

Contents

Preface	xi
1 Definitions and rough detection of tail heaviness	1
1.1 Definitions and basic properties of classes of heavy-tailed distributions	1
1.2 Tail index estimation	6
1.2.1 Estimators of a positive-valued tail index	6
1.2.2 The choice of k in Hill's estimator	8
1.2.3 Estimators of a real-valued tail index	13
1.2.4 On-line estimation of the tail index	17
1.3 Detection of tail heaviness and dependence	27
1.3.1 Rough tests of tail heaviness	27
1.3.2 Analysis of Web traffic and TCP flow data	30
1.3.3 Dependence detection from univariate data	42
1.3.4 Dependence detection from bivariate data	49
1.3.5 Bivariate analysis of TCP flow data	51
1.4 Notes and comments	56
1.5 Exercises	57
2 Classical methods of probability density estimation	61
2.1 Principles of density estimation	61
2.2 Methods of density estimation	70
2.2.1 Kernel estimators	70
2.2.2 Projection estimators	74
2.2.3 Spline estimators	76
2.2.4 Smoothing methods	76
2.2.5 Illustrative examples	83
2.3 Kernel estimation from dependent data	85
2.3.1 Statement of the problem	86
2.3.2 Numerical calculation of the bandwidth	89
2.3.3 Data-driven selection of the bandwidth	91
2.4 Applications	91
2.4.1 Finance: evaluation of market risk	91

2.4.2	Telecommunications	93
2.4.3	Population analysis	94
2.5	Exercises	95
3	Heavy-tailed density estimation	99
3.1	Problems of the estimation of heavy-tailed densities	100
3.2	Combined parametric–nonparametric method	101
3.2.1	Nonparametric estimation of the density by structural risk minimization	103
3.2.2	Illustrative examples	107
3.2.3	Web data analysis by a combined parametric–nonparametric method	109
3.3	Barron’s estimator and χ^2 -optimality	111
3.4	Kernel estimators with variable bandwidth	113
3.5	Retransformed nonparametric estimators	117
3.6	Exercises	119
4	Transformations and heavy-tailed density estimation	123
4.1	Problems of data transformations	123
4.2	Estimates based on a fixed transformation	124
4.3	Estimates based on an adaptive transformation	128
4.3.1	Estimation algorithm	128
4.3.2	Analysis of the algorithm	129
4.3.3	Further remarks	133
4.4	Estimating the accuracy of retransformed estimates	135
4.5	Boundary kernels	136
4.6	Accuracy of a nonvariable bandwidth kernel estimator	139
4.7	The D method for a nonvariable bandwidth kernel estimator	141
4.8	The D method for a variable bandwidth kernel estimator	142
4.8.1	Method and results	142
4.8.2	Application to Web traffic characteristics	144
4.9	The ω^2 method for the projection estimator	147
4.10	Exercises	149
5	Classification and retransformed density estimates	151
5.1	Classification and quality of density estimation	151
5.2	Convergence of the estimated probability of misclassification ..	154
5.3	Simulation study	155
5.4	Application of the classification technique to Web data analysis	160
5.4.1	Intelligent browser	160
5.4.2	Web data analysis by traffic classification	161
5.4.3	Web prefetching	161
5.5	Exercises	161

6	Estimation of high quantiles	163
6.1	Introduction	163
6.2	Estimators of high quantiles	164
6.3	Distribution of high quantile estimates	167
6.4	Simulation study	169
6.4.1	Comparison of high quantile estimates in terms of relative bias and mean squared error	169
6.4.2	Comparison of high quantile estimates in terms of confidence intervals	170
6.5	Application to Web traffic data	175
6.6	Exercises	176
7	Nonparametric estimation of the hazard rate function	179
7.1	Definition of the hazard rate function	180
7.2	Statistical regularization method	182
7.3	Numerical solution of ill-posed problems	185
7.4	Estimation of the hazard rate function of heavy-tailed distributions	187
7.5	Hazard rate estimation for compactly supported distributions	188
7.5.1	Estimation of the hazard rate from the simplest equations	188
7.5.2	Estimation of the hazard rate from a special kernel equation	193
7.6	Estimation of the ratio of hazard rates	197
7.6.1	Failure time detection	199
7.6.2	Hormesis detection	200
7.7	Hazard rate estimation in teletraffic theory	207
7.7.1	Teletraffic processes at the packet level	207
7.7.2	Estimation of the intensity of a nonhomogeneous Poisson process	208
7.8	Semi-Markov modeling in teletraffic engineering	210
7.8.1	The Gilbert–Elliott model	210
7.8.2	Estimation of a retrial process	212
7.9	Exercises	217
8	Nonparametric estimation of the renewal function	219
8.1	Traffic modeling by recurrent marked point processes	220
8.2	Introduction to renewal function estimation	221
8.3	Histogram-type estimator of the renewal function	224
8.4	Convergence of the histogram-type estimator	225
8.5	Selection of k by a bootstrap method	228
8.6	Selection of k by a plot	232
8.7	Simulation study	234
8.8	Application to the inter-arrival times of TCP connections	245

8.9 Conclusions and discussion 247
8.10 Exercises 248

Appendices

A Proofs of Chapter 2 251
B Proofs of Chapter 4 253
C Proofs of Chapter 5 267
D Proofs of Chapter 6 271
E Proofs of Chapter 7 275
F Proofs of Chapter 8 285
List of Main Symbols and Abbreviations 291
References 295
Index 307