

Luc Devroye  
Gábor Lugosi

# Combinatorial Methods in Density Estimation



Springer

# Contents

**Preface** vii

**1. Introduction** 1

◦ §1.1. References 3

**2. Concentration Inequalities** 4

§2.1. Hoeffding's Inequality 4

§2.2. An Inequality for the Expected Maximal Deviation 7

§2.3. The Bounded Difference Inequality 7

§2.4. Examples 9

§2.5. Bibliographic Remarks 10

§2.6. Exercises 11

§2.7. References 13

**3. Uniform Deviation Inequalities** 17

§3.1. The Vapnik–Chervonenkis Inequality 17

§3.2. Covering Numbers and Chaining 19

§3.3. Example: The Dvoretzky–Kiefer–Wolfowitz Theorem 22

§3.4. Bibliographic Remarks 23

§3.5. Exercises 23

§3.6. References 25

**4. Combinatorial Tools** 27

§4.1. Shatter Coefficients 27

§4.2. Vapnik–Chervonenkis Dimension and Shatter Coefficients 28

§4.3. Vapnik–Chervonenkis Dimension and Covering Numbers 30

§4.4. Examples 31

§4.5. Bibliographic Remarks 33

§4.6. Exercises 33

§4.7. References 35

**5. Total Variation** 38

§5.1. Density Estimation 38

§5.2. The Total Variation 39

§5.3. Invariance 39

§5.4. Mappings 40

§5.5. Convolutions 41

§5.6. Normalization 41

§5.7. The Lebesgue Density Theorem 42

- §5.8. LeCam's Inequality 43
  - §5.9. Bibliographic Remarks 43
  - §5.10. Exercises 43
  - §5.11. References 46
- 6. Choosing a Density Estimate 47**
- §6.1. Choosing Between Two Densities 47
  - §6.2. Examples 49
  - §6.3. Is the Factor of Three Necessary? 51
  - §6.4. Maximum Likelihood Does not Work 52
  - §6.5.  $L_2$  Distances Are To Be Avoided 52
  - §6.6. Selection from  $k$  Densities 53
  - §6.7. Examples Continued 55
  - §6.8. Selection from an Infinite Class 55
  - §6.9. Bibliographic Remarks 56
  - §6.10. Exercises 56
  - §6.11. References 57
- 7. Skeleton Estimates 58**
- §7.1. Kolmogorov Entropy 58
  - §7.2. Skeleton Estimates 58
  - §7.3. Robustness 60
  - §7.4. Finite Mixtures 60
  - §7.5. Monotone Densities on the Hypercube 61
  - §7.6. How To Make Gigantic Totally Bounded Classes 64
  - §7.7. Bibliographic Remarks 66
  - §7.8. Exercises 66
  - §7.9. References 68
- 8. The Minimum Distance Estimate: Examples 70**
- §8.1. Problem Formulation 70
  - §8.2. Series Estimates 71
  - §8.3. Parametric Estimates: Exponential Families 72
  - §8.4. Neural Network Estimates 73
  - §8.5. Mixture Classes, Radial Basis Function Networks 74
  - §8.6. Bibliographic Remarks 76
  - §8.7. Exercises 76
  - §8.8. References 77
- 9. The Kernel Density Estimate 79**
- §9.1. Approximating Functions by Convolutions 79
  - §9.2. Definition of the Kernel Estimate 80
  - §9.3. Consistency of the Kernel Estimate 81
  - §9.4. Concentration 82
  - §9.5. Choosing the Bandwidth 83
  - §9.6. Choosing the Kernel 84
  - §9.7. Rates of Convergence 85

- §9.8. Uniform Rate of Convergence 86
- §9.9. Shrinkage, and the Combination of Density Estimates 88
- §9.10. Bibliographic Remarks 90
- §9.11. Exercises 90
- §9.12. References 95
- 10. Additive Estimates and Data Splitting 98**
  - §10.1. Data Splitting 98
  - §10.2. Additive Estimates 99
  - §10.3. Histogram Estimates 103
  - §10.4. Bibliographic Remarks 105
  - §10.5. Exercises 105
  - §10.6. References 107
- 11. Bandwidth Selection for Kernel Estimates 108**
  - §11.1. The Kernel Estimate with Riemann Kernel 108
  - §11.2. General Kernels, Kernel Complexity 110
  - §11.3. Kernel Complexity: Univariate Examples 111
  - §11.4. Kernel Complexity: Multivariate Kernels 113
  - §11.5. Asymptotic Optimality 114
  - §11.6. Bibliographic Remarks 115
  - §11.7. Exercises 115
  - §11.8. References 116
- 12. Multiparameter Kernel Estimates 118**
  - §12.1. Multivariate Kernel Estimates—Product Kernels 118
  - §12.2. Multivariate Kernel Estimates—Ellipsoidal Kernels 121
  - §12.3. Variable Kernel Estimates 122
  - §12.4. Tree-Structured Partitions 124
  - §12.5. Changepoints and Bump Hunting 125
  - §12.6. Bibliographic Remarks 127
  - §12.7. Exercises 127
  - §12.8. References 132
- 13. Wavelet Estimates 134**
  - §13.1. Definitions 134
  - §13.2. Smoothing 135
  - §13.3. Thresholding 136
  - §13.4. Soft Thresholding 138
  - §13.5. Bibliographic Remarks 139
  - §13.6. Exercises 139
  - §13.7. References 140
- 14. The Transformed Kernel Estimate 142**
  - §14.1. The Transformed Kernel Estimate 142
  - §14.2. Box-Cox Transformations 143
  - §14.3. Piecewise Linear Transformations 146
  - §14.4. Bibliographic Remarks 148

§14.5. Exercises	148
§14.6. References	149
<b>15. Minimax Theory</b>	<b>150</b>
§15.1. Estimating a Density from One Data Point	150
§15.2. The General Minimax Problem	152
§15.3. Rich Classes	154
§15.4. Assouad's Lemma	156
§15.5. Example: The Class of Convex Densities	159
§15.6. Additional Examples	162
§15.7. Tuning the Parameters of Variable Kernel Estimates	163
§15.8. Sufficient Statistics	166
§15.9. Bibliographic Remarks	168
§15.10. Exercises	169
§15.11. References	174
<b>16. Choosing the Kernel Order</b>	<b>177</b>
§16.1. Introduction	177
§16.2. Standard Kernel Estimate: Riemann Kernels	179
§16.3. Standard Kernel Estimates: General Kernels	181
§16.4. An Infinite Family of Kernels	184
§16.5. Bibliographic Remarks	187
§16.6. Exercises	188
§16.7. References	188
<b>17. Bandwidth Choice with Superkernels</b>	<b>190</b>
§17.1. Superkernels	190
§17.2. The Trapezoidal Kernel	192
§17.3. Bandwidth Selection	193
§17.4. Bibliographic Remarks	194
§17.5. Exercises	194
§17.6. References	196
<b>Author Index</b>	<b>199</b>
<b>Subject Index</b>	<b>203</b>