Z_n = (T - 0.5n)/\sqrt{0.25n} = (T - 10)/\sqrt{5} \sim N(0, 1)$$

we obtain the asymptotic $P$-value

$$pv_a = P\{T \geq 13\} = P\left\{\frac{T - 10}{\sqrt{5}} \geq \frac{13 - 10}{\sqrt{5}}\right\} \approx$$