

Contents

Part I Background

1	Introduction to Probability Theory and Statistics	3
1.1	Basic Probability Theory	4
1.1.1	Introduction	4
1.1.2	Marginal and Conditional Distributions	5
1.1.3	Basic Distributions	6
1.1.4	Order Statistics	12
1.1.5	Convergence	14
1.2	Statistical Inference	15
1.2.1	Estimation and Confidence Intervals	15
1.2.2	Testing Hypotheses	18
1.2.3	Likelihood-Based Test Statistics: Without a Nuisance Parameter	20
1.2.4	Likelihood-Based Test Statistics: With a Nuisance Parameter	21
1.2.5	Multiple Testing	22
1.3	The Delta Method	24
1.4	The Newton-Raphson Method	25
1.5	The EM Algorithm	25
1.6	Sample Size and Power	27
1.7	Asymptotic Relative Efficiency	28
1.8	Bibliographical Comments	29
1.9	Problems	30
2	Introduction to Genetic Epidemiology	33
2.1	Basic Genetic Terminology	34
2.2	Genetic Association Studies	36
2.2.1	Linkage Disequilibrium and Association Studies	36
2.2.2	Population-Based Designs	38
2.2.3	Family-Based Designs	39
2.2.4	Other Designs	39

2.3	Hardy-Weinberg Principle	40
2.3.1	What Is Hardy-Weinberg Equilibrium?	40
2.3.2	Testing Hardy-Weinberg Equilibrium Proportions	43
2.3.3	Impact of Hardy-Weinberg Equilibrium or Disequilibrium	48
2.4	Population Substructure	48
2.4.1	Population Stratification	49
2.4.2	Cryptic Relatedness	50
2.5	Odds Ratio and Relative Risk	51
2.5.1	Odds Ratios	51
2.5.2	Relative Risks	53
2.6	Logistic Regression for Case-Control Studies	55
2.6.1	Prospective Case-Control Design	55
2.6.2	Retrospective Case-Control Design	55
2.7	Bibliographical Comments	57
2.8	Problems	58

Part II Single-Marker Analysis for Case-Control Data

3	Single-Marker Analysis for Unmatched Case-Control Data	61
3.1	Penetrance and Genotype Relative Risks	62
3.2	Genetic Models	63
3.3	Genotype-Based Tests	65
3.3.1	Cochran-Armitage Trend Tests	65
3.3.2	Trend Test Obtained from the Logistic Regression Model	67
3.3.3	Pearson's Chi-Squared Test	69
3.3.4	Pearson's Test Obtained from the Logistic Regression Model	69
3.3.5	Other Likelihood-Based Tests	70
3.4	Allele-Based Test	71
3.4.1	Test Statistics	71
3.4.2	Comparison of the Allele-Based Test with the Trend Test	72
3.5	Exact Tests	74
3.5.1	Exact Tests	75
3.5.2	Permutation Approach	76
3.5.3	Parametric Bootstrap Approach	77
3.6	Hardy-Weinberg Disequilibrium Trend Test	78
3.7	Combining the HWDTT and the CATT	79
3.8	Estimates of Odds Ratios	83
3.9	Simulating Case-Control Samples	84
3.9.1	Without Covariates	84
3.9.2	With Covariates	85
3.10	Adjusting out Covariates	86
3.11	Examples and Case Studies	87
3.11.1	Data from Genome-Wide Association Studies	87
3.11.2	Association Tests	87
3.11.3	Estimates of Odds Ratios	88
3.12	Bibliographical Comments	90
3.13	Problems	92

4	Single-Marker Analysis for Matched Case-Control Data	95
4.1	Notation and Models	96
4.2	Conditional Likelihoods for Matched Case-Control Data	99
4.3	Matched Trend Tests	99
4.3.1	1 : m Matching	100
4.3.2	A Variable Number of Controls and a Case Are Matched	102
4.4	Matching Disequilibrium Tests	103
4.4.1	A Variable Number of Controls and a Case Are Matched	103
4.4.2	1 : m Matching	104
4.5	A Model-Free Test	105
4.5.1	A Variable Number of Controls and a Case Are Matched	105
4.5.2	1 : m Matching	106
4.6	Multiple Cases and Multiple Controls Are Matched	109
4.7	Simulating Matched Case-Control Data	110
4.8	Performance of the Three Test Statistics	112
4.9	Estimates of Odds Ratios and Relative Risks	113
4.9.1	Conditional Odds Ratios	114
4.9.2	Relative Risks	116
4.10	Bibliographical Comments	117
4.11	Problems	117
5	Bayes Factors for Case-Control Association Studies	121
5.1	Introduction	122
5.2	Bayes Factor	123
5.2.1	Definition	123
5.2.2	Interpreting Bayes Factors	123
5.2.3	Approximations of Bayes Factors	124
5.3	Bayes Factor for Genetic Association Studies	127
5.3.1	Laplace Approximation	128
5.3.2	An Example	130
5.3.3	Coding the Genetic Effect	132
5.4	Approximate Bayes Factor for Genetic Association Studies I	132
5.4.1	No Covariates	133
5.4.2	With Covariates	134
5.4.3	An Alternative Derivation	135
5.4.4	Coding the Genetic Effect	135
5.4.5	An Example	136
5.5	Approximate Bayes Factor for Genetic Association Studies II	136
5.5.1	With Covariates	137
5.5.2	No Covariates	139
5.5.3	An Example	140
5.6	Prior Specification	141
5.6.1	Prior for Using a Single Parameter	141
5.7	Simulation Studies Using Approximate Bayes Factors	143
5.8	Bibliographical Comments	145
5.9	Problems	147

6	Robust Procedures	151
6.1	Robust Hypothesis Testing	152
6.1.1	Discrete Numbers of Alternative Hypotheses	152
6.1.2	Alternative Hypothesis Indexed by an Interval	153
6.1.3	Maximin Efficiency	153
6.2	Maximin Efficiency Robust Test	156
6.2.1	The MERT as a Robust Test	156
6.2.2	The MERT Versus a Single Trend Test for Genetic Association	159
6.3	Max Statistics	161
6.3.1	MAX3	162
6.3.2	Monte-Carlo Approaches for MAX3	164
6.3.3	Asymptotic Distribution of MAX3	166
6.3.4	Approximation of the Tail of the Distribution of MAX3: The Rhombus Formula	169
6.3.5	MAX	170
6.3.6	Comparing MAX2, MAX3 and MAX	170
6.3.7	Relationship Among MAX, Trend Test and Pearson's Test	171
6.3.8	Examples	173
6.4	MIN2	176
6.4.1	Joint Distribution and P-Value	176
6.4.2	MIN2 Versus the P-Value of MIN2	177
6.4.3	Examples	177
6.5	The Constrained Likelihood Ratio Test	178
6.5.1	Restricted Maximum Likelihood Estimates of Penetrances	179
6.5.2	Examples	179
6.6	Genetic Model Selection	180
6.6.1	Hardy-Weinberg Disequilibrium and Genetic Models	181
6.6.2	Performance of the Genetic Model Selection	182
6.6.3	Testing Association After the Genetic Model Selection	184
6.6.4	Examples	187
6.6.5	Choice of Thresholds for the Genetic Model Selection	188
6.6.6	Simulating the Null Distribution	190
6.7	Genetic Model Exclusion	191
6.7.1	Reducing the Genetic Model Space	191
6.7.2	The MERT-Based Genetic Model Exclusion Test	192
6.7.3	Examples	195
6.7.4	The MAX-Based Genetic Model Exclusion Test	196
6.8	Simulation Studies with Robust Tests	196
6.8.1	Critical Values and Type I Errors	196
6.8.2	Empirical Power	198
6.8.3	Discussion	199
6.9	MAX3 for Matched Pair Designs	200
6.10	Bibliographical Comments	202
6.11	Problems	204

Part III Multi-marker Analyses for Case-Control Data

7	Haplotype Analysis for Case-Control Data	209
7.1	Haplotype Inference	210
7.1.1	Phase and Phase Ambiguity	210
7.1.2	Haplotype Reconstruction	211
7.1.3	Estimating Haplotype Frequencies	212
7.2	Linkage Disequilibrium	215
7.2.1	Linkage Disequilibrium Coefficients	215
7.2.2	Testing for Linkage Equilibrium	216
7.2.3	Haplotype Block and Haplotype-Tagging SNPs	217
7.3	Haplotype-Based Population Association Analysis	220
7.3.1	Likelihood Ratio Test	220
7.3.2	Regression Method	221
7.3.3	Haplotype Similarity	224
7.4	Linkage Disequilibrium Contrast Tests	225
7.4.1	Composite LD Measure	225
7.4.2	Contrasting LD Measures	228
7.5	Examples	229
7.6	Bibliographical Comments	229
7.7	Problems	232
8	Gene-Gene Interactions	235
8.1	Two-Locus Association Analysis with Interactions	236
8.1.1	Saturated Logistic Regression Model	236
8.1.2	Incorporating Two-Locus Genetic Models	238
8.2	Association Analysis with Higher-Order Interactions	239
8.2.1	Multifactor Dimensionality Reduction	240
8.2.2	Logic Regression	242
8.3	Test for Two-Locus Interactions	243
8.3.1	A Representation of Two-Locus Interaction Effects	246
8.3.2	Contrasting Log-Odds Ratios	246
8.3.3	Relationship with the Log-Linear Model	248
8.3.4	Contrasting LD Measures	248
8.3.5	Test for Second-Order Interactions for Multiple Loci	251
8.3.6	Representation of Higher-Order Interactions	251
8.4	Bibliographical Comments	253
8.5	Problems	255

Part IV Topics Related to Analysis of Case-Control Association

9	Population Structure	259
9.1	Population Structure	260
9.2	Impact of Population Stratification	262
9.2.1	A Model for Population Stratification	262
9.2.2	Impact on Trend Tests	265
9.3	Correcting for Population Stratification	266

9.3.1	Genomic Control	266
9.3.2	Structural Association	269
9.3.3	Principal Components and Clustering	271
9.3.4	Multidimensional Scaling Plots	278
9.4	Selection of Marker Loci	279
9.5	Simulating Data in the Presence of Population Stratification	280
9.6	Comparison of Methods	281
9.7	Bibliographical Comments	283
9.8	Problems	285
10	Gene-Environment Interactions	287
10.1	Introduction	287
10.2	Gene-Environment Interactions and Inference	288
10.2.1	A $2 \times 2 \times 2$ Table	288
10.2.2	An Example	292
10.2.3	More General G and E	294
10.2.4	Gene-Environment Independence	299
10.3	Test Statistics for Gene-Environment Interaction	301
10.3.1	Likelihood Ratio Test	301
10.3.2	Score Test	302
10.3.3	Wald Test	303
10.3.4	Examples	304
10.4	Bibliographical Comments	306
10.5	Problems	307
11	Power and Sample Size Calculations	309
11.1	Single Marker Analysis Using Trend Tests	310
11.1.1	Power and Sample Size Formulas	310
11.1.2	Perfect Linkage Disequilibrium	312
11.1.3	Imperfect Linkage Disequilibrium	313
11.2	Pearson's Chi-Squared Test	316
11.2.1	Example	318
11.2.2	Asymptotic Power of Pearson's Chi-Squared Test and the Trend Test	318
11.3	Using Odds Ratios	319
11.4	Using a Power Program	320
11.4.1	Specifications	320
11.4.2	Examples	321
11.4.3	Limitations	324
11.5	Testing Interactions	324
11.5.1	Score Statistic for an Interaction	324
11.5.2	Gene-Environment Interactions	326
11.5.3	Gene-Gene Interactions	330
11.6	Bibliographical Comments	331
11.7	Problems	332

Part V Introduction to Genome-Wide Association Studies

12 Genome-Wide Association Studies	337
12.1 Introduction	337
12.2 Quality Control	338
12.3 Analysis of GWAS	342
12.3.1 Genome-Wide Scans and Ranking	342
12.3.2 Haplotype Analysis and Interactions	346
12.4 Replication and Follow-Up	347
12.5 Bibliographical Comments	348
12.6 Problems	349

Part VI Introduction to Family-Based Association Studies

13 Analysis of Family Data	353
13.1 Model-Based Methods for Linkage Analysis	354
13.2 Model-Free Methods for Linkage Analysis	357
13.2.1 Estimating Marker Identity by Descent	357
13.2.2 Interval Mapping	358
13.2.3 The Original Haseman-Elston (HE) Regression for a Quantitative Trait	360
13.2.4 The New HE Regression	360
13.2.5 Maximum Likelihood Variance Component Model	361
13.2.6 Qualitative Traits	364
13.3 Transmission/Disequilibrium Test	366
13.4 Robust Methods	372
13.4.1 Linkage Analysis Using Affected Sibpairs	372
13.4.2 Association Analysis Using Trios	374
13.5 Family-Based Methods for Linkage and Association Analysis: FBAT	376
13.5.1 A General FBAT	376
13.5.2 Application to Parent-Offspring Trios	377
13.5.3 A General Pedigree	378
13.5.4 FBAT Website and Software	383
13.6 Bibliographical Comments	384
13.7 Problems	386
References	389
Index	405

Analysis of Genetic Association Studies
Zheng, G.; Yang, Y.; Zhu, X.; Elston, R.C.
2012, XXII, 414 p., Hardcover
ISBN: 978-1-4614-2244-0