



## CONTENTS

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| PREFACE                                                             | v   |
| List of examples                                                    | 1   |
| 1. INTRODUCTION                                                     | 3   |
| 1.1 Outline                                                         | 3   |
| 1.2 Language                                                        | 3   |
| 1.3 Two philosophies                                                | 4   |
| 1.4 Notation                                                        | 6   |
| 2. BASIC CONCEPTS IN PROBABILITY                                    | 9   |
| 2.1 Definitions of probability                                      | 9   |
| 2.1.1 <i>Frequency limit definition</i>                             | 9   |
| 2.1.2 <i>Modern definition</i>                                      | 9   |
| 2.2 Properties of probability                                       | 10  |
| 2.2.1 <i>Addition law for sets of elementary events</i>             | 10  |
| 2.2.2 <i>Conditional probability and independence</i>               | 11  |
| 2.2.3 <i>Example of the addition law: scanning efficiency</i>       | 11  |
| 2.2.4 <i>Bayes theorem for discrete events</i>                      | 12  |
| 2.2.5 <i>Bayesian use of Bayes theorem</i>                          | 13  |
| 2.2.6 <i>Random variable</i>                                        | 14  |
| 2.3 Continuous random variables                                     | 14  |
| 2.3.1 <i>Probability density function</i>                           | 15  |
| 2.3.2 <i>Change of variable</i>                                     | 17  |
| 2.3.3 <i>Cumulative, marginal and conditional distributions</i>     | 17  |
| 2.3.4 <i>Bayes theorem for continuous variables</i>                 | 19  |
| 2.3.5 <i>Bayesian use of Bayes theorem for continuous variables</i> | 19  |
| 2.4 Properties of distributions                                     | 20  |
| 2.4.1 <i>Expectation, mean and variance</i>                         | 20  |
| 2.4.2 <i>Covariance and correlation</i>                             | 21  |
| 2.4.3 <i>Linear functions of random variables</i>                   | 23  |
| 2.4.4 <i>Ratio of random variables</i>                              | 25  |
| 2.4.5 <i>Approximate variance formulae</i>                          | 26  |
| 2.4.6 <i>Moments</i>                                                | 27  |
| 2.5 Characteristic function                                         | 28  |
| 2.5.1 <i>Definition and properties</i>                              | 29  |
| 2.5.2 <i>Cumulants</i>                                              | 32  |
| 2.5.3 <i>Probability generating function</i>                        | 32  |
| 2.5.4 <i>Sums of a random number of random variables</i>            | 33  |
|                                                                     | vii |

|        |                                                         |    |
|--------|---------------------------------------------------------|----|
| 3.     | CONVERGENCE AND THE LAW OF LARGE NUMBERS                | 36 |
| 3.1    | The Tchebycheff theorem and its corollary               | 36 |
| 3.1.1  | <i>Tchebycheff theorem</i>                              | 36 |
| 3.1.2  | <i>Bienaymé-Tchebycheff inequality</i>                  | 36 |
| 3.2    | Convergence                                             | 38 |
| 3.2.1  | <i>Convergence in distribution</i>                      | 38 |
| 3.2.2  | <i>The Paul Levy theorem</i>                            | 38 |
| 3.2.3  | <i>Convergence in probability</i>                       | 38 |
| 3.2.4  | <i>Stronger types of convergence</i>                    | 39 |
| 3.3    | The law of large numbers                                | 39 |
| 3.3.1  | <i>Monte Carlo integration</i>                          | 40 |
| 3.3.2  | <i>The Central Limit theorem</i>                        | 40 |
| 3.3.3  | <i>Example: Gaussian random number generator</i>        | 42 |
| 4.     | PROBABILITY DISTRIBUTIONS                               | 44 |
| 4.1    | Discrete distributions                                  | 44 |
| 4.1.1  | <i>Binomial distribution</i>                            | 44 |
| 4.1.2  | <i>Multinomial distribution</i>                         | 46 |
| 4.1.3  | <i>Poisson distribution.</i>                            | 48 |
| 4.1.4  | <i>Compound Poisson distribution</i>                    | 52 |
| 4.1.5  | <i>Geometric distribution</i>                           | 54 |
| 4.1.6  | <i>Negative binomial distribution</i>                   | 56 |
| 4.2    | Continuous distributions                                | 58 |
| 4.2.1  | <i>Normal one-dimensional (univariate normal)</i>       | 58 |
| 4.2.2  | <i>Normal many-dimensional (multivariate normal)</i>    | 62 |
| 4.2.3  | <i>Chi-square distribution</i>                          | 64 |
| 4.2.4  | <i>Student's t-distribution</i>                         | 67 |
| 4.2.5  | <i>Fisher-Snedecor F and Z distributions</i>            | 70 |
| 4.2.6  | <i>Uniform distribution</i>                             | 72 |
| 4.2.7  | <i>Triangular distribution</i>                          | 72 |
| 4.2.8  | <i>Beta distribution</i>                                | 73 |
| 4.2.9  | <i>Exponential distribution</i>                         | 74 |
| 4.2.10 | <i>Gamma distribution</i>                               | 76 |
| 4.2.11 | <i>Cauchy, or Breit-Wigner, distribution</i>            | 77 |
| 4.2.12 | <i>Log-normal distribution</i>                          | 79 |
| 4.2.13 | <i>Extreme value distribution</i>                       | 81 |
| 4.2.14 | <i>Weibull distribution</i>                             | 83 |
| 4.2.15 | <i>Asymptotic relationships between distributions</i>   | 84 |
| 4.3    | Handling of real life distributions                     | 86 |
| 4.3.1  | <i>General applicability of the normal distribution</i> | 86 |
| 4.3.2  | <i>Johnson empirical distributions</i>                  | 87 |
| 4.3.3  | <i>Truncation</i>                                       | 88 |
| 4.3.4  | <i>Experimental resolution</i>                          | 89 |
| 4.3.5  | <i>Examples of variable experimental resolution</i>     | 90 |

|                                                                        |     |
|------------------------------------------------------------------------|-----|
| 5. INFORMATION                                                         | 93  |
| 5.1 Basic concepts in information theory                               | 93  |
| 5.1.1 <i>Likelihood function</i>                                       | 93  |
| 5.1.2 <i>Statistic</i>                                                 | 94  |
| 5.2 Information of R.A. Fisher                                         | 94  |
| 5.2.1 <i>Definition of information</i>                                 | 94  |
| 5.2.2 <i>Properties of information</i>                                 | 94  |
| 5.3 Sufficient statistics                                              | 95  |
| 5.3.1 <i>Sufficiency</i>                                               | 95  |
| 5.3.2 <i>Examples</i>                                                  | 96  |
| 5.3.3 <i>Minimal sufficient statistics</i>                             | 97  |
| 5.3.4 <i>Darmois theorem</i>                                           | 98  |
| 5.4 Information and sufficiency                                        | 100 |
| 5.5 Example of experimental design                                     | 101 |
| 6. DECISION THEORY                                                     | 103 |
| 6.1 Basic concepts in decision theory                                  | 103 |
| 6.1.1 <i>Subjective probability, Bayesian approach</i>                 | 103 |
| 6.1.2 <i>Definitions and terminology</i>                               | 104 |
| 6.2 Choice of decision rules                                           | 105 |
| 6.2.1 <i>Classical choice: pre-ordering rules</i>                      | 105 |
| 6.2.2 <i>Bayesian choice</i>                                           | 106 |
| 6.2.3 <i>Minimax decisions</i>                                         | 107 |
| 6.3 Usual decision problems in statistics                              | 107 |
| 6.3.1 <i>Point estimation</i>                                          | 108 |
| 6.3.2 <i>Interval estimation</i>                                       | 108 |
| 6.3.3 <i>Tests of hypothesis</i>                                       | 109 |
| 6.4 Examples: adjustment of an apparatus                               | 111 |
| 6.4.1 <i>Adjustment given an estimate of the apparatus performance</i> | 111 |
| 6.4.2 <i>Adjustment with estimation of the optimum adjustment</i>      | 113 |
| 6.5 Conclusion: indeterminacy in classical and Bayesian decisions      | 114 |
| 7. THEORY OF ESTIMATORS                                                | 115 |
| 7.1 Desirable properties of estimators                                 | 115 |
| 7.1.1 <i>Consistency and convergence</i>                               | 115 |
| 7.1.2 <i>Bias and consistency</i>                                      | 116 |
| 7.2 Usual methods of constructing consistent estimators                | 117 |
| 7.2.1 <i>The moments method</i>                                        | 118 |
| 7.2.2 <i>Implicitly defined estimators</i>                             | 119 |
| 7.2.3 <i>The maximum likelihood method</i>                             | 121 |
| 7.2.4 <i>The least squares methods</i>                                 | 121 |

|       |                                                                                            |     |
|-------|--------------------------------------------------------------------------------------------|-----|
| 7.3   | Asymptotic distributions of estimates                                                      | 124 |
| 7.3.1 | <i>Asymptotic normality</i>                                                                | 124 |
| 7.3.2 | <i>Asymptotic expansion of moments of estimates</i>                                        | 126 |
| 7.3.3 | <i>Asymptotic bias and variance of the usual estimators</i>                                | 128 |
| 7.4   | Information and the precision of an estimator                                              | 130 |
| 7.4.1 | <i>Lower bounds for the variance -- Cramér-Rao inequality</i>                              | 131 |
| 7.4.2 | <i>Efficiency and minimum variance</i>                                                     | 132 |
| 7.4.3 | <i>Cramér-Rao inequality for several parameters</i>                                        | 134 |
| 7.4.4 | <i>The Gauss-Markov theorem</i>                                                            | 135 |
| 7.4.5 | <i>Asymptotic efficiency</i>                                                               | 136 |
| 7.5   | Bayesian inference                                                                         | 136 |
| 7.5.1 | <i>Choice of prior density</i>                                                             | 136 |
| 7.5.2 | <i>Inference about the mean when the variance is known</i>                                 | 138 |
| 7.5.3 | <i>Inference about the variance when the mean is known</i>                                 | 139 |
| 7.5.4 | <i>Inference about the mean and the variance</i>                                           | 141 |
| 7.5.5 | <i>Summary</i>                                                                             | 142 |
| 8.    | POINT ESTIMATION IN PRACTICE                                                               | 143 |
| 8.1   | Choice of estimator                                                                        | 143 |
| 8.1.1 | <i>Desirable properties of estimators</i>                                                  | 143 |
| 8.1.2 | <i>Compromise between statistical merits</i>                                               | 144 |
| 8.1.3 | <i>Cures to obtain simplicity</i>                                                          | 145 |
| 8.1.4 | <i>Economic considerations</i>                                                             | 147 |
| 8.1.5 | <i>Summary of properties of estimators</i>                                                 | 148 |
| 8.2   | The method of moments                                                                      | 152 |
| 8.2.1 | <i>Orthogonal functions</i>                                                                | 152 |
| 8.2.2 | <i>Comparison of likelihood and moment methods</i>                                         | 154 |
| 8.3   | The maximum likelihood method                                                              | 155 |
| 8.3.1 | <i>Summary of properties</i>                                                               | 155 |
| 8.3.2 | <i>Example: determination of the lifetime of a strange particle in a restricted volume</i> | 157 |
| 8.3.3 | <i>Academic example of a poor maximum likelihood estimate</i>                              | 158 |
| 8.3.4 | <i>Constrained parameters</i>                                                              | 159 |
| 8.4   | The least squares method (Chi-square)                                                      | 163 |
| 8.4.1 | <i>The linear model</i>                                                                    | 163 |
| 8.4.2 | <i>The polynomial model</i>                                                                | 165 |
| 8.4.3 | <i>Constrained parameters in the linear model</i>                                          | 166 |
| 8.4.4 | <i>Normally distributed data in nonlinear models</i>                                       | 169 |
| 8.4.5 | <i>Estimation from histograms; comparison of likelihood and least squares methods</i>      | 17  |
| 8.5   | Weights and detection efficiency                                                           | 172 |
| 8.5.1 | <i>Ideal method - maximum likelihood</i>                                                   | 172 |
| 8.5.2 | <i>Approximate method</i>                                                                  | 174 |
| 8.5.3 | <i>Exclusion of events with large weight</i>                                               | 176 |
| 8.5.4 | <i>Least squares method</i>                                                                | 178 |

|        |                                                                               |     |
|--------|-------------------------------------------------------------------------------|-----|
| 8.6    | Reduction of bias                                                             | 181 |
| 8.6.1  | <i>Exact distribution of the estimate known</i>                               | 181 |
| 8.6.2  | <i>Exact distribution of the estimate unknown</i>                             | 182 |
| 8.7    | Robust (distribution-free) estimation                                         | 184 |
| 8.7.1  | <i>Robust estimation of the centre of a distribution</i>                      | 184 |
| 8.7.2  | <i>Trimming and Winsorization</i>                                             | 186 |
| 8.7.3  | <i>Generalized <math>p^{th}</math>-power norms</i>                            | 187 |
| 8.7.4  | <i>Estimates of location for asymmetric distributions</i>                     | 188 |
| 9.     | INTERVAL ESTIMATION                                                           | 190 |
| 9.1    | Normally distributed data                                                     | 191 |
| 9.1.1  | <i>Confidence intervals for the mean</i>                                      | 191 |
| 9.1.2  | <i>Confidence intervals for several parameters</i>                            | 192 |
| 9.1.3  | <i>Interpretation of the covariance matrix</i>                                | 198 |
| 9.2    | The general case in one dimension                                             | 200 |
| 9.2.1  | <i>Confidence intervals and belts</i>                                         | 200 |
| 9.2.2  | <i>Confidence bounds (upper and lower limits)</i>                             | 201 |
| 9.3    | Use of the likelihood function                                                | 203 |
| 9.3.1  | <i>Parabolic log-likelihood function</i>                                      | 203 |
| 9.3.2  | <i>Non-parabolic likelihood functions</i>                                     | 204 |
| 9.4    | Use of asymptotic approximations                                              | 206 |
| 9.4.1  | <i>Asymptotic normality of the maximum likelihood estimate</i>                | 206 |
| 9.4.2  | <i>Asymptotic normality of <math>\partial \ln L / \partial \theta</math></i>  | 206 |
| 9.4.3  | <i>Many-dimensional confidence regions</i>                                    | 207 |
| 9.5    | Finite sample behaviour of three general methods of interval estimation       | 208 |
| 9.6    | Bayesian approach                                                             | 213 |
| 10.    | TESTS OF HYPOTHESES                                                           | 215 |
| 10.1   | Formulation of a test                                                         | 215 |
| 10.1.1 | <i>Basic concepts in testing</i>                                              | 215 |
| 10.1.2 | <i>Example: separation of two classes of events</i>                           | 216 |
| 10.2   | Comparison of tests                                                           | 219 |
| 10.2.1 | <i>Power</i>                                                                  | 219 |
| 10.2.2 | <i>Consistency</i>                                                            | 220 |
| 10.2.3 | <i>Bias</i>                                                                   | 220 |
| 10.2.4 | <i>Distribution-free tests</i>                                                | 221 |
| 10.2.5 | <i>Choice of tests</i>                                                        | 222 |
| 10.3   | Tests of simple hypotheses                                                    | 224 |
| 10.3.1 | <i>The Neyman-Pearson test</i>                                                | 224 |
| 10.3.2 | <i>Example: normal theory test versus sign test</i>                           | 225 |
| 10.4   | Tests of composite hypotheses                                                 | 227 |
| 10.4.1 | <i>Existence of a Uniformly Most Powerful test for the exponential family</i> | 227 |

|        |                                                                                |     |
|--------|--------------------------------------------------------------------------------|-----|
| 10.4.2 | <i>One- and two-sided tests</i>                                                | 228 |
| 10.4.3 | <i>Maximizing local power</i>                                                  | 229 |
| 10.5   | <b>Likelihood ratio test</b>                                                   | 230 |
| 10.5.1 | <i>Test statistic</i>                                                          | 230 |
| 10.5.2 | <i>Asymptotic distribution for continuous families of hypotheses</i>           | 230 |
| 10.5.3 | <i>Asymptotic power for continuous families of hypotheses</i>                  | 233 |
| 10.5.4 | <i>Examples</i>                                                                | 233 |
| 10.5.5 | <i>Small sample behaviour</i>                                                  | 236 |
| 10.5.6 | <i>Example of separate families of hypotheses</i>                              | 239 |
| 10.5.7 | <i>General methods for testing separate families</i>                           | 242 |
| 10.6   | <b>Tests and decision theory</b>                                               | 245 |
| 10.6.1 | <i>Bayesian choice between families of distributions</i>                       | 245 |
| 10.6.2 | <i>Sequential tests for optimum number of observations</i>                     | 249 |
| 10.6.3 | <i>Sequential probability ratio test for a continuous family of hypotheses</i> | 253 |
| 10.7   | <b>Summary of optimal tests</b>                                                | 254 |
| 11.    | <b>GOODNESS-OF-FIT TESTS</b>                                                   | 255 |
| 11.1   | <b>Likelihood ratio test</b>                                                   | 256 |
| 11.2   | <b>Pearson's <math>\chi^2</math> test</b>                                      | 257 |
| 11.2.1 | <i>Moments of the Pearson statistic</i>                                        | 258 |
| 11.2.2 | <i><math>\chi^2</math> test with estimation of parameters</i>                  | 258 |
| 11.2.3 | <i>Choosing optimal bin size</i>                                               | 259 |
| 11.3   | <b>Other tests on binned data</b>                                              | 263 |
| 11.3.1 | <i>Run test</i>                                                                | 263 |
| 11.3.2 | <i>Empty cell test, order statistics</i>                                       | 263 |
| 11.3.3 | <i>Neyman-Barton "smooth" test</i>                                             | 265 |
| 11.4   | <b>Tests free of binning</b>                                                   | 268 |
| 11.4.1 | <i>Smirnov-Cramér-von Mises test</i>                                           | 268 |
| 11.4.2 | <i>Kolmogorov test</i>                                                         | 269 |
| 11.4.3 | <i>Use of likelihood function</i>                                              | 271 |
| 11.5   | <b>Applications</b>                                                            | 272 |
| 11.5.1 | <i>Detection of fine structure</i>                                             | 272 |
| 11.5.2 | <i>Combining independent estimates</i>                                         | 275 |
| 11.5.3 | <i>Comparing distributions</i>                                                 | 279 |
| 11.6   | <b>Combining independent tests</b>                                             | 282 |
| 11.6.1 | <i>Independence of tests</i>                                                   | 282 |
| 11.6.2 | <i>Significance level of the combined test</i>                                 | 282 |
| 11.7   | <b>Orthogonal polynomials (Appendix)</b>                                       | 285 |
| 11.7.1 | <i>Definitions</i>                                                             | 285 |
| 11.7.2 | <i>Use of orthogonal functions</i>                                             | 286 |
|        | <b>BIBLIOGRAPHY</b>                                                            | 288 |
|        | <b>SUBJECT INDEX</b>                                                           | 292 |

