

Program

**The 2nd Workshop on Biostatistics and Bioinformatics:
Celebrating the International Year of Statistics**

Department of Mathematics and Statistics

Georgia State University

May 10-12, 2013

Contents

Sponsor 3

Organizer 3

Keynote and Invited Speakers 3

Acknowledgements 4

Conference Schedule 5

Keynote and Invited Talks10

Poster Abstracts 19

Sponsor

The Georgia State University Research Foundation, and the Department of Mathematics and Statistics in the Georgia State University

Organizer

Yichuan Zhao

Department of Mathematics and Statistics

Georgia State University

Keynote Speaker

Runze Li, Penn State University

Invited Speaker

Ash Abebe, Auburn University

Jeongyoun Ahn, University of Georgia

Guang Cheng, Purdue University

Don Edwards, University of South Carolina

Varghese George, Medical College of Georgia (Georgia Regents University)

Don Hong, Middle Tennessee State University

Jianhua Huang, Texas A&M University

Pengsheng Ji, University of Georgia

Abhyuday Mandal, University of Georgia

Alexander McLain, University of South Carolina

Qi Long, Emory University

Steve Qin, Emory University

Yanqing Sun, University of North Carolina, Charlotte

Chang Yu, Vanderbilt University

Donglin Zeng, University of North Carolina, Chapel Hill

Peng Zeng, Auburn University

Acknowledgements

The organizer thanks Earnestine Collier-Jones, Sandra Ahuama-Jonas, Yvonne Pierce, Leslie Meadows, Xin Qi and other volunteers for their great efforts in setting up this workshop and making it run successfully.

Conference Schedule

All workshop sessions meet in the room 150, College of Education Building, 30 Pryor Street, Atlanta, GA 30303.

Friday, May 10, 2013

2:00-6:00 pm **Registration:** on the 7th Floor, College of Education Building, 30 Pryor Street.

Saturday, May 11, 2013

8:00-8:30 am **Registration:** on the 1st Floor, College of Education Building, 30 Pryor Street.

8:30-8:45 am Conference Welcome: Guantao Chen, Chair of the dept., Georgia State University
Opening Remarks: William Long, Dean of Arts & Science, Georgia State University

8:45-9:45 am **Session 1 (Keynote Talk): Chair:** *Yichuan Zhao, Georgia State University*

**Feature Selection for Varying Coefficient Models With
Ultrahigh Dimensional Covariates**

Runze Li, Penn State University

9:45-10:00 am **Break:** Refreshments

10:00-11:50 am **Session 2: Chair:** *Xin Qi, Georgia State University*

Exact Analysis of Closed Sequential or Multistage Experiments with Binary Response

Don Edwards, University of South Carolina

Profile Local Linear Estimation of Generalized Semiparametric Regression Model for Longitudinal Data

Yanqing Sun, University of North Carolina, Charlotte

Iterative Rank Estimation for Longitudinal Data

Ash Abebe, Auburn University

11:50-1:45 pm **Lunch Time**

1:45-3:30 pm **Session 3: Chair:** *Ruiyan Luo, Georgia State University*

Auxiliary marker-assisted classification in the absence of class identifiers

Donglin Zeng, University of North Carolina, Chapel Hill

Statistical Approximation Methods for Imaging Mass Spectrometry Data Processing

Don Hong, Middle Tennessee State University

Multi-objective optimal experimental designs for event-related fMRI studies

Abhyuday Mandal, University of Georgia

3:30-3:45 pm **Break:** Refreshments

3:45-5:15 pm **Session 4: Chair:** *Alexander McLain, University of South Carolina*

Towards the understanding of the three-dimensional genome

Organization Statistical challenges and opportunities for analyzing

Hi-C data

Steve Qin, Emory University

Covariance Adjustment for Batch Effect in Gene Expression Data

Jeongyoun Ahn, University of Georgia

Models for sums of dependent Bernoulli random variables

Chang Yu, Vanderbilt University

5:15-5:30 pm **Break:** Refreshments

5:30-6:45 pm **Poster Session: Chair:** Remus Osan, Georgia State University:

Room 150, College of Education Building

7:00-9:30 pm **Workshop Banquet:** Alma Cocina : One Ninety One Peachtree Tower (adjacent to downtown Ritz-Carlton) 191 Peachtree Street NE Atlanta, GA 30303.

All workshop sessions meet in the room 150, College of Education Building

Sunday, May 12, 2013

8:00-8:30 am **Registration:** on the 1st Floor, College of Education Building, 30 Pryor Street.

8:30-10:15 am **Session 5: Chair:** *Jianhua Huang, Texas A&M University*

The Long March Towards Joint Asymptotics: My 1st Steps...

Guang Cheng, Purdue University

Semi-parametric grouped backward recurrence Cox model for the analysis of current duration data with preferential reporting

Alexander McLain, University of South Carolina

Joint Modeling of Cancer Risk and Correlated Functional Biomarkers Measured with Error

Qi Long, Emory University

10:15-10:30 am **Break:** Refreshments

10:30-12:05pm **Session 6: Chair:** *Guang Cheng, Purdue University*

Regularized Matrix Decomposition and its Applications

Jianhua Huang, Texas A&M University

Linearly Constrained Generalized Lasso

Peng Zeng, Auburn University

***UPS delivers optimal phase diagram in high-dimensional
variable selection***

Pengsheng Ji, University of Georgia

12:05-12:10pm **Final Remarks** by Yichuan Zhao

Keynote Talk

Feature Selection for Varying Coefficient Models With Ultrahigh Dimensional Covariates

Runze Li

Penn State University

This paper is concerned with feature screening and variable selection for varying coefficient models with ultrahigh dimensional covariates. We propose a new feature screening procedure for these models based on conditional correlation coefficient. We systematically study the theoretical properties of the proposed procedure, and establish their sure screening property and the ranking consistency. To enhance the finite sample performance of the proposed procedure, we further develop an iterative feature screening procedure. Monte Carlo simulation studies were conducted to examine the performance of the proposed procedures. In practice, we advocate a two-stage approach for varying coefficient models. The two stage approach consists of (a) reducing the ultrahigh dimensionality by using the proposed procedure and (b) applying regularization methods for dimension-reduced varying coefficient models to make statistical inferences on the coefficient functions. We illustrate the proposed two-stage approach by a real data example.

Invited Talks

Iterative Rank Estimation for Longitudinal Data

Ash Abebe

Auburn University

An iterative rank-based method is proposed for the estimation of the parameters of a generalized linear model with longitudinal responses. The correlation structure does not need to be specified and the estimator does not depend on prior estimation of the covariance matrix of the response vector. Our estimator is obtained through iterative minimization of a Wilcoxon type dispersion function that accommodates within subject dependency. Consistency and asymptotic normality are established under minimal assumptions. The procedure results in estimators that are robust in the response space but not in factor space. Thus the procedure is more appropriate for controlled studies.

Covariance Adjustment for Batch Effect in Gene Expression Data

Jeongyoun Ahn

University of Georgia

Batch bias has been found in many microarray studies that involve multiple batches of samples. Currently available methods for batch effect removal are mainly based on gene-by-gene analysis. However, there has been relatively little development on multivariate approach to batch adjustment, mainly because of the analytical difficulty due to the high dimensional nature of gene expression data. We propose a multivariate batch adjustment method that effectively eliminates inter-genes batch effects. The proposed method utilizes high dimensional sparse covariance estimation based on a factor model and a hard-thresholding technique. Another important aspect of the proposed method is that if there exists an ideally obtained batch, other batches can be adjusted so that they resemble the target batch. We study theoretical properties of the proposed estimator. Using two gene expression data sets we demonstrate the effectiveness of the proposed method and compare with other approaches in terms of both homogeneity of adjusted batches and cross-batch prediction performance.

The Long March Towards Joint Asymptotics: My 1st Steps...

Guang Cheng

Purdue University

We consider the joint asymptotics and inferences for the semi-nonparametric models where an Euclidean parameter and an infinite dimensional parameter are both of interest. The joint inferences, e.g., the prediction interval, are very useful in practice but its theoretical validity are very challenging to establish due to the different natures of two model components: parametric v.s. nonparametric. In this talk, we present the first comprehensive studies on the joint local/global inferences in a unified asymptotic framework. For simplicity of illustration, we use the generalized partly linear models as the prototypical example. Two types of regularizations: (i) penalized estimation; and (ii) local polynomial estimation, are considered. Some long standing conjectures in the semiparametric/nonparametric statistics have been solved as a by-product. The inference optimality/efficiency issues are also carefully addressed. The theoretical foundation of all our results is the novel joint Bahadur representation, which is also of independent interest. This talk can be viewed as the first step to exploring the intriguing joint asymptotics phenomena in statistics.

Exact Analysis of Closed Sequential or Multistage Experiments with Binary Responses

Don Edwards

University of South Carolina

The talk is about closed sequential or multistage sampling with binary responses (e.g. toxic response, DLT). It is shown that exact inference on the probability of response using closed (bounded) sequential or multistage procedures with general pre-specified elimination boundaries is completely tractable and not at all inconvenient using modern statistical software. Relevant theory is outlined and functions for this purpose written in R are described. A detailed example, a sharpening of a closed version of Wald's Sequential Probability Ratio Test, is given; applications to Medicare fraud investigations and election audits are outlined.

Statistical Approximation Methods for Imaging Mass Spectrometry Data Processing

Don Hong

Department of Mathematical Sciences and Center of Computational Sciences

Middle Tennessee State University

Murfreesboro, Tennessee, USA

In proteomics study, Imaging Mass Spectrometry (IMS) is an emerging and very promising new technique for protein analysis from intact biological tissues. Though it has shown great potential and is very promising for rapid mapping of protein localization and the detection of sizeable differences in protein expression, challenges remain in data processing due to the difficulty of high dimensionality. Advanced mathematical tools and statistical techniques can not only provide significance analysis of experimental data sets but also can help in finding new data features/patterns, guiding biological experiments designs, as well as leading computational tools development. In this talk, we would like to report some recent progress using statistical approximation methods for IMS cancer data analysis. We'll introduce the IMSmining, a software package newly developed by a research group at MTSU for imaging mass spectrometry processing using computational statistical methods.

Regularized Matrix Decomposition and its Applications

Jianhua Huang

Texas A&M University

In this talk, I will review some recent works on regularized matrix decomposition. Depending on the application, the matrix in consideration can be the data matrix, the latent canonical parameter matrix of an exponential family distribution, or the regression coefficient matrix of a multivariate regression. I will discuss use of various penalty functions for regularization purpose, including sparsity-inducing penalty, roughness penalty, and their combinations. Governed by the structure of the problem, the penalty can be designed for one-way or two-way regularization. I will illustrate the key ideas using applications in functional principal components analysis, biclustering, reconstruction of MEG/EEG source signals, and protein structure clustering using protein backbone angular distributions. This talk is based on joint works with Andreas Buja, Xin Gao, Seokho Lee, Mehdi Maadooliat, Haipeng Shen, Siva Tian, and Lan Zhou.

UPS delivers optimal phase diagram in high-dimensional variable selection

Pengsheng Ji

University of Georgia

Consider the linear regression model $Y = X_{\{n,p\}} \beta + z$ where both p and n are large but $p \gg n$. We find for the setting where both the signal β and the Gram matrix $X'X$ are sparse, neither the lasso nor the subset selection is asymptotically optimal in terms of the Hamming errors, or the phase diagram. We propose the Univariate Penalization Screening (UPS) for variable selection. This is a screen and clean method where we screen with univariate thresholding, and clean with penalized MLE. It has two important properties: sure screening and separable after screening. These properties enable us to reduce the original regression problem to many small-size regression problems that can be fitted separately. The UPS is effective both in theory and in computation. Some connections with optimality in multiple testing are established.

Joint Modeling of Cancer Risk and Correlated Functional Biomarkers

Measured with Error

Qi Long

Emory University

In some biomedical studies, biomarkers are measured repeatedly along some spatial structure or over time and are subject to measurement error. In these studies, it is often of interest to evaluate associations between a clinical endpoint and these biomarkers (also known as functional biomarkers). There are potentially two levels of correlation in such data, namely, between repeated measurements of a biomarker from the same subject and between multiple biomarkers from the same subject; none of the existing methods accounts for correlation between multiple functional biomarkers. We propose a Bayesian joint modeling approach to model a clinical outcome of interest (e.g., risk for colorectal cancer) in the presence of multiple functional biomarkers while accounting for potential correlation and measurement error. The proposed approach is applied to a study of biomarkers of risk for colorectal cancer and our results show that the risk for colorectal cancer is associated with two functional biomarkers, APC and TGF- α , in particular, with their values in the region between the proliferating zone and differentiating zone of colorectal crypts. Our simulation study also shows that the proposed approach achieves good performance in finite samples.

Multi-objective optimal experimental designs for event-related fMRI studies

Abhyuday Mandal

Department of Statistics

University of Georgia

Functional magnetic resonance imaging (fMRI) is an advanced technology for studying brain functions. Due to the complexity and high cost of fMRI experiments, high quality multi-objective fMRI designs are in great demand. We propose an efficient approach to find optimal experimental designs for event-related functional magnetic resonance imaging (ER-fMRI).

We consider multiple objectives, including estimating the hemodynamic response function (HRF), detecting activation, circumventing psychological confounds and fulfilling customized requirements. Taking into account these goals, we formulate a family of multi-objective design criteria and develop a genetic-algorithm-based technique to search for optimal designs. Our proposed technique incorporates existing knowledge about the performance of fMRI designs, and its usefulness is shown through simulations. We also consider a nonlinear model to accommodate a wide spectrum of feasible HRF shapes, and propose an approach for obtaining maximin efficient designs.

Our approach involves a reduction in the parameter space and an efficient search algorithm. The designs that we obtain are much more robust against misspecified HRF shapes than designs widely used by researchers. This research is joint with Ming-Hung (Jason) Kao, Dibyen Majumdar and John Stufken.

Semi-parametric grouped backward recurrence Cox model for the analysis of current duration data with preferential reporting

Alexander McLain

University of South Carolina

Current duration data arise in cross-sectional studies from questions on the length of time from an initiating event to the time of interview. For example in the National Survey on Family Growth (NSFG), women who were considered at risk of pregnancy at the time of interview were asked (a) "Are you currently attempting pregnancy," and (b) "If yes, how long have you been attempting to get pregnant." It is of interest to make inference on the distribution of the unobserved total length of pregnancy attempt based on responses to (b), referred to as the women's current duration of pregnancy attempt. Current durations are obtained only from women attempting pregnancy at the time of interview. Thus,

the observations are length-biased since those with longer durations are more likely to answer yes to question (a), and be included in the sample. Existing methods to estimate the distribution of the total durations from the observed current durations include continuous time nonparametric methods, and continuous time parametric methods that allow for the inclusion of covariates. In this article, we propose a semi-parametric grouped backward recurrence Cox model, and a piecewise constant baseline model (used to control for digit preference) to draw inference on current duration data. We discuss, and investigate through simulation studies, the consequences and robustness properties of our grouped methods when digit preference is present. Lastly, we present an analysis of the NSFG data.

Towards the understanding of the three-dimensional genome organization

Statistical challenges and opportunities for analyzing Hi-C data

Steve Qin

Emory University

Understanding how chromosomes fold provides insights into transcription regulation hence functional state of the cell. Recently, chromosomal conformation capture (3C)-related technologies have been developed to provide a genome-wide view of chromatin organization. Despite great technologies, multiple layers of noise and uncertainties stem from the sophisticated experiments, coupled with various sequencing-related artifacts, making the analysis of such data extremely challenging. Here using Hi-C as an example, we review the critical issues of analyzing this latest type of genomics data, including normalization, modeling and inference. We describe a novel Bayesian probabilistic approach, denoted “Bayesian 3D constructor for Hi-C data” (BACH), to infer chromosome three-dimensional (3D) structures from Hi-C data. We also discuss the observations we made when applying BACH to real Hi-C datasets. This is a collaboration with Ming Hu, Ke Deng, Jesse Dixon, Siddarth Selvaraj, Jennifer Fang, Bing Ren and Jun Liu.

Profile Local Linear Estimation of Generalized Semiparametric Regression

Model for Longitudinal Data

Yanqing Sun

University of North Carolina, Charlotte

Coauthors: Liuquan Sun and Jie Zhou

This paper studies the generalized semiparametric regression model for longitudinal data where the covariate effects are constant for some and time-varying for others. Different link functions can be used to allow more flexible modelling of longitudinal data. The nonparametric components of the model are estimated using a local linear estimating equation and the parametric components are estimated through a profile estimating function. The method automatically adjusts for heterogeneity of sampling times, allowing the sampling strategy to depend on the past sampling history as well as possibly time-dependent covariates without specifically model such dependence. A K-fold cross-validation bandwidth selection is proposed as a working tool for locating an appropriate bandwidth. A criteria for selecting the link function is proposed to provide better fit of the data. Large sample properties of the proposed estimators are investigated. Large sample pointwise and simultaneous confidence intervals for the regression coefficients are constructed. Formal hypothesis testing procedures are proposed to check for the covariate effects and whether the effects are time-varying. A simulation study is conducted to examine the finite sample performances of the proposed estimation and hypothesis testing procedures. The methods are illustrated with a data example.

Models for sums of dependent Bernoulli random variables

Chang Yu

Department of Biostatistics

Vanderbilt University

In many epidemiologic studies the first indication of an environmental or genetic contribution to the risk of disease is the way the disease cases cluster within the same family units. We provide an exact test of the initial hypothesis of no familial link with the disease, conditional on the number of diseased cases and the sizes of the various family units. Then we develop new discrete distributions that describe the behavior of the sum of dependent Bernoulli random variables. General results for these models include recursive relationships for their mass functions and moments.

Auxiliary marker-assisted classification in the absence of class identifiers

Donglin Zeng

University of North Carolina, Chapel Hill

Constructing classification rules for accurate diagnosis of a disorder is an important goal in medical practice. In many clinical applications, there is no clinically significant anatomical or physiological deviation exists to identify the gold standard disease status to inform development of classification algorithms. Despite absence of perfect disease class identifiers, there are usually one or more disease-informative auxiliary markers along with feature variables comprising known symptoms. Existing statistical learning approaches do not effectively draw information from auxiliary prognostic markers. We propose a large margin classification method, with particular emphasis on the support vector machine (SVM), assisted by available informative markers in order to classify disease without knowing a subject's true disease status. We view this task as statistical learning in the presence of missing data, and introduce a pseudo-EM algorithm to the classification. We also propose a sparse variable selection method embedded in the pseudo-EM algorithm. Theoretical examination shows that the proposed classification rule is Fisher consistent, and that under a linear rule, the proposed selection has an oracle variable selection property and the estimated coefficients are asymptotically normal. We apply the methods to build decision rules for including subjects in clinical trials of a new psychiatric disorder and present four applications to data available at the UCI Machine Learning Repository.

Linearly Constrained Generalized Lasso

Peng Zeng

Department of Mathematics and Statistics

Auburn University

Lasso and its variants have been widely used in regression analysis for high-dimensional data. As a way to incorporate prior knowledge in applications, we propose linearly constrained generalized lasso for simultaneous estimation and variable selection. The dual of the linearly constrained generalized lasso is derived. An efficient coordinate descent algorithm is proposed to estimate the dual. In order to measure the complexity of the model, a formula for the degrees of freedom is derived and an unbiased estimator of degrees of freedom is obtained. AIC and BIC criteria are used to select tuning parameters. Simulation studies demonstrate the excellent performance. Theoretical results are also discussed for some special scenarios.

Poster Abstracts

Estimating and Kernel Smoothing the Hazard Function for Right Truncated Data

Haci Akcin and Xu Zhang

Centers for Disease Control and Prevention (CDC)

Office of Surveillance, Epidemiology and Laboratory Services (OSELs)

Public Health Surveillance Program Office (PHSPO)

Division of Behavioral Surveillance (DBS)

Two types of truncation, left and right truncation, coexist in a truncated sample. Earlier researches focused on left truncation and it had been studied extensively. Lagakos et al. (1988) pioneered to study right truncated data in a clever way. They proposed transforming right truncated data to left truncated one and applying similar methods for left truncation to analyze the data. Interpretation of survival quantities, more specifically hazard rate, in reverse-time is not natural. Although it is the most direct way, researchers seldom use forward-time hazard. In this paper, we studied the nonparametric inference for the hazard rate of right truncated data. We estimate crude hazard rate for right truncated data in forward-time. Kernel smoothing techniques were used to get smoothed estimates. Three commonly used kernels, Uniform, Epanechnikov and biweight kernels were applied in this study on the AIDS data.

Empirical likelihood procedure for the percentile residual life function

Yueheng An

Georgia State University

The α -percentile ($0 < \alpha < 1$) residual life function at time x is defined as the α -percentile of the remaining life given survival up to time t . The decreasing α -percentile residual life (DPRL- α) is one class of life distributions defined by the α -percentile residual life function. In this poster, empirical likelihood method is applied to obtain an interval estimator of the α -percentile residual life function. Under some regularity conditions, the log empirical likelihood ratio asymptotically approaches to a chi-square distribution. Simulation studies carried out for several continuous distributions with α -percentile residual life.

Asymptotics of the Signed--Rank Estimator under Dependent Observations

Frazier BINDELE

University of South Alabama

In this paper, we consider a signed--rank estimator of nonlinear regression coefficients under stochastic errors. These errors include a wide array of applications in economic literature such as serial correlation, heteroscedasticity, autoregression, etc. General conditions for strong consistency and asymptotic normality of the resulting estimator are established.

A β_1 -adrenergic control of the action potential in mouse ventricular myocytes

Vladimir E. Bondarenko

Department of Mathematics and Statistics

Georgia State University, Atlanta GA

We employed a model for a compartmentalized β_1 -adrenergic signaling system in mouse ventricular myocytes to investigate behavior of the action potential and calcium dynamics. The model cell is divided into three major compartments: caveolae, extracaveolae, and cytosol. Each of the compartments contains β_1 -adrenergic receptors (β_1 -ARs), which are activated by isoproterenol. Activation of β_1 -ARs stimulates G_s -proteins to further activate adenylyl cyclases, which elevate the cAMP level and increase the activity of protein kinase A (PKA). cAMP concentration is also controlled by degradation with phosphodiesterases. There are nine major PKA targets in different cellular compartments: the fast sodium current, the L-type calcium current, ryanodine receptors, and the sodium-potassium pump are located in the caveolae; the ultrarapidly activating delayed rectifier current, the rapidly inactivating transient outward potassium current, and the time-independent potassium current are in the extracaveolae; and troponin I and phospholamban are in the cytosol. Activation of β_1 -ARs leads to a moderate prolongation of action potential in mouse ventricular cells. The computer model elucidates the mechanism of this prolongation, which results from two opposite effects: an increase in the ultrarapidly activating delayed rectifier current, on the one hand, and a decrease in the rapidly inactivating transient outward potassium current and an increase in the L-type calcium current, on the other hand.

Jackknife Empirical Likelihood for the Failure Time Model with Censored Data

Maxime Bouadoumou and Yichuan Zhao

Georgia State University

Kendall and Gehan estimating functions are used to estimate the regression parameter in accelerated failure time (AFT) models with censored observations. The accelerated failure time model is the preferred survival analysis method because it maintains a consistent association between the covariate and the survival time. The jackknife empirical likelihood method is used because it overcomes computation difficulty by circumventing the construction of the nonlinear constraint. U-statistic approach is used to construct the confidence intervals for the regression parameter. We conduct a simulation study to compare the Wald-type procedure, the empirical likelihood, and the jackknife empirical likelihood in terms of coverage probability and average length of confidence intervals. Jackknife empirical likelihood method has a better performance and overcomes the under-coverage problem of the Wald-type method. A real data is also used to illustrate the proposed methods.

In Search of Desirable Compounds

Adrijo Chakraborty

University of Georgia

Drug discovery scientific teams evaluate a compound's performance across a number of endpoints, often including its ability to bind to target(s), as well as its absorption, distribution, metabolism, excretion, and safety characteristics. A popular way to prioritize compounds is through desirability scoring across the selected endpoints. In addition to desirability scores, drug discovery teams can measure or compute many other compound descriptors that are thought to be related to the endpoints of interest. Scientists would like to identify a small number of such descriptors which can be tweaked to improve the performance of a compound. To that end, one primary objective of these teams is to identify the most crucial descriptors and relationships among the descriptors that impact compound desirability. Finding these relationships can help teams understand relationships between predictors and compounds' desirability, and can possibly suggest molecular alterations that could improve compound desirability. In this work, we propose a novel method based on the desirability scoring framework for identifying important descriptors. We also explore ways to visualize high dimensional predictor spaces that can effectively inform teams' decisions in synthesizing new compounds.

Escalation with Overdose Control using All Toxicities and Time to Toxicity

Data for Cancer Phase I Clinical Trials

*Ye Cui*¹

*Taofeek K Owonikoko*²

Zhibo Wang^{3,4}

*Fadlo R Khuri*²

Jeanne Kowalski ^{3,4}

Zhengjia Chen^{3,4}

1. *Department of Mathematics & Statistics, Georgia State University, Atlanta, GA 30303, USA.*
2. *Department of Hematology and Medical Oncology, Emory University, Atlanta, GA 30322.*
3. *Biostatistics and Bioinformatics Shared Resource at Winship Cancer Institute GA 30322.*
4. *Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322.*

The primary purpose of cancer Phase I clinical trial which is a critical step in development of new drug against cancer is to determine the maximum tolerated dose (MTD) and schedule of new drug. It is usually a small study with limited data so that fully utilizations of all toxicities and time to toxicity data are essential to improve the trial efficiency and accuracy of MTD estimation. Chen et al. (2010) proposed a novel normalized the equivalent toxicity score (NETS) system which can fully utilize multiple toxicities per patient instead of a binary indicator of dose limiting toxicity (DLT). Cheung et al. (2000) developed the time of event (TITE) approach to incorporate time to toxicity data. Escalation with Overdose Control (EWOC) is an adaptive Bayesian Phase I design which can allow rapid dose escalation while controlling the probability of overdosing patients. In this study, we use EWOC as a framework and integrate it with the NETS system and TITE approach to develop an advanced Phase I design entitled EWOC-NETS-TITE. Simulation studies have been conducted to compare its operating characteristics with those of EWOC and standard 3+3 design. Simulation results demonstrate that EWOC-NETS-TITE can not only substantially improve the trial efficiency and MTD accuracy, but also allow patients to be entered in a staggered fashion and shorten trial length. A user-friendly software of EWOC-NETS-TITE is under development and will be available in the future.

SMALL IMPROVEMENT TO THE KOLMOGOROV-SMIRNOV TEST

XING DONG

Health Informatics & Technology Solutions

ICF INTERNATIONAL

The Kolmogorov-Smirnov (K-S) test is widely used as a goodness-of-fit test. This thesis consists of two parts to describe ways to improve the classical K-S test in both 1-dimensional and 2-dimensional data. The first part is about how to improve the accuracy of the classical K-S goodness-of-fit test in 1-dimensional data. We replace the p-values estimated by the asymptotic distribution by near-exact p-values. In the second part, we propose two new methods to increase power of the widely used 2-dimensional two-sample Fasano and Franceschini test. Simulation studies show the new methods are significantly more powerful than the Fasano and Franceschini's test.

Semiparametric Bayesian Quantile Regression for Panel Data

Woo Sung Jang

Department of Statistics, North Carolina State University

Quantile regression is a valuable regression technique that is able to provide a comprehensive picture of the relationship between a response and covariates. The estimation and inference of quantile regression for cross sectional data have been well studied in the literature. However, little work has been carried out for quantile regression with random effects. The main challenges are that quantile regression requires no parametric distributional assumptions, and that quantiles are not additive. In this paper, we develop a new semiparametric Bayesian quantile regression method for panel data. Instead of making any parametric assumptions on the error distribution, we only assume that the conditional quantiles of the response given covariates and the random effects are linear in covariates. Then the likelihood can be approximated by linearly interpolating the quantile functions. For tail areas with sparse data, we adopt the extreme value theory to estimate the tail density. In addition, we propose a Metropolis-within-Gibbs algorithm to update fixed and random coefficients. Through numerical studies, we demonstrate that the proposed method is competitive or more efficient than other existing methods for quantile regression for panel data with moderate sample sizes. The performance of the proposed method is illustrated through simulation and the analysis of an apena duration data of elderly women.

Analyses of Mouse RPE Morphology Give Discriminatory Categories

Y Jiang, X Qi, T Jiang, Y Cheng, J Zhang, MA Chrenek, C Gardner, HE Grossniklaus, JM Nickerson

Georgia State University

Age-related macular degeneration (AMD) is the main cause of vision loss in the elderly and is a looming epidemic in our aging society. Presently there is no way to determine how a patient's eye will progress, and no effective treatment for AMD. Retinal pigment epithelium is a crucial site of AMD pathology and undergoes morphological changes as the eye ages and AMD progresses. We collect RPE morphological data from mouse eyes and conduct statistical studies to establish the relationship between the RPE morphology and the age and disease progression of the eye. This work provides a foundation for a potential diagnostic and prognostic tool for AMD.

New empirical likelihood inference for the bivariate survival function under univariate censoring

Ali Jinnah

Georgia State University

Lin and Ying (1993) proposed a non-parametric estimator of the bivariate survival function of paired failure times under univariate censoring. Recently Lu and Burke (2008) proposed estimators for bivariate distribution of paired failure times under univariate censoring. In this paper, we apply a variant of plug-in empirical likelihood by estimating the survival function of censored time by Kaplan-Meier estimator. A new empirical likelihood (NEW) confidence interval for the bivariate survival function is obtained. Furthermore, we conduct a simulation study to evaluate the performance of the proposed NEL method and the normal approximation (NA) in terms of coverage probability and confidence interval (CI) length under the setting mentioned in Lin and Ying (1993). The simulation study shows the coverage probabilities using the NEL method obtained are close to the nominal levels. We compare the results with the probability coverage and CI lengths produced using Lu and Burke (2008) setting.

Joint work with Yichuan Zhao

Modeling Extracellular Matrix in Breast Cancer

Byoungkoo Lee¹, Kevin W Eliceiri², Patricia J. Keely³, Alissa Weaver⁴, Yi Jiang⁵

Georgia State University

Breast cancer is a fatal disease, but the detailed mechanism of metastasis is unclear since extracellular matrix (ECM) affects many aspects of cellular behavior, including migration, invasion, and proliferation. Furthermore, the properties of ECM itself are also complex, with diverse topographies and mechanical properties depending on the density, alignment, polymerization, and crosslinking. We use *in vitro* and *in silico* ECM models to study how biophysical parameters of individual fibers affect the mechanical property of ECM network. The computational ECM model will be a key step toward simulating and investigating the invasive behavior of aggressive breast cancer cells.

Keywords — extracellular matrix, breast cancer

An Introduction to Bayesian Interim Analysis for Clinical Trials

Hal Li

Accenture (Octagon Research Solutions, Inc.)

The theory and methods for the Bayesian Interim Analysis for Clinical Trials is the motivation for this presentation. An introduction to the Bayesian statistics and Interim Analysis is provided. Examples are provided to inform the process used for the Bayesian Interim Analysis. Discussions will be given to the FDA current guidelines on the Bayesian Analysis.

In this presentation, we will discuss the following topics:

- Bayesian theory
- Normally distributed data
- Binary data

- Poisson distributed data
- Bayesian Inference
- Summary of the posterior distribution
- Interval estimation approach
- Predictive probability approach
- Bayesian Interim Analysis for Clinical Trials
- Simulation in Bayesian Inference
- MCMC Method
- FDA's guidance and opinion on Bayesian analysis

Jackknife empirical likelihood for σ^2 in the linear regression model

Hui-Ling Lin

Georgia State University

The variance of a random variable is σ^2 . It is the measure of spread from the center. Therefore, how to accurately estimate σ^2 has always been an important topic in recent years. In this paper, we consider linear regression model which is the most popular model in practice. We use jackknife empirical likelihood (JEL) method to estimate σ^2 in the regression model. The log-ratio of the proposed JEL converges to the standard chi-squared distribution. The simulation study is carried out to compare the JEL method and standard methods in terms of coverage probability and interval length for the confidence intervals of σ^2 from linear regression models. The proposed JEL has better performance. We also illustrate the proposed methods using two real data sets. Joint work with Yichuan Zhao.

Structural properties of Mg²⁺ binding sites based on a PDB survey

Yunmei Lu

Department of Computer Science, Georgia State University

Magnesium (Mg²⁺), the most abundant intracellular divalent metal ion, plays important roles in cell physiology through binding with various proteins and nucleotides. Functionally, Mg²⁺ interacts with DNA and RNA to regulate their synthesis and maintain their conformation, and will typically bind with most di- and tri-phospho nucleotides, including nicotinamide and flavin dinucleotides, under physiological conditions. Of interest to us are the cases of multiple protein-supplied ligands, as these, intuitively, constitute a class of more intimate protein binding pockets for Mg²⁺. By doing an analysis based on a PDB survey, we know that protein and water contribute about 45.3% and 49.12% ligands to Mg²⁺ binding sites, respectively, and other small molecular only contribute about 5.6%. Mg²⁺ binding sites prefer water very much, which is very different from Ca²⁺ binding sites. The maximal number of ligands at each Mg²⁺ binding site is 8, and most of them are 6. Besides these, the second structure, average distance between Mg²⁺ ion and ligand atoms, contiguous and discontiguous sites, geometry formed by ligands and Mg²⁺ ions are also analyzed in the current work.

Jackknife Empirical Likelihood for Absolute Deviation

Xueping Meng and Yichuan Zhao

Georgia State University

In asymptotic analysis, when the sample size becomes large, we focus on describing the properties of estimators of the absolute deviation. In this poster, we focus on using jackknife empirical likelihood (JEL) method to construct confidence intervals and find the coverage probability and average length for the empirical estimator. The empirical log-likelihood ratio statistic is derived. The asymptotic distribution is proved to be a standard chi-squared distribution. The results of comprehensive simulation studies show the comparison of the average length and coverage probability for various sample sizes by using the JEL method and the traditional normal approximation based method.

INDEX WORDS: Confidence interval, Coverage probability, Jackknife empirical likelihood

Rank Based Group Variable Selection

Guy-vanie Miakonkana

Auburn University

A robust rank based estimator for variable selection in linear models, with grouped variables, is studied. The proposed estimation procedure extends the existing rank based variable selection methods (Johnson and Peng (2008)) and the ww-scad (Wang and Li (2009)) to linear regression models with grouped predictors. The resulting estimator is robust to contamination or deviations in both the response and the design space. The oracle property and asymptotic normality of the estimator are established under some regularity conditions. Simulations studies and a real data analysis reveal that the proposed method performs better than the existing rank based methods for grouped variables selection. This estimation procedure also outperforms the adaptive Hlasso (Ahou and Zhu (2010)) in the presence of local contamination in the design space or for heavy tailed error distribution.

A model for mouse ventricular myocyte contraction

Paula D. Mullins and Vladimir E. Bondarenko

Department of Mathematics and Statistics

Georgia State University, Atlanta, GA

A model for mouse ventricular myocyte contraction is presented. The model was developed from a previously published model on electrical activity and calcium handling in mouse cardiac cells. The resulting model consists of 51 differential equations, which describe action potential, gating variables and states of ionic channels, contraction proteins, and ionic concentrations. The model equations were solved by the 4th-order Runge-Kutta method with the time step 0.0001 ms. Simulations were compared with previously published experimental data on contraction obtained for mouse ventricular myocytes. The model described well the steady-state force-calcium relationships for different sarcomere lengths, time course for the force development and cell shortening, the frequency dependence of peak force and cell shortening, time-to-peak and time to 50% relaxation of the force. Our simulation data points to the importance of using variable sarcomere length in the model for myocyte contraction.

MULTIDIMENSIONAL INDICES NONLINEAR SIGNED-RANK REGRESSION

Brice Merlin Nguelifack

Auburn University

We consider a nonlinear regression model when the index variable is multidimensional. Sufficient conditions on the non-linear function are given under which the Signed-Rank estimators are strongly consistent and asymptotically normally distributed. These sufficient conditions are satisfied by harmonic type functions, which are also of interest in the one dimensional index case where Wu's (Asymptotic theory of non-linear least-squares estimation, *Ann. Statist.* 9 (1981) 501 – 513) and Jennrich's (Asymptotic properties of nonlinear least-squares estimators, *Ann. Math. Statist.* 40 (1969) 633 – 643) sufficient conditions are not applicable.

Assessment of the Sustained Financial Impact of Risk Engineering Service on Insurance Claims Costs

Bobby I. Parker

Statistical Analysis/ Risk Engineering Consulting at Atlanta

This research paper creates a comprehensive statistical model, relating financial impact of risk engineering activity, and insurance claims costs. Specifically, the model shows important statistical relationships among six variables including: types of risk engineering activity, risk engineering dollar cost, duration of risk engineering service, and type of customer by industry classification, dollar premium amounts, and dollar claims costs.

We accomplish this by using a large data sample of approximately 15,000 customer-years of insurance coverage, and risk engineering activity. Data sample is from an international casualty/property insurance company and covers four years of operations, 2006-2009. The choice of statistical model is the linear mixed model, as presented in SAS 9.2 software. This method provides essential capabilities, including the flexibility to work with data having missing values, and the ability to reveal time-dependent statistical associations.

Decipher the Gene Regulatory Network using Comprehensive Microarray

Expression Data

Po-Chih Shen and Li-Yu Liu

Department of Agronomy, Biometrics Division

National Taiwan University

The diverse genomic data through genome sequencing and gene expression microarray technology have facilitated the researches in determining gene regulatory events in the biological processes. They can be used to construct the gene regulatory network. However, the data may be noisy. For example, AT1G24260 and AT4G36920 are two TFs and AT1G53160 is the target gene of the two TFs. We find that most of the expression levels of AT1G53160 increase in AT1G24260 that is a positive regulation and decrease in AT4G36920 that is a negative regulation. But there are a part of the expression levels of AT1G53160 not increasing and decreasing in AT1G24260 and AT4G36920. In this study, we purpose to integrate the knowledge about the transcription factor binding sites and the gene expression microarray data to construct gene regulatory network using a set of microarray expression data under different treatments. A normal mixture model was used to describe the gene expression levels either regulated by a transcription factor or not. The Galton Pearson's correlation coefficient (GPCC) and the coefficient of intrinsic dependence (CID) were further adopted to determine the strength of the interactions between genes. The proposed method was demonstrated using a toy example as well as real expression data.

Evaluating Fluctuations in Sex Ratios at Birth in Small Areas

Michele Steinmetz, J. Gregory Robinson, Heather King and Dan Litwok

Georgia State University

The sex ratio at birth is an important analytical measure across several disciplines such as biology, genetics, sociology and demography. As a demographic measure it can provide valuable information regarding other demographic trends. For example, the doubling time of a population rises as the ratio of males to females at birth increases. As the number of males in an area increases over time, it can lead to a marriage squeeze and slow the overall natural increase in a population. Data on sex ratio at birth has also been used to understand trends in low birth weight and infant mortality since male babies have higher infant mortality rates. (Mathews and Hamilton, 2005) The sex ratio at birth can also be indicative of other societal trends. An extremely imbalanced sex ratio at birth has been associated with a son preference in some Asian countries (Anderson and Silver, 1995; Belanger et al., 2003; Goodkind, 2004). While in Nordic countries, there is some evidence of a relationship between birth order and gender preference (Andersson et al., 2006). Globally, national sex ratios at birth tend to fluctuate very little. While this contributes to the usefulness of the sex ratio at birth as an analytical measure, it has left relatively unexamined how sex ratios at birth in smaller geographies can fluctuate considerably from year to year. There has been little research to document or understand the magnitude of variations in sex ratios below a national level. A reliable evaluation measure of the sex ratio at birth at lower levels of geography could prove to be very helpful in identifying any potential data quality or methodological concerns in research utilizing sex ratio at birth as an analytical measure.

Using a theoretical model to calculate probability ranges of sex ratios at birth, this research examines how well actual birth data from several countries over a series of years matched the model's probability ranges, given the randomness of the outcome of a male birth compared to a female birth for any given number of births. The model assumes a binomial distribution of births, using the expected outcome of 105 male births for every 205 total births. Then, setting bounds for the probability ranges, the actual birth data are matched against the probability ranges to examine how the birth data compare to the probabilities in the theoretical model. At the lower bound, the probability range was set at 90%. At the upper bound, the probability range was set at 99%. The expectation of the model is one to ten percent of the geographies examined in the data would have sex ratios at birth that would fall outside the expected number of births in any given year. However, since an assumption of this research is sex at

birth is a random event, there should be no pattern in which geographies fall beyond the bounds set by the model.

The model is tested with vital statistics and/or census data from the United States, Denmark and Kenya. U.S. Census data at the county level from 2000 and 2010 and Kenyan Census data at the division or constituency level from 1989, 1999, and 2009 on reported zero year olds by sex are compared to the theoretical model. U.S. vital statistics data on births by sex at the county level from 2000 – 2009 and Danish vital statistics data on births by sex at the municipality level from 2000 – 2009 are also compared to the theoretical model. If time permits, additional years of vital statistics from the U.S. and Denmark will be included as well as additional countries for a more robust analysis of the model.

Initial findings indicate the model can provide meaningful bounds of “statistical significance” in sex ratios at birth. The model has provided insight into how sex ratios at birth in smaller geographies behave, where births are infrequent events in any given year. Annual fluctuations in sex ratios at birth, in many counties, previously thought to be problematic, are found to be within the expected bounds set by the probability model. Conversely, the sex ratio at birth in areas with a very high number of births are found to have a much narrower range of acceptable fluctuation than previously understood. For example, the model indicates it takes almost 58,000 births to achieve an expected sex ratio range of 104 – 106, at the 90 percent threshold in the model. There are only 4 counties in the United States that regularly have approximately that number of births each year. The majority of counties in the United States have less than 500 births a year. At the 90 percent threshold for 500 births, the model predicts a sex ratio at birth between 91 and 121 would be acceptable.

The number of male births compared to the total number of births can also be adjusted in the model to reflect a different national sex ratio at birth for a country and the model still provides statistically sound results.

A model such as this can be a valuable evaluation tool in research and estimates production. It would allow researchers to focus on sex ratio outliers that may indicate real data problems rather than those due to random chance.

Sparse Factor Analysis by choice of norm

Yanhong Wang

Georgia State University

Factor analysis(FA) is a popular multivariate analysis method which is used to describe observed variables as linear combinations of hidden factors. One method of classic exploratory factor analysis(EFA) often draw factor loadings by scaled principal components called principal component factor analysis(PCFA). However, the factor loadings getting from this method suffers from the fact that factor loadings often not in the desirable sparse form, thus it is often difficult to interpret the results. In this article, we reinvestigate an existing sparse principal component analysis (SPCABP) method, and propose to use it as a basis for sparse principal factor analysis by choice of norm (SPFABP) in which factor loadings naturally adopt a sparse representation, greatly facilitating the interpretation of the fitted factor model. Experiments on simulated and real datasets show that the proposed approach performs better than existing FA methods while providing a useful representation of the latent factors.

Joint work with Dr. Xin Qi, Georgia State University

Analytical methods for evaluations of targeting performances for a simplified stochastic model of neural arborization

Jun Xia¹, Jie Zhang¹, Adil Virani¹, Gloria Sanin¹, Emily Su², Troy Shinbrot², Remus Osan^{1,3}

¹*Department of Mathematics and Statistics, Georgia State University, Atlanta, GA*

²*Department of Biomedical Engineering, Rutgers University, Piscataway, NJ*

³*Neuroscience Institute, Georgia State University, Atlanta, GA*

Spinal cord injury can lead to permanent damage of the ascending and descending neural pathways, which can lead to a loss of sensory feeling and motor control in the limbs, respectively. The growth of neuron can potentially lead to experimental therapies, but the issue of targeting the exact regions

remains problematic. While traditionally mathematical and computational research has been focused on faithful characterization of the morphological properties of the neurons, less emphasis has been placed on understanding how the growth rules of the neurons translate in effective targeting of desired neural regions. To quantitatively analyze and optimize targeting strategies, we have developed side-by-side *in silico* and *in vitro* experimental studies of neuron growth, branching and pruning using chicken DRG neurons as a prototypical model system [1].

In the current work we used a simplified stochastic model to investigate how branching influence the probability of successfully connecting to neurons located at different locations away from the initiation point. The rules for generating a simplified neural tree are as follows. Initially there is one active branch. Then, at each step of the neural growth, each active branch can further extend or it may branch out to create an additional new neurite. We use stochastic method to model the branching process, with the probability to branch taken to be p and the probability to create a new neurite taken to be $q = 1-p$. Thus, for each step, we can generate the full the distribution of the probability for each possible outcome. An illustration for this up to three time steps is shown in Figure 1.

References

[1] Osan R, Su E, Shinbrot T. **The interplay between branching and pruning on neuronal target search during developmental growth: functional role and implications.** *PLoS One* 2011, **6(10)**:e25135.

[2] Van Pelt J, and Uylings, **Natural variability in the geometry of dendritic branching patterns**, HBM (2005). in *Modeling in the Neurosciences*, ed: Reeke GN, Poznanski RR, Lindsay KA, Rosenberg JR and Sporns O (CRC Press, Boca Raton, FL) pp 89-116.

Numerical Study of Mucociliary Clearance in the Airway

Ling Xu and Yi Jiang

Georgia State University

We study the mucociliary clearance in human airway using numerical simulations. In our model, the beating cilia immersed in the viscous mucus provide the driving force. Each cilium is modeled as a rigid rod with prescribed motions. Parameters, such as fluid viscosity, cilia length, cilia beating frequency, and the orientation of cilium, are examined in order to understand their effects on the transport property of the mucociliary flow. Deviation of these parameter values from normal physiological range corresponds to various respiratory problems, e.g., primary cilia dyskinesia, cystic fibrosis.

INTEGRATIVE ANALYSIS OF GENOME AND TRANSCRIPTOME SEQUENCING DATA FOR GENOMIC REARRANGEMENT DISCOVERY IN MULTIPLE MYELOMA

RENDONG YANG

EMORY UNIVERSITY

We present a pipeline to perform integrative analysis of mate-pair and paired-end genomic DNA sequencing together with paired-end transcriptome sequencing data. Our pipeline employs the Burrows-Wheeler Aligner (BWA) algorithm to map short sequencing reads and classify the mapped read pairs into properly paired and discordantly paired categories based on their orientations and inferred insert sizes. Our pipeline detects inter- and intra-chromosomal rearrangements using discordant reads mapped to different chromosomes or separated with the insert size beyond the inferred limitation. Our pipeline takes into account gene annotation and sequence repeat information to increase specificity. Application of our pipeline to whole-genome and transcriptome sequencing data from three multiple myeloma cell line (KMS-11, MM-1S and RPMI-8226) identified known TRAF3 and CDC42BPB intra-chromosomal rearrangement as well as novel SPI1 and ZNF287 inter-chromosomal rearrangement in the RPMI cell line.

Identification of Differential Gene Pathways with Sparse Principal

Component Analysis

Yichao Yin¹, and Ruiyan Luo²

¹Department of Mathematics and Statistics, ²Institute of Public Health

Georgia State University

The development of the technology makes it possible to measure large amount of genes expressions simultaneously. Since biological functions are mostly coordinated by multiple genes, called “gene pathway”, it is interesting to identify differential gene pathways which are associated with clinical phenotype. Principal component analysis has been proposed to identify differential gene pathways in several literatures, while sparse principal component analysis (SPCA) has not drawn any attention. We proposed to use SPCA to identify differential gene pathways. The results show that, comparing to PCA, SPCA could identify more differential expressed gene pathways, especially when the higher-order interactions among genes are considered.

Approximation series for activity propagation in neural tissue in presence of periodic inhomogeneities

Jie Zhang

Georgia State University

The study of traveling waves of activity in neural tissue can provide deep insights into the functions of the brain during sensory processing or during abnormal states such as epilepsy, migraines or hallucinations. Computational models of these systems usually describe the tissue as a vast interconnected network of neurons comprised of large number of units with similar properties, for example integrate and fire neurons. It is also widely assumed that while the strength of the connections between neurons changes as a function of distance, this interaction does not depend on other local parameters.

These assumptions allow for formulation of a set of integro-differential equations describing the propagation of the traveling wave fronts in a one-dimensional integrate-and-fire network of synaptically coupled neurons, allowing for investigation of the network dynamics during wave initiation and propagation. Equations for the transition between initiation and transition toward constant speed traveling waves have been derived for Gaussian connectivity [1] and finite support connectivity [2]. These results have been also confirmed through numerical simulations, leading to methods for optimizing and improving simulations of large-scale networks [3]. These results have been extended beyond the simpler case of one-spike activity propagation, deriving equations for constant speed waves with a finite and infinite number of spikes [4]. This framework has produced insight on the mechanisms of stable constant-speed traveling wave solutions, but the study of inhomogeneities in synaptic connections likely to exist in the brain tissue has received much less attention since not surprisingly, the presence of inhomogeneities vastly increases the complexity of the mathematical models. However, recent work [5,6] used homogenization theory to determine how inhomogeneities can induce propagation failure.

We extended our previous models that exhibit constant-speed traveling waves to investigate how the presence of these inhomogeneities affects the relationship between the speed of the activity propagation and its acceleration. We determine that the estimates from homogenization theory do not accurately capture the conditions for propagation failure. More precisely, just prior to stopping, the

activity propagates at a higher average speed than predicted from the theoretical results of the homogenization theory. We derive more precise estimates for the conditions when propagation failure occurs. Furthermore, our study points to directions where researchers can obtain additional tools for analyzing experimental data in order to infer details of synaptic connectivity.