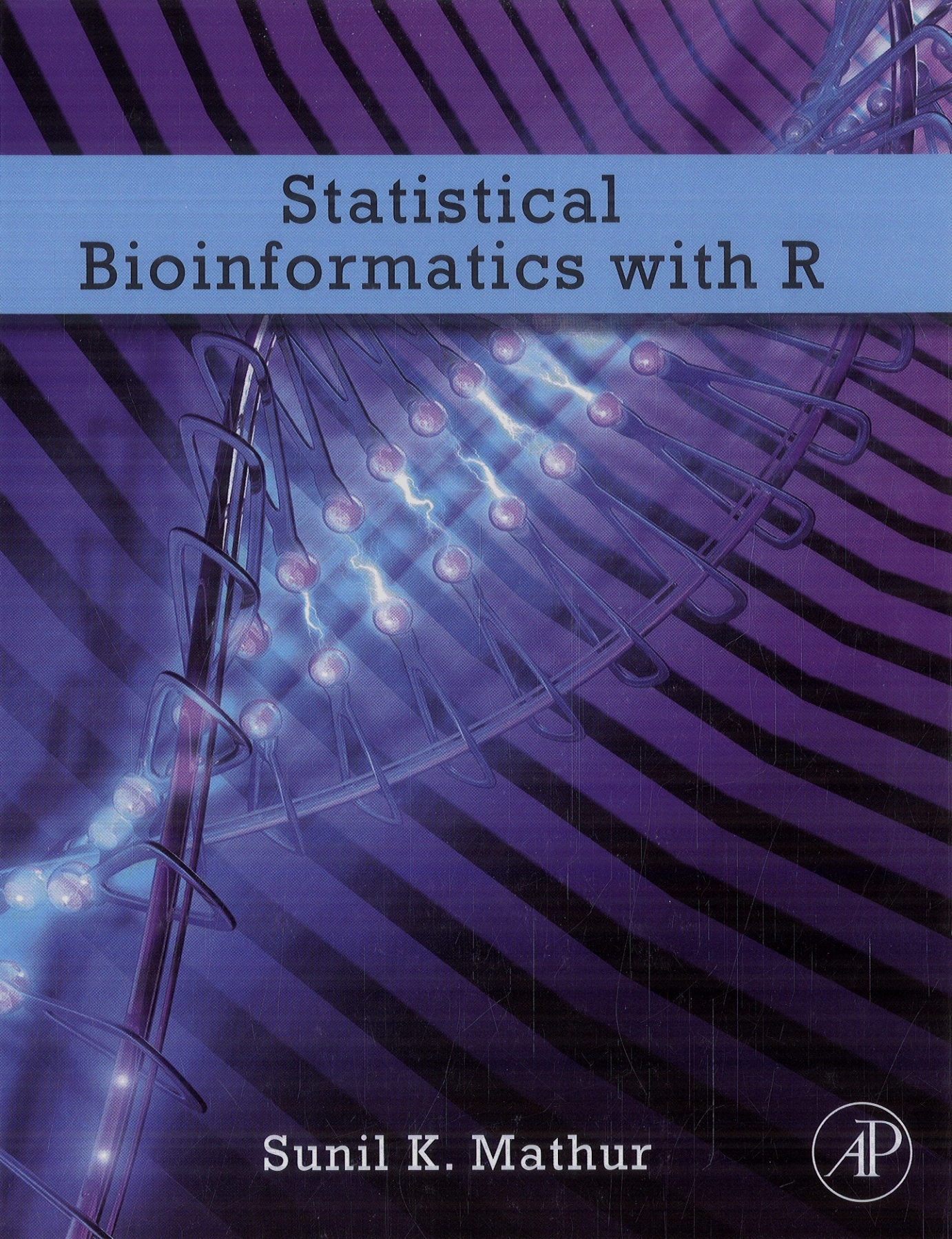




Statistical Bioinformatics with R

Sunil K. Mathur



Contents

PREFACE.....	ix
ACKNOWLEDGMENTS	xv
CHAPTER 1 Introduction	1
1.1 Statistical Bioinformatics	1
1.2 Genetics	3
1.3 Chi-Square Test	6
1.4 The Cell and Its Function	9
1.5 DNA	12
1.6 DNA Replication and Rearrangements	14
1.7 Transcription and Translation	15
1.8 Genetic Code	16
1.9 Protein Synthesis	19
Exercise 1	20
Answer Choices for Questions 1 through 15	21
CHAPTER 2 Microarrays	23
2.1 Microarray Technology	23
2.2 Issues in Microarray	25
2.3 Microarray and Gene Expression and Its Uses	29
2.4 Proteomics	30
Exercise 2	31
CHAPTER 3 Probability and Statistical Theory	33
3.1 Theory of Probability	34
3.2 Mathematical or Classical Probability	36
3.3 Sets	38
3.3.1 Operations on Sets	39
3.3.2 Properties of Sets	40
3.4 Combinatorics	41
3.5 Laws of Probability	44

3.6	Random Variables	53
3.6.1	Discrete Random Variable	55
3.6.2	Continuous Random Variable	56
3.7	Measures of Characteristics of a Continuous Probability Distribution	57
3.8	Mathematical Expectation	57
3.8.1	Properties of Mathematical Expectation	60
3.9	Bivariate Random Variable	62
3.9.1	Joint Distribution	62
3.10	Regression	71
3.10.1	Linear Regression	72
3.10.2	The Method of Least Squares	73
3.11	Correlation	78
3.12	Law of Large Numbers and Central Limit Theorem	80
CHAPTER 4	Special Distributions, Properties, and Applications	83
4.1	Introduction	83
4.2	Discrete Probability Distributions	84
4.3	Bernoulli Distribution	84
4.4	Binomial Distribution	84
4.5	Poisson Distribution	87
4.5.1	Properties of Poisson Distribution	88
4.6	Negative Binomial Distribution	89
4.7	Geometric Distribution	92
4.7.1	Lack of Memory	93
4.8	Hypergeometric Distribution	94
4.9	Multinomial Distribution	95
4.10	Rectangular (or Uniform) Distribution	99
4.11	Normal Distribution	100
4.11.1	Some Important Properties of Normal Distribution and Normal Probability Curve	101
4.11.2	Normal Approximation to the Binomial	106
4.12	Gamma Distribution	107
4.12.1	Additive Property of Gamma Distribution	108
4.12.2	Limiting Distribution of Gamma Distribution	108
4.12.3	Waiting Time Model	108

4.13	The Exponential Distribution	109
4.13.1	Waiting Time Model	110
4.14	Beta Distribution	110
4.14.1	Some Results	111
4.15	Chi-Square Distribution.....	111
4.15.1	Additive Property of Chi-Square Distribution	112
4.15.2	Limiting Distribution of Chi-Square Distribution	112
CHAPTER 5	Statistical Inference and Applications.....	113
5.1	Introduction	113
5.2	Estimation.....	115
5.2.1	Consistency.....	115
5.2.2	Unbiasedness	116
5.2.3	Efficiency	118
5.2.4	Sufficiency.....	120
5.3	Methods of Estimation.....	121
5.4	Confidence Intervals.....	122
5.5	Sample Size	132
5.6	Testing of Hypotheses	133
5.6.1	Tests about a Population Mean	138
5.7	Optimal Test of Hypotheses	150
5.8	Likelihood Ratio Test.....	156
CHAPTER 6	Nonparametric Statistics.....	159
6.1	Chi-Square Goodness-of-Fit Test.....	160
6.2	Kolmogorov-Smirnov One-Sample Statistic	163
6.3	Sign Test.....	164
6.4	Wilcoxon Signed-Rank Test	166
6.5	Two-Sample Test	169
6.5.1	Wilcoxon Rank Sum Test.....	169
6.5.2	Mann-Whitney Test.....	171
6.6	The Scale Problem.....	174
6.6.1	Ansari-Bardley Test.....	175
6.6.2	Lepage Test	178
6.6.3	Kolmogorov-Smirnov Test.....	180

6.7	Gene Selection and Clustering of Time-Course or Dose-Response Gene Expression Profiles	182
6.7.1	Single Fractal Analysis	184
6.7.2	Order-Restricted Inference	186
CHAPTER 7	Bayesian Statistics	189
7.1	Bayesian Procedures	189
7.2	Empirical Bayes Methods	192
7.3	Gibbs Sampler	193
CHAPTER 8	Markov Chain Monte Carlo	203
8.1	The Markov Chain	204
8.2	Aperiodicity and Irreducibility	213
8.3	Reversible Markov Chains	218
8.4	MCMC Methods in Bioinformatics	220
CHAPTER 9	Analysis of Variance	227
9.1	One-Way ANOVA	228
9.2	Two-Way Classification of ANOVA	241
CHAPTER 10	The Design of Experiments	253
10.1	Introduction	253
10.2	Principles of the Design of Experiments	255
10.3	Completely Randomized Design	256
10.4	Randomized Block Design	262
10.5	Latin Square Design	270
10.6	Factorial Experiments	278
10.6.1	2^N -Factorial Experiment	279
10.7	Reference Designs and Loop Designs	286
CHAPTER 11	Multiple Testing of Hypotheses	293
11.1	Introduction	293
11.2	Type I Error and FDR	294
11.3	Multiple Testing Procedures	297
REFERENCES		305
INDEX		315