

CURRICULUM VITAE

Dabao Zhang

Assistant Professor

Department of Statistics
Purdue University
150 N. University Street
West Lafayette, IN 47907

Office: MATH 534
Tel: (765)494-6046
Fax: (765)494-0558
Email: zhangdb@stat.purdue.edu

EDUCATION

Ph.D. **Statistics** (with Minors in Biometry and Management), Cornell University,
May 2003.

HONORS AND Awards

- National Science Foundation CAREER Award, 2009
- College of Science Interdisciplinary Award, 2009
- Liu Memorial Award, 2003

RESEARCH INTEREST

- **Statistical Theory and Methodology**
Bayesian Statistics; Generalized Threshold Estimates; Graphical Models; Model Selections; Multivariate Extreme Values; Multivariate Statistics; Shrinkage Estimates; Survival Analysis; Variable Selection with High-Dimensional Data.
- **Statistical Genetics and Bioinformatics**
Genetic Epidemiology; Genome-Wide Association Study; Identification of Molecular Signatures; Omics Data Analysis (e.g., Microarray Data and Mass Spectrometry Data); Quantitative Trait Loci Mapping; Variable Selection in Systems Biology.

PROFESSIONAL EXPERIENCE

- **Academic Positions**
Affiliated Member, Oncological Science Center, Purdue University, since 2006.
Affiliated Member, Bindley Bioscience Center, Purdue University, since 2006.
Assistant Professor of Statistics, Purdue University, since 2005.
Assistant Professor of Biostatistics and Computational Biology, University of Rochester Medical Center, August 2003 - August 2005.
- **Statistical Consulting**
The Nathan Shock Center for Aging and Developmental Biology, University of Rochester, Fall 2003 - Spring 2005.

Purdue Statistical Consulting Service Center, Spring 2006 - Spring 2009.

- **Teaching Experience**

Advanced Statistical Methodology, Purdue University.

Applied Regression Analysis, Purdue University.

Applied Survival Analysis, Purdue University.

Current Topics in Bioinformatics, University of Rochester.

Intermediate Statistical Methodology, Purdue University.

- **Managerial Experience**

Assistant Director: Statistical Consulting Service Center, Department of Statistics, Purdue University, January 2006 - May 2009.

RESEARCH FUNDING

- *CAREER: A new regularization framework for identifying composite signatures* (PI: Dabao Zhang), NSF. 6/1/2009-5/31/2014. Total Award: \$433,291.
- *New Statistical Approaches to Enhance Genome-Wide Single Nucleotide Polymorphism Association Mapping to Identify Genetic Markers for Common Complex Human Diseases* (PI: Min Zhang), Purdue Discovery Park. Role: Co-PI, 04/2009 – 03/2010. Total Award: \$39,442.
- *Cook Med Institute Grant from Cook Med Institute Incorporated* (PI: Dabao Zhang). 9/1/2006 – Open End. Total Award: \$25,942/year (financial support of a student).
- *Use of External Cueing to Treat Hypophonia in Parkinson's Disease* (PI: Jessica E. Huber), NIH/NIDCD. Role: Key Personnel, 4/1/2008 – 3/31/2013. Total Award: \$386,027 (year 1).
- *Warfighter Cancer Care Engineering* (PI: Pat Loehrer), DOD/USARMC. Role: Co-Investigator. 7/1/2008 – 1/31/2011. Total Award: \$2,200,000.
- *Empirical Bayes Variable Selection with $p \gg n$ Data*. Summer Faculty Grant from Purdue Research Foundation (PI: Dabao Zhang). Summer 2007; Total Award: \$7,000.

PAPERS

- D. Zhang, Y. Lin and M. Zhang (2009). Penalized orthogonal-components regression for large p small n data. *The Electronic Journal of Statistics*, 3: 781-796.
- M. Zhang, Y. Lin, L. Wang, V. Pungpapong, J. C. Fleet, and D. Zhang (2009). Case-control genome-wide association study of rheumatoid arthritis from GAW16 using POCRE-LDA. Accepted by *BMC Proceedings*.
- Y. Lin, M. Zhang, L. Wang, V. Pungpapong, J. C. Fleet, and D. Zhang (2009). Simultaneous genome-wide association studies for rheumatoid arthritis using the penalized orthogonal-components regression. Accepted by *BMC Proceedings*.
- M. Zhang, D. Zhang and M.T. Wells (2009). Generalized thresholding estimators for high-dimensional location parameters. Accepted by *Statistica Sinica*.
- N. Liu, D. Zhang and H. Zhao (2008). Detection of genotyping errors without replicates in samples from unrelated individuals. *Human Heredity*, 67: 154-162.

- D. Zhang, X. Huang, F.E. Regnier and M. Zhang (2008). Two-dimensional correlation optimized warping algorithm for aligning GC×GC-MS data. *Analytical Chemistry*, 80: 2664-2671.
- M. Zhang, D. Zhang, M. T., Wells (2008). Variable selection with large p small n regression models: mapping QTL with epistasis. *BMC Bioinformatics*, 9:251.
- D. Zhang, M. T. Wells and L. Peng (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *Journal of Multivariate Analysis*, 99: 577-588.
- T.R. Mhyre, R. Loy, P.N. Tariot, L.A. Profenno, K.A. Maguire-Zeiss, D. Zhang, P.D. Coleman and H.J. Federoff (2008). Proteomic analysis of peripheral leukocytes in Alzheimer's disease patients treated with divalproex sodium. *Neurobiology of Aging*, 29: 1631-1643.
- N.-H. Chan, L. Peng and D. Zhang (2007). Empirical likelihood based confidence intervals for conditional variance in heteroscedastic regression models. *Accepted by Econometric Theory*
- D. Zhang and M. Zhang (2007). Bayesian profiling of molecular signatures to predict event times. *Theoretical Biology & Medical Modelling*, 4:3, doi:10.1186/1742-4682-4-3.
- D. Zhang, M. Zhang, M. T. Wells (2006). Multiplicative background correction for spotted microarrays to improve reproducibility. *Genetical Research*, 87: 195-206.
- D. Zhang, M. T. Wells, B. W. Turnbull, D. Sparrow and P. A. Cassano (2005). Hierarchical Graphical Models: An Application to Pulmonary Function and Cholesterol Levels in the Normative Aging Study. *Journal of American Statistics Association*, 100: 719-727.
- M. Zhang, K. L. Montooth, M. T. Wells, A. G. Clark and D. Zhang (2005). Mapping Multiple Quantitative Trait Loci by Bayesian Classification, *Genetics*, 169: 2305-2318.
- D. Zhang, M. T. Wells, C. D. Smart and W. E. Fry (2005). Bayesian Normalization and Inference for Differential Gene Expression Data. *Journal of Computational Biology*, 12: 391-406.
- S. W. Perry, J. P. Norman, A. Litzburg, D. Zhang, S. Dewhurst and H. A. Gelbard (2005). HIV-1 Tat Induces Mitochondrial Hyperpolarization and Synaptic Stress Leading to Apoptosis, *Journal of Immunology*, 174: 4333-4344.
- Complex Traits Consortium (including D. Zhang). (2004). The Collaborative Cross: A Community Resource for the Genetic Analysis of Complex Traits. *Nature Genetics*, 36: 1133-1137.
- M. Zhang, X. Wang, D. Zhang, G. Xu, H. Dong, Y. Yu and J. Han (2004). Orphanin FQ antagonizes the inhibition of Ca²⁺ currents induced by mu-opioid receptors. *Journal of Molecular Neuroscience*, 25: 21-27.
- D. Zhang, S. He and Z. Xie (1993). Outlier Detection and Intervention for $ARIMA(p, d, 0)$. *Proceedings of First Asian Conference on Statistical Computation*.

PRESENTATIONS/INVITED TALKS

- *Penalized Orthogonal-Components Regression for Large p Small n Data*. Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, September 9, 2009.

- *Penalized Orthogonal-Components Regression for Large p Small n Data*. Department of Statistics, Purdue University, September 3, 2009.
- *Penalized Orthogonal-Components Regression for Large p Small n Data*. Department of Mathematics & Statistics, Georgia State University, August 28, 2009.
- *Penalized Orthogonal-Components Regression for Large p Small n Data*. School of Mathematics, Georgia Institute of Technology, August 27, 2009.
- *Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC \times GC-MS Data*. The International Chinese Statistical Association (ICSA) Applied Statistics Symposium, San Francisco, CA, June 24, 2009.
- *Penalized Orthogonal-Components Regression for Large p Small n Data*. Yale Statistics Workshop, May 16, 2009.
- *Systems Biology and Genome-Wide Association Study*. Systems Biology Luncheon at Computational Research Institute, Purdue University, April 14, 2009.
- *Some Multivariate Statistical Methods for Genome-Wide Association Study*. StatDay, Purdue University, April 2, 2009.
- *Some Multivariate Statistical Methods for Genome-Wide Association Study*. Genetic Analysis Workshop 16, St. Louis, Missouri, September 19, 2008.
- *Nonparametric Estimation of the Dependence Function for a Multivariate Extreme Value Distribution*. Department of Bioinformatics and Biostatistics, School of Public Health, University of Louisville, November 7, 2008.
- *On Preprocessing and Analyzing GC \times GC-MS Data*. Bioinformatics Seminar, Purdue University, West Lafayette, IN. September 9, 2008.
- *On Preprocessing and Analyzing GC \times GC-MS Data*. The Monthly Cancer Career Engineering Meeting, Purdue University, West Lafayette, IN. May 22, 2008.
- *Generalized Shrinkage Estimators Adaptive to Sparsity and Asymmetry of High Dimensional Parameter Spaces*. Department of Statistics, University of Illinois at Urbana-Champaign, September 20, 2007.
- *Generalized Shrinkage Estimators Adaptive to Sparsity and Asymmetry of High Dimensional Parameter Spaces*. Seminar on Bayesian Inference in Econometrics and Statistics, St. Louis, Missouri, May 4, 2007.
- *Background Correction and Normalization of Spotted Microarray*. Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, October 27, 2006.
- *Nonparametric Estimation of the Dependence Function for a Multivariate Extreme Value Distribution*. VIGRE Seminar, Department of Statistics, Purdue University, West Lafayette, IN. September 13, 2006.
- *Nonparametric Estimation of the Dependence Function for a Multivariate Extreme Value Distribution*. Ninth Meeting of New Researchers in Statistics and Probability, Seattle, WA, August 2006.
- *Background Correction and Normalization of Spotted Microarray*. Department of Statistics, Iowa State University. April 26, 2006.
- *Identifying Molecular Signatures for Survival Outcomes and Hierarchical Graphical Models*. Presented to Genentech Visitors, Department of Statistics, Purdue University. March 20, 2006.

- *Bayes, Empirical Bayes and Adaptive Bayes*. VIGRE Seminar, Department of Statistics, Purdue University. October 5, 2005.
- *Double Shrinkage Estimator and Its Implementation*. JSM, Minneapolis, MN, 2005.
- *Multivariate Statistics with Application to Bioinformatics*. Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, 2005.
- *Hierarchical Graphical Models: An Application to Pulmonary Function and Cholesterol Levels in the Normative Aging Study*. Center for Statistical Sciences, Brown University, 2005.
- *Multivariate Statistics with Applications*. Department of Statistics, Purdue University, 2005.
- *Multivariate Statistics with Application to Bioinformatics*. Department of Statistics and Probability, Michigan State University, 2005.
- *Bayesian Identification of Prognostic Molecular Signatures for Survival Phenotypes*. ENAR, Austin, TX, 2005.
- *Microarray, QTL Mapping and Identifying Molecular Markers/Signatures*, Center for Aging and Developmental Biology, University of Rochester Medical Center, 2005.
- *Double-Shrinkage Estimator and Its Implementation*. Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 2004.
- *QTLBayes: Mapping Multiple Quantitative Trait Loci Using Bayesian Variable Selection*. The Third Annual Conference of Complex Trait Consortium, Bar Harbor, ME, 2004.
- *Bayesian Normalization and Identification for Differential Gene Expression Data*. International Conference on Analysis of Genomic Data, Boston, MA, 2004.
- *Bayesian Variable Selection to Identify Quantitative Trait Loci*. Poster at the Eighth Annual International Conference on Research in Computational Molecular Biology, San Diego, CA, 2004.
- *Inference for Nonrecursive Graphical Model Mixed with Discrete and Censored Continuous Clustered Responses*. JSM, San Francisco, CA, 2003.
- *Bayesian Normalization and Identification for Differential Gene Expression Data*. Department of Biostatistics and Computational Biology, University of Rochester Medical Center, 2003.
- *Bayesian Normalization and Identification for Differential Gene Expression Data*. The Computational and Applied Genomics Program, Duke University Medical Center, 2003.
- *Bayesian Normalization and Identification for Differential Gene Expression Data*. Department of Biostatistics, University of Minnesota, 2003.
- *Bayesian Normalization and Inference on Differential Gene Expression Data*. Biometrics Unit Discussion Series, Cornell University, 2002.

SOFTWARE DEVELOPED

- *POCRE*: Penalized Orthogonal-Components Regression for Variable Selection with High-Dimensional Data.
- *2DCOW*: Two-Dimensional Correlation Optimized Warping Algorithm for Aligning GC×GC-MS Data.

- *QTLBayes*: Mapping Multiple QTL with Bayesian Classification.
- *MicroBayes*: Bayesian Normalization of cDNA Microarray Data.
- *SEMMIX*: Simultaneous Equation Models with Mixed Coefficients.

JOURNAL REFEREEING

American Statistician, Bioinformatics, Biometrical Journal, BMC Bioinformatics, Genetics, Human Molecular Genetics, Journal of American Statistical Association, Journal of Machine Learning Research, Journal of Multivariate Analysis, Journal of Nonparametric Statistics, Neurobiology of Aging, Statistical Applications in Genetics and Molecular Biology, Test, The Plant Cell.

PROFESSIONAL AFFILIATIONS

American Society of Human Genetics

American Statistical Association

Institute of Mathematical Statistics

International Chinese Statistical Association

Institute of Mathematical Statistics