

CONTENTS

Preface to the Second Edition	v
Preface to the First Edition	vii
Chapter 1. Introduction	1
1.1. Outline	1
1.2. Language	2
1.3. Two Philosophies	3
1.4. Notation	4
Chapter 2. Basic Concepts in Probability	9
2.1. Definitions of Probability	9
2.1.1. Mathematical probability	10
2.1.2. Frequentist probability	10
2.1.3. Bayesian probability	11
2.2. Properties of Probability	12
2.2.1. Addition law for sets of elementary events	12
2.2.2. Conditional probability and independence	13
2.2.3. Example of the addition law: scanning efficiency	14
2.2.4. Bayes theorem for discrete events	15
2.2.5. Bayesian use of Bayes theorem	16
2.2.6. Random variable	17
2.3. Continuous Random Variables	18
2.3.1. Probability density function	19
2.3.2. Change of variable	20

2.3.3. Cumulative, marginal and conditional distributions	21
2.3.4. Bayes theorem for continuous variables	22
2.3.5. Bayesian use of Bayes theorem for continuous variables . . .	22
2.4. Properties of Distributions	24
2.4.1. Expectation, mean and variance	24
2.4.2. Covariance and correlation	26
2.4.3. Linear functions of random variables	28
2.4.4. Ratio of random variables	30
2.4.5. Approximate variance formulae	32
2.4.6. Moments	33
2.5. Characteristic Function	34
2.5.1. Definition and properties	34
2.5.2. Cumulants	37
2.5.3. Probability generating function	38
2.5.4. Sums of a random number of random variables	39
2.5.5. Invariant measures	41
Chapter 3. Convergence and the Law of Large Numbers	43
3.1. The Tchebycheff Theorem and Its Corollary	43
3.1.1. Tchebycheff theorem	43
3.1.2. Bienaymé–Tchebycheff inequality	44
3.2. Convergence	45
3.2.1. Convergence in distribution	45
3.2.2. The Paul Levy theorem	46
3.2.3. Convergence in probability	46
3.2.4. Stronger types of convergence	47
3.3. The Law of Large Numbers	47
3.3.1. Monte Carlo integration	48
3.3.2. The Central Limit theorem	49
3.3.3. Example: Gaussian (Normal) random number generator . .	51
Chapter 4. Probability Distributions	53
4.1. Discrete Distributions	53
4.1.1. Binomial distribution	53
4.1.2. Multinomial distribution	56
4.1.3. Poisson distribution	57
4.1.4. Compound Poisson distribution	60
4.1.5. Geometric distribution	62

4.1.6. Negative binomial distribution	63
4.2. Continuous Distributions	64
4.2.1. Normal one-dimensional (univariate Gaussian)	64
4.2.2. Normal many-dimensional (multivariate Gaussian)	67
4.2.3. Chi-square distribution	70
4.2.4. Student's t -distribution	73
4.2.5. Fisher–Snedecor F and Z distributions	77
4.2.6. Uniform distribution	79
4.2.7. Triangular distribution	79
4.2.8. Beta distribution	80
4.2.9. Exponential distribution	82
4.2.10. Gamma distribution	83
4.2.11. Cauchy, or Breit–Wigner, distribution	84
4.2.12. Log-Normal distribution	85
4.2.13. Extreme value distribution	87
4.2.14. Weibull distribution	89
4.2.15. Double exponential distribution	89
4.2.16. Asymptotic relationships between distributions	90
4.3. Handling of Real Life Distributions	91
4.3.1. General applicability of the Normal distribution	91
4.3.2. Johnson empirical distributions	92
4.3.3. Truncation	93
4.3.4. Experimental resolution	94
4.3.5. Examples of variable experimental resolution	95
Chapter 5. Information	99
5.1. Basic Concepts	100
5.1.1. Likelihood function	100
5.1.2. Statistic	100
5.2. Information of R.A. Fisher	101
5.2.1. Definition of information	101
5.2.2. Properties of information	101
5.3. Sufficient Statistics	103
5.3.1. Sufficiency	103
5.3.2. Examples	104
5.3.3. Minimal sufficient statistics	105
5.3.4. Darmais theorem	106
5.4. Information and Sufficiency	108

5.5. Example of Experimental Design	109
Chapter 6. Decision Theory	111
6.1. Basic Concepts in Decision Theory	112
6.1.1. Subjective probability, Bayesian approach	112
6.1.2. Definitions and terminology	113
6.2. Choice of Decision Rules	114
6.2.1. Classical choice: pre-ordering rules	114
6.2.2. Bayesian choice	115
6.2.3. Minimax decisions	116
6.3. Decision-theoretic Approach to Classical Problems	117
6.3.1. Point estimation	117
6.3.2. Interval estimation	118
6.3.3. Tests of hypotheses	118
6.4. Examples: Adjustment of an Apparatus	121
6.4.1. Adjustment given an estimate of the apparatus performance	121
6.4.2. Adjustment with estimation of the optimum adjustment . .	123
6.5. Conclusion: Indeterminacy in Classical and Bayesian Decisions . .	124
Chapter 7. Theory of Estimators	127
7.1. Basic Concepts in Estimation	127
7.1.1. Consistency and convergence	128
7.1.2. Bias and consistency	129
7.2. Usual Methods of Constructing Consistent Estimators	130
7.2.1. The moments method	131
7.2.2. Implicitly defined estimators	132
7.2.3. The maximum likelihood method	135
7.2.4. Least squares methods	137
7.3. Asymptotic Distributions of Estimates	139
7.3.1. Asymptotic Normality	139
7.3.2. Asymptotic expansion of moments of estimates	141
7.3.3. Asymptotic bias and variance of the usual estimators	144
7.4. Information and the Precision of an Estimator	146
7.4.1. Lower bounds for the variance — Cramér–Rao inequality . .	147
7.4.2. Efficiency and minimum variance	149
7.4.3. Cramér–Rao inequality for several parameters	151
7.4.4. The Gauss–Markov theorem	152
7.4.5. Asymptotic efficiency	153

7.5. Bayesian Inference	154
7.5.1. Choice of prior density	154
7.5.2. Bayesian inference about the Poisson parameter	156
7.5.3. Priors closed under sampling	157
7.5.4. Bayesian inference about the mean, when the variance is known	157
7.5.5. Bayesian inference about the variance, when the mean is known	159
7.5.6. Bayesian inference about the mean and the variance	161
7.5.7. Summary of Bayesian inference for Normal parameters	162
Chapter 8. Point Estimation in Practice	163
8.1. Choice of Estimator	163
8.1.1. Desirable properties of estimators	164
8.1.2. Compromise between statistical merits	165
8.1.3. Cures to obtain simplicity	166
8.1.4. Economic considerations	168
8.2. The Method of Moments	170
8.2.1. Orthogonal functions	170
8.2.2. Comparison of likelihood and moments methods	172
8.3. The Maximum Likelihood Method	173
8.3.1. Summary of properties of maximum likelihood	173
8.3.2. Example: determination of the lifetime of a particle in a restricted volume	175
8.3.3. Academic example of a poor maximum likelihood estimate	177
8.3.4. Constrained parameters in maximum likelihood	179
8.4. The Least Squares Method (Chi-Square)	182
8.4.1. The linear model	183
8.4.2. The polynomial model	185
8.4.3. Constrained parameters in the linear model	186
8.4.4. Normally distributed data in nonlinear models	190
8.4.5. Estimation from histograms; comparison of likelihood and least squares methods	191
8.5. Weights and Detection Efficiency	193
8.5.1. Ideal method maximum likelihood	194
8.5.2. Approximate method for handling weights	196
8.5.3. Exclusion of events with large weight	199
8.5.4. Least squares method	201

8.6. Reduction of Bias	204
8.6.1. Exact distribution of the estimate known	204
8.6.2. Exact distribution of the estimate unknown	206
8.7. Robust (Distribution-free) Estimation	207
8.7.1. Robust estimation of the centre of a distribution	208
8.7.2. Trimming and Winsorization	210
8.7.3. Generalized p^{th} -power norms	211
8.7.4. Estimates of location for asymmetric distributions	213
Chapter 9. Interval Estimation	215
9.1. Normally distributed data	216
9.1.1. Confidence intervals for the mean	216
9.1.2. Confidence intervals for several parameters	218
9.1.3. Interpretation of the covariance matrix	223
9.2. The General Case in One Dimension	225
9.2.1. Confidence intervals and belts	225
9.2.2. Upper limits, lower limits and flip-flopping	227
9.2.3. Unphysical values and empty intervals	229
9.2.4. The unified approach	229
9.2.5. Confidence intervals for discrete data	231
9.3. Use of the Likelihood Function	233
9.3.1. Parabolic log-likelihood function	233
9.3.2. Non-parabolic log-likelihood functions	234
9.3.3. Profile likelihood regions in many parameters	236
9.4. Use of Asymptotic Approximations	238
9.4.1. Asymptotic Normality of the maximum likelihood estimate	238
9.4.2. Asymptotic Normality of $\partial \ln L / \partial \theta$	238
9.4.3. $\partial L / \partial \theta$ confidence regions in many parameters	240
9.4.4. Finite sample behaviour of three general methods of interval estimation	240
9.5. Summary: Confidence Intervals and the Ensemble	246
9.6. The Bayesian Approach	248
9.6.1. Confidence intervals and credible intervals	249
9.6.2. Summary: Bayesian or frequentist intervals?	250
Chapter 10. Test of Hypotheses	253
10.1. Formulation of a Test	254
10.1.1. Basic concepts in testing	254

10.1.2. Example: Separation of two classes of events	255
10.2. Comparison of Tests	257
10.2.1. Power	257
10.2.2. Consistency	259
10.2.3. Bias	260
10.2.4. Choice of tests	261
10.3. Test of Simple Hypotheses	263
10.3.1. The Neyman–Pearson test	263
10.3.2. Example: Normal theory test versus sign test	264
10.4. Tests of Composite Hypotheses	266
10.4.1. Existence of a uniformly most powerful test for the exponential family	267
10.4.2. One- and two-sided tests	268
10.4.3. Maximizing local power	269
10.5. Likelihood Ratio Test	270
10.5.1. Test statistic	270
10.5.2. Asymptotic distribution for continuous families of hypotheses	271
10.5.3. Asymptotic power for continuous families of hypotheses . .	273
10.5.4. Examples	274
10.5.5. Small sample behaviour	279
10.5.6. Example of separate families of hypotheses	282
10.5.7. General methods for testing separate families	285
10.6. Tests and Decision Theory	287
10.6.1. Bayesian choice between families of distributions	287
10.6.2. Sequential tests for optimum number of observations . . .	292
10.6.3. Sequential probability ratio test for a continuous family of hypotheses	297
10.7. Summary of Optimal Tests	298
Chapter 11. Goodness-of-Fit Tests	299
11.1. GOF Testing: From the Test Statistic to the P-value	299
11.2. Pearson’s Chi-square Test for Histograms	301
11.2.1. Moments of the Pearson statistic	302
11.2.2. Chi-square test with estimation of parameters	303
11.2.3. Choosing optimal bin size	304
11.3. Other Tests on Binned Data	308
11.3.1. Runs test	308

11.3.2. Empty cell test, order statistics	309
11.3.3. Neyman–Barton smooth test	311
11.4. Tests Free of Binning	313
11.4.1. Smirnov–Cramér–von Mises test	314
11.4.2. Kolmogorov test	316
11.4.3. More refined tests based on the EDF	317
11.4.4. Use of the likelihood function	317
11.5. Applications	318
11.5.1. Observation of a fine structure	318
11.5.2. Combining independent estimates	323
11.5.3. Comparing distributions	327
11.6. Combining Independent Tests	330
11.6.1. Independence of tests	330
11.6.2. Significance level of the combined test	331
References	335
Subject Index	341