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 *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