

David Gerard

University of Chicago

Department of Human Genetics

920 E. 58th Street, room 410

Cummings Life Science Center (CLSC)

Chicago, IL 60637

U.S.A.

Homepage: <https://dcgerard.github.io/>

Github: <https://github.com/dcgerard>

Phone: (773)834-3936

Email: dcgerard@uchicago.edu

Education

Ph.D. Statistics, University of Washington, June 2015.

M.S. Statistics, The Ohio State University, June 2012.

B.S. Mathematics, The Ohio State University, June 2010.

B.S. Molecular Genetics, The Ohio State University, June 2010.

Articles

In Preparation

Gerard, D., & Stephens, M. Empirical Bayes Shrinkage Estimation, and False Discovery Rate Estimation, Allowing For Unwanted Variation

Abstract:

In this paper we combine two important ideas in the analysis of large-scale genomics experiments (e.g. experiments that aim to identify genes that are differentially expressed between two conditions). The first idea is use of Empirical Bayes (EB) methods to handle the large number of potentially-sparse effects, and estimate false discovery rates and related quantities. The second is use of factor analysis methods to deal with sources of unwanted variation such as batch effects and unmeasured confounders. By combining the key ideas from both these lines of research we provide new and powerful EB methods for analyzing genomics experiments that can account for both large numbers of potentially-sparse effects and for sources of unwanted variation that can otherwise confound inference. In realistic simulations, these new methods provide significant gains in power and calibration over competing methods. In real data analysis we highlight that different methods, while often conceptually similar, can vary widely in their assessments of statistical significance, highlighting the need for care in both choice of methods and interpretation of results.

Gerard, D., Ferrão, L. F. V., & Stephens, M. Harnessing Empirical Bayes and Mendelian Segregation for Genotyping Autopolyploids with Messy Sequencing Data.

Abstract:

Genotyping by next-generation sequencing (NGS) is a powerful and popular way to detect single nucleotide polymorphisms (SNP's) and classify the alleles of individuals. Genotyping diploid individuals using NGS is a well-studied field and similar methods for polyploid individuals are just emerging. However, there are many aspects of NGS data that remain unexplored by most methods. We provide two main contributions in this paper: (1) many datasets feature related individuals, and so we use the structure of Mendelian segregation to build an empirical Bayes approach for

genotyping polyploid individuals; (2) we additionally draw attention to and then model common aspects of NGS data: sequencing error, read-mapping bias, overdispersion, and outlying observations. We verify our method in simulations and apply it to a dataset of hexaploid sweet potatoes (*Ipomoea batatas*).

Preprints

Gerard, D., & Stephens, M. (2017). Unifying and Generalizing Methods for Removing Unwanted Variation Based on Negative Controls. *arXiv preprint arXiv:1705.08393*. [[Link to arXiv](#)]

Gerard, D., & Hoff, P. (2015). Adaptive higher-order spectral estimators. *arXiv preprint arXiv:1505.02114*. [[Link to arXiv](#)] (Accepted for publication by the *Electronic Journal of Statistics*)

Published

Gerard, D., & Hoff, P. (2016). A higher-order LQ decomposition for separable covariance models. *Linear Algebra and its Applications*, 505, 57-84. [[Link to LAA](#)] [[Link to arXiv](#)]

Gerard, D., & Hoff, P. (2015). Equivariant minimax dominators of the MLE in the array normal model. *Journal of Multivariate Analysis*, 137, 32-49. [[Link to JMVA](#)] [[Link to arXiv](#)]

Pollack, J. D., **Gerard, D.**, & Pearl, D. K. (2013). Uniquely Localized Intra-Molecular Amino Acid Concentrations at the Glycolytic Enzyme Catalytic/Active Centers of Archaea, Bacteria and Eukaryota are Associated with Their Proposed Temporal Appearances on Earth. *Origins of Life and Evolution of Biospheres*, 43(2), 161-187. [[Link to Springer](#)]

Gerard, D., Gibbs, H. L., & Kubatko, L. (2011). Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC evolutionary biology*, 11(1), 291. [[Link to BMC](#)]

Teaching Experience

Lecturer of introductory statistics 06/2013 - 09/2013

Developed my own lesson plans and slides from which I lectured three times a week.

Created material for two teaching assistants during their small-group sections.

Course content included comparisons of good sampling designs versus bad sampling designs, conditions under which claims of causality may be made, confounding variables, good and bad measures for answering questions of interest, describing relationships between variables, and good and bad graphics to describe data.

Graduate Teaching Assistant 09/2012 - 06/2013

Led four one-hour recitation sections of introductory statistics every week.

These were small group discussions that acted primarily as support for the main content taught during lecture.

Lecturer of introductory statistics 09/2011 - 12/2011

Lectured to large groups of students using a pre-defined course plan.

Created my own slides to follow along the lesson plans.

Phylogenetic Workshop Assistant 06/2011 - 08/2011

Led a group of undergraduate researchers studying HIV. We studied the association between evolutionary distance (measured by branch distance along estimated trees) and clinical variables collected by an HIV treatment center in Belgium.

Phylogenetic Workshop Assistant 06/2010 - 08/2010

Led lab sessions on Unix and various phylogenetic program packages.

Led a group of undergraduates researching the evolutionary origins of flightless birds.

Individually mentored an undergraduate student on phylogenetic research and statistical analysis with R. We tried to incorporate protein structure into our gene tree estimates while studying the evolutionary relationships of HPV.

Undergraduate Teaching Assistant: 09/2008 - 12/2008

Taught four one-hour recitation sections of intermediate algebra every week.

Research Experience

Postdoctoral Scholar under Professor [Matthew Stephens](#): 09/2015 - Present

Developing empirical Bayes methods to account for hidden confounding in large scale gene expression studies.

Efficiently implementing these methods in R and C++.

Rigorously testing these methods using Monte Carlo methods.

Leading/mentoring graduate students on their individual projects.

Authoring reports and academic papers.

Research Assistant Under Professor [Peter Hoff](#): 09/2013 - 06/2015

Studied optimal estimation of covariance matrices in Gaussian arrays using the notions of equivariance.

Developed novel tensor decompositions, creating higher-order versions of well-known matrix decompositions.

Developed singular value shrinkage mean estimators for tensor-variate data.

Research Assistant Under Professor [Joe Verducci](#): 09/2011 - 06/2012

Studied monotone association between gene expression and drug resistance within unspecified subtypes of cancer.

Undergraduate Research Fellow Under Professors [Laura Kubatko](#) and [H. Lisle Gibbs](#): 04/2009 - 06/2010

Studied the hybridization of two subspecies of Missouri Rattlesnakes and two populations of Black Ratsnakes.

Duties included sequencing work in the lab, computer programming using Perl, and data analysis with various programs.

Awards

Z.W. Birnbaum Award for the best general exam of the 2013-2014 academic year. This is the only research award presented by the University of Washington Statistics Department.

Distinguished University Fellowship (September 2010 - June 2012).

Graduated Summa Cum Laude (June 2010).

Award for excellent research presentation at the BMPS Poster Forum (23 April 2010).

Medalist Scholarship (September 2006 - June 2010):

For academic excellence and essay writing skills.

President's Salute Nominee (April 2010):

For academic excellence among peers from the college of Arts and Sciences.

Coding

R Packages: `tenstr`, `hose`, `vicar`, `seqgendiff`, `updog`.

Advanced training: R, \LaTeX .

Intermediate training: C++, Shell.

Familiar: Python, Matlab/Octave, Mathematica.

Data Analysis

Data Munging (tidyverse).

Data Visualization (ggplot2).

Reproducible Data Analysis (rmarkdown/make).

Version Control (Git/Github).

Continuous Integration (Travis CI/AppVeyor).

R package development.

Last updated: September 1, 2017