
Contents

Preface, xi

CHAPTER 1 ■ Basic Probability and Statistics	1
1.1 BIOLOGICAL BACKGROUND	1
1.1.1 Genomics Terminology	1
1.1.2 Microarray Gene Expression	2
1.2 PROBABILITY DISTRIBUTIONS	3
1.3 PROBABILITY FUNCTIONS	6
1.4 CONTINGENCY TABLES	8
1.5 HYPOTHESIS TESTS AND <i>P</i> VALUES	11
1.6 BIBLIOGRAPHICAL NOTES	12
1.7 EXERCISES (PS1–PS3)	12
CHAPTER 2 ■ Introduction to Likelihood	13
2.1 LIKELIHOOD FUNCTION DEFINED	13
2.2 ODDS AND PROBABILITY: WHAT'S THE DIFFERENCE?	15
2.3 BAYESIAN USES OF LIKELIHOOD	16
2.3.1 Bayesian Updating	16
2.3.2 Bounds from <i>p</i> Values	17
	vii

2.4	BIBLIOGRAPHICAL NOTES	19
2.5	EXERCISES (L1–L3)	20
CHAPTER 3 ■ False Discovery Rates		21
3.1	INTRODUCTION	21
3.2	LOCAL FALSE DISCOVERY RATE	22
3.2.1	Local False Discovery Rate Defined	22
3.2.2	The LFDR, Posterior Odds, and Likelihood	24
3.3	NONLOCAL AND LOCAL FALSE DISCOVERY RATES	25
3.4	COMPUTING THE LFDR ESTIMATE	31
3.5	BIBLIOGRAPHICAL NOTES	31
3.6	EXERCISES (L4; A–B)	32
CHAPTER 4 ■ Simulating and Analyzing Gene Expression Data		35
4.1	SIMULATING GENE EXPRESSION WITH DICE	36
4.2	DE GAMES	38
4.2.1	Basic DE	39
4.2.2	Advanced DE	40
4.3	EFFECTS AND ESTIMATES (E&E)	42
4.3.1	Contrast with the DE Games	42
4.3.2	Generating Mean Expression Levels	42
4.3.3	Estimation of Effect Sizes	44
4.3.4	Scoring	45
4.4	UNDER THE HOOD: NORMAL DISTRIBUTIONS	46
4.4.1	Advanced DE's Normal Distributions	46
4.4.2	E&E's Normal Distributions	47

4.5	BIBLIOGRAPHICAL NOTES	49
4.6	EXERCISES (C-E; G1-G4)	49
CHAPTER 5 ■ Variations in Dimension and Data		57
5.1	INTRODUCTION	57
5.2	HIGH-DIMENSIONAL GENETICS	58
5.3	SUBCLASSES AND SUPERCLASSES	60
5.4	MEDIUM NUMBER OF FEATURES	63
5.4.1	Empirical Bayes Methods	63
5.4.2	Applications to Data of Medium Dimension	65
5.5	BIBLIOGRAPHICAL NOTES	70
5.6	EXERCISE (G5)	71
CHAPTER 6 ■ Correcting Bias in Estimates of the False Discovery Rate		73
BACKGROUND		73
6.1	WHY CORRECT THE BIAS IN ESTIMATES OF THE FALSE DISCOVERY RATE?	74
6.2	A MISLEADING ESTIMATOR OF THE FALSE DISCOVERY RATE	75
6.3	CORRECTED AND RE-RANKED ESTIMATORS OF THE LOCAL FALSE DISCOVERY RATE	77
6.4	APPLICATION TO GENE EXPRESSION DATA ANALYSIS	78
6.5	BIBLIOGRAPHICAL NOTES	78
6.6	EXERCISES (CFDR0-CFDR3)	79

CHAPTER 7 ■ The $\mathcal{L}$ Value: An Estimated Local False Discovery Rate to Replace a p Value	81
7.1 WHAT IF I ONLY HAVE ONE p VALUE? AM I DOOMED?	82
7.2 THE $\mathcal{L}$ VALUE TO THE RESCUE!	83
7.3 THE MULTIPLE-TEST $\mathcal{L}$ VALUE	84
7.4 BIBLIOGRAPHICAL NOTES	85
7.5 EXERCISES (LV1–LV9)	86
CHAPTER 8 ■ Maximum Likelihood and Applications	89
8.1 NON-BAYESIAN USES OF LIKELIHOOD	89
8.1.1 Maximum Likelihood Estimation	89
8.1.2 Likelihood-Based p Values	90
8.1.3 Likelihood-Based Confidence Intervals	91
8.2 EMPIRICAL BAYES USES OF LIKELIHOOD	94
8.2.1 Error Model for -omics Data	94
8.2.2 Likelihood for -omics Data	95
8.2.3 MLE for -omics Data	97
8.3 BIBLIOGRAPHICAL NOTES	97
8.4 EXERCISES (M1–M2)	98
Appendix A: Generalized Bonferroni Correction Derived from Conditional Compatibility, 101	
Appendix B: How to Choose a Method of Hypothesis Testing, 107	
BIBLIOGRAPHY, 111	
INDEX, 117	