

Highlights of Qualifications:

- Motivated, Diligent and hardworking person well trained in statistics, math and computer science.
- Demonstrated extensive experience in statistical modeling and forecasting, linear regression, survey sampling, experiment design, Bayesian analysis, machine learning, time series, etc.
- Versatile wit with exceptional computer skills including R, SAS, SQL, Java, C++, MATLAB, Linux, etc.

Education:

- PhD, Statistics, **Iowa State University (ISU)** Expected June 2013
- BS, Statistics, **University of Science and Technology of China** June 2008

Projects:

Salmonella Shedding Phenotypic Classes Prediction

- Examined data quality with the R package “arrayQualityMetrics”; uncovered and dropped a problematic chip.
- Invented non-statistical methods to filter genes/variables/features; alleviated the problem of large number of variables (genes) and small number of observations (samples).
- Integrated the **lasso** method with **the logistic regression** model to further select genes; identified a small list of biologically interesting genes which is under investigation using qPCR technology.

Modeling and Inference for Equivalence Classes of 3-D Orientations

- Introduced the UARS model for equivalence classes of orientations; the model has a nice geometric interpretation.
- Proposed a Metropolis-Hastings algorithm for MCMC and implemented it in parallel MATLAB code.
- Demonstrated that the proposed Bayesian method beats standard likelihood-based approaches.

Clustering Equivalence Classes of 3-D Orientations

- Proposed a hierarchical method for clustering orientations with spatial information; the method has several “knobs” for tuning clusters.
- Originated a series of Markov chains on partitions (named **Du Process**); proved properties of the Du process; illustrated use of the Du Process in model-based Bayesian clustering.

Miscellaneous Statistical Genetics Projects

- Communicated with non-statisticians, learned their problems, understood their intentions and explained statistical concepts and analysis results to them.
- Analyzed data using **generalized/mixed linear models**; adapted methods and solved problems of low expression, dependency structure, heterogeneous variance, confound factors and numerical issues.

Estimation of False Discovery Rate (FDR) using Sequential Permutation P-values

- Proposed a histogram-based method for estimating FDR using sequential permutation p-values; the method does not lose much power compared to methods based on regular permutation p-values.
- Implemented the method in **parallel** C++ code and developed the R package “dclong.spt” wrapping the C++ code.

Teaching Experience:

- Acted as teaching assistant and lab assistant for several graduate level courses including Statistical Methods, Advanced Probability Theory, Advanced Statistical Inference and Bayesian analysis.
- Stimulated students’ interests; cultivated and enhanced their understanding in statistical concepts.
- Tutored a student during summer 2009 on probability theory; improved and expedited her study.

Selected Publications:

- T. Bancroft, **C. Du** and D. Nettleton (2013). Estimation of False Discovery Rate Using Sequential Permutation p -Values. *Biometrics*. doi: 10.1111/j.1541-0420.2012.01825.x
- E. M. Takacs, J. Li, **C. Du**, L. Ponnala, D. Janick-Buckner, J. Yu, G. J. Muehlbauer, P. S. Schnable, M. C.P. Timmermans, Q. Sun, D. Nettleton and M. J. Scanlon. Ontogeny of the Maize Shoot Apical Meristem. *The Plant Cell Online, Am Soc Plant Biol*, 2012, 24, 3219-3234.
- **C. Du**, S. Vardeman and D. Nordman (2013). One-Sample Bayes Inference for a New Class of Distributions on Equivalence Classes of 3-D Orientations Defined by Crystallographic Symmetries. *Technometrics*, tentatively accepted.
- **C. Du** (2012). A Series of Stationary and Ergodic Markov Chains Defined on Partitions with Applications in Bayesian Clustering. (submitted to *Bayesian Analysis*)
- **C. Du** (2013). A Method for Identifying Grains in EBSD Scans of Material Specimens Using Spatially Informed Clustering of 3-D Orientations. (working)

Extracurricular Activities:

- Organized Career Day 2012 for the Dept. of Statistics and the Dept. of Math at ISU.
- Lunched a software learning group and taught R, MATLAB and Mathematica to members.
- Revitalized the mailing list of CSSA in Ames and maintained it for 2 years.
- Initiated an “English Only” campaign among international students in the Dept. of Statistics at ISU.