

Preface

We started writing this book two years ago targeting it not only as a graduate level text book in statistical genetics and genetic epidemiology, especially for genetic association studies, but also as a reference book for the analysis of genetic association studies. As a text book for graduate students in statistics, biostatistics, genetics and genetic epidemiology, in addition to covering various topics in this subject, we wanted to cover details of the various derivations as well as illustrate detailed step-by-step applications through both real examples and simulations. We hope this book can serve as a bridge from taking classes in statistical genetics and genetic epidemiology to conducting independent research in this area. As a reference, we wanted to cover a broad range of topics in genetic association studies, both population-based and family-based, but we focus mostly on population-based case-control association studies. The book should also be useful for other statisticians or readers who are not familiar with the subject.

The book covers many technical details, and the breadth of coverage gives the option to pick and choose what interests the reader most. The book contains thirteen chapters in six parts and we give here a brief introduction to each part. In the first part, we have two introductory chapters. The probability and statistical background required for this book is covered in the first chapter, while the second chapter covers the basic genetic and genetic epidemiology terminology necessary to understand the rest of the book. Readers who are familiar with the material in either of these two background chapters can skip one or both of them.

Part II of the book comprises four chapters. Chapters 3 and 4 cover single-marker analysis for case-control data in unmatched and matched designs, respectively. In Chap. 3, we introduce both genotype-based tests (including trend tests and Pearson's chi-squared test) and the allele-based test, and inference in terms of odds ratios. Their relation to each other and their relation to a logistic regression model are discussed. Exact tests for association and tests using the deviation from Hardy-Weinberg proportions to detect association are studied. How to simulate case-control data with or without covariates is also studied. In Chap. 4, we focus on matched designs under 1: m or variable matching. Results for the matched trend test and the matched Hardy-Weinberg disequilibrium test are derived. Chapter 5 covers Bayesian analysis of case-control genetic association studies. Bayesian analysis

plays an important role in the analysis of genetic association studies, especially in reporting results from genome-wide association studies. We focus on calculating Bayes factors and their approximations, derivations of approximate Bayes factors with or without covariates, coding genotypes in Bayesian analysis, and the choice of priors. We assume the underlying genetic model is known in all these three chapters. In Chap. 6, however, we assume the genetic model is unknown and study robust procedures for association studies. The maximin efficiency robust test, maximum-type statistics (including MAX3), constrained likelihood ratio test, tests based on genetic model selection or exclusion, and minimum p-values are considered.

Part III, comprising Chaps. 7 and 8, covers multi-marker analysis. We study haplotype analysis in Chap. 7 and gene-gene interactions in Chap. 8. Part IV contains three additional chapters on related topics. Population stratification is an important topic in the analysis of case-control data and is covered in Chap. 9. The impact of population stratification and various approaches to correct for it are discussed. Chapter 10 discusses gene-environment interactions with different genetic models, illustrated with real examples. Power and sample size calculations are important when designing an association study. In Chap. 11 we consider the power and sample size calculations for single marker analysis using the trend test with perfect or imperfect linkage disequilibrium, and for Pearson's chi-squared test. We also cover power for gene-gene and gene-environment interactions using an existing publicly available Power Program. An introduction to genome-wide association studies, popular since 2005, is presented as Chap. 12 in Part V. This brief introduction discusses quality control, analysis strategy, genome-wide scans, ranking, and replication.

An introduction to family-based association studies is given in Chap. 13, the last part of the book. Although we still focus on association studies, we also briefly discuss linkage analysis, including the original and revised Haseman-Elston regression models and linkage studies using affected sibpairs. We focus on the transmission disequilibrium test (TDT) and family-based association tests (FBAT). Both binary and quantitative traits are studied.

One challenge in writing this book has been how to balance the overall coverage, technical details and applications. Although we have tried to cover most topics of association studies, some topics, especially those recently developed since we started writing this book, including the analysis of imputed SNPs, copy number variants and the detection of rare variants, are not covered. The analysis of family data is reduced to one chapter. Moreover, the book focuses more on technical details and presenting application results than on demonstrating them with programs or the use of software. Almost all illustrations presented in the book, including figures and tables, were obtained by running our own programs, which were written using a combination of SAS, R, S-Plus, Maple, S.A.G.E., and other existing programs or software. Therefore it is not easy to present all the programs used in this book, although some illustrations are given. Selected materials can be used for one-semester or a one-year course in statistical genetics together with other supplementary reading materials.

We would like to thank John Kimmel, the former senior editor of statistics at Springer, and Marc Strauss, the current editor, for their support and unending patience. We are grateful to Han Zhang for collecting some references and Huizhen

Qin for producing a figure; to Prakash Gorroochurn, Neal Jeffries, Jungnam Joo, Qizhai Li, and Yong Zang for reading and commenting on some of the chapters. Gang Zheng thanks Joseph Gastwrith and Nancy Geller for their strong support of this project, and Zhaohai Li for teaching him statistical genetic at George Washington University. Yaning Yang also acknowledges the continuous encouragement from Zhiliang Ying and Jurg Ott.

Maryland, Anhui and Cleveland

Gang Zheng
Yaning Yang
Xiaofeng Zhu
Robert C. Elston

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