

1. Record Nr.	UNISA996466407003316
Titolo	Statistical analysis of microbiome data // Somnath Datta and Subharup Guha, editors
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2021] ©2021
ISBN	3-030-73351-3
Descrizione fisica	1 online resource (348 pages)
Collana	Frontiers in probability and the statistical sciences
Disciplina	579
Soggetti	Microbiology - Statistical methods Microbiologia Estadística matemàtica Microorganismes Genètica microbiana Llibres electrònics
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Preface -- Acknowledgments -- Contents -- Part I Preprocessing and Bioinformatics Pipelines -- Denoising Methods for Inferring Microbiome Community Content and Abundance -- 1 Introduction -- 2 Common Algorithmic Denoising Strategies -- 3 Model-Based Denoising -- 3.1 Hierarchical Divisive Clustering -- 3.2 Finite Mixture Model -- 3.3 Denoising Long-Read Technology -- 4 Model Assessment -- 4.1 With Known Truth -- 4.1.1 Accuracy in ASV Identification -- 4.1.2 Accuracy in Read Assignments -- 4.2 With Unknown Truth -- 4.2.1 Assessment with UMIs -- 4.2.2 Clustering Stability -- 5 Conclusions -- References -- Statistical and Computational Methods for Analysis of Shotgun Metagenomics Sequencing Data -- 1 Introduction -- 2 Methods for Species Identification and Quantification of Microorganisms -- 3 Metagenome Assembly and Applications -- 3.1 de Bruijn Assembly of a Single Genome -- 3.2 Modification for Metagenome and Metagenome- Assembled Genomes -- 3.3 Compacted de Bruijn Graph -- 4 Estimation of Growth Rates for Metagenome-Assembled Genomes

(MAGs) -- 5 Methods for Identifying Biosynthetic Gene Clusters -- 5.1 A Hidden Markov Model-Based Approach -- 5.2 A Deep Learning Approach -- 5.3 BGC Identification Based on Metagenomic Data -- 6 Future Directions -- References -- Bioinformatics Pre-Processing of Microbiome Data with An Application to Metagenomic Forensics -- 1 Introduction -- 2 Bioinformatics Pipeline -- 2.1 Microbiome Data -- 2.2 Quality Control -- 2.3 Taxonomic Profiling -- 2.3.1 MetaPhlAn2 -- 2.3.2 Kraken2 -- 2.3.3 Kaiju -- 2.4 Computing facilities -- 3 Methodology -- 3.1 Pre-Processing and Feature Selection -- 3.2 Exploration of Candidate Classifiers -- 3.3 The Ensemble Classifier -- 3.4 Class Imbalance -- 3.5 Performance Measures -- 3.6 Data Analysis -- 4 Results -- 5 Discussion -- 6 Data Acknowledgement -- 7 Code Availability.

References -- Part II Exploratory Analyses of Microbial Communities -- Statistical Methods for Pairwise Comparison of Metagenomic Samples -- 1 Introduction -- 2 Microbial Community Comparison Methods Based on OTU Abundance Data -- 3 Microbial Community Comparison Measures Based on a Phylogenetic Tree -- 3.1 The Fst Statistic and Phylogenetic Test for Comparing Communities -- 3.2 UniFrac, W-UniFrac, VAW-UniFrac, and Generalized UniFrac for Comparing Microbial Communities -- 3.3 VAW-UniFrac for Comparing Communities -- 4 Alignment-Free Methods for the Comparison of Microbial Communities -- 5 A Tutorial on the Use of UniFrac Type and Alignment-Free Dissimilarity Measures for the Comparison of Metagenomic Samples -- 5.1 Analysis Steps for UniFrac, W-UniFrac, Generalized UniFrac, and VAW-UniFrac -- 5.2 Analysis Steps for the Comparison of Microbial Communities Based on Alignment-Free Methods -- 6 Discussion -- References -- Beta Diversity and Distance-Based Analysis of Microbiome Data -- 1 Introduction -- 2 Quantifying Dissimilarity: Common Beta Diversity Metrics -- 3 Ordination and Dimension Reduction -- 3.1 Principal Coordinates Analysis -- 3.2 Double Principal Coordinate Analysis -- 3.3 Biplots -- 3.4 Accounting for Compositionality -- 3.5 Model-Based Ordination Using Latent Variables -- 4 Distance-Based Hypothesis Testing -- 4.1 Permutation Tests -- 4.2 Kernel Machine Regression Tests -- 4.3 Sum of Powered Score Tests -- 4.4 Adaptive Tests -- 4.5 Comparison of Distance-Based Tests -- 5 Strengths, Weaknesses, and Future Directions -- References -- Part III Statistical Models and Inference -- Joint Models for Repeatedly Measured Compositional and Normally Distributed Outcomes -- 1 Introduction -- 2 Motivating Data -- 3 Statistical Models -- 3.1 The Multinomial Logistic Mixed Model (MLMM) -- 3.2 Dirichlet-Multinomial Mixed Model (DMMM). -- 3.3 Goodness of Fit -- 4 Simulation Studies -- 4.1 Simulation Setting -- 4.2 Simulation Results -- 5 Data Analysis -- 6 Discussion -- 7 Software -- Appendix -- References -- Statistical Methods for Feature Identification in Microbiome Studies -- 1 Introduction -- 2 Differential Abundance Analysis -- 2.1 Compositional Methods -- 2.2 Count-Based Methods -- 2.3 Additional Notes -- 3 Mediation Analysis -- 4 Feature Identification Adjusting for Confounding -- 4.1 Covariate Adjustment -- 4.2 Model-Based Standardization -- 5 Summary -- References -- Statistical Methods for Analyzing Tree-Structured Microbiome Data -- 1 Introduction -- 2 Modeling Multivariate Count Data -- 2.1 Dirichlet-Multinomial Model -- 2.2 Dirichlet-Tree Multinomial Model -- 2.3 Implementation and Illustration -- 3 Estimating Microbial Compositions -- 3.1 Empirical Bayes Normalization -- 3.2 Phylogeny-Aware Normalization -- 3.3 Statistical Analysis of Compositional Data -- 3.4 Implementation and Illustration -- 4 Regression with Compositional Predictors -- 4.1 Constrained Lasso and Log-Ratio Lasso -- 4.2

Subcomposition Selection -- 4.3 Phylogeny-Aware Subcomposition
 Selection -- 4.4 Linear Regression and Variable Fusion -- 5 Additional
 References -- 6 Discussion -- References -- A Log-Linear Model for
 Inference on Bias in Microbiome Studies -- 1 Introduction -- 2 Methods
 -- 2.1 The Brooks' Data -- 2.2 Setup and Estimation -- 2.3 Inference
 -- 2.4 Testability of the Hypothesis -- 2.4.1 Example: Testable
 Hypotheses for Main Effects -- 2.4.2 Example: Testable Hypotheses for
 Interaction Effects -- 3 Simulations -- 3.1 Main Effect Simulation -- 3.2
 Interaction Effect Simulation Based on the Brooks Data -- 4 Results --
 4.1 Simulation Results -- 4.2 Do Interactions Between Taxa Affect Bias
 in the Brooks' Data? -- 4.3 Plate and Sample Type Effects in the Brooks'
 Data -- 5 Discussion -- Appendix.
 References -- Part IV Bayesian Methods -- Dirichlet-Multinomial
 Regression Models with Bayesian Variable Selection for Microbiome
 Data -- 1 Introduction -- 2 Methods -- 2.1 Dirichlet-Multinomial
 Regression Models for Compositional Data -- 2.2 Variable Selection
 Priors -- 2.3 Network Priors -- 2.3.1 Unknown G -- 2.4 Dirichlet-Tree
 Multinomial Models -- 2.5 Posterior Inference -- 3 Simulated Data --
 3.1 Simulation Study for DM Regression Models -- 3.2 DM Sensitivity
 Analysis -- 3.3 Simulation Study for DTM Regression Models -- 3.4
 DTM Sensitivity Analysis -- 4 Applications -- 4.1 Multi-omics
 Microbiome Study-Pregnancy Initiative (MOMS-PI) -- 4.2 Gut
 Microbiome Study -- 5 Conclusion -- References -- A Bayesian
 Approach to Restoring the Duality Between Principal Components of a
 Distance Matrix and Operational Taxonomic Units in Microbiome
 Analyses -- 1 Introduction -- 1.1 Motivating Datasets -- 1.2 Nonlinear
 or Stochastic Distances -- 1.3 Limitations of SVD-Based Approaches --
 2 A Bayesian Formulation -- 2.1 Posterior Density -- 3 Model Sum of
 Squares and Biplots -- 4 Posterior Inference -- 4.1 Gibbs Sampler --
 4.2 Dimension Reduction: Skinny Bayesian Technique -- 4.2.1
 Subsetted Data Matrix -- 4.2.2 Lower Dimensional Parameters and
 Induced Posterior -- 4.2.3 Faster Inference Procedure -- 4.3 Model
 Parameter Estimates -- 5 Simulation Study -- 5.1 Generation Strategy
 -- 6 Data Analysis -- 6.1 Tobacco Data -- 6.2 Subway Data -- 7 Data
 Acknowledgement -- 8 Discussion -- Supplementary Materials --
 Appendix -- Proof of Lemma 1 -- References -- Part V Special Topics
 -- Tree Variable Selection for Paired Case-Control Studies with
 Application to Microbiome Data -- 1 Introduction -- 2 Gini Index --
 2.1 Simulation Analysis -- 3 Multivariate Gini Index -- 3.1 Conditional
 Gini Index -- 4 Variable Importance -- 5 Analysis of Obesity Using
 Microbiome Data.
 6 Discussion -- Appendix -- References -- Networks for
 Compositional Data -- 1 Introduction -- 2 Methods -- 2.1 Learning
 Networks from Marginal Associations -- 2.1.1 ReBoot -- 2.1.2 SparCC
 -- 2.1.3 CCLasso -- 2.1.4 COAT -- 2.2 Learning Networks from
 Conditional Associations -- 2.2.1 SPIEC-EASI -- 2.2.2 gCoda -- 2.2.3
 SPRING -- 3 Data-Generating Models -- 3.1 Null Models -- 3.2 Copula
 Models -- 3.3 Logistic-Normal Model -- 4 Results -- 4.1 Spurious
 (Partial) Correlations -- 4.2 Performance in Network Discovery -- 4.3
 Case Studies in R -- 5 Future Directions -- References -- Index.
