A.V. KRYANEV, G.V. LUKIN

THE METRIC ANALYSIS
AND DATA PROCESSING

The Contents

Abstract

The book first of all is intended for people who use mathematical methods of data processing at the solution of the various maintenance applied problems.

Along with known and used mathematical methods of processing of uncertain data in the book new effective algorithms of numerical realization of these methods, and also new mathematical methods of data processing and the algorithms of their numerical realization stated only in journal articles or in our books [1,2], published while only in Russian are represented. For example, the basic maintenance of chapters 2, 4, 8, 9, 11, 12, and also completely the maintenance of chapters 13-15 while is published only in journal articles.

Besides, the book maintenance is of interest for students, post-graduate students and teachers of the universities, wishing to familiarize with modern mathematical methods of data processing, including forecasting methods.

In chapter 1 base elements of mathematical statistics are presented. The basic properties of estimations and a parity between them, in particular efficiency and robust estimations are analyzed. The traditional material including estimation interval a mathematical expectation and a dispersion of normally distributed random variable is stated. Two universal methods within the limits of the parametrical statistics are considered: a method of the moments and a maximum likelihood method (MLM). The method of the moments is stated in general allowing essentially to expand a framework of its application. By means of Fisher-Kramer-Rao inequality efficiency of estimations is analyzed. The basic properties of MLM-estimations are presented. The statement of theoretical results is accompanied by examples by which consideration likelihood functions are designed. Fisher's information and Fisher's information matrix, MLM-estimations and their characteristics are found. The special attention is given to multidimensional normal distribution.

In 2nd chapter methods of the account of the additional aprioristic information within the limits of the parametrical statistics are considered. Four methods of the account of the aprioristic information depending on its kind are presented. If the aprioristic information on required parameters has stochastic character it is offered to use two methods: Bayes's method and the generalized maximum likelihood method (GMLM) with the task of aprioristic sample. If the aprioristic information on required parameters u is set in the form of an accessory to aprioristic u set it is offered R_a to use two methods — minimax and GMLM taking into account a parity, $u \in R_a$ both from which use the concepts of metric character connected, first of all, with use of distance in space of a required vector u. Schemes of application of methods of the account of the aprioristic information and algorithms of their numerical realization are analyzed. Examples of methods application, in particular a case when members of initial sample submit to the normal law, are considered.

In 3rd chapter robust methods for estimation of position parameter in conditions of the outliers presence are presented. It is underlined that the scheme of the position parameter robust estimation can be generalized without essential qualitative changes on a multidimensional variant of the regressive model parameters estimation. For a basis of robust estimation minimax Huber's approach is taken. The iterative method of a numerical finding robust estimations of the position parameter based on geometrical position of required parameter robust estimation is offered. The general scheme of the robust M-estimations reception based on use of influence function is considered.

In chapter 4 some methods of function and distribution density restoration often used for the solution of applied problems nonparametric statistics are presented. The problem of distribution density restoration on sample unlike a similar problem for distribution function is underlined which belongs to a class of incorrectly tasks in view and consequently can be effectively solved only by using the additional aprioristic information on required density. The form and volume of the aprioristic information can be different and due to it various methods of density distribution restoration are used. In the chapter 5 methods of the density distribution restoration using various kinds and levels of the aprioristic information are described. In methods of histograms Rozenblatt - Parzen and a root - estimation of density the aprioristic information is

used for definition of suitable values of the factors similar to regularized parameter. In projective methods the aprioristic information can be used in the form of the aprioristic reference density definition, in regularized method of histograms the aprioristic information on required density is set in the form of an aprioristic class of density.

In chapter 5 the summary of schemes of hypotheses control about the restored of distribution law within the limits of parametrical and nonparametric statistics is given. Kolmogorov's fitting criterions $\omega 2$, $\chi 2$ are presented.

Chapter 6 is devoted numerical methods of statistical modelling. The means of random variables modelling, in particular transducers of regular distributed random variable γ, based on Lemer's and Neumann's methods are presented in this chapter. Application of a method of statistical modelling for calculation of certain integrals and the solution of integrated Fredholm's equations of the second kind are given.

In chapter 7 the method of the least squares (MLS) for linear models with uncertain data is stated. Classical MLS - scheme and its generalizations are presented. The most significant properties of MLS – estimations and their generalizations are given. In the end of the chapter the linear forecasting model using MLS at a stage of its training is stated.

In chapter 8 robust schemes for linear models with uncertain data are presented. All robust schemes estimation for linear models are constructed on the basis of influence functions and M- estimation. The special attention is given to Huber's M-estimation. Iterative numerical schemes for finding nonlinear robust estimations of linear models parameters, in particular MLS - iterative method are offered. For finding Huber's robust estimations the effective iterative procedure converging to robust estimation for limited number of iterations is offered.

In chapter 9 schemes of the account of the additional aprioristic information in linear models with uncertain data are stated. It is stated that under conditions when Fisher's information matrix of initial linear model is singular or is approximate to singular, the estimation parameters problem of linear model belong to a class of incorrectly problems in view and without the additional aprioristic information, it is impossible to receive comprehensible estimations on accuracy of required parameters. In this chapter various schemes of the account of the aprioristic information based on Bayes's method, minimax - method and GMLM- method are presented. For GMLM two estimation schemes depending on a kind of the aprioristic information are given. At application of first scheme GMLM it is possible to consider the aprioristic information of stochastic character set in the form of aprioristic sample. In this case it is considered that the joint probabilities density of sample members depend on a required vector u linear model. At application of second scheme GMLM the aprioristic information of the determined kind is considered, given in the form of aprioristic determined set to which obviously poses a required vector u. In § 9.5 it is considered also a regularized least squares method allowing to consider an error in elements of matrix A of initial linear model $Au = f - \varepsilon$.

In chapter 10 in the compressed form the method of the least squares for nonlinear models is stated. Newton-Gauss iterative method of a numerical finding of MLS – estimations of nonlinear models solutions is resulted. Levenberg - Markvardt's regularized modifications of Newton-Gauss's iterative processes are given.

Chapters 11, 12 are devoted the analysis and forecasting of time series. In chapter 11 methods of allocation of determined components from the time series are presented. All methods presented in the chapters are based on representation determined components in the form of decomposition on system of basic functions and an estimation of decomposition factors by

means of MLS or by robust schemes. As system of basic functions are undertaken: 1) system of polynomials, orthogonal on set of the fixed values of argument; 2) linear splines; 3) cubic splines; 4) wavelets. Effective numerical schemes of required factors calculation of determined components are offered. In chapter 12 some rather new methods of time processes forecasting are presented. The first two methods are based on the account of aprioristic expert estimations of predicted values and application of schemes of determined component allocation stated in chapter 11. The third of the forecasting methods presented in chapter 12 is based on the singular - spectral analysis and allocation of the main component of an investigated time series. The considerable part of chapter 12 is devoted to the non-stationary singular-spectral analysis.

In chapter 13 bases of a new direction of data processing - the metric analysis are considered, allowing to solve problems of interpolation, extrapolation and forecasting of functions of one and many variables. The concept of a matrix of metric uncertainty with which help are defined interpolated and predicted values of investigated functional dependence is entered. Within the limits of computing schemes of the metric analysis it is possible to consider in the optimum image of uncertainty both metric, and stochastic characters, receiving effective forecasting values. Chapter 13 maintenance corresponds to the maintenance of work of authors [51].

Chapter 14 is devoted to determined chaos intensively developed in last years. In the chapter examples of processes with the determined chaos widely quoted in the scientific literature devoted to determined chaos are resulted. The most significant known properties are resulted in present time of the determined chaotic processes.

In chapter 15 short introduction to planning of optimum measurements is given at restoration of functional dependences. The considered problem of planning is especially important in carrying out of physical experiments when the experimenter has possibility of a choice of argument values at which measurements of restored function are made. Various types of an optimality of experiment plans are introduced and methods of the optimization problems solutions of measurements plans for considered types of an optimality including algorithms of their numerical solution are stated.

1.   Kryanev (Крянев) A.V., Lukin G.V. *Mathematical methods of uncertain data processing.* Moscow, Science Publishers, 2006 (second edition, supplemented, in Russian).
2.   Kryanev (Крянев) A.V., Lukin G.V. *Metric analysis and data processing.* Moscow, Science Publishers, 2010 (in Russian).

The Sigma metric analysis process leads naturally to a quality control (QC) design scheme using quantitative and graphic tools to determine the necessary quality control procedures for routine monitoring of methods and instruments. Adopting Six Sigma as the Goal for Laboratory Testing. Six Sigma is a widely-accepted quality management system, perhaps best known outside of healthcare as the product of innovation at General Electric and Motorola.1 Six Sigma is also well known for the colorful titles of its practitioners â€" green belt (part-time Six Sigma worker), black belt (full-time Six Sigma w...Â Even when data on method bias is missing, a modified Sigma metric can be calculated. Data analysis is a process of inspecting, cleansing, transforming and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively. This process generated a data set containing 5,000 designs, including the input values for each design and its score along the six metrics. One approach at this stage would be to filter the dataset by the metric scores and directly select a few high-performing designs for further analysis. But more can be learned about the nature of the problem as a whole. In order to investigate this process and gain a deeper understanding of the design space, we can study how the designs perform along the four performance metrics by plotting them using scatter plots. Metric space analysis. Not just a single