Ben Chuanlong Du

Email: longendu@gmail.com Phone: 515-203-1787
Blog: legendu.net/

Highlights of Qualifications:

- Statistics/Machine learning skills: random forest, SVM, logistic regression, naïve Bayesian algorithm, hierarchical clustering, k-means clustering, linear regression, time series, Bayesian analysis, experiment design, optimization, etc.
- Programming skills: R, SQL, Java, Python, C/C++, NoSQL, Hadoop, MATLAB, VBA, SAS, Linux, data structure, etc.

Education:

- PhD, Statistics, *Iowa State University* (ISU) Jan 2014
- BS, Statistics, *University of Science and Technology of China* (USTC) Jun 2008

Working Experience:

- AVP; Quantitative Financial Analyst at Bank of America Apr 2014 – now
- Student Intern at Union Bank Jun 2013 – Sep 2013

Projects:

- ***Modeling*** *and* ***Inference*** *for Equivalence Classes of 3-D Orientations (dissertation)*
  - UARS models, equivalence class of orientations, Metropolis-Hastings, Bayesian inference
  - parallel computing, MATLAB, Mathematica, R
- ***Clustering*** *Equivalence Classes of 3-D Orientations (dissertation)*
  - Spatial information, hierarchical clustering, Markov chains on partition, Bayesian clustering
- *Salmonella Shedding Phenotypic Classes* ***Prediction*** *(research)*
  - large p small n, feature selection, lasso logistic regression, cross validation
- *Perspective to LinkedIn's Custom Churn (personal)*
  - random forest, feature selection, data integrity, outlier detection, cross validation
  - business relationship, customer behavior, customer properties
- *Miscellaneous Statistical Genetics* ***Consulting*** *Projects (research)*
  - Talk to client, explain statistics concept, interpret results,
  - generalize/mixed linear models, dependency structure, heterogeneous variance, confound factors
- ***Sampling*** *Design for Estimating Cash Dollar Amount (BOA)*
  - stratified sampling, Horvitz-Thompson estimation, sample size calculation
- *Online Transaction Monitoring – Next Day and 3 Days (N3D) Scenario (BOA)*
  - anti-money laundering, scenarios dev, risk factors, SAS, SQL, big data, data integrity
- *Validation of Risk Rating of Low Default Portfolios (Union Bank)*
  - Low default rate, Bayesian model, simulate distribution of number of defaults
- *Implementation of Statistical Methods in* ***C++*** *(research)*
  - adaptive rejection sampling, sequential permutation test, high performance, template, parallel computing, R package "dclong.spt"
- *Job Transition of Users on LinkedIn in* ***Java*** *(personal)*
  - big data, parallel computing, large file splitting, merge sort
- *Implementation of Ising model for* ***Clustering*** *in* ***Java*** *(course)*
  - lattice, Gibbs Sampling, generic programming
- *Simulation of Corn Kernel Distribution in* ***Java*** *(research)*

- parallel programming, generic programming

- *Implementation of Sequential Permutation Test in **Java** (research)*
  - sequential permutation test, parallel computing
- *Blog Managing System in **Python** (personal)*
  - Pelican/Octopress, fuzzy search, text mining, hash
- *Notes Taking App in **Python** (personal)*
  - NoSQL, mongodb, tagging, notes searching
- *Comprehensive **Linux** Configuration Management in **Bash** (personal)*
  - quick reinstallation, quick configuration, bash script
  - Debian, Ubuntu, Mint, antiX, Cygwin and MobaXterm support
- *Intelligent Email System in **VBA** (personal)*
  - MS Access, Outlook, VBA, contact database, holiday/birthday/agenda reminder, lending system, email parser, interact with database via email
- *Solve the Sum and Product **Puzzle** using **Mathematica** (personal)*
  - impossible puzzle, mathematical abstraction, parallel computing, upper limit exploration
- *Knowledge **Sharing** with Colleagues (BOA)*
  - developed and shared Linux shell scripts, video tutorial, automation tools (AutoHotkey, Sikuli, etc.)

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.
- Launched 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.