Springer Series in Statistics

Dudoit · van der Laar

Multiple Testing Procedures with Applications to Genomics

This book establishes the theoretical foundations of a general methodology for multiple hypothesis testing and discusses its software implementation in R and SAS. The methods are applied to a range of testing problems in biomedical and genomic research, including the identification of differentially expressed and co-expressed genes in high-throughput gene expression experiments, such as micro-array experiments; tests of association between gene expression measures and biological annotation metadata (e.g., Gene Ontology); sequence analysis; and the genetic mapping of complex traits using single nucleotide polymorphisms.

The book is aimed at both statisticians interested in multiple testing theory and applied scientists encountering high-dimensional testing problems in their subject matter area. Specifically, the book proposes resampling-based single-step and stepwise multiple testing procedures for controlling a broad class of Type I error rates, defined as tail probabilities and expected values for arbitrary functions of the numbers of Type I errors and rejected hypotheses (e.g., false discovery rate). Unlike existing approaches, the procedures are based on a test statistics joint null distribution and provide Type I error control in testing problems involving general data generating distributions (with arbitrary dependence structures among variables), null hypotheses, and test statistics. The multiple testing results are reported in terms of rejection regions, parameter confidence regions, and adjusted p-values.

Sandrine Dudoit is Associate Professor of Biostatistics and Statistics at the University of California, Berkeley (<u>www.stat.berkeley.edu/~sandrine</u>). Her research and teaching activities concern the development and application of statistical and computational methods for the analysis of highdimensional biomedical and genomic data. She is a founding core developer of the Bioconductor Project and is an Associate Editor for six journals, including the *Annals of Applied Statistics* and *Statistical Applications in Genetics and Molecular Biology*.

Mark J. van der Laan is Hsu/Peace Professor of Biostatistics and Statistics at the University of California, Berkeley (<u>www.stat.berkeley.edu/~laan</u>). His research concerns causal inference, adjusting for missing and censored data, and simultaneous estimation and testing based on high-dimensional observational and experimental biomedical and genomic data. He is co-author with James Robins of *Unified Methods for Censored Longitudinal Data and Causality* (Springer, 2003). He is a recipient of the 2005 COPSS Presidents' and Snedecor Awards and is an active Associate Editor for five journals, including the *Annals of Statistics* and the *International Journal of Biostatistics*.

Multiple Testing

Sandrine Dudoit · Mark J. van der Laan

Springer Series in Statistics

Procedures with Applications to Genomics

Springer

STATISTICS



> springer.com