

# Probability for Statistics and Machine Learning:

Fundamentals and Advanced Topics

Anirban DasGupta

# Contents

|          |                                                                       |           |
|----------|-----------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Review of Univariate Probability</b>                               | <b>10</b> |
| 1.1      | Experiments and Sample Spaces . . . . .                               | 10        |
| 1.2      | Conditional Probability and Independence . . . . .                    | 13        |
| 1.3      | Integer Valued and Discrete Random Variables . . . . .                | 15        |
| 1.3.1    | CDF and Independence . . . . .                                        | 16        |
| 1.3.2    | Expectation and Moments . . . . .                                     | 19        |
| 1.4      | Inequalities . . . . .                                                | 24        |
| 1.5      | Generating and Moment Generating Functions . . . . .                  | 27        |
| 1.6      | * Applications of Generating Functions to a Pattern Problem . . . . . | 31        |
| 1.7      | Standard Discrete Distributions . . . . .                             | 32        |
| 1.8      | Poisson Approximation to Binomial . . . . .                           | 37        |
| 1.9      | Continuous Random Variables . . . . .                                 | 39        |
| 1.10     | Functions of a Continuous Random Variable . . . . .                   | 44        |
| 1.10.1   | Expectation and Moments . . . . .                                     | 47        |
| 1.10.2   | Moments and the Tail of a CDF . . . . .                               | 50        |
| 1.11     | Moment Generating Function and Fundamental Inequalities . . . . .     | 51        |
| 1.11.1   | * Inversion of an MGF and Post's Formula . . . . .                    | 53        |
| 1.12     | Some Special Continuous Distributions . . . . .                       | 53        |
| 1.13     | Normal Distribution and Confidence Interval for a Mean . . . . .      | 60        |
| 1.14     | Stein's Lemma . . . . .                                               | 64        |
| 1.15     | * Chernoff's Variance Inequality . . . . .                            | 66        |
| 1.16     | * Various Characterizations of Normal Distributions . . . . .         | 67        |
| 1.17     | Normal Approximations and Central Limit Theorem . . . . .             | 68        |
| 1.17.1   | Binomial Confidence Interval . . . . .                                | 71        |
| 1.17.2   | Error of the CLT . . . . .                                            | 72        |
| 1.18     | Normal Approximation to Poisson and Gamma . . . . .                   | 75        |
| 1.18.1   | Confidence Intervals . . . . .                                        | 76        |
| 1.19     | * Convergence of Densities and Edgeworth Expansions . . . . .         | 77        |
| 1.20     | Exercises . . . . .                                                   | 78        |
| 1.21     | References . . . . .                                                  | 86        |
| <b>2</b> | <b>Multivariate Discrete Distributions</b>                            | <b>88</b> |
| 2.1      | Bivariate Joint Distributions and Expectations of Functions . . . . . | 88        |
| 2.2      | Conditional Distributions and Conditional Expectations . . . . .      | 92        |
| 2.2.1    | Examples on Conditional Distributions and Expectations . . . . .      | 93        |
| 2.3      | Using Conditioning to Evaluate Mean and Variance . . . . .            | 96        |
| 2.4      | Covariance and Correlation . . . . .                                  | 98        |
| 2.5      | Multivariate Case . . . . .                                           | 102       |
| 2.5.1    | Joint MGF . . . . .                                                   | 102       |
| 2.5.2    | Multinomial Distribution . . . . .                                    | 104       |
| 2.6      | * The Poissonization Technique . . . . .                              | 106       |

|          |                                                                          |            |
|----------|--------------------------------------------------------------------------|------------|
| 2.7      | Exercises . . . . .                                                      | 107        |
| <b>3</b> | <b>Multidimensional Densities</b>                                        | <b>111</b> |
| 3.1      | Joint Density Function and Its Role . . . . .                            | 111        |
| 3.2      | Expectation of Functions . . . . .                                       | 118        |
| 3.3      | Bivariate Normal . . . . .                                               | 121        |
| 3.4      | Conditional Densities and Expectations . . . . .                         | 125        |
| 3.4.1    | Examples on Conditional Densities and Expectations . . . . .             | 127        |
| 3.5      | Posterior Densities, Likelihood Functions, and Bayes Estimates . . . . . | 131        |
| 3.6      | Bivariate Normal Conditional Distributions . . . . .                     | 135        |
| 3.7      | * Useful Formulas and Characterizations for Bivariate Normal . . . . .   | 136        |
| 3.7.1    | Computing Bivariate Normal Probabilities . . . . .                       | 137        |
| 3.8      | * Conditional Expectation Given a Set and Borel's Paradox . . . . .      | 138        |
| 3.9      | Exercises . . . . .                                                      | 140        |
| 3.10     | References . . . . .                                                     | 143        |
| <b>4</b> | <b>Advanced Distribution Theory</b>                                      | <b>144</b> |
| 4.1      | Convolutions and Examples . . . . .                                      | 144        |
| 4.2      | Products and Quotients and the $t$ and $F$ Distribution . . . . .        | 148        |
| 4.3      | Transformations . . . . .                                                | 152        |
| 4.4      | Applications of Jacobian Formula . . . . .                               | 153        |
| 4.5      | Polar Coordinates in Two Dimensions . . . . .                            | 154        |
| 4.6      | * $n$ -Dimensional Polar and Helmert's Transformation . . . . .          | 156        |
| 4.6.1    | Efficient Spherical Calculations with Polar Coordinates . . . . .        | 156        |
| 4.6.2    | Independence of Mean and Variance in Normal Case . . . . .               | 159        |
| 4.6.3    | The $t$ Confidence Interval . . . . .                                    | 160        |
| 4.7      | The Dirichlet Distribution . . . . .                                     | 161        |
| 4.7.1    | * Picking a Point from the Surface of a Sphere . . . . .                 | 163        |
| 4.7.2    | * Poincaré's Lemma . . . . .                                             | 163        |
| 4.8      | * Ten Important High Dimensional Formulas for Easy Reference . . . . .   | 163        |
| 4.9      | Exercises . . . . .                                                      | 165        |
| 4.10     | References . . . . .                                                     | 168        |
| <b>5</b> | <b>Multivariate Normal and Related Distributions</b>                     | <b>169</b> |
| 5.1      | Definition and Some Basic Properties . . . . .                           | 169        |
| 5.2      | Conditional Distributions . . . . .                                      | 171        |
| 5.3      | Exchangeable Normal Variables . . . . .                                  | 174        |
| 5.4      | Sampling Distributions Useful in Statistics . . . . .                    | 175        |
| 5.4.1    | * Wishart Expectation Identities . . . . .                               | 176        |
| 5.4.2    | * Hotelling's $T^2$ and Distribution of Quadratic Forms . . . . .        | 178        |
| 5.4.3    | * Distribution of Correlation Coefficient . . . . .                      | 179        |
| 5.5      | * Noncentral Distributions . . . . .                                     | 180        |
| 5.6      | * Some Important Inequalities for Easy Reference . . . . .               | 181        |
| 5.7      | Exercises . . . . .                                                      | 183        |

|          |                                                                    |            |
|----------|--------------------------------------------------------------------|------------|
| 5.8      | References . . . . .                                               | 185        |
| <b>6</b> | <b>Finite Sample Theory of Order Statistics and Extremes</b>       | <b>186</b> |
| 6.1      | Basic Distribution Theory . . . . .                                | 186        |
| 6.2      | More Advanced Distribution Theory . . . . .                        | 189        |
| 6.3      | Quantile Transformation and Existence of Moments . . . . .         | 192        |
| 6.4      | Spacings . . . . .                                                 | 196        |
| 6.4.1    | Exponential Spacings and Rényi's Representation . . . . .          | 196        |
| 6.4.2    | Uniform Spacings . . . . .                                         | 198        |
| 6.5      | Conditional Distributions and Markov Property . . . . .            | 198        |
| 6.6      | Some Applications . . . . .                                        | 200        |
| 6.6.1    | * Records . . . . .                                                | 201        |
| 6.6.2    | The Empirical CDF . . . . .                                        | 202        |
| 6.7      | * Distribution of the Multinomial Maximum . . . . .                | 205        |
| 6.8      | Exercises . . . . .                                                | 206        |
| 6.9      | References . . . . .                                               | 209        |
| <b>7</b> | <b>Essential Asymptotics and Applications</b>                      | <b>210</b> |
| 7.1      | Some Basic Notation and Convergence Concepts . . . . .             | 210        |
| 7.2      | Laws of Large Numbers . . . . .                                    | 214        |
| 7.3      | Convergence Preservation . . . . .                                 | 218        |
| 7.4      | Convergence in Distribution . . . . .                              | 220        |
| 7.5      | Preservation of Convergence and Statistical Applications . . . . . | 225        |
| 7.5.1    | Slutsky's Theorem . . . . .                                        | 225        |
| 7.5.2    | Delta Theorem . . . . .                                            | 226        |
| 7.5.3    | Variance Stabilizing Transformations . . . . .                     | 229        |
| 7.6      | Convergence of Moments . . . . .                                   | 231        |
| 7.6.1    | Uniform Integrability . . . . .                                    | 231        |
| 7.6.2    | The Moment Problem and Convergence in Distribution . . . . .       | 233        |
| 7.6.3    | Approximation of Moments . . . . .                                 | 233        |
| 7.7      | Convergence of Densities and Scheffé's Theorem . . . . .           | 236        |
| 7.8      | Exercises . . . . .                                                | 240        |
| 7.9      | References . . . . .                                               | 245        |
| <b>8</b> | <b>Characteristic Functions and Applications</b>                   | <b>247</b> |
| 8.1      | Characteristic Functions of Standard Distributions . . . . .       | 248        |
| 8.2      | Inversion and Uniqueness . . . . .                                 | 251        |
| 8.3      | Taylor Expansions, Differentiability, and Moments . . . . .        | 255        |
| 8.4      | Continuity Theorems . . . . .                                      | 256        |
| 8.5      | Proof of the CLT and the WLLN . . . . .                            | 257        |
| 8.6      | * Producing Characteristic Functions . . . . .                     | 258        |
| 8.7      | Error of the Central Limit Theorem . . . . .                       | 260        |
| 8.8      | Lindeberg-Feller Theorem for General Independent Case . . . . .    | 262        |
| 8.9      | * Infinite Divisibility and Stable Laws . . . . .                  | 265        |

|           |                                                           |            |
|-----------|-----------------------------------------------------------|------------|
| 8.10      | * Some Useful Inequalities . . . . .                      | 267        |
| 8.11      | Exercises . . . . .                                       | 269        |
| 8.12      | References . . . . .                                      | 272        |
| <b>9</b>  | <b>Asymptotics of Extremes and Order Statistics</b>       | <b>273</b> |
| 9.1       | Central Order Statistics . . . . .                        | 273        |
| 9.1.1     | Single Order Statistic . . . . .                          | 273        |
| 9.1.2     | Two Statistical Applications . . . . .                    | 274        |
| 9.1.3     | Several Order Statistics . . . . .                        | 275        |
| 9.2       | Extremes . . . . .                                        | 277        |
| 9.2.1     | Easily Applicable Limit Theorems . . . . .                | 277        |
| 9.2.2     | The Convergence of Types Theorem . . . . .                | 280        |
| 9.3       | * Fisher-Tippett Family and Putting it Together . . . . . | 281        |
| 9.4       | Exercises . . . . .                                       | 282        |
| 9.5       | References . . . . .                                      | 285        |
| <b>10</b> | <b>Markov Chains and Applications</b>                     | <b>286</b> |
| 10.1      | Notation and Basic Definitions . . . . .                  | 286        |
| 10.2      | Examples and Various Applications as a Model . . . . .    | 287        |
| 10.3      | Chapman-Kolmogorov Equation . . . . .                     | 291        |
| 10.4      | Communicating Classes . . . . .                           | 295        |
| 10.5      | Gambler's Ruin . . . . .                                  | 297        |
| 10.6      | First Passage, Recurrence and Transience . . . . .        | 298        |
| 10.7      | Long Run Evolution and Stationary Distributions . . . . . | 303        |
| 10.8      | Exercises . . . . .                                       | 309        |
| 10.9      | References . . . . .                                      | 317        |
| <b>11</b> | <b>Random Walks</b>                                       | <b>318</b> |
| 11.1      | Random Walk on the Cubic Lattice . . . . .                | 318        |
| 11.1.1    | Some Distribution Theory . . . . .                        | 320        |
| 11.1.2    | Recurrence and Transience . . . . .                       | 322        |
| 11.1.3    | * Pólya's Formula for the Return Probability . . . . .    | 324        |
| 11.2      | First Passage Time and Arc Sine Law . . . . .             | 325        |
| 11.3      | * The Local Time . . . . .                                | 328        |
| 11.4      | Practically Useful Generalizations . . . . .              | 330        |
| 11.5      | Wald's Identity . . . . .                                 | 331        |
| 11.6      | * Fate of a Random Walk . . . . .                         | 333        |
| 11.7      | Chung-Fuchs Theorem . . . . .                             | 334        |
| 11.8      | * Six Important Inequalities . . . . .                    | 336        |
| 11.9      | Exercises . . . . .                                       | 337        |
| 11.10     | References . . . . .                                      | 340        |

|           |                                                                      |            |
|-----------|----------------------------------------------------------------------|------------|
| <b>12</b> | <b>Brownian Motion and Gaussian Processes</b>                        | <b>341</b> |
| 12.1      | Preview of Connections to the Random Walk . . . . .                  | 341        |
| 12.2      | Basic Definitions . . . . .                                          | 343        |
| 12.2.1    | Condition for a Gaussian Process to be Markov . . . . .              | 346        |
| 12.2.2    | * Explicit Construction of Brownian Motion . . . . .                 | 346        |
| 12.3      | Basic Distributional Properties . . . . .                            | 347        |
| 12.3.1    | Reflection Principle and Extremes . . . . .                          | 348        |
| 12.3.2    | Path Properties and Behavior Near Zero and Infinity . . . . .        | 350        |
| 12.3.3    | * Fractal Nature of Level Sets . . . . .                             | 352        |
| 12.4      | The Dirichlet Problem and Boundary Crossing Probabilities . . . . .  | 353        |
| 12.4.1    | Recurrence and Transience . . . . .                                  | 355        |
| 12.5      | * The Local Time of Brownian Motion . . . . .                        | 356        |
| 12.6      | Invariance Principle and Statistical Applications . . . . .          | 357        |
| 12.7      | Strong Invariance Principle and the KMT Theorem . . . . .            | 361        |
| 12.8      | Brownian Motion with Drift and Ornstein-Uhlenbeck Process . . . . .  | 363        |
| 12.8.1    | Negative Drift and Density of Maximum . . . . .                      | 363        |
| 12.8.2    | * Transition Density and the Heat Equation . . . . .                 | 364        |
| 12.8.3    | * The Ornstein-Uhlenbeck Process . . . . .                           | 365        |
| 12.9      | Exercises . . . . .                                                  | 366        |
| 12.10     | References . . . . .                                                 | 370        |
| <b>13</b> | <b>Poisson Processes and Applications</b>                            | <b>372</b> |
| 13.1      | Notation . . . . .                                                   | 372        |
| 13.2      | Defining a Homogeneous Poisson Process . . . . .                     | 373        |
| 13.3      | Important Properties and Uses as a Statistical Model . . . . .       | 374        |
| 13.4      | * Linear Poisson Process and Brownian Motion: A Connection . . . . . | 381        |
| 13.5      | Higher Dimensional Poisson Point Processes . . . . .                 | 382        |
| 13.5.1    | The Mapping Theorem . . . . .                                        | 384        |
| 13.6      | One Dimensional Nonhomogeneous Processes . . . . .                   | 385        |
| 13.7      | * Campbell's Theorem and Shot Noise . . . . .                        | 387        |
| 13.7.1    | Poisson process and Stable Laws . . . . .                            | 389        |
| 13.8      | Exercises . . . . .                                                  | 390        |
| 13.9      | References . . . . .                                                 | 393        |
| <b>14</b> | <b>Discrete Time Martingales and Concentration Inequalities</b>      | <b>395</b> |
| 14.1      | Illustrative Examples and Applications in Statistics . . . . .       | 395        |
| 14.2      | Stopping Times and Optional Stopping . . . . .                       | 399        |
| 14.2.1    | Stopping Times . . . . .                                             | 399        |
| 14.2.2    | Optional Stopping . . . . .                                          | 401        |
| 14.2.3    | Sufficient Conditions for Optional Stopping Theorem . . . . .        | 402        |
| 14.2.4    | Applications of Optional Stopping . . . . .                          | 404        |
| 14.3      | Martingale and Concentration Inequalities . . . . .                  | 406        |
| 14.3.1    | Maximal Inequality . . . . .                                         | 407        |

|           |                                                                  |            |
|-----------|------------------------------------------------------------------|------------|
| 14.3.2    | * Inequalities of Burkholder, Davis, and Gundy . . . . .         | 409        |
| 14.3.3    | Inequalities of Hoeffding and Azuma . . . . .                    | 411        |
| 14.3.4    | * Inequalities of McDiarmid and Devroye . . . . .                | 413        |
| 14.3.5    | The Upcrossing Inequality . . . . .                              | 415        |
| 14.4      | Convergence of Martingales . . . . .                             | 417        |
| 14.4.1    | The Basic Convergence Theorem . . . . .                          | 417        |
| 14.4.2    | Convergence in $L_1$ and $L_2$ . . . . .                         | 419        |
| 14.5      | * Reverse Martingales . . . . .                                  | 420        |
| 14.6      | Martingale Central Limit Theorem . . . . .                       | 422        |
| 14.7      | Exercises . . . . .                                              | 424        |
| 14.8      | References . . . . .                                             | 428        |
| <b>15</b> | <b>Probability Metrics</b>                                       | <b>430</b> |
| 15.1      | Standard Probability Metrics Useful in Statistics . . . . .      | 430        |
| 15.2      | Basic Properties of the Metrics . . . . .                        | 432        |
| 15.3      | Metric Inequalities . . . . .                                    | 438        |
| 15.4      | * Differential Metrics for Parametric Families . . . . .         | 442        |
| 15.4.1    | * Fisher Information and Differential Metrics . . . . .          | 442        |
| 15.4.2    | * Rao's Geodesic Distances on Distributions . . . . .            | 444        |
| 15.5      | Exercises . . . . .                                              | 445        |
| 15.6      | References . . . . .                                             | 447        |
| <b>16</b> | <b>Empirical Processes and VC Theory</b>                         | <b>448</b> |
| 16.1      | Basic Notation and Definitions . . . . .                         | 448        |
| 16.2      | Classic Asymptotic Properties of the Empirical Process . . . . . | 450        |
| 16.2.1    | Invariance Principle and Statistical Applications . . . . .      | 451        |
| 16.2.2    | * Weighted Empirical Process . . . . .                           | 453        |
| 16.2.3    | The Quantile Process . . . . .                                   | 455        |
| 16.2.4    | Strong Approximations of the Empirical Process . . . . .         | 455        |
| 16.3      | Vapnik-Chervonenkis Theory . . . . .                             | 456        |
| 16.3.1    | Basic Theory . . . . .                                           | 457        |
| 16.3.2    | Concrete Examples . . . . .                                      | 458        |
| 16.4      | CLTs for Empirical Measures and Applications . . . . .           | 461        |
| 16.4.1    | Notation and Formulation . . . . .                               | 461        |
| 16.4.2    | Entropy Bounds and Specific CLTs . . . . .                       | 462        |
| 16.4.3    | Concrete Examples . . . . .                                      | 464        |
| 16.5      | Maximal Inequalities and Symmetrization . . . . .                | 465        |
| 16.6      | * Connection to the Poisson Process . . . . .                    | 468        |
| 16.7      | Exercises . . . . .                                              | 469        |
| 16.8      | References . . . . .                                             | 473        |

|                                                                      |            |
|----------------------------------------------------------------------|------------|
| <b>17 Large Deviations</b>                                           | <b>475</b> |
| 17.1 Large Deviations for Sample Means . . . . .                     | 475        |
| 17.1.1 The Cramér-Chernoff Theorem in $\mathcal{R}$ . . . . .        | 476        |
| 17.1.2 Properties of the Rate Function . . . . .                     | 479        |
| 17.1.3 Cramér’s Theorem for General Sets . . . . .                   | 480        |
| 17.2 The Gärtner-Ellis Theorem . . . . .                             | 481        |
| 17.3 The $t$ -statistic . . . . .                                    | 483        |
| 17.4 Lipschitz Functions and Talagrand’s Inequality . . . . .        | 485        |
| 17.5 * Large Deviations in Continuous Time . . . . .                 | 487        |
| 17.5.1 * Continuity of a Gaussian Process . . . . .                  | 488        |
| 17.5.2 * Metric Entropy of $T$ and Tail of the Supremum . . . . .    | 490        |
| 17.6 Exercises . . . . .                                             | 491        |
| 17.7 References . . . . .                                            | 493        |
| <b>18 The Exponential Family and Statistical Applications</b>        | <b>495</b> |
| 18.1 One Parameter Exponential Family . . . . .                      | 495        |
| 18.1.1 Definition and First Examples . . . . .                       | 495        |
| 18.2 The Canonical Form and Basic Properties . . . . .               | 499        |
| 18.2.1 Convexity Properties . . . . .                                | 501        |
| 18.2.2 Moments and Moment Generating Function . . . . .              | 501        |
| 18.2.3 Closure Properties . . . . .                                  | 504        |
| 18.3 Multiparameter Exponential Family . . . . .                     | 505        |
| 18.4 * Sufficiency and Completeness . . . . .                        | 508        |
| 18.5 * Neyman-Fisher Factorization and Basu’s Theorem . . . . .      | 510        |
| 18.5.1 * Applications of Basu’s Theorem to Probability . . . . .     | 511        |
| 18.6 Curved Exponential Family . . . . .                             | 514        |
| 18.7 Exercises . . . . .                                             | 515        |
| 18.8 References . . . . .                                            | 518        |
| <b>19 Simulation and Markov Chain Monte Carlo</b>                    | <b>519</b> |
| 19.1 The Ordinary Monte Carlo . . . . .                              | 520        |
| 19.1.1 Basic Theory and Examples . . . . .                           | 521        |
| 19.1.2 Monte Carlo $P$ -values . . . . .                             | 526        |
| 19.1.3 Rao-Blackwellization . . . . .                                | 527        |
| 19.2 Textbook Simulation Techniques . . . . .                        | 528        |
| 19.2.1 Quantile Transformation and Accept-Reject . . . . .           | 528        |
| 19.2.2 Importance Sampling and its Asymptotic Properties . . . . .   | 532        |
| 19.2.3 Optimal Importance Sampling Distribution . . . . .            | 535        |
| 19.2.4 Algorithms for Simulating from Common Distributions . . . . . | 536        |
| 19.3 Markov Chain Monte Carlo . . . . .                              | 538        |
| 19.3.1 Reversible Markov Chains . . . . .                            | 541        |
| 19.3.2 Metropolis Algorithms . . . . .                               | 544        |
| 19.4 The Gibbs Sampler . . . . .                                     | 546        |

|           |                                                                  |            |
|-----------|------------------------------------------------------------------|------------|
| 19.5      | Convergence of MCMC and Bounds on Errors . . . . .               | 551        |
| 19.5.1    | Spectral Bounds . . . . .                                        | 553        |
| 19.5.2    | Dobrushin’s Inequality and Diaconis-Fill-Stroock Bound . . . . . | 556        |
| 19.5.3    | Drift and Minorization Methods . . . . .                         | 558        |
| 19.6      | MCMC on General Spaces . . . . .                                 | 560        |
| 19.6.1    | General Theory and Metropolis Schemes . . . . .                  | 560        |
| 19.6.2    | Convergence . . . . .                                            | 563        |
| 19.6.3    | Convergence of the Gibbs Sampler . . . . .                       | 567        |
| 19.7      | Practical Convergence Diagnostics . . . . .                      | 569        |
| 19.8      | Exercises . . . . .                                              | 571        |
| 19.9      | References . . . . .                                             | 579        |
| <b>20</b> | <b>Useful Tools for Statistics and Machine Learning</b>          | <b>582</b> |
| 20.1      | The Bootstrap . . . . .                                          | 582        |
| 20.1.1    | Consistency of the Bootstrap . . . . .                           | 584        |
| 20.1.2    | Further Examples . . . . .                                       | 587        |
| 20.1.3    | Higher Order Accuracy of the Bootstrap . . . . .                 | 590        |
| 20.1.4    | Bootstrap for Dependent Data . . . . .                           | 592        |
| 20.2      | The EM Algorithm . . . . .                                       | 594        |
| 20.2.1    | The Algorithm and Examples . . . . .                             | 595        |
| 20.2.2    | Monotone Ascent and Convergence of EM . . . . .                  | 600        |
| 20.2.3    | Modifications of EM . . . . .                                    | 602        |
| 20.3      | Kernels and Classification . . . . .                             | 603        |
| 20.3.1    | Smoothing by Kernels . . . . .                                   | 603        |
| 20.3.2    | Some Common Kernels in Use . . . . .                             | 605        |
| 20.3.3    | Kernels for Statistical Classification . . . . .                 | 606        |
| 20.3.4    | Reproducing Kernel Hilbert Spaces . . . . .                      | 609        |
| 20.3.5    | Mercer’s Theorem and Feature Maps . . . . .                      | 614        |
| 20.3.6    | Support Vector Machines . . . . .                                | 617        |
| 20.4      | Exercises . . . . .                                              | 619        |
| 20.5      | References . . . . .                                             | 625        |

## Suggested Courses with Different Themes

| Duration | Theme                                        | Chapters                  |
|----------|----------------------------------------------|---------------------------|
| 15 weeks | Beginning Graduate                           | 2-7, 9                    |
| 15 weeks | Advanced Graduate                            | 7, 8, 10, 11, 12, 13, 14  |
| 15 weeks | Special topics for Statistics students       | 9, 10, 15, 16, 17, 18, 20 |
| 15 weeks | Special topics for Computer science students | 4, 11, 14, 16, 17, 18, 19 |
| 8 weeks  | Summer course for Statistics students        | 11, 12, 14, 20            |
| 8 weeks  | Summer course for Computer science students  | 14, 16, 18, 20            |
| 8 weeks  | Summer course on Modelling and Simulation    | 4, 10, 13, 19             |