

---

# CONTENTS

---

## PREFACE

xxi

### **Introduction: The Problem of Induction and Statistical Inference**

1

- 0.1 Learning Paradigm in Statistics / 1
- 0.2 Two Approaches to Statistical Inference: Particular (Parametric Inference) and General (Nonparametric Inference) / 2
- 0.3 The Paradigm Created by the Parametric Approach / 4
- 0.4 Shortcoming of the Parametric Paradigm / 5
- 0.5 After the Classical Paradigm / 6
- 0.6 The Renaissance / 7
- 0.7 The Generalization of the Glivenko–Cantelli–Kolmogorov Theory / 8
- 0.8 The Structural Risk Minimization Principle / 10
- 0.9 The Main Principle of Inference from a Small Sample Size / 11
- 0.10 What This Book is About / 13

## **I THEORY OF LEARNING AND GENERALIZATION**

### **1 Two Approaches to the Learning Problem**

19

- 1.1 General Model of Learning from Examples / 19
- 1.2 The Problem of Minimizing the Risk Functional from Empirical Data / 21
- 1.3 The Problem of Pattern Recognition / 24
- 1.4 The Problem of Regression Estimation / 26
- 1.5 Problem of Interpreting Results of Indirect Measuring / 28
- 1.6 The Problem of Density Estimation (the Fisher–Wald Setting) / 30

vii

- 1.7 Induction Principles for Minimizing the Risk Functional on the Basis of Empirical Data / 32
- 1.8 Classical Methods for Solving the Function Estimation Problems / 33
- 1.9 Identification of Stochastic Objects: Estimation of the Densities and Conditional Densities / 35
  - 1.9.1 Problem of Density Estimation. Direct Setting / 35
  - 1.9.2 Problem of Conditional Probability Estimation / 36
  - 1.9.3 Problem of Conditional Density Estimation / 37
- 1.10 The Problem of Solving an Approximately Determined Integral Equation / 38
- 1.11 Glivenko–Cantelli Theorem / 39
  - 1.11.1 Convergence in Probability and Almost Sure Convergence / 40
  - 1.11.2 Glivenko–Cantelli Theorem / 42
  - 1.11.3 Three Important Statistical Laws / 42
- 1.12 Ill-Posed Problems / 44
- 1.13 The Structure of the Learning Theory / 48

## **Appendix to Chapter 1: Methods for Solving Ill-Posed Problems**

51

- A1.1 The Problem of Solving an Operator Equation / 51
- A1.2 Problems Well-Posed in Tikhonov’s Sense / 53
- A1.3 The Regularization Method / 54
  - A1.3.1 Idea of Regularization Method / 54
  - A1.3.2 Main Theorems About the Regularization Method / 55

## **2 Estimation of the Probability Measure and Problem of Learning**

59

- 2.1 Probability Model of a Random Experiment / 59
- 2.2 The Basic Problem of Statistics / 61
  - 2.2.1 The Basic Problems of Probability and Statistics / 61
  - 2.2.2 Uniform Convergence of Probability Measure Estimates / 62
- 2.3 Conditions for the Uniform Convergence of Estimates to the Unknown Probability Measure / 65
  - 2.3.1 Structure of Distribution Function / 65
  - 2.3.2 Estimator that Provides Uniform Convergence / 68
- 2.4 Partial Uniform Convergence and Generalization of Glivenko–Cantelli Theorem / 69

- 2.4.1 Definition of Partial Uniform Convergence / 69
- 2.4.2 Generalization of the Glivenko–Cantelli Problem / 71
- 2.5 Minimizing the Risk Functional Under the Condition of Uniform Convergence of Probability Measure Estimates / 72
- 2.6 Minimizing the Risk Functional Under the Condition of Partial Uniform Convergence of Probability Measure Estimates / 74
- 2.7 Remarks About Modes of Convergence of the Probability Measure Estimates and Statements of the Learning Problem / 77

### **3 Conditions for Consistency of Empirical Risk Minimization Principle**

79

- 3.1 Classical Definition of Consistency / 79
- 3.2 Definition of Strict (Nontrivial) Consistency / 82
  - 3.2.1 Definition of Strict Consistency for the Pattern Recognition and the Regression Estimation Problems / 82
  - 3.2.2 Definition of Strict Consistency for the Density Estimation Problem / 84
- 3.3 Empirical Processes / 85
  - 3.3.1 Remark on the Law of Large Numbers and Its Generalization / 86
- 3.4 The Key Theorem of Learning Theory (Theorem About Equivalence) / 88
- 3.5 Proof of the Key Theorem / 89
- 3.6 Strict Consistency of the Maximum Likelihood Method / 92
- 3.7 Necessary and Sufficient Conditions for Uniform Convergence of Frequencies to Their Probabilities / 93
  - 3.7.1 Three Cases of Uniform Convergence / 93
  - 3.7.2 Conditions of Uniform Convergence in the Simplest Model / 94
  - 3.7.3 Entropy of a Set of Functions / 95
  - 3.7.4 Theorem About Uniform Two-Sided Convergence / 97
- 3.8 Necessary and Sufficient Conditions for Uniform Convergence of Means to Their Expectations for a Set of Real-Valued Bounded Functions / 98
  - 3.8.1 Entropy of a Set of Real-Valued Functions / 98
  - 3.8.2 Theorem About Uniform Two-Sided Convergence / 99
- 3.9 Necessary and Sufficient Conditions for Uniform Convergence of Means to Their Expectations for Sets of Unbounded Functions / 100
  - 3.9.1 Proof of Theorem 3.5 / 101

- 3.10 Kant's Problem of Demarcation and Popper's Theory of Nonfalsifiability / 106
- 3.11 Theorems About Nonfalsifiability / 108
  - 3.11.1 Case of Complete Nonfalsifiability / 108
  - 3.11.2 Theorem About Partial Nonfalsifiability / 109
  - 3.11.3 Theorem About Potential Nonfalsifiability / 110
- 3.12 Conditions for One-Sided Uniform Convergence and Consistency of the Empirical Risk Minimization Principle / 112
- 3.13 Three Milestones in Learning Theory / 118

#### **4 Bounds on the Risk for Indicator Loss Functions 121**

- 4.1 Bounds for the Simplest Model: Pessimistic Case / 122
  - 4.1.1 The Simplest Model / 123
- 4.2 Bounds for the Simplest Model: Optimistic Case / 125
- 4.3 Bounds for the Simplest Model: General Case / 127
- 4.4 The Basic Inequalities: Pessimistic Case / 129
- 4.5 Proof of Theorem 4.1 / 131
  - 4.5.1 The Basic Lemma / 131
  - 4.5.2 Proof of Basic Lemma / 132
  - 4.5.3 The Idea of Proving Theorem 4.1 / 134
  - 4.5.4 Proof of Theorem 4.1 / 135
- 4.6 Basic Inequalities: General Case / 137
- 4.7 Proof of Theorem 4.2 / 139
- 4.8 Main Nonconstructive Bounds / 144
- 4.9 VC Dimension / 145
  - 4.9.1 The Structure of the Growth Function / 145
  - 4.9.2 Constructive Distribution-Free Bounds on Generalization Ability / 148
  - 4.9.3 Solution of Generalized Glivenko–Cantelli Problem / 149
- 4.10 Proof of Theorem 4.3 / 150
- 4.11 Example of the VC Dimension of the Different Sets of Functions / 155
- 4.12 Remarks About the Bounds on the Generalization Ability of Learning Machines / 160
- 4.13 Bound on Deviation of Frequencies in Two Half-Samples / 163

#### **Appendix to Chapter 4: Lower Bounds on the Risk of the ERM Principle**

169

- A4.1 Two Strategies in Statistical Inference / 169
- A4.2 Minimax Loss Strategy for Learning Problems / 171

- A4.3 Upper Bounds on the Maximal Loss for the Empirical Risk Minimization Principle / 173
  - A4.3.1 Optimistic Case / 173
  - A4.3.2 Pessimistic Case / 174
- A4.4 Lower Bound for the Minimax Loss Strategy in the Optimistic Case / 177
- A4.5 Lower Bound for Minimax Loss Strategy for the Pessimistic Case / 179

## **5 Bounds on the Risk for Real-Valued Loss Functions 183**

- 5.1 Bounds for the Simplest Model: Pessimistic Case / 183
- 5.2 Concepts of Capacity for the Sets of Real-Valued Functions / 186
  - 5.2.1 Nonconstructive Bounds on Generalization for Sets of Real-Valued Functions / 186
  - 5.2.2 The Main Idea / 188
  - 5.2.3 Concepts of Capacity for the Set of Real-Valued Functions / 190
- 5.3 Bounds for the General Model: Pessimistic Case / 192
- 5.4 The Basic Inequality / 194
  - 5.4.1 Proof of Theorem 5.2 / 195
- 5.5 Bounds for the General Model: Universal Case / 196
  - 5.5.1 Proof of Theorem 5.3 / 198
- 5.6 Bounds for Uniform Relative Convergence / 200
  - 5.6.1 Proof of Theorem 5.4 for the Case  $p > 2$  / 201
  - 5.6.2 Proof of Theorem 5.4 for the Case  $1 < p \leq 2$  / 204
- 5.7 Prior Information for the Risk Minimization Problem in Sets of Unbounded Loss Functions / 207
- 5.8 Bounds on the Risk for Sets of Unbounded Nonnegative Functions / 210
- 5.9 Sample Selection and the Problem of Outliers / 214
- 5.10 The Main Results of the Theory of Bounds / 216

## **6 The Structural Risk Minimization Principle 219**

- 6.1 The Scheme of the Structural Risk Minimization Induction Principle / 219
  - 6.1.1 Principle of Structural Risk Minimization / 221
- 6.2 Minimum Description Length and Structural Risk Minimization Inductive Principles / 224
  - 6.2.1 The Idea About the Nature of Random Phenomena / 224

- 6.2.2 Minimum Description Length Principle for the Pattern Recognition Problem / 224
- 6.2.3 Bounds for the Minimum Description Length Principle / 226
- 6.2.4 Structural Risk Minimization for the Simplest Model and Minimum Description Length Principle / 227
- 6.2.5 The Shortcoming of the Minimum Description Length Principle / 228
- 6.3 Consistency of the Structural Risk Minimization Principle and Asymptotic Bounds on the Rate of Convergence / 229
  - 6.3.1 Proof of the Theorems / 232
  - 6.3.2 Discussions and Example / 235
- 6.4 Bounds for the Regression Estimation Problem / 237
  - 6.4.1 The Model of Regression Estimation by Series Expansion / 238
  - 6.4.2 Proof of Theorem 6.4 / 241
- 6.5 The Problem of Approximating Functions / 246
  - 6.5.1 Three Theorems of Classical Approximation Theory / 248
  - 6.5.2 Curse of Dimensionality in Approximation Theory / 251
  - 6.5.3 Problem of Approximation in Learning Theory / 252
  - 6.5.4 The VC Dimension in Approximation Theory / 254
- 6.6 Problem of Local Risk Minimization / 257
  - 6.6.1 Local Risk Minimization Model / 259
  - 6.6.2 Bounds for the Local Risk Minimization Estimator / 262
  - 6.6.3 Proofs of the Theorems / 265
  - 6.6.4 Structural Risk Minimization Principle for Local Function Estimation / 268

## **Appendix to Chapter 6: Estimating Functions on the Basis of Indirect Measurements**

271

- A6.1 Problems of Estimating the Results of Indirect Measurements / 271
- A6.2 Theorems on Estimating Functions Using Indirect Measurements / 273
- A6.3 Proofs of the Theorems / 276
  - A6.3.1 Proof of Theorem A6.1 / 276
  - A6.3.2 Proof of Theorem A6.2 / 281
  - A6.3.3 Proof of Theorem A6.3 / 283

**7 Stochastic Ill-Posed Problems****293**

- 7.1 Stochastic Ill-Posed Problems / 293
- 7.2 Regularization Method for Solving Stochastic Ill-Posed Problems / 297
- 7.3 Proofs of the Theorems / 299
  - 7.3.1 Proof of Theorem 7.1 / 299
  - 7.3.2 Proof of Theorem 7.2 / 302
  - 7.3.3 Proof of Theorem 7.3 / 303
- 7.4 Conditions for Consistency of the Methods of Density Estimation / 305
- 7.5 Nonparametric Estimators of Density: Estimators Based on Approximations of the Distribution Function by an Empirical Distribution Function / 308
  - 7.5.1 The Parzen Estimators / 308
  - 7.5.2 Projection Estimators / 313
  - 7.5.3 Spline Estimate of the Density. Approximation by Splines of the Odd Order / 313
  - 7.5.4 Spline Estimate of the Density. Approximation by Splines of the Even Order / 314
- 7.6 Nonclassical Estimators / 315
  - 7.6.1 Estimators for the Distribution Function / 315
  - 7.6.2 Polygon Approximation of Distribution Function / 316
  - 7.6.3 Kernel Density Estimator / 316
  - 7.6.4 Projection Method of the Density Estimator / 318
- 7.7 Asymptotic Rate of Convergence for Smooth Density Functions / 319
- 7.8 Proof of Theorem 7.4 / 322
- 7.9 Choosing a Value of Smoothing (Regularization) Parameter for the Problem of Density Estimation / 327
- 7.10 Estimation of the Ratio of Two Densities / 330
  - 7.10.1 Estimation of Conditional Densities / 333
- 7.11 Estimation of Ratio of Two Densities on the Line / 334
- 7.12 Estimation of a Conditional Probability on a Line / 337

**8 Estimating the Values of Function at Given Points****339**

- 8.1 The Scheme of Minimizing the Overall Risk / 339
- 8.2 The Method of Structural Minimization of the Overall Risk / 343
- 8.3 Bounds on the Uniform Relative Deviation of Frequencies in Two Subsamples / 344
- 8.4 A Bound on the Uniform Relative Deviation of Means in Two Subsamples / 347

- 8.5 Estimation of Values of an Indicator Function in a Class of Linear Decision Rules / 350
- 8.6 Sample Selection for Estimating the Values of an Indicator Function / 355
- 8.7 Estimation of Values of a Real Function in the Class of Functions Linear in Their Parameters / 359
- 8.8 Sample Selection for Estimation of Values of Real-Valued Functions / 362
- 8.9 Local Algorithms for Estimating Values of an Indicator Function / 363
- 8.10 Local Algorithms for Estimating Values of a Real-Valued Function / 365
- 8.11 The Problem of Finding the Best Point in a Given Set / 367
  - 8.11.1 Choice of the Most Probable Representative of the First Class / 368
  - 8.11.2 Choice of the Best Point of a Given Set / 370

## II SUPPORT VECTOR ESTIMATION OF FUNCTIONS

### 9 Perceptrons and Their Generalizations

375

- 9.1 Rosenblatt's Perceptron / 375
- 9.2 Proofs of the Theorems / 380
  - 9.2.1 Proof of Novikoff Theorem / 380
  - 9.2.2 Proof of Theorem 9.3 / 382
- 9.3 Method of Stochastic Approximation and Sigmoid Approximation of Indicator Functions / 383
  - 9.3.1 Method of Stochastic Approximation / 384
  - 9.3.2 Sigmoid Approximations of Indicator Functions / 385
- 9.4 Method of Potential Functions and Radial Basis Functions / 387
  - 9.4.1 Method of Potential Functions in Asymptotic Learning Theory / 388
  - 9.4.2 Radial Basis Function Method / 389
- 9.5 Three Theorems of Optimization Theory / 390
  - 9.5.1 Fermat's Theorem (1629) / 390
  - 9.5.2 Lagrange Multipliers Rule (1788) / 391
  - 9.5.3 Kühn-Tucker Theorem (1951) / 393
- 9.6 Neural Networks / 395
  - 9.6.1 The Back-Propagation Method / 395
  - 9.6.2 The Back-Propagation Algorithm / 398



- 9.6.3 Neural Networks for the Regression Estimation Problem / 399
- 9.6.4 Remarks on the Back-Propagation Method / 399

## **10 The Support Vector Method for Estimating Indicator Functions 401**

- 10.1 The Optimal Hyperplane / 401
- 10.2 The Optimal Hyperplane for Nonseparable Sets / 408
  - 10.2.1 The Hard Margin Generalization of the Optimal Hyperplane / 408
  - 10.2.2 The Basic Solution. Soft Margin Generalization / 411
- 10.3 Statistical Properties of the Optimal Hyperplane / 412
- 10.4 Proofs of the Theorems / 415
  - 10.4.1 Proof of Theorem 10.3 / 415
  - 10.4.2 Proof of Theorem 10.4 / 415
  - 10.4.3 Leave-One-Out Procedure / 416
  - 10.4.4 Proof of Theorem 10.5 and Theorem 9.2 / 417
  - 10.4.5 The Key Lemma / 418
  - 10.4.6 Proof of Theorems 10.6 and 10.6a / 420
  - 10.4.7 Proof of Theorem 10.7 / 421
- 10.5 The Idea of the Support Vector Machine / 421
  - 10.5.1 Generalization in High-Dimensional Space / 422
  - 10.5.2 Hilbert–Schmidt Theory and Mercer Theorem / 423
  - 10.5.3 Constructing SV Machines / 424
- 10.6 One More Approach to the Support Vector Method / 426
  - 10.6.1 Minimizing the Number of Support Vectors / 426
  - 10.6.2 Generalization for the Nonseparable Case / 427
  - 10.6.3 Linear Optimization Method for SV Machines / 427
- 10.7 Selection of SV Machine Using Bounds / 428
- 10.8 Examples of SV Machines for Pattern Recognition / 430
  - 10.8.1 Polynomial Support Vector Machines / 430
  - 10.8.2 Radial Basis Function SV Machines / 431
  - 10.8.3 Two-Layer Neural SV Machines / 432
- 10.9 Support Vector Method for Transductive Inference / 434
- 10.10 Multiclass Classification / 437
- 10.11 Remarks on Generalization of the SV Method / 440

## **11 The Support Vector Method for Estimating Real-Valued Functions 443**

- 11.1  $\varepsilon$ -Insensitive Loss Functions / 443
- 11.2 Loss Functions for Robust Estimators / 445

- 11.3 Minimizing the Risk with  $\varepsilon$ -Insensitive Loss Functions / 448
  - 11.3.1 Minimizing the Risk for a Fixed Element of the Structure / 449
  - 11.3.2 The Basic Solutions / 452
  - 11.3.3 Solution for the Huber Loss Function / 453
- 11.4 SV Machines for Function Estimation / 454
  - 11.4.1 Minimizing the Risk for a Fixed Element of the Structure in Feature Space / 455
  - 11.4.2 The Basic Solutions in Feature Space / 456
  - 11.4.3 Solution for Huber Loss Function in Feature Space / 458
  - 11.4.4 Linear Optimization Method / 459
  - 11.4.5 Multi-Kernel Decomposition of Functions / 459
- 11.5 Constructing Kernels for Estimation of Real-Valued Functions / 460
  - 11.5.1 Kernels Generating Expansion on Polynomials / 461
  - 11.5.2 Constructing Multidimensional Kernels / 462
- 11.6 Kernels Generating Splines / 464
  - 11.6.1 Spline of Order  $d$  with a Finite Number of Knots / 464
  - 11.6.2 Kernels Generating Splines with an Infinite Number of Knots / 465
  - 11.6.3  $B_d$ -Spline Approximations / 466
  - 11.6.4  $B_d$  Splines with an Infinite Number of Knots / 468
- 11.7 Kernels Generating Fourier Expansions / 468
  - 11.7.1 Kernels for Regularized Fourier Expansions / 469
- 11.8 The Support Vector ANOVA Decomposition (SVAD) for Function Approximation and Regression Estimation / 471
- 11.9 SV Method for Solving Linear Operator Equations / 473
  - 11.9.1 The SV Method / 473
  - 11.9.2 Regularization by Choosing Parameters of  $\varepsilon_i$ -Insensitivity / 478
- 11.10 SV Method of Density Estimation / 479
  - 11.10.1 Spline Approximation of a Density / 480
  - 11.10.2 Approximation of a Density with Gaussian Mixture / 481
- 11.11 Estimation of Conditional Probability and Conditional Density Functions / 484
  - 11.11.1 Estimation of Conditional Probability Functions / 484
  - 11.11.2 Estimation of Conditional Density Functions / 488
- 11.12 Connections Between the SV Method and Sparse Function Approximation / 489
  - 11.12.1 Reproducing Kernels Hilbert Spaces / 490
  - 11.12.2 Modified Sparse Approximation and its Relation to SV Machines / 491

**12 SV Machines for Pattern Recognition 493**

- 12.1 The Quadratic Optimization Problem / 493
  - 12.1.1 Iterative Procedure for Specifying Support Vectors / 494
  - 12.1.2 Methods for Solving the Reduced Optimization Problem / 496
- 12.2 Digit Recognition Problem. The U.S. Postal Service Database / 496
  - 12.2.1 Performance for the U.S. Postal Service Database / 496
  - 12.2.2 Some Important Details / 500
  - 12.2.3 Comparison of Performance of the SV Machine with Gaussian Kernel to the Gaussian RBF Network / 503
  - 12.2.4 The Best Results for U.S. Postal Service Database / 505
- 12.3 Tangent Distance / 506
- 12.4 Digit Recognition Problem. The NIST Database / 511
  - 12.4.1 Performance for NIST Database / 511
  - 12.4.2 Further Improvement / 512
  - 12.4.3 The Best Results for NIST Database / 512
- 12.5 Future Racing / 514
  - 12.5.1 One More Opportunity. The Transductive Inference / 518

**13 SV Machines for Function Approximations, Regression Estimation, and Signal Processing 521**

- 13.1 The Model Selection Problem / 521
  - 13.1.1 Functional for Model Selection Based on the VC Bound / 522
  - 13.1.2 Classical Functionals / 524
  - 13.1.3 Experimental Comparison of Model Selection Methods / 525
  - 13.1.4 The Problem of Feature Selection Has No General Solution / 526
- 13.2 Structure on the Set of Regularized Linear Functions / 530
  - 13.2.1 The  $L$ -Curve Method / 532
  - 13.2.2 The Method of Effective Number of Parameters / 534
  - 13.2.3 The Method of Effective VC Dimension / 536
  - 13.2.4 Experiments on Measuring the Effective VC Dimension / 540
- 13.3 Function Approximation Using the SV Method / 543

- 13.3.1 Why Does the Value of  $\varepsilon$  Control the Number of Support Vectors? / 546
- 13.4 SV Machine for Regression Estimation / 549
  - 13.4.1 Problem of Data Smoothing / 549
  - 13.4.2 Estimation of Linear Regression Functions / 550
  - 13.4.3 Estimation of Nonlinear Regression Functions / 556
- 13.5 SV Method for Solving the Positron Emission Tomography (PET) Problem / 558
  - 13.5.1 Description of PET / 558
  - 13.5.2 Problem of Solving the Radon Equation / 560
  - 13.5.3 Generalization of the Residual Principle of Solving PET Problems / 561
  - 13.5.4 The Classical Methods of Solving the PET Problem / 562
  - 13.5.5 The SV Method for Solving the PET Problem / 563
- 13.6 Remark About the SV Method / 567

### III STATISTICAL FOUNDATION OF LEARNING THEORY

#### **14 Necessary and Sufficient Conditions for Uniform Convergence of Frequencies to Their Probabilities 571**

- 14.1 Uniform Convergence of Frequencies to their Probabilities / 572
- 14.2 Basic Lemma / 573
- 14.3 Entropy of the Set of Events / 576
- 14.4 Asymptotic Properties of the Entropy / 578
- 14.5 Necessary and Sufficient Conditions of Uniform Convergence. Proof of Sufficiency / 584
- 14.6 Necessary and Sufficient Conditions of Uniform Convergence. Proof of Necessity / 587
- 14.7 Necessary and Sufficient Conditions. Continuation of Proving Necessity / 592

#### **15 Necessary and Sufficient Conditions for Uniform Convergence of Means to Their Expectations 597**

- 15.1  $\varepsilon$  Entropy / 597
  - 15.1.1 Proof of the Existence of the Limit / 600
  - 15.1.2 Proof of the Convergence of the Sequence / 601
- 15.2 The Quasicube / 603
- 15.3  $\varepsilon$ -Extension of a Set / 608

- 15.4 An Auxiliary Lemma / 610
- 15.5 Necessary and Sufficient Conditions for Uniform Convergence.  
The Proof of Necessity / 614
- 15.6 Necessary and Sufficient Conditions for Uniform Convergence.  
The Proof of Sufficiency / 618
- 15.7 Corollaries from Theorem 15.1 / 624

**16 Necessary and Sufficient Conditions for Uniform One-Sided  
Convergence of Means to Their Expectations **629****

- 16.1 Introduction / 629
- 16.2 Maximum Volume Sections / 630
- 16.3 The Theorem on the Average Logarithm / 636
- 16.4 Theorem on the Existence of a Corridor / 642
- 16.5 Theorem on the Existence of Functions Close to the Corridor  
Boundaries (Theorem on Potential Nonfalsifiability) / 651
- 16.6 The Necessary Conditions / 660
- 16.7 The Necessary and Sufficient Conditions / 666

**Comments and Bibliographical Remarks **681****

**References **723****

**Index **733****